APOLLONIAN CIRCLE PACKINGS AND CLOSED HOROSPHERES ON HYPERBOLIC 3-MANIFOLDS

ALEX KONTOROVICH AND HEE OH
(WITH APPENDIX BY HEE OH AND NIMISH SHAH)

ABSTRACT. We show that for a given bounded Apollonian circle packing $\mathcal{P}$, there exists a constant $c > 0$ such that the number of circles of curvature at most T is asymptotic to $c \cdot T^\alpha$ as $T \to \infty$. Here $\alpha \approx 1.30568(8)$ is the residual dimension of the packing. For $\mathcal{P}$ integral, let $\pi^{\mathcal{P}}(T)$ denote the number of circles with prime curvature less than T. Similarly let $\pi_2^{\mathcal{P}}(T)$ be the number of pairs of tangent circles with prime curvatures less than T. We obtain the upper bounds $\pi^{\mathcal{P}}(T) \ll T^\alpha / \log T$ and $\pi_2^{\mathcal{P}}(T) \ll T^\alpha/(\log T)^2$, which are sharp up to constant multiple. The main ingredient of our proof is the effective equidistribution of expanding closed horospheres in the unit tangent bundle of a geometrically finite hyperbolic 3-manifold $\Gamma\backslash\mathbb{H}^3$ under the assumption that the critical exponent of Γ exceeds one.

CONTENTS

2000 *Mathematics Subject Classification.* Primary22E40.
Key words and phrases. Apollonian circle packing, Horospheres, Kleinian group.
Kontorovich is supported by an NSF Postdoc, grant DMS 0802998.
Oh is partially supported by NSF grant DMS 0629322.

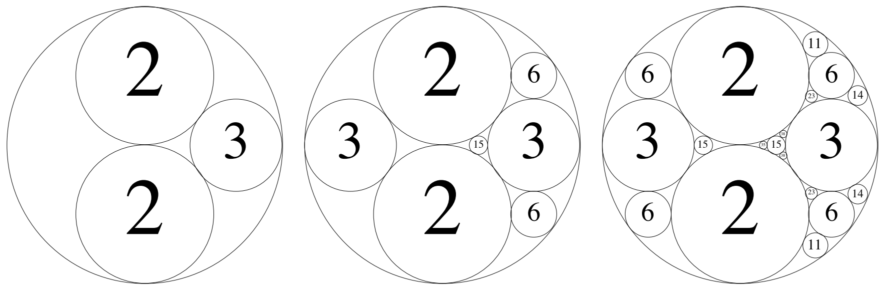

Figure 1. Generations 1-3.

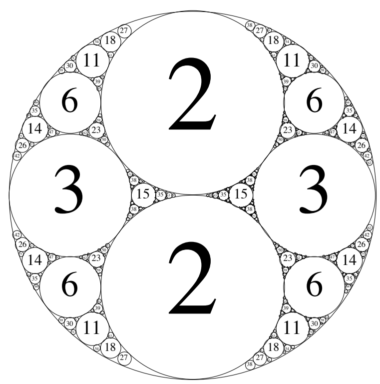

Figure 2. A bounded Apollonian circle packing.

1. Introduction

A set of four mutually tangent circles in the plane with distinct points of tangency is called a *Descartes configuration*. Given a Descartes configuration, one can construct four new circles, each of which is tangent to three of the given ones. Continuing to repeatedly fill the interstices between mutually tangent circles with further tangent circles, we arrive at an infinite circle packing. It is called an *Apollonian circle packing*, after the great geometer Apollonius of Perga (262-190 BC).

See Figure 1 showing the first three generations of this procedure, where each circle is labeled with its curvature (that is, the reciprocal of its radius). Unlike the inner circles, the bounding circle is oriented so that its "outward" normal vector points into the packing. In Figure 2, the outermost circle has curvature -1 (the sign conveys its orientation).

The astute reader would do well to peruse the lovely series of papers by Graham, Lagarias, Mallows, Wilks, and Yan on this beautiful topic, especially [16] and [15], as well as the recent letter of Sarnak to Lagarias [40] which inspired this paper.

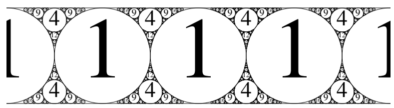

FIGURE 3. An Apollonian circle packing bounded by two parallel lines.

Counting circles in an Apollonian packing: Let $\mathcal{P}$ be either a bounded Apollonian circle packing or an unbounded one which is congruent to the packing in Figure 3.

For $\mathcal{P}$ bounded, denote by $N^{\mathcal{P}}(T)$ the number of circles in $\mathcal{P}$ in the packing whose curvature is at most T, i.e., whose radius is at least $1/T$. For $\mathcal{P}$ congruent to the packing in Figure 3, one alters the definition of $N^{\mathcal{P}}(T)$ to count circles in a fixed period.

It is easy to see that $N^{\mathcal{P}}(T)$ is finite for any given $T > 0$. The main goal of this paper is to obtain asymptotic formula for $N^{\mathcal{P}}(T)$ as T tends to infinity. To describe our results, recall that the residual set of $\mathcal{P}$ is defined to be the subset of the plane remaining after the removal of all of the interiors of circles in $\mathcal{P}$ (where the circles are oriented so that the interiors are disjoint). Let $\alpha = \alpha_{\mathcal{P}}$ denote the Hausdorff dimension of the residual set of $\mathcal{P}$. As any Apollonian packing can be moved to any other by a Möbius transformation, α does not depend on $\mathcal{P}$. The current record towards the exact value of $\alpha = 1.30568(8)$ is due to McMullen [29].

Boyd [6] showed in 1982 that

$$\lim_{T\to\infty} \frac{\log N^{\mathcal{P}}(T)}{\log T} = \alpha.$$

This confirmed Wilker's prediction [48] based on computer experiments.

Regarding an asymptotic formula for $N^{\mathcal{P}}(T)$, it was not clear from the literature whether one should conjecture a strictly polynomial growth rate. In fact, Boyd's numerical experiments led him to wonder whether "perhaps a relationship such as $N^{\mathcal{P}}(T) \sim c \cdot T^{\alpha}(\log(T/c'))^{\beta}$ might be more appropriate" (see [6] page 250).

In this paper, we show that $N^{\mathcal{P}}(T)$ has purely polynomial asymptotic growth. By $f(T) \sim g(T)$ with $T \to \infty$, we mean that $\lim_{T\to\infty} \frac{f(T)}{g(T)} = 1$

Theorem 1.1. *Given an Apollonian circle packing $\mathcal{P}$ which is either bounded or congruent to Figure 3, there exists $c = c(\mathcal{P}) > 0$ such that as $T \to \infty$,*

$$N^{\mathcal{P}}(T) \sim c \cdot T^{\alpha}.$$

In [6], Boyd actually considered the more general problem of counting those circles in a packing which are contained in a curvilinear triangle $\mathcal{R}$; let $N^{\mathcal{R}}(T)$ count the circles having curvature at most T (see Fig. 4). For this

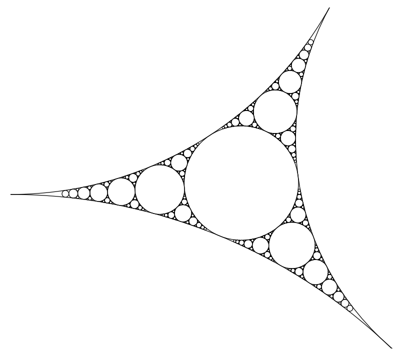

FIGURE 4. Circles in a curvilinear triangle.

question, it does not matter whether or not the full packing is bounded; the counting function $N^{\mathcal{R}}(T)$ is always well-defined. Since two such triangles are bi-Lipschitz equivalent, it follows from Theorem 1.1 that there exist constants $c_1, c_2 > 0$ such that for all $T \gg 1$,

$$c_1 \cdot T^\alpha \le N^{\mathcal{R}}(T) \le c_2 \cdot T^\alpha.$$

Though we believe that the asymptotic formula $N^{\mathcal{R}}(T) \sim c \cdot T^\alpha$ always holds, our techniques cannot yet establish this in full generality.

Primes and twin primes in an integral packing: A quadruple (a, b, c, d) of the curvatures of four circles in a Descartes configuration is called a *Descartes quadruple*. The Descartes circle theorem (see e.g. [9]) states that any Descartes quadruple (a, b, c, d) satisfies the quadratic equation[1]:

$$a^2 + b^2 + c^2 + d^2 = \frac{1}{2}(a + b + c + d)^2. \tag{1.2}$$

Given any three mutually tangent circles with distinct points of tangency and curvatures a, b and c, there are exactly two circles which are tangent to all of the given ones, having curvatures d and d', say. It easily follows from (1.2) that

$$d + d' = 2(a + b + c). \tag{1.3}$$

In particular, this shows that if a Descartes quadruple (a, b, c, d) corresponding to the initial four circles in the packing $\mathcal{P}$ is integral, then every circle in $\mathcal{P}$ also has integral curvature, as first observed by Soddy in 1937 [43]. Such a packing is called *integral*.

It is natural to inquire about the Diophantine properties of an integral Apollonian packing, such as how many circles in $\mathcal{P}$ have prime curvatures.

[1]Arguably the most elegant formulation of this theorem is the following excerpt from the poem "The Kiss Precise" by Nobel Laureate Sir Fredrick Soddy [44]:

Four circles to the kissing come. / The smaller are the bender. /
The bend is just the inverse of / The distance from the center. /
Though their intrigue left Euclid dumb / There's now no need for rule of thumb. /
Since zero bend's a dead straight line / And concave bends have minus sign, /
The sum of the squares of all four bends / Is half the square of their sum.

By rescaling, we may assume that $\mathcal{P}$ is *primitive*, that is, the greatest common divisor of the curvatures is one. We call a circle *prime* if its curvature is a prime number. A pair of prime circles which are tangent to each other will be called *twin prime circles*. It is easy to see that a primitive integral packing is either bounded or the one pictured in Fig. 3.

For $\mathcal{P}$ bounded, denote by $\pi^{\mathcal{P}}(T)$ the number of prime circles in $\mathcal{P}$ of curvature at most T, and by $\pi_2^{\mathcal{P}}(T)$ the number of twin prime circles in $\mathcal{P}$ of curvatures at most T. For $\mathcal{P}$ congruent to the packing in Figure 3, one alters the definition of $\pi^{\mathcal{P}}(T)$ and $\pi_2^{\mathcal{P}}(T)$ to count prime circles in a fixed period.

Sarnak showed in [40] that there are infinitely many prime and twin prime circles in any primitive integral packing $\mathcal{P}$, and that

$$\pi^{\mathcal{P}}(T) \gg \frac{T}{(\log T)^{3/2}}.$$

Using the recent results of Bourgain, Gamburd and Sarnak in [4] and [5] on the uniform spectral gap property of Zariski dense subgroups of $\mathrm{SL}_2(\mathbb{Z}[i])$, together with the Selberg's upper bound sieve, we prove:

Theorem 1.4. *Given a primitive integral Apollonian circle packing $\mathcal{P}$,*

(1) $\pi^{\mathcal{P}}(T) \ll \frac{T^\alpha}{\log T}$;
(2) $\pi_2^{\mathcal{P}}(T) \ll \frac{T^\alpha}{(\log T)^2}$.

Remarks:

(1) The number of pairs of tangent circles in $\mathcal{P}$ of curvatures at most T turn out to be equal to $3N_{\mathcal{P}}(T)$ up to an additive constant (see Lemma 2.5). Therefore, in light of Theorem 1.1, the upper bounds in Theorem 1.4 are only off by a constant multiple from the expected asymptotics.
(2) A suitably modified version of Conjecture 1.4 in [4], a generalization of Schinzel's hypothesis, implies that for some $c, c_2 > 0$,

$$\pi^{\mathcal{P}}(T) \sim c \cdot \frac{T^\alpha}{\log T} \quad \text{and} \quad \pi_2^{\mathcal{P}}(T) \sim c_2 \cdot \frac{T^\alpha}{(\log T)^2}$$

(see the discussion in [4, Ex D]). The constants c and c_2 are detailed in [13].

Orbital counting of a Kleinian group in a cone: By Descartes' theorem (1.2), any Descartes quadruple (a, b, c, d) lies on the cone $Q(x) = 0$, where Q denotes the quadratic form

$$Q(a, b, c, d) = a^2 + b^2 + c^2 + d^2 - \frac{1}{2}(a + b + c + d)^2.$$

In light of (1.3), one can "flip" the quadruple (a, b, c, d) into (a, b, c, d') via multiplication by S_4, where

$$S_1 = \begin{pmatrix} -1 & 2 & 2 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & -1 & 2 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$S_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 2 & -1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 2 & -1 \end{pmatrix}.$$

Let $\mathcal{A}$ denote the so-called Apollonian group generated by these reflections, that is $\mathcal{A} = \langle S_1, S_2, S_3, S_4 \rangle$, and let O_Q be the orthogonal group preserving Q. One can check that $\mathcal{A} < \mathrm{O}_Q(\mathbb{Z})$ and that Q has signature $(3, 1)$. Therefore $\mathcal{A}$ is a Kleinian group; moreover $\mathcal{A}$ turns out to be of infinite index in $\mathrm{O}_Q(\mathbb{Z})$.

For a fixed packing $\mathcal{P}$, there is a labeling by the Apollonian group $\mathcal{A}$ of all the (unordered) Descartes quadruples in $\mathcal{P}$. Moreover the counting problems for $N^{\mathcal{P}}(T)$ and $N_2^{\mathcal{P}}(T)$ for $\mathcal{P}$ bounded can be reduced to counting elements in the orbit $\xi_{\mathcal{P}} \cdot \mathcal{A}^t \subset \mathbb{R}^4$ of maximum norm at most T, where $\xi_{\mathcal{P}}$ is the unique root quadruple of $\mathcal{P}$ (see Def. 2.2 and Lemma 2.5). For $\mathcal{P}$ congruent to Figure 3, the same reduction holds with ξ_P given by $(0, 0, c, c)$ where c is the curvature of the largest circle in $\mathcal{P}$.

We prove the following more general counting theorem: Let $\iota : \mathrm{PSL}_2(\mathbb{C}) \to \mathrm{SO}_F(\mathbb{R})$ be a real linear representation, where F is a real quadratic form in 4 variables with signature $(3, 1)$. Let $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ be a geometrically finite Kleinian group. The limit set $\Lambda(\Gamma)$ of Γ is the set of accumulation points of Γ-orbits in the ideal boundary $\partial_\infty(\mathbb{H}^3)$ of the hyperbolic space $\mathbb{H}^3$. We assume that the Hausdorff dimension δ_Γ of $\Lambda(\Gamma)$ is strictly bigger than one.

Theorem 1.5. *Let $v_0 \in \mathbb{R}^4$ be a non-zero vector lying in the cone $F = 0$ with a discrete orbit $v_0\Gamma \subset \mathbb{R}^4$. Then for any norm $\|\cdot\|$ on $\mathbb{R}^4$, as $T \to \infty$,*

$$\#\{v \in v_0\Gamma : \|v\| < T\} \sim c \cdot T^{\delta_\Gamma}$$

where $c > 0$ is explicitly given in Theorem 7.1.

There are two main difficulties preventing existing counting methods from tackling the above asymptotic formula. The first is that Γ is not required to be a lattice in $\mathrm{PSL}_2(\mathbb{C})$ (recall that the Apollonian group $\mathcal{A}$ has infinite index in $\mathrm{O}_Q(\mathbb{Z})$), so Patterson-Sullivan theory enters in the spectral decomposition of the hyperbolic manifold $\Gamma\backslash\mathbb{H}^3$. The second difficulty noted by Sarnak in [40] stems from the fact that the stabilizer of v_0 in Γ may not have enough unipotent elements; in the application to Apollonian packings, the stabilizer is indeed either finite or a rank one abelian subgroup, whereas the stabilizer of v_0 in the ambient group G is a compact extension of a rank two abelian subgroup.

In the similar situation of an infinite area hyperbolic *surface*, that is when $\Gamma < \mathrm{PSL}_2(\mathbb{R})$, the counting problem in a cone with respect to a Euclidean norm was solved in Kontorovich's thesis [21], under the assumption that the stabilizer of v_0 in Γ is co-compact in the stabilizer of v_0 in $\mathrm{PSL}_2(\mathbb{R})$. The methods developed here are quite different. Our approach to the counting problem is via the equidistribution of expanding closed horospheres on hyperbolic 3-manifolds (see the next subsection for more detailed discussion). Our proof of the equidistribution works for hyperbolic surfaces as well, and in particular solves the counting problem in [21] for any norm and without assumptions on the stabilizer. In [22], we apply the methods of this paper to the problem of thin orbits of Pythagorean triples having few prime factors.

Equidistribution of expanding horospheres in hyperbolic 3-manifolds: Let Γ be a geometrically finite torsion-free discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. The main ingredient of our proof of Theorem 1.5 is the equidistribution of expanding closed horospheres in $\Gamma\backslash\mathbb{H}^3$.

The group $G = \mathrm{PSL}_2(\mathbb{C})$ is the group of orientation preserving isometries of the hyperbolic space $\mathbb{H}^3 = \{(x_1, x_2, y) : y > 0\}$. The invariant measure on $\mathbb{H}^3$ for the action of G and the Laplace operator are given respectively by

$$\frac{dx_1 dx_2 dy}{y^3} \quad \text{and} \quad \Delta = -y^2\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right) + y\frac{\partial}{\partial y}.$$

Set

$$\begin{aligned} N &= \{n_x := \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{C}\}, \\ A &= \{a_y := \begin{pmatrix} \sqrt{y} & 0 \\ 0 & \sqrt{y}^{-1} \end{pmatrix} : y > 0\}, \\ K &= \{g \in G : \bar{g}^t g = I\} \quad \text{and} \\ M &= \{\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} : \theta \in \mathbb{R}\}. \end{aligned} \tag{1.6}$$

By the Iwasawa decomposition $G = NAK$, any element $g \in G$ can be written uniquely as $g = n_x a_y k$ with $n_x \in N$, $a_y \in A$, and $k \in K$. Via the map

$$n_x a_y (0, 0, 1) = (x, y),$$

the hyperbolic space $\mathbb{H}^3$ and its unit tangent bundle $\mathrm{T}^1(\mathbb{H}^3)$ can be identified with the quotients G/K and G/M respectively.

Denoting by $[u]$ the image of $u \in G$ under the quotient map $G \to G/M$, the horospheres correspond to N-leaves $[u]N = [uN]$ in G/M; note this is well-defined as M normalizes N. For a closed horosphere $\Gamma\backslash\Gamma[u]N$, the translates $\Gamma\backslash\Gamma[u]Na_y$ represent closed horospheres $\Gamma\backslash\Gamma[ua_y]N$ which are expanding as $y \to 0$.

In the case of a hyperbolic surface $\Gamma\backslash\mathbb{H}^2$ of finite area, it is a theorem of Sarnak's [39] that such long horocycle flows are equidistributed with respect

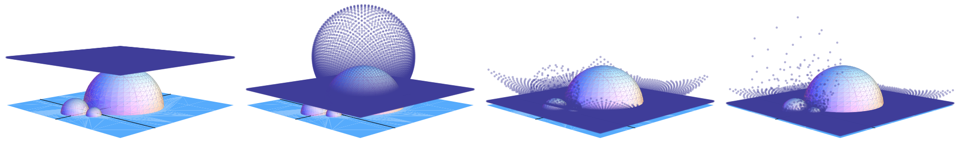

FIGURE 5. Equidistribution of expanding horospheres

to the Haar measure. The equidistribution of expanding horospheres for any *finite volume* hyperbolic manifold can also be proved using the mixing of geodesic flows. This approach appears already in Margulis's 1970 thesis [25]; see also [12].

In what follows we describe our equidistribution result for expanding closed horospheres on any geometrically finite hyperbolic 3 manifolds.

Assuming $\delta_\Gamma > 1$, Sullivan [45] showed, generalizing the work of Patterson [34], that there exists a unique positive eigenfunction ϕ_0 of the Laplacian operator Δ on $\Gamma\backslash\mathbb{H}^3$ of lowest eigenvalue $\delta_\Gamma(2-\delta_\Gamma)$ and of unit L^2-norm, that is, $\int_{\Gamma\backslash\mathbb{H}^3} \phi_0(x,y)^2 \frac{1}{y^3} dxdy = 1$. Moreover the base eigenvalue-value $\delta_\Gamma(2-\delta_\Gamma)$ is isolated in the L^2-spectrum of Δ by Lax-Phillips [24].

Note that the closed leaf $\Gamma\backslash\Gamma N$ inside $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ is an embedding of one of the following: a complex plane, a cylinder, or a torus. As $\phi_0 > 0$, it is a priori not clear whether the integral

$$\phi_0^N(a_y) := \int_{n_x \in (N\cap\Gamma)\backslash N} \phi_0(x,y)dx$$

converges. We show that for any $y > 0$, the integral $\phi_0^N(a_y)$ does converge and is of the form

$$\phi_0^N(a_y) = c_{\phi_0} y^{2-\delta_\Gamma} + d_{\phi_0} y^{\delta_\Gamma}$$

for some constants $c_{\phi_0} > 0$ and $d_{\phi_0} \geq 0$. By $f(y) \sim g(y)$ with $y \to 0$, we mean that $\lim_{y\to 0} \frac{f(y)}{g(y)} = 1$. The measure dn on N is given by $dn_x = dx$.

Theorem 1.7. *Let $\Gamma < G$ be a geometrically finite torsion-free discrete subgroup with $\delta_\Gamma > 1$ and let $\Gamma\backslash\Gamma N$ be closed. There exists $\epsilon > 0$ such that for any $\psi \in C_c^\infty(\Gamma\backslash G)^K = C_c(\Gamma\backslash\mathbb{H}^3)$, and for all small $y > 0$*

$$\int_{(N\cap\Gamma)\backslash N} \psi(\Gamma\backslash\Gamma n a_y)\, dn = c_{\phi_0} \cdot \langle \psi, \phi_0\rangle_{L^2(\Gamma\backslash\mathbb{H}^3)} \cdot y^{2-\delta_\Gamma}(1 + O(y^\epsilon))$$

where the implied constant depends only on the Sobolev norm of ψ.

Thus as $y \to 0$, the integral of any function $\psi \in C_c(\Gamma\backslash G)^K$ along the orthogonal translate $\Gamma\backslash\Gamma N a_y$ converges to 0 with the speed of order $y^{2-\delta_\Gamma}$. It also follows that for $\psi \in C_c(\Gamma\backslash G)^K$, as $y \to 0$,

$$\int_{(N\cap\Gamma)\backslash N} \psi(\Gamma\backslash\Gamma n a_y)\, dn \sim \langle \psi, \phi_0\rangle \cdot \int_{(N\cap\Gamma)\backslash N} \phi_0(\Gamma\backslash\Gamma n a_y)\, dn.$$

We denote by $\tilde{\Omega}_\Gamma$ the set of vectors $(p, \vec{v})$ in the unit tangent bundle $\mathrm{T}^1(\mathbb{H}^3)$ such that the end point of the geodesic ray tangent to $\vec{v}$ belongs to the limit set $\Lambda(\Gamma)$ and by $\hat{\Omega}_\Gamma$ its image under the projection of $\mathrm{T}^1(\mathbb{H}^3)$ to $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$.

Roblin [38], generalizing the work of Burger [7], showed that, up to a constant multiple, there exists a unique Radon measure $\hat{\mu}$ on $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ invariant for the horospherical foliations which is supported on $\hat{\Omega}_\Gamma$ and gives zero measure to all closed horospheres.

In the appendix A written jointly by Shah and the second named author, the following theorem is deduced from Theorem 1.7, based on the aforementioned measure classification of Burger and Roblin. In view of the isomorphism $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3) = \Gamma\backslash G/M$, the following theorem shows that the orthogonal translations of closed horospheres in the expanding direction are equidistributed in $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ with respect to the Burger-Roblin measure $\hat{\mu}$.

Theorem 1.8. *For any $\psi \in C_c(\Gamma\backslash G)^M$, as $y \to 0$,*

$$\int_{(N\cap\Gamma)\backslash N} \psi(\Gamma\backslash\Gamma n a_y)\, dn \sim c_{\phi_0} \cdot y^{2-\delta_\Gamma} \cdot \hat{\mu}(\psi)$$

where $\hat{\mu}$ is normalized so that $\hat{\mu}(\phi_0) = 1$.

Theorem 1.8 was proved by Roblin [38] when $(N\cap\Gamma)\backslash N$ is compact with a different interpretation of the constant c_{ϕ_0}. His proof does not yield an effective version as in Theorem 1.7 but works for any $\delta_\Gamma > 0$.

We also remark that the quotient type equidistribution results for *non-closed* horocycles for geometrically finite surfaces were established by Schapira [41].

We conclude the introduction by giving a brief outline of the proof of Theorem 1.7. By Dal'bo (Theorem 3.3), we have the following classification of closed horospheres in terms of its base point in the boundary $\partial_\infty(\mathbb{H}^3)$: $\Gamma\backslash\Gamma N$ is closed if and only if either $\infty \notin \Lambda(\Gamma)$ or ∞ is a bounded parabolic fixed point (see Def. 3.1). This classification is used repeatedly in our analysis establishing the following facts:

(1) For any bounded subset $B \subset (N\cap\Gamma)\backslash N$ which properly covers $(N\cap\Gamma)\backslash(\Lambda(\Gamma) - \{\infty\})$ (cf. Def. 4.5),

$$\int_{(N\cap\Gamma)\backslash N} \phi_0(n a_y)\, dn = \int_B \phi_0(n a_y)\, dn + O(y^{\delta_\Gamma}).$$

(2) Denoting by $\rho_{B,\epsilon} \in C_c(\Gamma\backslash G)$ the ϵ-approximation of B in the transversal direction,

$$\int_B \phi_0(n a_y)\, dn = \langle a_y \phi_0, \rho_{B,\epsilon}\rangle + O(\epsilon y^{2-\delta_\Gamma}) + O(y^{\delta_\Gamma}).$$

(3) For $\psi \in C_c(\Gamma\backslash\mathbb{H}^3)$, there exists a compact subset $B = B(\mathrm{supp}(\psi)) \subset (N \cap \Gamma)\backslash N$ such that for all $0 < y < 1$

$$\int_{(N\cap\Gamma)\backslash N} \psi(na_y)\, dn = \int_B \psi(na_y)\, dn.$$

These facts allow us to focus on the integral of ψ over a compact region, say, B, of $(N \cap \Gamma)\backslash N$ instead of the whole space. In approximating $\int_B \psi(na_y)\, dn$ with $\langle a_y\psi, \rho_{B,\epsilon}\rangle$, the usual argument based on the contracting property along the stable horospheres is not sufficient: the error terms combine with those coming from the spectral gap to overtake the main term! We develop a recursive argument which improves the error upon each iteration, and halts in finite time, once the main term is dominant. Using the spectral theory of $L^2(\Gamma\backslash G)$ along with the assumption $\delta_\Gamma > 1$, we get a control of the main term of $\langle a_y\psi, \rho_{B,\epsilon}\rangle$ as $\langle \psi, \phi_0\rangle \cdot \langle a_y\phi_0, \rho_{B,\epsilon}\rangle$.

For the application to Apollonian packings, we need to consider the max norm, which necessitates the extension of our argument to the unit tangent bundle, that is, the deduction of Theorem 1.8 from 1.7 as done in Appendix A.

We finally remark that the power savings error term in (1) of Theorem 1.7 is crucial to prove Theorem 1.4.

After this paper was submitted, the asymptotic formula for counting circles in a curvilinear triangle of any Apollonian packing has been obtained in [33]. See also [32] for similar counting results for hyperbolic and spherical Apollonian circle packings. We also refer to [31] for a survey on recent progress on counting circles.

Acknowledgments. We are grateful to Peter Sarnak for introducing us to this problem and for helpful discussions. We also thank Yves Benoist, Jeff Brock and Curt McMullen for useful conversations.

2. Reduction to orbital counting

2.1. **Apollonian group.** In a quadruple of mutually tangent circles the curvatures a, b, c, d satisfy the Descartes equation:

$$a^2 + b^2 + c^2 + d^2 = \frac{1}{2}(a + b + c + d)^2$$

as observed by Descartes in 1643 (see [9] for a proof).

Any quadruple (a, b, c, d) satisfying this equation is called a Descartes quadruple. A set of four mutually tangent circles with disjoint interiors is called a Descartes configuration.

We denote by Q the Descartes quadratic form given by

$$Q(a, b, c, d) = a^2 + b^2 + c^2 + d^2 - \frac{1}{2}(a + b + c + d)^2.$$

Hence $v = (a, b, c, d)$ is a Descartes quadruple if and only if $Q(v) = 0$. The orthogonal group corresponding to Q is given by

$$\mathrm{O}_Q = \{g \in \mathrm{GL}_4 : Q(vg^t) = Q(v) \text{ for all } v \in \mathbb{R}^4\}.$$

One can easily check that the Apollonian group $\mathcal{A} := \langle S_1, S_2, S_3, S_4\rangle$ defined in the introduction is a subgroup of $\mathrm{O}_Q(\mathbb{Z}) := \mathrm{O}_Q \cap \mathrm{GL}_4(\mathbb{Z})$.

Definition 2.1. (1) For $\mathcal{P}$ bounded, denote by $N^{\mathcal{P}}(T)$ the number of circles in $\mathcal{P}$ in the packing whose curvature is at most T, i.e., whose radius is at least $1/T$. Denote by $N_2^{\mathcal{P}}(T)$ the number of pairs of tangent circles in $\mathcal{P}$ of curvatures at most T.

(2) For $\mathcal{P}$ congruent to the packing in Figure 3, $N^{\mathcal{P}}(T)$ denotes the number of circles between two largest tangent circles including the lines and the largest circles. Similarly, $N_2^{\mathcal{P}}(T)$ denotes the number of unordered pairs of tangent circles between two largest tangent circles including the pairs containing the lines and the largest circles.

(3) For $\mathcal{P}$ bounded, denote by $\pi^{\mathcal{P}}(T)$ the number of prime circles in $\mathcal{P}$ of curvature at most T. Denote by $\pi_2^{\mathcal{P}}(T)$ the number of twin prime circles in $\mathcal{P}$ of curvatures at most T.

(4) For $\mathcal{P}$ congruent to the packing in Figure 3, one alters the definition of $\pi^{\mathcal{P}}(T)$ and $\pi_2^{\mathcal{P}}(T)$ to count prime circles in a fixed period.

We will interpret $N^{\mathcal{P}}(T)$ and $N_2^{\mathcal{P}}(T)$ as orbital counting functions on $\xi\mathcal{A}^t$ for a carefully chosen Descartes quadruple ξ of $\mathcal{P}$.

Definition 2.2. A Descartes quadruple $v = (a, b, c, d)$ with $a+b+c+d > 0$ is a root quadruple if $a \le 0 \le b \le c \le d$ and $a + b + c \ge d$.

If $\mathcal{P}$ is bounded, Theorem 3.2 in [15] shows that $\mathcal{P}$ contains a unique Descartes root quadruple $\xi := (a, b, c, d)$ with $a < 0$.

Theorem 2.3. [15, Thm 3.3] *The set of curvatures occurring in $\mathcal{P}$, counted with multiplicity, consists of the four entries in ξ, together with the largest entry in each vector $\xi\gamma^t$ as γ runs over all non-identity elements of the Apollonian group $\mathcal{A}$.*

In [15], this is stated only for an integral Apollonian packing, but the same proof works for any bounded packing. Let $w^{(n)}$ be a non-returning walk away from the root quadruple ξ along the Apollonian group, i.e., $w^{(n)} = \xi S_{i_1}^t \cdots S_{i_n}^t$ with $S_{i_k} \ne S_{i_{k+1}}$, $1 \le k \le n-1$. Then the key observation in the proof of above theorem is that $w^{(n)}$ is obtained from $w^{(n-1)}$ by changing one entry, and moreover the new entry inserted is always the largest entry in the new vector.

This theorem yields that for $T \gg 1$,

$$N^{\mathcal{P}}(T) = \#\{\gamma \in \mathcal{A} : \|\xi\gamma^t\|_{\max} < T\} + 3.$$

Consider the repeated generations of $\mathcal{P}$ with initial 4 circles given by the root quadruple. Then a geometric version of Theorem 2.3 is that for $n \ge 1$, each reduced word $\gamma = S_{i_n} \cdots S_{i_1}$ of length n corresponds to exactly one new circle, say C_γ, added at the n-th generation and the curvature of C_γ is the maximum among the entries of the quadruple $\xi\gamma^t$ (cf. [16, Section 4]). Thus the correspondence $\phi : \gamma \mapsto C_\gamma$ establishes a bijection between the set of all non-identity elements of $\mathcal{A}$ and the set of all circles not in the zeroth generation, and the set $\{C_\gamma | \gamma \neq e, \|\xi\gamma^t\| < T\}$ gives all circles (excluding those 4 initial circles) of curvature at most T.

For each $\gamma \neq e$ in $\mathcal{A}$, set

$$\phi_2(\gamma) = \{\{C_\gamma, C_\gamma(1)\}, \{C_\gamma, C_\gamma(2)\}, \{C_\gamma, C_\gamma(3)\}\}$$

where $C_\gamma(i)$, $i = 1, 2, 3$, are the three circles corresponding to the quadruple $\xi\gamma^t$ besides C_γ. Noting that each $(C_\gamma, C_\gamma(i))$ gives a pair of tangent circles, we claim that every pair of tangent circles arise as one of the triples in the image of ϕ_2, provided one of the circles in the pair does not come from the zeroth stage. If C and D form such a pair, they are from different stages of generation as no two circles in the same generation level touch each other, except for the initial stage. If, say, D is generated earlier than C, then for the element $\gamma \in \mathcal{A}$ giving $C = C_\gamma$, which is necessarily a non-identity element, D must be one of $C_\gamma(i)$'s. This is because it is clear from the construction of the packing that every circle is tangent *only* to three circles from previous generations.

Therefore ϕ_2 yields a one to three correspondence from $\mathcal{A} \setminus \{e\}$ to the set of all unordered pairs of tangent circles in $\mathcal{P}$ at least one circle of whose pair does not correspond to the root quadruple. Since there are 6 pairs arising from the initial 4 circles, we deduce that for $T \gg 1$,

$$N_2^{\mathcal{P}}(T) = 3 \cdot \#\{\gamma \in \mathcal{A} : \|\xi\gamma^t\|_{\max} < T\} + 3. \tag{2.4}$$

The above argument establishing (2.4) was kindly explained to us by Peter Sarnak.

If $\mathcal{P}$ lies between two parallel lines, that is, congruent to Figure 3, there exists the unique $c > 0$ such that $\mathcal{P}$ contains a Descartes quadruple $\xi := (0, 0, c, c)$.

In this case, the stabilizer of ξ in $\mathcal{A}^t$ is generated by two reflections S_3^t and S_4^t. One can directly verify that for all $T \gg 1$,

$$N^{\mathcal{P}}(T) = \#\{v \in \xi\mathcal{A}^t : \|v\|_{\max} < T\} + 3;$$

and

$$N_2^{\mathcal{P}}(T) = 3 \cdot \#\{v \in \xi\mathcal{A}^t : \|v\|_{\max} < T\} + 3.$$

Lemma 2.5. *Let $\mathcal{P}$ be either bounded or congruent to Fig. 3.*

(1) *For all $T \gg 1$,*

$$N^{\mathcal{P}}(T) = \begin{cases} \#\operatorname{Stab}_{\mathcal{A}^t}(\xi) \cdot \#\{v \in \xi\mathcal{A}^t : \|v\|_{\max} < T\} + 3 & \textit{for } \mathcal{P} \textit{ bounded} \\ \#\{v \in \xi\mathcal{A}^t : \|v\|_{\max} < T\} + 3 & \textit{otherwise.} \end{cases}$$

(2) *For all* $T \gg 1$,

$$N_2^{\mathcal{P}}(T) = 3N^{\mathcal{P}}(T) - 6.$$

(3) *The orbit* $\xi\mathcal{A}^t$ *is discrete in* $\mathbb{R}^4$.

(4) *For all* $T \gg 1$,

$$\pi^{\mathcal{P}}(T) \ll \sum_{i=1}^{4} \#\{v = (v_1, v_2, v_3, v_4) \in \xi\mathcal{A}^t : \|v\|_{\max} < T, v_i \textit{ is prime}\}.$$

(5) *For all* $T \gg 1$,

$$\pi_2^{\mathcal{P}}(T) \ll \sum_{1 \le i \ne j \le 4} \#\{v \in \xi\mathcal{A}^t : \|v\|_{\max} < T, v_i, v_j \textit{ are primes}\}.$$

Proof. The first two claims follow immediately from the discussion above, noting that the stabilizer of ξ in $\mathcal{A}^t$ is finite for $\mathcal{P}$ bounded. The third claim follows from the fact that $N^{\mathcal{P}}(T) < \infty$ for any $T > 0$. For claims (4) and (5), note that for $\mathcal{P}$ bounded,

$$\begin{aligned}\pi^{\mathcal{P}}(T) &\le 3 + \#\{\gamma \in \mathcal{A} : \|\xi\gamma^t\|_{\max} \text{ is prime } < T\} \\ &\ll \sum_{i=1}^{4} \#\{v = (v_1, v_2, v_3, v_4) \in \xi\mathcal{A}^t : \|v\|_{\max} < T, v_i \text{ is prime}\},\end{aligned}$$

and

$$\begin{aligned}\pi_2^{\mathcal{P}}(T) &\le 5 + \#\{\gamma \in \mathcal{A} : \|\xi\gamma^t\|_{\max} \text{ is prime } < T, \text{one more entry of } \xi\gamma^t \text{ is prime}\} \\ &\ll \sum_{i=1}^{4} \sum_{j \ne i} \#\{v = (v_1, v_2, v_3, v_4) \in \xi\mathcal{A}^t : \|v\|_{\max} < T, v_i, v_j \text{ are primes}\}.\end{aligned}$$

The claim (4) and (5) for $\mathcal{P}$ congruent to Fig. 3 can be shown similarly.

□

We remark that there are bounded packings which are not multiples of integral packings: for instance, $\xi = (3 - 2\sqrt{3}, 1, 1, 1)$ is a Descartes root quadruple which defines a bounded Apollonian packing. This is obvious from the viewpoint of geometry, but it is not at all clear *a priori* that the orbit $\xi\mathcal{A}^t$ should be discrete. Note also that there are other unbounded packings: by applying a suitably chosen Möbius transformation to a given packing, one can arrive at a packing which spreads uncontrollably to the entire plane, or one which is fenced off along one side by a single line.

2.2. **The residual set.** We consider the upper half-space model for the hyperbolic space:

$$\mathbb{H}^3 = \{(x_1, x_2, y) \in \mathbb{R}^3 : y > 0\}$$

with metric given by $\frac{dx_1^2 + dx_2^2 + dy^2}{y^3}$. A discrete subgroup of $\mathrm{Isom}(\mathbb{H}^3)$ is called a Kleinian group.

The ideal boundary $\partial_\infty(\mathbb{H}^3)$ of $\mathbb{H}^3$ can be identified with the set of geodesic rays emanating from a fixed point $x_0 \in \mathbb{H}^3$. The topology on $\partial_\infty(\mathbb{H}^3)$ is defined via the angles between corresponding rays: two geodesic rays are close if and only if the angle between the corresponding rays is small.

In the upper half-space model, we can identify $\partial_\infty(\mathbb{H}^3)$ with the extended complex plane $\{(x_1, x_2, 0)\} \cup \{\infty\} = \mathbb{C} \cup \{\infty\}$, which is homeomorphic to the sphere $\mathbb{S}^2$. The space $\mathbb{H}^3$ has the natural compactification $\mathbb{H}^3 \cup \partial_\infty(\mathbb{H}^3)$ (cf. [20, 3.2]).

Definition 2.6. For a Kleinian group Γ, the limit set $\Lambda(\Gamma)$ of Γ consists of limit points of an orbit Γz, $z \in \mathbb{H}^3$ in the ideal boundary $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. We denote by δ_Γ the Hausdorff dimension of $\Lambda(\Gamma)$.

By Sullivan, δ_Γ is equal to the critical exponent of Γ for geometrically finite Γ. In this subsection, we realize the action of $\mathcal{A}$ on Descartes quadruples arising from $\mathcal{P}$ as the action of a subgroup, $G_{\mathcal{A}}(\mathcal{P})$, of the Möbius transformations on $\hat{\mathbb{C}}$ in a way that the residual set of $\mathcal{P}$ coincides with the limit set of $G_{\mathcal{A}}(\mathcal{P})$. The residual set $\Lambda(\mathcal{P})$ is defined to be the closure of all the circles in $\mathcal{P}$, or equivalently, the complement in $\hat{\mathbb{C}}$ of the interiors of all circles in the packing $\mathcal{P}$ (where the circles are oriented so that the interiors are disjoint).

An oriented Descartes configuration is a Descartes configuration in which the orientations of the circles are compatible in the sense that either the interiors of all four oriented circles are disjoint or the interiors are disjoint when all the orientations are reversed. Given an ordered configuration $\mathcal{D}$ of four oriented circles (C_1, C_2, C_3, C_4) with curvatures (b_1, b_2, b_3, b_4) with centers $\{(x_i, y_i) : 1 \le i \le 4\}$, set

$$W_{\mathcal{D}} := \begin{pmatrix} \bar{b}_1 & b_1 & b_1 x_1 & b_1 y_1 \\ \bar{b}_2 & b_2 & b_2 x_2 & b_2 y_2 \\ \bar{b}_3 & b_3 & b_3 x_3 & b_3 y_3 \\ \bar{b}_4 & b_4 & b_4 x_4 & b_4 y_4 \end{pmatrix}$$

where $\bar{b}_i$ is the curvature of the circle which is the reflection of C_i through the unit circle centered at the origin, i.e., $\bar{b}_i = b_i(x_i^2 + y_i^2) - b_i^{-1}$ if $b_i \neq 0$. If one of the circles, say C_i, is a line, we interpret the center (x_i, y_i) as the outward unit normal vector and set $b_i = \bar{b}_i = 0$.

Then by [16, Thm. 3.2], for any ordered and oriented Descartes configuration $\mathcal{D}$, the map $g \mapsto W_{\mathcal{D}}^{-1} g W_{\mathcal{D}}$ gives an isomorphism

$$\psi_{\mathcal{D}} : \mathrm{O}_Q(\mathbb{R}) \to \mathrm{O}_{Q_W}(\mathbb{R})$$

where Q_W is the Wilker quadratic form:

$$Q_W = \begin{pmatrix} 0 & -4 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

On the other hand, if Möb(2) denotes the group of Möbius transformations and $\mathrm{GM}^*(2) := \text{Möb}(2) \times \{\pm I\}$ denotes the extended Möbius group, we have the following by Graham et al:

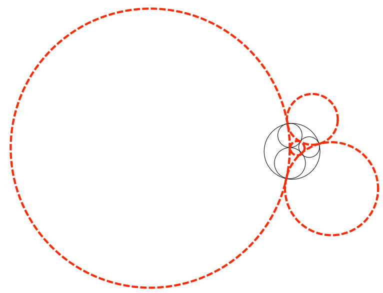

FIGURE 6. Action of $\mathcal{A}$ as Möbius transformations

Theorem 2.7. [16, Thm. 7.2] *There exists a unique isomorphism*

$$\pi : \mathrm{GM}^*(2) \to \mathrm{O}_{Q_W}(\mathbb{R})$$

such that for any ordered and oriented Descartes configuration $\mathcal{D}$,

(1) $W_{\gamma(\mathcal{D})} = W_{\mathcal{D}}\pi(\gamma)^{-1}$ *for any* $\gamma \in$ Möb(2).
(2) $W_{-\mathcal{D}} = -W_{\mathcal{D}}$.

Let $\mathcal{D}_0 = (C_1, C_2, C_3, C_4)$ denote the ordered and oriented Descartes configuration corresponding to the root quadruple ξ of $\mathcal{P}$. We obtain the following isomorphism:

$$\Phi_{\mathcal{D}_0} := \pi^{-1} \circ \psi_{\mathcal{D}_0} : \mathrm{O}_Q(\mathbb{R}) \to \mathrm{GM}^*(2).$$

Denote by $\mathfrak{s}_i := \mathfrak{s}_i(\mathcal{D}_0)$ the Möbius transformation given by the inversion in the circle, say, $\hat{C}_i$, determined by the three intersection points of the circles C_j, $j \neq i$. Figure 6 depicts the root quadruple $(C_1, \dots, C_4)$ as solid-lined circles and the corresponding dual quadruple $(\hat{C}_1, \dots, \hat{C}_4)$ as dotted-lined circles.

Note that $\mathfrak{s}_i$ fixes C_j, $j \neq i$ and moves C_i to the unique other circle that is tangent to C_j's, $j \neq i$.

We set

$$G_{\mathcal{A}}(\mathcal{P}) := \langle \mathfrak{s}_1, \mathfrak{s}_2, \mathfrak{s}_3, \mathfrak{s}_4 \rangle.$$

Lemma 2.8. *For each* $1 \le i \le 4$,

$$\Phi_{\mathcal{D}_0}(S_i) = \mathfrak{s}_i;$$

hence

$$\Phi_{\mathcal{D}_0}(\mathcal{A}) = G_{\mathcal{A}}(\mathcal{P}).$$

Proof. If $\gamma_i := \Phi_{\mathcal{D}_0}(S_i)$, then by (2.7),

$$W_{\gamma_i(\mathcal{D}_0)} = W_{\mathcal{D}_0}(\psi_{\mathcal{D}_0} S_i)^{-1} = S_i W_{\mathcal{D}_0}.$$

On the other hand, by [16, 3.25],

$$W_{\mathfrak{s}_i(\mathcal{D}_0)} = S_i W_{\mathcal{D}_0}.$$

Therefore

$$\gamma_i(\mathcal{D}_0) = \mathfrak{s}_i(\mathcal{D}_0);$$

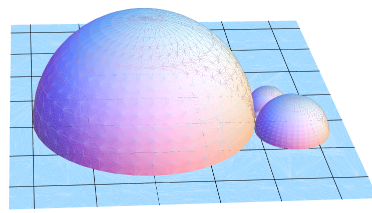

FIGURE 7. Fundamental domain

and hence $\gamma_i = \mathfrak{s}_i$. $\square$

Each inversion $\mathfrak{s}_i$ extends uniquely to an isometry of the hyperbolic space $\mathbb{H}^3$, corresponding to the inversion with respect to the hemisphere whose boundary is $\hat{C}_i$. The intersection of the exteriors of these hemispheres is a fundamental domain for the action of $G_{\mathcal{A}}(\mathcal{P})$ on $\mathbb{H}^3$.

The Hausdorff dimension of $\Lambda(\mathcal{P})$, say α, is independent of $\mathcal{P}$. Hirst showed in [18] that α is strictly between one and two. For our purpose, we only need to know that $\alpha > 1$, though much more precise estimates were made by Boyd in [6] and McMullen [29].

Proposition 2.9. (1) *$G_{\mathcal{A}}(\mathcal{P})$ is geometrically finite and discrete.*
(2) *We have $\Lambda(G_{\mathcal{A}}(\mathcal{P})) = \Lambda(\mathcal{P})$, and hence $\alpha = \delta_{G_{\mathcal{A}}(\mathcal{P})}$.*

Proof. For simplicity set $\Gamma := G_{\mathcal{A}}(\mathcal{P})$. The group Γ is geometrically finite (that is, it admits a finite sided Dirichlet domain) by [20, Thm.13.1] and discrete since $\mathcal{A}$ is discrete and $\Phi_{\mathcal{D}_0}$ is a topological isomorphism.

Clearly, Γ is non-elementary. It is well-known that $\Lambda(\Gamma)$ is same as the set of all accumulation points in the orbit x_0 under Γ for any (fixed) $x_0 \in \hat{\mathbb{C}}$. On the other hand, by [16, Thm 4.2], $\Lambda(\mathcal{P})$ is equal to the closure of all tangency points of circles in $\mathcal{P}$ and is invariant under Γ. This immediately yields

$$\Lambda(\Gamma) \subset \Lambda(\mathcal{P}).$$

If $x_0 \in \Lambda(\mathcal{P})$, then any neighborhood, say U, of x_0 contains infinitely many circles. Since $\cup_{1\le i\le 4}\Gamma(C_i)$ is the set of all circles in $\mathcal{P}$, there exist j and an infinite sequence $\gamma_i \in \Gamma$ such that $\gamma_i(C_j) \subset U$ for all i. Therefore $x_0 \in \Lambda(\Gamma)$. This proves that $\Lambda(\mathcal{P}) \subset \Lambda(\Gamma)$. $\square$

Set $\Gamma_{\mathcal{P}}$ to be the subgroup of holomorphic elements, that is,

$$\Gamma_{\mathcal{P}} := G_{\mathcal{A}}(\mathcal{P}) \cap \mathrm{PSL}_2(\mathbb{C}),$$

since Möb(2) is the semidirect product of complex conjugation with the subgroup $\text{Möb}_+(2) = \mathrm{PSL}_2(\mathbb{C})$ of orientation preserving transformations.

Then the above lemma holds with $\Gamma_{\mathcal{P}}$ in place of $G_{\mathcal{A}}(\mathcal{P})$ as both properties are inherited by a subgroup of finite index. By the well known Selberg's lemma, we can further replace $\Gamma_{\mathcal{P}}$ a torsion-free subgroup of finite index.

2.3. **Reduction to orbital counting for a Kleinian group.** Let $G := \mathrm{PSL}_2(\mathbb{C})$ and $\Gamma < G$ be a geometrically finite, torsion-free Kleinian subgroup. Suppose we are given a real linear representation $\iota : G \to \mathrm{SO}_F(\mathbb{R})$ where F is a real quadratic form in 4 variables with signature $(3, 1)$. As we prefer to work with a right action, we consider $\mathbb{R}^4$ as the set of row vectors and the action is given by $vg := \iota(g)v^t$ for $g \in G$ and $v \in \mathbb{R}^4$.

By Lemma 2.5 and the discussions in the previous subsection, Theorem 1.1 follows from:

Theorem 2.10. *Suppose $\delta_\Gamma > 1$. Let $v_0 \in \mathbb{R}^4$ be a non-zero vector lying in the cone $F = 0$ with a discrete orbit $v_0\Gamma \subset \mathbb{R}^4$. Then for any norm $\|\cdot\|$ on $\mathbb{R}^4$, there exists $c > 0$ such that*

$$\#\{v \in v_0\Gamma : \|v\| < T\} \sim c \cdot T^{\delta_\Gamma}.$$

It also follows from Lemma 2.5 that this theorem implies

$$N_2^{\mathcal{P}}(T) \sim (3c) \cdot T^{\alpha}$$

for the same c as in Theorem 1.1.

3. Geometry of closed horospheres on $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$

Let $G = \mathrm{PSL}_2(\mathbb{C})$ and $\mathbb{H}^3 = \{(x_1, x_2, y) : y > 0\}$. We use the notations N, A, K, M, n_x and a_y as defined in (1.6). Throughout this section, we suppose that $\Gamma < G$ is a torsion-free, non-elementary (i.e., $\Lambda(\Gamma)$ consists of more than 2 points) and geometrically finite Kleinian group.

Definition 3.1.
(1) A point $\xi \in \Lambda(\Gamma)$ is called a *parabolic fixed point* if ξ is fixed by a parabolic isometry of Γ (that is, an element of trace ± 2). The *rank* of a parabolic fixed point ξ is defined to be rank of the abelian group Γ_ξ which stabilizes ξ.
(2) A parabolic fixed point ξ is called a *bounded parabolic* if it is of rank 2 or if there exists a pair of two open disjoint discs in $\hat{\mathbb{C}}$ whose union, say U, is precisely Γ_ξ-invariant (that is, U is invariant by Γ_ξ and $\gamma(U) \cap U \neq \emptyset$ for all $\gamma \in \Gamma$ not belonging to Γ_ξ). Or equivalently, ξ is bounded parabolic if $\Gamma_\xi\backslash(\Lambda(\Gamma) - \{\xi\})$ is compact.
(3) A point $\xi \in \Lambda(\Gamma)$ is called a *point of approximation* if there exist a sequence $\gamma_i \in \Gamma$ and $z \in \mathbb{H}^3$ such that $\gamma_i z \to \xi$ and $\gamma_i z$ is within a bounded distance from the geodesic ray ending at ξ.

Since Γ is geometrically finite, it is known by the work of Beardon and Maskit [2] that any limit point $\xi \in \Lambda(\Gamma)$ is either a point of approximation or a bounded parabolic fixed point.

We denote by Vis the visual map from $\mathrm{T}^1(\mathbb{H}^3)$ to the ideal boundary $\partial_\infty(\mathbb{H}^3)$ which maps the vector $(p, \vec{v})$ to the end of the geodesic ray tangent to $\vec{v}$.

Definition 3.2.
- We denote by $\tilde{\Omega}_\Gamma \subset \mathrm{T}^1(\mathbb{H}^3)$ the set of vectors $(p, \vec{v})$ whose the image under the visual map belongs to $\Lambda(\Gamma)$.
- We denote by $\hat{\Omega}_\Gamma$ the image of $\tilde{\Omega}_\Gamma$ under the projection $\mathrm{T}^1(\mathbb{H}^3) \to \mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$.

We note that $\tilde{\Omega}_\Gamma$ is a closed set consisting of horospheres. For comparison, if Γ has finite co-volume, then $\Lambda_\Gamma = \mathbb{C} \cup \{\infty\}$ and $\hat{\Omega}_\Gamma = \mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$.

For $u \in \mathrm{PSL}_2(\mathbb{C})$, we denote by $[u]$ its image in $\mathrm{PSL}_2(\mathbb{C})/M$, which we may consider as a vector in $\mathrm{T}^1(\mathbb{H}^3)$. The horosphere determined by the vector $[u]$ in $\Gamma\backslash \mathrm{T}^1(\mathbb{H}^3)$ corresponds to $\Gamma\backslash\Gamma uNM/M$, which will be simply denoted by $\Gamma\backslash\Gamma[u]N$. We note that $u(\infty)$ is identified with $\mathrm{Vis}[u]$.

Theorem 3.3 (Dal'Bo)**.** [10] *Let $u \in \mathrm{PSL}_2(\mathbb{C})$.*

(1) *If $u(\infty) \in \Lambda(\Gamma)$ is a point of approximation, then $\Gamma\backslash\Gamma[u]N$ is dense in $\hat{\Omega}_\Gamma$.*
(2) *The orbit $\Gamma\backslash\Gamma uN$ is closed in $\Gamma\backslash G$ if and only if either $u(\infty) \notin \Lambda(\Gamma)$ or $u(\infty)$ is a bounded parabolic fixed point for Γ.*

The first claim was proved in [10, Prop. C and Cor. 1] under the condition that the length spectrum of Γ is not discrete in $\mathbb{R}$. As remarked there, this condition holds for non-elementary Kleinian groups by [17]. The second claim follows from [10, Prop. C and Cor. 1] together with the following lemma.

Lemma 3.4. (1) *We have $\Gamma \cap NM = \Gamma \cap N$.*
(2) *The orbit $\Gamma\backslash\Gamma NM$ is closed if and only if $\Gamma\backslash\Gamma N$ is closed.*

Proof. If $\gamma \in \Gamma \cap NM$, but not in N, then $\mathrm{Tr}^2(\gamma)$ is real and $0 \le \mathrm{Tr}^2(\gamma) < 4$ where $\mathrm{Tr}(\gamma)$ denotes the trace of γ. Hence γ is an elliptic element. As Γ is discrete, γ must be of finite order. However Γ is torsion-free, which forces $\gamma = e$. This proves the first claim. The second claim follows easily since the first claim implies the inclusion map $\Gamma\cap N\backslash N \to \Gamma\cap NM\backslash NM$ is proper. $\square$

It follows that if $\Gamma\backslash\Gamma N$ is closed in $\Gamma\backslash \mathrm{PSL}_2(\mathbb{C})$, then $\Gamma\backslash\Gamma N$ is either

(1) the embedding of a plane, if $\infty \notin \Lambda(\Gamma)$;
(2) the embedding of a cylinder, if ∞ is a bounded parabolic fixed point of rank one; or
(3) the embedding of a torus, if ∞ is a bounded parabolic fixed point of rank two.

For $X \subset \mathbb{H}^3 \cup \partial_\infty(\mathbb{H}^3)$, $\bar{X}$ denotes its closure in $\mathbb{H}^3 \cup \partial_\infty(\mathbb{H}^3)$.

Proposition 3.5. *Assume that $\Gamma\backslash\Gamma N$ is closed. There exists a finite-sided fundamental polyhedron $\mathcal{F} \subset \mathbb{H}^3$ for the action of Γ, and also a fundamental domain $\mathcal{F}_N \subset \mathbb{C}$ for the action of $N \cap \Gamma$ such that for some $r \gg 1$ and for some finite subset $I_\Gamma \subset \Gamma$,*

$$\{(x_1, x_2, y) \in \mathbb{H}^3 : x_1 + ix_2 \in \mathcal{F}_N,\ x_1^2 + x_2^2 + y^2 > r\} \subset \cup_{\gamma \in I_\Gamma} \gamma\mathcal{F}.$$

Proof. Choose a finite-sided fundamental polyhedron $\mathcal{F}$ for Γ with $\infty \in \bar{\mathcal{F}}$. If $\infty \notin \Lambda(\Gamma)$, then ∞ lies in the interior of $\cup_{\gamma \in I_\Gamma} \gamma\mathcal{F}$ for some finite $I_\Gamma \subset \Gamma$. As the exteriors of hemispheres form a basis of neighborhoods of ∞ in $\mathbb{H}^3 \cup \partial_\infty(\mathbb{H}^3)$, and $\mathcal{F}_N = \mathbb{C}$, the claim follows.

Now suppose that ∞ is a bounded parabolic fixed point of rank one. Let $n_{v_1} \in \Gamma \cap N$ be a generator for $\Gamma \cap N$ for $v_1 \in \mathbb{C}$, and fix a vector $v_2 \in \mathbb{C}$ perpendicular to v_1. The set $\mathcal{F}_N := \{s_1 v_1 + s_2 v_2 \in \mathbb{C} : 0 \le s_1 < 1\}$ is a fundamental domain in $\mathbb{C}$ for the action of $\Gamma \cap N$. By replacing $\mathcal{F}$ if necessary, we may assume that $\mathcal{F} \subset \mathcal{F}_N \times \mathbb{R}_{>0}$. Then by [26, Prop. A. 14 in VI], for all large $c > 1$, the sets

$$S(c) := \{(s_1 v_1 + s_2 v_2, y) \in \mathbb{H}^3 : y > c \text{ or } |s_2| > c\}$$

is precisely invariant by $\Gamma \cap N$ and $\mathcal{F} - S(c)$ is bounded away from ∞. Since $\mathcal{F} \cap S(c) \neq \emptyset$ as $\infty \in \bar{\mathcal{F}}$ and $S(c)$ is precisely invariant by $N \cap \Gamma$, it follows that

$$\{(x_1, x_2, y) \in S(c) : x_1 + ix_2 \in \mathcal{F}_N\} \subset \mathcal{F}. \tag{3.6}$$

Since

$$\{(x_1, x_2, y) : x_1 + ix_2 \in \mathcal{F}_N, x_1^2 + x_2^2 + y^2 > r\} \subset S(c)$$

for some large $c > 0$, the claim follows in this case.

If ∞ is a bounded parabolic of rank two, $S(c) := \{(x_1, x_2, y) : y > c\}$ is precisely invariant by $\Gamma \cap N$ and $\mathcal{F} - S(c)$ is bounded away from ∞ by [26, Prop. A. 13 in VI]. Choose $\mathcal{F}_N$ so that $\mathcal{F} \subset \mathcal{F}_N \times \mathbb{R}_{>0}$. Then (3.6) holds for the same reason, and since $\mathcal{F}_N$ is bounded,

$$\{(x_1, x_2, y) : x_1 + ix_2 \in \mathcal{F}_N, x_1^2 + x_2^2 + y^2 > r\} \subset S(c)$$

for some large $c > 0$. This proves the claim. $\square$

Considering the action of $N \cap \Gamma$ on $\partial_\infty(\mathbb{H}^3) \setminus \{\infty\} = \mathbb{C}$, we denote by $\Lambda_N(\Gamma)$ the image of $\Lambda(\Gamma) \setminus \{\infty\}$ in the quotient $(N \cap \Gamma)\backslash\mathbb{C}$, that is,

$$\Lambda_N(\Gamma) := \{[n_x] \in (N \cap \Gamma)\backslash N : x \in \Lambda(\Gamma) \setminus \{\infty\}\}.$$

Proposition 3.7. *If $\Gamma\backslash\Gamma N$ is closed, the set $\Lambda_N(\Gamma)$ is bounded.*

Proof. This is clear if ∞ is a bounded parabolic of rank two, as $N \cap \Gamma\backslash N$ is compact. If $\infty \notin \Lambda(\Gamma)$, then $\Lambda(\Gamma)$ is a compact subset of $\mathbb{C}$, and hence the claim follows. In the case when ∞ is a bounded parabolic of rank one, the claim follows from the well-known fact that $\Lambda(\Gamma)$ lies in a strip of finite width (see [47, Pf. of Prop. 8.4.3]). This can also be deduced from Proposition 3.5 using the fact that the intersection of the convex core of $\Gamma\backslash\mathbb{H}^3$ and the thick part of the manifold $\Gamma\backslash\mathbb{H}^3$ is compact for a geometrically finite group. $\square$

Proposition 3.8. *Assume that $\Gamma\backslash\Gamma N$ is closed. For any compact subset $J \subset \Gamma\backslash G$, the following set is bounded:*

$$N(J) := \{[n] \in (N \cap \Gamma)\backslash N \ : \Gamma\backslash\Gamma n A \cap J \neq \emptyset\}. \tag{3.9}$$

Proof. Let $\mathcal{F}_N$ and $\mathcal{F}$ be as in Proposition 3.5. If $\mathcal{F}_N$ is bounded, there is nothing to prove. Hence by Theorem 3.3, we may assume that either $\infty \notin \Lambda(\Gamma)$ or ∞ is a bounded parabolic fixed point of rank one.

To prove the proposition, suppose on the contrary that there exist sequences $n_j \in \mathcal{F}_N \to \infty$, $a_j \in A$, $\gamma_j \in \Gamma$ and $w_j \in J$ such that $n_j a_j = \gamma_j w_j$. As J is bounded, we may assume $w_j \to w \in G$ by passing to a subsequence. Then $\gamma_j^{-1} n_j a_j \to w$. Let $\gamma_0 \in \Gamma$ be such that $\gamma_0 w(0,0,1) \in \bar{\mathcal{F}}$. By the geometric finiteness of Γ, there exists a finite union, say $\mathcal{F}'$, of translates of $\mathcal{F}$ such that $\gamma_0 w_j(0,0,1) \in \mathcal{F}'$ for all large j.

As $n_j \to \infty$, the Euclidean norm of $n_j a_j(0,0,1)$ goes to infinity, and hence by Proposition 3.5,

$$n_j a_j(0,0,1) \in \mathcal{F}' \quad \text{for all large } j,$$

by enlarging $\mathcal{F}'$ if necessary.

Therefore for all large j,

$$\gamma_0 \gamma_j^{-1} n_j a_j(0,0,1) = \gamma_0 w_j(0,0,1) \in \mathcal{F}' \cap \gamma_0 \gamma_j^{-1}(\mathcal{F}').$$

Since $\mathcal{F}' \cap \gamma_0 \gamma_j^{-1}(\mathcal{F}') \neq \emptyset$ for only finitely many γ_j's, we conclude that $\{\gamma_j\}$ must be a finite set. As $n_j a_j = \gamma_j w_j \in \gamma_j J$, $n_j a_j$ must be a bounded sequence, contradicting $n_j \to \infty$. □

Note that $N(J)$ is defined so that for all $y > 0$,

$$(N \cap \Gamma)\backslash N a_y \cap J \subset N(J).$$

Corollary 3.10. *Let $\psi \in C_c(\Gamma\backslash G)$ have support J.*

1. *The set $N(J)$ defined in* (3.9) *is bounded.*
2. *For any function $\eta \in C_c(N \cap \Gamma\backslash N)$ with $\eta|_{N(J)} \equiv 1$, we have for all $y > 0$,*

$$\int_{(N\cap\Gamma)\backslash N} \psi(na_y) dn = \int_{(N\cap\Gamma)\backslash N} \psi(na_y)\, \eta(n)\, dn.$$

Proof. The first claim is immediate from the above proposition. For (2), it suffices to note that $\psi(na_y) \equiv 0$ for n outside of $N(J)$, and hence

$$\int_{(N\cap\Gamma)\backslash N} \psi(na_y)\, \eta(n)\, dn = \int_{N(J)} \psi(na_y)\, \eta(n)\, dn.$$

Using $\eta \equiv 1$ on $N(J)$, the claim follows. □

4. The base eigenfunction ϕ_0

In this section, we assume that $\Gamma < G = \mathrm{PSL}_2(\mathbb{C})$ is a geometrically finite torsion-free discrete subgroup and that the Hausdorff dimension δ_Γ of the limit set $\Lambda(\Gamma)$ is strictly bigger than one. Assume also that $\Gamma\backslash\Gamma N$ is closed.

By Sullivan [45], there exists a positive L^2-eigenfunction ϕ_0, unique up to a scalar multiple, of the Laplace operator Δ on $\Gamma\backslash\mathbb{H}^3$ with smallest eigenvalue

$\delta_\Gamma(2-\delta_\Gamma)$ and which is square-integrable, that is,

$$\int_{\Gamma\backslash\mathbb{H}^3} \phi_0(x,y)^2 \frac{1}{y^3} dxdy < \infty.$$

In this section, we study the properties of ϕ_0 along closed horospheres. For the base point $o=(0,0,1)$, we denote by ν_o a weak limit as $s\to\delta_\Gamma$ of the family of measures on $\overline{\mathbb{H}}^3$:

$$\nu(s) := \frac{1}{\sum_{\gamma\in\Gamma} e^{-sd(o,\gamma o)}} \sum_{\gamma\in\Gamma} e^{-sd(o,\gamma o)} \delta_{\gamma o}.$$

The measure ν_o is indeed the unique weak limit of $\{\nu(s)\}$ as $s\to\delta_\Gamma$. Sullivan [45] showed that ν_o is the unique finite measure (up to a constant multiple) supported on $\Lambda(\Gamma)$ with the property: for any $\gamma\in\Gamma$,

$$\frac{d(\gamma_*\nu_o)}{d\nu_o}(\xi) = e^{-\delta_\Gamma \beta_\xi(\gamma(o),o)} \tag{4.1}$$

where $\beta_\xi(z_1,z_2) := \lim_{z\to\xi} d(z_1,z) - d(z_2,z)$ is the Busemann function and $\gamma_*\nu_o(A) := \nu_o(\gamma^{-1}(A))$. The measure ν_o, called the Patterson-Sullivan measure, has no atoms.

The base eigenfunction ϕ_0 can be explicitly written as the integral of the Poisson kernel against ν_o:

$$\phi_0(x,y) = \int_{u\in\Lambda(\Gamma)} e^{-\delta_\Gamma \beta_u((x,y),o)} d\nu_o(u) \tag{4.2}$$

$$= \int_{u\in\Lambda(\Gamma)\setminus\{\infty\}} \left(\frac{(\|u\|^2+1)y}{\|x-u\|^2+y^2}\right)^{\delta_\Gamma} d\nu_o(u), \tag{4.3}$$

where $\Lambda(\Gamma)$ is identified as a subset of $\mathbb{C}\cup\{\infty\}$. The main goal of this section is to study the average of ϕ_0 along the translates $\Gamma\backslash\Gamma N a_y$.

Definition 4.4. For a given $\psi\in C(\Gamma\backslash G)^K$, define

$$\psi^N(a_y) := \int_{n\in(N\cap\Gamma)\backslash N} \psi(na_y)\, dn.$$

We will show that the integral $\phi_0^N(a_y)$ above converges, and moreover that all of the action in this integral takes place only over the following bounded set (see Proposition 3.7):

$$\Lambda_N(\Gamma) = \{[n_x]\in(N\cap\Gamma)\backslash N : x\in\Lambda(\Gamma)\setminus\{\infty\}\}.$$

If ∞ is a bounded parabolic fixed point of rank one, then there exists a vector $v_1\in\mathbb{R}^2$ such that n_{v_1} is a generator for $N\cap\Gamma$. Fixing vector v_2 perpendicular to v_1, we can decompose $x\in\mathbb{R}^2$ as $x=x_1v_1+x_2v_2$.

Definition 4.5. We say that an open subset $B\subset(N\cap\Gamma)\backslash N$ *properly covers* $\Lambda_N(\Gamma)$ if the following holds:

(1) if $\infty\notin\Lambda(\Gamma)$, then $\epsilon_0(B) := \inf_{u\in\Lambda_N(\Gamma), x\notin B} \|x-u\| > 0$

(2) if ∞ is a bounded parabolic of rank one, then

$$\epsilon_0(B) := \inf_{x \notin B, u \in \Lambda_N(\Gamma)} |u_2 - x_2| > 0.$$

(3) if ∞ is a bounded parabolic of rank two, then $B = (N \cap \Gamma)\backslash N$.

Proposition 4.6. (1) *$\phi_0^N(a_y) \gg y^{2-\delta_\Gamma}$ for all $0 < y \ll 1$.*

(2) *For any open subset $B \subset (N \cap \Gamma)\backslash N$ which properly covers $\Lambda_N(\Gamma)$, and all small $y > 0$,*

$$\phi_0^N(a_y) = \int_B \phi_0(n_x a_y)\, dx + O_{\epsilon_0(B)}(y^{\delta_\Gamma}). \tag{4.7}$$

Proof. Choose a fundamental domain $\mathcal{F}_N \subset \mathbb{C}$ for $(N \cap \Gamma)\backslash N$. We first show that $\phi_0^N(a_y) \gg y^{2-\delta_\Gamma}$ for all $0 < y \ll 1$. For a subset $Q \subset \mathcal{F}_N$, the notation Q^c denotes the complement of Q in $\mathcal{F}_N$. Observe that for a subset $Q \subset \mathcal{F}_N$,

$$\begin{aligned} &\int_{x \in Q} \phi_0(n_x a_y)\, dx \\ &= \int_{u \in \Lambda(\Gamma)} (\|u\|^2 + 1)^{\delta_\Gamma} \int_{x \in Q} \left(\frac{y}{\|x - u\|^2 + y^2} \right)^{\delta_\Gamma} dx d\nu_o(u) \\ &= \int_{u \in \Lambda(\Gamma)} \left(\int_{x \in Q - u} (\|u\|^2 + 1)^{\delta_\Gamma} \left(\frac{y}{\|x\|^2 + \|y\|^2} \right)^{\delta_\Gamma} dx \right) d\nu_o(u) \end{aligned} \tag{4.8}$$

where the interchange of orders is justified since everything is nonnegative.

We observe that by changing the variable $w = \frac{x}{y}$,

$$\begin{aligned} \int_{x \in \mathbb{R}^2} \left(\frac{y}{\|x\|^2 + \|y\|^2} \right)^{\delta_\Gamma} dx &= y^{2-\delta_\Gamma} \int_{w \in \mathbb{R}^2} \left(\frac{1}{\|w\|^2 + 1} \right)^{\delta_\Gamma} dw \\ &= y^{2-\delta_\Gamma} 2\pi \int_{r > 0} \frac{r}{(r^2 + 1)^{\delta_\Gamma}}\, dr \\ &= \frac{\pi}{\delta_\Gamma - 1} y^{2-\delta_\Gamma}. \end{aligned} \tag{4.9}$$

Case I: $\infty \notin \Lambda(\Gamma)$. In this case, we have

$$\omega_0 = \int_{u \in \Lambda(\Gamma)} (\|u\|^2 + 1)^{\delta_\Gamma} d\nu_o(u) \ll \nu_o(\Lambda(\Gamma)) < \infty.$$

Therefore for $Q = \mathbb{C}$, we obtain from (4.8) and (4.9) that

$$\begin{aligned} \phi_0^N(a_y) &= \int_{u \in \Lambda(\Gamma)} (\|u\|^2 + 1)^{\delta_\Gamma} d\nu_o(u) \cdot \int_{x \in \mathbb{R}^2} \left(\frac{y}{\|x\|^2 + \|y\|^2} \right)^{\delta_\Gamma} dx \\ &= \frac{\omega_0 \cdot \pi}{\delta_\Gamma - 1} y^{2-\delta_\Gamma}. \end{aligned} \tag{4.10}$$

This proves (1). To prove (2), let B be an open subset of $\mathbb{C}$ which properly covers $\Lambda_N(\Gamma)$. Then we have

$$\epsilon_0 := \inf_{u \in \Lambda_N(\Gamma), x \in B^c} \|x - u\| = \inf\{\|x\| : x \in B^c - \Lambda_N(\Gamma)\} > 0.$$

By setting $Q = B^c$ in (4.8), we deduce

$$
\begin{aligned}
\int_{x\in B^c} \phi_0(n_x a_y)\, dx &\leq \omega_0 \cdot \int_{\|x\|>\epsilon_0} \left(\frac{y}{\|x\|^2+\|y\|^2}\right)^{\delta_\Gamma} dx \\
&\leq \omega_0 y^{2-\delta_\Gamma} \cdot \int_{\|w\|>\epsilon_0 y^{-1}} \left(\frac{1}{\|w\|^2+1}\right)^{\delta_\Gamma} dw \\
&= 2\pi\omega_0 y^{2-\delta_\Gamma} \cdot \int_{r>\epsilon_0 y^{-1}} \frac{r}{(r^2+1)^{\delta_\Gamma}}\, dr \\
&= 2\pi\omega_0 y^{2-\delta_\Gamma} \left(\frac{\epsilon_0^2}{y^2}+1\right)^{1-\delta_\Gamma} \\
&\ll y^{\delta_\Gamma}.
\end{aligned}
$$

Case II: ∞ is a bounded parabolic fixed point of rank one.

We may assume without loss of generality that $\Gamma \cap N$ is generated by $(1,0)$, that is, $x \mapsto x+1$, and $\mathcal{F}_N = \{(x_1,x_2) : 0 \leq x_1 < 1, x_2 \in \mathbb{R}\}$. There exists $T > 0$ such that $\Lambda(\Gamma) \subset \mathbb{R} \times [-T,T]$ by [47, Pf. of Prop. 8.4.3].

By computing the Busemann function, we deduce from (4.1) for $k = (k_1,k_2) \in \Gamma \cap N$ that

$$
d((n_{-k})_*\nu_o)(u) = \left(\frac{\|u\|^2+1}{\|k-u\|^2+1}\right)^{\delta_\Gamma} d\nu_o(u). \tag{4.11}
$$

We have by changing orders of integrations of x_1 and u_1 that

$$
\begin{aligned}
&\int_{x\in\mathcal{F}_N} \phi_0(x,y)dx \\
&= \int_{u\in\mathbb{R}\times[-T,T]} \int_{x\in([0,1]\times\mathbb{R})-u} \left(\frac{(\|u\|^2+1)y}{\|x\|^2+y^2}\right)^{\delta_\Gamma} dx d\nu_o(u) \\
&= \int_{x\in\mathbb{R}^2} \left(\frac{y}{\|x\|^2+y^2}\right)^{\delta_\Gamma} \int_{u\in[-x_1,1-x_1]\times[-T,T]} (\|u\|^2+1)^{\delta_\Gamma} d\nu_o(u)dx.
\end{aligned} \tag{4.12}
$$

We set c_1 and c_2 to be, respectively, the infimum and the supremum of

$$
e_{x_1} := \int_{u\in[-x_1,1-x_1]\times[-T,T]} (\|u\|^2+1)^{\delta_\Gamma} d\nu_o(u)
$$

over all $x_1 \in \mathbb{R}$. We claim that

$$
0 < c_1 \leq c_2 < \infty.
$$

For any $k_1 \in \mathbb{Z}$, by changing the variable $u_1 \mapsto u_1 + k_1$ and recalling (4.11), we deduce

$$\begin{aligned} e_{x_1} &= \int_{u \in [-x_1, 1-x_1] \times [-T,T]} (\|u\|^2+1)^{\delta_\Gamma} d\nu_o(u) \\ &= \int_{u_1=k_1-x_1}^{u_1=k_1+1-x_1} \int_{u_2=-T}^{u_2=T} (\|u-k\|^2+1)^{\delta_\Gamma} d((n_{-k})_* \nu_o)(u) \\ &= \int_{u_1=k_1-x_1}^{u_1=k_1+1-x_1} \int_{u_2=-T}^{u_2=T} (\|u\|^2+1)^{\delta_\Gamma} d\nu_o(u). \end{aligned}$$

Choosing k_1 so that $x_1 \in [k_1-1, k_1)$, we have $e_{x_1} \ll \nu_o([0,2] \times [-T,T])$ and hence $c_2 < \infty$. Note that $e_{x_1} \asymp e_{x_1+1}$ where the implied constants is independent of $x_1 \in \mathbb{R}$, and that

$$e_{x_1} + e_{x_1+1} = \int_{u_1=k_1-x_1-1}^{u_1=k_1+1-x_1} \int_{u_2=-T}^{u_2=T} (\|u\|^2+1)^{\delta_\Gamma} d\nu_o(u) \gg \nu_o([0,1] \times [-T,T]).$$

On the other hand, since the $N \cap \Gamma$-translates of $[0,1] \times [-T,T]$ cover the support of ν_o except for ∞ and ν_o is atom-free, we have $\nu_o([0,1] \times [-T,T]) > 0$. This proves $c_1 = \inf e_{x_1} > 0$.

We now deduce from (4.12) and (4.9) that

$$\int_{x \in \mathcal{F}_N} \phi_0(x,y) dx \gg c_1 \int_{x \in \mathbb{R}^2} \left(\frac{y}{\|x\|^2 + y^2} \right)^{\delta_\Gamma} dx \gg y^{2-\delta_\Gamma},$$

proving (1).

If B is an open subset of $\mathcal{F}_N = [0,1] \times \mathbb{R}$ which properly covers $\Lambda_N(\Gamma)$, we have

$$\epsilon_0 := \inf_{x \in B^c, u \in \Lambda(\Gamma)} |u_2 - x_2| > 0.$$

Hence $\{(x,u) \in \mathbb{R}^2 \times \Lambda_N(\Gamma) : x \in B^c - u\}$ is contained in the set

$$\{u_1 \in \mathbb{R}, -u_1 < x_1 < 1 - u_1, |u_2| < T, |x_2| > \epsilon_0\}.$$

Since

$$\{u_1 \in \mathbb{R}, \quad -u_1 < x_1 < 1 - u_1\} = \{x_1 \in \mathbb{R}, \quad -x_1 < u_1 < 1 - x_1\},$$

we deduce from (4.8) by changing the order of integrations that

$$\begin{aligned} &\int_{x \in B^c} \phi_0(x,y) dx \\ &= \int_{u \in \Lambda(\Gamma)} \int_{x \in B^c - u} \left(\frac{(\|u\|^2+1) y}{\|x\|^2 + y^2} \right)^{\delta_\Gamma} dx d\nu_o(u) \\ &\le \int_{x_1 \in \mathbb{R}, |x_2| > \epsilon_0} \left(\frac{y}{\|x\|^2 + y^2} \right)^{\delta_\Gamma} \left(\int_{u_1=-x_1}^{u_1=1-x_1} \int_{u_2=-T}^{u_2=T} (\|u\|^2+1)^{\delta_\Gamma} d\nu_o(u) \right) dx_2 dx_1 \\ &\le c_2 \cdot \int_{x_1 \in \mathbb{R}, |x_2| > \epsilon_0} \left(\frac{y}{\|x\|^2 + y^2} \right)^{\delta_\Gamma} dx_2 dx_1. \end{aligned}$$

The x_1 integral can be evaluated explicitly and yields:

$$\begin{aligned}\int_{x\in B^c}\phi_0(x,y)dx &\ll \frac{\sqrt{\pi}\cdot\Gamma(\delta_\Gamma-1/2)}{\Gamma(\delta_\Gamma)}y^{\delta_\Gamma}\int_{|x_2|>\epsilon_0}\left(\frac{1}{x_2^2+y^2}\right)^{\delta_\Gamma-1/2}dx_2\\ &\ll \frac{\sqrt{\pi}\cdot\Gamma(\delta_\Gamma-1/2)}{\Gamma(\delta_\Gamma)}y^{\delta_\Gamma}\int_{|x_2|>\epsilon_0}\left(\frac{1}{x_2^2}\right)^{\delta_\Gamma-1/2}dx_2\\ &\ll y^{\delta_\Gamma}.\end{aligned}$$

Hence (2) is proved.

Case III: ∞ is a bounded parabolic fixed point of rank two. In this case (2) holds for a trivial reason. But we still need to show (1). The argument is similar to the case II. Without loss of generality, we assume that $N\cap\Gamma$ is generated by $(1,0)$ and $(0,1)$, so that $\mathcal{F}_N=[0,1]\times[0,1]$.

Similarly to (4.12) we have

$$\phi_0^N(a_y)=\int_{x\in\mathbb{R}^2}\left(\frac{y}{\|x\|^2+y^2}\right)^{\delta_\Gamma}\int_{u\in[-x_1,1-x_1]\times[-x_2,1-x_2]}(\|u\|^2+1)^{\delta_\Gamma}d\nu_o(u)dx$$

By (4.9), it suffices to show that

$$c_0:=\inf_{x_1,x_2\in\mathbb{R}}\int_{u\in[-x_1,1-x_1]\times[-x_2,1-x_2]}(\|u\|^2+1)^{\delta_\Gamma}d\nu_o(u)>0.$$

For $(x_1,x_2)\in[k_1-1,k_1)\times[k_2-1,k_2)$, by changing the variables $u_i\to u_i+k_i$ for $k_i\in\mathbb{Z}$ and by using (4.11), we deduce

$$\begin{aligned}&\int_{u_1=-x_1}^{u_1=1-x_1}\int_{u_2=-x_2}^{u_2=1-x_2}(\|u\|^2+1)^{\delta_\Gamma}d\nu_o(u)\\ &\qquad=\int_{u_1=k_1-x_1}^{u_1=k_1+1-x_1}\int_{u_2=k_2-x_2}^{u_2=k_2+1-x_2}(\|u\|^2+1)^{\delta_\Gamma}d\nu_o(u).\end{aligned}$$

Hence

$$c_0\gg\inf_{k_i-1\le x_i\le k_i}\nu_o([k_i-x_i,k_i-x_i+1-x_i])\gg\nu_o([0,1]^2)>0.$$

$\square$

Corollary 4.13. *For any $y>0$, there exist $c_{\phi_0}>0$ and $d_{\phi_0}\ge 0$ such that*

$$\phi_0^N(a_y)=c_{\phi_0}y^{2-\delta_\Gamma}+d_{\phi_0}y^{\delta_\Gamma}.$$

If $\infty\notin\Lambda(\Gamma)$, then $d_{\phi_0}=0$.

Proof. Since $\Delta\phi_0=\delta_\Gamma(2-\delta_\Gamma)\phi_0$, it follows that

$$-y^2\frac{\partial^2}{\partial y^2}\phi_0^N+y\frac{\partial}{\partial y}\phi_0^N=\delta_\Gamma(2-\delta_\Gamma)\phi_0^N.$$

As both y^{δ_Γ} and $y^{2-\delta_\Gamma}$ satisfy the above differential equation, we have

$$\phi_0^N(a_y)=c_{\phi_0}y^{2-\delta_\Gamma}+d_{\phi_0}y^{\delta_\Gamma}$$

for some $c_{\phi_0},d_{\phi_0}\ge 0$. Proposition 4.6 (1) implies that $c_{\phi_0}>0$.

In the case when $\mathcal{F}_N = \mathbb{C}$, the last claim is proved in (4.10). $\square$

5. Spherical functions and spectral bounds

We keep the notations set up in section 3. Let $0 \le s \le 2$, and consider the character χ_s on the subgroup $B := AMN$ of G defined by

$$\chi_s(a_y mn) = y^s$$

where $a_y \in A$ is given as before, and $m \in M$ and $n \in N$.

The unitarily induced representation $(\pi_s := \operatorname{Ind}_B^G \chi_s, V_s)$ admits a unique K-invariant unit vector, say v_s.

By the theory of spherical functions,

$$f_s(g) := \langle \pi_s(g) v_s, v_s \rangle = \int_K v_s(kg)dk$$

is the unique bi K-invariant function of G with $f_s(e) = 1$ and with $\mathcal{C} f_s = s(2-s) f_s$ where $\mathcal{C}$ is the Casimir operator of G. Moreover, there exist $c_s > 0$ and $\epsilon > 0$ such that for all y small

$$f_s(a_y) = c_s \cdot y^{2-s}(1 + O(y^\epsilon)) \tag{5.1}$$

by [14, 4.6].

Since the Casimir operator is equal to the Laplace operator Δ on K-invariant functions, this implies:

Theorem 5.2. *Let $\Gamma < G$ be a discrete subgroup. Let $\phi_s \in L^2(\Gamma\backslash G)^K \cap C^\infty(\Gamma\backslash G)$ satisfy $\Delta\phi_s = s(2-s)\phi_s$ and $\|\phi_s\|_2 = 1$. Then there exist $c_s > 0$ and $\epsilon > 0$ such that for all small $0 < y < 1$,*

$$\langle a_y \phi_s, \phi_s \rangle_{L^2(\Gamma\backslash G)} = c_s \cdot y^{2-s}(1 + O(y^\epsilon)).$$

In the unitary dual of G, the spherical part consists of the principal series and the complimentary series. We use the parametrization of $s \in \{1 + i\mathbb{R}, [1, 2]\}$ so that $s = 2$ corresponds to the trivial representation and the vertical line $1 + i\mathbb{R}$ corresponds to the tempered spectrum. Then the complimentary series is parametrized by V_s, $1 < s \le 2$ defined before.

Let $\{X_i\}$ denote an orthonormal basis of the Lie algebra of K with respect to an Ad-invariant scalar product, and define $\omega := 1 - \sum X_i^2$. This is a differential operator in the center of the enveloping algebra of $\operatorname{Lie}(K)$ and acts as a scalar on each K- isotypic component of V_s.

Proposition 5.3. *Fix $1 < s_0 < 2$. Let (V, π) be a representation of G which does not weakly contain any complementary series representation V_s with parameter $s \ge s_0$. Then for any $\epsilon > 0$, there exists $c_\epsilon > 0$ such that for any smooth vectors $w_1, w_2 \in V$, and $y < 1$,*

$$|\langle a_y w_1, w_2 \rangle| \le c_\epsilon \cdot y^{2-s_0-\epsilon} \cdot \|\omega(w_1)\| \cdot \|\omega(w_2)\|.$$

Proof. (We refer to [42] for the arguments below) As a G-representation, π has a Hilbert integral decomposition $\pi = \int_{z\in\hat{G}} \oplus^{m_z} \rho_z d\nu(z)$ where $\hat{G}$ denotes the unitary dual of G, ρ_z is irreducible and m_z is the multiplicity of ρ_z, and ν is the spectral measure on $\hat{G}$. By the assumption on π, for almost all z, ρ_z is either tempered or isomorphic to π_s for $1 \le s \le s_0$. As $1 < 3-s_0 < 2$, there exists the complementary series (V_{3-s_0}, π_{3-s_0}). We claim that the tensor product $\rho_z \otimes \pi_{3-s_0}$ is a tempered representation. Recall that a unitary representation of G is tempered if and only if there exists a dense subset of vectors whose matrix coefficients are $L^{2+\epsilon}$-integrable for any $\epsilon > 0$. If ρ_z is tempered, it is clearly tempered. If ρ_z is isomorphic to π_s for some $1 \le s \le s_0$, and v_z denotes the spherical vector of ρ_z of norm one, then the matrix coefficient $g \mapsto \langle \rho_z(g)v_z, v_z\rangle$ is $L^{2/(2-s_0)+\epsilon}$-integrable for any $\epsilon > 0$, by (5.1) together with the fact that the Haar measure on G satisfies $d(k_1 a_y k_2) \asymp y^{-3} dk_1 dy dk_2$ for all $0 < y \le 1$. Since ρ_z is irreducible, it follows that there exists a dense set of vectors whose matrix coefficients are $L^{2/(2-s_0)+\epsilon}$-integrable for any $\epsilon > 0$. Similarly there exists a dense set of vectors in V_{3-s_0} whose matrix coefficients are $L^{2/(s_0-1)+\epsilon}$-integrable for any $\epsilon > 0$. Hence by the Hölder inequality, there exists a dense set of vectors in $\rho_z \otimes \pi_{3-s_0}$ whose matrix coefficients are $L^{2+\epsilon}$-integrable for any $\epsilon > 0$, implying that $\rho_z \otimes \pi_{3-s_0}$ is tempered.

Since $\pi \otimes \pi_{3-s_0} = \int_z \oplus^{m_z} (\rho_z \otimes \pi_{3-s_0}) d\nu(z)$, we deduce that $\pi \otimes \pi_{3-s_0}$ is tempered.

We now claim that for any $\epsilon > 0$, there is a constant $c_\epsilon > 0$ such that any K-finite unit vectors w_1 and w_2, we have

$$|\langle a_y w_1, w_2\rangle| \le c_\epsilon \cdot y^{2-s_0-\epsilon} \cdot \prod_i \sqrt{(\dim\langle K w_i\rangle)}. \tag{5.4}$$

Noting that the K-span of $w_i \otimes v_{3-s_0}$ has the same dimension as the K-span of w_i, the temperedness of $\pi \otimes \pi_{3-s_0}$ implies that for any $\epsilon > 0$, there exists a constant $c_\epsilon > 0$ such that

$$\begin{aligned}\langle a_y(w_1 \otimes v_{3-s_0}), (w_2 \otimes v_{3-s_0})\rangle &= \langle a_y w_1, w_2\rangle \cdot \langle a_y v_{3-s_0}, v_{3-s_0}\rangle \\ &\le c_\epsilon \cdot y^{1+\epsilon} \cdot \prod_i \sqrt{(\dim\langle K w_i\rangle)}.\end{aligned}$$

As $\langle a_y v_{3-s_0}, v_{3-s_0}\rangle = c \cdot y^{-1+s_0}(1 + O(y_0^{\epsilon}))$ for some $\epsilon_0 > 0$, the claim (5.4) follows. Passing from the above bounds of (6.6) for K-finite vectors to those for smooth vectors has been detailed in [28, Pf. of Thm 6]. In particular, in the case of $G = \mathrm{SL}_2(\mathbb{C})$, the above degree of Sobolev norm suffices. □

Definition 5.5. For a geometrically finite discrete subgroup Γ of G with $\delta_\Gamma > 1$, we fix $1 < s_\Gamma < \delta_\Gamma$ so that there is no eigenvalue of Δ between $s_\Gamma(2 - s_\Gamma)$ and the base eigenvalue $\lambda_0 = \delta_\Gamma(2 - \delta_\Gamma)$ in $L^2(\Gamma\backslash G)$.

By the theorem of Lax-Phillips [24], the Laplace spectrum on $L^2(\Gamma\backslash G)^K$ has only finitely many eigenvalues outside the tempered spectrum. Therefore

$1 < s_\Gamma < \delta_\Gamma$ exists. The maximum difference between δ_Γ and s_Γ will be referred to as the spectral gap for Γ.

Let $\{Z_1, \cdots, Z_6\}$ denote an orthonormal basis of the Lie algebra of G. Let $\Gamma < G$ be a discrete subgroup of G. For $f \in C^\infty(\Gamma\backslash G) \cap L^2(\Gamma\backslash G)$, we consider the following Sobolev norm $\mathcal{S}_m(f)$:

$$\mathcal{S}_m(f) = \max\{\|Z_{i_1} \cdots Z_{i_m}(f)\|_2 : 1 \leq i_j \leq 6\}.$$

Corollary 5.6. *Let Γ be a geometrically finite discrete subgroup of G with $\delta_\Gamma > 1$. Then for any $\psi_1 \in L^2(\Gamma\backslash G) \cap C^\infty(\Gamma\backslash G)^K$, $\psi_2 \in C_c^\infty(\Gamma\backslash G)$ and $0 < y < 1$,*

$$\langle a_y\psi_1, \psi_2\rangle = \langle \psi_1, \phi_0\rangle\langle a_y\phi_0, \psi_2\rangle + O(y^{2-s_\Gamma}\mathcal{S}_2(\psi_1) \cdot \mathcal{S}_2(\psi_2)).$$

Here $\phi_0 \in L^2(\Gamma\backslash G)^K$ is the unique eigenfunction of Δ with eigenvalue $\delta_\Gamma(2-\delta_\Gamma)$ with unit L^2-norm.

Proof. We have

$$L^2(\Gamma\backslash G) = W_{\delta_\Gamma} \oplus V$$

where W_{δ_Γ} is isomorphic to V_{δ_Γ} as a G-representation and V does not contain any complementary series V_s with parameter $s > s_\Gamma$. Write $\psi_1 = \langle\psi_1, \phi_0\rangle\phi_0 + \psi_1^\perp$. Since ϕ_0 is the unique K-invariant vector in W_{δ_Γ} up to a constant multiple, we have $\psi_1^\perp \in V^K$. Hence by Proposition 5.3, for any $\epsilon > 0$ and $y \leq 1$,

$$\begin{aligned}\langle a_y\psi_1, \psi_2\rangle &= \langle \psi_1, \phi_0\rangle\langle a_y\phi_0, \psi_2\rangle + \langle a_y\psi_1^\perp, \psi_2\rangle \\ &= \langle \psi_1, \phi_0\rangle\langle a_y\phi_0, \psi_2\rangle + O(y^{2-s_\Gamma}\mathcal{S}_2(\psi_2) \cdot \mathcal{S}_2(\psi_2))\end{aligned}$$

since $\mathcal{S}_2(\psi_i^\perp) \ll \mathcal{S}(\psi_i)$. $\square$

6. Equidistribution of expanding closed horospheres with respect to the Burger-Roblin measure

Let $\Gamma < \mathrm{PSL}_2(\mathbb{C})$ be a geometrically finite Kleinian group with $\delta_\Gamma > 1$. Assume that $\Gamma\backslash\Gamma N$ is closed. Let μ denote the Haar measure on $G = NAK$ given by

$$d\mu(n_x a_y k) = y^{-3} dx dy dk$$

where dk is the probability Haar measure on K. We normalize ϕ_0 so that

$$\int_{\Gamma\backslash\mathbb{H}^3} \phi_0(x,y)^2 \frac{1}{y^3} dx dy = 1.$$

By Corollary 4.13, we have

$$\int_{n_x \in (N\cap\Gamma)\backslash N} \phi_0(x,y)\, dx = c_{\phi_0} y^{2-\delta_\Gamma} + d_{\phi_0} y^{\delta_\Gamma}$$

where $c_{\phi_0} > 0$ and $d_{\phi_0} \geq 0$.

In this section, we aim to prove the following theorem.

Theorem 6.1. *For any $\psi \in C_c^\infty(\Gamma\backslash G)^K$,*

$$\int_{n_x \in (N\cap\Gamma)\backslash N} \psi(n_x a_y)\, dx = \langle \psi, \phi_0\rangle \cdot c_{\phi_0} \cdot y^{2-\delta_\Gamma}(1 + O(y^{\frac{2}{7}(\delta_\Gamma - s_\Gamma)}))$$

where the implied constant depends only on the Sobolev norms of ψ, the volume of $N(supp(\psi))$ and the volume of an open subset of $(N\cap\Gamma)\backslash N$ which properly covers $\Lambda_N(\Gamma)$.

Most of this section is devoted to a proof of Theorem 6.1.

Definition 6.2. For a given $\psi \in C^\infty(\Gamma\backslash G)^K$ and $\eta \in C_c((N\cap\Gamma)\backslash N)$, define the function $\mathrm{I}_\eta(\psi)$ on G by

$$\mathrm{I}_\eta(\psi)(a_y) := \int_{n_x \in (N\cap\Gamma)\backslash N} \psi(n_x a_y)\eta(n_x)\, dx.$$

We denote by N^- the strictly lower triangular subgroup of G:

$$N^- := \{\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} : x \in \mathbb{C}\}.$$

The product map $N \times A \times N^- \times M \to G$ is a diffeomorphism at a neighborhood of e. Let ν be a smooth measure on AN^-M such that $dn \otimes \nu(an^-m) = d\mu$.

Fix a left-invariant Riemannian metric d on G and denote by U_ϵ the ball of radius ϵ about e in G.

Definition 6.3.

- We fix a non-negative function $\eta \in C_c^\infty((N\cap\Gamma)\backslash N)$ with $\eta = 1$ on a bounded open subset of $\mathcal{F}_N$ which properly covers $\Lambda_N(\Gamma)$.
- Fix $\epsilon_0 > 0$ so that for the ϵ_0-neighborhood U_{ϵ_0} of e, the multiplication map

  $$\mathrm{supp}(\eta) \times (U_{\epsilon_0} \cap AN^-M) \to \mathrm{supp}(\eta)(U_{\epsilon_0} \cap AN^-M) \subset \Gamma\backslash G$$

  is a bijection onto its image.
- For each $\epsilon < \epsilon_0$, let r_ϵ be a non-negative smooth function in AN^-M whose support is contained in

  $$W_\epsilon := (U_\epsilon \cap A)(U_{\epsilon_0} \cap N^-)(U_{\epsilon_0} \cap M)$$

  and $\int_{W_\epsilon} r_\epsilon\, d\nu = 1$.
- We define the following function $\rho_{\eta,\epsilon}$ on $\Gamma\backslash G$ which is 0 outside $\mathrm{supp}(\eta)U_{\epsilon_0}$ and for $g = n_x a n^- m \in \mathrm{supp}(\eta)(U_{\epsilon_0} \cap AN^-M)$,

  $$\rho_{\eta,\epsilon}(g) := \eta(n_x) \otimes r_\epsilon(an^-m).$$

Recall from Proposition 4.6 that

$$\phi_0^N(a_y) = \mathrm{I}_\eta(\phi_0)(a_y) + O(y^{\delta_\Gamma}).$$

Proposition 6.4. *We have for all small $0 < \epsilon \ll \epsilon_0$ and for all $0 < y < 1$,*

$$\phi_0^N(a_y) = \langle a_y\phi_0, \rho_{\eta,\epsilon}\rangle_{L^2(\Gamma\backslash G)} + O_\eta(\epsilon \cdot y^{2-\delta_\Gamma}) + O(y^{\delta_\Gamma})$$

where the first implied constant depends on the Lipschitz constant for η.

Proof. Let $h = a_{y_0} n_x^- m \in W_\epsilon$. Then for $n \in N$ and $y > 0$, we have

$$nha_y = na_{yy_0} n_{yx}^- m.$$

As the product map $A \times N \times K \to G$ is a diffeomorphism and hence a bi-Lipschitz map at a neighborhood of e, there exists $\ell > 0$ such that the ϵ-neighborhood of e in G is contained in the product $A_{\ell\epsilon} N_{\ell\epsilon} K_{\ell\epsilon}$ of $\ell\epsilon$-neighborhoods for all small $\epsilon > 0$.

Therefore we may write

$$n_{yx}^- = a_{y_1} n_{x_1} k_1 \in A_{\ell y \epsilon_0} N_{\ell y \epsilon_0} K$$

and hence

$$\begin{aligned} nha_y &= na_{yy_0y_1} n_{x_1} k_1 m \\ &= n(a_{yy_0y_1} n_{x_1} a_{yy_0y_1}^{-1}) a_{yy_0y_1} k_1 m = n(n_{x_1 yy_0y_1}) a_{yy_0y_1} k_1 m. \end{aligned}$$

As ϕ_0 is K-invariant and dn is N-invariant,

$$\begin{aligned} \int_{N\cap\Gamma\backslash N} \phi_0(nha_y) \cdot \eta(n) dn &= \int_{N\cap\Gamma\backslash N} \phi_0(n(n_{x_1 yy_0y_1}) a_{yy_0y_1}) \cdot \eta(n) dn \\ &= \int_{N\cap\Gamma\backslash N} \phi_0(na_{yy_0y_1})(\eta(n) + O(\epsilon)) dn \end{aligned}$$

as $\eta(n) - \eta(nn') = O_\eta(\epsilon)$ for all $n \in N$ and $n' \in N \cap U_\epsilon$. By Corollary 4.13, we deduce

$$\begin{aligned} &\int_{N\cap\Gamma\backslash N} \phi_0(nha_y) \cdot \eta(n) dn \\ &= \int_{N\cap\Gamma\backslash N} \phi_0(na_{yy_0y_1}) \eta(n) dn + O_\eta(\epsilon \phi_0^N(a_{yy_0y_1})) \\ &= c_{\phi_0}(yy_0y_1)^{2-\delta_\Gamma} + d_{\phi_0}(yy_0y_1)^{\delta_\Gamma} + O_\eta(\epsilon y^{2-\delta_\Gamma}) \\ &= c_{\phi_0} y^{2-\delta_\Gamma}(1 + O(1 - (y_0y_1)^{2-\delta_\Gamma})) + O_\eta(\epsilon y^{2-\delta_\Gamma}) + O(y^{\delta_\Gamma}) \\ &= c_{\phi_0} y^{2-\delta_\Gamma}(1 + O(\epsilon)) + O(y^{\delta_\Gamma}) \end{aligned}$$

as $|y_0 - 1| = O(\epsilon)$ and $|y_1 - 1| = O(y\epsilon)$.

As $\int r_\epsilon d\nu(h) = 1$, we deduce

$$\begin{aligned} \langle a_y \phi_0, \rho_{\eta,\epsilon} \rangle &= \int_{W_\epsilon} r_\epsilon(h) \left(\int_{N\cap\Gamma\backslash N} \phi_0(nha_y) \eta(n)\, dn \right) d\nu(h) \\ &= c_{\phi_0} y^{2-\delta_\Gamma}(1 + O_\eta(\epsilon)) + O(y^{\delta_\Gamma}) \\ &= \phi_0^N(a_y) + O_\eta(\epsilon y^{2-\delta_\Gamma}) + O(y^{\delta_\Gamma}). \end{aligned}$$

□

Lemma 6.5. *For $\psi \in C_c^\infty(\Gamma\backslash G)^K$, there exists $\hat{\psi} \in C_c^\infty(\Gamma\backslash G)^K$ such that*

(1) *for all small* $\epsilon > 0$ *and* $h \in U_\epsilon$,
$$|\psi(g) - \psi(gh)| \le \epsilon \cdot \hat{\psi}(g) \quad \text{for all } g \in \Gamma\backslash G.$$
(2) $\mathcal{S}_m(\hat{\psi}) \ll \mathcal{S}_5(\psi)$ *for any* $m \in \mathbb{N}$, *where the implied constant depends only on* $\operatorname{supp}(\psi)$.

Proof. Fix $\epsilon_0 > 0$. Let $f_0 \in C_c^\infty(\Gamma\backslash G)^K$ such that $f_0(g) = 1$ for all $g \in \operatorname{supp}(\psi)U_{\epsilon_0}^{-1}K$ and $f_0(g) = 0$ for all $g \in \Gamma\backslash G - \operatorname{supp}(\psi)U_{2\epsilon_0}^{-1}K$.

Set $C_\psi := \sup_{g\in\operatorname{supp}(\psi)} \sum_{i=1}^6 |X_i(\psi)(g)|$. Then there exists a constant $c_0 \ge 1$ such that for all $g \in \Gamma\backslash G$ and $h \in U_\epsilon$ for $\epsilon < \epsilon_0$,
$$|\psi(g) - \psi(gh)| \le \epsilon \cdot c_0 C_\psi.$$

Hence if we define $\hat{\psi} \in C_c^\infty(\Gamma\backslash G)^K$ by $\hat{\psi}(g) = c_0 C_\psi f_0(g)$ for $g \in \Gamma\backslash G$, then (1) holds.

Now by the Sobolev imbedding theorem (cf. [1, Thm. 2.30]), we have
$$C_\psi \le \mathcal{S}_5(\psi).$$
Since $\mathcal{S}_m(\hat{\psi}) \ll C_\psi$, this proves (2). □

Proposition 6.6. *Let* $\psi \in C^\infty(\Gamma\backslash G)^K$. *Then for any* $0 < y < 1$ *and any small* $\epsilon > 0$,
$$|\operatorname{I}_\eta(\psi)(a_y) - \langle a_y\psi, \rho_{\eta,\epsilon}\rangle| \ll (\epsilon + y) \cdot \operatorname{I}_\eta(\hat{\psi})(a_y).$$

Proof. If $an^-m \in W_\epsilon = (U_\epsilon \cap A)(U_{\epsilon_0} \cap N^-)(U_{\epsilon_0} \cap M)$, then
$$(an^-m)a_y = a_y a(a_{y^{-1}} n^- a_y)m \in a_y W_\epsilon$$
since $a_{y^{-1}}$ contracts N^- by conjugation as $0 < y < 1$.

As ψ is M-invariant, for any $h = an^-m \in W_\epsilon$, there exists an $h' \in (U_\epsilon \cap A)(U_{y\epsilon_0} \cap N^-)$ such that
$$|\psi(na_y) - \psi(nha_y)| = |\psi(na_y) - \psi(na_y h')| \ll \hat{\psi}(na_y)(\epsilon + y).$$

Hence
$$|\psi(na_y) - \int_{h\in W_\epsilon} \psi(nha_y) r_\epsilon(h) d\nu(h)| \ll \hat{\psi}(na_y)(\epsilon + y).$$
Therefore
$$|\operatorname{I}_\eta(\psi)(a_y) - \langle a_y\psi, \rho_\epsilon\rangle_{L^2(\Gamma\backslash G)}| \ll (\epsilon + y) \cdot \int_{(N\cap\Gamma)\backslash N} \hat{\psi}(na_y)\eta(n)dn.$$
□

Proof of Theorem 6.1: Recall that $\eta, \epsilon_0, \rho_{\eta,\epsilon} = \eta \otimes r_\epsilon$ are as in Def 6.3. For simplicity, we set $\rho_\epsilon = \rho_{\eta,\epsilon}$. Noting that r_ϵ is essentially an ϵ-approximation only in the A-direction, we obtain that $\mathcal{S}_2(\rho_\epsilon) = O_\eta(\epsilon^{-5/2})$.

We may further assume that $\eta = 1$ on $N(\operatorname{supp}(\psi))$ by Corollary 3.10 so that
$$\int_{(N\cap\Gamma)\backslash N} \psi(na_y)\, dn = \operatorname{I}_\eta(\psi)(a_y).$$

By Proposition 4.6, we also have

$$\phi_0^N(a_y) = \mathrm{I}_\eta(\phi_0)(a_y) + O(y^{\delta_\Gamma}).$$

Set $p = 5/2$. Fix $\ell \in \mathbb{N}$ so that

$$\ell > \frac{(2-\delta_\Gamma)(p+1)}{(\delta_\Gamma - s_\Gamma)}.$$

Setting $\psi_0(g) := \psi(g)$, we define for $1 \le i \le \ell$, inductively

$$\psi_i(g) := \hat{\psi}_{i-1}(g)$$

where $\hat{\psi}_{i-1}$ is given by Lemma 6.5.

Applying Proposition 6.6 to each ψ_i, we obtain for $0 \le i \le \ell - 1$

$$\mathrm{I}_\eta(\psi_i)(a_y) = \langle a_y\psi_i, \rho_\epsilon\rangle + O((\epsilon + y)\,\mathrm{I}_\eta(\psi_{i+1})(a_y))$$

and

$$\mathrm{I}_\eta(\psi_\ell)(a_y) = \langle a_y\psi_\ell, \rho_\epsilon\rangle + O_\eta((\epsilon + y)\mathcal{S}_2(\psi_\ell))$$

where the implied constant in the O_η notation depends on $\int \eta\, dn$.

Note that by Corollary 5.6, we have for each $1 \le i \le \ell$,

$$\begin{aligned}
\langle a_y\psi_i, \rho_\epsilon\rangle &= \langle \psi_i, \phi_0\rangle\langle a_y\phi_0, \rho_\epsilon\rangle + O(y^{2-s_\Gamma}\mathcal{S}_2(\rho_\epsilon)\mathcal{S}_2(\psi_i)) \\
&= O(\langle a_y\phi_0, \rho_\epsilon\rangle \cdot \|\psi_i\|_2) + O(y^{2-s_\Gamma}\mathcal{S}_2(\rho_\epsilon)\mathcal{S}_2(\psi_i)) \\
&= O(\langle a_y\phi_0, \rho_\epsilon\rangle \cdot \mathcal{S}_5(\psi)) + O(y^{2-s_\Gamma}\epsilon^{-p}\mathcal{S}_5(\psi)).
\end{aligned}$$

Hence for any $y < \epsilon$,

$$\begin{aligned}
&\mathrm{I}_\eta(\psi)(a_y) = \langle a_y\psi, \rho_\epsilon\rangle + \sum_{k=1}^{\ell-1} O(\langle a_y\psi_k, \rho_\epsilon\rangle(\epsilon+y)^k) + O_\eta(\mathcal{S}_5(\psi)(\epsilon+y)^\ell) \\
&= \langle a_y\psi, \rho_\epsilon\rangle + O(\langle a_y\phi_0, \rho_\epsilon\rangle\epsilon\mathcal{S}_5(\psi)) + O(\epsilon\mathcal{S}_5(\psi)y^{2-s_\Gamma}\epsilon^{-p}) + O(\mathcal{S}_5(\psi)\epsilon^\ell) \\
&= \langle \psi, \phi_0\rangle \cdot \langle a_y\phi_0, \rho_\epsilon\rangle + O(\langle a_y\phi_0, \rho_\epsilon\rangle\epsilon) + O(y^{2-s_\Gamma}\epsilon^{-p}) + O(\epsilon^\ell) \\
&= \langle \psi, \phi_0\rangle \cdot \phi_0^N(a_y) + O(y^{\delta_\Gamma}) + O(\epsilon y^{2-\delta_\Gamma}) + O(y^{2-s_\Gamma}\epsilon^{-p}) + O(\epsilon^\ell)
\end{aligned}$$

by Proposition 6.4, where the implied constants depend on the Sobolev norm $\mathcal{S}_5(\psi)$ and $\int \eta\, dn$.

Equating the two error terms $O(\epsilon y^{2-\delta_\Gamma})$ and $O(y^{2-s_\Gamma}\epsilon^{-p})$ gives the choice $\epsilon = y^{(\delta_\Gamma - s_\Gamma)/(p+1)}$. By the condition on ℓ, we then have $\epsilon^\ell \ll y^{2-\delta_\Gamma + \frac{2}{7}(\delta_\Gamma - s_\Gamma)}$. Hence we deduce:

$$\int_{(N\cap\Gamma)\backslash N} \psi(na_y)\, dn = \mathrm{I}_\eta(\psi)(a_y) = \langle \psi, \phi_0\rangle \cdot \phi_0^N(a_y) \cdot (1 + O(y^{\frac{2}{7}(\delta_\Gamma - s_\Gamma)})).$$

Note that the implied constant depends on the Sobolev norm $\mathcal{S}_5(\psi)$ of ψ and the L^1-norm $\int \eta\, dn$, which in turn depends only on the volumes $N(\mathrm{supp}(\psi))$ and any open subset of $(N \cap \Gamma)\backslash N$ which properly covers $\Lambda_N(\Gamma)$.

Burger-Roblin measure $\hat{\mu}$: In identifying $\partial_\infty(\mathbb{H}^3)$ with K/M, we may define the following measure $\hat{\mu}$ on $\mathrm{T}^1(\mathbb{H}^3) = G/M$: for $\psi \in C_c(G/M)$,

$$\hat{\mu}(\psi) = \int_{k\in K}\int_{a_y n_x \in AN} \psi(k a_y n_x) y^{\delta-1} dy dx d\nu_o(k)$$

where we consider the Patterson Sullivan measure ν_o in section 4 as a measure on K via the projection $K \to K/M$: for $f \in C(K)$, $\nu_o(f) = \int_{k\in K/M}\int_{m\in M} f(km) dm d\nu_o(k)$ for the probability invariant measure dm on M.

By the conformal property of ν_o, the measure $\hat{\mu}$ is left Γ-invariant and hence induces a Radon measure on $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ via the canonical projection.

Lemma 6.7. *For a K-invariant function $\psi \in C_c(G)$, we have*

$$\hat{\mu}(\psi) = \langle \psi, \phi_0\rangle. \tag{6.8}$$

Proof. It is easy to compute that for the base point $o = (0,0,1) \in \mathbb{H}^3$,

$$\beta_\infty(a_y n_x o, o) = -\log y$$

for any $0 < y \le 1$ and $x \in \mathbb{C}$. We note that $dg = y^{-1} dy dx dk$ for $g = a_y n_x k$ is a Haar measure on G. Therefore, using ψ is K-invariant,

$$\begin{aligned}
\hat{\mu}(\psi) &= \int_K \int_{AN} \int_{k_0 \in K} \psi(k a_y n_x k_0) d(k_0) y^{\delta-1} dy dx d\nu_o(k) \\
&= \int_{g\in G}\int_K \psi(kg) e^{-\delta\beta_\infty(go,o)} d\nu_o(k) dg \\
&= \int_G \psi(g) \int_{k\in K/M} e^{-\delta\beta_\infty(k^{-1}go,o)} d\nu_o(k) dg \\
&= \int_G \psi(g) \int_{K/M} e^{-\delta\beta_{k(\infty)}(go,o)} d\nu_o(k) dg = \langle \psi, \phi_0\rangle
\end{aligned}$$

as $\phi_0(g) = \int_{K/M} e^{-\delta\beta_{k(\infty)}(go,o)} d\nu_o(k)$. □

Generalizing Burger's result [7], Roblin [38, Thm 6.4] proved:

Theorem 6.9. *The measure $\hat{\mu}$ is, up to constant multiple, the unique Radon measure on $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ invariant for the horospherical foliations whose support is $\hat{\Omega}_\Gamma$ and which gives zero measure to closed horospheres.*

We call the measure $\hat{\mu}$ the Burger-Roblin measure.

In Appendix A, the following theorem is deduced from Theorem 6.1.

Theorem 6.10. *For $\psi \in C_c(\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3))$,*

$$\int_{n_x \in (N\cap\Gamma)\backslash N} \psi(n_x a_y)\, dx \sim c_{\phi_0} \cdot \hat{\mu}(\psi) \cdot y^{2-\delta_\Gamma} \quad \textit{as } y \to 0.$$

7. Orbital counting for a Kleinian group

Let $\iota : G = \mathrm{PSL}_2(\mathbb{C}) \to \mathrm{SO}_F(\mathbb{R})$ be a real linear representation where F is a real quadratic form in 4 variables of signature $(3,1)$. Let $\Gamma < G$ be a geometrically finite torsion-free discrete subgroup with $\delta_\Gamma > 1$. Let $v_0 \in \mathbb{R}^4$ be a non-zero (row) vector with $F(v_0) = 0$ such that the orbit $v_0\Gamma$ is discrete in $\mathbb{R}^4$.

Since the orthogonal group $\mathrm{O}_F(\mathbb{R})$ acts transitively on the set of non-zero vectors of the cone $F = 0$, there exists $g_0^t \in \mathrm{O}_F(\mathbb{R})$ such that the stabilizer of $g_0^t v_0^t$ is equal to $N^- M$ where N^- is the strictly lower triangular subgroup. In fact, $g_0^t v_0^t$ is unique up to homothety. Set

$$\Gamma_0 = g_0^{-1} \Gamma g_0.$$

As $v_0\Gamma$ is discrete, it follows that $\Gamma_0 \backslash \Gamma_0 NM$ is closed. Hence by Lemma 3.4, the orbit $\Gamma \backslash \Gamma g_0 N$ is closed, equivalently $\Gamma_0 \backslash \Gamma_0 N$ is closed.

Denote by $\phi_0 \in L^2(\Gamma_0 \backslash \mathbb{H}^3)$ the unique positive eigenfunction of Δ with eigenvalue $\delta_\Gamma(2 - \delta_\Gamma)$ and of unit L^2-norm $\int_{\Gamma_0 \backslash \mathbb{H}^3} \phi_0(x,y)^2 y^{-3} dx dy = 1$. By Corollary 4.13, we have

$$\phi_0^N(a_y) = c_{\phi_0} y^{2-\delta_\Gamma} + d_{\phi_0} y^{\delta_\Gamma}$$

where $c_{\phi_0} > 0$ and $d_{\phi_0} \geq 0$. Recall that the Patterson-Sullivan measure ν_o on K which is normalized so that

$$\phi_0(x,y) = \int_{u \in \Lambda(\Gamma) \backslash \{\infty\}} \left(\frac{(\|u\|^2 + 1) y}{\|x - u\|^2 + y^2} \right)^{\delta_\Gamma} d\nu_o(u).$$

Theorem 7.1. *For any norm $\|\cdot\|$ on $\mathbb{R}^4$, we have, as $T \to \infty$,*

$$\#\{v \in v_0\Gamma : \|v\| < T\} \sim \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot \left(\int_{k \in K} \|v_0(g_0 k^{-1} g_0^{-1})\|^{-\delta_\Gamma} d\nu_o(k) \right) \cdot T^{\delta_\Gamma}.$$

If $\|\cdot\|$ is $g_0 K g_0^{-1}$-invariant, then there exists $\epsilon > 0$ such that

$$\#\{v \in v_0\Gamma : \|v\| < T\} = \delta_\Gamma^{-1} \cdot \phi_0(e) \cdot c_{\phi_0} \cdot \|v_0\|^{-\delta_\Gamma} \cdot T^{\delta_\Gamma}(1 + O(T^{-\epsilon}))$$

where ϵ depends only on the spectral gap $\delta_\Gamma - s_\Gamma$ and the implied constant depends only on $\Lambda_N(\Gamma)$.

By replacing Γ with $\Gamma_0 = g_0^{-1}\Gamma g_0$, we may assume henceforth that $g_0 = e$, and thus the stabilizer of v_0 in G is NM. By Lemma 3.4, the stabilizer of v_0 in Γ is simply $\Gamma \cap N$.

Note that $N^- M$ fixes v_0^t, and that the highest weight β of the (irreducible) representation of ι is given by $\beta(a_y) = y^{-1}$. It follows that $\iota(a_y) v_0^t = y^{-1} v_0^t$, and hence $v_0 a_y = y^{-1} v_0$.

Set

$$B_T := \{v \in v_0 G : \|v\| < T\}.$$

Define the following function on $\Gamma \backslash G$:

$$F_T(g) := \sum_{\gamma \in (N \cap \Gamma) \backslash \Gamma} \chi_{B_T}(v_0 \gamma g).$$

Since $v_0\Gamma$ is discrete, $F_T(g)$ is well-defined and

$$F_T(e) = \#\{v \in v_0\Gamma : \|v\| < T\}.$$

We use the notation: for $\psi \in C_c(\Gamma\backslash G)$,

$$\psi^N(a_y) := \int_{(N\cap\Gamma)\backslash N} \psi(na_y)\, dn.$$

Lemma 7.2. *For any $\psi \in C_c(\Gamma\backslash G)$ and $T > 0$,*

$$\langle F_T, \psi\rangle = \int_{k\in M\backslash K}\int_{y>T^{-1}\|v_0k\|} \psi_k^N(a_y) y^{-3} dy dk$$

where $\psi_k(g) = \int_{m\in M} \psi(gmk)dm$.

Proof. Observe:

$$\begin{aligned}
\langle F_T, \psi\rangle &= \int_{\Gamma\backslash G} \sum_{\gamma\in(N\cap\Gamma)\backslash\Gamma} \chi_{B_T}(v_0\gamma g)\psi(g)dg \\
&= \int_{g\in(N\cap\Gamma)\backslash G} \chi_{B_T}(v_0 g)\psi(g)dg \\
&= \int_{k\in M\backslash K}\int_{y:\|v_0a_yk\|<T}\int_{n_xm\in(N\cap\Gamma)\backslash NM} \psi(n_xma_yk)y^{-3}\, dx\, dm\, dy\, dk \\
&= \int_{k\in M\backslash K}\int_{y>T^{-1}\|v_0k\|} \left(\int_{n_x\in(N\cap\Gamma)\backslash N}\int_{m\in M} \psi(n_xa_ymk)\, dm\, dx\right) y^{-3}dy\, dk \\
&= \int_{k\in M\backslash K}\int_{y>T^{-1}\|v_0k\|} \left(\int_{n_x\in(N\cap\Gamma)\backslash N} \psi_k(n_xa_y)dx\right) y^{-3}dy\, dk \\
&= \int_{k\in M\backslash K}\int_{y>T^{-1}\|v_0k\|} \psi_k^N(a_y)\, y^{-3}dy\, dk.
\end{aligned}$$

∎

Define a function $\xi_{v_0} : K \to \mathbb{R}$ by

$$\xi_{v_0}(k) = \|v_0k\|^{-\delta_\Gamma}.$$

For $\psi \in C_c(\Gamma\backslash G)$, the convolution $\xi_{v_0} * \psi$ is then given by

$$\xi_{v_0} * \psi(g) = \int_{k\in K} \psi(gk)\|v_0k\|^{-\delta_\Gamma} dk.$$

Corollary 7.3. *For any $\psi \in C_c(\Gamma\backslash G)$, we have, as $T \to \infty$,*

$$\langle F_T, \psi\rangle \sim \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot T^{\delta_\Gamma} \cdot \hat{\mu}(\xi_{v_0} * \psi).$$

For $\psi \in C_c^\infty(\Gamma\backslash G)^K$ and $\|\cdot\|$ K-invariant,

$$\langle F_T, \psi\rangle = \langle \psi, \phi_0\rangle \cdot \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot T^{\delta_\Gamma} \cdot \|v_0\|^{-\delta_\Gamma}(1 + O(T^{-\frac{2}{7}(\delta_\Gamma - s_\Gamma)}))$$

where s_Γ is as in Def. 5.5 and the implied constant depends only on the Sobolev norm of ψ, $\Lambda_N(\Gamma)$ and $N(supp(\psi))$.

Proof. Note that for any $\psi \in C_c(\Gamma\backslash G)$ and $k \in K$, the function ψ_k, defined in Lemma 7.2, is M-invariant. Applying Theorem 6.10 to ψ_k, we obtain that as $y \to 0$,

$$\int_{n_x \in (N\cap\Gamma)\backslash N} \psi_k(n_x a_y) dx \sim c_{\phi_0} \cdot y^{2-\delta_\Gamma} \cdot \hat{\mu}(\psi_k). \tag{7.4}$$

Hence by applying Lemma 7.2, inserting the definition of ψ_k, and evaluating the y-integral, we get

$$\begin{aligned}\langle F_T, \psi\rangle &\sim c_{\phi_0} \cdot \int_{M\backslash K} \int_{y > T^{-1}\|v_0 k\|} y^{-1-\delta_\Gamma} \hat{\mu}(\psi_k)\, dy\, dk \\ &= \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot T^{\delta_\Gamma} \int_{M\backslash K} \int_M \hat{\mu}(\psi(gmk)) \cdot \|v_0 k\|^{-\delta_\Gamma}\, dmdk \\ &= \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot T^{\delta_\Gamma} \cdot \hat{\mu}(\xi_{v_0} * \psi),\end{aligned}$$

proving the first claim.

Now suppose that both ψ and the norm $\|\cdot\|$ are K-invariant.

As $\psi_k = \psi$, by Theorem 6.1, we can replace (7.4) by an asymptotic formula with power savings error term:

$$\int_{n_x \in (N\cap\Gamma)\backslash N} \psi(n_x a_y) dx = c_{\phi_0} y^{2-\delta_\Gamma} \langle \psi, \phi_0\rangle (1 + O(y^{\frac{2}{7}(\delta_\Gamma - s_\Gamma)}))$$

and the implied constant depends on the Sobolev norm of ψ and $\Lambda_N(\Gamma)$.

On the other hand,

$$\xi_{v_0} * \psi = \|v_0\|^{-\delta_\Gamma} \cdot \psi,$$

and hence

$$\hat{\mu}(\xi_{v_0} * \psi) = \|v_0\|^{-\delta_\Gamma} \cdot \langle \psi, \phi_0\rangle.$$

Therefore

$$\begin{aligned}\langle F_T, \psi\rangle &= c_{\phi_0} \int_{y > T^{-1}\|v_0\|} y^{-1-\delta_\Gamma} \langle \psi, \phi_0\rangle (1 + O(y^{\frac{2}{7}(\delta_\Gamma - s_\Gamma)}))\, dy \\ &= \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot T^{\delta_\Gamma} \cdot \|v_0\|^{-\delta_\Gamma} (1 + O(T^{-\frac{2}{7}(\delta_\Gamma - s_\Gamma)}))\end{aligned}$$

where the implied constant depends only on the Sobolev norm of ψ and the set $\Lambda_N(\Gamma)$. □

Theorem 7.5. *As $T \to \infty$,*

$$F_T(e) \sim \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot \left(\int_K \|v_0 k^{-1}\|^{-\delta_\Gamma}\, d\nu_o(k) \right) \cdot T^{\delta_\Gamma}.$$

If $\|\cdot\|$ is K-invariant, then for some $\epsilon > 0$ (depending only on the spectral gap $\delta_\Gamma - s_\Gamma$),

$$F_T(e) = \delta_\Gamma^{-1} \cdot \phi_0(e) \cdot c_{\phi_0} \cdot \|v_0\|^{-\delta_\Gamma} \cdot T^{\delta_\Gamma} (1 + O(T^{-\epsilon})).$$

where the implied constant depends only on $\Lambda_N(\Gamma)$.

Proof. For all small $\epsilon > 0$, we choose a symmetric ϵ- neighborhood U_ϵ of e in G, which injects to $\Gamma\backslash G$, such that for all $T \gg 1$ and all $0 < \epsilon \ll 1$,

$$B_T U_\epsilon \subset B_{(1+\epsilon)T} \quad \text{and} \quad B_{(1-\epsilon)T} \subset \cap_{u \in U_\epsilon} B_T u.$$

For $\epsilon > 0$, let $\phi_\epsilon \in C_c^\infty(G)$ denote a non-negative function supported on U_ϵ and with $\int_G \phi_\epsilon \, dg = 1$. We lift ϕ_ϵ to $\Gamma\backslash G$ by averaging over Γ:

$$\Phi_\epsilon(\Gamma g) = \sum_{\gamma \in \Gamma} \phi_\epsilon(\gamma g).$$

Then

$$\langle F_{(1-\epsilon)T}, \Phi_\epsilon \rangle \le F_T(e) \le \langle F_{(1+\epsilon)T}, \Phi_\epsilon \rangle. \tag{7.6}$$

Note that

$$\langle F_{(1\pm\epsilon)T}, \Phi_\epsilon \rangle \sim \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot (T(1 \pm \epsilon))^{\delta_\Gamma} \cdot \hat{\mu}(\xi_{v_0} * \Phi_\epsilon). \tag{7.7}$$

Considering the function $R_{v_0} : G \to \mathbb{R}$ given by

$$R_{v_0}(g) := y^{\delta_\Gamma} \xi_{v_0}(k_0)$$

for $g = a_y n_x k_0 \in ANK$, we have

$$\begin{aligned}
\hat{\mu}(\xi_{v_0} * \Phi_\epsilon) &= \int_{g \in U_\epsilon} \int_{k_0 \in K} \phi_\epsilon(g k_0) \xi_{v_0}(k_0) d(k_0) d\hat{\mu}(g) \\
&= \int_{k a_y n_x \in KAN \cap U_\epsilon} \int_{k_0 \in K} \phi_\epsilon(k a_y n_x k_0) \xi_{v_0}(k_0) y^{\delta - 1} d(k_0) dy dx d\nu_o(k) \\
&= \int_{k \in K} \int_{g \in G} \phi_\epsilon(kg) R(g) dg d\nu_o(k) \\
&= \int_{g \in U_\epsilon} \phi_\epsilon(g) \int_{k \in K} R(k^{-1} g) d\nu_o(k) dg \\
&= \int_{k \in K} R(k^{-1}) d\nu_o(k) + O(\epsilon) \\
&= \int_{k \in K/M} \|v_0 k^{-1}\|^{-\delta_\Gamma} d\nu_o(k) + O(\epsilon).
\end{aligned}$$

as $\int_G \phi_\epsilon \, dg = 1$, where the implied constant depends only on the Lipschitz constant for R.

As $\epsilon > 0$ is arbitrary, we deduce that

$$F_T(e) \sim \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot \int_{k \in K} \|v_0 k^{-1}\|^{-\delta_\Gamma} d\nu_o(k) \cdot T^{\delta_\Gamma}.$$

If $\|\cdot\|$ is K-invariant, we may take both U_ϵ and ϕ_ϵ to be K-invariant. Hence by Corollary 7.3, we may replace (7.7) by

$$\langle F_{(1\pm\epsilon)T}, \Phi_\epsilon \rangle = \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot \|v_0\|^{-\delta_\Gamma} \cdot (T(1 \pm \epsilon))^{\delta_\Gamma} \cdot \langle \phi_0, \Phi_\epsilon \rangle \cdot (1 + O(T^{-\frac{2}{7}(\delta_\Gamma - s_\Gamma)}))$$

where the implied constant depends on $\mathcal{S}_5(\Phi_\epsilon) = \mathcal{S}_5(\phi_\epsilon)$, and the sets $\Lambda_N(\Gamma)$ and $N(U_\epsilon) := \{[n] \in (N \cap \Gamma)\backslash N : \Gamma n A \cap U_\epsilon \neq \emptyset\}$.

Since

$$\langle \phi_0, \Phi_\epsilon \rangle = \phi_0(e) + O(\epsilon),$$

we have

$$\langle F_{(1\pm\epsilon)T}, \Phi_\epsilon \rangle = \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot \|v_0\|^{-\delta_\Gamma} \cdot T^{\delta_\Gamma} \cdot \phi_0(e) + O(\epsilon T^{\delta_\Gamma}) + O(\epsilon^{-q} T^{\delta_\Gamma - \frac{2}{7}(\delta_\Gamma - s_\Gamma)})$$

for some $q > 0$ depending on $\mathcal{S}_5(\phi_\epsilon)$. Therefore by setting $\epsilon^{1+q} = T^{-\frac{2}{7}(\delta_\Gamma - s_\Gamma)}$,

$$F_T(e) = \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot \|v_0\|^{-\delta_\Gamma} \cdot T^{\delta_\Gamma} \cdot \phi_0(e) \cdot (1 + O(T^{-\epsilon'}))$$

for $\epsilon' = \frac{2}{7(1+q)}(\delta_\Gamma - s_\Gamma)$, where the implied constant depends only on $\Lambda_N(\Gamma)$. □

8. The Selberg sieve and circles of prime curvature

8.1. **Selberg's sieve.** We first recall the Selberg upper bound sieve. Let A denote the finite sequence of real non-negative numbers $\mathsf{A} = \{a_n\}$, and let P be a finite product of distinct primes. We are interested in an upper bound for the quantity

$$S(\mathsf{A}, P) := \sum_{(n,P)=1} a_n.$$

To estimate $S(\mathsf{A}, P)$ we need to know how A is distributed along certain arithmetic progressions. For $d \nmid P$, define

$$\mathsf{A}_d := \{a_n \in \mathsf{A} : n \equiv 0(d)\}$$

and set $|\mathsf{A}_d| := \sum_{n \equiv 0(d)} a_n$. We record

Theorem 8.1. [19, Theorem 6.4] *Suppose that there exists a finite set S of primes such that P has no prime factor in S. Suppose that there exist $\mathcal{X} > 1$ and a function g on square-free integers with $0 < g(p) < 1$ for $p|P$ and g is multiplicative outside S (i.e., $g(d_1 d_2) = g(d_1)g(d_2)$ if d_1 and d_2 are square-free integers with no factors in S) such that for all $d \nmid P$ square-free,*

$$|\mathsf{A}_d| = g(d)\mathcal{X} + r_d(\mathsf{A}). \tag{8.2}$$

Let h be the multiplicative function on square-free integers (outside S) given by $h(p) = \frac{g(p)}{1-g(p)}$. Then for any $D > 1$, we have that

$$S(\mathsf{A}, P) \leq \mathcal{X} \left(\sum_{d < \sqrt{D}, d|P} h(d) \right)^{-1} + \sum_{d < D, d|P} \tau_3(d) \cdot |r_d(\mathsf{A})|,$$

where $\tau_3(d)$ denotes the number of representations of d as the product of three natural numbers.

8.2. **Executing the sieve.** Recall from section 2 that Q denotes the Descartes quadratic form and $\mathcal{A}$ denotes the Apollonian group in $\mathrm{O}_Q(\mathbb{Z})$. Fix a primitive integral Apollonian packing $\mathcal{P}$ with its root quadruple $\xi \in \mathbb{Z}^4$. Then $\mathcal{P}$ is either bounded or given by the one in Figure 3.

To execute the sieve, it is important to work with a simply connected group. Hence we will set $\Gamma_{\mathcal{A}}$ to be the preimage of $\mathrm{SO}_Q(\mathbb{R})^\circ \cap \mathcal{A}$ in the spin double cover $\mathrm{Spin}_Q(\mathbb{R})$ of $\mathrm{SO}_Q(\mathbb{R})$. Recall that α denotes the Hausdorff dimension of the residual set of the packing $\mathcal{P}$. As shown in section 2.5, α is equal to $\delta_{\Gamma_{\mathcal{A}}}$, the Hausdorff dimension of the limit set of $\Gamma_{\mathcal{A}}$. As $\mathrm{Spin}_Q(\mathbb{R})$ is isomorphic to $\mathrm{SL}_2(\mathbb{C})$, we have a real linear representation $\iota : \mathrm{SL}_2(\mathbb{C}) \to \mathrm{SO}_Q(\mathbb{R})$ which factors through the quotient map $\mathrm{SL}_2(\mathbb{C}) \to \mathrm{PSL}_2(\mathbb{C})$. By setting Γ to be the preimage of $\Gamma_{\mathcal{A}}$ under ι, the counting results in the previous section are all valid for Γ.

As before, B_T denotes the ball in the cone:

$$B_T := \{v \in \mathbb{R}^4 : Q(v^t) = 0, \ \ \|v\| < T\}$$

Since we are only looking for an upper bound, we may assume $\|\cdot\|$ is $g_0 K g_0^{-1}$-invariant, where $g_0 \in \mathrm{O}_Q(\mathbb{R})$ is such that the stabilizer of $g_0^t v_0^t$ is equal to $N^- M$. We fix a small $\epsilon > 0$ and let $\phi_\epsilon \in C_c^\infty(\mathbb{H}^3) = C_c^\infty(G)^K$ be as in the proof of Theorem 7.5 with $G = \mathrm{SL}_2(\mathbb{C})$ and $K = \mathrm{SU}(2)$.

Definition 8.3 (Weight)**.** *We define the smoothed weight to each $\gamma \in \Gamma$,*

$$w_T(\gamma) := \int_{G/K} \chi_{B_T}(\xi\gamma g)\phi_\epsilon(g)\, d\mu(g).$$

Let f be a primitive integral polynomial in 4 variables. Consider the sequence $\mathsf{A}(T) = \{a_n(T)\}$ where

$$a_n(T) := \sum_{\substack{\gamma \in \mathrm{Stab}_\Gamma(\xi)\backslash\Gamma \\ f(\xi\gamma) = n}} w_T(\gamma).$$

Clearly,

$$|\mathsf{A}(T)| = \sum_n a_n(T) = \sum_{\gamma \in \mathrm{Stab}_\Gamma(\xi)\backslash\Gamma} w_T(\gamma)$$

and

$$|\mathsf{A}_d(T)| = \sum_{n \equiv 0 (d)} a_n(T) = \sum_{\substack{\gamma \in \Gamma \\ f(\xi\gamma) \equiv 0 (d)}} w_T(\gamma). \tag{8.4}$$

For any subgroup Γ_0 of Γ with

$$\mathrm{Stab}_\Gamma(\xi) = \mathrm{Stab}_{\Gamma_0}(\xi),$$

we define

$$F_T^{\Gamma_0}(g) := \sum_{\gamma \in \mathrm{Stab}_\Gamma(\xi)\backslash\Gamma_0} \chi_{B_T}(\xi\gamma g),$$

and for each element $\gamma_1 \in \Gamma$, we also define a function on $\Gamma_0\backslash\mathbb{H}^3$ by

$$\Phi^{\Gamma_0}_{\epsilon,\gamma_1}(g) := \sum_{\gamma\in\Gamma_0} \phi_\epsilon(\gamma_1^{-1}\gamma g)$$

which is an ϵ-approximation to the identity about $[\gamma_1^{-1}]$ in $\Gamma_0\backslash\mathbb{H}^3$. The function $\Phi^{\Gamma_0}_{\epsilon,e}$ is simply the lift of ϕ_ϵ to $\Gamma_0\backslash\mathbb{H}^3$ and will be denoted by $\Phi^{\Gamma_0}_\epsilon$.

For $d \in \mathbb{Z}$, let $\Gamma_\xi(d)$ be the subgroup of Γ which stabilizes $\xi \mod d$, i.e.,

$$\Gamma_\xi(d) := \{\gamma \in \Gamma : \xi\gamma \equiv \xi(d)\}.$$

Note that

$$\mathrm{Stab}_\Gamma(\xi) = \mathrm{Stab}_{\Gamma_\xi(d)}(\xi).$$

Lemma 8.5. (1) $|\mathsf{A}(T)| = \left\langle F^\Gamma_T, \Phi^\Gamma_\epsilon \right\rangle_{L^2(\Gamma\backslash\mathbb{H}^3)}$;

(2) *for any integer d,*

$$|\mathsf{A}_d(T)| = \sum_{\substack{\gamma_1\in\Gamma_\xi(d)\backslash\Gamma \\ f(\xi\gamma_1)\equiv 0(d)}} \left\langle F^{\Gamma_\xi(d)}_T, \Phi^{\Gamma_\xi(d)}_{\epsilon,\gamma_1} \right\rangle_{L^2(\Gamma_\xi(d)\backslash\mathbb{H}^3)}.$$

Proof. We have

$$\begin{aligned}
|\mathsf{A}(T)| &= \sum_{\gamma\in\mathrm{Stab}_\Gamma(\xi)\backslash\Gamma} w_T(\gamma) \\
&= \sum_{\gamma\in\mathrm{Stab}_\Gamma(\xi)\backslash\Gamma} \int_{G/K} \chi_{B_T}(\xi\gamma g)\phi_\epsilon(g)d\mu(g) \\
&= \int_{G/K} F^\Gamma_T(g)\phi_\epsilon(g)\, d\mu(g) \\
&= \int_{\Gamma\backslash G/K} F^\Gamma_T(g)\Phi^\Gamma_\epsilon(g)\, d\mu(g) \\
&= \left\langle F^\Gamma_T, \Phi^\Gamma_\epsilon \right\rangle_{L^2(\Gamma\backslash\mathbb{H}^3)}
\end{aligned}$$

Expand (8.4) as

$$|\mathsf{A}_d(T)| = \sum_{\substack{\gamma_1\in\Gamma_\xi(d)\backslash\Gamma \\ f(\xi\gamma_1)\equiv 0(d)}} \sum_{\gamma\in\mathrm{Stab}_\Gamma(\xi)\backslash\Gamma_\xi(d)} w_T(\gamma\gamma_1).$$

The inner sum is

$$\begin{aligned}
\sum_{\gamma\in\mathrm{Stab}_\Gamma(\xi)\backslash\Gamma_\xi(d)} w_T(\gamma\gamma_1) &= \sum_{\gamma\in\mathrm{Stab}_\Gamma(\xi)\backslash\Gamma_\xi(d)} \int_{G/K} \chi_{B_T}(\xi\gamma g)\phi_\epsilon(\gamma_1^{-1}g)\, d\mu(g) \\
&= \int_{G/K} F^{\Gamma_\xi(d)}_T(g)\phi_\epsilon(\gamma_1^{-1}g)\, d\mu(g) \\
&= \int_{\Gamma_\xi(d)\backslash G/K} F^{\Gamma_\xi(d)}_T(g)\Phi^{\Gamma_\xi(d)}_{\epsilon,\gamma_1}(g)\, d\mu(g).
\end{aligned}$$

Thus

$$\sum_{\gamma \in \mathrm{Stab}_\Gamma(\xi)\backslash\Gamma_\xi(d)} w_T(\gamma\gamma_1) = \left\langle F_T^{\Gamma_\xi(d)}, \Phi_{\epsilon,\gamma_1}^{\Gamma_\xi(d)} \right\rangle_{L^2(\Gamma_\xi(d)\backslash\mathbb{H}^3)}.$$

□

Denote by $\mathrm{Spec}(\Gamma\backslash\mathbb{H}^3)$ the spectrum of the Laplace operator on $L^2(\Gamma\backslash\mathbb{H}^3)$. As mentioned before, the work of Sullivan, generalizing Patterson's, implies that $\mathrm{Spec}(\Gamma\backslash\mathbb{H}^3)\cap[0,1) \neq \emptyset$ in which case $\lambda_0 = \alpha(2-\alpha)$ is the base eigenvalue of Δ. For the principal congruence subgroup

$$\Gamma(d) := \{\gamma \in \Gamma : \gamma = I \pmod{d}\}$$

of Γ of level d, we have

$$\mathrm{Spec}(\Gamma\backslash\mathbb{H}^3) \subset \mathrm{Spec}(\Gamma(d)\backslash\mathbb{H}^3).$$

The following is obtained by Bourgain, Gamburd and Sarnak in [4] and [5].

Theorem 8.6. *Let Γ be a Zariski dense subgroup of $\mathrm{Spin}_Q(\mathbb{Z})$ with $\delta_\Gamma > 1$. Then there exist $1 \le \theta < \delta_\Gamma$ such that we have for all square-free integers d,*

$$\mathrm{Spec}(\Gamma(d)\backslash\mathbb{H}^3) \cap [\theta(2-\theta), \delta_\Gamma(2-\delta_\Gamma)] = \{\delta_\Gamma(2-\delta_\Gamma)\}.$$

In order to control the error term for $|\mathsf{A}_d(T)|$, we need a version of Corollary 7.3 uniform over all congruence subgroups $\Gamma_\xi(d)$ of Γ.

Proposition 8.7. *There exists $\epsilon_0 > 0$, uniform over all square-free integers d, such that for any $\gamma_1 \in \Gamma$ and for any congruence subgroup $\Gamma_\xi(d)$ of Γ, we have*

$$\left\langle F_T^{\Gamma_\xi(d)}, \Phi_{\epsilon,\gamma_1}^{\Gamma_\xi(d)} \right\rangle_{L^2(\Gamma_\xi(d)\backslash\mathbb{H}^3)} = \frac{c_{\phi_0}\cdot d_\epsilon}{\delta_\Gamma \cdot [\Gamma : \Gamma_\xi(d)]} \cdot \|\xi\|^\alpha \cdot T^\alpha + O(\epsilon^{-3} T^{\alpha-\epsilon_0})$$

for some $d_\epsilon > 0$ where the implied constant depends only on $\Lambda_N(\Gamma)$.

Proof. Note that the congruence subgroup $\Gamma(d)$ of level d is a finite index subgroup of $\Gamma_\xi(d)$. Since

$$\mathrm{Spec}(\Gamma_\xi(d)\backslash\mathbb{H}^3) \subset \mathrm{Spec}(\Gamma(d)\backslash\mathbb{H}^3)$$

the spectral gap Theorem 8.6 holds for the family $\Gamma_\xi(d)$, d square-free, as well.

As we are assuming $\|\cdot\|$ is $g_0 K g_0^{-1}$-invariant, by Corollary 7.3, we have

$$\langle F_T, \Phi_{\epsilon,\gamma_1}^{\Gamma_\xi(d)}\rangle = \langle \Phi_{\epsilon,\gamma_1}^{\Gamma_\xi(d)}, \phi_0^{\Gamma_\xi(d)}\rangle \cdot \delta_{\Gamma_\xi(d)}^{-1} \cdot c_{\phi_0^{\Gamma_\xi(d)}} \cdot T^{\delta_\Gamma} \cdot \|\xi\|^\alpha \cdot (1 + O(T^{-\epsilon_0})), \tag{8.8}$$

where ϵ_0 depends only on the spectral gap for $L^2(\Gamma_\xi(d)\backslash\mathbb{H}^3)$ and the implied constant depends only on

$$\Lambda_N(g_0^{-1}\Gamma_\xi(d)g_0) = \{[n_x] \in (N \cap g_0^{-1}\Gamma_\xi(d)g_0)\backslash N : x \in \Lambda(g_0^{-1}\Gamma_\xi(d)g_0)\}.$$

As $\Gamma_\xi(d)$ is a subgroup of finite index in Γ, $\delta_{\Gamma_\xi(d)} = \delta_\Gamma$ and

$$\Lambda(g_0^{-1}\Gamma_\xi(d)g_0) = \Lambda(g_0^{-1}\Gamma g_0),$$

and moreover by the definition of $\Gamma_\xi(d)$, we have

$$\mathrm{Stab}_\Gamma(\xi) = \mathrm{Stab}_{\Gamma_\xi(d)}(\xi),$$

implying

$$N \cap g_0^{-1}\Gamma_\xi(d) g_0 = N \cap g_0^{-1}\Gamma g_0.$$

Hence

$$\Lambda_N(g_0^{-1}\Gamma_\xi(d) g_0) = \Lambda_N(g_0^{-1}\Gamma g_0),$$

yielding that the implied constant in (8.8) is independent of d.

By Theorem 8.6, ϵ_0 can be taken to be uniform over all d. If $\Gamma_0 < \Gamma$ is a subgroup of finite index, the base eigenfunction in $L^2(\Gamma_0\backslash\mathbb{H}^3)$ with the unit L^2-norm is given by

$$\phi_0^{\Gamma_0} := \frac{1}{\sqrt{[\Gamma : \Gamma_0]}}\phi_0$$

with $\phi_0 = \phi_0^\Gamma$. Therefore

$$\langle \Phi_{\epsilon,\gamma_1}^{\Gamma_\xi(d)}, \phi_0^{\Gamma_\xi(d)}\rangle \cdot c_{\phi_0^{\Gamma_\xi(d)}} = \langle \Phi_{\epsilon,e}, \phi_0\rangle \cdot c_{\phi_0} \cdot \frac{1}{[\Gamma : \Gamma_\xi(d)]}.$$

This finishes the proof as $\delta_\Gamma = \alpha$, by setting $d_\epsilon = \langle \Phi_{\epsilon,e}, \phi_0\rangle$. □

Setting

$$\mathcal{X} = \delta_\Gamma^{-1} \cdot c_{\phi_0} \cdot d_\epsilon \cdot \|\xi\|^\alpha \cdot T^\alpha,$$

the following corollary is immediate from Lemma 8.5 and Proposition 8.7.

Corollary 8.9. *There exists $\epsilon_0 > 0$ uniform over all square-free integers d such that*

$$|\mathsf{A}_d(T)| = \mathcal{O}_f^0(d)\left(\frac{1}{[\Gamma : \Gamma_\xi(d)]}\mathcal{X} + O(\mathcal{X}^{1-\epsilon_0})\right),$$

where

$$\mathcal{O}_f^0(d) = \sum_{\substack{\gamma \in \Gamma_\xi(d)\backslash\Gamma \\ f(\xi\gamma)\equiv 0(d)}} 1.$$

8.3. **Proof of Theorem 1.4.** We set

$$f_1(x_1, x_2, x_3, x_4) = x_1 \quad \text{and} \quad f_2(x_1, x_2, x_3, x_4) = x_1 x_2.$$

For d square-free and $i = 1, 2$, set

$$g_i(d) = \mathcal{O}_{f_i}^0(d)/[\Gamma : \Gamma_\xi(d)].$$

Proposition 8.10. *There exists a finite set S of primes such that:*

(1) *for any square-free integer $d = d_1 d_2$ with no prime factors in S and for each $i = 1, 2$,*

$$g_i(d_1 d_2) = g_i(d_1) \cdot g_i(d_2);$$

(2) *for any prime p outside S,*

$$g_1(p) \in (0, 1) \quad \textit{and} \quad g_1(p) = p^{-1} + O(p^{-3/2}).$$

(3) *for any prime p outside S,*

$$g_2(p) \in (0,1) \quad \textit{and} \quad g_2(p) = 2p^{-1} + O(p^{-3/2}).$$

Proof. According to the theorem of Matthews, Vaserstein and Weisfeiler [27], there exists a finite set of primes S so that

- for all primes p outside S, Γ projects onto $G(\mathbb{F}_p)$
- for $d = p_1 \cdots p_t$ square-free with $p_i \notin S$, the diagonal reduction

$$\Gamma \to G(\mathbb{Z}/d\mathbb{Z}) \to G(\mathbb{F}_{p_1}) \times \cdots G(\mathbb{F}_{p_t})$$

  is surjective.

Enlarge S so that $G(\mathbb{F}_p)$'s have no common composition factors for different p's outside S. This is possible because $G = \mathrm{Spin}(Q)$ can be realized as SL_2 over $\mathbb{Q}[\sqrt{-1}]$. Hence there exists a finite set S of primes such that for p outside S,

$$G(\mathbb{F}_p) = \begin{cases} \mathrm{SL}_2(\mathbb{F}_p) \times \mathrm{SL}_2(\mathbb{F}_p) & \text{for } p \equiv 1(4) \\ \mathrm{SL}_2(\mathbb{F}_{p^2}) & \text{for } p \equiv 3(4). \end{cases} \tag{8.11}$$

It then follows from Goursat's lemma [23, p.75] that Γ surjects onto $G(\mathbb{Z}/d_1\mathbb{Z}) \times G(\mathbb{Z}/d_2\mathbb{Z})$ for any square-free d_1 and d_2 with no prime factors in S. This implies that for $d = d_1 d_2$ square-free and without any prime factors in S, the orbit of ξ mod d, say $\mathcal{O}(d)$, is equal to $\mathcal{O}(d_1) \times \mathcal{O}(d_2)$ in $(\mathbb{Z}/d_1\mathbb{Z})^4 \times (\mathbb{Z}/d_2\mathbb{Z})^4 = (\mathbb{Z}/d\mathbb{Z})^4$. It also follows that $\mathcal{O}^0(d)$ is equal to $\mathcal{O}^0(d_1) \times \mathcal{O}^0(d_2)$. Therefore $g(d) = g(d_1)g(d_2)$ as desired.

Denote by V the cone defined by $Q = 0$ minus the origin, i.e.,

$$V = \{(x_1, x_2, x_3, x_4) \neq 0 : \sum_{i=1}^4 2x_i^2 - (\sum_{i=1}^4 x_i)^2 = 0\}.$$

Note that

$$W_1 := \{x \in V : f_1(x) = 0\} = \{(0, x_2, x_3, x_4) \neq 0 : \sum_{i=2}^4 2x_i^2 - (\sum_{i=2}^4 x_i)^2 = 0\}.$$

Since both quadratic forms

$$Q(x_1, x_2, x_3, x_4) = \sum_{i=1}^4 2x_i^2 - (\sum_{i=1}^4 x_i)^2 \text{ and } Q(0, x_2, x_3, x_4) = \sum_{i=2}^4 2x_i^2 - (\sum_{i=2}^4 x_i)^2$$

are absolutely irreducible, we have by [3, Thm. 1.2.B],

$$\#V(\mathbb{F}_p) = p^3 + O(p^{5/2}), \quad \#W_1(\mathbb{F}_p) = p^2 + O(p^{3/2}).$$

Since V is a homogeneous space of G with a connected stabilizer, by [35, Prop 3.22],

$$\mathcal{O}(p) = V(\mathbb{F}_p), \text{ and hence } \mathcal{O}^0_{f_1}(p) = W_1(\mathbb{F}_p).$$

Therefore we deduce $g_1(p) = \frac{\#\mathcal{O}^0_{f_1}(p)}{\#\mathcal{O}(p)} = p^{-1} + O(p^{-3/2})$, proving (1).

Now $W_2 := \{x \in V : f_2(x) = 0\}$ is the union of two quadrics given by $V \cap \{x_1 = 0\}$ and $V \cap \{x_2 = 0\}$. Hence

$$\#W_2(\mathbb{F}_p) = 2p^2 + O(p^{3/2}).$$

This yields that $g(p) = 2p^{-1} + O(p^{-3/2})$. $\square$

We are now ready to prove Theorem 1.4. First consider $f_1(x_1, x_2, x_3, x_4) = x_1$ so that $\mathsf{A}(T) = \{a_n(T)\}$ where

$$a_n(T) := \sum_{\substack{\gamma \in \mathrm{Stab}_\Gamma(\xi)\backslash\Gamma \\ f_1(\xi\gamma) = n}} w_T(\gamma)$$

is a smoothed count for the number of vectors (x_1, x_2, x_3, x_4) in the orbit $\xi\mathcal{A}^t$ of max norm bounded above by T and $x_1 = n$.

By Lemma 8.9 and as $\#\mathcal{O}^0_{f_1}(d)$ is multiplicative with $\#\mathcal{O}^0_{f_1}(p) = p^2 + O(p^{3/2})$, the quantity $r_d(\mathsf{A})$ in the decomposition (8.2) satisfies for some $\epsilon_0 > 0$,

$$r_d(\mathsf{A}) \ll d^2 T^{\alpha - \epsilon_0}.$$

Thus for any $\varepsilon_1 > 0$,

$$\sum_{d < D, d|P} \tau_3(d)|r_d(\mathsf{A})| \ll_{\varepsilon_1} D^{3+\varepsilon_1} T^{\alpha - \epsilon_0},$$

which is $\ll T^\alpha / \log T$ for $D = T^{\epsilon_0/4}$, say. The key here is that D can be taken as large as a fixed power of T. Let P be the product of all primes $p < D = T^{\epsilon_0/4}$ outside of the bad set S.

As h is a multiplicative function defined by $h(p) = \frac{g_1(p)}{1 - g_1(p)}$ for p primes with g_1 in Proposition 8.10, we deduce that (cf. [19, 6.6])

$$\sum_{d < \sqrt{D}, d|P} h(d) \gg \log D \gg \log T,$$

and Theorem 8.1 gives

$$S(\mathsf{A}(T), P) \ll T^\alpha / \log T.$$

Therefore

$$\begin{aligned} &\#\{(x_1, x_2, x_3, x_4) \in \xi\mathcal{A}^t : \max_{1 \le i \le 4} |x_i| < T,\ (x_1, \prod_{p < T^{\epsilon_0/4}} p) = 1\} \\ &\ll S(\mathsf{A}((1+\epsilon)T), P) \ll T^\alpha / \log T. \end{aligned}$$

Hence

$$\#\{(x_1, x_2, x_3, x_4) \in \xi\mathcal{A}^t : \max_{1 \le i \le 4} |x_i| < T,\ x_1 : \text{prime}\} \ll T^\alpha / \log T.$$

Since this argument is symmetric in the x_i's, we have

$$\#\{(x_1, x_2, x_3, x_4) \in \xi\mathcal{A}^t : \max_{1 \le i \le 4} |x_i| = \text{prime at most T}\} \ll T^\alpha / \log T.$$

By Lemma 2.5, this proves

$$\pi^{\mathcal{P}}(T) \ll \frac{T^\alpha}{\log T}.$$

In order to prove

$$\pi_2^{\mathcal{P}}(T) \ll \frac{T^\alpha}{(\log T)^2}, \tag{8.12}$$

we proceed the same way with the polynomial $f_2(x_1, x_2, x_3, x_4) = x_1 x_2$ and with the sequence $\mathsf{A}(T) = \{a_n(T)\}$ where

$$a_n(T) := \sum_{\substack{\gamma \in \mathrm{Stab}_\Gamma(\xi)\backslash\Gamma \\ f_2(\xi\gamma)=n}} w_T(\gamma).$$

is a smoothed count for the number of vectors (x_1, x_2, x_3, x_4) in the orbit $\xi\mathcal{A}^t$ of max norm bounded above by T and $x_1 x_2 = n$.

Note that $g_2(p) = 2p^{-1} + O(p^{-3/2})$ by Proposition 8.10, and that

$$\sum_{d<\sqrt{D}, d|P} h(d) \gg (\log D)^2 \gg (\log T)^2$$

for $h(p) = \frac{g_2(p)}{1-g_2(p)}$ for p primes.

Therefore Theorem 8.1 gives

$$S(\mathsf{A}(T), P) \ll T^\alpha/(\log T)^2$$

which implies

$$\#\{(x_1, x_2, x_3, x_4) \in \xi\mathcal{A}^t : \max_{1\le i\le 4} |x_i| < T, x_1, x_2 : \text{ primes}\} \ll T^\alpha/(\log T)^2.$$

Again by the symmetric property of x_i's, we have

$$\#\{x \in \xi\mathcal{A}^t : \max_{1\le i\le 4} |x_i| < T, x_i, x_j : \text{primes for some } i \neq j\} \ll T^\alpha/(\log T)^2.$$

By Lemma 2.5, this proves

$$\pi_2^{\mathcal{P}}(T) \ll T^\alpha/(\log T)^2.$$

A. Appendix: Non-accumulation of expanding closed horospheres on singular tubes (by Hee Oh and Nimish Shah)

In this appendix, we deduce Theorem A.1 from Theorem 6.1: Recall that $\Gamma < G = \mathrm{PSL}_2(\mathbb{C})$ is a torsion free discrete geometrically finite group with $\delta_\Gamma > 1$ and that $\phi_0 \in L^2(\Gamma\backslash\mathbb{H}^3)$ denotes the positive base eigenfunction of Δ of eigenvalue $\delta_\Gamma(2-\delta_\Gamma)$ and of norm $\int_{\Gamma\backslash\mathbb{H}^3} \phi_0^2(g)\, d\mu(g) = 1$.

We continue the notations N, a_y, N^-, etc., from section 6. We assume that $\Gamma\backslash\Gamma N$ is closed. By Corollary 4.13, for some $c_{\phi_0} > 0$ and $d_{\phi_0} \ge 0$,

$$\int_{\Gamma\backslash\Gamma N} \phi_0(na_y)dn = c_{\phi_0} y^{2-\delta_\Gamma} + d_{\phi_0} y^{\delta_\Gamma}.$$

We also recall the Burger-Roblin measure $\hat{\mu}$ defined in Theorem 6.10 which is normalized so that $\hat{\mu}(\phi_0) = 1$.

Theorem A.1. *For $\psi \in C_c(\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3))$,*

$$\int_{n_x \in (N\cap\Gamma)\backslash N} \psi(n_x a_y)\, dx \sim c_{\phi_0} \cdot \hat{\mu}(\psi) \cdot y^{2-\delta_\Gamma} \quad \textit{as } y \to 0.$$

Proof. The idea of proof is motivated by the proof of Lemma 2.1 in [37]. For each $0 < y \le 1$, define the measure μ_y on $\Gamma\backslash G/M = \mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ by

$$\mu_y(\psi) = c_{\phi_0}^{-1} y^{-2+\delta_\Gamma} \int_{(N\cap\Gamma)\backslash N} \psi(\Gamma\backslash\Gamma[n_x a_y])\, dx$$

for $\psi \in C_c(\Gamma\backslash G/M)$.

Consider the family

$$\mathcal{M} := \{\mu_y : 0 < y < 1\}.$$

We claim that $\mathcal{M}$ is relatively compact in the set of locally finite Borel measures on $\Gamma\backslash G/M$ with respect to the weak*-topology. For any compact subset $C \subset \Gamma\backslash G/M$, let ψ be a K-invariant smooth non-negative function of compact support which is one over C. Then

$$\mu_y(C) \le \mu_y(\psi).$$

As $\mu_y(\psi) \to \hat{\mu}(\psi) = \langle \psi, \phi_0 \rangle$ by Theorem 6.1, the claim follows.

It now suffices to show that every accumulation point of $\mathcal{M}$ is equal to $\hat{\mu}$. Let μ_0 be an accumulation point of $\mathcal{M}$, which is clearly N-invariant.

We denote by $\mathcal{E}_P$ the set of vectors $v \in \mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ the horospheres determined by which is closed. In the identification of $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^3)$ with $\Gamma\backslash G/M$, the set $\mathcal{E}_P$ corresponds to the image under the projection $\Gamma\backslash G \to \Gamma\backslash G/M$ of the sets $\Gamma\backslash\Gamma gNA$ for $\Gamma\backslash\Gamma gN$ closed.

Fix a bounded parabolic fixed point ξ_0 of Γ. If $\xi_0 = g_0(\infty)$ for $g_0 \in \mathrm{PSL}_2(\mathbb{C})$, then a cusp, say, $D(\xi_0)$, at ξ_0 is the image of $\cup_{y>y_1} g_0 N a_y$ for some $y_1 \gg 1$ under the projection $\pi : G \to \Gamma\backslash G/M$.

There exists $c_0 > 1$ such that for any $z = \Gamma\backslash\Gamma g_0 n a_y \in D(\xi_0)$,

$$c_0^{-1} y^{r-\delta} \le \phi_0(z) \le c_0 y^{r-\delta} \tag{A.2}$$

where $r \in \{1,2\}$ is the rank of ξ_0 (see [46, Sec. 5] as well as the proof of [8, Lem. 3.5]).

Noting that $D_0 = D_0(\xi_0) \subset \mathcal{E}_P$, we first claim that for any weak limit, say, μ_0, of $\mathcal{M}$,

$$\mu_0(D_0) = 0. \tag{A.3}$$

Let Q be a relatively compact open subset of D_0. For $y_\epsilon > 1$, setting $D_\epsilon := D_0 a_{y_\epsilon} \subset D_0$, we have that

$$\int_{D_\epsilon K} \phi_0^2 \, d\mu(g) \ll y_\epsilon^{r-2\delta}. \tag{A.4}$$

For the part of $D_\epsilon K$ inside the unit neighborhood of the convex core of Γ, this estimate follows from the proof of [8, Lem. 4.2]. When $r = 1$, the integral of ϕ_0^2 over the part of $D_\epsilon K$ outside the unit neighborhood of the convex core of Γ is comparable to

$$\int_{\log y_\epsilon}^{\infty} \int_{x=e^t}^{\infty} e^{-2\delta t} dx dt \asymp y_\epsilon^{1-2\delta}.$$

Fixing a neighborhood U of e in $N^- M$ such that $D_0 U \subset D_0 K$, the set $D_\epsilon U a_{y_\epsilon}^{-1}$ is a neighborhood of Q. Therefore if μ_{y_i} weakly converges to μ_0,

$$\begin{aligned} c_{\phi_0} \cdot \mu_0(\phi_0 \cdot \chi_Q) &\le \lim_{y_i \to 0} \frac{1}{y_i^{2-\delta}} \int_{\Gamma \backslash \Gamma N} (\phi_0 \cdot \chi_{D_\epsilon U a_{y_\epsilon}^{-1}})(n a_{y_i}) \, dn \\ &= \lim_{y_i \to 0} \frac{1}{y_i^{2-\delta}} \int_{\Gamma \backslash \Gamma N} \phi_0(n a_{y_i}) \cdot \chi_{D_\epsilon U}(n a_{y_i} a_{y_\epsilon}) \, dn. \end{aligned}$$

As $y_\epsilon > 1$ and $U \subset N^- M$,

$$a_{y_\epsilon} U a_{y_\epsilon^{-1}} \subset U.$$

Hence if $n a_{y y_\epsilon} \in D_\epsilon U$, then

$$n a_y \in D_\epsilon a_{y_\epsilon^{-1}} (a_{y_\epsilon} U a_{y_\epsilon^{-1}}) \subset D_\epsilon a_{y_\epsilon^{-1}} U \subset D_0 U.$$

Moreover if $n a_{y y_\epsilon} \in \Gamma g_0 n' a_{y'} M U$ for some $n' \in N$ and $y' > y_1$,

$$n a_y \in \Gamma g_0 n' a_{y'} a_{y_\epsilon}^{-1} (a_{y_\epsilon} M U a_{y_\epsilon}^{-1}) \subset g_0 n' a_{y' y_\epsilon^{-1}} M U. \tag{A.5}$$

Using the formula

$$\phi_0(g) = \int_{\xi \in \Lambda(\Gamma)} e^{-\delta \beta_\xi(go,o)} d\nu_0(\xi)$$

for $g \in G$ (see (4.2)), it is easy to check that

$$e^{-\delta d(uo,o)} \phi_0(g) \le \phi_0(gu) \le e^{\delta d(uo,o)} \phi_0(g)$$

for any $g, u \in G$. Therefore we can deduce from (A.2) and (A.5) that for some constant $c' > 1$,

$$\phi_0(n a_y) \le c' \cdot y_\epsilon^{\delta - r} \cdot \phi_0(n a_y a_{y_\epsilon})$$

for all $na_y a_{y_\epsilon} \in D_\epsilon U$. Hence

$$\begin{aligned}
&c_{\phi_0} \cdot \mu_0(\phi_0 \cdot \chi_Q) \\
&\le c_{\phi_0} c' {y_\epsilon}^{2-r} \lim_{y_i \to 0} \frac{1}{(y_i y_\epsilon)^{2-\delta}} \int_{\Gamma\backslash\Gamma N} (\phi_0 \cdot \chi_{D_\epsilon U})(\Gamma\backslash\Gamma n a_{y_i} a_{y_\epsilon})\, dn \\
&\le c_{\phi_0} c' y_\epsilon^{2-r} \int_{\Gamma\backslash\mathbb{H}^3} (\phi_0^2 \cdot \chi_{D_\epsilon K})(g)\, d\mu(g) \quad \text{by Theorem 6.1} \\
&\ll y_\epsilon^{2-2\delta} \quad \text{by (A.4).}
\end{aligned}$$

Therefore

$$\mu_0(\phi_0 \cdot \chi_Q) \ll y_\epsilon^{2-2\delta}.$$

As $\delta > 1$, $y_\epsilon > 1$ is arbitrary, and $\min_{g\in Q} \phi_0(g) > 0$, we have $\mu_0(Q) = 0$. This proves the claim (A.3).

We now claim that $\mu_0(\mathcal{E}_P) = 0$ for any weak limit μ_0: $\mu_{y_i} \to \mu_0$. Suppose not. Since there are only finitely many cusps, there exist relatively compact open subset $Q \subset \mathcal{E}_P$ and a bounded parabolic fixed point $\xi_0 \in \Lambda(\Gamma)$ such that $\mu_0(Q) > 0$ and its image $a_y(Q) = Qa_y$ under the geodesic flow converges to ξ_0 as $y \to \infty$. Fix $y_0 > 1$ such that

$$Qa_{y_0} \subset D_0(\xi_0).$$

By passing to a subsequence, we may assume $\mu_{y_i y_0}$ is convergent with a weak limit, say, μ_0'. Since $Qa_{y_0} \subset D_0(\xi_0)$, by (A.3), we have

$$\mu_0'(Qa_{y_0}) = 0.$$

Therefore for any $\epsilon > 0$, there exists a neighborhood $U_\epsilon \subset N^- M$ of e such that

$$\mu_0'(Qa_{y_0} U_\epsilon) < {y_0}^{-2+\delta}\epsilon. \tag{A.6}$$

Noting that $Qa_{y_0} U_\epsilon a_{y_0}^{-1}$ is a neighborhood of Q, we have

$$\begin{aligned}
c_{\phi_0} \cdot \mu_0(Q) &\le \lim_{y_i \to 0} \frac{1}{y_i^{2-\delta}} \int_{\Gamma\backslash\Gamma N} \chi_{Qa_{y_0} U_\epsilon a_{y_0}^{-1}}(n a_{y_i}) dn \\
&= y_0^{2-\delta} \lim_{y_i \to 0} \frac{1}{(y_0 y_i)^{2-\delta}} \int_{\Gamma\backslash\Gamma N} \chi_{Qa_{y_0} U_\epsilon}(n a_{y_i} a_{y_0}) dn \\
&= y_0^{2-\delta} \mu_0'(Qa_{y_0} U_\epsilon) \\
&\le \epsilon \quad \text{by (A.6).}
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, $\mu_0(Q) = 0$. This proves

$$\mu_0(\mathcal{E}_P) = 0.$$

We deduce from Theorem 6.9 that

$$\mu_0 = \alpha_1 \hat{\mu}$$

for some $\alpha_1 \ge 0$. On the other hand, by Theorem 6.1,

$$\mu_0(\phi_0) = \hat{\mu}(\phi_0) = 1.$$

It follows that $\alpha_1 = 1$. □

References

[1] Thierry Aubin. *Nonlinear analysis on manifolds. Monge-Ampère equations*, volume 252 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982.

[2] Alan F. Beardon and Bernard Maskit. Limit points of Kleinian groups and finite sided fundamental polyhedra. *Acta Math.*, 132:1–12, 1974.

[3] A. I. Borevich and I. R. Shafarevich. *Number theory.* Translated from the Russian by Newcomb Greenleaf. Pure and Applied Mathematics, Vol. 20. Academic Press, New York, 1966.

[4] Jean Bourgain, Alex Gamburd, and Peter Sarnak. Affine linear sieve, expanders and sum-product, 2008. To appear in Inventiones.

[5] Jean Bourgain, Alex Gamburd, and Peter Sarnak. Generalization of Selberg's theorem and Selberg's sieve, 2009. Preprint.

[6] David W. Boyd. The sequence of radii of the Apollonian packing. *Math. Comp.*, 39(159):249–254, 1982.

[7] Marc Burger. Horocycle flow on geometrically finite surfaces. *Duke Math. J.*, 61(3):779–803, 1990.

[8] Kevin Corlette and Alessandra Iozzi. Limit sets of discrete groups of isometries of exotic hyperbolic spaces. *Trans. Amer. Math. Soc.*, 351(4):1507–1530, 1999.

[9] H. S. M. Coxeter. The problem of Apollonius. *Amer. Math. Monthly*, 75:5–15, 1968.

[10] F. Dal'bo. Topologie du feuilletage fortement stable. *Ann. Inst. Fourier (Grenoble)*, 50(3):981–993, 2000.

[11] S. G. Dani. Invariant measures and minimal sets of horospherical flows. *Invent. Math.*, 64(2):357–385, 1981.

[12] Alex Eskin and C. T. McMullen. Mixing, counting, and equidistribution in Lie groups. *Duke Math. J.*, 71(1):181–209, 1993.

[13] E. Fuchs. 2010. Ph. D. Thesis, Princeton University.

[14] Ramesh Gangolli and V. S. Varadarajan. *Harmonic analysis of spherical functions on real reductive groups*, volume 101 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin, 1988.

[15] Ronald L. Graham, Jeffrey C. Lagarias, Colin L. Mallows, Allan R. Wilks, and Catherine H. Yan. Apollonian circle packings: number theory. *J. Number Theory*, 100(1):1–45, 2003.

[16] Ronald L. Graham, Jeffrey C. Lagarias, Colin L. Mallows, Allan R. Wilks, and Catherine H. Yan. Apollonian circle packings: geometry and group theory. I. The Apollonian group. *Discrete Comput. Geom.*, 34(4):547–585, 2005.

[17] Y. Guivarc'h and A. Raugi. Products of random matrices: convergence theorems. In *Random matrices and their applications (Brunswick, Maine, 1984)*, volume 50 of *Contemp. Math.*, pages 31–54. Amer. Math. Soc., Providence, RI, 1986.

[18] K. E. Hirst. The Apollonian packing of circles. *J. London Math. Soc.*, 42:281–291, 1967.

[19] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[20] Michael Kapovich. *Hyperbolic manifolds and discrete groups*, volume 183 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 2001.

[21] Alex Kontorovich. The hyperbolic lattice point count in infinite volume with applications to sieves. *Duke Math. J.*, 149(1):1–36, 2009.

[22] Alex Kontorovich and Hee Oh. Almost prime Pythagorean triples in thin orbits. *Preprint, 2009*.

[23] Serge Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2002.

[24] Peter D. Lax and Ralph S. Phillips. The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces. *J. Funct. Anal.*, 46(3):280–350, 1982.

[25] Gregory Margulis. *On some aspects of the theory of Anosov systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2004. With a survey by Richard Sharp: Periodic orbits of hyperbolic flows, Translated from the Russian by Valentina Vladimirovna Szulikowska.

[26] Bernard Maskit. *Kleinian groups*, volume 287 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1988.

[27] C. Matthews, L. Vaserstein, and B. Weisfeiler. Congruence properties of Zariski-dense subgroups. *Proc. London Math. Soc*, 48:514–532, 1984.

[28] F. Maucourant. Homogeneous asymptotic limits of Haar measures of semisimple linear groups and their lattices. *Duke Math. J.*, 136(2):357–399, 2007.

[29] C. T. McMullen. Hausdorff dimension and conformal dynamics. III. Computation of dimension. *Amer. J. Math.*, 120(4):691–721, 1998.

[30] Amir Mohammadi and Alireza Salehi Golsefidy. Translates of horospherical measures and counting problems. *Preprint*, 2008.

[31] Hee Oh. Dynamics on Geometrically finite hyperbolic manifolds with applications to Apollonian circle packings and beyond. *Proc. of ICM (Hyderabad, 2010)*.

[32] Hee Oh and Nimish Shah. Equidistribution and counting for orbits of geometrically finite hyperbolic groups. *Preprint*.

[33] Hee Oh and Nimish Shah. The asymptotic distribution of circles in the orbits of Kleinian groups. *Preprint*.

[34] S.J. Patterson. The limit set of a Fuchsian group. *Acta Mathematica*, 136:241–273, 1976.

[35] Vladimir Platonov and Andrei Rapinchuk. *Algebraic groups and number theory*, volume 139 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA, 1994. Translated from the 1991 Russian original by Rachel Rowen.

[36] Marina Ratner. On Raghunathan's measure conjecture. *Ann. of Math. (2)*, 134(3):545–607, 1991.

[37] Marina Ratner. Raghunathan's topological conjecture and distributions of unipotent flows. *Duke Math. J.*, 63(1):235–280, 1991.

[38] Thomas Roblin. Ergodicité et équidistribution en courbure négative. *Mém. Soc. Math. Fr. (N.S.)*, (95):vi+96, 2003.

[39] Peter Sarnak. Asymptotic behavior of periodic orbits of the horocycle flow and eisenstein series. *Comm. Pure Appl. Math.*, 34(6):719–739, 1981.

[40] Peter Sarnak. Letter to J. Lagarias, 2007. available at www.math.princeton.edu/ sarnak.

[41] Barbara Schapira. Equidistribution of the horocycles of a geometrically finite surface. *Int. Math. Res. Not.*, (40):2447–2471, 2005.

[42] Yehuda Shalom. Rigidity, unitary representations of semisimple groups, and fundamental groups of manifolds with rank one transformation group. *Ann. of Math. (2)*, 152(1):113–182, 2000.

[43] F. Soddy. The bowl of integers and the hexlet. *Nature*, 139:77–79, 1937.

[44] F. Soddy. The kiss precise. *Nature*, 137:1021, 1937.

[45] Dennis Sullivan. The density at infinity of a discrete group of hyperbolic motions. *Inst. Hautes Études Sci. Publ. Math.*, (50):171–202, 1979.

[46] Dennis Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Math.*, 153(3-4):259–277, 1984.

[47] William Thurston. *The Geometry and Topology of Three-Manifolds.* available at www.msri.org/publications/books. Electronic version-March 2002.
[48] J. B. Wilker. Sizing up a solid packing. *Period. Math. Hungar.*, 8(2):117–134, 1977.

Mathematics department, Brown University, Providence, RI
E-mail address: alexk@math.brown.edu

Mathematics department, Brown university, Providence, RI and Korea Institute for Advanced Study, Seoul, Korea
E-mail address: heeoh@math.brown.edu

Department of Mathematics, The Ohio State University, Columbus, OH
E-mail address: shah@math.ohio-state.edu