ISSN 1364-0380 (on line) 1465-3060 (printed)

Geometry & Topology
Volume 8 (2004) 1301–1359
Published: 19 October 2004

On groups generated by two positive multi-twists: Teichmüller curves and Lehmer's number

Christopher J Leininger

Department of Mathematics, Columbia University
2990 Broadway MC 4448, New York, NY 10027, USA

Email: `clein@math.columbia.edu`

Abstract

From a simple observation about a construction of Thurston, we derive several interesting facts about subgroups of the mapping class group generated by two positive multi-twists. In particular, we identify all configurations of curves for which the corresponding groups fail to be free, and show that a subset of these determine the same set of Teichmüller curves as the non-obtuse lattice triangles which were classified by Kenyon, Smillie, and Puchta. We also identify a pseudo-Anosov automorphism whose dilatation is Lehmer's number, and show that this is minimal for the groups under consideration. In addition, we describe a connection to work of McMullen on Coxeter groups and related work of Hironaka on a construction of an interesting class of fibered links.

AMS Classification numbers Primary: 57M07, 57M15

Secondary: 20H10, 57M25

Keywords: Coxeter, Dehn twist, Lehmer, pseudo-Anosov, mapping class group, Teichmüller

Proposed: Benson Farb Received: 16 February 2004
Seconded: Walter Neumann, Joan Birman Revised: 17 August 2004

1 Introduction

Let S be a connected finite type oriented surface. In $\mathrm{Mod}(S)$, the mapping class group of S, a particularly tractable class of elements (or *automorphisms*) are the *positive multi-twists*. These are products of positive Dehn twists about disjoint essential simple closed curves. For a given positive multi-twist, the union of these simple closed curves is a closed essential 1–manifold, and the set of positive multi-twists is in a one-to-one correspondence with $\mathcal{S}'(S)$, the set of isotopy classes of essential 1–manifolds on S. Given $A \in \mathcal{S}'(S)$, we let T_A denote the positive multi-twist which is the product of positive Dehn twists about the components of A.

This paper is concerned with subgroups of $\mathrm{Mod}(S)$ generated by two positive multi-twists and is based on a construction of Thurston [54] (see also Long [36] and Veech [56]). When $A \cup B$ *fills* the surface (that is, every essential curve intersects A or B) Thurston constructs a certain type of Euclidean cone metric, which we refer to as a *flat structure*, for which $\langle T_A, T_B \rangle$ acts by affine homeomorphisms. The derivative of this action defines a discrete homomorphism $\mathrm{DAf}\colon \langle T_A, T_B \rangle \to \mathrm{PSL}_2(\mathbb{R})$ with finite kernel. This homomorphism is determined by a single number, $\mu(A \cup B)$, depending on the geometric intersection numbers of the components of A with those of B.

The novelty in our approach to studying these groups is Proposition 5.1 in which we show that $\mu(A \cup B)$ is the spectral radius of the *configuration graph*, $\mathcal{G}(A \cup B)$. This graph has a vertex for each component of A and of B and an edge for every point of intersection between corresponding components (see Figure 1).

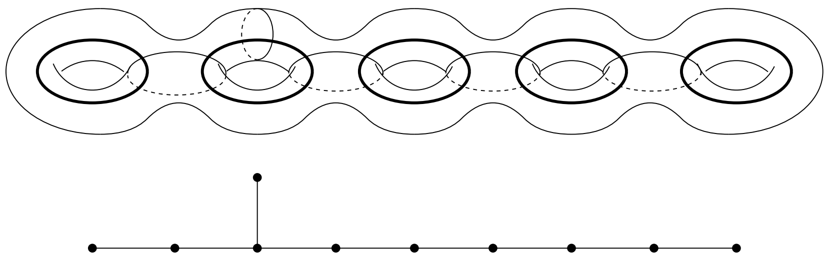

Figure 1: 1–manifolds A_L and B_L with configuration graph $\mathcal{G}(A_L \cup B_L) = \mathcal{E}h_{10}$

This observation, along with some elementary hyperbolic geometry and well-known graph theoretic results, has many interesting consequences.

1.1 Freeness

The graphs of type $\mathcal{A}_c$ $(c \geq 1)$, $\mathcal{D}_c$ $(c \geq 4)$, $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$ play an important role in our work, and we refer to them as *recessive* graphs (see Figures 2–4). Any graph which is not recessive will be called *dominant.*

Figure 2: $\mathcal{A}_c$, $c \geq 1$

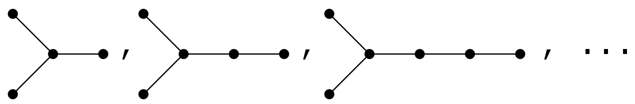

Figure 3: $\mathcal{D}_c$, $c \geq 4$

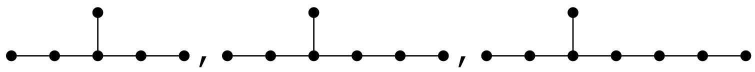

Figure 4: $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$

Theorem 6.1 *$\langle T_A, T_B\rangle \cong \mathbb{F}_2$ if and only if $\mathcal{G}(A \cup B)$ contains a dominant component.*

This theorem was inspired by the work of Hamidi-Tehrani in [26] where sufficient conditions for $\langle T_A, T_B\rangle \cong \mathbb{F}_2$ in terms of intersection numbers of components of A with those of B are given. In [54], Thurston remarks (without proof) that a necessary and sufficient condition for this group to be free is that $\mu(A \cup B) \geq 2$, and this is the basis for Theorem 6.1.

1.2 Teichmüller curves

The proof of Theorem 6.1 reduces to the case that $A \cup B$ fills S (see Proposition 10.1 and Section 10), and we assume this to be the case for the remainder of Section 1.

The flat structure on S determines a quadratic differential and thus a Teichmüller disk (see Section 3). The group $\langle T_A, T_B\rangle$ stabilizes this disk, though in general this group has infinite index in the full stabilizer. The quotient of a Teichmüller disk by its stabilizer is called a Teichmüller curve when it has finite area, that is when the stabilizer is a lattice. In this case, the Teichmüller curve isometrically immerses into the moduli space of S, and we say that the Teichmüller disk covers a Teichmüller curve.

In Zemljakov and Katok [61] it is shown how to associate a Riemann surface and quadratic differential to a rational polygon in such a way that billiards trajectories in the polygon correspond to geodesics for the associated flat structure (see Section 3.4). A theorem of Veech [56] implies that the billiards in a polygon have optimal dynamical properties if the corresponding Teichmüller disk covers a Teichmüller curve, in which case the polygon is called a *lattice polygon*. In particular, understanding and classifying Teichmüller curves and lattice polygons is an interesting problem which has received much attention (see eg Veech [56], [57], Harvey [27], Gutkin and Judge [25], Kenyon and Smillie [32], Puchta [48], McMullen [44], and Calta [13]). Our second main theorem provides a complete classification for a certain class of Teichmüller curves.

Theorem 7.1 *The Teichmüller curves for which the associated stabilizers contain a group generated by two positive multi-twists with finite index are precisely those defined by $A \cup B$ filling S, where $\mathcal{G}(A \cup B)$ is critical or recessive.*

The critical graphs are those of type $\mathcal{P}_{2c}$ $(c \geq 1)$, $\mathcal{Q}_c$ $(c \geq 5)$, $\mathcal{R}_7$, $\mathcal{R}_8$, and $\mathcal{R}_9$ (see Figures 5 – 7).

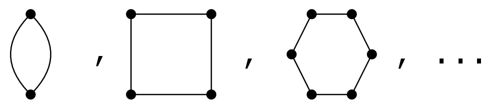

Figure 5: $\mathcal{P}_{2c}$, $c \geq 1$

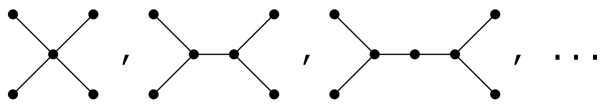

Figure 6: $\mathcal{Q}_c$, $c \geq 5$

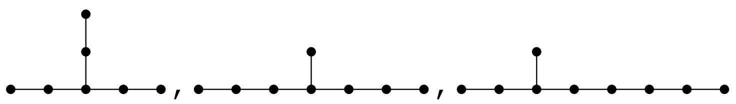

Figure 7: $\mathcal{R}_7$, $\mathcal{R}_8$, and $\mathcal{R}_9$

A classification of right and acute lattice triangles was initiated by Kenyon and Smillie in [32] and completed by Puchta in [50] (see Theorem 3.5). Using this classification, we prove the following theorem.

Theorem 7.2 *The Teichmüller curves determined by the right and acute lattice triangles have associated stabilizers containing a finite index subgroup of the form $\langle T_A, T_B \rangle$ with $\mathcal{G}(A \cup B)$ recessive.*

Moreover, all the flat structures associated to $\mathcal{G}(A \cup B)$ of type $\mathcal{A}_c$ and $\mathcal{D}_c$ are affine equivalent to structures which can be tiled by one or two regular Euclidean polygons (see Section 7). These were all studied by Veech in [56] and [57] and by Earle and Gardiner in [17]. The remaining three cases where $\mathcal{G}(A \cup B)$ has type $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$ correspond to the exceptional triangles mentioned above and were studied by Veech in [56], Vorobets in [58], and Kenyon and Smillie in [32]. In particular, in those cases that $\langle T_A, T_B \rangle$ fails to be free, the aforementioned references provide a description of these groups.

Theorem 7.3 *If $\mathcal{G}(A \cup B)$ is recessive, then* DAf *maps $\langle T_A, T_B \rangle$ onto a Fuchsian triangle group with finite central kernel of order at most 2. The signature of the triangle group is described by the following table.*

configuration graph	*signature*	*configuration graph*	*signature*
$\mathcal{D}_c$, $c \geq 4$	$(c-1, \infty, \infty)$	$\mathcal{E}_6$	$(6, \infty, \infty)$
$\mathcal{A}_{2c+1}$, $c \geq 1$	$(c+1, \infty, \infty)$	$\mathcal{E}_7$	$(9, \infty, \infty)$
$\mathcal{A}_{2c}$, $c \geq 1$	$(2, 2c+1, \infty)$	$\mathcal{E}_8$	$(15, \infty, \infty)$

1.3 Lehmer's number and Coxeter groups

The original purpose of Thurston's construction was not to study groups generated by two positive multi-twists, but rather to construct explicit examples of *pseudo-Anosov* automorphisms. Indeed, in these groups, pseudo-Anosov automorphisms are generic (see Proposition 6.4). Associated to a pseudo-Anosov automorphism, ϕ, is an algebraic integer $\lambda(\phi) > 1$ called the *dilatation* reflecting certain dynamical properties (see Section 2.5).

Theorem 6.2 *For any surface S, any $A, B \in \mathcal{S}'(S)$, and any pseudo-Anosov element*

$$\phi \in \langle T_A, T_B \rangle < \mathrm{Mod}(S)$$

we have $\lambda(\phi) \geq \lambda_L \approx 1.1762808$. Moreover, $\lambda(\phi) = \lambda_L$ precisely when S has genus 5 (with at most one marked point), $\{A, B\} = \{A_L, B_L\}$ as in Figure 1 (up to homeomorphism), and ϕ is conjugate to $(T_A T_B)^{\pm 1}$.

Here λ_L is *Lehmer's number* which is the largest real root of *Lehmer's polynomial*:

$$x^{10}+x^9-x^7-x^6-x^5-x^4-x^3+x+1 \tag{1}$$

λ_L was discovered by Lehmer in 1933 [35] and is the smallest known Salem number and Mahler measure of an integral polynomial (see Section 9.1).

Given any $\mathcal{G}=\mathcal{G}(A\cup B)$, we can view this as a Coxeter graph, and we let $\mathfrak{C}(\mathcal{G})$ and $\mathfrak{A}(\mathcal{G})$ denote the associated *Coxeter group* and *Artin group*, respectively, with $\pi_{ac}\colon \mathfrak{A}(\mathcal{G})\to\mathfrak{C}(\mathcal{G})$ the canonical epimorphism. We let Θ denote the geometric action of $\mathfrak{C}(\mathcal{G})$ on $(\mathbb{R}^K,\Pi_{\mathcal{G}})$ (see Section 8).

Theorem 6.2 is strikingly similar to the main theorem of [43] (see Theorem 8.4) in which McMullen shows that the minimal spectral radius of any essential element in a Coxeter group with respect to Θ is either 1 or else bounded below by λ_L. Moreover, λ_L is achieved precisely when the associated Coxeter graph is $\mathcal{E}h_{10}$.

We say $\mathcal{G}$ has *small type* if there are no multiple edges between vertices. It is well known that there is a homomorphism

$$\Psi\colon \mathfrak{A}(\mathcal{G}(A\cup B))\to \mathrm{Mod}(S)$$

sending the standard generators to the corresponding Dehn twists in the A and B curves, when $\mathcal{G}(A\cup B)$ has small type (see Section 8.2). This provides the first link with the groups under consideration. The following describes the connection with McMullen's Theorem 8.4.

Theorem 8.1 *Let $\mathcal{G}(A\cup B)$ be non-critical dominant with small type. Then $\sigma_A\sigma_B$ is sent by Ψ to a pseudo-Anosov with dilatation equal to the spectral radius of its image under $\Theta\circ\pi_{ac}$. Moreover, among all essential elements in $\langle\sigma_A,\sigma_B\rangle$, $\sigma_A\sigma_B$ minimizes both dilatation as well as spectral radius for the respective homomorphisms.*

In this theorem, $\sigma_A\sigma_B$ is the *bicolored Coxeter element* (see Section 8.1). Inspired by the work of Hironaka in [28] (see Theorem 8.5), we find that under some additional hypothesis, (part of) the action of T_AT_B on $H_1(S;\mathbb{R})$ is almost semi-conjugate to the geometric action of this Coxeter element.

Theorem 8.2 *Let $\mathcal{G}(A\cup B)$ have small type and suppose that A and B can be oriented so that all intersections of A with B are positive. Then there exists a homomorphism*

$$\eta\colon \mathbb{R}^K\to H_1(S;\mathbb{R})$$

such that

$$(T_A T_B)_* \circ \eta = -\eta \circ \Theta(\sigma_A \sigma_B).$$

Moreover, $\Theta(\sigma_A\sigma_B)|_{\ker(\eta)} = -I$ *and* η *preserves spectral radii.*

To say that η preserves spectral radii, we simply mean that $\Theta(\sigma_A\sigma_B)$ and $(T_AT_B)_*$ have the same spectral radius (modulus of leading eigenvalue). If we wish to relate T_AT_B to $\sigma_A\sigma_B$, this theorem is likely the best we can do. For, unlike Hironaka's Theorem 8.5, there is no relation between $\dim(H_1(S;\mathbb{R}))$ and K.

Remarks

(1) Theorem 6.1 allows the possibility that $A \cup B$ does not fill S, while we implicitly assume this for the other theorems. We also note that Theorem 6.1 holds when S has nonempty boundary (see Section 10).

(2) We caution the reader that the connection to Coxeter groups we have described is only valid when the configuration graph has small type. This is easily explained by the fact that the adjacency matrix for a graph is the same as the *Coxeter adjacency matrix* only when the graph has small type.

The paper is organized as follows. Sections 2 and 3 contain definitions and theorems regarding surface topology, mapping class groups, and Teichmüller space. We recall the relevant facts concerning matrices and graphs in Section 4. In Section 5 we give Thurston's construction and prove Proposition 5.1 relating this to the spectral radius of the configuration graph. Next we discuss some basics of Fuchsian groups and use them to prove Theorems 6.1 and 6.2 in Section 6. In Section 7 we discuss in more detail the groups corresponding to the critical and recessive configurations and prove Theorems 7.1, 7.2, and 7.3. We then turn to Coxeter and Artin groups in Section 8, describe the connection with groups generated by two positive multi-twists, and prove Theorems 8.1 and 8.2. In Section 9 we provide a few applications of the theorems and indicate some interesting open questions.

We have also included two appendices. The first, Section 10, reduces the proof of Theorem 6.1 to the filling case, as well as extending it to the situation of surfaces with boundary. The second, Section 11, addresses a construction of pseudo-Anosov automorphisms given by Penner which extends Thurston's construction. For completeness, we show that the lower bound given by Theorem 6.2 holds for this class as well.

Acknowledgements

I would like to thank Norbert A'Campo, Joan Birman, Hessam Hamidi-Tehrani, Curt McMullen, Walter Neumann, and Alan Reid for helpful conversations and some very useful and interesting references which not only simplified the exposition, but which led to some of the connections to other areas of mathematics. Thanks also to Dan Margalit and the referees for carefully reading earlier versions of this paper and providing many helpful suggestions. The author was partially supported by an N.S.F. postdoctoral fellowship.

2 Surface topology and mapping class groups

For more details on the material reviewed in this section, see Thurston [54], Birman [8], Ivanov [31], and the lecture notes [18].

2.1 Surfaces and essential 1–manifolds

Let $S = S_{g,p}$ be a smooth, compact, connected, oriented, genus–g surface with p marked points. We will ignore the trivial cases, and hence from this point on assume $S \neq S_{0,p}$ for $p \leq 3$. Denote by $\dot{S}$ the surface S with the p marked points removed.

We denote the set of isotopy classes of essential simple closed curves on S by $\mathcal{S}(S)$. That is, an element of $\mathcal{S}(S)$ is an isotopy class of homotopically essential simple closed curves on $\dot{S}$ not isotopic to a puncture of $\dot{S}$. The geometric intersection number for a pair of elements $a, b \in \mathcal{S}(S)$, denoted $i(a,b)$, is the minimal number of transverse intersection points among all representatives of a and of b.

Let $\mathcal{S}'(S)$ denote the set of isotopy classes of essential, closed 1–manifolds embedded in S. An element $A \in \mathcal{S}'(S)$ is an embedded 1–submanifold of $\dot{S}$, for which every component is homotopically essential in $\dot{S}$ and not isotopic to a puncture, well-defined up to isotopy. We will make no distinction between 1–manifolds and the isotopy classes they represent when convenient. We refer to the components of A as elements of $\mathcal{S}(S)$. Whenever we write $A = a_1 \cup \cdots \cup a_n$ it will be assumed that $a_i \in \mathcal{S}(S)$, for each $i = 1, \ldots, n$.

Note that an element of $\mathcal{S}'(S)$ is allowed to have several of its components parallel (isotopic) to one another.

When considering elements $A, B \in \mathcal{S}'(S)$ as representative 1–manifolds of their isotopy classes, we will always assume that they meet transversely and minimally. When this is done, a component a of A and b of B meet in exactly $i(a,b)$ points. Consequently, the configuration graph $\mathcal{G}(A \cup B)$ depends only on A and B, and its components are in a one-to-one correspondence with those of $A \cup B$, thought of as a subset of S.

We further note that $\mathcal{G}(A \cup B)$ is *bipartite*. That is, the vertices may be colored by two colors (call them A and B) so that no two vertices of the same color are adjacent.

When $A \cup B$ fills S (see Section 1) it follows that the components of $S \setminus (A \cup B)$ are disks (each with at most one marked point). Note that $\mathcal{G}(A \cup B)$ is connected when $A \cup B$ fills S.

2.2 Uniqueness

Given a bipartite graph $\mathcal{G}$, there may be several different pairs of 1–manifolds having $\mathcal{G}$ as the configuration graph. Indeed, $\mathcal{G}$ need not even determine the homeomorphism type of the underlying surface. However, there are instances in which one does have uniqueness.

Proposition 2.1 *Suppose $A_i, B_i \in \mathcal{S}'(S_i)$, $A_i \cup B_i$ filling S_i, $i = 1, 2$, and $\mathcal{G} = \mathcal{G}(A_1 \cup B_1) = \mathcal{G}(A_2 \cup B_2)$. If $\mathcal{G}$ is a tree with only one vertex of valence at most three (in particular, if it is recessive), then there is a homeomorphism from S_1 to S_2 taking $\{A_1, B_1\}$ to $\{A_2, B_2\}$, up to adding marked points.*

Sketch of proof Let $\mathcal{N}(A_i \cup B_i)$ denote a regular neighborhood of $A_i \cup B_i$ in S_i. S_i is obtained from $\mathcal{N}(A_i \cup B_i)$ by adding disks with zero or one marked point each, and there is just one way to do this, up to homeomorphism. So, to find a homeomorphism from S_1 to S_2, it suffices to find a homeomorphism from $\mathcal{N}(A_1 \cup B_1)$ to $\mathcal{N}(A_2 \cup B_2)$. We view $\mathcal{N}(A_i \cup B_i)$ as a union of annular neighborhoods of the components of A_i and B_i with pairs of annuli intersecting in squares if the corresponding curves intersect, and otherwise not at all (see Figure 8). Thus, in each surface, we have an annulus associated to each vertex of $\mathcal{G}(A_i \cup B_i)$ and a square of intersections of annuli for each edge.

We first define a homeomorphism on the annulus corresponding to the three-valent vertex, v, (if it exists, otherwise, we can start at any vertex). We do this so that the three (or fewer) squares corresponding to the edges meeting v

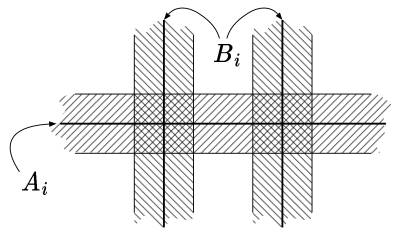

Figure 8: Pieces of annuli intersecting in squares

are taken to the squares corresponding to the same three edges. This is possible because homeomorphisms of the circle act transitively on ordered triples of points (and so homeomorphisms of an annulus act transitively on ordered triples of disjoint squares). Next we extend the homeomorphism over the annuli corresponding to the vertices adjacent to v, again preserving the squares corresponding to the edges meeting those vertices. The homeomorphism is already defined on one of these squares and now there is at most one other square (since these vertices have valence at most 2), so this is possible as well. We may continue in this way extending over annuli corresponding to adjacent vertices, preserving intersection squares. At each stage, we only encounter vertices with valence at most 2, so this is always possible.

Since there are only finitely many vertices, after finitely many steps we obtain a homeomorphism from $\mathcal{N}(A_1 \cup B_1)$ to $\mathcal{N}(A_2 \cup B_2)$, taking annuli to annuli. The union of the cores of these annuli is precisely $A_i \cup B_i$, so applying an isotopy if necessary, we may assume that $A_1 \cup B_1$ is taken to $A_2 \cup B_2$. $\square$

2.3 Automorphisms

We say that a homeomorphism $\phi\colon S \to S$ is *allowable* if it preserves the marked points. We denote the group of allowable, orientation preserving homeomorphisms of S by $\mathrm{Homeo}_+(S)$ and the identity component by $\mathrm{Homeo}_0(S)$. The *mapping class group* is defined to be the quotient group

$$\mathrm{Mod}(S) = \mathrm{Homeo}_+(S)/\mathrm{Homeo}_0(S).$$

An element of $\mathrm{Mod}(S)$ is referred to as an *automorphism* of S, and by definition is a homeomorphism, well-defined up to isotopy. When no confusion can arise, we will make no distinction between a homeomorphism and the automorphism it determines.

2.4 Multi-twists

Given $a \in \mathcal{S}(S)$, a *positive Dehn twist* is the isotopy class of a homeomorphism supported in an annular neighborhood of a described as follows. If we identify the annular neighborhood of a with the annulus $\mathbb{R}/\tau\mathbb{Z} \times [0,\sigma]$ by an orientation preserving homeomorphism, then with respect to the obvious coordinates on this annulus, (t,s), the Dehn twist is given by

$$(t,s) \mapsto \left(t + s\frac{\tau}{\sigma}, s\right) \tag{2}$$

We note that this makes the Dehn twist affine with respect to the natural Euclidean metric on the annulus for any $\tau, \sigma > 0$.

Given $A = a_1 \cup \cdots \cup a_n \in \mathcal{S}'(S)$, a *multi-twist* along A is the product

$$T_{a_1}^{\epsilon_1} \cdots T_{a_n}^{\epsilon_n}$$

where $\epsilon_i \in \{\pm 1\}$. The *positive multi-twist* along A, written T_A, is given by the above product where $\epsilon_i = 1$ for each $i = 1, \dots, n$. The map $A \mapsto T_A$ determines a bijection between $\mathcal{S}'(S)$ and the set of positive multi-twists.

In the definition of T_A the order of the product does not matter since Dehn twists in disjoint curves obviously commute. In fact, for $a, b \in \mathcal{S}(S)$, we have

$$\begin{array}{rcr} i(a,b) = 0 & \Rightarrow & T_aT_b = T_bT_a \\ i(a,b) = 1 & \Rightarrow & T_aT_bT_a = T_bT_aT_b \end{array} \tag{3}$$

The second equality is the well known *braid relation* and easily follows from the calculation that $T_aT_b(a) = b$.

Remark One often requires only that $\epsilon_i \in \mathbb{Z}_+$ in the definition of a positive multi-twist. However, we may replace a power of a Dehn twist about a curve a by a product of Dehn twist about several isotopic copies of a, so there is no loss in generality in taking only the powers 1 in our definition.

2.5 Pseudo-Anosov automorphisms

An automorphism $[\phi] \in \mathrm{Mod}(S)$ is called *pseudo-Anosov* if there is a representative ϕ which leaves invariant a pair of mutually transverse singular foliations with the following property. These foliations admit transverse measures and ϕ multiplies one measure by a factor $\lambda > 1$ and the other by λ^{-1}. The number $\lambda = \lambda([\phi]) = \lambda(\phi) > 1$ is an algebraic integer called the *dilatation* of $[\phi]$.

The dilatation has the following geometric description. For any $a \in \mathcal{S}(S)$ and any complete hyperbolic metric on S, the length of the geodesic representative of $\phi^n(a)$ grows like λ^n. That is, $\lambda^{-n}\mathrm{length}(\phi^n(a))$ converges to a nonzero number.

2.6 Reduction to the filling case

As was mentioned in the introduction, Theorem 6.1 is valid for any surface and any pair of 1–manifolds, but the proof reduces to the filling case. A proof of the following is given in the Appendix, in Section 10.

Proposition 10.1 *It suffices to prove Theorem 6.1 for $A \cup B$ filling S.*

Convention For the remainder of this paper (excluding Section 10) we shall assume that every pair of essential 1–manifolds is filling.

3 Teichmüller and moduli spaces

For more details on Teichmüller space and quadratic differentials see Gardiner and Lakic [23], Masur [40], Masur and Tabachnikov [41], Earle and Gardiner [17], and McMullen [44].

Consider the space of complex structures on S, with orientation compatible with the given orientation. $\mathrm{Homeo}_+(S)$ acts on this space, and the quotient is called the *moduli space* of S and is denoted $\mathcal{M}(S)$. If we quotient by the action of the subgroup $\mathrm{Homeo}_0(S)$ the resulting space is called the *Teichmüller space* of S, and is denoted $\mathcal{T}(S)$. $\mathcal{T}(S)$ is the universal orbifold covering of $\mathcal{M}(S)$, with covering group $\mathrm{Mod}(S)$.

Given $[J_0], [J_1] \in \mathcal{T}(S)$, the Teichmüller distance is defined by

$$d([J_0],[J_1]) = \frac{1}{2} \inf_{f \simeq Id_S} \log\left(K\left(f\colon\ (S,J_0) \to (S,J_1)\right)\right),$$

where the infimum is taken over all quasi-conformal homeomorphisms f isotopic to the identity, and $K(f)$ is the dilatation of f. The action of $\mathrm{Mod}(S)$ is by isometries, and so the metric pushes down to $\mathcal{M}(S)$.

3.1 Quadratic differentials

Let $[J] \in \mathcal{T}(S)$, and consider the space $\mathcal{Q}(S,J)$ of integrable meromorphic quadratic differentials on (S,J) which are holomorphic on $\dot{S}$. Any $q \in \mathcal{Q}(S,J)$ determines a singular Euclidean metric $|q|$ on S with cone-type singularities having cone angles $k\pi$ for $k \in \mathbb{Z}_{\geq 3}$ at non-marked points and $k \in \mathbb{Z}_{\geq 1}$ at marked points. It also defines a singular measured foliation $\mathcal{F}_h$, called the *horizontal foliation*, whose leaves are geodesic with respect to $|q|$. These leaves are precisely the injectively immersed 1–manifolds γ satisfying $q(\gamma'(t)) \geq 0$. We refer to this structure as a *flat structure*, and will also denote it by q.

When all the leaves of $\mathcal{F}_h$ are compact, the complement of the singular leaves is a disjoint union of annuli. In this situation, we say that q (or $\mathcal{F}_h$) determines an annular decomposition of S.

Suppose now we are given a flat structure q. That is, we have a singular Euclidean metric $|q|$ (having the above types of singularities) and a singular foliation $\mathcal{F}_h$ with geodesic leaves. This defines a complex structure J and quadratic differential which can be described as follows. The singular Euclidean metric is given by an atlas of charts into $\mathbb{C}$ on the complement of the singularities for which the transition functions are Euclidean isometries. This defines a complex structure on the complement of the singularities which then extends over this finite set. Requiring that the leaves of $\mathcal{F}_h$ be sent to horizontal lines by our charts restricts our transition functions to be of the form $z \mapsto \pm z + \xi$, for some $\xi \in \mathbb{C}$. We refer to such an atlas of charts as a *preferred atlas* for q. The form dz^2 is invariant under the transition functions and pulls back to the desired quadratic differential. The horizontal foliation is precisely $\mathcal{F}_h$.

We also obtain a locally defined orthonormal basis e_1, e_2 for the tangent space to any non-singular point such that e_1 is tangent to $\mathcal{F}_h$. Away from the singularities, this basis is globally well-defined by this condition, up to sign (ie by replacing $\{e_1, e_2\}$ by $\{-e_1, -e_2\}$). The dual basis $\{e^1, e^2\}$ locally defines a holomorphic 1–form $\omega = e^1 + ie^2$. Although ω is not in general globally well-defined, its square is, and this is precisely the quadratic differential $q = \omega^2$. Note that ω is globally defined precisely when the metric has no holonomy.

3.2 Teichmüller disks and curves

Given $[J] \in \mathcal{T}(S)$, and $q \in \mathcal{Q}(S, J)$, there exists a map

$$\widetilde{f}\colon \mathrm{SL}_2\mathbb{R} \to \mathcal{T}(S)$$

which sends $\gamma \in \mathrm{SL}_2\mathbb{R}$ to a point in $\mathcal{T}(S)$ obtained by deforming $[J]$ according to γ as follows. An element $\gamma \in \mathrm{SL}_2\mathbb{R}$ defines a new atlas by composing each chart in the preferred atlas with γ (here we are identifying $\mathbb{C}$ with $\mathbb{R}^2$ and γ is the obvious $\mathbb{R}$–linear map). The transition functions for the new atlas are again of the form $z \mapsto \pm z + \xi$, and we obtain a new complex structure $\gamma \cdot J$ and quadratic differential $\gamma \cdot q$. We define $\widetilde{f}(\gamma) = \gamma \cdot J$.

Note that SO(2) does not change the underlying complex structure, and so $\tilde{f}$ factors through a map

$$f\colon \mathbb{H}^2 \cong \mathrm{SO}_2 \setminus \mathrm{SL}_2(\mathbb{R}) \to \mathcal{T}(S).$$

After scaling the hyperbolic metric this is a holomorphic isometric embedding and is called a *Teichmüller disk.*

Given a Teichmüller disk $f\colon \mathbb{H}^2 \to \mathcal{T}(S)$, we have the stabilizer of $f(\mathbb{H}^2)$

$$\mathrm{Stab}(f(\mathbb{H}^2)) < \mathrm{Mod}(S).$$

Conjugating by f we obtain a subgroup of $PSL_2(\mathbb{R})$ which we denote

$$\mathrm{Stab}(f) = f^{-1}\mathrm{Stab}(f(\mathbb{H}^2))f.$$

Forming the quotient by $\mathrm{Stab}(f)$, f then descends to a map

$$\widehat{f}\colon \mathbb{H}^2/\mathrm{Stab}(f) \to \mathcal{M}(S).$$

When $\mathbb{H}^2/\mathrm{Stab}(f)$ has finite area, its image $\widehat{f}(\mathbb{H}^2/\mathrm{Stab}(f))$ is an algebraic curve totally geodesically immersed in $\mathcal{M}(S)$ called a *Teichmüller curve.*

If f is a Teichmüller disk defined by q, then every automorphism of $\mathrm{Stab}(f(\mathbb{H}^2))$ can be realized by an affine automorphism with respect to the flat structure. The derivative with respect to the basis $\{e_1, e_2\}$ defines a discrete representation

$$\mathrm{DAf}\colon \mathrm{Stab}(f(\mathbb{H}^2)) \to \mathrm{PSL}_2(\mathbb{R})$$

(this is into $\mathrm{PSL}_2(\mathbb{R})$, rather than $\mathrm{SL}_2(\mathbb{R})$ because the basis is only defined up to sign).

An element of the kernel of DAf leaves the complex structure and the quadratic differential invariant. It follows that such an element fixes the Teichmüller disk pointwise. Because the action on $\mathcal{T}(S)$ is properly discontinuous, the kernel of DAf is finite.

We collect these and other facts into the following theorem for ease of reference. Parts of this theorem have appeared in several different locations (see eg the lecture notes [18], Thurston [54], Kra [34], Long [36], and Veech [56]).

Theorem 3.1 (Thurston, Kra, Veech) *Let $f\colon \mathbb{H}^2 \to \mathcal{T}(S)$ be a Teichmüller disk. Then*

$$\mathrm{DAf}\colon \mathrm{Stab}(f(\mathbb{H}^2)) \to \mathrm{PSL}_2\mathbb{R}$$

is discrete, with finite kernel. For $\phi \in \mathrm{Stab}(f(\mathbb{H}^2)) \setminus \{1\}$ the following is true:

(1) *if $\mathrm{DAf}(\phi)$ is elliptic or the identity, then ϕ has finite order,*

(2) *if $\mathrm{DAf}(\phi)$ is parabolic, then ϕ is reducible and some power of ϕ is a positive multi-twist, and*

(3) *if $\mathrm{DAf}(\phi)$ is hyperbolic, then ϕ is pseudo-Anosov and the dilatation is given by $\lambda(\phi) = \exp\left(\frac{1}{2}L(\mathrm{DAf}(\phi))\right)$ where $L(\mathrm{DAf}(\phi))$ is the translation length of $\mathrm{DAf}(\phi)$ on $\mathbb{H}^2$.*

There is a strong converse to part 3 of the theorem which is essentially Bers' description of pseudo-Anosov automorphisms.

Theorem 3.2 (Bers) *Given any pseudo-Anosov automorphism ϕ, there is a unique Teichmüller disk which it stabilizes.*

The quotients $\mathbb{H}^2/\mathrm{DAf}(\mathrm{Stab}(f(\mathbb{H}^2)))$ and $\mathbb{H}^2/\mathrm{Stab}(f)$ are essentially the same (see eg [17] or [44]).

Proposition 3.3 *$\mathbb{H}^2/\mathrm{Stab}(f)$ is isometric to $\mathbb{H}^2/\mathrm{DAf}(\mathrm{Stab}(f(\mathbb{H}^2)))$. In particular, $\mathrm{Stab}(f)$ has finite co-area if and only if $\mathrm{DAf}(\mathrm{Stab}(f(\mathbb{H}^2)))$ does.*

3.3 Homology representation

As noted above, q is the square of a holomorphic 1–form $\omega = e^1 + ie^2$ if and only if the metric has no holonomy. Now suppose $q = \omega^2$ and let $f\colon \mathbb{H}^2 \to \mathcal{T}(S)$ denote the associated Teichmüller disk. In this case, the two-dimensional subspace $\langle e^1, e^2 \rangle \subset H^1(S;\mathbb{R})$ is left invariant by the action of $\mathrm{Stab}(f(\mathbb{H}^2))$ since the $\mathbb{R}$–span of the vector fields $\{e_1, e_2\}$ is invariant. Moreover, the action on $\langle e^1, e^2 \rangle$ is dual to the action on $\langle e_1, e_2 \rangle \cong \mathbb{R}^2$ given by DAf. In particular, the induced action on $H^1(S;\mathbb{R})$ has a finite kernel. Since finite order automorphisms can never act trivially on (co)homology, we obtain the following result.

Proposition 3.4 *Suppose $q = \omega^2$ and $f\colon \mathbb{H}^2 \to \mathcal{T}(S)$ is the corresponding Teichmüller disk. Then the action of $\mathrm{Stab}(f(\mathbb{H}^2))$ on homology is faithful.*

3.4 Example: Billiards

Let $P \subset \mathbb{R}^2$ be a compact *rational* polygon, that is, the angle at every vertex is a rational multiple of π. One can naturally associate the data of a surface with flat structure (S_P, q_P) so that the geodesics correspond to trajectories of billiards in P. We give a very brief discussion of this and refer the reader to Zemlyakov and Katok [61], Kerckhoff, Masur, and Smillie [33], and Masur and Tabachnikov [41] for more details.

To construct S_P, first consider the dihedral group D_{2k} generated by reflections in the lines through the origin in $\mathbb{R}^2$, parallel to the sides of P. Let

$$\mathcal{P} = \coprod_{\gamma \in D_{2k}} \gamma P$$

be the disjoint union of the images of P under the linear action by D_{2k}. We view the components γP as having a well-defined embedding in $\mathbb{R}^2$, up to translation. The group D_{2k} acts on $\mathcal{P}$ in an obvious way, and we form the surface S_P as the quotient of $\mathcal{P}$ obtained by identifying an edge e of $\mathcal{P}$ with its image γe, for $\gamma \in D_{2k}$, if e and γe are parallel (with the same orientation). dz^2 is defined on each polygon and pieces together to give a well-defined quadratic differential q_P on S_P. In fact, dz is invariant, so that $q = \omega^2$ for a globally defined 1–form ω. The polygon P is said to be a *lattice polygon* if (S_P, q_P) defines a Teichmüller curve in $\mathcal{M}(S_P)$.

The right and isosceles lattice triangles have been classified by Kenyon and Smillie in [32]. They conjectured that there were exactly three non-isosceles, acute lattice triangles and proved this for a large number of examples. The conjecture was proven by Puchta in [48]. We collect these facts together in the following.

Theorem 3.5 (Kenyon–Smillie, Puchta) *The right lattice triangles are those with smallest angle $\frac{\pi}{k}$, $k \in \mathbb{Z}_{\geq 4}$. The acute isosceles lattice triangles are those with smallest angle $\frac{\pi}{k}$, $k \in \mathbb{Z}_{\geq 3}$. There are precisely three acute, non-isosceles lattice triangles, namely those with angles*

$$(1) \quad \left(\frac{\pi}{4}, \frac{\pi}{3}, \frac{5\pi}{12}\right) \qquad (2) \quad \left(\frac{2\pi}{9}, \frac{\pi}{3}, \frac{4\pi}{9}\right) \qquad (3) \quad \left(\frac{\pi}{5}, \frac{\pi}{3}, \frac{7\pi}{15}\right)$$

We will discuss the surfaces and quadratic differentials associated to these lattice triangles in more detail in Section 7.

4 Matrices and graphs

4.1 Non-negative matrices

Let M be a square, $n \times n$ matrix with real entries. The *spectral radius of M* is the maximum of the moduli of its eigenvalues, and we denote this by $\mu(M)$.

We say that M is *non-negative* (respectively, *positive*) if the entries of M are non-negative (respectively, positive) and in this case we write $M \geq 0$ (respectively, $M > 0$). Say that $M \geq 0$ is *irreducible* if for every $1 \leq i, j \leq n$ there is some power, M^k, so that $(M^k)_{ij} > 0$ (see [22]). If $M, M' \geq 0$, then write $M \leq M'$ if $M_{ij} \leq M'_{ij}$ for every $1 \leq i, j \leq n$ and write $M < M'$ if in addition this inequality is strict for *some* $1 \leq i, j \leq n$. We similarly define $\vec{V} \geq 0$, $\vec{V} > 0$, $\vec{V} \leq \vec{V}'$, and $\vec{V} < \vec{V}'$ for vectors $\vec{V}$ and $\vec{V}'$ in $\mathbb{R}^n$.

The following theorem on irreducible matrices will be useful (see [22] for a proof).

Theorem 4.1 (Perron–Frobenius) *Suppose that $M \geq 0$ is irreducible. Then M has a unique (up to scaling) non-negative eigenvector $\vec{V}$. This vector is positive with eigenvalue $\mu = \mu(M) > 0$. Moreover, for any non-negative vector $\vec{U} \neq 0$, we have*

$$\min_{1\leq i\leq n}\left(\frac{(M\vec{U})_i}{\vec{U}_i}\right) \leq \mu \leq \max_{1\leq i\leq n}\left(\frac{(M\vec{U})_i}{\vec{U}_i}\right)$$

with either inequality being an equality if and only if $\vec{U}$ is a multiple of $\vec{V}$.

Remark We define $\frac{(M\vec{U})_i}{\vec{U}_i} = +\infty$ whenever $\vec{U}_i = 0$.

When M is irreducible, we refer to the eigenvalue $\mu(M)$ (equal to the spectral radius) as the *Perron–Frobenius eigenvalue* (briefly, PF eigenvalue) of M and an associated eigenvector, as in the theorem, is called a *Perron–Frobenius eigenvector* (briefly, PF eigenvector) for M.

4.2 Graphs

For more details on spectral radii of graphs, see the survey article of Cvetković and Rowlinson [16]. The author is thankful to Curt McMullen for pointing out this reference which greatly simplified the exposition.

Given any finite graph $\mathcal{G}$, one can associate to $\mathcal{G}$ a matrix, $\mathcal{A}d(\mathcal{G})$, called the *adjacency matrix*, as follows. Labeling the vertices of $\mathcal{G}$ by $x_1, \ldots, x_n$, the (i,j)–entry of $\mathcal{A}d(\mathcal{G})$ is defined to be the number of edges connecting x_i to x_j. The *spectral radius of* $\mathcal{G}$ is defined to be $\mu(\mathcal{G}) = \mu(\mathcal{A}d(\mathcal{G}))$. Note that when $\mathcal{G}$ is connected, $\mathcal{A}d(\mathcal{G})$ is irreducible. Indeed, $((\mathcal{A}d(\mathcal{G}))^k)_{ij}$ is the number of combinatorial paths of length k from the ith vertex to the jth. The following is an elementary consequence of Theorem 4.1.

Theorem 4.2 *If $\mathcal{G}_0 \subset \mathcal{G}$ is a subgraph of a connected graph $\mathcal{G}$, then $\mu(\mathcal{G}_0) \leq \mu(\mathcal{G})$, with equality if and only if $\mathcal{G}_0 = \mathcal{G}$.*

From this theorem one easily obtains a proof of the following (which is a special case of the classical result of Smith [53]).

Theorem 4.3 (Smith) *The set of connected bipartite graphs $\mathcal{G}$ with $\mu(\mathcal{G}) < 2$ are precisely the recessive graphs, and those with $\mu(\mathcal{G}) = 2$ are precisely the critical graphs.*

Proof An explicit calculation (see eg [16]) shows that the spectral radius of every critical graph is 2. Any connected bipartite graph $\mathcal{G}$ contains or is contained in one of the critical graphs. To see this, we note that if $\mathcal{G}$ is not a tree, then it contains a cycle (of even length since $\mathcal{G}$ is bipartite). Hence $\mathcal{P}_{2c} \subset \mathcal{G}$ for some c. If $\mathcal{G}$ is a tree, then one of the following holds

(1) $\mathcal{G}$ is homeomorphic to an interval (and thus contained in some $\mathcal{Q}_c$),

(2) $\mathcal{G}$ contains a vertex with valence at least 4 (and so contains $\mathcal{Q}_5$),

(3) $\mathcal{G}$ has at least two vertices of valence at least 3 (and so contains some $\mathcal{Q}_c$), or

(4) $\mathcal{G}$ has exactly one vertex of valence 3 and all other vertices of valence at most 2.

In case (4), by inspection, $\mathcal{G}$ is either contained in one of $\mathcal{Q}_c$, $\mathcal{R}_7$, $\mathcal{R}_8$, or $\mathcal{R}_9$, or else it contains $\mathcal{R}_7$, $\mathcal{R}_8$, or $\mathcal{R}_9$.

The only connected proper subgraphs of the critical graphs are the recessive graphs, and so any other connected graph contains some critical graph. The theorem now follows from Theorem 4.2. □

There is also a classification of graphs, similar to Smith's, for graphs having spectral radius in the interval $(2, \sqrt{2+\sqrt{5}}]$ due to Cvetković, Doob, and Gutman [15] and Brouwer and Neumaier [12]. From this, we easily obtain the following.

Theorem 4.4 (Cvetković, Doob, Gutman, Brouwer, Neumaier) *Given any bipartite graph $\mathcal{G}$ with $\mu(\mathcal{G}) > 2$, we have*

$$\mu(\mathcal{G}) \geq \mu_L \approx 2.0065936$$

with equality if and only if $\mathcal{G} = \mathcal{E}h_{10}$.

Here μ_L is the square root of the unique largest root of

$$x^5 - 9x^4 + 27x^3 - 31x^2 + 12x - 1 \tag{4}$$

This polynomial is the square root of the characteristic polynomial for the matrix $(\mathcal{A}d(\mathcal{E}h_{10}))^2$.

Proof Appealing to the aforementioned classification (see [16]), one can verify by explicit calculation that $\mathcal{E}h_{10}$ uniquely minimizes spectral radius among graphs in the list, and that its spectral radius is μ_L. The classification is for graphs without multiple edges, so we verify directly that a graph $\mathcal{G}$ with multiple edges has $\mu(\mathcal{G}) > \mu_L$. Such a graph must contain one of the graphs shown in Figure 9. These each have spectral radius at least $\sqrt{5} > \mu_L$, and so the theorem follows from Theorem 4.2. □

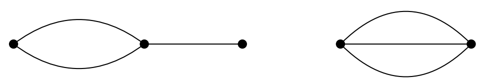

Figure 9: Subgraphs of a graph with multiple edges and $\mu > 2$

5 Affine actions for groups generated by two positive multi-twists

5.1 Constructing the flat structure

In this section, we recall the construction of Thurston [54]. Slight variations are also described in Long [36], Veech [56], and a special case in the lecture notes [18].

Viewing $A \cup B$ as a graph on S, the components of $S \setminus A \cup B$ are then the (interiors of) faces of this graph (actually, we are viewing S as a 2–complex with $A \cup B$ as the 1–skeleton). Thus, each face is a disk (with at most one marked point) which we may view as a $2k$–gon for some $k \in \mathbb{Z}$. Since A and B are assumed to intersect minimally, any face containing no marked points must have at least four edges. Write $A = a_1 \cup \cdots \cup a_n$ and $B = b_1 \cup \cdots \cup b_m$.

Let $\Gamma_{A,B}$ be the dual graph to $A \cup B$ embedded in S so that the vertex of $\Gamma_{A,B}$ dual to a face with a marked point is that marked point. $\Gamma_{A,B}$ defines a cell division of S, which we also denote by $\Gamma_{A,B}$, each 2–cell of which is a rectangle. Every rectangle contains a single arc of some a_i and a single arc of some b_j intersecting in one point (see Figure 10). Note that every vertex which is *not* a marked point of S must have valence at least 4 by the previous paragraph.

One can now use $\Gamma_{A,B}$ to define a Euclidean cone metric on S by declaring each rectangle to be a Euclidean rectangle. The choice of Euclidean rectangles

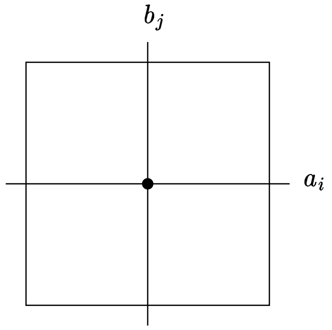

Figure 10: The local picture in any rectangle

is of course subject to the condition that whenever two rectangles meet along an edge, the shared edge must have the same length in each rectangle. It follows that we obtain one real parameter for each component of A and of B, corresponding to the length of the edges which that component meets. This defines a flat structure having orthonormal basis $\pm\{e_1, e_2\}$ with e_1 parallel to the edges which B transversely intersects, and e_2 parallel to the edges which A intersects.

Since we want $\langle T_A, T_B \rangle$ to act by affine transformations with respect to this structure, we choose these rectangle parameters as follows. Define $N = N_{A,B}$ to be the $n \times m$ matrix whose (i,j)–entry is $i(a_i, b_j)$. The connectivity of $A \cup B$ guarantees that NN^t is irreducible (here N^t is the matrix transpose of N). Let $\vec{V}$ be a PF eigenvector for $\mu = \mu(NN^t)$. Notice that for the same reason, N^tN is also irreducible, and setting $\vec{V}' = \mu^{-\frac{1}{2}} N^t \vec{V} \geq 0$, we see that

$$N^tN\vec{V}' = N^tN\mu^{-\frac{1}{2}}N^t\vec{V} = \mu^{-\frac{1}{2}}N^t(NN^t\vec{V}) = \mu^{-\frac{1}{2}}N^t\mu\vec{V} = \mu\mu^{-\frac{1}{2}}N^t\vec{V} = \mu\vec{V}'$$

so that $\mu(N^tN) = \mu = \mu(NN^t)$. With this choice of $\vec{V}$ and $\vec{V}'$, note that we also have $\vec{V} = \mu^{-\frac{1}{2}}N\vec{V}'$. We write $\mu(A \cup B)$ to denote $\sqrt{\mu(NN^t)}$ (the reason for the square root will soon become evident).

We now make any rectangle of $\Gamma_{A,B}$ containing arcs of a_i and b_j into a Euclidean rectangle for which the sides transverse to a_i have length $\vec{V}_i$ and the sides transverse to b_j have length $\vec{V}'_j$ (see Figure 11). For any component a_i of A, the rectangles containing arcs of a_i fit together to give a Euclidean annulus (which is a neighborhood of a_i). The length of this annulus is $\vec{V}_i$, and to see what the girth is, note that for each $j = 1, \ldots, m$ and for each intersection point of a_i with b_j there is a rectangle of width $\vec{V}'_j$ in the annulus. So, for

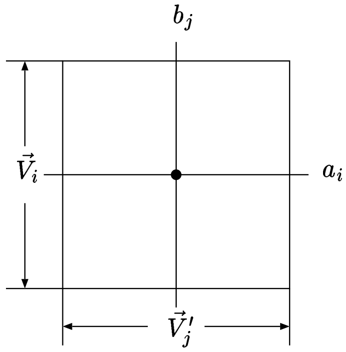

Figure 11: The Euclidean rectangle

each $j = 1, \ldots, m$, there is a contribution of $i(a_i, b_j)$ rectangles of width $\vec{V}'_j$. Therefore, the girth is

$$\sum_{j=1}^{m} i(a_i, b_j)\vec{V}'_j = (N\vec{V}')_i = \mu^{\frac{1}{2}}\vec{V}_i = \mu(A \cup B)\vec{V}_i$$

Similarly, the rectangles containing arcs of b_j fit together to give a Euclidean annulus of length $\vec{V}'_j$ and girth $\mu(A \cup B)\vec{V}'_j$.

We now verify that T_A and T_B are represented by affine transformations with respect to this structure. The derivative of the affine map for T_A (in terms of $\pm\{e_1, e_2\}$) is given by

$$\mathrm{DAf}(T_A) = \begin{pmatrix} 1 & \mu(A \cup B) \\ 0 & 1 \end{pmatrix}$$

To see this, first construct the affine twist on each of the Euclidean annuli described above and note that it has the desired derivative (see (2) and Section 2.4). Since each of the twists is the identity on the boundary of its defining annulus, they all piece together to give a well-defined affine homeomorphism with the correct derivative. Similarly, the derivative of the affine representative of T_B is

$$\mathrm{DAf}(T_B) = \begin{pmatrix} 1 & 0 \\ -\mu(A \cup B) & 1 \end{pmatrix}$$

We note that, by construction, all vertices of $\Gamma_{A,B}$ are fixed by every element of $\langle T_A, T_B \rangle$.

5.2 $\mu(A \cup B)$ vs $\mu(\mathcal{G}(A \cup B))$

Given $A \cup B$ filling S, we have associated two positive numbers, $\mu(A \cup B)$ and $\mu(\mathcal{G}(A \cup B))$. Not surprisingly, these are the same numbers.

Proposition 5.1 *With $N = N_{A,B}$ as in the previous section, we have*

$$\mathcal{A}d(\mathcal{G}(A \cup B)) = \begin{pmatrix} 0 & N \\ N^t & 0 \end{pmatrix}$$

In particular, $\mu(A \cup B) = \sqrt{\mu(NN^t)} = \mu(\mathcal{G}(A \cup B))$.

Proof Let us denote the vertices of $\mathcal{G}(A \cup B)$ (and curves of A and B) by both $a_1, \ldots, a_n, b_1, \ldots, b_m$ and $x_1, \ldots, x_{n+m}$ where $x_i = a_i \subset A$, $i = 1, \ldots, n$, and $x_{j+n} = b_j \subset B$, $j = 1, \ldots, m$. Then, the (i,j)–entry of $\mathcal{A}d(\mathcal{G}(A \cup B))$ is the number of edges from x_i to x_j. By definition of $\mathcal{G}(A \cup B)$, this is equal to $i(x_i, x_j)$. Since the (i,j)–entry of N (respectively N^t) is $i(a_i, b_j)$ (respectively $i(b_i, a_j)$), it is immediate that

$$\mathcal{A}d(\mathcal{G}(A \cup B)) = \begin{pmatrix} 0 & N \\ N^t & 0 \end{pmatrix}$$

To see the second statement, note that $\mu(\mathcal{A}d(\mathcal{G}(A \cup B)))^2 = \mu((\mathcal{A}d(\mathcal{G}(A \cup B)))^2)$ and that

$$(\mathcal{A}d(\mathcal{G}(A \cup B)))^2 = \begin{pmatrix} NN^t & 0 \\ 0 & N^tN \end{pmatrix}$$ □

6 Fuchsian groups

Here we note a few lemmas concerning the Fuchsian groups which occur as the images of groups $\langle T_A, T_B \rangle$ under DAf. The proofs are routine exercises in hyperbolic geometry, and we refer to Beardon's text [6], Ratcliffe's text [51], and the Thurston's notes [55] for more background on hyperbolic geometry and Fuchsian groups. The following two theorems are then easily derived from these lemmas.

Theorem 6.1 $\langle T_A, T_B \rangle \cong \mathbb{F}_2$ *if and only if $\mathcal{G}(A \cup B)$ contains a dominant component.*

Theorem 6.2 *For any surface S, any $A, B \in \mathcal{S}'(S)$, and any pseudo-Anosov element*

$$\phi \in \langle T_A, T_B \rangle < \mathrm{Mod}(S)$$

we have $\lambda(\phi) \geq \lambda_L \approx 1.1762808$. Moreover, $\lambda(\phi) = \lambda_L$ precisely when S has genus 5 (with at most one marked point), $\{A, B\} = \{A_L, B_L\}$ as in Figure 1 (up to homeomorphism), and ϕ is conjugate to $(T_A T_B)^{\pm 1}$.

Remark I would like to thank Joan Birman for pointing out that the number obtained as the minimal dilation was Lehmer's number.

6.1 Groups generated by two parabolics

For $\mu > 0$, set

$$\gamma_1(\mu) = \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}, \quad \gamma_2(\mu) = \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix} \in \mathrm{PSL}_2\mathbb{R}$$

and

$$\Gamma_\mu = \langle \gamma_1(\mu), \gamma_2(\mu) \rangle < \mathrm{PSL}_2\mathbb{R}$$

Note that DAf maps $\langle T_A, T_B \rangle$ onto $\Gamma_{\mu(A\cup B)}$. We write $\mathcal{O}_\mu = \mathbb{H}^2/\Gamma_\mu$.

Recall that the convex core of a hyperbolic manifold M, denoted $C(M)$, is the smallest convex sub-manifold for which the inclusion is a homotopy equivalence. The *signature* of a triangle orbifold, $\mathcal{O}$, (or, equivalently, its associated triangle group) is a triple (p, q, r) with $p, q, r \in \mathbb{Z}_{\geq 2} \cup \{\infty\}$. $\mathcal{O}$ is a sphere with cone points of order p, q, and r, where a cone point of infinite order is a cusp.

Lemma 6.3 *If Γ_μ is discrete, then $\Gamma_\mu \cong \mathbb{F}_2$ if and only if $\mu \geq 2$. Moreover,*

1. *for $\mu > 2$, $\mathcal{O}_\mu$ has infinite area and $C(\mathcal{O}_\mu)$ is a twice punctured disk,*
2. *for $\mu = 2$, $\mathcal{O}_\mu \cong \dot{S}_{0,3}$, and*
3. *for $\mu < 2$, $\mathcal{O}_\mu$ is a triangle orbifold with signature*
    - *(q, ∞, ∞) if γ_1 and γ_2 are not conjugate and*
    - *$(2, q, \infty)$ if γ_1 and γ_2 are conjugate*

    where q is the order of $\gamma_1\gamma_2$.

In particular, Γ_μ has finite co-area if and only if $\mu \leq 2$.

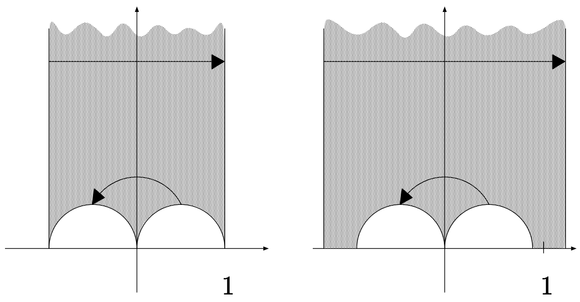

Figure 12: Fundamental domains for Γ_μ with $\mu = 2$ (left) and $\mu > 2$ (right)

Proof If $\mu \geq 2$, one can construct a fundamental domain for the action of Γ_μ on $\mathbb{H}^2$, as is shown in Figure 12 in the upper half-plane model. Identifying the faces of this fundamental domain as indicated gives the quotient $\mathcal{O}_\mu$. Therefore $\mathcal{O}_\mu \cong \dot{S}_{0,3}$ for $\mu = 2$, and $C(\mathcal{O}_\mu)$ is a twice-punctured disk (with punctures represented by cusps) for $\mu > 2$. In particular, $\Gamma_\mu \cong \mathbb{F}_2$ if $\mu \geq 2$.

When $\mu < 2$, we note that $\mathrm{Tr}(\gamma_1\gamma_2) = 2 - \mu^2 \in (-2, 2)$, and so $\gamma_1\gamma_2$ is elliptic. Because Γ_μ is discrete, $\gamma_1\gamma_2$ must have finite order, and so $\Gamma_\mu \not\cong \mathbb{F}_2$.

Using the Dirichlet domain construction centered at the point $2i$, one can check that the resulting fundamental domain must be contained in the set, P, shown in Figure 13.

$$P = \left\{ z \in \mathbb{H}^2 \mid d(z, 2i) \leq d(z, \gamma(2i)) \text{ for } \gamma = \gamma_1(\mu)^{\pm 1}, \gamma_2(\mu)^{\pm 1} \right\}$$

P is a polygon with two finite vertices at points $z_\pm \in \mathbb{H}^2$ ($Re(z_\pm) = \frac{\pm\mu}{2}$ and $|z_\pm| = 1$) and two infinite vertices at 0 and ∞.

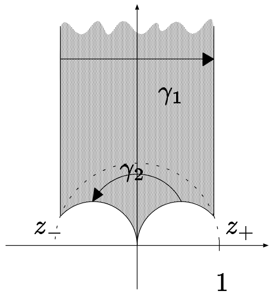

Figure 13: The set P containing the fundamental domain

The point z_+ is fixed by $\gamma_1\gamma_2$ and z_- is fixed by the conjugate $\gamma_2\gamma_1$. Let θ denote the interior angle at z_+, which is equal to the angle at z_-. A calculation shows that $2\cos\left(\frac{\theta}{2}\right) = \mu$ and that $\gamma_1\gamma_2$ is a *clockwise* rotation about the point z_+ through an angle 2θ.

If $2\theta = \frac{2\pi}{q}$ for some integer q, then $\mathcal{O}$ is obtained from P by identifying the two pairs of edges according to γ_1 and γ_2. In this case, $\mathcal{O}_\mu$ has two cusps, one corresponding to each of $\langle\gamma_1\rangle$ and $\langle\gamma_2\rangle$ and one cone point of order $q = \frac{2\theta}{2\pi}$ (which is the order of $\gamma_1\gamma_2$). The maximal parabolic subgroups $\langle\gamma_1\rangle$ and $\langle\gamma_2\rangle$ are not conjugate since they represent different cusps. In this case $\mathcal{O}_\mu$ is a triangle orbifold with signature (q,∞,∞), as required.

If $2\theta = \frac{4\pi}{q}$ for some odd integer q, then $\gamma_1\gamma_2$ generates a cyclic subgroup of order q in Γ_μ. This subgroup also contains the clockwise rotation $\rho = (\gamma_1\gamma_2)^{\frac{q+1}{2}}$ about z_+ through an angle $\theta = \frac{2\pi}{q}$. Consider the element $\delta = \gamma_1^{-1}\rho$ which takes 0 to ∞ and fixes i. If we intersect P with the complement of the unit disk in $\mathbb{C}$, we obtain a fundamental domain for Γ_μ as shown in Figure 14, with quotient given by the identifications by γ_1 and δ as shown. In this case $\mathcal{O}_\mu$ is a triangle orbifold with signature $(2,q,\infty)$ and γ_1 and γ_2 are conjugate by δ, as required.

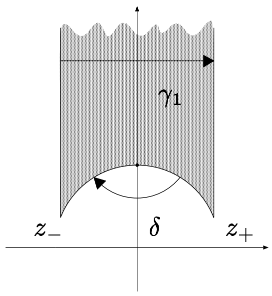

Figure 14: The fundamental domain when $2\theta = \frac{4\pi}{q}$, q odd

Finally, suppose $2\theta \notin \left\{\frac{4\pi}{q} \mid q \in \mathbb{Z}_+\right\}$. In this case the cyclic subgroup generated by $\gamma_1\gamma_2$ contains a rotation about z_+ through an angle less than θ. In particular, the Dirichlet domain construction based at an appropriate point on the imaginary axis gives a *compact* fundamental domain contained in P. This contradicts the fact that Γ_μ contains parabolics and hence any fundamental domain is noncompact. $\square$

6.2 Freeness at last

Proof of Theorem 6.1 Proposition 10.1 shows that it suffices to assume that $A \cup B$ fills S. According to Theorem 4.3, $\mu(\mathcal{G}(A \cup B)) \geq 2$ if and only if $\mathcal{G}(A \cup B)$ is dominant. Now, by Proposition 5.1, $\mu(A \cup B) = \mu(\mathcal{G}(A \cup B))$, so it suffices to show that $\langle T_A, T_B \rangle$ is free if and only if $\mu(A \cup B) \geq 2$. This latter equivalence is precisely what Thurston had suggested in [54].

Suppose $\mu(A \cup B) \geq 2$. By Lemma 6.3, DAf is a surjection to $\mathbb{F}_2$. Since $\langle T_A, T_B \rangle$ is generated by two elements, we have a surjection

$$\chi \colon \mathbb{F}_2 \to \langle T_A, T_B \rangle$$

$\mathrm{DAf} \circ \chi$ is therefore a surjection from $\mathbb{F}_2$ onto itself. Free groups are Hopfian (see [38]), hence this is an isomorphism. Therefore, χ is also an isomorphism, proving that $\langle T_A, T_B \rangle \cong \mathbb{F}_2$.

If $\mu(A \cup B) < 2$, Lemma 6.3 implies $\mathrm{DAf}(T_A T_B)$ has finite order, so Theorem 3.1 says $T_A T_B$ also has finite order, hence $\langle T_A, T_B \rangle \ncong \mathbb{F}_2$. □

The next proposition is included for it own interest. It makes precise the statement that most elements in $\langle T_A, T_B \rangle$ are pseudo-Anosov.

Proposition 6.4 *Suppose $A \cup B$ fills S and $\mathcal{G}(A \cup B)$ is dominant. Then every element of $\langle T_A, T_B \rangle$ is pseudo-Anosov except conjugates of powers of T_A and T_B, and also $T_A T_B$ when $\mathcal{G}(A \cup B)$ is critical.*

Proof The hyperbolic elements of $\Gamma_\mu = \pi_1(\mathbb{H}^2/\Gamma_\mu)$ are precisely those elements corresponding to loops that are freely homotopic to closed geodesics. All loops that are not homotopic to cusps have such a representative, and so the corollary follows from Theorem 3.1, Lemma 6.3, and the fact that DAf is an isomorphism in this case. □

Remark When Γ_μ is discrete and $\mu < 2$, it also has precisely 3 conjugacy classes of cyclic subgroups which make up all non-hyperbolic elements. However, DAf is not necessarily an isomorphism in this case, so we only know that all non-pseudo-Anosov elements map by DAf to one of 3 cyclic subgroups, up to conjugacy.

6.3 Dilatation bounds and Lehmer's number

The connection between translation lengths and dilatations is provided by Theorem 3.1. This is the basis for the proof of Theorem 6.2, and so to apply this we will need a few elementary facts concerning translation lengths in the Fuchsian groups Γ_μ. As Lemma 6.3 shows, for $\mu \leq 2$, $\mathbb{H}^2/\Gamma_\mu$ is a triangle orbifold. Furthermore, triangle orbifolds can have no closed embedded geodesics, so the following is an immediate consequence of Theorem 11.6.8 of Beardon [6].

Proposition 6.5 (Beardon) *For any $\mu \leq 2$, the smallest translation length of a hyperbolic element of Γ_μ is bounded below by*

$$2\sinh^{-1}\left(\sqrt{\cos\left(\frac{3\pi}{7}\right)}\right)$$

For the remaining cases, we have the following.

Lemma 6.6 *When $\mu > 2$, the smallest translation length of a hyperbolic element of Γ_μ is realized (uniquely up to conjugacy) by $(\gamma_1(\mu)\gamma_2(\mu))^{\pm 1}$, and is given by $2\log(\lambda_\mu)$, where λ_μ is the larger root of*

$$x^2 + x(2 - \mu^2) + 1 \tag{5}$$

Remark The larger root λ_μ of (5) defines an increasing function of $\mu > 2$.

Proof Since $C(\mathbb{H}^2/\Gamma_\mu)$ is a twice-punctured disk, any geodesic γ determines a conjugacy class represented by an element which we also call $\gamma \in \Gamma_\mu$. The translation length of γ is the length of the geodesic with the same name, and is given by $2\log(\lambda)$ where λ is the spectral radius of a matrix representative of $\gamma \in \Gamma_\mu$. The only simple closed geodesic is the boundary of the convex core. Moreover, for any other closed geodesic, one can cut and paste a collection of arcs of this geodesic to obtain a curve homotopic to this boundary curve. It follows that the boundary geodesic is the unique shortest geodesic. This is represented by $\gamma_1(\mu)\gamma_2(\mu)$.

The natural representation of the projective class of $\gamma_1(\mu)\gamma_2(\mu)$ by a matrix (given the matrices for $\gamma_1(\mu)$ and $\gamma_2(\mu)$ we have chosen) has $Tr(\gamma_1(\mu)\gamma_2(\mu)) = 2 - \mu^2 < 0$, so we see that $-\lambda$ satisfies the characteristic equation. Therefore, λ is the larger root of (5). □

Corollary 6.7 *Let $\phi \in \langle T_A, T_B \rangle$ be any pseudo-Anosov automorphism. If $\mu = \mu(A \cup B) > 2$, then*

$$\lambda(\phi) \geq \lambda_\mu$$

where λ_μ is a root of (5). Equality holds if and only if $\phi = (T_A T_B)^{\pm 1}$ up to conjugacy. If $\mu \leq 2$, then

$$\lambda(\phi) > 1.47$$

Proof Suppose $\mu > 2$. By Theorem 3.1 $\lambda(\phi) = \exp(\frac{1}{2}L(\mathrm{DAf}(\phi)))$, hence the smallest dilatation occurs precisely when $\mathrm{DAf}(\phi)$ has smallest translation length (dilation is an increasing function of translation length). By Lemma 6.6, this is precisely when $\mathrm{DAf}(\phi)$ is conjugate to $\mathrm{DAf}(T_A T_B)^{\pm 1}$. As the proof of Theorem 6.1 shows, DAf is an isomorphism, and so this happens if and only if ϕ is conjugate to $(T_A T_B)^{\pm 1}$. In this case, we have $\lambda(\phi) = \lambda(T_A T_B) = \exp\left(\frac{1}{2} 2 \log(\lambda_\mu)\right) = \lambda_\mu$.

When $\mu \leq 2$, by similar reasoning (appealing now to Proposition 6.5 rather than Lemma 6.6) we obtain

$$\lambda(\phi) \geq \exp\left(\frac{1}{2} 2 \sinh^{-1}\left(\sqrt{\cos\left(\frac{3\pi}{7}\right)}\right)\right) > 1.47 \qquad \square$$

Proof of Theorem 6.2 By Corollary 6.7, we need only consider $\mu = \mu(A \cup B) > 2$, and it suffices to show that $\lambda_\mu \geq \lambda_L$ with equality if and only if $\{A, B\} = \{A_L, B_L\}$. The remark following Lemma 6.6 tells us that to minimize λ_μ, we need only minimize $\mu(A \cup B)$. Theorem 4.4 says that μ is minimized uniquely by μ_L when $\mathcal{G}(A \cup B) = \mathcal{E}h_{10}$. Here μ_L^2 is the largest root of the polynomial (4). By Corollary 6.7, λ_{μ_L} is a root of (5) with $\mu = \mu_L$. Thus, $(\lambda_{\mu_L}, \mu_L) = (x, y)$ satisfies

$$\begin{cases} x^2 + x(2 - y^2) + 1 = 0 \\ y^5 - 9y^4 + 27y^3 - 31y^2 + 12y - 1 = 0 \end{cases}$$

Eliminating y from this pair we find that λ_{μ_L} is a root of (1), and so $\lambda_{\mu_L} = \lambda_L$.

The only configuration which minimizes μ is $\mathcal{E}h_{10}$. Since $\mathcal{E}h_{10}$ is a tree with one vertex having valence at most three, Proposition 2.1 completes the proof of the theorem. $\square$

7 Teichmüller curves, triangle groups, and billiards

The first theorem of this section is the following:

Theorem 7.1 *The Teichmüller curves for which the associated stabilizers contain a group generated by two positive multi-twists with finite index are precisely those defined by $A \cup B$ filling S, where $\mathcal{G}(A \cup B)$ is critical or recessive.*

Proof We must show that $\langle T_A, T_B \rangle$ has finite co-area if and only if $\mathcal{G}(A \cup B)$ is recessive or critical. Proposition 3.3 implies that the former happens if and only if $\Gamma_{\mu(A\cup B)} = \langle \mathrm{DAf}(T_A), \mathrm{DAf}(T_B) \rangle$ has finite co-area. By Lemma 6.3, $\Gamma_{\mu(A\cup B)}$ has finite co-area if and only if $\mu(A \cup B) \leq 2$. Proposition 5.1 implies $\mu(A \cup B) = \mu(\mathcal{G}(A \cup B))$, and Theorem 4.3 implies that $\mu(\mathcal{G}(A \cup B)) \leq 2$ if and only if $\mathcal{G}(A \cup B)$ is recessive or critical. □

The Teichmüller curves obtained from this theorem are most interesting when $\mathcal{G}(A \cup B)$ are recessive (the critical configurations all give Teichmüller curves covered by the thrice-punctured sphere). To better understand these curves, we describe another construction for surfaces and flat structures studied by Veech [56] and Earle and Gardiner [17]. Embedded in this construction is the billiard construction for all but three of the lattice triangles from Theorem 3.5. To complete the billiard picture we describe the remaining three exceptional triangles and verify the following.

Theorem 7.2 *The Teichmüller curves determined by the right and acute lattice triangles have associated stabilizers containing a finite index subgroup of the form $\langle T_A, T_B \rangle$ with $\mathcal{G}(A \cup B)$ recessive.*

The Teichmüller curves determined by these lattice triangles do not account for all Teichmüller curves determined by recessive configuration. However, the constructions described below are general enough to take care of all of these. From this we obtain a complete description of the non-free groups generated by two positive multi-twists.

Theorem 7.3 *If $\mathcal{G}(A \cup B)$ is recessive, then* DAf *maps $\langle T_A, T_B \rangle$ onto a Fuchsian triangle group with finite central kernel of order at most 2. The signature of the triangle group is described by the following table.*

configuration graph	*signature*	*configuration graph*	*signature*
$\mathcal{D}_c$, $c \geq 4$	$(c-1, \infty, \infty)$	$\mathcal{E}_6$	$(6, \infty, \infty)$
$\mathcal{A}_{2c+1}$, $c \geq 1$	$(c+1, \infty, \infty)$	$\mathcal{E}_7$	$(9, \infty, \infty)$
$\mathcal{A}_{2c}$, $c \geq 1$	$(2, 2c+1, \infty)$	$\mathcal{E}_8$	$(15, \infty, \infty)$

We will explain the proof of Theorems 7.2 and 7.3 in Section 7.1. The strategy in all cases is the same, but requires verification on a case-by-case basis. This is done for the $\mathcal{D}_c$ graphs in Section 7.2, the $\mathcal{A}_c$ graphs in Section 7.3 for c odd and Section 7.4 for c even, and finally the three graphs $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$ in Section 7.5.

7.1 Proof outlined

In sections 7.2 through 7.5 we will describe a flat structure and positive multi-twists in 1–manifolds A and B which act as affine automorphisms with respect to this structure. This will generally be a different structure than the one we constructed in Section 5.1 for A and B. However, it is affine equivalent to that one, ie they define the same Teichmüller disk. This follows from the uniqueness in Theorem 3.2 since both Teichmüller disks are stabilized by any pseudo-Anosov automorphism in $\langle T_A, T_B \rangle$.

For each of the flat structures under consideration, and each of the pairs of 1–manifolds A and B, we will see that $\mathcal{G}(A \cup B)$ is recessive. In particular, to prove Theorem 7.2 it suffices to recognize those flat structures coming from the billiard construction for the right and acute lattice triangles among those which we describe. This is verified in Section 7.5.

Proof of Theorem 7.3 DAf maps $\langle T_A, T_B \rangle$ onto a Fuchsian triangle group by Lemma 6.3 and the fact that $\mu(A \cup B) < 2$. To see that the kernel is central, we note that by definition, any element of the kernel has derivative $\pm I$. In particular, this must leave both the A–annuli and the B–annuli invariant, and hence also each of A and B are invariant. Any automorphism which leaves a 1–manifold invariant must commute with the associated multi-twist. In particular, we see that every element of the kernel commutes with the generators T_A and T_B, and so is central.

Since any element of the kernel leaves each of A and B invariant, it induces an automorphism of the graph $A \cup B$ which leaves the A edges and B edges invariant. This in turn induces an automorphism of the graph $\mathcal{G}(A \cup B)$ preserving the bicoloring. Said differently, we obtain a homomorphism

$$\delta\colon\ \ker(\mathrm{DAf}) \to \mathrm{Aut}_{bc}(\mathcal{G}(A \cup B))$$

Here $\mathrm{Aut}_{bc}(\mathcal{G}(A \cup B))$ is the automorphism group of the graph preserving the bicoloring (which has index at most 2 in the full automorphism group).

Claim $\ker(\delta)$ *has order at most* 2.

Proof of Claim We note that any $\phi \in \ker(\delta)$ leaves each component of A and of B invariant. Relabeling A and B and renumbering the components if necessary, we may assume that $a_1 \subset A$ is a component corresponding to a 1–valent vertex of $\mathcal{G}(A \cup B)$. This a_1 has only a single point of intersection with $(A \setminus \{a_1\}) \cup B$, and hence a_1 is the closure of an edge, e_1, of the graph $A \cup B$. Since a_1 is invariant by ϕ, so is e_1. Thus, $\ker(\delta)$ consists of isotopy classes of (orientation preserving) homeomorphisms of S leaving the 1–skeleton $(A \cup B)$ of a cell structure invariant and fixing the edge e_1 (not necessarily pointwise). Such a group has order at most two. □

Now, when $\mathcal{G}(A \cup B)$ is of type $\mathcal{A}_{2c}$, $\mathrm{Aut}_{bc}(\mathcal{G}(A \cup B))$ is trivial, hence ker(DAf) has order at most 2 by the claim. When $\mathcal{G}(A \cup B)$ is of type $\mathcal{A}_{2c+1}$, $c \geq 1$, $\mathrm{Aut}_{bc}(\mathcal{G}(A \cup B))$ has order two. However, in this case we claim that $\ker(\delta)$ is trivial. This is because the two possible elements in this group are the identity and a hyperelliptic involution. The latter is in the full stabilizer, but not in $\langle T_A, T_B \rangle$ because it does not fix the two vertices of $\Gamma_{A,B}$. This implies ker(DAf) has order at most two in this case also.

When $\mathcal{G}(A \cup B)$ is of type $\mathcal{D}_c$, $\ker(\delta)$ is again trivial. This is because on the curve corresponding to the valence three vertex, any $\phi \in \ker(\delta)$ must fix the three points of intersection with the curves corresponding to the three adjacent vertices. It follows that ϕ is the identity on that curve, hence on all of S. When $c \geq 5$, there is only one non-trivial automorphism of $\mathcal{D}_c$, and so ker(DAf) has order at most two, and we are done in this case. When $c = 4$, we again use the fact that all vertices of $\Gamma_{A,B}$ are fixed to see that the ker(DAf) is trivial.

For the three exceptional cases we note that $\mathrm{Aut}(\mathcal{G}(A \cup B))$ is trivial when $\mathcal{G}(A \cup B) = \mathcal{E}_7$ or $\mathcal{E}_8$, so ker(DAf) has order at most two by the claim. In the one remaining case that $\mathcal{G}(A \cup B) = \mathcal{E}_6$, we note that $\mathrm{Aut}(\mathcal{G}(A \cup B))$ has order two. However, the non-trivial element is induced by an automorphism of the surface which does not fix the vertices of $\Gamma_{A,B}$, hence is not in $\langle T_A, T_B \rangle$.

All that remains is to verify that $\langle \mathrm{DAf}(T_A), \mathrm{DAf}(T_B) \rangle$ has the required signature. We will check below that $\mathrm{DAf}(T_A T_B)$ has order given by the larger of the two finite numbers listed in the signature. For the cases of $\mathcal{G}(A \cup B)$ of type $\mathcal{D}_c$, $\mathcal{A}_{2c+1}$, $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$ this will prove that the signature is as listed by showing that there must be two cusps in these cases. For then, the signature is (q, ∞, ∞), where q is the order of the product of the two parabolic generators by Lemma 6.3.

Suppose that in the cases listed there were only one cusp. By Lemma 6.3 there is an element $\mathrm{DAf}(\phi)$ conjugating $\mathrm{DAf}(T_A)$ to $\mathrm{DAf}(T_B)$. Up to an element of the kernel, ϕ would conjugate T_A to T_B. Because the kernel is central with order at most two, we obtain a conjugation of T_A^2 to T_B^2. This cannot happen since this would imply a homeomorphism taking a union of two copies of A to a union of two copies of B which is not possible for the given configurations.

Finally, for the remaining cases $\mathcal{G}(A \cup B) = \mathcal{A}_{2c}$, we find an element in $\langle \mathrm{DAf}(T_A), \mathrm{DAf}(T_B) \rangle$ conjugating $\mathrm{DAf}(T_A)$ to $\mathrm{DAf}(T_B)$. Lemma 6.3 completes the proof, modulo finding this conjugating element and verifying the orders of $T_A T_B$. This is carried out in the next four sections.

7.2 The $\mathcal{D}_c$ configurations

The following is described in more detail by Earle and Gardiner in [17].

Consider a regular $2k$–gon, Δ_{2k}, with $k \in \mathbb{Z}_{\geq 2}$, embedded in the plane with two vertical edges. Identifying opposite edges by Euclidean translations we obtain a surface S of genus $\lfloor \frac{k}{2} \rfloor$. Because the gluings are by isometry, we obtain a Euclidean cone metric on S, and the foliation by horizontal lines provides a holomorphic quadratic differential q (this restricts to dz^2 on Δ_{2k}). Let $\alpha_{2k} = \frac{\pi}{k}$ and $\beta_{2k} = \frac{\alpha_{2k}}{2}$.

Note first that the counter-clockwise rotation about the center of Δ_{2k} through an angle α_{2k} defines an isometry of S of order $2k$. We denote this by ρ_{2k}.

We also see that the horizontal foliation of q has all closed leaves, decomposing S into $\lceil \frac{k}{2} \rceil$ annuli. Let B be the essential 1–manifold which is the union of the cores of the annuli, taking two parallel copies of the core of the annulus meeting the two vertical sides of Δ_k (see Figure 15). T_B acts by an affine transformation leaving this foliation invariant, having derivative

$$\mathrm{DAf}(T_B) = \begin{pmatrix} 1 & 2\cot(\beta_{2k}) \\ 0 & 1 \end{pmatrix}.$$

Rotate the horizontal foliation by an angle β_{2k} (ie multiply q by $e^{i\alpha_{2k}}$). This rotated foliation also has all closed leaves, decomposing S into $\lfloor \frac{k}{2} \rfloor$ annuli in another way. Let A be the union of the cores of these annuli. T_A also acts by an affine transformation, with derivative $\mathrm{DAf}(T_A)$ given by

$$\begin{pmatrix} \cos(\beta_{2k}) & -\sin(\beta_{2k}) \\ \sin(\beta_{2k}) & \cos(\beta_{2k}) \end{pmatrix} \begin{pmatrix} 1 & 2\cot(\beta_{2k}) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\beta_{2k}) & \sin(\beta_{2k}) \\ -\sin(\beta_{2k}) & \cos(\beta_{2k}) \end{pmatrix}.$$

One can now verify that $T_A T_B = \rho_{2k}^{k+1}$ (eg consider the action on the line segment from the center of the polygon to the vertex at the top of the left

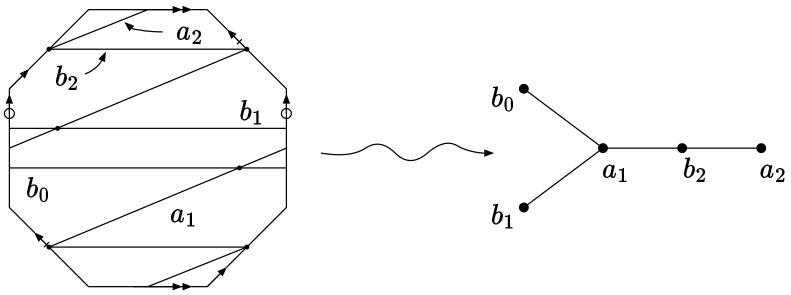

Figure 15: $A \cup B$ in S and $\mathcal{G}(A \cup B)$, when $k = 4$

vertical edge and on the horizontal line segment from the center to the midpoint of the right vertical edge). We also note that $\mathrm{DAf}(\rho_{2k}) \in \mathrm{PSL}_2\mathbb{R}$ has order k (which is half its order in $\mathrm{SL}_2\mathbb{R}$).

As is indicated in Figure 15, $\mathcal{G}(A \cup B)$ is the graph $\mathcal{D}_{k+1}$, when $k \geq 3$ (for $k = 2$, we get $\mathcal{A}_3$). This proves Theorem 7.3 for the graphs $\mathcal{D}_c$.

7.3 The $\mathcal{A}_c$ configurations I: c odd

The examples below where studied by Veech in [56] and [57]. However, we follow the discussion of Earle and Gardiner in [17].

Let S, A, and B be as in the previous section and assume that b_0 and b_1 were the parallel components of B. Write $A = a_1 \cup \cdots \cup a_n$ and $B = b_0 \cup \cdots \cup b_m$. Note that $n + m = k$. Let $B' = B \setminus b_0$ and note that we may replace T_B by the following isotopic homeomorphism:

$$T_B \simeq \widehat{T}_{B'} = T_{b_1}^2 T_{b_2} \cdots T_{b_m}$$

Now construct a 2–fold cover $\pi\colon \widetilde{S} \to S$ (which is a branched cover when k is odd) for which all components of A and of $B' \setminus b_1$ lift to loops, but the preimage of b_1 is a connected double cover of b_1. Writing $\widetilde{A} = \pi^{-1}(A)$ and $\widetilde{B} = \pi^{-1}(B')$, one can check that $T_{\widetilde{A}}$ and $T_{\widetilde{B}}$ cover T_A and $\widehat{T}_{B'}$, respectively. Moreover, these act as affine transformations with respect to $\pi^*(q)$ with derivatives

$$\mathrm{DAf}(T_{\widetilde{A}}) = \mathrm{DAf}(T_A) \quad \text{and} \quad \mathrm{DAf}(T_{\widetilde{B}}) = \mathrm{DAf}(\widehat{T}_{B'}) = \mathrm{DAf}(T_B).$$

So, $\mathrm{DAf}(T_{\widetilde{A}} T_{\widetilde{B}})$ has order k.

$\mathcal{G}(\widetilde{A} \cup \widetilde{B})$ is of type $\mathcal{A}_c$ since each curve of A and B intersects at most two other curves. $\widetilde{B}$ has $2m-1$ components, $\widetilde{A}$ has $2n$ components, so $\mathcal{G}(\widetilde{A} \cup \widetilde{B}) = \mathcal{A}_{2k-1}$. This proves Theorem 7.3 for the graphs $\mathcal{A}_c$ with c odd.

Remark In Veech's description of these examples, he explicitly constructed the surface $\widetilde{S}$ from two regular $2k$–gons in the plane, identified along an edge. $\widetilde{S}$ is obtained by identifying opposite sides of the resulting non-convex polygon.

7.4 The $\mathcal{A}_c$ configurations II: c even

The following construction is due to Veech [56], [57].

For $k \in \mathbb{Z}_{\geq 1}$ we consider two regular $(2k+1)$–gons, Δ^0_{2k+1} and Δ^1_{2k+1}, in the plane sharing a horizontal edge, and denote the non-convex polygon which is their union by Δ_{2k+1}. Identifying opposite sides of Δ_{2k+1} we obtain a genus–k surface, which we denote by S.

In the same fashion as above, we obtain a flat structure q on S, which restricts to dz^2 on Δ_{2k+1}. Let $\alpha_{2k+1} = \frac{2\pi}{2k+1}$ and $\beta_{2k+1} = \frac{\alpha_{2k+1}}{2}$.

The counter-clockwise rotations through an angle α_{2k+1} about the centers of Δ^0_{2k+1} and Δ^1_{2k+1} define an isometry ρ_{2k+1} of S of order $2k+1$. There is also an involution σ_{2k+1} obtained by rotating Δ_{2k+1} about the center of the edge shared by Δ^0_{2k+1} and Δ^1_{2k+1}. Note that σ_{2k+1} is in the kernel of DAf.

The horizontal foliation has all closed leaves, and so decomposes S into k annuli. Let B be the union of the cores of these annuli. Then T_B acts on S by affine transformations with derivative

$$\mathrm{DAf}(T_B) = \begin{pmatrix} 1 & 2\cot(\beta_{2k+1}) \\ 0 & 1 \end{pmatrix}$$

Next, we let $A = \rho^{k+1}_{2k+1}(B)$. Equivalently, A is obtained as follows. Rotate the horizontal foliation of q through an angle $(k+1)\alpha_{2k+1}$. This has the same effect as rotating through an angle $\beta_{2k+1} = (k+1)\alpha_{2k+1} - \pi$ (and hence multiplying q by $e^{i\alpha_{2k+1}}$). This foliation has all closed leaves and decomposes S into annuli, the union of the cores of which are precisely A.

Now one can check that

$$T_A T_B = \rho_{2k+1}\sigma_{2k+1}$$

(eg one can verify that this holds on appropriately chosen segments). So that we see

$$\mathrm{DAf}(T_A) = \mathrm{DAf}(T_A T_B)^{k+1}\mathrm{DAf}(T_B)\mathrm{DAf}(T_A T_B)^{-(k+1)}.$$

Thus, $\mathrm{DAf}(T_A)$ and $\mathrm{DAf}(T_B)$ are conjugate in $\mathrm{DAf}(\langle T_A, T_B\rangle)$ and $\mathrm{DAf}(T_A T_B)$ has order $2k+1$. One can check that $\mathcal{G}(A \cup B) = \mathcal{A}_{2k}$, thus proving Theorem 7.3 for this class of graphs.

7.5 Billiards and $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$

The constructions of the previous three sections provide a description of the surfaces and quadratic differentials coming from the billiard construction (see Section 3.4) for the right and acute isosceles lattice triangles given in Theorem 3.5, as we shall now explain.

Consider first the case where P is an acute, isosceles triangle with apex angle of the form $\frac{\pi}{k}$, $k \in \mathbb{Z}_{\geq 3}$, and k odd. P tiles the regular $2k$–gon, Δ_{2k}, with all apex vertices at the center of Δ_{2k}. Take any copy of P, call it P', in this tiling. Reflecting P' across the edge opposite the apex gives a copy, P'', exactly opposite P' through the center of Δ_{2k}, up to translation (see Figure 16). It follows that if we identify opposite sides of Δ_{2k} as in Section 7.2 we get exactly the surface and quadratic differential (up to a complex multiple) from the billiard construction for P.

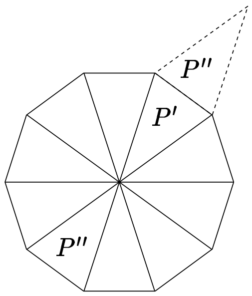

Figure 16: Reflecting P' in the side opposite the apex gives a translate of P''

Similarly, the construction from Section 7.3 gives the billiard surface and quadratic differential for the acute, isosceles triangle with apex angle $\frac{\pi}{k}$, $k \in \mathbb{Z}_{\geq 3}$, and k even. When P is a right triangle with smallest angle of the form $\frac{\pi}{k}$, $k \in \mathbb{Z}_{\geq 4}$, the construction using a regular k–gon in Section 7.2 for k even, and two regular k–gons in Section 7.4 for k is odd, give the billiard surface and quadratic differential for P.

We have thus proved Theorem 7.2, with the exception of the three non-isosceles lattice triangles of Theorem 3.5, and Theorem 7.3, except for the cases $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$.

Now, one may directly verify that $\mu(A \cup B)$ is given by $2\cos\left(\frac{\pi}{12}\right)$, $2\cos\left(\frac{\pi}{18}\right)$, and $2\cos\left(\frac{\pi}{30}\right)$, for $\mathcal{G}(A \cup B) = \mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$, respectively. The orders of $\mathrm{DAf}(T_A T_B)$ are thus, respectively, 6, 9, and 15. This shows that the signatures are as required and completes the proof of Theorem 7.3. $\square$

We now consider the billiard construction for the three exceptional lattice triangles. We refer to the triangles with angles $\left(\frac{\pi}{4}, \frac{\pi}{3}, \frac{5\pi}{12}\right)$, $\left(\frac{2\pi}{9}, \frac{\pi}{3}, \frac{4\pi}{9}\right)$, and $\left(\frac{\pi}{5}, \frac{\pi}{3}, \frac{7\pi}{15}\right)$, as Δ_6, Δ_7, and Δ_8, respectively.

The following two surfaces and flat structures are described by Vorobets in [58]. Consider first the non-convex polygon shown in Figure 17. This is a union of three squares and four equilateral triangles as shown. The surface, S, obtained by identifying parallel sides as indicated has genus three, and a flat structure q. S is tiled by 24 copies of Δ_6, no two of which are parallel. These triangles are obtained by considering the centers of the squares and equilateral triangles and the singular point (there is just one), and appropriately "connecting the dots". One of the tiles is shown in the Figure 17. Since 24 is the order of the dihedral group generated by reflections in lines through the origin parallel to the three sides of Δ_6, it follows that q is the flat structure from the billiard construction for Δ_6.

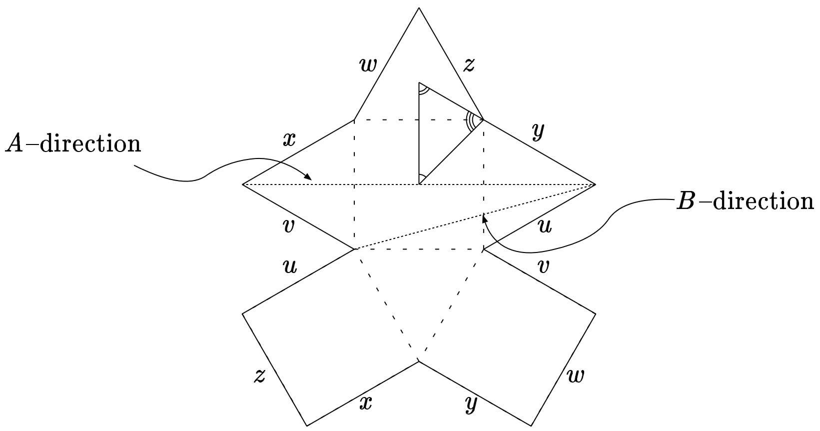

Figure 17: Δ_6 surface

The foliations parallel to the two line segments shown in Figure 17 have all closed leaves, decomposing S into annuli in two different ways. One direction shown is horizontal and the other makes an angle $\frac{\pi}{12}$ with the first. Appealing to some trigonometry we see that these foliations do indeed define annular decompositions. Moreover, for each of the two annular decomposition, the product of a single Dehn twist in each annulus acts as an affine transformation. Let A denote the union of the cores of the horizontal annuli and B the union of the cores of the other annuli. One can then verify that $\mathcal{G}(A \cup B) = \mathcal{E}_6$.

In a completely analogous fashion, we can consider the non-convex polygon shown in Figure 18 which is a union of three regular pentagons and five equilateral triangles. Identifying parallel sides as indicated gives a surface, S, of genus four with a flat structure, q. This is tiled by 30, pairwise nonparallel copies of Δ_8, again obtained by appropriately joining centers of pentagons, regular triangles and the singular point. As above, we see that q is the flat structure coming from the billiard construction for Δ_8.

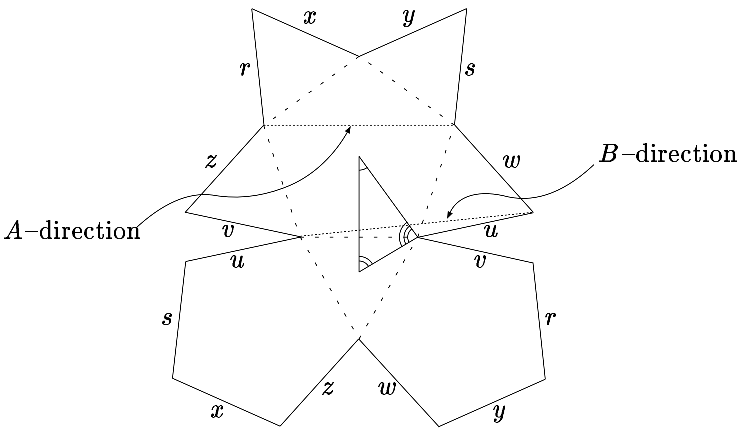

Figure 18: Δ_8 surface

One of the directions shown is horizontal and the other makes an angle $\frac{\pi}{30}$ with the horizontal. Again some trigonometry shows that we obtain annular decompositions in these directions having cores A and B, respectively, and T_A and T_B act as affine transformations. In this case $\mathcal{G}(A \cup B) = \mathcal{E}_8$.

Finally, for the triangle Δ_7, we briefly describe the construction of Kenyon and Smillie in [32]. We identify sides of the polygon in Figure 19 as indicated. The result is a surface, S, of genus three and a flat structure q. S is tiled by 18 copies of Δ_7 is as indicated and again q is the flat structure from the billiard construction for Δ_7.

The foliations in the two directions indicated (one horizontal, the other at an angle $\frac{\pi}{18}$ from horizontal) define annular decompositions. For each annular decomposition, the product of a single Dehn twist in each annulus acts as an affine transformation. Denoting the union of the cores of the horizontal annuli by A and the union of the other cores by B, one can check that $\mathcal{G}(A \cup B) = \mathcal{E}_7$.

Therefore, the three exceptional lattice triangles Δ_6, Δ_7, and Δ_8 define the same Teichmüller curves as the configurations with graph $\mathcal{E}_6$, $\mathcal{E}_7$, and $\mathcal{E}_8$, respectively. This completes the proof of Theorem 7.2. □

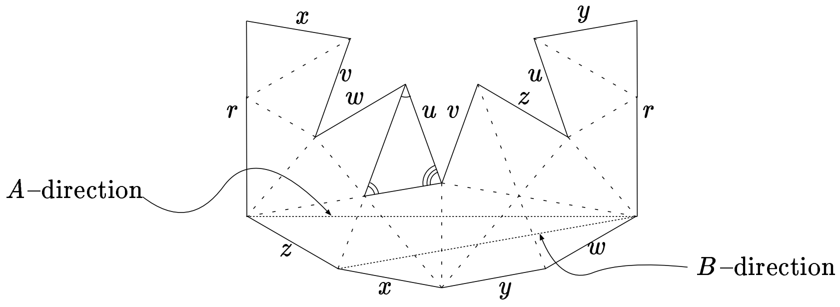

Figure 19: Δ_7 surface

We note that for each of the realizations of recessive configurations on surfaces we have described over the last four sections, the corresponding quadratic differentials are squares of holomorphic 1–forms. Therefore, the following is a consequence of Proposition 2.1 and Proposition 3.4.

Corollary 7.4 *If $\mathcal{G}(A \cup B)$ is recessive, then the action of $\langle T_A, T_B\rangle$ on homology is faithful.*

8 Coxeter and Artin groups

In this section we recall a few facts about Coxeter groups and Artin groups which indicate a connection with groups generated by two positive multi-twists. We then state McMullen's Theorem 8.4 and verify the following.

Theorem 8.1 *Let $\mathcal{G}(A \cup B)$ be non-critical dominant with small type. Then $\sigma_A\sigma_B$ is sent by Ψ to a pseudo-Anosov with dilatation equal to the spectral radius of its image under $\Theta \circ \pi_{ac}$. Moreover, among all essential elements in $\langle \sigma_A, \sigma_B\rangle$, $\sigma_A\sigma_B$ minimizes both dilatation as well as spectral radius for the respective homomorphisms.*

We next examine Hironaka's Theorem 8.5 and use her ideas, along with Theorem 8.6 of Howlett, to prove the following:

Theorem 8.2 *Let $\mathcal{G}(A \cup B)$ have small type and suppose that A and B can be oriented so that all intersections of A with B are positive. Then there exists a homomorphism*

$$\eta\colon \mathbb{R}^K \to H_1(S;\mathbb{R})$$

such that

$$(T_A T_B)_* \circ \eta = -\eta \circ \Theta(\sigma_A \sigma_B)$$

Moreover, $\Theta(\sigma_A\sigma_B)|_{\ker(\eta)} = -I$ *and* η *preserves spectral radii.*

This is a strengthening of the special case of Theorem 8.1 in which A and B can be oriented as in the theorem. For, in this situation the flat structure defined by $A \cup B$ in Section 5.1 has no holonomy, and the dilatation of a pseudo-Anosov is equal to the spectral radius of the action on homology (see McMullen [44]).

Remark I would like to thank Walter Neumann for first indicating the connection with Coxeter groups which led to Proposition 8.3.

8.1 Graphs and groups

Let $\mathcal{G}$ be any finite graph without loops (cycles of length one), which we refer to as a *Coxeter graph.* Let $\Sigma = \{s_1, \ldots, s_K\}$ be the vertices of $\mathcal{G}$ (throughout Section 8, K will denote the number of vertices of $\mathcal{G}$). For each $1 \le i < j \le K$, let m_{ij} be 2 plus the number of edges connecting s_i to s_j, and set $m_{ii} = 1$. So, when $\mathcal{G}$ has small type, $m_{ij} \in \{1,2,3\}$ for all i,j. This will be the primary case of interest for us. We will restrict ourselves to the case that $\mathcal{G}$ is connected (see Humphreys [30] for more details).

Remark We note that one usually allows the possibility that some vertices are connected by infinitely many edges, but because we are only interested in the small type case, we have not bothered to include this in the discussion. Also, a common convention is to consider Coxeter graphs as graphs without multiple edges, where the edge between s_i and s_j is labeled with $m_{ij} \in \mathbb{Z}_{\ge 3} \cup \{\infty\}$. The convention we have adopted is more suitable to our situation.

Given a Coxeter graph, $\mathcal{G}$, there are two groups associated to it: the *Coxeter group*

$$\mathfrak{C}(\mathcal{G}) = \langle s_i \in \Sigma \mid (s_i s_j)^{m_{ij}} = 1, 1 \le i \le j \le K \rangle$$

and the *Artin group*

$$\mathfrak{A}(\mathcal{G}) = \left\langle s_i \in \Sigma \;\middle|\; (s_i s_j)^{\frac{m_{ij}}{2}} = (s_j s_i)^{\frac{m_{ij}}{2}}, 1 \le i < j \le K \right\rangle$$

where for m odd we define $(xy)^{\frac{m}{2}} = (xy)^{\frac{m-1}{2}} x$ (eg if $m_{ij} = 3$, the relation is the braid relation $s_i s_j s_i = s_j s_i s_j$).

We will discuss both groups, and certain definitions are the same in each. However, when we wish to distinguish between an element of $\mathfrak{A}(\mathcal{G})$ and $\mathfrak{C}(\mathcal{G})$, we will denote the former with a "prime". Thus, $s_i \in \mathfrak{C}(\mathcal{G})$ and $s_i' \in \mathfrak{A}(\mathcal{G})$.

$\mathfrak{C}(\mathcal{G})$ is obtained from $\mathfrak{A}(\mathcal{G})$ by adding the relation $s_i^2 = 1$. This defines an epimorphism

$$\pi_{ac}\colon \mathfrak{A}(\mathcal{G}) \to \mathfrak{C}(\mathcal{G})$$

obtained by sending s_i' to s_i.

Given a Coxeter or Artin group a *special subgroup* is any subgroup generated by a subset $\Sigma_0 \subset \Sigma$. These special subgroups are precisely the Coxeter and Artin groups, respectively, associated to the largest subgraph of $\mathcal{G}$ having Σ_0 as its vertex set. An element of $\mathfrak{C}(\mathcal{G})$ is said to be *essential* if it is not conjugate into any proper special subgroup, and we call an element $\mathfrak{A}(\mathcal{G})$ essential if its image by π_{ac} is essential.

The product of all the generators (in any order) is called a *Coxeter element.* For any bipartite graph $\mathcal{G}$, there is a special Coxeter element defined as follows. Since the graph is bipartite there exists a partition $\Sigma = A \cup B$ so that no two A–vertices (respectively, B–vertices) are adjacent. The product of the elements of A (respectively, B) defines

$$\sigma_A = \prod_{s_i \in A} s_i \quad \text{and} \quad \sigma_B = \prod_{s_j \in B} s_j.$$

The product $\sigma_A \sigma_B$ is called the *bi-colored Coxeter element.*

8.2 Artin groups and mapping class groups

Let $A, B \in \mathcal{S}'(S)$ with $\mathcal{G} = \mathcal{G}(A \cup B)$ of small type. There is a nice relationship between $\mathfrak{A}(\mathcal{G})$ and $\mathrm{Mod}(S)$. The vertices $s_1, \ldots, s_K$ of $\mathcal{G}$ can be identified with the components of A and B as well as generators of $\mathfrak{A}(\mathcal{G})$. By (3) from Section 2.4 and the definition of $\mathfrak{A}(\mathcal{G})$, we can define a homomorphism

$$\Psi\colon \mathfrak{A}(\mathcal{G}) \to \mathrm{Mod}(S)$$

by sending the generator s_i to the Dehn twist about the curve corresponding to the vertex s_i. Note that, after relabeling if necessary, we have $\Psi(\sigma_A) = T_A$ and $\Psi(\sigma_B) = T_B$.

Remark This construction can be carried out for any graph of small type (not necessarily bipartite). Indeed, to such a graph one can (nonuniquely) associate a surface and a set of curves, pairwise intersecting at most once, and define a homomorphisms Ψ as above. For more on this see eg Crisp and Paris [14], Wajnryb [59], [60], Perron and Vannier [49], and A'Campo [1].

8.3 Geometric representations of Coxeter groups

Suppose that $\mathcal{G}$ is a connected Coxeter graph. There is an associated quadratic form $\Pi_{\mathcal{G}}$ on $\mathbb{R}^K$ and a faithful representation

$$\Theta\colon\ \mathfrak{C}(\mathcal{G}) \to \mathrm{O}(\Pi_{\mathcal{G}})$$

where $\mathrm{O}(\Pi_{\mathcal{G}})$ is the orthogonal group of the quadratic form $\Pi_{\mathcal{G}}$, and each generator $s_i \in \Sigma$ is represented by a reflection.

Up to equivalence over $\mathbb{R}$, there are precisely four possibilities for the form $\Pi_{\mathcal{G}}$ (see [30]). These are characterized by the signature, $sgn(\Pi_{\mathcal{G}})$, and K. Accordingly, the group $\mathfrak{C}(\mathcal{G})$ is said to be:

- *spherical* if $sgn(\Pi_{\mathcal{G}}) = (K, 0)$,
- *affine* if $sgn(\Pi_{\mathcal{G}}) = (K-1, 0)$,
- *hyperbolic* if $sgn(\Pi_{\mathcal{G}}) = (p, 1)$ and $p + 1 \leq K$, and
- *higher-rank* if $sgn(\Pi_{\mathcal{G}}) = (p, q)$ and $p + q \leq K$, $q \geq 2$.

When $\mathcal{G}$ has small type (our only case of interest), this quadratic form is easily described in terms of $\mathcal{A}d(\mathcal{G})$, the associated adjacency matrix (see Section 4.2). The form $\Pi_{\mathcal{G}}$ is then defined by the matrix

$$2I - \mathcal{A}d(\mathcal{G}).$$

Because $\mathcal{G}$ is connected, $\mathcal{A}d(\mathcal{G})$ is an irreducible matrix. Moreover, one can easily see that the group $\mathfrak{C}(\mathcal{G})$ is spherical or affine if and only if $\mu(\mathcal{G}) < 2$ or $\mu(\mathcal{G}) = 2$, respectively. The following is therefore a consequence of Theorem 4.3.

Proposition 8.3 *For $\mathcal{G} = \mathcal{G}(A \cup B)$ of small type we have*

- *$\mathcal{G}$ is recessive if and only if $\mathfrak{C}(\mathcal{G})$ is spherical,*
- *$\mathcal{G}$ is critical if and only if $\mathfrak{C}(\mathcal{G})$ is affine, and*
- *$\mathcal{G}$ is non-critical dominant if and only if $\mathfrak{C}(\mathcal{G})$ is hyperbolic or higher-rank.*

This proposition and the construction mentioned in the previous section begin to shed light on an interesting connection between Coxeter and Artin groups and the work presented so far. The following result of McMullen [43] indicates that the connection is stronger still (see also [7] and [2]).

Theorem 8.4 (McMullen) *Suppose $\mathcal{G}$ is bipartite and has small type. Then over all essential $\phi \in \mathfrak{C}(\mathcal{G})$, the spectral radius of $\Theta(\phi)$ is minimized by the bi-colored Coxeter element $\phi = \sigma_A \sigma_B$, and is given as the larger absolute value of a root of the polynomial*

$$x^2 + x(2 - (\mu(\mathcal{G}))^2) + 1$$

For spherical or affine Coxeter groups, this minimum is 1, and among all hyperbolic or higher-rank Coxeter groups, the minimal spectral radius is uniquely minimized for the Coxeter group $\mathfrak{C}(\mathcal{E}h_{10})$, and the minimal spectral radius is precisely λ_L.

Theorem 8.1 is an immediate consequence of Theorem 6.2, Proposition 8.3, and Theorem 8.4. □

Remarks

(1) McMullen's theorem does not require $\mathcal{G}$ to have small type, although in that case the appearance of $\mu(\mathcal{G})$ in the theorem is replaced by the spectral radius of the *Coxeter adjacency matrix.* The bipartite assumption is also unnecessary, however in this case, the number given by the theorem is only a lower bound, not the minimum.

(2) We note that for a random element of $\langle \sigma_A, \sigma_B \rangle < \mathfrak{A}(\mathcal{G})$, there is no connection between is spectral radius under $\Theta \circ \pi_{ac}$ and its dilatation under Ψ. In particular, solving the minimization problem for one does not solve it for the other.

8.4 Coxeter links and Coxeter actions

In [28], Hironaka provides a construction of a fibered link in S^3 which depends on a Coxeter graph of small type (as well as some additional data). This expresses the link complement as the mapping torus of an automorphism of the fiber called the *monodromy.* The main theorem of [28] states that, up to sign, the action on homology of the monodromy is conjugate to the geometric action of a certain Coxeter element (see below for the precise statement). As we shall see, under certain additional hypotheses, the monodromy is of the form $T_A T_B$ for appropriate $A \cup B$ filling the fiber.

We now describe Hironaka's construction (for more details, see [28]). A *chord diagram* is a collection of straight arcs $\mathcal{L} = \{l_1, \ldots, l_K\}$, called *chords*, in the unit disk $\mathbb{D} \subset \mathbb{R}^2$ connecting mutually disjoint pairs of points on the boundary of $\mathbb{D}$. The chord diagram defines a Coxeter graph $\mathcal{G}$ of small type as follows. The vertices Σ are identified with the chords of $\mathcal{L}$, and two vertices s_i and s_j are joined by an edge if and only if the corresponding chords intersect non-trivially (see Figure 20). For a chord diagram $\mathcal{L}$ defining a Coxeter graph $\mathcal{G}$, we say that $\mathcal{L}$ is a *chord realization* of $\mathcal{G}$, and that the graph is *chord-realizable.*

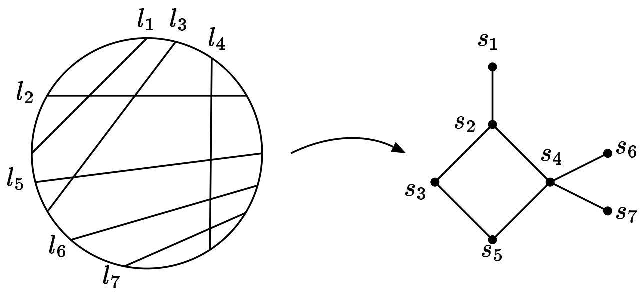

Figure 20: A Coxeter graph from a chord diagram

Suppose that $\mathcal{G}$ is a chord-realizable Coxeter graph. An ordering on the vertices $\Sigma = \{s_1, \ldots, s_K\}$ (equivalently, an ordering on the chords $\mathcal{L} = \{l_1, \ldots, l_K\}$) gives rise to a fibered link as follows. Recall that a *Hopf band* H is an annulus spanning a Hopf link L (see Figure 21). For each chord we plumb a right-handed Hopf band onto the disk in S^3 so that the core of the band agrees with the chord in the disk. We do this in the order specified by the ordering of the chords (see Figure 21). We denote the resulting surface by S, and its boundary by $L = \partial S$. It is well known that L is a fibered link (see [21]). We also note that the ordering of the vertices also specifies a Coxeter element $c = s_1 s_2 \cdots s_K$.

Finally, if we orient the chords in a chord diagram so that the ordering is *compatible* with the orientation, then the resulting link is said to be a *Coxeter link.* The compatibility here simply means that for $i < j$, the chord s_i must intersect the chord s_j positively (if at all). For example, the ordering of the chords in Figure 21 is compatible with the orientations.

The following is proved in [28].

Theorem 8.5 (Hironaka) *Given an oriented, ordered chord diagram with associated Coxeter graph $\mathcal{G}$, Coxeter link $L = \partial S$, fiber S, and monodromy ϕ, there exists an isomorphism*

$$\nu \colon \mathbb{R}^K \to H_1(S; \mathbb{R})$$

such that $\phi_* \circ \nu = -\nu \circ \Theta(c)$ where c is the Coxeter element determined by the ordering. If the spectral radius of ϕ_* is greater than 1, then it is bounded below by λ_L.

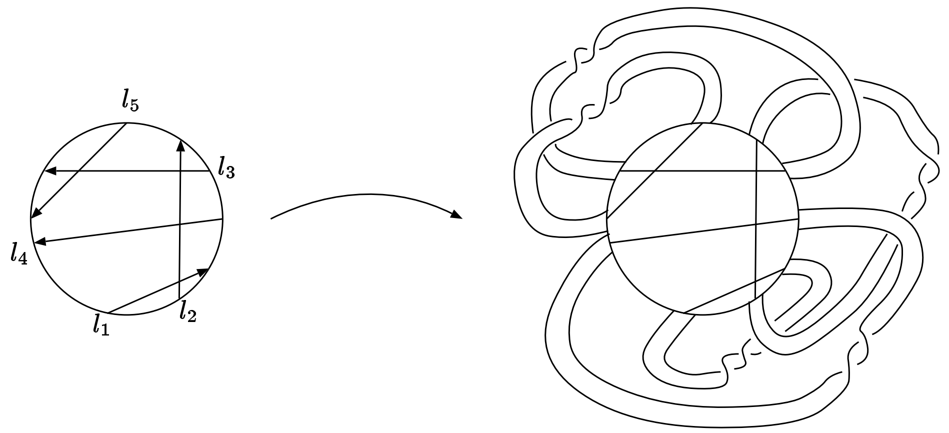

Figure 21: Plumbing Hopf bands onto a chord diagram

Hironaka's proof uses the following interpretation of a theorem of Howlett [29] in the case that $\mathcal{G}$ has small type.

Theorem 8.6 (Howlett) *If $\mathcal{G}$ is of small type, then the Coxeter element c is given by*

$$c = -(I - \mathcal{A}d(\mathcal{G})^+)^{-1}(I - \mathcal{A}d(\mathcal{G})^+)^t$$

where $\mathcal{A}d(\mathcal{G})^+$ is the upper triangular part of the adjacency matrix $\mathcal{A}d(\mathcal{G})$.

To prove her theorem, Hironaka shows that the Seifert matrix for the link is given by $I - \mathcal{A}d(\mathcal{G})^+$. It then follows from classical knot theory (see eg [52]) that the action of the monodromy on homology is given by c, as required.

Remark There is another construction of fibered links for which Dehn twists and Coxeter diagrams appear very naturally. This is described by A'Campo in [3], [4], and the references contained therein. Although we have not fully investigated this, it seems likely that this construction is closely related to the one described above.

To relate Hironaka's Theorem to our work, we recall that according to Gabai [20] the monodromy of a fibered link obtained by (generalized) plumbing of two fibers is the composition of the two monodromies (see [20] for a more precise statement). The monodromy for a Hopf link (with fiber a right-handed Hopf band) is a positive Dehn twist about the core of the band. It follows that the monodromy of the Coxeter link constructed above is the product of Dehn twists about the cores of the plumbed on Hopf bands (the product is taken, from left to right, in the order given by the ordering of the chords).

Suppose now that a chord diagram has bipartite Coxeter graph $\mathcal{G}$ with vertices Σ colored by A and B, and there is an ordering of the vertices so that for all $s_i \in A$ and $s_j \in B$, we have $i < j$. We call this a *bi-colored ordering* with respect to A and B. The cores of the Hopf bands associated to A (respectively, B) give an essential 1–manifold we also denote by A (respectively, B) in the surface S. It is easy to see that $\mathcal{G} = \mathcal{G}(A \cup B)$.

The previous two paragraphs imply the following theorem.

Theorem 8.7 *In the setting of Theorem 8.5, if we further assume that the ordering is a bi-colored ordering with respect to A and B, then $\phi = T_A T_B$. In particular, the action of $(T_A T_B)_*$ on $H_1(S;\mathbb{R})$ is conjugate to the action of $-\Theta(\sigma_A \sigma_B)$ on $\mathbb{R}^K$.*

It is not hard to see that this implies Theorem 8.2 in the special case that $T_A T_B$ is the monodromy for a Coxeter link. We now give the proof in the general case.

Proof of Theorem 8.2 Suppose $\mathcal{G}(A \cup B)$ has small type and we have oriented A and B so that all intersections of A with B are positive. So, for any pair of components $a_i \subset A$ and $b_j \subset B$, we have

$$a_i \cdot b_j = i(a_i, b_j) = -b_j \cdot a_i. \tag{6}$$

Let $\mathcal{N}(A \cup B)$ be a regular neighborhood of $A \cup B$ in $\mathcal{S}$, and let us denote the inclusion into S by $\iota\colon \mathcal{N}(A \cup B) \to S$.

We may define a monomorphism

$$\eta_0\colon \mathbb{R}^K \to H_1(\mathcal{N}(A \cup B);\mathbb{R})$$

by sending each basis element to the homology class of the oriented curve it determines. We view $a_1, \ldots, a_n, b_1, \ldots, b_m$ as an ordered basis for both $\mathbb{R}^K$ as well as the subspace $V \subset H_1(\mathcal{N}(A \cup B);\mathbb{R})$ which they span. To see that these are indeed linearly independent in $H_1(\mathcal{N}(A \cup B);\mathbb{R})$, we note that for any one of these, say a_1, we can easily find an arc α which intersects a_1 once, but misses all the others. This arc determines an element of $H_1(\mathcal{N}(A \cup B), \partial\mathcal{N}(A \cup B);\mathbb{R})$, which by Poincaré duality, is identified with the dual space of $H_1(\mathcal{N}(A \cup B);\mathbb{R})$ via intersection numbers. Thus there is a functional vanishing on all the vectors except a_1. Since a_1 was arbitrary, the vectors are linearly independent.

The action on homology of a Dehn twist T_a is given by

$$(T_a)_*(x) = x + (a \cdot x)a.$$

By (6), the matrix for the actions of T_A and T_B on V with respect to the basis $a_1, \ldots, a_n, b_1, \ldots, b_m$ is thus given by

$$(T_A)_* = \begin{pmatrix} I & N \\ 0 & I \end{pmatrix} \quad \text{and} \quad (T_B)_* = \begin{pmatrix} I & 0 \\ -N^t & I \end{pmatrix}$$

where $N_{ij} = i(a_i, b_j)$ as in Section 5.1.

Now, by Theorem 8.6 and Proposition 5.1, we have:

$$\begin{aligned} \Theta(\sigma_A \sigma_B) &= -(I - \mathcal{A}d(\mathcal{G})^+)^{-1}(I - \mathcal{A}d(\mathcal{G})^+)^t \\ &= -\begin{pmatrix} I & -N \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} I & -N \\ 0 & I \end{pmatrix}^t = -\begin{pmatrix} I & N \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ -N^t & I \end{pmatrix} \\ &= -(T_A)_*(T_B)_* = -(T_A T_B)_* \end{aligned}$$

The matrices are given with respect to $a_1, \ldots, a_n, b_1, \ldots, b_m$ in both $\mathbb{R}^K$ and $H_1(\mathcal{N}(A \cup B); \mathbb{R})$. Because η_0 is "the identity" with respect to this basis, we see that

$$(T_A T_B)_* \circ \eta_0 = -\eta_0 \circ \Theta(\sigma_A \sigma_B). \tag{7}$$

We now obtain the required homomorphism

$$\eta = \iota_* \circ \eta_0 \colon \mathbb{R}^K \to H_1(S; \mathbb{R}).$$

Because $T_A T_B$ is supported on $\mathcal{N}(A \cup B)$, we have $\iota \circ T_A T_B = T_A T_B \circ \iota$, and hence

$$(T_A T_B)_* \circ \eta = -\eta \circ \Theta(\sigma_A \sigma_B).$$

This proves Theorem 8.2, except for the last sentence.

To see this, note that the kernel of η is the image of the kernel of ι_* by η_0^{-1}. Since S is obtained from $\mathcal{N}(A \cup B)$ by gluing disks to the boundary, we see that the kernel of ι_* consists of the span of the homology classes of the boundary. However, $T_A T_B$ fixes the boundary pointwise, and so acts as the identity on this span in H_1. Therefore, by (7), $\Theta(\sigma_A \sigma_B)$ acts as $-I$ on the kernel of η.

Finally, we see that spectral radii are preserved in the pseudo-Anosov case by Theorem 8.1 and the fact that the dilatation is equal to the spectral radius (see [44]). The only other case is when $\mathcal{G}(A \cup B)$ is recessive or critical. In these cases, the spectral radius (which is 1) is necessarily preserved. □

9 Applications and questions

Here we provide a few applications of our work and state a few interesting questions.

9.1 Lehmer's question, Salem numbers, and Teichmüller curves

The interest in Lehmer's number stems from a problem in number theory known as Lehmer's question (see [35]). To state it, we recall that given a monic integral polynomial $p(x) \in \mathbb{Z}[x]$, the *Mahler measure* of p is defined by

$$\Omega(p) = \prod_{p(\theta)=0} \max\{1, |\theta|\}.$$

Question 9.1 (Lehmer) *Is there an $\epsilon > 1$ such that if $\Omega(p) > 1$, then $\Omega(p) \geq \epsilon$?*

At present, the smallest known Mahler measure greater than 1 occurs for Lehmer's polynomial (1), Section 1.3, and is equal to λ_L. One may view Theorem 6.2 as a resolution of Lehmer's question in a particular situation. More precisely, if we let $\mathfrak{D}_2$ denote the set of all dilatations of pseudo-Anosov elements in groups generated by two positive multi-twists, then Theorem 6.2 implies the following.

Corollary 9.2 *The Mahler measure of the minimal polynomial of any element of $\mathfrak{D}_2$ is bounded below by λ_L.*

A *Salem number* is an algebraic integer $\lambda > 1$, such that the Galois conjugates include λ^{-1} and all (except λ) lie in the unit disk. Note that a Salem number is equal to the Mahler measure of its minimal polynomial. In particular, an affirmative answer to the following (see [10] and [24]) would be a consequence of such an answer to Lehmer's question.

Question 9.3 *Is there an $\epsilon > 1$ such every Salem number λ satisfies $\lambda \geq \epsilon$?*

Lehmer's number is a Salem number, so of course the best guess for ϵ is λ_L. Because of this question, one is generally interested in "small" Salem numbers. There are currently 47 known Salem numbers less than 1.3, including λ_L (see [10], [11], [45], and also [19]). However, we only obtain 5 small Salem numbers as elements of $\mathfrak{D}_2$. This set consists of all but 1 of the Salem numbers obtained by McMullen in [43] as spectral radii of certain elements of Coxeter groups. This is not surprising, given Theorem 6.2 and the fact that 5 of the 6 small Salem numbers obtained by McMullen come from bicolored Coxeter elements. On the other hand, the only dilatations in $\mathfrak{D}_2$ which can occur in the interval $(1, 1.3)$ are of the form $\lambda(T_A T_B)$, for $\mathcal{G}(A \cup B)$ of small type (see the proof of

Theorem 4.4 and use Proposition 6.5 along with the fact that there is exactly one simple closed geodesic when $\mathcal{G}(A \cup B)$ is non-critical dominant).

The elements of $\mathfrak{D}_2$ are not at all representative of the general case of dilatations of pseudo-Anosov automorphisms which are *not* bounded away from 1 (see Penner [48], Bauer [5], and McMullen [42]). In particular, we ask the following:

Question 9.4 *Which Salem numbers occur as dilatations of pseudo-Anosov automorphisms?*

Question 9.5 *Is there some topological condition on a pseudo-Anosov which guarantees that its dilatation is a Salem number?*

In the same vein as Questions 9.1 and 9.3, we ask the following:

Question 9.6 *Is there an $\epsilon > 1$, such that if ϕ is a pseudo-Anosov automorphism in a finite co-area Teichmüller disk stabilizer, then $\lambda(\phi) \geq \epsilon$?*

Given that the dilatations we are obtaining are naturally occurring as spectral radii of hyperbolic elements in certain non-elementary Fuchsian groups, we would be remiss not to mention the following (see [46], [37], and also [24]).

Theorem 9.7 (Neumann–Reid) *The Salem numbers are precisely the spectral radii of hyperbolic elements of arithmetic Fuchsian groups derived from quaternion algebras.*

However, because the non-cocompact arithmetic Fuchsian groups are necessarily commensurable with $PSL_2\mathbb{Z}$, relatively few of the groups generated by two positive multi-twists even inject into arithmetic groups.

9.2 Unexpected multi-twists and the 3−chain relation

The work in this paper has a connection to a problem posed by McCarthy at the 2002 AMS meeting in Ann Arbor, MI. This was to determine the extent to which the lantern relation in the mapping class group is characterized by its algebraic properties (in particular the intersection patterns of the defining curves). Two different solutions to this were obtained, independently by Hamidi-Tehrani in [26], and by Margalit in [39].

This question asks us to decide when an element in a group generated by two Dehn twists can be a multi-twist. One could ask the same question more generally, ie for positive multi-twists. The answer is given, to a certain extent, by Proposition 6.4, Theorem 7.3, and Theorem 3.1. We do not spell this out here, but instead provide a partial answer to a related question posed by Margalit in [39]. I am grateful to Joan Birman for pointing this out to me.

Margalit asks to what extent the n–*chain relation* can be characterized. This is the relation

$$(T_{a_1}T_{a_2}\cdots T_{a_n})^k = M$$

where:

- $a_1,\ldots,a_n$ are essential simple closed curves on a surface with $i(a_i,a_{i+1}) = 1$, $i=1,\ldots,n-1$, and all other intersection numbers 0,
- M is either T_d or $T_{d_1}T_{d_2}$, where d or $d_1\cup d_2$ is the boundary of a regular neighborhood of $a_1\cup\cdots\cup a_n$ (depending on whether n is even or odd, respectively), and
- $k=2n+2$ for n even, and $k=n+1$ for n odd.

Margalit gives a characterization for $n=2$, which we state here.

Theorem 9.8 (Margalit) *Suppose $M=(T_xT_y)^k$, where M is a multi-twist and $k\in\mathbb{Z}$, is a non-trivial relation between powers of Dehn twists in* $\mathrm{Mod}(S)$, *and $[M,T_x]=1$. Then the given relation is the 2–chain relation, ie $M=T_c^j$, where c is the boundary of a neighborhood of $x\cup y$, $i(x,y)=1$, and $k=6j$.*

We note that although our work has been primarily concerned with groups generated by two multi-twists, we can in fact obtain a similar characterization of the 3–chain relation.

Theorem 9.9 *Suppose $M=(T_xT_yT_z)^k$, where M is a multi-twist and $k\in\mathbb{Z}$, is a non-trivial relation between powers of Dehn twists in* $\mathrm{Mod}(S)$, *and $[M,T_x]=[T_x,T_z]=1$. Then the given relation is the 3–chain relation, ie $M=(T_cT_d)^j$, where $c\cup d$ is the boundary of a neighborhood of $x\cup y\cup z$, $i(x,y)=i(y,z)=1$, and $k=4j$.*

The non-triviality here means that $i(x,y)\neq 0$, $i(y,z)\neq 0$, and $k\neq 0$.

Proof Since $[T_x,M]=1$, conjugating by T_x^{-1}, we obtain

$$M=T_x^{-1}(T_xT_yT_z)^kT_x=(T_yT_zT_x)^k=(T_y(T_zT_x))^k.$$

Also, T_zT_x is a positive multi-twist since $[T_x, T_z] = 1$ implies $i(x,z) = 0$.

Proposition 6.4 tells us that $\mathcal{G}(y \cup (z \cup x))$ is recessive or critical (otherwise $T_yT_zT_x$, and all of its powers, would be pseudo-Anosov on the subsurface filled by $y \cup (z \cup x)$, contradicting the fact that some power is a multi-twist). The only such graph with 3 vertices is $\mathcal{A}_3$. □

10 Appendix A: The nonfilling case

Here we provide a proof of the following.

Proposition 10.1 *It suffices to prove Theorem 6.1 for $A \cup B$ filling S.*

As the proof will require us to deal with surfaces having nonempty boundary, we can also allow S to have boundary with no added complications. In particular, Theorem 6.1 remains true in this setting. Now, an allowable homeomorphism is one which leaves the marked points invariant and fixes the boundary components pointwise. The definition of $\mathrm{Mod}(S)$ is as in Section 2.3.

On S, consider two elements $A, B \in \mathcal{S}'(S)$. Let $\mathcal{N}(A \cup B)$ denote the regular neighborhood of $A \cup B$ in S. Write $\overline{S} = \overline{S}_{A\cup B}$ for the subsurface of S obtained by taking the union of $\mathcal{N}(A \cup B)$ with any open disks, once-marked open disks, and half open annuli in the complement of $\mathcal{N}(A \cup B)$.

Next, let $\widehat{S}$ be the surface obtained from $\overline{S}$ by gluing a disk with one marked point to each boundary component, and write

$$\varepsilon\colon \overline{S} \to \widehat{S}$$

for the inclusion. We let A and B denote the images under ϵ of the 1–manifolds of the same name. The components of $\widehat{S}$ bijectively correspond to the components, $A_1 \cup B_1, \ldots, A_k \cup B_k$, of $A \cup B$, and we write these as $\widehat{S}_1, \ldots, \widehat{S}_k$.

Note that $A_r \cup B_r$ fills each component $\widehat{S}_r$, except when $\widehat{S}_r \cong S_{0,2}$. In this situation $A_r \cup B_r$ is a single closed curve which is not essential in $\widehat{S}_r$. However, it should be clear from what follows that this technicality may be ignored.

The groups we need to consider are

$$G = \langle T_A, T_B \rangle < \mathrm{Mod}(S) \qquad \overline{G} = \langle T_A, T_B \rangle < \mathrm{Mod}(\overline{S})$$
$$\widehat{G} = \langle T_A, T_B \rangle < \mathrm{Mod}(\widehat{S}) \qquad \widehat{G}_r = \langle T_{A_r}, T_{B_r} \rangle < \mathrm{Mod}(\widehat{S}_r)$$

for each $r = 1, \ldots, k$. G is the group from Theorem 6.1.

Proposition 10.1 follows easily from the next proposition since $A_r \cup B_r$ fills each of $\widehat{S}_r$.

Proposition 10.2 *For G and $\widehat{G}_1, \dots, \widehat{G}_k$ as above*

$$G \cong \mathbb{F}_2 \Leftrightarrow \widehat{G}_r \cong \mathbb{F}_2$$

for some $r = 1, \dots, k$.

Proof The map ε induces an epimorphism

$$\varepsilon_* \colon \operatorname{Mod}(\overline{S}) \to \operatorname{Mod}(\widehat{S}).$$

Moreover, the kernel of ε_* is generated by Dehn twists about curves parallel to the boundary components of $\overline{S}$, which defines the following central extension [9]

$$0 \to \mathbb{Z}^{|\partial \overline{S}|} \to \operatorname{Mod}(\overline{S}) \to \operatorname{Mod}(\widehat{S}) \to 0.$$

We write

$$\overline{\varepsilon}_* \colon \overline{G} \to \widehat{G}$$

to denote the restricted epimorphism.

The inclusion

$$i \colon \overline{S} \to S$$

also induces a homomorphism

$$i_* \colon \operatorname{Mod}(\overline{S}) \to \operatorname{Mod}(S).$$

One can show that the kernel of i_* is contained in the kernel of ε_*. We write

$$\overline{i}_* \colon \overline{G} \to G$$

to denote the restriction of i_* to $\overline{G}$. By construction, $\overline{i}_*$ is surjective.

We also note that

$$\operatorname{Mod}(\widehat{S}) \cong \prod_{r=1}^{k} \operatorname{Mod}(\widehat{S}_r)$$

which allows us to view $\widehat{G}$ as a subgroup of the direct product

$$\widehat{G} < \prod_{r=1}^{k} \widehat{G}_r.$$

Denote the projection onto the r^{th} factor by

$$\pi_r \colon \widehat{G} \to \widehat{G}_r$$

and note that this is surjective.

Suppose now that there exists an isomorphism $\widehat{G}_r \to \mathbb{F}_2$ for some r. $\overline{G}$ is two-generator, hence a quotient of $\mathbb{F}_2$, so we have:

$$\begin{array}{ccccccccc} \mathbb{F}_2 & \longrightarrow & \overline{G} & \xrightarrow{\overline{\varepsilon}_*} & \widehat{G} & \xrightarrow{\pi_r} & \widehat{G}_r & \longrightarrow & \mathbb{F}_2 \\ & & {\scriptstyle \overline{i}_*}\big\downarrow & & & & & & \\ & & G & & & & & & \end{array}$$

All the arrows are surjections, and free groups are Hopfian (see [38]), so the composition of all the horizontal arrows is an isomorphism. Therefore, all horizontal arrows are isomorphisms, and in particular $\overline{G} \cong \mathbb{F}_2$.

Since $\overline{i}_*$ is surjective, we'll have $G \cong \mathbb{F}_2$ if $\overline{i}_*$ is also injective. The kernel of $\overline{i}_*$ is contained in the kernel of ε_*, and is therefore contained in the center of $\overline{G}$. Since $\overline{G} \cong \mathbb{F}_2$, the center is trivial and so $\overline{i}_*$ is injective.

Now suppose that $G \cong \mathbb{F}_2$, and note that this guarantees that $\overline{G} \cong \mathbb{F}_2$, again appealing to the Hopfian property. Because the kernel of $\overline{\varepsilon}_*$ is central, it follows that $\widehat{G} \cong \mathbb{F}_2$. We need to verify that $\widehat{G}_r \cong \mathbb{F}_2$ for some r. If this were not the case, then $K_r = \ker(\pi_r)$ is a non-trivial normal subgroup of $\widehat{G}$ for each r. An easy induction argument shows that the commutator group

$$[\dots[[[K_1, K_2], K_3], K_4], \dots, K_k]$$

is contained in each K_r, and hence must be trivial in $\widehat{G}$. Since $\widehat{G} \cong \mathbb{F}_2$, any commutator subgroup of non-trivial normal subgroups must be non-trivial. This contradiction proves the proposition. □

11 Appendix B: Penner's construction

In [47], Penner gives a generalization of a special case of Thurston's construction for pseudo-Anosov automorphisms. In this section, we show that the lower bound λ_L remains valid for this class of pseudo-Anosov automorphisms. In fact, we show that the the dilatations of pseudo-Anosov automorphisms obtained from this construction are bounded below by $\sqrt{5} > \lambda_L$.

We begin by describing Penner's construction. Consider $A, B \in \mathcal{S}'(S)$, label the components $A = a_1 \cup \cdots \cup a_n$ and $B = b_1 \cup \cdots \cup b_m$, and suppose that $A \cup B$ fills S. Consider the semi-group $\mathfrak{G}(A, B)$ consisting of all automorphisms of the form

$$\prod_{k=0}^{N} T_{a_{l_k}}^{\epsilon_k} T_{b_{s_k}}^{-\delta_k} \tag{8}$$

where $N, \epsilon_k, \delta_k \in \mathbb{Z}_{\geq 0}$. That is, $\mathfrak{G}(A,B)$ consists of all possible products of positive Dehn twists about components of A and negative Dehn twists about components of B.

There is a subsemigroup $\mathfrak{G}_0(A,B)$ consisting of all elements of $\mathfrak{G}(A,B)$ for which every component of A and B is twisted along non-trivially at least once in the above product. In [47], Penner proves:

Theorem 11.1 (Penner) *$\mathfrak{G}_0(A,B)$ consists entirely of pseudo-Anosov automorphisms.*

Note that $\mathfrak{G}_0(A,B)$ contains all the elements of $\langle T_A, T_B \rangle$ representable as words in T_A and T_B where T_A (respectively, T_B) appears with all positive (respectively, negative) exponents. However, this is a relatively small subset of $\mathfrak{G}_0(A,B)$, as most elements of $\mathfrak{G}_0(A,B)$ do not obviously lie in $\langle T_A, T_B \rangle$. Thus, Penner's construction generalizes a particular case of the construction we have been considering.

The method which Penner uses to prove Theorem 11.1 allows one to easily obtain the following bound.

Theorem 11.2 *The dilatation of any element of $\mathfrak{G}_0(A,B)$ is bounded below by $\sqrt{5}$.*

The proof we give uses the methods described in [47]. We refer the reader to that paper for a more complete description of those techniques. We also note that the estimates we give are rough, and this bound is likely not sharp, though we do not prove this.

Proof Fix an element $\phi \in \mathfrak{G}_0(A,B)$. As in [47], we will consider ϕ^2 instead of ϕ. Since $\lambda(\phi^2) = (\lambda(\phi))^2$, it suffices to prove that $\lambda(\phi^2) \geq 5$.

At each intersection point of a component a_l of A with a component b_s of B, apply a homotopy of b_s so that it meets a_l as in Figure 22. The union of the resulting curves is a *bigon track*, τ (this is essentially a train track except we have weakened the non-degeneracy condition on complementary regions, allowing bigons). Let us denote the branches by $\beta_1, \ldots, \beta_K$.

Next, we represent ϕ^2 as a product of Dehn twists in a particular way so that $\phi^2(\tau)$ is easily seen to be carried by τ. For each component c of A and of B one takes two push-offs, $c^{\pm}$, one on each side of c. We then express ϕ^2 as a product of twists along the push-offs, rather than the original curves. Because

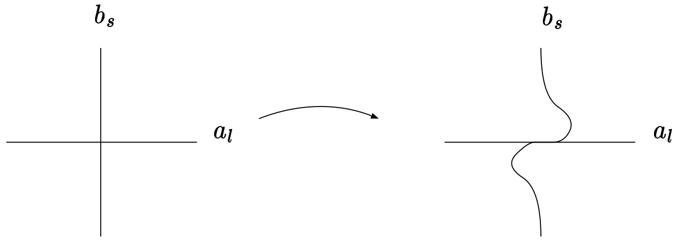

Figure 22: Modifying intersection points

every curve which we twist along in ϕ shows up twice as many times in ϕ^2, we can arrange that we twist along *both* push-offs in ϕ^2. We do this so that we twist along all positive push-offs in the first application of ϕ and then along negative push-offs in the second application. Thus, if ϕ is given by the product in (8), we have

$$\phi^2 = \prod_{k=0}^{N} T_{a_{l_k}^-}^{\epsilon_k} T_{b_{s_k}^-}^{-\delta_k} \prod_{k=0}^{N} T_{a_{l_k}^+}^{\epsilon_k} T_{b_{s_k}^+}^{-\delta_k}.$$

For each a_l^+, $T_{a_l^+}(\tau)$ is carried by τ, as is indicated by Figure 23 in the case that $i(a_l, B) = 1$. Let us write $M^{a_l^+}$ to denote the incidence matrix describing how τ carries $T_{a_l^+}(\tau)$. One can verify that $M^{a_l^+}$ has the form

$$M^{a_l^+} = I + R^{a_l^+}$$

where I is the $K \times K$ identity matrix, and $R^{a_l^+}$ is a non-negative integral matrix. Moreover, if β_p is any branch contained in a_l and β_q is a branch contained in B which intersects a_l^+, then the (p,q)–entry satisfies $(R^{a_l^+})_{pq} = 1$. Similar statements hold for push-offs a_l^-, b_s^+, and b_s^-.

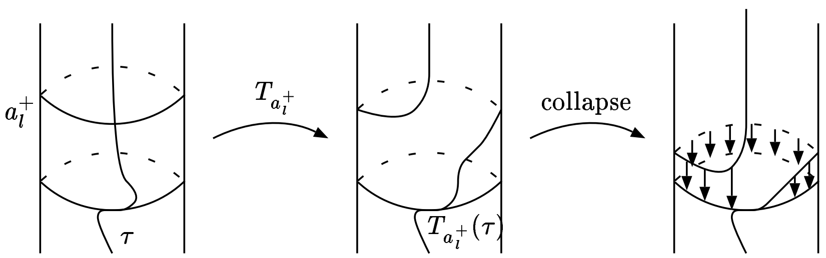

Figure 23: τ carrying $T_{a_l^+}(\tau)$

In particular, suppose that a_l and b_s intersect in at least one point ξ, and let $\beta_{i^+}, \beta_{i^-}, \beta_{j^+}, \beta_{j^-}$ be the branches of τ around ξ as indicated in Figure 24. The (p,q)–entries of $R^{c^\pm}$ for $c = a_l$ or b_s satisfy:

$$\begin{array}{llll} (R^{a_l^+})_{pq} = 1 & \text{for } p = i^\pm,\ q = j^+ & (R^{a_l^-})_{pq} = 1 & \text{for } p = i^\pm,\ q = j^- \\ (R^{b_s^+})_{pq} = 1 & \text{for } p = j^\pm,\ q = i^+ & (R^{b_s^-})_{pq} = 1 & \text{for } p = j^\pm,\ q = i^- \end{array}$$

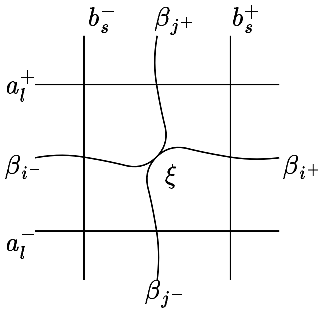

Figure 24: The branches around the intersection point ξ

Now, the incidence matrix M describing τ carrying $\phi^2(\tau)$ is given by the product

$$M = \prod_{k=0}^{N} (M^{a^-_{l_k}})^{\epsilon_k} (M^{b^-_{s_k}})^{-\delta_k} \prod_{k=0}^{N} (M^{a^+_{l_k}})^{\epsilon_k} (M^{b^+_{s_k}})^{-\delta_k}.$$

It is not hard to see that one of $M^{b_s^+}$ or $M^{b_s^-}$ occurs between $M^{a_l^-}$ and $M^{a_l^+}$ in this product (these matrices all occur since $\phi \in \mathfrak{G}_0(A,B)$). So, we may write

$$M = X_0 M^{a_l^-} X_1 M^{b_s^\sigma} X_2 M^{a_l^+} X_3$$

where $X_t = I + Y_t$, and Y_t is a non-negative integral matrix, for $0 \le t \le 3$, and $\sigma \in \{+,-\}$. Expanding this out, we see that

$$\begin{aligned} M &= (I + Y_0)(I + R^{a_l^-})(I + Y_1)(I + R^{b_s^\sigma})(I + Y_2)(I + R^{a_l^+})(I + Y_3) \\ &= I + R^{a_l^-} + R^{a_l^+} + R^{a_l^-} R^{b_s^\sigma} + R^{a_l^-} R^{b_s^\sigma} R^{a_l^+} + Z \end{aligned}$$

where Z is a non-negative integral matrix. Using the above values for $(R^{c^\pm})_{pq}$, one can check that each of the first 5 matrices in this last sum has a positive entry in the $(i^\pm)$th rows. It follows that the sum of the entries in each of the $(i^\pm)$th rows of M is at least 5.

The $\beta_{i^\pm}$ were arbitrary branches contained in A: a_l and b_s were arbitrary, and every branch in A is adjacent to some intersection point (eg $(R^{a_l^-})_{i^\pm j^-} = 1$,

$(R^{b_s^\sigma})_{j^- i^\sigma} = 1$, so the $i^\pm i^\sigma$ entry of the third term is at least 1). Therefore, every row of M with index corresponding to a branch in A has the sum of its entries being at least 5. A similar argument can be made for branches contained in B, and thus it follows that every row of M has sum at least 5. Appealing to Theorem 4.1, we see that the PF eigenvalue of M is at least 5: take $\vec{U}$ to be the vector with all entries equal to 1, and apply the first inequality of the theorem.

The following lemma, which is implicit in the proof of Theorem 11.1 given in [47] completes the proof. □

Lemma 11.3 (Penner) *The PF eigenvalue of M is $\lambda(\phi^2)$.*

References

[1] **N A'Campo**, *Le groupe de monodromie du déploiement des singularités isolées de courbes planes. I*, Math. Ann. 213 (1975) 1–32

[2] **N A'Campo**, *Sur les valeurs propres de la transformation de Coxeter*, Invent. Math. 33 (1976) 61–67

[3] **N A'Campo**, *Generic immersions of curves, knots, monodromy and gordian number*, Publ. Math., Inst. Hautes Étud. Sci. 88 (1998) 151–169

[4] **N A'Campo**, *Planar trees, slalom curves and hyperbolic knots* Publ. Math., Inst. Hautes Étud. Sci. 88 (1998) 171–180

[5] **M Bauer**, *An upper bound for the least dilatation*, Trans. Am. Math. Soc. 330 (1992) 361–370

[6] **A F Beardon**, *The Geometry of Discrete Groups*, Graduate Texts in Mathematics 91, Springer–Verlag, New York (1983)

[7] **S Berman**, **Y S Lee**, **R V Moody**, *The spectrum of a Coxeter transformation, affine Coxeter transformations, and the defect map*, J. Algebra 121 (1989) 339–357

[8] **J Birman**, *Braids, links, and mapping class groups*, Ann. Math. Stud. 82, Princeton University Press, Princeton, NJ (1974)

[9] **J Birman**, *Mapping Class Groups*, Notes from a Columbia University course, Fall 2002

[10] **D W Boyd**, *Small Salem numbers*, Duke Math. J. 44 (1977) 315–328

[11] **D W Boyd**, *Pisot and Salem numbers in intervals of the real line*, Math. Comput. 32 (1978) 1244–1260

[12] **A E Brouwer**, **A Neumaier**, *The graphs with spectral radius $\sqrt{2+\sqrt{5}}$*, Linear Algebra Appl. 114/115 (1989) 273–276

[13] **K Calta**, *Veech surfaces and complete periodicity in genus* 2, J. Am. Math. Soc. 17 (2004) 871–908

[14] **J Crisp**, **L Paris**, *The solution to a conjecture of Tits on the subgroup generated by the squares of the generators of an Artin group*, Invent. Math. 145 (2001) 19–36

[15] **D Cvetković**, **M Doob**, **I Gutman**, *On graphs whose spectral radius does not exceed* $\sqrt{2+\sqrt{5}}$, Ars Combin. 14 (1982) 225–239

[16] **D Cvetković**, **P Rowlinson**, *The largest eigenvalue of a graph: A survey*, Linear and Multilinear Algebra 28 (1990) 3–33

[17] **C J Earle**, **F P Gardiner**, *Teichmüller disks and Veech's $\mathcal{F}$–structures*, from "Extremal Riemann surfaces (San Francisco, CA, 1995)", Contemp. Math. 201, Am. Math. Soc., Providence, RI (1997) 165–189

[18] **A Fathi**, **F Laudenbach**, **V Poenaru**, et. al., *Travaux de Thurston sur les surfaces*, Astérisque 66–67 (1979)

[19] **V Flammang**, **M Grandcolas**, **G Rhin**, *Small Salem numbers*, from "Number theory in progress, (Zakopane-Kościelisko, 1997)", de Gruyter, Berlin (1999) 165–168

[20] **D Gabai**, *The Murasugi sum is a natural geometric operation II*, from " Combinatorial methods in topology and algebraic geometry (Rochester, N.Y., 1982)", Contemp. Math. 44 (1985) 93–100

[21] **D Gabai**, *Detecting fibered links in* S^3, Comment. Math. Helv. 61 (1986) 519–555

[22] **F Gantmacher**, *The theory of matrices*, vol. 2, Chelsea (1959)

[23] **F P Gardiner**, **N Lakic**, *Quasiconformal Teichmüller theory*, Math. Surv. Monogr. 76, Am. Math. Soc., Providence, RI (2000)

[24] **E Ghate**, **E Hironaka**, *The arithmetic and geometry of Salem numbers*, Bull. Am. Math. Soc. (New Ser.) 38 (2001) 293–314

[25] **E Gutkin**, **C Judge**, *Affine mappings of translation surfaces: geometry and arithmetic*, Duke Math. J. 103 (2000) 191–213

[26] **H Hamidi-Tehrani**, *Groups generated by positive multi-twists and the fake lantern problem*, Algebr. Geom. Topol. 2 (2002) 1155–1178

[27] **W J Harvey**, *On certain families of compact Riemann surfaces*, from "Mapping class groups and moduli spaces of Riemann surfaces (Göttingen, 1991/Seattle, WA, 1991)", Contemp. Math. 150, Amer. Math. Soc., Providence, RI (1993) 137–148

[28] **E Hironaka**, *Chord diagrams and Coxeter links*, J. Lond. Math. Soc. (2) 69 (2004) 243–257

[29] **R B Howlett**, *Coxeter groups and M–matrices*, Bull. London Math. Soc. 14 (1982) 137–141

[30] **J E Humphreys**, *Reflection Groups and Coxeter Groups*, Cambr. Stud. Adv. Math. 29, Cambridge University Press (1990)

[31] **N V Ivanov**, *Subgroups of Teichmüller modular groups*, Transl. Math. Monogr. 115, Am. Math. Soc., Providence, RI (1992)

[32] **R Kenyon**, **J Smillie**, *Billiards on rational-angled triangles*, Comment. Math. Helv. 75 (2000) 65–108

[33] **S Kerckhoff**, **H Masur**, **J Smillie**, *Ergodicity of billiard flows and quadratic differentials*, Ann. Math. (2), 124 (1986) 293–311

[34] **I Kra**, *On the Nielsen–Thurston–Bers type of some self-maps of Riemann surfaces*, Acta. Math. 146 (1981) 231–270

[35] **D H Lehmer**, *Factorization of certain cyclotomic functions*, Ann. Math. 34 (1933) 461–479

[36] **D D Long**, *Constructing pseudo-Anosov map*, from "Knot theory and manifolds (Vancouver, BC, 1983)", Lecture Notes in Mathematics 1144, Springer–Verlag, Berlin (1985) 108–114

[37] **C Maclachlan**, **A W Reid**, *The arithmetic of hyperbolic 3–manifolds*, Graduate Texts in Mathematics 219, Springer–Verlag, New York (2003)

[38] **W Magnus**, **A Karrass**, **D Solitar**, *Combinatorial Group Theory*, Dover Publications (1976)

[39] **D Margalit**, *A lantern lemma*, Algebr. Geom. Topol. 2 (2002) 1179–1195

[40] **H Masur**, *Transitivity properties of the horocyclic and geodesic flows on moduli space*, J. Anal. Math. 39 (1981) 1–10

[41] **H Masur**, **S Tabachnikov**, *Rational billiards and flat structures*, from "Handbook of dynamical systems, Vol. 1A", North-Holland, Amsterdam (2002) 1015–1089

[42] **C T McMullen**, *Polynomial invariants for fibered 3–manifolds and Teichmüller geodesics for foliations*, Ann. Sci. Éc. Norm. Supér. (4) 33 (2000) 519–560

[43] **C T McMullen**, *Coxeter groups, Salem numbers and the Hilbert metric*, Publ. Math., Inst. Hautes Étud. Sci. 95 (2002) 151–183

[44] **C T McMullen**, *Billiards and Teichmüller curves on Hilbert modular surfaces*, J. Am. Math. Soc. 16 (2003) 857–885

[45] **M J Mossinghoff**, *Polynomials with small Mahler measure*, Math. Comput. 67 (1998) 1697–1705

[46] **W D Neumann**, **A W Reid**, *Arithmetic of hyperbolic manifolds*, from "Topology '90 (Columbus, OH, 1990)", Ohio State Univ. Math. Res. Inst. Publ. 1, de Gruyter, Berlin (1992) 273–310

[47] **R C Penner**, *A construction of pseudo-Anosov homeomorphisms*, Trans. Am. Math. Soc. 310 (1988) 179–197

[48] **R C Penner**, *Bounds on least dilatations*, Proc. Am. Math. Soc. 113 (1991) 443–450

[49] **B Perron**, **J P Vannier**, *Groupe de monodromie géométrique des singularités simples*, C. R. Acad. Sci. Paris Sér. I Math. 315 (1992) 1067–1070

[50] **J-C Puchta**, *On triangular billiards*, Comment. Math. Helv. 76 (2001) 501–505

[51] **J Ratcliffe**, *Foundations of Hyperbolic Manifolds*, Graduate Texts in Mathematics 149, Springer–Verlag, New York (1994)

[52] **D Rolfsen**, *Knots and Links*, Mathematics Lecture Series 7, Publish or Perish, Inc., (1990)

[53] **J H Smith**, *Some properties of the spectrum of a graph*, from "Combinatorial Structures and Their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969)", R Guy, et al. (editors), Gordon and Breach, New York (1970) 403–406

[54] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Am. Math. Soc. (New Ser.) 19 (1988) 417–431

[55] **W P Thurston**, *The Geometry and Topology of Three–Manifolds*, Princeton University course notes, `http://www.msri.org/publications/books/gt3m/` (1980)

[56] **W A Veech**, *Teichmüller curves in moduli space, Einstein series and an application to triangular billiards*, Invent. Math. 97 (1989) 553–583

[57] **W A Veech**, *The billiard in a regular polygon*, Geom. Funct. Anal. 2 (1992) 341–379

[58] **Y B Vorobets**, *Plane structures and billiards in rational polygons: the Veech alternative*, Russ. Math. Surv. 51 (1996) 779–817

[59] **B Wajnryb**, *A simple presentation for the mapping class group of an orientable surface*, Israel J. Math. 45 (1983) 157–174

[60] **B Wajnryb**, *Artin groups and geometric monodromy*, Invent. Math. 138 (1999) 563–571

[61] **A N Zemljakov**, **A B Katok**, *Topological transitivity of billiards in polygons*, Matem. Zametki 18 (1975) 291–300