STATISTICS AND COMPRESSION OF SCL

DANNY CALEGARI AND JOSEPH MAHER

Abstract. In a hyperbolic group, a random word of length n in the commutator subgroup has stable commutator length of order $n/\log n$. In any finitely generated group, either stable commutator length vanishes identically, or the result of a random walk of length n conditioned to lie in the commutator subgroup has stable commutator length bounded above by order $n/\log n$ and below by order $\sqrt{n}$. The upper bounds are obtained by explicit estimates on random words and random geodesics in free and hyperbolic groups. The lower bounds are obtained from properties of the statistical distribution of values of quasimorphisms. One result we prove of independent interest is a central limit theorem for values of the rotation quasimorphism on random walks in semigroups of homeomorphisms of the circle.

1. Introduction

Bounded cohomology, as introduced by Gromov in [25], is (among other things) a functor from the category of groups and homomorphisms to the category of normed vector spaces and norm-decreasing linear maps. One of the main virtues of this functor is its *monotonicity*: if the metric invariants associated to a group G are (in some sense) "smaller" than the invariants associated to a group H there are no "interesting" homomorphisms from G to H. As a well-known example, Bestvina–Fujiwara [2] used 2-dimensional bounded cohomology to show that every homomorphism from a higher rank lattice to a mapping class group factors through a finite group (this fact was known earlier by work of Farb–Masur [20], building on work of Kaimanovich–Masur [31]).

To this date, such tools have generally been used somewhat crudely, because of the enormous difficulty in computing bounded cohomology, or deriving useful invariants from it. Most authors have concentrated on bounded cohomology in dimension 2, and have focused almost exclusively on a trichotomous distinction: namely for a given group G whether $H^2_b(G)$ is trivial, finite dimensional, or infinite dimensional. In this paper, we initiate a study of more quantitative *probabilistic* invariants of (2-dimensional) bounded cohomology, via its relation to *homogeneous quasimorphisms* and *stable commutator length*.

For any group G there is an exact sequence of real vector spaces

$$0 \to H^1(G) \to Q(G) \to H^2_b(G) \to H^2(G)$$

where H^2_b denotes bounded cohomology in dimension 2, and Q denotes the vector space of homogeneous quasimorphisms (see § 2 for a precise definition, and e.g. [10] Thm. 2.50 for a proof). For a finitely presented group, H^1 and H^2 are finite dimensional, so H^2_b and Q contain (almost) the same information. Moreover, H^2_b and Q/H^1 are Banach spaces, and the map $Q/H^1 \to H^2_b$ is 2-bilipschitz. In this

Date: version 0.19; September 1, 2010.

paper we focus more on Q/H^1, though our results can be interpreted as statements about H^2_b.

Bavard [1] interpreted Q/H^1 dually, in terms of an algebraic invariant called *stable commutator length* (hereafter scl). For any $g \in G'$, Bavard derived a formula

$$\mathrm{scl}(g) = \sup_{\phi \in Q/H^1} \phi(g)/2D(\phi)$$

where $D(\cdot)$ (the "defect") is a natural norm on Q/H^1. The function scl is monotone nonincreasing under homomorphisms between groups. Therefore, quantitative features of scl (e.g. spectral gaps, statistical distribution, etc.) represent a sharper tool than $\dim(H^2_b)$ with which to do geometry.

This paper studies the statistical distribution of scl in two important (but related) contexts: as a function on random *geodesics*, and as a function on random *walks*, in either case of fixed length and in a given generating set. Our first results concern the distribution of scl on random elements of word length n in *hyperbolic groups*. Before we state our theorems, it is important to observe that scl is only defined on elements of the commutator subgroup G' of a given group G. To make use of statistical theorems conditioned on elements lying in G' it is important to understand elements in G that are not in G', but have image in H' under some homomorphism $G \to H$. This issue is addressed in this paper by means of assigning to each $g \in G$ a *correction*, namely an element $hg \in G'$ together with an estimate of $\mathrm{scl}(hg)$ (see Lemma 3.20). Then under any homomorphism $\varphi : G \to H$ for $\varphi(g) \in H'$ one can estimate $\mathrm{scl}(\varphi(g))$ from $\mathrm{scl}(hg)$ together with an estimate of the length $|h|$. With this in mind, our first main theorem is as follows:

Hyperbolic Theorem 3.39. *Let G be a hyperbolic group. For any positive C_1 there are positive constants C_2, C_3 so that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(g) \leq C_3 n/\log n \mid g \in G'_n) \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hg) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon} \mid g \in G_n) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.20.

For potential applications, it is significant that our results are inherited by quasiconvex subgroups, and we prove:

Quasiconvex Theorem 3.42. *Let G be a hyperbolic group G, and let H be a quasiconvex subgroup. Fix a finite generating set S for H_n, and let H_n denote the set of words in H of length n in the generators S. Then for any positive C_1 there are positive constants C_2, C_3 so that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(g) \leq C_3 n/\log n \mid g \in H_n \cap G') \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hg) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon} \mid g \in H_n) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.20.

Experimental evidence (see § 3.2) suggests that (at least in free groups) there is a constant C so that $\mathrm{scl}(g) - Cn/\log n = o(n/\log n)$ in probability, as $n \to \infty$. It would be very interesting to prove this, at least for some class of hyperbolic groups, and to give a precise formula for the constant C. More generally, it would

be interesting to compute explicit bounds for C_2 and C_3 in the theorems above, for specific classes of groups.

The ability to prove statistical theorems about random elements of fixed word length in a hyperbolic group depends on the important fact that the language of geodesics in a hyperbolic group is *regular*. This property does not hold for many groups of interest, so it is important to be able to filter scl by another invariant than word length. Instead, one studies *random walks* on a group G (with respect to a fixed probability measure, which in our case is supported on a symmetric finite generating set) and looks at the distribution of scl on random walks of a given length, conditioned to land in G'.

For hyperbolic groups, the situation for random walks and for random geodesics is comparable. Our main result is:

Hyperbolic Random Walk Theorem 4.13. *Let w_n be the nearest neighbor random walk of length n on the Cayley graph of a hyperbolic group G with respect to a symmetric generating set. Then for any positive C_1 there are positive constants C_2, C_3 such that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(w_n) \leq C_3 n/\log n \mid w_n \in G') \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hw_n) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon}) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.20.

If S_n is a sequence of random variables, we use the informal notation $S_n \sim f(n)$ to describe the "order of magnitude" of a typical element in the distribution S_n (see § 4.5 for a precise definition). With this notation, our main theorem concerning random walks in arbitrary groups is as follows:

Random Walk Theorem 4.25. *Let G be an arbitrary group, and H a finitely generated subgroup of G with finite symmetric generating set S. Suppose Q/H^1 is nonzero (so that* scl *does not vanish identically on G'). For any sufficiently large constant C, let* scl_n *denote the random variable equal to the value of* scl *on a random walk of length contained in $[n - C, n + C]$ in H, conditioned to land in H'. Then*

$$\sqrt{n} \precsim \mathrm{scl}_n \precsim n/\log n.$$

The upper bound comes from our results on random walks in free groups, and the monotonicity of scl under homomorphisms. The lower bound comes from a central limit theorem (due to Björklund–Hartnick [3], of which we learned during the writing of this paper) for the distribution of values of a quasimorphism, together with Bavard duality.

One technical drawback of Björklund–Hartnick is that their results only hold for *symmetric* random walk. For many interesting geometric applications, it is important to be able to deal with *semigroups* of transformations. So we consider it significant that we can show the following:

Rotation Number Central Limit Theorem 4.14. *Let S be a finite subset of* $\mathrm{Homeo}^+(S^1)$*, and $\widetilde{S}$ a collection of lifts of elements of S to* $\mathrm{Homeo}^+(\mathbb{R})$*. Let $X := X_0, X_1, \cdots$ be a random process taking values in* $\mathrm{Homeo}^+(\mathbb{R})$*, where $X_0 = \mathrm{Id}$,*

and $X_{n+1} = X_n \widetilde{s}$ *where* $\widetilde{s}$ *is chosen from* $\widetilde{S}$ *with the uniform distribution. Then* $X_n(0)$ *satisfies a central limit theorem; i.e. there are constants* E, σ *so that*

$$n^{-1/2}(X_n(0) - nE) \xrightarrow{pr} N(0, \sigma)$$

where $N(0, \sigma)$ *denotes the normal distribution with standard deviation* σ*.*

In the special case that S is symmetric, this theorem is superseded by [3]. Nevertheless, even in this case our proof is of independent interest.

We hope that this paper will be of interest to geometers, group theorists, and ergodic theorists. Because of the diversity of our potential readership, we have included proofs of some facts that might ordinarily be left as background.

2. Background

We recall here some standard definitions and facts for the convenience of the reader. A basic reference for the material in this section is [10].

2.1. Stable commutator length.

Definition 2.1. Let G be a group, and G' its commutator subgroup. Given $g \in G'$, the *commutator length* of g, denoted $\mathrm{cl}(g)$, is the least number of commutators in G whose product is g, and the *stable commutator length*, denoted $\mathrm{scl}(g)$, is the limit

$$\mathrm{scl}(g) = \lim_{n \to \infty} \frac{\mathrm{cl}(g^n)}{n}$$

If X is a space with $\pi_1(X) = G$, conjugacy classes in G correspond to homotopy classes of loops in X, and conjugacy classes in G' correspond to homotopy classes of homologically trivial loops in X.

For an element $g \in G'$ associated to the homotopy class of $\gamma : S^1 \to X$, we have $\mathrm{cl}(g) \leq n$ if and only if there is an oriented surface S of genus n with one boundary component, and a map $f : S \to X$ such that $\partial f : \partial S \to X$ is in the free homotopy class of γ. A more sophisticated relationship between scl and maps of surfaces with multiple boundary components can be formulated (see [10], Chapter 2), but it is superfluous for this paper.

Stable commutator length extends to a pseudo-norm on the real vector space $B_1^H(G) := B_1(G)/H$ where $B_1(G)$ is the vector space of real group 1-boundaries (in the bar complex), and H is the real subspace spanned by expressions of the form $g - hgh^{-1}$ and $g^n - ng$. For general groups, scl is not a norm on $B_1^H(G)$, however it is a norm on this space when G is (word) hyperbolic. See [10], § 2.6.

Occasionally it is necessary to estimate the scl of a product in terms of scl of the factors. One has the following estimate:

Lemma 2.2. *Let* G *be a group, and* h*,* g *elements of* G'*. Then there is an inequality*

$$|\mathrm{scl}(hg) - \mathrm{scl}(g)| \leq \mathrm{scl}(h) + 1/2$$

Proof. In a free group with free generators x and y, one has $\mathrm{scl}(yx - x - y) = 1/2$, where $yx - x - y$ is thought of as an element of $B_1^H(F)$. But scl is monotone non-increasing under homomorphisms, so $\mathrm{scl}(hg - g - h) \leq 1/2$ and therefore $|\mathrm{scl}(hg) - \mathrm{scl}(g + h)| \leq 1/2$. By the properties of a pseudo-norm, $|\mathrm{scl}(g + h) - \mathrm{scl}(g)| \leq \mathrm{scl}(h)$, and the lemma is proved. □

Another elementary estimate is as follows:

Lemma 2.3. *Let G be a group with finite symmetric generating set S. Let $g \in G'$. Then there is a constant C_1 so that*

$$\mathrm{scl}(g) \leq C_1|g|$$

where $|\cdot|$ denotes word length with respect to S.

Proof. Since G is finitely generated, so is $H_1(G)$. For the sake of argument, assume $H_1(G)$ is torsion free (the general case is very similar). We can therefore choose a new generating set S' with generators s_i, r_j where the s_i generate independent $\mathbb{Z}$ summands in $H_1(G)$, and the r_j are trivial in $H_1(G)$. Let $n = \max \mathrm{cl}(r_j)$. Rewriting a word in the S generators as a word in the S' generators multiplies length by at most a constant, so without loss of generality we can assume the generators have the desired property.

For any words u, v, w we have

$$uvw = vu[u^{-1}, v^{-1}]w = vuw[w^{-1}u^{-1}w, w^{-1}v^{-1}w]$$

In other words, we may move a specific letter to anywhere in the word at the cost of adding a commutator. Hence we may rewrite any word g (not necessarily in G) as $s_1^{a_1}s_2^{a_2}\cdots r_1^{b_1}r_1^{b_2}\cdots h$ where h is a product of at most $|g|$ commutators, and the a_i, b_i are the exponent sums of the respective generators in g. By hypothesis, each a_i is zero, and each $r_i^{b_i}$ can be written as a sum of at most nb_i commutators, so $\mathrm{cl}(g) \leq (n+1)|g|$, and therefore certainly $\mathrm{scl}(g) \leq (n+1)|g|$. □

2.2. **Quasimorphisms.** There is a duality between stable commutator length and certain functions on G called *homogeneous quasimorphisms*.

Definition 2.4. Let G be a group. A function $\phi : G \to \mathbb{R}$ is a *quasimorphism* if there is some least non-negative number $D(\phi)$ called the *defect*, so that for all $g, h \in G$, there is an inequality

$$|\phi(gh) - \phi(g) - \phi(h)| \leq D(\phi)$$

A quasimorphism is *homogeneous* if, further, it satisfies $\phi(g^n) = n\phi(g)$ for all $g \in G$ and all $n \in \mathbb{Z}$.

Denote the vector space of all quasimorphisms on G by $\widehat{Q}(G)$, and the subspace of homogeneous quasimorphisms by $Q(G)$. Given any $\psi \in \widehat{Q}$, the *homogenization* $\overline{\psi}$, defined by

$$\overline{\psi}(g) = \lim_{n\to\infty} \psi(g^n)/n$$

exists and satisfies $|\overline{\psi} - \psi| \leq D(\psi)$. Moreover, $\overline{\psi}$ is a homogeneous quasimorphism with $D(\overline{\psi}) \leq 2D(\psi)$. See [10], Lem. 2.21 and 2.58. It follows that the vector space $Q(G)$ can be identified as the quotient $\widehat{Q}/C_b^1(G)$, where $C_b^1(G)$ denotes the vector space of L^∞ functions on G.

A quasimorphism has defect zero if and only if it is a homomorphism. Hence $H^1(G)$ is a vector subspace of $Q(G)$, and $\widehat{Q}/H^1$ and Q/H^1 are Banach spaces with respect to the norm $D(\cdot)$; see [10], Cor. 2.57.

Generalized Bavard Duality ([10], Thm. 2.79) is the statement that Q/H^1 with the norm $2D(\cdot)$ is the isometric dual of $B_1^H(G), \mathrm{scl}$. For the case of individual elements $g \in G'$, this was proved by Bavard [1].

By (generalized) Bavard duality, a quasimorphism can be used to give a lower bound on stable commutator length. This observation is elaborated further in § 4.

3. Random geodesics

In this section we study the distribution of scl on the set of elements in G' with word length n in a fixed generating set S for G. In fact, to get meaningful results, it is necessary to fix a positive constant C (depending on G and the generating set S) and study the distribution of scl on the set of elements in G' with word length contained in a window $[n-C, n+C]$. This is because the condition that $g \in G'$ may nontrivially constrain the word length of g mod m, for some m. For example, in a free group F with free generating set, every element of F' has even word length.

For this and other reasons, it is quite difficult to make statistical assertions about the properties of words in G' directly. Our strategy therefore is to make statistical assertions about the properties of words in G, and then infer properties of words in G' by using estimates on the relative proportion of G' in G.

One remark is in order. We state our results in the form "the proportion of words in G_n for which condition X holds is at least/most $f(n)$" for various conditions X and functions n. This is a shorthand for the mathematical statement

$$\mathbf{P}(g \text{ satisfies } X \mid g \in G_n) \geq \text{ or } \leq f(n)$$

where G_n (or whatever finite set is being investigated) has the uniform probability measure. We make the convention that this inequality is required to hold *for all sufficiently large* n, unless we explicitly say otherwise.

3.1. Free groups.

Our discussion begins with free groups for two reasons. Firstly, the statements and proofs are relatively clean, and secondly because these statements can be used to derive conclusions about the behavior of scl under *random walks* in free groups, and thereby about the behavior of scl on random walks in *arbitrary* groups.

In the sequel, fix a free group F of rank k with finite symmetric free generating set S (so that $|S| = 2k$), and let $|\cdot|$ denote word length with respect to S. Let F_n denote the set of elements of word length n (equivalently, the set of reduced words in the generators) and $F'_n := F' \cap F_n$. Note that F'_n is empty unless n is even. For $g \in F$, let $|g|$ denote the word length of g. Since there is a natural bijection between elements of F and reduced words in the generating set, we feel free to move back and forth between elements and words as necessary.

For free groups, Sharp proved the following asymptotic estimate:

Theorem 3.1 (Sharp [39], Thm. 1)**.** *Let F be a free group of rank $k \geq 2$. Then there is an explicit constant σ depending on k so that*

$$\lim_{n\to\infty,\ n \text{ even}} \left| \sigma^k n^{k/2} \frac{|F'_n|}{|F_n|} - \frac{2}{(2\pi)^{k/2}} \right| = 0$$

where the limit is taken over even *positive integers n.*

More qualitatively, the proportion of reduced words of length in the interval $[n, n+1]$ that are in F' is $n^{-k/2}$. Consequently, if we can prove a statistical property of words in F that holds for all but $o(n^{-k/2})$ of the words of length n, it must hold for all but $o(1)$ of the words in F' of length approximately n.

Sharp proved an analogous theorem for hyperbolic groups, that we shall state and use in § 3.3.

For an arbitrary element g of F, it does not make sense to talk about $\mathrm{scl}(g)$. Let $\alpha : F \to \mathbb{Z}^k$ denote the abelianization map, where the standard generators of F are

taken to the co-ordinate generators of $\mathbb{Z}^k$. By abuse of notation, let $|\cdot|$ denote the L^1 norm on $\mathbb{Z}^k$. The following lemma lets us move back and forth between reduced words in F and cyclically reduced homologically trivial words.

Lemma 3.2. *Let g be an element in F, representing $\alpha(g)$ in $H_1(F)$. Then there is an element h in F so that the following holds:*

1. $|h| \leq |\alpha(g)| + 4$;
2. $hg \in F'$;
3. *the product of the reduced representatives of h and g is cyclically reduced.*

If we identify elements with their reduced representatives, then bullet 3 can be restated as saying that hg is cyclically reduced. We call hg the *correction* of g.

Proof. Let x, y and so on be free generators for F. There is a unique reduced word f of the form $x^a y^b \cdots$ with length $\alpha(g)$ and such that $\alpha(f) = -\alpha(g)$. In most cases, either f or a word obtained from it by permuting the terms corresponding to the generators will suffice as h. There are a few exceptional cases, which can be treated in an ad hoc manner. The "worst" case is when $\alpha(g) = 0$ and there is a generator x so that (the reduced representative of) g starts with x^{-1} and ends with x; in this case one must take h to be (for instance) the commutator $xyx^{-1}y^{-1}$. □

Now, let g be a *cyclically reduced* word in F, and suppose we can express g^n as a product of $\mathrm{cl}(g^n)$ commutators. Then by Culler [18] (also see [9]) there is a fatgraph Γ with underlying surface S, so $\chi(\Gamma) = \chi(S) = 1 - 2\mathrm{cl}(g^n)$, and whose boundary component is labeled by the cyclic word g^n in such a way that the labels on opposite sides of a fat edge of Γ are inverse words.

The vertices of Γ divide the boundary cyclic word g^n into segments, so that each segment and its inverse appears an equal number of times. A vertex of Γ of valence d contributes $2 - d$ to χ; it follows that each segment of g^n contributes at least $(d-2)/d \geq 1/3$ to $-\chi$. Since $\mathrm{cl}(g^n) = \mathrm{genus}(S) = (1 - \chi(\Gamma))/2$ it follows that to obtain a lower bound on scl, we only need to obtain an upper bound on the length of the segments, equivalently, to obtain an upper bound on the length of the pair of subwords in g^n that are inverse in F. Explicitly, we make the following definition:

Definition 3.3. If g is a reduced word in a free group, let $\sigma(g)$ denote the maximum length of a pair of mutually inverse subwords of g, and $\sigma_c(g)$ the maximum length of a pair of mutually inverse cyclic subwords of g.

Then we have an estimate:

Lemma 3.4. *Let $g \in G'$ be cyclically reduced. Then* $\mathrm{scl}(g) \geq |g|/6\sigma_c(g)$.

Proof. We have

$$\mathrm{scl}(g) = \lim_{n\to\infty} \mathrm{cl}(g^n)/n = \lim_{n\to\infty} (1 - \chi(\Gamma))/2n$$

where Γ is a fatgraph bounding g^n. By the definition of $\sigma_c(g)$ we have at least $n|g|/\sigma_c(g)$ segments in $\partial\Gamma$, each contributing at least $1/3$ to $-\chi$. Taking n arbitrarily large, we get the desired inequality. □

On the other hand, we will study scl in words that are reduced but not necessarily cyclically reduced. For such words we have the following lemma:

Lemma 3.5. *If g is reduced but not cyclically reduced, and hg is the correction of g, then $\sigma_c(hg) \leq 2\sigma(g) + |h|$.*

Proof. At most one of an inverse pair of cyclic words can intersect h in hg, and at least half of its length in g is part of an inverse pair of words in g. □

It is easier to estimate σ in a random reduced word than σ_c in a random cyclically reduced word, since reduced words (unlike cyclically reduced words) are generated by a stationary Markov process.

Lemma 3.6. *Let F be a free group. For any positive C_1 there is a positive constant C_2 so that the proportion of the set of reduced words g in F_n with $\sigma(g) \geq C_2 \log n$ is less than n^{-C_1}.*

Proof. Let g be a reduced word in F of length n, and let w be a subword of g. A copy of w^{-1} in g must be disjoint from w. Let $g = uwv$ as reduced words. Conditioned on containing w in its given location, the subword v (resp. u) read forwards (resp. backwards) can be obtained by a Markov process, where each successive letter is chosen from S with equal probability subject to there being no cancellation. Starting at each given prefix of v, the probability that the following substring of v of length $|w|$ (if it exists) will be equal to w^{-1} is precisely $(2k-1)^{-|w|}$. So the probability that some copy of w^{-1} exists is at most $(n-2|w|)(2k-1)^{-|w|}$. There are $n+1-m$ subwords of length m for any m, and therefore the probability that $\sigma(g) \geq m$ is less than $n^2(2k-1)^{-m}$. Hence

$$\mathbf{P}(\sigma(g) \geq C_2 \log n) < n^2 C_3^{-C_2 \log n} \leq n^{-C_1}$$

for C_2 sufficiently large. □

Putting these estimates together, we get a lower bound on scl for most words in F'_n.

Proposition 3.7. *For any positive C_1 there is a positive constant C_2 so that the proportion of the set of reduced words g in F'_n (with n even) with $\mathrm{scl}(g) \leq C_2 n/\log n$ is less than n^{-C_1}.*

Proof. By Lemma 3.6, for all but a set of proportion n^{-C_1} of g in F_n, we have $\sigma(g) \leq C_2 \log n$. By Theorem 3.1 we have $\sigma(g) \leq C_2 \log n$ for all but a set of proportion $C_3 n^{k/2-C_1}$ in F'_n. By Lemma 3.2, each such g has a correction hg of length at most $n+4$, and by Lemma 3.5 we have $\sigma_c(hg) \leq 2C_2 \log n + 4$. It follows that $\mathrm{scl}(hg)$ is at least $(n+4)/(12C_2 \log n + 24)$ for such hg, and therefore $\mathrm{scl}(g)$ is at least $(n+4)/(12C_2 \log n + 24) - C_3$. Relabeling constants, the proposition is proved. □

To get upper bounds on scl of a random word in F'_n requires a different argument. Let g be a reduced word in F'. For any integer m the *m-words* of g are the successive subwords of g of length m. Denote these by $g_1, g_2, \cdots, g_{\lfloor n/m \rfloor}$. This defines an integer-valued measure μ_g on F_m with total mass $\lfloor n/m \rfloor$, where for each $v \in F_m$, we set $\mu_g(v)$ equal to the number of m-words equal to v (as a string) (the measure μ_g depends on m, but we suppress this in our notation). The following lemma is key:

Lemma 3.8. *For any positive C_1 there are positive constants C_2, C_3 so that for $m = C_3 \log n$ the proportion of words g in F_n for which there is an inequality*

$$\sum_{v \in F_m} |\mu_g(v) - \mu_g(v^{-1})| > n^{1-C_2}$$

is at most n^{-C_1}.

On the other hand, Lemma 3.6 says that for big enough C_3, and for $m = C_3 \log n$, with high probability at most one of $\mu_g(v)$ and $\mu_g(v^{-1})$ is non-zero for all v, and therefore

$$\sum_{v \in F_m} |\mu_g(v) - \mu_g(v^{-1})| = \lfloor n/m \rfloor \sim n/\log n$$

The proof of Lemma 3.8 requires a technical estimate on rate of mixing in Markov chains (Theorem 3.9, below). But first we give a rough idea of the proof using only very crude estimates, and then explain what steps are required to make the argument rigorous.

We think of the sequence of m-words g_i as a (stationary ergodic aperiodic) Markov process with states F_m. This Markov process determines a random walk in $\mathbb{Z}^d$ with $d = |F_m|/2$ where the coordinates are identified with inverse pairs $v, v^{-1} \in F_m$ and the coefficients at each stage are the number of v's minus the number of v^{-1}'s among the g_i seen so far.

Assume for the moment that the g_i are independent, and uniformly distributed among the possible v. Then the random walk is called a Polya random walk on $\mathbb{Z}^d$. Let S_n be the result of a Polya random walk of length n on $\mathbb{Z}^d$ starting at the origin. We claim that for any positive ϵ, δ, there is an estimate

$$\mathbf{P}(|S_n|_1 \geq n^{1/2+\epsilon}) \leq n^{-\delta}$$

where $|\cdot|_1$ denotes the L^1 norm, and $n \geq \max(d^{C_4}, C_5)$ where C_4 and C_5 are constants depending only on ϵ and δ.

To see this, let e_i be the unit coordinate vectors for $i = 1$ to d, let v_i be the number of e_i's in the walk, and v_{-i} the number of $-e_i$'s. If we let $k_i = |v_i + v_{-i}|$ then each k_i is nonnegative, and $\sum k_i = n$. By independence, the value of the difference $v_i - v_{-i}$ can be thought of as the result of a Polya random walk of length k_i on $\mathbb{Z}$ starting at the origin.

Now, for any positive ϵ', δ' and any $k \leq n$ it is certainly true that

$$\mathbf{P}(|v_i - v_{-i}| > n^{1/2+\epsilon'} \mid |v_i + v_{-i}| = k) < n^{-\delta'}$$

whenever $n \geq C_5$ depending on ϵ' and δ'. Since the right hand side does not depend on k, we have an unconditional estimate

$$\mathbf{P}(|v_i - v_{-i}| > n^{1/2+\epsilon'} \text{ for some } i) < dn^{-\delta'}$$

for $n \geq C_5$. If $d = n^\alpha$ then

$$\mathbf{P}(\sum_i |v_i - v_{-i}| > n^{1/2+\epsilon'+\alpha}) < n^{\alpha-\delta'}$$

So if we choose α sufficiently small so that $\epsilon > \epsilon' + \alpha$ and $\delta < \delta' - \alpha$, the claim is proved.

Of course, the g_i are not strictly independent; nevertheless, the correlations decay exponentially at a rate depending on the second biggest eigenvalue of the adjacency matrix of the Markov chain.

We fix the following notation. Consider a stationary aperiodic ergodic Markov chain with finite state space V. Let M be the probability transition matrix. Then 1 is an eigenvalue of M, and by the Perron–Frobenius theorem, every other eigenvalue has absolute value bounded below 1. Let $\lambda_1 < 1$ be the absolute value of the biggest eigenvalue other than 1. Let π be the stationary probability on V, and let v_0 be a particular state of the Markov chain. Define $a = \lambda_1 + \pi(v_0) - \lambda_1 \pi(v_0)$ and let

$1 < \beta < 1/\pi(v)$. Let S_n be the number of returns to the state v_0, for some given initial distribution $q(\cdot)$. Then for any n, there is the following estimate:

Theorem 3.9 (Kahale [29], Thm. 4.2)**.**

$$\mathbf{P}(S_n \geq \beta\pi(v_0)n) \leq \frac{\beta}{1-a+\beta a}\sqrt{\sum_{v\in V}\frac{q(v)^2}{\pi(v)}}\exp\left(-(\beta-1)^2\pi(v_0)(1-\lambda_1)n\right)$$

In our context, we have $\pi(v) = 1/d$ for all v (in particular $\pi(v_0) = 1/d$), and want to take $\beta = 1 + n^{-1/2-\epsilon}$. There is no harm in taking $q(\cdot)$ to be the Dirac distribution concentrated on some (any) initial state, in which case the square root term is $\sqrt{d}$. It remains to estimate λ_1.

Lemma 3.10. *Let X be an ergodic, recurrent Markov chain on a finite state space V, with transition probability P_{ij} to move from state i to state j. Let X' be the associated Markov chain on the space whose states are the possible words of length n in the Markov chain X, where $v \to w$ is an allowable transition in X' if the suffix of v of length $(n-1)$ is equal to the prefix of w of length $(n-1)$, and where the probability of the transition P'_{vw} is equal to P_{ij}, where i is the last letter of v and j is the last letter of w. Then with notation as above, $\lambda_1(X) = \lambda_1(X')$.*

Proof. Let M be the transition matrix of X, and M' the transition matrix of X', so that $\lambda_1(X)$ means the second biggest eigenvalue of M, and $\lambda_1(X')$ means the second biggest eigenvalue of M'. We will prove the stronger result that the spectrum of M' is equal to the spectrum of M, padded by zeroes. This follows by showing that the traces of M^l and $(M')^l$ are equal for every integer l.

The trace of M^l is equal to the probability that a path of length l in X is a *cycle*; i.e. that it starts and ends at the same state. But for any l there is an obvious (probability-preserving) bijection between the cycles of length l in X and the cycles of length l in X'. Consequently the traces of M^l and $(M')^l$ are equal for all l, and the proof follows. □

We can now prove Lemma 3.8:

Proof. Let X be the Markov chain producing a random reduced word in F, and let X' be as in Lemma 3.10, with transition probabilities P'. The successive m-words in g are governed by a Markov chain X^m with the same states as X', but with transition probability $(P')^m$. Hence by Lemma 3.10, we have $\lambda_1(X^m) = (\lambda_1(X))^m < \lambda_1(X) < 1$. In particular, these eigenvalues are bounded away from 1, independently of m. Note that $d = 4 \cdot 3^{m-1}$ is the number of states of X^m.

Taking $\beta = 1 + n^{-1/2-\epsilon}$, by Theorem 3.9 and Lemma 3.10, we can estimate

$$\mathbf{P}(|v_i - v_{-i}| > 2n^{1/2+\epsilon}) \leq C_1 d^{1/2} e^{-n^{2\epsilon}d^{-1}C_2}$$

where C_1 and C_2 are independent of n and d. If $d = n^\alpha$ then choosing $\alpha < 2\epsilon$ gives the desired result for sufficiently large n (depending on α). □

Remark 3.11. If one takes ϵ arbitrarily close to $1/2$, then we can also take α arbitrarily close to $1/2$. In the case of a free group of rank k, we have $d = (2k)\cdot(2k-1)^{m-1}$ so setting $d \sim n^{1/2}$ gives $m \sim \log n/2\log(2k-1)$.

Proposition 3.12. *For any positive C_1 there is a positive constant C_2 so that except for a set of proportion n^{-C_1}, every $g \in F_n$ has a correction hg with $|h| \leq n^{1/2+\epsilon}$ and $\mathrm{scl}(hg) \leq C_2 n/\log n$.*

Proof. By Lemma 3.8, except for a set of words of proportion n^{-C_1} we can assume that the m-words in g can all be paired except for a subset of total length $C_3 n^{1-C_2} \log n$. Moreover, again except for a set of words of proportion n^{-C_4}, we can assume the length of $|h|$, which is approximately equal to the norm of $\alpha(g)$, is $\leq n^{1/2+\epsilon}$. For all words with both properties, g can be written as a product of $C_5 \log n$ commutators with a word of length $C_5 \log n$. Applying Lemma 2.3 completes the proof. □

Putting Proposition 3.12 together with Proposition 3.7 and Theorem 3.1 we obtain the following:

Theorem 3.13. *In a free group F of finite rank, for any positive C_1 there are positive constants C_2, C_3 so that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(g) \leq C_3 n/\log n \mid g \in F_n') \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hg) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon} \mid g \in F_n) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.2.

Remark 3.14. Although scl is only defined in F', it is important for later applications to be able to simultaneously estimate scl of the correction hg together with the length of the h term. The point is that under a homomorphism of groups $\rho : F \to G$, a homologically nontrivial element g may become homologically trivial, and we would like to estimate (from above) scl of $\rho(g)$. This can be done if we can control $\mathrm{scl}(hg)$ in F, and $|h|$ in F, since scl can only go down under homomorphisms, and the length of h can only expand by a constant factor. Applying Lemma 2.2 then lets us estimate $\mathrm{scl}(\rho(g))$. This will be especially important in § 4, where we are concerned with the statistical distribution of scl on words in G obtained by a *random walk* of length n in a fixed symmetric generating set.

3.2. **Experimental data.** We computed the mean and the standard deviation of $\mathrm{scl}\log n/n$ on a random sample of 50,000 cyclically reduced words of length n in F_2' in the standard generating set, for $n = 12k$ and $1 \leq k \leq 6$, using the program `scallop` [13], implementing the algorithm described in [9] (modified as explained in [10], § 4.1.7). These results are tabulated in Table 1.

TABLE 1. random cyclically reduced words in F_2' of length n

n	$\mathbf{E}(\mathrm{scl})\log n/n$	$\sigma(\mathrm{scl})\log n/n$	$\sigma(\mathrm{scl})\sqrt{n}/\mathbf{E}(\mathrm{scl})$
12	0.1299	0.03711	0.9888
24	0.1478	0.02040	0.6753
36	0.1497	0.01374	0.5508
48	0.1506	0.01154	0.5310
60	0.15104	0.009800	0.5026
72	0.15138	0.008699	0.4876

It appears from the data that $\mathrm{scl}\log n/n$ approaches a limiting value of approximately 0.152 a.s. as $n \to \infty$. Moreover it seems plausible that $\sigma(\mathrm{scl})\sqrt{n}/\mathbf{E}(\mathrm{scl})$ approaches a positive limit as $n \to \infty$. The shape of the distribution of $\log(\mathrm{scl})$

appears to be almost, but not quite, Gaussian for $n = 72$. For various reasons, the value of scl (which is always rational in a free group by [9]) is more likely than not to have a small denominator. The histogram of values on words of length 72 has spikes at 5/2, 3 and 7/2; see Figure 1.

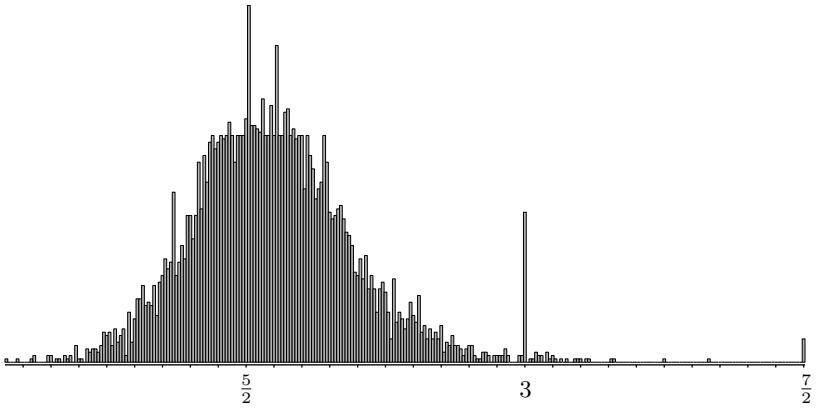

FIGURE 1. Histogram of scl on random words of length 72 in F_2'

Experiments in higher rank are much more computationally costly, and we were only able to test words of length up to about 48 in F_3'; see Table 2. Nevertheless, the data behaves qualitatively similar to the rank 2 case (as one would expect).

TABLE 2. random cyclically reduced words in F_3' of length n

n	$\mathbf{E}(\text{scl}) \log n/n$	$\sigma(\text{scl}) \log n/n$	$\sigma(\text{scl})\sqrt{n}/\mathbf{E}(\text{scl})$
12	0.2118	0.05466	0.8941
24	0.2306	0.02603	0.5529
36	0.2346	0.01808	0.4625
48	0.2354	0.01485	0.4371

This motivates the following conjecture:

Conjecture 3.15. *Let F be a free group of finite rank. For n even, let scl_n denote the value of* scl *on a random element of F_n'. Then there is a positive constant C so that there is convergence in probability*

$$\text{scl}_n \log n/n \xrightarrow{pr} C$$

The constant C should depend (of course) on the rank of F. It would be interesting to give a formula for C (or even to estimate its rate of growth) in terms of the rank.

3.3. **Hyperbolic groups.** We now explain how to generalize the arguments of the previous section to (Gromov) hyperbolic groups. For a basic introduction to hyperbolic groups, see [26]. Our arguments in the remainder of this section will make use of geometric properties of geodesics and surface subgroups in such groups.

It is convenient to use so-called *pleated surfaces*, first introduced by Thurston to study Kleinian groups, and adapted to hyperbolic groups in [11] using Mineyev's flow space [35].

One caveat is that we are studying arbitrary elements in G, rather than conjugacy classes. However, it is a fact that in a hyperbolic group, almost every element is very close to being a shortest representative in its conjugacy class, and therefore this issue can addressed.

We fix the following notation in the sequel. Let G be a hyperbolic group, and S a symmetric generating set. Let $C_S(G)$ (or C for short, if S and G are understood) be the Cayley graph for G with respect to G. We think of C as a geodesic metric space in the usual way, by making each edge isometric to the unit interval. Let ∂C denote the Gromov boundary of C (which is also the Gromov boundary of G).

3.4. **Pleated surfaces and Mineyev's flow space.** Mineyev's *flow space* $\overline{\mathcal{F}}(C)$ is a metric space which is G-equivariantly quasi-isometric to C, and homeomorphic to the disjoint union of unoriented isometrically embedded copies of $\mathbb{R}$ (called the *flowlines*) for each pair of distinct points in ∂C. The key feature of this space for our purposes is the *synchronous* exponential convergence of flowlines asymptotic to a common ideal point, a property enjoyed by complete simply-connected manifolds of negative curvature, but not by Cayley graphs of hyperbolic groups in general. See [35], especially § 3.2, § 8.6, § 13 and Theorems 44 and 57 for details.

Remark 3.16. In fact, Mineyev constructs a bigger flow space $\mathcal{F}$ which is the disjoint union of *oriented* flowlines. This oriented flow space admits an $\mathbb{R}$ action that translates each flowline, and an anti-commuting $\mathbb{Z}/2\mathbb{Z}$ action that interchanges each flowline with its oppositely oriented counterpart. The caveat is that the bigger flow space is merely a *pseudometric* space, in which the $\mathbb{Z}/2\mathbb{Z}$ action moves points a distance 0. This pseudometric descends to a genuine metric on the unoriented quotient.

Definition 3.17. A pleated surface (possibly with boundary) in $\overline{\mathcal{F}}(C)$ consists of the following data:

(1) a hyperbolic surface S with geodesic boundary, and a full geodesic lamination L (i.e. one whose complementary domains are all ideal triangles) with $\partial S \subset L$;
(2) a homomorphism $\rho : \pi_1(S) \to G$;
(3) if $\widetilde{L}$ denotes the preimage of L in the universal cover $\widetilde{S}$, a $\pi_1(S)$-equivariant map $\iota : \widetilde{L} \to \overline{\mathcal{F}}(C)$ which multiplies lengths by a fixed constant κ (depending only on G) on each leaf of $\widetilde{L}$

Lemma 3.18 (Calegari–Fujiwara [11]; also [10] Lem. 3.40)**.** *Suppose G is hyperbolic, and $\rho : \pi_1(S) \to G$ is injective on simple loops. Then there is a pleated surface realizing ρ in the sense of Definition 3.17. Moreover, ι is* 1*-Lipschitz.*

An expression of g^n as a product of commutators gives rise to a homomorphism from a surface group $\pi_1(S)$ to G sending the boundary element to g^n. If the surface is not injective on simple loops, it can be compressed, reducing its complexity. Therefore we can construct a pleated representative. The components of $\partial\widetilde{S}$ map to flowlines in $\overline{\mathcal{F}}$ stabilized by conjugates of g.

However, geodesic segments corresponding to elements g in the Cayley graph are not necessarily close to the axes, since a given element may not be (almost) shortest in its conjugacy class.

Definition 3.19. Given $g \in G$, the *translation length* of g, denoted $\tau(g)$ is the limit

$$\tau(g) = \lim_{n\to\infty} |g^n|/n$$

where $|\cdot|$ denotes word length with respect to S.

Note that $\tau(g) \leq |g|$, and in fact the sequence $|g^n|$ is subadditive by the triangle inequality, and therefore the limit in Definition 3.19 exists. One thinks of $\tau(g)$ as being the "shortest" word length in the conjugacy class of g, in an ideal sense. This ideal cannot be reached, but it can be approximated, and it is generally sufficient to bound $|g| - \tau(g)$.

Lemma 3.20. *Let g be a geodesic word in a hyperbolic group G, representing $\alpha(g)$ in $H_1(G)$. Then there is a geodesic word h in G so that the following holds:*

1. *$|h| \leq C_1|\alpha(g)| + C_2$;*
2. *$hg \in G'$;*
3. *hg is almost shortest in its conjugacy class; i.e. $|hg| - \tau(hg) < C_3$.*

Moreover, the constants above depend only on G, and not on g.

Proof. Amongst all the geodesics h' with $\alpha(h') = -\alpha(g)$, choose one of shortest length. Since G is hyperbolic it is finitely generated, and therefore so is $H_1(G)$. Choosing generators for $H_1(G)$, one sees that there is a constant C_1 so that we can choose $|h'| \leq C_1|\alpha(g)|$. Since the action of G on $\partial G \times \partial G - \Delta$ is ergodic (see e.g. [26]), one can find words v_1, v_2 of uniformly bounded length so that $h'v_1g$ and gv_2h' are uniformly quasigeodesic. Then $v_2h'v_1 = h$ satisfies the conclusion of the lemma. □

Note if $g \in G'$ then $|h| \leq C_2$. Moreover, $hg \in G'$ so $h \in G'$. It follows that $|\mathrm{scl}(hg) - \mathrm{scl}(g)| \leq C_4$ for such g. Note that Lemma 3.20 is the analog of Lemma 3.2 for hyperbolic groups.

Lemma 3.20 does not produce geodesics, but only "almost geodesics". In order to keep the notation to a minimum, we will fix, once and for all, a constant C_3 as in Lemma 3.20, and say that for a geodesic word g, the infinite power $\dot{g}$ is "almost geodesic" if $|g| - \tau(g) < C_3$.

Definition 3.21. Fix a constant C_1. If g is geodesic, define $\sigma(g, C_1)$ to be the maximal common length of a pair of segments ℓ, ℓ' in g for which there is some $a \in G$ so that ℓ and $a(\ell')$ are Hausdorff distance at most C_1 apart, and oppositely aligned (with respect to a fixed orientation on g). If $\dot{g}$ is almost geodesic, define $\sigma_c(g, C_1)$ to be the maximal common length of a pair of segments ℓ, ℓ' in $\dot{g}$ for which there is some $a \in G$ so that ℓ and $a(\ell')$ are Hausdorff distance at most C_1 apart, and oppositely aligned (with respect to a fixed orientation on $\dot{g}$).

Generalizing Lemma 3.4 we have an estimate:

Lemma 3.22. *Let $g \in G'$, and let $\dot{g}$ be almost geodesic. Then if C_1 is sufficiently big, depending on G and the constant in the definition of almost geodesic, but not on g, then there is a constant C_2 so that* $\mathrm{scl}(g) \geq |g|/C_2\sigma_c(g, C_1)$.

Proof. We work in Mineyev's flow space. The argument very closely follows the proof of [11], Thm. A, and the reader is referred to that proof for details. Let S be a pleated surface of genus m bounding g^n, which exists by Lemma 3.18. Let l be a geodesic axis for g^n in Mineyev's flow space, and let α be a fundamental domain for the translation action of g^n. As in [11], observe that apart from a subset of α of length at most $(4m-2)C_3$, every point in $p \in \alpha$ can be joined by an arc of length at most C_1 to a point in an oppositely aligned translate l'. A surface of genus m with one boundary component contains at most $6m-3$ disjoint nonparallel proper essential arcs, and each such arc has 2 endpoints. It follows that there is some pair ℓ, ℓ' of segments in l and l' which are oppositely aligned, and Hausdorff distance at most C_1 apart, and satisfying

$$\text{length}(\ell) \geq \frac{\tau(g^n) - (4m-2)C_4}{12m-6}$$

Increasing C_1 slightly, one obtains segments in $\dot{g}$ certifying $\sigma_c(g, C_1) \geq \text{length}(\ell) - C_5$ for some C_5. Since $\tau(g^n) = n\tau(g)$, and since $|g| - \tau(g) < C_3$ (by the definition of almost geodesic), the lemma is proved. □

In the sequel, we fix C_1 as in Lemma 3.22, and abbreviate $\sigma(\cdot, C_1)$ and $\sigma_c(\cdot, C_1)$ by $\sigma(\cdot)$ and $\sigma_c(\cdot)$ respectively. The analog of Lemma 3.5 for hyperbolic groups is straightforward, and therefore estimating a lower bound on $\text{scl}(g)$ reduces to estimating an upper bound on $\sigma(g)$ for a random $g \in G'$ of length n. In the case of free groups, this bound followed from the fact that there was a stationary Markov process which generated random reduced words in F. The analog of this fact for hyperbolic groups is well-known, and is the subject of the next subsection.

3.5. **Combings.** For an introduction to the theory of regular languages and its connection to geometric group theory, see [19].

Definition 3.23. Let G be a hyperbolic group, and S a symmetric finite generating set. A *combing* for G is a regular language $L \subset S^*$ with the following properties:

(1) L is prefix closed;
(2) the evaluation map $L \to G$ is a bijection;
(3) elements of L are geodesics in the word metric determined by S.

Remark 3.24. Many variations on the definition of combing exist in the literature. We use one of the more restrictive definitions, since it has nicer properties.

The following result is essentially due to Cannon, though he used different language to express it.

Theorem 3.25 (Cannon [14])**.** *Let G be a hyperbolic group, and S a finite symmetric generating set. Choose a total order on the elements of S, and extend this to a lexicographic total order on S^*. The language of lexicographically first geodesics is a combing of G (in the sense of Definition 3.23).*

Since L is regular and prefix closed, there is a finite directed graph Γ with edges labeled by elements of S, and with a distinguished "initial vertex", so that words in L are in bijection with directed paths in Γ starting at the initial vertex. By the definition of a combing, there is a natural bijection between G_n and paths of length n in Γ starting at the initial vertex. By taking limits, this bijection extends to a *map* from the set of (right) infinite paths in Γ starting at the initial vertex, to

∂G. This map is surjective, and bounded to one (see e.g. [17] for a comprehensive introduction to this theory).

Notation 3.26. Given $g \in G$, the *path* (in Γ) associated to g is the path of length $|g|$, starting at the initial vertex associated to the unique element of L that evaluates to g. We denote this path by γ_g, or just γ if g is understood.

We refer to the endpoint of the path γ_g as the *image* of g in Γ. We also sometimes refer to this image as the *cone type* of g.

For any directed graph Γ, one defines a new directed graph $C(\Gamma)$ as follows. The vertices of $C(\Gamma)$ are the *components* of Γ; i.e. the maximal subgraphs in which any distinct ordered pair of vertices are joined by a directed path (in the theory of Markov chains one says that the states *communicate*). Then $C(\Gamma)$ is obtained from Γ by contracting each component to a vertex. By construction, $C(\Gamma)$ is itself a directed graph with no directed loops.

Each vertex Γ_i of $C(\Gamma)$ corresponds to a subgraph of Γ. By abuse of notation, we use the same notation for the subgraph of Γ and the vertex of $C(\Gamma)$.

Definition 3.27. For each vertex Γ_i of $C(\Gamma)$, let $\lambda(\Gamma_i)$ be the biggest eigenvalue of the adjacency matrix of Γ_i, and let $\lambda(\Gamma)$ be the supremum of $\lambda(\Gamma_i)$ over all vertices Γ_i. Call a component Γ_i *big* if $\lambda(\Gamma_i) = \lambda(\Gamma)$.

By hypothesis, the adjacency matrix of each Γ_i is Perron-Frobenius, so the biggest eigenvalue is real and positive (for components Γ_i of Γ corresponding to a single vertex of Γ in a singleton communicating class, $\lambda(\Gamma_i) = 0$).

A theorem of Coornaert [16] easily implies the following structure theorem for $C(\Gamma)$ (again, Coornaert did not use this language):

Theorem 3.28 (Coornaert [16])**.** *Let G be hyperbolic with finite symmetric generating set S, and let Γ be a graph parameterizing a combing. Let $C(\Gamma)$ be as above. Then each directed path in $C(\Gamma)$ contains at most one big component.*

In other words, distinct big components cannot occur in series, but only in parallel. This is equivalent to the main theorem of Coornaert in [16] that for a hyperbolic group G, there are constants C_1, C_2 and $\lambda > 1$ so that the number of words $|G_n|$ of length n satisfies

$$C_1 \lambda^n \leq |G_n| \leq C_2 \lambda^n$$

A hyperbolic group for which $C(\Gamma)$ consists of a single big component is said to be *strongly Markov*.

3.6. **Patterson–Sullivan measures.** Instead of considering the uniform measure on G_n, it is convenient to study closely related measures depending on the combing. The following discussion is entirely standard, and therefore we gloss over the details; a basic reference is Coornaert [16] or [23]; also see [12], especially § 4.

For each n, let ν_n be the probability measure on G defined by

$$\nu_n = \frac{\sum_{|g| \leq n} \lambda^{-|g|} \delta_g}{\sum_{|g| \leq n} \lambda^{-|g|}}$$

where δ_g is the Dirac measure on the element g. The measure ν_n extends (by zero) to a probability measure on $G \cup \partial G$.

Definition 3.29. A weak limit $\nu := \lim_{n\to\infty} \nu_n$ is a *Patterson–Sullivan* measure associated to the combing.

It is a standard fact that the limit ν actually exists, and has support contained in ∂G; see e.g. [16]. We will make use of the following fact, also due to Coornaert:

Theorem 3.30 (Coornaert [16] Thm. 7.7)**.** *The action of G on ∂G is ergodic with respect to ν.*

For each $g \in G$, let $\mathrm{Cone}(g)$ denote the subset of ∂G consisting of endpoints of infinite paths in the combing starting at g. There is a natural Radon measure κ on G defined by $\kappa(g) = \nu(\mathrm{Cone}(g))$. By definition, the restriction of κ to G_n is a probability measure for each n.

There is a natural stationary Markov chain with underlying graph Γ, so that the bijection between elements of G_n in the κ measure, and random paths of length n in Γ starting at the initial vertex, is *measure-preserving*. If $g \in G_n$ and $h \in \mathrm{Cone}(g) \cap G_{n+1}$ have images v and w respectively in Γ, the probability of a transition from v to w at each step of the Markov chain is $\nu(\mathrm{Cone}(h))/\nu(\mathrm{Cone}(g)) = \kappa(h)/\kappa(g)$. Given v and w, this ratio does not depend on n or the choice of g and h, and is therefore well-defined. Informally, a random path of length n in Γ for this Markov chain is distributed like the initial segment of length n of a random path (in the uniform measure) of length N with $N \gg n$.

The following lemma shows that for the purposes of estimating scl, it is just as good to use the κ measure as the uniform measure.

Lemma 3.31. *The ratio of the κ measure to the uniform measure takes only finitely many values. Moreover, there is a uniform constant C so that every point in G_n is within distance C of some point in G_n in the support of κ.*

Proof. The ratio of κ to the uniform measure on a point g depends only on the image of g in Γ (i.e. the cone type), proving the first claim. Moreover, there is a constant C (approximately equal to 2δ) so that for any point p there is an open subset $U \subset \partial G$ so that every geodesic from the basepoint to U must pass within C of p. The open set U has positive ν measure, by Theorem 3.30, so some point within distance C of p has a cone with growth rate $C\lambda^n$, and therefore has κ positive. □

We are now in a position to generalize the main results of § 3.1 to general hyperbolic groups.

Lemma 3.32. *Let G be a hyperbolic group. For any positive C_1 there is a positive constant C_2 so that the proportion of the set of reduced words g in G_n with $\sigma(g) \geq C_2 \log n$ is less than n^{-C_1}.*

Proof. The proof is essentially the same as that in Lemma 3.6. A word representing a random element of G_n (in the κ measure) is obtained by a stationary Markov process with state space Γ and suitable transition probabilities.

We now show that there is a constant C_3 so that apart from a subset of G_n of proportion less than n^{-C_1}, every word corresponds to a path contained entirely in a big component of Γ, apart from a prefix and suffix of length at most $C_3 \log n$. This follows from Coornaert's theorem, and its corollary Theorem 3.28, as we now explain. Let μ be the biggest absolute value of an eigenvalue of a non-maximal component. By finiteness, this absolute value exists, and $\mu < \lambda$, the common

biggest eigenvalue of the maximal components. There are at most $\mu^k p(k)$ words of length k not in a big component, for some polynomial p. Therefore the proportion of words of length n with a prefix and suffix of total length at least k, not in a big component, is at most $\mu^k p(k)/\lambda^k$. Consequently, if k is at least $C_3 \log(n)$ for some large C_3, this ratio is at most n^{-C_1} for some other C_1, and as C_3 gets big, so does C_1.

Taking $C_2 > C_3$, we ignore this prefix and suffix, and pay attention to the subset of the walk contained in a single big component Γ_i. Because the component is recurrent, and has $\lambda(\Gamma_i) = \lambda(G) > 1$, any random walk of length C_4 has a positive probability of wandering a distance $\geq C_1$ from any given bi-infinite geodesic in G. The Lemma now follows by the same argument as Lemma 3.6. □

From this lemma we can estimate $\mathrm{scl}(g) \geq C_2 n/\log n$ for an element of G'_n, with probability at least $1 - n^{-C_1}$.

Remark 3.33. In fact, since each directed path in Γ intersects at most one big component, once a path in the stationary Markov process enters a big component, it can never leave again.

3.7. **Upper bounds and coarse cancellation.** To estimate an upper bound, we need to prove an analog of Lemma 3.8.

Let g be an element of G_n. The combing lets us identify the element g with a specific geodesic γ of length n and for some m let $\gamma_1, \gamma_2, \cdots, \gamma_{\lfloor n/m \rfloor}$ be the successive subpaths of length m. These are distributed somehow among the allowable paths of length m in Γ. For each $v \in G_m$, let $\mu_g(v)$ be the number of γ_i for which the value (in G) of γ_i is v.

The central difficulty is to obtain the precise cancellation between $\mu_g(v)$ and $\mu_g(v^{-1})$. This difficulty is twofold. Firstly, even if Γ has a unique big component (i.e. if G is strongly Markov) there is no reason to expect $\mu_g(v)$ and $\mu_g(v^{-1})$ to be almost equal for almost all g. For, although the set of geodesic words of length n is permuted by taking inverses, the set of geodesic words accepted by the regular language L is typically *not*. Secondly, in general the distribution of $\mu_g(v)$ depends *a priori* on which big component the path associated to g spends most of its time in.

The remedy is to study the distribution of μ_g for words g conditioned to spend most of their time in a given big component Γ_i, and then to show that for such words, the mass of μ_g can be partitioned into subsets supported on pairs of *almost inverse* words (we call this *coarse cancellation*).

Now, each big component Γ_i of Γ is itself the underlying graph of a stationary *ergodic* Markov chain. The stationary measure for this Markov chain determines a probability measure on the set of paths in Γ_i of length m, and by evaluation, a probability measure $\mu_{i,m}$ on G_m. A specific random path γ of length n in Γ_i determines a measure μ_γ on G_m of total mass $\lfloor n/m \rfloor$. We would like to compare the measures $\lfloor n/m \rfloor \mu_{i,m}$ and μ_γ.

Lemma 3.34. *For any big component Γ_i of Γ, and for any positive C_1, there are positive constants C_2, C_3 so that for $m = C_3 \log n$ the proportion of random paths γ in Γ_i of length n for which there is an inequality*

$$|\lfloor n/m \rfloor \mu_{i,m} - \mu_\gamma| > n^{1-C_2}$$

is at most n^{-C_1}, where $|\cdot|$ denotes total mass of a signed measure.

Proof. This follows immediately from Theorem 3.9 and Lemma 3.10, which concern general finite stationary ergodic Markov chains, and not the special case of a free group. □

For fixed m, the measures $\mu_{i,m}$ depend on the choice of big component Γ_i. Let p_i be the probability that a random path in Γ enters Γ_i, so that $\sum p_i = 1$. Define a measure μ_m on G_m by $\mu_m = \sum p_i \mu_{i,m}$. As remarked above, it is not typically true that $\mu_m(v) = \mu_m(v^{-1})$ for all $v \in G_m$, but something of the sort is true (the precise statement is given in Lemma 3.36 below). So to get the desired coarse cancellation, it will suffice to show that the measures $\mu_{i,m}$ are all approximately equal to each other, and therefore to μ_m. The key idea is to use the ergodicity of G on ∂G for the Patterson–Sullivan measure to compare $\mu_{i,m}$ and $\mu_{j,m}$ for different i, j.

Lemma 3.35. *There is a constant C so that the following is true: for each pair of big components Γ_i, Γ_j of Γ, there is a map $f_{i,j} : G_m \to Prob(G_m)$ satisfying*

1. *for all g and all h in the support of $f_{i,j}(g)$ there is $a \in G$ with $|a| \leq C$ so that $d(g, aha^{-1}) \leq C$; and*
2. *$\mu_{j,m}(h) = \sum_g \mu_{i,m}(g) f_{i,j}(g)(h)$.*

We express bullet (2) by saying that $\mu_{j,m}$ is obtained by convolving *$\mu_{i,m}$ with $f_{i,j}$.*

Proof. In words, this lemma says that each probability distribution $\mu_{i,m}$ can be "smeared out" (i.e. "convolved" with $f_{i,j}$) a bounded amount to obtain the distribution $\mu_{j,m}$. In fact, the proof is a very simple trick, which is a variation on the main trick of [12]. By Kakutani's random ergodic theorem (see [32]; alternately this follows trivially from Theorem 3.9) and the definition of $\mu_{i,m}$, almost every infinite path in Γ that enters Γ_i is composed of subpaths of length m that are distributed in G_m according to $\mu_{i,m}$. Call an infinite path Γ_i-*typical* if its subpaths have this property, and let γ_i be Γ_i-typical for each i. By abuse of notation, we also think of γ_i as an infinite geodesic path in G, limiting to some point in ∂G. By ergodicity of the action of G on ∂G (i.e. Theorem 3.30), there are typical γ_i and γ_j and some $a \in G$ so that $a \cdot \gamma_i$ and γ_j have the same endpoint, and therefore outside some compact subset, they synchronously fellow-travel. It follows that we can subdivide $a \cdot \gamma_i$ and γ_j into subpaths of length m, throwing away finitely many at the start, so that corresponding subpaths are each contained in the δ-neighborhood of each other. If two associated subpaths evaluate to g and h in G_m respectively, then $g = a_1 h a_2$ where each of a_1, a_2 has length at most δ. Hence $d(g, a_1 h a_1^{-1}) \leq 2\delta$. So define $f_{i,j}(g)(h)$ to be the probability that a subpath of γ_i evaluating to g fellow-travels (as above) a subpath of γ_j evaluating to h. □

We conclude that each individual measure $\mu_{i,m}$ can be convolved (in the sense above) by a weighted average $\sum p_j f_{i,j}$ to obtain μ_m. Now, though it might not be true that $\mu_m(v) = \mu_m(v^{-1})$ for all $v \in G_m$, this is approximately true, up to convolution in the sense of Lemma 3.35:

Lemma 3.36. *There is a constant C so that the following is true: there is a map $f : G_m \to Prob(G_m)$ satisfying*

1. *for all g and all h is the support of $f(g)$ there is $a \in G$ with $|a| \leq C$ so that $d(g, ah^{-1}a^{-1}) \leq C$; and*
2. *$\mu_m(h) = \sum_g \mu_m(g) f(g)(h)$.*

Proof. Choose some $N \gg m$, so that with probability $1 - N^{-C_1}$, a random path γ in Γ of length N determines μ_γ for which $||\lfloor N/m \rfloor \mu_{i,m} - \mu_\gamma| \leq N^{1-C_2}$ for some $\mu_{i,m}$. We call a γ for which such an estimate holds (for some i) *almost typical*. By Lemma 3.31, for all but at most $1 - C_3 \cdot N^{-C_1}$ almost typical paths γ, there is an almost typical path γ' so that γ and γ' evaluate to g and g' in G_N respectively, and $d(g^{-1}, g') < C_4$. Let γ^{-1} denote the path in G from Id to g^{-1} obtained by reversing and translating γ (note that γ^{-1} will not typically be a path in Γ). Since subpaths of γ^{-1} are in bijection with subpaths of γ but oriented oppositely, $\mu_\gamma(v) = \mu_{\gamma^{-1}}(v^{-1})$ for all $v \in G_m$. Since γ^{-1} and γ' synchronously fellow-travel, after throwing away suffixes of finite length, we can pair subpaths of γ^{-1} and γ' of length m so that corresponding subpaths are each contained in the δ-neighborhood of each other. Taking $N \to \infty$ (for fixed m) the proof follows. □

We can now deduce the following analog of Proposition 3.12:

Proposition 3.37. *For any positive C_1 there is a positive constant C_2 such that except for a set of proportion n^{-C_2} every $g \in G_n$ has a correction hg with $|h| \leq C_2 n/\log n$ and $\mathrm{scl}(hg) \leq C_2 n/\log n$.*

Proof. By Lemma 3.31, it suffices to show this for the κ measure in place of the uniform measure, since multiplying a word by a constant will not change either the correction or scl by much. As in the proof of Lemma 3.32, apart from a prefix of length at most $C_3 \log n$, we can assume the path associated to g is contained in a single big component Γ_i. By Lemma 3.34, the subwords of g in Γ_i of length m are distributed nearly according to the law $\mu_{i,m}$, apart from a subset of length $n^{1-\epsilon}$. By Lemma 3.35 and Lemma 3.36 these subwords can be partitioned into pairs b, b' so that there are elements a of length at most 2δ for which $d(ab'a^{-1}, b^{-1}) \leq 4\delta$. Thus g can be written as a product of $C_4 n/\log n$ commutators times a word of length $C_5 n/\log n$. Relabeling constants, the proof follows. □

Finally, Sharp proved the following analog of Theorem 3.1 for strongly Markov groups:

Theorem 3.38 (Sharp [39], Thm. 3)**.** *Let G be a strongly Markov group with $\dim(H_1(G)) = k$. Then there are explicit positive constants C_1 and C_2 so that*

$$\lim_{n\to\infty} \left| n^{k/2} \sum_{m \in [n-C_1, n+C_1]} \frac{|G'_m|}{|G_m|} - C_2 \right| = 0$$

In fact, Sharp found a precise formula for the constants. For a group with several large components, one gets such an estimate for words corresponding to walks that enter a specific large component of Γ, with an *a priori* different constant C_2 for each component. In any case, the only point we want to take from this theorem is that the relative proportion of G'_m in G_m for $m \in [n - C_1, n + C_1]$ is of order $n^{-k/2}$.

Putting this together, we get the main theorem of this section:

Theorem 3.39. *Let G be a hyperbolic group. For any positive C_1 there are positive constants C_2, C_3 so that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(g) \leq C_3 n/\log n \mid g \in G'_n) \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hg) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon} \mid g \in G_n) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.20.

Emboldened by the experimental results in § 3.2, and some experience in the study of scl in hyperbolic groups, we make the following conjecture:

Conjecture 3.40. *Let G be a hyperbolic group. There is a constant C_1 so that for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}(|\mathrm{scl}(hg)\log n/n - C_1| > \epsilon \wedge |h| < n^{1/2+\epsilon} \mid g \in G_n) \to 0$$

3.8. **Quasiconvex subgroups.**

Definition 3.41. A subspace Y of a geodesic metric space X is *quasiconvex* if there exists a $K \geq 0$ such that, for all $y_1, y_2 \in T$ and for all x contained in a geodesic segment from y_1 to y_2, one has $d(x, Y) \leq K$.

Let G be a hyperbolic group. A subgroup H is *quasiconvex* if H is quasiconvex as a subset of the Cayley graph $C_S(G)$ of G with respect to any finite generating set S.

Since hyperbolic groups are finitely generated, a quasiconvex subgroup H of G is also necessarily finitely generated. It follows that H is a quasiconvex subgroup of G if and only if the inclusion of H in G is a quasi-isometric embedding (with respect to any choice of word metrics on G and H). Note that a quasiconvex subgroup of a hyperbolic group is itself hyperbolic, as an abstract group. It is no more effort to generalize Theorem 3.39 to the following:

Theorem 3.42. *Let G be a hyperbolic group G, and let H be a quasiconvex subgroup. Fix a finite generating set S for H_n, and let H_n denote the set of words in H of length n in the generators S. Then for any positive C_1 there are positive constants C_2, C_3 so that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(g) \leq C_3 n/\log n \mid g \in H_n \cap G') \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hg) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon} \mid g \in H_n) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.20.

Proof. Since scl can only go down under homomorphisms from one group to another, the upper bound follows immediately from Theorem 3.39 applied to H. Note that $H_n \cap G'$ is typically bigger than H'_n, so this follows from the second inequality in Theorem 3.39.

To obtain a lower bound, first observe that an (intrinsic) geodesic in H is uniformly quasigeodesic in G. Just as in Lemma 3.32, a random element of H_n is obtained by a Markov process, so that at every stage there is a uniform positive probability of wandering a distance $\geq C_1$ from any given bi-infinite geodesic in G (and therefore in H). Hence for every C_1 there is a C_2 such that

$$\mathbf{P}(\sigma(g) \geq C_2 \log n \mid g \in H_n) \leq n^{-C_1}$$

where $\sigma(g)$ is measured in G; and from this and Lemma 3.22 we obtain the desired lower bound. □

4. Random walks

One of the most significant features of stable commutator length is its *monotonicity* under homomorphisms. Unfortunately, if $\rho : G \to H$ is a homomorphism between finitely generated groups, it is very hard to compare the pushforward of the uniform measures on G_n with the uniform measures on H_m, and therefore this monotonicity cannot easily be translated into statistical statements about the values of scl on random elements of G_n and H_m.

The remedy is to study the distribution of scl with respect to measures μ and $\rho_*(\mu)$. A natural class of measures on groups that behaves well with respect to pushforward are the measures associated to *random walks*. As in § 3, it is convenient for both technical and expository reasons to work out the theory first in the case of free groups, and then to generalize to other classes of groups.

4.1. **Free groups.** A useful reference for random walks on groups is Woess [41], but we recall some basic definitions for the convenience of the reader. We may think of constructing a random word of length n by multiplying n elements of the group together, chosen independently and with the same distribution, from some fixed subset of the group. The simplest and most important case is to take the uniform probability measure on a generating set. Equivalently, we may take the nearest neighbor random walk of length n on the Cayley graph for the group with respect to the generating set.

If we consider the free group F of rank k with respect to the standard generating set, then the Cayley graph is the infinite $2k$-valent tree T_{2k}, and random words of length n correspond to nearest neighbor random walks of length n in T_{2k} starting at the origin. The nearest neighbor random walk on T_{2k} is spherically symmetric; i.e. the distribution of random words of length n is a weighted sum of uniform distributions on F_m, the sphere of radius m, for $m \leq n$. This immediately gives an upper bound of order $O(n/\log n)$ on the growth rate of scl under random walk in a free group, and to establish a lower bound it is sufficient to show that a random walk in T_{2k} makes linear progress from the origin (with sufficient control on the error). For technical reasons, it is convenient to obtain exponential control over the probability of a large deviation. Such an estimate is standard, but it is not much work to give a rigorous proof which will serve as a model for the analogous case of hyperbolic groups.

Let w_n denote random walk on F_k. The word length $|w_n|$ of w_n in F is a stationary Markov process on $\mathbb{Z}_{\geq 0}$, with transition probabilities given by

$$\begin{aligned} p(0,1) &= 1 \\ p(n,n+1) &= (2k-1)/2k \text{ if } n > 0 \\ p(n,n-1) &= 1/2k \text{ if } n > 0 \\ p(i,j) &= 0 \text{ otherwise.} \end{aligned}$$

Here $p(i,j)$ is the probability to move from state i to state j at each step of the process.

Set λ to be the expected change in length for $|w_n| > 1$, i.e. $\lambda = (k-1)/k$. We will now show that as n becomes large, nearly all of the random walks have lengths concentrated in the interval $[n(\lambda - \epsilon), n(\lambda + \epsilon)]$.

Lemma 4.1. *Let w_n be the nearest neighbor random of length n on the Cayley graph of a free group of rank k with respect to the standard generating set. Then for any $\epsilon > 0$ there is a constant $C > 0$ such that*

$$\mathbf{P}(\left|\tfrac{1}{n}\,|w_n| - \lambda\right| \geq \epsilon) \leq e^{-Cn}.$$

This will follow from the Bernstein inequality for exponentially mixing random variables. We say that the sequence of random variables X_n is *exponentially strongly mixing* if the correlations of events occurring at least time n apart decays exponentially in n. More precisely, let X_n be a sequence of real-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. A cylindrical set is an element of $\mathcal{F}$ corresponding to $X_k^{-1}(U)$ for some measurable set U. Let A_k be an element of $\mathcal{F}$ generated by the cylindrical sets for $X_1, \ldots X_k$, and let B_k be an element of $\mathcal{F}$ generated by the cylindrical sets $X_k, X_{k+1}, \ldots$. Define

$$\alpha(n) = \sup_{k, A_k, B_k} \left|\mathbf{P}(A_k \cap B_{k+n}) - \mathbf{P}(A_k)\mathbf{P}(B_{k+n})\right|.$$

Then we say the sequence X_n is exponentially strongly mixing if there is a constant $C > 0$ such that $\alpha(n) \leq e^{-Cn}$ for all n.

We will use the following Bernstein inequality for sums of random variables that satisfy an exponential mixing condition, due to Merlevède, Peligrad and Rio [34]. We state a simplified version of their result which suffices for our purposes.

Theorem 4.2. [34] *Let X_n be a sequence of bounded real-valued random variables, which are exponentially mixing, and such that $\mathbf{E}(X_i) = 0$ for each i. Let $S_n = X_1 + \cdots + X_n$. Then for any $\epsilon > 0$, there is a constant $C > 0$, such that*

$$\mathbf{P}(|S_n| \geq \epsilon n) \leq e^{-Cn}.$$

We now prove Lemma 4.1.

Proof. We may construct the word length Markov chain by taking a sequence of independent identically distributed Bernoulli variables X_n with values in $\{\pm 1\}$, and $\mathbf{P}(X_n = -1) = 1/2k$, and $\mathbf{P}(X_n = +1) = (2k-1)/2k$. Define a new sequence of random variables by

$$Y_n = \begin{cases} X_n & \text{if } Y_1 + \cdots + Y_{n-1} > 0 \\ +1 & \text{if } Y_1 + \cdots + Y_{n-1} = 0. \end{cases}$$

Then the Y_n's are the increments of the Markov chain, i.e. $Y_n = |w_n| - |w_{n-1}|$.

Let Z_n be the random variable which is the indicator function for $|w_n| = 0$, i.e.

$$Z_n = \begin{cases} +1 & \text{if } |w_n| = 0 \\ 0 & \text{if } |w_n| > 0. \end{cases}$$

As $X_n \leq Y_n \leq X_n + 2Z_n$, this implies

$$\sum_{i=1}^{n} X_i \leq |w_n| \leq \sum_{i=1}^{n} X_i + 2Z_i.$$

The Markov chain is transient, and in particular, using Theorem 3.9 for example, one may show

$$\mathbf{P}(|w_n| = 0) \leq Ce^{-\lambda^2 n},$$

where $C = 2\sqrt{2}k^{5/2}/(3k-1)$. Therefore, $\sum_{i=n}^{\infty} \mathbf{E}(Z_i)$ is finite, and as the expected values of the sum of n Bernoulli random variables is λn, this implies

$$\lambda n \leq \mathbf{E}(|w_n|) \leq \lambda n + K,$$

for some fixed K, and so in particular $\frac{1}{n}\mathbf{E}(|w_n|) \to \lambda$ as $n \to \infty$.

As the X_i are independent, we may apply the Bernstein inequality for independent variables, obtaining:

$$\mathbf{P}(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \lambda\right| \geq \epsilon) \leq e^{-Cn}.$$

As $\sum_{i=1}^n X_i$ is a lower bound for $|w_n|$, this implies that $\mathbf{P}(\sum_{i=1}^n X_i \leq t) \geq \mathbf{P}(|w_n| \leq t)$ for any t. Therefore $\mathbf{P}(\frac{1}{n}|w_n| - \lambda \leq -\epsilon) \leq e^{-Cn}$, giving one half of the desired inequality.

To apply the Bernstein inequality to $\sum_{i=1}^n X_i + 2Z_i$, we need to show that the X_i and the Z_i are exponentially strongly mixing. As $\mathcal{F}$ is generated by cylindrical sets, it suffice to consider

$$|\mathbf{P}(X_k \in A, Z_{k+n} \in B) - \mathbf{P}(X_k \in A)\mathbf{P}(Z_{n+k} \in B)|$$

and

$$|\mathbf{P}(Z_k \in A, Z_{k+n} \in B) - \mathbf{P}(Z_k \in A)\mathbf{P}(Z_{n+k} \in B)|,$$

where A and B are subsets of either $\{\pm 1\}$ or $\{0, 1\}$ as appropriate. In the case of Z_k and X_{k+n}, the random variable Z_k only depends on $X_1, \ldots X_k$, so the random variables are independent and the correlations are zero. However, by enumerating the cases, one can see that the correlation terms above are all bounded above by a constant multiple of $\mathbf{P}(Z_{k+n} = 1)$, which decays exponentially in n, as required.

So by the Bernstein inequality for weakly dependent random variables (i.e. Theorem 4.2) we obtain:

$$\mathbf{P}(\left|\sum_{i=1}^{n}(X_i + Z_i) - \lambda n - \sum_{i=1}^{n}\mathbf{E}(Z_i))\right| \leq \epsilon n) \leq e^{-Cn},$$

for a different constant C. As $\sum_{i=1}^n (X_i + Z_i)$ is an upper bound for $|w_n|$, this implies that $\mathbf{P}(\sum_{i=1}^n (X_i + Z_i) \geq t) \geq \mathbf{P}(|w_n| \geq t)$. Therefore, as there is an upper bound K for $\sum_{i=1}^n \mathbf{E}(Z_i)$,

$$\mathbf{P}(\tfrac{1}{n}|w_n| - \lambda - \tfrac{1}{n}K \geq \epsilon) \leq e^{-Cn}.$$

This gives the other half of the inequality, possibly for a different choice of ϵ and C. □

The length estimates from Lemma 4.1, together with Theorem 3.13 immediately give the following growth rates for the scl of random walks on free groups.

Theorem 4.3. *Let w_n be the nearest neighbor random walk of length n on the Cayley graph of a free group F of finite rank with respect to a symmetric free generating set. Then for any positive C_1 there are positive constants C_2, C_3 such that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(w_n) \leq C_3 n/\log n \mid w_n \in F') \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hw_n) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon}) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.2.

We remark that our argument here will give different constants from the constants in Theorem 3.13, as in particular the constants in Theorem 4.3 will depend on the constants in Theorem 3.13, and also the rate of escape λ, and the values of ϵ and C in Lemma 4.1.

4.2. **Hyperbolic groups.** We now generalize our results to hyperbolic groups. Let G be a group and S a finite symmetric generating set. By monotonicity of scl, there is an upper bound of order $n/\log n$ for the value of scl in G on symmetric random walk in S. We now show that when G is hyperbolic, there is also a *lower* bound of order $n/\log n$.

Before proceeding, we give a brief overview of the argument. Let w_n be the location of the random walk at time n, and let γ_n be a geodesic from 1 to w_n. Blachére and Brofferio [4] define a natural metric on G which they call the *Green metric*. This metric has the following key properties (see Blachére, Haïssinksy and Mathieu [6]): firstly, the volume of the ball of radius r grows like e^r; and secondly, for any given geodesic ν starting at the origin, the probability that an initial segment of γ_n fellow travels ν for length r decays like e^{-r}. From this it readily follows that the probability that γ_n has any subsegment which (after translating the initial point to the origin) fellow travels a given geodesic ν for length r decays like ne^{-r}. If the geodesic γ_n has two disjoint subsegments ℓ and ℓ' of length r such that ℓ fellow travels with the reverse of ℓ', then we may apply this estimate to the smallest subsegment containing both ℓ and ℓ'. If the central segment has length s, then the probability of getting such a match decays as ne^{-2r-s}, but the number of words of length s grows as e^s, and the number of choices for ℓ grows as e^r. Therefore, the probability that the geodesic γ_n has segments ℓ and ℓ' of length r, such that ℓ fellow travels with the reverse of ℓ' decays as n^2e^{-r}. This gives an estimate for $\sigma(w_n)$, which enables us to deduce the lower bound for the growth rate of scl.

In the rest of this section we first review some basic properties of hyperbolic groups, then discuss the properties of random walks and the Green metric which we will need, before finally obtaining the estimates described above.

We briefly review some well-known properties of δ-hyperbolic spaces. It will be convenient to consider different, but quasi-isometric, metrics on the δ-hyperbolic group, and if necessary we will distinguish them with superscripts. We will write ∂G for the Gromov boundary of G, and we will write $\overline{G}$ for $G \cup \partial G$. The Gromov boundary only depends on the quasi-isometry class of the metric, and all the results below hold for any hyperbolic metric space, though the constants K_i in the results may differ, as the constants all depend on the constant of hyperbolicity, which depends on the metric.

Let (G, d) be a δ-hyperbolic space. Given a point $z \in G$, the Gromov product based at z is defined to be

$$(x \cdot y)_z = \tfrac{1}{2}(d(z,x) + d(z,y) - d(x,y)),$$

and this is equal to the distance from z to a geodesic from x to y, up to bounded error K_1, which only depends on δ. We may extend this definition to points on the boundary by

$$(x \cdot y)_z = \sup \liminf_{i,j \to \infty} (x_i \cdot y_i)_z$$

where the supremum is taken over all sequences $x_i \to x$ and $y_j \to y$. This supremum is finite unless x and y are the same point in ∂G.

Let R be a positive real number, and x a point in G. We define the *shadow* of x based at z with parameter R, denoted $S_z(x,R)$ to be

$$S_z(x,R) = \{y \in \overline{G} \mid (x \cdot y)_z \geq d(z,x) - R\}.$$

We warn the reader that this definition differs from that of other authors, for example [6], who define the shadow to be $S_1(x,R) \cap \partial G$, in our notation.

We will also use the fact that nearest point projection onto a geodesic γ is coarsely well defined, i.e. there is a constant K_2, which only depends on δ, such that if p and q are nearest points on γ to x, then $d(p,q) \leq K_2$.

The distances between any finite collection of points in a Gromov hyperbolic space may be realized by an embedded tree in the space, up to an additive error depending on the number of points (see e.g. [23]). We will write $d^T(x,y)$ for the distance between two points x and y in a tree T, with respect to the metric in T.

Lemma 4.4 (Approximate tree)**.** *Let $x_1, \ldots, x_n$ be a finite set of points in a δ-hyperbolic space (G,d). There is a constant K_3, which depends only on δ, and a geodesic tree T in X containing the x_i, such that*

$$d(x,y) - K_3 n \leq d^T(x,y) \leq d(x,y)$$

for all points x and y in T. We call T the approximate tree *determined by the x_i.*

Let $(x \cdot y)_z^T$ denote the Gromov product in a tree T with metric d^T. Suppose x, y and z all lie in an approximate tree determined by at most n points in G. From Lemma 4.4 we obtain the following inequality:

$$(x \cdot y)_z - 2K_3 n \leq (x \cdot y)_z^T \leq (x \cdot y)_z + K_3 n,$$

Furthermore, in the tree T, there is a unique point t such that $d^T(z,t) = (x \cdot y)_z$, namely the unique point which lies in the intersection of the three geodesics connecting the points x, y and z. We shall call this point the *center* of x, y and z (it depends on the approximate tree T, but its location is well-defined up to an error depending only on δ and n).

The next three lemmas are elementary exercises in δ-hyperbolic geometry, but we give the details for the sake of completeness. We first show that we may produce shadow sets that are strictly nested.

Lemma 4.5. *There is a constant K, which depends only on δ, such that for all positive constants A and R, and any $x, z \in G$ with $d(x,z) \geq A + R + 2K$, the shadow $S_z(x,R)$ is disjoint from the complement of the shadow $S_z(x, R+A+K)$. Furthermore for any pair of points $a, b \in G$ such that $a \in S_z(x,R)$ and $b \in G \setminus S_z(x, R+A+K)$, the distance between a and b is at least A.*

Proof. Let $a, b \in G$ such that $a \in S_z(x,R)$ and $b \in G \setminus S_z(x, R+A+K)$. Let T be an approximate tree containing z, x, a and b. Let p be the center in T of the triple x, z, a, and let q be the center in T of the triple x, z, b, so both p and q lie on the geodesic from z to x in T. We shall choose $K = 37K_3$, where K_3 is the approximate tree constant.

As $a \in S_z(x,R)$, the Gromov product $(a \cdot x)_z$ is at least $d(z,x) - R$, and so the distance in T from x to p is at most $R + 12K_3$. Similarly, as $b \notin S_z(x, R+A+K)$, the Gromov product $(b \cdot x)_z$ is at most $d(z,x) - R - A - K$, and as $d(z,x) \geq R + A + 2K$, the distance from x to q is at least $R + A + K - 12K_3$. Therefore the distance in T between p and q is at least $A + K - 24K_3$. As the geodesic in T between a and b

passes through p and q, the distance between a and b in T is at least $A+K-24K_3$. This implies that the distance in G between a and b is at least $A+K-36\delta$. As a and b were arbitrary points in $S_z(x,R)$ and $G \setminus S_z(x, R+A+K)$, for the choice $K = 37K_3$ the distance in G between the two sets is at least A, as claimed.

Finally, we observe that as a geodesic from a to b in T passes through p, the Gromov product $(a \cdot b)_z$ is bounded for all $a \in S_z(x,R)$ and $b \in G \setminus S_z(x, R+A+K)$, and so the boundary points of the two sets are disjoint. □

We now show that the shadow of x based at z with parameter R is roughly the complement of the shadow of z based at x with parameter $d(z,x) - R$.

Lemma 4.6. *There is a constant K, which only depends on δ, such that for all constants $R \geq 2K$, and all $x, z \in G$ with $d(x,z) \geq R + K$,*

$$S_x(z, d(x,z) - R - K) \subset G \setminus S_z(x,R) \subset S_x(z, d(x,z) - R + K).$$

Proof. We shall choose K to be $25K_3$, where K_3 is the approximate tree constant.

Let $y \in S_x(z, d(x,z) - R - K)$, and let T be an approximating tree for the points x, y and z. Let p be the center in T for the triple x, y, z. As $y \in S_x(z, d(x,z) - R - K)$ the Gromov product $(y \cdot z)_x$ is at least $R + K$, and so $d(x,p) \geq R + K - 12K_3$. Therefore the Gromov product $(x \cdot y)_z$ is at most $d(x,z) - R - K + 24K_3$, and as $K > 24K_3$, this implies that $y \in G \setminus S_z(x,R)$, which gives the left hand inclusion.

Now suppose that $y \in G \setminus S_z(x,R)$, and let T be an approximating tree for the points x, y and z. Let p be the center in T for the triple x, y, z. As $y \notin S_z(x,R)$ the Gromov product $(x \cdot y)_z$ is at most $d(x,z) - R$, and so $d(x,p) \geq R - 12K_3$. Therefore the Gromov product $(y \cdot z)_x$ is at least $R - 24K_3$, and as $K = 25K_3$, this implies that $y \in S_x(z, d(x,z) - R - K)$, which gives the right hand inclusion. □

Finally, we show that under certain conditions we may "change the basepoint" of a shadow.

Lemma 4.7. *There are constants K_4 and K_5, which only depend on δ, such that for any three points $x, y, z \in G$ with $(x \cdot y)_z \leq R + K_4$, there is an inclusion of shadows:*

$$S_z(x,R) \subset S_y(x, R + K_5).$$

Proof. Let t be a point in $S_z(x,R)$, and let T be an approximating tree containing the points x, y, z and t. Let p be the center in T of x, y, z, and let q be the center in T of x, z, t, so both p and q lie on the geodesic in T from z to x.

As $t \in S_z(x,R)$, the Gromov product $(x \cdot t)_z$ is at least $d(x,z) - R$, so the center q of the triple x, z, t is distance at most $R + 12K_3$ from x, where K_3 is the approximate tree constant. If we choose $K_4 = 25K_3$, then the distance from p to x in T is greater than the distance from q to x in T, so in particular p and q are distinct, and the path from z to x passes first through p and then through q. Therefore, the Gromov product $(x \cdot t)_y$ is at least $d(y,q) - 37K_3$, and so $t \in S_y(x, R + K_5)$, where $K_5 = 49K_3$, which depends only on δ, as required. □

We now wish to use these properties of hyperbolic spaces to study random walks, and we now review some useful properties of random walks on hyperbolic groups. Let μ be a probability distribution on a hyperbolic group G, and consider the probability space $(G, \mu)^{\mathbb{Z}_+}$ whose points correspond to sequences of independent random variables chosen from G according to the distribution μ. The distribution of random walks of length n is given by the n-fold convolution of μ with itself,

which we shall denote μ_n. We shall call $(G,\mu)^{\mathbb{Z}_+}$ the *step space* for the random walk. Each sequence of steps gives rise to a path in G, where the location w_n at time n is give by $w_n = s_1 s_2 \dots s_n$. The map $\rho\colon \{s_i\} \mapsto \{w_i\}$ takes a step sequence to a location sequence, also known as a *trajectory*, and we shall call the collection of trajectories the *path space* $(G^{\mathbb{Z}_+}, \nu)$, where the measure ν is $\mu \circ \rho^{-1}$. We shall write ν_g for the harmonic measure starting from g instead of 1, so $\nu_g(X) = \nu(g^{-1}X)$. We will make essential use of the following fundamental result of Kaimanovich.

Theorem 4.8 (Kaimanovich [30], Thms. 7.6 and 7.7)**.** *Let μ be a probability measure with finite first moment on a non-elementary hyperbolic group G, such that the support of μ generates G. Then almost every path $w_n(\omega)$ of the random walk generated by μ converges to a point $\partial(\omega)$ in the Gromov boundary ∂G, and furthermore ∂G with the hitting measure is isomorphic to the Poisson boundary of the random walk.*

As the map from $(G^{\mathbb{Z}_+}, \nu)$ to the boundary is a measure isomorphism, we shall denote the harmonic measure on ∂G by ν as well. The harmonic measure ν is the weak limit of the convolution measures μ_n, and is the unique μ-stationary measure on ∂G.

Blachère, Haïssinsky and Mathieu [6] show that the harmonic measure ν is a quasiconformal metric for the *Green metric* on G, which we now define. Let $F(x,y)$ be the probability that a random walk starting at x hits y. Up to a constant factor, this is equal to the Green function

$$G(x,y) = \sum_{n=0}^{\infty} \mu_n(x^{-1}y).$$

Blachère and Brofferio [4] defined the Green metric to be

$$d^\mu(x,y) = -\log F(x,y),$$

and showed that if μ has finite support then (G, d^μ) is a δ-hyperbolic space, quasi-isometric to G with the word metric (see also Blachère, Haïssinsky and Mathieu [5]). We will write $(x \cdot y)^\mu_z$ for the Gromov product with respect to the metric d^μ, and similarly we will write $S^\mu_z(g,R)$ for the shadow with respect to the metric d^μ. We will use the Green metric on G for the rest of this section, and as the constant of hyperbolicity δ^μ for the Green metric depends on μ, we will just write that some constant depends on μ to include the case in which it depends on δ^μ and μ.

Blachère, Haïssinsky and Mathieu estimate the harmonic measures of shadows.

Theorem 4.9 (Blachère-Haïssinsky-Mathieu [6] Thm. 2.3 and Lem. 2.4)**.** *There is a constant R_0, which only depends on μ, such that for all $R > R_0$ there is a constant C, which depends on R and μ, such that*

$$\tfrac{1}{C} e^{-d^\mu(1,g)} \leq \nu(S^\mu_1(g,R)) \leq C e^{-d^\mu(1,g)}.$$

Note that as the harmonic measure ν is supported on ∂G, our alternative definition of shadows makes no difference in the corollary above. In fact, Blachère, Haïssinsky and Mathieu [6] show that the harmonic measure ν is a quasiconformal (in fact conformal) measure on the boundary ∂G, with respect to the visual measure defined by d^μ on ∂G, and so ν is Ahlfors regular. Furthermore, they show that with respect to the Green metric, the rate of escape is equal to the asymptotic entropy h, and the logarithm of the volume growth rate is equal to 1. Recall that

Coornaert showed that the number of group elements of length n is equal to λ^n up to a multiplicative constant, where λ is the exponential growth rate of the number of elements in the ball of radius n. As the growth rate in the Green metric is e, this implies that there are constants C_1 and C_2 such that

$$C_1 e^n \leq |G_n^\mu| \leq C_2 e^n, \tag{4.2.1}$$

where G_n^μ is the number of group elements whose lengths in the Green metric lie in the interval $(n-1, n]$.

We now show an upper bound estimate for the convolution measures μ_n of shadows, which is independent of n.

Lemma 4.10. *There is a constant R_0, which only depends on μ, such that for any constant $R \geq R_0$, there is a constant C, which only depends on R and μ, such that for all $g \in G$, and for all positive integers n,*

$$\mu_n(S_1^\mu(g, R)) \leq C e^{-d^\mu(1,g)}.$$

Proof. The basic idea is that if you are in a shadow $S_1^\mu(g, R)$ at time n, then it is very likely that the limit point of the trajectory lies in a slightly larger neighborhood of the shadow. The estimate for the harmonic measure of a shadow then gives an upper bound for the time n measure μ_n. We now make this argument precise.

Fix a constant $M < 1$, and let R_0 be the constant from Theorem 4.9. Then there is a constant R_1 such that $\nu(S_1^\mu(g, R_0)) \leq M < 1$ for all g with $d^\mu(1, g) \geq R_1$. Using the nesting properties of shadow sets, Lemma 4.5, there is a constant K, which only depends on μ, such that for any g with $d^\mu(1, g) \geq K + R_1$, then $d(S_1^\mu(g, R), G \setminus S_1^\mu(g, R + K)) \geq R_1$. In particular, for any $h \in S_1^\mu(g, R)$, there is a shadow complement $G \setminus S_1^\mu(g, R + K)$ which is contained in $S_h^\mu(1, K)$, and so $\nu_h(G \setminus S_h^\mu(1, K)) \leq M < 1$.

As ν is μ_n-stationary,

$$\nu(S_1^\mu(g, R + K)) = \sum_{h \in G} \mu_n(x) \nu_h(S_1^\mu(g, R + K)).$$

We may restrict the sum to elements of h in $S_1^\mu(g, R)$, so

$$\nu(S_1^\mu(g, R + K)) \geq \sum_{h \in S_1^\mu(g,R)} \mu_n(x)(1 - M).$$

Therefore $\mu_n(S_1^\mu(g, R)) \leq \nu(S_1^\mu(g, R+K))/(1-M)$, where K is the constant above, which only depends on μ. Finally, the harmonic measure estimate from Theorem 4.9, implies that $\mu_n(S_1^\mu(g, R)) \leq C e^{-d^\mu(1,g)}$, for a constant C which only depends on R and μ, as required. □

We would like to know whether a geodesic from 1 to w_n fellow travels with a particular geodesic word. Given a particular geodesic word, choose a translate a of the word such that 1 is the midpoint, and label the endpoints a_+ and a_-. A geodesic γ fellow-travels with a if there is a subgeodesic of γ which is Hausdorff distance at most $2\delta^\mu$ from a. In this case there is a constant R_2, which only depends on δ^μ, such that one endpoint of γ lies in $S_1^\mu(a_+, R_2)$, and the other endpoint lies in $S_1^\mu(1, a_-, R_2)$. Without loss of generality, we shall assume that the constant R_2 is larger than the constant R_0 from Theorem 4.9. We say a translate ga of a is *perpendicular* to 1 if the closest point on ga to 1 is within distance R of the midpoint g, where R is the diameter of the support of μ. Perpendicular translates

have the property that the path consisting to a geodesic from 1 to g, followed by a geodesic from g to ga_+ is almost a geodesic. To be precise, there is a constant K, which only depends on μ, as R also depends on μ, such that

$$d^\mu(1,g) + d^\mu(g, ga_+) - K \leq d^\mu(1, ga_+) \leq d^\mu(1,g) + d^\mu(g, ga_+), \tag{4.2.2}$$

and the same statement holds with a_+ replace by a_-. The right hand inequality is the triangle inequality, and the left hand inequality follows from choosing an approximate tree containing $1, g, ga_+$, and the closest point on ga to 1. Let P_a be the set of all perpendicular translates. The *matching set* for a is the following subset of $\overline{G} \times \overline{G}$,

$$M_a = \bigcup_{g \in P_a} gS_1^\mu(a_+, R_2) \times gS_1^\mu(a_-, R_2),$$

The matching set M_a detects fellow-traveling with a, as we now describe. At time k, let p_k be the closest point to w_k on a geodesic $[1, w_n]$. If the point p_k is within distance R of the midpoint of a segment of $[1, w_n]$ which fellow travels with a translate ga, then $(w_k^{-1}w_n, w_k^{-1}) \in gS_1^\mu(a_+, R_2) \times gS_1^\mu(a_-, R_2) \subset M_a$. The location w_k at time k only depends on the steps $s_1, \dots s_k$, and $w_k^{-1}w_n$ only depends on $s_{k+1}, \dots s_n$, so the pair $(w_k^{-1}w_n, w_k^{-1})$ is distributed as the product $\mu_{n-k} \times \mu_k$. Therefore the probability that a geodesic from 1 to w_n fellow travels with a at time k is bounded above by $\mu_{n-k} \times \mu_n(M_a)$. We now show that this decays exponentially in the length of a.

Lemma 4.11. *There are constants C and K, which only depend on μ, such that for all k and l, and all elements $a \in G$ with $d^\mu(1,a) \geq K$,*

$$\mu_k \times \mu_l(M_a) \leq Ce^{-d^\mu(1,a)}.$$

Proof. By the definition of the matching set M_a,

$$\mu_k \times \mu_l(M_a) \leq \sum_{g \in P_a} \mu_k(gS_1^\mu(a_+, R_2))\mu_l(gS_1^\mu(a_-, R_2))$$

The translate of a shadow $gS_1^\mu(a_+, R_2)$ is the shadow $S_g^\mu(ga, R_2)$. Using Lemma 4.7, this shadow is contained in a shadow with a different basepoint, i.e. $S_g^\mu(ga, R_2) \subset S_1^\mu(ga_+, R_2 + K)$, for some constant K which only depends on μ. We now use the convolution measure estimate from Lemma 4.10, to obtain

$$\mu_n(gS_1^\mu(a_+, R_2)) \leq C_3 e^{-d^\mu(1, ga_+)},$$

for all n, where C_3 only depends on μ, as R_2 only depends on μ. A similar argument shows that $gS_1^\mu(a_-, R_2) \subset S_1^\mu(ga_-, R_2 + K)$, and therefore

$$\mu_n(gS_1^\mu(a_-, R_2)) \leq C_3 e^{-d^\mu(1, ga_-)},$$

and again this holds for all n, and C_3 only depends on μ, and in fact we may choose it be the same as the preceding constant. Combining these two estimates, and using the estimate (4.2.2), and the fact that $d^\mu(1, a_-) + d^\mu(1, a_+) = d^\mu(1, a)$, we obtain

$$\mu_k(gS_1^\mu(a_+, R_2))\mu_l(gS_1^\mu(a_-, R_2)) \leq C_4 e^{-d^\mu(1,a) - 2d^\mu(1,g)},$$

where again C_4 only depends on μ. We may sum over all words by first summing over all words g in G_n^μ, and then summing over all n.

$$\mu_k \times \mu_l(M_a) \leq \sum_{n=0}^{\infty} \sum_{g \in G_n^\mu} C_4 e^{-d^\mu(1,a) - 2n},$$

As $|G_n^\mu| \leq C_2 e^n$, where C_2 is the constant from line (4.2.1), this implies

$$\mu_k \times \mu_l(M_a) \leq C_5 e^{-d^\mu(1,a)} \sum_{n=0}^{\infty} e^{-n},$$

where C_5 only depends on μ. The sum is finite, and so

$$\mu_k \times \mu_l(M_a) \leq C_6 e^{-d^\mu(1,a)},$$

where C_6 only depends on μ, as required. □

We now show that the probability that a random word w_k has a pair of matching segments of length m decays exponentially in m.

Lemma 4.12. *There is a constant C, which only depends on μ, such that*

$$\mathbf{P}(\sigma(w_n) = m) \leq Cn^2 e^{-m}.$$

Proof. Let w_n be the position of a trajectory of the random walk at time n, and let γ be a geodesic from 1 to w_n. Let R be the range of the random walk, i.e. the diameter of the support of μ, so the length of γ is at most Rn. For some $l \geq 0$, let γ' be a subsegment of γ, divided in to the union of three segments, γ_1, γ_2 and γ_3 , such that γ_1 and γ_3 have length m, and γ_2 has length l. Let p_k be the nearest point projection of w_k to γ. By Lemma 4.11, the probability that γ' fellow travels with a particular word v of length $l + 2m$, with p_k distance at most R from the midpoint of the segment, is at most Ke^{-l-2m}, where K only depends on μ. Therefore, the probability that γ' fellow travels with a particular word v of length $l + 2m$, within distance R of any p_k, is at most Kne^{-l-2m}. If v has an initial segment of length m that fellow travels with a, and a final segment of length m which fellow travels with a^{-1}, then the length l of the central segment is at most Rn, and for a given l, by (4.2.1), there are at most $C_2 e^l$ choices for the word corresponding to the central segment γ_2. Therefore, for a fixed word a of length m, the probability that γ has disjoint subsegments γ_1 and γ_3 which fellow travel with a and a^{-1} is at most $KC_2 Rn^2 e^{-2m}$. Again, by 4.2.1, there are at most $C_2 e^m$ words of length m, so the probability that γ has segments γ_1 and γ_3 of length m, such that γ_1 fellow travels with the inverse of γ_3, is at most $KC_2^2 Rn^2 e^{-m}$, where K, C_2 and R only depend on μ, as required. □

In particular, for any positive C_1, there is a positive constant C_2 such that

$$\mathbf{P}(\sigma(w_n) \geq C_2 \log n) \leq Kn^{-C_1},$$

and so this gives the lower bound on scl (i.e. the analogs of Lemma 3.6 and Lemma 3.32) in the context of random walks on hyperbolic groups. The upper bound follows from the fact that random walks on G are homomorphic images of random walks on the free group in the generators of G, scl can only go down under homomorphisms and the correction h grows by at most a constant factor. So we have shown:

Theorem 4.13. *Let w_n be the nearest neighbor random walk of length n on the Cayley graph of a hyperbolic group G with respect to a symmetric generating set. Then for any positive C_1 there are positive constants C_2, C_3 such that*

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(w_n) \leq C_3 n/\log n \mid w_n \in G') \geq 1 - n^{-C_1}$$

Moreover,

$$\mathbf{P}(C_2 n/\log n \leq \mathrm{scl}(hw_n) \leq C_3 n/\log n \wedge |h| < n^{1/2+\epsilon}) \geq 1 - n^{-C_1}$$

where hg is any correction to g provided by Lemma 3.20.

4.3. **Random walk in** $\mathrm{Homeo}^+(S^1)$**.** We abruptly shift our focus from geometry to dynamics. We are concerned now with general groups, and are interested in obtaining lower bounds on scl from quasimorphisms. In this section, we study the statistics of *rotation quasimorphisms* on semigroups of homeomorphisms of the circle.

A basic reference for the theory of groups acting on the circle is [22], [36], or [8], Ch. 2.

We think of S^1 as $\mathbb{R}$ mod $\mathbb{Z}$ in what follows. Let $h \in \mathrm{Homeo}^+(S^1)$. Under the covering projection $\mathbb{R} \to S^1$ the homeomorphism h lifts to an element $\widetilde{h} \in \mathrm{Homeo}^+(\mathbb{R})$, unique up to composition with an integer translation of $\mathbb{R}$ (with which it commutes). Poincaré's *rotation number* is the limit

$$\mathrm{rot}(\widetilde{h}) = \lim_{n\to\infty} \frac{\widetilde{h}^n(0)}{n}$$

Different choices of lift change the value of $\mathrm{rot}(\widetilde{h})$ by an integer, so $\mathrm{rot}(h) := \mathrm{rot}(\widetilde{h})$ (mod $\mathbb{Z}$) is well-defined, taking values in $\mathbb{R}/\mathbb{Z} = S^1$.

The real-valued function rot is in fact a homogeneous quasimorphism on the subgroup of $\mathrm{Homeo}^+(\mathbb{R})$ centralized by integer translations; i.e. on the central extension of $\mathrm{Homeo}^+(S^1)$ consisting of all lifts of all elements to $\mathrm{Homeo}^+(\mathbb{R})$. Given a subgroup G of $\mathrm{Homeo}^+(S^1)$, let $\widetilde{G}$ denote the preimage in $\mathrm{Homeo}^+(\mathbb{R})$. Then rot is a homogeneous quasimorphism on $\widetilde{G}$. Moreover, by construction, for any $\widetilde{h} \in \widetilde{G}$, there is an estimate

$$|\widetilde{h}(0) - \mathrm{rot}(\widetilde{h})| < 1$$

In the sequel we adopt the (nonstandard) notation $R(\widetilde{h}) := \widetilde{h}(0)$. Then R is an inhomogeneous quasimorphism on $\widetilde{G}$, whose homogenization is rot.

A finite set S of homeomorphisms of S^1 generates a subgroup G of $\mathrm{Homeo}^+(S^1)$, which we can think of as the image of an abstract free group F generated by the set S under a homomorphism $\varphi : F \to \mathrm{Homeo}^+(S^1)$. A choice of lift $\widetilde{s} \in \mathrm{Homeo}^+(\mathbb{R})$ for each $s \in S$ determines a lift $\widetilde{\varphi} : F \to \mathrm{Homeo}^+(\mathbb{R})$, and composition with the rotation quasimorphism defines a quasimorphism rot_φ on F, unique up to elements of $H^1(F)$ (which parameterize the choices of lift). Denote by $\widetilde{S}$ the union of the $\widetilde{s}$. The quasimorphism rot_φ factors to a quasimorphism on the central extension $\widetilde{G}$ generated by the lifts $\widetilde{s}$. We can think of a random word in the generators $\widetilde{S}$ as determining a random walk in F, or just as well as determining a random walk in $\widetilde{G}$.

For the sake of interesting geometric applications, we explicitly allow the case that S is *not* symmetric, in which case the support of the random walk will only be a *semigroup* in general.

The purpose of this subsection is to prove the following:

Theorem 4.14. *Let S be a finite subset of* $\mathrm{Homeo}^+(S^1)$*, and $\widetilde{S}$ a collection of lifts of elements of S to* $\mathrm{Homeo}^+(\mathbb{R})$*. Let $X := X_0, X_1, \cdots$ be a random process taking values in* $\mathrm{Homeo}^+(\mathbb{R})$*, where $X_0 = \mathrm{Id}$, and $X_{n+1} = X_n \widetilde{s}$ where $\widetilde{s}$ is chosen from*

$\widetilde{S}$ with the uniform distribution. Then $X_n(0)$ satisfies a central limit theorem; i.e. there are constants E, σ so that

$$n^{-1/2}(X_n(0) - nE) \xrightarrow{pr} N(0, \sigma)$$

where $N(0,\sigma)$ denotes the normal distribution with standard deviation σ.

Proof. Let F be the semigroup generated by S. Let μ be a stationary measure for the action; i.e. a probability measure on S^1 that is a fixed point for the convolution operator on probability measures:

$$D : \nu \to \frac{1}{|S|} \sum_s s_*(\nu)$$

Since the set of such invariant measures is a convex compact set, we can further insist that μ is an extremal point in this set, i.e. an ergodic stationary measure. By construction, the action of F on S^1 is absolutely continuous with respect to μ.

Since μ is ergodic, if it contains atoms it is finite, and the action is semi-conjugate to an action that factors through a finite cyclic group of rotations. In this case, rot is a homomorphism, and the central limit theorem holds for straightforward reasons. So without loss of generality (and for the sake of exposition) we will assume μ contains no atoms.

The property of being stationary means that for any measurable subset $I \subset S^1$ (in particular, for all intervals I), there is an equality

$$\frac{1}{|S|} \sum_{s \in S} \mu(s(I)) = \mu(I)$$

We pull back μ to a Radon measure on $\mathbb{R}$ invariant under integer translation. By abuse of notation, we denote this pulled-back measure also by μ, and observe that it satisfies $\mu([t, t+1]) = 1$ for all $t \in \mathbb{R}$. Moreover, since each $\widetilde{s}$ commutes with integer translation, the pulled-back measure μ is also stationary for $\widetilde{S}$; i.e. $1/|S| \sum_{\widetilde{s} \in \widetilde{S}} \mu(\widetilde{s}(I)) = \mu(I)$ for all measurable subsets $I \subset \mathbb{R}$.

Define $m : \mathbb{R} \to \mathbb{R}$ by $m(t) = \mu([0, t])$ if t is positive, and $m(t) = -\mu([t, 0])$ if t is negative. Evidently, $|m(t) - t| \leq 1$ for all $t \in \mathbb{R}$. For any $t \in \mathbb{R}$, define

$$f(t) = \left(\frac{1}{|S|} \sum m(\widetilde{s}(t)) \right) - m(t)$$

If $u \in \mathbb{R}$ is arbitrary, and I is the interval from t to u, then

$$f(u) = f(t) + \left(\frac{1}{|S|} \sum \mu(\widetilde{s}(I)) \right) - \mu(I) = f(t)$$

Hence $f(\cdot)$ is *constant*, and equal to some fixed E — the *drift* of X. Define a random sequence Y by $Y_n = m(X_n(0)) - nE$. Since $|m(t) - t| \leq 1$ for any $t \in \mathbb{R}$, to complete the proof it suffices to show that $n^{-1/2} Y_n$ converges in probability to a Gaussian. The key point is that Y is a (bounded) *martingale*: the boundedness is the trivial estimate $|Y_{n+1} - Y_n| \leq C_1$ for some uniform constant C_1 independent of n (and depending only on $\widetilde{S}$), and the martingale property is the equality $\mathbf{E}(Y_{n+1} \mid Y_0, Y_1, \cdots, Y_n) = Y_n$ for all n. See e.g. Hall–Heyde [27] for an introduction to the theory of martingales.

The particular version of the martingale Central Limit Theorem we use is (a special case of) Thm. 3.2. (p. 58) from [27]. If we write $V_n = \sum_{i=1}^{n} \mathbf{E}((Y_{n+1} - Y_n)^2 \mid Y_n)$ then it suffices to show that V_n/n converges in probability to σ^2, for some

constant σ (which will be the standard deviation in the limiting Gaussian). In our situation, the increments $(Y_n - Y_{n-1})^2$ depend continuously only on the value of $X_{n-1} \bmod \mathbb{Z}$. Since by hypothesis the measure μ on $\mathbb{R}/\mathbb{Z}$ is ergodic for the operator D, the random ergodic theorem (see [21], Thm. 3.1) implies such convergence in probability (even in L^1), thus completing the proof. □

Remark 4.15. Different choices of lifts $\widetilde{S}$ influence the value of the drift E in a straightforward manner. Two different lifts $\widetilde{s}$ and $\widetilde{s}'$ differ by translation by some integer m. If E and E' denote the associated drifts, then $E' - E = m/|S|$.

Remark 4.16. There are two natural contexts in which $E = 0$. Firstly, if $S = S^{-1}$ (i.e. symmetric random walk in a group), and $\widetilde{S}$ is chosen with this same property, then the random process X is invariant under taking inverses, and therefore $E = 0$. Secondly, if $\widetilde{S}$ is conjugated to itself (as a subset of $\mathrm{Homeo}^+(\mathbb{R})$) under some reflection $t \to 2c - t$, then $E = -E = 0$ by symmetry.

Bavard duality lets us compute scl in central extensions of certain subgroups of $\mathrm{Homeo}^+(S^1)$.

Corollary 4.17. *Let G be a finitely generated subgroup of $\mathrm{Homeo}^+(S^1)$, and let $\widetilde{G}$ be the central extension consisting of preimages of elements of G in $\mathrm{Homeo}^+(\mathbb{R})$. Suppose that* scl *vanishes identically on G. Then the value of* $\mathrm{scl}/\sqrt{n}$ *on a random word of length n in a finite symmetric generating set $\widetilde{S}$ satisfies*

$$\lim_{n\to\infty} \mathbf{P}(a < \mathrm{scl}_n/\xi\sqrt{n} < b) = \frac{2}{2\pi}\int_a^b \mathbf{1}_{[0,\infty)} e^{-t^2/2} dt$$

Proof. For any group G, there is a short exact sequence

$$0 \to H^1(G) \to Q(G) \to H^2_b(G) \to H^2(G)$$

Since scl vanishes identically on G, by Bavard duality (see [1] or [10], Thm. 2.70) we have $Q/H^1 = 0$. The coboundary of the rotation quasimorphism exists, however, as an element of $H^2_b(G)$ whose image in $H^2(G)$ is nontrivial (and equal to the *Euler class*). The group $\widetilde{G}$ is the central extension associated to the class in $H^2(G)$, so $Q(\widetilde{G})/H^1$ is one dimensional, and spanned by rot. The defect satisfies $D(\mathrm{rot}) = 1/|Z|$ where $|Z|$ is the order of the centralizer of a monotone equivalent action, and is finite (see e.g. [10], Rmk. 4.69). So $\mathrm{scl}(g) = |Z| \cdot |\mathrm{rot}(g)|/2$ on (the commutator subgroup of) $\widetilde{G}$, and the conclusion follows from Theorem 4.14. □

Example 4.18. Corollary 4.17 applies to many naturally occurring families of groups, including Hilbert modular groups $\mathrm{SL}(2, \mathcal{O}(n))$ where $\mathcal{O}(n)$ is the ring of integers in $\mathbb{Q}(\sqrt{n})$ for n square-free, $\mathrm{SL}(2, \mathbb{Z}[1/2])$, Thompson's circle group F and certain generalized Stein-Thompson groups, and many others. The fact that scl vanishes identically on these groups follows from the stronger property that they are *boundedly generated by commutators*.

For Hilbert modular groups, this is a consequence of a deep theorem of Carter–Keller–Paige, namely [40] Thm. 6.1 which says that if A is the ring of integers in a number field K containing infinitely many units, and T is an element of $\mathrm{SL}(2, A)$ which is not a scalar matrix, then $\mathrm{SL}(2, A)$ has a finite index normal subgroup which is boundedly generated by conjugates of T.

The case $\mathrm{SL}(2,\mathbb{Z}[1/2])$ is due to Liehl [33] who proves that the group is boundedly generated by elementary matrices (which are themselves products of commutators of bounded length).

The case of Thompson's group is due to Ghys–Sergiescu [24], and some generalizations are due to Zhuang [42]. For an introduction to Thompson's groups and their properties, see [15].

An application of Theorem 4.14 of independent interest is to the distribution of certain natural geometric quantities under certain random walks in the hyperbolic plane. Consider the following process. Let S be a finite set of elements in $\mathrm{PSL}(2,\mathbb{R})$. Define a random process p_i in $\mathbb{H}^2$ where p_0 is a basepoint, $p_1 = s_1(p_0)$, $p_2 = s_1 s_2(p_0)$ and so on, where successive s_i are chosen randomly from S (with some fixed probability distribution). For any three points a, b, c in $\mathbb{H}^2$, let $\Delta(a,b,c)$ denote the signed area of the hyperbolic triangle they bound. Then define the random sum $A_n = \sum_{i=1}^{n-1} \Delta(p_0, p_i, p_{i+1})$.

Corollary 4.19. *With notation as above, the A_n satisfy a central limit theorem.*

Proof. Let P be the (immersed) infinite polygonal path with successive vertices p_i, and for any n let P_n be the immersed polygon obtained by closing up the subpath of P from p_0 to p_n by adding the geodesic $p_n p_0$. At each vertex p_i, let α_i denote the external angle (i.e. the turning angle) of P (or P_n) at the vertex p_i, normalized to take values in $[-\pi, \pi)$. Note that each α_i depends only on s_i, and the α_i are bounded i.i.d. random variables. The Gauss–Bonnet theorem for immersed hyperbolic polygons says that the $\sum_{i=0}^{n} \alpha_i - \mathrm{area}(P_n)$ is equal to 2π times the winding number of ∂P_n. This winding number is (up to a constant) equal to the value of the rotation quasimorphism on the random product $\widetilde{s}_1 \widetilde{s}_2 \cdots \widetilde{s}_n$ for fixed lifts of the s_i to $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ determined implicitly by the normalization of the α_i. By definition $A_n = \mathrm{area}(P_n)$. So the ordinary central limit theorem for the sums of the i.i.d. variables α_i together with Theorem 4.14 proves a central limit theorem for A_n, as claimed. □

Remark 4.20. For certain subsets S of $\mathrm{PSL}(2,\mathbb{R})$, the polygonal paths P will be uniformly quasigeodesic. In this case, there is another quite different proof of a central limit theorem for A_n, based on the thermodynamic formalism; see e.g. [38]. As a very concrete and attractive example, take S to consist of two elements, one of which (call it L) is the composition of a rotation (centered at the origin) through angle α with a translation (with axis through the center) through distance ℓ, and the other (call it R) is the composition of a rotation (centered at the origin) through angle $-\alpha$ with the same translation. The path P has all edges of length ℓ, and at each vertex it turns left or right through angle α with equal probability. If α is small enough compared to ℓ, the path P is uniformly quasigeodesic. In this case, the successive terms $\Delta(p_0, p_n, p_{n+1})$ define a Hölder continuous function on the one-sided shift space on the alphabet $\{L, R\}$. The Hölder continuity amounts to the observation that for four points a, b, c, d in $\mathbb{H}^2$ with $d(a,b) = d(c,d) = \ell$, the difference of $\Delta(a,c,d)$ and $\Delta(b,c,d)$ is bounded by a constant that decays exponentially fast in the distance from b to c. The central limit theorem for Hölder functions on shift spaces gives a different proof in this case. This kind of argument is implicit in [37] and a related argument is pursued in [12, 28]. However, if S is symmetric (so that the semigroup it generates is a group), the paths P periodically

make long backtracks, so the function is *never* Hölder continuous on the shift space, and this argument does not apply.

Remark 4.21. It is interesting to study how the constants E and σ (the drift and the standard deviation) vary as a function of the set $\widetilde{S}$. As in Remark 4.20, fix an angle α, and for each ℓ let P_ℓ be a random polygonal path with edges of length ℓ, making a turn left or right through angle α at each vertex. For each n, close the polygon up after n steps and let $W_\ell(n)$ denote the winding number. By the left-right symmetry, the drift of W_ℓ is zero for all ℓ. But the standard deviation $\sigma(\ell)$ undergoes a phase transition: it is zero for $\ell \in [2\cosh^{-1}(1/\sin(\alpha/2)), \infty)$, and increases monotonically to α as ℓ decreases to 0. In this case $\sigma(\ell)$ is real analytic (as a function of ℓ) on $[0, 2\cosh^{-1}(1/\sin(\alpha/2)))$.

4.4. **Random value of quasimorphisms.** In fact, very shortly after proving Theorem 4.14, we learned that a completely general statement has independently been obtained for *symmetric* random walk by Björklund–Hartnick [3]. They prove the following theorem (in fact, their results hold in considerably greater generality):

Theorem 4.22 (Björklund–Hartnick [3], Thm. 1.1)**.** *Let G be a finitely generated group, and S a finite symmetric generating set. Let $f : G \to \mathbb{R}$ be a quasimorphism, and $X_n := s_n \cdots s_1$ an i.i.d. left-random walk on G in the generating set S. Then f satisfies a central limit theorem with respect to X_n. Moreover, if the homogenization of f is nonzero, then the central limit is non-degenerate.*

Corollary 4.23. *Let G be any finitely generated group, and S a finite symmetric generating set. Suppose $Q(G)/H^1(G)$ is finite dimensional but nonzero. Then the value of* scl$/\sqrt{n}$ *on a random walk of length n in S is distributed (for large n) like the supremum of the absolute value of finitely many Gaussians. In particular, for any ϵ there are positive constants a, b depending on ϵ such that $\mathbf{P}(a < \text{scl}/\sqrt{n} < b) \geq 1-\epsilon$.*

For symmetric generating sets, Theorem 4.14 is, of course, a special case of Theorem 4.22. Under very general conditions, Björklund–Hartnick show that a quasimorphism f on a group G with a probability measure μ (satisfying some conditions) has a *(bi-) harmonic representative*; i.e. there is a function f' that differs from f by a bounded amount, and with the property that f' is invariant under convolution with μ (the existence of such a harmonic representative was also proved in quite a different way by Burger–Monod [7]). It is this harmonic feature of f' that lets one prove a CLT using the martingale CLT, as in the proof of Theorem 4.14.

4.5. **Compression.** We introduce the following notation.

Definition 4.24. If S_n is a sequence of random variables, we write $S_n \sim f(n)$ if the variable $S_n/f(n)$ converges in probability to a probability distribution on $\mathbb{R}$ not concentrated at 0. Write $S_n \precsim f(n)$ if $S_n/f(n)$ converges in probability to some probability distribution.

We obtain the following theorem:

Theorem 4.25. *Let G be an arbitrary group, and H a finitely generated subgroup of G with finite symmetric generating set S. Suppose Q/H^1 is nonzero (so that* scl *does not vanish identically on G'). For any sufficiently large constant C, let* scl_n

denote the random variable equal to the value of scl *on a random walk of length contained in* $[n-C, n+C]$ *in* H*, conditioned to land in* H'*. Then*

$$\sqrt{n} \precsim \mathrm{scl}_n \precsim n/\log n.$$

The upper bound holds because scl can only go down under homomorphisms. There is a surjection from a finite rank free group onto H, and scl grows at a rate at most $\log n/n$ in finite rank free groups. The lower bound holds because either scl in G vanishes identically on H, or else there is a homogeneous quasimorphism ϕ on G which restricts to a nontrivial homogeneous quasimorphism on H. Therefore the central limit theorem for quasimorphisms shows that ϕ satisfies a a CLT on $H \cap G'$.

It seems reasonable to make the following conjecture:

Conjecture 4.26. *We have* $\sqrt{n} \sim \mathrm{scl}_n$ *if and only if* Q/H^1 *is finite dimensional.*

The "if" direction of this conjecture is Corollary 4.23. Moreover, we ask the following question:

Question 4.27. *Is there a finitely generated group* G *for which* scl_n *grows strictly faster than* $\sqrt{n}$ *and strictly slower than* $n/\log n$*?*

5. Acknowledgments

Danny Calegari was supported by NSF grants DMS 0707130 and DMS 1005246. Joseph Maher was supported by NSF grant DMS 0706764, and would like to thank the Hausdorff Research Institute for Mathematics for their hospitality while working on this paper. We would like to thank Michael Björklund, Sean Cleary, David Fisher, Alex Furman, Nikolai Makarov, Curt McMullen, Andrés Navas, Jay Rosen, Richard Sharp, Brian Simanek and Alden Walker for some useful conversations about this material.

References

[1] C. Bavard, *Longeur stable des commutateurs*, L'Enseign. Math. **37** (1991), 109–150
[2] M. Bestvina and K. Fujiwara, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. **6** (2002), 69–89
[3] M. Björklund and T. Hartnick, *On the random geometry of quasimorphisms*, Geom. Topol. to appear
[4] S. Blachère, S. Brofferio, *Internal diffusion limited aggregation on discrete groups having exponential growth*, Probab. Theory Related Fields **137** (2007), no. 3–4, 323–343.
[5] S. Blachère, P. Haïssinsky and P. Mathieu, *Asymptotic entropy and Green speed for random walks on countable groups*, Ann. Probab. **36** (2008), no. 3, 1134–1152.
[6] S. Blachère, P. Haïssinsky and P. Mathieu, *Harmonic measures versus quasiconformal measures for hyperbolic groups*, preprint, arXiv:0806.3915
[7] M. Burger and N. Monod, *Bounded cohomology of lattices in higher rank Lie groups*, J. Eur. Math. Soc. **1** (1999), no. 2, 199–235
[8] D. Calegari, *Foliations and the geometry of 3-manifolds*, Oxford Mathematical Monographs. Oxford University Press, Oxford, 2007
[9] D. Calegari, *Stable commutator length is rational in free groups*, Jour. AMS **22** (2009), no. 4, 941–961
[10] D. Calegari, *scl*, MSJ Memoirs, **20**. Mathematical Society of Japan, Tokyo, 2009
[11] D. Calegari and K. Fujiwara, *Stable commutator length in word hyperbolic groups*, Groups Geom. Dyn. **4** (2010), no. 1, 59–90
[12] D. Calegari and K. Fujiwara, *Combable functions, quasimorphisms and the central limit theorem*, Ergodic Theory Dynam. Systems, to appear

[13] D. Calegari and A. Walker, `scallop`, computer program available from the author's webpage, and from `computop.org`
[14] J. Cannon, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Ded. **16** (1984), no. 2, 123–148
[15] J. Cannon, W. Floyd and W. Parry, *Introductory notes on Richard Thompson's groups*, Enseign. Math. (2) **42** (1996), no. 3-4, 215–256
[16] M. Coornaert, *Mesures de Patterson-Sullivan sure le bord d'un espace hyperbolique au sens de Gromov*, Pac. J. Math. **159** (1993), no. 2, 241–270
[17] M. Coornaert and A. Papadopoulos, *Symbolic dynamics and hyperbolic groups*, Lecture Notes in Mathematics **1539**, Springer-Verlag 1993
[18] M. Culler, *Using surfaces to solve equations in free groups*, Topology **20** (1981), no. 2, 133–145
[19] D. Epstein, J. Cannon, D. Holt, S. Levy, M. Paterson and W. Thurston, *Word processing in groups*, Jones and Bartlett, Boston, MA, 1992
[20] B. Farb and H. Masur, *Superrigidity and mapping class groups*, Topology **37** (1998), no. 6, 1169–1176
[21] A. Furman, *Random walks on groups and random transformations*, Handbook of dynamical systems, Vol. 1a, 931–1014, North-Holland, Amsterdam, 2002
[22] É. Ghys, *Groups acting on the circle*, Enseign. Math. (2) **47** (2001), no. 3-4, 329–407
[23] É. Ghys and P. de la Harpe (éditeurs), *Sur les groupes hyperboliques d'aprés Mikhael Gromov*, Progress in Mathematics, **83**, Birkhäuser, 1990.
[24] É. Ghys and V. Sergiescu, *Sur un groupe remarquable de diffomorphismes du cercle*, Comment. Math. Helv. **62** (1987), no. 2, 185–239
[25] M. Gromov, *Volume and bounded cohomology*, IHES Publ. Math. **56** (1982), 5–99
[26] M. Gromov, *Hyperbolic groups*, Essays in group theory, MSRI Publ. **8**, 75–263, Springer, New York, 1987
[27] P. Hall and C. Heyde, *Martingale limit theory and its application*, Academic Press, New York, 1980
[28] M. Horsham and R. Sharp, *Lengths, quasi-morphisms and statistics for free groups*, Spectral analysis in geometry and number theory, 219–237, Contemp. Math., 484, Amer. Math. Soc., Providence, RI, 2009
[29] N. Kahale, *Large deviation bounds for Markov chains*, Comb. Prob. and Computing **6** (1997), 465–474
[30] V. Kaimanovich, *The Poisson formula for groups with hyperbolic properties*, Annals of Mathematics, **152** (2000), 659–692
[31] V. Kaimanovich and H. Masur, *The Poisson boundary of the mapping class group*, Invent. Math. **125** (1996), no. 2, 221–264
[32] S. Kakutani, *Random ergodic theorems and Markoff processes with a stable distribution*, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, University of California Press, Berkeley (1951), 247–261
[33] B. Liehl, *Beschränkte Wortlänge in* SL_2, Math. Z. **186** 1984), no. 4, 509–524
[34] F. Merlevède, M. Peligrad and E. Rio, *Bernstein inequality and moderate deviations under strong mixing conditions*, IMS Collections, High Dimensional Probability V: The Luminy Volume, Vol. 5 (2009) 273–292
[35] I. Mineyev, *Flows and joins of metric spaces*, Geom. Topol. **9** (2005), 403–482
[36] A. Navas, *Grupos de difeomorfismos del círculo*, Ensaios Matemáticos [Mathematical Surveys], **13**, Sociedade Brasileira de Matemática, Rio de Janeiro, 2007
[37] J.-C. Picaud, *Cohomologie bornée des surfaces et courants géodésiques*, Bull. Soc. Math. France **125** (1997), no. 1, 115–142
[38] D. Ruelle, *Thermodynamic formalism*, Addison-Wesley Publishing Co., Reading, Mass., 1978
[39] R. Sharp, *Local limit theorems for free groups*, Math. Ann. **321** (2001), 889–904
[40] D. Witte-Morris, *Bounded generation of* $\mathrm{SL}(n, A)$ *(after D. Carter, G. Keller, and E. Paige)*, New York J. Math. **13** (2007), 383–421
[41] W. Woess *Random walks on infinite graphs and groups*, Cambridge University Press
[42] D. Zhuang, *Irrational stable commutator length in finitely presented groups*, J. Mod. Dyn. **2** (2008), no. 3, 499–507

Department of Mathematics, Caltech, Pasadena CA, 91125
E-mail address: dannyc@its.caltech.edu

Department of Mathematics, CUNY College of Staten Island, Staten Island, NY 10314
E-mail address: joseph.maher@csi.cuny.edu