Math. Proc. Camb. Phil. Soc. (2023), **174**, 569–584
doi:10.1017/S0305004122000421
First published online 15 November 2022

Curves on the torus intersecting at most k times

BY TARIK AOUGAB
Department of Mathematics, Hilles Hall 2nd floor
Haverford College, Haverford, PA, 19041*, U.S.A.*
e-mail: taougab@haverford.edu

AND JONAH GASTER
*Department of Mathematical Sciences, EMS Building E*403
University of Wisconsin-Milwaukee, Milwaukee, WI, 53211*, U.S.A.*
e-mail: gaster@uwm.edu

(*Received* 10 *January* 2022*; accepted* 03 *August* 2022)

Abstract

We show that any set of distinct homotopy classes of simple closed curves on the torus that pairwise intersect at most k times has size $k + O(\sqrt{k}\log k)$. Prior to this work, a lemma of Agol, together with the state of the art bounds for the size of prime gaps, implied the error term $O(k^{21/40})$, and in fact the assumption of the Riemann hypothesis improved this error term to the one we obtain $O(\sqrt{k}\log k)$. By contrast, our methods are elementary, combinatorial, and geometric.

2020 Mathematics Subject Classification: 20F65 (Primary); 57M15, 05D99 (Secondary)

1. *Introduction*

Let $T \approx \mathbb{R}^2/\mathbb{Z}^2$ be the closed oriented surface of genus one. We indicate the homotopy class of an embedding of S^1 briefly by 'curve'. By pulling a curve tight and lifting it to the universal cover, the collection of curves on T is in one-to-one correspondence with slopes $\mathbb{Q} \cup \{\infty\}$. From this vantage point, the *intersection number* of a pair of curves on T (that is, the minimum possible number of intersection points among representatives from the pair of homotopy classes) can be computed explicitly via

$$\iota\left(\frac{p}{q}, \frac{a}{b}\right) = |\, pb - qa \,|\,.$$

A collection of curves is called a *k-system* when any pair of curves has intersection number at most k.

Let $\eta_S(k)$ equal the maximum size of a k-system on the closed surface S. It was first shown by [**20**] that $\eta_S(k)$ goes to infinity with k. The determination of the growth rate of $\eta_S(k)$, as a function of both k and the genus g of S, is a subtle counting problem, about which much remains unknown [**1, 5, 16, 17, 24**]. Notably, Greene has used probabilistic methods, leveraging the hyperbolic geometric bounds of Przytycki, to obtain $\eta_S(k) = O(g^{k+1}\log g)$, when k is fixed and g grows [**16**, Theorem 3].

© The Author(s), 2022. Published by Cambridge University Press on behalf of Cambridge Philosophical Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

For the study of $\eta_S(k)$ with g fixed, the simplest nontrivial case is evidently $S = T$. A simple mod p counting argument, as far as we know first observed by Agol, demonstrates that $\eta_T(k)$ is at most one more than the smallest prime greater than k [4]. Together with the elementary observation $\eta_T(k) \geqslant k+2$ and the Prime Number Theorem, this implies $\eta_T(k)/k \to 1$ as $k \to \infty$ [3].

More can be said. The size of prime gaps, large and small, is a major field of study. The currently best upper bound is due to Baker–Harman–Pintz, which, together with Agol's observation, implies that $\eta_T(k) = k + O(k^{21/40})$ [7]. Cramér showed that a positive resolution of the Riemann hypothesis would provide $\eta_T(k) = k + O(\sqrt{k}\log k)$ [10], and he formulated a stronger conjecture that would imply $\eta_T(k) = k + O((\log k)^2)$ [11]; although there seems to be general suspicion in the analytic number theory community that Cramér's error term should be replaced by $O((\log k)^{2+\epsilon})$ [15, 22].

All of these estimates pass through Agol's remarkable prime number bound, but it is reasonable to be skeptical about whether estimation of $\eta_T(k)$ should depend on such notoriously subtle and difficult questions. The purpose of this paper is to sharpen currently available estimates, without reference to fine data about the distribution of the primes. Our methods are elementary, combinatorial and hyperbolic.

THEOREM 1·1. *There is a constant $C > 0$ so that $\eta_T(k) \leqslant k + C\sqrt{k}\log k$.*

As for sharpness, it deserves remarking that there is a dearth of nontrivial lower bounds for $\eta_T(k)$. In fact, we are unaware of any example of a k-system of size $k+7$ on the torus (cf. [14, p. 116]).

The function $\eta_T(k)$ admits a dual formulation: let $\kappa_S(n)$ indicate the minimum, taken over collections of n curves on S, of the maximum pairwise intersection. (For clarity, we will often use 'n' to indicate the size of a set of curves and 'k' for an intersection number.) It is not hard to see that

$$\eta_S(k) = \max\{n : \kappa_S(n) \leqslant k\}\ , \text{ and } \ \kappa_S(n) = \min\{k : \eta_S(k) \geqslant n\}.$$

Our path towards Theorem 1·1 will be to first estimate $\kappa_T(n)$ from below.

There is a kind of convexity to exploit in the study of κ_T, originally observed by Agol. As remarked above, curves on T are in correspondence with slopes $\mathbb{Q} \cup \{\infty\}$. The latter form the vertices of the *Farey complex* $\mathcal{F}$, in which a set of slopes form a simplex when they pairwise intersect once. Any collection of n curves that is maximal with respect to inclusion among k-systems determines a collection of vertices in $\mathcal{F}$ so that the induced simplicial complex is a triangulated n-gon (see Lemma 2·4 for detail).

Conversely, any triangulation of an n-gon can be realised as a subcomplex of $\mathcal{F}$, in a way that is unique up to the action of $\mathrm{PSL}(2,\mathbb{Z}) \curvearrowright \mathcal{F}$ by simplicial automorphism. As $\mathrm{PSL}(2,\mathbb{Z})$ preserves the intersection form, the (multi-)set of pairwise intersection numbers is a well-defined function on the set of triangulations of an n-gon. The set of triangulations of an n-gon forms the vertex set of a well-studied simplicial complex in combinatorics called the *associahedron* $\mathcal{A}_n$ (there is a slight indexing issue; the object we refer to as $\mathcal{A}_n$ is the $(n-2)$-dimensional associahedron). One obtains a 'max intersection' function $\kappa : \mathcal{A}_n \to \mathbb{N}$ induced by the intersection form on $\mathcal{F}$, and the above discussion leads to $\kappa_T(n) = \min \kappa$ (see Proposition 2·5).

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

Theorem 1·1 follows from the following:

THEOREM 1·2. *There is a constant $C > 0$ so that, for any $\tau \in \mathcal{A}_n$, we have $\kappa(\tau) \geqslant n - C\sqrt{n}\log n$.*

We briefly describe the proof of this theorem. The Farey complex $\mathcal{F}$ admits a natural embedding into a compactification of the hyperbolic plane $\mathbb{H}^2 \cup \partial_\infty \mathbb{H}^2$, so that the vertices of $\mathcal{F}$ embed naturally as $\mathbb{Q} \cup \{\infty\} \hookrightarrow \mathbb{R} \cup \{\infty\} \approx \partial_\infty \mathbb{H}^2$, with edges between vertices mapping to geodesics. The hyperbolic plane $\mathbb{H}^2$ admits a maximal $\mathrm{PSL}(2, \mathbb{Z})$-invariant horospherical packing $\{H_{p/q} : p/q \in \mathbb{Q} \cup \{\infty\}\}$, where $H_{p/q}$ is centered at $p/q \in \mathbb{R} \cup \{\infty\}$, so that a set of slopes span a simplex in $\mathcal{F}$ precisely when the corresponding horospheres are pairwise tangent. (The nomenclature for $H_{p/q}$ is standard in hyperbolic geometry [**29**, Chapter 3] – these are also called *Ford circles* in the literature [**8, 12, 13**]).

A sketch of our proof of Theorem 1·2 is as follows:

(i) locate a 'nice horoball' H for τ, so that $\mathrm{ht}(\tau, H)$, the *height of τ relative to H*, is $O\big(\sqrt{\kappa(\tau)}\big)$. See Definition 3 and Proposition 5·1;

(ii) use H to construct a convex combination of pairwise intersection numbers for τ whose sum is at least $n - O(h \log h)$, where $h = \mathrm{ht}(\tau, H)$. It follows that there is a pair of horoballs of τ with intersection number at least $n - O(h \log h)$. See Proposition 4·2.

The proof of Theorem 1·2 is now immediate: if $\kappa(\tau) \leqslant n$, then $\kappa(\tau) \geqslant n - C\sqrt{n}\log n$.

The first step above uses the hyperbolic geometry of $\mathbb{H}^2$ in an essential way, in which we exploit a simple relationship between intersection numbers, hyperbolic geometry, and Ford circles (see Lemma 2·3).

Organisation. We describe the reduction from $\kappa_T(n)$ to $\kappa : \mathcal{A}_n \to \mathbb{N}$ in Section 2, analyse several examples in Section 3, bound $\kappa(\tau)$ from below in Section 4, locate a good horoball for τ in Section 5, and prove Theorem 1·1 in Section 6.

2. *Preliminaries*

We collect here some useful facts about intersection numbers and the Farey graph $\mathcal{F}$. For more of the beautiful connections between hyperbolic geometry, the Farey graph, continued fractions, and Diophantine approximation, we suggest the reader consult [**18, 25, 26, 27**].

2·1. *Horoballs in trees, Farey labellings, and intersection numbers*

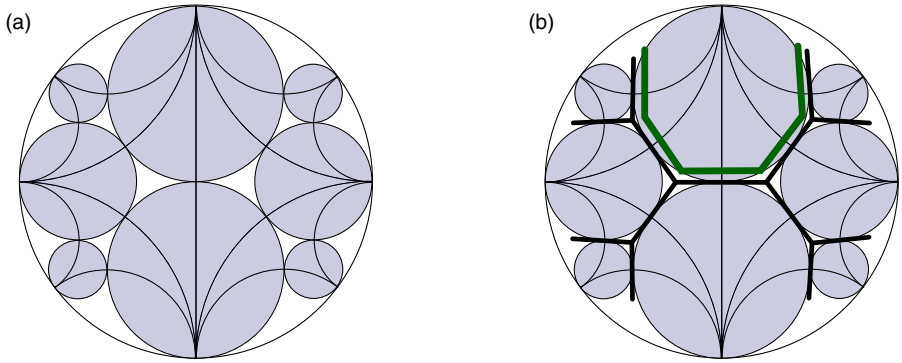
Dual to a triangulation of an n-gon $\tau \in \mathcal{A}_n$ there is a trivalent tree with n leaves embedded in the plane, which we refer to as τ^*. Because τ^* is embedded in the plane, the three edges incident to vertices of τ^* are cyclically ordered. Hence any non-backtracking path in τ^* induces a sequence of left-right turns.

The vertices of τ (that is, the slopes of the k-system) correspond to 'horoballs' in τ^*:

Definition 1 (Horoballs in trees). A *horoball* is a union of edges in a path of the dual tree τ^* that is composed of uni-directional turns (that is, only left or only right turns), which is moreover maximal with respect to inclusion among all such uni-directional subsets of τ^*.

Note that horoballs may be alternately considered as subsets of τ, τ^*, or $\mathbb{H}^2$. We leave the intended meaning to be implied contextually. See Figure 1 for an illustration.

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press



A triangulation of an octagon embeds in $\mathcal{F}$, with corresponding Farey horoballs.

Combinatorial horoballs in the dual tree.

Fig. 1. Horoballs in $\mathbb{H}^2$ and horoballs in trees.

A triangulation $\tau \in \mathcal{A}_n$ admits an embedding into $\mathcal{F}$, and any such embedding gives a map from horoballs of τ to $\mathbb{Q} \cup \{\infty\}$ that records the center of the corresponding horoball in $\mathbb{H}^2$.

Definition 2 (Farey labellings and intersection numbers). A *Farey labelling* of τ is the map from horoballs to $\mathbb{Q} \cup \{\infty\}$ obtained from an embedding of τ in $\mathcal{F}$. The intersection number $\iota(H_1, H_2)$ of a pair of horoballs H_1 and H_2 is given by the intersection number of the slopes corresponding to H_1 and H_2 in a Farey labelling of τ.

We refer to the introduction for an explanation of why intersection numbers of τ are well-defined.

Farey labellings are especially pleasant because the vertices spanning a simplex of $\mathcal{F}$ satisfy a remarkably simple relationship. Namely, if p/q and a/b span an edge of $\mathcal{F}$, then the two other vertices of $\mathcal{F}$ that span a triangle with p/q and a/b are $(p+a)/(q+b)$ and $(p-a)/(q-b)$; this is the 'Farey addition' rule. Note that the choice of labels $1/0$, $0/1$, and $1/1$ for three horoballs incident to a non-leaf vertex of τ^* induces a Farey labelling by using Farey addition to deduce labels of neighbouring horoballs.

2·2. *Monotonicity of intersection numbers and left-right sequences*

The intersection number $\iota(H_1, H_2)$ admits a description more intrinsic to the structure of τ^*, which we now describe. There is a unique (possibly degenerate) non-backtracking path σ between the pair of horoballs H_1 and H_2, and this path determines a sequence of left-right turns $(\ell_1, \ell_2, \ldots, \ell_s)$, where σ makes ℓ_1 turns in the same direction, followed by ℓ_2 turns in the opposite direction, etc. The quantity $\iota(H_1, H_2)$ is given by the numerator of the continued fraction with coefficients $(\ell_1, \ell_2, \ldots, \ell_s)$ [**14**, Theorem 5·3].

Remark. Observe that there is ambiguity in this computation of $\iota(H_1, H_2)$. For one, the non-backtracking path σ may go either from H_1 to H_2, or from H_2 to H_1. Moreover, one must declare that σ is starting with either 'left' or 'right' at its origin vertex, so that it can be observed whether σ is switching directions or not at later vertices. These choices may be

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

made arbitrarily and independently, and this ambiguity has no affect on the calculation of $\iota(H_1, H_2)$. See [**14**, figure 3, example 1].

This viewpoint suggests a certain monotonicity.

LEMMA 2·1 ('Monotonicity of intersection numbers'). *Suppose that σ and σ' are non-backtracking paths with respective left-right sequences $(\ell_1, \ldots, \ell_s)$ and $(\ell'_1, \ldots, \ell'_{s'})$. If $s' \geqslant s$ and $\ell_i' \geqslant \ell_i$ for each $i = 1, \ldots, s$, then the intersection number determined by σ' is at least that determined by σ.*

Proof. This lemma is almost exactly [**14**, lemma 5·5], with the sole difference that we may have $s' > s$. Therefore to prove the claim we may assume that σ' contains σ as an initial subpath.

Choose a Farey labelling with label $1/0$ at the horoball forming the origin of σ (and σ'), and labels $0/1$ and $1/1$ at the two neighbouring horoballs that intersect σ. Compare the denominators of the Farey labels of the horoballs at the terminuses of σ and σ'; because these labels are computed using Farey addition, it is evident that the denominator of the horoball for σ' is at least that of σ. The intersection number of any horoball with $1/0$ is given by the denominator of its Farey label, so the claim follows.

2·3. *Intersection numbers and hyperbolic distance*

The quotient of $\mathbb{H}^2$ by $\mathrm{PSL}(2, \mathbb{Z})$ is a hyperbolic orbifold with one cusp and two orbifold points, one of order 2 and one of order 3. The preimage of the maximal horoball neighborhood of the cusp under the covering projection $\mathbb{H}^2 \to \mathbb{H}^2/\mathrm{PSL}(2, \mathbb{Z})$ is $\mathcal{H} = \{H_{p/q} : p/q \in \mathbb{Q} \cup \{\infty\}\}$, a $\mathrm{PSL}(2, \mathbb{Z})$-invariant collection of horoballs centered at the completed rationals. The following lemma is an exercise in hyperbolic geometry.

LEMMA 2·2. *Every point in $\mathbb{H}^2$ is within $\log 2/\sqrt{3}$ of a horoball in $\mathcal{H}$.*

There is a simple well-known fundamental relationship between intersection numbers of curves on the torus and hyperbolic distance between the corresponding horoballs (see e.g. [**27**, proposition 6·2]).

LEMMA 2·3. *We have $d_{\mathbb{H}^2}\left(H_{p/q}, H_{a/b}\right) = 2 \log \iota\left(\frac{p}{q}, \frac{a}{b}\right)$ for any $p/q, a/b \in \mathcal{F}$.*

Proof. Applying an element of $\mathrm{PSL}(2, \mathbb{Z})$, we may assume that $p/q = \infty$ in the upper half-plane model for $\mathbb{H}^2 \cup \partial_\infty \mathbb{H}^2$. The horosphere $H_{a/b}$ is given by $\{z : \left|z - (a/b + i/2b^2)\right| = 1/2b^2\}$ (see e.g. [**2**]), so

$$d_{\mathbb{H}^2}\left(H_{p/q}, H_{a/b}\right) = d_{\mathbb{H}^2}\left(a/b + \frac{i}{b^2}, a/b + i\right) = 2 \log b = 2 \log \iota\left(\frac{p}{q}, \frac{a}{b}\right).$$

2·4. *Width and height for horoballs*

The interior of each edge of τ^* is incident to exactly two horoball regions. Thus, for any choice of horoball H in τ^*, there are exactly two other horoball regions, distinct from H, that are incident to the interiors of the extreme edges of H. Call these H_1 and H_2.

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

Definition 3. The width of τ relative to H is $w = \iota(H_1, H_2)$. The height of τ relative to H is

$$\mathrm{ht}(\tau, H) = \max\{\ \iota(H, H') : H' \text{ is a horoball of } \tau\ \}.$$

For the remainder of this article, we will suppress the difference between the triangulation $\tau \in \mathcal{A}_n$ and its dual tree τ^*. The translation between them is quite natural, and the difference can henceforth be understood from context.

2·5. *The two kappas*

Recall from the introduction the quantity $\kappa_T(n)$, which is the minimum, taken over collections Γ of n curves on S, of the maximum pairwise intersection number of curves in Γ.

Definition 4 ('Max Intersection Function'). The function $\kappa : \mathcal{A}_n \to \mathbb{N}$ is defined by

$$\kappa(\tau) = \max\{\ \iota(H_1, H_2) : H_1 \text{ and } H_2 \text{ are horoballs of } \tau\ \}.$$

As noted in the introduction, we claim that $\kappa_T(n) = \min\{\kappa(\tau) : \tau \in \mathcal{A}_n\}$.

That $\kappa_T(n) \leqslant \min \kappa$ is easy: for any $\tau \in \mathcal{A}_n$, choose a Farey labelling. The quantity $\kappa(\tau)$ is equal to the maximum pairwise intersection number of the n slopes obtained in this Farey labelling, and $\kappa_T(n)$ is the minimum of the maximum pairwise intersection of any n slopes, so $\kappa_T(n) \leqslant \kappa(\tau)$ for each $\tau \in \mathcal{A}_n$.

The reverse inequality is slightly less obvious, and relies on a certain convexity of maximal k-systems in $\partial_\infty \mathbb{H}^2$. The following lemma makes this precise.[1]

LEMMA 2·4. *If Γ is a k-system on T which is maximal with respect to inclusion among k-systems, then $\mathcal{F}$ induces a triangulation of the n-gon which forms the convex hull of $\Gamma \subset \partial_\infty \mathbb{H}^2$.*

Proof. Let g_{ab} indicate the geodesic in $\mathbb{H}^2$ with endpoints $a, b \in \partial_\infty \mathbb{H}^2$. Suppose that $\alpha, \beta \in \Gamma \subset \mathcal{F}$, that Δ is a Farey triangle intersecting $g_{\alpha\beta}$, and that δ is a vertex of Δ (and, hence, of $\mathcal{F}$). Because the path σ that computes the intersection number $\iota(\alpha, \delta)$ is a subpath of the path σ' that computes $\iota(\alpha, \beta)$ (and similarly for $\iota(\delta, \beta)$), Lemma 2·1 implies that both $\iota(\delta, \alpha)$ and $\iota(\delta, \beta)$ are at most k. For any $\gamma \in \Gamma$, either $g_{\alpha\gamma}$ or $g_{\beta\gamma}$ intersect Δ, so it follows that $\iota(\gamma, \delta) \leqslant k$ as well. Maximality of Γ implies that $\delta \in \Gamma$, so the convex hull of Γ is equal to the union of Farey triangles spanning elements of Γ.

This demonstrates that $\kappa_T(n) \geqslant \min \kappa$, so we may conclude:

PROPOSITION 2·5. *We have* $\kappa_T(n) = \min\limits_{\tau \in \mathcal{A}_n} \kappa$.

3. *Illustrative examples*

There are several natural elements of $\mathcal{A}_n$ that we can use to observe $\kappa : \mathcal{A}_n \to \mathbb{N}$.

Remark. Technically, the vertices of the associahedron correspond to triangulations of a *labelled* convex polygon. Notice however that the max intersection function $\kappa : \mathcal{A}_n \to \mathbb{N}$ is

[1] We are grateful to Ian Agol for sharing an unpublished note with us which contained Lemma 2·4 [**3**].

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

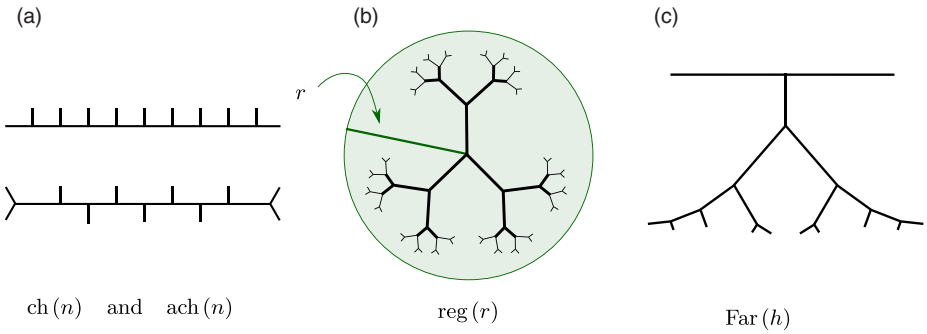


Fig. 2. Several dual trees of elements in $\mathcal{A}_n$.

invariant under permutation of labels, so we can safely refer to $\kappa(\tau)$ for elements τ of $\mathcal{A}_n$ without reference to a particular ordering of the horoballs of τ.

(i) The element $\mathrm{ch}(n) \in \mathcal{A}_n$ (for 'chain') contains a horoball of width n. We have $\kappa(\mathrm{ch}(n)) = n$, and the height relative to the horoball of width n is 1.

(ii) The element $\mathrm{ach}(n) \in \mathcal{A}_n$ (for 'alternating chain') contains a path of length $n-3$ that switches direction $n-4$ times. Here we have $\kappa(\mathrm{ach}(n)) = F_n$, the nth Fibonacci number, the largest width horoball of $\mathrm{ach}(n)$ is 3, and the height relative to any horoball is at least $F_{\lfloor \frac{n}{2} \rfloor}$.

(iii) The element $\mathrm{reg}(r) \in \mathcal{A}_n$ (for 'regular'), with $n = 3 \cdot 2^{r-1}$, is formed by choosing the subtree of the homogeneous (infinite) trivalent tree that is induced on all vertices at combinatorial distance at most r from a fixed vertex. The tree $\mathrm{reg}(r)$ contains the alternating chain $\mathrm{ach}(2r+1)$ as a subtree, and in fact we have $\kappa(\mathrm{reg}(r)) = \kappa(\mathrm{ach}(2r+1)) = F_{2r+1}$. The largest width of a horoball of $\mathrm{reg}(r)$ is given by $2r-1$, and the height of $\mathrm{reg}(r)$ relative to this horoball is F_{r+1}.

(iv) The element $\mathrm{Far}(h) \in \mathcal{A}_n$ (for the 'Farey series'), with $n = 2 + \sum_{k \leqslant h} \phi(k)$, where ϕ is Euler's totient function, is the subgraph of $\mathcal{F}$ induced on fractions in $\mathbb{Q} \cap [0, 1]$ that can be written with denominator $\leqslant h$, together with $1/0$. Observe that $\kappa(\mathrm{Far}(h)) = h^2 - 2h$, the largest width of a horoball of $\mathrm{Far}(h)$ is given by h, while the height relative to this horoball is given by $h-1$.

We collect this information in Figure 2 and Table 1.

Remark. Observe the large difference between $\kappa(\mathrm{ach}(n)) \approx ((1+\sqrt{5})/2)^n$ and $\kappa(\mathrm{ch}(n)) = n$. Though we have not discussed it, there is a natural simplicial structure on $\mathcal{A}_n$, with edges between triangulations of an n-gon that differ by a single diagonal flip. It is not hard to see that the diameter of $\mathcal{A}_n$ is at most $2n$ (in fact, this quantity can be determined precisely [**23, 28**]), so it follows that the change in κ across an edge of $\mathcal{A}_n$ can be arbitrarily large.

4. *Estimating kappa using heights*

Let $\tau \in \mathcal{A}_n$. The strategy to obtain a lower bound for $\kappa(\tau)$ is to find a set of pairwise intersection functions $\{I_\alpha\}$ for τ, and estimate the convex combination

$$\sum_\alpha r_\alpha I_\alpha \geqslant n - \epsilon(n), \tag{4·1}$$

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

Table 1. *Some data for attractive elements of $\mathcal{A}_n$. Some entries include only leading-order terms, ignoring multiplicative constants. Note that $\Phi = 1+\sqrt{5}/2$.*

τ	$\kappa(\tau)$	Largest width horoball H	$\mathrm{ht}(\tau, H)$
$\mathrm{ch}(n)$	n	$n-2$	1
$\mathrm{ach}(n)$	Φ^n	3	$\geqslant \Phi^{n/2}$
$\mathrm{reg}(r)$	$n^{2\log_2 \Phi}$	$2\log_2 n$	$n^{\log_2 \Phi}$
$\mathrm{Far}(h)$	$\frac{\pi^2}{3}n$	$\sqrt{n}$	$\sqrt{n}$

for some set of non-negative weights $\{r_\alpha\}$ with $\sum r_\alpha = 1$, and error term $\epsilon(n)$. Of course, we have $\kappa(\tau) \geqslant I_\alpha$ for all α, and by convexity we must have $I_\alpha \geqslant n - \epsilon(n)$ for some α.

We will rely on three facts from analytic number theory, which we group together in a single lemma for convenience. Below we indicate the interval $\{1, \ldots, m\}$ by $[m]$ and the subset of $\{1, \ldots, m\}$ relatively prime to s by $[m]_s$, e.g. $\phi(m) = \#[m]_m$. The number of divisors of n is indicated by $d(n)$.

LEMMA 4·1. *We have the following estimates:*

$$\#[m]_n = \frac{m}{n}\phi(n) + O(d(n)), \tag{4·2}$$

$$\sum_{k \leqslant h} d(k) = h \log h + O(h) \text{ and} \tag{4·3}$$

$$\sum_{k \leqslant h} \frac{\phi(k)}{k} = O(h). \tag{4·4}$$

The estimate (4·2) is a standard application of Möbius inversion [9, lemma 3·4]. The second estimate (4·3) is a weaker version of a famous theorem of Dirichlet [**6**, chapter 3], and a more precise form of (4·4) can be found in [**30**]. (Note that the error term in (4·2) is in fact $O(\vartheta(n))$, where $\vartheta(n)$ is the number of square-free divisors of n. However, in the sum $\sum_{j \leqslant h} \vartheta(j)$, one finds the same order of growth as $\sum_{j \leqslant h} d(j)$ [**21**], so in our application, Lemma 4·3, this improvement is immaterial.)

In this section we will show:

PROPOSITION 4·2. *There is a constant $C > 0$ so that, for any $\tau \in \mathcal{A}_n$ and any horoball H of τ with $h = \mathrm{ht}(\tau, H)$, we have $\kappa(\tau) \geqslant n - Ch \log h$.*

As in Section 2·4, let H_1 and H_2 be the two horoballs that are incident to H along its two extreme edges. By construction, the non-backtracking path σ from H_1 to H_2 is contained in H; we indicate the vertices σ passes through in order by $p_1, \ldots, p_w$. See Figure 3.

For each j, the complement $\tau \setminus p_j$ consists of three components, and we indicate (the closure of) the unique such component that doesn't intersect H as the jth *branch* $B(j)$.

Label H with $1/0$, label the horoball neighbouring H along the edge $\overline{p_j p_{j+1}}$ with $0/1$, and label the horoball neighbour of H along $\overline{p_{j-1} p_j}$ with $1/1$. Now Farey addition determines how to fill in labels for the remaining horoballs, and let $\mathrm{ht}_H B(j)$ indicate the maximum denominator among Farey labels for horoballs intersecting $B(j)$.

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

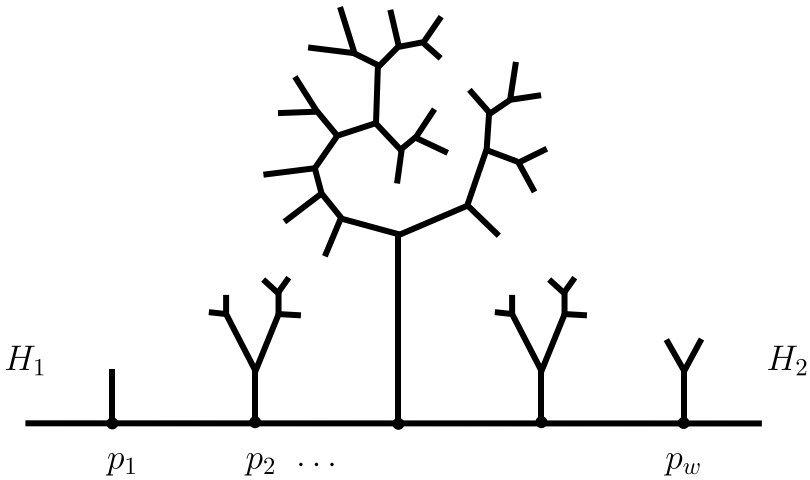


Fig. 3. The horoball H determines branches for τ, and extreme horoballs H_1 and H_2.

We may count the n horoball regions of τ by filtering them according to their heights on the branches. That is, for each $k \in \mathbb{N}$, let $X(k) = \{j : \mathrm{ht}_H\, B(j) \geqslant k\}$ (that is, the set of indices j where the jth branch has height at least k), and let $x_k = \#X(k)$. See Figure 6.

Suppose that $k \leqslant \mathrm{ht}_H\, B(j)$. We now define 'leftmost' and 'rightmost' horoballs $L(j)_k$ and $R(j)_k$ at height at least k on the branch $B(j)$, and corresponding 'numerator invariants' $\ell(j)_k$ and $r(j)_k$. Consider the subset of horoballs of $B(j)$ whose relative heights to H are $\geqslant k$, nonempty by assumption. We declare $L(j)_k$ and $R(j)_k$ to be the elements of this set with maximal and minimal Farey labels, respectively. (Among horoballs of $B(j)$ at height $\geqslant k$ from H, the horoball $L(j)_k$ appears furthest to the *left* from the viewpoint of H; similarly $R(j)_k$ appears furthest to the *right*.) Now we define 'numerator invariants' $\ell(j)_k$ and $r(j)_k$ for $B(j)$ as follows: Suppose the Farey labels of $L(j)_k$ and $R(j)_k$ are a_l/b_l and a_r/b_r respectively. We let $\ell(j)_k := \lfloor k \cdot (a_l/b_l) \rfloor$ and $r(j)_k := \lceil k \cdot (a_r/b_r) \rceil$. See Figure 4 and Figure 5.

Consider j_k^+ (resp. j_k^-), the maximal (resp. minimal) index of $X(k)$. Below, we indicate $\lambda_k = \ell(j_k^-)_k$ and $\rho_k = r(j_k^+)_k$; roughly speaking, λ_k (resp. ρ_k) is the 'leftmost (resp. rightmost) height k numerator invariant'. See Figure 7.

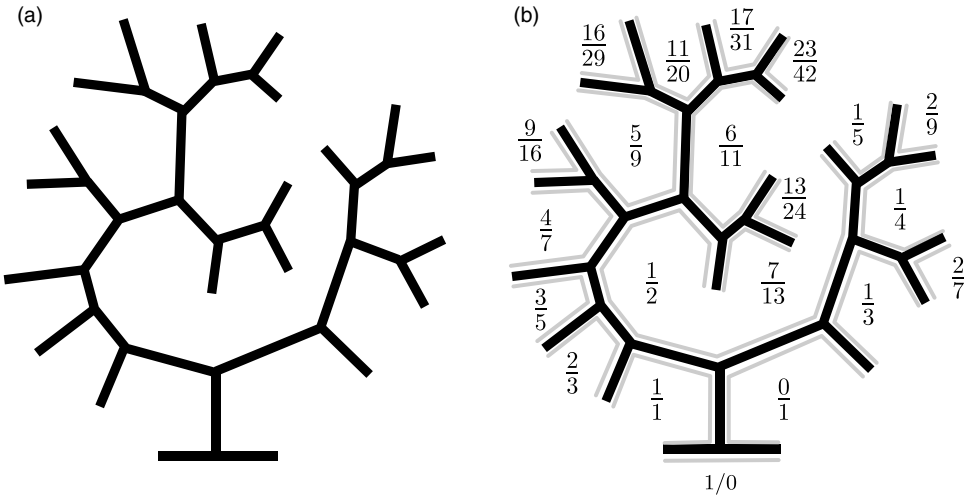
LEMMA 4·3. *We have*

$$\sum_{k=1}^{h} \phi(k) x_k \geqslant n + \sum_{k=1}^{h} \phi(k) + \sum_{k=1}^{h} \frac{\phi(k)}{k} (\rho_k - \lambda_k) - O\,(h \log h)\,.$$

Proof. Observe that the sum of the number of horoballs of τ at height k relative to H, as k goes from 1 to h, is exactly $n - 1$.

The number of horoballs of $B(j)$ at height k relative to H is at most $\phi(k)$, so the total number of horoballs of τ at height k from H is at most $\phi(k)x_k$. Observe that we may count the vertices of $B(j_k^-)$ and $B(j_k^+)$ with slightly more care: Provided the numerators of horoballs of $B(j)$ at height k are between ℓ and r, the number of horoballs of $B(j)$ at height k relative to H is at most $\phi(k) - \#[r-1]_k$, and at most $\#[\ell]_k$. Therefore the total number of horoballs of

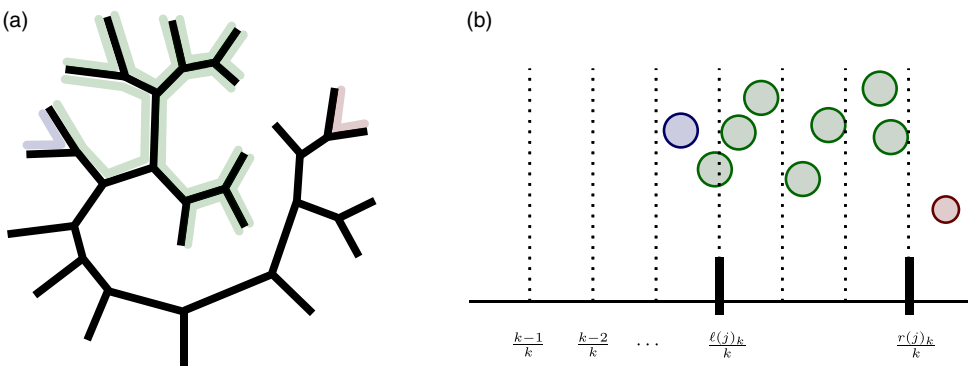
https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press



A branch $B(j)$

Farey labels for the horoballs of $B(j)$

Fig. 4. A branch acquires Farey labels.



$L(j)_k$ in blue and $R(j)_k$ in red, for $k = 9$

The horoballs at heights $\geqslant k$ determine the numerator invariants $\ell(j)_k$ and $r(j)_k$.

Fig. 5. The numerator invariants $\ell(j)_k$ and $r(j)_k$ of $B(j)$.

τ at height k relative to H is at most

$$\phi(k)x_k - (\underbrace{\phi(k) - \#[\lambda_k]_k}_{\text{overcount in } B(j_k^-)}) - \underbrace{\#[\rho_k - 1]_k}_{\text{overcount in } B(j_k^+)} .$$

(The reader is reassured that the overcounts above are distinct when $j_k^- = j_k^+$; in that case $x_k = 1$ and the expression simplifies to $\#[\lambda_k]_k - \#[\rho_k - 1]_k$.) By (4·2), the count above is at most

$$\phi(k)x_k - \phi(k) + \frac{\lambda_k}{k}\phi(k) - \frac{\rho_k - 1}{k}\phi(k) + O(d(k)).$$

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

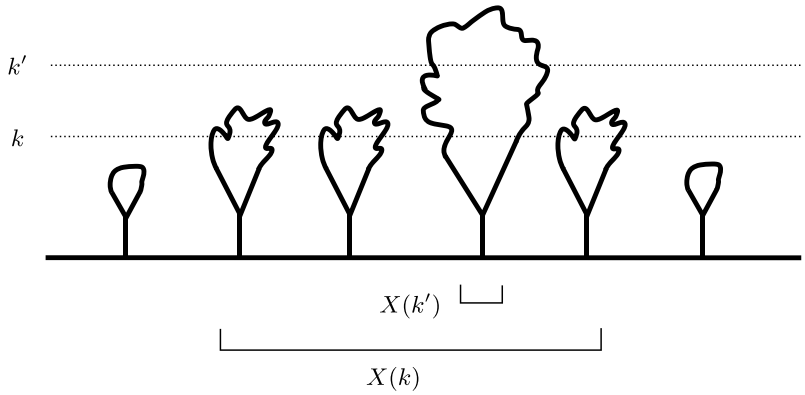


Fig. 6. The horoballs of τ may be filtered according to heights from H.

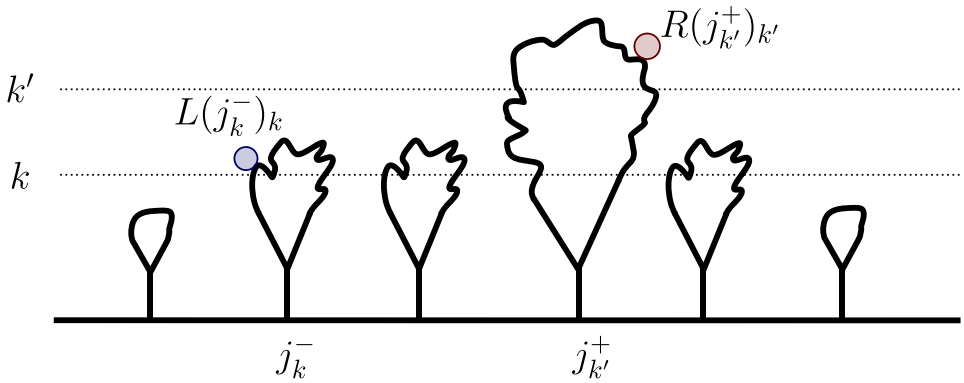


Fig. 7. The leftmost horoball at height $\geqslant k$ and rightmost horoball at height $\geqslant k'$ are used to build intersection number $I_{kk'}$.

The sum of this expression as k goes from 1 to h is $n-1$, so rearranging we find that

$$\sum_{k=1}^{h} \phi(k)x_k \geqslant n + \sum_{k=1}^{h} \phi(k) + \sum_{k=1}^{h} \frac{\phi(k)}{k}(\rho_k - \lambda_k) - 1 - \sum_{k=1}^{h} \frac{\phi(k)}{k} - C\sum_{k=1}^{h} d(k).$$

By (4·3) and (4·4), the last three terms can be replaced by $O(h \log h)$.

The reader may observe how Lemma 4·3 is somewhat suggestive of (4·1).

Given a choice of two heights k and k', consider the intersection number $I_{kk'}$ between $L(j^-_k)_k$ and $R(j^+_{k'})_{k'}$. See Figure 7. It is straightforward to compute

$$I_{kk'} \geqslant kk'|j^+_{k'} - j^-_k| + k'\lambda_k - k\rho_{k'}.$$

Notice that $|j^+_{k'} - j^-_k| + |j^+_k - j^-_{k'}| \geqslant x_k + x_{k'} - 2$, so upon dividing by kk' we find

$$\frac{1}{kk'}(I_{kk'} + I_{k'k}) \geqslant x_k + x_{k'} - 2 + \left(\frac{\lambda_k}{k} - \frac{\rho_k}{k}\right) + \left(\frac{\lambda_{k'}}{k'} - \frac{\rho_{k'}}{k'}\right). \tag{4·5}$$

With Lemma 4·3 in mind, we would like to choose pairs $\{k, k'\} \subset [h]$ so that the sum over the choices made of the terms '$x_k + x_{k'}$' on the right-hand side of (4·5) is equal to $\sum \phi(k)x_k$;

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

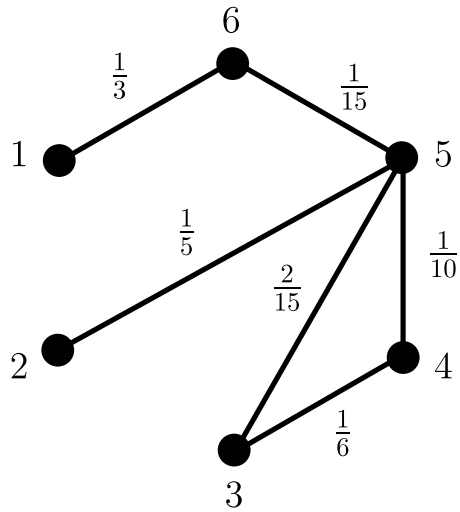


Fig. 8. The graph Γ_6 as described in Proposition 4·4. Edge weights sum to 1.

meanwhile, the sum of coefficients on the left-hand side should be equal to 1. The following proposition makes this idea feasible.

PROPOSITION 4·4. *For each $h \in \mathbb{N}$, there is a graph Γ_h satisfying*:

(i) the vertex set of Γ_h is given by $\{1, \ldots, h\}$;

(ii) the valence of vertex k is $\phi(k)$;

(iii) the sum $\sum_{k \sim k'} \frac{2}{kk'}$ over the edges of Γ_h is equal to 1.

See Figure 8 for a picture of Γ_6.

Proof. Declare $k \sim k'$ when $\gcd(k, k') = 1$ and $k + k' > h$.

For property (ii), choose a vertex k. Each integer i relatively prime to k can be shifted by k to $i + k$, another integer relatively prime to k. For $1 \leqslant i \leqslant k$, we may choose the maximum n so that $i + nk \leqslant h$. The result is a bijection of $[k]_k$, the integers in $[k]$ relatively prime to k, with the set of integers k', relatively prime to k, less than h, and so that $k + k' > h$. Therefore the valence of k is $\#[k]_k = \phi(k)$.

For property (iii), observe that, to transform Γ_h into Γ_{h+1}, the edges $k \sim k'$ with $k + k' = h + 1$ are deleted and replaced by edges $k \sim (h + 1)$ and $k' \sim (h + 1)$. Because $k + k' = h + 1$, the edge weight $2/kk'$ in Γ_h is equal to the sum of edge weights $2/k(h + 1) + 2/k'(h + 1)$ in Γ_{h+1}, so the sum $\sum_{k \sim k'} 2/kk'$ is independent of h. For the base case $h = 2$, observe that $2/1 \cdot 2 = 1$.

Let E indicate the set of edges of Γ_h. Now Proposition 4·4 and (4·5) imply:

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

$$\sum_{(k,k')\in E} \frac{1}{kk'}\,(I_{kk'}+I_{k'k}) \geqslant \sum_{(k,k')\in E} \left[x_k + x_{k'} - 2 + \left(\frac{\lambda_k}{k} - \frac{\rho_k}{k}\right) + \left(\frac{\lambda_{k'}}{k'} - \frac{\rho_{k'}}{k'}\right)\right]$$
$$= \sum_k \phi(k)x_k \;-\; \sum_k \phi(k) \;+\; \sum_k \frac{\phi(k)}{k}\,(\lambda_k - \rho_k)$$

Applying Lemma 4·3, we find that

$$\sum_{(k,k')\in E} \frac{1}{kk'}\,(I_{kk'}+I_{k'k}) \geqslant n - O(h\log h)\,. \tag{4·6}$$

Because $\sum_E 2/kk' = 1$ by Proposition 4·4, the left-hand side of (4·6) is a convex combination of intersection numbers for τ. Therefore inequality (4·6) implies that there is a pair of horoballs whose intersection number is at least the right-hand side, proving Proposition 4·2.

5. *Finding a horoball of controlled relative height*

For many $\tau \in \mathcal{A}_n$, there exist H so that the $\mathrm{ht}(\tau, H)$ is $O(1)$, so Proposition 4·2 with height demonstrates that $\kappa(\tau) = n - O(1)$. However, such a horoball need not exist, e.g. every horoball of $\mathrm{ach}(n)$ has height at least $\approx ((1+\sqrt{5})/2)^{n/2}$. Nonetheless, $\kappa(\mathrm{ach}(n))$ is quite large (on the order $(1+\sqrt{5})/2)^n$), so one might hope that it is always possible to find horoballs of small height relative to $\kappa(\tau)$. We show:

PROPOSITION 5·1. *There exists a constant $C > 0$ so that, for any $\tau \in \mathcal{A}_n$, there exists a horoball H of τ so that the height of τ relative to H is controlled as* $\mathrm{ht}(\tau, H) \leqslant C\sqrt{\kappa(\tau)}$.

Remark. The reader can observe that the conclusion above fits the data in Table 1. It is tempting to hope for an improvement of Proposition 5·1 along the following lines: as $\kappa(\tau)$ gets closer to $\min \kappa$ (e.g. if $\kappa(\tau) \leqslant n$), one should be able to find horoballs of τ with relative heights $\ll \sqrt{n}$.

On the other hand, the row containing $\tau = \mathrm{Far}(h)$, with $n \approx 3/\pi^2 h^2$, makes this hope seem quite remote. Indeed, $\kappa(\mathrm{Far}(h))$ is greater than n only by the innocuous looking linear factor $\frac{\pi^2}{3} \approx 3.3$, yet every horoball has relative height $\geqslant h \approx \sqrt{n}$.

Proof. Let K_1, K_2 be horoballs of τ so that $\iota(K_1, K_2) = \kappa(\tau)$, and let $r = \log \kappa(\tau)$. By Lemma 2·3, we have $d_{\mathbb{H}^2}(K_1, K_2) = 2r$.

Let $x \in \mathbb{H}^2$ be the midpoint of the geodesic from K_1 to K_2. By Lemma 2·2 there is a horoball $H \in \mathcal{H}$ so that $d_{\mathbb{H}^2}(H, x) \leqslant \log(2/\sqrt{3})$. The horoball H corresponds to a horoball in the dual tree to $\mathcal{F}$ which is, by construction, incident to the geodesic between K_1 and K_2. By Lemma 2·4, H is a horoball in τ as well. (Strictly speaking, one could distinguish notationally between Farey horoballs in $\mathcal{H}$, horoballs in the dual tree to $\mathcal{F}$, and horoballs in τ; because we feel it only obscures the relevant choices, we leave this distinction to be implied by the reader.)

We claim that H satisfies the requisite bound. Let K be any other horoball of τ. Because $\mathbb{H}^2$ is δ-hyperbolic, the point x is within δ of the geodesic segment between K and K_i for i equal to either 1 or 2. A standard application of the triangle inequality (see Figure 9) then yields

$$d_{\mathbb{H}^2}(K, K_i) \geqslant d_{\mathbb{H}^2}(K, x) + d_{\mathbb{H}^2}(x, K_i) - 2\delta = d_{\mathbb{H}^2}(K, x) + r - 2\delta.$$

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

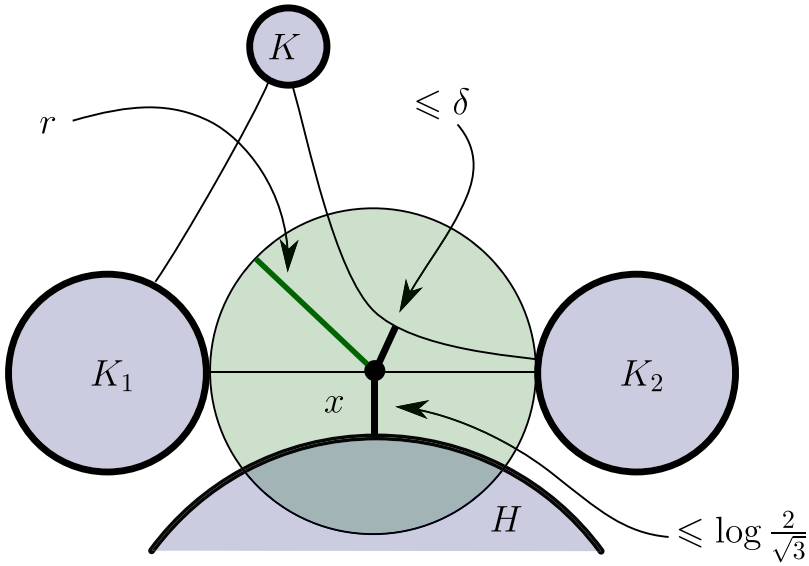


Fig. 9. A horoball K not far from the K_i is not far from H as well.

Because $d_{\mathbb{H}^2}(K, K_i) \leqslant 2r$, we conclude that $d_{\mathbb{H}^2}(K, x) \leqslant r + 2\delta$, and

$$d_{\mathbb{H}^2}(H, K) \leqslant d_{\mathbb{H}^2}(H, x) + d_{\mathbb{H}^2}(x, K) \leqslant \log \frac{2}{\sqrt{3}} + r + 2\delta.$$

By Lemma 2·3 we conclude that

$$\iota(H, K) = e^{\frac{1}{2} d_{\mathbb{H}^2}(H,K)} \leqslant \frac{2}{\sqrt{3}} e^{2\delta} \sqrt{\kappa(\tau)}.$$

Remark. Ian Agol has suggested a slightly different version of the above proof: choose Farey labels for the horoballs in τ, and enlarge the horoballs by $\log \kappa(\tau) + \log(2/\sqrt{3})$. A variation on Lemma 2·2 together with Lemma 2·3 implies that every trio of these horoballs mutually intersect, so by Helly's theorem there is a point x in their common intersection [**19**], and one may finish as above.

6. *From kappa to eta*

As stated in the introduction, it is an exercise to show that

$$\eta_T(k) = \max\{n : \kappa_T(n) \leqslant k\}.$$

By Theorem 1·2, we may conclude $\eta_T(k) \leqslant \max\{n : n - C\sqrt{n} \log n \leqslant k\}$.

Theorem 1·1 now follows from the following lemma:

LEMMA 6·1. *Suppose that $C > 0$ is a constant, and that $f : \mathbb{R} \to \mathbb{R}$ is an increasing, sublinear function, with $f(x) = o(x)$. There is a $D > 0$ so that for any $k \geqslant 1$ we have* $\max\{x : x - Cf(x) \leqslant k\} \leqslant k + Df(k)$.

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

Proof. Because $f(x) = o(x)$, there is a $C_1 > 0$ large enough so that $C_1 - 1 > Cf(C_1)$. Because f is sublinear, for any $k \geqslant 1$ we have

$$(C_1 - 1)k > Ckf(C_1) \geqslant Cf(C_1 k).$$

Adding k to both sides and rearranging we find

$$C_1 k - Cf(C_1 k) > k. \tag{6·1}$$

Let $F(k) = \max\{x : x - Cf(x) \leqslant k\}$. By (6·1) we have $F(k) < C_1 k$. Of course, by definition of $F(k)$ we have $F(k) - Cf(F(k)) \leqslant k$. Because f is increasing and sublinear, we find

$$F(k) \leqslant k + Cf(F(k)) \leqslant k + Cf(C_1 k) \leqslant k + CC_1 f(k),$$

as claimed.

Because $\sqrt{x}\log x$ is increasing and sublinear, this completes the proof of Theorem 1·1.

Remark. The conclusion of Lemma 6·1 holds under much weaker assumptions. For instance, sublinearity of f can be replaced by the assumption that there is some $C_2 > 0$ so that $f(x + y)$ is at most $C_2 f(x) + C_2 f(y)$.

Acknowledgements. The authors thank Josh Greene and Ian Agol for valuable feedback on an early draft of this paper, and an anonymous referee for their fastidious reading. We are especially grateful to Ian Agol for sharing with us an unpublished note that contained Lemma 2·4, and for suggesting an alternative to our proof of Proposition 5·1 (see Figure 5).

REFERENCES

[1] T. Aougab, I. Biringer and J. Gaster. Packing curves on surfaces with few intersections. *Internat. Math. Res. Notices*, **2019**(16) (2019), 5205–5217.
[2] J. Athreya, S. Chaubey, A. Malik and A. Zaharescu. Geometry of Farey-Ford polygons. *New York J. Math.* **21** (2015), 637–656.
[3] I. Agol. Intersections of curves on the torus. Unpublished.
[4] I. Agol. Bounds on exceptional Dehn filling. *Geom. Topol.*, **4** (2000), 431–449.
[5] T. Aougab. Local geometry of the k-curve graph. *Trans. Amer. Math. Soc.*, **370**(4) (2018), 2657–2678.
[6] T. M. Apostol. *Introduction to Analytic Number Theory* (Springer Science and Business Media, 2013).
[7] R. C Baker, G. Harman and J. Pintz. The difference between consecutive primes, ii. *Proc. London Math. Soc.*, **83**(3) (2001), 532–562.
[8] J. H Conway and R. Guy. *The Book of Numbers* (Springer Science and Business Media, 2012).
[9] E. Cohen. Arithmetical functions associated with the unitary divisors of an integer. *Math. Z.*, **74**(1) (1960), 66–80.
[10] H. Cramér. Some theorems concerning prime numbers. *Arkiv för Mathematik, Astronom o Fysik* (1920).
[11] H. Cramér. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arith.*, **2** (1936), 23–46.
[12] F. Bonahon. Funny fractions and Ford circles. YouTube video (2015), available at: www.youtube.com/watch?v=0hlvhQZIOQw
[13] L. R. Ford. Fractions. *Amer. Math. Monthly*, **45**(9) (1938), 586–601.
[14] J. Gaster M. Lopez, E. Rexer, Z. Riell and Y. Xiao. Combinatorics of k -farey graphs. *Rocky Mountain J. Math.*, **50**(1) (2020), 135–151.
[15] A. Granville. Harald Cramér and the distribution of prime numbers. *Scand. Actuar. J.*, **1995**(1) (1995), 12–28.

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press

[16] J. E. GREENE. On curves intersecting at most once, ii. arXiv preprint arXiv:1811.01413, (2018).
[17] J. E. GREENE. On loops intersecting at most once. *Geom. Funct. Anal.*, **29**(6) (2019), 1828–1843.
[18] A. HATCHER. Topology of numbers. Unpublished manuscript, in preparation, (2002).
[19] E. HELLY. Über mengen konvexer körper mit gemeinschaftlichen punkte. *Jahresber. Deutsch. Math.-Verein.*, **32** (1923), 175–176.
[20] M. JUVAN, A. MALNIČ and B. MOHAR. Systems of curves on surfaces. *J. Combin. Theory Ser. B*, **68**(1) (1996), 7–22.
[21] F. MERTENS. Über einige asymptotische gesetze der zahlentheorie. *J. Reine Angew. Math.*, **1874**(77) (1874), 289–338.
[22] T. OLIVEIRA E SILVA, S. HERZOG and S. PARDI. Empirical verification of the even Goldbach conjecture and computation of prime gaps up to 4×10^{18}. *Math. Comput.* **83**(288) (2014), 2033–2060.
[23] L. POURNIN. The diameter of associahedra. *Adv. Math.*, **259** (2014), 13–42.
[24] P. PRZYTYCKI. Arcs intersecting at most once. *Geom. Funct. Anal.*, **25**(2) (2015), 658–670.
[25] C. SERIES. The geometry of Markoff numbers. *Math. Intelligencer*, **7**(3) (1985), 20–29.
[26] C. SERIES. The modular surface and continued fractions. *J. London Math. Soc.*, **2**(1) (1985), 69–80.
[27] B. SPRINGBORN. The hyperbolic geometry of Markov's theorem on Diophantine approximation and quadratic forms. *Enseign. Math.*, **63**(3) (2018), 333–373.
[28] D. D. SLEATOR, R. E. TARJAN and W. P. THURSTON. Rotation distance, triangulations, and hyperbolic geometry. *J. Amer. Math. Soc.*, **1**(3) (1988), 647–681.
[29] W. P. THURSTON. Geometry and topology of 3-manifolds. Lecture notes (Princeton University, 1978).
[30] A. WALFISZ. Weylsche exponentialsummen in der neueren zahlentheorie. *VEB Deutscher Verlag der Wissenschaften* (1963).

https://doi.org/10.1017/S0305004122000421 Published online by Cambridge University Press