THE ROYAL
SWEDISH
ACADEMY OF
SCIENCES

INSTITUT
MITTAG-LEFFLER

Auravägen 17, SE-182 60 Djursholm, Sweden
Tel. +46 8 622 05 60 Fax. +46 8 622 05 89
info@mittag-leffler.se www.mittag-leffler.se

IML Workshop on Growth and Mahler Measures in Geometry and Topology

REPORT No. 1, 2013/2013, summer4

ISSN 1103-467X
ISRN IML-R- -1-13/13- -SE

Growth and Mahler Measures in Geometry and Topology

Institut Mittag-Leffler Workshop, July 1 – 5, 2013

Preprint Collection

Support by the Institut Mittag-Leffler (Djursholm, Sweden) is gratefully acknowledged.

IML Workshop: "Growth and Mahler measures in geometry and topology"

Abstract: The goal of the workshop was to bring together geometers, topologists and number theorists with the aim of exploring connections between low-dimensional geometry and topology and number theory. In the focus were topics such as growth of fundamental groups of low-dimensional space forms, cohomological torsion growth, entropy of surface automorphisms, geometry of Teichmüller space, Mahler measures of A-polynomials and volumes of hyperbolic knots and manifolds. Closely related are developments in quantum topology and quantum knot invariants.

Key words: Growth, Mahler measure; hyperbolic manifolds; mapping class theory; Teichmüller theory, knot polynomials; Betti numbers, hyperbolic volume; Salem numbers, Lehmer's problem

Organisers:

Eriko HIRONAKA	Ruth KELLERHALS
Florida State University	University of Fribourg
hironaka@math.fsu.edu	ruth.kellerhals@unifr.ch

Workshop Participants:

Mikhael Belolipetsky	IMPA Rio de Janeiro, Brazil
Michelle Bucher-Karlsson	Université de Genève, Switzerland
Abhijit Champanerkar	College of Staten Island, NY, USA
Pierre Dehornoy	ENS de Lyon, France
Vincent Emery	Stanford Univerisity, CA, USA
Yohei Komori	Waseda University, Tokyo, Japan
Thang Le	Georgia Tech. University, GA, USA
Chris Leininger	University of Illinois, Urbana-Champaign, IL, USA
Alex Lubotzky	Einstein Institute Jerusalem, Israel
Wolfgang Lück	Hausdorff-Institut, Bonn, Germany
Werner Müller	Universität Bonn, Bonn, Germany
Kate Petersen	Florida State University, FL, USA
Jean Raimbault	Université Paris VI-Jussieu, France
Peter Shalen	University of Illinois Chicago, IL, USA
Dan Silver	University of South Alabama, AL, USA
Christopher Sinclair	University of Oregon, Eugene, OR, USA
Chris Smyth	University of Edinburgh, UK
Susan Williams	University of South Alabama, AL, USA

Contents

SMALL DILATATION HOMEOMORPHISMS AS MONODROMIES OF LORENZ KNOTS

PIERRE DEHORNOY

Abstract. We exhibit low-dilatation families of surface homeomorphisms among monodromies of Lorenz knots.

Pseudo-Anosov homeomorphisms are topological/dynamical objects that can be seen as geometric counterparts of non-cyclotomic irreducible polynomials. In this dictionnary, Mahler measure becomes what is called *geometrical dilatation*. A natural task is then to exhibit (or better, to classify) homeomorphisms with low dilatation. There exist several constructions of such low-dilatation families (see the census [Hir11]): for example using fibered faces of the Thurston norm ball [McM00], or using mixed-sign Coxeter diagrams [Hir12]. The goal of this note is to exhibit an additional construction, that comes from *Lorenz knots*, that is, periodic orbits of the Lorenz vector field.

This text contains few new results. Most of the content comes from the article [Deh14] where we investigated the homological dilatation of Lorenz knots. The interest here is (*i*) to restrict our attention to subfamilies of Lorenz knots for which a stronger statement can be obtained with much less technical efforts, (*ii*) to notice that what was proven for homological dilatation in [Deh14] can also be proven for geometrical dilatation.

1. Introduction

It is known since Thurston [FLP79, Thu88] that every homeomorphism of a surface is isotopic to either a periodic homeomorphism, or to a pseudo-Anosov one, or to a reducible one. A *pseudo-Anosov homeomorphism* of a surface S is a homeomorphism h such that S admits two transverse measured foliations, called *stable* and *unstable* and usually denoted by $(\mathcal{F}^s, \mu^s)$, and $(\mathcal{F}^u, \mu^u)$, that are invariant under h, and such that there exists a positive real $\lambda(h)$, called the *geometrical dilatation* of h, such that μ^s and μ^u are uniformly multiplied by $\lambda(h)^{-1}$ and $\lambda(h)$ under h respectively. Another property of h is that all closed curves on S are stretched at speed $\lambda(h)$: for any auxiliary metric on S and for any closed curve γ on S, we have $\lim_{n\to\infty} \log(\|h^n(\gamma)\|)/n = \log(\lambda(h))$. The most standard example is given by the action of a hyperbolic matrix in $\mathrm{SL}_2(\mathbb{Z})$ on the torus $\mathbb{R}^2/\mathbb{Z}^2$. In this case the invariant foliations are given by the two eigendirections of the matrix and the dilatation is the largest eigenvalue. On surfaces of higher genus, the foliations have prong-type singularities (see Figure 1.1). A reducible homeomorphism is one that admits invariant curves. Such curves divide the surface into elementary pieces where the dynamics is either periodic or of pseudo-Anosov type.

This decomposition result of Thurston can be compared with the fact that every polynomial is either cycloctomic, or has positive Mahler measure, or is reducible. The dilatation factor for a pseudo-Anosov

---

Date: February 8, 2014.

Thanks to Eko Hironaka and Ruth Kellerhals for organizing the conference *Growth and Mahler measures in geometry and topology* in July 2013 and to the Institut Mittag-Leffler for hosting this conference.
Partially supported by SNF project 137548 *Knots and surfaces*.

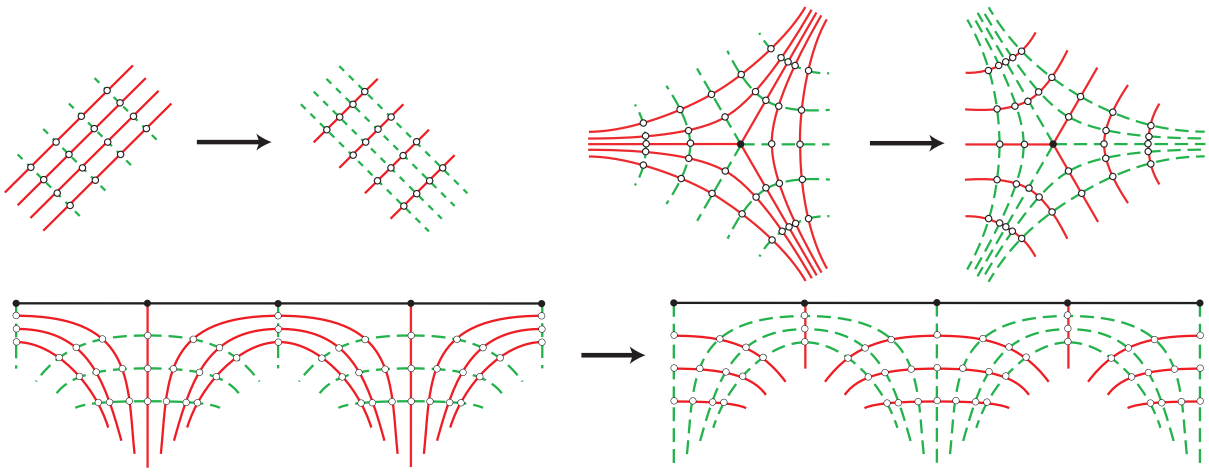

Figure 1.1. The stable and unstable foliations of a pseudo-Anosov homeomorphism (in red and green respectively). On the top left around a regular point, on the top right around a 3-prong singularity, and on the bottom around a boundary component with several singularities.

homeomorphism, or better its normalized version $\lambda(h)^{|\chi(S)|}$, is a natural counterpart of the Mahler measure. In particular, on a given surface, it is easy to find homeomorphisms with arbitrarily high dilatation (for example by iterating a fixed homeomorphism), but low-dilatation homeomorphisms are harder to construct. For $g \geq 1$ and $s \geq 0$, one usually defines $\delta_{g,s}$ as the infimum of the dilatation of a pseudo-Anosov homeomorphism on a closed surface of genus g with s boundary components. It is *a priori* not clear whether $\delta_{g,s}$ is 1 or larger, and in the latter case whether it is a minimum or not. D. Penner showed [Pen91] that $\delta_{g,s}$ is actually a minimum larger than 1, and that there exist two positive constansts c_1, c_2 such that for all g, one has $\frac{c_1}{g} \leq \log \delta_{g,0} \leq \frac{c_2}{g}$ (similar results hold for $s \geq 1$). The optimal value of c_1 is not known, and the only known values of $\delta_{g,0}$ are $\delta_{1,0}$ and $\delta_{2,0}$ [CH08]. It also follows from the work of Penner that, for any positive D and for a fixed surface, only finitely many mapping classes have a dilatation smaller than D. It is then natural to study what happens when the genus tends to infinity.

A family $(h_n, S_n)_{n\in\mathbb{N}}$, where S_n is a closed orientable surface and h_n a pseudo-Anosov homeomorphism of S_n, is said to be of *low-dilatation* if the sequence $\log(\lambda(h_n))|\chi(S_n)|$ is bounded.

Low-dilatation families are well understood in the context of 3-manifolds, where a theorem of B. Farb, C. Leininger and D. Margalit [FLM11] states that the punctured suspensions of a low-dilatation family live in some fibered faces of the Thurston norm ball of a *finite* number of 3-manifolds. However it is still unknown how different homeomorphisms having the same suspensions are related.

Question 1.1 (Farb-Leininger-Margalit [FLM11])**.** Given a positive number D, what can be said about the dynamics of those homeomorphisms with normalized dilatation smaller than D (*i.e.*, those satisfying $\log(\lambda(h))|\chi(S)| \leq \log D$)? Are they all obtained by some stabilization of the elements of a finite list?

For example, E. Hironaka showed [Hir07] that the polynomials of smallest Mahler measure in degrees 2, 4, 6, 8, and 10 all arise as dilatations of monodromies of fibered links obtained by Hopf or trefoil plumbings on some torus links, so that the monodromies are small perturbations of some periodic surface homeomorphisms.

What we do here is to exhibit low-dilatation families of dynamical origin, by considering certain subfamilies of the set of Lorenz knots, that is, periodic orbits of the (geometric model of the) Lorenz flow. These knots are fibered, so that they give rise to surface homeomorphisms, most of which are of pseudo-Anosov type. Our statement here is a variant of Theorem A of [Deh14], where we restrict our attention to subfamilies of Lorenz knots for which we obtain better bounds on the dilatation. Denote by $\mathcal{L}orenz_{b,k}$ the set of Lorenz knots described by a hanging Young diagram (see later) made of a rectangle of width b at the bottom of which is attached a diagram with at most k cells (see Figure 1.2).

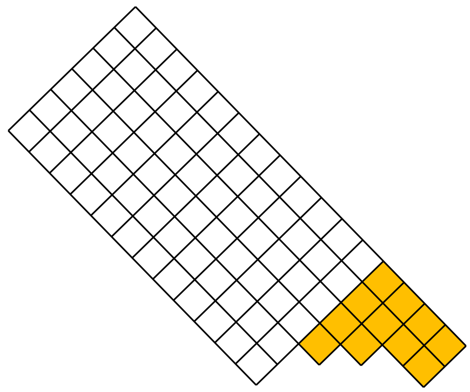

FIGURE 1.2. A diagram coding a knot in $\mathcal{L}orenz_{b,k}$, for $b = 6$ and $k = 11$. The *mixing zone* (in yellow) corresponds to those additional k cells.

Theorem 1.2. *The dilatation λ of the monodromy knot in $\mathcal{L}orenz_{b,k}$ of Euler characteristics χ satisfies* $\log(\lambda) \leq \frac{b \log k}{|\chi| - k}$. *In particular, for all b and k, the monodromies of the elements of $\mathcal{L}orenz_{b,k}$ form a low-dilatation family.*

For these families of Lorenz knots, Question 1.1 has a positive answer: the monodromies act like periodic homeomorphisms on a huge part of the surface (corresponding to the rectangular part of the associated Young diagram), and the non-periodicity is concentrated in a part of the surface of bounded size (corresponding to the additional k cells). Indeed, the rectangular part of the diagram corresponds exactly to a torus link, which is known to have periodic monodromy.

2. LORENZ KNOTS AS ITERATED MURASUGI SUMS

Lorenz knots are defined as periodic orbits of the (geometric) Lorenz flow. They have been introduced and first studied by J. Birman and R. Williams [BW81]. We refer to the original article or to [Deh11] for more details. Let us just mention that Lorenz knots form a family that contains all torus knots and is stable under cabling, so that it also contains all algebraic knots. Also, Lorenz knots are fibered, so that to each of them is canonically associated its monodromy, a homeomorphism of the genus-minimizing spanning surface. As Lorenz knots can be considered as perturbations of torus knots, it is natural to investigate the dilatation of the monodromies of those Lorenz knots which are hyperbolic.

2.a. **Young diagrams, Lorenz knots, and canonical spanning surfaces.** There are several ways of enumerating Lorenz knots and links. The most convenient from our point of view is using Young diagrams (introduced in this context in [Deh11]). The procedure is shown on Figure 2.3.

Starting from a Young diagram D, one puts its bottom-left corner on top (we call this *hanging position*). Then, by desingularizing evering intersection point into a positive braid crossing, one associates

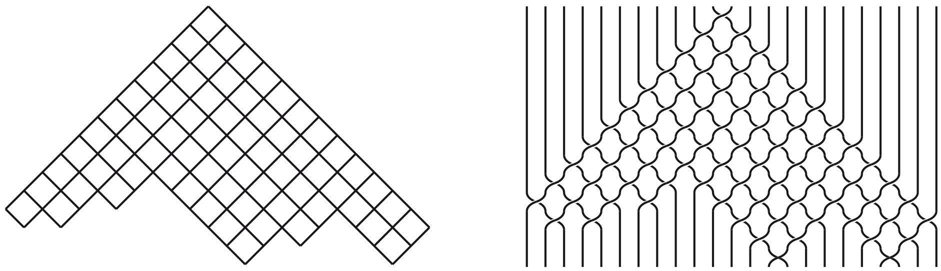

Figure 2.3. To every hanging Young diagram D (on the left), one associates a braid $\beta(D)$ (on the right) whose closure is the Lorenz link $K(D)$.

a braid, called a *Lorenz braid* and denoted by $\beta(D)$. Its closure forms a Lorenz link, that we denote by $K(D)$. All Lorenz links can be obtained in this way.

Now, to the closure of every braid is associated a canonical spanning surface, obtained by gluing a disc behind every strand and a ribbon at every crossing. Applying this construction to $\beta(D)$ yields a canonical spanning surface for $K(D)$, that we denote by $S(D)$. One can check that the Euler characteristics of $S(D)$ is the number of cells of D, hence denoted by $\chi(D)$.

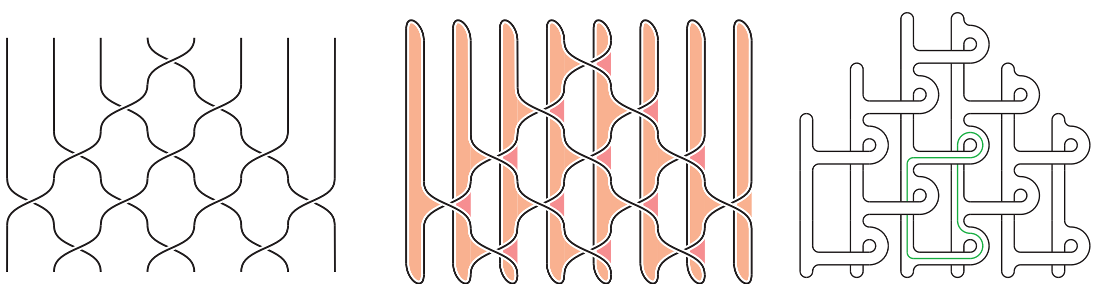

Figure 2.4. To every Lorenz braid $\beta(D)$ (on the left), one associates a link $K(D)$ and a canonical spanning surface $S(D)$ (in the middle). This surface can actually be immersed into the plane (following S. Baader [BD13], on the right). In this representation, the correspondance between elementary curves (in green) and cells of the diagram is straightforward.

2.b. **Monodromy.** In this section, we describe an inductive construction of the surface $S(K)$ for every Lorenz knot K, called the Murasugi sum. This procedure ensures that K is a fibered knot with fiber $S(K)$, and yields a decomposition of the associated monodromy $h(K)$ as an explicit product of Dehn twists.

By construction, for every cell c of a Young diagram D a simple close curve on $S(D)$ that winds once around c is canonically associated. We call it a *elementary curve* and denote it by $\gamma(c)$ (see Figure 2.4 right).

Proposition 2.1. *Let D be a Young diagram. Then the Lorenz link $K(D)$ is fibered with fiber $S(D)$, and its monodromy $h(D)$ is the product of all Dehn twists around all elementary curves of $S(D)$, in the order prescribed on Figure 2.5.*

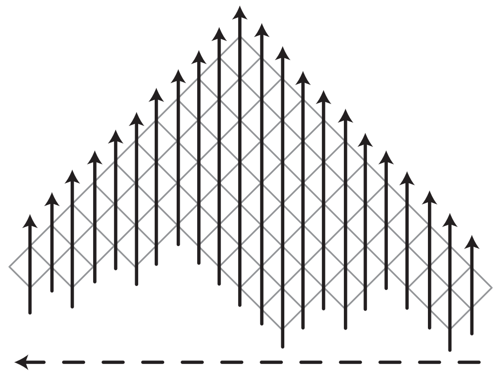

FIGURE 2.5. Decomposition of the monodromy $h(D)$ as the product of the Dehn twists around the $\chi(D)$ elementary curves on $S(D)$ (that is, those curves that turns once around the cells of D). The order is from right to left, and in every column from bottom to top.

Note that if $K(D)$ is a multi-component link, the fiber surface may not be unique, as well as the monodromy. However, if $K(D)$ is a knot, we have uniqueness of the fiber surface and of the monodromy homeomorphism.

We will only sketch the proof of Proposition 2.1 and refer to the survey [Deh11] for more details. The starting point is the 2-component Hopf link, which is the Lorenz link associated to the Young diagram with one cell only. The 2-component Hopf link is known to be fibered, the fiber surface being a twisted annulus that we call the *Hopf annulus*, and the monodromy being a right-handed Dehn twist.

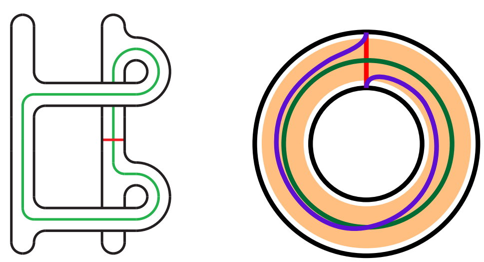

FIGURE 2.6. A Hopf annulus in $\mathbb{S}^3$(on the left), with the associated elementary curve (in green). The action of the monodromy on the annulus (on the right, seen on an abstract annulus) is a Dehn twist on the green curve: the red segment is sent to the purple one.

The induction step for proving Proposition 2.1 is done using the *Murasugi sum* of surfaces [Mur58]. This is an operation that takes two surfaces with boundary S_1, S_2, depends on a choice of a $2n$-gon in each of them, and associates a new surface with boundary $S_1 \sharp S_2$ that contains S_1 and S_2 as subsurfaces (see Figure 2.7). This operation preserves the fibered character, in the following sense: if S_1, S_2 are two fibered surfaces in $\mathbb{S}^3$ with monodromies h_1, h_2, then the Murasugi sum $S_1 \sharp S_2$, where S_1 is glued on top, is fibered with monodromy $h_1 \circ h_2$ (see the proof of D. Gabai [Gab83] or an expanded version in [Deh11]).

In particular, Murasugi gluing a Hopf annulus to a fibered surface yields another fibered surface. In this way, starting from the canonical Seifert surface associated to a hanging Young diagram D, we obtain that the surface associated to the diagram D' obtained from D by adding a cell on the bottom-right border

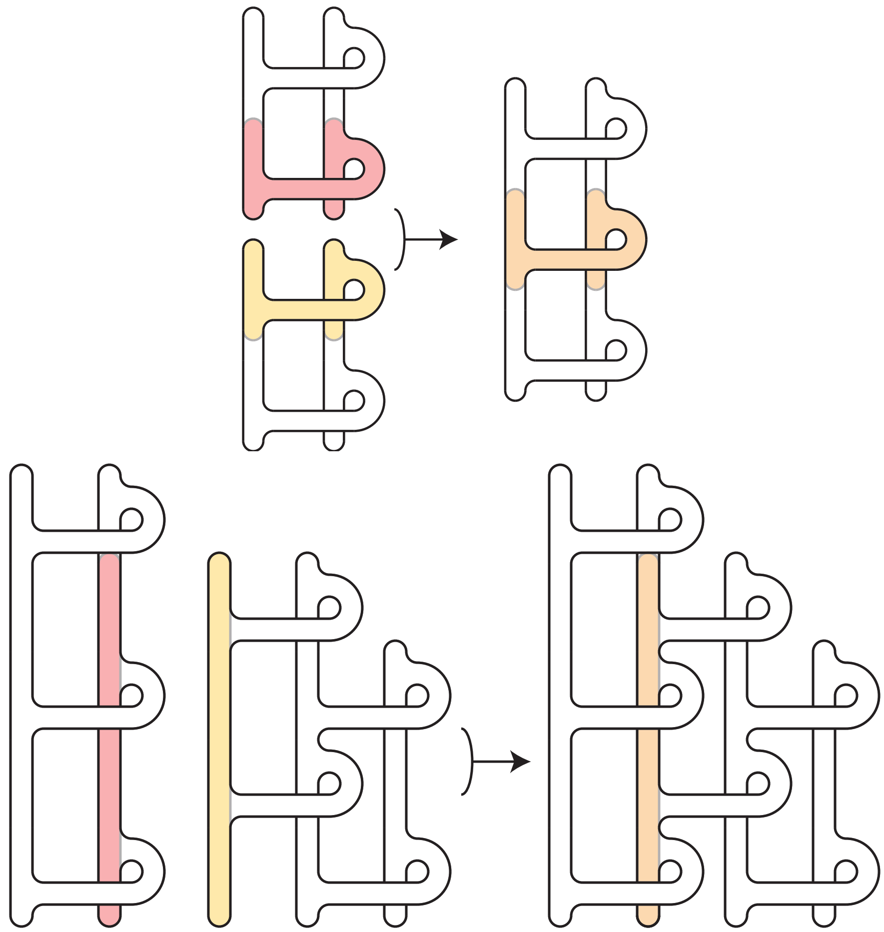

FIGURE 2.7. Two examples of Murasugi sums of canonical surfaces associated to positive braids. Observe that in the second example the left summand is glued on top.

of D is also fibered. Moreover, the monodromy associated to a Hopf annulus is a right-handed Dehn twist so that the monodromy associated to the surface $S(D)$ is a product of Dehn twists along the cores of the glued Hopf annuli. The order of the product is determined by the order of the gluing. The latter needs to preserve the respective positions of the Hopf annuli, namely one should glue first an annulus that is on top of another one. The order given on Figure 2.5 obeys this constaint. This completes the (sketch of) proof of Proposition 2.1.

2.c. **Action of the monodromy on elementary curves.** The dilatation of a pseudo-Anosov homeomorphism can be read on its action on curves. So, in order to bound the dilatation, one should bound the stretching of curves under the homeomorphism. Cutting the canonical surface $S(D)$ associated to a diagram D along all elementary curves reduces $S(D)$ to a neighborhood of its boundary, so that elementary curves contain all the information on $S(D)$. In particular for Lorenz knots, it is enough to estimate the stretching of elementary curves under $h(D)$ is order to control the dilatation of $h(D)$.

Now come the two key observations. For some orientation reason, the second observation works only when considering $h^{-1}(D)$ instead of $h(D)$. Therefore we consider the inverse of the monodromy, which makes little difference.

We say that a cell c of a hanging Young diagram is *internal* if there is a cell, say c', in North-West position with respect to c (see Figure 2.8). Otherwise it is called *external*.

Lemma 2.2. *Assume that D is a Young diagram, $L(D)$ is the associated Lorenz link, $S(D)$ is the canonical Seifert surface for $L(D)$, and $h(D)$ is the associated monodromy. Let d be an internal cell of D and a be the cell in NW position with respect to c. Then we have $h(D)^{-1}(\gamma(d)) = \gamma(a)$.*

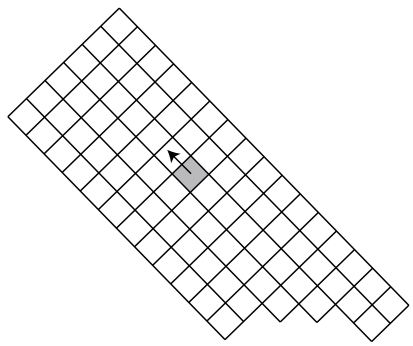

FIGURE 2.8. An internal cell and its image under the (inverse of the) monodromy.

The proof is displayed on Figure 2.9, where the successive images of the curve $\gamma(d)$ under consecutive Dehn twists are depicted (see also [BD13]).

In order to fully control $h(D)^{-1}$, we need to know what happens to external cells when iterating (backwards) the monodromy. For a general Lorenz link, the behaviour is rather hard to control (this is the reason of the heavy computations in [Deh14]). However, if we suppose that the diagram we are considering lies in $\mathcal{L}orenz_{b,k}$, things become simpler. In particular the image of an elementary curve corresponding to an external cell is not so simple, but its second image is.

For D a diagram in $\mathcal{L}orenz_{b,k}$, we call *mixing zone* of D the set of those k cells that are outside the main rectangle of D (see Figure 1.2). We also assume that we have an auxiliary metric on S for which all elementary curves have length at most 1.

Lemma 2.3. *Assume that D is a Young diagram in $\mathcal{L}orenz_{b,k}$, and that $h(D)$ is the monodromy associated to the canonical surface $S(D)$. Let c be an external cell of D. Then $h(D)^{-2}(\gamma(c))$ is a curve of length at most k that lies entirely in the mixing zone.*

The proof is depicted on Figure 2.10. The idea is that, with arguments similar to the proof of Lemma 2.2, one can describe the curve $h(D)^{-1}(\gamma(c))$: it is the concatenation of one external curve, and many internal curves. When iterating $h(D)^{-1}$ once more, the different contributions cancel, except in the mixing zone.

2.d. **Proof of Theorem 1.2.** Assume that D is a Young diagram in $\mathcal{L}orenz_{b,k}$, that $S(D)$ is the associated canonical surface, and that $h(D)$ is the corresponding monodromy. Denote by $l(D)$ the length of the long rectangle in D (that is, the complement of the mixing zone). We also take an auxiliary metric on $S(D)$ for which all elementary curves have length at most 1.

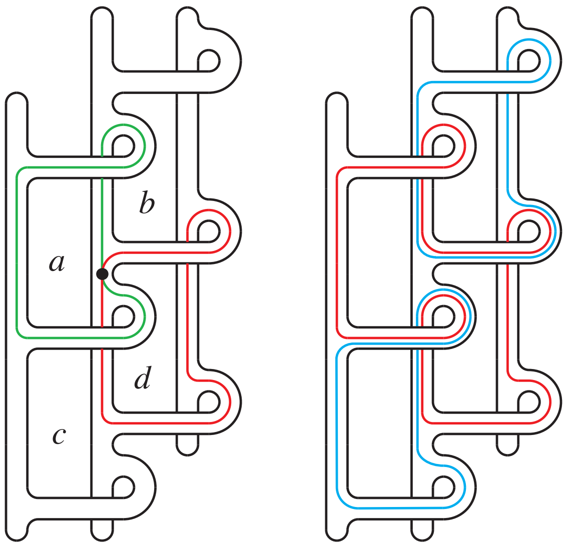

FIGURE 2.9. Proof of Lemma 2.2. The image of the elementary curve $\gamma(d)$ associated to an internal cell d under the inverse of the monodromy $h(D)^{-1}$: first the Dehn twists associated to cells that are distant from d do not modify $\gamma(d)$. Then it is changed by the Dehn twist around a into a curve that encircles both a and d. This curve is then unchanged (in particular it does not intersect the two blue curves $\gamma(b)$ and $\gamma(c)$ on the second picture). Finally it is changed by the Dehn twist around d into $\gamma(a)$. Subsequent twists to not modify it any more.

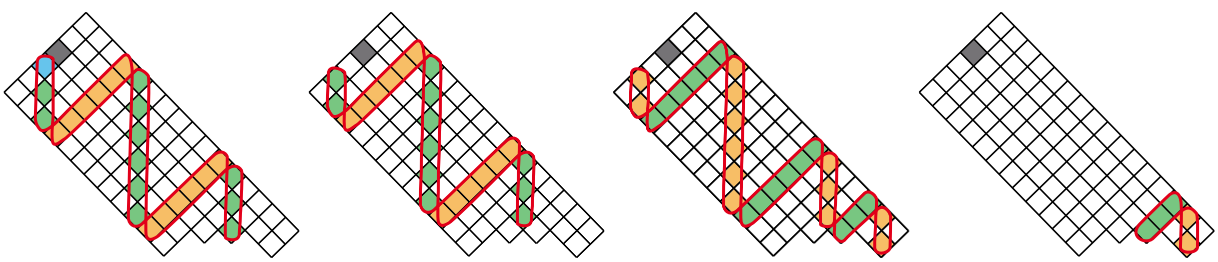

FIGURE 2.10. Proof of Lemma 2.3: an external cell c and its image under $h(D)^{-2}$. On the left, $h(D)^{-1}(\gamma(c))$ is a curve that turns positively around blocks of orange cells and negatively around blocks of green (and blue) cells. On the center left, the image under $h(D)^{-1}$ of the curve $h(D)^{-1}(\gamma(c))$, except the part that winds around the blue cell. On the center right, the image under $h(D)^{-1}$ of the part of the curve $h(D)^{-1}(\gamma(c))$ that winds around the blue cell. On the right, the concatenation of those two parts is the curve $h(D)^{-2}(\gamma(c))$, it is a curve that only winds around some cells of the mixing zone of D.

Let c be an arbitrary cell in the mixing zone on $S(D)$. By Lemma 2.2, the $l(D)$ first images of the curve $\gamma(c)$ under $h(D)^{-1}$ all correspond to internal cells, hence have length one. After a few more iteration, the image is then an elementary external curve, and after two more iterations, it is a curve in the mixing zone of length at most k. Then the process goes on: the next l iterations yield a curve of length at most k. Summarizing, the length of $h(D)^{-n}(\gamma(c))$ grows by a factor at most k every $l(D)$ steps. Therefore, the growth rate of $\gamma(c)$ is bounded by $\log k/l(D)$.

Now, the same argument works for any cell, not just in the mixing zone, except that the initial dilatation arises earlier. But this does not change the growth rate, hence $h(D)^{-1}$ asymptotically streches all curves on $S(D)$ by a factor at most $\log k/l(D)$.

Finally, an elementary computation shows that the Euler characteristics of $S(D)$ is $b\cdot l(D)+k$, therefore the dilatation is smaller than $b\log k/(|\chi(D)| - k)$.

References

[BD13] Sebastian Baader, Pierre Dehornoy, Trefoil plumbing, preprint arXiv:1308.5866.

[BW81] Joan Birman, Robert Williams, Knotted Periodic Orbits in Dynamical Systems I: Lorenz System, in S. Lomonaco Jr. ed., Low Dimensional Topology, *Contemp. Math.* **20** (1981), 1–60.

[CH08] Jin-Hwan Cho, Ji-Young Ham, The minimal dilatation of a genus-two surface, *Experiment. Math.* **17** (2008), 257–267.

[Deh11] Pierre Dehornoy, Les nœuds de Lorenz, *Enseign. Math. (2)* **57** (2011), 211–280.

[Deh14] Pierre Dehornoy, On the zeroes of the Alexander polynomial of a Lorenz knot, to appear in *Ann. Inst. Fourier*, arXiv:1110.4178.

[FLM11] Benson Farb, Chris Leininger, and Dan Margalit, Small dilatation pseudo-Anosov homeomorphisms and 3-manifolds, *Adv. Math.* **228** (2011), 1466–1502.

[FLP79] Albert Fathi, François Laudenbach, Valentin Poenaru, Travaux de Thurston sur les surface, *Astérisque* **66-67** (1979).

[Gab83] David Gabai, The Murasugi sum is a natural geometric operation, *Contemp. Math.* **20** (1983), 131–143. (2004), 243–257.

[HK06] Eriko Hironaka and Eiko Kin, A family of pseudo-Anosov braids with small dilatation, *Alg. Geom. Top.* **6** (2006), 699–738.

[Hir06] Eriko Hironaka, Salem–Boyd sequences and Hopf plumbing, *Osaka J. Math.* **43** (2006), 497–516.

[Hir07] Eriko Hironaka, On hyperbolic perturbations of algebraic links and small Mahler measure, in Singularities in Geometry and Topology (Sapporo, 2004), *Adv. Stud. Pure Math.* **46** (2007), 77–94.

[Hir10] Eriko Hironaka, Small dilatation pseudo-Anosov mapping classes coming from the simplest hyperbolic braid, *Alg. and Geom. Top.* **10** (2010), 2041–2060.

[Hir11] Eriko Hironaka, Mapping classes associated to mixed-sign Coxeter graphs, preprint arXiv:1110.1013

[Hir12] Eriko Hironaka, Small dilatation pseudo-Anosov mapping classes, in Intelligence of low-dimensional topology *RIMS* **1812** (2012), 25–33.

[LT11] Erwan Lanneau, Jean-Luc Thiffeault, On the minimum dilatation of pseudo-Anosov homeomorphisms on surfaces of small genus, *Ann. Inst. Fourier* **61** (2011), 105–144.

[Lei04] Chris Leininger, On groups generated by two positive multi-twists: Teichmller curves and Lehmer's number, *Geom. Topol.* **8** (2004), 1301–1359.

[McM00] Curtis T. McMullen, Polynomial invariants for fibered 3-manifolds and Teichmller geodesics for foliations, *Ann. Sci. Éc. Norm . Supér. (4)* **33** (2000), 510–560.

[Mur58] Kunio Murasugi, On the genus of the alternating knot, I, *J. Math. Soc. Japan* **10** (1958), 94–105.

[Pen91] Robert C. Penner, Bounds on least dilatations, *Proc. Amer. Math. Soc.* **113** (1991), 443–450.

[Thu88] William P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces, *Bull. Amer. Math. Soc.* **19** (1988), 417–431.

Institut Fourier, Université Joseph Fourier-Grenoble 1, 100 rue des mathématiques, BP 74, 38402 Saint-Martin-d'Hères, France

E-mail address: `pierre.dehornoy@ujf-grenoble.fr`

BOUNDS FOR TORSION HOMOLOGY OF ARITHMETIC GROUPS

VINCENT EMERY

ABSTRACT. This note presents as a survey some ideas used in my paper "Torsion homology of arithmetic lattices and K_2 of imaginary fields". It is essentially based on my talk at the conference "Growth and Mahler measures in geometry and topology", Djursholm, July 2013.

1. INTRODUCTION

Let $X = G/K$ be a symmetric space of noncompact type, i.e., G is a semisimple Lie group without compact or Euclidean factors and $K \subset G$ is a maximal compact subgroup. For example, the case $G = \mathrm{PSL}_2(\mathbb{C})$ corresponds to $X = \mathbf{H}^3$, the hyperbolic space of dimension 3. A *lattice* $\Gamma \subset G$ is a discrete subgroup of finite covolume in G (with respect to any Haar measure on G). It defines an *orbifold* $\Gamma\backslash X$, which has finite volume with respect to the G-invariant Riemannian metric on X. We say that Γ is *uniform* if the quotient $\Gamma\backslash X$ is compact. If Γ is torsion-free then the quotient $M = \Gamma\backslash X$ is a manifold, locally isometric to X. By Selberg's lemma, any quotient orbifold of finite volume is covered with finite degree by a manifold. The study of lattices of G (or orbifolds) can be reduced to *irreducible* ones, i.e., those which are not commensurable with a product of two subgroups that are lattices in factors of G. If G is simple, then trivially all lattices are irreducible.

According to a theorem of Gromov, the Betti numbers of the locally symmetric spaces $M = \Gamma\backslash X$ can be bounded linearly in the volume $\mathrm{vol}(M)$. This shows that the complexity of the topology of the manifold M is controlled by a simple geometric invariant, the volume. Gromov's theorem is valid in a more generic context than for symmetric spaces, the relevant condition being the negative curvature. Moreover, it has been recently extended to the case of orbifolds by Samet [6]. For a fixed semisimple Lie group G as above (of noncompact type, without compact factors and without center), the result can be formulated as the following. Here $b_j(\Gamma)$ denotes the rank of the group homology $H_j(\Gamma)$ (with coefficient in $\mathbb{Z}$) of Γ.

Theorem 1 (Samet)**.** *There exists a constant C_G such that for any irreducible lattice $\Gamma \subset G$ we have*

$$b_j(\Gamma) \leq C_G \,\mathrm{vol}(\Gamma\backslash X),$$

for any j.

A natural question is to ask whether a similar result can be proved for the torsion part of the homology $H_j(\Gamma)$. A motivation is the growing interest in torsion homology of arithmetic lattices due to connection with number theory; see [2, 3]. An *arithmetic group* is a group of the form (or more precisely commensurable to) $H(\mathbb{Z})$ where H is an algebraic $\mathbb{Q}$-group. By a theorem of Borel and Harish-Chandra, such a group is always a lattice in $H(\mathbb{R})$ provided the latter is semisimple. Moreover, Margulis proved as a consequence of its superrigity theorem that every irreducible lattice in a semisimple Lie group of real rank at least 2 is arithmetic.

This note presents a short introduction to some ideas used in [4] to obtain upper bounds on the torsion homology of arithmetic lattices; here we do not discuss the aspects related to K_2 though. It is essentially based on my talk at the conference "Growth and Mahler measures in geometry and topology", Djursholm, July 2013. I would like to thank again both organizers, Eriko Hironaka and Ruth Kellerhals, for having invited me to this very nice workshop. I also thank the staff of the Mittag-Leffler institute for the perfect organization.

2. Consequences of Lehmer's conjecture

Lehmer's conjecture asserts that the Mahler measure of a noncyclotomic polynomial with integral coefficients is bounded away from 1. A well-known consequence of this conjecture is the following (cf. [5, Section 10]).

Conjecture 2 (Short geodesic conjecture)**.** *Let X be a symmetric space of noncompact type. There exists $\epsilon > 0$ such that the length of any closed geodesic on an arithmetic locally symmetric space $M = \Gamma\backslash X$ is greater than ϵ.*

Let us restrict ourselves to compact manifolds $M = \Gamma\backslash X$. Assuming Conjecture 2 we can embed around any $x \in M$ a geodesic ball $B_r(x)$ of radius $r = \epsilon/2$. One needs about $\mathrm{vol}(M)/\mathrm{vol}(B_r(x))$ balls to cover all M. Using the notion of *nerve* of an open covering like this, one can show that M is homotop to simplicial complex whose size (i.e, the total number of simplices) is bounded linearly in $\mathrm{vol}(M)$. Thus, the following statement is a consequence of Conjecture 2.

Conjecture 3. *There exists a constant β_X such that any compact arithmetic manifold $M = \Gamma\backslash X$ is homotop to a simplicial complex of size bounded above by $\beta_X \mathrm{vol}(M)$.*

A statement like this is exactly what we need to bound the torsion homology of M. In general, in a complex of abelian groups of bounded ranks the torsion homology does not need to be bounded (take $\mathbb{Z} \xrightarrow{N} \mathbb{Z}$ with N arbitrarily large). However, in a simplicial complex the "boundary maps" are concretely given by the boundary of simplices. This observation can be used to bound the torsion homology. More precisely, we can use the following result of Gabber to do so. For a proof see [7, §2.1]. For an abelian group A, we denote by A_{tors} its subgroup of torsion elements.

Lemma 4 (Gabber)**.** *Let $A = \mathbb{Z}^a$ with the standard basis $(e_i)_{i=1,\dots,a}$ and $B = \mathbb{Z}^b$, so that $B \otimes \mathbb{R}$ is equipped with the standard Euclidean norm $\|\cdot\|$. Let $\phi : A \to B$ be a $\mathbb{Z}$-linear map such that $\|\phi(e_i)\| \leq \alpha$ for each $i = 1,\dots,a$. If we denote by Q the cokernel of ϕ, then*

$$|Q_{\mathrm{tors}}| \leq \alpha^{\min\{a,b\}}\,.$$

Thus, applying Lemma 4 to Conjecture 3 we would obtain a bound, for any j,

$$\log |H_j(M)_{\mathrm{tors}}| \leq C_X \,\mathrm{vol}(M) \tag{2.1}$$

for some constant C_X.

Remark 5. For any $n \geq 2$, Conjecture 2 fails for *nonarithmetic* hyperbolic n-manifolds (see [1]). Moreover, Conjecture 3 fails for nonarithmetic hyperbolic 3-manifolds: the statement in the conjecture implies the finiteness of the number of manifolds of bounded volume, which is known to fail by Thurston-Jørgensen description of the volume spectrum of 3-manifolds. Note that restricted to the class of arithmetic manifolds (or lattices) this finiteness property holds for any symmetric space X, by a theorem of Borel and Prasad.

3. The case of noncompact manifolds

The upper bound for torsion homology (2.1) that would follow from Lehmer's conjecture in the case of compact arithmetic manifolds can actually be proved unconditionally for *noncompact* arithmetic manifolds. This is due to the classification of nonuniform arithmetic lattices, described in the next proposition (cf. [5, Lemma 5.2]).

Proposition 6. *Let G be a semisimple real Lie group as in Section 1. If $\Gamma \subset G$ is an irreducible nonuniform arithmetic subgroup, then it can be written as a subgroup $\Gamma \subset H(\mathbb{Z})$ of finite index, where H is an algebraic $\mathbb{Q}$-group of same (real) dimension as G.*

Example 7. Every nonuniform arithmetic lattice in $\mathrm{PSL}_2(\mathbb{C})$ is commensurable to a Bianchi group $\mathrm{PSL}_2(\mathcal{O}_F)$ for some imaginary quadratic field F. The corresponding $\mathbb{Q}$-group is then given by Weil's restriction of scalars as $H = \mathrm{Res}_{F/\mathbb{Q}}(\mathrm{PSL}_2/F)$, which has dimension $6 = \dim(\mathrm{PSL}_2)\cdot[F:\mathbb{Q}]$. To obtain all uniform lattices instead, one needs to take into account fields F of (arbitrarily) large degrees, which will give raise to algebraic $\mathbb{Q}$-groups of larger dimensions.

As a consequence of Proposition 6, any element in a nonuniform arithmetic group $\Gamma \subset G$ can be written as a matrix whose characteristic polynomial has integral coefficients and fixed degree equal to $\dim(G)$. But Lehmer's conjecture is known to hold when the degrees of the polynomials is fixed (or bounded), and this will show that the length of closed geodesics is bounded away from 0. In other words:

Proposition 8. *Conjecture 2 holds for noncompact arithmetic manifolds $M = \Gamma\backslash X$.*

From this result, one can use the same argument to obtain a version of the statement of Conjecture 3 valid for noncompact manifolds. The problem is that the argument used there assumed that the manifolds were compact. To deal with noncompact manifolds one needs to remove some neighbourhoods of their unbounded part and check that the complexity of the topology does not become too bad where the cuts are performed. This is a difficult task, which could be done by Gelander in his thesis. He obtained the following result (see [5]):

Theorem 9 (Gelander)**.** *There exists a constant β_X such that any noncompact arithmetic manifold $M = \Gamma\backslash X$ is homotop to a simplicial complex of size bounded above by $\beta_X \mathrm{vol}(M)$.*

Together with Lemma 4 one then obtains the following.

Corollary 10. *There exists a constant C_X such that for any noncompact arithmetic manifold $M = \Gamma\backslash X$, we have for any j:*

$$\log|H_j(M)_{\mathrm{tors}}| \leq C_X \,\mathrm{vol}(M).$$

4. An extension to the case of orbifolds

For arithmetic application especially, it is useful to have a version a Corollary 10 where the arithmetic subgroups $\Gamma \subset G$ may contain torsion, i.e., so that the quotient $\Gamma \backslash X$ is in general an orbifold. Such an extension was obtained in [4] for G respecting some conditions, and it was applied to obtain upper bound for K_2 of the ring of integers of totally imaginary number fields.

Theorem 11 (Emery)**.** *Let G be a semisimple Lie group as in Section 1 and such that for any (nonuniform arithmetic) irreducible lattice $\Gamma_0 \subset G$ we have $H_q(\Gamma_0, \mathbb{Q}) = 0$ for $q = 1, \ldots, j$. Then, there exists a constant C_G such that for any nonuniform arithmetic irreducible lattice $\Gamma \subset G$ we have:*

$$\log |H_j(\Gamma)_{\mathrm{tors}}| \leq C_G \operatorname{vol}(\Gamma \backslash G).$$

Proof. Using Proposition 6 one sees that we can construct for any Γ as torsion-free normal subgroup $\Gamma_0 \subset G$ whose index is bounded by a constant depending only on $\dim(G)$. Then, the idea is to use Lyndon-Hochschild-Serre spectal sequence

$$E^2_{pq} = H_p(\Gamma/\Gamma_0, H_q(\Gamma_0)) \implies H_{p+q}(\Gamma),$$

together with Corollary 10, which gives an upper bound for $H_q(\Gamma_0)$. We refer to [4] for the details. □

For example, if G has real rank at least 2 then superrigidity implies at once vanishing of the first Betti number and arithmeticity of lattices. Thus, we get the following result for torsion homology in degree one.

Corollary 12. *Let G be a semisimple real Lie group without compact factor and of real rank at least 2. Then, there exists a constant C_G such that for any nonuniform irreducible lattice $\Gamma \subset G$ we have:*

$$\log |H_1(\Gamma)_{\mathrm{tors}}| \leq C_G \operatorname{vol}(\Gamma \backslash G).$$

References

1. Mikhail Belolipetsky and Scott Thomson, *Systoles of hyperbolic manifolds*, Algebr. Geom. Topol. **11** (2011), no. 3, 1455–1469.
2. Nicolas Bergeron and Akshay Venkatesh, *The asymptotic growth of torsion homology for arithmetic groups*, J. Inst. Math. Jussieu, to appear, preprint `arXiv:1104.1083`.
3. Frank Calegari and Akshay Venkatesh, *A torsion Jacquet-Langlands correspondence*, preprint `arXiv:1212.3847v1`.
4. Vincent Emery, *Torsion homology of arithmetic lattices and K_2 of imaginary fields*, preprint `arXiv:1303.6132`.
5. Tsachik Gelander, *Homotopy type and volume of locally symmetric manifolds*, Duke Math. J. **124** (2004), no. 3, 459–515.
6. Iddo Samet, *Betti numbers of finite volume orbifolds*, Geom. Topol. **17** (2013), no. 2, 1113–1147.
7. Christophe Soulé, *Perfect forms and the Vandiver conjecture*, J. Reine Angew. Math. **517** (1999), 209–221.

SMALL DILATATION PSEUDO-ANOSOV MAPPING CLASSES AND SHORT CIRCUITS ON TRAIN TRACK AUTOMATA

ERIKO HIRONAKA

Abstract. This note is a survey of recent results surrounding the minimum dilatation problem for pseudo-Anosov mapping classes. In particular, we give evidence for the conjecture that the minimum accumulation point of the genus normalized dilatations of pseudo-Anosov mapping classes on closed surfaces equals the square of the golden ratio. We also find explicit fat train track maps determining a sequence of pseudo-Anosov mapping classes whose normalized dilatations converge to this limit.

1. Introduction

Let S be a compact surface of genus g with b boundary components. A *mapping class* ϕ on S is a self-homeomorphism of S considered up to isotopy. The map $\phi : S \to S$ is *pseudo-Anosov* if S admits a pair of ϕ-invariant transverse measured singular foliations, called the *unstable foliation* $(\mathcal{F}^u, \nu^u)$ and *stable foliation* $(\mathcal{F}^s, \nu^s)$, so that the action of ϕ stretches ν^u by a constant $\lambda > 1$, and contracts ν^s by $\frac{1}{\lambda}$. The constant λ has the property that $\log(\lambda)$ is the minimal topological entropy of elements in the isotopy class of ϕ and is called the *dilatation* of ϕ. The theory of pseudo-Anosov mapping classes is developed in detail in [FLP], [CB] and [Thu2].

In a 1991 paper, Penner [Pen] proved that as a function of genus $g \geq 2$, the minimum dilatation δ_g for pseudo-Anosov mapping classes on closed genus g surfaces satisfies

$$\log \delta_g \asymp \frac{1}{g}. \tag{1}$$

Penner's paper has brought recent interest to the minimum dilatation problem, which asks what are the values of δ_g for $g \geq 2$, and what are the mapping classes that realize these values. So far the exact value of the minimum dilatation δ_g is known only for $g = 2$ [CH]. In this paper we give a brief survey of the minimum dilatation problem and its relations to the study of train track maps, digraphs, polynomials and algebraic integers, and give an illustrative example.

1.1. **Lehmer's problem and dilatations.** Questions surrounding the values of δ_g are closely analogous to Lehmer's problem on Mahler measures. Dilatations of pseudo-Anosov mapping classes are special algebraic integers called Perron numbers. These are real algebraic integers $\lambda > 1$ all of whose algebraic conjugates are strictly smaller in complex norm. Furthermore, dilatations have the property that λ^{-1} is also an algebraic integer, and hence λ is an algebraic unit. The Mahler measure $m(\lambda)$ of an algebraic integer λ is the absolute value of the product of its conjugates outside the unit circle. In [Leh] Lehmer asks: is there is a positive gap between 1 and the next largest Mahler measure? A negative answer would mean that the set of Mahler measures is dense in the interval $[1, \infty)$. Lehmer's question leads immediately to several others.

For each fixed degree n, any bound on Mahler measure bounds the size of the coefficients of the minimal polynomial, and hence the Mahler measures greater than one for algebraic integers of fixed degree n achieve a minimum $m_n > 1$. It is not known how m_n behaves as n goes to infinity, nor about properties of the algebraic integers achieving m_n. For example: is there a bound on the number of algebraic conjugates outside the unit circle?

The complex norm $h(\lambda)$ of the largest conjugate of an algebraic integer λ is called the *house* of λ. The *normalized house*

$$h(\lambda)^{d_{\mathrm{alg}}}$$

is the house raised to the degree of the minimal polynomial. It is not known whether this coarse upper bound for Mahler measure is bounded away from one for non-cyclotomic algebraic integers (cf. [Dob]).

1.2. **Perron numbers.** For Perron numbers, there is an alternative way to normalize house, other than algebraic degree. Each Perron number is the spectral radius of a *Perron-Frobenius matrix*: a $d \times d$ matrix M with non-negative integer entries such that for some power $k \geq 1$, M^k has strictly positive entries. The minimum such d, which is an upper bound for d_{alg}, is the degree of the characteristic polynomial of M, called the *Perron-Frobenius degree* of the Perron number. McMullen recently showed in [McM2] that for Perron units λ with Perron-Frobenius degree d_{PF}, we have

$$\lambda^{d_{\mathrm{PF}}} \geq \gamma_0^4, \tag{2}$$

where γ_0 is the golden ratio.

1.3. **Normalized dilatations.** It is an open question whether all Perron units are dilatations of pseudo-Anosov mapping classes (partial results in this direction were found by Thurston in [Thu3]). Define the *genus-normalized* dilatation to be $\lambda(\phi)^g$ and let $\ell_g = (\delta_g)^g$, the minimum genus-normalized dilatation for fixed genus g. Penner's result (1) is equivalent to the statement that there are constants c and C so that

$$1 < c \leq \ell_g \leq C.$$

It is an open problem to determine sharp bounds for c and C, or to find the limit of ℓ_g as g goes to infinity.

McMullen's result (2) on normalized Perron units is evidence for the following conjecture.

Conjecture 1.1. *The smallest accumulation point for the sequence ℓ_g is γ_0^2.*

For the pseudo-Anosov mapping classes (S_g, ϕ_g) that we later describe in this paper, the surfaces S_g have genus g, the normalized dilatations $\lambda(\phi_g)^g$ converge to γ_0^2, hence γ_0^2 is an upper bound for the smallest accumulation point. This together with McMullen's result (2) is not enough to prove the conjecture, however, since in general both d_{alg} and $2g$ can be strictly smaller than d_{PF}, and the latter can be as large as $6g - 6$ [Pen].

Conjecture 1.1 was originally inspired by a question of Lanneau and Thiffeault posed in [LT]. An orientable pseudo-Anosov mapping class is one where the stable and unstable foliations are orientable. Lanneau and Thiffeault ask whether for orientable pseudo-Anosov mapping classes on surfaces of even genus, the minimum dilatation is the largest real root of the polynomial

$$LT_n(x) = x^{2n} - x^{n+1} - x^n - x^{n-1} + 1.$$

If λ_n is the largest root of $LT_n(x)$, then it is not hard to show that $(\lambda_n)^n$ is a monotone decreasing sequence converging to γ_0^2.

1.4. **Main example.** In this paper, we explicitly define a sequence of pseudo-Anosov mapping classes whose genus normalized dilatations define a strictly monotone decreasing sequence converging to γ_0^2. The existence of such sequences was already proved in [Hir] [AD] and [KT2], but the description we give here, using the language of fat train track maps and digraphs, is the first constructive one, and serves to give a glimpse of what small dilatation mapping classes look like in general.

We show the following.

Theorem 1.2. *There is a sequence of pseudo-Anosov mapping classes (S_n, ϕ_n) described by fat train track maps $f_n : \tau_n \to \tau_n$, $n \geq 2$ with the following properties:*

(1) S_n is a closed orientable surface of genus $g = n$ if 3 doesn't divide n and genus $g = n - 1$ if 3 divides n,
(2) $\lambda(\phi_n)$ is the largest real root of $LT_n(x)$,
(3) the genus-normalized dilatations of (S_n, ϕ_n) converge to γ_0^2.
(4) (S_n, ϕ_n) is an orientable mapping class if and only if n is even,
(5) (S_n, ϕ_n) have the smallest dilatation among orientable pseudo-Anosov mapping classes of genus $g = n$ when $n = 2, 4, 8$, and of genus $g = 5$ when $n = 6$.
(6) the train track maps f_n have folding decompositions corresponding to length 3 circuits on fat train track automata, and
(7) the topological type of the digraph associated to the train track map f_n is fixed for $n \geq 2$.

Corollary 1.3. *The square of the golden mean γ_0^2 is an accumulation point for normalized dilatations of orientable pseudo-Anosov mapping classes.*

Sequences satisfying properties (1)–(5) were also found in [Hir] as mapping classes associated to a convergent sequence on a fibered face. The difference in this paper is that our description is constructive.

1.5. **Organization.** Thurston's fibered face theory [Thu1], Fried's results about cross-sections of pseudo-Anosov flows [Fri], McMullen's theory of Teichmüller polynomials [McM1] and the universal finiteness theorem of Farb, Leininger and Margalit [FLM] together imply that the problem of finding minimum dilatations reduces to understanding the roots of families of polynomials arising as specializations of a finite list of multivariable polynomials. We recall these results in Section 2. In Section 3 we describe the restriction of Lehmer's problem to Perron units, and its recent partial solution by McMullen [McM2]. The special case of orientable pseudo-Anosov mapping classes, and the Lanneau-Thiffeault question is discussed in Section 4. In Section 5 we define fat train track maps, and their automata. We also explain how to compute Both the Teichmüller and Alexander polynomials in this context. In Section 6, we describe a sequence of fat train track maps whose Teichmüller polynomial specializes to the LT polynomials, and prove Theorem 1.2.

2. Fibered faces, dilatations and polynomials

Fibered face theory gives a natural way to partition the set of pseudo-Anosov mapping classes into families that are in one-to-one correspondence with rational points on convex Euclidean polyhedra (possibly single points). Each family contains mapping classes defined on different surfaces, but having related dynamics. In particular, the normalized dilatation varies continuously with respect to the induced Euclidean metric. Furthermore, each set has an associated Teichmüller polynomial, whose specialization at each point in the set determines the dilatation of the associated mapping class.

2.1. **Fibered face theory.** In [Thu1], Thurston defines a norm $|| \; ||$ on $H^1(M;\mathbb{R})$ as follows. Given a surface $(S,\partial S) \subset (M,\partial M)$, let

$$\chi_-(S) = \sum_{S' \subset S} \max\{-\chi(S'),0\},$$

where the sum is taken over connected components S' of S. Given $\alpha \in H^1(M;\mathbb{Z})$, let

$$||\alpha|| = \min\{\chi_-(S) \;:\; (S,\partial S) \subset (M,\partial M) \text{ is Poincaré dual to } \alpha\}.$$

Then $|| \; ||$ extends to a unique norm on $H^1(M;\mathbb{R})$. Furthemore, the unit norm ball is a convex polyhedron, and the convex hull of rational vertices. The norm $|| \; ||$ is called the *Thurston norm*, and the unit ball is called the *Thurston norm ball*.

An element of $H^1(M;\mathbb{Z})$ is called *fibered* if it is dual to the fiber of a fibration $\psi_\alpha : M \to S^1$ over the circle.

Theorem 2.1 (Thurston [Thu1]). *For every open top-dimensional face F of the unit Thurston norm ball, either every integral point in the cone $F \cdot \mathbb{R}^+$ over F is fibered, or none of them are.*

If the integral points on $F \cdot \mathbb{R}^+$ are fibered, we say F is a *fibered face* and $F \cdot \mathbb{R}^+$ is a *fibered cone*.

Circle fibrations of M are in one-to-one correspondence with mapping classes (S,ϕ) with the property that M is the mapping torus of (S,ϕ):

$$M \simeq S \times [0,1]/(x,1) \sim (\phi(x),0),$$

where S is homeomorphic to the *fiber* of the fibration. The mapping class (S,ϕ) is called the *monodromy* of the fibration.

A *primitive integral element* in $H^1(M;\mathbb{Z})$ is a point with relatively prime integral coefficients. Given a fibered element $\alpha \in H^1(M;\mathbb{Z})$, any positive integer multiple $m\alpha$ has the property that $\psi_{m\alpha}$ is the composition of ψ_α with the m-fold cyclic covering of the circle to itself. If follows that primitive integral elements on fibered cones correspond to fibrations of M over the circle with connected fibers.

A key theorem of Thurston that connects the classification of mapping classes and that of fibered 3-manifolds is the following.

Theorem 2.2 (Thurston [Thu2])**.** *A mapping class is pseudo-Anosov if and only if its mapping torus is a hyperbolic 3-manifold.*

It follows that there is a one-to-one correspondence between pseudo-Anosov mapping classes (S, ϕ) on surfaces S and rational points on fibered faces of hyperbolic 3-manifolds whose denominator equals $|\chi(S)|$.

2.2. **Removing singularities.** To study the dynamical properties of a pseudo-Anosov mapping class it is natural to remove the singularities of the invariant stable and unstable foliations. This process preserves essential information about the surface (e.g., genus) and the dynamics of the mapping class (e.g., dilatation). In many cases, this process increases the first Betti number of the mapping torus, and hence the dimension of the associated fibered face.

Lemma 2.3. *Let S be a compact surface with boundary, and ϕ a pseudo-Anosov map on S. The first Betti number of the mapping torus of (S, ϕ) is $r + 1$, where r is the rank of the ϕ-invariant homology $H_1(S, \partial S; \mathbb{Z})$.*

Proof. See, for example, [McM1]. □

Define the *singularities* of a pseudo-Anosov mapping class (S, ϕ) to be the set of singularities of the stable and unstable ϕ-invariant foliations. The union of singularities on S is a finite set of points closed under the action of ϕ. Let S^0 be the complement of small neighborhoods of the singular points. There is a unique pseudo-Anosov mapping class ϕ^0 defined on S^0 determined up to isotopies that fix the boundary component pointwise. Correspondingly, there is a well-defined way to define invariant foliations for ϕ^0 whose extensions to S are the original invariant foliations of ϕ, so that certain leaves terminate at the boundary. The leaves terminating at a boundary component are called prongs, and the degree of the singularity equals the number of prongs minus 2.

By this construction, the dilatations $\lambda(\phi)$ and $\lambda(\phi^0)$ are stretching factors of the same maps on the same foliations, and hence are equal. Furthermore, (S, ϕ) can be recovered from (S^0, ϕ^0) by closing off the boundary components with disks.

Corollary 2.4. *Suppose (S, ϕ) is a pseudo-Anosov mapping class such that the number of orbits of boundary components and the number of orbits of singularities add up to at least 2. Then the first Betti number of the mapping torus of (S^0, ϕ^0) is greater than or equal to 2, and hence $(S^0, \phi^)$ corresponds to a point on a fibered face of positive dimension.*

Proof. For any mapping class ϕ on a surface with boundary S, the sum γ of loops around the orbits of a boundary component determines a ϕ^0-invariant element $[\gamma]$ in $H_1(S^0, \partial S^0; \mathbb{Z})$. If there is more than one orbit, then $[\gamma]$ is non-trivial. The rest follows from Lemma 2.3. □

2.3. **Normalized dilatations.** The *normalized dilatation* of a pseudo-Anosov mapping class (S, ϕ) is defined by

$$L(S, \phi) = \lambda(\phi)^{|\chi(S)|}.$$

Given a fibered element $\alpha \in H^1(M; \mathbb{Z})$ with monodromy (S_α, ϕ_α) define

$$\mathcal{H}(\alpha) = \log(\lambda(\phi_\alpha)).$$

When α is an integral element, $\mathcal{H}(\alpha)$ is the *topological entropy* of ϕ_α.

Theorem 2.5 (Fried [Fri], McMullen [McM1])**.** *The function $\mathcal{H}(\alpha)$ extends to a real analytic, convex function that is homogeneous of degree -1 on each fibered cone $F \cdot \mathbb{R}^+$ and goes to infinity toward the boundary of the fibered face F.*

Given a primitive integral point $\alpha \in F \cdot \mathbb{R}^+$, let $\overline{\alpha} = \alpha/q$ be its projection onto F.

Corollary 2.6. *The function on the rational points of a fibered face F that sends $\overline{\alpha}$ to $L(S_\alpha, \phi_\alpha)$ extends to a real analytic, strictly convex function on F that goes to infinity toward the boundary of F.*

Proof. By homogeneity, we have

$$\log(L(S_\alpha, \phi_\alpha)) = ||\alpha|| \log(\lambda(\phi_\alpha)) = \mathcal{H}(\overline{\alpha}).$$

□

Remark 2.7. Strict convexity of $\mathcal{H}$ and its behavior toward the boundary of F imply that this function has a unique minimum on F. The minimum, however, does not necessarily occur at a rational point, and hence it may not be realized by the monodromy of a circle fibration [Sun].

Corollary 2.8. *Any convergent sequence on the interior of a fibered face that is not eventually constant corresponds to a family of pseudo-Anosov mapping classes with unbounded Euler characteristic and bounded normalized dilatation.*

Farb, Leininger and Margalit prove the following partial converse.

Theorem 2.9 (Universal Finiteness Theorem [FLM])**.** *Let Φ be a family of pseudo-Anosov mapping classes with the property that for some constant $C > 1$, we have*

$$L(S, \phi) < C$$

for all (S, ϕ) in $\mathcal{F}$. Then there is a finite set of manifolds $\mathcal{M} = \{M_1, \ldots, M_k\}$ so that the mapping torus (S^0, ϕ^0) corresponding to each element of Φ is an element of $\mathcal{M}$.

It follows that to study the dynamics of a family of mapping classes with bounded normalized dilatation, it suffices to look at a finite collection of fibered faces of hyperbolic 3-manifolds.

2.4. **Teichmüller polynomials.** In [McM1], McMullen defined, for each fibered hyperbolic 3-manifold M, and fibered face $F \subset H^1(M;\mathbb{R})$, an element $\Theta_F \in \mathbb{Z}G$, called the *Teichmüller polynomial* where $\mathbb{Z}G$ is the group ring over $G = H_1(M;\mathbb{Z})/\text{torsion}$. Since G is a free abelian group, we can identify elements with monomials in the generators of G, and think of elements of $\mathbb{Z}G$ as polynomials in several variables with integer coefficients. Given an element $\theta \in \mathbb{Z}G$, written

$$\theta = \sum_{g \in G} a_g g,$$

and $\alpha \in H^1(M;\mathbb{Z})$, the *specialization* of θ at α is defined by

$$\theta^{(\alpha)}(t) = \sum_{g \in G} a_g t^{\alpha(g)}.$$

Theorem 2.10 (McMullen [McM1])**.** *Let F be the fibered face of a hyperbolic 3-manifold. Then for each integral $\alpha \in F \cdot \mathbb{R}^+$, the dilatation of (S_α, ϕ_α) equals the house of the specialization*

$$\lambda(\phi_\alpha) = |\Theta_F^{(\alpha)}|.$$

Combining the Universal Finiteness Theorem (Theorem 2.9) with Penner's result on the asymptotic behavior of minimum dilatations given in Equation (1), it follows that there are a finite number of fibered faces that contain points corresponding to mapping classes whose closures (obtained by filling in punctures) give rise to mapping classes (S_g, ϕ_g) realizing $\lambda(\phi_g) = \delta_g$. Theorem 2.10 shows further that there is a finite set of group ring elements $\Theta_i \in \mathbb{Z}G_i$, $i = 1, \dots, k$, so that the dilatations of these maps equal the house of specializations of these elements.

We now change notation, and think of group rings $\mathbb{Z}G$ as Laurent polynomial rings. That is, if G has generators $t_1, \dots, t_k$, then there is a natural isomorphism of $\mathbb{Z}G$ with the Laurent polynomial ring $\Lambda(t_1, \dots, t_k) = \mathbb{Z}[t_1^{\pm 1}, \dots, t_k^{\pm 1}]$, where each element of G is considered as a monomial in $t_1, \dots, t_k$. Similarly, there is an isomorphism of $\mathbb{Z}^k$ with $\mathrm{Hom}(G;\mathbb{Z})$ where $\mathbf{m} = (m_1, \dots, m_k)$ corresponds to the map that sends t_i to t^{m_i}, where we think of t as the generator of $\mathbb{Z}$. By these identifications, the specialization of $p(t_1, \dots, t_k) \in \Lambda(t_1, \dots, t_k)$, at $\mathbf{m}$ is defined by

$$p^{(\mathbf{m})}(t) = p(t^{m_1}, \dots, t^{m_k}).$$

Putting the Universal Finiteness Theorem (Theorem 2.9) together with Theorem 2.10, we have the following.

Theorem 2.11 (Universal Finiteness Theorem II)**.** *For any constant C, there is a finite list of Laurent polynomials $p_1, \dots, p_r \in \mathbb{Z}[[t_1, \dots, t_k]]$ so that if (S, ϕ) satisfies $L(S, \phi) < C$, then*

$$\lambda(\phi) = |p_i^{(\mathbf{m})}(t)|$$

for some $i = 1, \dots, r$ and $\mathbf{m} \in \mathbb{Z}^{\mathbf{k}}$.

2.5. **The magic manifold.** All of the known minimum dilatation examples for punctured as well as closed surfaces are associated, after possibly adding or removing punctures, to points on the fibered face of the magic manifold (see [KT1] [KKT]). This is the 3-cusped hyperbolic 3-manifold that is topologically equal to the complement of the link drawn in Figure 1 in the 3-sphere S^3. The name *magic manifold* appears also in the context of hyperbolic 3-manifolds which admit many non-hyperbolic Dehn fillings, and is the 3-cusped hyperbolic 3-manifold with smallest volume [Gor].

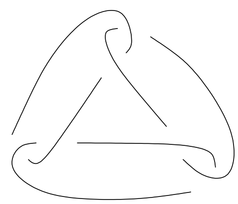

FIGURE 1. Magic Manifold as complement of links in S^3.

The first homology group $G = H_1(M;\mathbb{Z})$ is a free group on 3 generators x, y, z corresponding to meridian loops around the component of the link. The symmetry of the link induces a symmetry on the Thurston norm. Let $\hat{x}, \hat{y}, \hat{z}$ be the dual elements. These form a basis for $H^1(M;\mathbb{R})$, and x, y, z define coordinate functions on $H^1(M;\mathbb{R})$. With respect to these coordinates, Thurston norm ball is the convex polytope with vertices $(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1), (\pm 1, \pm 1, \pm 1)$. Consider the face F defined by the convex hull of $(1, 0, 0), (1, 1, 1), (0, 1, 0), (0, 0, -1)$. The cone over F can be characterized by the property

$$x + y - z > \max\{x, y, x - z, y - z, 0\},$$

and F is given by

$$\{(x, y, z) \ : \ x + y - z = 1, x > 0, y > 0, x > z, y > z\}.$$

We switch to multiplicative notation by replacing x, y, z with t^x, t^y, t^z. Then, the Teichmüller polynomial for F is given by

$$P(t^x, t^y, t^z) \quad = \quad t^{x+y-z} - t^x - t^y - t^{x-z} - t^{y-z} + 1. \tag{3}$$

2.6. **Dehn Fillings.** Let M be a hyperbolic 3-manifold with cusps. Each cusp looks topologically like

$$S^1 \times S^1 \times (0, \infty),$$

and we can think of M as the interior of a 3-manifold M^u with torus boundary components. A *Dehn filling* of M^u at a torus boundary component is the 3-manifold given by attaching a solid torus by identifying boundaries. The filled 3-manifold is determined up to homeomorphism type by the image of the contracting loop on the surface of the solid torus in $\pi_1(M)$. This can be specified by a slope when M is a knot or link complement in S^3 as follows. The meridian μ is the element of the fundamental group of the torus boundary component that contracts in S^3, and the longitude γ is the element whose linking number with the knot in S^3 equals zero. Then Dehn fillings are determined by rational numbers $\frac{p}{q}$, where $q\mu + p\gamma$ is the contracting loop. If the component of the link is clear, we write the Dehn filling as $M(\frac{b}{a})$. Thus, for example, if M is the complement of a knot in S^3, then $M(0) = S^3$. If M' is obtained from the complement M of a link with k components $\ell_1, \dots, \ell_k$ with meridians μ_i and longitudes γ_i, then we write M' as $M' = M(\frac{p_1}{q_1}; \dots; \frac{p_k}{q_k})$.

If M has a circle fibration $\psi : M \to S^1$ with monodromy (S, ϕ), then the intersection of S with a cusp of M determines a Dehn filling M' of M along the cusp. Let F be the fibered face of M containing the dual element α_S of S. The map $H^1(M';\mathbb{R}) \to H^1(M;\mathbb{R})$ defined by the inclusion $M \hookrightarrow M'$ is one-to-one, since every loop on M' can be pushed off into M. Let F' be the preimage of F in $H^1(M';\mathbb{R})$. Since the map $H_1(M;\mathbb{R}) \to H_1(M';\mathbb{R})$ has kernel generated by the contracting loop of the Dehn filling, we have the following.

Proposition 2.12. *If the boundary slope is a finite order element of $H^1(M;\mathbb{R})$, then the inclusion $F' \hookrightarrow F$ is a bijection. Otherwise, F' maps to a co-dimension one linear section of F.*

The elements of F' inherit many of the properties of F.

Proposition 2.13. *Let α' be a rational element of F', and α its image in F.*

1. *The boundary slopes defined by the intersection of the dual surface S_α with the cusp are all homologically equivalent to that defined by S.*
2. *The intersections S'_α with the filled cusp define a periodic orbit of ϕ'_α.*
3. *If the points in the periodic orbit do not come from poles of the quadratic differential on S determined (up to scalar multiple) by the stable and unstable foliations associated to ϕ_α, then (S_α, ϕ_α) is pseudo-Anosov and*

$$\lambda(\phi'_\alpha) = \lambda(\phi_\alpha).$$

The proof of parts (1) and (2) of Proposition 2.13 is an easy consequence of the definitions. Part (3) follows from the fact that the stable and unstable foliations of (S_α, ϕ_α) also form stable and unstable foliations for $(S'_\alpha, \phi'_\alpha)$ as long as the periodic orbit does not consist of poles.

Remark 2.14. In the case of poles, it is possible that $(S'_\alpha, \phi'_\alpha)$ is not pseudo-Anosov. In this case, by Theorem 2.2, it follows that the Dehn filling M' is not hyperbolic, and hence $(S'_\alpha, \phi'_\alpha)$ is not pseudo-Anosov for all rational $\alpha' \in F'$. Such a Dehn filling is called an *exceptional Dehn filling*, and it was shown by Thurston that there are only a finite number of boundary slopes with this property.

Let $\Theta \in \mathbb{Z}G$ be the Teichmüller polynomial for F and $\Theta' \in \mathbb{Z}G'$ the Teichmüller polynomial for F', where $G = H_1(M;\mathbb{Z})/\text{torsion}$ and $G' = H_1(M';\mathbb{Z})/\text{torsion}$.

Proposition 2.15. *If no periodic orbit contains poles, then the Teichmüller polynomial of F' is a factor of the specialization of the Teichmüller polynomial for F defined by the map $i_* : G \to G'$ induced by the inclusion $i : M \to M'$, that is, if*

$$\Theta = \sum_g a_g g,$$

then Θ' *divides* $\sum_g a_g i_*(g)$.

Remark 2.16. Assuming the case that the periodic orbit does not consist of poles, the effect of Dehn filling on normalized dilatation is more complicated than for the dilatation itself. For example, if α' is a rational element of F' and α is its image in F, then

$$\chi(S_\alpha) = \chi(S'_\alpha) - s_\alpha,$$

where s_α is the number of components in the intersection of S_α with the cusp, and depends on α. Thus, the normalized dilatation function L on F' is not the pull back of the normalized dilatation function on F, and the effect of pull back on the minimizer of normalized dilatation is not obvious.

2.7. **Fibered faces of the manifold $M_{\mathrm{m}}(\frac{1}{-2})$.** The minimum dilatation orientable pseudo-Anosov mapping class of genus 8 is the monodromy of a fibration of $M_s = M_{\mathrm{m}}(\frac{1}{-2})$ (see [Hir]). The manifold M_s is homeomorphic to the complement of the encircled closure of the braid $\sigma_1\sigma_2^{-1}$, where σ_1 and σ_2 are the standard braid generators of the braid group on 3-strands. This two component link, known as 6_2^2 in Rolfsen's knot table [Rolf], is symmetric in the two components and can be drawn in two ways (see Figure 2).

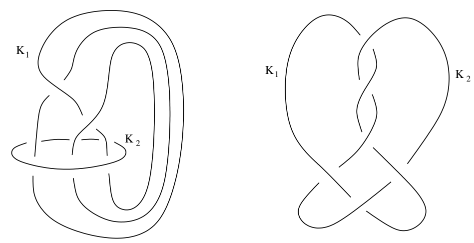

FIGURE 2. Two drawings of the 6_2^2 link.

Let M_{m} be the magic manifold described in Section 2.5. Assume that the Dehn filling is done on the cusp of M_{m} corresponding to the coordinate function y. Then inclusion map $M_{\mathrm{m}} \to M_s$ induces the surjection

$$H_1(M_{\mathrm{m}};\mathbb{R}) \to H_1(M_s;\mathbb{R})$$

has kernel generated by $t^{y+2(x+z)}$. Substituting $x = b$, $z = a$ and $y = -2(b+a)$ in Equation 3 gives

$$\begin{aligned} P(t^a,t^b) &= t^{3b+a} - t^{2b+2a} - t^b - t^{b-a} - t^{a+2b} + 1 \\ &= (t^{b+a}+1)(t^{2b} - t^{b+a} - t^b - t^{b-a} + 1). \end{aligned}$$

Let F_{m} be the fibered face described in Section 2.5. In [Hir], we show that the fibered face F_s of M_s corresponding to F_{m} is the locus

$$F_s = \{(x,z) \ : \ x = 1, \ -1 < z < 1\},$$

and the Teichmüller polynomial equals

$$\theta_s(t^a, t^b) = t^{2b} - t^{b+a} - t^b - t^{b-a} + 1.$$

The Alexander polynomial of M_s equals [Rolf]

$$\Delta_s(t^a, t^b) = t^{2b} - t^{b+a} + t^b - t^{b-a} + 1.$$

Let $\alpha(a,b)$ denote the element of $H^1(M : \mathbb{R}$ that sends x to b and z to a. If b is even, and a is odd, then

$$|\theta_s(t^a, t^b)| = |\Delta_s(t^a, t^b)|$$

and we have the following.

Proposition 2.17. *On the fibered face F_s of M_s, the monodromy of $\alpha(a,b)$ is orientable if and only if b is even and a is odd, and in particular, it is orientable when b is even and $a = 1$.*

The monodromy $(S_{(a,b)}, \phi_{(a,b)})$ associated to a rational point on F_s whose primitive element has coordinates (a,b) has topological Euler characteristic equal to minus the degree of the Alexander polynomial. Thus, the genus of $S_{(a,b)}$ is given by

$$g(a,b) = 1 + b - \frac{s}{2},$$

where s is the number of punctures of $S_{(a,b)}$.

Let K_1 and K_2 be the connected components of the 6_2^2-link, and let μ_i and γ_i be their meridian and longitude for $i = 1,2$. Then μ_1 and μ_2 generate $H_1(M_s;\mathbb{Z})$ and

$$\gamma_1 = 3\mu_2 \qquad \gamma_2 = 3\mu_1.$$

Take any integral $(a,b) \in F_s \cdot \mathbb{R}^+$, and let $\alpha = \alpha(a,b)$. Let B_i be the boundary tori of tubular neighborhoods of K_i in M_s. For $i = 1,2$, let $m_i = \alpha(\mu_i)$ and $\ell_i = \alpha(\gamma_i)$ be the images of the meridians and longitudes of K_i. Let

$$d_1 = \gcd(a, 3b) \qquad \text{and} \qquad d_2 = \gcd(3a, b).$$

Then d_i is the index of the image of $\pi_1(B_i)$ in $\mathbb{Z}$, and hence is equal to the number of connected components of $S_{(a,b)} \cap B_i$.

In the particular case where $(a,b) = (1,n)$, we have the following.

Lemma 2.18. *The number of punctures s of $S_{(1,n)}$ is given by*

$$s = \begin{cases} 2 & \text{if } 3 \text{ doesn't divide } n \\ 4 & \text{if } 3 \text{ divides } n \end{cases}$$

Corollary 2.19. *The monodromies $(S_{1,g}, \phi_{1,g})$, where $g = 2,4 \pmod 6$, have the property that*

(1) $S_{1,g}$ has genus g;
(2) $S_{1,g}$ has two singularities of degrees $3g-2$ and $g-2$, respectively;
(3) $(S_{1,g}, \phi_{1,g})$ is orientable; and
(4) $\lambda(\phi_{1,g}) = |LT_{1,g}|$.

By Fried's theorem (Theorem 2.5), the function $L(S,\phi)$ extends to a continuous convex function on F that goes to infinity toward the boundary. Thus, it has a unique minimum in F_s. The Teichmüller polynomial is invariant under the involution on $H_1(M_s;\mathbb{R})$ given by sending z to $-z$. It follows that $\lambda(S,\phi))$ is symmetric around the $z=0$ axis, and the minimum of L on F occurs at the rational point $\frac{\alpha(0,1)}{||\alpha(0,1)}$, and is given by

$$\lambda(\phi_{(0,1)}) = |t^3 - 3t + 1| = \frac{3+\sqrt{5}}{2} = \gamma_0^2.$$

Thus the conjectural minimum accumulation point for genus normalized dilatations of pseudo-Anosov mapping classes (Conjecture 1.1).

Concretely $(S_{(0,1)}, \phi_{(0,1)})$ is the mapping class known as the *simplest hyperbolic braid.* Using the left diagram in Figure 2, consider the three times punctured disk D bounded by the encircling link K_2. Then D is Poincare dual to μ_2 considered as an element of $H_1(M_s;\mathbb{Z})$, and hence is the dual surface to $\alpha(0,1)$. The mondromy is defined by considering M_s as the complement of the braid defined by K_1 in a solid torus given by the complement of a thickened K_2 in S^3. The solid torus fibers uniquely up to isotopy over S^1 with fiber D, and the monodromy is the braid monodromy defined by K_2, namely the one defined by $\sigma_1\sigma_2^{-1}$, where σ_1 and σ_2 are the braid generators.

The points $\alpha(1,n)$ in $H^1(M_s;\mathbb{R})$ define rays converging to the ray through $\alpha(0,1)$, and hence the sequence $L(S_{(1,n)}, \phi_{(1,n)})$ converges to $\to L(S_{(0,1)}, \phi_{(0,1)})$. Since $\chi(D) = -2$, we have

$$\lambda(\phi_{(1,g)})^{2g} = L(S_{(1,g)}, \phi_{(1,g)}) \to L(S_{(0,1)}, \phi_{(0,1)}) = \gamma_0^4.$$

This leads to the more general version of Conjecture 1.1.

Conjecture 2.20. *The smallest accumulation point for normalized dilatations is* γ_0^4.

The minimum dilatation orientable pseudo-Anosov mapping classes of genus 7 was found independently in [AD] and [KT2] and is the monodromy of $M_w = M_m(\frac{3}{-2})$, which is the complement of the $(-2,3,8)$-pretzel link, also known as the Whitehead sister-link in S^3. The minimum dilatations of pseudo-Anosov mapping classes arising as monodromies of circle fibrations of M_w are all of the form $|LT_{a,b}|$, where $a \in \{3,7,13,17\}$ and $b = g+2$. Putting together the examples above, we have the following.

Proposition 2.21. *For all g*

$$\delta_g \leq |LT_{1,g}|,$$

and hence

$$\limsup(\delta_g)^g \leq \gamma_0^2$$

and

$$\limsup L(S,\phi) \leq \gamma_0^4.$$

Let $\lambda_{(a,b)} = |LT_{(a,b)}|$, and let $\lambda_{(x,y,z)} = |P(t^x, t^y, t^z)|$. In Table 1, we show the smallest known dilatations for orientable and unconstrained pseudo-Anosov mapping classes on closed surfaces of genus 2 through 12. These put together the results in [AD] (Table 1.9), [KT2] (Thm 1.6, 1.7, 1.12, and Prop. 4.3.7), [KKT] (Table 1) and [Hir] (Prop 4.7).

g	orientable	unconstrained
2	$\lambda_{(1,2)} \approx 1.72208$	same
3	$\lambda_{(3,4)} \approx 1.40127$	same
4	$\lambda_{(1,4)} \approx 1.28064$	$\lambda_{(3,5)} \approx 1.26123$
5	$\lambda_{(1,6)} \approx 1.17628$	$\lambda_{(1,7)} \approx 1.14879$
6	$\lambda_{(10,8,3)} \approx 1.20189$	$\lambda_{(1,8)} \approx 1.12876$
7	$\lambda_{(2,9)} \approx 1.11548$	same
8	$\lambda_{(1,8)} \approx 1.12876$	$\lambda_{(18,17,7)} \approx 1.10403$
9	$\lambda_{(2,11)} \approx 1.09282$	same
10	$\lambda_{(1,10)} \approx 1.10149$	$\lambda_{(1,12)} \approx 1.08377$
11	$\lambda_{(1,12)} \approx 1.08377$	$\lambda_{(1,13)} \approx 1.07705$
12	$\lambda_{(12,20,3)} \approx 1.10240$	$\lambda_{(3,14)} \approx 1.07266$

TABLE 1. Smallest known dilatations for genus $g \leq 12$.

2.8. **Dilatations of pseudo-Anosov mapping classes.** We are particularly interested in the subclass of pseudo-Anosov mapping classes whose stable and unstable foliations are orientable. This is equivalent to the condition that the *homological dilatation* $\lambda_{\text{hom}}(\phi)$, which is the spectral radius of the action of ϕ on the first homology of S, is equal to the geometric dilatation $\lambda(\phi)$. Such mapping classes are called *orientable*. Let δ_g^+ be the minimum dilatation for orientable pseudo-Anosov mapping classes on S_g. By the results in [Pen] and [HK], δ_g^+ has the same asymptotic behavior as δ_g:

$$\log(\delta_g^+) \asymp \frac{1}{g}.$$

In the orientable case, δ_g^+ has been computed for $g = 2, 3, 4, 5, 7, 8$ beginning with work by Lanneau and Thiffeault in [LT] and continuing with [Hir], [AD] [KT2]. In [LT] Lanneau and Thiffeault also gave the first attempt to describe the behavior of minimum dilatation explicitly as a function of g. Given a polynomial $p(t)$, the *house* of $p(t)$ is given by

$$|p| = \max\{|\mu| \ : \ p(\mu) = 0\}.$$

Question 2.22. *Let*

$$p_n(t) = t^{2n} - t^{n+1} - t^n - t^{n-1} + 1.$$

Then for even genus $g \geq 2$,

$$\delta_g^+ = |p_g|.$$

If the answer to Question 2.22 is affirmative, then

$$\liminf_{g \to \infty} (\delta_g^+)^g \leq \gamma_0^2,$$

where γ_0 is the golden mean. This suggests the following conjecture (cf. Conjecture 1.1).

Conjecture 2.23. *The genus-normalized minimum dilatations satisfy*

$$\liminf_{g\to\infty}(\delta_g^+)^g = \gamma_0^2.$$

3. Digraphs and Perron units

The dynamics of a pseudo-Anosov mapping class $\phi : S \to S$, in particular, the structure of the stable and unstable invariant foliations, can be captured in terms of an associated directed graph, via an associated train track map. The train track map defines a Perron-Frobenius linear map T that preserves a symplectic bilinear form, and the dilatation of the mapping class equals the Perron-Frobenius eigenvalue of T. It follows that dilatations are Perron units. The minimum dilatation problem for pseudo-Anosov mapping classes is closely related in spirit to Lehmer's problem for Mahler measures of monic integer polynomials posed in [Leh]. In this section, we review Lehmer's question on the distribution of algebraic integers, and focus on the particular case of Perron units.

3.1. **Mahler measure and Lehmer's question.**

Given a monic integer polynomial

$$p(t) = t^d + a_{d-1}t^{d-1} + \cdots + a_0, \qquad a_i \in \mathbb{Z}$$

the *Mahler measure* is given by

$$\mathcal{M}(p) = \prod_{p(\mu)=0} \max\{1, |\mu|\}.$$

In [Leh], Lehmer asks: *is there a positive gap between 1 and the next smallest Mahler measure?*

The smallest known Mahler measure greater than one is called *Lehmer's number*

$$\lambda_L \approx 1.17628,$$

and its minimal polynomial for λ_L is

$$p_L(t) = t^{10} + t^9 - t^7 - t^6 - t^5 - t^4 - t^3 + t + 1.$$

By a result of Smyth [Smy], the smallest Mahler measure of a non-reciprocal irreducible polynomial is approximately $\lambda_S = 1.32472$, which is greater than λ_L. Thus to solve Lehmer's problem it suffices to look at reciprocal polynomials.

3.2. **Normalized house.**

The *house* of a polynomial is given by

$$|p| = \max\{|\mu| \; : \; p(\mu) = 0\}.$$

We have the inequalities

$$|p| \le \mathcal{M}(p) \le |p|^d. \tag{4}$$

We call $|p|^d$ the *normalized house* of $p(t)$. It is an open question whether there is a positive gap between 1 and the next smallest normalized house. If the answer is no, it would imply that there are sequences of Mahler measures converging to 1 from above.

Lehmer's polynomial p_L has only one root outside the unit circle, and hence we have the first inequality in Equation (4),

$$|p_L| = \mathcal{M}(p_L).$$

The second inequality is also sharp (e.g., take $p(t) = t^n - 2$).

3.3. **Perron numbers.** A *Perron-Frobenius matrix* T is an $n \times n$ matrix whose entries are all non-negative real numbers, and such that for some k_0, the entries of T^k are all positive all $k \geq k_0$. Given a non-negative matrix $T = [a_{i,j}]$, one can define an associated directed graph, or *digraph*, D with n vertices $v_1, \dots, v_n$ and $a_{i,j}$ directed edges from v_i to v_j. By this correspondence T is Perron-Frobenius if and only if D is *strongly connected*, i.e., there is a directed path between any two vertices, and *aperiodic*, the path lengths of cycles have no common divisor greater than one [Kit]. By the Perron-Frobenius theorem, if T is Perron-Frobenius, then there is a vector v with positive entries such that $Tv = \lambda v$, for some $\lambda > 1$, and λ is completely determined by these properties. This λ is called the *Perron-Frobenius eigenvalue* of T, or *dilatation* of D.

A *Perron number* is a real algebraic integer $\lambda > 1$ such that all algebraic conjugates have complex norm strictly less than λ. An algebraic integer is a Perron number if and only if it is the Perron-Frobenius eigenvalue of a matrix. Pisot and Salem numbers are examples of Perron numbers. A Pisot number is an algebraic integer greater than one all of whose other algebraic conjugates lie strictly inside the unit circle. A Salem number is an algebraic integer greater than one all of whose other algebraic conjugates lie on or inside the unit circle with at least one on the unit circle. The smallest Pisot number is the smallest Mahler measure λ_S for non-reciprocal polynomials found by Smyth. It is not known whether there are Salem numbers arbitrarily close to 1 or whether the infimum of all Mahler measures greater than 1 is a Salem numbers. The smallest known Salem number is Lehmer's number λ_L.

Graph theory provides an answer to the minimum normalized house problem for Perron numbers and their defining polynomials. Recalling the correspondence between Perron-Frobenius matrices and digraphs, one notes that the smallest dilatation digraph has the form given in Figure 3 (see [Pen]). The characteristic polynomial of the digraph is

$$p_n(t) = t^n - t - 1,$$

for $n \geq 4$. The polynomial is interesting also in the case $n = 2$, since $|p_2| = \gamma_0$ is the golden mean, and in the case $n = 3$, since $p_3 = x^3 - x - 1$ is the Smyth polynomial defining λ_S. We also have

$$\lim_{n \to \infty} |p_n|^n = 2,$$

where the convergence is from above.

Properties of the normalized house of reciprocal Perron numbers were recently studied in [McM2], showing that any Perron unit α of degree n satisfies the inequality

$$\alpha^n \geq \gamma_0^4,$$

where γ_0 is the golden mean (see Theorem 3.2).

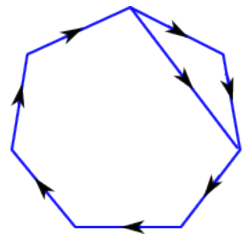

FIGURE 3. Minimum dilatation digraph.

3.4. **Complexity of digraphs.** The *complexity* c of a digraph is the number of edges minus the number of vertices of the graph (or minus the topological Euler characteristic).

Lemma 3.1 (Ham-Song [HS]). *If λ is the spectral radius of M, then c satisfies the inequality*

$$c \leq \lambda^{2n} - 1.$$

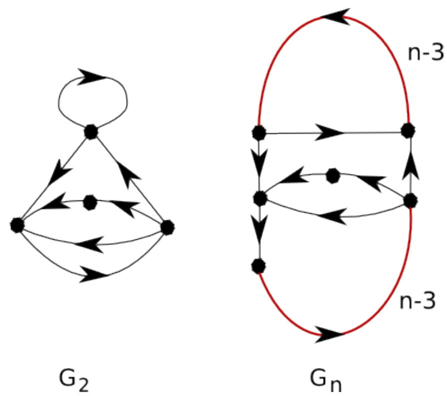

FIGURE 4. Digraphs realizing $LT_{1,n}$.

Figure 4 shows a family of directed graphs whose characteristic polynomials are given by $LT_{1,3}$. In the Figure, an edge labeled m is subdivided into a chain of m edges and $m-1$ additional vertices. Other examples of digraphs with the same dilatation were found in [Bir]. The ones shown in Figure 4 have the additional property that they are defined from the transition matrix of train track maps associated to pseudo-Anosov mapping classes (see Section 6).

The LT polynomials satisfy

$$|LT_{1,n}| \leq |LT_{a,n}|$$

for all $1 \leq a < n$, and for any fixed $0 < a$,

$$\lim_{n \to \infty} |LT_{a,n}|^{2n} = \left(\frac{3+\sqrt{5}}{2}\right)^2 = \gamma_0^4.$$

Thus to find the smallest Perron units, it suffices to consider only those with $\lambda < \lambda_n = |LT_{1,n}|$. It follows that to solve the minimum dilatation problem it suffices to look at mapping classes whose corresponding digraphs have complexity $c \leq 5$.

3.5. **Dilatations of digraphs whose matrices preserve a symplectic form.** It is well-known that any Perron number can be realized as the spectral radius of a Perron Frobenius matrix. Furthermore, any Perron unit is the dilatation of a Perron Frobenius matrix that preserves a symplecitc form. It is not known, however, whether every Perron unit is a dilatation of pseudo-Anosov mapping class.

Given a Perron unit λ, we define its *PF-degree* to be the minimum dimension of a Perron Frobenius matrix realizing λ. McMullen has recently announced the following result giving further support to Conjecture 1.1.

Theorem 3.2 (McMullen [McM2])**.** *Let p_d be the minimum Perron unit of Perron degree d. Then*

(1) $(p_n)^n \geq \gamma_0^4$ *for all* $n \geq 1$*, and*
(2) $\lim_{n\to\infty}(p_n)^n = \gamma_0^4$.

4. Orientable pseudo-Anosov mapping classes

In [LT] Lanneau and Thiffeault studied potential defining polynomials for δ_g^+ in the cases $g = 2, \ldots, 8$, and found lower bounds for δ_g^+ for these g. Using known examples whose dilatations match these lower bounds they determined δ_g^+ for $g = 2, 3, 4, 5$. From the results of Cho and Ham in [CH], it follows that $\delta_2 = \delta_2^+$. Lanneau and Thiffeault's lower bound for $g = 6$ agrees with δ_5^+, showing that δ_g^+ is not strictly monotone decreasing. An example realizing δ_7^+ was found in [AD] and in [KT2], and an example realizing δ_8^+ was found in [Hir]. The exact value for δ_6^+ is not known.

The minimum dilatations of orientable pseudo-Anosov mapping classes for low genus are given in Table 2. The associated *PF-polynomial* is the characteristic polynomial of an associated Perron-Frobenius matrix. This is not necessarily irreducible. In Table 2 we repeatedly see the cyclotomic factor $\sigma(t) = t^2 - t + 1$.

g	$\delta_g^+ \approx$	PF polynomial	factorization
2	1.72208	$t^4 - t^3 - t^2 - t + 1$	irreducible
3	1.40127	$t^8 - t^7 - t^4 - t + 1$	$\sigma(t)(t^6 - t^4 - t^3 - t^2 + 1)$
4	1.28064	$t^8 - t^5 - t^4 - t^3 + 1$	irreducible
5	1.17628	$t^{12} - t^7 - t^6 - t^5 + 1$	$\sigma(t)(t^{10} + t^9 - t^7 - t^6 - t^5 - t^4 - t^3 + t + 1)$
7	1.11548	$t^{18} - t^{11} - t^9 - t^7 + 1$	$\sigma(t)(t^{14} + t^{13} - t^9 - t^8 - t^7 - t^6 - t^5 + t + 1)$
8	1.12876	$t^{16} - t^9 - t^8 - t^7 + 1$	irreducible

Table 2. List of minimum dilatations and their PF polynomials.

For $a, b \in \mathbb{Z}$, define the *Lanneau-Thiffeault polynomial* $LT_{a,b}$ to be the polynomial

$$LT_{a,b}(t) = t^{2b} - t^{b+a} - t^b - t^{b-a} + 1.$$

As can be seen from Table 2, for $g = 2, 3, 4, 5, 7, 8$, the PF polynomial for the minimum dilatations of orientable pseudo-Anosov mapping classes is a Lanneau-Thiffeault polynomial.

Question 2.22 can be rephrased as follows.

Question 4.1 (Lanneau-Thiffeault Question). *For even $g \geq 2$ is it true that*

$$\delta_g^+ = |LT_{1,g}|$$

where $|LT_{1,g}|$ is the house of $LT_{1,g}(t)$?

By the following result, $|LT_{1,g}|$ is an upper bound for δ_g^+ for g ranging in an arithmetic sequence or even integers.

Theorem 4.2. *[*[Hir]*] For each $g \equiv 2, 4 \pmod 6$, there is an orientable pseudo-Anosov mapping class on a genus g closed surface with dilatation equal to $|LT_{1,g}|$.*

5. Fat train track maps and automata

For each pseudo-Anosov mapping class, one can associate a fat train track map that encodes essential geometric information, including information about singularities, the invariant stable foliation, and dilatations. In this section, we give relevant background and definitions.

5.1. **Train tracks and train track maps.** A train track is a finite topological graph τ (or 1-complex) with no double edges or vertices of degree one. A *smoothing* of τ at a vertex v is a choice of tangent directions for the half edges of τ that meet at v, that is if e_1 and e_2 meet at a vertex, then they meet either smoothly or in a cusp.

In Figure 5, e_3 meets e_1 and e_2 smoothly, while e_1 and e_2 meet at a cusp.

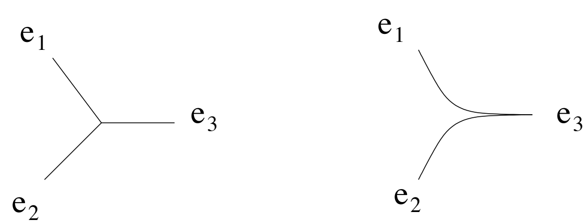

Figure 5. Smoothing at a trivalent vertex

Figure 6 shows a smoothing of a degree four vertex.

For our examples, we will consider train tracks consisting of a $3b$-gon whose edges meet in cusps and $2b$-edges attached smoothly to the vertices of the $3b$-gon in one of the ways shown in Figure 5 and Figure 6.

By a *fat graph*, we mean a graph such that at any vertex v, there is a cyclic ordering of the half edges that meet at v. This gives a local embedding of the half edges meeting at v

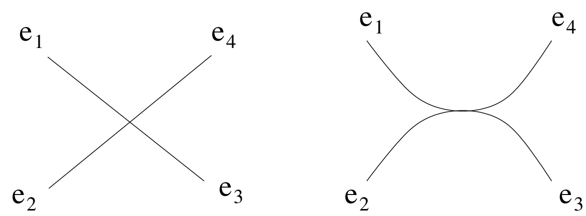

FIGURE 6. Smoothing at a degree 4 vertex

into a disk centered at v. Given any fat graph Γ, there is a canonical orientable surface S_Γ with boundary on which Γ embeds so that

(1) at each vertex the ordering of the edges corresponds to the counterclockwise ordering on the surface; and
(2) S_Γ deformation retracts to the image of Γ under the embedding.

Each boundary component is one boundary component of an annular complementary component of τ on S_Γ. Consider the edges surrounding the other *interior* boundary component. Each time two adjacent edges meet in a cusp, we call it a *vertex of the polygon formed by τ around the boundary component.* If the number of vertices of the polygon is k, we say the boundary component is *contained in a k-gon* of τ.

A fat train track τ embedded on a surface S *fills* S if S is obtained from S_τ by filling in some subset (possibly empty) of the boundary components of S_τ with disks.

A *train track map* $f : \tau \to \tau$ is a local embedding so that vertices map to vertices, and edges map to edge-paths on τ so that no subinterval of an edge passes across two half edges meeting at a cusp. We consider train track maps up to isotopy on τ.

A train track map f determines a linear transformation $\mathbb{R}^{\mathcal{E}}$ to itself as followis. Let $\mathcal{E}$ be the set of (unoriented) edges of τ. Given $e \in \mathcal{E}$, let

$$f_*(e) = \sum_{e'} a_{e'} e',$$

where $a_{e'}$ is the number of times $f(e)$ passes over e'. Define $T : \mathbb{R}^{\mathcal{E}} \to \mathbb{R}^{\mathcal{E}}$, where for each $w \in \mathbb{R}^{\mathcal{E}}$,

$$T(w)(e) = w(f_*(e)),$$

where w extends linearly. The transformation T is called the *transition map* defined by f.

The *weight space* W_τ of a train track τ is the subspace of $\mathbb{R}^{\mathcal{E}}$ consisting of edge labels so that if three half edges e_1, e_2 and e_3 meet at a vertex as in Figure 5, then

$$w(e_1) + w(e_2) = w(e_3),$$

and if e_1, e_2, e_3 and e_4 meet as in Figure 6, then

$$w(e_1) + w(e_2) = w(e_3) + w(e_4).$$

An edge labeling w determines a labeling on edge paths, which we also denote by w. Given a train track map f with transition map T, we have $T(W_\tau) = W_\tau$.

A train track $\tau \subset S$ and train track map $f : \tau \to \tau$ is *compatible* with a mapping class (S, ϕ), if τ fills S and the induced map ϕ_* on τ equals f.

Theorem 5.1. *If (S, ϕ) is pseudo-Anosov, then*

(1) (S, ϕ) has a compatible train track τ and train track map $f : \tau \to \tau$;
(2) the induced map f_* on W_τ is Perron-Frobenius, and preserves a symplectic form; and
(3) $\lambda(\phi)$ is the spectral radius of f_*.

In the examples that follow, it is possible to find a subcollection of edges in $\mathcal{E}$ whose duals in $\mathbb{R}^{\mathcal{E}}$ form a basis for W_τ. We call these the *real* edges of τ and the complementary set of edges the *infinitessimal* edges.

5.2. **Simplest hyperbolic braid.** Figure 7 gives an example of a fat train track and train track map compatible with the simplest hyperbolic braid. The weights in the weight space are determined by their labels on the two longer edges of the train track, and the three encircling loops are the corresponding infinitessimal edges. The action of the simplest hyperbolic braid monodromy defined by $\sigma_1\sigma_2^{-1}$ acts on the real edges according to the matrix

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix},$$

and the dilatation is the largest eigenvalue $\frac{3+\sqrt{5}}{2} = \gamma_0^2$.

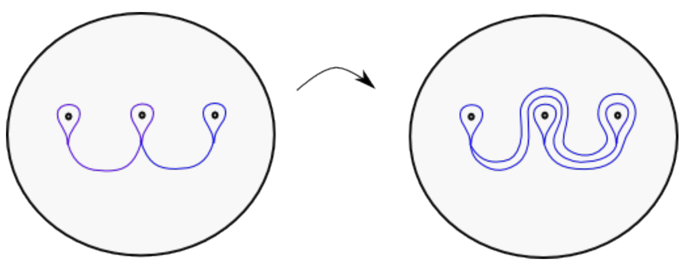

FIGURE 7. Train track for simplest hyperbolic braid monodromy

5.3. **Orientable train tracks.** Each train track on S determines a foliation on S as follows. For each complementary region of τ on S surrounded by a k-gon, the foliation has a k-pronged singularity. A train track is orientable, if there is an orientation on the edges so that if two edges meet smoothly at a vertex, the orientations are compatible.

Figure 8 sketches the foliation around a boundary component of S corresponding to a hexagon on a fat train track. The orientation on the train track determines an orientation on the foliations.

Thus, we have the following.

Proposition 5.2. *A pseudo-Anosov map (S, ϕ) that has a compatible train track map $f : \tau \to \tau$, where τ is orientable, is orientable.*

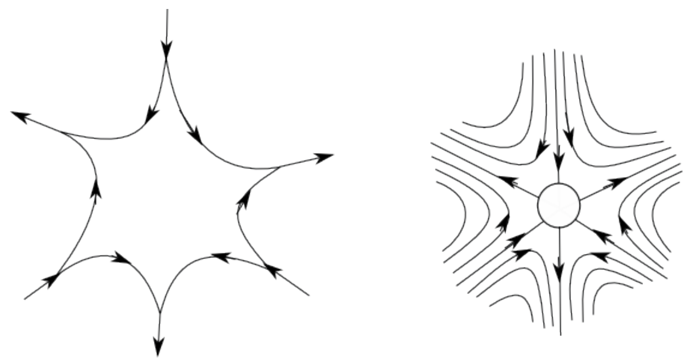

FIGURE 8. A hexagon on a fat traintrack, and corresponding foliations.

5.4. **Train track automaton.** Given two fat train tracks τ_1 and τ_2, a *folding map* $\mathfrak{f}$: $\tau_1 \to \tau_2$ is a quotient map obtained by identifying edge-segments of a pair of edge in τ_1 as follows. Take two edges e_1 and e_2 on τ_1 with half edges that meet at a cusp at a vertex v, and that are adjacent in the fat graph ordering. Then the folding map of e_1 over e_2 is obtained by identifying the embedded image of a closed interval in e_1 with endpoint v with e_2 by a homeomorphism sending v to v. The fat train track automaton is the set of all fat train tracks with a directed edge from one train track to another if there is a folding map between them.

Each folding map is a homotopy equivalence of graphs and defines a linear transformation between edge labels, and between weight spaces. A circuit in the fat train track automaton corresponds to a composition of folding maps together with an homeomorphism of train tracks. Thus, the transition matrix for the train track map corresponds to a composition of transition matrices for folding maps and a permutation matrix.

A train track automaton is a directed graph whose vertices are train tracks and edges are folding maps.

Proposition 5.3 (Stallings [Sta], Ham-Song [HS])**.** *Any pseudo-Anosov mapping class can be represented by a circuit on a train track automaton.*

6. Small dilatation examples

In this section, we define train track maps for mapping classes (S_n, ϕ_n) for all integers $n \geq 2$, and describe corresponding circuits in the train track folding automaton, and digraphs. These train track maps define mapping classes with the same genus, boundary components, and dilatations as $(S_{1,n}, \phi_{1,n})$.

We begin with a fat train track map defining (S_2, ϕ_2) in Figure 9 . One can check that all of the train tracks in the circuit shown in Figure 9 fix a genus two surface with two complementary disk components, one bounded by the central hexagon, and the other bounded by the edges of the hexagon and by each side of the four real edges. The train track map defined by composing the folded mapping classes described in the circuit corresponds to the orientable pseudo-Anoosv mapping classes whose dilatation realizes $\delta_2 = \delta_2^+$.

The center hexagon is made up of infinitessimal edges and the other four edges are real edges. Starting at the upper left train track in the the automaton, we first fold edge a over edge c and the following adjacent infinitessimal edge. In the next step we fold b over the new edge a. Then we fold the new edge b over c. Finally by a rotation, we return to the original train track.

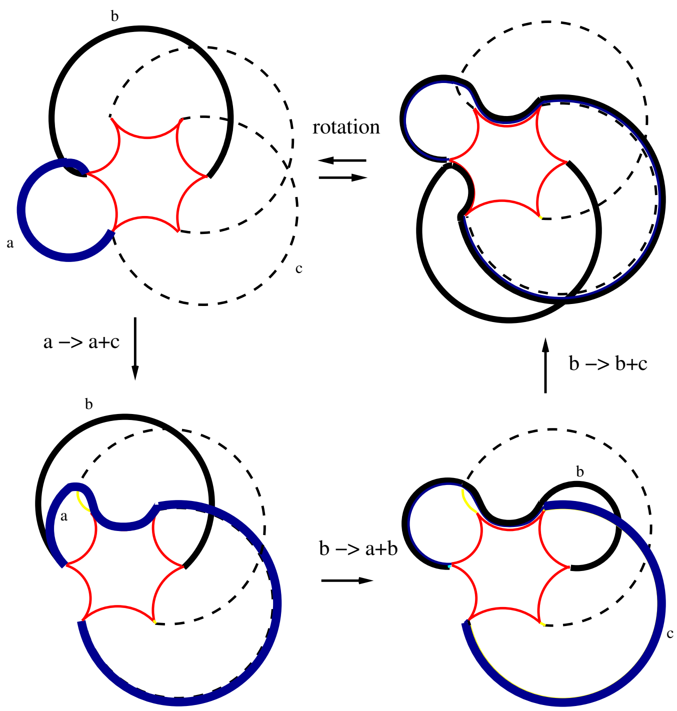

FIGURE 9. Train track circuit for example realizing δ_2^+ and δ_2.

The transition matrices for the folding diagrams starting at the top left and going around counter-clockwise are:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The composition is given by

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

and its characteristic polynomial is $x^4 - x^3 - x^2 - x + 1$. This gives

$$\delta_2 = \delta_2^+ = |x^4 - x^3 - x^2 - x + 1| \approx 1.72208.$$

The train track in Figure 9 generalizes to the one in Figure 10.

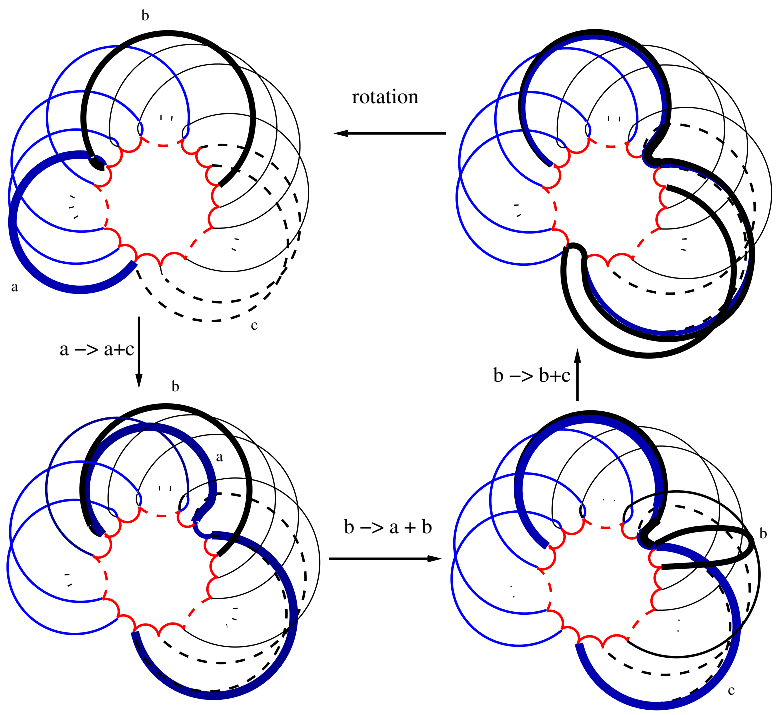

FIGURE 10. Circuit in train track automaton for (S_n, ϕ_n)

Let G_n be the digraphs in Figure 4. The "shape" of the train track map and folding maps for (S_n, ϕ_n) are related to each other in a systematic way, and one observes the following.

Proposition 6.1. *The digraphs associated to the transition matrices for the train track maps of (S_n, ϕ_n) are G_n, and hence the dilatations of (S_n, ϕ_n) are given by*

$$\lambda(\phi_n) = |LT_{1,n}|.$$

The genus of S_n can be determined from the topological Euler characteristic of G_n, $\chi(G_n) = 2n$ and the number of boundary components of the fat graph. There is one component for the central $3n$-gon, and either one or three other boundary components, depending on whether n is divisible by 3. This implies the following.

Proposition 6.2. *The surface S_n has genus $g = n$ if $n = 1, 2 \pmod 3$, and has genus $g = n - 1$ if $n = 0 \pmod 3$.*

From the train track maps, we can also determine when the mapping classes are orientable, for this is exactly when the train tracks themselves are orientable as seen in the next proposition.

Proposition 6.3. *The mapping class (S_n, ϕ_n) is orientable if and only if n is even.*

Proof. The complementary region of (S_n, ϕ_n) splits into a central $3n$-gon and either one n-gon, or three $n/3$-gons, depending on whether or not n is divisible by 3. In order for the train track to be orientable, we need to have each polygon have an even number of sides. Thus, n must be even.

When n is even, there are two possible ways to orient the central $3n$-gon. Each extends to a compatible orientation on the entire train track. (An example is shown in Figure 11). □

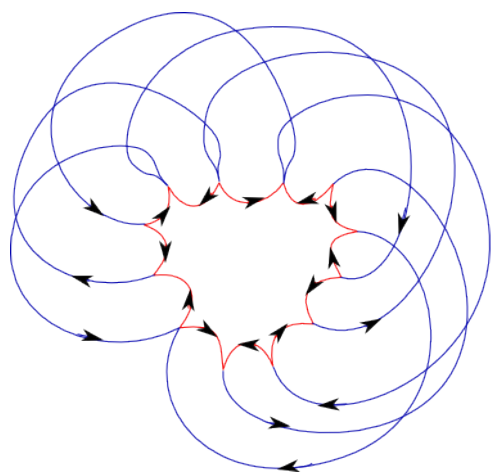

FIGURE 11. Oriented train track for $n = 4$.

REFERENCES

[AD] J. Aaber and N. Dunfield. Closed surface bundles of least volume. *Algebr. Geom. Topology* **10** (2010), 2315–2342.

[Bir] J. Birman. On pseudo-Anosov mapping classes with minimum dilatation and Lanneau-Thiffeault numbers. *arxiv:1101.2383v1* (2011).

[CB] A. Casson and S. Bleiler. *Automorphisms of surfaces after Nielsen and Thurston.* Cambridge University Press, 1988.

[CH] J. Cho and J. Ham. The minimal dilatation of a genus two surface. *Experiment. Math.* **17** (2008), 257–267.

[Dob] E. Dobrowolski. On a question of Lehmer and the number of irreducible factors of a polynomial. *Acta. Arith.* **34** (1979), 391–401.

[FLM] B. Farb, C. Leininger, and D. Margalit. Small dilatation pseudo-Anosovs and 3-manifolds. *Adv. Math* **228** (2011), 1466–1502.

[FLP] A. Fathi, F. Laudenbach, and V. Poénaru. Some dynamics of pseudo-Anosov diffeomorphisms. In *Travaux de Thurston sur les surfaces*, volume 66-67 of *Astérisque*. Soc. Math. France, Paris, 1979.

[Fri] D. Fried. Flow equivalence, hyperbolic systems and a new zeta function for flows. *Comment. Math. Helvetici* **57** (1982), 237–259.

[Gor] C. Gordon. Small surfaces and Dehn filling. In *Proceedings of the Kirbyfest (Berkeley, CA, 1998)*, volume 2, pages 177–199. Coventry, 1999.

[HS] J-Y Ham and W. T. Song. The minimum dilatation of pseudo-Anosov 5-braids. *Experimental Mathematics* **16** (2007), 167,180.

[Hir] E. Hironaka. Small dilatation pseudo-Anosov mapping classes coming from the simplest hyperbolic braid. *Algebr. Geom. Topol.* **10** (2010), 2041–2060.

[HK] E. Hironaka and E. Kin. A family of pseudo-Anosov braids with small dilatation. *Algebr. Geom. Topol.* **6** (2006), 699–738.

[KT1] E. Kin and M. Takasawa. Pseudo-Anosov braids with small entropy and the magic 3-manifold. *Comm. Anal. Geo.* **19** (2011), 705–758.

[KT2] E. Kin and M. Takasawa. Pseudo-Anosovs on closed surfaces having small entropy and the Whitehead sister link exterior. *J. Math. Soc. Japan* **65** (2013), 411–446.

[Kit] B. Kitchens. *Symbolic dynamics: one-sided, two-sided and countable state Markov shifts.* Springer, 1998.

[KKT] S. Kojima, E. Kin, and M. Takasawa. Minimal dilatations of pseudo-Anosovs generated by the magic 3-manifold and their asymptotic behavior. *Algebraic and Geometric Topology* **13** (2013), 3537–3602.

[LT] E. Lanneau and J.-L. Thiffeault. On the minimum dilatation of pseudo-Anosov homeomorphisms on surfaces of small genus. *Ann. de l'Inst. Four.* **61** (2009), 164–182.

[Leh] D. H. Lehmer. Factorization of certain cyclotomic functions. *Ann. of Math.* **34** (1933), 461–469.

[McM1] C. McMullen. Polynomial invariants for fibered 3-manifolds and Teichmüller geodesics for foliations. *Ann. Sci. École Norm. Sup.* **33** (2000), 519–560.

[McM2] C. McMullen. Entropy and the clique polynomial. *Preprint* (2013).

[Pen] R. Penner. Bounds on least dilatations. *Proceedings of the A.M.S.* **113** (1991), 443–450.

[Rolf] D. Rolfsen. *Knots and Links*. Publish or Perish, Inc, Berkeley, 1976.

[Smy] C. J. Smyth. On the product of the conjugates outside the unit circle of an algebraic integer. *Bull. London Math. Soc.* **3** (1971), 169–175.

[Sta] J. R. Stallings. Topology of finite graphs. *Invent. Math.* **71** (1983), 551–565.

[Sun] H. Sun. A transcendental invariant of pseudo-Anosov maps. *preprint* (2012).

[Thu1] W. Thurston. A norm for the homology of 3-manifolds. *Mem. Amer. Math. Soc.* **339** (1986), 99–130.

[Thu2] W. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)* **19** (1988), 417–431.

[Thu3] W. Thurston. *arXiv:1402.2008 [math.DS]* (2014).

Department of Mathematics, Florida State University, 1017 Academic Way, Tallahassee, FL 32306-4510. email: hironaka@math.fsu.edu

COXETER GARLANDS IN $\mathbf{H}^4$ AND 2-SALEM NUMBERS

YOHEI KOMORI

1. Introduction

One idea to construct a series of hyperbolic Coxeter polytopes is to paste copies of an initial Coxeter polytope along their orthogonal facets. (cf.[20]). T. Zehrt and C. Zehrt [21] studied such Coxeter polytopes in 4-dimensional hyperbolic space $\mathbb{H}^4$, called *Coxeter garlands* whose initial truncated simplex has 2 orthogonal facets (cf.[14]). They computed the growth functions $G^n(t)$ of hyperbolic Coxeter groups G^n associated to Coxeter garlands and showed that they have two reciprocal pairs of poles on the positive real axis and remaining poles are on the unit circle.

On the other hand Cannon and Wagreich [1] and Parry [11] proved that the growth functions of cocompact 2 and 3-dimensional hyperbolic Coxeter groups have one reciprocal pairs of poles on the positive real axis and remaining poles are on the unit circle. As a consequence, the growth rates of cocompact 2 and 3-dimensional hyperbolic Coxeter groups are Salem numbers or quadratic units.

By means of results in [21] we will show that the growth rate of a Coxeter Garland G^n is always a *2-Salem number*, which is a generalization of a Salem number (cf.[13, 10]). Numerical calculations were performed by using Mathematica 9.

Y. Umemoto considered different series of hyperbolic Coxeter polytopes in $\mathbb{H}^4$, called *Coxeter dominoes* and proved that infinite many Coxeter dominoes have 2-Salem numbers as their growth rates [18]. She also showed that a Coxeter Garland G^n has a 2-Salem number as its growth rates when n is congruent to 1 modulo 15 [19].

The paper is organized as follows. In section 2 we collect basic definitions and results of hperbolic Coxeter groups and their growth functions. 2-Salem number is defined in section 3, and in section 4 we review a work of T. Zehrt and C. Zehrt [21] on Coxeter garlands. We will show our main result in section 5.

The author would like to thank Ruth Kellerhals and Eriko Hironaka for their wonderful organization of the conference on Growth and Mahler measure in geometry and topology at the Mittag-Leffler Institute on July, 2013. Thanks also to Ruth Kellerhals again and Yuriko Umemoto for helpful discussion and stimulative comments. This work was partially supported by Grant-in-Aid for Scientific Research(C) (19540194), Ministry of Education, Science and Culture of Japan.

2. Hyperbolic Coxeter groups and their growth functions

A convex polytope P with finite number of facets in the n-dimensional hyperbolic space $\mathbb{H}^n$ is called a *hyperbolic Coxeter polytope* if its dihedral angles are submultiples of π (cf.[12]). We can associate to P a graph called the *Coxeter diagram* Γ of P as follows: each vertex of Γ represents a facet of P and two vertices are connected by an edge with number $m \geq 3$ if the dihedral angles between two facets corresponding to two vertices is equal to π/m. If two facets are orthogonal then the corresponding two vertices are not connected, while if two facets are ultra-parallel then the corresponding two vertices are connected by a dotted edge (see Figures 1 and 2). The set S of reflections with respects to facets of P generates an infinite discrete group W acting on $\mathbb{H}^n$, and we call (W, S) a *n-dimensional hyperbolic Coxeter group.* The *growth function* $W(t)$ of (W, S) is the formal power series $\sum_{k=0}^{\infty} a_k t^k$ where a_k is the number of elements $g \in W$ whose word length with respect to S is equal to k (cf.[5]). Since the cardinality of W is infinite and that of S is finite, the *growth rate* of (W, S) $\omega := \limsup_{k\to\infty} \sqrt[k]{a_k}$ is bigger than or equal to 1 while it is less than or equal to the cardinality $|S|$ of S, since $a_k \leq |S|^k$, i.e. $1 \leq \omega \leq |S|$. By means of Cauchy-Hadamard formula, the *radius of convergence* R of $W(t)$ is the reciprocal of ω, i.e. $1/|S| \leq R \leq 1$. Therefore $W(t)$ is not only a formal power series but also an analytic function of $t \in \mathbb{C}$ on the open disk $|t| < R$. In practice the analytic function $W(t)$ on $|t| < R$ extends to a rational function $P(t)/Q(t)$ on $\mathbb{C}$ by analytic continuation where $P(t), Q(t) \in \mathbb{Z}[t]$ are relatively prime. We also have precise formulas due to Solomon [16] and Steinberg [17] to calculate the rational function $P(t)/Q(t)$ from the Coxeter diagram of (W, S) ([16, 17]. See also [6, 8]).

Theorem 1 (Steinberg formula). *Let us denote by (W_T, T) the Coxeter subgroup of (W, S) generated by the subset $T \subseteq S$, and let its growth function be $W_T(t)$. Set $\mathcal{F} = \{T \subseteq S \ : \ W_T \text{ is finite } \}$. Then*

$$\frac{1}{W(t^{-1})} = \sum_{T\in\mathcal{F}} \frac{(-1)^{|T|}}{W_T(t)}.$$

Theorem 2 (Solomon formula). *The growth function $G(t)$ of an irreducible finite Coxeter group (G, T) can be written as $G(t) = \prod_{i=1}^{k}[m_i + 1]$ where $[n] := 1 + t + \cdots + t^{n-1}$ and $\{m_1, m_2, \cdots, m_k\}$ is the set of exponents of (G, T).*

Since a_k is a natural number for all $k \geq 0$, $t = R$ is a pole of $W(t)$ on the circle $|t| = R$. Hence R is a real zero of the denominator $Q(t)$ closest to the origin $0 \in \mathbb{C}$ of all zeros of $Q(t)$. Theorem 2 implies that $P(0) = \pm 1$. Hence $a_0 = 1$ implies that $Q(0) = \pm 1$. Moreover de la Harpe [4] proved that Γ is of exponential growth, i.e. $\omega > 1$. Therefore the growth rate ω, the reciprocal of R, becomes a real algebraic integer grater than 1 whose conjugates have moduli less than or equal to the modulus of ω.

The growth function $W(t)$ of a *cocompact* hyperbolic Coxeter group has special symmetries: Serre [15] and Charney and Davis [1] proved that for cocompact n-dimensional hyperbolic Coxeter groups, $W(t)$ is *reciprocal*, i.e. $W(1/t) = W(t)$ when n is even, and *anti-reciprocal*, i.e. $W(1/t) = -W(t)$ when n is odd.

3. 2-Salem numbers

In this section we review the notion of 2-Salem numbers following Samet [13] and Kerada [10]. We start from the definition of Salem numbers. A real algebraic integer $\tau > 1$ is called a *Salem number* if τ^{-1} is a conjugate of τ and all conjugates of τ other than τ and τ^{-1} lie on the unit circle, and at least one of them is on the unit circle. We remark that a quadratic unit is not a Salem number in this sense. It can be proved that the minimal polynomial of τ over $\mathbf{Z}$ is palindromic of even degree, where a polynomial $Q(t)$ of degree d is called *palindromic* if $Q(t)$ satisfies $Q(t) = t^d Q(1/t)$. Cannon and Wagreich [1] and Parry [11] proved

Theorem 3 (Cannon, Wagreich and Parry)**.** *The growth rates of cocompact 2 and 3-dimensional hyperbolic Coxeter groups are Salem numbers or quadratic units.*

It seems a natural question whether quadratic units can be realized actually as growth rates of cocompact 2 and 3-dimensional hyperbolic Coxeter groups. We answer in the affirmative by giving concrete examples.

Example 1. *The growth function for hyperbolic pentagon angles* $\pi/2, \pi/4, \pi/4, \pi/4, \pi/4$ *is*

$$f_S(x) = \frac{(x+1)^2(x^2+1)}{(x^2-4x+1)(x^2+x+1)}.$$

Therefore its growth rate is a quadratic unit.

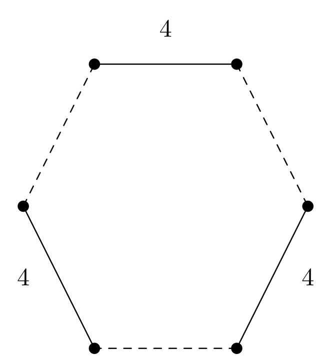

Figure 1

Example 2. *The growth function for 3-dimensional hyperbolic Lambert cube* (cf.[7, 9]) *defined by the Coxeter diagram of Figure 1 is*

$$f_S(x) = \frac{(x+1)^3(x^2+1)}{(x-1)(x^2-3x+1)(x^2+x+1)}.$$

Therefore its growth rate is a quadratic unit.

Now we come to the definition of 2-Salem numbers.

Definition 1 (cf.[13, 10])**.** *A real algebraic integer $\alpha > 1$ is called a 2-Salem number if it has a real conjugate $\beta > 1$ while other conjugates ω satisfy $|\omega| \leq 1$ and at least one of them is on the unit circle.*

It can be proved that $1/\alpha < 1$ and $1/\beta < 1$ are also conjugate of α and other conjugates are on the unit circle. Also the minimal polynomial of α over $\mathbf{Z}$ is palindromic of even degree.

4. Coxeter garlands and their growth functions

In this section we review a work [21] of T. Zehrt and C. Zehrt on growth functions of Coxeter garlands. The Coxeter diagram of Figure 2 represents a compact 4-dimensional hyperbolic Coxeter polytope with 2 orthogonal facets named by a and b. They took 2 copies of it, pasted them along the

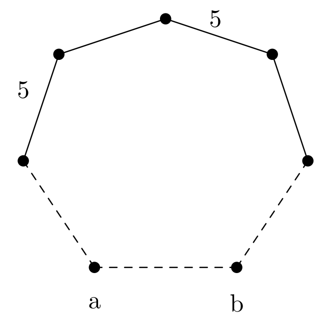

Figure 2

facet of type a and got a new polytope P with 2 orthogonal facets of type b. They calculated the growth function for the geometric Coxeter group G corresponding to P as follows

$$\frac{[2]^2[5][6](t^5+1)}{t^{16}-4t^{15}+t^{14}+t^{12}+t^{11}+2t^9+2t^7+t^5+t^4+t^2-4t+1}.$$

Let G^n be the geometric Coxeter group corresponding to the Coxeter polytope P^n constructed from n copies of P by (n-1)- gluings along orthogonal facets of type b.. They called G^n a Coxeter garland in $\mathbf{H}^4$. Then by means of the gluing formula ([21] Corollary 2), they showed

Theorem 4 ([21] Theorem 1)**.** *The growth function $G^n(t)$ of G_n is equal to $[2][2][5][6](t^5+1)/Z_n(t)$ where the denominator is a palindromic polynomial of degree 16 defined by*

$$\begin{aligned}D_n(t) &= t^{16} - 2(n+1)t^{15} + t^{14} + (n-1)t^{13} + t^{12} + nt^{11}\\ &+(n-1)t^{10} + 2t^9 + 2(n-1)t^8 + 2t^7 + (n-1)t^6\\ &+nt^5 + t^4 + (n-1)t^3 + t^2 - 2(n+1)t + 1.\end{aligned}$$

About the pole distribution of $G^n(t)$, they proved

Theorem 5 ([21] Theorem 2)**.** *$D_n(t)$ has six reciprocal pairs of roots on the unit circle, and two reciprocal pairs of roots on the positive real axis.*

Therefore the growth rates of Coxeter garlands seem to be 2-Salem numbers, but it might be happened that $D_n(t)$ is a product of Salem polynomials and cyclotomic polynomials.

5. Irreducibility of the denominator polynomials $D_n(t)$

In this section we prove that the growth rates of Coxeter garlands are 2-Salem numbers. More precisely

Theorem 6. *For any $n \in \mathbf{N}$, the denominator polynomial $D_n(t)$ of the growth function $G^n(t)$ of a Coxeter garland G^n is irreducible over $\mathbf{Z}$, hence $D_n(t)$ is a 2-Salem polynomial.*

Because of Theorem 5, if $D_n(t)$ is reducible over $\mathbf{Z}$, hence can be written as a product of polynomials over $\mathbf{Z}$, then each factor should be a palindromic polynomial of even degree. The next observation will play a key role for computations below:

Proposition 1. *For any $n \in \mathbf{N}$, $D_n(i) = 2$.*

First we assume that $D_n(t)$ has a palindromic polynomial of degree 2 as its factor over $\mathbf{Z}$:

$$\begin{aligned}D_n(t) &= (t^2+pt+1)(t^{14} + at^{13} + bt^{12} + ct^{11} + dt^{10} + et^9 + ft^8 + gt^7\\ &\quad +ft^6 + et^5 + dt^4 + ct^3 + bt^2 + at + 1).\end{aligned}$$

$D_n(i) = 2$ implies $p(2a - 2c + 2e - g) = -2$, hence $p = 2, 1, -1, -2$.

Suppose that $p = 2$. Since the quadratic factor is $t^2 + 2t + 1$, $t = -1$ should be a root of $D_n(t)$ while $D_n(-1) = 4(1+n) \neq 0$, a contradiction. Therefore $p = 2$ cannot be happened.

Suppose that $p = -2$. Since the quadratic factor is $t^2 - 2t + 1$, $t = 1$ should be a root of $D_n(t)$ while $D_n(1) = 4n \neq 0$, a contradiction. Therefore $p = -2$ cannot be happened.

Suppose that $p = 1$. Since the quadratic factor is $t^2 + t + 1$, $t = \frac{-1\pm\sqrt{3}i}{2}$ should be a root of $D_n(t)$ while $Z(\frac{-1\pm\sqrt{3}i}{2}) = 2(-1 \mp \sqrt{3})n \neq 0$, a contradiction. Therefore $p = 1$ cannot be happened.

Suppose that $p = -1$. Since the quadratic factor is $t^2 - t + 1$, $t = \frac{1 \pm \sqrt{3}i}{2}$ should be a root of $D_n(t)$ while $Z(\frac{1 \pm \sqrt{3}i}{2}) = (1 \mp \sqrt{3})(1+n) \neq 0$, a contradiction. Therefore $p = -1$ cannot be happened.

The above arguments conclude that $D_n(t)$ cannot have a quadratic factor.

Next we assume that $D_n(t)$ has a palindromic polynomial of degree 4 as its factor over $\mathbf{Z}$:

$$\begin{aligned} D_n(t) &= (t^4 + pt^3 + qt^2 + pt + 1)(t^{12} + at^{11} + bt^{10} + ct^9 + dt^8 + et^7 \\ &\quad + ft^6 + et^5 + dt^4 + ct^3 + bt^2 + at + 1). \end{aligned}$$

$D_n(i) = 2$ implies $(2-q)(2-2b+2d-f) = 2$, hence $q = 0, 1, 3, 4$.

Suppose that $q = 0$:

$$\begin{aligned} D_n(t) &= (t^4 + pt^3 + pt + 1)(t^{12} + at^{11} + bt^{10} + ct^9 + dt^8 + et^7 \\ &\quad + ft^6 + et^5 + dt^4 + ct^3 + bt^2 + at + 1). \end{aligned}$$

Comparing coefficients of degrees from 1 to 6 of both sides, a, b, c, d, e and f can be written as polynomials in p and n:

$$\begin{aligned} a &= -2(1+n) - p \\ b &= 1 + 2(1+n)p + p^2 \\ c &= -1 + n - 2p - 2(1+n)p^2 - p^3 \\ d &= (1 - n + 2(1+n))p + 3p^2 + 2(1+n)p^3 + p^4 \\ e &= n + 2(1+n) + (-1 + n - 4(1+n))p^2 - 4p^3 - 2(1+n)p^4 - p^5 \\ f &= -2 + n + (1 - 2n - 4(1+n))p + p^2 + (1 - n + 6(1+n))p^3 \\ &\quad +5p^4 + 2(1+n)p^5 + p^6. \end{aligned}$$

Substitute them in coefficients of degree 7 of both sides, we have

$$\begin{aligned} &1 + 4p + 7p^2 + p^3 - 7p^4 - 5p^5 - 2p^6 - p^7 \\ &+n(-4 - p + 10p^2 - 5p^4 - 2p^6) = 0. \end{aligned} \tag{1}$$

Also substitute them in coefficients of degree 8 of both sides, we have

$$\begin{aligned} &-2 - 10p - 6p^2 + 6p^3 + 6p^4 + 4p^5 + 2p^6 \\ &+n(2 - 8p + 2p^3 + 4p^5) = 0. \end{aligned} \tag{2}$$

Eliminating n from equations (1) and (2), we have the following equation in p:

$$(-1+p)(1+p)(-3 - 21p - 19p^2 + 12p^3 + 14p^4) = 0.$$

Drawing the graph of the factor of degree 4, integer solutions of this equation are $p = \pm 1$. On the other hand substituting $p = 1$ in the equation (1), we have $-2(1+n) \neq 0$, a contradiction. Also substituting $p = -1$ in the equation (2), we have $4n \neq 0$, a contradiction. Therefore $q = 0$ cannot be happened.

Since similar arguments also work for the remaining cases $q = 1, 3, 4$, $D_n(t)$ cannot have a quartic factor.

Next we assume that $D_n(t)$ has a palindromic polynomial of degree 6 as its factor over $\mathbf{Z}$:

$$\begin{aligned} D_n(t) &= (t^6+pt^5+qt^4+rt^3+qt^2+pt+1)(t^{10}+at^9+bt^8+ct^7 \\ &\quad +dt^6+et^5+dt^4+ct^3+bt^2+at+1). \end{aligned}$$

$D_n(i) = 2$ implies $(2p-r)(2a-2c+e) = -2$, hence $2p-r = 2, 1, -1, -2$.

Suppose that $2p-r = 2$, then $2a-2c+e = 1$:

$$\begin{aligned} D_n[t] &= (t^6+pt^5+qt^4+(2p-2)t^3+qt^2+pt+1)(t^{10}+at^9+bt^8+ct^7 \\ &\quad +dt^6+(2c-2a+1)t^5+dt^4+ct^3+bt^2+at+1). \end{aligned}$$

Comparing coefficients of degrees from 1 to 4 of both sides, a, b, c and d can be written as polynomials in p, q and n:

$$\begin{aligned} a &= -2(1+n)-p \\ b &= 1+2p+2np+p^2-q \\ c &= 1+n-3p-2p^2-2np^2-p^3+2q+2nq+2pq \\ d &= -3-4n+p+3np+5p^2+2p^3+2np^3+p^4-2q-4pq \\ &\quad -4npq-3p^2q+q^2. \end{aligned}$$

Substitute them in coefficients of degree 5 of both sides, we have

$$\begin{aligned} &-3+8p+p^2-5p^3-2p^4-p^5-5q+4pq+6p^2q+4p^3q-2q^2-3pq^2 \\ &+n(-5+8p-3p^2-2p^4-3q+6p^2q-2q^2) = 0. \end{aligned} \tag{3}$$

Substitute them in coefficients of degree 6 of both sides, we have

$$\begin{aligned} &3-12p+2p^2+4p^3+3p^4+8q-3pq-11p^2q-2p^3q-p^4q+2q^2+4pq^2 \\ &+3p^2q^2-q^3+n(7-9p-4p^2+6p^3+8q-9pq-2p^3q+4pq^2) = 0. \end{aligned} \tag{4}$$

Substitute them in coefficients of degree 7 of both sides, we have

$$\begin{aligned} &-3+14p+7p^2-11p^3-4p^4-3p^5-12q+4pq+12p^2q+12p^3q-4q^2 \\ &-9pq^2+n(-7+18p-9p^2+4p^3-6p^4-9q-8pq+18p^2q-6q^2) = 0. \end{aligned} \tag{5}$$

Eliminating n from equations (3) and (4), we have the following equation in p and q:

$$\begin{aligned} &-6-p+62p^2-126p^3+70p^4+p^5-10q+49pq-57p^2q+38p^3q \\ &-32p^4q-14q^2+33pq^2-9p^2q^2+q^3-12pq^3+15p^2q^3+q^4-2q^5 = 0. \end{aligned} \tag{6}$$

Eliminating n from equations (3) and (5), we have the following equation in p and q:

$$\begin{aligned} &6-16p+78p^2-100p^3+48p^4-16p^5-7q-8pq-21p^2q+64p^3q \\ &+2p^4q+15q^2-32pq^2-36p^2q^2+12q^3+4q^4 = 0. \end{aligned} \tag{7}$$

Computing the resultant of equations (6) and (7) as $\mathbf{Z}[p]$-polynomials,

$$\begin{aligned}
&-8(-1+p)(-80+296p-861p^2+1162p^3-713p^4-268p^5+468p^6\\
&+16p^7-96p^8+4p^{10})(-5475312+34083042p-134485023p^2\\
&+707826723p^3-2742855623p^4+5867659300p^5-6730355143p^6\\
&+3718368253p^7-323615337p^8-613351370p^9+214464326p^{10}\\
&+22488900p^{11}-17440236p^{12}+106568p^{13}+464888p^{14})=0.
\end{aligned}$$

Also computing the resultant of equations (6) and (7) as $\mathbf{Z}[q]$-polynomials,

$$\begin{aligned}
&-q(194820-63992q+45985q^2-166553q^3+84898q^4-45052q^5\\
&+52944q^6-28976q^7+7088q^8-784q^9+32q^{10})(-2738917152\\
&+13492543248q-36191712384q^2+82537797960q^3-142602958866q^4\\
&+137995520901q^5-29199693741q^6-82067358493q^7+101493662948q^8\\
&-58969044416q^9+20178685032q^{10}-4201236128q^{11}+516039832q^{12}\\
&-34091288q^{13}+929776q^{14})=0.
\end{aligned}$$

Drawing the graphs of the factors of 10 and 14 for these two equations, we see that $p=1, q=0$ is the unique integer solution of them. On the other hand substituting $p=1, q=0$ in the equation (3), we have $-2(1+n)\neq 0$, a contradiction. Therefore $2p-r=2$ cannot be happened.

Since similar arguments also work for the remaining cases $2p-r=1,-1,2$, $D_n(t)$ cannot have a factor of degree 6.

Finally we assume that $D_n(t)$ has a palindromic polynomial of degree 8 as its factor over $\mathbf{Z}$:

$$\begin{aligned}
D_n(t) &= (t^8+pt^7+qt^6+rt^5+st^4+rt^3+qt^2+pt+1)\\
&\quad (t^8+at^7+bt^6+ct^5+dt^4+ct^3+bt^2+at+1).
\end{aligned}$$

$D_n(i)=2$ implies $(s-2q+2)(d-2b+2)=2$, hence $s-2q=0,-1,-3,-4$.

Suppose that $s=2q$, then $d=2b-1$:

$$\begin{aligned}
D_n[t] &= (t^8+pt^7+qt^6+rt^5+2qt^4+rt^3+qt^2+pt+1)\\
&\quad (t^8+at^7+bt^6+ct^5+(2b-1)t^4+ct^3+bt^2+at+1).
\end{aligned}$$

Comparing coefficients of degrees from 1 to 3 of both sides, a,b and c can be written as polynomials in p,q,r and n:

$$\begin{aligned}
a &= -2(1+n)-p\\
b &= 1-(-2(1+n)-p)p-q\\
c &= -1+n-(1-(-2(1+n)-p)p-q)p-(-2(1+n)-p)q-r.
\end{aligned}$$

Substitute them in coefficients of degree 4 of both sides, we have

$$\begin{aligned}
&-5p-4p^2-p^3-q-pq-p^2q+q^2+2r+pr\\
&+n(-4p-2p^2-2pq+2r)=0.
\end{aligned} \tag{8}$$

Substitute them in coefficients of degree 5 of both sides, we have

$$\begin{aligned}(9)\qquad &-1-3p-5p^2-2p^3+4q+2pq-p^2q+q^2-2r-2pr-p^2r+qr\\ &+n(1-2p-4p^2+4q-2pq-2pr)=0.\end{aligned}$$

Substitute them in coefficients of degree 6 of both sides, we have

$$\begin{aligned}(10)\qquad &-2-3p-3p^2-p^3-3q-7pq-4p^2q+4q^2+r-pr-p^2r+qr\\ &+n(1-2p-2p^2-8pq+2r-2pr)=0.\end{aligned}$$

Eliminating n from equations (8) and (9), we have the following equation in p,q and r:

$$\begin{aligned}(11)\ &(-5p-4p^2-p^3-q-pq-p^2q+q^2+2r+pr)(1-2p-4p^2+4q\\ &-2pq-2pr)-(-1-3p-5p^2-2p^3+4q+2pq-p^2q+q^2-2r-2pr\\ &-p^2r+qr)(-4p-2p^2-2pq+2r)=0.\end{aligned}$$

Eliminating n from equations (8) and (10), we have the following equation in p,q and r:

$$\begin{aligned}(12)\quad &(-5p-4p^2-p^3-q-pq-p^2q+q^2+2r+pr)(1-2p-2p^2-8pq\\ &+2r-2pr)-(-2-3p-3p^2-p^3-3q-7pq-4p^2q+4q^2+r-pr\\ &-p^2r+qr)(-4p-2p^2-2pq+2r)=0.\end{aligned}$$

Computing the resultant of equations (11) and (12) as $\mathbf{Z}[p,q]$-polynomials,

$$\begin{aligned}&4(-1+q)(p+q)(8-106p-261p^2-109p^3-36p^4+18p^5+2p^6\\ &+117q+17pq-77p^2q+88p^3q+158p^4q+127q^2+448pq^2+796p^2q^2\\ &+186p^3q^2-70p^4q^2-284q^3-486pq^3-618p^2q^3-124p^3q^3-44q^4\\ &+80pq^4+128p^2q^4+144q^5+40pq^5-32q^6)=0.\end{aligned}$$

Computing the resultant of equations (11) and (12) as $\mathbf{Z}[p,r]$-polynomials,

$$\begin{aligned}&-4(2p-r)(3p+p^2-r)(-9+1426p+6429p^2+13270p^3+15364p^4\\ &+13804p^5+2012p^6-624r-2058pr-5740p^2r-14504p^3r-14780p^4r\\ &+528p^5r-49r^2-458pr^2+2394p^2r^2-4104p^3r^2-6380p^4r^2+342r^3\\ &+1074pr^3+8660p^2r^3+4848p^3r^3-268r^4-3016pr^4-1024p^2r^4\\ &+320r^5-72pr^5+32r^6)=0.\end{aligned}$$

Computing the resultant of equations (11) and (12) as $\mathbf{Z}[q,r]$-polynomials,

$$\begin{aligned}&8(-1+q)^2(2q+r)(-184-3105q-5779q^2+7036q^3\\ &+11074q^4-14200q^5+4024q^6+923r-2539qr-5778q^2r\\ &+8548q^3r-1376q^4r-816q^5r+2216r^2-1193qr^2-7152q^2r^2\\ &+8210q^3r^2-2352q^4r^2+997r^3-1116qr^3-352q^2r^3+460q^3r^3\\ &+444r^4-742qr^4+276q^2r^4-20qr^5-4r^6)=0.\end{aligned}$$

From these three equations, the possibilities of integer solutions (p,q,r) are $q=1, r=2p$ or $q=1, r=3p+p^2$ or $q=-p, r=2p$.

Substituting $q = 1, r = 2p$ in equations (9) and (10), we have

$$4 - 3p - 10p^2 - 4p^3 + n(5 - 4p - 8p^2) = 0$$
$$-1 - 6p - 9p^2 - 3p^3 + n(1 - 6p - 6p^2) = 0.$$

Eliminating n,

$$9 - p - 3p^2 + 5p^3 = 0$$

which has no integer solution, a contradiction.

Substituting $q = 1, r = 3p + p^2$ in equations (9) and (10), we have

$$4 - 4p - 13p^2 - 7p^3 - p^4 + n(5 - 4p - 10p^2 - 2p^3 = 0)$$
$$-1 - 4p - 8p^2 - 5p^3 - p^4 + n(1 - 4p - 6p^2 - 2p^3) = 0.$$

Eliminating n,

$$9 - 4p - 7p^2 + 12p^3 + 10p^4 + 2p^5 = 0$$

which has no integer solution, a contradiction.

Substituting $q = -p, r = 2p$ in equations (9) and (10), we have

$$-1 - 11p - 12p^2 - 3p^3 + n(1 - 6p - 6p^2) = 0$$
$$-2 + 2p + 4p^2 + p^3 + n(1 + 2p + 2p^2) = 0.$$

Eliminating n,

$$1 - 27p - 40p^2 - 14p^3 = 0$$

which has no integer solution, a contradiction. Therefore the case that $s = 2q$ cannot be happened.

Since similar arguments also work for the remaining cases $s - 2q = -1, -3, -4$, $D_n(t)$ cannot have a factor of degree 8.

References

[1] J. Cannon and P. Wagreich, Growth functions of surface groups, Math. Ann. **293** (1992), no. 2, 239-257.

[2] R. Charney, M. Davis, Reciprocity of growth functions of Coxeter groups, Geom. Dedicata **39** (1991), 373-378.

[3] W. Floyd, Growth of planar Coxeter groups, P.V. numbers, and Salem numbers, Math. Ann. **293** (1992), no. 3, 475-483.

[4] P. de la Harpe, Groupes de Coxeter infinis non affines, Exposition. Math 5 (1987), 91-96.

[5] P. de la Harpe, *Topics in geometric group theory*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 2000.

[6] G. Heckman, The volume of hyperbolic Coxeter polytopes of even dimension, Indag. Math. (N.S.) **6** (1995), 189-196.

[7] HCh. Im Hof, Napier cycles and hyperbolic Coxeter groups. Bull Soc. Math. Belgique **42**(1990), 523-545.

[8] J. Humphreys, *Reflection groups and Coxeter groups*, Cambridge Studies in Advanced Mathematics, 29, Cambridge Univ. Press, Cambridge, 1990.

[9] R. Kellerhals, On the volume of hyperbolic polyhedra, Math. Ann. bf 285(1989), 541-569.

[10] M. Kerada, Une caractérisation de certaines classes d'entiers algébriques généralisant les nombres de Salem, Acta Arithmetica **72** (1995), 55-65.

[11] W. Parry, Growth series of Coxeter groups and Salem numbers, J. Algebra **154** (1993), no. 2, 406-415.

[12] J. Ratcliffe, *Foundations of hyperbolic manifolds*, Grad. Texts in Math., 149, Springer, New York, 1994.

[13] P. Samet, Algebraic integers with two conjugates outside the unit circle, Algebraic integers with two conjugates outside the unit circle, Proc. Cambridge Philos. Soc. **49** (1953), 421-436.

[14] L. Schlettwein, Hyperbolische Simplexe, Diplomarbeit, Universität Basel, 1995.

[15] J. Serre, Cohomologie des groupes discrets, Prospects Math., Ann. Math. Stud. 70, 77-169, 1971.

[16] L. Solomon, The orders of the finite Chevalley groups, J. Algebra **3**(1966), 376-393.

[17] R. Steinberg, *Endomorphisms of linear algebraic groups*, Memoirs of the American Mathematical Society, No. 80, Amer. Math. Soc., Providence, RI, 1968.

[18] Y. Umemoto, The growth function of Coxeter dominoes and 2-Salem numbers, to appear in Algebraic and Geometric Topology.

[19] Y. Umemoto, Ph.D dissertation, in preparation.

[20] E. Vinberg, Hyperbolic reflection groups, Russ. Math. Surv. **40** (1985), 31-75.

[21] T. Zehrt and C Zehrt-Liebendörfer, The growth function of Coxeter garlands in $\mathbf{H}^4$, Beitr Algebra Geom. **53** (2012), 451-460.

Department of Mathematics, School of Education, Waseda University, Shinjuku, Tokyo 169-8050, Japan

E-mail address: `ykomori@waseda.jp`

A RECIPE TO COMPUTE MAHLER MEASURES

LÜCK, W.

Abstract. We give an algorithm to compute the Mahler measure of a polynomial which does only depend on the coefficients, does not need any informations about the roots, and comes with an explicit estimate of the error term. We also prove the positivity of the Novikov-Shubin invariants for matrices over the complex group ring of $\mathbb{Z}^d$.

0. Introduction

The main result of this paper is the following result, explanation will follow.

Theorem 0.1. *Let p be an element in $\mathbb{C}[\mathbb{Z}^d] = \mathbb{C}[z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1}]$ which is not constant. Define positive constants which depend only on d, the width $\operatorname{wd}(p)$, the leading coefficient $\operatorname{lead}(p)$ and the L^1-norm $||p||_{L^1}$*

$$\begin{aligned} C &:= \frac{12 \cdot \sqrt{3}}{\sqrt{47}} \cdot \big(d \cdot \operatorname{wd}(p)\big)^2 \cdot \left(\frac{||p||_{L^1}^2}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d \cdot \operatorname{wd}(p)}} + \frac{3 \cdot e}{2} \cdot d \cdot \operatorname{wd}(p); \\ \beta &:= \frac{1}{3 \cdot d \cdot \operatorname{wd}(p)}. \end{aligned}$$

Then there is a monotone decreasing sequence of positive real numbers $c(p, ||p||_{L^1})_k$, called characteristic sequence, such that for all integers numbers $L \geq 1$ we get for the Mahler measure $M(p)$

$$0 \leq \ln\big(||p||_{L^1}\big) - \ln(M(p)) - \sum_{k=1}^{L} \frac{c(p, ||p||_{L^1})_k}{2k} \leq C \cdot L^{-\beta}.$$

The *Mahler measure* of p is defined to be

$$M(p) := \exp\left(\int_{T^d} \ln\big(|(p(z_1, z_2, \ldots, z_d)|\big) \, d\mu_{T^d}\right),$$

where μ_{T^d} is the Haar measure of the d-dimensional torus T^d.

For a survey on the Mahler measure and its intriguing connections to number theory, topology and geometry, were refer for instance to [**?**, **?**, **?**, **?**]. The width $\operatorname{wd}(p)$ and the leading coefficient $\operatorname{lead}(p)$ are explained in Subsection **??**, whereas the L^1-norm $||p||_{L^1}$ and the characteristic sequence are introduced in Section **??**.

Here are some remarks on the algorithm.

Remark 0.2 (Dependency on the coefficients)**.** The width $\operatorname{wd}(p)$, the leading coefficient $\operatorname{lead}(p)$, the L^1-norm $||p||_{L^1}$ and the characteristic sequence $c(p, ||p||_{L^1})_k$ can be computed directly from the coefficients of p, one does not need any information about the roots of p.

Remark 0.3 (Estimate of the error term)**.** Theorem **??** provides an algorithm to compute the Mahler measure $M(p)$ of a non-constant element $p \in \mathbb{C}[\mathbb{Z}^d]$ up to a given upper bound on the error term. Namely, fix $\epsilon > 0$. Choose a natural number L satisfying

$$L \geq \left(\frac{\frac{12 \cdot \sqrt{3}}{\sqrt{47}} \cdot \big(d \cdot \operatorname{wd}(p)\big)^2 \cdot \left(\frac{||p||_{L^1}^2}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d \cdot \operatorname{wd}(p)}} + \frac{3 \cdot e}{2} \cdot d \cdot \operatorname{wd}(p)}{\epsilon}\right)^{3 \cdot d \cdot \operatorname{wd}(p)},$$

where the right hand side depends only on d, the width $\operatorname{wd}(p)$, the L^1-norm $||p||_{L^1}$ and the upper bound on the error term ϵ. Then we get

$$\ln(||p||_{L^1}) - \sum_{k=1}^{L} \frac{c(p, ||p||_{L^1})_k}{2k} - \epsilon \leq \ln(M(p)) \leq \ln(||p||_{L^1}) - \sum_{k=1}^{L} \frac{c(p, ||p||_{L^1})_k}{2k}.$$

Remark 0.4 (Continuity of the Mahler measure)**.** Fix constants $K_w, K_1, K_l > 0$.

At least it gives, by letting L run, a decreasing sequence of upper bounds for $\ln(M(p)$ because of

$$0 \leq \ln(M(p)) \leq \ln\big(||p||_{L^1}\big) - \sum_{k=1}^{L} \frac{c(p, ||p||_{L^1})_k}{2k}.$$

Our estimates are sometimes very crude, we have not tried to give optimal estimates.

Finally we mention the following direct consequence of Theorem **??**

Corollary 0.6. *Let $A \in M_{m,n}(\mathbb{C}[\mathbb{Z}^d])$ be any matrix. Then the Novikov-Shubin invariant of the bounded $\mathbb{Z}^d$-equivariant operator $r_A^{(2)} \colon L^2(\mathbb{Z}^d)^m \to L^2(\mathbb{Z}^d)^n$ given by right multiplication with A is positive.*

In this context we mention the unpublished preprint [**?**], where examples of groups G and matrices $A \in M_{m,n}(\mathbb{Z}G)$ are constructed for which the Novikov-Shubin invariant of $r_A^{(2)} \colon L^2(A)^m \to L^2(A)^n$ is zero, disproving a conjecture of Lott-Lück [**?**, Conjecture 7.2].

Acknowledgments. The author wants to thank the Mittag-Leffler institute for its hospitality during and the organizers for running the workshop *Growth and Mahler measures in geometry and topology* in July 2013. This paper is financially supported by the Leibniz-Preis of the author granted by the Deutsche Forschungsgemeinschaft DFG.

The paper is organized as follows: Contents

1. Some basic notions

Consider a non-zero element $p = p(z_1^{\pm 1}, \ldots, z_d^{\pm 1})$ in $\mathbb{C}[\mathbb{Z}^d] = \mathbb{C}[z_1^{\pm 1}, \ldots, z_d^{\pm 1}]$ for some integer $d \geq 1$.

1.1. **The width and the leading coefficient.** There are integers n_d^- and n_d^+ and elements $q_n(z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1})$ in $\mathbb{C}[\mathbb{Z}^{d-1}] = \mathbb{C}[z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1}]$ uniquely determined by the properties that

$$\begin{array}{rcl} n_d^- & \leq & n_d^+; \\ q_{n_d^-}(z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1}) & \neq & 0; \\ q_{n_d^+}(z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1}) & \neq & 0; \\ p(z_1^{\pm 1}, \ldots, z_d^{\pm 1}) & = & \displaystyle\sum_{n=n_d^-}^{n_d^+} q_n(z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1}) \cdot z_d^n. \end{array}$$

Define inductively elements $p_i(z_1^{\pm 1}, \ldots, z_{d-i}^{\pm 1})$ in $\mathbb{C}[\mathbb{Z}^{d-i}] = \mathbb{C}[z_1^{\pm 1}, \ldots, z_{d-i}^{\pm 1}]$ and integers $w_i(p) \geq 0$ for $i = 0, 1, 2, \ldots, d$ by

$$\begin{array}{rcl} p_0(z_1^{\pm 1}, \ldots, z_d^{\pm 1}) & := & p(z_1^{\pm 1}, \ldots, z_d^{\pm 1}); \\ p_1(z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1}) & := & q_{n_d^+}(z_1^{\pm 1}, \ldots, z_{d-1}^{\pm 1}); \\ p_i & := & (p_{i-1})_1 \quad \text{for } i = 1, 2 \ldots, d; \\ w_0(p) & := & n_d^+ - n_d^-; \\ w_i(p) & := & w_0(p_i) \quad \text{for } i = 1, 2 \ldots, d. \end{array}$$

Define the *width* of $p = p(z_1^{\pm 1}, \ldots, z_d^{\pm 1})$ to be

$$\operatorname{wd}(p) = \max\{w_0(p), w_1(p), \ldots, w_d(p)\}, \tag{1.1}$$

and the leading *coefficient of* p to be

$$\operatorname{lead}(p) \quad = \quad p_d. \tag{1.2}$$

Obviously we have

$$\operatorname{wd}(p) \geq \operatorname{wd}(p_1) \geq \operatorname{wd}(p_2) \geq \cdots \geq \operatorname{wd}(p_d) = 0;$$
$$\operatorname{lead}(p) = \operatorname{lead}(p_1) = \ldots = \operatorname{lead}(p_0) \neq 0.$$

Remark 1.3 (Leading coefficient)**.** The name "leading coefficient" comes from the following alternative definition. Equip $\mathbb{Z}^d$ with the lexicographical order, i.e., we put $(m_1, \ldots, m_d) < (n_1, \ldots, n_d)$, if $m_d < n_d$, or if $m_d = n_d$ and $m_{d-1} < n_{d-1}$, or if $m_d = n_d$, $m_{d-1} = n_{d-1}$ and $m_{d-2} < n_{d-2}$, or if $\ldots$, or if $m_i = n_i$ for $i = d, (d-1), \ldots, 2$ and $m_1 < n_1$. We can write p as a finite sum with complex coefficients $a_{n_1,\ldots,n_d}$

$$p(z_1^{\pm}, \ldots, z_d^{\pm}) = \sum_{(n_1,\ldots,n_d)\in\mathbb{Z}^d} a_{n_1,\ldots,n_d} \cdot z_1^{n_1} \cdot z_2^{n_2} \cdot \cdots \cdot z_d^{n_d}.$$

Let $(m_1, \ldots m_d) \in \mathbb{Z}^d$ be maximal with respect to the lexicographical order among those elements $(n_1, \ldots, n_d) \in \mathbb{Z}^d$ for which $a_{n_1,\ldots,n_d} \neq 0$. Then the leading coefficient of p is $a_{m_1,\ldots,m_d}$.

Example 1.4 ($d = 1$)**.** In the special case $d = 1$, we can write

$$p(z^{\pm 1}) = \sum_{n=n_-}^{n^+} a_n \cdot z^n$$

for integers n^- and n^+ with $n^- \leq n^+$ and complex numbers a_n with $a_{n^-} \neq 0$ and $a_{n^+} \neq 0$, and we get $\operatorname{wd}(p) = n^+ - n^-$ and $\operatorname{lead}(p) = a_{n^+}$.

Remark 1.5 (Dependence on the ordering of the variables)**.** Notice that p_i, $\operatorname{wd}(p)$ and $\operatorname{lead}(p)$ do depend on the ordering of the variables $z_1, \ldots, z_d$. For instance we get for $d = 2$, $p(z_1, z_2) = z_1^3 \cdot z_2 + 2 \cdot z_1 \cdot z_2^2 - 1$ and the element $q(z_1, z_2) = z_2^3 \cdot z_1 + 2 \cdot z_2 \cdot z_1^2 - 1$ obtained from p by interchanging z_1 and z_2

$$\begin{aligned}
\operatorname{wd}(p) &= 2;\\
p_1(z_1) &= 2 \cdot z_1;\\
\operatorname{lead}(p) &= 2;\\
\operatorname{wd}(q) &= 3;\\
q_1(z_1) &= z_1;\\
\operatorname{lead}(q) &= 1.
\end{aligned}$$

The same remark applies to the passage to the inverse of each variables, i.e., we get different values if we replace z_i by z_i^{-1} for some i or all i.

Notice that the Mahler measure does not change by these operations on p.

1.2. **The spectral density function.** If we consider p as an element in $\mathbb{C}[\mathbb{Z}^d]$, multiplication with p induces a bounded $\mathbb{Z}^d$-equivariant operator $r_p^{(2)} \colon L^2(\mathbb{Z}^d) \to L^2(\mathbb{Z}^d)$. We will denote by

$$F(p) \colon [0, \infty) \quad \to \quad [0, \infty) \tag{1.6}$$

its *spectral density function* in the sense of [**?**, Definition 2.1 on page 73]. In the special situation considered here, it can be computed in terms of volumes of subsets of the d-torus T^d equipped with its Haar measure, see [**?**, Example 2.6 on page 75]

$$F(p)(\lambda) \quad = \quad \operatorname{vol}\bigl(\{(z_1, \ldots, z_d) \in T^d \mid |p(z_1, \ldots, z_d)| \leq \lambda\}\bigr). \tag{1.7}$$

2. Mahler measures and Fuglede-Kadison determinants

The following theorem allows us to apply results about Fuglede-Kadison determinants which appear for instance in [**?**, Chapter 3] to Mahler measures.

Theorem 2.1 (Mahler measure and Fuglede-Kadison determinants over $\mathbb{Z}^d$)**.** *Consider a non-zero element $p = p(z_1^{\pm 1}, \ldots, z_d^{\pm 1})$ in $\mathbb{C}[\mathbb{Z}^d] = \mathbb{C}[z_1^{\pm 1}, \ldots, z_d^{\pm 1}]$ for some natural number d. It defines a bounded $\mathbb{Z}^d$-equivariant operator $r_p^{(2)} \colon L^2(\mathbb{Z}^d) \to L^2(\mathbb{Z}^d)$ by multiplication with p.*

Then the Fuglede-Kadison determinant $\det^{(2)}_{\mathcal{N}(\mathbb{Z})}(r_p^{(2)})$ of $r_p^{(2)}$ agrees with the Mahler measure $M(p)$ of p.

Proof. This follows from [**?**, Example 3.13 on page 128] since the volume of the set $\{(z_1, \ldots, z_d) \in T^d \mid p(z_1, \ldots, z_d) = 0\}$ is zero. $\square$

The relation between the Fuglede-Kadison determinant and the Mahler measures is also considered in [**?**] and [**?**].

3. The recipe

For $d \geq 1$ consider $p = p(z_1^{\pm 1}, \ldots, z_d^{\pm 1}) \in \mathbb{C}[\mathbb{Z}^d] = \mathbb{C}[z_1^{\pm 1}, \ldots, z_d^{\pm 1}]$. We can write

$$p(z_1^{\pm 1}, \ldots, z_d^{\pm 1}) = \sum_{(n_1, \ldots, n_d) \in \mathbb{Z}^d} a_{n_1, \ldots, n_d} \cdot z_1^{n_1} \cdot \cdots \cdot z_d^{n_d}.$$

Define

$$\begin{aligned} \overline{p} &:= \sum_{(n_1, \ldots, n_d) \in \mathbb{Z}^d} \overline{a_{n_1, \ldots, n_d}} \cdot z_1^{-n_1} \cdot \cdots \cdot z_d^{-n_d}; \\ ||p||_{L^1} &:= \sum_{(n_1, \ldots, n_d) \in \mathbb{Z}^d} |a_{n_1, \ldots, n_d}|; \\ \operatorname{tr}_{\mathbb{C}\mathbb{Z}^d}(p) &:= a_{0, \ldots, 0}. \end{aligned}$$

Choose $K \geq ||r_p^{(2)}||$, where $||r_p^{(2)}||$ is the operator norm of $r_p^{(2)} \colon L^2(\mathbb{Z}^d) \to L^2(\mathbb{Z}^d)$ which is the supremum of the set $\{|p(z_1, \ldots, z_d)| \mid (z_1, \ldots, z_d) \in T^d\}$. An example for K is $||p||_{L^1}$. Define

$$c(p, K)_k \quad := \quad \operatorname{tr}_{\mathbb{C}\mathbb{Z}^d}\left(\left(1 - K^{-2} \cdot p \cdot \overline{p}\right)^k\right) \in [0, 1). \tag{3.1}$$

Then we get for the logarithm of the Mahler measure of p

$$\ln(M(p)) = \ln(K) - \textstyle\sum_{k=1}^{\infty} \frac{c(p,K)_k}{2k}. \tag{3.2}$$

Let $\alpha(p)$ be the Novikov-Shubin invariant of p which is a rational number with $0 < \alpha(p) \leq 1$ or is ∞^+, see Section **??**. Then for any choice of real number $0 < \alpha < \alpha(p)$ there exists a constant C independent of k such that for all $k \geq 1$ we have

$$0 \leq \ln(K) - \ln(M(p)) - \textstyle\sum_{k=1}^{\infty} \frac{c(p,K)_k}{2k} \leq \frac{C}{\alpha \cdot L^{\alpha}}. \tag{3.3}$$

A possible choice for C is

$$C = \sup\{k^{\alpha} \cdot c_k(p, K) \mid k \geq 1\}, \tag{3.4}$$

where the supremum is finite since one knows $\lim_{k \to \infty} k^{\alpha} \cdot c_k(p, K) = 0$. All these claims above are proved in [**?**, Theorem 3.172 on page 195].

It remains to get a concrete estimate of the constant C in terms of p. This requires some preparation.

4. Uniform estimate on spectral density functions

The main result of this section is the following

Theorem 4.1 (Uniform spectral density estimate)**.** *Consider an element $p = p(z_1^{\pm 1}, \dots, z_d^{\pm 1})$ in $\mathbb{C}[\mathbb{Z}^d] = \mathbb{C}[z_1^{\pm 1}, \dots, z_d^{\pm 1}]$ with $\mathrm{wd}(p) \geq 1$.*

Then we get for its spectral density function

$$F(p)(\lambda) \leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot d \cdot \mathrm{wd}(p) \cdot \left(\frac{\lambda}{|\operatorname{lead}(p)|} \right)^{\frac{1}{d \cdot \mathrm{wd}(p)}} \quad \text{for } \lambda \in [0, \infty).$$

For the case $d = 1$ and p a monic polynomial, a similar estimate of the shape $F(p)(\lambda) \leq C_k \cdot \lambda^{\frac{1}{k-1}}$ can be found in [**?**, Theorem 1], where $k \geq 2$ is the number of non-zero coefficients, and the sequence of real numbers $(C_k)_{k \geq 2}$ is recursively defined and satisfies $C_k \geq k - 1$.

4.1. Degree one.

In this subsection we deal with Theorem **??** in the case $d = 1$.

We get from the Taylor expansion of $\cos(x)$ around 0 with the Lagrangian remainder term that for any $x \in \mathbb{R}$ there exists $\theta(x) \in [0, 1]$ such that

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{\cos(\theta(x) \cdot x)}{4!} \cdot x^4.$$

This implies for $x \neq 0$ and $|x| \leq 1/2$

$$\left| \frac{2 - 2\cos(x)}{x^2} - 1 \right| = \left| \frac{2 \cdot \cos(\theta(x) \cdot x)}{4!} \cdot x^2 \right| \leq \left| \frac{2 \cdot \cos(\theta(x) \cdot x)}{4!} \right| \cdot |x|^2 \leq \frac{1}{12} \cdot \frac{1}{4} = \frac{1}{48}.$$

Hence we get for $x \in [-1/2, 1/2]$

$$\frac{47}{48} \cdot x^2 \leq 2 - 2\cos(x). \tag{4.2}$$

Lemma 4.3. *For any complex number $a \in \mathbb{Z}$ we get for the spectral density function of $(z - a) \in \mathbb{C}[\mathbb{Z}] = \mathbb{C}[z, z^{-1}]$*

$$F(z - a)(\lambda) \leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot \lambda \quad \text{for } \lambda \in [0, \infty).$$

Proof. We compute using (**??**), where $r := |a|$.

$$\begin{aligned} F(z-a)(\lambda) &= \operatorname{vol}\{z \in S^1 \mid |z - a| \leq \lambda\} \\ &= \operatorname{vol}\{z \in S^1 \mid |z - r| \leq \lambda\} \\ &= \operatorname{vol}\{\phi \in [-1/2, 1/2] \mid |\cos(\phi) + i \sin(\phi) - r| \leq \lambda\} \\ &= \operatorname{vol}\{\phi \in [-1/2, 1/2] \mid |\cos(\phi) + i \sin(\phi) - r|^2 \leq \lambda^2\} \\ &= \operatorname{vol}\{\phi \in [-1/2, 1/2] \mid (\cos(\phi) - r)^2 + \sin(\phi)^2 \leq \lambda^2\} \\ &= \operatorname{vol}\{\phi \in [-1/2, 1/2] \mid r \cdot (2 - 2\cos(\phi) + (r - 1)^2 \leq \lambda^2\}. \end{aligned}$$

We estimate using (**??**) for $\phi \in [-1/2, 1/2]$

$$r \cdot (2 - 2\cos(\phi)) + (r - 1)^2 \geq r \cdot (2 - 2\cos(\phi)) \geq \frac{47}{48} \cdot \phi^2.$$

This implies for $\lambda \geq 0$

$$\begin{aligned}
F(z-a)(\lambda) &= \operatorname{vol}\{\phi \in [-1/2, 1/2] \mid r \cdot (2 - 2\cos(\phi) + (r-1)^2 \leq \lambda^2\} \\
&\leq \operatorname{vol}\{\phi \in [-1/2, 1/2] \mid \frac{47}{48} \cdot \phi^2 \leq \lambda^2\} \\
&= \operatorname{vol}\left\{\phi \in [-1/2, 1/2] \,\middle|\, |\phi| \leq \sqrt{\frac{48}{47}} \cdot \lambda\right\} \\
&\leq 2 \cdot \sqrt{\frac{48}{47}} \cdot \lambda \\
&= \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot \lambda.
\end{aligned}$$

□

Lemma 4.4. *Let $p(z) \in \mathbb{C}[\mathbb{Z}] = \mathbb{C}[z, z^{-1}]$ be an element with $\operatorname{wd}(p) \geq 1$. Then we get for its spectral density function*

$$F(p)(\lambda) \leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot \operatorname{wd}(p) \cdot \left(\frac{\lambda}{|\operatorname{lead}(p)|}\right)^{\frac{1}{\operatorname{wd}(p)}} \quad \textit{for } \lambda \in [0, \infty).$$

Proof. We can write $p(z)$ as a product

$$p(z) = \operatorname{lead}(p) \cdot z^k \cdot \prod_{i=1}^{r} (z - a_i)$$

for an integer $r \geq 0$, non-zero complex numbers $a_1, \ldots, a_r$ and an integer k.

Since for any polynomial p and complex number $c \neq 0$ we have for all $\lambda \in [0, \infty)$

$$F(c \cdot p)(\lambda) = F(p)\left(\frac{\lambda}{|c|}\right),$$

we can assume without loss of generality $\operatorname{lead}(p) = 1$. Since the width, the leading coefficient and the spectral density functions of $p(z)$ and $z^{-k} \cdot p(z)$ agree, we can assume without loss of generality $k = 0$, or equivalently, that $p(z)$ has the form

$$p(z) = \prod_{i=1}^{r} (z - a_i).$$

We proceed by induction over r. The case $r = \operatorname{wd}(p) = 1$ is taken care of by Lemma **??**. The induction step from $r - 1 \geq 1$ to r is done as follows.

Put $q(z) = \prod_{i=1}^{r-1} (z - a_i)$. Then $p(z) = q(z) \cdot (z - a_r)$. The following inequality for elements $q_1, q_2 \in \mathbb{C}[z, z^{-1}]$ and $s \in (0, 1)$ is a special case of [**?**, Lemma 2.13 (3) on page 78]

$$F(q_1 \cdot q_2)(\lambda) \leq F(q_1)(\lambda^{1-s}) + F(q_2)(\lambda^s). \tag{4.5}$$

We conclude from (**??**) applied to $p(z) = q(z) \cdot (z - a_r)$ in the special case $s = 1/r$

$$F(p)(\lambda) \leq F(q)(\lambda^{\frac{r-1}{r}}) + F(z - a_r)(\lambda^{1/r}).$$

We conclude from the induction hypothesis for $\lambda \in [0, \infty)$

$$\begin{aligned}
F(q)(\lambda) &\leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot (r-1) \cdot \lambda^{\frac{1}{r-1}}; \\
F(z - a_r)(\lambda) &\leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot \lambda.
\end{aligned}$$

This implies for $\lambda \in [0, \infty)$

$$\begin{aligned}
F(p)(\lambda) &\leq F(q)(\lambda^{\frac{r-1}{r}}) + F(z - a_r)(\lambda^{1/r}) \\
&\leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot (r-1) \cdot \left(\lambda^{\frac{r-1}{r}}\right)^{\frac{1}{r-1}} + \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot \lambda^{\frac{1}{r}} \\
&\leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot (r-1) \cdot \lambda^{\frac{1}{r}} + \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot \lambda^{\frac{1}{r}} \\
&= \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot r \cdot \lambda^{\frac{1}{r}}.
\end{aligned}$$

□

4.2. **The induction step.** Now we finish the proof of Theorem **??** by induction over d. The induction beginning $d = 1$ has been taken care of in Subsection **??**, the induction step from $d-1$ to $d \geq 2$ is done as follows.

Since $F(\lambda) \leq 1$, the claim is obviously true for $\frac{\lambda}{|\operatorname{lead}(p)|} \geq 1$. Hence we can assume in the sequel $\frac{\lambda}{|\operatorname{lead}(p)|} \leq 1$.

We conclude from (**??**) and Fubini's Theorem applied to $T^d = T^{d-1} \times S^1$, where χ_A denotes the characteristic function of a subset A and $p_1(z_1^{\pm}, \ldots, z_{d-1}^{\pm 1})$ has been defined in Subsection **??**

(4.6)

$$\begin{aligned}
&F(p)(\lambda) \\
&= \operatorname{vol}\big(\{(z_1, \ldots, z_d) \in T^d \mid |p(z_1, \ldots, z_d)| \leq \lambda\}\big) \\
&= \int_{T^d} \chi_{\{(z_1,\ldots,z_d)\in T^d \mid |p(z_1,\ldots,z_d)| \leq \lambda\}} \, d\mu_{T^n} \\
&= \int_{T^{d-1}} \left(\int_{S^1} \chi_{\{(z_1,\ldots,z_d)\in T^d \mid |p(z_1,\ldots,z_d)| \leq \lambda\}} \, d\mu_{S^1} \right) \, d\mu_{T^{d-1}} \\
&= \int_{T^{d-1}} \chi_{\{(z_1,\ldots,z_{d-1})\in T^{d-1} \mid |p_1(z_1,\ldots,z_{d-1})\leq |\operatorname{lead}(p)|^{1/d}\cdot\lambda^{(d-1)1/d}\}} \\
&\qquad \cdot \left(\int_{S^1} \chi_{\{(z_1,\ldots,z_d)\in T^d \mid |p(z_1,\ldots,z_d)| \leq \lambda\}} \, d\mu_{S^1} \right) \, d\mu_{T^{d-1}} \\
&\qquad + \int_{T^{d-1}} \chi_{\{(z_1,\ldots,z_{d-1})\in T^{d-1} \mid |p_1(z_1,\ldots,z_{d-1})>|\operatorname{lead}(p)|^{1/d}\cdot\lambda^{(d-1))/d}\}} \\
&\qquad\quad \cdot \left(\int_{S^1} \chi_{\{(z_1,\ldots,z_d)\in T^d \mid |p(z_1,\ldots,z_d)| \leq \lambda\}} \, d\mu_{S^1} \right) \, d\mu_{T^{d-1}} \\
&\leq \int_{T^{d-1}} \chi_{(z_1,\ldots,z_{d-1}) \mid |p_1(z_1,\ldots,z_{d-1})| \leq |\operatorname{lead}(p)|^{1/d}\cdot\lambda^{(d-1)1/d}\}} + \\
&\qquad \max \left\{ \int_{S^1} \chi_{\{(z_1,\ldots,z_d)\in T^d \mid |p(z_1,\ldots,z_d)| \leq \lambda\}} \, d\mu_{S^1} \,\middle|\, (z_1, \ldots, z_{d-1}) \in T^{d-1} \right. \\
&\qquad\qquad \left. \text{with } |p_1(z_1, \ldots, z_{d-1})| > |\operatorname{lead}(p)|^{1/d} \cdot \lambda^{(d-1)/d} \right\}.
\end{aligned}$$

We get from the induction hypothesis applied to $p_1(z_1, \ldots, z_{d-1})$ and (**??**) since $\frac{\lambda}{|\operatorname{lead}(p)|} \leq 1$, $\operatorname{wd}(p_1) \leq \operatorname{wd}(p)$ and $\operatorname{lead}(p) = \operatorname{lead}(p_1)$

(4.7)

$$\begin{aligned}
&\int_{T^{d-1}} \chi_{(z_1,\dots,z_{d-1})|\ |p_1(z_1,\dots,z_{d-1})|\le|\operatorname{lead}(p)|^{1/d}\cdot\lambda^{(d-1)1/d}\}}\\
&\quad = \int_{T^{d-1}} \chi_{(z_1,\dots,z_{d-1})|\ |p_1(z_1,\dots,z_{d-1})|\le|\operatorname{lead}(p_1)|^{1/d}\cdot\lambda^{(d-1)1/d}\}}\\
&\quad = F(p_1)\big(|\operatorname{lead}(p_1)|^{1/d}|\cdot\lambda^{(d-1)/d}\big)\\
&\quad \le \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot(d-1)\cdot\operatorname{wd}(p_1)\cdot\left(\frac{|\operatorname{lead}(p_1)|^{1/d}\cdot\lambda^{(d-1)/d}}{|\operatorname{lead}(p_1)|}\right)^{\frac{1}{(d-1)\cdot\operatorname{wd}(p_1)}}\\
&\quad = \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot(d-1)\cdot\operatorname{wd}(p_1)\cdot\left(\frac{\lambda}{|\operatorname{lead}(p_1)|}\right)^{\frac{1}{d\cdot\operatorname{wd}(p_1)}}\\
&\quad = \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot(d-1)\cdot\operatorname{wd}(p_1)\cdot\left(\frac{\lambda}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d\cdot\operatorname{wd}(p_1)}}\\
&\quad \le \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot(d-1)\cdot\operatorname{wd}(p)\cdot\left(\frac{\lambda}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d\cdot\operatorname{wd}(p_1)}}\\
&\quad \le \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot(d-1)\cdot\operatorname{wd}(p)\cdot\left(\frac{\lambda}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d\cdot\operatorname{wd}(p)}}.
\end{aligned}$$

Fix $(z_1,\dots,z_{d-1})\in T^{d-1}$ with $|p_1(z_1,\dots,z_{d-1})| > \operatorname{lead}(p)^{1/d}\cdot\lambda^{(d-1)/d}$. Consider the element $f(z_d^{\pm 1}) := p(z_1,\dots z_{d-1},z_d^{\pm})\in\mathbb{C}[z_d^{\pm}]$. It has the shape

$$f(z_d^{\pm}) = \sum_{n=n^-}^{n^+} q_n(z_1,\dots,z_{d-1})\cdot z_d^n.$$

The leading coefficient of $f(z_d^{\pm 1})$ is $p_1(z_1,\dots z_{d-1}) = q_{n_+}(z_1,\dots,z_{d-1})$. Hence we get from Lemma **??** applied to $f(z_d^{\pm 1})$ and (**??**) since $\frac{\lambda}{|\operatorname{lead}(p)|}\le 1$, $\operatorname{wd}(f)\le\operatorname{wd}(p)$ and $|\operatorname{lead}(f)| = |p_1(z_1,\dots z_{d-1}))| > |\operatorname{lead}(p)|^{1/d}\cdot\lambda^{(d-1)/d}$

(4.8)

$$\begin{aligned}
&\int_{S^1} \chi_{\{(z_1,\dots,z_d)\in T^d|\ |p(z_1,\dots,z_d)|\le\lambda\}}\ d\mu_{S^1}\\
&\quad = \int_{S^1} \chi_{\{z_d\in S^1|\ |f(z_d)|\le\lambda\}}\ d\mu_{S^1}\\
&\quad = \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot\operatorname{wd}(f)\cdot\left(\frac{\lambda}{\operatorname{lead}(f)}\right)^{\frac{1}{\operatorname{wd}(f)}}\\
&\quad \le \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot\operatorname{wd}(f)\cdot\left(\frac{\lambda}{\operatorname{lead}(p)^{1/d}\cdot\lambda^{(d-1)/d}}\right)^{\frac{1}{\operatorname{wd}(f)}}\\
&\quad = \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot\operatorname{wd}(f)\cdot\left(\frac{\lambda}{\operatorname{lead}(p)}\right)^{\frac{1}{d\cdot\operatorname{wd}(f)}}\\
&\quad \le \frac{8\cdot\sqrt{3}}{\sqrt{47}}\cdot\operatorname{wd}(p)\cdot\left(\frac{\lambda}{\operatorname{lead}(p)}\right)^{\frac{1}{d\cdot\operatorname{wd}(p)}}.
\end{aligned}$$

Combining (**??**), (**??**) and (**??**) yields for λ with $\frac{\lambda}{|\operatorname{lead}(p)|} \leq 1$

$$\begin{aligned} F(p)(\lambda) &\leq \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot (d-1) \cdot \operatorname{wd}(p) \cdot \left(\frac{\lambda}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d \cdot \operatorname{wd}(p)}} \\ &\qquad\qquad + \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot \operatorname{wd}(p) \cdot \left(\frac{\lambda}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d \cdot \operatorname{wd}(p)}} \\ &= \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot d \cdot \operatorname{wd}(p) \cdot \left(\frac{\lambda}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d \cdot \operatorname{wd}(p)}}. \end{aligned}$$

This finishes the proof of Theorem **??**.

4.3. **Positivity of Novikov-Shubin invariants.** For the definition and basic properties about Novikov-Shubin invariants we refer to [**?**, Chapter 2].

Theorem 4.9 (Positivity of the Novikov-Shubin invariants over $\mathbb{C}[\mathbb{Z}^d]$)**.** *Consider any natural number d and any matrix $A \in M_{m,n}(\mathbb{C}[\mathbb{Z}^d])$. Choose $I \subseteq \{1, 2, \dots, m\}$ and $J \subseteq \{1, 2, \dots, n\}$ of the same cardinality $|I| = |J|$ such that for the corresponding square submatrix $A[I, J]$ of A we have $\det_{\mathbb{C}[\mathbb{Z}^d]}(A[I, J]) \neq 0$ and for any other choice of subsets $I' \subseteq \{1, 2, \dots, m\}$ and $J' \subseteq \{1, 2, \dots, n\}$ with $|I'| = |J'|$ and $\det_{\mathbb{C}[\mathbb{Z}^d]}(A[I', J']) \neq 0$ we have $|I'| \leq |I|$. (Such a choice always exists.)*

Then the Novikov-Shubin invariant of the bounded $\mathbb{Z}^d$-equivariant operator $r_A^{(2)} \colon L^2(\mathbb{Z}^d)^m \to L^2(\mathbb{Z}^d)^n$ given by right multiplication with A satisfies

$$\alpha(r_A^{(2)}) \geq \frac{1}{d \cdot \operatorname{wd}\bigl(\det_{\mathbb{C}[\mathbb{Z}^d]}(A[I, J])\bigr)},$$

and is in particular positive.

Proof. We first treat the special case, where $m = n$ and $\det_{\mathbb{C}[\mathbb{Z}^d]}(A) \neq 0$. We get directly from Theorem **??**

$$\alpha\bigl(r^{(2)}_{\det_{\mathbb{C}[\mathbb{Z}^d]}(A)} \colon L^2(\mathbb{Z}^d) \to L^2(\mathbb{Z}^d)\bigr) \quad \geq \quad \frac{1}{d \cdot \operatorname{wd}\bigl(\det_{\mathbb{C}[\mathbb{Z}^d]}(A)\bigr)}.$$

We can find by Cramer's rule a matrix $B \in M_{m,n}(\mathbb{C}[\mathbb{Z}^d])$ with $AB = \det_{\mathbb{C}[\mathbb{Z}^d]}(A) \cdot I_n$. The kernel of $r_B^{(2)}$ is trivial by [**?**, Lemma 1.34 (1) on page 35]. We conclude from [**?**, Lemma 2.14 (2) on page 79 and Lemma 2.15 (1) on page 80] for the Novikov-Shubin invariant of $r_A^{(2)} \colon L^2(\mathbb{Z}^d)^n \to L^2(\mathbb{Z}^d)^n$

$$\begin{aligned} \alpha(r_A^{(2)}) &\geq \alpha(r_B^{(2)} \circ r_A^{(2)}) \\ &= \alpha(r_{AB}^{(2)}) \\ &= \alpha(r_{\det_{\mathbb{C}[\mathbb{Z}^d]}(A) \cdot I_n}) \\ &= \alpha(r_{\det_{\mathbb{C}[\mathbb{Z}^d]}(A)}). \end{aligned}$$

Hence the claim follows in the special case $m = n$ and $\det_{\mathbb{C}[\mathbb{Z}^d]}(A) \neq 0$.

Next we deal with the general case of a matrix $A \in M_{m,n}(\mathbb{C}[\mathbb{Z}^d])$. Recall that we have chosen $I \subseteq \{1, 2, \dots, m\}$ and $J \subseteq \{1, 2, \dots, n\}$ of the same cardinality $|I| = |J|$ such that for the corresponding square submatrix $A[I, J]$ of A we have $\det_{\mathbb{C}[\mathbb{Z}^d]}(A[I, J]) \neq 0$ and for any other choice of subsets $I' \subseteq \{1, 2, \dots m\}$ and $J' \subseteq \{1, 2, \dots n\}$ with $|I'| = |J'|$ and $\det_{\mathbb{C}[\mathbb{Z}^d]}(A[I', J']) \neq 0$ we have $|I'| \leq |I|$.

Put $k = |I| = |J|$. Let $i^{(2)} \colon L^2(\mathbb{Z}^d)^k \to L^2(\mathbb{Z}^d)^m$ be the inclusion corresponding to $I \subseteq \{1, 2, \dots, m\}$ and let $\operatorname{pr}^{(2)} \colon L^2(\mathbb{Z}^d)^n \to L^2(\mathbb{Z}^d)^k$ be the projection corresponding to $J \subseteq \{1, 2, \dots, n\}$. Then $r^{(2)}_{A[I,J]} \colon L^2(\mathbb{Z}^d)^k \to L^2(\mathbb{Z}^d)^k$ agrees with the composite

$$r^{(2)}_{A[I,J]}\colon L^2(\mathbb{Z}^d)^k \xrightarrow{i^{(2)}} L^2(\mathbb{Z}^d)^m \xrightarrow{r_A^{(2)}} L^2(\mathbb{Z}^d)^n \xrightarrow{\mathrm{pr}^{(2)}} L^2(\mathbb{Z}^d)^k.$$

Let $p^{(2)}\colon L^2(G)^m \to \overline{\ker(r_A^{(2)})^{\perp}}$ be the orthogonal projection onto $\overline{\ker(r_A^{(2)})^{\perp}} \subseteq L^2(G)^m$. Let $j^{(2)}\colon \overline{\operatorname{im}(r_A^{(2)})} \to L^2(G)^n$ be the inclusion of the closure $\overline{\operatorname{im}(r_A^{(2)})}$ of the image of $r_A^{(2)}$. Let $(r_A^{(2)})^{\perp}\colon \ker(r_A^{(2)})^{\perp} \to \overline{\operatorname{im}(r_A^{(2)})}$ be the $\mathbb{Z}^d$-equivariant bounded operator uniquely determined by

$$r_A^{(2)} \quad = \quad j^{(2)} \circ (r_A^{(2)})^{\perp} \circ p^{(2)}.$$

Let $\mathbb{C}[\mathbb{Z}^d]_{(0)}$ be the quotient field of $\mathbb{C}[\mathbb{Z}^d]$. The $\mathbb{C}[\mathbb{Z}^d]_{(0)}$-rank of the matrix $A \in M_{m,n}(\mathbb{C}[\mathbb{Z}^d]_{(0)})$ is equal to k. Therefore the dimension over $\mathbb{C}[\mathbb{Z}^d]_{(0)}$ of the image of $r_A\colon \mathbb{C}[\mathbb{Z}^d]_{(0)}^m \to \mathbb{C}[\mathbb{Z}^d]_{(0)}^n$ is k. Hence the von Neumann dimension of the closure of the image of $r_A^{(2)}\colon L^2(\mathbb{Z}^d)^m \to L^2(\mathbb{Z}^d)^n$ is k by [**?**, Lemma 1.34 (1) on page 35]. Again by [**?**, Lemma 1.34 (1) on page 35] we conclude that the von Neumann dimension of the kernel of $r^{(2)}_{A[I,J]} = \mathrm{pr}^{(2)} \circ r_A^{(2)} \circ i^{(2)}\colon L^2(\mathbb{Z})^k \to L^2(\mathbb{Z})^k$ is zero and the von Neumann dimension of the closure of its image is k. In particular

$$r^{(2)}_{A[I,J]} = \mathrm{pr}^{(2)} \circ r_A^{(2)} \circ i^{(2)} = \mathrm{pr}^{(2)} \circ j^{(2)} \circ (r_A^{(2)})^{\perp} \circ p^{(2)} \circ i^{(2)}\colon L^2(\mathbb{Z})^k \to L^2(\mathbb{Z})^k$$

is injective and hence dense image. This implies that $p^{(2)} \circ i^{(2)}\colon L^2(\mathbb{Z})^k \to \overline{\ker(r_A^{(2)})^{\perp}}$ is injective and $\mathrm{pr}^{(2)} \circ j^{(2)}\colon \overline{\operatorname{im}(r_A^{(2)})} \to L^2(\mathbb{Z}^d)^k$ has dense image. The morphism $(r_A^{(2)})^{\perp}\colon \ker(r_A^{(2)})^{\perp} \to \overline{\operatorname{im}(r_A^{(2)})}$ is by construction a weak isomorphism, i.e., has dense image and is injective. We conclude from the additivity of the von Neumann dimension, see [**?**, Theorem 1.12 (1) on page 21] that all three morphisms $p^{(2)} \circ i^{(2)}\colon L^2(\mathbb{Z})^k \to \ker(r_A^{(2)})^{\perp}$, $(r_A^{(2)})^{\perp}\colon \ker(r_A^{(2)})^{\perp} \to \overline{\operatorname{im}(r_A^{(2)})}$ and $\mathrm{pr}^{(2)} \circ j^{(2)} \to \overline{\operatorname{im}(r_A^{(2)})}\colon L^2(\mathbb{Z}^d)^k$ are weak isomorphisms. We conclude from [**?**, Lemma 2.11 (9) on page 77] and [**?**, Lemma 2.14 (2) on page 79]

$$\begin{aligned}
\alpha(r_A^{(2)}) &= \alpha\bigl((r_A^{(2)})^{\perp}\bigr)\\
&\geq \alpha\bigl(\mathrm{pr}^{(2)} \circ j^{(2)} \circ (r_A^{(2)})^{\perp} \circ p^{(2)} \circ i^{(2)}\bigr)\\
&= \alpha\bigl(\mathrm{pr}^{(2)} \circ r_A^{(2)} \circ i^{(2)}\bigr)\\
&= \alpha\bigl(r^{(2)}_{A[I,J]}\colon L^2(\mathbb{Z}^d)^k \to L^2(\mathbb{Z}^d)^k\bigr).
\end{aligned}$$

We get from the already proved special case applied to $A[I,J]$

$$\alpha\bigl(r^{2)}_{A[I,J]}\bigr) \geq \frac{1}{d \cdot \mathrm{wd}\bigl(\det_{\mathbb{C}[\mathbb{Z}^d]}(A[I,J])\bigr)}.$$

This finishes the proof of Theorem **??**. $\square$

It is known that the Novikov-Shubin invariants of a matrix over $\mathbb{Z}[\mathbb{Z}^d]$ is a rational numbers larger than zero. This follows from [**?**, Proposition 39 on page 494]. (The author of [**?**] informed us that his proof of this statement is correct when $d = 1$ but has a gap when $d > 1$. The nature of the gap is described in [**?**, page 16]. The proof in this case can be completed by the same basic method used in [**?**].)

In the case $d = 1$ the Novikov-Shubin invariant $\alpha(p)$ is explicitly known. Namely, we can write

$$p(z) = a_0 \cdot z^{r_0} \cdot \prod_{i=1}^{s} (z - a_i)^{r_i}$$

for $a_0 \in \mathbb{C}$ with $a_0 \neq 0$, $r_0 \in \mathbb{Z}$, $s \in \mathbb{Z}$ with $s \geq 0$, $a_i \in \mathbb{C}$ with $a_i \neq 0$ and $a_i \neq a_j$ for $i \neq j$, and $r_i \in \mathbb{Z}$ with $r_i \geq 1$. Then we get from [**?**, Lemma 2.58 on page 100]

$$(4.10)\quad \alpha(p) \quad := \quad \begin{cases} \min\left\{ \frac{1}{r_i} \;\middle|\; i = 1, 2, \ldots, s, |a_i| = 1 \right\} & \text{if p has a root on } S^1; \\ \infty^+ & \text{otherwise.} \end{cases}$$

Example 4.11 (Irreducible polynomial)**.** Let $p \in \mathbb{Q}[z]$ be an irreducible polynomial. Then all its roots have multiplicity 1. This implies

$$\alpha(p) \quad := \quad \begin{cases} 1 & \text{if p has a root on } S^1; \\ \infty^+ & \text{otherwise.} \end{cases}$$

So one can choose α in the recipe appearing in Section **??** to be any number $0 < \alpha < 1$ if p has a root on S^1 or to be any number $0 < \alpha$ otherwise. This is better than the choice of α as $\frac{1}{3 \cdot d \cdot \mathrm{wd}(p)}$ appearing in Theorem **??**. However, in Theorem **??** we do have an a priori estimate on the constant C and not only the expression (**??**).

5. Estimating the characteristic sequence

5.1. **The basic estimate.** Consider an element $p = p(z_1^{\pm 1}, \ldots, z_d^{\pm 1})$ in $\mathbb{C}[\mathbb{Z}^d] = \mathbb{C}[z_1^{\pm 1}, \ldots, z_d^{\pm 1}]$ with $\mathrm{wd}(p) \geq 1$. Let β and γ be real numbers satisfying

$$(5.1)\qquad 0 < \beta \cdot d \cdot \mathrm{wd}(p) < \gamma < 1.$$

Lemma 5.2. *Let K be a real number greater or equal to $||r_p^{(2)}||$, e.g., $K = ||p||_{L^1}$. Then we obtain for every natural number k with $k \geq 1$ the inequality*

$$\begin{aligned} 0 \quad &\leq \quad k^\beta \cdot c(p, K)_k \\ &\leq \quad \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot d \cdot \mathrm{wd}(p) \cdot \left(\frac{K^2}{|\,\mathrm{lead}(p)|} \right)^{\frac{1}{d \cdot \mathrm{wd}(p)}} \cdot k^{\beta - \frac{\gamma}{d \cdot \mathrm{wd}(p)}} + k^\beta \cdot (1 - k^{-\gamma})^k. \end{aligned}$$

Proof. Since $F(p)(0) = 0$, we conclude from [**?**, Lemma 3.179 on page 196] for $\lambda \in [0, 1]$.

$$0 \leq c(p, K)_k \leq F(p)(K^2 \cdot \lambda) + (1 - \lambda)^k.$$

If we put $\lambda = k^{-\gamma}$ and multiply with k^β, we obtain for any integer k with $k \geq 1$

$$0 \leq k^\beta \cdot c(p, K)_k \leq k^\beta \cdot \left(F(p)(K^2 \cdot k^{-\gamma}) + (1 - k^{-\gamma})^k \right).$$

Combining this with Theorem **??** yields for $k \geq 1$ the inequality

$$\begin{aligned} 0 \quad &\leq \quad k^\beta \cdot c(p, K)_k \\ &\leq \quad k^\beta \cdot \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot d \cdot \mathrm{wd}(p) \cdot \left(\frac{K^2 \cdot k^{-\gamma}}{|\,\mathrm{lead}(p)|} \right)^{\frac{1}{d \cdot \mathrm{wd}(p)}} + k^\beta \cdot (1 - k^{-\gamma})^k. \\ &= \quad \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot d \cdot \mathrm{wd}(p) \cdot \left(\frac{K^2}{|\,\mathrm{lead}(p)|} \right)^{\frac{1}{d \cdot \mathrm{wd}(p)}} \cdot k^{\beta - \frac{\gamma}{d \cdot \mathrm{wd}(p)}} + k^\beta \cdot (1 - k^{-\gamma})^k. \end{aligned}$$

□

We get using l'Hospital's rule

$$\begin{aligned} \lim_{x \to \infty} x \cdot \ln(1 - x^{-\gamma}) \quad &= \quad -\infty; \\ \lim_{x \to \infty} \frac{\ln(x)}{x \cdot \ln(1 - x^{-\gamma})} \quad &= \quad 0; \\ \lim_{x \to \infty} \left(\frac{\beta \cdot \ln(x)}{x \cdot \ln(1 - x^{-\gamma})} + 1 \right) \quad &= \quad 1; \\ \lim_{x \to \infty} \beta \cdot \ln(x) + x \cdot \ln(1 - x^{-\gamma}) \quad &= \quad -\infty; \\ \lim_{x \to \infty} x^\beta \cdot (1 - x^{-\gamma})^x \quad &= \quad 0. \end{aligned}$$

Hence can choose a real number $D(\beta,\gamma)$ such that

$$k^\beta \cdot (1-k^{-\gamma})^k \quad \leq \quad D(\beta,\gamma) \quad \text{for } k \in \mathbb{Z}, k \geq 1. \tag{5.3}$$

Since $\beta - \frac{\gamma}{d \cdot \operatorname{wd}(p)} < 0$, we have

$$k^{\beta - \frac{\gamma}{d\cdot \operatorname{wd}(p)}} \quad \leq \quad 1 \quad \text{for } k \geq 1. \tag{5.4}$$

We conclude from Lemma **??** together with (**??**) and (**??**)

Lemma 5.5. *Let K be an real number greater or equal to $||r_p^{(2)}||$, e.g., $K = ||p||_{L^1}$. Then we obtain for every natural number k with $k \geq 1$ the inequality*

$$\begin{aligned} 0 &\leq k^\beta \cdot c(p,K)_k \\ &\leq \frac{8\cdot\sqrt{3}}{\sqrt{47}} \cdot d \cdot \operatorname{wd}(p) \cdot \left(\frac{K^2}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d\cdot\operatorname{wd}(p)}} + D(\beta,\gamma). \end{aligned}$$

From now one we fix the choice

$$\begin{aligned} \gamma &= \frac{1}{2}; \\ \beta &= \frac{1}{3\cdot d\cdot \operatorname{wd}(p)}. \end{aligned}$$

We leave it to the reader to verify that we can arrange

$$D\left(\frac{1}{3\cdot d\cdot\operatorname{wd}(p)}, \frac{1}{2}\right) \quad \leq \quad e. \tag{5.6}$$

We conclude from Lemma **??** and (**??**)

Lemma 5.7. *Let p be a non-zero element in $\mathbb{C}[\mathbb{Z}^d]$. Let K be an real number greater or equal to $||r_p^{(2)}||$, e.g., $K = ||p||_{L^1}$. Then we obtain for every natural number k with $k \geq 1$ the inequality*

$$\begin{aligned} 0 &\leq k^{\frac{1}{3\cdot d\cdot\operatorname{wd}(p)}} \cdot c(p,K)_k \\ &\leq \frac{8\cdot\sqrt{3}}{\sqrt{47}} \cdot d \cdot \operatorname{wd}(p) \cdot \left(\frac{K^2}{|\operatorname{lead}(p)|}\right)^{\frac{1}{d\cdot\operatorname{wd}(p)}} + e. \end{aligned}$$

5.2. **Proof of Theorem ??.**

Proof. We conclude from Theorem **??** that the Novikov-Shubin invariant of $r_p^{(2)}$ satisfies

$$\alpha(r_p^{(2)}) \geq \frac{1}{d\cdot\operatorname{wd}(p)}.$$

With our choice $\beta = \frac{1}{3\cdot d\cdot\operatorname{wd}(p)}$, this implies $\beta < \alpha(r_p^{(2)})$. Put $K = ||p||_{L^1}$. We conclude from [**?**, Theorem 3.172 (5) on page 195] by inspecting its proof, see [**?**, page 200], that for any real number D satisfying

$$k^\beta \cdot c(p, ||p||_{L^1})_k \leq D \quad \text{for } k \geq 1,$$

we get for all $L \geq 1$ the inequality

$$0 \leq 2\cdot\ln(||p||_{L^1}) - 2\cdot\ln(M(p)) - \sum_{k=1}^{L} \frac{c(p,||p||_{L^1})_k}{k} \leq \frac{D}{\beta}\cdot L^{-\beta},$$

and hence

$$0 \leq \ln(||p||_{L^1}) - \ln(M(p)) - \sum_{k=1}^{L} \frac{c(p,||p||_{L^1})_k}{2k} \leq \frac{3\cdot d\cdot\operatorname{wd}(p)\cdot D}{2}\cdot L^{-\beta}.$$

Because of Lemma **??** we can choose

$$D = \frac{8 \cdot \sqrt{3}}{\sqrt{47}} \cdot d \cdot \mathrm{wd}(p) \cdot \left(\frac{||p||_{L^1}^2}{|\operatorname{lead}(p)|} \right)^{\frac{1}{d \cdot \mathrm{wd}(p)}} + e.$$

Since $C = \frac{3 \cdot d \cdot \mathrm{wd}(p) \cdot D}{2}$, we conclude or all $L \geq 1$

$$0 \leq \ln(||p||_{L^1}) - \ln(M(p)) - \sum_{k=1}^{\infty} \frac{c(p, ||p||_{L^1})_k}{2k} \leq C \cdot L^{-\beta}.$$

This finishes the proof of Theorem **??**. □

References

[1] D. Boyd, D. Lind, F. R. Villegas, and C. Deninger. The many aspects of Mahler's measure. Report about the workshop the many aspects of Mahler measure, April 2003, Birs, Banff, 2003.

[2] D. W. Boyd. Speculations concerning the range of Mahler's measure. *Canad. Math. Bull.*, 24(4):453–469, 1981.

[3] D. W. Boyd. Uniform approximation to Mahler's measure in several variables. *Canad. Math. Bull.*, 41(1):125–128, 1998.

[4] M. Carrizosa. Survey on Lehmer problems. *São Paulo J. Math. Sci.*, 3(2):317–327, 2009.

[5] O. T. Dasbach and M. N. Lalín. On the recurrence of coefficients in the Lück-Fuglede-Kadison determinant. In *Proceedings of the "Segundas Jornadas de Teoría de Números"*, Bibl. Rev. Mat. Iberoamericana, pages 119–134. Rev. Mat. Iberoamericana, Madrid, 2008.

[6] O. T. Dasbach and M. N. Lalin. Mahler measure under variations of the base group. *Forum Math.*, 21(4):621–637, 2009.

[7] B. Grabowksy, L.and Virág. Random walks on Lamplighters via random Schrödinger operators. Preprint, 2013.

[8] W. M. Lawton. A problem of Boyd concerning geometric means of polynomials. *J. Number Theory*, 16(3):356–362, 1983.

[9] J. Lott. Heat kernels on covering spaces and topological invariants. *J. Differential Geom.*, 35(2):471–510, 1992.

[10] J. Lott. Delocalized L^2-invariants. *J. Funct. Anal.*, 169(1):1–31, 1999.

[11] J. Lott and W. Lück. L^2-topological invariants of 3-manifolds. *Invent. Math.*, 120(1):15–60, 1995.

[12] W. Lück. *L^2-Invariants: Theory and Applications to Geometry and K-Theory*, volume 44 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 2002.

[13] C. Smyth. The Mahler measure of algebraic numbers: a survey. In *Number theory and polynomials*, volume 352 of *London Math. Soc. Lecture Note Ser.*, pages 322–349. Cambridge Univ. Press, Cambridge, 2008.

Mathematicians Institut der Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany

E-mail address: `wolfgang.lueck@him.uni-bonn.de`

URL: `http://www.him.uni-bonn.de/lueck`

ARITHMETIC GROUPS AND LEHMER'S CONJECTURE

KATHLEEN L. PETERSEN

1. Introduction

Arithmetic groups are a rich class of groups where connections between topology and number theory are showcased in a particularly striking way. One construction of these groups is motivated by the modular group, $\mathrm{PSL}_2(\mathbb{Z})$. The group of orientation preserving isometries of the hyperbolic upper half plane, $\mathbb{H}^2$, is isomorphic to $\mathrm{PSL}_2(\mathbb{R})$. Since $\mathbb{Z}$ is a discrete subgroup of $\mathbb{R}$ it follows that $\mathrm{PSL}_2(\mathbb{Z})$ is discrete in $\mathrm{PSL}_2(\mathbb{R})$. The modular group acts on $\mathbb{H}^2$ by linear fractional transformations, and the quotient $\mathbb{H}^2/\mathrm{PSL}_2(\mathbb{Z})$ is a finite volume hyperbolic orbifold.

The modular group has deep connections to many branches of mathematics and to number theory in particular. The modular group encodes the moduli space of elliptic curves. Modular forms, which are analytic functions on $\mathbb{H}^2$ satisfying a functional equation with respect to the modular group, have far-reaching connections between geometry, number theory, and analysis. In particular, Wiles' proof of the Taniyama Shimura conjecture (the modularity theorem) established a proof of Fermat's Last Theorem, one of the most famous conjectures of our time.

The geometry of the action of the modular group on $\mathbb{H}^2$ can also be used to provide a proof of Roth's theorem (the Thue-Siegel-Roth theorem). This theorem essentially says that an algebraic integer (which is not in $\mathbb{Z}$) does not have many 'good' rational approximations. Precisely, Roth's theorem says that if α is an irrational algebraic integer, then for any $\epsilon > 0$

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{2+\epsilon}}$$

has only finitely many solutions where $p, q \in \mathbb{Z}$ are co-prime.

Arithmetic groups are essentially subgroups of matrix groups defined over integer rings. For example, the groups $\mathrm{SL}_n(\mathcal{O}_K)$ are arithmetic, where $\mathcal{O}_K$ is the ring of integers of a number field K. The arithmetic groups we will concentrate on in the manuscript are a class of arithmetic groups which generalize the modular group, and act on (products of) hyperbolic spaces. There are many similarities between these arithmetic groups and the modular group, but there are also many difference. These differences showcase the dichotomy between lattices of low rank and higher-rank lattices. Some of this behavior can be seen algebraically, for example by the congruence subgroup property, and Kazhdan's property (T).

One interesting connection between the underlying number theory of these groups and the topology of their quotients is that through the distance formula, lengths of geodesics correspond traces of matrices. Because of the arithmeticity, these traces correspond to special kinds of algebraic integers. As we discuss in § 9, in the case of arithmetic Fuchsian groups, these algebraic integers are Salem numbers. We will outline a proof of the equivalence of the Salem conjecture and the short geodesic conjecture for arithmetic hyperbolic surfaces. See [11] for this and other connections between Salem numbers and geometry.

The purpose of this manuscript is to provide an introduction to this class of arithmetic groups motivated by the modular group, and outline the proof of this correspondence between the geodesic length and the Mahler measure.

2. The Modular Group

One way to deconstruct the modular group is as follows. From a geometric viewpoint, we wish to construct a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$; such a group will be discrete in $\mathrm{Isom}(\mathbb{H}^2)$ and act on $\mathbb{H}^2$ by linear fractional transformations. The quotient by this action will be an orbifold, a manifold with some well behaved singularities. (If the subgroup is torsion free, it will be a manifold.) We also want to ensure that the subgroup is large enough, so that the quotient has finite volume.

We begin with $M_2(\mathbb{Q})$, the 2×2 matrices with rational coefficients; the field $\mathbb{Q}$ introduces the arithmeticity since it is the quotient field of $\mathbb{Z}$. We then take $M_2(\mathbb{Z})$ which is discrete in $M_2(\mathbb{Q})$. We require a subgroup of $\mathrm{PSL}_2(\mathbb{R})$, so we take the norm one elements, $\mathrm{SL}_2(\mathbb{Z})$ and then projectivize. Happily, the resulting group $\mathrm{PSL}_2(\mathbb{Z})$ is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$; this follows from the fact that $M_2(\mathbb{Z})$ is discrete in $M_2(\mathbb{R})$, which is due to the discreteness of $\mathbb{Z}$ in $\mathbb{R}$. As a result, the quotient $\mathbb{H}^2/\mathrm{PSL}_2(\mathbb{Z})$ is a finite volume orbifold. We will generalize the construction

$$M_2(\mathbb{Q}) \dashrightarrow M_2(\mathbb{Z}) \dashrightarrow \mathrm{SL}_2(\mathbb{Z}) \dashrightarrow \mathrm{PSL}_2(\mathbb{Z})$$

to produce more discrete subgroups in $\mathrm{PSL}_2(\mathbb{R})$, and then further generalize this to produce discrete groups in $\mathrm{PSL}_2(\mathbb{C})$ and products of $\mathrm{PSL}_2(\mathbb{R})$ and $\mathrm{PSL}_2(\mathbb{C})$.

3. Quaternion Algebras

Let K be a number field with r_1 real places and r_2 complex places, so $[K : \mathbb{Q}] = r_1 + 2r_2$. We will label the real embeddings as $\sigma_1, \ldots, \sigma_{r_1}$ and the complex embeddings as $\tau_1, \overline{\tau_1}, \ldots, \tau_{r_2}, \overline{\tau_{r_2}}$. Let $\mathcal{O}_K$ be the ring of integers in K, elements in K which are roots of a monic polynomial in $\mathbb{Z}[x]$.

3.1. **Hilbert Symbols.** Let $\mathcal{Q}$ be a quaternion algebra over a field F. That is, $\mathcal{Q}$ is a four dimensional central simple algebra over F. If F does not have characteristic two, we can encode the data defining $\mathcal{Q}$ using a Hilbert symbol. We now assume that $\mathrm{char}(F) \neq 2$. For non-zero elements $a, b \in F$ the Hilbert symbol $\left(\frac{a,b}{F}\right)$ defines the quaternion algebra

$$\left(\frac{a,b}{F}\right) = \{r_1 + r_2 i + r_3 j + r_4 k : i^2 = a, j^2 = b, ij = -ji = k\}$$

where r_1, r_2, r_3, and r_4 are elements of F. It follows that $k^2 = -ab$. Using this notation, Hamilton's quaternions are

$$\mathcal{H} = \left(\frac{-1,-1}{\mathbb{R}}\right).$$

The Hilbert symbol $\left(\frac{1,1}{F}\right)$ defines a quaternion algebra isomorphic to $M_2(F)$ as can be seen by the map

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ i \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ j \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ k \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The Hamiltonians and $M_2(\mathbb{R})$ are not isomorphic. In particular, the Hamiltonians are a division algebra, but $M_2(\mathbb{R})$ has zero divisors; the non-zero elements of determinant zero are all zero divisors. It is a consequence of the Wedderburn-Artin theorem that a quaternion algebra over F is either isomorphic to $M_2(F)$ or is a division algebra. We say that $\mathcal{Q}$ is *ramified* if it is isomorphic to a division algebra, and *split* if it is isomorphic to a matrix algebra. Frobenius showed that the Hamiltonians are the only ramified quaternion algebra over $\mathbb{R}$.

Different Hilbert symbols often define isomorphic quaternion algebras. In particular, for $\mathcal{Q} = \left(\frac{a,b}{F}\right)$ and any non-zero $u \in F$,

$$\mathcal{Q} \cong \left(\frac{b,a}{F}\right) \cong \left(\frac{au^2,b}{F}\right) \cong \left(\frac{a,-ab}{F}\right).$$

The isomorphisms between the last three algebras and $\mathcal{Q}$ can be seen by switching the roles of i and j, by $i \mapsto iu^{-1}$ and $j \mapsto j$, and by the map $i \mapsto i$, $j \mapsto k$, respectively. This shows that over $\mathbb{R}$ a quaternion algebra is isomorphic to $\mathcal{H}$ exactly when a and b are negative, and that all quaternion algebras over $\mathbb{C}$ are isomorphic to $M_2(\mathbb{C})$.

If $[L : F] = 2$ one can often embed L as a quaternion algebra over F. For example, if $L = F(\sqrt{a})$ then $L \hookrightarrow \left(\frac{a,b}{F}\right)$ identifying i with $\sqrt{a}$.

3.2. **Norm and Trace.** For $q = r_1 + r_2 i + r_3 j + r_4 k \in \mathcal{Q} = \left(\frac{a,b}{F}\right)$, define the conjugate of q to be $\overline{q} = r_1 - r_2 i - r_3 j - r_4 k$. This is well-defined independent of the choice of basis since the center of $\mathcal{Q}$ is F. We define the *reduced norm* of q to be

$$n(q) = q \cdot \overline{q} = r_1^2 - ar_2^2 - br_3^2 + abr_4^2.$$

Similarly, the *reduced trace* is $t(q) = q + \overline{q}$. Let the superscript one denote elements of norm one. The norm is preserved by homomorphism, so if $\mathcal{Q} \cong M_2(F)$ then the image of $\mathcal{Q}^1$ is $\mathrm{SL}_2(F)$.

We extend this discussion to the following classification lemma.

Lemma 3.1. *For the quaternion algebra $\mathcal{Q} = \left(\frac{a,b}{F}\right)$, the following are equivalent:*

1. $\mathcal{Q} \cong \left(\frac{1,1}{F}\right) \cong M_2(F)$.
2. *$\mathcal{Q}$ is not a division algebra.*
3. *The quadratic form $ax^2 + by^2 = 1$ has a solution $(x, y) \in F \times F$.*

Proof. It suffices to show the equivalence of the third. An element $q \in \mathcal{Q}$ is invertible exactly when $n(q) \neq 0$. Consider $q_1 = r_1 + r_2 i + r_3 j + r_4 k$. If $r_1 \neq 0$ then letting $q_2 = b_2 i + b_3 j + b_4 k$ with

$$b_2 = b(r_1 r_4 + r_2 r_3),\ b_3 = a(r_2^2 - br_4^2),\ b_4 = (r_1 r_2 + br_3 r_4)$$

we see that if $n(q_1) = 0$ then $n(q_2) = 0$ as well. The norm of q_2 is

$$-ar_2^2 - br_3^2 + abr_4^2 = 0.$$

Therefore, $\mathcal{Q}$ is not a division algebra exactly when there is some non-zero element $0 \neq q = b_2 i + b_3 j + b_4 k$ with zero norm.

Assume that $\mathcal{Q}$ is not a division algebra. If any two of b_2, b_3, and b_4 are zero then $q = 0$. If $b_4 \neq 0$ then let $x = b_3/ab_4$, $y = b_2/ab_4$. If $b_4 = 0$ then let $x = (1 + a)/2a$, $y = b_3(1 - a)/2ab_2$. Hence we have solutions to $ax^2 + by^2 = 1$.

Assume there is a solution to $ax^2 + by^2 = 1$. If $x = 0$ then $b = c^2$ for some $c \in F$. Then $(c + j)(c - j) = 0$ and $\mathcal{Q}$ is not a division algebra. If $x \neq 0$ then $a + b(y/x)^2 = (1/x)^2$, so $a = (1/x)^2 - b(y - x)^2$ and it follows that the norm of $(1/x) + i + (y/x)j$ is zero, so $\mathcal{Q}$ is not a division algebra.

□

3.3. **Extension of Scalars.** In our construction, we begin with $\mathcal{Q} = \left(\frac{a,b}{K}\right)$ a quaternion algebra over a number field K. To create a Fuchsian group, the goal is to construct a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$, so we need a well behaved map to $M_2(\mathbb{R})$.

If $F \subset F'$ we can extend the scalars of $\mathcal{Q} = \left(\frac{a,b}{F}\right)$ by

$$\left(\frac{a,b}{F}\right) \otimes_F F' \cong \left(\frac{a,b}{F'}\right).$$

Similarly, if $\iota : F \to F'$ is an injection then we define the quaternion algebra

$$\mathcal{Q}^\iota = \left(\frac{\iota(a), \iota(b)}{\iota(F)}\right)$$

by

$$r_1 + r_2 i + r_3 j + r_4 k \mapsto \iota(r_1) + \iota(r_2)i' + \iota(r_3)j' + \iota(r_4)k'$$

where $1, i, j, k$ are the the basis elements for $\mathcal{Q}$ and $1, i', j', k'$ are the basis elements for $\mathcal{Q}^\iota$. If ν is a place of K with completion K_ν then we can extend scalars to K_ν as

$$\mathcal{Q}^\nu = \left(\frac{a,b}{K}\right) \otimes_K K_\nu \cong \left(\frac{a,b}{K_\nu}\right).$$

We say that $\mathcal{Q}$ is split at ν if $\mathcal{Q}^\nu$ is isomorphic to $M_2(K_\nu)$ and ramified if it is isomorphic to a division algebra. By the Hasse-Minkowski theorem a quaternion algebra $\mathcal{Q} = \left(\frac{a,b}{K}\right)$ is isomorphic to $M_2(K)$ if and only if for all places ν this extension by scalars is split. When ν is an infinite place, this extension of scalars is isomorphic to either $\left(\frac{a,b}{\mathbb{R}}\right)$ or $\left(\frac{a,b}{\mathbb{C}}\right)$. We will use split extensions to produce maps from $\mathcal{Q}$ to $M_2(\mathbb{R})$ or $M_2(\mathbb{C})$. For discreteness, we need to be mindful of the other infinite places of K. That is, if σ is an embedding of K into $\mathbb{R}$ we need to understand the ramification of $\mathcal{Q}^\sigma \cong \left(\frac{a,b}{\mathbb{R}}\right)$. (If τ is a complex embedding $\mathcal{Q}^\tau \cong \left(\frac{a,b}{\mathbb{C}}\right)$ is always split.)

Let $\mathcal{Q} = \left(\frac{a,b}{K}\right)$. In view of Lemma 3.1 if $K \subset L$ and $\mathbb{Q} \otimes_K L$ is split then there is an $(x, y) \in L \times L$ such that $ax^2 + by^2 = 1$. If $y = 0$ then a is a square, so we can explicitly see the map to $M_2(K)$ defined by

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \mapsto \sqrt{a}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad j \mapsto \begin{pmatrix} 1 & b \\ 0 & -1 \end{pmatrix}.$$

For any Galois automorphism σ, $\sigma(a)$ must also be a square. It follows that if $\sigma(K) \subset L$, then $\mathcal{Q}^\sigma \otimes_{\sigma(K)} L$ is also split and σ acts on the image of $\mathcal{Q}$ in $M_2(L)$, sending it to the image of $\mathcal{Q}^\sigma \subset M_2(L)$. Otherwise, if $y \neq 0$ then a map from $\mathcal{Q}$ to $M_2(L)$ can explicitly be given by

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \mapsto \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix}, \quad j \mapsto \begin{pmatrix} y^{-1} & -axy^{-1} \\ xy^{-1} & -y^{-1} \end{pmatrix}.$$

If $\mathcal{Q}^\sigma \otimes_{\sigma(K)} L$ is also split then there are x', y' in L such that $\sigma(a)(x')^2+\sigma(b)(y')^2 = 1$ then the map between matrix groups can be seen by

$$\begin{pmatrix} y^{-1} & -axy^{-1} \\ xy^{-1} & -y^{-1} \end{pmatrix} \mapsto \begin{pmatrix} (y')^{-1} & -\sigma(a)x'(y')^{-1} \\ x'(y')^{-1} & -(y')^{-1} \end{pmatrix}$$

and extending the map on K by the Galois automorphism. We will use ρ to denote such a map from $\mathcal{Q}$ to $M_2(K)$.

For a quaternion algebra $\mathcal{Q}$ over a number field K, the number of places (finite and infinite) where $\mathcal{Q}$ is ramified is even. Moreover, for any even subset of places of K there is a quaternion algebra that is ramified at this set. This quaternion algebra is unique up to isomorphism.

3.4. **Orders.** In the construction of the modular group, we chose $M_2(\mathbb{Z}) \subset M_2(\mathbb{Q})$ to ensure discreteness and to get a quotient of finite volume; we generalize this idea using orders. Let $\mathcal{Q}$ be a quaternion algebra over the number field K. For any vector space V over K an $\mathcal{O}_K$ *lattice* L in V is a finitely generated $\mathcal{O}_K$ module contained in V. It is *complete* if $L \otimes_{\mathcal{O}_K} K \cong V$. An *order* $\mathcal{O}$ in the quaternion algebra $\mathcal{Q}$ is a complete $\mathcal{O}_K$ lattice which is also a ring with unity. An order is called maximal if it is maximal with respect to inclusion. If $\mathcal{O}$ is an order in $\mathcal{Q}$ defined over K, then since it is a lattice if $\alpha \in \mathcal{O}$ then both $\mathrm{tr}(\alpha)$ and $n(\alpha)$ lie in $\mathcal{O}_K$. (See [16] Lemma 2.2.4 page 83, for example.)

In the construction of the modular group $K = \mathbb{Q}$, and $\mathcal{O}_K = \mathbb{Z}$. If $V = M_2(\mathbb{Q})$ then since $M_2(\mathbb{Z}) \otimes_{\mathbb{Q}} \mathbb{Q} \cong M_2(\mathbb{Q})$ it is a complete lattice and we conclude that $M_2(\mathbb{Z})$ is an order. Similarly, for any number field K, $M_2(\mathcal{O}_K)$ is an order in $M_2(K)$. The order $\mathcal{O}' = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}\left(\frac{1+i+j+k}{2}\right)$ is contained in $\mathcal{O} = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}k$ and so $\mathcal{O}'$ is not maximal. By the Skolem-Noether theorem two isomorphic orders in $\mathcal{Q}$ are conjugate. The number of conjugacy classes of maximal orders is finite and called the *type number* of $\mathcal{Q}$. The type number of $M_2(K)$ is finite and equals $|\mathrm{Cl}_K/\mathrm{Cl}_K^{(2)}|$ where Cl_K is the class number of K and $\mathrm{Cl}_k^{(2)}$ is the subgroup generated by squares. In the quaternion algebra $\mathcal{Q} = \left(\frac{-1,-11}{\mathbb{Q}}\right)$, define

$$\tau = \frac{-1 + \frac{i+k}{2}}{2}, \quad z = \frac{i+j}{2}$$

so that $\tau^3 = 1$ and $z^2 - z + 3 = 0$. The maximal orders

$$\mathcal{O}_\tau = \mathbb{Z}[\tau] + j\mathbb{Z}[\tau] \quad \text{and} \quad \mathcal{O}_z = \mathbb{Z}[z] + i\mathbb{Z}[z]$$

are not isomorphic. Regardless, the intersection of two maximal orders is an order, so we often focus on the order $\mathcal{O}_k[1, i, j, k]$.

4. Construction of Arithmetic Fuchsian Groups

Let $\mathcal{Q}$ quaternion algebra over a number field K with a maximal order $\mathcal{O}$. Assume that $\mathcal{Q}$ is split at at least one real embedding of K. (In fact, for ease we often call this the identity embedding.) Next, take the norm one elements of $\mathcal{O}$, $\mathcal{O}^1$. The set $\mathcal{O}^1$ is a maximal discrete group of norm one elements in our quaternion algebra, and the split place produces a mapping from this group to $\mathrm{SL}_2(\mathbb{R})$ as seen by the extension of scalars. It remains to show that the image is discrete in $\mathrm{SL}_2(\mathbb{R})$ and has finite co-area.

To ensure discreteness of the image, we impose the condition that all other infinite places are real and $\mathcal{Q}$ is ramified at these places. Recall that the standard isomorphism from $\mathcal{Q}$ to $M_2(\mathbb{R})$ is denoted as ρ. We will use ρ to denote this map restricted to $\mathcal{O}^1$ as well. We now sketch a proof that $\rho(\mathcal{O}^1)$ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{C})$. This will follow from the following two results.

Lemma 4.1. *The norm one elements in the Hamiltonians, $\mathcal{H}^1 = \left(\frac{-1-1}{\mathbb{R}}\right)^1$ are a compact set.*

Proof. The norm of $q = (r_1 + r_2 i + r_3 j + r_4 k) \in \mathcal{H}$ is

$$n(q) = q\bar{q} = r_1^2 + r_2^2 + r_3^2 + r_4^2.$$

It follows that $\mathcal{H}^1$ is isomorphic to S^3 and is compact. □

Lemma 4.2. *Let $C \subset \mathbb{C}$ be a compact set, and K a number field. Then there are only finitely many algebraic integers $\alpha \in \mathcal{O}_K$ such that α and all of its conjugates lie in C.*

Proof. If $\sigma_1, \ldots, \sigma_{r_1}$ are all real places of K and $\tau_1, \ldots, \tau_{r_2}$ are all complex places of K, then the injection $\phi : \mathcal{O}_K \to \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ defined by

$$\alpha \in K \mapsto (\sigma_1(\alpha), \ldots, \sigma_{r_1}(\alpha), \tau_1(\alpha), \ldots, \tau_{r_2}(\alpha))$$

sends $\mathcal{O}_K$ to a lattice. Any compact subset of $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ can contain only finitely many lattice points, and therefore its preimage under ϕ contains only finitely many integers α such that α and all of its conjugates are in the set.

□

The quaternion algebra $\mathcal{Q}$ is defined over the totally real number field K which is split at the identity embedding. Therefore, $\mathcal{Q} \otimes_K \mathbb{R}$ is isomorphic to $M_2(\mathbb{R})$. Call this isomorphism ρ. Consider a convergent sequence $\{q_n\} \subset \mathcal{O}^1 \subset \mathcal{Q}$. Under the mapping ρ we can assume that

$$\rho(q_n) \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

To show discreteness, it suffices to show that for n large enough the q_n are all equal. Let $q_n = r_{1,n} + r_{2,n} i + r_{3,n} j + r_{4,n} k$. The images $\rho(q_n)$ converge to the identity, and since ρ is a homomorphism, $q_n \to 1$, the identity in $\mathcal{Q}$. Therefore

$$r_{1,n} \to 1, \quad r_{2,n} \to 0, \quad r_{3,n} \to 0, \quad r_{4,n} \to 0,$$

and so there is an N_0 such that for all $n > N_0$ $r_{2,n}, r_{3,n}$, and $r_{4,n}$ are within ϵ of 0, and $r_{1,n}$ is within ϵ of 1.

The number field K is totally real, and the quaternion algebra $\mathcal{Q}$ is ramified at all non-identity places σ of K. Therefore $\mathcal{Q}^\sigma \otimes_K \mathbb{R} \cong \mathcal{H}$ for all of these places. It follows that if σ is a ramified real place the induced map takes $\mathcal{O}^1$ into $\mathcal{H}^1$, the norm one elements of the Hamiltonians. Since $\mathcal{H}^1$ is compact by Lemma 4.1, all of these conjugates of $r_{1,n}, r_{2,n}, r_{3,n}$, and $r_{4,n}$ are all bounded. The numbers $r_{1,n}, r_{2,n}, r_{3,n}$, and $r_{4,n}$ are bounded by the above discussion of the identity place. Therefore, discreteness follows by Lemma 4.2. For a general maximal order the values $r_{1,n}, r_{2,n}, r_{3,n}$, and $r_{4,n}$ may not be algebraic integers, but they are 'almost' algebraic integers and discreteness follows.

This construction yields what is called a Fuchsian group derived from a quaternion algebra. For a broader family of groups, we introduce the notion of commensurability.

5. Commensurability

If A and B are both subgroups of a group G we say that A and B are *commensurable* if the intersection $A \cap B$ has finite index in both A and B. We say that A and B are *commensurable in the wide sense* if a conjugate of A is commensurable with B. This parallels the notion of commensurability of manifolds (or orbifolds). Two manifolds M and N are commensurable if they share a finite sheeted cover.

Definition 5.1. A *Fuchsian group derived from a quaternion algebra* is a finite index subgroup of $P\rho(\mathcal{O}^1)$ where $\mathcal{O}$ is a maximal order in a quaternion algebra over a totally real number field which is unramified in exactly one place. An *arithmetic Fuchsian group* is a subgroup of $\mathrm{PSL}_2(\mathbb{R})$ which is commensurable (in the wide sense) to a Fuchsian group derived from a quaternion algebra.

There is a precise relationship between arithmetic and derived groups. Define $\Gamma^{(2)}$ as $\langle \gamma^2 : \gamma \in \Gamma \rangle$, so $\Gamma^{(2)}$ is a (finite index) subgroup of Γ. The group Γ is arithmetic if and only if $\Gamma^{(2)}$ is derived (see [16] Corollary 8.3.5).

As we have seen, $\mathrm{PSL}_2(\mathbb{Z})$ is an arithmetic Fuchsian group with torsion. For example $\pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has order two. Therefore, the quotient is an orbifold. By Selberg's Lemma, such an orbifold has a finite sheeted (branched) cover which is a manifold. That is, an arithmetic Fuchsian group with torsion has a finite index subgroup which is torsion free.

Example 5.2. Let K be the splitting field of the biquadratic polynomial $p(x) = x^4 - 5x^2 + 2$. Then $p(x)$ has four real roots,

$$\pm\sqrt{\frac{5 \pm \sqrt{17}}{2}}.$$

Consider the quaternion algebra

$$\mathcal{Q} = \left(\frac{\sqrt{5+\sqrt{17}} - 2, -1}{K} \right).$$

The integer -1 is fixed by all elements of the Galois group of K. The other conjugates of $\sqrt{5+\sqrt{17}}-2$ are $-\sqrt{5+\sqrt{17}}-2$, $-\sqrt{5-\sqrt{17}}-2$ and $\sqrt{5-\sqrt{17}}-2$. All four of these conjugates are real; $\sqrt{5+\sqrt{17}}-2$ is positive, but the other conjugates are negative. Therefore the quaternion algebra is split at the identity embedding, but is ramified at all three non-identity embeddings. It follows that if $\mathcal{O}$ is a maximal order, $P\rho(\mathcal{O}^1) \subset \mathrm{PSL}_2(\mathbb{C})$ is a Fuchsian group derived from a quaternion algebra. Specifically, $P\rho(\mathcal{O}_K[1,i,j,k])$ is such a group. In fact (see Theorem 2) this is a co-compact group.

6. Arithmetic Kleinian Groups

The construction of arithmetic Kleinian groups is very similar to the construction of the arithmetic Fuchsian groups. In this case, we begin with K, a number field with exactly one complex place. If $\mathcal{Q}$ is a quaternion algebra over K then $\mathcal{Q} \otimes_K \mathbb{C} \cong M_2(\mathbb{C})$. Denote this map to $M_2(\mathbb{C})$ as ρ as above. If $\mathcal{Q}$ is unramified at all real places, then for any (maximal) order $\mathcal{O} \subset \mathcal{Q}$ the proof that $\rho(\mathcal{O}^1)$ is a finite covolume discrete subgroup of $\mathrm{SL}_2(\mathbb{C})$ is analogous to the Fuchsian case.

Definition 6.1. A *Kleinian group derived from a quaternion algebra* is a finite index subgroup of $P\rho(\mathcal{O}^1)$ where $\mathcal{O}$ is a maximal order in a quaternion algebra over a number field with exactly one complex place which is ramified in all real places. An *arithmetic Kleinian group* is a subgroup of $\mathrm{PSL}_2(\mathbb{C})$ which is commensurable (in the wide sense) to a Kleinian group derived from a quaternion algebra.

The Bianchi groups are natural analogs of the modular group in the Kleinian setting. Let K be an imaginary quadratic number field. Then $\mathcal{O} = \mathcal{O}_K[1, i, j, k]$ is an order in the quaternion algebra $\left(\frac{1,1}{K}\right)$. Under the map ρ, $\rho(\mathcal{O}) \subset M_2(\mathcal{O}_K)$ and the image of the norm one elements is contained in $\mathrm{SL}_2(\mathcal{O}_K)$. As there is just one (complex) place and $M_2(\mathcal{O}_K)$ is unramified at the identity place we see that $\mathrm{SL}_2(\mathcal{O}_K)$ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{C})$. (In fact, discreteness directly follows from the fact that $\mathcal{O}_K$ is discrete is $\mathbb{C}$.) The groups $\mathrm{PSL}_2(\mathcal{O}_K)$ are called the Bianchi groups. The quotient $Q_K = \mathbb{H}^3/\mathrm{PSL}_2(\mathcal{O}_K)$ is a cusped hyperbolic 3-orbifold. Hurwitz showed that the number of cusps is equal to the class number of $\mathcal{O}_K$. The figure-8 knot complement can be realized as $\mathbb{H}^3/\Gamma$ where $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ is generated by

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -\omega \\ 0 & 1 \end{pmatrix}$$

with $\omega = \frac{1}{2}(-1+\sqrt{-3})$. The group Γ is an index 12 subgroup of the Bianchi group $\mathrm{PSL}_2(\mathcal{O}_{\mathbb{Q}(\sqrt{-3})})$. Reid [22] proved that the figure-8 is the only arithmetic knot complement (in S^3). Cuspidal cohomology computations show that any arithmetic link complement in S^3 must be of the form $\mathbb{H}^3/\Gamma$ where Γ is commensurable with the Bianchi group $\mathrm{PSL}_2(\mathcal{O}_{\mathbb{Q}(\sqrt{-d})})$ for

$$d \in \{1, 2, 3, 5, 6, 7, 11, 15, 19, 23, 31, 39, 47, 71\}.$$

The Whitehead link is arithmetic, and the fundamental group of the complement is a finite index subgroup of $\mathrm{PSL}_2(\mathcal{O}_{\mathbb{Q}(\sqrt{-1})})$. In fact, Baker [1] showed that all links are sub links of arithmetic links.

7. Properties of Arithmetic Fuchsian and Kleinian groups

An alternate definition of arithmetic groups, due to Margulis, is that a Kleinian group Γ is arithmetic if it has infinite index in its commensurator. The commensurator of Γ is

$$\mathrm{Comm}(\Gamma) = \{x \in \mathrm{PSL}_2(\mathbb{C}) : x^{-1}\Gamma x \text{ is commensurable with } \Gamma\}.$$

A similar statement is true for Fuchsian groups.

For any Fuchsian or Kleinian group Γ, the field $\mathbb{Q}(\mathrm{tr}(\gamma) : \gamma \in \Gamma)$ is a number field. In the arithmetic case, if $\Gamma = P\rho(\mathcal{O}^1)$ where $\mathcal{O}$ is an order in a quaternion algebra defined over K, then $\mathbb{Q}(\mathrm{tr}(\gamma)) = K$. In the general setting the *invariant trace field* $\mathbb{Q}(\mathrm{tr}^2(\gamma))$ is an invariant of the commensurability class. (For arithmetic groups these fields coincide.) For any Fuchsian or Kleinian group, one can construct a quaternion algebra as well. One distinguishing characteristic of the arithmetic Fuchsian and Kleinian groups is that all traces are algebraic integers, since they correspond to traces of elements in an order. In fact, a Kleinian group Γ is arithmetic if and only if the invariant trace field is a number field with one complex place, the traces are all algebraic integers, and the associated quaternion algebra is ramified at all real places. A similar statement is true in the Fuchsian case. (See [16] for details.)

If $\mathcal{O}$ is a maximal order, the co-area of the derived Fuchsian group $P\rho(\mathcal{O}^1)$, the area of $\mathbb{H}^2/P\rho(\mathcal{O}^1)$, is given by

$$\frac{8\pi\Delta_K^{\frac{3}{2}}\zeta_K(2)}{(4\pi^2)^{[K:\mathbb{Q}]}}\prod_{\mathcal{P}|\Delta(\mathcal{Q})}(N(\mathcal{P})-1)$$

where Δ_K is the absolute discriminant of K, $\Delta(\mathcal{Q})$ is the (reduced) discriminant of $\mathcal{Q}$, and ζ_K is the Dedekind zeta function of K [5]. Similar to the Fuchsian case, if $\mathcal{O}$ is a maximal order in the quaternion algebra $\mathcal{Q}$ over K, the co-volume of the derived group $P\rho(\mathcal{O}^1)$ is

$$\frac{4\pi^2|\Delta_K|^{\frac{3}{2}}\zeta_K(2)}{(4\pi^2)^{[K:\mathbb{Q}]}}\prod_{\mathcal{P}|\Delta(\mathcal{Q})}(N(\mathcal{P})-1).$$

7.1. **Co-compactness.** The modular group and the Bianchi groups are non-co-compact. That is, the quotients are non compact 2- and 3-orbifolds. One way to see this is that each contains the parabolic element $\pm\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, the image of the norm one element $1+i+j$. In fact, the commensurability classes containing these groups are precisely the non-co-compact arithmetic Fuchsian and Kleinian groups.

It is not difficult to determine which arithmetic Fuchsian and Kleinian groups are co-compact, and which are not. (See [16] Theorem 8.2.3.)

Theorem 1. Let Γ be an arithmetic Kleinian group commensurable with the derived Kleinian group $P\rho(\mathcal{O}^1)$, where $\mathcal{O}$ is an order in the quaternion algebra Q defined over K. Then the following are equivalent.

(1) Γ is non-cocompact.
(2) $K=\mathbb{Q}(\sqrt{-d})$ and $Q\cong M_2(K)$
(3) Γ is commensurable with a Bianchi group, $\mathrm{PSL}_2(\mathcal{O}_{\mathbb{Q}(\sqrt{-d})})$.

Proof. Since Γ is a Kleinian group, $G=P\rho(\mathcal{O}^1)$ is as well, and we must have that for $Q=\left(\frac{a,b}{K}\right)$ that K has exactly one complex place, τ, and if $\sigma_1,\dots,\sigma_{r_1}$ are the real places, $Q^{\sigma_\ell}\otimes_{\sigma_\ell(K)}\mathbb{R}\cong\mathcal{H}$.

Co-compactness is a commensurability invariant, so Γ is non-cocompact exactly when $G=P\rho(\mathcal{O}^1)$ is compact. Therefore, if γ is not co-compact G contains a parabolic element which is conjugate to some $\pm\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ for $x\neq 0$ and can be written as

$$\pm\Big(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}+x\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\Big).$$

Since the identity is in the quaternions algebra Q, and maps to the identity matrix, we see that up to an isomorphism, $x\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is in Q. This element has norm 0 and corresponds to a zero divisor. Therefore, Q is not a division algebra and must be isomorphic to $M_2(K)$. It follows that K has no real places. Therefore, 1 implies 2.

Assuming 2, notice that $M_2(\mathcal{O}_{\mathbb{Q}(\sqrt{-d})})$ is an order in $\mathcal{Q} = M_2(\mathbb{Q}(\sqrt{-d}))$. The intersection of a maximal order $\mathcal{O}$ with $M_2(\mathcal{O}_{\mathbb{Q}(\sqrt{-d})})$ is an order in $\mathcal{Q}$ and it follows that the subgroups containing norm one elements are commensurable.

Assuming 3, notice that any Bianchi group contains the element $\pm \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and therefore the quotient is not co-compact. Compactness is a commensurability invariant. Therefore, 3 implies 1.

□

Similarly, we have the following for Fuchsian groups.

Theorem 2. Let Γ be an arithmetic Fuchsian group commensurable with the derived Fuchsian group $P\rho(\mathcal{O}^1)$, where $\mathcal{O}$ is an order in the quaternion algebra Q defined over K. Then the following are equivalent.

(1) Γ is non-cocompact.
(2) $K = \mathbb{Q}$ and $Q \cong M_2(K)$
(3) Γ is commensurable with the Modular group, $\mathrm{PSL}_2(\mathbb{Z})$.

Example 7.1. Consider the quaternion algebra

$$\mathcal{Q} = \left(\frac{-t, t^2-7}{\mathbb{Q}(t)}\right)$$

where t is a root of $x^3 - 7$. The field $\mathbb{Q}(t)$ has one real place, corresponding to the real root $\sqrt[3]{7}$ and one complex place corresponding to the conjugate roots $\omega\sqrt[3]{7}$, and $\omega^2\sqrt[3]{7}$ where $\omega = (-1+\sqrt{-3})/2$ is a primitive third root of unity. Therefore $\mathcal{Q}$ is split at the complex place, but is ramified at the real place since $-\sqrt[3]{7}$ and $\sqrt[3]{7}-7$ are both negative. It follows that if $\mathcal{O}$ is a maximal order in $\mathcal{Q}$, $P\rho(\mathcal{O}^1)$ is a finite co-volume derived Kleinian group. By Theorem 1 this is a co-compact group.

8. General Construction

The construction of the derived Fuchsian and Kleinian groups are special cases of a more general construction. Let a and b be non-negative integers, with at least one positive. The product $[\mathbb{H}^2]^a \times [\mathbb{H}^3]^b$ carries a metric inherited from the metric on $\mathbb{H}^2$ and $\mathbb{H}^3$. It follows that the group

$$[\mathrm{PSL}_2(\mathbb{R})]^a \times [\mathrm{PSL}_2(\mathbb{C})]^b$$

is a subgroup of the group of orientation preserving isometries of $[\mathbb{H}^2]^a \times [\mathbb{H}^3]^b$.

Let K be a number field with r_1 real places and r_2 complex places. Let $\mathcal{Q}$ be a quaternion algebra over K. Let $\sigma_1, \ldots, \sigma_l$ be the real places where $\mathcal{Q}$ is unramified, and $\sigma_{l+1}, \ldots, \sigma_{r_1}$ be the real places where $\mathcal{Q}$ is ramified. Let $\tau_1, \ldots \tau_{r_2}$ be the complex places. Assume that there is some infinite place where $\mathcal{Q}$ is ramified. (This is called the *Eichler condition*. The quaternion algebras $\left(\frac{-1,-1}{K}\right)$ and $\left(\frac{\sqrt{2}-4,-1}{\mathbb{Q}(\sqrt{2})}\right)$ do not satisfy the Eicher condition, for example.) Then for $\ell = 1 \ldots l$

$$\mathcal{Q}^{\sigma_\ell} \otimes_{\sigma_\ell(K)} \mathbb{R} \cong M_2(\mathbb{R})$$

and for $\ell = l+1 \ldots r_1$

$$\mathcal{Q}^{\sigma_\ell} \otimes_{\sigma_\ell(K)} \mathbb{R} \cong \mathcal{H}$$

and for $\ell = 1 \ldots r_2$

$$\mathcal{Q}^{\sigma_\ell} \otimes_{\tau_\ell(K)} \mathbb{C} \cong M_2(\mathbb{C}).$$

Using the explicit maps in § 3.3 this gives a map $\rho : \mathcal{Q} \to M_2(\mathbb{R})^l \times M_2(\mathbb{C})^{r_2}$. Choose a maximal order $\mathcal{O}$ in $\mathcal{Q}$ and take the norm one elements, $\mathcal{O}^1$ in $\mathcal{O}$. The restriction of ρ is the map (which we will also call ρ)

$$\rho : \mathcal{O}^1 \to [\mathrm{SL}_2(\mathbb{R})]^l \times [\mathrm{SL}_2(\mathbb{C})]^{r_2}$$

defined in each coordinate by q mapping to the image in $M_2(\mathbb{R})$ or $M_2(\mathbb{C})$ by the above. The fact that the image is full rank (finite co-volume) follows from the fact that the order $\mathcal{O}$ in $\mathcal{Q}$ was chosen to be full rank. It remains to address discreteness. This is similar to the Fuchsian case.

Consider a convergent sequence $\{q_n\}_{n=1}^{\infty}$ where $q_n = r_{1,n} + r_{2,n}i + r_{3,n}j + r_{4,n}k \in \mathcal{O}^1$ and $\rho(q_n) = (M_{n,1}, \ldots, M_{n,l+r_2}) \in [\mathrm{SL}_2(\mathbb{R})]^l \times [\mathrm{SL}_2(\mathbb{C})]^{r_2}$. By composition, we may assume that this sequence converges to the product of identity matrices. Since the map to matrices is defined by sending each $(\mathcal{O}^1)^{\sigma_\ell}$ $(1 \le \ell \le l)$ to one $\mathrm{SL}_2(\mathbb{R})$ in the product, and each $(\mathcal{O}^1)^{\tau_\ell}$ $(1 \le \ell \le r_2)$ to one $\mathrm{SL}_2(\mathbb{C})$ in the product, we conclude that for each of these places, the image $M_{n,\ell} \in \mathrm{SL}_2(\mathbb{R})$ or $\mathrm{SL}_2(\mathbb{C})$ is converging to the identity matrix. That is, for all such embeddings ψ, we have

$$\psi(q_n) = \psi(r_{1,n}) + \psi(r_{2,n})i + \psi(r_{3,n})j + \psi(r_{4,n})k \to 1.$$

(These i, j, and k correspond to the basis elements for the quaternion algebra $\mathcal{Q}^{\psi}$.) We conclude that

$$\psi(r_{1,n}) \to 1, \quad \psi(r_{2,n}), \psi(r_{3,n}), \psi(r_{4,n}) \to 0.$$

Therefore, for all split places, the conjugate of $r_{1,n}$ is a bounded distance from 1 and the conjugates of $r_{2,n}$, $r_{3,n}$ and $r_{4,n}$ are a bounded distance from 0.

Now, consider a ramified (real) place σ. The extension of scalars of Q^{σ} is isomorphic to $\mathcal{H}$. Under this identification, the elements of norm one in Q^{σ} map to $\mathcal{H}^1$, the norm one elements of the Hamiltonians. This set is compact by Lemma 4.1. We conclude that the group $P\rho(\mathcal{O}^1)$ is discrete by Lemma 4.2.

Example 8.1. Let K be the splitting field of the biquadratic polynomial $p(x) = x^4 - 5x^2 + 4$. Then the roots of $p(x)$ can be determined by the quadratic formula and are

$$\pm\sqrt{\frac{5 \pm \sqrt{41}}{2}}.$$

This has two real roots, $\pm\sqrt{\frac{5+\sqrt{41}}{2}}$ and two complex conjugate roots, $\pm\sqrt{\frac{5-\sqrt{41}}{2}}$. Any quaternion algebra is split at the complex place.

The quaternion algebra $\mathcal{Q}_1 = \left(\frac{1,1}{K}\right) \cong M_2(K)$ and order $\mathcal{O} = M_2(\mathcal{O}_K)$ correspond to $\mathrm{PSL}_2(\mathcal{O}_K)$ embedded in $[\mathrm{PSL}_2(\mathbb{R})]^2 \times [\mathrm{PSL}_2(\mathbb{C})]$.

Alternately, the quaternion algebra $\mathcal{Q}_2 = \left(\frac{-1,-1}{K}\right)$ is ramified at both real places. Therefore, if $\mathcal{O}$ is a maximal order, $P\rho(\mathcal{O}^1)$ is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. In fact, since K is not a quadratic number field, we conclude that this group is cocompact.

Now, Consider the quaternion algebra $\mathcal{Q}_3 = \left(\frac{\sqrt{41}-7,\sqrt{41}}{K}\right)$. At the identity embedding, the extension of scalars gives $\left(\frac{\sqrt{41}-7,\sqrt{41}}{\mathbb{R}}\right)$. The automorphism corresponding to the other real place, σ_2, sends $\sqrt{41}$ to $-\sqrt{41}$. Therefore the extension of scalars corresponding to this place is

$$\left(\frac{\sigma(\sqrt{41}-7),\sigma(\sqrt{41})}{\mathbb{R}}\right)=\left(\frac{-\sqrt{41}-7,-\sqrt{41}}{\mathbb{R}}\right).$$

The quaternion algebra is ramified at σ_2 and split at the identity, σ_1. As a result, if $\mathcal{O}$ is a maximal order in $\mathcal{Q}_3$ then $P\rho(\mathcal{O}^1)$ is a co-compact, finite co-volume discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})\times\mathrm{PSL}_2(\mathbb{C})$.

8.1. **The groups $\mathbf{PSL}_2(\mathcal{O}_K)$.** The simplest examples of this construction are the groups $\mathrm{PSL}_2(\mathcal{O}_K)$ which correspond to the order $M_2(\mathcal{O}_K)$ in the quaternion algebra $M_2(K)$. If K has r_1 real places and r_2 complex places then the extension by scalars corresponding to each place is split. We obtain the mapping

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \to \prod_{\psi} \begin{pmatrix} \psi(a) & \psi(b) \\ \psi(c) & \psi(d) \end{pmatrix}$$

where the product is over all infinite places of K.

For any number field K with r_1 real places and r_2 complex places, the quotient

$$[\mathrm{PSL}_2(\mathbb{R})]^{r_1}\times[\mathrm{PSL}_2(\mathbb{C})]^{r_2}/\mathrm{PSL}_2(\mathcal{O}_K)$$

is not co-compact. This is evident as $\pm\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\in\mathrm{PSL}_2(\mathcal{O}_K)$. The quotient has a finite number of finite volume topological ends, called cusps. Each cusp cross section is a Euclidean co-dimension one manifold. It is not difficult to show that the number of cusps of $\mathrm{PSL}_2(\mathcal{O}_K)$ equals the class number of $\mathcal{O}_K$.

To see this, first notice that the cusps are equivalence classes of elements of $(K\cup\infty)\subset\mathbb{C}$ under the action of $\mathrm{PSL}_2(\mathcal{O}_K)$. (Consider $K\cup\infty$ corresponding to the identity place of K in the product.) Two elements $p_1=\alpha_1/\beta_1$ and $p_2=\alpha_2/\beta_2$ are equivalent if there is a $M\in\mathrm{SL}_2(\mathcal{O}_K)$ such that $M(p_1)=p_2$.

The ideals (α_1,β_1) and (α_2,β_2) are equivalent in the class group if there is a $\gamma\in K$ such that

$$(\alpha_1,\beta_1)=(\gamma)(\alpha_2,\beta_2),$$

so $(\alpha_1,\beta_1)=(\gamma\alpha_2,\gamma\beta_2)=I$. The whole number ring is $\mathcal{O}_K=II^{-1}$, and so $1=\alpha_iI^{-1}+\beta_iI^{-1}$. That is, there is are elements s_i and t_i in I^{-1} such that

$$\alpha_is_i-\beta_it_i=1.$$

It follows that with

$$M_i=\begin{pmatrix} \alpha_i & t_i \\ \beta_i & s_i \end{pmatrix}$$

$M_i(\infty)=p_i$. The matrix $\pm M_2M_1^{-1}\in\mathrm{PSL}_2(\mathcal{O}_K)$ and takes p_1 to p_2. Conversely, if the two ideals are in different elements of the class group no such matrix exists. (See [27].)

There are some striking differences between the groups $\mathrm{PSL}_2(\mathcal{O}_K)$ when K is neither $\mathbb{Q}$ nor an imaginary quadratic number field and the modular group and Bianchi groups. This is a specific manifestation of the difference between higher rank arithmetic groups and lower rank groups. For these groups, this difference can be tied to the existence of infinitely many units in $\mathcal{O}_K$; by Dirichlet's unit Theorem, the rank of the unit group of $\mathcal{O}_K$ is r_1+r_2-1. One such difference involves the subgroup structure of these groups. Let I be a non-zero ideal of $\mathcal{O}_K$. The reduction modulo I map defines a map from $\mathrm{PSL}_2(\mathcal{O}_K)$ to the finite group

$\mathrm{PSL}_2(\mathcal{O}_K/I)$. The kernel of this map is called the principal congruence subgroup of level I, and is denoted $\Gamma(I)$. These are finite index subgroups of $\mathrm{PSL}_2(\mathcal{O}_K)$. A finite index subgroup of $\mathrm{PSL}_2(\mathcal{O}_K)$ is called a congruence subgroup if it contains some principal congruence subgroup. The group $\mathrm{PSL}_2(\mathcal{O}_K)$ is said to have the congruence subgroup property (CSP) if all finite index subgroups are congruence subgroups. Fricke [10] and Pick [21] showed that there are finite index subgroups of the Modular group which are not congruence subgroups. Serre [25] showed that $\mathrm{PSL}_2(\mathcal{O}_K)$ has the CSP precisely when K is not $\mathbb{Q}$ or an imaginary quadratic. In fact, this difference between the subgroup structure of $\mathrm{PSL}_2(\mathcal{O}_K)$ depending on whether K has positive unit rank can be seen topologically by looking at minimally cusped quotients [20, 18, 19]

9. Lehmer's Conjecture and Geodesics

9.1. **Geodesics and Systoles.** One natural way to measure a manifold is by the lengths of its geodesics. The *length spectrum* of a manifold is the collection of lengths of all closed geodesics, including multiplicities. For non-cocompact manifolds we consider only the lengths of non-boundary parallel curves as the length of a boundary parallel curve is not well-defined. In some sense, this is akin to studying a number field by its zeta function, which encodes the norms of all ideals. If M is a hyperbolic 3-manifold one uses the set of complex lengths (complex numbers encoding lengths and rotations for loxodromic elements). Surprisingly, there are isospectral manifolds which are not isometric [28], similar to the existence of number fields with the same zeta function. For arithmetic hyperbolic 2- or 3- manifolds, isospectrality is known to imply commensurability [23].

For a compact Riemannian manifold the spectrum of the Laplacian consists of the eigenvalues of the Laplace operator. For hyperbolic surfaces, via the Selberg trace formula (see [13, 14]), this spectrum and the length spectrum encode the same data (see [9]). One conjecture in this direction is Selberg's eigenvalue conjecture which states that the first non-zero eigenvalue of a principal congruence subgroup (the kernel of the modulo n map) of the modular group is bounded by 1/4 [24].

The smallest non-zero term in the length spectrum corresponds to the length of the shortest geodesic, the *systole*. The length of the systole is connected to the overall geometry of the manifold. Notably, Gromov [12] showed that in each dimension n there is a universal constant C_n such that for any Riemannian n-manifold M

$$\mathrm{length}(\mathrm{systole}(M)) \leq C_n \mathrm{volume}(M).$$

For general hyperbolic 2-manifolds it is not difficult to see that there are hyperbolic surfaces where the length of the shortest geodesic gets arbitrarily small, by constructing bar bell surfaces, for example. In fact, one can do this for any genus. However, it is conjectured that this length is universally bounded away from zero for the arithmetic Fuchsian and Kleinian groups.

Conjecture 9.1 (Short Geodesic Conjecture)**.** *The length of any geodesic in an arithmetic hyperbolic 2- or 3-manifold is universally bounded away from zero.*

It is not difficult to see that this is true for the non-cocompact groups. If Γ is a non-compact arithmetic hyperbolic 2- or 3-manifold it is enough to bound the length of the systole of $\Gamma^{(2)} < \Gamma$ which is derived. Therefore $\Gamma^{(2)}$ is either a subgroup of the modular group or a Bianchi group. As outlined below, the length of a geodesic corresponds to the Mahler measure of the trace. In this case the trace is an algebraic integer in $\mathbb{Q}$ or an imaginary quadratic number field. Using Dobrowolski's bound [8] (for example) these Mahler measures are bounded, and so is the systole length. One can obtain sharper results (see [16] Theorem 12.3.6) by direct computation and show, for example, that if $\mathbb{H}^3/\Gamma$ is a cusped hyperbolic 3-manifold with a systole of length less that 0.431277313 then Γ is not arithmetic.

9.2. **Lehmer's Conjecture.** Let α be an algebraic number with minimal polynomial

$$p(x) = a(x - r_1) \ldots (x - r_n).$$

The Mahler measure of α is

$$M(\alpha) = \frac{1}{2\pi} \int_0^{2\pi} \log |p(e^{i\theta})| \, d\theta = |a| \prod_{i=1}^{n} \max\{1, |r_i|\}.$$

As the Mahler measure is an invariant of the polynomial, we often refer to the Mahler measure of a polynomial in $\mathbb{Z}[x]$ as the Mahler measure of any of its roots and we write $M(p)$. It is elementary to see that the Mahler measure of any product of cyclotomic polynomials is one. Conversely, Kronecker showed that any monic polynomial in $\mathbb{Z}[x]$ all of whose roots lie on or inside the unit circle must be a product of cyclotomics and factors of x.

In 1933 Lehmer [15] asked whether there was a universal bound $\mu > 1$ such that if $p(x) \in \mathbb{Z}[x]$ is not a product of cyclotomics, then $M(p) > \mu$. This is often called Lehmer's conjecture, or Lehmer's question. The polynomial with the smallest known Mahler measure bigger than one was discovered by Lehmer. It is known as Lehmer's polynomial and is

$$l(x) = x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1.$$

The Mahler measure of Lehmer's polynomial is

$$M(l) = 1.176280818 \ldots.$$

(In fact, one can construct many polynomials with this Mahler measure.) A strong version of Lehmer's conjecture is that this is the smallest Mahler measure amongst polynomials in $\mathbb{Z}[x]$ which are not products of cyclotomics and powers of x.

There are bounds for the Mahler measure which depend on the degree of the polynomial (see the papers by Blanksby and Montgomery [4] and Dobrowolski [8]). So, if Lehmer's conjecture is not true, then the degrees of the polynomials with small Mahler measure must increase. Additionally, Lehmer's conjecture has been proven for certain special types of polynomials. Smyth [26] showed that Lehmer's conjecture is true for non-reciprocal polynomials. Reciprocal polynomials are those whose coefficients read the same forwards as backwards; a polynomial is reciprocal when if r is a root then $1/r$ is also a root. Borwein, Dobrowolski, and Mossinghoff [6] showed that the conjecture holds for a large class of polynomials which includes the Littlewood polynomials (those with coefficients in $\{-1, 1\}$). (See also, [7], [3], and [2].)

We say that a monic irreducible polynomial $p(x) \in \mathbb{Z}[x]$ is a *Salem polynomial* if all but two roots of p lie off the unit circle, and these roots are real numbers r and $1/r$. Additionally, if $r > 1$ is a root of a Salem polynomial, we call r a Salem number. For the purposes of this note, we will call a monic irreducible polynomial $p(x) \in \mathbb{Z}[x]$ a *complex Salem number* if exactly four roots of p are off the unit circle and these roots are complex numbers of the form $z, 1/z, \overline{z}$, and $1/\overline{z}$. We will call the numbers $z, 1/z, \overline{z}$, and $1/\overline{z}$ complex Salem numbers. The Salem conjecture asserts that the Mahler measure of any Salem polynomial is uniformly bounded away from 1. In some sense, this is the simplest case of Lehmer's conjecture. A complex Salem conjecture can be formulated similarly.

9.3. **Lengths and Mahler Measure.** A geodesic in $M = \mathbb{H}^2/\Gamma$ corresponds to a hyperbolic element $\gamma \in \Gamma$ since the axis of a hyperbolic element in Γ projects to a geodesic in $\mathbb{H}^2/\Gamma$, and every non-peripheral closed curve is freely homotopic to a unique closed geodesic corresponding to one of these axes. Up to conjugation,

$$\gamma^{\pm 1} = \pm \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$$

with $\lambda > 1$ so that $\pm\mathrm{tr}(\gamma) = \lambda + \lambda^{-1}$. It is a straightforward application of the hyperbolic distance formula that the translation length of γ, length(γ) is related to λ by

$$\mathrm{length}(\gamma) = 2 \log |\lambda|$$

so that

$$\cosh(\tfrac{1}{2}\mathrm{length}(\gamma)) = \tfrac{1}{2}|\lambda + \lambda^{-1}|.$$

It follows that the length of the geodesic is bounded away from zero if and only if the (absolute value of the) trace of γ is bounded away from two. The 3-dimensional case is similar, using complex length.

Now we establish a correspondence between short geodesics and the Salem conjecture, due to Neumann and Reid [17].

Theorem 3. The short geodesic conjecture for arithmetic hyperbolic 2-manifolds is equivalent to Salem's conjecture. The short geodesic conjecture for arithmetic hyperbolic 3-manifolds is equivalent to the complex Salem conjecture.

We sketch a proof Theorem 3 in the Fuchsian case. We refer the reader to [16] for a detailed treatment, especially in the Kleinian case.

We reduce to the case where $M = \mathbb{H}^2/\Gamma$ and Γ is derived, since if Γ_1 is arithmetic then $\Gamma_1^{(2)}$ is a finite index subgroup of a derived group. First we show that lengths of geodesics correspond to Salem numbers.

Claim 9.2. *Let Γ be a derived Fuchsian group, and let $\gamma \in \Gamma$ be a hyperbolic element. Then $|tr(\gamma)| = \lambda + \lambda^{-1}$ where $\lambda > 1$ is a Salem number.*

Proof. (sketch) By construction, $\Gamma = P\rho(\mathcal{O}^1)$ where $\mathcal{O}$ is an order in the quaternion algebra $\mathcal{Q} = \left(\frac{a,b}{K}\right)$ and K is a totally real number field and$\mathcal{Q}$ is ramified at all non-identity (real) places. Let $\gamma \in \Gamma$ with $\mathrm{tr}(\gamma) = \lambda + \lambda^{-1}$. Let

$$p(x) = x^2 - (\lambda + \lambda^{-1})x + 1.$$

The element $(\lambda + \lambda^{-1}) \in \mathcal{O}_K$ because $|\lambda + \lambda^{-1}| = \pm\mathrm{tr}(\gamma)$ and corresponds to $\mathrm{tr}(\alpha)$ for some element $\alpha \in \mathcal{O}$ and therefore lies in $\mathcal{O}_K$ as remarked on earlier. Moreover, λ^{-1} is a conjugate of λ since are both roots of the polynomial $p(x) \in \mathcal{O}_K[x]$. Let L denote the quadratic extension of K determined by $p(x)$ so that $\lambda, \lambda^{-1} \in L$.

Let ψ be a non-trivial Galois automorphism of K; ψ extends to automorphisms of L. The automorphisms of L corresponding to the identity place are the identity and the map that exchanges λ and λ^{-1}. Since K is real and $\lambda + \lambda^{-1} \in K$ we conclude that λ is either real or on the unit circle. But γ is hyperbolic and so $|\mathrm{tr}(\lambda + \lambda^{-1})| > 2$, ensuring that λ is real and not on the unit circle.

If ψ is a non-identity automorphism, then ψ induces a map from $\mathcal{Q}$ to $\mathcal{H}$ and by restriction $\mathcal{O}^1$ maps into $\mathcal{H}^1$, so the trace of $\psi(\lambda + \lambda^{-1})$ must have absolute value less than two. Extending ψ to L,

$$\psi(\lambda + \lambda^{-1}) = [\psi(\lambda)] + [\psi(\lambda)]^{-1}.$$

Since K is totally real, this is in $\mathbb{R}$ so that either $\psi(\lambda)$ is real or $\psi(\lambda)$ is on the unit circle. If $\psi(\lambda)$ were real, then $|\psi(\lambda + \lambda^{-1})| = |\psi(\lambda) + \psi(\lambda)^{-1}| < 2$, is equivalent to $\big(\psi(\lambda) - 1\big)^2 < 0$, which is impossible. Therefore, $\psi(\lambda)$ is on the unit circle.

□

Consider the case when λ_n is a Salem number corresponding to the geodesic γ_n. By the above discussion on lengths and traces, the following are equivalent: a sequence of Salem numbers $\{\lambda_n\}$ is bounded away from one, the Mahler measure of each term in $\{\lambda_n\}$ is bounded away from one, the sequence $\{\lambda_n + \lambda_n^{-1}\}$ is bounded from two, the geodesic lengths $\{\mathrm{length}(\gamma_n)\}$ are all bounded away from zero.

It suffices to show that any Salem number λ corresponds to a hyperbolic element γ in some arithmetic Fuchsian group.

Claim 9.3. *Let λ be a Salem number. Then there is a derived Fuchsian group Γ and a hyperbolic element $\gamma \in \Gamma$ so that $|tr(\gamma)| = \lambda + \lambda^{-1}$.*

Proof. (sketch) The only conjugate of λ which lies off the unit circle is λ^{-1}. It follows that the field $\mathbb{Q}(\lambda + \lambda^{-1}) = K$ is totally real and $L = \mathbb{Q}(\lambda)$ is a quadratic extension of K. We want to construct a quaternion algebra over K which is split at exactly one place. Moreover, we need to ensure that $\lambda + \lambda^{-1}$ appears as a trace of a norm one element.

By controlling the ramification set, we can construct a quaternion algebra $\mathcal{Q}$ over K, which is ramified at all non-identity real places of K, in which L embeds. The element $\lambda \in L$ is an algebraic integer since $|\lambda + \lambda^{-1}| = |\mathrm{tr}(\gamma)|$ is an algebraic integer and λ satisfies $x^2 - (\lambda + \lambda^{-1})x + 1$. Moreover, the relative trace and norm are $\mathrm{tr}_{K/L}(\lambda) = \lambda + \lambda^{-1} \in \mathcal{O}_K$ and $N_{K/L} = \lambda\lambda^{-1} = 1$. In the embedding $L \hookrightarrow \mathcal{Q}$ these correspond to the reduced norm and trace of an element q. It suffices to take a maximal order $\mathcal{O}$ containing q.

□

References

1. Mark D. Baker, *All links are sublinks of arithmetic links*, Pacific J. Math. **203** (2002), no. 2, 257–263. MR 1897900 (2003a:57008)
2. M.-J. Bertin, A. Decomps-Guilloux, M. Grandet-Hugot, M. Pathiaux-Delefosse, and J.-P. Schreiber, *Pisot and Salem numbers*, Birkhäuser Verlag, Basel, 1992, With a preface by David W. Boyd. MR 1187044 (93k:11095)
3. Marie-José Bertin and Martine Pathiaux-Delefosse, *Conjecture de Lehmer et petits nombres de Salem*, Queen's Papers in Pure and Applied Mathematics, vol. 81, Queen's University, Kingston, ON, 1989. MR 986490 (90f:11090)
4. P. E. Blanksby and H. L. Montgomery, *Algebraic integers near the unit circle*, Acta Arith. **18** (1971), 355–369. MR 0296021 (45 #5082)
5. A. Borel, *Commensurability classes and volumes of hyperbolic* 3*-manifolds*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **8** (1981), no. 1, 1–33. MR 616899 (82j:22008)
6. Peter Borwein, Edward Dobrowolski, and Michael J. Mossinghoff, *Lehmer's problem for polynomials with odd coefficients*, Ann. of Math. (2) **166** (2007), no. 2, 347–366. MR 2373144 (2008j:11153)

7. David W. Boyd, *Small Salem numbers*, Duke Math. J. **44** (1977), no. 2, 315–328. MR 0453692 (56 #11952)
8. E. Dobrowolski, *On a question of Lehmer and the number of irreducible factors of a polynomial*, Acta Arith. **34** (1979), no. 4, 391–401. MR 543210 (80i:10040)
9. J. Elstrodt, F. Grunewald, and J. Mennicke, *Groups acting on hyperbolic space*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998, Harmonic analysis and number theory. MR 1483315 (98g:11058)
10. R. Fricke, *uber die substitutionsgruppen welche zu den aus dem legendreschen integralmodul gezogenen wurzeln gehoren*, Math Annalen **28** (1886), 99Ð118.
11. Eknath Ghate and Eriko Hironaka, *The arithmetic and geometry of Salem numbers*, Bull. Amer. Math. Soc. (N.S.) **38** (2001), no. 3, 293–314. MR 1824892 (2002c:11137)
12. Mikhael Gromov, *Filling Riemannian manifolds*, J. Differential Geom. **18** (1983), no. 1, 1–147. MR 697984 (85h:53029)
13. Dennis A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2,R)$*. Vol. I*, Lecture Notes in Mathematics, Vol. 548, Springer-Verlag, Berlin, 1976. MR 0439755 (55 #12641)
14. ———, *The Selberg trace formula for* $\mathrm{PSL}(2, \mathbf{R})$*. Vol. 2*, Lecture Notes in Mathematics, vol. 1001, Springer-Verlag, Berlin, 1983. MR 711197 (86e:11040)
15. D. H. Lehmer, *Factorization of certain cyclotomic functions*, Ann. of Math. (2) **34** (1933), no. 3, 461–479. MR 1503118
16. Colin Maclachlan and Alan W. Reid, *The arithmetic of hyperbolic 3-manifolds*, Graduate Texts in Mathematics, vol. 219, Springer-Verlag, New York, 2003. MR 1937957 (2004i:57021)
17. Walter D. Neumann and Alan W. Reid, *Arithmetic of hyperbolic manifolds*, Topology '90 (Columbus, OH, 1990), Ohio State Univ. Math. Res. Inst. Publ., vol. 1, de Gruyter, Berlin, 1992, pp. 273–310. MR 1184416 (94c:57024)
18. Kathleen L. Petersen, *One-cusped congruence subgroups of Bianchi groups*, Math. Ann. **338** (2007), no. 2, 249–282. MR 2302062 (2008b:20063)
19. ———, *Counting cusps of subgroups of* $\mathrm{PSL}_2(\mathcal{O}_K)$, Proc. Amer. Math. Soc. **136** (2008), no. 7, 2387–2393. MR 2390505 (2008m:11091)
20. Hans Petersson, *Über die Konstruktion zykloider Kongruenzgruppen in der rationalen Modulgruppe*, J. Reine Angew. Math. **250** (1971), 182–212. MR 0294255 (45 #3324)
21. G. Pick, *uber gewisse ganzzahlige lineare substitutionen welche sich nicht durch algebraisches congruenzen erklaren lassen*, Math Annalen **28** (1886), 119–124.
22. Alan W. Reid, *Arithmeticity of knot complements*, J. London Math. Soc. (2) **43** (1991), no. 1, 171–184. MR 1099096 (92a:57011)
23. ———, *Isospectrality and commensurability of arithmetic hyperbolic* 2- *and* 3*-manifolds*, Duke Math. J. **65** (1992), no. 2, 215–228. MR 1150584 (93b:58158)
24. Atle Selberg, *On the estimation of Fourier coefficients of modular forms*, Proc. Sympos. Pure Math., Vol. VIII, Amer. Math. Soc., Providence, R.I., 1965, pp. 1–15. MR 0182610 (32 #93)
25. Jean-Pierre Serre, *Le problème des groupes de congruence pour SL2*, Ann. of Math. (2) **92** (1970), 489–527. MR 0272790 (42 #7671)
26. C. J. Smyth, *On the product of the conjugates outside the unit circle of an algebraic integer*, Bull. London Math. Soc. **3** (1971), 169–175. MR 0289451 (44 #6641)
27. Gerard van der Geer, *Hilbert modular surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 16, Springer-Verlag, Berlin, 1988. MR 930101 (89c:11073)
28. Marie-France Vignéras, *Variétés riemanniennes isospectrales et non isométriques*, Ann. of Math. (2) **112** (1980), no. 1, 21–32. MR 584073 (82b:58102)

TORSION HOMOLOGY OF THREE–MANIFOLDS

JEAN RAIMBAULT

Abstract. We review various heuristics, questions and conjectures about the torsion part of the homology of compact three–manifolds. We also present a theorem on growth of torsion homology in congruence covers of arithmetic manifolds and give an informal introduction to it's proof.

1. Introduction and overview

Let M be a compact three–manifold; one of the simplest topological invariants of M is given by it's homology groups $H_p(M;\mathbb{Z})$ which can be computed using any cell structure for M, for example from a Heegard decomposition or from a presentation of M as a Dehn surgery on a link in the three–sphere. The groups $H_p(M;\mathbb{Z})$ are finitely generated abelian groups and as such they decompose as a direct sum

$$H_p(M;\mathbb{Z}) = \mathbb{Z}^{b_p(M)} \oplus T_p(M)$$

where $T_p(M)$ is the torsion subgroup, which is finite. Moreover, at least for closed manifolds and manifolds whose boundary is a disjoint union of tori one can easily see that H_1 determines the others. A basic question which will interest us here is the following: given a compact manifold M, what is the range of $b_1(M')$ and $t_1(M') = |T_1(M')|$ for M' a finite cover of M? Related to this one can ask what the behaviour of these numbers is in specific sequences of finite covers of M. This note is a expanded version of the talk given by the author at the workshop "Growth and Mahler measure in geometry and topology" which was held at the institute Mittag-Leffler from July 1 to 5, 2013. As its title indicates it is mainly focused on the torsion part of the homology; moreover, our main concern will be with hyperbolic manifolds. Our main aim is to provide an informal introduction to the contents of the author's papers [25],[26] and (to a lesser extent, since these contain much more than is talked about here) to the seminal paper of N. Bergeron and A. Venkatesh [4] and to the joint work of the author with M. Abèrt, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov and I. Samet [1].

Let us now describe in some detail what is to be found here. In the first part we will quickly review various results and conjectures about general three–manifolds: first we talk about growth of torsion in cyclic covers, then we explain how to relate homological torsion growth to ℓ^2-invariants in the context of three–manifolds, and finally we present probabilistic results on the homology and volume of random Heegard splittings. All of this motivates the belief that finite–volume hyperbolic manifolds should often have a large torsion subgroup in their first homology, and more precisely that its size should be (in "nice" situations) close to a certain exponent of the volume. The second part is dedicated to put this vague heuristic statement in a more rigorous form, which is achieved through stating various conjectures and some positive results in specific contexts; in particular we conclude the section with a survey of the homology growth in congruence covers of arithmetic manifolds, for which we can actually state some proven results. The last section explains informally the analytic methods which are used to prove the results explained in the previous.

Acknowledgments. I am happy to thank the organizers of the workshop, Eriko Hironaka and Ruth Kellerhals, for inviting me to Mittag Leffler. I also wish to thank this institution for it's hospitality and Jenny Wiklund for organizing the practical aspects of my visit.

2. Growth of torsion homology in finite covers

Let M be a compact three–manifold; then by trivial considerations, if M' is a finite cover of M of degree d we have $b_1(M) \leq Cd$ where C is the smallest number of 1-simplices in a triangulation of M. There is a similar bound for torsion: by Lemma 5 in [11] there is a constant C' depending on the number of 1-simplices in a triangulation of M and the degree of its 2-simplices such that $\log t_1(M') \leq C'd$. The rough behaviour of the homology in a sequence of finite covers $M_n \to M$ can thus be studied through the behaviour of the numerical sequences

$$\frac{b_1(M_n)}{[\pi_1(M) : \pi_1(M_n)]} \text{ and } \frac{\log t_1(M_n)}{[\pi_1(M) : \pi_1(M_n)]}.$$

2.1. Cyclic covers.

The only case in which there are complete results on the exponential growth rate of torsion is that of cyclic covers, which already provides some interesting examples.

2.1.1. *Exponential growth of torsion.* The setting in this section is as follows: we have a compact three–manifold M with an epimorphism $\pi_1(M) \to \mathbb{Z}$, and we study the sequence of cyclic covers M_n corresponding to the surjections $\pi_1(M) \to \mathbb{Z} \to \mathbb{Z}/n$. Associated to the morphism $\pi_1(M) \to \mathbb{Z}$ (or to the corresponding infinite cyclic cover) is a certain sequence of Laurent polynomials $\Delta_0, \dots, \Delta_k, \dots$, called the Alexander polynomials of the covering, such that Δ_{i+1} divides Δ_i and $\Delta_i = 1$ for large i. The exponential growth rate of the sequence $t_1(M_n)$ is completely understood in terms of the Δ_k; the following result has been proved independantly by T. Le in [15] and by the author in [27] (see also [29], [13], [33] and [4, Section 7]).

Theorem 2.1. *Notations as above, let r be the smallest index such that $\Delta_r \neq 0$. Then we have*

$$\lim_{n\to+\infty} \frac{\log t_1(M)}{n} = m(\Delta_r) := \int_0^1 \log|\Delta_r(e^{2i\pi\theta})| d\theta.$$

It is a well-known result of L. Kronecker that the right-hand side above (which is the logarithmic Mahler measure of Δ_r) is zero if and only if Δ_r is a cyclotomic polynomial; in general it equals the sum of the $\log|\alpha|$ over the roots α of Δ_r with $|\alpha| > 1$. We will present various examples where Δ_r is explicitely computed in the sequel; for now let us indicate how to define the Alexander polynomials.

Let $R = \mathbb{Z}[t^{\pm 1}]$ and V be a finitely generated R-module; let $A \in M_{l,m}(R)$ be a presentation matrix for V, then the ith Alexander polynomial $\Delta_i(V)$ of V is a greatest common divisor for the $(l-i)$-minors of A (it does not depend on A). If $\widehat{M} \to M$ is an infinite cyclic covering then the homology group $H_1(\widehat{M};\mathbb{Z})$ is a R-module (where t is a generator for the covering group) and we put $\Delta_i = \Delta_i(H_1(\widehat{M};\mathbb{Z}))$. It is defined up to multiplication by a unit $\pm t^k, k \in \mathbb{Z}$, and we see that $r = \mathrm{rk}_R(H_1(M;\mathbb{Z}))$.

2.1.2. *Fibered manifolds.* We suppose here that M fibers over the circle, i.e. there is a surface S and an homeomorphism ϕ of S such that

$$M \cong S \times [0,1]/\sim \text{ where } (x,0) \sim (\phi(x),1).$$

Then $\pi_1(S)$ is a normal subgroup of $\pi_1(M)$ with quotient $\pi_1(M)/\pi_1(S) = \mathbb{Z}$, the corresponding infinite cyclic covering has $\Delta_r = \Delta_0 = \det(1 - t\phi_*)$.

This example allows to exhibit fibered manifolds which have cyclic covers with exponential growth of homology. For example let M be the fibered Sol-manifold given by the above construction with $S = \mathbb{T}^2$ and $\phi = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Then $\Delta_0 = t^2 - 3t + 1$ and $m(\Delta_0) = \log\left(\frac{3+\sqrt{5}}{2}\right)$. One can also give hyperbolic examples with $m(\Delta_0) > 0$, although they are more complicated. Some were computed by T. Koberda, and he made the following conjecture (which if true would imply that every finite–volume hyperbolic manifold has a sequence of covers where t_1 grows exponentially with the volume).

Conjecture 2.2. *Let S be an hyperbolic surface and ϕ a pseudo-Anosov homeomorphism of S. There exists a finite cover S' of S and a lift ϕ' of ϕ to S' such that* $\det(1 - t\phi_*)$ *is not cyclotomic.*

2.1.3. *Knot complements.* Let k be a knot in $\mathbb{S}^3$ and $M = \mathbb{S}^3 - k$. Then $H_1(M) = \mathbb{Z}$ and thus there is a unique cyclic cover M_n of degree n of M. The Alexander polynomial Δ_0 of the infinite cyclic cover of M can be computed from a diagram of the knot (this is actually how it was originally introduced), and it is always nonzero. In this setting Theorem 2.1 is due to D. Silver and S. Williams. For the simplest knots we have:

- if k is the trefoil then $\Delta_0 = t^2 + t - 1$ is cyclotomic;
- if k is the figure-eight then $\Delta_0 = t^2 - 3t + 1$.

Since the complement of the figure-eight is hyperbolic this gives examples of a sequence of covers of a hyperbolic manifold with exponential growth of torsion.

2.1.4. *More general results.* Theorem 2.1 admits a generalization to non-cyclic abelian covers ([15], see also [32] and [27]) which is less precise (although [15] gives the best possible result conditionnally to a number–theoretical conjecture). There is no other, more general type of covers where such a result is known to hold.

2.2. **Relations with ℓ^2-invariants.** Here we state without justification the more or less conjectured relations between homology of covers and the so–called ℓ^2–invariants in the special case of three–manifolds; we will recast them in analytic context in Section 4 below. The reader is refered to [18] for the definition of ℓ^2-invariants; the original paper on approximation is [16].

2.2.1. *Approximation and homology growth.* Suppose that M is a three–manifold with a CW-structure; there is then defined an invariant (which a priori depends on the chosen CW-structure) called Reidemeister torsion and denoted $\tau(M)$. If $M_\infty \to M$ is an infinite normal cover then there is also defined an ℓ^2-Reidemeister torsion which we will denote $\tau^{(2)}(M_\infty)$. The problem of approximation is then the following: let $\Gamma = \pi_1(M)$, $\Lambda = \pi_1(M_\infty)$; if Γ/Λ is residually finite we can choose a nested sequence of finite-index normal subgroups $\Gamma = \Gamma_0 \supset \ldots \supset \Gamma_n \supset \ldots$ such that $\bigcap_{n\geq 0} \Gamma_n = \Lambda$; letting M_n be the cover of M corresponding to Γ_n with the lifted CW-structure, do we have

$$\lim_{n\to+\infty} \tau(M_n)^{\frac{1}{[\Gamma:\Gamma_n]}} = \tau^{(2)}(M_\infty)? \tag{2.1}$$

This problem is linked to homology growth in the torsion as follows: we have $\tau(M) = t_1(M)^{-1} \times R(M)$ where $R(M)$ comes from the characteristic 0 homology. In some sequences as above it is hoped that $R(M_n)$ has subexponential growth rate with the index, so that if valid (2.1) would yield the exponential growth rate of torsion in the sequence $t_1(M_n)$. This is a way to prove the result about homology growth in cyclic covers (Theorem 2.1).

2.2.2. *Geometrized manifolds.* When $M_\infty = \widetilde{M}$ is the universal cover of M the ℓ^2-torsion is computable in terms of the geometric JSJ-decomposition of M: if $X_1, \ldots, X_m$ are the hyperbolic pieces then

$$\log \tau^{(2)}(\widetilde{M}) = \frac{1}{6\pi} \sum_{i=1}^{m} \operatorname{vol} X_i. \tag{2.2}$$

This does not depend on the choice of a CW-structure on M.

Finally, let us note that the conjectural picture above has been clarified by W. Lück in the case of covers of Seifert fibered manifolds; we will cite the following result from [17] (see Corollary 1.13 there).

Theorem 2.3. *Let M be a compact Seifert fibered three–manifold, and M_n a tower of finite normal covers of M with $\bigcap_n \pi_1(M_n)$ trivial. Then*

$$\lim_{n\to+\infty} \frac{\log t_1(M_n)}{[\pi_1(M):\pi_1(M_n)]} = 0.$$

2.3. **Random manifolds.** Let S be a closed surface, $\mathrm{Mod}(S)$ the mapping class group of S. Let g_k be the kth step of a uniform random walk on $\mathrm{Mod}(S)$ (i.e. g_{k+1} is given by $g_k g$ where g is chosen in a finite symmetric generating set for $\mathrm{Mod}(S)$ with respect to the uniform distribution). Then the manifold M_k obtained by gluing two handlebodies according to X_k (a "random Heegard splitting") is called a random Dunfield–Thurston manifold (after the paper [10] where they were studied first), and the study of the statistical properties of the homology of M_k as $k \to +\infty$ may provide insight. The following theorem is a compilation of results by J. Maher [19], J. Brock–J. Souto (mostly unwritten) and E. Kowalski ([14], which is the main source for this subsection).

Theorem 2.4. *Suppose that the genus of S is at least two, then:*

(i) *M_k is hyperbolic with asymptotic probability 1;*
(ii) *there are $0 < c_1 < c_2$ such that $c_1 k \le \mathrm{vol}\, M_k \le c_2 k$ with asymptotic probability 1;*
(iii) *for any sequence u_k such that $u_k \to +\infty$ we have $t_1(M_k) > e^{\frac{k}{u_k}}$ with asymptotic probability 1; moreover, there are $C, \alpha > 0$ such that the expectation of $t_1(M_k)$ is $\ge Ce^{\alpha k}$ for large enough k.*
(iv) *$b_1(M_k) = 0$ with asymptotic probability 1.*

Combining (ii) and (iii) above we see that for this model of random manifold, the torsion homology is larger than a certain exponent of the volume for a positive proportion of manifolds when the complexity k goes to infinity. This is another motivation to study further the relation between $\log t_1$ and vol for hyperbolic three–manifolds.

3. Hyperbolic manifolds

In this section M will always be a finite–volume hyperbolic manifold (which may vary from one occurence to the other).

3.1. **A conjecture.** The behaviour of random manifolds makes it clear that the torsion homology of hyperbolic three–manifolds is an interesting object of study. It exhibits a wide variety of behaviours in finite covers: as we saw in 2.1.3 and 2.1.2 there are sequences where $t_1(M)$ grows as fast as possible, namely exponentially. On the other hand it has been proven by M. Baker, M. Boileau and S. Wang [2] that there exists finite–volume hyperbolic manifolds which have towers of finite covers which are homology spheres. Nevertheless, in view of the expected links between approximation of ℓ^2–invariants and homology growth (2.1),(2.2) we can ask the following question which predicts some kind of uniform behaviour for all hyperbolic manifolds.

Conjecture 3.1. *For any finite-volume hyperbolic three–manifold M there is a tower of finite covers $\ldots \to M_n \to \ldots \to M_0 = M$ with $\bigcap_n \pi_1(M_n) = \{1\}$ and*

$$\lim_{n\to+\infty} \frac{\log t_1(M_n)}{\mathrm{vol}\, M_n} = \frac{1}{6\pi}.$$

It is not expected (at least not by everybody) that for *any* tower satisfying the assumptions of the question the growth of torsion homology will be as indicated; indeed there are numerical computations suggesting that for certain sequences this might not be the case [5, Figure 4.5], [31] . On the other hand, the same computations suggest that if a tower as in the conjecture satisfies the additional assumption that $H_1(M_n;\mathbb{Q}) = 0$ for all n then the conclusion holds for that sequence.

3.2. **Convergence of hyperbolic manifolds and homology.** We will put the sequence of covers for which the conjecture was made above in a more general geometric context. Note that if M is a compact manifold, then the condition in Question 3.1 is equivalent to $\mathrm{inj}(M_n) \to +\infty$ for a tower. It is thus natural to consider sequences of compact hyperbolic manifolds with $\mathrm{inj}(M_n) \to +\infty$; we will be interested in sequence satisfying the following weaker condition, which was introduced in [1, Definition 1.1] under the name of "Benjamini–Schramm"[1] convergence to $\mathbb{H}^3$[2]:

$$\forall R > 0, \lim_{n\to+\infty} \frac{\mathrm{vol}\{x \in M_n : \mathrm{inj}_x(M_n) \le R\}}{\mathrm{vol}\, M_n} \xrightarrow[n\to+\infty]{} 0. \tag{3.1}$$

That this is a good setting in which to study the asymptotics of homology is illustrated by the following result [1, Theorem 1.8], [25, Proposition C]:

Theorem 3.2. *Let M_n be a sequence of finite-volume hyperbolic three–manifolds which BS-converges to $\mathbb{H}^3$; then we have*

$$\lim_{n\to+\infty} \frac{b_1(M_n)}{\mathrm{vol}\, M_n} = 0.$$

Note that even for sequences of covers the convergence to 0 can be arbitrarily slow [12]. For torsion the picture is less clear; it is known by work of J. Brock and N. Dunfield that BS-convergence to $\mathbb{H}^3$ is far from implying exponential growth of torsion (in [5] they construct a sequence M_n which satisfies (3.1) but has $H_1(M_n; \mathbb{Z}) = 0$ for all n). The following conjecture can nevertheless be made.

Conjecture 3.3. *Let M_n be BS-convergent to $\mathbb{H}^3$ and such that the Cheeger constants of the M_n are bounded away from 0 and $H_1(M_n; \mathbb{Q}) = 0$ for all n; then*

$$\lim_{n\to+\infty} \frac{\log t_1(M_n)}{\mathrm{vol}\, M_n} = \frac{1}{6\pi}.$$

3.3. **Congruence manifolds.**

3.3.1. *Congruence groups.* Let G be a $\mathbb{Q}$–form of $\mathrm{SL}_2(\mathbb{C}) \times G'$ where G' is a compact group. Let $\mathbb{A}_f$ be the ring of finite adèles of $\mathbb{Q}$; then $\mathrm{G}(\mathbb{Q})$ is dense in $\mathrm{G}(\mathbb{A}_f)$, and a subgroup $\Gamma \subset \mathrm{G}(\mathbb{Q})$ is said to be a congruence group in $\mathrm{G}(\mathbb{Q})$ if the closure K_f of Γ in $\mathrm{G}(\mathbb{A}_f)$ is compact and open and moreover $\Gamma = \mathrm{G}(\mathbb{Q}) \cap K_f$. The following theorem is a slight generalization of [1, Theorem 1.12] and [26, Theorem B].

Theorem 3.4. *Let G as above and Γ_n be the images in $\mathrm{SL}_2(\mathbb{C})$ of a sequence of pairwise distinct congruence groups in $\mathrm{G}(\mathbb{Q})$; then the sequence of orbifolds $M_n = \Gamma_n\backslash\mathbb{H}^3$ is BS-convergent to $\mathbb{H}^3$.*

We will call a hyperbolic three–manifold a congruence manifold if its fundamental group is conjugated in $\mathrm{PSL}_2(\mathbb{C})$ to the image of a congruence group in some $\mathrm{G}(\mathbb{Q})$. The Cheeger constant of congruence three–manifolds is known to be bounded below by a uniform constant [9], but the rational homology of a congruence manifold can be nonzero, so that Conjecture 3.3 does not necessarily apply to a sequence of congruence manifolds. The following conjecture is nevertheless believed to be true.

Conjecture 3.5. *Let $\mathrm{G}, \Gamma_n, M_n$ be as in Theorem 3.4. Then we have*

$$\lim_{n\to+\infty} \frac{\log t_1(M_n)}{\mathrm{vol}\, M_n} = \frac{1}{6\pi}.$$

[1] After the work [3] of these authors on a similar notion for regular graphs.

[2] It fits into a more general notion of convergence for finite–volume manifolds, which also incorporates the convergence towards an infinite Galois cover; see[1, Definition 3.1 and Lemma 3.5]

This is supported by the computations of M.H. Şengun [30] and of Brock–Dunfield [5, Figure 4.4]. It is also motivated by arithmetic considerations, see [4].

3.3.2. *Example: the Bianchi groups.* The simplest (to describe) example of congruence groups is given by the Bianchi groups. Let $F = \mathbb{Q}(\sqrt{-d})$ be an imaginary quadratic field and $\mathcal{O}_F$ its ring of integers. We get a $\mathbb{Q}$-form of $\mathrm{PSL}_2(\mathbb{C})$ by taking the Weil restriction G of PSL_2/F to $\mathbb{Q}$. At finite places we get:

(i) $\mathrm{G}(\mathbb{Q}_p) = \mathrm{PSL}_2(\mathbb{Q}_p) \times \mathrm{PSL}_2(\mathbb{Q}_p)$ if p is split in $F/\mathbb{Q}$;
(ii) $\mathrm{G}(\mathbb{Q}_p) = \mathrm{PSL}_2(F_p)$ where $F_p = \mathbb{Q}_p(\sqrt{-d})$ is a quadratic extension of $\mathbb{Q}_p$ if p is inert or ramified in $F/\mathbb{Q}$.

Let $\Gamma = \mathrm{PSL}_2(\mathcal{O}_F)$; then the closure of Γ in $\mathrm{PSL}_2(\mathbb{Q}_p)$ is equal to $\mathrm{PSL}_2(\mathbb{Z}_p) \times \mathrm{PSL}_2(\mathbb{Z}_p)$ in case (i) and to $\mathrm{PSL}_2(\mathcal{O}_p)$ in case (ii) (where $\mathcal{O}_p$ is the closure of $\mathcal{O}_F$ in F_p), and we see that Γ is a congruence group in $\mathrm{G}(\mathbb{Q})$. We can define many other congruence subgroups inside Γ: for an ideal $\mathfrak{I}$ in $\mathcal{O}_F$ let $\Gamma(\mathfrak{I})$ be the kernel of the reduction morphism $\mathrm{PSL}_2(\mathcal{O}_F) \to \mathrm{PSL}_2(\mathcal{O}_F/\mathfrak{I})$, then $\Gamma(\mathfrak{I})$ is easily seen to be a congruence group, and it follows that the preimage in Γ of any subgroup in $\mathrm{PSL}_2(\mathcal{O}_F/\mathfrak{I})$ is also a congruence groups; it is usual to denote by $\Gamma_0(\mathfrak{I})$ the preimage of the upper triangular matrices and by $\Gamma_1(\mathfrak{I})$ the preimage of the subgroup of unipotent matrices.

3.3.3. *Local coefficients.* Conjecture 3.5 is wide open at present; however there is a scheme of proof which we will describe in the next section and allows to get a better understanding of what is involved in the conjecture; moreover it actually succeeds in proving results similar to the conjecture, by replacing the trivial local system $\mathbb{Z}$ by other $\pi_1(M)$-modules. We will describe here a generalized version of Conjecture 3.5 and state the results obtained in [4], [1] and [26] in this direction.

We will consider here lattices in $\mathrm{SL}_2(\mathbb{C})$ rather than in $\mathrm{PSL}_2(\mathbb{C})$ for reasons that will soon be apparent; in any case, torsion free lattices in $\mathrm{PSL}_2(\mathbb{C})$ at least lift to isomorphic lattices in $\mathrm{SL}_2(\mathbb{C})$ so there is no loss of generality when considering manifolds. We let $\Gamma \subset \mathrm{SL}_2(\mathbb{C})$ be a lattice and ρ a representation of Γ in $\mathrm{SL}(L)$ for some free, finitely generated $\mathbb{Z}$-module L. There is then a ℓ^2-torsion associated to the chain complex $C_*(M; \mathbb{Z}[\Gamma] \otimes_{\mathbb{Z}} L)$ and a torsion homology $H_1(M; L)_{\mathrm{tors}}$ and we can ask how the growth of the latter in covers of M relates to the former, in the spirit of (2.1). We will formulate a conjecture in the case where Γ is arithmetic and ρ comes from a representation of $\mathrm{SL}_2(\mathbb{C})$. The real representations of $\mathrm{SL}_2(\mathbb{C})$ are on the spaces

$$V_{m,q} = \mathrm{Sym}^m(\mathbb{C}^2) \otimes \mathrm{Sym}^q(\overline{\mathbb{C}}^2)$$

where $\mathrm{SL}_2(\mathbb{C})$ acts on $\overline{\mathbb{C}}^2$ by conjugate matrices. For any lattice Γ the ℓ^2-torsion of $C_*(M; \mathbb{Z}[\Gamma] \otimes_{\mathbb{Z}} V)$ is given for $V = V_{m,q}$ by

$$\log \tau^{(2)}(\widetilde{M}; V) = \mathrm{vol}(M) t^{(2)}(V),$$

where

$$t^{(2)}(V) = \frac{-1}{48\pi}\big((m+q+2)^3 - |m-q|^3 + 3|m-q|(m+q+2)(m+q+2-|m-q|)\big) \tag{3.2}$$

has been computed in [4, 5.9.3, Example (3)].

Conjecture 3.6. *Let Γ be a congruence groups and Γ_n a sequence of pairwise distinct congruence groups contained in Γ; let ρ, V be a representation of $\mathrm{SL}_2(\mathbb{C})$ and suppose*[3] *that there is a lattice L in V which is preserved by $\rho(\Gamma)$. Then we have*

$$\lim_{n\to+\infty} \frac{\log t_1(M_n; L)}{\mathrm{vol}\, M_n} = -t^{(2)}(V).$$

[3] This is always the case is Γ is defined over a quadratic imaginary field.

As we promised, here are results proving this conjecture for a large portion of these coefficients systems. The first such result is due to N. Bergeron and A. Venkatesh [4] for cocompact lattices; for congruence covers of the Bianchi orbifolds there were previous result on exponential growth of homology in this context (due independantly to J. Pfaff [24] and the author [28, Section 6.5], both relying in some way on [20]) but they are far less precise than the following statement (and their proof cannot be expected to gice such a result).

Theorem 3.7. *Let Γ be an arithmetic lattice, $V = V_{m,q}$ and suppose there is a lattice L in V preserved by Γ. Let Γ_n be a sequence of pairwise distinct, torsion–free congruence subgroups of Γ. Suppose furthermore that $m \neq q$.*

[1] *If Γ is cocompact then*

$$\lim_{n\to+\infty} \frac{\log t_1(M_n; L)}{\operatorname{vol} M_n} = -t^{(2)}(V).$$

[26] *If Γ is a Bianchi group and the sequence Γ_n is moreover cusp-uniform then the same conclusion holds.*

Being cusp-uniform means that the conformal structures on the boundary components do not degenerate. We will give the proof of the first result in the next section (note that in this form it is not explicitely stated in [1]) and explain how to deal with the second case, when Γ is not cocompact.

4. Trace formula and analytic torsion

4.1. **Analytic torsion and an attempt to prove Conjecture 3.3 in the compact case.** We suppose here that M is a compact hyperbolic manifold and let E be a flat bundle with a Euclidean metric on M; the Hodge-Laplace operators $\Delta^p[M]$ on p-forms on M then have a self-adjoint extension to L^2-forms and the space $L^2\Omega^p(M;E)$ of these forms decomposes as a Hilbert sum of finite-dimensional eigenspaces. Let $0 < \lambda_1 \leq \ldots \leq \lambda_k \leq \ldots$ be the positive eigenvalues of $\Delta^p[M]$; then by Weyl's law for the asymptotics of λ_k the series

$$\zeta_p(s) = \sum_{k\geq 1} \lambda_k^{-s}$$

defines a holomorphic function in the half-plane $\mathrm{Re}(s) > 3/2$. The Minakshisundaram–Pleijel expansion for the trace of the heat kernel allows, via (4.4) below, to prove that it in fact extends to a meromorphic function on $\mathbb{C}$ which is regular at 0. One then defines the Ray–Singer analytic torsion of M with coefficients in E by the expression

$$T(M;E) = \prod_{p=1}^{3} \exp(\zeta_p'(0)). \tag{4.1}$$

This is useful to study homological torsion in view of the Cheeger–Müller Theorem[4], which amazingly relates this spectrally defines invariant to a topological one, in the spirit of Hodge–de Rham theory. We will state it first for $E = \mathbb{R}$ the trivial line bundle; it then amounts to an equality

$$T(M;\mathbb{R}) = \frac{R^1(M)}{\mathrm{vol}(M)\cdot |H_1(M;\mathbb{Z})_{\mathrm{tors}}|} \tag{4.2}$$

where $R^1(M)$ is defined to be the covolume of the lattice of non-torsion integral cohomology classes inside the space of harmonic forms. Now for another case of interest to us: suppose $E = E_\rho$ as

[4]The original result by Cheeger [8] and Müller is proven by both for orthogonal bundles; Cheeger's proof has been extended by Müller [21] to cover also unimodular bundles, and an even more general result has been proven by J.M. Bismut and W. Zhang.

above, Γ is arithmetic and L is a lattice in V which is preserved by $\rho(\Gamma)$. Suppose in addition that $H^*(M; E_\rho) = 0$; then

$$T(M; E_\rho) = \frac{|H_2(M; L)| \cdot |H_0(M; L)|}{|H_1(M; L)|}. \tag{4.3}$$

Thus the cocompact part of Theorem 3.7 follows from the following results:

Theorem 4.1. *Notations as in Theorem 3.7, if Γ is cocompact we have*

$$\lim_{n \to +\infty} \frac{\log T(M; E_\rho)}{\operatorname{vol} M_n} = t^{(2)}(V).$$

Lemma 4.2. *If Γ_n is a sequence of congruence subgroups of Γ then for $p = 0, 2$*

$$\log |H_p(M; L)_{\mathrm{tors}}| = o(\operatorname{vol} M_n).$$

Lemma 4.2 is proven in [4, Section 8.6] in a more general context; a very short proof for subgroups of Bianchi groups is given in [26, Lemma 6.5].

4.1.1. *Strong acyclicity and the proof of Theorem 4.1.* To study analytic torsion it is convenient to use an expression for $\zeta_p'(0)$ in terms of the trace of the heat kernel $e^{-t\Delta^p[M]}$ of M (with coefficients in some bundle E); it is defined to be

$$\operatorname{Tr} e^{-t\Delta^p[M]} := \dim \ker \Delta^p[M] + \sum_{k \geq 1} e^{-t\lambda_k}$$

A formal computation shows that we have

$$\zeta_p(s) = \frac{1}{\Gamma(s)} \int_0^{+\infty} \operatorname{Tr} e^{-t\Delta^p[M]} t^s \frac{dt}{t} \tag{4.4}$$

where $\Gamma(s) = \int_0^{+\infty} e^{-t} t^{s-1} dt$ is Euler's function. In particular, for any $t_0 > 0$ it follows that:

$$\zeta_p'(0) = \frac{d}{ds} \left(\frac{1}{\Gamma(s)} \int_0^{t_0} \operatorname{Tr} e^{-t\Delta^p[M]} t^s \frac{dt}{t} \right)_{s=0} + \int_{t_0}^{+\infty} \operatorname{Tr} e^{-t\Delta^p[M]} \frac{dt}{t}. \tag{4.5}$$

There is also an heat operator $e^{-t\Delta^p[\mathbb{H}^3]}$ on the hyperbolic space, and it has a "L^2-trace" $\operatorname{Tr} e^{-t\Delta^p[\mathbb{H}^3]}$ which we will define below; one defines the L^2–analytic torsion for $\mathbb{H}^3$ with coefficients in E using (4.5) with the heat operator $e^{-t\Delta^p[M]}$ replaced by $e^{-t\Delta^p[\mathbb{H}^3]}$, and a computation using the Harish-Chandra Plancherel formula yields (3.2) for its logarithm $t^{(2)}(V)$. The first part of the proof of Theorem 4.1 is thus the following result, the proof of which we will sketch in 4.1.3 below.

Lemma 4.3. *Let M_n be a sequence of compact hyperbolic three–manifolds which is BS-convergent to $\mathbb{H}^3$ and such that there exists a $\delta > 0$ for which $\operatorname{inj}(M_n) \geq \delta$ for all n. Then for any $t_0 > 0$ we have*

$$\frac{1}{\operatorname{vol} M_n} \frac{d}{ds} \left(\frac{1}{\Gamma(s)} \int_0^{t_0} \left(\operatorname{Tr} e^{-t\Delta^p[M_n]} - (\operatorname{vol} M_n) \operatorname{Tr} e^{-t\Delta^p[\mathbb{H}^3]} \right) t^s \frac{dt}{t} \right)_{s=0} \xrightarrow[n \to +\infty]{} 0.$$

To conclude the proof of Theorem 4.1 one needs to show that

$$\limsup_{t_0 \to +\infty} \left(\sup_n \int_{t_0}^{+\infty} \operatorname{Tr} e^{-t\Delta^p[M_n]} \frac{dt}{t} \right) = 0. \tag{4.6}$$

To prove this one needs to estimate *in a uniform manner in n* the exponential decay as $t \to +\infty$ of $\operatorname{Tr} e^{-t\Delta^p[M_n]}$. Taking for granted that the trace at a given time (say $t = 1$) $\operatorname{Tr} e^{-\Delta^p[M_n]}$ is bounded, this follows from the following result [4, Lemma 4.1].

Lemma 4.4. *If $m \neq q$, then there is a $\lambda_0 > 0$ such that for any discrete subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{C})$ and $M = \Gamma \backslash \mathbb{H}^3$, the Laplace operators $\Delta^p[M]$ do not have eigenvalues beneath λ_0.*

4.1.2. *A Selberg-type conjecture.* The existence of a uniform spectral gap is not a necessary condition for (4.6) to hold. Indeed, one hopes that it stays true in the case of trivial coefficients, where is is known that there has to appear eigenvalues very close to zero on 1-forms. We propose the following conjecture, which if true implies that Theorem 4.1 extends to trivial coefficients. For $\lambda > 0$ let $m(\lambda; M_n) = \dim\ker(\Delta^p[M_n] - \lambda\,\mathrm{Id})$ and let $m_p([0,\delta]; M_n) = (\sum_{\lambda\in[0,\lambda_1]} m_p(\lambda; M_n)$ be the number of eigenvalues of $\Delta^p[M_n]$ below λ_1.

Question 4.5. *Does there exists $\lambda_0 > 0$ such that for any $\varepsilon > 0$ there is a $C_\varepsilon > 0$ such that for all n and $\lambda_1 \leq \lambda_0$ we have*

$$\frac{\sum_{\lambda\in[0,\lambda_1]} m_1(\lambda; M)}{\mathrm{vol}\, M} \leq C_\varepsilon \delta^{1+\varepsilon} \,\mathrm{vol}\, M \quad ?$$

(Or less precisely, does this hold for some exponent $c > 0$ in place of $1 + \varepsilon$ on the right-hand side?)

Of course, the conjecture is that the answer is positive when M_n is a sequence of congruence covers of an arithmetic manifold.

4.1.3. *The trace formula and the proof of the Main Lemma, compact case.* If M is a compact hyperbolic three–manifolds the "trace formula" for M is in its crudest form the equality

$$\mathrm{Tr}\, e^{-t\Delta^p[M]} = \int_M \mathrm{tr}\, K_t^p[M](x,x)dx$$

where $K_t^p[M]$ is a kernel on $M \times M$ with coeficients in the right bundle[5], such that the operator $e^{-t\Delta^p[M]}$ is given by convolution with K_t^p. There is a similar kernel k_t^p on $\mathbb{H}^3$, and we will show unsing the formula above that for any $t_0 > 0$ there is a $C > 0$ such that for all $t \in]0, t_0]$

$$\mathrm{tr}\, K_t^p[M](x,x) - \mathrm{tr}\, k_t^p(x,x) \leq C\, \mathrm{inj}_x(M)^{-3} t^{-\frac{3}{2}} e^{-\frac{\mathrm{inj}_x(M)^2}{10t}} \tag{4.7}$$

for all $x \in M$. Now since the L^2-trace is given by

$$\mathrm{Tr}\, e^{-t\Delta^p[\mathbb{H}^3]} = \mathrm{tr}\, k_t^p(x_0, x_0) \tag{4.8}$$

for any $x_0 \in \mathbb{H}^3$, the proof of Lemma 4.3 follows without difficulties from (4.7).

The proof of (4.7) follows [4, (4.5.1)], [1, Lemma 8.23]; the principle is as follows: there is an expansion

$$K_t^p[M](x,x) = \sum_{\gamma\in\Gamma} k_t^p(x, \gamma x)$$

so we have to show that the sum over nontrivial elements is bounded by the right-hand side of (4.7); this is an immediate consequence of the estimate $|k_t^p(x,y)| \leq Ct^{-\frac{3}{2}} e^{-\frac{d(x,y)^2}{5t}}$ and of the well-known bound

$$|\{\gamma \in \Gamma :\ d(x,\gamma x) \leq r\}| \leq C\, \mathrm{inj}_x(M)^{-3} e^{cr}.$$

4.2. **The trace formula and analytic torsion in the non-compact case.** In the non-compact case the Laplace operator has continuous spectrum, the heat kernel is not trace–class and we cannot define the analytic torsion immediately as in the compact case. What is to be done is to define the trace using the Selberg trace formula and analyse it to get the analytic continuation and regularity at 0 as in the compact case. This is done in [23], [22] and [25], we will follow the approach of the latter paper which is better suited to the problems at hand.

[5] So that $K_t^p[M](x,y)$ is a linear map from $\wedge^p T_y^* M \otimes E_y$ to $\wedge^p T_x^* M \otimes E_x$.

4.2.1. *Selberg trace formula.* Let us try first to explain the Selberg Trace formula in our case (i.e. cusped hyperbolic three–manifolds) in some detail. Let M be a finite-volume, non-compact hyperbolic three–manifold; as in the compact case there is on M a kernel $K_t^p[M]$ the convolution with which is the heat operator on p-forms. The non-compactness of M implies that the function $\operatorname{tr} K_t^p[M]$ is not integrable on M, and the trace formula results from an asymptotic estimate of the integral of it on certain compact subsets which exhaust M.

More precisely, there exists a compact subset $M^1 \subset M$ which is a submanifold with smooth boundary consisting of h flat (with the induced metric) tori $T_1, \dots, T_h$ and such that $M - M^1$ is the disjoint union of $T_j \times [1, +\infty[$ with the metric $\frac{dx^2+dy_j^2}{y_j^2}$ where dx^2 is the flat metric on T_j and $y_j = \log d(\cdot, T_j)$ (d being the hyperbolic distance on M). Now putting $M^Y = \{x \in M \ \max_j y_j(x) \leq Y\}$ for $Y \in [1, +\infty[$, one can express the integral of $\operatorname{tr} K_t^p[M]$ on M^Y in two ways, using either the geometric expansion of the heat kernel or the spectral expansion; this yield two expansions as $Y \to \infty$ which are of the form $A \log Y + B + o(1), A' \log Y + B' + o(1)$ for some A, B, A', B' which we will describe presentely, and the trace formula is the equality $B = B'$.

4.2.2. *Geometric side.* Here we quickly explain [25, Proposition 3.4]. The heat kernel of M is written as

$$K_t^p[M](x,y) = \sum_{\gamma \in \Gamma} k_t^p(x, \gamma y);$$

the sum over $\gamma \in \Gamma$ can be separated into three summands: the one corresponding to $\gamma = 1$, the sum over elements in Γ with trace in $]2, +\infty[$ (usually called loxodromic elements, so we'll denote the set of them by Γ_{lox}) and the sum over nontrivial unipotents elements. The two first sums yield integrable functions on Γ: the integral of $\operatorname{tr} k_t^p(x,x)$ over M equals $\operatorname{Tr} e^{-t\Delta^p[\mathbb{H}^3]} \cdot \operatorname{vol} M$ (where x_0 is any point in $\mathbb{H}^3$). We have an inequality similar to (4.7) for the sum over loxodromics:

$$\sum_{\gamma \in \Gamma_{\mathrm{lox}}} \operatorname{tr} k_t^p(x, \gamma x) \leq C t^{-\frac{3}{2}} \ell(x)^{-3} e^{-\frac{\ell_x(M)^2}{10t}}$$

where C depends on Γ and is uniform for t in a compact set, and $\ell(x)$ is the smallest length of a closed curve through x which is (freely) homotopic to a losed geodeic (in particular it is bounded below for $x \in M$). It follows that this summand is bounded on M and thus integrable; we will denote its integral by $G_t^p(x)$:

$$G_t^p(x) = \int_M \sum_{\gamma \in \Gamma_{\mathrm{lox}}} \operatorname{tr} k_t^p(x, \gamma x) dx. \tag{4.9}$$

The sum over unipotent terms is not integrable over M, because the displacement of a unipotent element goes to 0 as one gets closer to its fixed point; however it is not hard to quantify the divergence: there is a smooth function h_t^p on $[0, +\infty)$ such that $\operatorname{tr} k_t^p(x, nx) = h_t^p(d(x, nx))$ for any $x \in \mathbb{H}^3$ and any unipotent $n \in \mathrm{SL}_2(\mathbb{C})$; moreover $d(x, nx)$ is given by $\ell(|n|)$ where $|n|$ is the Euclidean norm of n in an horosphere passing through x. Now if $\{\Lambda\}$ is the conjugacy class of a unipotent subgroup of Γ then we have when $\min_j Y_j \to +\infty$ the following asymptotic expansion

$$\int_{M^Y} \sum_{\gamma \in \{\Lambda\}, \gamma \neq 1} \operatorname{tr} k_t^p(x, \gamma x) = \sum_{j=1}^h \log(Y_j) \int_0^{+\infty} r \log(r) h_t^p(\ell(r)) + \sum_{j=1}^h \kappa_j \operatorname{vol} \Lambda_j \int_0^{+\infty} r h_t^p(\ell(r)) dr + o(1)$$

where κ_j is a constant which depends only on the geometry of T_j. More precisely, we have

$$\kappa_j \operatorname{vol} \Lambda_j = \kappa_j' + \log \alpha_1(T_j)$$

where κ_j' depends only on the conformal structure of T_j, thus only on M and not on the choice of M^1, and $\alpha_1(T_j)$ is the (Euclidean) systole of T_j (and thus does depend on M^1). The geometric expression for the "trace" of the heat kernel is finally given by

$$\operatorname{Tr}_R K_t^p[M] = \operatorname{Tr} e^{-t\Delta^p[\mathbb{H}^3]} \cdot \operatorname{vol} M + \int_M G_t^p(x)dx + \sum_{j=1}^h \kappa_j \operatorname{vol} \Lambda_j \int_0^{+\infty} r h_t^p(\ell(r))dr \tag{4.10}$$

(the index R in Tr_R stands for "regularized" since this is not a bona fide trace).

4.2.3. *Spectral side.* The spectral decomposition for the space $L^2\Omega^p(M;E)$ can be written in its roughest form as

$$L^2\Omega^p(M;E) = L^2_{\mathrm{disc}}\Omega^p(M;E) \oplus L^2_{\mathrm{cont}}\Omega^p(M;E)$$

where $L^2_{\mathrm{disc}}\Omega^p(M;E)$ is the closure of the space generated by eigenforms of $\Delta^p[M]$ and $L^2_{\mathrm{cont}}\Omega^p(M;E)$ its orthogonal complement. If $\lambda_1 \leq \ldots \leq \lambda_k \leq \ldots$ are the positive eigenvalues of $\Delta^p[M]$ in $L^2_{\mathrm{disc}}\Omega^p(M;E)$ it is known that the sum $\sum_{k\geq 1} e^{-t\lambda_k}$ is convergent for $t > 0$. On the other hand there is an exact description of the continuous part in terms of the so-called Eisenstein series, and it yields an expansion

$$\int_{M^Y} \operatorname{tr} K_t^p[M](x,x)dx = T \cdot \sum_{j=1}^h \log Y_j + \dim\ker \Delta^p[M] + \sum_{k\geq 1} e^{-t\lambda_k} + S_t^p + o(1) \tag{4.11}$$

where S_t^p, T are computed in terms of the heat kernel and certain "intertwining" operators on V (see [25, (3.15)] for $p = 0$). Thus the spectral expression for the regularized trace is given by

$$\operatorname{Tr}_R e^{-t\Delta^p[M]} - \dim\ker \Delta^p[M] = \sum_{k\geq 1} e^{-t\lambda_k} + S_t^p. \tag{4.12}$$

4.2.4. *Regularized analytic torsion.* Using the geometric side of the trace formula one can study the asymptotics of $\operatorname{Tr}_R e^{-t\Delta^p[M]}$ as $t \to 0$, and this yields an expension similar to Minakshisundaram-Pleijel with additional terms in $t^{\frac{k}{2}} \log t$ for $k \geq -1$ [26, Proposition 5.4]; it allows to deduce that the zeta function defined as in (4.4) with traces repaced by regularized traces is a meromorphic function which is regular at 0, and one then defines regularized analytic torsion as in (4.1); see [26, 5.3.1] for details. The following result then replaces Theorem 4.1 in the non-compact case.

Theorem 4.6. *Let ρ be a strongly acyclic representation of $\mathrm{SL}_2(\mathbb{C})$, Γ a Bianchi group and Γ_n a sequence of pairwise distinct torsion-free congruence subgroups of Γ. Suppose in addition that $M_n = \Gamma_n\backslash\mathbb{H}^3$ is cusp-uniform. Then we have*

$$\lim_{n\to+\infty} \frac{\log T_R(M_n;E_\rho)}{\operatorname{vol} M_n} = t^{(2)}(V).$$

The proof goes as in the compact case, with some additional difficulties: for the small–time part of the proof one has to deal with the terms coming from unipotent elements on the geometric side; this part of the proof applies to any sequence of hyperbolic manifolds which is BS convergent to $\mathbb{H}^3$ and cusp-uniform (the cusp-uniformity allows to control the κ_j', see the proof of Theorem 4.5 in [25]). For the large-time part one has to control the term S_t^p: this is more technical and this is where we use the fact that the manifolds M_n are congruence (which allows to compute more or less explicitely the "intertwining operators" from which this term comes). The proof of this limit is given in *loc. cit*, Section 5.4 under an assumption on the intertwining operators, and the latter is shown to hold for sequence of congruence lattices in [26, 3.3].

4.3. **Asymptotic Cheeger–Müller and homology growth.** This is actually the hardest part in the proof of Theorem 3.7: there is currently no extension of the Cheeger–Müller theorem to the non-compact case. The result which is proven in these two papers is that one can define a "Reidemeister" torsion $\tau(M; L)$ such that asymptotic equality with analytic torsion holds: the statement of Theorem 5.1 in [26] is that we have

$$\lim_{n\to+\infty} \frac{\log T(M_n; E) - \log \tau(M_n; L)}{\operatorname{vol} M_n} = 0 \tag{4.13}$$

for a sequence of (manifold) congruence covers of a Bianchi orbifold. We will not detail the proof here: the main ingredients are an asymptotic equality of regularized torsion with an analytic torsion defined for the compact truncated manifolds $M_n^{Y^n}$ for an explicit sequence Y^n (Theorem 6.1 in [25]), a Cheeger–Müller equality for the latter proven in the generality we need by J. Brüning and X. Ma ([6], see [25, (6.3)]), and ideas originating in [7] (Proposition 5.4 in [26]) to conclude.

The torsion τ appearing in (4.13) is given by

$$\tau(M; V_{\mathbb{Z}}) = \frac{|H^1(M; L)_{\mathrm{tors}}|}{\operatorname{vol} H^1(M; L)_{\mathrm{free}}} \times \frac{\operatorname{vol} H^2(M; L)_{\mathrm{free}}}{|H^2(M; L)_{\mathrm{tors}}|} \tag{4.14}$$

where the covolume are taken for the metric on $H^*(M; E)$ coming from the embedding $i^* : H^*(M; E) \to \bigoplus_j H^*(T_j; L)$ (that this is a sensible choice is justified by (4.13)); the last part of the proof of Theorem 3.7 is to show that all terms but $|H^2(M; L)_{\mathrm{tors}}|$ are subexponential in the volume, and again we will not give any detail about the proof here, referring the reader to [26, Section 6] instead.

References

[1] M. Abert, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault, and I. Samet. On the growth of L^2-invariants for sequences of lattices in Lie groups. *ArXiv e-prints*, October 2012, 1210.2961.

[2] Mark Baker, Michel Boileau, and Shicheng Wang. Towers of covers of hyperbolic 3-manifolds. *Rend. Istit. Mat. Univ. Trieste*, 32(suppl. 1):35–43 (2002), 2001. Dedicated to the memory of Marco Reni.

[3] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp. (electronic), 2001.

[4] Nicolas Bergeron and Akshay Venkatesh. The asymptotic growth of torsion homology for arithmetic groups. *Journal de l'Institut de Math. de Jussieu*, 12(2):391–447, 2013.

[5] J. F. Brock and N. M. Dunfield. Injectivity radii of hyperbolic integer homology 3-spheres. *ArXiv e-prints*, April 2013, 1304.0391.

[6] Jochen Brüning and Xiaonan Ma. On the gluing formula for the analytic torsion. *preprint*, 2011.

[7] F. Calegari and A. Venkatesh. A torsion Jacquet–Langlands correspondence. *ArXiv e-prints*, December 2012, 1212.3847.

[8] Jeff Cheeger. Analytic torsion and the heat equation. *Ann. of Math. (2)*, 109(2):259–322, 1979.

[9] Laurent Clozel. Démonstration de la conjecture τ. *Invent. Math.*, 151(2):297–328, 2003.

[10] Nathan M. Dunfield and William P. Thurston. Finite covers of random 3-manifolds. *Invent. Math.*, 166(3):457–521, 2006.

[11] V. Emery. Torsion homology of arithmetic lattices and K2 of imaginary fields. *ArXiv e-prints*, March 2013, 1303.6132.

[12] S. Friedl. A note on the growth of Betti numbers and ranks of 3-manifold groups. *ArXiv e-prints*, January 2013, 1301.4025.

[13] Francisco González-Acuña and Hamish Short. Cyclic branched coverings of knots and homology spheres. *Rev. Mat. Univ. Complut. Madrid*, 4(1):97–120, 1991.

[14] Emmanuel Kowalski. On the complexity of Dunfield–Thurston random three–manifolds. *available on the author's webpage*, 2009.

[15] Thang Le. Homology torsion growth and Mahler measure. *ArXiv e-prints*, October 2010, 1010.4199.

[16] W. Lück. Approximating L^2-invariants by their finite-dimensional analogues. *Geom. Funct. Anal.*, 4(4):455–481, 1994.

[17] W. Lück. Approximating L^2-Invariants and Homology Growth. *Geom. Funct. Anal.*, 23(2):622–663, 2013.

[18] Wolfgang Lück. *L^2-invariants: theory and applications to geometry and K-theory*, volume 44 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge.* Springer-Verlag, Berlin, 2002.
[19] Joseph Maher. Random Heegaard splittings. *J. Topol.*, 3(4):997–1025, 2010.
[20] P. Menal-Ferrer and J. Porti. Higher dimensional Reidemeister torsion invariants for cusped hyperbolic 3-manifolds. *ArXiv e-prints*, October 2011, 1110.3718.
[21] Werner Müller. Analytic torsion and R-torsion for unimodular representations. *J. Amer. Math. Soc.*, 6(3):721–753, 1993.
[22] Werner Müller and Jonathan Pfaff. Analytic torsion of complete hyperbolic manifolds of finite volume. *J. Funct. Anal.*, 263(9):2615–2675, 2012.
[23] Jinsung Park. Analytic torsion and Ruelle zeta functions for hyperbolic manifolds with cusps. *J. Funct. Anal.*, 257(6):1713–1758, 2009.
[24] J. Pfaff. Exponential growth of homological torsion for towers of congruence subgroups of Bianchi groups. *ArXiv e-prints*, February 2013, 1302.3079.
[25] J. Raimbault. Asymptotics of analytic torsion for hyperbolic three–manifolds. *ArXiv e-prints*, December 2012, 1212.3161.
[26] J. Raimbault. Analytic, Reidemeister and homological torsion for congruence three–manifolds. *ArXiv e-prints*, July 2013, 1307.2845.
[27] Jean Raimbault. Exponential growth of torsion in abelian coverings. *Algebr. Geom. Topol.*, 12(3):1331–1372, 2012.
[28] Jean Raimbault. *Torsion homologique dans les revêtements finis*. PhD thesis, Université Pierre et Marie Curie, 2012.
[29] Robert Riley. Growth of order of homology of cyclic branched covers of knots. *Bull. London Math. Soc.*, 22(3):287–297, 1990.
[30] Mehmet Haluk Şengün. On the integral cohomology of Bianchi groups. *Exp. Math.*, 20(4):487–505, 2011.
[31] Mehmet Haluk Şengün. On the torsion homology of non-arithmetic hyperbolic tetrahedral groups. *Int. J. Number Theory*, 8(2):311–320, 2012.
[32] Daniel S. Silver and Susan G. Williams. Mahler measure, links and homology growth. *Topology*, 41(5):979–991, 2002.
[33] Daniel S. Silver and Susan G. Williams. Torsion numbers of augmented groups with applications to knots and links. *Enseign. Math. (2)*, 48(3-4):317–343, 2002.

Institut de Mathématiques de Jussieu, Unité Mixte de Recherche 7586 du CNRS, Université Pierre et Marie Curie, 4, place Jussieu 75252 Paris Cedex 05, France.
E-mail address: raimbault@math.jussieu.fr

TORSION IN THE HOMOLOGY OF ARITHMETIC KLEINIAN GROUPS

MEHMET HALUK ŞENGÜN

Contents

1. Introduction

These are the extended notes of a talk I gave at the Mittag-Leffler Institute workshop "Growth and Mahler measures in geometry and topology" in July 2013. I thank the organizers and the Mittag-Leffler Institute for this most stimulating workshop and the wonderful environment.

In these notes, I will start with some background on arithmetic Kleinian group, their associated modular forms and their cohomology groups as Hecke modules. Afterwards, I will consider the torsion in the integral homology and its importance for number theory. Finally, I will briefly sketch some recent work on the asymptotic behaviour of torsion, including my recent work with N.Bergeron and A. Venkatesh.

2. Background

2.1. **Quaternion Algebras.** The definition of arithmetic hyperbolic 3-manifold relies on the notion of quaternion algebra. In this section we collect results about quaternion algebras that we will need.

A **quaternion algebra** D over a field K (denoted D/K) is a necessarily non-commutative ring with an injective ring map $K \to D$ such that

(1) the image of K is the centre of D,
(2) the dimension of D, considered as a K-vector space, is 4,
(3) D has no non-trivial two-sided ideals.

It is well-known that if the characteristic of K is not 2, then every quaternion algebra is isomorphic to one given in the following form:

$$\left(\frac{a,b}{K}\right) := \{1K \oplus iK \oplus jK \oplus ijK \mid i^2 = a1,\ j^2 = b1,\ ij = -ji\}.$$

Here the multiplication rules are extended via linearity to the whole vector space. The canonical example of a quaternion algebra is the 2×2 matrix algebra $M_2(K) \simeq (\frac{1,1}{K})$. A well-known example is the so called *Hamiltonians* $\mathbb{H} \simeq (\frac{-1,-1}{\mathbb{R}})$.

In fact, every quaternion algebra over a field K is either a skew-field (a.k.a. division ring) or isomorphic to $M_2(K)$. If K is algebraically closed, then the only quaternion algebra D/K is, up to isomorphism, $M_2(K)$. This implies that when we base change a quaternion algebra D/K to the algebraic closure $\overline{K}$ of K, it will necessarily become isomorphic to $M_2(\overline{K})$. Hence, we may regard D/K as a subalgebra of the matrix algebra $M_2(\overline{K})$. In particular, we can define trace and norm maps on D/K using the trace and determinant maps on $M_2(\overline{K})$. In fact, one does not need to go to the algebraic closure to get to the matrix algebra, it is known that there is always a degree two extension L/K such that $D \otimes_K L \simeq M_2(L)$.

When $K = \mathbb{R}$, there are, up to isomorphism, two quaternion algebras D/K, namely $\mathbb{H}$ and $M_2(\mathbb{R})$. Similarly, when K is a finite extension of $\mathbb{Q}_p$, there are again, up to isomorphism, only two, namely $M_2(K)$ and a skew-field which can be explicitly given (but we will not here) using the unique unramified degree 2 extension of K.

Given a quaternion algebra D over a number field K, we say that D **ramifies** over a place v of K, if the base-change quaternion algebra $D \otimes_K K_v$, where K_v is the completion of K at the place v, is a skew-field. Note that if $v : K \hookrightarrow K_v$ is a place of K and $D/K \simeq (\frac{a,b}{K})$, then

$$D \otimes_K K_v \simeq \left(\frac{v(a), v(b)}{K_v}\right).$$

Observe that D/K cannot ramify over a complex place v since then $K_v \simeq \mathbb{C}$ and there is only one quaternion algebra over $\mathbb{C}$. On the contrary, ramification behaviour may vary among the real places. The **discriminant** $\delta(D)$ of D/K is the product of all prime ideals of $\mathcal{O}_K$ corresponding to the finite places of K over which D ramifies. Let us call the set of places of K over which D/K ramifies $\mathbf{S(D)}$. It turns out that $S(D)$ provides all the information one needs about D/K. It is well known that

(1) S(D) has finite even cardinality,

(2) if $S(D) = S(E)$, then $D/K \simeq E/K$,
(3) for any set S of places of K of even cardinality that contains no complex place, there is a quaternion algebra D/K such that $S(D) = S$.

Hence there is a bijection between quaternion algebras D/K and finite sets S of places of K of even size which do not contain complex places.

An **order** $\mathcal{O}$ in a quaternion algebra D over a number field K with ring of integers $\mathcal{O}_K$ is a finitely generated $\mathcal{O}_K$-module in D such that $\mathcal{O} \otimes_{\mathcal{O}_K} K = D$ which is also a ring with 1. The number of conjugacy classes of maximal (with respect to inclusion) order of D/K is finite. In particular, if D/K is a skew-field, the elements of integral norm form a maximal order of D and it is the unique maximal order up to conjugation.

2.2. **Hyperbolic 3-space.** Hyperbolic 3-space $\mathcal{H}_3$ is the unique connected, simply connected Riemannian manifold of dimension 3 with constant sectional curvature -1. A standard model for $\mathcal{H}_3$ is the upper half space model

$$\{(x, y) \in \mathbb{C} \times \mathbb{R} \mid y > 0\} = \{(x_1, x_2, y) \in \mathbb{R}^3 \mid y > 0\}$$

with the metric coming from the line element

$$ds^2 = \frac{dx_1^2 + dx_2^2 + dy^2}{y^2}$$

under which the distance $d(P, P')$ between two points $P = (x_1, x_2, y)$,$P' = (x_1', x_2', y')$ is given by

$$\cosh d(P, P') = 1 + \frac{(x_1 - x_1')^2 + (x_2 - x_2')^2 + (y - y')^2}{2yy'}.$$

The orientation-preserving isometries of $\mathcal{H}_3$ can be identified with the group $\mathrm{PSL}_2(\mathbb{C}) \simeq \mathrm{PGL}_2(\mathbb{C})$. The action on $\mathcal{H}_3$ is given as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \left(\frac{(ax + b)\overline{(cx + d)} + a\bar{c}y^2}{|cx + d|^2 + |c|^2y^2}, \frac{y}{|cx + d|^2 + |c|^2y^2} \right)$$

where $z = (x, y) \in \mathcal{H}_3$.

It is computationally convenient to regard $\mathcal{H}_3$ inside the Hamiltonians $\mathbb{H}$. We embed $\mathcal{H}_3$ into $\mathbb{H}$ via $(x_1, x_2, y) \mapsto (x_1, x_2, y, 0)$. In other words, $(x, y) \mapsto x + yj$. In particular, the element $j \in \mathbb{H}$ corresponds to $(0, 0, 1) \in \mathbb{R}^3$. Then the action of $\mathrm{PSL}_2(\mathbb{C})$ on $\mathcal{H}_3$ takes the familiar form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = (az + b)(cz + d)^{-1}$$

where the inverse of $cz + d$ is taken inside the skew-field $\mathbb{H}$.

To see this transition, consider $\mathbb{H}$ as a subalgebra of $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{C} \simeq M_2(\mathbb{C})$ as follows

$$\mathbb{H} \hookrightarrow M_2(\mathbb{C}), \quad x + yj \mapsto \begin{pmatrix} x & -y \\ \bar{y} & \bar{x} \end{pmatrix}.$$

In particular, if $x \in \mathbb{C}$, we have $x \mapsto \left(\begin{smallmatrix} x & 0 \\ 0 & \bar{x} \end{smallmatrix}\right)$. This embedding respects the ring operations of $\mathbb{H}$. It is now trivial to verify the equivalence between the two descriptions of the action of $\mathrm{PSL}_2(\mathbb{C})$ on $\mathcal{H}_3$ that we gave above when we consider this matrix representation of $\mathcal{H}_3 \subset \mathbb{H}$.

2.3. **Arithmetic Kleinian Groups.** Consider a quaternion algebra D/K over a number field K with $n \geq 0$ real places and a unique complex place. Assume that D ramifies at *all* the real places of K. We will need the groups $\mathrm{PGL}_1(D)$ and $\mathrm{PSL}_1(D)$, the group of invertibles elements of D modulo its center and the goup of norm one elements modulo its center, respectively. For example, we have $\mathrm{PSL}_1(M_2(K)) \simeq \mathrm{PSL}_2(K)$ and $\mathrm{PSL}_1(\mathbb{H}) \simeq SU(2)/\{\pm 1\} \simeq SO(3)$, a compact Lie group. Let $\mathcal{O}$ be an order in D. Then $\mathrm{PSL}_1(\mathcal{O})$ is discrete when considered as a subgroup

$$\mathrm{PSL}_1(\mathcal{O}) \hookrightarrow \prod_{v \text{ infinite}} \mathrm{PSL}_1(D \otimes_K K_v) = \mathrm{PSL}_2(\mathbb{C}) \times SO(3)^n.$$

As the other factors are compact, the image of $\mathrm{PSL}_1(\mathcal{O})$ under the projection map onto the factor $\mathrm{PSL}_2(\mathbb{C})$ is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. The map from $\mathrm{PSL}_1(\mathcal{O})$ onto its discrete image in $\mathrm{PSL}_2(\mathbb{C})$ has finite kernel. We will, for convenience, denote this image with $\mathrm{PSL}_1(\mathcal{O})$ as well.

A discrete subgroup Γ of $\mathrm{PSL}_2(\mathbb{C})$ is called a **Kleinian group**. If the associated orbifold (manifold if Γ is torsion-free) $\Gamma \backslash \mathcal{H}_3$ has finite volume, we say that Γ has finite covolume. Two discrete subgroups Γ_1,Γ_2 are *commensurable* (denoted $\Gamma_1 \equiv \Gamma_2$) if their intersection $\Gamma_1 \cap \Gamma_2$ is finite index in both Γ_1 and Γ_2. We say that they are *widely commensurable* if they are commensurable after possibly conjugation. We define the *commensurator* $Comm(\Gamma)$ of a discrete subgroup Γ as

$$Comm(\Gamma) = \{\gamma \in \mathrm{PSL}_2(\mathbb{C}) \mid \gamma \Gamma \gamma^{-1} \equiv \Gamma\}.$$

This is a group that clearly contains Γ.

A Kleinian group Γ is called **arithmetic** if it is widely commensurable with some discrete subgroup $\mathrm{PSL}_1(\mathcal{O})$ that arises in the above way from some order $\mathcal{O}$ in some suitable quaternion algebra D/K over a suitable number field K. In this case, we call K the *field of definition* of Γ and call D/K the *defining quaternion algebra*. It follows from a general result in the theory of arithmetic groups that every arithmetic Kleinian group has finite covolume. Moreover, the associated orbifold is non-compact if and only if $S(D) = \emptyset$. Thus Γ is non-cocompact if and only if K is imaginary quadratic and $D/K \simeq M_2(K)$.

The simplest examples of arithmetic Kleinian groups are the so called *Bianchi groups*. Take K to be imaginary quadratic and $D/K \simeq M_2(K)$. Then $M_2(\mathcal{O}_K)$ is a maximal order and $\mathrm{PSL}_1(M_2(\mathcal{O}_K)) \simeq \mathrm{PSL}_2(\mathcal{O}_K)$ is the Bianchi group associated to K. As mentioned above, any non-cocompact arithmetic lattice in $\mathrm{PSL}_2(\mathbb{C})$ is widely commensurable with a Bianchi group.

Another well-known example of an arithmetic lattice is the complement of the figure eight knot. In fact, Alan Reid proved in [11] that this is the only arithmetic *knot complement*. Its fundamental group is isomorphic to an index 12 subgroup of the Bianchi group $\mathrm{PSL}_2(\mathcal{O}_K)$ associated to $K = \mathbb{Q}(\sqrt{-3})$.

Other well-known examples of arithmetic lattices are provided by the *hyperbolic tetrahedral groups*. A hyperbolic tetrahedral group is the index 2 subgroup consisting of orientation-preserving isometries in the discrete group generated by reflections in the faces of a hyperbolic tetrahedron (with possible ideal vertices) whose dihedral angles are submultiples of π. Lannér proved in 1950 [8] that there are 32 such hyperbolic tetrahedra. It is known that 23 of these groups are arithmetic. They are discussed in detail in [9, 4].

2.4. **Automorphic Forms on $\mathcal{H}_3$.** Let us introduce the automorphic forms that are associated to arithmetic Kleinian groups. These are *vector valued* real analytic functions on $\mathcal{H}_3$ with certain transformation properties satisfying certain differential equations and growth properties.

Given $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{C})$ and $z = x + yj \in \mathcal{H}_3$, let us introduce the multiplier system

$$J(\gamma, z) := \begin{pmatrix} cx + d & -cy \\ \bar{c}y & \overline{cx + d} \end{pmatrix}$$

Given a function $F : \mathcal{H}_3 \to \mathbb{C}^{k+1}$ and $\gamma \in \mathrm{PSL}_2(\mathbb{C})$, we define the *slash operator*

$$(F|_k\gamma)(z) := \sigma^k(J(\gamma, z)^{-1})F(\gamma z)$$

where σ^k is the *symmetric k^{th} power* of the standard representation of $\mathrm{PSL}_2(\mathbb{C})$ on $\mathbb{C}^2$.

The case $k = 2$ will be especially important for us. In this case we have $F : \mathcal{H}_3 \to \mathbb{C}^3$ and

$$(F|_2\gamma)(z) = \frac{1}{|r|^2 + |s|^2} \begin{pmatrix} \bar{r}^2 & 2\bar{r}s & s^2 \\ -\bar{r}\bar{s} & |r|^2 - |s|^2 & rs \\ \bar{s}^2 & -2r\bar{s} & r^2 \end{pmatrix} F(\gamma z)$$

where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $r = cx + d$ and $s = cy$.

The center of the universal enveloping algebra of the Lie algebra associated to the real Lie group $\mathrm{PSL}_2(\mathbb{C})$ is generated by two elements Ψ, Ψ'. These act on real analytic functions $F : \mathcal{H}_3 \to \mathbb{C}^{k+1}$ as differential operators.

Let Γ be an arithmetic lattice in $\mathrm{PSL}_2(\mathbb{C})$ with defining field K and defining quaternion algebra D/K. An **automorphic form** for Γ with weight k and eigenvalues (λ, λ') is a real analytic function $F : \mathcal{H}_3 \to \mathbb{C}^{k+1}$ with the following properties.

(1) $F|_k\gamma = F$ for every $\gamma \in \Gamma$,

(2) $\Psi F = \lambda F$ and $\Psi' F = \lambda' F$,
(3) if Γ is non-cocompact, then F has at worst polynomial growth at each cusp.

The set $M(\Gamma, k, \lambda, \lambda')$ of automorphic forms for Γ with weight k and eigenvalues (λ, λ') is a finite dimensional complex vector space.

Let $\beta_1 := -\frac{dx}{y}, \beta_2 := \frac{dy}{y}, \beta_3 := \frac{d\bar{x}}{y}$ be a basis of differential 1-forms on $\mathcal{H}_3$. A differential form ω is *harmonic* if $\Delta\omega = 0$ where $\Delta = d \circ \delta + \delta \circ d$ is the usual Laplacian with d being the exterior derivative and δ the codifferential operator. Then $\mathrm{PSL}_2(\mathbb{C})$ acts on the space of differential 1-forms as

$$\gamma \cdot {}^t(\beta_1, \beta_2, \beta_3)_{(z)} = \sigma^2(J(\gamma, z))^t(\beta_1, \beta_2, \beta_3)_{(z)}.$$

A **weight 2 cuspidal modular form** for Γ is a real analytic function $F = (F_1, F_2, F_3) : \mathcal{H}_3 \to \mathbb{C}^3$ with the following properties.

(1) $F_1\beta_1 + F_2\beta_2 + F_3\beta_3$ is a harmonic differential 1-form on $\mathcal{H}_3$ that is Γ-invariant,
(2) If Γ is non-cocompact, then $\int_{\mathbb{C}/\mathcal{O}_K} (F|_2\gamma)(x, y)dx = 0$ for every $\gamma \in \mathrm{PSL}_2(\mathcal{O}_K)$.

The last condition, in which case K is necessarily an imaginary quadratic field, is equivalent to saying that the constant coefficient in the *Fourier-Bessel expansion* of $F|\gamma$ is equal to zero for every $\gamma \in \mathrm{PSL}_2(\mathcal{O}_K)$. Let us explain this. As Γ is non-cocompact, it contains parabolic elements. The Γ-invariance of F, which is implicit in condition (1), implies that F is a periodic function in the $x = (x_1, x_2)$-variable. It follows that the F has a Fourier-Bessel expansion of the form

$$F(x, y) = \sum_{0 \neq \alpha \in \mathcal{O}_K} c(\alpha) y^2 \mathbb{K}\left(\frac{4\pi|\alpha|y}{\sqrt{|\triangle|}}\right) \psi\left(\frac{\alpha x}{\sqrt{\triangle}}\right)$$

where

$$\psi(x) = e^{2\pi(x+\bar{x})}$$

and

$$\mathbb{K}(t) = \left(-\frac{i}{2}K_1(y), K_0(y), \frac{i}{2}K_1(y)\right)$$

with K_0, K_1 are the hyperbolic Bessel functions satisfying the differential equation

$$\frac{dK_j}{dy^2} + \frac{1}{y}\frac{dK_j}{dy} - \left(1 + \frac{1}{y^{2j}}\right)K_j = 0, \qquad j = 0, 1$$

and decreases rapidly at infinity.

The space of weight 2 cuspidal modular forms for a fixed Γ form a finite dimensional complex vector space which we will denote with $S_2(\Gamma)$. As automorphic forms for Γ, we have $S_2(\Gamma) \subset M(\Gamma, 2, 0, 0)$.

2.5. **Hecke Operators on the Cohomology.** In this section we will consider the cohomology groups $H^1(\Gamma\backslash\mathcal{H}_3,\mathbb{Z})$ for Kleinian groups Γ. When Γ is arithmetic, there is a very special infinite family of operators, called *Hecke operators*, that act on these cohomology groups. We will study these operators.

Let Γ be any cofinite Kleinian group and L a field. Given $g \in Comm(\Gamma)$, let us consider the 3-folds M, M_g, M^g associated to the lattices $\Gamma, \Gamma_g := \Gamma \cap g\Gamma g^{-1}, \Gamma^g := \Gamma \cap g^{-1}\Gamma g$ respectively. We have *finite* coverings, induced by inclusion of fundamental groups,

$$r_g : M_g \to M, \qquad r^g : M^g \to M$$

and an isometry

$$\tau : M_g \to M^g$$

induced by conjugation by g isomorphism between Γ_g and Γ^g. The composition $s_g := r^g \circ \tau$ gives us a second finite covering from M_g to M. The coverings r_g induce linear maps between the homology groups

$$r_g^i : H^i(M,L) \to H^i(M_g,L).$$

The process of summing the finitely many preimages in M_g of a point of M under s_g leads to

$$s_g^* : H^i(M_g,L) \to H^i(M,L).$$

Note that s_g^* is equivalent to the composition

$$H^i(M_g,L) \to H^i(M^g,L) \to H^i(M,L)$$

where the first arrow is induced by τ, and the second arrow is simply the corestriction map (which corresponds to the transfer map of group cohomology). We define the **Hecke operator** T_g associated to $g \in Comm(\Gamma)$ as the composition

$$T_g := s_g^* \circ r_g^* : H^i(M,L) \to H^i(M,L).$$

There is a notion of isomorphism of Hecke operators that we shall not present. It turns out that up to isomorphism, T_g depends only on the double coset $\Gamma g \Gamma$.

One can define Hecke operators using the above process for homology groups as well.

It is a classical result of Margulis that the commensurator of Γ is dense in $\mathrm{PSL}_2(\mathbb{C})$ if and only if Γ is arithmetic. The commensurator of a Bianchi group $\mathrm{PSL}_2(\mathcal{O}_K)$ is $\mathrm{PGL}_2(K)$. More generally, the commensurator of an arithmetic lattice Γ in $\mathrm{PSL}_2(\mathbb{C})$ with defining quaternion algebra D/K is $\mathrm{PGL}_1(D) \subset \mathrm{PSL}_2(\mathbb{C})$.

Especially important is the subfamily $\mathbb{T}$ of Hecke operators associated to the prime elements of the ring of integers $\mathcal{O}_K$ of the definition field K. More precisely, let π be a prime element of $\mathcal{O}_K$. Then we have $\left(\begin{smallmatrix}\pi & 0\\ 0 & 1\end{smallmatrix}\right) \in Comm(\Gamma)$ (note that here we regard $D/K \subset M_2(\overline{K})$). If T_π is the Hecke operator associated to $\left(\begin{smallmatrix}\pi & 0\\ 0 & 1\end{smallmatrix}\right)$, then we set

$$\mathbb{T} = \{T_\pi \mid \pi \in \mathcal{O}_K\}.$$

2.6. **Cohomology and Automorphic Forms.** Let Γ be an arithmetic lattice in $\mathrm{PSL}_2(\mathbb{C})$ with associated 3-fold Y.

Assume that Y is non-compact. Without loss of generality, we assume that Γ is finite index subgroup of a Bianchi group. Borel and Serre constructed in [3] a compact 3-fold X with boundary (known as the *Borel-Serre compactification* of Y), such that the interior of X is homeomorphic to Y and the embedding $Y \hookrightarrow X$ is a homotopy equivalence. This implies that $H^i(X,\mathbb{C}) \simeq H^i(Y,\mathbb{C})$ for every $i \geq 0$. The boundary ∂X of X is a disjoint union of 2-tori[1], each one corresponding to a cusp of Y. In fact, topologically, we can think of X as the manifold obtained by attaching a 2-torus at infinity to each cusp of Y. It is a classical result that the number of cusps of Y when Γ is the Bianchi group associated to K is equal to the class number of K.

Consider the map $res : H^i(X,\mathbb{C}) \to H^i(\partial X,\mathbb{C})$ given by the restriction to the boundary. The kernel of this map gives a subspace of $H^i(Y,\mathbb{C})$ which is called the *cuspidal cohomology*, denoted $H^i_{cusp}(Y,\mathbb{C})$. Harder constructed in [6] a certain section of the above restriction map which gives a subspace of $H^i(Y,\mathbb{C})$ which is called the *Eisenstein cohomology* such that the decomposition

$$H^i(Y,\mathbb{C}) = H^i_{cusp}(Y,\mathbb{C}) \oplus H^i_{Eis}(Y,\mathbb{C})$$

is invariant under the action of the special family $\mathbb{T}$ of Hecke operators. Observe that when Y is compact, we do not have the Eisenstein cohomology anymore, that is $H^i(Y,\mathbb{C}) = H^i_{cusp}(Y,\mathbb{C})$. In a natural sense, while the Eisenstein cohomology comes from the contribution of the boundary of the Borel-Serre compactification X, the cuspidal cohomology belongs to the interior of X.

While one has a good control over the Eisenstein cohomology, the cuspidal part is very mysterious. The importance of cuspidal cohomology comes from the fact that it can be identified with certain types of automorphic forms called "cohomological".

Now let us get back to the general case where Y is not necessarily non-compact. Let $S_2(\Gamma)$ denote the space of weight 2 cuspidal modular forms for an arithmetic lattice Γ as discussed above. Then there is an isomorphism, called **generalized Eichler-Shimura Isomorphism**

$$S_2(\Gamma) \simeq H^1_{cusp}(\Gamma,\mathbb{C}) \simeq H^2_{cusp}(\Gamma,\mathbb{C})$$

which respects the Hecke action on $S_2(\Gamma)$ and the action of $\mathbb{T}$ on $H^i_{cusp}(\Gamma,\mathbb{C})$. The fact that this is not just an isomorphism of vector spaces but of Hecke modules is crucial for number theory.

[1]Except when $K = \mathbb{Q}(\sqrt{-1}), \mathbb{Q}(\sqrt{-3})$ where one has units other than ± 1 in $\mathcal{O}_K$.

Roughly speaking, the above isomorphism comes from the facts that weight 2 cuspidal modular forms are essentially harmonic differential 1-forms and every cohomology class in the de Rahm cohomology

$$H^1_{dR}(\Gamma\backslash\mathbb{H},\mathbb{R}) \hookrightarrow H^1(\Gamma\backslash\mathbb{H},\mathbb{R}) \simeq H^1(\Gamma,\mathbb{R})$$

can be represented by a harmonic differential 1-form. The general connection between (vector valued) harmonic differential k-forms and cohomological automorphic forms was studied by Matsushima and Murakami, see [10], in the cocompact setting.

All of the above was studied more generally by Harder in [6, 7] for the algebraic group $\mathrm{Res}_{K/\mathbb{Q}}(\mathrm{PGL}_2)$ for any number field K. In [5] Franke generalized this sort of connection to the fullest, that is, to the case of arithmetic lattices in general real Lie groups and their associated modular forms. Today most of the popular methods for computing with modular forms, such as the modular symbols method, is based on this passage to the (co)homology.

3. Torsion

In the rest of this note, we will focus on the torsion in the homology groups $H_1(\Gamma,\mathbb{Z})$. First, let us motivate this shift of focus from the complex cohomology to integral cohomology.

Let Γ be a arithmetic Kleinian group. Then $H_1(\Gamma,\mathbb{Z})$ is a finitely generated abelian group which is isomorphic to $Tor \oplus \mathbb{Z}^\beta$ where $Tor \simeq H_1(\Gamma,\mathbb{Z})_{tors}$ is a finite abelian group and β is the Betti number of the associated 3-fold. Consider the diagram

$$\begin{tikzcd} H_1(\Gamma,\mathbb{Z}) \arrow[r, "\otimes\mathbb{C}"] \arrow[d, "\otimes\mathbb{F}_p"] & H_1(\Gamma,\mathbb{C}) \\ H_1(\Gamma,\mathbb{F}_p) & \end{tikzcd}$$

Notice that $\mathbb{F}_p$-vector space $H_1(\Gamma,\mathbb{F}_p)$ admits an action of the Hecke operators $\mathbb{T}$ as well. It has been believed since the first computations of Fritz Grunewald, and later work of Avner Ash in the setting of $\mathrm{GL}_3(\mathbb{Z})$ that the Hecke eigenvalue systems captured in these $\mathbb{F}_p$-vector spaces are intimately related to mod p Galois representations. Indeed, a recently announced result [12] of Peter Scholze proves this to be the case more generally for GL_n over CM fields.

Theorem 3.1. *(Scholze) Let $\Psi = \{\Psi(\mathfrak{p})\}$ be a Hecke eigenvalue system captured in some $H_1(\Gamma,\mathbb{F}_p)$. Then there is a continuous representation*

$$\rho : Gal(\overline{\mathbb{Q}}/K) \to GL_2(\overline{\mathbb{F}}_p)$$

such that except for finitely many prime ideals, one has

$$Tr(\rho(Frob_{\mathfrak{p}})) = \Psi(\mathfrak{p}).$$

The issue with the torsion is that the very possible existence of p-torsion in $H_1(\Gamma, \mathbb{Z})$ can give rise to situations where a Hecke eigenvalue system captured in $H_1(\Gamma, \mathbb{F}_p)$ does not arise from a mod p reduction of a Hecke eigenvalue system captured in $H_1(\Gamma, \mathbb{C})$. In a slogan form,

"*torsion in the homology gives rise to* genuine *mod p arithmetic data.*"

Thus we are interested in understanding the nature of torsion, starting with its mere size.

3.1. **Asymptotics of Torsion.** How much torsion is there? The short answer is "A lot!". Explicit computations show that the size of torsion is supported by sporadic primes of astronomical sizes. Here is a sample. Let $K = \mathbb{Q}(\sqrt{-11})$. Then $4999 \cdot \mathcal{O}_K = \mathfrak{p}\bar{\mathfrak{p}}$ and a machine computation shows that

$$\underbrace{3527\ldots 5847}_{\text{a 53 digit prime}} \mid \#H_1(\Gamma_0(\mathfrak{p}), \mathbb{Z})_{tors}.$$

In recent work [2], Bergeron and Venkatesh spell out a conjectural picture that describes the behaviour of torsion asymptotically. In fact, in our case they prove the following.

Theorem 3.2. *(Bergeron-Venkatesh) Let* $\{\Gamma_n\}_{n\geq 1}$ *be a decreasing tower of cocompact congruence type arithmetic Kleinian groups such that* $\bigcap_n \Gamma_n = \{1\}$. *Let* E *be a* Γ_0*-invariant lattice in one of the standard representations* $Sym^k \otimes \overline{Sym}^\ell$ *of the real Lie group* $\mathrm{PSL}_2(\mathbb{C})$ *with* $k \neq \ell$. *Then*

$$\lim_{n\to\infty} \frac{\log |H_1(\Gamma_n, E)_{tors}|}{vol(\Gamma_n \backslash \mathcal{H}_3)} = c_{k,\ell},$$

where

$$c_{k,\ell} := \frac{1}{6\pi} \cdot \frac{1}{2^3} \cdot \Big((k+\ell+2)^3 - |k-\ell|^3 + 3(k+\ell+2)(k-\ell)(k+\ell+2-|k-\ell|)\Big).$$

In a nutshell, the theorem says that

"*Torsion grows exponentially with the volume.*"

Experiments carried out in strongly suggested that the above theorem should stay valid in the setting of non-compact arithmetic Kleinian groups with $E = \mathbb{Z}$ (note that this lattice lives in the representation $\mathbb{C} \simeq Sym^0 \otimes \overline{Sym}^0$ which violates the above hypothesis $k \neq \ell$).

In recent work [1] with Bergeron and Venkatesh, we formulated a conjecture on the topological complexity of arithmetic hyperbolic 3-manifolds (these are hyperbolic 3-manifolds whose fundamental groups are arithmetic Kleinian groups) which implies an asymptotic growth result like the above for $E = \mathbb{Z}$ (note that in this case we have $c_{0,0} = 1/(6\pi)$).

Conjecture 3.3. *There is a constant $C = C(M_0)$ such that, for any arithmetic congruence hyperbolic 3-manifold $M \to M_0$ of volume V, there exist immersed surfaces S_i of genus $\leq V^C$ such that the $[S_i]$ span $H_2(M, \mathbb{R})$.*

Theorem 3.4. *Let $(M_i \to M_0)_{i\in\mathbb{N}}$ be a sequence of arithmetic congruence hyperbolic 3-manifolds s.t. M_0 is compact and $V_i = vol(M_i)$ goes to infinity. Assume the following two conditions are satisfied:*

(i) *'Few small eigenvalues': For every $\varepsilon > 0$ there exists some positive real number c such that*

$$\limsup_{i\to\infty} \frac{1}{V_i} \sum_{0<\lambda\leq c} |\log \lambda| \leq \varepsilon.$$

Here λ ranges over eigenvalues of the first Laplacian Δ on M_i.

(ii) *'Small Betti numbers': $\beta(M_i, \mathbb{Q}) = o(\frac{V_i}{\log V_i})$.*

Then, if Conjecture 3.3 holds, as $i \to \infty$, we have:

$$\lim_{n\to\infty} \frac{\log \# H_1(M_i, \mathbb{Z})_{tors}}{V_i} = \frac{1}{6\pi}.$$

We prove Conjecture 3.3 in two cases.

Theorem 3.5. *Conjecture 3.3 is true in the two following cases:*

(i) *When M_0 arises from a division algebra $D \otimes F$ where D is a quaternion algebra over $\mathbb{Q}$ and F is an imaginary quadratic field, $\pi_1 M$ is a principal congruence subgroup such that all of $S_2(\pi_1 M)$ arise from classical elliptic modular forms over $\mathbb{Q}$ via the "Base-Change" functoriality of Langlands;*

(ii) *When M_0 is a Bianchi manifold, and $S_2(\pi_1 M)$ is 1-dimensional, associated with a non-CM elliptic curve, for which we assume the equivariant B-SD conjecture and the Frey–Szpiro conjecture.*

Furthermore, we carry out explicit machine computations which reveal that the above two cases in fact occur very often.

References

[1] N.Bergeron, M.H.Şengün and A. Venkatesh. *Torsion Homology Growth and Cycle Complexity of Arithmetic Manifolds.* preprint (2013).

[2] N.Bergeron and A.Venkatesh. *The Asymptotic Growth of Torsion Homology for Arithmetic Groups.* J. Inst. Math. Jussieu 12.2 (2013): 391–447.

[3] A.Borel and J.-P. Serre. *Corners and Arithmetic Groups.* Comment. Math. Helv. 48 (1973): 436–91.

[4] J.Elstrodt, F.Grunewald and J. Mennicke. *Groups Acting on Hyperbolic Space.* Berlin: Springer-Verlag, 1998.

[5] J.Franke. *Harmonic Analysis in Weighted L_2-Spaces.* Ann. Sci. École Norm. Sup. (4) 31.2 (1998): 181–279.

[6] G.Harder. *On the Cohomology of $SL(2, \mathcal{O})$.* Lie groups and their representations (Proc. Summer School on Group Representations of the Bolyai János Math. Soc., Budapest, 1971) (1975): 139–50.

[7] G.Harder. *Eisenstein Cohomology of Arithmetic Groups. The Case GL_2.* Invent. Math. 89.1 (1987): 37–118.

[8] F. Lannér. *On Complexes With Transitive Groups of Automorphisms.* Comm. Sem. Math. Univ. Lund. **11** (1950).

[9] C.Maclachlan and A.W.Reid. *The Arithmetic of Hyperbolic 3-Manifolds.* Vol. 219. New York: Springer-Verlag, 2003.

[10] Y.Matsushima and S.Murakami. *On Vector Bundle Valued Harmonic Forms and Automorphic Forms on Symmetric Riemannian Manifolds.* Ann. of Math. (2) 78 (1963): 365–416.

[11] A.W.Reid. *Arithmeticity of Knot Complements.* J. London Math. Soc. (2) 43.1 (1991): 171–84.

[12] P.Scholze. *On Torsion in the Cohomology of Locally Symmetric Varieties.* preprint (2013).

E-mail address: `M.H.Sengun@warwick.ac.uk`
URL: `http://warwick.ac.uk/haluksengun`

MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY, UK

SPLITTINGS OF KNOT GROUPS

STEFAN FRIEDL, DANIEL S. SILVER, AND SUSAN G. WILLIAMS

Abstract. Let K be a knot of genus g. If K is fibered, then it is well known that the knot group $\pi(K)$ splits only over a free group of rank $2g$. We show that if K is not fibered, then $\pi(K)$ splits over non-free groups of arbitrarily large rank. Furthermore, if K is not fibered, then $\pi(K)$ splits over every free group of rank at least $2g$. However, $\pi(K)$ cannot split over a group of rank less than $2g$. The last statement is proved using the recent results of Agol, Przytycki–Wise and Wise.

1. Introduction

We start out with a few definitions from group theory. Let π be a group. We say that π *splits over the subgroup* B if π admits an HNN decomposition with base group A and amalgamating subgroup B. More precisely, π splits over the subgroup B if there exists an isomorphism

$$\pi \xrightarrow{\cong} \langle A, t \,|\, \varphi(b) = tbt^{-1} \text{ for all } b \in B\rangle,$$

where $B \subset A$ are subgroups of π and $\varphi\colon B \to A$ is a monomorphism. In this notation, relations of A are implicit. We will write such a presentation more compactly as $\langle A, t \,|\, \varphi(B) = tBt^{-1}\rangle$.

In this paper we are interested in splittings of knot groups. Given a knot $K \subset S^3$ we denote the knot group $\pi_1(S^3 \setminus K)$ by $\pi(K)$. We denote by $g(K)$ the genus of the knot, the minimal genus of a Seifert surface Σ for K. It follows from the Loop Theorem and the Seifert-van Kampen theorem that we can split the knot group $\pi(K)$ over the free group $\pi_1(\Sigma)$ of rank $2g(K)$. The *rank* $\mathrm{rk}(G)$ of a group G is the minimal size of a set of generators for G.

It is well known that if K is a fibered knot, that is, the knot complement $S^3 \setminus K$ fibers over S^1, then the group $\pi(K)$ splits only over free groups of rank $2g(K)$. (See, for example, Lemma 3.1.) We show that this property characterizes fibered knots. In fact, we can say much more.

Theorem 1.1. *Let K be a non-fibered knot. Then $\pi(K)$ splits over non-free groups of arbitrarily large rank.*

Date: August 29, 2013.

The second and third authors were partially supported by grants #245671 and #245615 from the Simons Foundation.

Neuwirth [Ne65, Problem L] asked whether there exists a knot K such that $\pi(K)$ splits over a *free* group of rank other than $2g(K)$. By the above, such a knot would necessarily have to be non-fibered. Lyon [Ly71, Theorem 2] showed that there does in fact exist a non-fibered genus-one knot K with incompressible Seifert surfaces of arbitrarily large genus. This implies in particular that there exists a knot K for which $\pi(K)$ splits over free groups of arbitrarily large rank. We give a strong generalization of this result.

Theorem 1.2. *Let K be a non-fibered knot. Then for any integer $k \geq 2g(K)$ there exists a splitting of $\pi(K)$ over a free group of rank k.*

Note that an incompressible Seifert surface gives rise to a splitting over a free group of *even rank*. The splittings over free groups of *odd rank* in the theorem are therefore not induced by incompressible Seifert surfaces.

Feustel and Gregorac [FG73] showed that if N is an aspherical, orientable 3-manifold such that $\pi = \pi_1(N)$ splits over the fundamental group of a *closed* surface $\Sigma \neq S^2$, then this splitting can be realized topologically by a properly embedded surface. (More splitting results can be found in [CS83, Proposition 2.3.1].) The fact that fundamental groups of non-fibered knots can be split over free groups of odd rank shows that the result of Feustel and Gregorac does not hold for splittings over fundamental groups of surfaces with boundary.

Theorems 1.1 and 1.2 can be viewed as strengthenings of Stallings's fibering criterion. We refer to Section 7 for a precise statement.

Our third main theorem shows that Theorem 1.2 is optimal.

Theorem 1.3. *If K is a knot, then $\pi(K)$ does not split over a group of rank less than $2g(K)$.*

The case $g(K) = 1$ follows from the Kneser Conjecture and work of Waldhausen [Wal68b], as we show in Section 8.1. However, to the best of our knowledge, the classical methods of 3-manifold topology do not suffice to prove Theorem 1.3 in the general case. We use the recent result [FV12a] that Wada's invariant detects the genus of any knot. This result in turn relies on the seminal work of Agol [Ag08, Ag12], Wise [Wi09, Wi12a, Wi12b], Przytycki–Wise [PW11, PW12a] and Liu [Liu11].

Theorem 1.3 is of interest for several reasons:

(1) It gives a completely group-theoretic chararcterization of the genus of a knot, namely

$$g(K) = \frac{1}{2}\min\{\mathrm{rk}(B) \mid \pi(K) \text{ splits over the group } B\}.$$

A different group-theoretic characterization was given by Calegari (see the proof of Proposition 4.4 in [Ca09]) in terms of the 'stable commutator length' of the longitude.

(2) Theorem 1.3 fits into a long sequence of results showing that minimal-genus Seifert surfaces 'stay minimal' even if one relaxes some conditions. For example, Gabai [Ga83] showed that the genus of an *immersed* surface cobounding a longitude of K is at least $g(K)$. Furthermore, minimal-genus Seifert surfaces give rise to surfaces of minimal complexity in the 0-framed surgery N_K (see [Ga87]) and in most S^1-bundles over N_K (see [Kr99, FV12b]).

(3) Given a closed 3-manifold N it is obvious that $\mathrm{rk}(\pi_1(N))$ is a lower bound for the Heegaard genus $g(N)$ of N. In light of Theorem 1.3 one might hope that this is in an equality; that is, that $\mathrm{rk}(\pi_1(N)) = g(N)$. This is not the case, though, as was shown by various authors (see [BZ84, ScW07] and [Li13]).

The paper is organized as follows. In Section 2 we discuss several basic facts about HNN decompositions of groups. In Section 3 we recall that incompressible Seifert surfaces give rise to HNN decompositions of knot groups and we characterize in Lemma 3.1 the splittings of fundamental groups of fibered knots. In Section 4 we consider the genus-one non-fibered knot $K = 5_2$. We give explicit examples of splittings of the knot group over a non-free group and over the free group F_3 of rank 3, and inequivalent splittings of the knot group over F_2.

Section 5 contains the proof of Theorem 1.1, and in Section 6 we give the proof of Theorem 1.2. In Section 7 we show that these two theorems strengthen Stallings's fibering criterion. In Section 8.1 we give a proof of Theorem 1.3 for genus-one knots. The proof relies mostly on the Kneser Conjecture and a theorem of Waldhausen. In Section 8.2 we review the definition of Wada's invariant of a group. Finally, in Section 8.3 we prove Theorem 8.5, which combined with the main result of [FV12a] provides a proof of Theorem 1.3 for all genera.

We conclude this introduction with two questions. The precise notions are explained in Section 2.

(1) Let π be a word hyperbolic group and let $\epsilon\colon \pi \to \mathbb{Z}$ be an epimorphism such that $\mathrm{Ker}(\epsilon)$ is not finitely generated. Does (π, ϵ) admit splittings over (infinitely many) pairwise non-isomorphic groups? (The group $\pi = \pi(K)$ satisfies these conditions if K is a non-fibered knot.)

(2) Let K be a non-fibered knot of genus g. Does $\pi(K)$ admit (infinitely many) inequivalent splittings over the free group F_{2g} on $2g$ generators?

Conventions and notations. All groups are assumed to be finitely presented unless we say specifically otherwise. All 3-manifolds are assumed to be connected, compact and orientable. Given a submanifold X of a 3-manifold N, we denote by $\nu X \subset N$ an open tubular neighborhood of X in N. Given $k \in \mathbb{N}$ we denote by F_k the free group on k generators.

Acknowledgments. The first author wishes to thank the University of Sydney for its hospitality. We are also very grateful to Eduardo Martinez-Pedroza, Saul Schleimer and Henry Wilton for very helpful conversations.

2. Hnn-decompositions and splittings of groups

2.1. **Splittings of groups.** An *HNN decomposition* of a group π is a 4-tuple (A, B, t, φ) consisting of subgroups $B \le A$ of π, a *stable letter* $t \in \pi$, and an injective homomorphism $\varphi\colon B \to A$, such that the natural inclusion maps induce an isomorphism from $\langle A, t \,|\, \varphi(B) = tBt^{-1}\rangle$ to π. Alternatively, a HNN-decomposition of π can be viewed as an isomorphism

$$f\colon \pi \xrightarrow{\cong} \langle A, t \,|\, \varphi(B) = tBt^{-1}\rangle$$

where $\varphi\colon B \to A$ is an injective map. We will frequently go back and forth between these two points of view.

We need a few more definitions:

(1) Given an HNN-decomposition (A, B, t, φ) we refer to the homomorphism $\epsilon\colon \pi \to \mathbb{Z}$ that is given by $\epsilon(t) = 1$ and $\epsilon(a) = 0$ for $a \in A$ as the *canonical epimorphism*.

(2) Let π be a group and let $\epsilon \in \mathrm{Hom}(\pi, \mathbb{Z})$ be an epimorphism. A *splitting of* (π, ϵ) *over a subgroup* B *(with base group* A*)* is an HNN decomposition (A, B, t, φ) of π such that ϵ equals the canonical epimorphism. With the alternative point of view explained above, a splitting of (π, ϵ) is an isomorphism

$$f\colon \pi \xrightarrow{\cong} \langle A, t \,|\, \varphi(B) = tBt^{-1}\rangle$$

such that the following diagram commutes:

$$\begin{tikzcd}
\pi \arrow[rr, "f"] \arrow[dr, "\epsilon"'] & & \langle A, t \,|\, \varphi(B) = tBt^{-1}\rangle \arrow[dl, "\psi"] \\
& \mathbb{Z} &
\end{tikzcd}$$

where ψ denotes the canonical epimorphism.

(3) Two splittings (A, B, t, φ) and (A', B', t', φ') of (π, ϵ) are called *weakly equivalent* if there exists an automorphism Φ of π with $\Phi(B) = B'$. If Φ can be chosen to be an inner automorphism of π, then the two HNN decompositions are said to be *strongly equivalent*.

We conclude this section with the following well-known lemma of [BS78]. It appears as Theorem B* in [Str84] where an elementary proof can be found.

Lemma 2.1. *Let π be a finitely presented group and let $\epsilon \in Hom(\pi, \mathbb{Z})$ be an epimorphism. Then there exists a splitting*

$$f : \pi \xrightarrow{\cong} \langle A, t \,|\, \varphi(B) = tBt^{-1}\rangle$$

of (π, ϵ) where A and B are finitely generated.

2.2. **Splittings of pairs (π, ϵ) with finitely generated kernel.** The following lemma characterizes splittings of pairs (π, ϵ) for which $\mathrm{Ker}(\epsilon)$ is finitely generated.

Lemma 2.2. *Let π be a finitely presented group, $\epsilon\colon \pi \to \mathbb{Z}$ an epimorphism, and t an element of π with $\epsilon(t) = 1$. If $Ker(\epsilon)$ is finitely generated, then there exists a canonical isomorphism*

$$\pi = \langle B, t \,|\, \varphi(B) = tBt^{-1}\rangle$$

where $B := Ker(\epsilon)$ and where $\varphi\colon B \to B$ is given by conjugation by t. Furthermore, any other splitting of (π, ϵ) is strongly equivalent to this splitting.

Proof. Let π be a finitely presented group and let $\epsilon\colon \pi \to \mathbb{Z}$ be an epimorphism such that $B = \mathrm{Ker}(\epsilon)$ is finitely generated. We have an exact sequence

$$1 \to B \to \pi \xrightarrow{\epsilon} \mathbb{Z} \to 0.$$

Let $t \in \pi$ with $\epsilon(t) = 1$. The map $n \mapsto t^n$ defines a right-inverse of ϵ, and we see that B is canonically isomorphic to the semi-direct product $\langle t\rangle \ltimes B$ where t^n acts on B by conjugation by t^n. That is, we have a canonical isomorphism

$$\pi = \langle B, t \,|\, \varphi(B) = tBt^{-1}\rangle.$$

We now suppose that we have another splitting $\pi = \langle C, s \,|\, \psi(D) = sDs^{-1}\rangle$ of (π, ϵ). By our hypothesis the group $B = \mathrm{Ker}(\epsilon)$ is finitely generated. On the other hand, it follows from standard results in the theory of graphs of groups (see [Se80]) that

$$\mathrm{Ker}(\epsilon) \cong \cdots C_k *_{D_k} C_{k+1} *_{D_{k+1}} C_{k+2} \cdots ,$$

where $C_i = C$ and $D_i = D$ for all $i \in \mathbb{Z}$ and each map $D_i \to C_{i+1}$ is given by ψ.

As in [Ne65], the fact that the infinite free product with amalgamation is finitely generated implies that $C_i = D_i = \psi(D_{i-1})$ for all $i \in \mathbb{Z}$. This, in turn, implies that each C_i and D_i is isomorphic to $D = \mathrm{Ker}(\epsilon)$. It is now clear that the identity on π already has the desired property relating the two splittings of (π, ϵ). □

2.3. **Induced splittings of groups.** Let

$$\pi = \langle A, t \,|\, \varphi(B) = tBt^{-1}\rangle$$

be an HNN-extension. Given $n \le m \in \mathbb{N}$ we denote by $A_{[n,m]}$ the result of amalgamating the groups t^iAt^{-i}, $i = n, \dots, m$ along the subgroups $t^i\varphi(B)t^{-i} = t^{i+1}Bt^{-i-1}$, $i = n, \dots, m-1$. In our notation,

$$A_{[n,m]} = \langle *_{i=m}^{n} t^iAt^{-i} \mid t^j\varphi(B)t^{-j} = t^{j+1}Bt^{-j-1} \ (j = n, \dots, m-1)\rangle.$$

Given any $k \le m \le n \le l$, we have a canonical map $A_{[m,n]} \to A_{[k,l]}$ which is a monomorphism (see, for example, [Se80] for details). If $\epsilon\colon \pi \to \mathbb{Z}$ is the canonical epimorphism, then it is well known that $\mathrm{Ker}(\epsilon)$ is given by the direct limit of the groups $A_{[-m,m]}$, $m \in \mathbb{N}$; that is,

$$\mathrm{Ker}(\epsilon) = \lim_{m\to\infty} A_{[-m,m]}.$$

The following well-known lemma shows that a splitting of a pair (π, ϵ) gives rise to a sequence of splittings.

Lemma 2.3. *Let*
$$\pi = \langle A, t \,|\, \varphi(B) = tBt^{-1} \rangle$$
be an HNN-extension. For any integer $n \geq 0$, let
$$\varphi_n : \pi_1(A_{[0,n]}) \to A_{[1,n+1]}$$
be the map that is given by conjugation by t. Then the obvious inclusion maps induce an isomorphism
$$\langle A_{[0,n+1]}, t \,|\, \varphi_n(A_{[0,n]}) = tA_{[0,n]}t^{-1} \rangle \xrightarrow{\iota} \pi = \langle A, t \,|\, \varphi(B) = tBt^{-1} \rangle.$$

Proof. We write
$$\Gamma = \langle A_{[0,n+1]}, t \,|\, \varphi_n(A_{[0,n]}) = tA_{[0,n]}t^{-1} \rangle.$$
We denote by π' (respectively Γ') the kernel of the canonical map from π (respectively Γ) to $\mathbb{Z}$. It is clear that it suffices to show that the restriction of $\iota : \Gamma \to \pi$ to $\pi' \to \Gamma'$ is an isomorphism.

For $i \in \mathbb{Z}$, we write $A_i := t^iAt^{-i}$ and $B_i := \varphi(t^{i+1}Bt^{-i-1})$. Note that Γ' is canonically isomorphic to
$$\cdots (A_0 *_{B_0} \cdots *_{B_n} A_{n+1}) *_{A_1 *_{B_1} \cdots *_{B_n} A_{n+1}} \big(A_1 *_{B_1} \cdots *_{B_{n+1}} A_{n+2}\big) *_{A_2 *_{B_2} \cdots *_{B_{n+1}} A_{n+2}} \cdots,$$
and π' is canonically isomorphic to
$$\cdots *_{B_0} A_{-1} *_{B_{-1}} A_0 *_{B_0} A_1 *_{B_1} * \cdots$$
It is now straightforward to see that ι does indeed restrict to an isomorphism $\Gamma' \to \pi'$. □

Note that the isomorphism in Lemma 2.3 is canonical. Throughout the paper we will therefore make the identification
$$\pi = \langle A_{[0,n+1]}, t \,|\, \varphi_n(A_{[0,n]}) = tA_{[0,n]}t^{-1} \rangle.$$
In the paper we will also write $A = A_{[0,0]}$.

3. Splittings of knot groups and incompressible surfaces

Now let $K \subset S^3$ be a knot, that is, an oriented embedded simple closed curve in S^3. We write $X(K) := S^3 \setminus \nu K$ and
$$\pi(K) := \pi_1(X(K)) = \pi_1(S^3 \setminus \nu K).$$
The orientation of K gives rise to a canonical epimorphism $\epsilon_K \colon \pi(K) \to \mathbb{Z}$ sending the oriented meridian to 1.

Let Σ be a Seifert surface of genus g for K; that is, a connected, orientable, properly embedded surface Σ of genus g in $X(K)$ such that $\partial\Sigma$ is an oriented longitude for K. Note that Σ is dual to the canonical epimorphism ϵ.

Suppose that Σ is incompressible. (Recall that a surface Σ in a 3-manifold N is called *incompressible* if the inclusion-induced map $\pi_1(\Sigma) \to \pi_1(N)$ is injective.) We pick a tubular neighborhood $\Sigma \times [-1,1]$. The manifold $X(K) \setminus \Sigma \times (-1,1))$ is the result of *cutting along* Σ. The Seifert–van Kampen theorem gives us a splitting

$$\pi_1(X(K)) = \langle \pi_1(X(K) \setminus \Sigma \times (-1,1)), t \mid \varphi(\pi_1(\Sigma \times -1) = t\pi_1(\Sigma \times 1)t^{-1} \rangle$$

of $(\pi(K), \epsilon_K)$, where φ is induced by the canonical homeomorphism $\Sigma \times -1 \to \Sigma \times 1$. We thus see that $\pi(K)$ splits over the free group $\pi_1(\Sigma)$ of rank $2g$.

Given a knot $K \subset S^3$, we denote by $g = g(K)$ the minimal genus of a Seifert surface. It follows from the Loop Theorem (see, for example, [He76, Chapter 4]) that a Seifert surface of minimal genus is incompressible. Hence $\pi(K)$ splits over a free group of rank $2g(K)$.

If two incompressible Seifert surfaces of a knot K are isotopic, then it is clear that the corresponding splittings of $\pi(K)$ are strongly equivalent. There are many examples of knots that admit non-isotopic minimal genus Seifert surfaces; see e.g. [Ly74b, Ei77a, Ei77b, Al12, HJS13]. We expect that these surfaces give rise to splittings that are not strongly equivalent.

On the other hand, if a knot is fibered, then it admits a unique minimal genus Seifert surface up to isotopy (see e.g. [EL83, Lemma 5.1]). It is therefore perhaps not entirely surprising that $\pi(K)$ admits a unique splitting up to strong equivalence. More precisely, we have the following well-known lemma, which is originally due to Neuwirth [Ne65].

Lemma 3.1. *Let K be a fibered knot of genus g with fiber Σ. Then any splitting of $\pi(K)$ is strongly equivalent to*

$$\langle \pi_1(X(K) \setminus \Sigma \times (-1,1)), t \mid \varphi(\pi_1(\Sigma \times -1) = t\pi_1(\Sigma \times 1)t^{-1} \rangle.$$

In particular $\pi(K)$ only splits over the free group of rank $2g$.

Proof. If Σ is a fiber surface for $X(K)$, then the infinite cyclic cover of $X(K)$ is diffeomorphic to $\Sigma \times \mathbb{R}$. Put differently, $\operatorname{Ker}(\epsilon_K) \cong \pi_1(\Sigma)$ which implies in particular that $\operatorname{Ker}(\epsilon_K)$ is finitely generated. The lemma is now a straightforward consequence of Lemma 2.2. $\square$

4. Splitting of the knot group for $K = 5_2$

In this section we give several explicit splittings of the knot group $\pi(K)$ where $K = 5_2$, the first non-fibered knot in the Alexander-Briggs table. We construct:

1. three splittings of $\pi(5_2)$ over the free group F_2, no two being weakly equivalent;
2. a splitting of $\pi(5_2)$ over the free group F_3 on three generators;
3. a splitting of $\pi(5_2)$ over a non-free group.

Note that neither the second nor the third splitting is induced by an incompressible surface. We will also see that at least two of the three splittings over F_2 are not induced by an incompressible surface.

Since K is a knot of genus one, a minimal-genus Seifert surface gives rise to a splitting of $\pi(K)$ over a free group of rank 2. In the following we will consider an explicit splitting that comes from a Wirtinger presentation of the knot group:

$$\pi(K) = \langle a, b, t \,|\, tat^{-1} = b,\, tb^{-1}ab^{-1}t^{-1} = (b^{-1}a)^2 \rangle.$$

Here the knot group has an HNN decomposition (A, B, t, φ), where A is the free group on a, b while B is the subgroup freely generated by a and $b^{-1}ab^{-1}$. The isomorphism φ sends $a \mapsto b$ and $b^{-1}ab^{-1} \mapsto (b^{-1}a)^2$. For the remainder of this section we identify $\pi(K)$ with $\langle A, t \,|\, \varphi(B) = tBt^{-1} \rangle$.

Proposition 4.1. *Consider the splittings:*

$$\begin{array}{rcl} \pi(K) & = & \langle A, t \,|\, \varphi(B) = tBt^{-1} \rangle, \\ \pi(K) & = & \langle A_{[0,1]}, t \,|\, \varphi_1(A) = tAt^{-1} \rangle, \\ \pi(K) & = & \langle A_{[0,2]}, t \,|\, \varphi_2(A_{[0,1]}) = tA_{[0,1]}t^{-1} \rangle \end{array}$$

where the latter two splittings are provided by Lemma 2.3. Then the following hold.

(i) Each is a splitting over a free group of rank two.
(ii) No two of the splittings of $(\pi(K), \epsilon_K)$ are weakly equivalent.
(iii) At least two of the splittings are not induced by an incompressible Seifert surface.

In the proof of Proposition 4.1 we will make use of the following lemma which is perhaps also of independent interest.

Lemma 4.2. *Let M be a hyperbolic 3-manifold with empty or toroidal boundary, and let G be a subgroup of $\pi := \pi_1(M)$. If $f\colon \pi \to \pi$ is an automorphism with $f(G) \subset G$, then $f(G) = G$.*

We do not know whether the conclusion of the lemma holds for any 3-manifold.

Proof. Let $f\colon \pi \to \pi$ be an automorphism with $f(G) \subset G$. Since M is hyperbolic, it is a consequence of the Mostow Rigidity Theorem that the group of outer automorphisms of π is finite. (See, for example, [BP92, Theorem C.5.6] and [Jo79, p. 213] for details.) Consequently, there exists a positive integer n and an element $x \in \pi$ such that $f^n(G) = xGx^{-1}$. It follows from [Bu07, Theorem 4.1] that $f^n(G) = G$. The assumption that $f(G) \subset G$ implies inductively that $f^n(G) \subset f(G)$. Hence $f(G) = G$. □

We can now turn to the proof of Proposition 4.1.

Proof of Proposition 4.1. It is clear that the first and the second splitting are over a free group of rank two. It remains to show that $A_{[0,1]}$ is a free group of rank two. First note that

$$A_{[0,1]} \cong \langle a_0, b_0, a_1, b_1 \,|\, a_1 = b_0,\, b_1^{-1}a_1b_1^{-1} = (b_0^{-1}a_0)^2 \rangle,$$

where a_i and b_i denote $t^i a t^{-i}$ and $t^i b t^{-i}$, respectively. Using the first relation to eliminate the generator b_0, we obtain $A_{[0,1]} \cong \langle a_0, a_1, b_1 \,|\, r\rangle$, where $r = (a_1^{-1}a_0)^2 b_1 a_1^{-1} b_1$. We let $c = a_1^{-1}a_0$ and $d = b_1 a_1^{-1}$. Clearly $\{c, d, r\}$ is a basis for the free group on a_0, a_1, b_1. Hence $A_{[0,1]} \cong \langle c, d, r \,|\, r\rangle \cong \langle c, d \,|\, \rangle$ is indeed a free group of rank 2. This concludes the proof of (i).

We turn to the proof of (ii). Since K is not fibered it follows from Stallings's theorem (see Theorem 7.1) that $\operatorname{Ker}(\epsilon_K) = \lim_{k\to\infty} A_{[-k,k]}$ is not finitely generated. It follows that easily that for any $l \geq k$ the map $A_{[0,k]} \to A_{[0,l]}$ is a proper inclusion. In particular, we have proper inclusions $A \subsetneq A_{[0,1]} \subsetneq A_{[0,2]}$. Since $S^3 \setminus \nu K$ is hyperbolic, the desired statement now follows from Lemma 4.2.

We prove (iii). It is well known (see, for example, [Ka05]) that any two minimal-genus Seifert surfaces of 5_2 are isotopic. This implies, in particular, that any two splittings of $\pi(K)$ induced by minimal-genus Seifert surfaces are strongly equivalent. It follows from (ii) that at least two of the three splittings are not induced by a minimal genus Seifert surface. $\square$

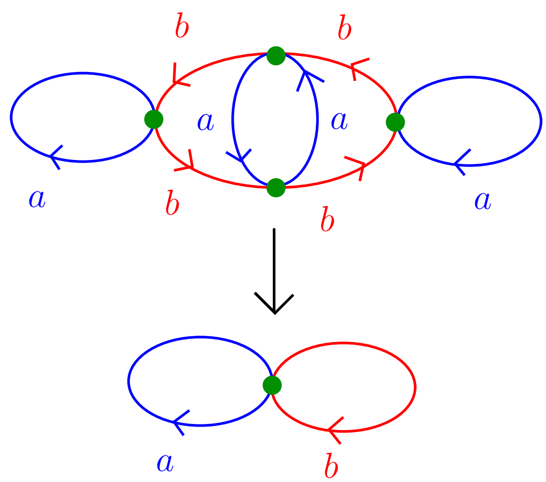

FIGURE 1. Covering graph.

We show that $\pi(K)$ admits a splitting over a free group of rank 3. In order to do so we note that there exists a canonical isomorphism

$$\begin{aligned} & \langle a, b, t \,|\, tat^{-1} = b,\ tb^{-1}ab^{-1}t^{-1} = (b^{-1}a)^2\rangle \\ \cong\ & \langle a, b, c, t \,|\, tat^{-1} = b,\ tb^{-1}ab^{-1}t^{-1} = (b^{-1}a)^2, tb^{-2}ab^{-2}t^{-1} = c\rangle. \end{aligned} \tag{1}$$

Let A' be the free group generated by a, b, c. Let B' be the subgroup of A' generated by $a, b^{-1}ab^{-1}, b^{-2}ab^{-2}$. The fundamental group of the covering graph in Figure 1 is free on $a, b^{-1}ab^{-1}, b^{-1}a^2b, b^{-2}ab^{-2}$, and b^4, and so B' is a free rank-3 subgroup of A'. The elements $b, b^{-1}ab^{-1}a, c$ of A' also generate a free group of rank 3, since they are free in the abelianization of A'. There exists therefore a unique homomorphism $\varphi' : B' \to A'$ such that $\varphi'(a) = b$, $\varphi'(b^{-1}ab^{-1}) = b^{-1}ab^{-1}a$ and $\varphi'(b^{-2}ab^{-2}) = c$. It follows that φ' is in fact a monomorphism. Hence from (1),

$$\langle A', t \,|\, tB't^{-1} = \varphi(B')\rangle$$

defines a splitting of $\pi(K)$ over the free group B' of rank three.

Finally we give an explicit splitting of $\pi(K)$ over a subgroup that is not free. Recall that by Lemma 2.3 the group $\pi(K)$ admits an HNN decomposition with the HNN base $A_{[0,2]}$ defined as the amalgamated product of A, tAt^{-1} and t^2At^{-2}. It suffices to prove the following claim.

Claim. The group $A_{[0,2]}$ is not free.

Note that $A_{[0,2]}$ has the presentation

$$\langle a_0, b_0, a_1, b_1, a_2, b_2 \,|\, a_1 = b_0,\ b_1^{-1}a_1b_1^{-1} = (b_0^{-1}a_0)^2,\ a_2 = b_1,\ b_2^{-1}a_2b_2^{-1} = (b_1^{-1}a_1)^2\rangle.$$

Using the first and third relations, we eliminate the generators b_0 and b_1. Thus

$$A_{[0,2]} \cong \langle a_0, a_1, a_2, b_2 \,|\, r_1,\ r_2\rangle,$$

where $r_1 = (a_1^{-1}a_0)^2a_2a_1^{-1}a_2$ and $r_2 = (a_2^{-1}a_1)^2b_2a_2^{-1}b_2$.

Let $\epsilon = a_1^{-1}a_0$ and $f = a_2a_1^{-1}$. One checks that $\{\epsilon, f, r_1, b_2\}$ is a basis for the free group $\langle a_0, a_1, a_2, b_2 \,|\, \rangle$. Using the substitutions

$$a_0 = f^{-2}\epsilon^{-2}r_1\epsilon,\ a_1 = f^{-2}\epsilon^{-2}r_1 \text{ and } a_2 = f^{-1}\epsilon^{-2}r_1,$$

we see

$$A_{[0,2]} \cong \langle \epsilon, f, b_2 \,|\, r_2\rangle \cong \langle \epsilon, f, b_2 \,|\, f^{-2}\epsilon^{-2}(b_2\epsilon^2)f(b_2\epsilon^2)\rangle.$$

We perform two more changes of variables. First we let $g = b_2\epsilon^2$ and eliminate b_2 to obtain

$$A_{[0,2]} \cong \langle \epsilon, f, g \,|\, \epsilon^{-2}(gf)^2f^{-3}\rangle, .$$

Second, we let $h = gf$ and we eliminate g:

$$A_{[0,2]} \cong \langle \epsilon, f, h \,|\, \epsilon^{-2}h^2 = f^3\rangle.$$

We thus see that $A_{[0,2]}$ is a free product of two free groups amalgamated over an infinite cyclic group. By Lemma 4.1 of [BF94] (see Example 4.2), if the group $A_{[0,2]}$ is free, then either $\epsilon^{-2}h^2$ or f^3 is a basis element in its respective factor. Since neither element is a basis element (seen for example by abelianizing), the group $A_{[0,2]}$ is not free. This concludes the proof of the claim.

5. Splittings of fundamental groups of non-fibered knots over non-free groups

In Section 4 we saw that we can split the knot group $\pi(5_2)$ over a group that is not free. We will now see that this example can be greatly generalized. We recall the statement of our first main theorem.

Theorem 5.1. *If K is a non-fibered knot, then $\pi(K)$ admits splittings over* non-free *subgroups of arbitrarily large rank.*

Proof. Let $\Sigma \subset X(K)$ be a Seifert surface of minimal genus. We write $A = \pi_1(X(K) \setminus \Sigma \times (-1,1)$ and $B = \pi_1(\Sigma \times -1)$, and we consider the corresponding splitting

$$\pi(K) = \langle A, t \,|\, \varphi(B) = tBt^{-1} \rangle$$

of $(\pi(K), \epsilon_K)$ over $\pi_1(\Sigma)$. Given $n \leq m$ we consider, as in Section 2.3, the group

$$A_{[n,m]} = \langle *_{i=m}^{n} t^i A t^{-i} \mid t^j \varphi(B) t^{-1} = t^{j+1} B t^{-j-1} \ (j = n, \dots, m-1) \rangle.$$

By Lemma 2.3 the group $\pi(K)$ splits over the group $A_{[0,n]}$ for any non-negative integer n.

Claim. There exists an integer m such that $A_{[0,n]}$ is not a free group for any $n \geq m$.

As we pointed out in Section 2.3, we have an isomorphism

$$\mathrm{Ker}(\epsilon_K \colon \pi(K) \to \mathbb{Z}) \cong \lim_{k \to \infty} A_{[-k,k]}$$

where the maps $A_{[-l,l]} \to A_{[-k,k]}$ for $l \leq k$ are monomorphisms. It follows from [FF98, Theorem 3] that $\mathrm{Ker}(\epsilon_K)$ is not locally free; that is, there exists a finitely generated subgroup of $\mathrm{Ker}(\epsilon_K)$ which is not a free group. But this implies that there exists $k \in \mathbb{N}$ such that $A_{[-k,k]}$ is not a free group. We have a canonical isomorphism $A_{[-k,k]} \cong A_{[0,2k]}$, and for any $n \geq 2k$ we have a canonical monomorphism $A_{[0,2k]} \to A_{[0,n]}$. It now follows that $A_{[0,n]}$ is not a free group for any $n \geq 2k$. This concludes the proof of the claim.

To complete the proof of Theorem 5.1 it remains to prove the following claim:

Claim. Writing $H_n := A_{[0,n]}$ we have

$$\lim_{n \to \infty} \mathrm{rk}(H_n) = \infty.$$

Since $\Sigma \subset X(K)$ is not a fiber it follows from [He76, Theorem 10.5] that there exists an element $g \in A \setminus B$. By work of Przytycki–Wise (see [PW12b, Theorem 1.1]) the subgroup $B = \pi_1(\Sigma \times -1) \subset \pi(K)$ is separable. This implies, in particular, that there exists an epimorphism $\alpha \colon \pi(K) \to G$ onto a finite group G such that $\alpha(g) \notin \alpha(B)$. Then

$$D := \alpha(B) \subsetneq C := \alpha(A).$$

Given $n \in \mathbb{N}$ we denote by α_n the restriction of α to $H_n \subset \pi(K)$ and we write $G_n := \alpha(H_n)$.

Note that in
$$H_n = A_0 *_{B_0} \cdots *_{B_{n-1}} A_n$$
the groups A_i, viewed as subgroups of $\pi(K)$, are conjugate. It follows that the groups $\alpha_n(A_i)$ are conjugate in G. In particular, each of the groups $\alpha_n(A_i)$ has order $|C|$. The same argument shows that each of the groups $\alpha_n(B_i)$ has order $|D|$. Standard arguments about fundamental groups of graphs of groups (see, for example, [Se80]) imply that $\mathrm{Ker}(\alpha_n : H_n \to G_n)$ is the fundamental group of a graph of groups, where the underlying graph $\tilde{\mathcal{G}}$ is a connected graph with $(n+1)\cdot |G_n|/|C|$ vertices and $n \cdot |G_n|/|D|$ edges. From the Reidemeister-Schreier theorem (see, for example, [MKS76, Theorem 2.8] and from the fact that $\mathrm{Ker}(\alpha_n : H_n \to G_n)$ surjects onto $\pi_1(\tilde{\mathcal{G}})$ it then follows that
$$\begin{array}{rcl} \mathrm{rk}(H_n) & \geq & \frac{1}{|G_n|}\,\mathrm{rk}(\mathrm{Ker}(\alpha_n : H_n \to G_n)) \\ & \geq & \frac{1}{|G_n|}\,\mathrm{rk}(\pi_1(\tilde{\mathcal{G}})) \\ & = & \frac{1}{|G_n|}\big(n\cdot |G_n|/|D| - (n+1)\cdot |G_n|/|C| + 1\big) \\ & \geq & (n+1)\left(\frac{1}{|D|} - \frac{1}{|C|}\right). \end{array}$$
But this sequence diverges to ∞ since $|D| < |C|$. □

6. Splittings of fundamental groups of non-fibered knots over free groups

6.1. **Statement of the theorem.** Lyon [Ly71, Theorem 2] showed that there exists a non-fibered knot K of genus one that admits incompressible Seifert surfaces of arbitrarily large genus (see also [Sce67, Gu81, Ts04] for related examples). By the discussion in Section 3, this implies that $\pi(K)$ splits over free groups of arbitrarily large rank.

Splitting along incompressible Seifert surfaces is a convenient way to produce knot group splittings. Yet there are many non-fibered knots that have unique incompressible Seifert surfaces (see, for example, [Wh73, Ly74a, Ka05]). For such a knot, Seifert surfaces gives rise to only one type of knot group splitting.

In Section 4 we saw an example of a splitting of a knot group over a free group that is not induced by an embedded surface. We generalize the example in our second main theorem. We recall the statement.

Theorem 6.1. *Let K be a non-fibered knot. Then for any integer $k \geq 2g(K)$ there exists a splitting of $\pi(K)$ over a free group of rank k.*

The key to extending the result in Section 4 is the following theorem, which we will prove in the next subsection.

Theorem 6.2. *Let K be a non-fibered knot. Then there exists a Seifert surface Σ of minimal genus such that for a given base point $p \in \Sigma = \Sigma \times 0$ there exists a nontrivial element $g \in \pi_1(S^3 \setminus \Sigma \times (0,1), p)$ such that the subgroup of $\pi(K)$ generated* *by* $\pi_1(\Sigma \times 0, p)$ *and* g *is the free product of* $\pi_1(\Sigma \times 0, p)$ *and the infinite cyclic group* $\langle g \rangle$*.*

Theorem 6.1 is now a consequence of Theorem 6.2 and the following proposition about HNN decompositions.

Proposition 6.3. *Assume that* (π, ϵ) *splits over a free group* F *of rank* n *with base group* A*. If there exists an element* $g \in A$ *such that the subgroup of* π *generated by* F *and* g *is the free product* $F * \langle g \rangle$*, then* (π, ϵ) *splits over free groups of every rank greater than* n*.*

Proof. By hypothesis we can identify π with

$$\langle A, t \mid \varphi(x_i) = t x_i t^{-1} \ (1 \leq i \leq n) \rangle,$$

where $x_1, \ldots, x_n$ generate the group F and where ϵ is given by $\epsilon(t) = 1$ and $\epsilon(A) = 0$.

The kernel of the second-factor projection $F * \langle g \rangle \to \langle g \rangle = \mathbb{Z}$ is an infinite free product $*\{g^i F g^{-i} \mid i \in \mathbb{Z}\}$. Let l be any positive integer. Choose a nontrivial element $z \in F$ and define $z_i = g^i z g^{-i}$, for $1 \leq i \leq l$. Then $F' = \langle F, z_1, \ldots, z_l \rangle$ is a free subgroup of $F * \langle g \rangle$ with rank $n + l$. By hypothesis F' is then also a free subgroup of A of rank $n + l$.

Note that π is canonically isomorphic to

$$\langle A, c_1, \ldots, c_l, t \mid \varphi(x_i) = t x_i t^{-1}, c_j = t z_j t^{-1} (1 \leq i \leq n, 1 \leq j \leq l) \rangle.$$

We denote by A' the free product of A and $\langle c_1, \ldots, c_l \rangle$, and we denote by φ' the unique homomorphism

$$\varphi' \colon F' = F * \langle z_1, \ldots, z_l \rangle \to A' = A * \langle c_1, \ldots, c_l \rangle$$

that extends φ and that maps each z_j to c_j. Since φ' is the free product of two isomorphisms, it is also an isomorphism. We then have a canonical isomorphism

$$\pi \cong \langle A', t \mid \varphi'(F') = t F' t^{-1} \rangle.$$

We have thus shown that (π, ϵ) splits over the free group F' of rank $n + l$. □

6.2. **Proof of Theorem 6.2.** To prove Theorem 6.2 we will need to discuss the JSJ pieces of knot complements. (See [AFW12] for exposition about JSJ decompositions.) It is therefore convenient to generalize a few notions for knots to more general 3-manifolds.

Given a 3-manifold N, we can associate to each class $\epsilon \in H^1(N; \mathbb{Z})$ its Thurston norm $x_N(\epsilon)$, which is defined as the minimal 'complexity' of a surface dual to ϵ. We say that a class $\epsilon \in H^1(N; \mathbb{Z})$ is *fibered* if there exists a fibration $p \colon N \to S^1$ such that the induced map $p_* \colon \pi_1(N) \to \pi_1(S^1) = \mathbb{Z}$ agrees with $\epsilon \in H^1(N; \mathbb{Z}) = \mathrm{Hom}(\pi_1(N), \mathbb{Z})$. It is well known that given a non-zero $d \in \mathbb{Z}$, the class ϵ is fibered if and only if $d\epsilon$ is fibered. Note that given a non-trivial knot $K \subset S^3$ we have $x_{X(K)}(\epsilon_K) = 2g(K) - 1$, and ϵ_K is fibered if and only if K is fibered. We refer to [Th86] for background and more information.

We will need the following theorem, which in particular implies Theorem 6.2 in the case that $S^3 \setminus \nu K$ is hyperbolic.

Theorem 6.4. *Let N be a hyperbolic 3-manifold and let Σ be a properly embedded, connected Thurston norm-minimizing surface that is not a fiber surface. We write $M = N \setminus \Sigma \times (0,1)$ and we pick a base point p on $\Sigma \times 0 = \Sigma$. Then there exists a nontrivial element $g \in \pi_1(M,p)$ such that the subgroup of $\pi_1(M,p)$ generated by $\pi_1(\Sigma,p)$ and g is the free product of $\pi_1(\Sigma,p)$ and $\langle g\rangle$.*

Proof. Let N be a hyperbolic 3-manifold. We denote by $T_1,\dots,T_k$ the boundary components of N. Let Σ be a properly embedded, connected Thurston norm-minimizing surface that is not a fiber surface. We write $M = N \setminus \Sigma \times (0,1)$ and we pick a base point p on $\Sigma \times 0 = \Sigma$. We now take all fundamental groups with respect to this base point. It follows again from the Loop Theorem and the fact that Σ is Thurston norm-minimizing that the inclusion-induced map $\Gamma := \pi_1(\Sigma) \to \pi_1(M)$ is a monomorphism. We will henceforth view $\Gamma = \pi_1(\Sigma)$ as a subgroup of $\pi_1(M)$.

We first suppose that Σ hits all boundary components of N. Since Σ is not a fiber surface, it follows from the Tameness Theorem of Agol [Ag04] and Calegari–Gabai [CG06] that $\pi_1(M)$ is word-hyperbolic and that $\Gamma = \pi_1(\Sigma)$ is a quasi-convex subgroup of $\pi_1(M)$. (We refer to [Wi12a, Sections 14 and 16] for more details.) It then follows from work of Gromov [Gr87, 5.3.C] (see also [Ar01, Theorem 1]) that there exists an element $g \in \pi_1(M)$ such that the subgroup of $\pi_1(M)$ generated by Γ and g is in fact the free product of Γ and $\langle g\rangle$.

We now suppose that there exists a boundary component T_i that is not hit by Σ. We pick a path in M connecting T_i to the chosen base point and we henceforth view $\pi_1(T_i)$ as a subgroup of $\pi_1(M)$. Note that $\pi_1(N)$ is hyperbolic relative to the subgroups $\pi_1(T_1),\dots,\pi_1(T_k)$. Since Σ is not a fiber surface, it follows from the Tameness Theorem and from work of Hruska [Hr10, Corollary 1.3] that Γ is a relatively quasi-convex subgroup of $\pi_1(N)$. Since Γ is a non-abelian surface group we can find an element $g \in \Gamma$ such that $\langle g\rangle \cap \pi_1(T_i)$ is trivial. We see again from the Tameness Theorem that $\langle g\rangle$ is a relatively quasi-convex subgroup of $\pi_1(N)$.

Summarizing, we have shown that $\pi_1(\Sigma)$ and $\langle g\rangle$ are two relatively quasi-convex subgroups of $\pi_1(N)$ which have trivial intersection with the parabolic subgroup $\pi_1(T_i)$. It now follows from Martinez-Pedroza [MP09, Theorem 1.2] that there exists a $h \in \pi_1(T_i)$ such that the subgroup of $\pi_1(N)$ generated by Γ and hgh^{-1} is the free product of Γ and $\langle hgh^{-1}\rangle$. The proposition now follows from the observation that according to our choices, both Γ and $\langle hgh^{-1}\rangle$ lie in $\pi_1(M)$. □

We can now prove Theorem 6.2. For the reader's convenience we recall the statement.

Theorem 6.5. *Let K be a non-fibered knot. Then there exists a Seifert surface Σ of minimal genus and a nontrivial element $g \in \pi_1(S^3 \setminus \Sigma \times (0,1))$ such that the subgroup* *generated by* $\pi_1(\Sigma \times 0)$ *and* g *is the free product of* $\pi_1(\Sigma \times 0)$ *and the infinite cyclic group* $\langle g \rangle$.

Proof. Let K be a non-fibered knot. We write $X = S^3 \setminus \nu K$. We denote by $X_v, v \in V$, the JSJ components, and we denote by $T_\epsilon, \epsilon \in E$, the JSJ tori of X. We let $\epsilon \in H^1(X;\mathbb{Z}) = \operatorname{Hom}(H_1(X);\mathbb{Z}),\mathbb{Z}) \cong \mathbb{Z}$ be the generator that corresponds to the canonical homomorphism $\epsilon_K \colon H_1(X;\mathbb{Z}) \to \mathbb{Z}$. For each $v \in V$, we denote by $\epsilon_v \in H^1(X_v;\mathbb{Z})$ the restriction of ϵ to X_v.

The pair (V, E) has a natural graph structure, since each JSJ torus cobounds two JSJ components. Since X is a knot complement, this graph is a based tree, where the base is the vertex $b \in V$ for which X_b contains the boundary torus. We now denote by T_b the boundary torus of X, and for each $v \neq b$ we denote by T_v the unique JSJ torus which is a boundary component of X_v and which separates X_v from X_b.

Claim. There exists an element $w \in V$ such that X_w is hyperbolic and such that $\epsilon_w \in H^1(X_w;\mathbb{Z})$ is not a fibered class.

We say that a vertex $v \in V$ is non-fibered if $\epsilon_v \in H^1(X_v;\mathbb{Z})$ is not a fibered class. Since $\epsilon = \epsilon_K$ is by assumption not fibered, it follows from [EN85, Theorem 4.2] that some vertex is not fibered. Let $w \in V$ be a non-fibered vertex of minimal distance to b.

Note that if $v \in V$ is fibered and if ϵ_v is non-trivial, then the restriction of ϵ_v to any boundary torus is also non-trivial. Since ϵ_b is non-trivial and since $w \in V$ is a non-fibered vertex of minimal distance to b, we conclude that the restriction of ϵ_w to T_w is non-trivial.

It follows from the Geometrization Theorem and from [JS79, Lemma VI.3.4] that X_w is one of the following:

(1) the exterior of a torus knot;
(2) a 'composing space', that is, a product $S^1 \times W_n$, where W_n is the result of removing n open disjoint disks from D^2;
(3) a 'cable space', that is, a manifold obtained from a solid torus $S^1 \times D^2$ by removing an open regular neighborhood in $S^1 \times \operatorname{Int}(D^2)$ of a simple closed curve c that lies in a torus $S^1 \times s$, where $s \subset \operatorname{Int}(D^2)$ is a simple closed curve and c is non-contractible in $S^1 \times D^2$;
(4) a hyperbolic manifold.

As we argued above, the restriction of $\epsilon_w \in H^1(X_w;\mathbb{Z})$ to one of the boundary tori, namely T_w, is non-trivial. It is well known that in each of the first three cases, this would imply that ϵ_w is a fibered class. Hence X_w must be hyperbolic. This concludes the proof of the claim.

In the following, given a vertex v with ϵ_v non-zero, we denote by $d_v \in \mathbb{N}$ the divisibility of $\epsilon_v \in H^1(X_v;\mathbb{Z})$. For all other vertices we write $d_v = 0$.

Claim. There exists a minimal genus Seifert surface Σ for K with the following properties:

(1) Σ intersects each T_e transversally;
(2) each intersection $\Sigma \cap T_e$ consists of a possibly empty union of parallel, non-null-homologous curves;
(3) for each v with $d_v \neq 0$ the surface $\Sigma_v := \Sigma \cap X_v$ is the union of d_v parallel copies of a surface Σ'_v.

For each v with $d_v \neq 0$ we pick a properly embedded Thurston norm-minimizing surface Σ'_v that represents $\frac{1}{d_v}\epsilon_v$. After possibly gluing in annuli and disks, we may assume that at each boundary torus T of N_v, all the components of $\Sigma'_v \cap T$ are parallel as oriented curves and no component of $\Sigma'_v \cap T$ is null-homologous. We now pick a tubular neighborhood $\Sigma'_v \times [-1, 2]$ of Σ'_v and we denote by Σ_v the union of $\Sigma'_v \times r_i$ where $r_i = \frac{i}{d_v}$ with $i = 0, \dots, d_v - 1$. For each v with $d_v = 0$ we denote by $\Sigma'_v = \Sigma_v$ the empty set.

The surfaces Σ_v are chosen such that at each JSJ torus the boundary curves are parallel. Since at a JSJ edge the adjacent surfaces have to represent the same homology class, at each JSJ torus the adjacent surfaces have exactly the same number of boundary components which furthermore represent the same homology class in the JSJ torus. After an isotopy in the neighborhood of the tori we can therefore glue the surfaces Σ_v together to obtain a properly embedded surface Σ. Since the Thurston norm is linear on rays, it follows from [EN85, Proposition 3.5] that Σ is a connected Thurston norm-minimizing surface representing ϵ. By construction, the intersection of Σ with ∂X consists of one curve, which is necessarily a longitude for K. We thus see that Σ is indeed a genus-minimizing Seifert surface for K. It is now clear that Σ has the desired properties. This concludes the proof of the claim.

Recall that $\epsilon_w \in H^1(X_w; \mathbb{Z})$ is not a fibered class. By the discussion at the beginning of this section, this implies that $\frac{1}{d_w}\epsilon_w$ is also not a fibered class, and so Σ'_w is not a fiber surface.

We pick a base point p_w on $\Sigma'_w = \Sigma'_w \times 0$, which is then also a base point for X_w. It follows from Theorem 6.4 that there exists an element $g \in \pi_1(X_w \setminus \Sigma'_w \times (0, 1), p_w)$ such that the subgroup of $\pi_1(X_w \setminus \Sigma'_w \times (0, 2], p_w)$ generated by $\pi_1(\Sigma'_w, p_w)$ and g is in fact the free product of $\pi_1(\Sigma'_w, p_w)$ and $\langle g \rangle$. It now remains to prove the following claim.

Claim. The subgroup of $\pi_1(X, p_w)$ generated by $\pi_1(\Sigma, p_w)$ and g is the free product of $\pi_1(\Sigma, p_w)$ and $\langle g \rangle$.

We may pick an oriented simple closed curve c in $X_w \setminus \Sigma'_w \times (0, 2]$ that intersects $\Sigma'_w = \Sigma'_w \times 0$ in precisely the base point p_w and that represents $g \in \pi_1(X_w \setminus \Sigma'_w \times (0, 2], p_w)$. Note that $\pi_1(\Sigma \cup c, p_w)$ is precisely the free product of $\pi_1(\Sigma, p_w)$ and $\langle g \rangle$. It thus suffices to show that the inclusion-induced map

$$\pi_1(\Sigma \cup c, p_w) \to \pi_1(X, p_w)$$

is injective.

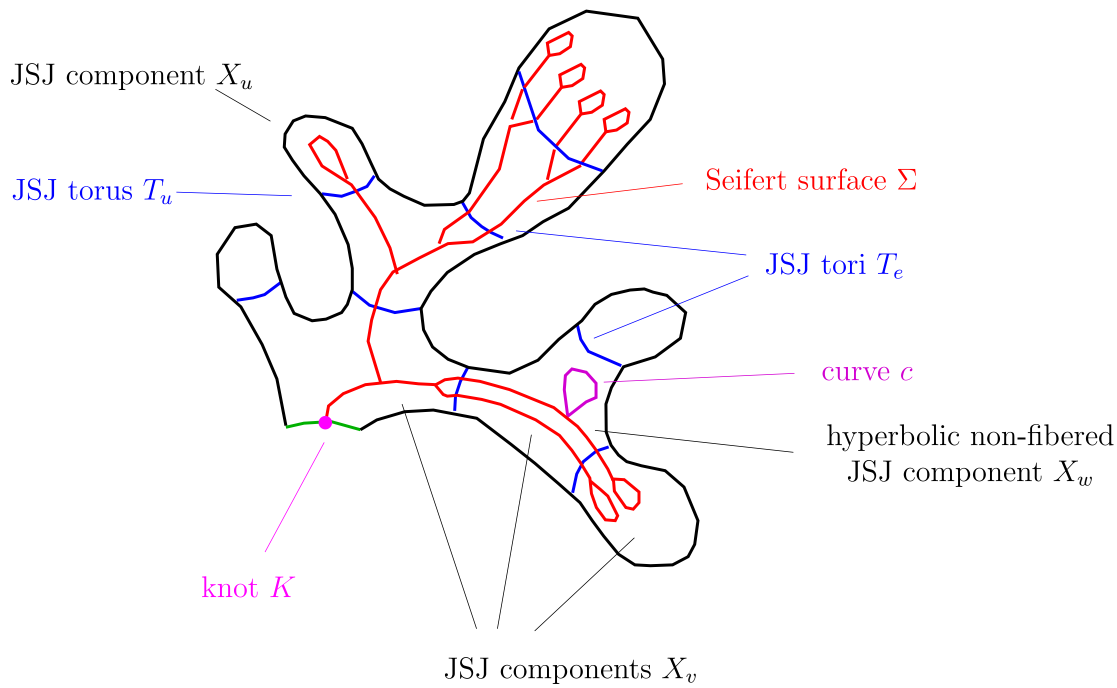

FIGURE 2. Schematic picture for the Seifert surface Σ and the curve c.

Let h be an element in the kernel of this map. We pick a representative curve d which intersects the JSJ tori transversally. We will show that h represents the trivial element in $\pi_1(\Sigma \cup c, p_w)$ by induction on

$$n(d) := \sum_{v \in V} \#\text{components of } d \cap X_v.$$

If $n(d) = 1$, then d lies in the component of $(\Sigma \cup w) \cap X_w = \Sigma_w \cup c$ that contains p_w. Then c lies completely in $\Sigma'_w \cup c$. But the map $\pi_1(\Sigma'_w \cup c, p_w) = \pi_1(\Sigma'_w, p_w) * \langle g \rangle \to \pi_1(X_w)$ is injective, and the map $\pi_1(X_w) \to \pi_1(X)$ is also injective. It thus follows that h is the trivial element.

We now consider the case that $n := n(d) > 1$. We then think of $\pi_1(X)$ as the fundamental group of the graph of groups $\pi_1(X_v)$. We can view the curve d as a concatenation of curves $d_1, \dots, d_n$ such that each curve d_i lies completely in some X_u. Recall that we assume that d represents the trivial element. A standard argument in the theory of fundamental groups of graph of groups (see e.g. [He87]) implies that there exists a d_i with the following two properties:

1. the two endpoints of d_i lie on the same boundary torus T of some X_u,
2. d_i is homotopic in X_u rel endpoints to a curve s_i that lies completely in T.

Note that the two endpoints of d_i lie on $T \cap \Sigma_u$. In fact we can prove a stronger statement.

Claim. The two endpoints of d_i lie on the same component of $T \cap \Sigma_u$.

We first make the following observation. Let S be a properly oriented embedded surface S in an oriented 3-manifold M and let a be an oriented embedded arc that does not intersect S at the endpoints. We can then associate to S and a the algebraic intersection number $S \cdot a \in \mathbb{Z}$, which has in particular the following two properties:

(1) for any properly oriented embedded arc b homotopic to a rel base points we have $S \cdot a = S \cdot b$,
(2) if a lies completely in a boundary component B of M, then $S \cdot a$ equals the algebraic intersection number of the oriented curve ∂S with the oriented arc a in B.

We now turn to the proof of the claim. We first note that there exists a homeomorphism $r\colon X_u \to X_u$ which is the identity on $X_u \setminus \Sigma_u' \times (-1,2)$, which has the property that for any $x \times t$ with $x \in \Sigma_u'$ and $t \in [0,1]$ we have

$$f(x \times t) = x \times (t - \frac{1}{2d_v})$$

and which is isotopic to the identity on X_u. More informally, r is a map that pushes everything on $\Sigma \times [0,1]$ slightly to the left. Note that r pushes everything on Σ_u off Σ_u. Furthermore, if $u = w$, then the intersection of $r(\Sigma_w \cup c)$ with Σ_w is also empty.

Since s_i and d_i are homotopic rel base points and since r is homotopic to the identity, the curves $r(s_i)$ and $r(d_i)$ are homotopic rel base points. It follows from the above that $\Sigma_u \cdot r(s_i) = \Sigma_u \cdot r(d_i)$. But the latter is clearly zero, since $r(d_i)$ does not intersect Σ_u. We now conclude that $\partial\Sigma_u \cdot r(s_i) = \Sigma_u \cdot r(s_i) = 0$. Since the curves $\partial\Sigma_u \cap T$ are all parallel it now follows that $r(s_i)$ does not intersect $\Sigma_u \cap T$ at all. But this means that the two endpoints of s_i, and thus also the two endpoints of d_i, have to lie on the same component of $T \cap \Sigma_u$. This concludes the proof of the claim.

We then make the following claim.

Claim. The curve d_i is homotopic in X_u rel end points to a curve d_i' that lies completely in $T \cap \Sigma_u$.

By the previous claim we know that the two endpoints of d_i lie on the same component of $T \cap \Sigma_u$. We denote the initial point of d_i by P, and the terminal point by Q. We denote by r the component of $\partial\Sigma_u$ that contains P. We endow r with an orientation. Note that r is homologically essential on T. The curve r thus defines a subsummand $\langle r\rangle$ of $\pi_1(T,P) \cong \mathbb{Z}^2$.

We also pick a curve t_i in $T \cap X_u$ from P to Q. The concatenation $s_i t_i^{-1}$ lies in T, and also lies in $(\Sigma \cup c) \cap X_u$. The curve $s_i t_i^{-1}$ thus represents an element in $\pi_1((\Sigma \cup c) \cap X_u, P) \cap \pi_1(T,P)$. But the group $\pi_1((\Sigma \cup c) \cap X_u, P)$ is free (regardless of whether c lies on the P-component of $(\Sigma \cup c) \cap X_u$ or not) whereas $\pi_1(T,P) \cong$

$\mathbb{Z}^2$. The two groups thus intersect in an infinite cyclic subgroup. Furthermore, the intersection contains the subsummand $\langle r\rangle$. It follows that the intersection equals $\langle r\rangle$. In particular, $s_i^{-1}t_i$ is homotopic rel P to r^k for some k. It now follows that relative to the end points we have the following homotopies:

$$d_i \sim d_i s_i^{-1} s_i \sim s_i \sim s_i t_i^{-1} t_i \sim r^k t_i.$$

But the curve $d_i' := r^k t_i$ lies completely in $T \cap \Sigma_u$. This concludes the proof of the claim.

We can thus replace $d = d_1 \dots d_{i-1} d_i d_{i+1} \dots d_l$ by $d_1 \dots d_{i-1} d_i' d_{i+1} \dots d_l$ and push d_i' slightly into the adjacent JSJ component of X. We have found a representative of h of smaller length than d. The claim that h represents the trivial element now follows by induction.

This concludes the proof that the subgroup of $\pi_1(X \setminus \Sigma \times (0,2], p_w)$ generated by $\pi_1(\Sigma, p_w)$ and g is the free product of $\pi_1(\Sigma, p_w)$ and $\langle g\rangle$. We are therefore done with the proof of Theorem 6.5. □

7. Comparison with Stallings's fibering criterion

Let K be a knot. Recall that we denote by $\epsilon_K \colon \pi(K) \to \mathbb{Z}$ the unique epimorphism that sends the oriented meridian to 1. Stallings [St62] proved the following theorem.

Theorem 7.1. *If K is not fibered, then Ker(ϵ_K) is not finitely generated.*

It follows from Lemma 2.2 that if Ker(ϵ_K) is finitely generated, then there exists precisely one group B such that $\pi(K)$ splits over B. Thus Stalling's theorem follows as a consequence of either Theroem 5.1 or Theorem 6.1.

On the other hand, a group π with an epimorphism $\epsilon \colon \pi \to \mathbb{Z}$ such that Ker(ϵ) is not finitely generated may still split over a unique group. The Baumslag-Solitar group, the semidirect product $\mathbb{Z} \ltimes \mathbb{Z}[\frac{1}{2}]$ where $n \in \mathbb{Z}$ acts on $\mathbb{Z}[\frac{1}{2}]$ by multiplication by 2^n, has abelianization $\mathbb{Z}$. The kernel of the abelianization $\epsilon : \pi \to \mathbb{Z}$ is the infinitely generated subgroup $\mathbb{Z}[\frac{1}{2}]$. Since every finitely generated subgroup of $\mathbb{Z}[\frac{1}{2}]$ is isomorphic to $\mathbb{Z}$, $\mathbb{Z} \ltimes \mathbb{Z}[\frac{1}{2}]$ splits only over subgroups isomorphic to $\mathbb{Z}$. (In fact, any two splittings are easily seen to be strongly equivalent.) This shows that the conclusions of Theorems 5.1 and 6.1 are indeed stronger than the conclusion of Theorem 7.1.

Stallings's fibering criterion has been generalized in several other ways. For example, if K is not fibered, then Ker(ϵ) can be written neither as a descending nor as an ascending HNN-extension [BNS87], Ker(ϵ) admits uncountably many subgroups of finite index (see [FV12c, Theorem 5.2], [SW09a] and [SW09b, Theorem 3.4]), the pair $(\pi(K), \epsilon_K)$ has 'positive rank gradient' (see [DFV12, Theorem 1.1]) and Ker(ϵ_K) admits a finite index subgroup which is not normally generated by finitely many elements (see [DFV12, Theorem 5.1]).

8. Proof of Theorem 1.3

In this section we will prove Theorem 1.3, i.e. we will show that if K is a knot, then $\pi(K)$ does not split over a group of rank less than $2g(K)$. We will first give a 'classical' proof for genus-one knots before we provide the proof for all genera.

8.1. **Genus-one knots.** In this subsection we prove:

Theorem 8.1. *If K is a genus-one knot, then $\pi(K)$ does not split over a free group of rank less than two.*

The main ingredients in the proof are two classical results from 3-manifold topology. First, we recall the statement of the Kneser Conjecture, which was first proved by Stallings [St59] in the closed case, and by Heil [Hei72, p. 244] in the bounded case.

Theorem 8.2. (Kneser Conjecture) *Let N be a 3-manifold with incompressible boundary. If there exists an isomorphism $\pi_1(N) \cong \Gamma_1 * \Gamma_2$, then there exist compact, orientable 3-manifolds N_1 and N_2 with $\pi_1(N_i) \cong \Gamma_i$, $i = 1, 2$ and $N \cong N_1 \# N_2$.*

In the following, we say that a properly embedded 2-sided annulus A in a 3-manifold N is *essential* if the inclusion map $A \hookrightarrow N$ induces a π_1-injection and if A is not properly homotopic into ∂N. The second classical result we will use is the following, which is a direct consequence of a theorem of Waldhausen [Wal68b] (see Corollary 1.2(i) of [Sco80]).

Theorem 8.3. *Let N be an irreducible 3-manifold with incompressible boundary. If $\pi_1(N)$ splits over $\mathbb{Z}$, then N contains an essential, properly embedded 2-sided annulus.*

We turn to the proof of Theorem 8.1.

Proof of Theorem 8.1. Let K be a genus-one knot. Since K is non-trivial, the Loop Theorem implies that $\partial X(K)$ is incompressible. Since knot complements are prime 3-manifolds, it now follows from the Kneser Conjecture that $\pi(K)$ can not split over the trivial group, i.e. $\pi(K)$ cannot split over a free group of rank zero.

Now suppose that J is a non-trivial knot such that $\pi(J)$ splits over a free group of rank one, that is, over a group isomorphic to $\mathbb{Z}$. From Theorem 8.3 we deduce that $X(J)$ contains an essential, properly embedded, 2-sided annulus A. Lemma 2 of [Ly74a] (an immediate consequence of [Wal68a]) implies that the knot J is either a composite or a nontrivial cable knot. If J is a composite knot, then it follows from the additivity of the knot genus (see, for example, [Ro90, p. 124]) that the genus of J is at least two. Moreover, a well-known result of Schubert [Sct53] (see Proposition 2.10 of [BZ85]) implies that the genus of any cable knot is greater than one. Thus in both cases we see that $g(J) \geq 2$.

We now see that for the genus-one knot K the group $\pi(K)$ cannot split over a free group of rank one. $\square$

8.2. **Wada's invariant.** For the proof of Theorem 1.3 we will need Wada's invariant, which is also known as the twisted Alexander polynomial or the twisted Reidemeister torsion of a knot.

We introduce the following convention. If π is a group and $\gamma : \pi \to \mathrm{GL}(k, R)$ a representation over a ring, then we denote by γ also the $\mathbb{Z}$-linear extension of γ to a map $\mathbb{Z}[\pi] \to M(k, R)$. Furthermore, if A is a matrix over $\mathbb{Z}[\pi]$ then we denote by $\gamma(A)$ the matrix given by applying γ to each entry of A.

Let π be a group, $\epsilon\colon \pi \to \mathbb{Z}$ an epimorphism, and $\alpha\colon \pi \to \mathrm{GL}(k, \mathbb{C})$ a representation. First note that α and ϵ give rise to a tensor representation

$$\begin{array}{rcl} \alpha \otimes \epsilon\colon \pi & \to & \mathrm{GL}(k, \mathbb{C}[t^{\pm 1}]) \\ g & \mapsto & t^{\epsilon(g)} \cdot \alpha(g). \end{array}$$

Now let

$$\pi = \langle g_1, \dots, g_k \,|\, r_1, \dots, r_l \rangle$$

be a presentation of π. By adding trivial relations if necessary, we may assume that $l \geq k - 1$. We denote by F_k the free group with generators $g_1, \dots, g_k$. Given $j \in \{1, \dots, k\}$ we denote by $\frac{\partial}{\partial g_j}\colon \mathbb{Z}[F_k] \to \mathbb{Z}[F_k]$ the Fox derivative with respect to g_j, i.e. the unique $\mathbb{Z}$-linear map such that

$$\begin{array}{rcl} \dfrac{\partial g_i}{\partial g_j} & = & \delta_{ij}, \\ \dfrac{\partial uv}{\partial g_j} & = & \dfrac{\partial u}{\partial g_j} + u \dfrac{\partial v}{\partial g_j} \end{array}$$

for all $i, j \in \{1, \dots, k\}$ and $u, v \in F_k$. We denote by

$$M := \left(\frac{\partial r_i}{\partial g_j} \right)$$

the $l \times k$-matrix over $\mathbb{Z}[\pi]$ of all the Fox derivatives of the relators. Given subsets $I = \{i_1, \dots, i_r\} \subset \{1, \dots, k\}$ and $J = \{j_1, \dots, j_s\} \subset \{1, \dots, l\}$ we denote by $M_{J,I}$ the matrix formed by deleting the columns $i_1, \dots, i_r$ and by deleting the rows $j_1, \dots, j_s$ of M.

Note that there exists at least one $i \in \{1, \dots, k\}$ such that $\epsilon(g_i) \neq 0$. It follows that

$$\det((\alpha \otimes \epsilon)(1 - g_i)) = \det\left(\mathrm{id}_k - t^{\epsilon(g_i)} \alpha(g_i)\right) \neq 0.$$

We define

$$Q_i := \gcd\{\det((\alpha \otimes \epsilon)(M_{J,\{i\}})) \,|\, J \subset \{1, \dots, l\} \text{ with } |J| = l + 1 - k\}.$$

(Note that each $M_{J,\{i\}}$ is a $(k-1) \times (k-1)$-matrix.) It is worth considering the special case that $l = k - 1$; that is, the case of a presentation of deficiency one. Then the only choice for J is the empty set, and hence

$$Q_i = \det((\alpha \otimes \epsilon)(M_{\emptyset,\{i\}})).$$

Wada [Wad94] introduced the following invariant of the triple (π, ϵ, α).

$$\Delta^{\alpha}_{\pi,\epsilon} := Q_i \cdot \det((\alpha \otimes \epsilon)(1 - g_i))^{-1} \in \mathbb{C}(t).$$

A priori, Wada's invariant depends on the various choices we made. The following theorem proved by Wada [Wad94, Theorem 1] shows that the indeterminacy is well controlled.

Theorem 8.4. *Let π be a group, let $\epsilon\colon \pi \to \mathbb{Z}$ be an epimorphism, and let $\alpha\colon \pi \to GL(k, \mathbb{C})$ be a representation. Then $\Delta^{\alpha}_{\pi,\epsilon}$ is well-defined up to multiplication by a factor of the form $\pm t^k r$, where $k \in \mathbb{Z}$ and $r \in \mathbb{C}^*$.*

Finally, let $K \subset S^3$ be a knot and let $\alpha\colon \pi(K) \to \mathrm{GL}(k, \mathbb{C})$ be a representation. As before, we denote by $\epsilon\colon \pi(K) \to \mathbb{Z}$ the epimorphism that sends the oriented meridian of K to 1. We write

$$\Delta^{\alpha}_{K} = \Delta^{\alpha}_{\pi,\epsilon}.$$

If $\alpha\colon \pi(K) \to \mathrm{GL}(1, \mathbb{C})$ is the trivial one-dimensional representation, then Wada's invariant is determined by the classical Alexander polynomial Δ_K. More precisely, we have

$$\Delta^{\alpha}_{K} = \frac{\Delta_K}{1 - t}.$$

Wada's invariant equals the twisted Reidemeister torsion of a knot, and is closely related to the twisted Alexander polynomial of a knot, which was first introduced by Lin [Lin01]. We refer to [Ki96, FV10] for more details about Wada's invariant, its interpretation as twisted Reidemeister torsion and its relationship to twisted Alexander polynomials.

8.3. **Proof of Theorem 1.3.** Before we provide the proof of Theorem 1.3 we need to introduce two more definitions. First, given a non-zero polynomial $p(t) = \sum_{i=r}^{s} a_i t^i \in \mathbb{C}[t^{\pm 1}]$ with $a_r \neq 0$ and $a_s \neq 0$, we write

$$\deg(p(t)) = s - r.$$

If $f(t) = p(t)/q(t) \in \mathbb{C}(t)$ is a non-zero rational function, we write

$$\deg(f(t)) = \deg(p(t)) - \deg(q(t)).$$

Note that if Wada's invariant of a triple (π, ϵ, α) is non-zero, then the degree of Wada's invariant $\Delta^{\alpha}_{\pi,\epsilon}$ is well defined.

We can now formulate the following theorem.

Theorem 8.5. *Let π be a group and let*

$$f : \pi \to \langle A, t \,|\, f(B) = tBt^{-1} \rangle$$

be a splitting. We denote by $\epsilon\colon \langle A, t \,|\, f(B) = tBt^{-1}\rangle \to \mathbb{Z}$ *the canonical epimorphism which is given by* $\epsilon(t) = 1$ *and* $\epsilon(a) = 0$ *for* $a \in A$*. If* $\alpha\colon \pi \to GL(k,\mathbb{C})$ *is a representation such that* $\Delta^\alpha_{\pi,\epsilon} \neq 0$*, then*

$$\deg \Delta^\alpha_{\pi,\epsilon} \leq k(\mathrm{rk}(B) - 1).$$

In [FKm06] (see also [Fr12]) it was shown that if K is a knot and $\alpha\colon \pi(K) \to \mathrm{GL}(k,\mathbb{C})$ is a representation such that $\Delta^\alpha_K \neq 0$, then

$$\deg \Delta^\alpha_K \leq k(2\,\mathrm{genus}(K) - 1). \tag{2}$$

In light of the discussion in Section 3, we can view Theorem 8.5 as a generalization of (2).

Proof. Let π be a group and let

$$\pi = \langle g_1, \dots, g_k, t \,|\, r_1, \dots, r_l, \varphi(b) = tbt^{-1} \text{ for all } b \in B\rangle$$

be a splitting, where $\varphi\colon B \to A$ is a monomorphism and B is a rank-d subgroup of $A = \langle g_1, \dots, g_k, t \,|\, r_1, \dots, r_l\rangle$. We pick generators $x_1, \dots, x_d$ for B. Note that

$$\begin{array}{rl} & \langle g_1, \dots, g_k, t \,|\, r_1, \dots, r_l, \varphi(b) = tbt^{-1} \text{ for all } b \in B\rangle \\ = & \langle g_1, \dots, g_k, t \,|\, r_1, \dots, r_l, \varphi(x_1)^{-1}tx_1t^{-1}, \dots, \varphi(x_d)^{-1}tx_dt^{-1}\rangle. \end{array}$$

We write $K := \mathrm{Ker}(\epsilon)$.

We denote by M the $(l+d) \times (k+1)$-matrix over $\mathbb{Z}[\pi]$ that is given by all the Fox derivatives of the relators. We make the following observations.

(1) The relators $r_1, \dots, r_l$ are words in $g_1, \dots, g_k$. The Fox derivatives of the r_i with respect to the g_j thus lie in $\mathbb{Z}[K]$.
(2) For any $i \in \{1, \dots, k\}$ and $j \in \{1, \dots, r\}$ we have
$$\frac{\partial}{\partial g_i}\left(\varphi(x_j)^{-1}tx_jt^{-1}\right) = \frac{\partial}{\partial g_i}\left(\varphi(x_j)^{-1}\right) + \varphi(x_j)^{-1}t\frac{\partial}{\partial g_i}x_j.$$
The same argument as in (1) shows that the first term lies in $\mathbb{Z}[K]$, and one can similarly see that the second term is of the form $t \cdot g$, where $g \in \mathbb{Z}[K]$.

Thus $M_{\emptyset,\{k+1\}}$, the matrix obtained from M by deleting the $(k+1)$-st column, is of the form

$$M_{\emptyset,\{k+1\}} = P + tQ,$$

where P and Q are matrices over $\mathbb{Z}[K]$, and where all but the last d rows of Q are zero.

Let $\alpha\colon \pi \to \mathrm{GL}(k,\mathbb{C})$ be a representation and $J \subset \{1, \dots, d+l\}$ a subset with $|J| = d + l - k$. It follows from the above that

$$M_{J,\{k+1\}} = P_J + tQ_J,$$

where P_J and Q_J are matrices over $\mathbb{Z}[K]$ and where at most d rows of Q_J are non-zero. We then see that

$$\det((\alpha \otimes \epsilon)(M_{J,\{k+1\}})) = \det(\alpha(P_J) + t\alpha(Q_J)),$$

where at most kr rows of $\alpha(Q_J)$ are non-zero. If $\det(\alpha(P_J) + t\alpha(Q_J))$ is non-zero, then it follows from an elementary argument that

$$\deg(\det(\alpha(P_J) + t\alpha(Q_J))) \leq kr.$$

We now consider

$$Q := \gcd\{\det((\alpha \otimes \epsilon)(M_{J,\{k+1\}})) \,|\, J \subset \{1, \dots, l\} \text{ with } |J| = d + l - k\}.$$

By the above, if $Q \neq 0$, then $\deg(Q) \leq kr$.

Since $\epsilon(t) = 1$,

$$\Delta^{\alpha}_{\pi,\epsilon} = Q \cdot \det((\alpha \otimes \epsilon)(1 - t))^{-1} = Q \cdot \det(\mathrm{id}_k - \alpha(t)t)^{-1} \in \mathbb{C}(t).$$

Finally, we suppose that $\Delta^{\alpha}_{\pi,\epsilon} \neq 0$. By the above, this implies that $Q \neq 0$. In particular, we see that

$$\begin{array}{rcl} \deg(\Delta^{\alpha}_{\pi,\epsilon}) & = & \deg\left(Q \cdot \det(\mathrm{id}_k - \alpha(t)t))\right) \\ & = & \deg(Q) - \deg(\det(\mathrm{id}_k - \alpha(t)t)) \\ & = & \deg(Q) - k \\ & \leq & kr - k = k(\mathrm{rk}\, B - 1). \end{array}$$

This concludes the proof of the theorem. □

The last ingredient in the proof of Theorem 1.3 is the following result from [FV12a]. The proof of the theorem builds on the virtual fibering theorem of Agol [Ag08] (see also [FKt12]), which applies for knot complements by the work of Liu [Liu11], Przytycki-Wise [PW11, PW12a] and Wise [Wi09, Wi12a, Wi12b].

Theorem 8.6. *Let K be a knot. Then there exists a representation $\alpha\colon \pi(K) \to GL(k, \mathbb{C})$ such that $\Delta^{\alpha}_K \neq 0$ and such that*

$$\deg \Delta^{\alpha}_K = k(2g(K) - 1).$$

In [FV12a, Theorem 1.2] an analogous statement is formulated for twisted Reidemeister torsion instead of Wada's invariant. The theorem, as stated, now follows from the interpretation (see, for example, [Ki96, FV10]) of Wada's invariant as twisted Reidemeister torsion.

We can now formulate and prove the following result, which is equivalent to Theorem 1.3.

Theorem 8.7. *Let K be a knot. If $\pi(K)$ splits over a group B, then* $\mathrm{rk}(B) \geq 2g(K)$.

Proof. Let K be a knot and let

$$f\colon \pi(K) \to \pi = \langle A, t \,|\, \varphi(B) = tBt^{-1} \rangle$$

be an isomorphism. We denote by $\epsilon\colon \langle A, t \,|\, \varphi(B) = tBt^{-1} \rangle \to \mathbb{Z}$ the canonical epimorphism which is given by $\epsilon(t) = 1$ and $\epsilon(a) = 0$ for $a \in A$.

Note that $\epsilon \circ f\colon \pi(K) \to \mathbb{Z}$ is an epimorphism. In particular, it sends the meridian to either 1 or -1. By possibly changing the orientation of the knot, we can assume that $\epsilon \circ f \colon \pi(K) \to \mathbb{Z}$ sends the meridian to 1. By Theorem 8.6, there exists a representation $\alpha \colon \pi(K) \to \mathrm{GL}(k, \mathbb{C})$ such that $\Delta_K^\alpha \neq 0$ and such that

$$\deg \Delta_K^\alpha = k(2g(K) - 1).$$

By definition, we have

$$\Delta_K^\alpha = \Delta_{\pi(K),\epsilon\circ f}^\alpha = \Delta_{\pi,\epsilon}^\alpha.$$

Theorem 8.5 implies that

$$\mathrm{rk}(B) \geq \frac{1}{k} \deg \left(\Delta_{\pi,\epsilon}^\alpha\right) + 1 = \frac{1}{k} \deg \left(\Delta_K^\alpha\right) + 1 = 2g(K).$$

□

References

[Ag04] I. Agol, *Tameness of hyperbolic 3-manifolds*, unpublished paper (2007), math/0405568.

[Ag08] I. Agol, *Criteria for virtual fibering*, Journal of Topology 1 (2008), 269-284.

[Ag12] I. Agol, *The virtual Haken conjecture*, with an appendix by I. Agol, D. Groves and J. Manning, preprint (2012).

[AFW12] M. Aschenbrenner, S. Friedl and H. Wilton, *3-Manifold groups*, preprint, 2012, arXiv:1206.0202

[Al12] I. Altman, *Sutured Floer homology distinguishes between Seifert surfaces*, Topology Appl. 159 (2012), no. 14, 3143–3155.

[Ar01] G. Arzhantseva, *On quasiconvex subgroups of word hyperbolic groups*, Geom. Dedicata 87 (2001), no. 1-3, 191–208.

[BP92] R. Benedetti and C. Petronio, *Lectures on hyperbolic geometry*, Universitext, Springer-Verlag, Berlin, 1992.

[BF94] M. Bestvina and M. Feighn, *Outer limits*, preprint (1994)

[BNS87] R. Bieri, W. Neumann and R. Strebel, *A geometric invariant of discrete groups*, Invent. Math. 90 (1987), 451–477.

[BS78] R. Bieri and R. Strebel, *Almost finitely presented soluble groups*, Comment. Math. Helv. 53 (1978), 258–278.

[BZ84] M. Boileau and H. Zieschang, *Heegaard genus of closed orientable Seifert 3-manifolds*, Invent. Math. 76 (1984), 455–468.

[BZ85] G. Burde and H. Zieschang, *Knots*, Walter de Gruyter, Berlin, 1985.

[Bu07] J. O. Button, *Mapping tori with first Betti number at least two*, J. Math. Soc. Japan 59 (2007), no. 2, 351–370

[Ca09] D. Calegari, *scl*, MSJ Memoirs, 20. Mathematical Society of Japan, Tokyo, 2009.

[CG06] D. Calegari and D. Gabai, *Shrinkwrapping and the taming of hyperbolic 3-manifolds*, J. Amer. Math. Soc. 19 (2006), no. 2, 385–446.

[CZ93] D. J. Collins and H. Zieschang, *Combinatorial group theory and fundamental groups*, Algebra, VII, 1–166, 233-240, Encyclopaedia Math. Sci., 58, Springer, Berlin, 1993.

[CS83] M. Culler and P. Shalen, *Varieties of group representations and splittings of 3-manifolds*, Ann. Math. 117 (1983), no. 1, 109–146.

[DFV12] J. DeBlois, S. Friedl and S. Vidussi, *Rank gradients of infinite cyclic covers of 3-manifolds*, Preprint (2012)

[EL83] A. Edmonds and C. Livingston, *Group actions on fibered three-manifolds*, Comment. Math. Helv. 58 (1983), no. 4, 529–542.

[EN85] D. Eisenbud and W. Neumann, *Three-dimensional link theory and invariants of plane curve singularities*, Annals of Mathematics Studies, 110. Princeton University Press, Princeton, NJ, 1985.

[Ei77a] J. Eisner, *A characterisation of non-fibered knots*, Mich. Math. J. 24 (1977), 41–44.

[Ei77b] J. Eisner, *Knots with infinitely many minimal spanning surfaces*, Trans. Amer. Math. Soc. 229 (1977), 329–349.

[Fa98] B. Farb, *Relatively hyperbolic groups*, Geom. Funct. Anal. 8, No. 5 (1998), 810–840.

[FG73] C. D. Feustel and R. J. Gregorac, *On realizing HNN groups in 3-manifolds*, Pacific J. Math. 46 (1973), 381–387.

[FF98] B. Freedman and M. Freedman, *Kneser-Haken finiteness for bounded* 3*-manifolds locally free groups, and cyclic covers*, Topology 37 (1998), no. 1, 133–147.

[Fr12] S. Friedl, *Twisted Reidemeister torsion, the Thurston norm and fibered manifolds*, Preprint (2012). To appear in Geometriae Dedicata.

[FKm06] S. Friedl and T. Kim, *The Thurston norm, fibered manifolds and twisted Alexander polynomials*, Topology 45 (2006), 929–953.

[FKt12] S. Friedl and T. Kitayama, *The virtual fibering theorem for* 3*-manifolds*, Preprint (2012)

[FV10] S. Friedl and S. Vidussi, *A survey of twisted Alexander polynomials*, The Mathematics of Knots: Theory and Application (Contributions in Mathematical and Computational Sciences), editors: Markus Banagl and Denis Vogel (2010), p. 45-94.

[FV12a] S. Friedl and S. Vidussi, *The Thurston norm and twisted Alexander polynomials*, Preprint (2012)

[FV12b] S. Friedl and S. Vidussi, *Minimal Genus on 4-manifolds with a Free Circle Action*, Preprint (2012)

[FV12c] S. Friedl and S. Vidussi, *A Vanishing Theorem for Twisted Alexander Polynomials with Applications to Symplectic 4-manifolds*, Journal of the Eur. Math. Soc., to appear.

[Ga83] D. Gabai, *Foliations and the topology of* 3*-manifolds*, J. Differ. Geom. 18 (1983), no. 3, 445–503.

[Ga87] D. Gabai, *Foliations and the topology of 3-manifolds. III.*, J. Differ. Geom. 26 (1987), 479–536.

[Gr87] M. Gromov, *Hyperbolic groups*, in S. M. Gersten (ed.) Essays in Group Theory, Math. Sci. Res. Inst. Publ. 8, Springer, New York (1987), 75–263.

[Gu81] R. Gustafson, *A simple genus-one knot with incompressible spanning surfaces of arbitrarily high genus*, Pacific J. Math. 96 (1981), no. 1, 81–98.

[HJS13] M. Hedden, A. Juhász and S. Sarkar, *On Sutured Floer homology and the equivalence of Seifert surfaces*, Alg. Geom. Topology 13 (2013), 505–548.

[Hei72] W. Heil, *On Kneser's conjecture for bounded* 3*-manifolds*, Proc. Cambridge Philos. Soc. 71 (1972), 243–246.

[He76] J. Hempel, 3*-Manifolds*, Ann. of Math. Studies, no. 86. Princeton University Press, Princeton, N. J., 1976.

[He87] J. Hempel, *Residual finiteness for* 3*-manifolds*, Combinatorial group theory and topology (Alta, Utah, 1984), pp. 379–396, Ann. of Math. Stud., 111, Princeton Univ. Press, Princeton, NJ, 1987.

[Hr10] C. Hruska, *Relative hyperbolicity and relative quasi-convexity for countable groups*, Alg. Geom. Topology 10 (2010), 1807–1856.

[JS79] W. Jaco and P. Shalen, *Seifert fibered spaces in* 3*-manifolds*, Mem. Amer. Math. Soc. 21 (1979), no. 220.

[Jo79] K. Johannson, *On the mapping class group of simple 3-manifolds*, Topology of Low-Dimensional Manifolds (Proc. Second Sussex Conf., Chelwood Gate, 1977), Lecture Notes in Math., vol. 722, Springer, Berlin, 1979, 48–66.

[Ka05] O. Kakimizu, *Classification of the incompressible spanning surfaces for prime knots of 10 or less crossings*, Hiroshima Math. J. 35 (2005), no. 1, 47–92.

[Ki96] T. Kitano, *Twisted Alexander polynomials and Reidemeister torsion*, Pacific J. Math. 174 (1996), no. 2, 431–442.

[Kr99] P. Kronheimer, *Minimal genus in* $S^1 \times M^3$, Invent. Math. 135, no. 1: 45–61 (1999).

[Li13] T. Li, *Rank and genus of* 3*-manifolds*, J. Amer. Math. Soc. 26 (2013), 777–829.

[Lin01] X. S. Lin, *Representations of knot groups and twisted Alexander polynomials*, Acta Math. Sin. (Engl. Ser.) 17 (2001), 361–380.

[Liu11] Y. Liu, *Virtual cubulation of nonpositively curved graph manifolds*, J. of Topology, to appear (2011).

[Ly71] H.C. Lyon, *Incompressible surfaces in knot spaces*, Trans. Am. Math. Soc. 157 (1971), 53–62.

[Ly74a] H.C. Lyon, *Simple knots with unique spanning surfaces*, Topology 13 (1974), 275–279.

[Ly74b] H. C. Lyon, *Simple knots without unique minimal surfaces*, Proc. Amer. Math. Soc. 43 (1974), 449–454.

[MKS76] W. Magnus, A. Karrass and D. Solitar, *Combinatorial Group Theory: Presentations of Groups in Terms of Generators and Relations*, Second revised edition. Dover Publications, Inc., New York (1976)

[MP09] E. Martinez-Pedroza, *Combination of quasiconvex subgroups of relatively hyperbolic groups*, Groups Geom. Dyn. 3 (2009), no. 2, 317–342.

[Ne65] L. Neuwirth, *Knot groups*, Annals of Mathematics Studies, No. 56 Princeton University Press, Princeton, N.J. 1965.

[PW11] P. Przytycki and D. Wise, *Graph manifolds with boundary are virtually special*, J. of Topology, to appear (2011).

[PW12a] P. Przytycki and D. Wise, *Mixed* 3*-manifolds are virtually special*, preprint (2012).

[PW12b] P. Przytycki and D. Wise, *Separability of embedded surfaces in* 3*-manifolds*, preprint (2012)

[Ro90] D. Rolfsen, *Knots and links*, Corrected reprint of the 1976 original. Mathematics Lecture Series, 7. Publish or Perish, Inc., Houston, TX, 1990.

[Sce67] C. B. Schaufele, *The commutator group of a doubled knot*, Duke Math. J. 34 (1967), 677–681.

[Sct53] H. Schubert, *Knoten und Vollringe*, Acta Math. 90 (1953), 131-286.

[ScW07] J. Schultens and R. Weidmann, *On the geometric and the algebraic rank of graph manifolds*, Pacific J. Math. 231 (2007), 481–510.

[Sco80] P. Scott, *A new proof of the annulus and torus theorems*, Amer. J. Math. 102 (1980), 241-277.

[ST80] H. Seifert and W. Threlfall, *A textbook of topology*, Academic Press, 1980.

[Se80] J.-P. Serre, *Trees*, Springer-Verlag, Berlin-New York, 1980.

[Si96] D. Silver, *HNN bases and high-dimensional knots*, Proc. Amer. Math. Soc. 124 (1996), no. 4, 1247–1252.

[Str84] R. Strebel, *Finitely presented soluble groups*, in Group Theory: Essays for Philip Hall, Academic Press, London, 1984.

[SW09a] D. Silver and S. Williams, *Nonfibered knots and representation shifts*, Proceedings of the Postnikov Memorial Conference, Banach Center Publications, 85 (2009), 101–107.

[SW09b] D. Silver and S. Williams, *Twisted Alexander Polynomials and Representation Shifts*, Bull. London Math. Soc., 41 (2009), 535–540.

[St59] J. Stallings, *J. Grushko's theorem II, Kneser's conjecture*, Notices Amer. Math. Soc. 6 (1959), No 559-165, 531–532.

[St62] J. Stallings, *On fibering certain 3–manifolds*, 1962 Topology of 3–manifolds and related topics (Proc. The Univ. of Georgia Institute, 1961), 95–100 Prentice-Hall, Englewood Cliffs, N.J. (1962)

[Th86] W. P. Thurston, *A norm for the homology of* 3*-manifolds*, Mem. Amer. Math. Soc. 339 (1986), 99–130.

[Ts04] Y. Tsutsumi, *Hyperbolic knots spanning accidental Seifert surfaces of arbitrarily high genus*, Math. Z. 246 (2004), no. 1-2, 167–175.

[Wad94] M. Wada, *Twisted Alexander polynomial for finitely presentable groups*, Topology 33 (1994), 241–256.

[Wal68a] F. Waldhausen, *The word problem in fundamental groups of sufficiently large irreducible 3-manifolds*, Ann. Math. (2) 88 (1968), 272-280.

[Wal68b] F. Waldhausen, *On irreducible 3-manifolds which are sufficiently large*, Ann. Math. (2) 87 (1968), 56-88.

[Wh73] W. Whitten, *Isotopy types of knot spanning surfaces*, Topology 12 (1973), 373–380.

[Wi09] D. Wise, *The structure of groups with a quasi-convex hierarchy*, Electronic Res. Ann. Math. Sci 16 (2009), 44–55.

[Wi12a] D. Wise, *The structure of groups with a quasi-convex hierarchy*, 189 pages, preprint (2012), downloaded on October 29, 2012 from
`http://www.math.mcgill.ca/wise/papers.html`

[Wi12b] D. Wise, *From riches to RAAGs:* 3*-manifolds, right–angled Artin groups, and cubical geometry*, CBMS Regional Conference Series in Mathematics, 2012.

Mathematisches Institut, Universität zu Köln, Germany
E-mail address: `sfriedl@gmail.com`

Department of Mathematics and Statistics, University of South Alabama
E-mail address: `silver@southalabama.edu`

Department of Mathematics and Statistics, University of South Alabama
E-mail address: `swilliam@southalabama.edu`

SALEM NUMBERS: A SURVEY

CHRIS SMYTH

Abstract. I survey results about, and recent applications of, Salem numbers.

1. Introduction

In this article I state and prove some basic results about Salem numbers, and then survey some of the literature about them. My intention is to complement other general treatises on these numbers, rather than to repeat their coverage. This applies particularly to the work of Bertin and her coauthors [6, 8, 9] and to the application-rich Salem number survey of Ghate and Hironaka [27]. I have, however, quoted some results from Salem's classical monograph [62].

Recall that a number is an *algebraic integer* if it is the zero of a polynomial with integer coefficients and leading coefficient 1. Then its (Galois) *conjugates* are the zeros of its *minimal polynomial*, which is the lowest degree polynomial of that type that it satisfies. This degree is the *degree* of the algebraic integer.

A *Salem number* is an algebraic integer $\tau > 1$ of degree at least 4, conjugate to τ^{-1}, all of whose conjugates, excluding τ and τ^{-1}, lie on $|z| = 1$. Then $\tau + \tau^{-1}$ is a real algebraic integer > 2, all of whose conjugates $\neq \tau + \tau^{-1}$lie in the real interval $(-2, 2)$. Such numbers are easy to find: an example is $\tau+\tau^{-1} = 1+\sqrt{2}$, giving $(\tau+\tau^{-1}-1)^2 = 2$, so that $\tau^4-2\tau^3+\tau^2-2\tau+1 = 0$ and $\tau = 1.8832\ldots$. We note that this polynomial is a so-called (self)-reciprocal polynomial: it satisfies the equation $z^{\deg P}P(z^{-1}) = P(z)$. This simply means that its coefficients form a palindromic sequence: they read the same backwards as forwards. This holds for the minimal polynomial of all Salem numbers. It is simply a consequence of τ and τ^{-1} having the same minimal polynomial.

Salem numbers are usually defined in an apparently more general way, as in the following proposition. The proposition shows that this apparent greater generality is illusory.

2010 *Mathematics Subject Classification.* Primary 11R06.
Key words and phrases. Salem number.

Lemma 1 (Salem [63, p.26])**.** *Suppose that $\tau > 1$ is an algebraic integer, all of whose conjugates $\neq \tau$ lie in the closed unit disc $|z| \leq 1$, with at least one on its boundary $|z| = 1$. Then τ is a Salem number (as defined above).*

Proof. Taking τ' to be a conjugate of τ on $|z| = 1$, we have that $\bar{\tau}' = \tau'^{-1}$ is also a conjugate τ'' say, so that $\tau'^{-1} = \tau''$. For any other conjugate τ_1 of τ we can apply a Galois automorphism mapping $\tau'' \mapsto \tau_1$ to deduce that $\tau_1 = \tau_2^{-1}$ for some conjugate τ_2 of τ. Hence the conjugates of τ occur in pairs τ', τ'^{-1}. Since τ itself is the only conjugate in $|z| > 1$, it follows that τ^{-1} is the only conjugate in $|z| < 1$, and so all conjugates of τ apart from τ and τ^{-1} in fact lie on $|z| = 1$. □

It is known that an algebraic integer lying with all its conjugates on the unit circle must be a root of unity (Kronecker [Kr]). So in some sense Salem numbers are the algebraic integers that are 'the nearest things to roots of unity'. And, like roots of unity, the set of all Salem numbers is closed under taking powers.

Lemma 2. *If τ is a Salem number of degree d, then so is τ^n for all $n \in \mathbb{N}$.*

Proof. If τ is conjugate to τ' then τ^n is conjugate to τ'^n. So τ^n will be a Salem number of degree d unless some of its conjugates coincide: say $\tau_1^n = \tau_2^n$ with $\tau_1 \neq \tau_2$. but then, by applying a Galois automorphism mapping $\tau_1 \mapsto \tau$, we would have $\tau^n = \tau_3^n$ say, where $\tau_3 \neq \tau$ is a conjugate of τ. But then $|\tau^n| > 1$ while $|\tau_3^n| \leq 1$, a contradiction. □

Which number fields contain Salem numbers? Of course one can simply choose a list of Salem numbers $\tau, \tau', \tau'', \ldots$ say, and then the number field $\mathbb{Q}(\tau, \tau', \tau'', \ldots)$ certainly contains $\tau, \tau', \tau'', \ldots$. However, if one is interested only in Salem numbers whose degree is that of the field, we can be much more specific.

Proposition 3 (Salem [61, p.169])**.** *Suppose that K is a number field with $[K : \mathbb{Q}] = d$. Then K contains a Salem number τ of degree d (equivalently, $K = \mathbb{Q}(\tau)$ for some Salem number τ) if and only if K has a totally real subfield K' of index 2, and $K = K'(\tau)$ with $\tau + \tau^{-1} = \alpha$, where $\alpha > 2$ is an algebraic integer in K', all of whose conjugates $\neq \alpha$ lie in $(-2, 2)$.*

If $K = \mathbb{Q}(\tau)$ for some Salem number τ of degree d, then there is a Salem number $\tau_1 \in K$ such that the set of Salem numbers of degree d in K consists of the powers of τ_1.

Proof. If K contains a Salem number τ of degree d, then clearly $K = \mathbb{Q}(\tau)$, and so the subfield $K' = \mathbb{Q}(\alpha)$ is totally real, where $\alpha = \tau + \tau^{-1} > 2$, with all its conjugates $\neq \alpha$ lying in $(-2,2)$. Since $\tau^2 - \alpha\tau + 1 = 0$, we have $[K : K'] = 2$.

Conversely, suppose that K has a totally real subfield K' of index 2, and $K = K'(\alpha)$, where $\alpha > 2$ is an algebraic integer in K', all of whose conjugates $\neq \alpha$ lie in $(-2,2)$. Then, defining τ by $\tau^2 - \alpha\tau + 1 = 0$ we have $K = \mathbb{Q}(\tau)$, where τ is a Salem number.

For the last part, consider the set of all Salem numbers of degree d in $K = \mathbb{Q}(\tau)$. Now the number of Salem numbers $< \tau$ in K is clearly finite, as there are only finitely many possibilities for the minimal polynomials of such numbers. Hence there is a smallest such number, τ_1 say. For any Salem number, τ' say, in K we can choose a positive integer r such that $\tau_1^r \leq \tau' < \tau_1^{r+1}$. But if $\tau_1^r < \tau'$ then $\tau'\tau_1^{-r}$ would be another Salem number in K which, moreover, would be less than τ_1, a contradiction. Hence $\tau' = \tau_1^r$. □

Here we have used the following lemma.

Lemma 4. *If $\tau' > \tau$ are both Salem numbers of degree $d = [K : \mathbb{Q}]$ in a number field K, then $\tau'\tau^{-1}$ is also a Salem number of degree d in K.*

Proof. Since τ has degree d, we have $K = \mathbb{Q}(\tau)$. Hence τ' is a polynomial in τ. Therefore any Galois automorphism taking $\tau \mapsto \tau^{-1}$ will map τ' to a real conjugate of τ', namely $\tau'^{\pm 1}$. But it cannot map τ' to itself for then, as τ is also a polynomial in τ', τ would be mapped to itself, a contradiction. So τ' is mapped to τ'^{-1} by this automorphism. Hence $\tau'\tau^{-1}$ is conjugate to its reciprocal. So the conjugates of $\tau'\tau^{-1}$ occur in pairs x, x^{-1}. Again, because τ' is a polynomial in τ, any automorphism fixing τ will also fix τ', and so fix $\tau'\tau^{-1}$. Likewise, any automorphism fixing τ' will also fix τ.

Next consider any conjugate of $\tau'\tau^{-1}$ in $|z| > 1$. It will be of the form $\tau_1'\tau_1^{-1}$, where τ_1' is a conjugate of τ' and τ_1 is a conjugate of τ. For this to lie in $|z| > 1$, we must either have $|\tau_1'| > 1$ or $|\tau_1| < 1$, i.e., $\tau_1' = \tau'$ or $\tau_1 = \tau^{-1}$. But in the first case, as we have seen, $\tau_1 = \tau$, so that $\tau_1'\tau_1^{-1} = \tau'\tau^{-1}$, while in the second case $\tau_1' = \tau'^{-1}$, giving $\tau_1'\tau_1^{-1} = \tau\tau'^{-1} \in |z| < 1$. Hence $\tau'\tau^{-1}$ itself is the only conjugate of $\tau'\tau^{-1}$ in $|z| > 1$. It follows that all conjugates of $\tau'\tau^{-1}$ apart from $(\tau'\tau^{-1})^{\pm 1}$ must lie on $|z| = 1$, making $\tau'\tau^{-1}$ a Salem number.

To show that $\tau'\tau^{-1}$ has degree d, consider d automorphisms that map τ to each of its d conjugates. Then, as we have seen, only the automorphism that maps τ to itself maps $\tau'\tau^{-1}$ to itself. However, if $\tau'\tau^{-1}$ has degree d/k then there are k such automorphisms mapping $\tau'\tau^{-1}$ to itself. Hence $k = 1$

and $\tau'\tau^{-1}$ has degree d. As τ is a unit, $\tau'\tau^{-1}$ is an algebraic integer, and so is a Salem number. $\square$

We now show that the powers of Salem numbers have an unusual property.

Proposition 5 (Salem [62]). *For every Salem number τ and every $\varepsilon > 0$ there is a real number $\lambda > 0$ such that the distance $\|\lambda\tau^n\|$ of $\lambda\tau^n$ to the nearest integer is less than ε for all $n \in \mathbb{N}$.*

Proof. We consider the standard embedding of the algebraic integers $\mathbb{Z}(\tau)$ as a lattice in $\mathbb{R}^d$ defined for $k = 0, 1, \ldots, d-1$ by the map

$$\tau^k \mapsto (\tau^k, \tau^{-k}, \operatorname{Re}\tau_2^k, \operatorname{Im}\tau_2^k, \operatorname{Re}\tau_3^k, \operatorname{Im}\tau_3^k, \ldots, \operatorname{Re}\tau_{d/2}^k, \operatorname{Im}\tau_{d/2}^k),$$

where $\tau^{\pm 1}, \tau_j^{\pm 1}\,(j = 2, \ldots, d/2)$ are the conjugates of τ. As this is a lattice of full dimension d, we know that for every $\varepsilon' > 0$ there are lattice points in the 'slice' $\{(x_1, \ldots, x_n) \in \mathbb{R}^n : |x_i| < \varepsilon'\,(i = 2, \ldots, d)\}$. Such a lattice point corresponds to an element $\lambda(\tau)$ of $\mathbb{Z}(\tau)$ with conjugates λ_i satisfying $|\lambda_i| < \sqrt{2}\varepsilon'\,(i = 2, \ldots, d)$.

Next, consider the sums

$$\sigma_n = \lambda(\tau)\tau^n + \lambda(\tau^{-1})\tau^{-n} + \lambda(\tau_2)\tau_2^n + \lambda(\tau_2^{-1})\tau_2^{-n} + \ldots \lambda(\tau_{d/2})\tau_{d/2}^n + \lambda(\tau_{d/2}^{-1})\tau_{d/2}^{-n},$$

where $\lambda(x) \in \mathbb{Z}[x]$. Since σ_n is a symmetric function of the conjugates of τ, it is rational. As it is an algebraic integer, it is in fact a rational integer. Since all terms $\lambda(\tau^{-1})\tau^{-n}$, $\lambda(\tau_2)\tau_2^n$, $\lambda(\tau_2^{-1})\tau_2^{-n}$, $\ldots$, $\lambda(\tau_{d/2})\tau_{d/2}^n$, $\lambda(\tau_{d/2}^{-1})\tau_{d/2}^{-n}$ are $< \sqrt{2}\varepsilon'$ in modulus, we see that

$$|\sigma_n - \lambda(\tau)\tau^n| < (d-1)\sqrt{2}\varepsilon'.$$

Hence, choosing $\varepsilon' = \varepsilon/((d-1)\sqrt{2})$, we have $\|\lambda\tau^n\| \le |\sigma_n - \lambda(\tau)\tau^n| < \varepsilon$. $\square$

In fact, this property essentially characterises Salem (and Pisot) numbers among all real numbers. Pisot [56] proved that if λ and τ are real numbers such that

$$\|\lambda\tau^n\| \le \frac{1}{2e\tau(\tau+1)(1+\log\lambda)}$$

for all integers $n \ge 0$ then $\tau \in S \cup T$ and $\lambda \in \mathbb{Q}(\tau)$. The denominator in this result was later improved by Cantor [18] to $2e\tau(\tau+1)(2+\sqrt{\log\lambda})$, and then by Decomps-Guilloux and Grandet-Hugot [20] to $e(\tau+1)^2(2+\sqrt{\log\lambda})$.

For further results concerning the distribution of the fractional parts of $\lambda\tau^n$ for τ a Salem number, see Dubickas [23] and Zaïmi [75, 76, 77].

2. A smallest Salem number?

Define the polynomial $L(z)$ by

$$L(z) = z^{10} + z^9 - z^7 - z^6 - z^5 - z^4 - z^3 + z + 1.$$

Then $L(z)$ is the minimal polynomial of a Salem number $\tau_{10} = 1.176\ldots$. This was discovered by D. H. Lehmer [33] in 1933. Curiously, the polynomial $L(-z)$ had appeared a year earlier in Reidemeister's book [57] as the Alexander polynomial of the $(-2, 3, 7)$ Pretzel knot. Lehmer's paper seems to be the first where what is now called the *Mahler measure* of a polynomial appears: the Mahler measure $M(P)$ of a monic one-variable polynomial P is the product $\prod_i \max(1, |\alpha_i|)$ over the roots α_i of the polynomial.

Lehmer also asked whether the Mahler measure of any nonzero noncyclotomic irreducible polynomial with integer coefficients is bounded below by some constant > 1. This is now commonly referred to as 'Lehmer's conjecture' — see [71]. If this were true, then there would be a smallest Salem number. The 'strong version' of 'Lehmer's conjecture' states that τ_{10} is that number. A consequence of this strong version, applied to the minimal polynomial of a Salem number, is the following.

Conjecture 6. *Suppose that $n \in \mathbb{N}$ and $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n$ are real numbers with $\alpha_0 \in (2, \tau_{10} + \tau_{10}^{-1})$ and $\alpha_1, \ldots, \alpha_n \in (-2, 2)$. Then $\prod_{i=0}^{n}(x + \alpha_i) \notin \mathbb{Z}[x]$.*

(Note that $\tau_{10} + \tau_{10}^{-1} = 2.026\ldots$.) For if there were a Salem number $\tau < \tau_{10}$ then $\alpha_0 = \tau + \tau^{-1}$ would lie in $(2, \tau_{10} + \tau_{10}^{-1})$ and its conjugates α_i for $i > 0$ would lie in $(-2, 2)$, with $\prod_{i=0}^{n}(x + \alpha_i) \in \mathbb{Z}[x]$. Thus the strong version of Lehmer's Conjecture would then be false.

3. Construction of Salem numbers

3.1. **Salem's method.** Salem [62, Theorem IV, p.30] found a simple way to construct infinite sequences of Salem numbers from Pisot numbers. Recall that a Pisot number is an algebraic integer greater than 1 all of whose conjugates, excluding itself, all lie in the open unit disc $|z| < 1$. Now if $P(z)$ is the minimal polynomial of a Pisot number, then, except possibly for some small values of n, the polynomials $S_{n,P,\pm1}(z) = z^n P(z) \pm z^{\deg P} P(z^{-1})$ factor as the minimal polynomial of a Salem number, possibly multiplied by some cyclotomic polynomials. In particular, for $P(z) = z^3 - z - 1$, the minimal polynomial of the smallest Pisot number, $S_{8,P,-1} = (z-1)L(z)$. Salem's construction shows that every Pisot number is the limit on both sides of a sequence of Salem numbers. (The construction has to be modified slightly when P is reciprocal.)

Boyd [11] proved that all Salem numbers could be produced by Salem's construction, in fact with $n = 1$. It turns out that many different Pisot numbers can be used to produce the same Salem number. These Pisot numbers can be much larger than the Salem number they produce. In particular, on taking $P(z) = z^3 - z - 1$ and $\varepsilon = -1$, the minimal polynomial of the smallest Pisot number $\theta_0 = 1.3247\ldots$, Salem's method shows that there are infinitely many Salem numbers less than θ_0. This fact motivates the next definition, due to Boyd.

Salem numbers less than 1.3 are called *small*. A table of 39 such numbers was compiled by Boyd [11], with later additions of four each by Boyd [13] and Mossinghoff [50], making 47 in all. See the table [51]. (The starred entries in this table are the four Salem numbers found by Mossinghoff. They include one of degree 46.) Further, it was determined by Flammang, Grandcolas and Rhin [26] that the table was complete up to degree 40. This was extended up to degree 44 by Mossinghoff, Rhin and Wu [52] as part of a larger project to find small Mahler measures.

In [14] Boyd showed how to find, for a given $n \geq 2$, $\varepsilon = \pm 1$ and real interval $[a, b]$, all Salem numbers in that interval that are roots of $S_{n,P,\varepsilon}(z) = 0$ for some Pisot number having minimal polynomial $P(z)$. In particular, of the four new small Salem numbers that he found, two were discovered by this method. The other two he found in [14] are not of this form: they are roots only of some $S_{1,P,\varepsilon}(z) = 0$.

Boyd and Bertin [7] investigated the properties of the polynomials $S_{1,P,\pm 1}(z)$ in detail. For a related, but interestingly different, way of constructing Salem numbers, see Boyd and Parry [17].

Let T denote the set of all Salem numbers (Salem's notation). (It couldn't be called S, because that is used for the set of all Pisot numbers. The notation S here is in honour of Salem, however: Salem [59] had proved the magnificent result that the Pisot numbers form a closed subset of the real line. And so, I suppose, is T!) Salem's construction shows that the derived set (set of limit points) of T contains S. Salem [62, p.31] wrote 'We do not know whether numbers of T have limit points other than S'. Boyd [11, p. 327] conjectured that there were no other such limit points, i.e., that the derived set of $S \cup T$ is S. (He had recently conjectured [12] that $S \cup T$ is closed – a conjecture that left open the possibility that some numbers in T could be limit points of T.)

3.2. **Salem numbers and matrices.** One strategy that has been used to try to prove Lehmer's Conjecture is to attach some combinatorial object (knot, graph, matrix,...) to an algebraic number (for example, to a Salem number). But it is not clear whether the object could throw light on the (e.g.) Salem number, or, on the contrary, that the Salem number could throw light on the object.

Typically, however, such attachment constructions seem to work only for a restricted class of algebraic numbers, and not in full generality. For example, McKee and Smyth [38] consider integer symmetric matrices as the objects for attachment. (These can be considered as generalisations of graphs: one can identify a graph with its adjacency matrix – an integer symmetric matrix having all entries 0 or 1, with only zeros on the diagonal.) The main tool for our work was the following classical result, which deserves to be better known.

Theorem 7 (Cauchy's Interlacing Theorem)**.** *Let M be a real $n \times n$ symmetric matrix, and M' be the matrix obtained from M by removing the ith row and column. Then the eigenvalues $\lambda_1, \dots, \lambda_n$ of M and the eigenvalues $\mu_1, \dots, \mu_{n-1}$ of M' interlace, i.e.,*

$$\lambda_1 \le \mu_1 \le \lambda_2 \le \mu_2 \le \cdots \le \mu_{n-1} \le \lambda_n.$$

We say that an $n \times n$ integer symmetric matrix M is *cyclotomic* if all its eigenvalues lie in the interval $[-2, 2]$. It is so-called because then its associated reciprocal polynomial

$$P_M(z) = z^n \det\left((z + z^{-1})I - M\right)$$

has all its roots on $|z| = 1$ and so (Kronecker again) is a product of cyclotomic polynomials. Here I is the $n \times n$ identity matrix.

The cyclotomic graphs are very familiar.

Theorem 8 (J.H. Smith 1969 [68])**.** *The connected cyclotomic graphs consist of the (not necessarily proper) induced subgraphs of the Coxeter graphs $\tilde{A}_n(n \ge 2)$, $\tilde{D}_n(n \ge 4)$, $\tilde{E}_6$, $\tilde{E}_7$, $\tilde{E}_8$, as in Figure 1.*

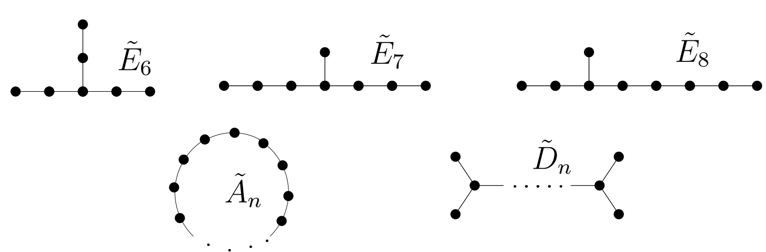

FIGURE 1. The Coxeter graphs $\tilde{E}_6, \tilde{E}_7, \tilde{E}_8, \tilde{A}_n(n \geqslant 2)$ and $\tilde{D}_n(n \geqslant 4)$. (The number of vertices is 1 more than the index.)

(These graphs also occur in the theory of Lie algebras, reflection groups, Lie groups, Tits geometries, surface singularities, subgroups of $\mathrm{SU}_2(\mathbb{C})$ (McKay correspondence),. . . .)

McKee and Smyth describe all the cyclotomic matrices, of which the cyclotomic graphs form a small subset. They prove that the strong version of Lehmer's conjecture is true for the set of polynomials P_M: namely, if M is not a cyclotomic matrix, then P_M has Mahler measure at least $\tau_{10} = 1.176\ldots$, the smallest known Salem number. In fact they show that the smallest three known Salem numbers are all Mahler measures of P_M for some integer symmetric matrix M, while the fourth smallest known Salem number is not.

For other construction methods for Salem numbers see Lakatos [33, 34, 35] and also [42, 38, 39, 40, 41, 69, 70]. In particular, in [33, 35] Lakatos shows that Salem numbers arise as the spectral radius of Coxeter transformations of certain oriented graphs containing no oriented cycles.

3.3. **Traces of Salem numbers.** McMullen [44, p.230] asked whether there are any Salem numbers of trace less than -1. McKee and Smyth [38, 39] found examples of Salem numbers of trace -2, and indeed showed that there are Salem numbers of every trace. It is known [39] that a Salem number of degree $d \geq 10$ has trace at least $\lfloor 1-d/9 \rfloor$. In particular, for $d = 22$ the trace is at least -2. (For this case this result was obtained earlier by McMullen [29, Cor.1.8], but with the extra restriction that the minimal polynomial $S(x)$ of the Salem number had $S(-1) = \pm 1$ and $S(1) = \pm 1$.)

3.4. **Distribution modulo 1 of the powers of a Salem number.** Let $\tau > 1$ be a Salem number. Salem [62, Theorem V, p.33] proved that although the powers $\tau^n \pmod 1$ of τ are everywhere dense on $(0, 1)$, they are not uniformly distributed on this interval. See also [21].

3.5. **Sumsets of Salem numbers.** Dubickas [22] shows that a sum of $m \geq 2$ Salem numbers cannot be a Salem number, but that for every $m \geq 2$ there are m Salem numbers whose sum is a Pisot number and also m Pisot numbers whose sum is a Salem number.

3.6. **Galois group of Salem number fields.** Lalande [36] and Christopoulos and McKee [19] studied the Galois group of a number field defined by a Salem number. Let τ be a Salem number of degree $2n$, $K = \mathbb{Q}(\tau)$ and L be its Galois closure. Then it is known that $G := \mathrm{Gal}(L/\mathbb{Q}) \leq C_2^n \rtimes S_n$. Conversely, if K is a real number field of degree $2n > 2$ with exactly 2 real embeddings, and, for its Galois closure L, that $G \leq C_2^n \rtimes S_n$, then Lalande proved that K is generated by a Salem number.

Now, for a Salem number τ, let $K' = \mathbb{Q}(\tau+\tau^{-1})$, L' be its Galois closure and $N \subset G$ be the fixing group of L'. Then Christopoulos and McKee showed that G is isomorphic to $N \rtimes \mathrm{Gal}(L'/\mathbb{Q})$, where N is isomorphic to either C_2^n or C_2^{n-1}. The latter case is possible only when n is odd.

Amoroso [1] found a lower bound, conditional on the Generalised Riemann Hypothesis, for the exponent of the class group of such number fields L.

3.7. **Other Salem number studies.** Salem [60], [63, p. 35] proved that every Salem number is the quotient of two Pisot numbers.

P. Borwein and Hare [10] studied the 'spectrum' of values $a_0 + a_1\tau + \cdots + a_n\tau^n$ when the $a_i \in \{-1, 0, 1\}$, $n \in \mathbb{N}$ and τ is a Salem number.

For connections between small Salem numbers and exceptional units, see Silverman [67].

Dubickas and Smyth [24] studied the lines passing through two conjugates of a Salem number.

For generalisations of Salem numbers, see Bertin [4, 5], Kerada [31], Meyer [49], Samet [63], Schreiber [65] and Smyth [69]. Note the correction made to [63] in [69].

4. Salem numbers outside Number Theory

The survey of Ghate and Hironaka [27] contains many applications of Salem numbers, for the period up to 1999. Only a few of the applications they describe are briefly recalled here, in subsections 4.1, 4.2 and 4.3. Otherwise, I concentrate on developments since their paper appeared.

For some of these applications, the restriction that Salem numbers should have degree at least 4 can be dropped: the results also hold for reciprocal Pisot numbers, whose minimal polynomials are $x^2 - ax + 1$ for $a \in \mathbb{N}$, $n \geq 3$. Some authors include these numbers in the definition of Salem numbers. Accordingly, I will allow these numbers to be Salem numbers in this section.

4.1. **Growth of groups.** For a group G with finite generating set $S = S^{-1}$, we define its *growth series* $F_{G,S}(x) = \sum_{n=0}^{\infty} a_n x^n$, where a_n is the number of elements of G that can be represented as the product of n elements of S, but not by fewer. For certain such groups, $F_{G,S}(x)$ is known to be a rational function. Then expanding $F_{G,S}(x)$ out in partial fractions leads to a closed formula for the a_n. See [27, Section 4] for a detailed description, including references. See also [2].

In particular, let G be a Coxeter group generated by reflections in $d \geq 3$ geodesics in the upper half plane, forming a polygon with angles π/p_i ($i = 1, 2, \ldots, d$), where $\sum_i \pi/p_i < \pi$. Taking S to be the set of these reflections, it is known (Cannon and Wagreich, Floyd and Plotnick, Parry) that then the denominator of $F_{G,S}(x)$ – call it $\Delta_{p_1,p_2,\ldots,p_d}(x)$ – is the minimal polynomial of a Salem number, τ say, possibly multiplied by some cyclotomic polynomials. Then the a_n grow exponentially with growth rate $\lim_{n\to\infty} a_{n+1}/a_n = \tau$. Hironaka [30] proved that among all such $\Delta_{p_1,p_2,\ldots,p_d}(x)$, the lowest growth rate was achieved for $\Delta_{p_1,p_2,p_3}(x)$, which is Lehmer's polynomial $L(x)$, with growth rate $\tau_{10} = 1.176\ldots$.

4.2. **Alexander Polynomials.** A result of Seifert tells us that a polynomial $P \in \mathbb{Z}[x]$ is the Alexander polynomial of some knot iff it is monic and reciprocal, and $P(1) = \pm 1$. In particular, Hironaka [30] showed that $\Delta_{p_1,p_2,\ldots,p_d}(-x)$ is the Alexander polynomial of the $(p_1, p_2, \ldots, p_d, -1)$ pretzel knot. Hence, from the result of the previous section, we see that Alexander polynomials are sometimes Salem polynomials (albeit in $-x$).

Indeed, Silver and Williams [66], in their study of Mahler measures of Alexander polynomials, found families of links whose Alexander polynomials had Mahler measure equal to a Salem number. The first family $l(q)$ was obtained [66, Example 5.1] from the link 7_1^2 by giving q full right-handed twists to one of the components as it passed through the other component (the trivial knot). The Mahler measure of the Alexander polynomials of these links produced a decreasing sequence of Salem numbers for $q = 1, 2, \ldots, 11$. For $q = 10$ the Salem number 1.18836... (the second-smallest known) was produced, with minimal polynomial

$$x^{18} - x^{17} + x^{16} - x^{15} - x^{12} + x^{11} - x^{10} + x^9 - x^8 + x^7 - x^6 - x^3 + x^2 - x + 1,$$

while $q = 11$ gave the Salem number $M(L(x)) = 1.17628\ldots$. For $q > 11$ Salem numbers were not produced. The second example was obtained in a similar way [66, Example 5.8], using the link formed from the knot 5_1 by an adding the trivial knot encircling two strands of the knot, and then giving these strands q full right-hand twists. For increasing $q \geq 3$ this gave a monotonically increasing sequence of Salem numbers tending to the smallest Pisot number $\theta_0 = 1.3247\ldots$. These Salem numbers are equal to $M(x^{2(q+1)}(x^3-x-1)+x^3+x^2-1)$. Furthermore, $M(x^n(x^3-x-1)+x^3+x^2-1)$ is also a Salem number for $n \geq 9$ and odd. Silver (private communication) has shown that these Salem numbers are also Mahler measures of Alexander polynomials: " Putting an odd number of half-twists in the rightmost arm of the pretzel knot produces 2-component links rather than knots. Their Alexander polynomials have two variables. However, setting the two variables equal to each other produces the so-called 1-variable Alexander polynomials, and indeed the 'odd' sequence of Salem polynomials ... results."

4.3. **Lengths of closed geodesics.** It is known that there is a bijection between the set of Salem numbers and the set of closed geodesics on certain arithmetic hyperbolic surfaces. Specifically, the length of the geodesic is $2\log\tau$, where τ is the Salem number corresponding to the geodesic. Thus there is a smallest Salem number iff there is a geodesic of minimal length among all closed geodesics on all arithmetic hyperbolic surfaces. See Ghate and Hironaka [27, Section 3.4] and also Maclachlan and Reid [43, Section 12.3] for details.

4.4. **Arithmetic Fuchsian groups.** Neumann and Reid [53, Lemmas 4.9, 4.10] have shown that Salem numbers are precisely the spectral radii of hyperbolic elements of arithmetic Fuchsian groups. See also [27], [43, pp. 378–380] and [37, Theorem 9.7].

The following result is related.

Theorem 9 (Sury [73]). *The set of Salem numbers is bounded away from 1 iff there is some neighbourhood U of the identity in $\mathrm{SL}_2(\mathbb{R})$ such that, for each arithmetic cocompact Fuchsian group Γ, the set $\Gamma\cap U$ consists only of elements of finite order.(A Fuchsian group is a group Γ discrete in $\mathrm{SL}_2(\mathbb{R})$ and such that $\Gamma\backslash H$ has finite volume.)*

4.5. **A dynamical system.** For given $\beta > 1$, define the map $T_\beta : [0,1) \to [0,1)$ by $T_\beta x = x - \lfloor x\rfloor$. Klaus Schmidt [64] conjectured that for β a Salem number, the orbit of $\beta - \lfloor\beta\rfloor$ is eventually periodic. This conjecture was proved by Boyd [15] to hold for Salem numbers of degree 4. However, using a heuristic model in [16], his results indicated that while Schmidt's conjecture was likely to also hold for Salem numbers of degree 6, it may be false for a positive proportion of Salem numbers of degree 8. As Boyd points out, the basic reason seems to be that, for β a Salem number of degree d, this orbit corresponds to a pseudorandom walk on a d-dimensional lattice. Under this model, but assuming true randomness, the probability of the walk intersecting itself is 1 for $d \le 6$, but is less than 1 for $d > 6$.

4.6. **Surface automorphisms.** A *K3 surface* is a simply-connected compact complex surface X with trivial canonical bundle. The intersection form (a nonvanishing holomorphic 2-form) makes $H^2(X,\mathbb{Z})$ into an even unimodular 22-dimensional lattice of signature $(3,19)$; see [48, p.17]. Now let $F : X \to X$ be an automorphism of positive entropy of a K3 surface X. Then the spectral radius $\lambda(F)$ (modulus of the largest eigenvalue) of F is a Salem number. More specifically, the characteristic polynomial $\chi(F)$ of F acting by pullback on this lattice is the minimal polynomial of a Salem number multiplied by $k \geq 0$ cyclotomic polynomials. Since $\chi(F)$ has degree 22, the degree of $\lambda(F)$ is at most 22. (If F is projective, $\chi(F)$ and so $\lambda(F)$ has degree at most 20.)

It is an interesting problem to descibe which Salem numbers arise in this way. Gross and McMullen [29] have shown that if the minimal polynomial $S(x)$ of a Salem number of degree 22 has $|S(-1)| = |S(1)| = 1$ (which they call the *unramified* case) then it is the characteristic polynomial an automorphism of some (non-projective) K3 surface X. (If the entropy of F is 0 then this characteristic polynomial is simply a product of cyclotomic polynomials.) It is known (see [44, p.211] and references given there) that the topological entropy $h(F)$ of F is equal to $\log \lambda(F)$, so is either 0 or the logarithm of a Salem number.

For each even $d \geq 2$ let τ_d be the smallest Salem number of degree d. McMullen [46, Theorem 1.2] has proved that if $F : X \to X$ is an automorphism of *any* compact complex surface X with positive entropy, then $h(F) \geq \log \tau_{10} = \log(1.176\ldots) = 0.162\ldots$. Bedford and Kim [3] have shown that this lower bound is realised by a particular rational surface automorphism. McMullen [47] showed that it was realised for a non-projective K3 surface automorphism, and later [48] that it was realised for a projective K3 surface automorphism. He showed that the value $\log \tau_d$ was realised for a projective K3 surface automorphism for $d = 2, 4, 6, 8, 10$ or 18, but not for $d = 14$, 16, or 20. (The case $d = 12$ is currently undecided.) In fact, if $h(F) = \log \tau$ for a K3 surface automorphism and Salem number τ, then [58, p. 475] τ has degree at most 22, and degree at most 20 if the surface is projective – see [29, Remark p. 268].

McMullen [44] found 10 Salem numbers of degree 22 and trace -1, also having some other properties, from which he was able to construct from each of these Salem numbers a K3 surface automorphism having a Siegel disc. (These were the first known examples having Siegel discs).

Oguiso [54] showed that, as for K3 surfaces (see above), the characteristic polynomial of an automorphism of a hyper-Kähler manifold is also the minimal polynomial of a Salem number multiplied by $k \geq 0$ cyclotomic polynomials. In another paper [55] he constructed an automorphism F of a (nonprojective) K3 surface with $\lambda(F) = \tau_{14}$. Here the K3 surface contained an E_8 configuration of rational curves, and the automorphism also had a Siegel disc.

Reschke [58] studied the automorphisms of two-dimensional complex tori. He showed that the entropy of such an automorphism, if positive, must be a Salem number of degree at most 6, and gave necessary and sufficient conditions for such a Salem number to arise in this way.

4.7. **Salem numbers and Coxeter systems.** Consider a Coxeter system (W, S), consisting of a multiplicative group W generated by a finite set $S = \{s_1, \ldots, s_n\}$, with relations $(s_i s_j)^{m_{ij}} = 1$ for each i, j, where $m_{ii} = 1$ and $m_{ij} \geq 2$ for $i \neq j$. The s_i act as reflections on $\mathbb{R}^n$. For any $w \in W$ let $\lambda(w)$ denote its spectral radius. This is the modulus of the largest eigenvalue of its action on $\mathbb{R}^n$. Then McMullen [45, Theorem 1.1] shows that $\lambda(w) \geq \tau_{10} = 1.176\ldots$. This could be interpreted as circumstantial evidence for λ_{10} indeed being the smallest Salem number.

The Coxeter diagram of (W, S) is the weighted graph whose vertices are the set S, and whose edges of weight m_{ij} join s_i to s_j when $m_{ij} \geq 3$. Denoting by $Y_{a,b,c}$ the Coxeter system whose diagram is a tree with 3 branches of lengths a, b and c, joined at a single node, McMullen also showed that the smallest Salem numbers of degrees 6, 8 and 10 coincide with $\lambda(w)$ for the Coxeter elements of $Y_{3,3,4}$, $Y_{2,4,5}$ and $Y_{2,3,7}$ respectively. In particular, $\lambda(w) = \tau_{10}$ for the Coxeter elements of $Y_{2,3,7}$.

4.8. **Dilatation of pseudo-Anosov automorphisms.** For a closed connected oriented surface $\mathcal{S}$ having a pseudo-Anosov automorphism that is a product of pairs of positive multi-twists, Leininger [37, Theorem 6.2] showed that its dilatation is at least τ_{10}. This follows from McMullen's work on Coxeter systems quoted above. The case of equality is explicitly described (in particular, $\mathcal{S}$ has genus 5). (However, on surfaces of genus g there are examples of pseudo-Anosov automorphisms having dilatations equal to $1+O(1/g)$ as $g \to \infty$. These are not Salem numbers when g is sufficiently large.)

4.9. **Bernoulli convolutions.** Following Solomyak [72], let $\lambda \in (0, 1)$, and $Y_\lambda = \sum_{n=0}^{\infty} \pm\lambda^n$, with the $\pm$ chosen independently '+' or '−' each with probability $\frac{1}{2}$. Let $\nu_\lambda(E)$ be the probability that $Y_\lambda \in E$, for any Borel set E. So it is the infinite convolution product of the means $\frac{1}{2}(\delta_{-\lambda^n} + \delta_{\lambda^n})$ for $n = 0, 1, 2, \ldots, \infty$, and so is called a *Bernoulli convolution.* Then $\nu_\lambda(E)$

satisfies the self-similarity property

$$\nu_\lambda(E) = \frac{1}{2}\left(\nu_\lambda(S_1^{-1}E) + \nu_\lambda(S_2^{-1}E)\right),$$

where $S_1 x = 1 + \lambda x$ and $S_2 x = 1 - \lambda x$. It is known that the support of ν_λ is a Cantor set of zero length when $\lambda \in (0, \frac{1}{2})$, and the interval $[-(1-\lambda)^{-1}, (1-\lambda)^{-1}]$ when $\lambda \in (\frac{1}{2}, 1)$. When $\lambda = \frac{1}{2}$, ν_λ is the uniform measure on $[-2, 2]$. Now the Fourier transform $\hat{\nu}_\lambda(\xi)$ of ν_λ is equal to $\prod_{n=0}^{\infty} \cos(\lambda^n \xi)$. Salem [63, p. 40] proved that if $\lambda \in (0, 1)$ and $1/\lambda$ is not a Pisot number, then $\lim_{\xi\to\infty} \hat{\nu}_\lambda(\xi) = 0$. This contrasts with an earlier result of Erdős that if $\lambda \neq \frac{1}{2}$ and $1/\lambda$ is a Pisot number, then $\hat{\nu}_\lambda(\xi)$ does not tend to 0 as $\xi \to \infty$. Recently Feng [25] has studied ν_λ when $1/\lambda$ is a Salem number, proving in this case that ν_λ the corresponding measure ν_λ is a multifractal measure satisfying the multifractal formalism in all of the increasing part of its multifractal spectrum.

Acknowledgements. This paper is an expanded version of a talk that I gave at the meeting 'Growth and Mahler measures in geometry and topology' at the Mittag-Leffler Institute, Djursholm, Sweden, in July 2013. I would like to thank the organisers Eriko Hironaka and Ruth Kellerhals for the invitation to attend the meeting, and to thank them, the Institute staff and fellow participants for making it such a stimulating and pleasant week.

I also thank David Boyd, Eriko Hironaka, Curtis McMullen, James McKee and Dan Silver for their very helpful comments and corrections on an earlier draft of this survey.

References

[1] Amoroso, Francesco, Une minoration pour l'exposant du groupe des classes d'un corps engendré par un nombre de Salem. Int. J. Number Theory **3** (2007), no. 2, 217-229.

[2] Bartholdi, Laurent; Ceccherini-Silberstein, Tullio G., Salem numbers and growth series of some hyperbolic graphs. Geom. Dedicata **90** (2002), 107-114.

[3] Bedford, Eric; Kim, Kyounghee, Periodicities in linear fractional recurrences: degree growth of birational surface maps. Michigan Math. J. **54** (2006), no. 3, 647-670.

[4] Bertin, M. J., K-nombres de Pisot et de Salem. Advances in number theory (Kingston, ON, 1991), 391-397, Oxford Sci. Publ., Oxford Univ. Press, New York, 1993.

[5] Bertin, M.-J., K-nombres de Pisot et de Salem. Acta Arith. **68** (1994), no. 2, 113-131.

[6] Bertin, M. J., Quelques nouveaux résultats sur les nombres de Pisot et de Salem. Number theory in progress, Vol. 1 (Zakopane-Kościelisko, 1997), 1-9, de Gruyter, Berlin, 1999.

[7] Bertin, M.-J.; Boyd, David W., A characterization of two related classes of Salem numbers. J. Number Theory **50** (1995), no. 2, 309-317.

[8] Bertin, M.-J.; Decomps-Guilloux, A.; Grandet-Hugot, M.; Pathiaux-Delefosse, M.; Schreiber, J.-P., Pisot and Salem numbers. Birkhäuser Verlag, Basel, 1992.

[9] Bertin, Marie-José; Pathiaux-Delefosse, Martine, Conjecture de Lehmer et petits nombres de Salem. Queen's Papers in Pure and Applied Mathematics, **81**. Queen's University, Kingston, ON, 1989.
[10] Borwein, Peter; Hare, Kevin G., Some computations on the spectra of Pisot and Salem numbers. Math. Comp. **71** (2002), no. 238, 767-780.
[11] Boyd, David W., Small Salem numbers. Duke Math. J. **44** (1977), no. 2, 315-328.
[12] Boyd, David W., Pisot sequences which satisfy no linear recurrence. Acta Arith. **32** (1977), no. 1, 89-98.
[13] Boyd, David W., Pisot and Salem numbers in intervals of the real line. Math. Comp. **32** (1978), no. 144, 1244-1260.
[14] Boyd, David W., Families of Pisot and Salem numbers. Seminar on Number Theory, Paris 1980-81 (Paris, 1980/1981), 19-33, Progr. Math., **22**, Birkhäuser, Boston, Mass., 1982.
[15] Boyd, David W., Salem numbers of degree four have periodic expansions. Théorie des nombres (Quebec, PQ, 1987), 57-64, de Gruyter, Berlin, 1989.
[16] Boyd, David W., On the beta expansion for Salem numbers of degree 6. Math. Comp. **65** (1996), no. 214, 861-875, S29-S31.
[17] Boyd, David W.; Parry, Walter, Limit points of the Salem numbers. Number Theory (Banff, AB, 1988), 27-35, de Gruyter, Berlin, 1990.
[18] Cantor, David G., On power series with only finitely many coefficients (mod 1): solution of a problem of Pisot and Salem. Acta Arith. 34 (1977/78), no. 1, 43-55.
[19] Christopoulos, Christos; McKee, James, Galois theory of Salem polynomials. Math. Proc. Cambridge Philos. Soc. **148** (2010), no. 1, 47-54.
[20] Decomps-Guilloux, Annette; Grandet-Hugot, Marthe, Nouvelles caractrisations des nombres de Pisot et de Salem. Acta Arith. **50** (1988), no. 2, 155-170.
[21] Doche, Christophe; Mendès France, Michel; Ruch, Jean-Jacques, Equidistribution modulo 1 and Salem numbers. Funct. Approx. Comment. Math. **39** (2008), part 2, 261-271.
[22] Dubickas, Artūras, Arithmetical properties of powers of algebraic numbers. Bull. London Math. Soc. **38** (2006), no. 1, 70-80.
[23] Dubickas, Artūras, Sumsets of Pisot and Salem numbers. Expo. Math. **26** (2008), no. 1, 85-91.
[24] Dubickas, Artūras; Smyth, Chris, On the lines passing through two conjugates of a Salem number. Math. Proc. Cambridge Philos. Soc. **144** (2008), no. 1, 29-37.
[25] Feng, De-Jun, Multifractal analysis of Bernoulli convolutions associated with Salem numbers. Adv. Math. **229** (2012), no. 5, 3052–3077.
[26] Flammang, V.; Grandcolas, M.; Rhin, G., Small Salem numbers. Number theory in progress, Vol. **1** (Zakopane-Kościelisko, 1997), 165-168, de Gruyter, Berlin, 1999.
[27] Ghate, Eknath; Hironaka, Eriko, The arithmetic and geometry of Salem numbers. Bull. Amer. Math. Soc. (N.S.) **38** (2001), no. 3, 293-314.
[28] Gromov, Mikhaïl, On the entropy of holomorphic maps. Enseign. Math. (2) **49** (2003), no. 3-4, 217-235.
[29] Gross, Benedict H.; McMullen, Curtis T., Automorphisms of even unimodular lattices and unramified Salem numbers. J. Algebra **257** (2002), no. 2, 265-290.
[30] Hironaka, Eriko, The Lehmer polynomial and pretzel links. Canad. Math. Bull. **44** (2001), no. 4, 440-451. Erratum: *ibid.* **45** (2002), no. 2, 231.
[31] Kerada, Mohamed, Une caractérisation de certaines classes d'entiers algbriques généralisant les nombres de Salem. Acta Arith. **72** (1995), no. 1, 55-65.
[32] Kronecker, Leopold, Zwei sätse über Gleichungen mit ganzzahligen Coefficienten, J. Reine Angew. Math. **53** (1857), 173–175. See also Werke. Vol. 1, 103–108, Chelsea Publishing Co., New York, 1968.
[33] Lakatos, Piroska, Salem numbers, PV numbers and spectral radii of Coxeter transformations. C. R. Math. Acad. Sci. Soc. R. Can. **23** (2001), no. 3, 71-77.

[34] Lakatos, Piroska, A new construction of Salem polynomials. C. R. Math. Acad. Sci. Soc. R. Can. **25** (2003), no. 2, 47-54.

[35] Lakatos, Piroska, Salem numbers defined by Coxeter transformation. Linear Algebra Appl. **432** (2010), no. 1, 144-154.

[36] Lalande, Franck, Corps de nombres engendrés par un nombre de Salem. Acta Arith. **88** (1999), no. 2, 191-200.

[37] Leininger, Christopher J., On groups generated by two positive multi-twists: Teichmüller curves and Lehmer's number. Geom. Topol.**8** (2004), 1301-1359.

[38] McKee, James; Smyth, Chris, Salem numbers of trace 2 and traces of totally positive algebraic integers. Algorithmic number theory, 327-337, Lecture Notes in Comput. Sci., **3076**, Springer, Berlin, 2004.

[39] McKee, James; Smyth, Chris, There are Salem numbers of every trace. Bull. London Math. Soc. **37** (2005), no. 1, 25-36.

[40] McKee, James; Smyth, Chris, Salem numbers, Pisot numbers, Mahler measure, and graphs. Experiment. Math. **14** (2005), no. 2, 211-229.

[41] McKee, James; Smyth, Chris, Salem numbers and Pisot numbers via interlacing. Canad. J. Math. **64** (2012), no. 2, 345-367.

[42] McKee, J. F.; Rowlinson, P.; Smyth, C. J., Salem numbers and Pisot numbers from stars. Number theory in progress, Vol. **1** (Zakopane-Kościelisko, 1997), 309-319, de Gruyter, Berlin, 1999.

[43] Maclachlan, Colin; Reid, Alan W., The arithmetic of hyperbolic 3-manifolds. Graduate Texts in Mathematics, **219**. Springer-Verlag, New York, 2003.

[44] McMullen, Curtis T., Dynamics on K3 surfaces: Salem numbers and Siegel disks. J. Reine Angew. Math. **545** (2002), 201–233.

[45] McMullen, Curtis T., Coxeter groups, Salem numbers and the Hilbert metric. Publ. Math. Inst. Hautes Études Sci. No. **95** (2002), 151–183.

[46] McMullen, Curtis T., Dynamics on blowups of the projective plane. Publ. Math. Inst. Hautes Études Sci. No. **105** (2007), 49-89.

[47] McMullen, Curtis T., K3 surfaces, entropy and glue. J. Reine Angew. Math. **658** (2011), 1–25.

[48] McMullen, Curtis T., Automorphisms of projective K3 surfaces with minimum entropy. Preprint.

[49] Meyer, Yves, Algebraic numbers and harmonic analysis. North-Holland Mathematical Library, Vol. **2**. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1972.

[50] Mossinghoff, Michael J., Polynomials with small Mahler measure. Math. Comp. **67** (1998), no. 224, 1697-1705, S11-S14.

[51] Mossinghoff, M., List of small Salem numbers.
`http://www.cecm.sfu.ca/~ mjm/Lehmer/lists/SalemList.html`

[52] Mossinghoff, Michael J.; Rhin, Georges; Wu, Qiang, Minimal Mahler measures. Experiment. Math. **17** (2008), no. 4, 451-458.

[53] Neumann, Walter D.; Reid, Alan W., Arithmetic of hyperbolic manifolds. Topology '90 (Columbus, OH, 1990), 273-310, Ohio State Univ. Math. Res. Inst. Publ., **1**, de Gruyter, Berlin, 1992.

[54] Oguiso, Keiji, Salem polynomials and birational transformation groups for hyper-Kähler manifolds. (Japanese) Sūgaku **59** (2007), no. 1, 1-23. English translation: Salem polynomials and the bimeromorphic automorphism group of a hyper-Kähler manifold. Selected papers on analysis and differential equations, Amer. Math. Soc. Transl. Ser. 2, **230** (2010), 201-227.

[55] Oguiso, Keiji, The third smallest Salem number in automorphisms of K3 surfaces. Algebraic geometry in East Asia-Seoul 2008, 331-360, Adv. Stud. Pure Math., **60**, Math. Soc. Japan, Tokyo, 2010.

[56] Pisot, Charles, Répartition (mod 1) des puissances successives des nombres réels. Comment. Math. Helv. **19** (1946), 153-160.

[57] Reidemeister, K., Knot theory. Translated from the German by Leo F. Boron, Charles O. Christenson and Bryan A. Smith. BCS Associates, Moscow, Idaho, 1983.

[58] Reschke, Paul, Salem numbers and automorphisms of complex surfaces. Math. Res. Lett. **19** (2012), no. 2, 475-482.

[59] Salem, R., A remarkable class of algebraic integers. Proof of a conjecture of Vijayaraghavan. Duke Math. J. **11** (1944), 103-108.

[60] Salem, R., Sets of uniqueness and sets of multiplicity. II. Trans. Amer. Math. Soc. **56** (1944), 32-49.

[61] Salem, R., Power series with integral coefficients. Duke Math. J. **12** (1945), 153-172.

[62] Salem, Raphaël, Algebraic numbers and Fourier analysis. D. C. Heath and Co., Boston, Mass., 1963.

[63] Samet, P. A., Algebraic integers with two conjugates outside the unit circle. I: Proc. Cambridge Philos. Soc. **49** (1953), 421-436; II: *ibid* **50** (1954), 346.

[64] Schmidt, Klaus, On periodic expansions of Pisot numbers and Salem numbers. Bull. London Math. Soc. **12** (1980), no. 4, 269-278.

[65] Schreiber, Jean-Pierre, Sur les nombres de Chabauty-Pisot-Salem des extensions algébriques de $\mathbb{Q}_p$. C. R. Acad. Sci. Paris Sér. A-B **269** (1969), A71-A73.

[66] Silver, Daniel S.; Williams, Susan G., Mahler measure of Alexander polynomials. J. London Math. Soc. (2) **69** (2004), no. 3, 767-782.

[67] Silverman, Joseph H., Small Salem numbers, exceptional units, and Lehmer's conjecture. Symposium on Diophantine Problems (Boulder, CO, 1994). Rocky Mountain J. Math. **26** (1996), no. 3, 1099-1114.

[68] Smith, John H., Some properties of the spectrum of a graph. Combinatorial Structures and their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969) pp. 403-406 Gordon and Breach, New York, 1970.

[69] Smyth, C. J., Generalised Salem numbers. Groupe de travail en théorie analytique et élémentaire des nombres, 1986-1987, 130-135, Publ. Math. Orsay, **88-01**, Univ. Paris XI, Orsay, 1988.

[70] Smyth, C. J., Salem numbers of negative trace. Math. Comp. **69** (2000), no. 230, 827-838.

[71] Smyth, Chris, The Mahler measure of algebraic numbers: a survey. Number theory and polynomials, 322-349, London Math. Soc. Lecture Note Ser., **352**, Cambridge Univ. Press, Cambridge, 2008.

[72] Solomyak, Boris, On the random series $\sum \pm\lambda^n$ (an Erdős problem). Ann. of Math. (2) **142** (1995), no. 3, 611-625.

[73] Sury, B., Arithmetic groups and Salem numbers. Manuscripta Math. **75** (1992), no. 1, 97–102.

[74] Vaz, Sandra; Martins Rodrigues, Pedro; Sousa Ramos, José, Pisot and Salem numbers in the β-transformation. Iteration theory (ECIT '06), 168-179, Grazer Math. Ber., **351**, Institut für Mathematik, Karl-Franzens-Universitüt Graz, Graz, 2007.

[75] Zaïmi, Toufik, On integer and fractional parts of powers of Salem numbers. Arch. Math. (Basel) **87** (2006), no. 2, 124-128.

[76] Zaïmi, Toufik, An arithmetical property of powers of Salem numbers. J. Number Theory **120** (2006), no. 1, 179-191.

[77] Zaïmi, Toufik, Comments on the distribution modulo one of powers of Pisot and Salem numbers. Publ. Math. Debrecen **80** (2012), no. 3-4, 417-426.

School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, Edinburgh, EH9 3JZ, Scotland

E-mail address: c.smyth@ed.ac.uk