Annals of Mathematics **177** (2013), 969–1027
http://dx.doi.org/10.4007/annals.2013.177.3.4

The Dehn function of $\mathrm{SL}(n;\mathbb{Z})$

By Robert Young

Abstract

We prove that when $n \geq 5$, the Dehn function of $\mathrm{SL}(n;\mathbb{Z})$ is quadratic. The proof involves decomposing a disc in $\mathrm{SL}(n;\mathbb{R})/\mathrm{SO}(n)$ into triangles of varying sizes. By mapping these triangles into $\mathrm{SL}(n;\mathbb{Z})$ and replacing large elementary matrices by "shortcuts," we obtain words of a particular form, and we use combinatorial techniques to fill these loops.

1. Introduction

The Dehn function of a group is a geometric invariant that measures the difficulty of reducing a word that represents the identity to the trivial word. Likewise, the Dehn function of a space measures the difficulty of filling a closed curve in a space with a disc. If a group acts properly discontinuously, cocompactly, and by isometries on a space, then the Dehn functions of the group and space grow at the same rate. Thus, for example, since a curve in the plane can be filled by a disc of quadratic area, the Dehn function of $\mathbb{R}^2$ grows like n^2 and the Dehn function of $\mathbb{Z}^2$, which acts on the plane, also grows like n^2.

Dehn functions can grow very quickly. For example, if Sol_3 is the space consisting of the 3-dimensional solvable Lie group

$$\mathrm{Sol}_3 = \left\{ \begin{pmatrix} e^t & 0 & x \\ 0 & e^{-t} & y \\ 0 & 0 & 1 \end{pmatrix} \middle| x, y, t \in \mathbb{R} \right\},$$

with a left-invariant metric, then its Dehn function grows exponentially. One reason for this is that Sol_3 is isomorphic to a horosphere in the rank 2 symmetric space $H^2 \times H^2$, where H^2 is the hyperbolic plane. This space contains 2-dimensional flats that intersect Sol_3 in large loops. Since they are contained in flats, the loops have fillings in $H^2 \times H^2$ of quadratic area, but since these fillings go far from Sol_3, they are exponentially difficult to fill in Sol_3.

Subsets of symmetric spaces of rank at least 3 often have smaller Dehn functions. For example, $(H^2)^3$ has a horosphere isometric to the solvable group

$$\mathrm{Sol}_5 = \left\{ \begin{pmatrix} e^{t_1} & 0 & 0 & x \\ 0 & e^{t_2} & 0 & y \\ 0 & 0 & e^{t_3} & z \\ 0 & 0 & 0 & 1 \end{pmatrix} \middle| x, y, z, t_i \in \mathbb{R}, t_1 + t_2 + t_3 = 0 \right\}.$$

As before, there are flats in $(H^2)^3$ that intersect Sol_5, but since $(H^2)^3$ has rank 3, the intersections may be spheres instead of loops. Indeed, loops contained in unions of flats have fillings contained in unions of flats and Sol_5 has a quadratic Dehn function. This result was first stated by Gromov [Gro93, 5.A$_9$]; a proof of a more general case along the lines stated here was given by Druţu [Dru04].

This suggests that the filling invariants of subsets of symmetric spaces depend strongly on rank. Some of the main test cases for this idea are lattices acting on high-rank symmetric spaces. If a lattice acts on a symmetric space with noncompact quotient, one can remove an infinite union of horoballs from the space to obtain a space on which the lattice acts cocompactly. When the symmetric space has rank 2, removing these horoballs may create difficult-to-fill holes in flats, as in Sol_3, but when the rank is 3 or more, Gromov conjectured

Conjecture 1.1 ([Gro93, 5.D(5)(c)]). *If Γ is an irreducible lattice in a symmetric space with $\mathbb{R}$-rank at least 3, then Γ has a polynomial Dehn function.*

See [BEW] for a more general conjecture, which generalizes the Lubotzky-Mozes-Raghunathan theorem. A special case of this conjecture is the following conjecture of Thurston (see [Ger93]).

Conjecture 1.2. *When $p \geq 4$, $\mathrm{SL}(p;\mathbb{Z})$ has a quadratic Dehn function.*

In this paper, we will prove Thurston's conjecture when $p \geq 5$.

Theorem 1.3. *When $p \geq 5$, $\mathrm{SL}(p;\mathbb{Z})$ has a quadratic Dehn function.*

When p is small, the Dehn function of $\mathrm{SL}(p;\mathbb{Z})$ is known; when $p = 2$, the group $\mathrm{SL}(2;\mathbb{Z})$ is virtually free, and thus hyperbolic. As a consequence, its Dehn function is linear. When $p = 3$, Epstein and Thurston [ECH$^+$92, Ch. 10.4] proved that the Dehn function of $\mathrm{SL}(3;\mathbb{Z})$ grows exponentially; Leuzinger and Pittet generalized this result to any noncocompact lattice in a rank 2 symmetric space [LP96]. This exponential growth has applications to finiteness properties of arithmetic groups as well; Bux and Wortman [BW07] describe a way that the constructions in [ECH$^+$92] lead to a proof that $\mathrm{SL}(3;\mathbb{F}_q[t])$ is not finitely presented (this fact was first proved by Behr [Beh79]), then generalize to a large class of lattices in reductive groups over function fields. The previous best known bound for the Dehn function of $\mathrm{SL}(p;\mathbb{Z})$ when $p \geq 4$

was exponential; this result is due to Gromov, who sketched a proof that the Dehn function of Γ is bounded above by an exponential function [Gro93, 5.A$_7$]. A full proof of this fact was given by Leuzinger [Leu04a].

Notable progress toward Conjecture 1.1 was made by Druţu [Dru04] in the case that Γ is a lattice in G with $\mathbb{Q}$-rank 1. In this case, Γ acts cocompactly on a subset of G constructed by removing infinitely many disjoint horoballs. Druţu showed that if G has $\mathbb{R}$-rank 3 or greater, then the boundaries of these horoballs satisfy a quadratic filling inequality and that if Γ has $\mathbb{Q}$-rank 1, then it enjoys an "asymptotically quadratic" Dehn function; i.e., its Dehn function is bounded by $n^{2+\epsilon}$ for any $\epsilon > 0$. More recently, Bestvina, Eskin, and Wortman [BEW] have made progress toward a higher-dimensional generalization of Conjecture 1.1 by proving filling estimates for S-arithmetic lattices.

The basic idea of the proof of Theorem 1.3 (we will give a more detailed sketch in Section 3) is to use fillings of curves in the symmetric space $\mathrm{SL}(p;\mathbb{R})/\mathrm{SO}(p)$ as templates for fillings of words in $\mathrm{SL}(p;\mathbb{Z})$. Fillings that lie in the thick part of $\mathrm{SL}(p;\mathbb{R})/\mathrm{SO}(p)$ correspond directly to fillings in $\mathrm{SL}(p;\mathbb{Z})$, but in general, an optimal filling of a curve in the thick part may have to go deep into the cusp of $\mathrm{SL}(p;\mathbb{Z})\backslash\mathrm{SL}(p;\mathbb{R})/\mathrm{SO}(p)$. Regions of this cusp correspond to parabolic subgroups of $\mathrm{SL}(p;\mathbb{Z})$, so we develop geometric techniques to cut the filling into pieces that each lie in one such region. This reduces the problem of filling the original word to the problem of filling words in parabolic subgroups of Γ. This step is fairly general, and these geometric techniques may be applied to a variety of groups. We fill these words using combinatorial techniques, especially the fact that Γ contains many overlapping solvable subgroups. This step is specific to $\mathrm{SL}(p;\mathbb{Z})$ and is the step that fails in the case $p = 4$.

In Section 2, we define some of the notation and concepts that will be used in the rest of the paper. Readers who are already familiar with Dehn functions may wish to skip parts of this section, but note that Section 2.3 introduces much of the notation we will use to describe subgroups and elements of $\mathrm{SL}(p)$ and that Section 2.4 introduces the new notion of "templates" for fillings.

In Section 3, we give an outline of the proof of Theorem 1.3. This outline reduces Theorem 1.3 to a series of lemmas that decompose words in $\mathrm{SL}(p;\mathbb{Z})$ into words in smaller and smaller subgroups of $\mathrm{SL}(p;\mathbb{Z})$. In Section 5 we describe the main geometric technique: a method for decomposing words in $\mathrm{SL}(p;\mathbb{Z})$ into words in maximal parabolic subgroups. Then, in Sections 6 and 7, we describe a normal form for elements of $\mathrm{SL}(p;\mathbb{Z})$ and prove several combinatorial lemmas giving ways to manipulate this normal form. Finally, in Sections 8 and 9, we apply these techniques to prove the lemmas, and in Section 10, we ask some open questions.

Some of the ideas in this work were inspired by discussions at the American Institute of Mathematics workshop, "The Isoperimetric Inequality for $\mathrm{SL}(n;\mathbb{Z})$," and the author would like to thank the organizers, Nathan Broaddus, Tim Riley, and Kevin Wortman; and participants, especially Mladen Bestvina, Alex Eskin, Martin Kassabov, and Christophe Pittet. The author would also like to thank Tim Riley, Yves de Cornulier, Kevin Wortman, and Mladen Bestvina for many helpful conversations while the author was visiting Bristol University, Université de Rennes, and University of Utah. Parts of this paper were completed while the author was a visitor at the Institut des Hautes Études Scientifiques and a Courant Instructor at New York University.

2. **Preliminaries**

In this section, we will describe some of the concepts and notation we will use throughout this paper.

We use a variant of big-O notation throughout this paper; the notation

$$f(x, y, \dots) = O(g(x, y, \dots))$$

means that there is a $c > 0$ such that $|f(x, y, \dots)| \leq cg(x, y, \dots) + c$ for all values of the parameters. In most cases, c will also depend implicitly on p.

If $f : X \to Y$ is Lipschitz, we say that f is c-Lipschitz if $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$, and we let $\mathrm{Lip}(f)$ be the infimal c such that f is c-Lipschitz.

2.1. *Words and curves.* If G is a group with finite generating set S, we call a formal product of elements of S and their inverses a *word* in S. By abuse of notation, we will also call this a word in G and leave S implicit. We denote the empty word by ε. There is a natural evaluation map taking words in G to G, and we say that a word *represents* its corresponding group element. If $w = s_1^{\pm 1} \cdots s_n^{\pm 1}$, we say that w has length $\ell(w) = n$.

If G acts on a space X, words in G correspond to curves in X. Let X be a connected simplicial complex or riemannian manifold, and let G act on X by maps of simplicial complexes or by isometries, respectively. Let S be a finite generating set for G and for all $s \in S$, and let $\gamma_s : [0, 1] \to X$ be a curve connecting x_0 to sx_0. Let γ_s^{-1} be the same curve with the reverse parametrization. If $w = s_1^{\pm 1} \cdots s_n^{\pm 1}$ is a word in S that represents g, we can construct a curve γ_w in X by concatenating translates of the $\gamma_{s_i}^{\pm 1}$'s. The resulting curve connects x_0 and gx_0, and its length is bounded by the length of w:

$$\ell(\gamma_w) \leq \ell(w) \max_{s \in S} \ell(\gamma_s).$$

2.2. *Dehn functions and the Filling Theorem.* A full introduction to the Dehn function can be found in [Bri02]. We will just summarize some necessary results and notation here. If

$$G = \langle h_1, \dots, h_d \mid r_1, \dots, r_s \rangle$$

is a finitely presented group and w is a word representing the identity, there is a sequence of steps that reduces w to the empty word, where each step is a free reduction or insertion or the application of a relator. We call the number of applications of relators in a sequence its *cost*, and we call the minimum cost of a sequence that reduces w to ε the *filling area* of w, denoted by $\delta_G(w)$. We then define the *Dehn function* of G to be

$$\delta_G(n) = \max_{\ell(w) \le n} \delta_G(w),$$

where the maximum is taken over words representing the identity. For convenience, if v, w are two words representing the same element of H, we define $\delta_G(v, w) = \delta_G(vw^{-1})$; this is the minimum cost to transform v to w.

Likewise, if X is a simply-connected riemannian manifold or simplicial complex (more generally a local Lipschitz neighborhood retract) and $\gamma : S^1 \to X$ is a Lipschitz closed curve, we define its filling area $\delta_X(\gamma)$ to be the infimal area of a Lipschitz map $D^2 \to X$ that extends γ. We can define the Dehn function of X to be

$$\delta_X(n) = \sup_{\ell(\gamma) \le n} \delta_X(\gamma),$$

where the supremum is taken over null-homotopic closed curves. As in the combinatorial case, if β and γ are two curves connecting the same points and that are homotopic with their endpoints fixed, we define $\delta_X(\beta, \gamma)$ to be the infimal area of a homotopy between β and γ, which fixes their endpoints.

Note that combinatorial fillings can be converted into geometric fillings. Gromov stated the following theorem connecting geometric and combinatorial Dehn functions, a proof of which can be found in [Bri02].

THEOREM 2.1 (Gromov's Filling Theorem). *If X is a simply connected riemannian manifold or simplicial complex and G is a finitely presented group acting properly discontinuously, cocompactly, and by isometries on M, then $\delta_G \sim \delta_M$.*

Here, $f \sim g$ if f and g grow at the same rate. Specifically, if $f, g : \mathbb{N} \to \mathbb{N}$, let $f \lesssim g$ if and only if there is a c such that

$$f(n) \le cg(cn + c) + c \text{ for all } n$$

and $f \sim g$ if and only if $f \lesssim g$ and $g \lesssim f$.

2.3. $\mathrm{SL}(p;\mathbb{R})$ *and* $\mathrm{SL}(p;\mathbb{Z})$. Let $\Gamma = \mathrm{SL}(p;\mathbb{Z})$, and let $G = \mathrm{SL}(p) = \mathrm{SL}(p;\mathbb{R})$. One of the main geometric features of G is that it acts on a nonpositively curved symmetric space, which we denote by $\mathcal{E}$. Let $\mathcal{E} = \mathrm{SL}(p;\mathbb{R})/\mathrm{SO}(p)$. We consider $\mathcal{E}$ with the metric obtained from the inner product $\langle u, v\rangle = \mathrm{trace}(u^{tr}v)$ on the space of symmetric matrices. Under this metric, $\mathcal{E}$ is a nonpositively curved symmetric space. The lattice Γ acts on $\mathcal{E}$ with finite covolume, but the action is not cocompact. Let $\mathcal{M} := \Gamma\backslash\mathcal{E}$. If $x \in G$, we write the equivalence class of x in $\mathcal{E}$ as $[x]_\mathcal{E}$; similarly, if $x \in G$ or $x \in \mathcal{E}$, we write the equivalence class of x in $\mathcal{M}$ as $[x]_\mathcal{M}$.

If $g \in G$ is a matrix with coefficients $\{g_{ij}\}$, we define

$$\|g\|_2 = \sqrt{\sum_{i,j} g_{ij}^2},$$

$$\|g\|_\infty = \max_{i,j} |g_{ij}|.$$

Note that for all $g, h \in G$, we have $\log \|g\|_2 = O(d_G(I, g))$.

One key fact about the geometry of $\mathrm{SL}(p;\mathbb{Z})$ is a theorem of Lubotzky, Mozes, and Raghunathan [LMR93].

THEOREM 2.2. *The word metric on* $\mathrm{SL}(p;\mathbb{Z})$ *for* $p \geq 3$ *is equivalent to the restriction of the riemannian metric of* $\mathrm{SL}(p;\mathbb{R})$ *to* $\mathrm{SL}(p;\mathbb{Z})$. *That is, there is a* c *such that for all* $g \in \mathrm{SL}(p;\mathbb{Z})$, *we have*

$$c^{-1} d_G(I, g) \leq d_\Gamma(I, g) \leq c\, d_G(I, g).$$

We define the subset of G on which Γ acts cocompactly by interpreting $\mathcal{E}$ as the set of unimodular bases of $\mathbb{R}^p$ up to rotation; if $v_1, \dots, v_n$ are the rows of a matrix g, then the point $[g]_\mathcal{E}$ corresponds to the basis $\{v_1, \dots, v_n\}$. An element of Γ acts on $\mathcal{E}$ by replacing the basis elements by integer combinations of basis elements. This preserves the lattice that they generate, so we can think of $\mathcal{M}$ as the set of unit-covolume lattices in $\mathbb{R}^p$ up to rotation. Nearby points in $\mathcal{M}$ or $\mathcal{E}$ correspond to bases or lattices that can be taken into each other by small linear deformations of $\mathbb{R}^p$. Note that this set is not compact — for instance, the injectivity radius of a lattice is a positive continuous function on $\mathcal{M}$, and there are lattices with arbitrarily small injectivity radiuses.

Let $\mathcal{E}(\epsilon)$ be the set of points that correspond to lattices with injectivity radius at least ϵ. This is invariant under Γ, and when $0 < \epsilon \leq 1/2$, it is contractible and Γ acts on it cocompactly [ECH+92]. We call $\mathcal{E}(\epsilon)$ the *thick part* of $\mathcal{E}$ and its preimage $G(\epsilon)$ in G the thick part of G. “Thick” here refers to the fact that the quotients $\Gamma\backslash\mathcal{E}(\epsilon)$ and $\Gamma\backslash G(\epsilon)$ have injectivity radius bounded below.

Epstein, et al. construct a Lipschitz deformation retraction from $\mathcal{E}$ to $\mathcal{E}(\epsilon)$, so $\mathcal{E}(\epsilon)$ is a local Lipschitz neighborhood retract in $\mathcal{E}$. The results of [Gro] imply that Gromov's Filling Theorem extends to such retracts, so proving a filling inequality for Γ is equivalent to proving one for $\mathcal{E}(\epsilon)$.

We will also define some subgroups of G. In the following, $\mathbb{K}$ represents either $\mathbb{Z}$ or $\mathbb{R}$. Let $z_1,\dots,z_p$ be the standard generators for $\mathbb{Z}^p$, and if $S \subset \{1,\dots,p\}$, let $\mathbb{R}^S = \langle z_s\rangle_{s\in S}$ be a subspace of $\mathbb{R}^p$. If $q \le p$, there are many ways to include SL(q) in SL(p). Let SL(S) be the copy of SL($\#S$) in SL(p) that acts on $\mathbb{R}^S$ and fixes z_t for $t \notin S$. If $S_1,\dots,S_n$ are disjoint subsets of $\{1,\dots,p\}$ and $S := \bigcup S_i$, let

$$U(S_1,\dots,S_n) \subset \mathrm{SL}(S;\mathbb{Z})$$

be the subgroup of matrices preserving the flag

$$\mathbb{R}^{S_i} \subset \mathbb{R}^{S_i \cup S_{i-1}} \subset \dots \subset \mathbb{R}^S$$

when acting on the right. If the S_i are sets of consecutive integers in increasing order, $U(S_1,\dots,S_n)$ is block upper-triangular. For example, $U(\{1\},\{2,3\})$ is the subgroup of SL($3;\mathbb{K}$) consisting of matrices of the form

$$\begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix}.$$

If $d_1,\dots,d_n > 0$, let $U(d_1,\dots,d_n)$ be the group of upper block triangular matrices with blocks of the given lengths, so that the subgroup illustrated above is $U(1,2)$. If $\sum_i d_i = p$, this is a parabolic subgroup of Γ; if $\sum_i d_i < p$, then it is a parabolic subgroup of SL($\sum_i d_i;\mathbb{Z}$). Let $\mathcal{P}$ be the set of groups $U(d_1,\dots,d_n)$ with $\sum_i d_i = p$, including $U(p) = \Gamma$. Any parabolic subgroup of Γ is conjugate to a unique such group.

One feature of SL($p;\mathbb{Z}$) is that it has a particularly simple presentation, the Steinberg presentation. If $1 \le i \ne j \le p$, let $e_{ij}(x) \in \Gamma$ be the identity matrix with the (i,j)-entry replaced by x; we call these *elementary matrices*. Let $e_{ij} := e_{ij}(1)$. When $p \ge 3$, there is a finite presentation which has the matrices e_{ij} as generators [Ste62, Mil71]:

$$\begin{aligned} \Gamma = \langle e_{ij} \mid [e_{ij}, e_{kl}] &= I && \text{if } i \ne l \text{ and } j \ne k, \\ [e_{ij}, e_{jk}] &= e_{ik} && \text{if } i \ne k, \\ (e_{ij}e_{ji}^{-1}e_{ij})^4 &= I\rangle, \end{aligned} \tag{1}$$

where we adopt the convention that $[x,y] = xyx^{-1}y^{-1}$. We will use a slightly expanded set of generators. Let

$$\Sigma := \{e_{ij} \mid 1 \le i \ne j \le p\} \cup D,$$

where $D \subset \Gamma$ is the set of diagonal matrices in SL($p;\mathbb{Z}$); note that this set is finite. If R is the set of relators given above with additional relations expressing each element of D as a product of elementary matrices, then $\langle \Sigma \mid R\rangle$ is a finite presentation of Γ with relations R. Furthermore, if $H = \mathrm{SL}(q;\mathbb{Z}) \subset \mathrm{SL}(p;\mathbb{Z})$ or if H is a subgroup of block-upper-triangular matrices, then H is generated by $\Sigma \cap H$.

2.4. *Templates and relative Dehn functions.* In this section, we introduce some new definitions which we will use in the proof of Theorem 1.3.

The *relative Dehn function*, $\delta^{\mathrm{rel}}_{H\subset G}$, of a subgroup $H \subset G$ describes the difficulty of filling words in H by discs in G. If $G = \{S|R\}$ is a finite presentation for G and $S_0 \subset S$ is a generating set for H and S_0^* represents the set of words in S_0, we define

$$\delta^{\mathrm{rel}}_{H\subset G}(n) = \max_{w\in S_0^*, \ell(w)\le n} \delta_G(w).$$

By definition, $\delta^{\mathrm{rel}}_{G\subset G}(n) = \delta_G(n)$.

If $\omega : G \to S^*$ is a map such that for all g, $\omega(g)$ is a word representing g which has length $\sim \ell(g)$, we say that ω is a *normal form* for G. We will define a *triangular relative Dehn function*, $\delta^{\mathrm{tri}}_{H,\omega}$, which describes the difficulty of filling "ω-triangles" with vertices in H. If $g_1, g_2, g_3 \in G$, we say that

$$\Delta_\omega(g_1, g_2, g_3) = \omega(g_1^{-1}g_2)\omega(g_2^{-1}g_3)\omega(g_3^{-1}g_1)$$

is the *ω-triangle* with vertices g_1, g_2, g_3. Then we can define

$$\delta^{\mathrm{tri}}_{H,\omega}(n) = \max_{\substack{h_1,h_2,h_3\in H\\ \mathrm{diam}\{h_1,h_2,h_3\}\le n}} \delta_G\left(\Delta_\omega(h_1, h_2, h_3)\right).$$

If $h \in H$, then even though $\omega(h)$ will have endpoints in H, it need not be a word in H, so upper bounds on $\delta^{\mathrm{rel}}_{H\subset G}$ might not lead to upper bounds on $\delta^{\mathrm{tri}}_{H,\omega}$. On the other hand, we can bound $\delta^{\mathrm{rel}}_{H\subset G}$ by decomposing words in H into ω-triangles. We can describe these decompositions using *templates*. Let τ be a triangulation of D^2 whose vertices are labeled by elements of G; this is a template. If the boundary vertices of τ are labeled (in order), $g_1, \dots, g_n$, we let

$$w_\tau = \omega(g_1^{-1}g_2)\cdots\omega(g_{n-1}^{-1}g_n)\omega(g_n^{-1}g_1)$$

and call w_τ the *boundary word* of τ. If $w = w_1 \cdots w_n$ is a word and the boundary of τ is an n-gon with labels I, w_1, $w_1w_2, \dots, w_1\cdots w_{n-1}$, we call τ a *template for w*.

We say that we can *break* a word w into some words w_i at cost C if each of the w_i's represent the identity and there exist words g_i such that

$$\delta_\Gamma\left(w, \prod_i g_i w_i g_i^{-1}\right) = C.$$

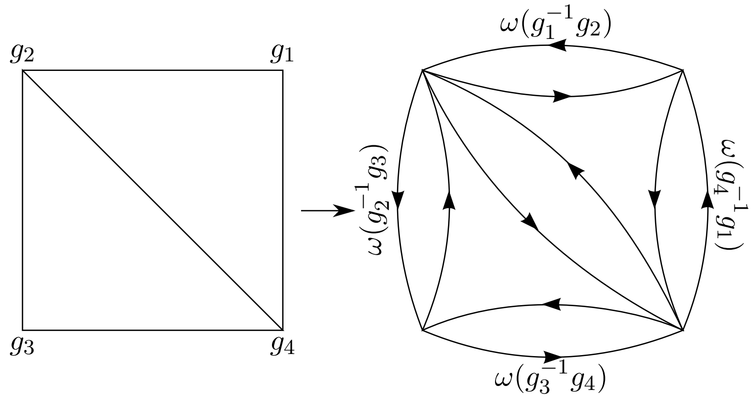

Figure 1. The boundary word of the template on the left is $w_\tau = \omega(g_1^{-1}g_2)\omega(g_2^{-1}g_3)\omega(g_3^{-1}g_4)\omega(g_4^{-1}g_1)$. On the right, we use the template to break w_τ into five ω-bigons of the form $\omega(g_i^{-1}g_j)\omega(g_j^{-1}g_i)$ and two ω-triangles of the form $\Delta_\omega(g_i, g_j, g_k)$.

In particular, this means that

$$\delta_\Gamma(w) \leq C + \delta_\Gamma\left(\prod_i g_i w_i g_i^{-1}\right) \leq C + \sum_i \delta_\Gamma(w_i).$$

If τ is a template, we can break w_τ into the ω-triangles and ω-bigons corresponding to faces and edges of τ at cost 0, as in Figure 1. If τ is a template for w, then w_τ can be transformed to w at cost $O(n)$, implying the following lemma.

Lemma 2.3. *Let $w = w_1 \cdots w_n$ be a word of length n and let τ be a template for w. If the i-th face of τ has vertices g_{i1}, g_{i2}, g_{i3} and the j-th edge of τ has vertices h_{j1}, h_{j2}, then*

$$\delta_G(w) \leq \sum_i \delta_G(\Delta_\omega(g_{i1}, g_{i2}, g_{i3})) + \sum_j \delta_G(\omega(h_{j1}^{-1}h_{j2})\omega(h_{j2}^{-1}h_{j1})) + O(n).$$

Many Dehn function bounds involve a divide-and-conquer strategy which breaks a complicated word into smaller, simpler words, and templates are useful to describe such strategies. For example, one divide-and-conquer strategy uses the template in Figure 2 to build a filling of arbitrary words in a group out of ω-triangles. A strategy like this is used, for instance, in [Gro93, 5.A$''_3$], [LP04], and [dCT10]; in fact, the following lemma is essentially equivalent to Lemma 4.3 in [dCT10].

Lemma 2.4. *If there is an $\alpha > 1$ such that for all $h_i \in H$ such that*

$$\delta^{\text{tri}}_{H,\omega}(n) \lesssim n^\alpha,$$

then

$$\delta^{\text{rel}}_{H\subset G}(n) \lesssim n^\alpha.$$

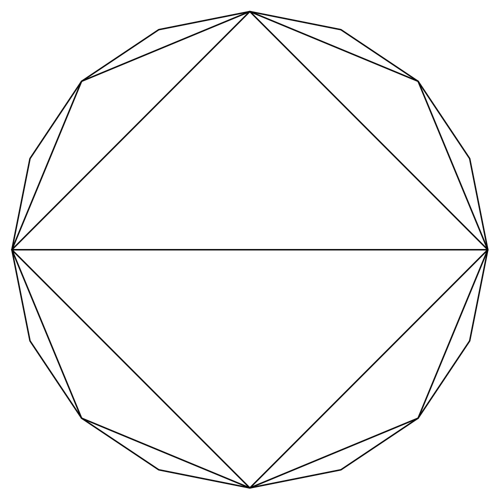

Figure 2. A dyadic template.

Proof. Let $S_0 \subset S$ be a generating set for H, as in the definition of $\delta^{\mathrm{rel}}_{H\subset G}$. Without loss of generality, we may assume that the identity I is in S_0. Let $w = w_1 \cdots w_n$. It suffices to consider the case that $n = 2^k$ for some $k \in \mathbb{Z}$; otherwise, we may pad w with the letter I until its length is a power of 2. Let $w(i) = w_1 \cdots w_i$. Let τ be the template consisting of $2^k - 2$ triangles as in Figure 2, where the vertices of τ are labeled by $w(i)$.

Each triangle of τ has vertices labeled

$$w(i2^j), w((i+1/2)2^j), w((i+1)2^j)$$

for some $1 \le j < k$ and $0 \le i < 2^{-j}n$, which are separated by distances at most 2^j. By the hypothesis, the corresponding ω-triangle has a filling of area $O(2^{\alpha j})$. Similarly, each edge has vertices labeled $w(i2^j)$ and $w((i+1)2^j)$ and corresponds to an ω-bigon that can be filled at cost $O(2^\alpha j)$. There are $\sim 2^{-j}n$ bigons and edges of size 2^j, so after summing all the contributions, we find that $\delta_H(w) \lesssim n^\alpha$. $\square$

3. **Sketch of proof**

Note that since $\mathrm{SL}(p;\mathbb{Z})$ is not hyperbolic when $p \ge 3$, its Dehn function is at least quadratic. To prove Theorem 1.3, it suffices to show that any word in $\mathrm{SL}(p;\mathbb{Z})$ has a quadratic filling. We proceed by induction on subgroups of $\mathrm{SL}(p;\mathbb{Z})$. Very roughly, we decompose words in $\mathrm{SL}(p;\mathbb{Z})$ into words in subgroups of $\mathrm{SL}(p;\mathbb{Z})$ and then repeat the process inductively to get a filling of the original word. We reduce in two main ways. First, a word in $\mathrm{SL}(p;\mathbb{Z})$ corresponds to a curve in the symmetric space $\mathcal{E} = \mathrm{SL}(p;\mathbb{R})/\mathrm{SO}(n)$, and since $\mathcal{E}$ is nonpositively curved, it has a filling of quadratic area. By breaking this filling into pieces lying in different horoballs, we can break the original word into pieces lying in maximal parabolic subgroups.

Second, a parabolic subgroup of $\mathrm{SL}(p;\mathbb{Z})$ is conjugate to an upper triangular subgroup and can be written as a semidirect product of a unipotent group (the off-diagonal part) and a product of $\mathrm{SL}(q;\mathbb{Z})$'s (the diagonal blocks). We use techniques like those used by Leuzinger and Pittet [LP04] to reduce words in $\mathrm{SL}(p;\mathbb{Z})$ to words in the diagonal blocks. Since each diagonal block is smaller than the original matrix, repeating these two steps eventually simplifies the word.

We describe this process more rigorously in the following lemmas. In all of these lemmas, ω will represent a normal form for $\mathrm{SL}(p;\mathbb{Z})$; we will define ω in Section 6. One key property of ω will be that it is a product of words representing elementary matrices, which we call *shortcuts*. These shortcuts are based on the constructions in [LMR93]. Lubotzky, Mozes, and Raghunathan showed that the transvection $e_{ij}(x)$ can be represented by a word of length logarithmic in x; we denote this word by $\hat{e}_{ij}(x)$ and call it a *shortcut for* $e_{ij}(x)$. If $H \subset \mathrm{SL}(p;\mathbb{Z})$, we say that w is a *shortcut word* in H if we can write $w = \prod_{i=1}^{n} w_i$, where each w_i is either a diagonal matrix in H or a shortcut $\hat{e}_{a_ib_i}(x_i)$ where $e_{a_ib_i}(x_i) \in H$. Our normal form ω will express elements $g \in \mathrm{SL}(q;\mathbb{Z})$ as shortcut words, and if $g \in \mathrm{SL}(q;\mathbb{Z})$ or $g \in U(s_1,\dots,s_k)$, then $\omega(g)$ will be a shortcut word in $\mathrm{SL}(q;\mathbb{Z})$ or in $U(s_1,\dots,s_k)$ respectively.

First, we will break loops in $\mathrm{SL}(q;\mathbb{Z}) \subset \mathrm{SL}(p;\mathbb{Z})$ into ω-triangles with vertices in maximal parabolic subgroups.

Lemma 3.1 (Reduction to maximal parabolics). *Let $p \geq 5$ and $2 < q \leq p$. There is a $c > 0$ such that if w is a word in $\mathrm{SL}(q;\mathbb{Z})$ of length ℓ, then there are words $w_1,\dots,w_k$ such that we can break w into the $w_1,\dots,w_k$ at cost $O(\ell)$; each w_i either has length $\leq c$ or is an ω-triangle with vertices in some $U(q_i, q-q_i)$; and*

$$\sum_i \ell(w_i)^2 = O(\ell^2).$$

As a consequence, if

$$\delta^{\mathrm{tri}}_{U(s,q-s),\omega}(n) \lesssim n^2$$

for $s = 1,\dots,q-1$, then

$$\delta^{\mathrm{rel}}_{\mathrm{SL}(q;\mathbb{Z})\subset\mathrm{SL}(p;\mathbb{Z})}(n) \lesssim n^2.$$

By our choice of ω, each w_i above is a shortcut word in some parabolic subgroup. Each parabolic subgroup is a semi-direct product of a unipotent subgroup and a (virtual) product of copies of $\mathrm{SL}(q_i;\mathbb{Z})$, so we fill the triangles obtained in the previous lemma by reducing them to shortcut words in the diagonal blocks and shortcut words in the unipotent subgroup. The word in the unipotent subgroup can be filled by combinatorial methods, leaving just the words in the diagonal blocks.

Remark 3.2. Ideally, we would be able to construct a projection from an ω-triangle in a parabolic subgroup P to shortcut words in each diagonal block, and thus break an ω-triangle w in P into one shortcut word for each diagonal block of P at cost $O(\ell(w)^2)$. When $P \neq U(p-1,1)$, this is possible, but when $P = U(p-1,1)$, a different method of proof is necessary.

LEMMA 3.3 (Reduction to diagonal blocks). *Let $p \geq 5$ and $q < p$. Let $1 \leq s_1, \dots, s_k \leq q$ be such that $\sum_i s_i \leq p$, and suppose that w is an ω-triangle with vertices in $U(s_1, \dots, s_k)$ of length ℓ. There are words $w_1, \dots, w_n$ such that we can break w into the w_i's at cost $O(\ell^2)$. Furthermore, for all i, there is a $q_i < q$ such that w_i is a shortcut word in $\mathrm{SL}(q_i; \mathbb{Z})$, and*

$$\sum_i \ell(w_i)^2 = O(\ell^2).$$

To apply Lemma 3.1 to these w_i and complete the induction, we need to replace these shortcut words with words in $\mathrm{SL}(q;\mathbb{Z})$. When q is sufficiently large, this can be done at quadratic cost.

LEMMA 3.4 (Moving shortcuts into subgroups). *Let $p \geq 5$ and $2 < q \leq p$. If w is a shortcut word in $\mathrm{SL}(q;\mathbb{Z})$, there is a word w' in $\mathrm{SL}(q;\mathbb{Z})$ such that $\ell(w') = O(\ell(w)$ and $\delta_\Gamma(w, w') = O(\ell(w)^2)$.*

Ultimately, the previous three lemmas break loops in $\mathrm{SL}(p;\mathbb{Z})$ into shortcut words in $\mathrm{SL}(2;\mathbb{Z})$. Even though $\mathrm{SL}(2;\mathbb{Z})$ is virtually free and has linear Dehn function, shortcut words may leave $\mathrm{SL}(2;\mathbb{Z})$ and may have quadratic fillings.

LEMMA 3.5 (Base case). *Let $p \geq 5$, and let w be a shortcut word in $\mathrm{SL}(2;\mathbb{Z})$ of length ℓ. Then*

$$\delta_\Gamma(w) = O(\ell^2).$$

These four lemmas prove Theorem 1.3.

Proof of Theorem 1.3. We claim that if w is a shortcut word in $\mathrm{SL}(q;\mathbb{Z})$, then

$$\delta_\Gamma(w) = O(\ell^2).$$

This implies that

$$\delta^{\mathrm{rel}}_{\mathrm{SL}(q;\mathbb{Z}) \subset \mathrm{SL}(p;\mathbb{Z})}(n) \lesssim n^2,$$

and when $q = p$, this proves the theorem.

We proceed by induction. When $q = 2$, the statement is Lemma 3.5. Otherwise, since $q > 2$, we can apply Lemma 3.4 to replace w by a word w' in $\mathrm{SL}(q;\mathbb{Z})$, and we apply Lemmas 3.1 and 3.3 to break w' into words w_i, each a shortcut word in some $\mathrm{SL}(q_i;\mathbb{Z})$'s, such that $\sum_i \ell(w_i)^2 = O(\ell^2)$. This has cost $O(\ell^2)$, and by the inductive hypothesis, the total filling area of the w_i's is also $O(\ell^2)$, as desired. $\square$

In the next two subsections, we will describe some of the ideas behind the proofs of these lemmas. Then, Lemma 3.1 will be proved in Section 5, Lemma 3.4 will be proved in Section 7.1, Lemma 3.3 will be proved in Section 8, and Lemma 3.5 will be proved in Section 9.

3.1. *Constructing templates from Lipschitz fillings.* One idea behind the proof of Theorem 1.3 is that we can use a Lipschitz filling of a curve in a symmetric space to construct a template for a filling of w. In this section, we will sketch how to use the pattern of intersections between the filling and the horoballs in the symmetric space to break w into ω-triangles lying in parabolic subgroups.

If w is a word in $\mathrm{SL}(q;\mathbb{Z})$, it corresponds to a curve γ_w in the nonpositively curved symmetric space $\mathcal{E} = \mathrm{SL}(q;\mathbb{R})/\mathrm{SO}(q)$ of length ℓ, and this curve has a quadratic filling. Indeed, if $D^2(\ell)$ is the disc $[0,\ell]\times[0,\ell]$, there is a filling $f : D^2(\ell) \to \mathcal{E}$ that has Lipschitz constant at most 2. We can construct f by choosing a basepoint on the curve and contracting the curve to the basepoint along geodesics. Choose a Siegel set $\mathcal{S}\subset\mathcal{E}$; this is a fundamental set for the action of $\mathrm{SL}(q;\mathbb{Z})$ on $\mathcal{E}$ (see Section 4). Each point of $\mathcal{E}$ lies in some translate of $\mathcal{S}$; we can define a map $\rho : \mathcal{E}\to\mathrm{SL}(q;\mathbb{Z})$ by sending each point x to a group element $\rho(x)$ such that $x\in\rho(x)\mathcal{S}$. Then, if τ is a triangulation of $D^2(\ell)$, we can label each vertex v by the element $\rho(f(v))$. This is a template, and if the boundary edges of τ each have length bounded by a constant, then the boundary word w_τ of the template is uniformly close to w.

As a simple application, we will show that for any q, the Dehn function of $\mathrm{SL}(q;\mathbb{Z})$ is bounded by an exponential function. It is straightforward to show that the injectivity radius of $z\in\mathcal{E}/\mathrm{SL}(q;\mathbb{Z})$ shrinks exponentially quickly as $z\to\infty$; that is, that there is a c such that if $x,y\in\mathcal{E}$, $d_{\mathcal{E}}(I,x)\le r$, and $d_{\mathcal{E}}(x,y)\le e^{-cr}$, then $d_{\mathrm{SL}(q;\mathbb{Z})}(\rho(x),\rho(y))\le c$. Let τ be a triangulation of $D^2(\ell)$ by triangles with side lengths at most $e^{-2c\ell}$. If an edge of τ connects vertices u and v, then $d_{\mathrm{SL}(q;\mathbb{Z})}(\rho(f(u)),\rho(f(v)))\le c$, so

$$\delta(w_\tau)\le F\delta(3c)+E\delta(2c),$$

where F is the number of faces of τ and E is the number of edges. Since we can construct τ to have at most exponentially many triangles, $\delta(w_\tau)\lesssim e^{\ell}$.

Triangulations with larger simplices lead to larger ω-triangles but potentially stronger bounds on the Dehn function; for example, in [You], we used a triangulation by triangles of diameter ~ 1 to prove a quartic bound on $\mathrm{SL}(q;\mathbb{Z})$ when $q\ge 5$. The basic idea behind that proof is that if $x,y\in\mathcal{E}$ are sufficiently close together, then either $\rho(x)^{-1}\rho(y)$ is bounded or it lies in a parabolic subgroup of $\mathrm{SL}(q;\mathbb{Z})$, so the methods above produce a template whose triangles all either have bounded size or lie in a parabolic subgroup. Furthermore, since each edge is short, the group elements corresponding to edges satisfy bounds that make the triangles easy to fill.

We use a similar idea to prove Lemma 3.1. One can show (see Corollary 4.8) that if x is deep in the cusp of $\mathcal{M}$, i.e., if $r(x) = d_{\mathcal{E}}(x, [\mathrm{SL}(p,\mathbb{Z})]_{\mathcal{E}})$ is large, then there is a ball around x of radius $\sim r(x)$ that is contained in a horoball corresponding to a maximal parabolic subgroup of $\mathrm{SL}(q;\mathbb{Z})$. In particular, if $d(x,y) \ll r(x)$, then $\rho(x)^{-1}\rho(y)$ lies in a maximal parabolic subgroup of $\mathrm{SL}(q;\mathbb{Z})$.

If $f : D^2(\ell) \to \mathcal{E}$ is a Lipschitz filling of a curve γ, we can construct a triangulation of $D^2(\ell)$ where the size of each triangle is proportional to its distance from the thick part. By labeling the vertices of this triangulation as above, we get a template made of triangles that are either "small" or "large." Small triangles are those whose image under f is in the thick part; since the injectivity radius of $\mathcal{E}$ is bounded away from zero in the thick part, the vertex labels of a small triangle are a bounded distance apart in $\mathrm{SL}(q;\mathbb{Z})$. Large triangles are those whose image is in the thin part. The image of a large triangle under f lies in a horoball, and its vertex labels lie in a conjugate of one of the maximal parabolic subgroups. Ultimately, this lets us break words in $\mathrm{SL}(q;\mathbb{Z})$ into ω-triangles with vertices in parabolic subgroups. This is a key step in the proofs of Lemmas 3.1 and 3.5 and in a special case of Lemma 3.3.

3.2. *Shortcuts in* $\mathrm{SL}(p;\mathbb{Z})$. Another idea behind the proof of Theorem 1.3 is the idea of *shortcuts*, words of length $\sim \log n$ that represent transvections in $\mathrm{SL}(p;\mathbb{Z})$ with coefficients of order n. These shortcuts are a key ingredient in the construction of the normal form ω. Transvections satisfy Steinberg relations, and one of the key combinatorial lemmas (Lemma 7.6) states that when these Steinberg relations are written in terms of shortcuts, the resulting words have quadratic fillings.

Our shortcuts are based on constructions of Lubotzky, Mozes, and Raghunathan [LMR93], who used them to show that distances in the word metric on $\mathrm{SL}(p;\mathbb{Z})$ are comparable to distances in the riemannian metric on the symmetric space $\mathrm{SL}(p;\mathbb{Z})/\mathrm{SO}(p)$ when $p \geq 3$. In particular, if $M \in \mathrm{SL}(p;\mathbb{Z})$ is a matrix with coefficients bounded by $\|M\|_\infty$, there is a word w that represents M as a product of $\sim \log \|M\|_\infty$ generators of $\mathrm{SL}(p;\mathbb{Z})$. They construct this w by decomposing M into a product of transvections with integer coefficients, then writing each transvection as a word in $\mathrm{SL}(p;\mathbb{Z})$. This can be done efficiently because unipotent subgroups of $\mathrm{SL}(p;\mathbb{Z})$ are exponentially distorted; a transvection with L^∞ norm N can be written as a word of length $\sim \log N$. In fact, a transvection can be written as a word of length $\sim \log N$ in many ways.

One advantage of working with $\mathrm{SL}(p;\mathbb{Z})$ instead of an arbitrary lattice in a high-rank Lie group is that shortcuts in $\mathrm{SL}(p;\mathbb{Z})$ can be written with just a few generators and that many of the generators of $\mathrm{SL}(p;\mathbb{Z})$ commute. For example, if we define $e_{ij}(x)$ to be the elementary matrix obtained by replacing the (i,j)-entry of the identity matrix by x, there is a word $\hat{e}_{13}(x)$ in the alphabet $\{e_{12}, e_{21}, e_{13}, e_{23}\}$ that represents $e_{13}(x)$ and has length $\sim \log|x|$. If w is a product of generators that commute with this alphabet, it is easy to fill words like $[\hat{e}_{13}(x), w]$. Furthermore, when $p \geq 5$, different ways of constructing shortcuts are close together; we can arrange things so that if $\hat{e}$ and $\hat{e}'$ are shortcuts for the same elementary matrix written in different alphabets, then

$$\delta(\hat{e}, \hat{e}') \lesssim \ell(\hat{e})^2. \tag{2}$$

This lets us write elementary matrices in terms of whichever alphabet is most convenient. The fact that (2) is not true when $p = 4$ is the biggest obstacle to extending these techniques to SL($4;\mathbb{Z}$); an analogue of (2) for SL($4;\mathbb{Z}$) would lead to a polynomial bound on its Dehn function.

Remark on notation. We will generally use hats to denote shortcuts, so $e_{ij}(x)$ and $u(V)$ will denote unipotent matrices and $\hat{e}_{ij}(x)$ and $\hat{u}(V)$ will denote words of logarithmic length that represent the corresponding matrices.

We prove Lemma 3.3 using these shortcuts. The normal form ω expresses elements of SL($p;\mathbb{Z}$) in terms of shortcuts, and the proof of Lemma 3.3 mostly consists of combinatorial calculations involving these shortcuts. For example, as mentioned above, one step in the proof involves constructing fillings of Steinberg relations. The elementary matrices $e_{ij}(x)$ satisfy relations like $[e_{ij}(x), e_{kl}(y)] = I$ and $[e_{ij}(x), e_{jk}(y)] = e_{ik}(xy)$, so the corresponding products of shortcuts (e.g., $[\hat{e}_{ij}(x), \hat{e}_{kl}(y)]$) are words representing the identity. By rewriting these shortcuts in appropriate alphabets, we can fill these words efficiently.

4. **Siegel sets and the depth function**

Let $G = \mathrm{SL}(p;\mathbb{R})$ and $\Gamma = \mathrm{SL}(p;\mathbb{Z})$. Given a fundamental set F for the action of Γ on $\mathcal{E}$, one can construct a map $\mathcal{E} \to \Gamma$ that sends each point x of $\mathcal{E}$ to an element $g \in \Gamma$ such that $x \in gF$. In general, this map need not be well behaved, but if F is a Siegel set, this map has many useful properties. In this section, we will define a Siegel set $\mathcal{S}$ and describe some of its properties. Note that the constructions in this section generalize to many reductive and semisimple Lie groups with the use of precise reduction theory, but we will only state the results for SL($p;\mathbb{Z}$), as stating the theorems in full generality requires a lot of additional background. (See Section 10 for some discussion of the general case.)

Let $\mathrm{diag}(t_1, \dots, t_p)$ be the diagonal matrix with entries $(t_1, \dots, t_p)$. Let A be the set of diagonal matrices in G, and if $\epsilon > 0$, let

$$A_\epsilon^+ = \Big\{ \mathrm{diag}(t_1, \dots, t_p) \mid \prod t_i = 1, t_i > 0, t_i \geq \epsilon t_{i+1} \Big\}.$$

Let $\mathcal{M} = \mathrm{SL}(p;\mathbb{Z})\backslash\mathcal{E}$. One of the main features of $\mathcal{M}$ is that it is Hausdorff equivalent to A^+_ϵ; our main goal in this section is to describe this Hausdorff equivalence and its "fibers." Let N be the set of upper triangular matrices with 1's on the diagonal, and let N^+ be the subset of N with off-diagonal entries in the interval $[-1/2, 1/2]$. Translates of the set $N^+A^+_\epsilon$ are known as Siegel sets. The following properties of Siegel sets are well known (see, for instance, [BHC62]).

Lemma 4.1. *There is an $1 > \epsilon_\mathcal{S} > 0$ such that if we let*

$$\mathcal{S} := [N^+A^+_{\epsilon_\mathcal{S}}]_\mathcal{E} \subset \mathcal{E},$$

then

- $\Gamma\mathcal{S} = \mathcal{E}$.
- *There are only finitely many elements $\gamma \in \Gamma$ such that $\gamma\mathcal{S} \cap \mathcal{S} \neq \emptyset$.*

In particular, the quotient map $\mathcal{S} \to \mathcal{M}$ is a surjection. We define $A^+ := A^+_{\epsilon_\mathcal{S}}$.

The inclusion $A^+ \hookrightarrow \mathcal{S}$ is a Hausdorff equivalence. That is, if we give A the riemannian metric inherited from its inclusion in G, so that

$$d_A(\mathrm{diag}(d_1, \dots, d_p), \mathrm{diag}(d'_1, \dots, d'_p)) = \sqrt{\sum_{i=1}^p \left|\log \frac{d'_i}{d_i}\right|^2},$$

then

Lemma 4.2 ([JM02]). *There is a c such that if $n \in N^+$ and $a \in A^+$, then $d_\mathcal{E}([na]_\mathcal{E}, [a]_\mathcal{E}) \leq c$. In particular, if $x \in \mathcal{S}$, then $d_\mathcal{E}(x, [A^+]_\mathcal{E}) \leq c$. Furthermore, if $x, y \in A^+$, then $d_A(x, y) = d_\mathcal{S}(x, y)$.*

Proof. For the first claim, note that if $x = [na]_\mathcal{E}$, then $x = [a(a^{-1}na)]_\mathcal{E}$, and $a^{-1}na \in N$. Furthermore,

$$\|a^{-1}na\|_\infty \leq \epsilon_\mathcal{S}^{-p},$$

so

$$d_\mathcal{E}([x]_\mathcal{E}, [a]_\mathcal{E}) \leq d_G(I, a^{-1}na)$$

is bounded independently of x.

For the second claim, we clearly have $d_A(x, y) \geq d_\mathcal{S}(x, y)$. For the reverse inequality, it suffices to note that the map $\mathcal{S} \to A^+$ given by $na \mapsto a$ for all $n \in N^+$, $a \in A^+$ is distance-decreasing. □

Siegel conjectured that the quotient map from $\mathcal{S}$ to $\mathcal{M}$ is also a Hausdorff equivalence; that is,

Theorem 4.3. *There is a c' such that if $x, y \in \mathcal{S}$, then*

$$d_\mathcal{E}(x, y) - c' \leq d_\mathcal{M}([x]_\mathcal{M}, [y]_\mathcal{M}) \leq d_\mathcal{E}(x, y).$$

Proofs of this conjecture can be found in [Leu04b], [Ji98], [Din94]. As a consequence, the natural quotient map $A^+ \to \mathcal{M}$ is a Hausdorff equivalence.

Since $\mathcal{S}$ is a fundamental set, any point $x \in \mathcal{E}$ can be written (possibly nonuniquely) as $x = [\gamma na]_\mathcal{E}$ for some $\gamma \in \Gamma$, $n \in N^+$, and $a \in A^+$. Theorem 4.3 implies that these different decompositions are a bounded distance apart.

Corollary 4.4 (see [JM02, Lemmas 5.13, 5.14]). *There is a constant c'' such that if $x, y \in \mathcal{M}$, $n, n' \in N^+$, and $a, a' \in A^+$ are such that $x = [na]_\mathcal{M}$ and $y = [n'a']_\mathcal{M}$, then*

$$|d_\mathcal{M}(x, y) - d_A(a, a')| \leq c.''$$

In particular, if $[\gamma na]_\mathcal{E} = [\gamma' n'a']_\mathcal{E}$ for some $\gamma, \gamma' \in \Gamma$, then $d_A(a, a') \leq c.''$

Proof. Note that by Lemma 4.2,

$$d_\mathcal{M}(x, [a]_\mathcal{M}) \leq d_\mathcal{E}([na]_\mathcal{E}, [a]_\mathcal{E}) \leq c$$

and likewise $d_\mathcal{M}(y, [a']_\mathcal{M}) \leq c$. Furthermore, by the theorem and the lemma,

$$d_A(a, a') - c' = d_\mathcal{S}(a, a') - c' \leq d_\mathcal{M}([a]_\mathcal{M}, [a']_\mathcal{M}) \leq d_A(a, a'),$$

so if we let $c'' = c' + 2c$, the corollary follows. $\square$

Let $\rho : \mathcal{E} \to \Gamma$ be a map such that $\rho(\mathcal{S}) = I$ and $x \in \rho(x)\mathcal{S}$ for all x. Any point $x \in \mathcal{E}$ can be uniquely written as $x = [\rho(x)na]_\mathcal{E}$ for some $n \in N^+$ and $a \in A^+$. Let $\phi : \mathcal{E} \to A^+$ be the map $[\rho(x)na]_\mathcal{E} \mapsto a$. There are many choices for ρ but, by Corollary 4.4, they only affect the definition of ϕ by a bounded amount. If $\phi(x) = \operatorname{diag}(a_1, \dots, a_p)$, let $\phi_i(x) = \log a_i$. If $x, y \in \mathcal{E}$, then $|\phi_i(x) - \phi_i(y)| \leq d_\mathcal{E}(x, y) + c''$; let $c_\phi := c''$.

Define the *depth function* $r : \mathcal{E} \to \mathbb{R}^+$, $r(x) = d_\mathcal{M}([x]_\mathcal{M}, [I]_\mathcal{M})$. This function measures the distance between x and the thick part of $\mathcal{E}$; the results above imply that

$$r(x) \sim \log \|\phi(x)\|_2 \sim \phi_1(x) - \phi_p(x).$$

Since the injectivity radius of the cusp decreases exponentially as one gets further away from Γ, the distortion of ρ depends on depth.

Lemma 4.5. *There is a c such that if $x, y \in \mathcal{E}$, then*

$$d_\Gamma(\rho(x), \rho(y)) \leq c(d_\mathcal{E}(x, y) + r(x) + r(y)) + c.$$

Proof. By Theorem 4.3, there is a c_0 such that $d_\mathcal{E}([\rho(x)]_\mathcal{E}, x) \leq r(x) + c_0$, so

$$d_\mathcal{E}([\rho(x)]_\mathcal{E}, [\rho(y)]_\mathcal{E}) \leq r(x) + r(y) + d_\mathcal{E}(x, y) + 2c_0.$$

The lemma follows by Theorem 2.2. $\square$

The depth function governs ρ in other ways as well. Recall that if $x \in \mathcal{E}$ and $\tilde{x} \in G$ is a representative of x, we can construct a lattice $\mathbb{Z}^p\tilde{x} \subset \mathbb{R}^p$ and a different choice of $\tilde{x}$ corresponds to a lattice that differs by a rotation. When $r(x)$ is large, then the lattice has short vectors. If y is close to x, then vectors that are short in $\mathbb{Z}^p\tilde{y}$ are also short in $\mathbb{Z}^p\tilde{x}$. These vectors define a subspace in $\mathbb{Z}^p$, and $\rho(x)^{-1}\rho(y)$ must preserve that subspace; i.e., $\rho(x)^{-1}\rho(y)$ must lie in a parabolic subgroup. The next lemmas make this argument formal. If $x \in \mathcal{E}$ and $\tilde{x} \in G$ is such that $x = [\tilde{x}]_{\mathcal{E}}$, let

$$V(x,r) = \langle v \in \mathbb{Z}^p \mid \|v\tilde{x}\|_2 \le r\rangle;$$

we call this the *r-short subspace of x*, and it is independent of the choice of $\tilde{x}$. Let $z_1, \dots, z_p \in \mathbb{Z}^p$ be the standard generators of $\mathbb{Z}^p$.

LEMMA 4.6. *There is a $c_V > 0$ depending only on p such that if $x = [\gamma na]_{\mathcal{E}}$, where $\gamma \in \Gamma$, $n \in N^+$,*

$$a = \operatorname{diag}(a_1, \dots, a_p) \in A^+,$$

and

$$e^{c_V} a_{k+1} < r < e^{-c_V} a_k,$$

then $V(x,r) = Z_k\gamma^{-1}$, where $Z_k := \langle z_{k+1}, \dots, z_p\rangle$.

Proof. Note that $V(\gamma' x', r) = V(x', r)\gamma'^{-1}$, so we may assume that $\gamma = I$ without loss of generality. Let $n = \{n_{ij}\} \in N^+$, and let $\tilde{x} = na$. We have

$$\begin{aligned} z_j\tilde{x} &= z_j na \\ &= a_j z_j + \sum_{i=j+1}^{p} n_{ji} z_i a_i. \end{aligned}$$

Since $a_{i+1} \le a_i \epsilon_{\mathcal{S}}^{-1}$, we have $a_i \le a_{k+1}\epsilon_{\mathcal{S}}^{-p}$ for $i \ge k+1$ and $a_i \ge a_k \epsilon_{\mathcal{S}}^{p}$ for $i \le k$. Since $|n_{ji}| \le 1/2$ when $i > j$, we have

$$\|z_j\tilde{x}\|_2 \le a_{k+1}\sqrt{p}\epsilon_{\mathcal{S}}^{-p}$$

when $j > k$. Thus,

$$V(x, a_{k+1}\sqrt{p}\epsilon_{\mathcal{S}}^{-p}) \supset Z_k.$$

On the other hand, assume that $v \notin Z_k$, and let $v = \sum_i v_i z_i$ for some $v_i \in \mathbb{Z}$. Let j be the smallest integer such that $v_j \neq 0$; by assumption, $j \le k$. The z_j-coordinate of $v\tilde{x}$ is $v_j a_j$, so

$$\|v\tilde{x}\|_2 \ge a_j > a_k \epsilon_{\mathcal{S}}^{p},$$

and thus if $t < a_k\epsilon_{\mathcal{S}}^{p}$, then $V(x,t) \subset Z_k$. Therefore, if

$$a_{k+1}\sqrt{p}\epsilon_{\mathcal{S}}^{-p} \le t < a_k\epsilon_{\mathcal{S}}^{p},$$

then $V(\tilde{x}, t) = Z_j$. We can choose $c_V = \log \sqrt{p}\epsilon_{\mathcal{S}}^{-p}$. $\square$

In particular, since different choices in the construction of $\rho(x)$ must still lead to the same $V(x,r)$, this means that if $\phi_k(x)-\phi_{k+1}(x)$ is sufficiently large, then different choices of $\rho(x)$ must differ by an element of $U(k,p-k)$. The next lemma extends this by noting that nearby points of $\mathcal{E}$ must have the same r-short subspaces.

Lemma 4.7. *Let*

$$B_j(c) := \{x \in \mathcal{E} \mid \phi_j(x) - \phi_{j+1}(x) > c\}.$$

There is a $c > 0$ depending only on p such that if $1 \le j < p$, $x, y \in \mathcal{E}$ are in the same connected component of $B_j(c)$, and $g, h \in \Gamma$ are such that $x \in g\mathcal{S}$, $y \in h\mathcal{S}$, then $g^{-1}h \in U(j, p-j)$. In particular, $\rho(x)^{-1}\rho(y) \in U(j,p-j)$.

Proof. Define $s(z) = \exp \frac{\phi_{j+1}(z)+\phi_j(z)}{2}$ so that if $z \in B_j(c)$, then

$$e^{c/2}e^{\phi_{j+1}(z)} < s(z) < e^{-c/2}e^{\phi_j(z)}.$$

We will show that if c is sufficiently large, then the function $z \mapsto V(z, s(z))$ is constant on each connected component of $B_j(c)$. Let $c = 2(c_V + c_\phi + 1)$.

Note that if $z, z' \in \mathcal{E}$, if $\tilde{z}, \tilde{z}' \in G$ are representatives of z and z', and if $v \in \mathbb{Z}^p$, then

$$\left|\log \|v\tilde{z}\|_2 - \log \|v\tilde{z}'\|_2\right| \le d_{\mathcal{E}}(z, z').$$

Furthermore, if $z, z' \in B_j(c)$, then

$$|\log s(z) - \log s(z')| \le d_{\mathcal{E}}(z,z') + c_\phi.$$

Fix z. Since $z = [\rho(z)n\phi(z)]_{\mathcal{E}}$ for some $n \in N^+$, Lemma 4.6 states that if

$$\exp(c_V + \phi_{j+1}(z)) < r < \exp(-c_V + \phi_j(z)),$$

then $V(z,r) = Z_j\rho(z)^{-1}$. In particular, if

$$s(z)e^{-c_\phi - 1} < r < s(z)e^{c_\phi + 1},$$

then $V(z,r) = Z_j\rho(z)^{-1}$. So if $z' \in \mathcal{E}$ is distance at most $1/2$ from z, then

$$|\log s(z) - \log s(z')| \le 1/2 + c_\phi,$$

and

$$V(z, s(z)e^{-c_\phi-1}) \subset V(z, s(z')e^{-1/2}) \subset V(z', s(z')) \subset V(z, s(z)e^{c_\phi 1}),$$

so $V(z', s(z')) = V(z, s(z))$. Thus the function $z \mapsto V(z, s(z))$ is locally constant at each point of $B_j(c)$, and thus it is constant on each connected component of $B_j(c)$.

Say that $x \in B_j(c)$ and that $x \in g\mathcal{S}$ for some $g \in \Gamma$. We can write $x = [gna]_{\mathcal{E}}$ for some $n \in N^+$, $a = \operatorname{diag}(a_1, \ldots, a_p) \in A^+$. Corollary 4.4 implies that $d_A(a, \phi(x)) \le c_\phi$, and thus $|\log a_i - \phi_i(x)| \le c_\phi$ for all i. In particular,

$$e^{c_V}a_{j+1} < s(x) < e^{-c_V}a_j,$$

so Lemma 4.6 shows that $V(x, s(x)) = Z_j g^{-1}$.

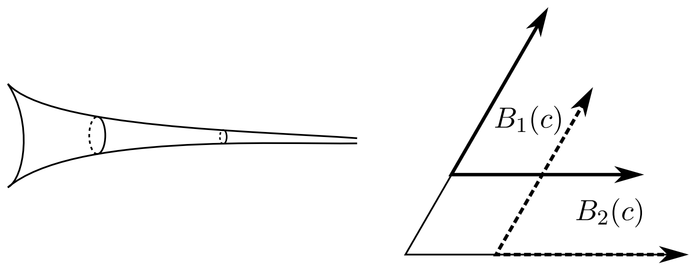

Figure 3. Left: $\mathcal{M}$ for $p = 2$. Right: A^+ for $p = 3$. When $p = 2$, A^+ is 1-dimensional, and the cusp has fundamental group $\mathbb{Z}$, conjugate to a parabolic subgroup. When $p = 3$, A^+ is 2-dimensional, and the cusp is more complicated. The marked regions correspond to $B_1(c)$ (bounded by solid lines) and $B_2(c)$ (bounded by dashed lines). The images in $\mathrm{SL}(3;\mathbb{Z})$ of the fundamental groups of $B_1(c)$ and $B_2(c)$ are parabolic subgroups of $\mathrm{SL}(3;\mathbb{Z})$.

In particular, if $y \in B_j(c)$ is in the same connected component as x and if $y \in h\mathcal{S}$, then $V(x, s(x)) = V(y, s(y))$, so $Z_j g^{-1} = Z_j h^{-1}$, and $g^{-1}h$ stabilizes Z_j. This implies $g^{-1}h \in U(j, p-j)$, as desired. $\square$

Since $r(x) \sim \phi_1(x) - \phi_p(x)$, if $r(x)$ is large, then $x \in B_j(c)$ for some j. As a consequence, if x is deep in the cusp of $\mathcal{M}$, there is a large ball B around x such that $\rho(B)$ is contained in a coset of a maximal parabolic subgroup (see Figure 3).

We claim

Corollary 4.8. *There is a $c' > 0$ such that if $x \in \mathcal{E}$, $r(x) > c'$, and $B \subset \mathcal{E}$ is the ball of radius $\frac{r(x)}{4p^2}$ around x, then $\rho(B) \subset gU(j, p-j)$ for some $g \in \Gamma$ and $1 \le j \le p-1$. Indeed, if $h \in \Gamma$ is such that $h\mathcal{S} \cap B \neq \emptyset$, then $h \in gU(j, p-j)$.*

Proof. We will find a c' such that if $r(x) > c'$, then

$$\frac{r(x)}{4p^2} < \frac{\phi_j(x) - \phi_{j+1}(x) - 2c_\phi - c}{2}$$

for some j, where c is as in Lemma 4.7. If $y \in B$, then x and y are connected by a geodesic segment of length at most $r(x)/4p^2$, and if z is a point on that segment, then

$$|(\phi_j(z) - \phi_{j+1}(z)) - (\phi_j(x) - \phi_{j+1}(x))| \le 2c_\phi + 2\frac{r(x)}{4p^2},$$

so $z \in B_j(c)$, and x and y satisfy the conditions of Lemma 4.7.

Since $\sum_i \phi_i(x) = 0$, we have

$$|\phi_i(x)| \leq p \max_j |\phi_j(x) - \phi_{j+1}(x)|$$

for all j. By Corollary 4.4,

$$r(x) \leq c_\phi + d_{A^+}(I, \phi(x)) \leq c_\phi + p^2 \max_j |\phi_j(x) - \phi_{j+1}(x)|,$$

so there is a j such that

$$|\phi_j(x) - \phi_{j+1}(x)| \geq \frac{r(x) - c_\phi}{p^2}.$$

However, by the definition of A^+, $\phi_j(x) - \phi_{j+1}(x) > \log \epsilon_{\mathcal{S}}$ (see Lemma 4.1), so if $r(x)$ is sufficiently large, then

$$\phi_j(x) - \phi_{j+1}(x) \geq \frac{r(x) - c_\phi}{p^2}.$$

If $r(x)$ is even larger, then

$$\frac{r(x)}{4p^2} < \frac{\phi_j(x) - \phi_{j+1}(x) - 2c_\phi - c}{2},$$

as desired. $\square$

In the next section, we will use this property of $r(x)$ to construct a template.

5. **Reducing to maximal parabolic subgroups**

In this section, we will prove Lemma 3.1 by constructing a disc in $\mathcal{E} = \mathrm{SL}(p;\mathbb{R})/\mathrm{SO}(p)$, a triangulation of that disc, and a template based on that triangulation. The basic idea of the proof is sketched in Section 3.1: any curve in $\mathcal{E}$ can be filled by a Lipschitz disc, which might travel through the thin part of $\mathcal{E}$. We triangulate the disc so that each triangle lies in a single horoball, label the triangulation to get a template, then bound the lengths of the words in the template.

Let $r : \mathcal{M} \to \mathbb{R}$ be the depth function defined in Section 4. We will prove the following

LEMMA 5.1. *Let $q \in \mathbb{Z}$ and $q \geq 2$. If $w = w_1 \cdots w_\ell$ is a word in Γ that represents the identity, then there is a triangulation τ of a square of side length $\sim \ell$ with straight-line edges and a labelling of the vertices of τ by elements of Γ such that the resulting template satisfies*

(1) *If g_1, g_2 are the labels of an edge e in the template, then*

$$d_\Gamma(g_1, g_2) = O(\ell(e)),$$

where $\ell(e)$ is the length of e as a segment in the square.

(2) *There is a* $c > 0$ *independent of* w *such that if* $g_1, g_2, g_3 \in \Gamma$ *are the labels of a triangle in the template, then either* $\operatorname{diam}\{g_1, g_2, g_3\} \leq c$ *or there is a* $1 \leq k < q$ *such that all of the* g_i *are contained in the same coset of* $U(k, q-k)$.

(3) τ *has* $O(\ell^2)$ *triangles, and if the* i*-th triangle of* τ *has vertices labeled* (g_{i1}, g_{i2}, g_{i3}), *then*

$$\sum_i (d_\Gamma(g_{i1}, g_{i2}) + d_\Gamma(g_{i1}, g_{i3}) + d_\Gamma(g_{i2}, g_{i3}))^2 = O(\ell^2).$$

Similarly, if the i*-th edge of* τ *has vertices labeled* h_{i1}, h_{i2}, *then*

$$\sum_i d_\Gamma(h_{i1}, h_{i2})^2 = O(\ell^2).$$

This immediately implies Lemma 3.1.

We construct this template in the way described in Section 3.1. We start with a filling of w by a Lipschitz disc $f : D^2 \to \mathcal{E}$ and then construct a template for w by triangulating the disc and labelling its vertices using ρ. We ensure that properties (1) and (2) hold by carefully controlling the lengths of edges. If edges are too long, then property (2) will not hold. On the other hand, if $x, y \in \mathcal{E}$, then $\rho(x)$ and $\rho(y)$ may be separated by up to $\sim r(x)+r(y)+d_\mathcal{E}(x, y)$, so if edges are too short, then (1) will not hold. For both these conditions, it suffices to construct a triangulation so that the triangle containing x has diameter roughly proportional to the depth function $r(x)$.

We will need the following lemma, which cuts a square of side 2^k into dyadic squares whose side lengths are comparable to a Lipschitz function h; that is, there is a $c > 0$ such that if S is one of the subsquares, with side length $\sigma(S)$, then

$$c^{-1} \min\left\{2^k, \min_{x\in S} h(x)\right\} \leq \sigma(S) \leq c \max_{x\in S} h(x).$$

This is similar to the decomposition used to prove the Whitney extension theorem, which, given a closed set K, decomposes $\mathbb{R}^n \setminus K$ into cubes such that for each cube S, the side length $\sigma(S)$ of S satisfies $\sigma(S) \sim d(S, K)$.

A *dyadic square* is a square of the form

$$S_{i,j,s} := [i2^s, (i+1)2^s] \times [j2^s, (j+1)2^s]$$

for some $i, j, s \in \mathbb{Z}$, $s \geq 0$. We denote the set of dyadic squares contained in $D^2(t)$ by $\mathcal{D}_t$. If S is a square, let $\sigma(S)$ be its side length.

Lemma 5.2. *Let* $t = 2^k$, $k \geq 0$, *let* $D^2(t) = [0, t] \times [0, t]$, *and let* $h : D^2(t) \to \mathbb{R}$ *be a* 1*-Lipschitz function such that* $h(x) \geq 1$ *for all* x. *There is a set of dyadic squares* U *such that*

(1) U *covers* $D^2(t)$, *and any two squares in* U *intersect only along their edges.*

(2) *If* $S \in U$, *then*

$$\min\left\{\frac{h(x)}{6}, \frac{t}{2}\right\} \leq \sigma(S) \leq h(x)$$

for all $x \in S$.

(3) *Each square in* U *neighbors no more than* 16 *other squares.*

Proof. The dyadic squares can be arranged in a rooted tree whose root is $D^2(t)$ so that the children of a dyadic square of side length 2^s, $s > 1$ are the four squares of side length 2^{s-1} which it contains. If S is a dyadic square, let $a(S)$ be its parent square. If S and T are dyadic squares whose interiors intersect, then one must be the ancestor of the other. That is, either $S \subset T$ and $T = a^k(S)$ for some k or vice versa.

Let

$$U_0 := \{S \mid S \in \mathcal{D}_t \text{ and } \sigma(S) \leq h(x) \text{ for all } x \in S\},$$

and let U be the set of maximal elements in U_0:

$$U := \{S \mid S \in U_0 \text{ and } a^k(S) \notin U_0 \text{ for all } k\}.$$

We claim that this is the desired cover.

First, we show that it is a cover of $D^2(t)$. If $x \in D^2(t)$, then $x \in S$ for some $S \in \mathcal{D}_t$ with $\sigma(S) = 1$. Since $h(z) \geq 1$ for all $z \in D^2(t)$, we know that $S \in U_0$. If n is the largest integer such that $a^n(S) \in U_0$, then $a^n(S) \in U$. So x is contained in a square of U, and since x was arbitrary, U is a cover of $D^2(t)$.

Furthermore, if $S, T \in U$ intersect along more than an edge, then one must be an ancestor of the other. Since S and T are maximal elements of U_0, this means that $S = T$.

Next, we prove property (2). By the definition of U_0, if $S \in U$, then $\sigma(S) \leq h(x)$ for all $x \in S$, so it remains to prove the lower bound on $\sigma(S)$. If $S = D^2(t)$, then $\sigma(S) \geq t/2$, so the bound holds; otherwise, if $S \in U$, then $a(S) \notin U_0$ by the definition of U, so there must be some $x_0 \in a(S)$ such that $h(x_0) < 2\sigma(S)$. If $x \in S$, then $d(x, x_0) \leq 4\sigma(S)$, so $h(x) \leq h(x_0) + d(x, x_0) < 6\sigma(S)$, as desired.

Finally, we prove property (3). Suppose that S and T neighbor each other, and let $x \in S \cap T$. By property (2), $\sigma(S) \leq h(x) \leq 6\sigma(T)$ and likewise $\sigma(T) \leq h(x) \leq 6\sigma(S)$. Indeed, since S and T are dyadic squares, we must have $\sigma(T) \leq 4\sigma(S) \leq 16\sigma(T)$, so each square in U can be neighbors with at most four other squares on each side, for a total of 16. □

As a corollary, we obtain

COROLLARY 5.3. *Let* $t = 2^k$, $k \geq 0$, *let* $D^2(t) = [0, t] \times [0, t]$, *and let* $h : D^2(t) \to \mathbb{R}$ *be a* 1*-Lipschitz function such that* $h(x) \geq 1$ *for all* x. *There is a triangulation* τ_h *of* $D^2(t)$ *such that*

(1) *All vertices of τ_h are lattice points, and τ_h contains no more than $2t^2$ triangles.*

(2) *If x and y are connected by an edge of τ_h, then*

$$\min\left\{\frac{h(x)}{6}, \frac{t}{2}\right\} \leq d(x,y) \leq \sqrt{2}h(x).$$

(3) *If we consider $\tau_h^{(2)}$ to be the set of triangles of τ_h, then*

$$\sum_{\Delta \in \tau_h^{(2)}} \operatorname{diam}(\Delta)^2 \leq 64t^2.$$

Likewise, if $\tau_h^{(1)}$ is the set of edges, then

$$\sum_{e \in \tau_h^{(1)}} \ell(e)^2 \leq 128t^2.$$

Proof. Let U be the partition into squares constructed in Lemma 5.2. Two adjacent squares in U need not intersect along an entire edge, so U is generally not a polyhedron. To fix this, we subdivide the edges of each square so that two distinct polygons in U intersect either in a vertex, in an edge, or not at all; call the resulting polyhedron U'. By replacing each n-gon in U' with $n-2$ triangles, we obtain a triangulation, which we denote τ_h. We claim that this τ_h satisfies the required properties.

The first property is clear; the vertices of any dyadic square are lattice points by definition, and the area of any triangle whose vertices are lattice points is at least $1/2$ by Pick's Theorem.

The second property follows from the corresponding property of U.

The third property follows from the fact that the number of neighbors of each square is bounded. Since we divide each edge in U into at most four edges of U', each square S of U corresponds to at most 16 triangles of τ_h, each with diameter at most $2\sigma(S)$. So if $\tau_h^{(2)}$ is the set of triangles of τ_h, then

$$\sum_{\Delta \in \tau_h^{(2)}} \operatorname{diam}(\Delta)^2 \leq \sum_{S \in U} 16(2\sigma(S))^2.$$

Since $\sum_{S \in U} \sigma(S)^2 = \operatorname{area} D^2(t)$, this is at most $64t^2$. Likewise, if $\tau_h^{(1)}$ is the set of edges, then

$$\sum_{e \in \tau_h^{(1)}} \ell(e)^2 \leq \sum_{S \in U} 32(2\sigma(S))^2 = 128t^2. \quad \square$$

We use this lemma to prove Lemma 5.1 by letting $h(x) \sim r(x)$.

Proof of Lemma 5.1. Let $w(i) = w_1 \cdots w_i$. Let $\alpha : [0, \ell] \to \mathcal{E}$ be the curve corresponding to w, parametrized so that $\alpha(i) = [w(i)]_{\mathcal{E}}$. If c_Σ is the maximum length of a curve corresponding to a generator, then α is c_Σ-Lipschitz. Let $t = 2^k$ be the smallest power of 2 larger than ℓ, and let $\alpha' : [0,t] \to \mathcal{E}$:

$$\alpha'(x) = \begin{cases} \alpha(x) & \text{if } x \leq \ell, \\ [I]_{\mathcal{E}} & \text{otherwise.} \end{cases}$$

Since $\mathcal{E}$ is nonpositively curved, we can use geodesics to fill α'. If $x, y \in \mathcal{E}$, let $\gamma_{x,y} : [0,1] \to \mathcal{E}$ be a geodesic parametrized so that $\gamma_{x,y}(0) = x$, $\gamma_{x,y}(1) = y$, and $\gamma_{x,y}$ has constant speed. We can define a homotopy $f : [0,t] \times [0,t] \to \mathcal{E}$ by

$$f(x,y) = \gamma_{\alpha'(x),\alpha'(0)}(y/t);$$

this sends three sides of $D := [0,t] \times [0,t]$ to $[I]_{\mathcal{E}}$ and is a filling of α. Since $\mathcal{E}$ is nonpositively curved, this map is $2c_\Sigma$-Lipschitz and has area $O(\ell^2)$.

Let $h : D \to \mathbb{R}$,

$$h(x) = \max\left\{1, \frac{r(f(x))}{16p^2 c_\Sigma}\right\}$$

This function is 1-Lipschitz. If $h(x)$ is sufficiently large and $B \in D$ is a disc of radius $2h(x)$ around x, then $f(B)$ is contained in a ball of radius $r(f(x))/(4p^2)$ around $f(x)$. By Corollary 4.8, $\rho(f(B))$ is contained in a coset of a maximal parabolic subgroup. Let τ_h be the triangulation of D constructed in Corollary 5.3.

If v is an interior vertex of τ_h, label it $\rho(f(v))$. If $(i,0)$ is a boundary vertex on the side of D corresponding to α' and $i \leq \ell$, label it by $w(i)$. Label all the rest of the boundary vertices by I. Note that for all vertices v, if g is the label of v, then $f(v) \in g\mathcal{S}$.

If x is a lattice point on the boundary of D, then $f(x) = [I]_{\mathcal{M}}$ and so $h(x) = 1$. In particular, each lattice point on the boundary of D is a vertex of τ_h, so the boundary of τ_h is a $4t$-gon with vertices labeled $I, w(1), \dots, w(n-1), I, \dots, I$. We identify vertices labeled I and remove self-edges to get a template τ for w.

First, property (1) follows from Lemma 4.5. That is, if v_1 and v_2 are the endpoints of an edge of τ, labeled by g_1 and g_2, then $d(v_1, v_2) \sim r(f(v_1)) \sim r(f(v_2))$, so by Lemma 4.5,

$$d_\Gamma(g_1, g_2) = O(d(v_1,v_2) + r(f(v_1)) + r(f(v_2))) = O(d(v_1,v_2)),$$

as desired.

Second, note that if x_1, x_2, and x_3 are the vertices of a triangle of τ, with labels g_1, g_2, and g_3, then Corollary 5.3 implies that $\operatorname{diam}\{x_1,x_2,x_3\} \leq 2h(x_1)$. If $h(x_1)$ is sufficiently large, then Corollary 4.8 shows that g_1, g_2, and g_3 are in the same coset of $U(j, p-j)$ for some j; otherwise, by property (1), g_1, g_2, and g_3 must be within bounded distance of one another.

Finally, property (3) follows from property (1) and the corresponding property of τ_h. □

Note that it is not necessary that $q \geq 5$ for this template to exist. In fact, a suitable generalization of the proposition should hold for any lattice in a semisimple Lie group.

6. **Shortcuts and normal forms**

As before, we let $\Gamma = \mathrm{SL}(p;\mathbb{Z})$, with the generating set Σ consisting of the unit transvections $e_{ij} = e_{ij}(1)$ and the diagonal matrices. In this section, we will define a normal form $\omega : \Gamma \to \Sigma^*$ that associates each element of Γ with a word in Γ that represents it. This normal form will use short representatives of unipotent elements like those constructed by Lubotzky, Mozes, and Raghunathan [LMR93].

6.1. *Shortcuts.* Recall that $e_{ij}(x)$, $i \neq j$ represents the matrix obtained from the identity matrix by replacing the (i,j)-entry with x. Lubotzky, Mozes, and Raghunathan noted that when $p \geq 3$, this group element can be represented by a word $\hat{e}_{ij}(x)$ of length $\sim \log|x|$, which we call a *shortcut*. Since the particular generators used to construct a shortcut will be important later on, we will define many different ways to shorten a given transvection: If $S \subset \{1,\dots,p\}$ is a set such that $i \in S$, $j \notin S$, and $\#S \geq 2$, then $\hat{e}_{ij;S}(x)$ will be a shortcut for $e_{ij}(x)$, which is a product of unit transvections lying in the parabolic subgroup $U(S,\{j\})$. More generally, recall that if $S,T \subset \{1,\dots,p\}$ are disjoint and $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, then $u(V)$ represents the unipotent matrix in $U(S,T)$ corresponding to V. If $\#S \geq 2$, we will define a curve $\hat{u}_S(V)$ that lies in a thick part of G, goes from I to $u(V)$, and has length $O(\log\|V\|_2)$.

We will provide a condensed version of the constructions of $\hat{e}_{ij;S}(x)$ and $\hat{u}_S(V)$; for more details, see [LMR93] or [Ril05]. We start by defining a solvable subgroup $H_{S,T} \subset U(S,T)$ for each pair of disjoint sets $S,T \subset \{1,\dots,p\}$. Without loss of generality, we may take $S=\{1,\dots,s\}$ and $T=\{s+1,\dots,s+t\}$. Let A and B be $\mathbb{R}$-split, $\mathbb{Q}$-anisotropic tori in $\mathrm{SL}(S)$ and $\mathrm{SL}(T)$ respectively; their integer points are isomorphic to $\mathbb{Z}^{s-1}$ and $\mathbb{Z}^{t-1}$ respectively. Let

$$H_{S,T} := (A \times B) \ltimes (\mathbb{R}^S \otimes \mathbb{R}^T),$$

where A acts on $\mathbb{R}^S$ on the left and B acts on $\mathbb{R}^T$ on the right. Without loss of generality, we may take $S = \{1,\dots,s\}$ and $T = \{s+1,\dots,s+t\}$ and write

$$H_{S,T} = \left\{ \begin{pmatrix} M & V & 0 \\ 0 & N & 0 \\ 0 & 0 & I \end{pmatrix} \middle| M \in A, N \in B, V \in \mathbb{R}^S \otimes \mathbb{R}^T \right\}.$$

Note that the integer points of $H_{S,T}$ form a cocompact lattice in $H_{S,T}$, so $H_{S,T}$ lies in a thick part of $\mathrm{SL}(p)$. We may conjugate $H_{S,T}$ so that A and B become the subgroups of diagonal matrices with positive coefficients; it follows that $\mathbb{R}^S \otimes \mathbb{R}^T$ is exponentially distorted in $H_{S,T}$ as long as $\#S \geq 2$ or $\#T \geq 2$.

Thus, given any $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, there is a curve in $H_{S,T}$ from I to $u(V)$ of length $\sim \log \|V\|_2$. Note that this curve may depend on S and T as well as V. The following construction removes the dependence on T: If $\#S \geq 2$, S^c is the complement of S, $V \in \mathbb{R}^S \otimes \mathbb{R}^{S^c}$, and $\{z_1, \dots, z_p\}$ is the standard basis of $\mathbb{R}^p$, we can write $V = \sum_{i \in S^c} v_i \otimes z_i$ for some vectors $v_i \in \mathbb{R}^S$. For all $i \in S^c$, define $u_i : [0,1] \to H_{S,S^c}$ to be a geodesic in $H_{S,\{i\}}$ that connects I to $u(v_i \otimes z_i)$ and define

$$\hat{u}_S(V) = \prod_{i \in S^c} u_i.$$

This is a curve connecting I to $u(V)$ that has length $O(\log \|V\|_2)$. Furthermore, if $T \subset \{1, p\}$ is such that $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, then $\hat{u}_S(V)$ is a curve in $H_{S,T}$. We think of it as the result of using a torus in SL(S) to "compress" $u(V)$.

If $\#S \geq 2$, $i \in S$, $j \notin S$, and $x \in \mathbb{Z}$, then $\hat{u}_S(xz_i \otimes z_j)$ is a curve in $H_{S,\{j\}} \subset U(S, \{j\})$ which connects I to $e_{ij}(x)$. Let $\hat{e}_{ij;S}(x)$ be a word in $U(S, \{j\})$ approximating $\hat{u}_S(xz_i \otimes z_j)$. Since $H_{S,\{j\}}$ lies in a thick part of $U(S, \{j\})$, the length of $\hat{e}_{ij;S}(x)$ is comparable to the length of $\hat{u}_S(xz_i \otimes z_j)$.

In many cases, the precise value of S does not matter, so for each pair (i,j), we choose a d_{ij} such that $d_{ij} \notin \{i,j\}$ and define $\hat{e}_{ij}(x) = \hat{e}_{ij;\{i,d_{ij}\}}(x)$. As a special case, for all i,j, we set $\hat{e}_{ij}(\pm 1) = e_{ij}^{\pm 1}$.

6.2. *A normal form.* Recall that if $H \subset \Gamma$, we say that w is a *shortcut word* in H if we can write $w = \prod_{i=1}^n w_i$, where each w_i is either a diagonal matrix in H or a shortcut $\hat{e}_{a_i b_i}(x_i)$ where $e_{a_i b_i}(x_i) \in H$. Note that any word in H is automatically a shortcut word, but not vice versa, since a shortcut for an element of H may use generators that are not in H.

We claim

Lemma 6.1. *There is a normal form $\omega : \Gamma \to \Sigma^*$ such that*

(1) *For all $g \in \Gamma$, $\ell(\omega(g)) = O(d_\Gamma(I, g))$.*
(2) *For all $i, j \in \{1, \dots, p\}$ with $i \neq j$ and $x \in \mathbb{Z}$,*

$$\omega(e_{ij}(x)) = \hat{e}_{ij}(x).$$

(3) *If $g \in P$ where $P = U(S_1, \dots, S_k)$ is a group of block upper-triangular matrices, then $\omega(g)$ is a product of a bounded number of shortcut words in the diagonal blocks of P, a bounded number of shortcuts corresponding to off-diagonal entries of P, and possibly one diagonal matrix.*

Proof. If $g = e_{ij}(x)$, we define $\omega(g) = \hat{e}_{ij}(x)$; this satisfies all three conditions.

Otherwise, let $g \in \Gamma$ and let $P = U(S_1, \dots, S_k) \in \mathcal{P}$ be the unique minimal $P \in \mathcal{P}$ containing g. Then g is a block-upper-triangular matrix that can be written as a product

$$g = \begin{pmatrix} m_1 & V_{12} & \cdots & V_{1k} \\ 0 & m_2 & \cdots & V_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_k \end{pmatrix} d, \tag{3}$$

where the i-th block of the matrix corresponds to S_i. Here, $V_{i,j} \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$, and $d \in D$ is a diagonal matrix chosen so that $\det m_i = 1$. If $P = \Gamma$, then there is only one block, and we take $m_1 = g$, and $d = I$. We can write g as a product:

$$\gamma_i := \left(\prod_{j=1}^{i-1} u(V_{ji})\right) m_i,$$
$$g = \gamma_k \cdots \gamma_1 d.$$

We will construct $\omega(g)$ by replacing the terms in this product decomposition with shortcut words.

First, consider the m_k. When $\#S_k \geq 3$, we can use Theorem 2.2 to replace m_k by a word in $\mathrm{SL}(S_k;\mathbb{Z})$, but the theorem does not apply when $\#S_k = 2$ because $\mathrm{SL}(2;\mathbb{Z})$ is exponentially distorted inside $\mathrm{SL}(p;\mathbb{Z})$. We thus use a variant of the Lubotzky-Mozes-Raghunathan theorem to write m_k as a shortcut word $\mathrm{SL}(2;\mathbb{Z})$.

Proposition 6.2 (cf. [LMR93]). *There is a constant c such that for all $g \in \mathrm{SL}(k;\mathbb{Z})$, there is a shortcut word in $\mathrm{SL}(k;\mathbb{Z})$ that represents g and has length*

$$\ell(w) \leq c \log \|g\|_2.$$

For $i = 1, \ldots, k$, let $\hat{m}_i$ be a shortcut word representing m_i as in Proposition 6.2. For $1 \leq i < j \leq k$ and $V \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$, let

$$\hat{n}(V) := \prod_{a \in S_i, b \in S_j} \hat{e}_{ab}(x_{ab}),$$

where x_{ab} is the (a,b)-coefficient of V; this is a shortcut word representing $u(V)$. Let

$$\hat{\gamma}_i := \left(\prod_{j=1}^{i-1} \hat{n}(V_{ji})\right) \hat{m}_i,$$
$$\omega(g) = \hat{\gamma}_k \cdots \hat{\gamma}_1 d.$$

This is a word in $\mathrm{SL}(p;\mathbb{Z})$ that represents g. It is straightforward to show that there is a constant c_ω independent of g such that $\ell(\omega(g)) \leq c_\omega d_\Gamma(I,g)$.

Furthermore, if $Q = U(T_1, \ldots, T_r)$ is a block upper-triangular subgroup and $g \in Q$, then $Q \supset P$ and for every i, there is a j such that $S_i \subset T_j$. In particular, $\hat{m}_i$ is a shortcut word in $\mathrm{SL}(T_j;\mathbb{Z})$, and each shortcut making up $\hat{V}_{ji}$ is either contained in $\mathrm{SL}(T_a)$ for some a or corresponds to an off-diagonal entry of Q. Since there are a bounded number of $\hat{m}_i$ and a bounded number of terms in the $\hat{V}_{ji}$, this normal form satisfies property (3). □

6.3. *Summary of notation.* For quick reference, we summarize the above constructions, which will be used throughout the rest of the paper.

- $u(V)$ is the unipotent element in $U(S,T)$ corresponding to V.
- $H_{S,T}$ is a solvable subgroup of $U(S,T)$, isomorphic to
$$(\mathbb{R}^{\#S-1}\times\mathbb{R}^{\#T-1})\ltimes(\mathbb{R}^S\otimes\mathbb{R}^T).$$
- $\hat{u}_S(V)$ is a curve in $H_{S,T}$ that connects I and $u(V)$.
- If $i\in S$ and $j\notin S$, then $\hat{e}_{ij;S}(x)$ is a word in $U(S,\{j\})$ that represents $e_{ij}(x)$.
- For each $1\le i\ne j\le p$, we choose some $d_{ij}\notin\{i,j\}$ and define $\hat{e}_{ij}(x):=\hat{e}_{ij;\{i,d_{ij}\}}(x)$.

7. Manipulating shortcuts

In Section 6, we constructed shortcuts $\hat{e}_{ij;S}(x)$; that is, words of logarithmic length that represent transvections $e_{ij}(x)$. The normal form ω is built using products of these shortcuts, and in this section, we will develop ways to manipulate such products.

7.1. *Moving shortcuts between solvable groups.* One of the main ideas behind these tools is that when p is large, we can construct shortcuts for $e_{ij}(x)$ that lie in small subgroups of Γ and we can construct quadratic-area homotopies from one to another.

This subsection is devoted to proving that when p is large, shortcuts that come from different solvable subgroups can be connected by quadratic-area homotopies. We will prove the following lemma.

LEMMA 7.1. *If $p\ge 5$ and if $S\subset\{1,\dots,p\}$ is such that $2\le\#S\le p-2$, $i\in S$, and $j\notin S$, then*
$$\delta_\Gamma(\hat{e}_{ij}(x),\hat{e}_{ij;S}(x))=O((\log|x|)^2).$$

We will use a special case of a theorem of Leuzinger and Pittet on Dehn functions of solvable groups. Recall that the curves $\hat{u}_S(V)$ used to define the $\hat{e}_{ij}$'s lie in solvable subgroups of the form
$$H_{S,T}=(A\times B)\ltimes(\mathbb{R}^S\otimes\mathbb{R}^T),$$
where A and B are $\mathbb{R}$-split, $\mathbb{Q}$-anisotropic tori in $\mathrm{SL}(S)$ and $\mathrm{SL}(T)$ respectively. These subgroups are contained in the thick part of G, and when either S or

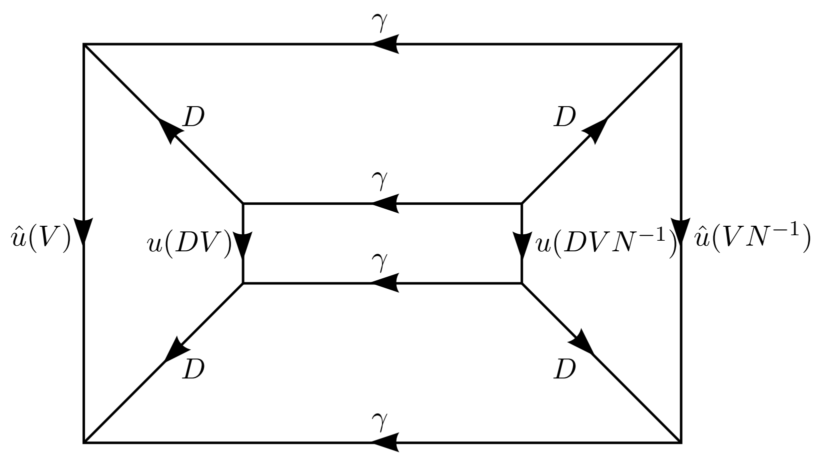

Figure 4. A filling of $\omega = \gamma\hat{u}_S(V)\gamma^{-1}\hat{u}_S(VN^{-1})^{-1}$.

T is large enough, results of Leuzinger and Pittet [LP04] imply that $H_{S,T}$ has quadratic Dehn function (see also [dCT10]).

THEOREM 7.2. *If $s = \#S \geq 3$ or $t = \#T \geq 3$, then $H_{S,T}$ has a quadratic Dehn function.*

We use manipulations in $H_{S,T}$ to prove the following lemma.

LEMMA 7.3. *Let $S \subset \{1, \dots, p\}$ be such that $2 \leq \#S \leq p-2$, and let $T = S^c$ be the complement of S. Let $0 < \epsilon < 1/2$ be sufficiently small that $H_{S,T} \subset G(\epsilon)$. If γ is a curve in the ϵ-thick part of $\mathrm{SL}(S) \times \mathrm{SL}(T)$ that connects (I, I) to (M, N), and if $V \in \mathbb{R}^S \otimes \mathbb{R}^T$, then*

$$\delta_{G(\epsilon')}(\gamma\hat{u}_S(V)\gamma^{-1}, \hat{u}_S(MVN^{-1})) = O((\ell(\gamma) + \log(\|V\|_2 + 2))^2),$$

where ϵ' is independent of γ and V.

Proof. Let $s = \#S$ and $t = \#T$. Let A and B be the tori used to define $H_{S,T}$, and let Let $\{v_1, \dots, v_s\} \subset \mathbb{R}^S$ and $\{w_1, \dots, w_t\} \subset \mathbb{R}^T$ be the corresponding eigenbases of $\mathbb{R}^S$ and $\mathbb{R}^T$.

We first consider the case that γ is a curve in $\mathrm{SL}(T)$ and that $V = xv_i \otimes w_j$. In this case, $M = I$ and we want to fill the curve

$$\omega := \gamma\hat{u}_S(V)\gamma^{-1}\hat{u}_S(VN^{-1})^{-1}.$$

Let $\ell = \ell(\omega)$. Let $\delta = \exp(-\ell(\gamma))$, and let $D \in A$ be such that $\|xDv_i\|_2 \leq \delta$; we can choose D so that $d_A(I, D) = O(\ell)$. If we conjugate ω by D, it becomes easy to fill. We will fill ω with a disc of the form shown in Figure 4.

This disc is comprised of four trapezoids and a central "thin rectangle." Each of the edges labeled D corresponds to a translate of the geodesic in A

that connects I to D, and each edge has length at most $O(\ell)$. Each trapezoid can be filled with quadratic area. The left and right trapezoids are contained in $H_{S,T}$, so they have quadratic filling area in $H_{S,T}$; furthermore, since $H_{S,T}$ lies in the thick part of G, the filling stays in the thick part. The top and bottom trapezoids each represent a commutator of a curve in $\mathrm{SL}(S)$ and a curve in $\mathrm{SL}(T)$ and so can be filled by the rectangle resulting from the product of those curves. Each curve stays in some thick part of $\mathrm{SL}(S)$ or $\mathrm{SL}(T)$, so the rectangle does as well.

The central thin rectangle can be filled by a disc of area $O(\ell)$. Call the edges labeled by γ the "long edges" of the rectangle. Since δ is small, these long edges synchronously fellow travel, and since they lie in the thick part, they can be filled by a disc in the thick part of area $O(\ell)$. This gives a quadratic filling of ω.

The same technique works if instead we have $\gamma : [0,1] \to \mathrm{SL}(S)$ and $V = xv_i \otimes w_j$. The main change is that D is now a matrix in B such that $\|xw_j D\|_2 \leq \delta$.

Now suppose γ is a curve in $\mathrm{SL}(S) \times \mathrm{SL}(T)$. It can be homotoped to a concatenation of curves $\gamma = \gamma_S\gamma_T$, where γ_S and γ_T are the projections of γ to each factor. This homotopy can be taken to have quadratic area and lie in the thick part of G, and the lemma can be applied to γ_S and γ_T separately. This proves the lemma in the case that $V = xv_i \otimes w_j$.

In general, we can decompose V as a sum of eigenvectors $V = \sum_{i,j} x_{ij} v_i \otimes w_j$, so we will use the quadratic Dehn function of $H_{S,T}$ to break $\hat{u}_S(V)$ up into pieces corresponding to each eigenvector and apply the lemma to each piece. We can construct a homotopy from $\gamma\hat{u}_S(V)\gamma^{-1}$ to $\hat{u}_S(MVN^{-1})$ that goes through the following stages:

$$\begin{aligned}
&\gamma\hat{u}_S(V)\gamma^{-1}, \\
&\gamma\Bigg(\prod_{i,j} \hat{u}_S(x_{ij}v_i \otimes w_j)\Bigg)\gamma^{-1} && \text{by Theorem 7.2,} \\
&\prod_{i,j} \gamma\hat{u}_S(x_{ij}v_i \otimes w_j)\gamma^{-1} && \text{by free insertions,} \\
&\prod_{i,j} \hat{u}_S(M(x_{ij}v_i \otimes w_j)N^{-1}) && \text{by the arguments above,} \\
&\hat{u}_S(MVN^{-1}) && \text{by Theorem 7.2.}
\end{aligned}$$

Each stage has quadratic area, so the homotopy as a whole has quadratic area. ∎

Let $\{z_1, \ldots, z_p\}$ be the standard basis of $\mathbb{R}^p$. Lemma 7.1 then follows from the following lemma.

Lemma 7.4. *There is an $0 < \epsilon < 1/2$ such that if $i \in S, S'$ and $j \notin S \cup S'$, where $2 \leq \#S, \#S' \leq p-2$, and if $x \in \mathbb{R}$, then*

$$\delta_{G(\epsilon)}(\hat{u}_S(xz_i \otimes z_j), \hat{u}_{S'}(xz_i \otimes z_j)) = O((\log|x|)^2).$$

In particular,

$$\delta_\Gamma(\hat{e}_{ij}(x), \hat{e}_{ij;S}(x)) = O((\log|x|)^2).$$

Proof. First, consider the case that $S \subset S'$. Let $T = S^c$ be the complement of S and $T' = (S')^c$ be the complement of S'. Then $\hat{u}_S(xz_i \otimes z_j)$ is a curve in $H_{S,T}$, and $H_{S,T}$ and $H_{S',T'}$ both have quadratic Dehn functions. Let $s = \#S$, $t = \#T$.

Recall that $A = \mathbb{R}^{s-1} \subset H_{S,T}$ is an $\mathbb{R}$-split, $\mathbb{Q}$-anisotropic torus in $\mathrm{SL}(S)$; let $A(\mathbb{Z})$ be the integer points of A. We can decompose xz_i as a sum of eigenvectors $xz_i = \sum_k v_k$ and "compress" each term in this sum using A. That is, there are vectors $y_k \in \mathbb{R}^S$ such that $\|y_k\|_2 \leq 1$ and elements $A_k \in A(\mathbb{Z})$ such that $v_k = A_k y_k$ and $d_A(I, A_k) = O(\log|x|)$. Then

$$e_{ij}(x) = \prod_k A_k u(y_k \otimes z_j) A_k^{-1}.$$

Let γ_k, $k = 1, \dots, s$ be a geodesic in A that connects I to A_k and has length $O(\log|x|)$. Let $\mathcal{U}_k : [0,1] \to G$ be the curve $\mathcal{U}_k(t) = u(ty_k \otimes z_j)$. We can then construct a curve

$$\omega = \prod_k \gamma_k \mathcal{U}_k \gamma_k^{-1}$$

that connects I to $e_{ij}(x)$.

We can use ω as an intermediate stage in a homotopy between $\hat{u}_S(xz_i \otimes z_j)$ and $\hat{u}_{S'}(xz_i \otimes z_j)$. On one hand, ω lies in $H_{S,T}$ and has length $O(\log|x|)$, so there is a quadratic-area homotopy from $\hat{u}_S(xz_i \otimes z_j)$ to ω. On the other hand, since γ_k lies in a thick part of $\mathrm{SL}(S)$, ω also lies in a thick part of $\mathrm{SL}(S')$, and we can apply Lemma 7.3 to each term of ω to construct a homotopy from ω to

$$\omega' = \prod_k \hat{u}_{S'}(v_k \otimes z_j).$$

This is a curve in $H_{S',T'}$ of length $O(\log|x|)$, so there is a quadratic-area homotopy from ω' to $\hat{u}_{S'}(xz_i \otimes z_j)$. Concatenating these homotopies produces a homotopy from $\hat{u}_S(xz_i \otimes z_j)$ to $\hat{u}_{S'}(xz_i \otimes z_j)$, as desired.

For the general case, let $k \in S$ and $k' \in S'$ be such that $i \neq k$, $i \neq k'$. Then, if $V = xz_i \otimes z_j$, we can use the argument above to construct a homotopy from $\hat{u}_S(V)$ to $\hat{u}_{S'}(V)$ that goes through the stages

$$\hat{u}_S(V) \to \hat{u}_{\{i,k\}}(V) \to \hat{u}_{\{i,k,k'\}}(V) \to \hat{u}_{\{i,k'\}}(V) \to \hat{u}_{S'}(V).$$

(If $k = k'$, the middle stages can be omitted.)

Since Γ acts geometrically on $G(\epsilon)$ and $\hat{e}_{ij;S}(x)$ is an approximation of $\hat{u}_S(xz_i \otimes z_j)$, this implies that

$$\delta_\Gamma(\hat{e}_{ij;S}(x), \hat{e}_{ij;S'}(x)) = O((\log|x|)^2)$$

and, in particular,

$$\delta_\Gamma(\hat{e}_{ij}(x), \hat{e}_{ij;S}(x)) = O((\log|x|)^2). \quad \square$$

We also note the following corollary, which will be useful later.

Corollary 7.5. *If $S, S', T \subset \{1, \dots, p\}$ are such that $2 \le \#S, \#S' \le p-2$, $S \cap T = \emptyset$, and $S' \cap T = \emptyset$, and if $V \in \mathbb{R}^{S\cap S'} \otimes \mathbb{R}^T$, then there is an $0 < \epsilon < 1/2$ such that*

$$\delta_{G(\epsilon)}(\hat{u}_S(V), \hat{u}_{S'}(V)) = O((\log\|V\|_2)^2).$$

Proof. For all $i \in S \cap S'$ and $j \in T$, let v_{ij} be the coefficient of V in the (i,j)-position. Let

$$\omega = \prod_{i,j} \hat{u}_S(v_{ij} z_i \otimes z_j),$$
$$\omega' = \prod_{i,j} \hat{u}_{S'}(v_{ij} z_i \otimes z_j).$$

Then ω is a curve in H_{S,S^c} with length $O(\log\|V\|_2)$, so there is a quadratic-area homotopy from $\hat{u}_S(V)$ to ω and likewise from ω' to $\hat{u}_{S'}(V)$. By Lemma 7.4, there is a quadratic-area homotopy from ω to ω'. Combining these homotopies proves the corollary. $\square$

Lemma 3.4 is then a corollary of Lemma 7.1.

Proof of Lemma 3.4. Suppose that w is a shortcut word in SL($q;\mathbb{Z}$) and $q \ge 5$. We can write $w = \prod_{i=1}^n w_i$, where each w_i is either a diagonal matrix in H or a shortcut $\hat{e}_{a_ib_i}(x_i)$ where $e_{a_ib_i}(x_i) \in \mathrm{SL}(q;\mathbb{Z})$. Since $q \ge 3$, we can use Lemma 7.1 to replace each shortcut $\hat{e}_{a_ib_i}(x_i)$ by a shortcut $\hat{e}_{a_ib_i;S_i}(x_i)$ that lies in SL($q;\mathbb{Z}$) at a total cost of order $O(\ell(w)^2)$. The result is a word w' in SL($q;\mathbb{Z}$) of length $O(\ell(w))$, and $\delta(w,w') = O(\ell(w)^2)$, as desired. $\square$

7.2. *The shortened Steinberg presentation.* The Steinberg presentation gives relations between products of elementary matrices; in this section, we will develop ways to manipulate the corresponding shortcut words.

This subsection is devoted to building an analogue of the Steinberg presentation for shortcut words. We will prove the following lemma.

Lemma 7.6 (The shortened Steinberg presentation). *If $x, y \in \mathbb{Z} \setminus \{0\}$, then*

(1) *If* $1 \leq i, j \leq p$ *and* $i \neq j$*, then*
$$\delta_\Gamma(\hat{e}_{ij}(x)\hat{e}_{ij}(y), \hat{e}_{ij}(x+y)) = O((\log|x| + \log|y|)^2).$$
In particular,
$$\delta_\Gamma(\hat{e}_{ij}(x)\hat{e}_{ij}(-x)) = \delta_\Gamma(\hat{e}_{ij}(x)^{-1}, \hat{e}_{ij}(-x)) = O((\log|x|)^2).$$
(2) *If* $1 \leq i, j, k \leq p$ *and* $i \neq j \neq k$*, then*
$$\delta_\Gamma([\hat{e}_{ij}(x), \hat{e}_{jk}(y)], \hat{e}_{ik}(xy)) = O((\log|x| + \log|y|)^2).$$
(3) *If* $1 \leq i, j, k, l \leq p$*,* $i \neq l$*, and* $j \neq k$
$$\delta_\Gamma([\hat{e}_{ij}(x), \hat{e}_{kl}(y)]) = O((\log|x| + \log|y|)^2).$$
(4) *Let* $1 \leq i, j, k, l \leq p$*,* $i \neq j$*, and* $k \neq l$*, and*
$$s_{ij} = e_{ji}^{-1} e_{ij} e_{ji}^{-1},$$
so that s_{ij} *represents*
$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathrm{SL}(\{i,j\}; \mathbb{Z}).$$
Then
$$\delta_\Gamma(s_{ij}\hat{e}_{kl}(x)s_{ij}^{-1}, \hat{e}_{\sigma(k)\sigma(l)}(\tau(k,l)x)) = O((\log|x| + \log|y|)^2),$$
where σ *is the permutation switching* i *and* j*, and* $\tau(k,l) = -1$ *if* $k = i$ *or* $l = i$ *and* 1 *otherwise.*

(5) *If* $b = \mathrm{diag}(b_1, \dots, b_p)$*, then*
$$\delta_\Gamma(b\hat{e}_{ij}(x)b^{-1}, \hat{e}_{ij}(b_i b_j x)) = O(\log|x|^2).$$

Proof. Part (1) follows from Theorem 7.2. Recall that there is a d_{ij} such that $\hat{e}_{ij}(x) = \hat{e}_{ij;\{i,d_{ij}\}}(x)$. Let $S = \{i, d_{ij}\}$. Then $\hat{e}_{ij}(x)\hat{e}_{ij}(y)\hat{e}_{ij}(x+y)^{-1}$ is an approximation of a closed curve in H_{S,S^c} of length $O((\log|x| + \log|y|)^2)$. Since H_{S,S^c} has quadratic Dehn function, it can be filled by a quadratic-area disc.

The rest of the parts of Lemma 7.6 involve conjugating a shortcut by a word. Most of the proofs below follow the same basic outline. To conjugate $\hat{e}_{ij}(x)$ by a word w representing a matrix M, first choose S such that $M \in \mathrm{SL}(S)$, where $i \in S$ and $j \notin S$. Next, replace w with a word w' in $\mathrm{SL}(S)$ and replace $\hat{e}_{ij}(x)$ by $\hat{e}_{ij;S}(x)$. Finally, use Lemma 7.3 to conjugate $\hat{e}_{ij;S}(x)$ by w'.

Part (2): Let $d \notin \{i, j, k\}$ and $S = \{i, j, d\}$, so that $\hat{e}_{ij;\{i,d\}}(x)$ is a word in $\mathrm{SL}(S;\mathbb{Z})$. We construct a homotopy going through the stages
$$\omega_0 = [\hat{e}_{ij}(x), \hat{e}_{jk}(y)]\hat{e}_{ik}(xy)^{-1},$$
$$\omega_1 = [\hat{e}_{ij;\{i,d\}}(x), \hat{u}_S(yz_j \otimes z_k)]\hat{e}_{ik;S}(xy)^{-1},$$
$$\omega_2 = \hat{u}_S((xyz_i + yz_j) \otimes z_k)\hat{u}_S(yz_j \otimes z_k)^{-1}\hat{u}_S(xyz_i \otimes z_k)^{-1}.$$

Here, we use Lemma 7.1 to construct a homotopy between ω_0 and ω_1. We apply Lemma 7.3 with $\gamma = \hat{e}_{ij;\{i,d\},\{j\}}(x)$ and $V = yz_j \otimes z_k$ to construct a homotopy between ω_1 and ω_2. Finally, ω_2 is a curve in H_{S,S^c} with length $O(\log|x| + \log|y|)$ and thus has filling area $O((\log|x| + \log|y|)^2)$. The total area used is $O((\log|x| + \log|y|)^2)$.

Part (3): Let $S = \{i,k,d\}$. We use the same techniques to construct a homotopy going through the stages

$$\begin{aligned}
&[\hat{e}_{ij}(x), \hat{e}_{kl}(y)], && \\
&[\hat{e}_{ij;S}(x), \hat{e}_{kl;S}(y)] && \text{by Lemma 7.1,} \\
&\varepsilon && \text{by Theorem 7.2 applied to } H_{S,S^c},
\end{aligned}$$

where ε represents the empty word. This homotopy has area $O((\log|x| + \log|y|)^2)$.

Part (4): We consider several cases depending on k and l. When i, j, k, and l are distinct, the result follows from part (3), since $s_{ij} = e_{ji}^{-1} e_{ij} e_{ji}^{-1}$, and we can use part (3) to commute each letter past $\hat{e}_{kl}(x)$. If $k = i$ and $l \neq j$, let $d \notin \{i,j,l\}$, and let $S = \{i,j,d\}$. There is a homotopy from

$$s_{ij} \hat{e}_{il}(x) s_{ij}^{-1} \hat{e}_{jl}(-x)^{-1}$$

to

$$s_{ij} \hat{u}_S(xz_i \otimes z_l) s_{ij}^{-1} \hat{e}_{jl}(xz_j \otimes z_l)$$

of area $O((\log|x|)^2)$, and since s_{ij} is a word in SL$(S;\mathbb{Z})$, the proposition follows by an application of Lemma 7.3. A similar argument applies to the cases $k = j$ and $l \neq i$; $k \neq i$ and $l = j$; and $k \neq j$ and $l = i$.

If $i = k$ and $j = l$, let $d \notin \{i,j\}$. There is a homotopy going through the stages

$$\begin{aligned}
&s_{ij} \hat{e}_{ij}(x) s_{ij}^{-1}, && \\
&s_{ij} [e_{id}, \hat{e}_{dj}(x)] s_{ij}^{-1} && \text{by part (2),} \\
&[s_{ij} e_{id} s_{ij}^{-1}, s_{ij} \hat{e}_{dj}(x) s_{ij}^{-1}] && \text{by free insertion,} \\
&[e_{jd}^{-1}, \hat{e}_{di}(x)] && \text{by previous cases,} \\
&\hat{e}_{ji}(-x) && \text{by part (2),}
\end{aligned}$$

and this homotopy has area $O((\log|x|)^2)$. One can treat the case that $i = l$ and $j = k$ the same way.

Since any diagonal matrix in Γ is the product of at most p elements s_{ij}, part (5) follows from part (4). $\square$

8. **Reducing to diagonal blocks**

In this section, we work to prove Lemma 3.3, which claims that we can break an ω-triangle in a block upper-triangular subgroup $U(S_1,\dots,S_k) \subset \mathrm{SL}(p;\mathbb{Z})$ into shortcut words in the blocks $\mathrm{SL}(S_i;\mathbb{Z})$ on the diagonal. As before, we let $G = \mathrm{SL}(p;\mathbb{R})$ and $\Gamma = \mathrm{SL}(p;\mathbb{Z})$.

Let $P := U(S_1,\dots,S_k)$, and let $P^+ \subset P$ be the finite-index subgroup consisting of matrices in P whose diagonal blocks all have determinant 1. Let $K := \times_i \mathrm{SL}(S_i;\mathbb{Z})$, and consider the map $\eta : P^+ \to K$ which sends an element of P^+ to its diagonal blocks. If $N = \ker\eta$, we can write P^+ as a semidirect product $P^+ = K \ltimes N$.

In most cases, one can prove Lemma 3.3 in two stages. First, break an ω-triangle in P^+ into shortcut words in K and N, then fill the resulting shortcut words. This is harder to do when $P = U(p-1,1)$ or $U(1,p-1)$, because we cannot use Lemma 7.3 to manipulate the shortcut words. It is possible to use Lemma 7.6 to conjugate unipotent matrices by words in $\mathrm{SL}(p-1)$ one generator at a time, but this produces a cubic bound on the Dehn function rather than a quadratic bound. Instead, we will use the methods of Section 5 to break words in $U(p-1,1)$ and $U(1,p-1)$ into ω-triangles in smaller parabolic subgroups. In the next subsection, we consider the case that $\#S_i \le p-2$ for all i, and in Section 8.2, we will consider the case of $U(p-1,1)$ and $U(1,p-1)$.

8.1. *Case* 1: *Small* S_i*'s.* Let P, P^+, K, and N be as above. The goal of this section is to prove

Proposition 8.1. *Let* $P = U(S_1,\dots,S_k)$, *where* $\#S_i \le p-2$ *for all* i. *If* $g_1, g_2, g_3 \in P$ *and* $g_1g_2g_3 = 1$, *let*

$$w = \omega(g_1)\omega(g_2)\omega(g_3).$$

Then we can break w *into words* $v_1,\dots,v_k$ *at cost* $O(\ell(w)^2)$, *where* v_i *is a shortcut word in* $\mathrm{SL}(S_i;\mathbb{Z})$ *and* $\ell(v_i) = O(\ell(w))$.

We will prove this by breaking w into a product of a shortcut word in K and a shortcut word in N, then filling each of these shortcut words.

If $g \in \Gamma$ and $Q = U(T_1,\dots,T_r)$ is the minimal element of $\mathcal{P}$ containing g, then $\omega(g)$ is a product of a shortcut word in $\mathrm{SL}(T_i)$ for each i, at most p^2 shortcuts $\hat{e}_{ij}(x_{ij})$ (one for each entry above the diagonal), and a diagonal matrix. If $g \in P$, then $Q \subset P$, so each T_i is a subset of some S_j. In particular, each shortcut word in $\mathrm{SL}(T_i)$ is also a shortcut word in some $\mathrm{SL}(S_j)$, and each transvection $e_{ij}(x)$ with $i > j$ is either an element of some $\mathrm{SL}(S_j)$ or an element of N. Consequently, we can consider w as a product of at most $3p^2$ shortcut words in the $\mathrm{SL}(S_i)$, at most $3p^2$ shortcuts $\hat{e}_{ij}(x_{ij})$ such that $e_{ij}(x_{ij}) \in N$, and three diagonal matrices.

If $V \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$, let

$$\hat{n}_{ij}(V) := \prod_{a\in S_i, b\in S_j} \hat{e}_{ab}(v_{ab}).$$

Call a shortcut word in one of the $\mathrm{SL}(S_i)$ a *diagonal word*, and call a word of the form $\hat{n}_{ab}(V)$ an *off-diagonal block*. If $e_{ij}(x) \in N$, then $\hat{e}_{ij}(x)$ is an off-diagonal block, so w is a product of up to 3 diagonal matrices, up to $3p^2$ diagonal words, and up to $3p^2$ off-diagonal blocks.

We can use this terminology to describe the proof of Proposition 8.1.

(1) Break w into w_K, a product of boundedly many diagonal words, and w_N, a product of boundedly many off-diagonal blocks (Corollary 8.3).
(2) Break w_K into one diagonal word for each S_i (Corollary 8.5).
(3) Fill w_N using Lemma 7.6 (Lemma 8.6).

First, though, we rid ourselves of the diagonal matrices in w. Lemma 7.6 lets us move diagonal matrices past shortcuts, so we can shift the diagonal matrices to the beginning of w using $O(\ell(w)^2)$ applications of relations. Since each diagonal word and off-diagonal block represents an element of P^+, the product of the diagonal matrices is a diagonal matrix that is an element of P^+. Replace the three diagonal matrices with the product of k diagonal matrices, one in each $\mathrm{SL}(S_i;\mathbb{Z})$. (We think of each of these as a diagonal word with one letter.) The resulting word, which we call w', is the product of up to $3p^2+p$ diagonal words and up to $3p^2$ off-diagonal blocks.

The next step is to separate the diagonal words and the off-diagonal blocks. We need the following lemma.

LEMMA 8.2. *Assume, as above, that $\#S_i \le p-2$ for all i. If v is a shortcut word in $\mathrm{SL}(S_a)$ that represents M and $V \in \mathbb{Z}^{S_b} \otimes \mathbb{Z}^{S_c}$ for some $1 \le b < c \le k$, then*

(1) *If $a = b$, then*

$$\delta_\Gamma(v\hat{n}_{bc}(V)v^{-1}, \hat{n}_{bc}(MV)) = O((\ell(v) + \log \|V\|_2)^2).$$

(2) *If $a = c$, then*

$$\delta_\Gamma(v\hat{n}_{bc}(V)v^{-1}, \hat{n}_{bc}(VM^{-1})) = O((\ell(v) + \log \|V\|_2)^2).$$

(3) *If a, b, and c are distinct, then*

$$\delta_\Gamma([v, \hat{n}_{bc}(V)]) = O((\ell(v) + \log \|V\|_2)^2).$$

Remark. If we instead assume that $3 \le \#S_i \le p-3$ for all i, proving the lemma becomes much simpler. For example, if $3 \le \#S_a \le p-3$, we can prove part (3) by replacing the shortcuts in v by words in $\mathrm{SL}(S_a;\mathbb{Z})$ and replacing the shortcuts in $\hat{n}_{bc}(V)$ by words in $\mathrm{SL}((S_a)^c;\mathbb{Z})$. Since the corresponding sets of generators commute, we can commute the words at quadratic cost. When $\#S_a$ is particularly large or small, though, we need to use more involved methods.

Proof. We may assume that $\#S_a \geq 2$; otherwise, v would be the empty word.

Parts (1) and (2): We will mainly consider part (1); part (2) is essentially symmetric. Since $\hat{n}_{bc}(V)$ is a product of shortcuts, we will show that

$$\delta_\Gamma(v\hat{e}_{ij}(x)v^{-1}, \hat{n}_{bc}(xMz_i \otimes z_j)) = O((\ell(v) + \log|x|)^2)$$

for every $x \in \mathbb{Z}$, $i \in S_b$, $j \in S_c$ and then apply that to each term of $\hat{n}_{bc}(V)$.

Let $\omega_0 = v\hat{e}_{ij}(x)v^{-1}$. First, we use Lemma 7.1 to replace the shortcuts in v and v^{-1} by words in $\mathrm{SL}(S)$ for some S. If $\#S_b \geq 3$, we can take $S = S_b$; otherwise, take $l \in \{1, \dots, p\}$ such that $l \notin S$, $l \neq j$, and let $S = S_b \cup \{l\}$. Call the resulting word v'. We can use the same lemma to replace $\hat{e}_{ij}(x)$ by $\hat{u}_S(xz_i \otimes z_j)$, transforming ω_0 to

$$\omega_1 = v'\hat{u}_S(xz_i \otimes z_j)(v')^{-1}.$$

Finally, Lemma 7.3 applies to ω_1, so we can transform it to $\hat{u}_S(xMz_i \otimes z_j)$. Since this is a curve in H_{S,S^c}, which has a quadratic Dehn function, we can use Theorem 7.2 and Lemma 7.1 to transform this to $\hat{n}_{bc}(xMz_i \otimes z_j)$.

Applying this result to each term of $\hat{n}_{bc}(V)$, we can transform $v\hat{n}_{bc}(V)v^{-1}$ to

$$\prod_{i,j} \hat{n}_{bc}(v_{ij}Mz_i \otimes z_j).$$

We can apply parts (1) and (3) of Lemma 7.6 to reduce this to $\hat{n}_{bc}(MV)$, as desired. Part (2) follows similarly.

Part (3): Since $\hat{n}_{bc}(V)$ is a product of at most p^2 shortcuts, it suffices to show that if $v \in \mathrm{SL}(S)$ and $m, n \notin S$, then

$$\delta_\Gamma([v, \hat{e}_{mn}(x)]) = O((\ell(v) + \log|x|)^2).$$

If $2 \leq \#S \leq p-3$, then we can use Lemma 7.1 to replace v with a word in $\mathrm{SL}(S \cup \{m\})$ and prove the lemma by applying Lemma 7.3 to $H_{S\cup\{m\},(S\cup\{m\})^c}$. It just remains to consider the case that $\#S = p-2$.

Without loss of generality, we may take $S = \{2, \dots, p-1\}$. Since $\#S \geq 3$, we can use Lemma 7.1 to replace v with a word v' in $\mathrm{SL}(S)$. We claim that

$$\delta_\Gamma([v', \hat{e}_{1p}(x)]) = O((\ell(v) + \log|x|)^2).$$

We will construct a homotopy from $v'\hat{e}_{1p}(x)(v')^{-1}$ to $\hat{e}_{1p}(x)$ through the curves

$$\begin{aligned}
&v'[e_{12}(1),\hat{e}_{2p}(x)](v')^{-1} && \text{by Lemmas 7.1 and 7.6,}\\
&[v'e_{12}(1)(v')^{-1}, v'\hat{e}_{2p}(x)(v')^{-1}] && \text{by free insertion,}\\
&\left[\prod_{i=2}^{p-1}\hat{e}_{1i}(m_i), \prod_{i=2}^{p-1}\hat{e}_{ip}(n_i)\right] && \text{by Lemma 7.3,}\\
&\hat{e}_{1p}\left(\sum_i m_i n_i\right) = \hat{e}_{1p}(x) && \text{by Lemma 7.6.}
\end{aligned}$$

Here, m_i and n_i are the coefficients of Mz_2 and z_2M^{-1} respectively. The total cost of these steps is at most $O((\ell(v)+\log|x|)^2)$. The last step needs some explanation. Let

$$w_1 = \prod_{i=2}^{p-1}\hat{e}_{1i}(m_i),$$
$$w_2 = \prod_{i=2}^{p-1}\hat{e}_{ip}(n_i),$$

so we are transforming $[w_1,w_2]$ to $\hat{e}_{1p}(x)$. Each term $\hat{e}_{1i}(m_i)$ of w_1 commutes with every term of w_1 and w_2 except for $\hat{e}_{ip}(n_i)$ and its inverse, and Lemma 7.6 lets us transform $[\hat{e}_{1i}(m_i),\hat{e}_{ip}(n_i)]$ to $\hat{e}_{1p}(m_in_i)$. This commutes with every term of w_1 and w_2. So, if we use Lemma 7.6 to move terms of w_1 past w_2, the only new terms that appear are of this form, so once we get rid of all of the terms of w_1 and w_2, we are left with

$$\prod_{i=2}^{p-1}\hat{e}_{1p}(m_in_i).$$

Since $\sum_i m_in_i = z_2M^{-1}\cdot Mxz_2 = x$, we can use Lemma 7.6 to convert this to $\hat{e}_{1p}(x)$. All of the coefficients in this process are bounded by $\|M\|_2^2x$, and $\|M\|_2$ is exponential in $\ell(w)$, so this step has cost $O((\ell(w)+\log|x|)^2)$. This concludes the proof. □

In particular, this lets us break w into a product of diagonal words and a product of off-diagonal blocks.

Corollary 8.3. *If $g_1,g_2,g_3\in P$, $g_1g_2g_3=1$, and*

$$w=\omega(g_1)\omega(g_2)\omega(g_3),$$

then there are words w_K and w_N such that w_K is a product of at most $3p^2+p$ diagonal words, w_N is a product of at most $3p^2$ off-diagonal blocks, $\ell(w_K)=O(\ell(w))$, $\ell(w_N)=O(\ell(w))$, and

$$\delta_\Gamma(w,w_Kw_N)=O(\ell(w)^2).$$

Proof. As we noted before Lemma 8.2, it takes $O(\ell(w)^2)$ applications of relations to replace w by a word w' that is a product of at most $3p^2+p$ diagonal words and at most $3p^2$ off-diagonal blocks. Lemma 8.2 lets us move diagonal words past off-diagonal blocks. This process will affect the coefficients of these off-diagonal blocks, but it is straightforward to check that these coefficients remain bounded by $e^{\ell(w)}$ throughout the entire process. Thus, moving a diagonal word past an off-diagonal block always has cost $O(\ell(w)^2)$. We start with a bounded number of diagonal words and off-diagonal blocks, and no additional terms are created in the process, so we use Lemma 8.2 only boundedly many times and the total cost remains $O(\ell(w)^2)$. The resulting word can be broken into a product of $3p^2+p$ diagonal words, which we call w_K, and a product of at most $3p^2$ off-diagonal blocks, which we call w_N. □

Next, we sort the diagonal words so that all the shortcut words in $\mathrm{SL}(S_i;\mathbb{Z})$ are grouped together for $i=1,\dots,k$. We use the following lemma.

LEMMA 8.4. *Let $S,T\subset\{1,\dots,p\}$ be disjoint subsets such that $\#S,\#T\leq p-2$. Let w_S be a shortcut word $\mathrm{SL}(S;\mathbb{Z})$, and let w_T be a shortcut word in $\mathrm{SL}(T;\mathbb{Z})$. Then*

$$\delta_\Gamma([w_S,w_T])=O((\ell(w_S)+\ell(w_T))^2).$$

Proof. If $\#S\geq 3$ and $\#T\geq 3$, we can use Lemma 7.1 to replace w_S and w_T by words in $\mathrm{SL}(S;\mathbb{Z})$ and $\mathrm{SL}(T;\mathbb{Z})$, then commute the resulting words letter by letter. Similarly, if $\#S=1$ or $\#T=1$, then w_S or w_T is trivial. It remains only to study the case that one of $\#S$ and $\#T$ is 2. Without loss of generality, we take $S=\{1,2\}$ and $T=S^c$. Consider the case that w_T is a word in $\mathrm{SL}(S^c;\mathbb{Z})$.

Let $v=w_T$ be a word in $\mathrm{SL}(S^c;\mathbb{Z})$, and consider $\delta_\Gamma([v,w_S])$. Since w_S is a shortcut word, we can write it as a product $w_S=w_1\cdots w_n$ of diagonal matrices and shortcuts. If w_i is a diagonal matrix, it commutes with each letter of v; otherwise, we can bound the filling area of $[v,w_i]$ using part (3) of Lemma 8.2, which states that

$$\delta_\Gamma([w_i,v])\leq c(\ell(w_i)+\ell(v))^2,$$

so

$$\delta_\Gamma([w_S,v])\leq\sum_i\delta_\Gamma([w_i,v])\leq cn(\ell(w_S)+\ell(v))^2.$$

To get rid of the extra n, we need a slightly better bound.

When $\ell(v)\geq\ell(w_i)$, we can get a stronger bound on $\delta_\Gamma([w_i,v])$ by breaking v into segments of length $\sim\ell(w_i)$. Let $v=v_1\cdots v_d$, where $\ell(w_i)\leq\ell(v_j)\leq 2\ell(w_i)$ for each j. Then $d\leq\ell(v)/\ell(w_i)$ and

$$\delta_\Gamma([w_i,v_j])\leq 9c\ell(w_i)^2,$$

so

$$\delta_\Gamma([w_i, v]) \le 9cd\ell(w_i)^2 \le 9c\ell(v)\ell(w_i).$$

On the other hand, if $\ell(v) < \ell(w_i)$, then

$$\delta_\Gamma([w_i, v]) \le 4c\ell(w_i)^2.$$

So we have

$$\begin{aligned}\delta_\Gamma([w_S, v]) &\le \sum_i \delta_\Gamma([w_i, v]) \\ &\le \sum_i 9c\ell(v)\ell(w_i) + 4c\ell(w_i)^2 \\ &\le 9c\ell(v)\ell(w_S) + 4c\ell(w_S)^2 = O((\ell(w_S) + \ell(v))^2).\end{aligned}$$

So if w_T is a word in SL($S^c;\mathbb{Z}$), the lemma holds. Otherwise, w_T is a shortcut word in SL($S^c;\mathbb{Z}$), and since $\#S^c \ge 3$, we can use Lemma 7.1 to replace w_T with a word in SL($S^c;\mathbb{Z}$) at cost $O(\ell(w_T)^2)$. The lemma follows. □

We use this lemma repeatedly to sort the shortcut words in w_K.

COROLLARY 8.5. *If v is a word representing the identity that is the product of at most c diagonal words, then we can break v into diagonal words $v_1, \dots, v_k$ at cost $O(c^2\ell(v)^2)$, where v_i is a shortcut word in* SL($S_i;\mathbb{Z}$)*,* $\ell(v_1 \cdots v_k) = \ell(v)$.

Proof. We just need to swap diagonal words in v until all the diagonal words in SL($S_1;\mathbb{Z}$) are at the beginning, followed by all the diagonal words in SL($S_2;\mathbb{Z}$) and so on. This takes at most c^2 swaps, and each swap has cost $O(\ell(v)^2)$. Since v represents the identity, each v_k also represents the identity. □

So we can break the original w into the $v_1, \dots, v_k$ and w_N at cost $O(\ell(w)^2)$, and the v_i each have $\ell(v_i) = O(\ell(w))$. To prove the lemma, it just remains to fill w_N. Recall that w_N is a product of at most $3p^2$ off-diagonal blocks.

LEMMA 8.6. *If $w = w_1 \cdots w_d$ is a product of d off-diagonal blocks which represents the identity, then*

$$\delta_\Gamma(w) = O(d\ell(w)^2).$$

Proof. Let g_i be the group element represented by w_i. We will define a normal form ω_N for N and then fill w by bounding

$$\delta_\Gamma(\omega_N(g_1 \cdots g_{i-1})\omega_N(g_i), \omega_N(g_1 \cdots g_i)).$$

We can combine these fillings into a filling for w.

If $m \in N$, we can write m in block upper-triangular form:

$$m = \begin{pmatrix} I & V_{12} & \cdots & V_{1k} \\ 0 & I & \cdots & V_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I \end{pmatrix},$$

where $V_{ij} \in \mathbb{Z}^{S_i} \otimes \mathbb{Z}^{S_j}$. We can decompose this into a product

$$\omega_N(m) = x_k(m) \cdots x_1(m),$$

where

$$x_i(m) = \hat{n}_{i,i+1}(V_{i,i+1}) \cdots \hat{n}_{i,k}(V_{i,k}).$$

This is a normal form for elements of N; note that it has one term for each block above the diagonal. Furthermore, each subword of w represents a single off-diagonal block, so there are a_i, b_i, and V_i such that for all i, $\omega_N(g_i) = \hat{n}_{a_i,b_i}(V_i)$.

So consider $\omega_N(g_1 \cdots g_{i-1})\hat{n}_{a_i,b_i}(V_i)$. The coefficients of $g_1 \cdots g_{i-1}$ are all bounded by $\exp(\ell(w))$, so $\omega_N(g_1 \cdots g_{i-1})$ and $n = \hat{n}_{a_i,b_i}(V_i)$ both have length $O(\ell(w))$. We will transform $\omega_N(g_1 \cdots g_{i-1})n$ to $\omega_N(g_1 \cdots g_i)$ by moving n to the left until we can combine it with the right term in $\omega_N(g_1 \cdots g_{i-1})$. We move w to the left by commuting it with other off-diagonal blocks using Lemma 7.6. That is, we make replacements:

$$\hat{n}_{ab}(V)n \to \begin{cases} n\hat{n}_{ab}(V) & \text{if } a \neq b_i \text{ and } b \neq a_i, \\ n\hat{n}_{ab}(V)\hat{n}_{a,b_i}(VV_i) & \text{if } b = a_i, \\ n\hat{n}_{ab}(V)\hat{n}_{a_i,b}(-V_iV) & \text{if } a = b_i. \end{cases}$$

One can use Lemma 7.6 to make each of these replacements at cost $O(\ell(w)^2)$. Each replacement moves n one term to the left and possibly creates one extra off-diagonal block. We make replacements until n is next to the $\hat{n}_{a_i,b_i}(V')$-term in $\omega_N(g_1 \cdots g_{i-1})$, then use Lemma 7.6 to combine the two terms into a single off-diagonal block. Note that since we stop at the $\hat{n}_{a_i,b_i}(V')$-term, we only move n past $\hat{n}_{ab}(V)$ when $a \leq a_i$, so we never need the third case.

During this process, we have inserted at most k additional off-diagonal blocks. Using a similar replacement process, we can move these new blocks to their places in $\omega_N(g_1 \cdots g_{i-1})$. One can check that this does not add any further new blocks and has cost $O(\ell(w)^2)$. Thus,

$$\delta_\Gamma(\omega_N(g_1 \cdots g_{i-1})n, \omega_N(g_1 \cdots g_i)) = O(\ell(w)^2).$$

Therefore,

$$\delta_\Gamma(w) \leq \sum_{i=1}^{d} \delta_\Gamma(\omega_N(g_1 \cdots g_{i-1})n, \omega_N(g_1 \cdots g_i)) = O(d\ell(w)^2),$$

as desired. $\square$

8.2. *Case* 2: *Large S_i's.* The goal of this section is to prove

PROPOSITION 8.7. *Let $P = U(p-1,1)$. If $g_1, g_2, g_3 \in P$ and $g_1g_2g_3 = 1$, let*

$$w = \omega(g_1)\omega(g_2)\omega(g_3).$$

Then we can break w into words $v_1, \dots, v_d$ at cost $O(\ell(w)^2)$, where v_i is a shortcut word in $\mathrm{SL}(p-1;\mathbb{Z})$ and $\sum \ell(v_i)^2 = O(\ell(w)^2)$.

Because some of our lemmas give poor bounds in this case, the proof of Proposition 8.7 is very different from the proof of Proposition 8.1. Instead of using combinatorial manipulations to construct a filling, we use a variation of the adaptive triangulation argument used to prove Lemma 3.1 to reduce the problem of filling w to the problem of filling ω-triangles in parabolic subgroups of P. We can then use Proposition 8.1 to fill such triangles.

Since P is a finite-index extension of $\mathrm{SL}(p-1;\mathbb{Z}) \ltimes \mathbb{Z}^{p-1}$, any word in P that represents the identity can be reduced to a word in $\mathrm{SL}(p-1;\mathbb{Z}) \ltimes \mathbb{Z}^{p-1}$ at cost linear in the length of the word. Druţu showed that if $p \geq 4$, then the group $H = \mathrm{SL}(p-1;\mathbb{R}) \ltimes \mathbb{R}^{p-1}$ has a quadratic Dehn function [Dru04], but we will need the stronger result that a curve of length ℓ can be filled by a Lipschitz map of a disc of radius ℓ.

Let $\mathcal{E}_H := H/\mathrm{SO}(p-1)$. The map $H \to \mathrm{SL}(p-1)$ induces a fibration of $\mathcal{E}_H$ over $\mathcal{E}_{p-1} := \mathrm{SL}(p-1)/\mathrm{SO}(p-1)$ with fiber $\mathbb{R}^{p-1}$. Let $m : \mathcal{E}_H \to \mathcal{E}_{p-1}$ be this projection map. If $x \in H$, let $[x]_{\mathcal{E}_H}$ be the corresponding point of $\mathcal{E}_H$. We will show

LEMMA 8.8. *If $p \geq 4$, there is a c_0 such that for any $\gamma : [0,\ell] \to \mathcal{E}_H$ that is a constant-speed parametrization of a closed curve of length ℓ, if $\ell \geq 1$, and if $D^2(\ell) := [0,\ell] \times [0,\ell]$, then there is a c_0-Lipschitz map $f : D^2(\ell) \to \mathcal{E}_H$ that agrees with γ on the boundary of $D^2(\ell)$.*

The proof of this lemma requires some involved geometric and combinatorial constructions, so we postpone it until the end of this section.

Assuming the lemma, we can prove Proposition 8.7. First, we need to translate the notions in Section 4 to the context of H. Let

$$\mathcal{M}_{p-1} := \mathrm{SL}(p-1;\mathbb{Z})\backslash\mathcal{E}_{p-1}.$$

Let $A^+, N^+ \subset \mathrm{SL}(p-1)$ be as in (4), so that N^+A^+ is a Siegel set in $\mathrm{SL}(p-1)$ and $\mathcal{S}_{p-1} := [N^+A^+]_{\mathcal{E}_{p-1}}$ is a fundamental set. Let $N_H^+ \subset H$ be the subset of $\mathbb{R}^{p-1} \in H$ consisting of vectors with components in $[-1/2, 1/2]$, and define $\mathcal{S}_H = [N_H^+N^+A^+]_{\mathcal{E}_H}$. This is a fundamental set for the action of P on $\mathcal{E}_H$. Just as we defined the map $\rho : \mathcal{E} \to \mathrm{SL}(p;\mathbb{Z})$ using $\mathcal{S}$, we can define a map $\rho_H : \mathcal{E}_H \to P$ such that $\rho_H(\mathcal{S}_H) = I$ and $x \in \rho_H(x)\mathcal{S}_H$ for all $x \in \mathcal{E}_H$. Note that, unlike the previous case, $\mathcal{S}_H$ is not Hausdorff equivalent to A^+.

Define a depth function $r : \mathcal{E}_{p-1} \to \mathbb{R}^+$ by letting

$$r(x) = d_{\mathcal{M}_{p-1}}([x]_{\mathcal{M}_{p-1}}, [I]_{\mathcal{M}_{p-1}}).$$

Since $\mathcal{E}_H$ fibers over $\mathcal{E}_{p-1}$, we can define a depth function $r_H : \mathcal{E}_H \to \mathbb{R}^+$ by letting $r_H(x) = r(m(x))$.

A lemma similar to Lemma 4.5 holds.

LEMMA 8.9. *There is a c such that if $x, y \in \mathcal{E}_H$, then*

$$d_\Gamma(\rho_H(x), \rho_H(y)) \leq c(d_{\mathcal{E}_H}(x, y) + r_H(x) + r_H(y)) + c.$$

Proof. If $x \in \mathcal{E}_H$ and $n_1 \in N_H^+$, $n_2 \in N^+$, and $a \in A^+$ are such that $x = [\rho_H(x)n_1n_2a]_{\mathcal{E}_H}$, then $r_H(x) = r([n_2a]_{\mathcal{E}_{p-1}})$, so

$$d_{\mathcal{E}_H}([\rho_H(x)]_{\mathcal{E}_H}, x) \leq d_H(I, n_1) + r_H(x) = r_H(x) + O(1).$$

In particular,

$$d_{\mathcal{E}_H}([\rho_H(x)]_{\mathcal{E}_H}, [\rho_H(y)]_{\mathcal{E}_H}) \leq 2\operatorname{diam}(N_H^+) + r_H(x) + r_H(y) + d_{\mathcal{E}_H}(x, y),$$

and since $p \geq 5$, the lemma follows from Theorem 2.2. □

In addition, balls that are deep in the cusp are contained in a translate of a parabolic subgroup. Specifically, Corollary 4.8 implies that there is a $c_1 > 0$ such that if $x \in \mathcal{E}_H$, $r_H(x) > c_1$, and B is the ball of radius $\frac{r_H(x)}{4(p-1)^2}$ around x, then $\rho_H(B) \subset gU(q, p-1-q, 1)$ for some $1 \leq q \leq p-2$ and some $g \in P$.

We can thus use Corollary 5.3 to prove the following (cf. Lemma 5.1).

LEMMA 8.10. *Let $p \geq 4$, and let $P = U(p-1, 1)$. There is a c_2 such that if w is a word in P that represents the identity and $\ell = \ell(w)$, then there is a template τ for w such that*

(1) *If $g_1, g_2, g_3 \in P$ are the labels of a triangle in the template, then either* $\operatorname{diam}\{g_1, g_2, g_3\} \leq c_2$ *or there is a q such that all of the g_i are contained in the same coset of $U(q, p-1-q, 1)$.*

(2) *If g_1, g_2 are the labels of an edge e in the template, then*

$$d_\Gamma(g_1, g_2) = O(\ell(e)).$$

(3) *τ has $O(\ell^2)$ triangles, and if the i-th triangle of τ has vertices labeled (g_{i1}, g_{i2}, g_{i3}), then*

$$\sum_i (d_\Gamma(g_{i1}, g_{i2}) + d_\Gamma(g_{i1}, g_{i3}) + d_\Gamma(g_{i2}, g_{i3}))^2 = O(\ell^2).$$

Similarly, if the i-th edge of τ has vertices labeled h_{i1}, h_{i2}, then

$$\sum_i d_\Gamma(h_{i1}, h_{i2})^2 = O(\ell^2).$$

Proof. The proof of the lemma proceeds in much the same way as the proof of Lemma 5.1. If t is the smallest power of 2 that is greater than ℓ, we can use Lemma 8.8 to construct a Lipschitz map $f : [0,t] \times [0,t] \to \mathcal{E}_H$ that takes the boundary of $D := [0,t] \times [0,t]$ to the curve corresponding to w. We assume that this is parametrized so that if x is a lattice point in ∂D, then $f(x) = [g]_{\mathcal{E}_H}$ for some $g \in P$. This map has a Lipschitz constant bounded independent of ℓ, say by c_3.

Let $h : D \to \mathbb{R}$,

$$h(x) = \max\left\{1, \frac{r_H(f(x))}{8(p-1)^2 c_3}\right\}.$$

This function is 1-Lipschitz, and we let τ_h be the triangulation of D constructed in Corollary 5.3. If v is a vertex of τ_h, we label it $\rho_H(f(v))$; this makes τ_h a template.

If x is a lattice point on the boundary of D, then $f(x) \in [P]_{\mathcal{E}_H}$ and so $h(x) = 1$. In particular, each lattice point on the boundary of D is a vertex of τ_h, so the boundary of τ_h is a $4t$-gon whose vertex labels fellow-travel with w. We can add $O(t)$ small triangles to τ_h to get a template τ for w.

To prove that this satisfies the conditions of the lemma, we first need to show that if v_1 and v_2 are the endpoints of an edge of τ, labeled by g_1 and g_2, then $d_\Gamma(g_1, g_2) \lesssim d(v_1, v_2)$. By Lemma 8.9, we know that

$$d_\Gamma(g_1, g_2) = O\left(d_{\mathcal{E}_H}(f(v_1), f(v_2)) + r_H(f(v_1)) + r_H(f(v_2)) + 1\right),$$

and each of these terms is $O(d(v_1, v_2))$, so $d_\Gamma(g_1, g_2) = O(d(v_1, v_2))$ as desired.

Part (2) of the lemma then follows from the bounds in Corollary 5.3. Part (1) of the lemma follows from the fact that if x_1, x_2, and x_3 are the vertices of a triangle of τ, with labels g_1, g_2, and g_3, then Corollary 5.3 implies that

$$\operatorname{diam}\{x_1, x_2, x_3\} \leq 2h(x_1) \leq \frac{r_H(f(x))}{4(p-1)^2 c_3}.$$

If $h(x_1)$ is sufficiently large, then the remark before the lemma implies that g_1, g_2, and g_3 are in the same coset of $U(q, p-1-q, 1)$ for some j. $\square$

Proposition 8.7 follows as a corollary.

Proof of Proposition 8.7. If w is a word in P, let τ be a template for w satisfying the properties in Lemma 8.10. We can use τ to break w into a set of ω-triangles and bigons. Each of these either has bounded length or is a shortcut word in some $U(q, p-1-q, 1)$. The ones with bounded length can be filled with bounded area. Since there are at most $O(\ell(w)^2)$ of these, this has total cost $O(\ell(w)^2)$. The shortcut words can be broken into smaller pieces by using Proposition 8.1. This also has total cost $O(\ell(w)^2)$, and the resulting shortcut words $v_1, \dots, v_d$ fulfill the conditions of the proposition. $\square$

So, to prove Proposition 8.7, it suffices to prove Lemma 8.8.

8.3. *Proof of Lemma* 8.8. In this section, we will show that Lipschitz curves in $\mathcal{E}_H$ can be filled with Lipschitz discs. We will proceed by decomposing a loop in $\mathcal{E}_H$ into simple pieces. First, we will recall an argument of Gromov that in order to fill a Lipschitz loop with a Lipschitz disc, it suffices to be able to fill a family of Lipschitz triangles with Lipschitz discs (similar to the arguments using templates in Section 3). Since H is a semidirect product, these triangles can be chosen so that each side is a concatenation of a geodesic in $\mathcal{E}_{p-1} := \mathrm{SL}(p-1;\mathbb{R})$ and a curve that represents an element of $\mathbb{R}^{p-1}$. We complete the proof by filling polygons whose sides consist of such curves with Lipschitz discs.

In this section, let $D^2(\ell) := [0,\ell]\times[0,\ell]$ and $D^2 := D^2(1)$, and let $S^1(\ell)$ and S^1 be the boundaries of $D^2(\ell)$ and D^2.

The following remarks and lemma about Lipschitz maps between polygons will be helpful.

Remark 8.11. If S is a convex polygon with nonempty interior and diameter at most 1, there is a map $S \to D^2$ whose Lipschitz constant varies continuously with the vertices of S.

Remark 8.12. For every ℓ, there is a $c(\ell)$ such that any closed curve in $\mathcal{E}_H$ of length ℓ can be filled by a $c(\ell)$-Lipschitz map $D^2(1) \to \mathcal{E}_H$. This follows from compactness and the homogeneity of $\mathcal{E}_H$.

LEMMA 8.13. *Let $\gamma : S^1(\ell) \to X$ be a closed curve, and let $\beta_0 : D^2(\ell) \to X$ be a map such that $\beta_0|_{S^1(\ell)}$ is some reparametrization of γ. Then there is a map $\beta : D^2(\ell) \to X$ such that $\gamma = \beta|_{S^1(\ell)}$ such that*

$$\mathrm{Lip}(\beta) = O(\max\{\mathrm{Lip}(\beta_0), \mathrm{Lip}(\gamma)\}).$$

Proof. Let $\gamma_0 = \beta_0|_{S^1(\ell)}$ and let $h : S^1(\ell)\times[0,\ell] \to X$ be a homotopy from γ to γ_0. Since the two curves differ by a reparametrization, we can choose h to have Lipschitz constant of order $O(\max\{\mathrm{Lip}\,\gamma_0, \mathrm{Lip}\,\gamma\})$. If we glue β_0 and h together, we get a map $\beta_1 : D' \to X$, where

$$D' = \left[D^2(\ell) \cup (S^1(\ell)\times[0,\ell])\right]/\sim$$

is the space obtained by identifying $S^1(\ell)\times\{1\}$ and $\partial D^2(\ell)$. This map is a filling of γ with Lipschitz constant $O(\max\{\mathrm{Lip}\,\gamma, \mathrm{Lip}\,\beta_0\})$, and since D' is bilipschitz equivalent to $D^2(\ell)$, there is a map $\beta : D^2(\ell) \to X$ that agrees with γ on its boundary and has Lipschitz constant $O(\max\{\mathrm{Lip}\,\gamma, \mathrm{Lip}\,\beta_0\})$. □

One application of this remark is to convert homotopies to discs; if $f_1, f_2 : [0,\ell] \to X$ are two maps with the same endpoints, and $h : [0,\ell]\times[0,\ell] \to X$ is a Lipschitz homotopy between f_1 and f_2 with endpoints fixed, then there is a disc filling $f_1 f_2^{-1}$ with Lipschitz constant $O(\mathrm{Lip}(h))$.

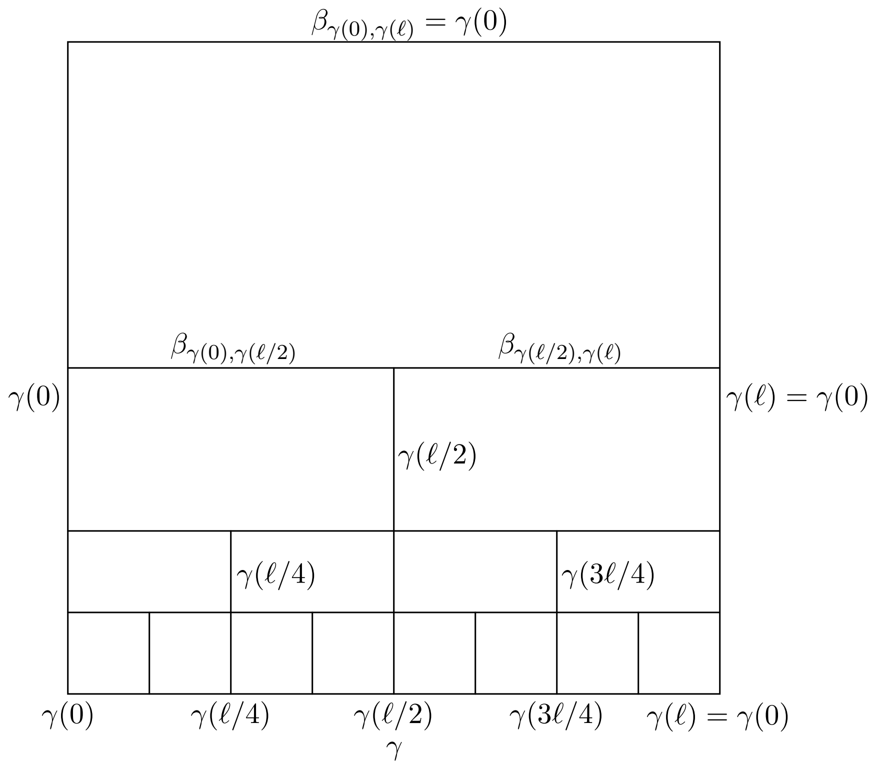

Figure 5. A filling of γ in X.

Suppose that X is a metric space and ω is a normal form for X. That is, suppose that if $x, y \in X$, then $\omega(x,y) : [0,1] \to X$ is a constant-speed curve connecting x and y and that there is a c such that $\ell(\omega(x,y)) \le cd(x,y) + c$. (Note that we do not require ω to satisfy any fellow-traveler properties.) We may assume that $\omega(x,x)$ is a constant curve for each x. Then, just as we built fillings of curves out of fillings of triangles in Section 3, we can build Lipschitz discs by gluing together Lipschitz triangles (cf. [Gro96]). Let Δ be the equilateral triangle with side length 1.

Proposition 8.14. *Let X be a homogeneous riemannian manifold or a simplicial complex with bounded complexity, and let ω be a normal form for X. Suppose that there is a c_1 such that for all $x, y, z \in X$, there is a map $f_{x,y,z} : \Delta \to X$ that takes the sides of the triangle to $\omega(x,y)$, $\omega(y,z)$, and $\omega(x,z)^{-1}$ and such that $\operatorname{Lip} f_{x,y,z} \le c_1 \operatorname{diam}\{x,y,z\} + c_1$. Then there is a C such that for every unit-speed Lipschitz closed curve $\gamma : [0,\ell] \to X$ of length $\ell \ge 1$, there is a map $g : D^2(\ell) \to X$ that agrees with γ on the boundary and has $\operatorname{Lip} g \le C$.*

Proof. We will construct a map $g : [0,\ell] \times [0,\ell] \to X$ that agrees with γ on one side and stays constant on the other three sides. The construction is essentially Gromov's construction of Lipschitz extensions from [Gro96]. Let $k = \lceil \log_2 \ell \rceil$. We will construct g as in Figure 5. The figure depicts a decomposition of $[0,\ell] \times [0,\ell]$ into $k+1$ rows of rectangles; the top row has one $\ell \times \frac{\ell}{2}$ rectangle, while the i-th from the top consists of 2^{i-1} rectangles of dimensions $2^{-i+1}\ell \times 2^{-i}\ell$. The bottom row is an exception, consisting of 2^k squares of side length $2^{-k}\ell$. Call the resulting complex D.

We label all the edges of D by curves in X. First, we label all the vertical edges by constant curves; the vertical edges with x-coordinate a are labeled by $\gamma(a)$. We label horizontal edges using the normal form: The edge from (x_1, y) to (x_2, y) is labeled by $\omega(\gamma(x_1), \gamma(x_2))$, except for the bottom edge of D, which is labeled γ. We can then define g on the 1-skeleton of D by sending each edge to the constant-speed parametrization of its label. It is easy to check that this construction is Lipschitz, with Lipschitz constant of order $O(\ell)$.

Let R be a rectangle in D. If R is in the bottom row of cells, then g maps the boundary of R to a curve of length bounded independently of γ, so we may extend g over R using Remark 8.12, and all these extensions have Lipschitz constant bounded independently of γ. Otherwise, suppose that R is a $2^{-i+1}\ell \times 2^{-i}\ell$ rectangle. Then, since g maps both its vertical edges to points, the restriction of g to the boundary of R is a curve of the form $\omega(x,y)\omega(y,z)\omega(x,z)^{-1}$, where $x = \gamma(t)$, $y = \gamma(t+2^{-i}\ell)$, and $z = \gamma(t+2^{-i+1}\ell)$. By assumption, there is a map $f_{x,y,z} : \Delta \to X$ that fills this curve and has Lipschitz constant $\le c_1 2^{-i-1}\ell + c_1$. Since R is bilipschitz equivalent to $D^2(2^{-i-1}\ell)$, we can reparametrize $f_{x,y,z}$ to get a map $R \to X$ that agrees with g on ∂R and has Lipschitz constant bounded independently of γ and i.

Defining extensions like this on every rectangle gives us a map $g : [0,\ell] \times [0,\ell] \to X$ whose boundary is a reparametrization of γ and whose Lipschitz constant is bounded independent by some C_0 independent of γ, so the proposition follows by applying Lemma 8.13. $\square$

Now we construct a normal form ω_H for $\mathcal{E}_H$. First, for each $h \in H$, we will construct a curve $\lambda_h : [0,1] \to H$ that connects I to h. We can write h as

$$h = \begin{pmatrix} M & v \\ 0 & 1 \end{pmatrix}$$

for some $M \in \mathrm{SL}(p-1)$ and $v \in \mathbb{R}^{p-1}$; we denote the corresponding unipotent matrix in H by $u(v)$. Let γ_M be a geodesic in $\mathrm{SL}(p-1)$ that connects I to M. We will construct λ_h by concatenating γ_M and a curve $\psi(v) : [0,1] \to H$ that connects I to $u(v)$.

If $v = 0$, let $\psi(v)$ be constant. If $v \in \mathbb{R}^{p-1}$, $v \neq 0$, we can write $v = \kappa\bar{v}$, where $\kappa := \max\{\|v\|_2, 1\}$ and $\bar{v} := \frac{v}{\kappa}$, so that $\|\bar{v}\|_2 \le 1$ and $0 \le \log\kappa =$

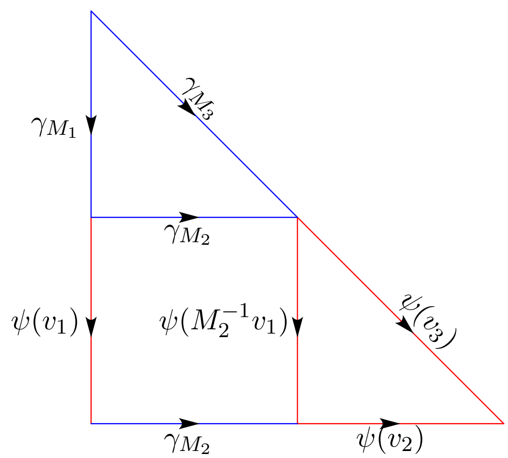

Figure 6. A quadratic filling of $\omega_H(M_1v_1)\omega_H(M_2v_2)\omega_H(M_3v_3)^{-1}$.

$O(d_H(I,u(v)))$. Let

$$v_1 = \frac{v}{\|v\|_2}, v_2, \dots, v_{p-1} \in \mathbb{R}^{p-1}$$

be an orthonormal basis of $\mathbb{R}^{p-1}$. Let $D(v)$ be the matrix that stretches v_1 by a factor of κ and shrinks the rest of the v_i by a factor of $\kappa^{1/(p-2)}$. This is positive definite and has determinant 1. Furthermore, $D(v)\bar{v} = v$, so $u(v) = D(v)u(\bar{v})D(v)^{-1}$. Let $\mathcal{D}(v)$ be the curve $t \mapsto D(v)^t$ for $0 \le t \le 1$, let $\mathcal{U}(v)$ be the curve $t \mapsto u(tv)$ for $0 \le t \le 1$, and define $\psi(v)$ to be the concatenation $\psi(v) = \mathcal{D}(v)\mathcal{U}(\bar{v})\mathcal{D}(v)^{-1}$. This has length $O(\overline{\log}\|v\|)$, where $\overline{\log}x = \max\{1, \log x\}$. Define $\lambda_h = \gamma_M\psi(v)$. It is easy to check that this satisfies the desired length bounds.

If $x, y \in \mathcal{E}_H$, choose lifts $\tilde{x}, \tilde{y}$ such that $[\tilde{x}]_{\mathcal{E}_H} = x$ and $[\tilde{y}]_{\mathcal{E}_H} = y$. Since $\mathrm{SO}(p-1)$ has bounded diameter, different choices of lift only differ by a bounded distance. Define $\omega_H(x,y)$ so that

$$\omega_H(x,y)(t) = [\tilde{x}\lambda_{\tilde{x}^{-1}\tilde{y}}(t)]_{\mathcal{E}_H}.$$

It is easy to check that this satisfies the desired length bounds.

Next, we will construct discs filling triangles whose sides are in normal form. We claim

Lemma 8.15. *If $h_1, h_2 \in H$ and*

$$w = \tilde{\omega}_H(h_1)\tilde{\omega}_H(h_2)\tilde{\omega}_H(h_1h_2)^{-1},$$

then there is a filling $f : D^2 \to \mathcal{E}_H$ of $[w]_{\mathcal{E}_H}$ such that $\mathrm{Lip}(f) = O(d(I,h_1) + d(I,h_2))$.

To prove this lemma, we will follow the template of Figure 6. The figure suggests that a filling for w can be constructed from fillings for two triangles and a rectangle. The following lemmas will construct those fillings.

LEMMA 8.16. *Let* $\gamma : [0,1] \to \mathrm{SL}(p-1)$ *be a curve connecting* I *and* M, *and let* $v \in \mathbb{R}^{p-1}$. *There is a map* $f : D^2 \to \mathcal{E}_H$ *that sends the boundary of* D^2 *to the curve* $[\gamma\psi(v)\gamma^{-1}\psi(Mv)^{-1}]_{\mathcal{E}_H}$ *and that has Lipschitz constant* $\operatorname{Lip} f = O(\overline{\log}\|v_1\|_2 + \ell(\gamma))$.

LEMMA 8.17. *Let* $v_1, v_2 \in \mathbb{R}^{p-1}$. *There is a map* $f : D^2 \to \mathcal{E}_H$ *that sends the boundary of* D^2 *to the curve* $[\psi(v_1)\psi(v_2)\psi(v_1+v_2)^{-1}]_{\mathcal{E}_H}$ *that has Lipschitz constant* $\operatorname{Lip} f = O(\overline{\log}\|v_1\|_2 + \overline{\log}\|v_2\|_2)$.

If we assume these two lemmas, then the proof of Lemma 8.15 follows easily.

Proof of Lemma 8.15. Let X be a triangle decomposed as in Figure 6. We put a metric on X so that the two small triangles are isoceles right triangles with legs of length 1 and the corner rectangle is a square with side length 1. Under this metric, X is bilipschitz equivalent to D^2. Let $\ell = \ell(w)$. Construct a map on the 1-skeleton of X so that if an edge is labeled by a curve γ in the figure, it is sent to a constant-speed parametrization of $[\gamma]_{\mathcal{E}_H}$. It is easy to check that the Lipschitz constant of this map is of order $O(\ell)$. We can use Lemmas 8.17 and 8.16 to construct maps from the lower right triangle and the corner square to $\mathcal{E}_H$ with Lipschitz constants of order $O(\ell)$.

It only remains to construct a filling of the upper triangle. The boundary of the upper triangle is a curve in $\mathcal{E}_{p-1} = \mathrm{SL}(p-1)/\mathrm{SO}(p-1)$, so we can construct a Lipschitz filling of the upper triangle by coning it off by geodesics. This filling also has Lipschitz constant of order $O(\ell)$, completing the construction. □

It remains to prove the two lemmas.

Proof of Lemma 8.16. Let $w = \gamma\psi(v)\gamma^{-1}\psi(Mv)^{-1}$. If $v = 0$, then $w = \gamma\gamma^{-1}$, so we may assume that $v \neq 0$. If $\ell(w) \leq 1$, we can use Remark 8.12 to fill w, so we also assume that $\ell(w) \geq 1$.

Recall that $\psi(v)$ is defined as the concatenation $\mathcal{D}(v)\mathcal{U}(\bar{v})\mathcal{D}(v)^{-1}$, where $\bar{v}$ is a vector parallel to v with length at most 1. We can thus write

$$w = \gamma\mathcal{D}(v)\mathcal{U}(\bar{v})\mathcal{D}(v)^{-1}\gamma^{-1}\mathcal{D}(Mv)\mathcal{U}(-\overline{Mv})\mathcal{D}(Mv)^{-1}.$$

Let

$$\gamma' = \mathcal{D}(Mv)^{-1}\gamma\mathcal{D}(v).$$

Changing the basepoint of w, we obtain the curve

$$w_1 = \gamma'\mathcal{U}(\bar{v})(\gamma')^{-1}\mathcal{U}(-\overline{Mv}).$$

This can be filled by a map of the form

$$\beta(x,t) = \gamma'(x)u(t \cdot \gamma'(x)^{-1}\overline{Mv})$$

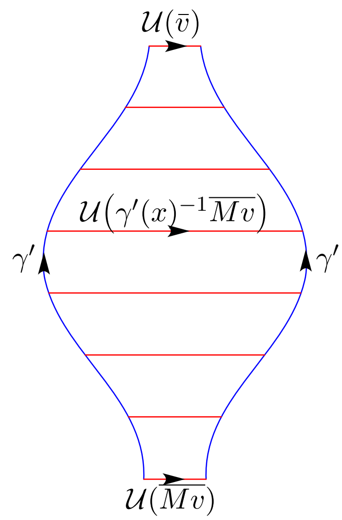

Figure 7. An exponential filling of $\gamma'\mathcal{U}(\bar{v})(\gamma')^{-1}\mathcal{U}(-\overline{Mv})$.

(see Figure 7). This filling has a foliation (horizontal curves in the figure) consisting of curves $\mathcal{U}(\gamma'(x)^{-1}\overline{Mv})$, but these may be exponentially large. We will use a homotopy in $\mathrm{SL}(p-1)$ to replace γ' by a curve σ such that the length of $\sigma(x)^{-1}\overline{Mv}$ is always bounded.

First, we construct σ. Let

$$S := \{m \mid m \in \mathrm{SL}(p-1), \|m^{-1}\overline{Mv}\|_2 \leq 1\},$$

and let

$$M' := D(Mv)^{-1}MD(v)$$

be the endpoint of γ'. Since $\bar{v} = (M')^{-1}\overline{Mv}$, the endpoint of γ' lies in S, and we will construct a curve σ in S that connects the identity to M'.

Consider the case that $\overline{Mv} = \bar{v}$, so that M' is in the stabilizer of $\overline{Mv}$, which we write $\mathrm{SL}(p-1)_{\overline{Mv}}$. This stabilizer is contained in S and is isomorphic to $\mathrm{SL}(p-2) \ltimes \mathbb{R}^{p-2}$. Furthermore, it is connected and since $p \geq 5$, its inclusion in $\mathrm{SL}(p-1)$ is undistorted, so we can let σ be a path in $\mathrm{SL}(p-1)_{\overline{Mv}}$ between I and M'.

To construct σ in the general case, it suffices to construct a curve in S that connects M' to a point in $\mathrm{SL}(p-1)_{\overline{Mv}}$; we can then apply the previous case. If $\|\overline{Mv}\|_2 = \|\bar{v}\|_2$, we can take this to be a curve in $\mathrm{SO}(p-1)$ of bounded length. Otherwise, we can construct a path of matrices that shrink (or grow) $\overline{Mv}$ and grow (or shrink) all the perpendicular directions; this path can be taken to lie in S, and its length is at most

$$O\left(\Big|\log\|\overline{Mv}\|_2 - \log\|\bar{v}\|_2\Big|\right) \leq O(\log\|M'\|_2).$$

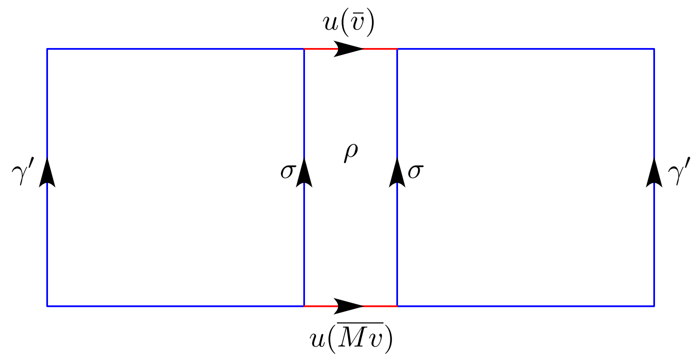

Figure 8. A quadratic filling of $\gamma'\mathcal{U}(\bar{v})(\gamma')^{-1}\mathcal{U}(-\overline{Mv})$.

We will use σ to construct a map $f : [0, 2\ell(w)+1] \times [0, \ell(w)] \to \mathcal{E}_H$ whose boundary is a parametrization of w. The domain of this map is divided into two $\ell(w) \times \ell(w)$ squares and a $\ell(w) \times 1$ rectangle (Figure 8); the squares will be homotopies in $\mathrm{SL}(p-1)$. We will map the boundaries of each of these shapes into $\mathcal{E}_H$ by Lipschitz maps and then construct Lipschitz discs in $\mathcal{E}_H$ with those boundaries.

Let f take the 1-skeleton of the rectangle into $\mathcal{E}_H$ as labeled in the figure, parametrizing each edge with constant speed. The boundaries of the shapes in the figure are then $[\sigma\gamma^{-1}]_{\mathcal{E}_H}$ and

$$w_2 = [\sigma\mathcal{U}(\bar{v})\sigma^{-1}\mathcal{U}(-\overline{Mv})]_{\mathcal{E}_H}.$$

The first curve, $\sigma\gamma^{-1}$, is a closed curve in $\mathrm{SL}(p-1)$ of length $O(\ell(w))$. Since $\mathrm{SL}(p-1)/\mathrm{SO}(p-1)$ is nonpositively curved, the projection to $\mathcal{E}_H$ has a filling in $\mathcal{E}_H$ with area $O(\ell(w)^2)$. This can be taken to be a c-Lipschitz map from $D^2(\ell(w))$, where c depends only on p.

The second curve is the boundary of a "thin rectangle." That is, there is a Lipschitz map

$$\rho : [0, \ell(w)] \times [0, 1] \to H,$$
$$\rho(x, t) = \sigma(x)u(t\sigma(x)^{-1}\overline{Mv}) = u(t\overline{Mv})\sigma(x),$$

which sends the four sides of the rectangle to $\sigma, \mathcal{U}(\bar{v}), \sigma^{-1}$, and $\mathcal{U}(-\overline{Mv})$. Projecting this disc to $\mathcal{E}_H$ gives a Lipschitz filling of w_2.

We glue these discs together to get a Lipschitz map from the rectangle to $\mathcal{E}_H$. The boundary of the rectangle is a Lipschitz reparametrization of $[w]_{\mathcal{E}_H}$, so we can use Lemma 8.13 to get a filling of $[w]_{\mathcal{E}_H}$ by a disc $D^2(\ell(w))$ with Lipschitz constant of order $O(1)$. Rescaling this gives a filling of $[w]_{\mathcal{E}_H}$ by the disc D^2 with Lipschitz constant of order $O(\ell(w))$, as desired. □

Proof of Lemma 8.17. Let $w = \psi(v_1)\psi(v_2)\psi(v_1 + v_2)^{-1}$. As before, we may assume that $\ell(w) > 3$. Let $S = \langle v_1, v_2 \rangle \subset \mathbb{R}^{p-1}$ be the subspace generated

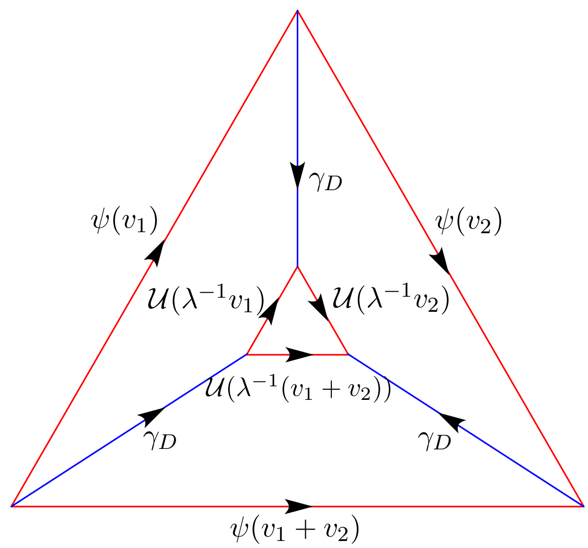

Figure 9. A quadratic filling of $\psi(v_1)\psi(v_2)\psi(v_1+v_2)^{-1}$.

by the v_i, and let $\lambda = \max\{\|v_1\|_2, \|v_2\|_2, \|v_1+v_2\|_2\}$. Let $D \in \mathrm{SL}(p-1)$ be the matrix such that $Ds = \lambda s$ for $s \in S$ and $Dt = \lambda^{-1/(p-1-\dim(S))}t$ for vectors t that are perpendicular to S; this is possible because $\dim(S) \leq 2$ and $p \geq 5$.

Let γ_D be the curve $t \mapsto D^t$ for $0 \leq t \leq 1$; this has length $O(\log \lambda) = O(\ell(w))$. We construct a filling of $[w]_{\mathcal{E}_H}$ based on a triangle with side length $\ell(w)$ as in Figure 9. The central triangle in the figure has side length 1; since $\ell(w) \geq 3$, the trapezoids around the outside are bilipschitz equivalent to discs $D^2(\ell)$ with Lipschitz constant bounded independently of w. Let f take each edge to H as labeled, and give each edge a constant-speed parametrization; f is Lipschitz on each edge, with a Lipschitz constant independent of the v_i. Let $\bar{f}$ be the projection of f to $\mathcal{E}_H$. We have defined $\bar{f}$ on the edges in the figure; it remains to extend it to the interior of each cell.

The map $\bar{f}$ sends the boundary of the center triangle to a curve of length at most 3, so we can use Remark 8.12 to extend $\bar{f}$ to its interior. The map $\bar{f}$ sends the boundary of each trapezoid to a curve of the form

$$[\psi(v_i)^{-1}\gamma_D\mathcal{U}(\lambda v_i)\gamma_D^{-1}]_{\mathcal{E}_H}. \tag{4}$$

Lemma 8.16 gives Lipschitz discs filling such curves. Each of these fillings has Lipschitz constant bounded independently of w, so the resulting map on the triangle is a filling of $[w]_{\mathcal{E}_H}$ by a triangle of side length $\ell(w)$ with Lipschitz constant bounded independently of w. By rescaling and mapping the triangle to D^2, we obtain a filling of $[w]_{\mathcal{E}_H}$ by the disc D^2 with Lipschitz constant of order $O(\ell(w))$ as desired. □

9. **The base case:** SL$(2;\mathbb{Z})$

In this section, we will prove Lemma 3.5, which states that if w is a shortcut word in SL$(2;\mathbb{Z})$, then

$$\delta_\Gamma(w) = O(\ell^2).$$

The proof uses the adaptive template methods developed in Section 5. The main change from Section 5 is that the curve that we fill will not be in the thick part of $\mathcal{E}_2$.

Let $w = w_1 \cdots w_n$ be a shortcut word representing the identity, where each w_i is either a diagonal matrix in SL$(2;\mathbb{Z})$ or a shortcut of the form $\hat{e}_{12}(x)$ or $\hat{e}_{21}(x)$. We first use Lemma 7.6 to replace all occurrences of $\hat{e}_{21}(x)$ in w by $g\hat{e}_{12}(-x)g^{-1}$, where g is a word representing a Weyl group element. This has cost $O(\ell(w)^2)$, and it lets us assume that $\hat{e}_{21}(x)$ does not occur in w for $|x| \geq 1$.

For each i, let $w(i)$ be the group element represented by $w_1 \cdots w_i$. Let $\mathcal{S}_2$ be a Siegel set for SL$(2;\mathbb{Z})$. For each i, we will construct a curve $\alpha_i : [0, \ell(w_i)] \to \mathcal{E}_2$ which connects $[w(i)]_{\mathcal{E}_2}$ to $[w(i+1)]_{\mathcal{E}_2}$ such that

- The curves α_i are uniformly Lipschitz, with Lipschitz constants bounded independently of w.
- There is an integer $t_i \in [0, \ell(w_i)]$ such that if $0 \leq j \leq t_i$ is an integer, then $\alpha_i(j) \in w(i)\mathcal{S}_2$ and if $t_i < j \leq \ell(w_i)$, then $\alpha_i(j) \in w(i+1)\mathcal{S}_2$.

If $\ell(w_i) < 3$, we define α_i on $[0,1]$ as the geodesic connecting $[w(i)]_{\mathcal{E}_2}$ and $[w(i+1)]_{\mathcal{E}_2}$ and we define α_i on $[1, \ell(w_i)]$ to be the constant value $[w(i+1)]_{\mathcal{E}_2}$. We let $t_i = 0$.

If $\ell(w_i) \geq 3$, let x be such that $w_i = \hat{e}_{12}(x)$. Let

$$D = \begin{pmatrix} e & 0 \\ 0 & e^{-1} \end{pmatrix},$$

and note that $[D^x]_{\mathcal{E}_2} \in \mathcal{S}_2$ for all $x \geq 0$. Let $t_i = \left\lceil \frac{\ell(w_i)}{3} \right\rceil$, and let $\beta : [0, \ell(w_i)] \to$ SL$(2;\mathbb{R})$ be the concatenation of geodesic segments connecting

$$\begin{aligned}
p_1 &= w(i), \\
p_2 &= w(i)D^{\log(|x|)/2}, \\
p_3 &= w(i)D^{\log(|x|)/2}e_{12}(\pm 1), \\
p_4 &= w(i)D^{\log(|x|)/2}e_{12}(\pm 1)D^{-\log(|x|)/2} = w(i)e_{12}(x) = w(i+1).
\end{aligned}$$

Here the sign of ± 1 is the same as the sign of x. Parametrize this curve so that $\beta|_{[0,t_1]}$ connects p_1 and p_2, $g|_{[t_1,t_1+1]}$ connects p_2 and p_3, and $g|_{[t_1+1,\ell(w_i)]}$ connects p_3 and p_4. Let $\alpha_i = [\beta]_{\mathcal{E}_2}$. This curve has velocity bounded independently of x. Furthermore, if $t \in \mathbb{Z}$, then $\alpha_i(t) \in w(i)\mathcal{S}_2$ if $t \leq t_1$ and $\alpha_i(t) \in w(i+1)\mathcal{S}_2$ if $t > t_1$.

Let $\alpha : [0,\ell(w)] \to \mathcal{E}_2$ be the concatenation of the α_i. From here, we largely follow the proof of Lemma 5.1; we construct a filling f of α, an adaptive triangulation τ, and a template based on τ so that a vertex x of τ is labeled by an element γ such that $f(x) \in \gamma\mathcal{S}_2$.

Let d be the smallest power of 2 larger than $\ell(w)$ and let $\alpha' : [0,d] \to \mathcal{E}_2$ be the extension of α to $[0,d]$, where $\alpha'(t) = [I]_{\mathcal{E}_2}$ when $t \geq \ell(w)$. Let $D^2(d) = [0,d] \times [0,d]$. We can map $\partial D^2(d)$ into $\mathcal{E}_2$ by sending one side to α' and sending the other three sides to $[I]_{\mathcal{E}_2}$, and since $\mathcal{E}_2$ is nonpositively curved, we can extend this map to all of $D^2(d)$ by coning it to a point along geodesics. Call the resulting map $f : D^2(d) \to \mathcal{E}_2$. This is c-Lipschitz for some c independent of w and has area $O(\ell(w)^2)$.

Let $\mathcal{M}_2 = \mathrm{SL}(2;\mathbb{Z})\backslash\mathcal{E}_2$, and define the depth function $r : \mathcal{E}_2 \to \mathbb{R}+$ by

$$r(x) = d_{\mathcal{M}_2}([x]_{\mathcal{M}_2}, [I]_{\mathcal{M}_2}).$$

Let $h : [0,d] \times [0,d]$ be

$$h(x) = \max\left\{1, \frac{r(f(x))}{32c}\right\}.$$

This is a 1-Lipschitz function, so we can use Corollary 5.3 to construct a triangulation τ_h of $[0,d] \times [0,d]$. It remains to convert this triangulation into a template.

For each vertex v of τ_h, we label v by an element $g \in \mathrm{SL}(2;\mathbb{Z})$ such that $f(v) \in g\mathcal{S}_2$. For the interior vertices, any such element suffices. For the boundary vertices, we must make choices that agree with w. Let $\ell_i = \sum_{j=1}^{i} \ell(w_j)$ so that α_i and α_{i+1} meet at $(\ell_i, 0)$. We have $\alpha'(\ell_i) = [w(i)]_{\mathcal{E}_2}$, so $h(\ell_i, 0) = 1$ and $(\ell_1, 0)$ must be a vertex of τ_h; we label it $w(i)$. Let $\beta_0 = 0$, $\beta_i = \ell_{i-1} + t_i$ for $i = 1, \dots, n$, and $\beta_{n+1} = d$, so that if $\beta_i < t \leq \beta_{i+1}$, then $f(t,0) = \alpha'(t) \in w(i)\mathcal{S}_2$. For all t with $\beta_i < t \leq \beta_{i+1}$, label the point $(t,0)$ by the element $w(i)$. Boundary vertices that are not of the form $(t,0)$ are all sent to $[I]_{\mathcal{E}_2}$ under f, and we label them by I. With this labeling, the boundary word of τ_h is w.

A filling of the triangles in τ_h thus gives a filling of w. As in Lemma 5.1, each triangle of τ_h either has short edges and thus a bounded filling area, or has vertices whose labels lie in a translate of a parabolic subgroup. In this case, that parabolic subgroup must be $U(1,1)$, and Lemma 7.6 allows us to fill any such triangle with quadratic area. Corollary 5.3(3) thus implies that $\delta(w) = O(\ell(w)^2)$, as desired.

10. **Open questions**

One natural open question is whether these results can be extended to a proof of Conjecture 1.1 or, as an important special case, whether they can be used to find a bound on the Dehn function of $\mathrm{SL}(4;\mathbb{Z})$. Some parts of the proof, especially the geometric lemmas in Section 5, extend naturally to other lattices in semisimple groups. That is, if Γ acts on a symmetric space $\mathcal{E}$, one can define a fundamental set $\mathcal{S}$ that is a union of Siegel sets, use $\mathcal{S}$ to define a map $\rho : \mathcal{E} \to \Gamma$, and show that if x and y are close together and deep in a cusp, then $\rho(x)$ and $\rho(y)$ lie in a coset of some parabolic subgroup. Using this fact, one can find various ways to construct templates whose triangles have vertices lying in parabolic subgroups.

For $\mathrm{SL}(p;\mathbb{Z})$, we filled these triangles using combinatorial lemmas, but these lemmas are hard to generalize to other groups. In general, appropriate analogues of Lemmas 7.6 and 7.1 should lead to a polynomial bound on the Dehn function of a lattice. One way to get such a bound is to use a template consisting of simplices all of size ~ 1, as in [You]. In this case, each edge can be labeled by a group element that lies in a parabolic subgroup. By the Langlands decomposition, this parabolic subgroup has a reductive part and a unipotent part, and the group element is the product of a bounded element of the reductive part and an exponentially large unipotent element. Lemmas that conjugate unipotent elements by reductive elements will then suffice to fill the resulting ω-triangles.

For $\mathrm{SL}(4;\mathbb{Z})$, we can say a little more. One of the main advantages of using $\mathrm{SL}(p;\mathbb{Z})$ here is that when p is large, it contains many solvable subgroups (the $H_{S,T}$'s defined in Section 6) with quadratic Dehn functions and large intersections; this is one thing allowing us to prove, for example Lemma 7.1. This gets more difficult for small p because the solvable groups and their intersections get smaller.

For example, when $p \geq 6$, Lemma 7.1 can be proved in a few lines: Let $\gamma_{S,T}$ be a geodesic connecting I and $e_{1,6}(x)$ in $H_{S,T}$. As long as $\#S \geq 2$ or $\#T \geq 2$, this has length $\sim \log|x|$. We can then construct a homotopy from, say, $\gamma_{\{1,2,3\},\{6\}}$ to $\gamma_{\{1,3,4\},\{6\}}$ that goes through the stages

$$\gamma_{\{1,2,3\},\{6\}} \to \gamma_{\{1\},\{5,6\}} \to \gamma_{\{1,3,4\},\{6\}}.$$

In the first step, we use the fact that both curves lie in $H_{\{1,2,3\},\{5,6\}}$, which has quadratic Dehn function; likewise, in the second step, we use the fact that both curves lie in $H_{\{1,3,4\},\{5,6\}}$. The full lemma can be proved in the same way. When $p = 5$, however, the lemma is more difficult to prove, because the overlaps between solvable subgroups are smaller, and when $p = 4$, the lemma is not known. In fact, Lemma 7.1 is the main obstacle to proving a polynomial bound on the Dehn function of $\mathrm{SL}(4;\mathbb{Z})$. In unpublished work, I have reduced the problem of bounding the Dehn function of the whole group by a polynomial to the problem of proving that $\delta_{\mathrm{SL}(4;\mathbb{Z})}(\hat{e}_{ij}(x), \hat{e}_{ij;S}(x))$ is bounded by a polynomial in the length of $\hat{e}_{ij}(x)$.

Similarly, one may ask about higher-order filling inequalities in arithmetic groups. These filling inequalities generalize the Dehn function, but instead of bounding the area of a disc filling a curve γ of a given length, they bound the $(k+1)$-volume of a chain filling a cycle of a given k-volume. Gromov stated a generalization of Conjecture 1.1 to this situation

CONJECTURE 10.1. *If Γ is an irreducible lattice in a symmetric space with $\mathbb{R}$-rank at least $k+2$, then any k-cycle of volume V has a filling by a k-chain of volume polynomial in V.*

Bestvina, Eskin, and Wortman [BEW] have made partial progress toward a more general conjecture stated in terms of volume distortion in lattices.

References

[Beh79] H. BEHR, $\mathrm{SL}_3(\mathbf{F}_q[t])$ is not finitely presentable, in *Homological Group Theory* (Proc. Sympos., Durham, 1977), *London Math. Soc. Lecture Note Ser.* no. 36, Cambridge Univ. Press, Cambridge, 1979, pp. 213–224. MR 0564424. Zbl 0434.20025.

[BEW] M. BESTVINA, A. ESKIN, and K. WORTMAN, Filling boundaries of coarse manifolds in semisimple and solvable arithmetic groups. arXiv 1106.0162.

[BHC62] A. BOREL and HARISH-CHANDRA, Arithmetic subgroups of algebraic groups, *Ann. of Math.* **75** (1962), 485–535. MR 0147566. Zbl 0107.14804. http://dx.doi.org/10.2307/1970210.

[Bri02] M. R. BRIDSON, The geometry of the word problem, in *Invitations to Geometry and Topology*, *Oxf. Grad. Texts Math.* **7**, Oxford Univ. Press, Oxford, 2002, pp. 29–91. MR 1967746. Zbl 0996.54507.

[BW07] K.-U. BUX and K. WORTMAN, Finiteness properties of arithmetic groups over function fields, *Invent. Math.* **167** (2007), 355–378. MR 2270455. Zbl 1126.20030. http://dx.doi.org/10.1007/s00222-006-0017-y.

[dCT10] Y. DE CORNULIER and R. TESSERA, Metabelian groups with quadratic Dehn function and Baumslag-Solitar groups, *Confluentes Math.* **2** (2010), 431–443. MR 2764812. Zbl 05849551. http://dx.doi.org/10.1142/S1793744210000235.

[Din94] J. T. DING, A proof of a conjecture of C. L. Siegel, *J. Number Theory* **46** (1994), 1–11. MR 1268360. Zbl 0797.11058. http://dx.doi.org/10.1006/jnth.1994.1001.

[Dru04] C. DRUŢU, Filling in solvable groups and in lattices in semisimple groups, *Topology* **43** (2004), 983–1033. MR 2079992. Zbl 1083.20033. http://dx.doi.org/10.1016/j.top.2003.11.004.

[ECH+92] D. B. A. EPSTEIN, J. W. CANNON, D. F. HOLT, S. V. F. LEVY, M. S. PATERSON, and W. P. THURSTON, *Word Processing in Groups*, Jones and Bartlett Publishers, Boston, MA, 1992. MR 1161694. Zbl 0764.20017.

[Ger93] S. M. Gersten, Isoperimetric and isodiametric functions of finite presentations, in *Geometric Group Theory*, *Vol.* 1 (Sussex, 1991), *London Math. Soc. Lecture Note Ser.* **181**, Cambridge Univ. Press, Cambridge, 1993, pp. 79–96. MR 1238517. Zbl 0829.20054. http://dx.doi.org/10.1017/CBO9780511661860.008.

[Gro] C. Groft, Generalized Dehn functions I. arXiv 0901.2303.

[Gro93] M. Gromov, Asymptotic invariants of infinite groups, in *Geometric Group Theory*, *Vol.* 2 (Sussex, 1991), *London Math. Soc. Lecture Note Ser.* **182**, Cambridge Univ. Press, Cambridge, 1993, pp. 1–295. MR 1253544. Zbl 0841.20039.

[Gro96] M. Gromov, Carnot-Carathéodory spaces seen from within, in *Sub-Riemannian Geometry*, *Progr. Math.* **144**, Birkhäuser, Basel, 1996, pp. 79–323. MR 1421823. Zbl 0864.53025. http://dx.doi.org/10.1007/978-3-0348-9210-0_2.

[JM02] L. Ji and R. MacPherson, Geometry of compactifications of locally symmetric spaces, *Ann. Inst. Fourier* (*Grenoble*) **52** (2002), 457–559. MR 1906482. Zbl 1017.53039. Available at http://aif.cedram.org/item?id=AIF_2002__52_2_457_0.

[Ji98] L. Ji, Metric compactifications of locally symmetric spaces, *Internat. J. Math.* **9** (1998), 465–491. MR 1635185. Zbl 0929.32017. http://dx.doi.org/10.1142/S0129167X98000208.

[LP96] E. Leuzinger and C. Pittet, Isoperimetric inequalities for lattices in semisimple Lie groups of rank 2, *Geom. Funct. Anal.* **6** (1996), 489–511. MR 1392327. Zbl 0856.22013. http://dx.doi.org/10.1007/BF02249261.

[LP04] E. Leuzinger and C. Pittet, On quadratic Dehn functions, *Math. Z.* **248** (2004), 725–755. MR 2103539. Zbl 1132.20021. http://dx.doi.org/10.1007/s00209-004-0678-4.

[Leu04a] E. Leuzinger, On polyhedral retracts and compactifications of locally symmetric spaces, *Differential Geom. Appl.* **20** (2004), 293–318. MR 2053916. Zbl 1052.22008. http://dx.doi.org/10.1016/j.difgeo.2004.03.001.

[Leu04b] E. Leuzinger, Tits geometry, arithmetic groups, and the proof of a conjecture of Siegel, *J. Lie Theory* **14** (2004), 317–338. MR 2066859. Zbl 1086.53073.

[LMR93] A. Lubotzky, S. Mozes, and M. S. Raghunathan, Cyclic subgroups of exponential growth and metrics on discrete groups, *C. R. Acad. Sci. Paris Sér. I Math.* **317** (1993), 735–740. MR 1244421. Zbl 0786.22016.

[Mil71] J. Milnor, *Introduction to Algebraic K-Theory*, *Ann. of Math. Studies* **72**, Princeton Univ. Press, Princeton, N.J., 1971. MR 0349811. Zbl 0237.18005.

[Ril05] T. R. Riley, Navigating in the Cayley graphs of $\mathrm{SL}_N(\mathbb{Z})$ and $\mathrm{SL}_N(\mathbb{F}_p)$, *Geom. Dedicata* **113** (2005), 215–229. MR 2171306. Zbl 1110.20026. http://dx.doi.org/10.1007/s10711-005-5230-0.

[Ste62] R. Steinberg, Générateurs, relations et revêtements de groupes algébriques, in *Colloq. Théorie des Groupes Algébriques* (Bruxelles, 1962), Librairie Universitaire, Louvain, 1962, pp. 113–127. MR 0153677. Zbl 0272.20036.

[You] R. Young, A polynomial isoperimetric inequality for SL(n,Z). arXiv 0903.2495.

(Received: December 15, 2009)
(Revised: May 21, 2012)

University of Toronto, Toronto, ON, Canada
E-mail: ryoung@math.toronto.edu