Forum of Mathematics, Sigma (2023), **Vol. 11**:e114 1–47
doi:10.1017/fms.2023.114

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Counting geodesics of given commutator length

Viveka Erlandsson[1] and Juan Souto[2]

[1]School of Mathematics, University of Bristol, Woodland Road, Bristol, BS81UG, UK, and
Department of Mathematics and Statistics, UiT The Arctic University of Norway, Norway; E-mail: v.erlandsson@bristol.ac.uk.
[2]CNRS, IRMAR - UMR 6625, Université de Rennes, Campus de Beaulieu, Rennes, 35042, France; E-mail: jsoutoc@gmail.com.

Received: 17 July 2023; **Revised:** 8 October 2023; **Accepted:** 20 October 2023

2020 Mathematics Subject Classification: *Primary* – 37D40; *Secondary* – 37A17

Abstract

Let Σ be a closed hyperbolic surface. We study, for fixed g, the asymptotics of the number of those periodic geodesics in Σ having at most length L and which can be written as the product of g commutators. The basic idea is to reduce these results to being able to count critical realizations of trivalent graphs in Σ. In the appendix, we use the same strategy to give a proof of Huber's geometric prime number theorem.

Contents

1. Introduction

In this paper, we will be interested in studying the growth of the number of certain classes of geodesics in a closed, connected and oriented hyperbolic surface $\Sigma = \Gamma\backslash\mathbb{H}^2$, which we assume to be fixed throughout the paper.

In [10], Huber proved that the cardinality of the set $\mathbf{C}(L)$ of nontrivial oriented periodic geodesics $\gamma \subset \Sigma$ of length $\ell_\Sigma(\gamma) \leqslant L$ behaves like

$$|\mathbf{C}(L)| \sim \frac{e^L}{L} \text{ as } L \to \infty. \tag{1.1}$$

Here, $\sim$ means that the ratio between both quantities tends to 1 as $L \to \infty$. Huber's result has been generalized in all possible directions, allowing, for example, for surfaces of finite area [9] and for higher-dimensional negatively curved manifolds or replacing periodic geodesic by orbits of Anosov flows [12].

Variations of Huber's result have also been obtained when one imposes additional restrictions on the geodesics one is counting. Here, one should mention Mirzakhani's work on counting simple geodesics [13] or more generally geodesics of given type [14 , 8], but what is closer to what we will care about in this paper are earlier results of Katsuda–Sunanda [11] and Phillips–Sarnak [16] giving the asymptotic behavior of the number of geodesics with length at most L and satisfying some homological condition. For example, Katsuda–Sunada [11] proved that

$$|\{\gamma \in \mathbf{C}(L) \text{ homologically trivial}\}| \sim (g_\Sigma - 1)^{g_\Sigma} \cdot \frac{e^L}{L^{g_\Sigma+1}}, \tag{1.2}$$

where g_Σ is the genus of Σ. Phillips–Sarnak [16] get this same result with an estimate on the error term.

Here, we will be counting certain kinds of homologically trivial curves. Every homologically trivial curve γ can be represented, up to free homotopy, as the product of commutators in $\pi_1(\Sigma)$. The smallest number of commutators needed is the *commutator length* of γ. This quantity agrees with the genus of the smallest connected oriented surface S with connected boundary ∂S for which there is a (continuous) map $S \longrightarrow \Sigma$ sending ∂S to γ. This explains why we think of the commutator length $cl(\gamma)$ of γ as the *genus of* γ.

In this paper, we study, for fixed but otherwise arbitrary $g \geqslant 1$, the asymptotic behavior of the cardinality of the set

$$\mathbf{B}_g(L) = \left\{\gamma \subset \Sigma \;\middle|\; \begin{array}{l}\text{closed geodesic with } \ell_\Sigma(\gamma) \leqslant L \text{ and} \\ \text{commutator length } cl(\gamma) = g\end{array}\right\} \tag{1.3}$$

This is our main result:

Theorem 1.1. *Let Σ be a closed, connected and oriented hyperbolic surface and for $g \geqslant 1$, and $L > 0$ let $\mathbf{B}_g(L)$ be as in Equation (1.3). We have*

$$|\mathbf{B}_g(L)| \sim \frac{2}{12^g \cdot g! \cdot (3g-2)! \cdot \mathrm{vol}(T^1\Sigma)^{2g-1}} \cdot L^{6g-4} \cdot e^{\frac{L}{2}}$$

as $L \longrightarrow \infty$.

Here, as we will throughout the paper, we have endowed the unit tangent bundle $T^1\Sigma$ with the Liouville measure, normalized in such a way that $\mathrm{vol}(T^1\Sigma) = 2\pi \cdot \mathrm{vol}(\Sigma) = -4\pi^2\chi(\Sigma)$. In particular, we have that, as in Equation (1.2), the quantities in Theorem 1.1 depend on the topology of the underlying surface Σ but there is no dependence on its geometry.

Note that by definition the curves in $\mathbf{B}_g(L)$ bound a surface of genus g but that this surface is just the image of a continuous, or if you want smooth, map. We prove that only a small but definite proportion of the elements in $\mathbf{B}_g(L)$ arise as the boundary of an immersed surface of genus g (and connected boundary):

Theorem 1.2. *Let Σ be a closed, connected and oriented hyperbolic surface and for $g \geqslant 1$, and $L > 0$ let $\mathbf{B}_g(L)$ be as in Equation (1.3). We have*

$$|\{\gamma \in \mathbf{B}_g(L) \textit{ bounds immersed surface of genus } g\}| \sim \frac{1}{2^{4g-2}} |\mathbf{B}_g(L)|$$

as $L \longrightarrow \infty$.

Theorem 1.2 should perhaps be compared with the immersion theorem in [4, Theorem 4.79]. The immersion theorem asserts that every homologically trivial geodesic γ in Σ virtually bounds an immersed surface, meaning that there is an immersion into Σ of a compact surface S in such a way that the boundary of S is mapped onto γ with positive degree. In some sense, Theorem 1.2 seems to suggest that, most of the time, the genus of S does not agree with the genus of γ.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

In the course of the proof of Theorem 1.1, we will need a different counting result that we believe has its own interest. Suppose namely that X is a compact trivalent graph. Under a critical realization of X in Σ, we understand a (continuous) map

$$\phi : X \to \Sigma$$

sending each edge to a nondegenerate geodesic segment in such a way that any two (germs of) edges incident to the same vertex are sent to geodesic segments meeting at angle $\frac{2\pi}{3}$ (see Lemma 2.2 for an explanation of the etymology of this terminology). Although it may not be completely evident to the reader at this point, the set

$$\mathbf{G}^X(L) = \left\{ \begin{array}{c} \phi : X \to \Sigma \text{ critical realization} \\ \text{with length } \ell_\Sigma(\phi) \leqslant L \end{array} \right\} \tag{1.4}$$

is finite, where the *length* of a critical realization is defined to be the sum

$$\ell_\Sigma(\phi) = \sum_{e \in \mathbf{edge}(X)} \ell_\Sigma(\phi(e))$$

of the lengths of the geodesic segments $\phi(e)$ when e ranges over the set of edges of X. The following is the key to Theorem 1.1:

Theorem 1.3. *Let Σ be a closed, connected and oriented hyperbolic surface. For every connected trivalent graph X, we have*

$$|\mathbf{G}^X(L)| \sim \left(\frac{2}{3}\right)^{3\chi(X)} \cdot \frac{\mathrm{vol}(T^1\Sigma)^{\chi(X)}}{(-3\chi(X)-1)!} \cdot L^{-3\chi(X)-1} \cdot e^L$$

as $L \to \infty$. Here, $\chi(X)$ is the Euler-characteristic of the graph X.

Let us sketch the proof of Theorem 1.3. Recall first that there are basically two approaches (that we know of) to establish Huber's theorem (1.1): Either one approaches it à la Huber [10], that is, from the point of spectral analysis or, as Margulis did later [12], exploiting the ergodic properties of the geodesic flow. Indeed, already the predecessor to Huber's theorem, namely Delsarte's lattice point counting theorem [7] can be approached from these two different points of view. In Section 3 below, we will sketch the argument to derive Delsarte's theorem from the fact that the geodesic flow is mixing, discussing also the count of those geodesic arcs in Σ going from x to y and whose initial and terminal speed are in predetermined sectors of the respective unit tangent spaces – see Theorem 3.1 for the precise statement. This is indeed all the dynamics we will need in the proof of the theorems above. To be clear, we do not make use of any of the slightly rarified refinements of the mixing property of the geodesic flow. Specifically, we do not need exponential mixing or such.

The basic idea of the proof of Theorem 1.3 is to note that, for a fixed graph X, the set of all its realizations in Σ, that is, maps $X \to \Sigma$ mapping each edge geodesically, is naturally a manifold. Generically, each connected component contains a unique critical realization. To count how many such critical realizations, there are with total length less than L we consider for small ε the set $\mathcal{G}_{\varepsilon-\mathrm{crit}}(L)$ of geodesic realizations of length at most L and where the angles, instead of being $\frac{2\pi}{3}$ on the nose, are in the interval of size ε around that number. Delsarte's theorem allows us to compute the volume $\mathrm{vol}(\mathcal{G}_{\varepsilon-\mathrm{crit}}(L))$ of that set of geodesic realizations. A little bit of hyperbolic geometry then shows that most of the connected components of $\mathcal{G}_{\varepsilon-\mathrm{crit}}(L)$ have basically the same volume. We thus know, with a small error, how many connected components we have, and thus how many critical realizations. This concludes the sketch of the proof of Theorem 1.3.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Remark. The strategy used to prove Theorem 1.3 can be used to give a pretty easy proof of Huber's theorem (1.1). We work this out in the appendix, and although there is no logical need of doing so, we encourage the reader to have a look at it before working out the details of the proof of Theorem 1.3.

Let us also sketch the proof of Theorem 1.1. A *fat graph* is basically a graph which comes equipped with a regular neighborhood $\mathbf{neigh}(X)$ homeomorphic to an oriented surface. Such a fat graph *has genus* g if this regular neighborhood is homeomorphic to a compact surface of genus g with connected boundary. The point of considering fat graphs is that whenever we have a realization $\phi : X \to \Sigma$ of the graph underlying a genus g fat graph then we get a curve, namely $\phi(\partial\,\mathbf{neigh}(X))$ which has at most genus g. The basic idea of the proof of Theorem 1.1 is to consider the set

$$\mathbf{X}_g = \left\{ (X,\phi) \,\middle|\, \begin{array}{c} X \text{ is a fat graph of genus } g \text{ and} \\ \phi : X \to \Sigma \text{ is a critical realization} \\ \text{of the underlying graph} \end{array} \right\} \Big/ _{\text{equiv}}$$

of (equivalence classes) of realizations (see Section 7 for details) and prove that the map

$$\Lambda : \mathbf{X}_g \to \mathbf{C}, \quad (X,\phi) \mapsto \text{ geodesic homotopic to } \phi(\partial X)$$

is basically bijective onto $\mathbf{B}_g$ and that generically, the geodesic $\Lambda(X,\phi)$ has length almost exactly equal to $2 \cdot \ell(\phi) - C$ for some explicit constant C. Once we are here, we get the statement of Theorem 1.1 from Theorem 1.3 together with a result by Bacher and Vdovina [1]. Other than proving Theorem 1.3, the bulk of the work is to establish the properties of the map Λ. The key step is to bound the number of curves with at most length L and which arise in two essentially different ways as the boundary of genus g surfaces:

Theorem 1.4. *For any g, there are at most* $\mathbf{const} \cdot L^{6g-5} \cdot e^{\frac{L}{2}}$ *genus g closed geodesics γ in Σ with length $\ell(\gamma) \leqslant L$ and with the property that there are two nonhomotopic fillings $\beta_1 : S_1 \to \Sigma$ and $\beta_2 : S_2 \to \Sigma$ of genus $\leqslant g$.*

Here, a genus g filling of γ is a continuous map $\beta : S \to \Sigma$ from a genus g surface S with connected boundary such that $\beta(\partial S) = \gamma$.

Theorem 1.4 may look kind of weak because we are only bounding by $\frac{\mathbf{const}}{L}$ the proportion of those elements in $\mathbf{B}_g(L)$ that we are double counting when we count surfaces of genus g instead of counting curves. It should however be noted that this is the order of the error term in Equation (1.2) (see [16]), and this is indeed the order of error term that we expect in the results we prove here. For what it is worth, it is not hard to show that the set of those γ in Theorem 1.4 is at least of the order of $\mathbf{const} \cdot L^{6g-10} \cdot e^{\frac{L}{2}}$. Indeed, if $\omega \in \pi_1(\Sigma)$ arises in two ways as a commutator, for example if we choose $\omega = [aabab, ba^{-1}a^{-1}] = [aaba, bba^{-1}]$ and if η is a randomly chosen product of $g-1$ generators, then $\omega\eta$ arises in two different ways as a product of g commutators and hence admits two nonhomotopic fillings, and there are $\mathbf{const} \cdot L^{6g-10} \cdot e^{\frac{L}{2}}$ many choices for η.

Section-by-section summary

In Section 2, we discuss realizations of graphs and the topology and geometry of spaces of realizations, and in Section 3 we discuss Delsarte's classical lattice point counting result and a couple of minimal generalizations thereof. At this point, we will have all tools needed to prove Theorem 1.3; this is done in Section 4. In Section 5 and Section 6, we work out the geometric aspects of the proof of Theorem 1.1, the main result being Theorem 1.4. Although we do not use any results about pleated surfaces, some of the arguments in these two sections will come pretty naturally to those readers used to working with such objects. In Section 7, we prove Theorem 1.1, combining the results of the previous two sections with Theorem 1.3. Theorem 1.2 is proved in Section 8. Finally, Section 9 is dedicated to discussing what parts of what we do hold if we soften the assumption that Σ is a closed hyperbolic surface. To conclude, we present in Appendix A a proof of Huber's theorem using the same idea as in the proof of Theorem 1.3.

Remark. After conclusion of this paper, we learned that the problem of calculating how many elements arise as commutators in some group has also been treated in other cases. More concretely, Park [15] uses Wicks forms to get asymptotics, as $L \to \infty$, for the number of elements in the free group $\mathbb{F}_r$ of rank r which arise as a commutator and have word length L. In the same paper, he also treats the case that the group is a free product of two nontrivial finite groups. It would be interesting to figure out if it is possible to apply the methods here to recover Park's beautiful theorems.

Notation

Under a graph X, we understand a one-dimensional CW-complex with finitely many cells. We denote by $\mathbf{vert} = \mathbf{vert}(X)$ the set of vertices, by $\mathbf{edge} = \mathbf{edge}(X)$ the set of edges of X, and by $\mathbf{half} = \mathbf{half}(X)$ the set of half-edges of a graph X – a half-edge is nothing other than the germ of an edge. Given a vertex $v \in \mathbf{vert}(X)$, we let $\mathbf{half}_v = \mathbf{half}_v(X)$ be the set of half-edges emanating out of v. Note that two elements of $\mathbf{half}_v$ might well correspond to the same edge – that is, X might have edges which are incident to the same vertex on both ends. The cardinality of $\mathbf{half}_v$ is the *degree* of X at v and we say that X is *trivalent* if its degree is 3 at every vertex. The reader can safely assume that the graphs they encounter are trivalent.

When it comes to surfaces, we will be working all the time with the same underlying surface, our fixed closed connected and oriented hyperbolic surface $\Sigma = \Gamma\backslash\mathbb{H}^2$. We identify $T^1\Sigma$ with $\Gamma\backslash T^1\mathbb{H}^2 = \Gamma\backslash \operatorname{PSL}_2 \mathbb{R}$ and endow $T^1\Sigma$ with the distance induced by a $\operatorname{PSL}_2 \mathbb{R}$ left-invariant Riemannian metric on $\operatorname{PSL}_2 \mathbb{R}$, say one that gives length 2π to each unit tangent space $T^1_{x_0}\Sigma$ and such that the projection $T^1\mathbb{H}^2 \to \mathbb{H}^2$ is a Riemannian submersion. This means that the unit tangent bundle has volume $\operatorname{vol}(T^1\Sigma) = 2\pi \cdot \operatorname{vol}(\Sigma) = 4\pi^2 \cdot |\chi(S)|$. Angles between tangent vectors based at the same point of Σ will always be unoriented, meaning that they take values in $[0, \pi]$ – note that this is consistent with the unit tangent spaces having length 2π.

Often we will denote the discrete sets we are counting by boldface capitals, such as $\mathbf{C}$ or $\mathbf{B}$. They will often come with a wealth of decorations such as for example $\mathbf{G}^X(L)$ or $\mathbf{B}_g(L)$. Often these sets arise as discrete subsets of larger spaces which will be denoted by calligraphic letters. Often the boldfaced and the calligraphic letters go together: $\mathbf{G}$ will be a subset of the space $\mathcal{G}$.

A comment about constants. In this paper, there are two kinds of constants: the ones whose value we are trying to actually compute and those about which we just need to know that they exist. Evidently, the first kind we have to track carefully. It would, however, be too painful, and for no clear gain, to do the same with all possible constants. And in general constants tend to breed more and more constants. This is why we we just write **const** for a constant whose actual value is irrelevant, allowing the precise value of **const** to change from line to line. We hope that this does not cause any confusion.

And now a comment about Euclidean vectors. All vector arising here, indicated with an arrow as in $\vec{v} = (v_1, \ldots, v_k)$, are *positive* in the sense that the entries v_i are positive. In other words, they live in $\mathbb{R}^k_+$ or maybe in $\mathbb{N}^k$. We will write

$$\|\vec{v}\| = |v_1| + \cdots + |v_k| = v_1 + \cdots + v_k$$

for the L^1-norm on vectors. This is the only norm we will encounter here.

2. Realizations of graphs in surfaces

In this section, we discuss certain spaces of maps of graphs into a surface. These spaces play a key rôle in this paper. Although long, the material here should be nice and pleasant to read – only the proof of Proposition 2.3 takes some amount of work.

Realizations

We will be interested in connected graphs living inside our hyperbolic surface Σ, or more precisely in continuous maps

$$\phi : X \to \Sigma \tag{2.1}$$

of graphs X into Σ which when restricted to each edge are geodesic. We will say that such a map (2.1) is a *realization of* X *in* Σ. We stress that realizations do not need to be injective and that in fact the map ϕ could be constant on certain edges or even on larger pieces of the graph. A *regular* realization is one whose restriction to every edge is nonconstant. If $\phi : X \to \Sigma$ is a regular realization and if $\vec{e} \in \mathbf{half}(X)$ is a half-edge incident to a vertex $v \in \mathbf{vert}(X)$, then we denote by $\phi(\vec{e}) \in T^1_{\phi(v)}\Sigma$ the unit tangent vector at $\phi(v)$ pointing in the direction of the image of $\vec{e}$.

We endow the set $\mathcal{G}^X$ of all realizations of the (always connected) graph X with the compact-open topology and note that unless X itself is contractible, the space $\mathcal{G}^X$ is not connected: The connected components of $\mathcal{G}^X$ correspond to the different possible free homotopy classes of maps of X into Σ. Indeed, pulling segments tight relative to their endpoints we get that any homotopy

$$[0,1] \times X \to \Sigma, \quad (t,x) \mapsto \varphi_t(x)$$

between two realizations is homotopic, relative to $\{0,1\} \times X$, to a homotopy, which we are still denoting by the same symbol, such that

1. $t \mapsto \varphi_t(v)$ is geodesic for every vertex $v \in \mathbf{vert}(X)$, and
2. $\varphi_t : X \to \Sigma$ is a realization for all t.

A homotopy $[0,1] \times X \to \Sigma$ satisfying (1) and (2) is said to be a *geodesic homotopy*.

Geodesic homotopies admit a different intrinsic description. Indeed, uniqueness of geodesic representatives in each homotopy class of arcs implies that each realization $\phi \in \mathcal{G}^X$ has a neighborhood which is parametrized by the image of the vertices. This implies that the map

$$\Pi : \mathcal{G}^X \to \Sigma^{\mathbf{vert}(X)}, \quad \phi \mapsto (\phi(v))_{v \in \mathbf{vert}(X)} \tag{2.2}$$

is a cover. Pulling back the product of hyperbolic metrics, we think of it as a manifold locally modeled on the product $(\mathbb{H}^2)^{\mathbf{vert}(X)} = \mathbb{H}^2 \times \cdots \times \mathbb{H}^2$ of $\mathbf{vert}(X)$ worth of copies of the hyperbolic plane. Geodesic homotopies are, from this point of view, nothing other than geodesics in $\mathcal{G}^X$.

Since all of this will be quite important, we record it here as a proposition:

Proposition 2.1. *Let X be a graph. The map (2.2) is a cover and geodesic homotopies are geodesics with respect to the pull-back metric.*

Length function

On the space $\mathcal{G}^X$ of realizations of the graph X in Σ, we have the *length function*

$$\ell_\Sigma : \mathcal{G}^X \to \mathbb{R}_{\geqslant 0}, \quad \ell_\Sigma(\phi) = \sum_{e \in \mathbf{edge}(X)} \ell_\Sigma(\phi(e)).$$

First, note that Arzela–Ascoli implies that ℓ_Σ is a proper function. It follows thus that the restriction of ℓ_Σ to any and every connected component of $\mathcal{G}^X$ has a minimum. Now, convexity of the distance function $d_{\mathbb{H}^2}(\cdot,\cdot)$ implies that ℓ_Σ is convex. More precisely, if (φ_t) is a geodesic homotopy between two realizations in Σ, then the function $t \mapsto \ell_\Sigma(\varphi_t)$ is convex. Indeed, it is strictly convex unless the image of $[0,1] \times X$ is contained in a geodesic in Σ, or rather if the image of some (and hence any) lift to the universal cover is contained in a geodesic in $\mathbb{H}^2$.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Note now also that the length function is smooth when restricted to the set of regular realizations – its derivative is given by the first variation formula

$$\frac{d}{dt}\ell_\Sigma(\phi_t) = \sum_{v\in\mathbf{vert}(X)}\sum_{\vec{e}\in\mathbf{half}_v(X)}\left\langle -\phi(\vec{e}), \frac{d}{dt}\phi_t(v)\right\rangle,$$

where $\phi(\vec{e}) \in T^1_v\Sigma$ is the unit tangent vector based at v and pointing as the image of the half-edge $\vec{e}$. It follows that a regular realization ϕ is a critical point for the length function $\ell_\Sigma : \mathcal{G}^X \to \mathbb{R}_{\geqslant 0}$ if and only if for every $v \in \mathbf{vert}(X)$ we have $\sum_{\vec{e}\in\mathbf{half}_v}\phi(\vec{e}) = 0$. Note that this implies, in the for us relevant case that X is trivalent, that the (unsigned) angle $\angle(\phi'(\vec{e}_1), \phi'(\vec{e}_2))$ between the images of any two half-edges incident to the same vertex is equal to $\frac{2\pi}{3}$.

Definition. A regular realization $\phi : X \to \Sigma$ of a trivalent graph X into Σ is *critical* if we have

$$\angle(\phi(\vec{e}_1), \phi(\vec{e}_2)) = \frac{2\pi}{3}$$

for every vertex $v \in \mathbf{vert}(X)$ and for any two distinct $\vec{e}_1, \vec{e}_2 \in \mathbf{half}_v(X)$.

We collect in the next lemma a few of the properties of critical realizations:

Lemma 2.2. *Let X be a trivalent graph. A regular realization $\phi \in \mathcal{G}^X$ is a critical point for the length function if and only if ϕ is a critical realization. Moreover, if $\phi \in \mathcal{G}^X$ is a critical realization and $\mathcal{G}^\phi$ is the connected component of $\mathcal{G}^X$ containing ϕ, then the following holds:*

1. *ϕ is the unique critical realization in $\mathcal{G}^\phi$.*
2. *ϕ is the global minimum of the length function on $\mathcal{G}^\phi$.*
3. *Besides ϕ, there are no other local minima in $\mathcal{G}^\phi$ of the length function.*
4. *The connected component $\mathcal{G}^\phi$ is isometric to the product $\mathbb{H}^2 \times \cdots \times \mathbb{H}^2$ of* $\mathbf{vert}(X)$ *many copies of the hyperbolic plane.*

Note that lack of global smoothness of the length function means that we cannot directly derive (2) and (3) from (1). This is why they appear as independent statements.

Proof. Statement (1) was actually discussed in the paragraph preceding the definition of critical realization. Let us focus in the subsequent ones. Recall that if

$$[0, 1] \to \mathcal{G}^\phi,\ t \mapsto \phi_t$$

is a (nonconstant) geodesic homotopy with $\phi_0 = \phi$, then the length function

$$t \mapsto \ell_\Sigma(\phi_t(X)) \tag{2.3}$$

is convex. In our situation it is strictly convex: Since ϕ is critical, we get that its image, or rather the image of its lifts to the universal cover, are not contained in a geodesic because if they were, then the angles between any two half-edges starting at the same vertex could only take the values 0 or π. Now, strict convexity and the fact that $\phi = \phi_0$ is a critical point of the length function implies that $t = 0$ is the minimum and only critical point of the function (2.3). From here, we get directly (2) and (3). To prove (4), note that if $\mathcal{G}^\phi$ were not simply connected, then there would be a nontrivial geodesic homotopy starting and ending at ϕ, contradicting the strict convexity of the length function. Note now that the restricting the cover (2.2) to $\mathcal{G}^\phi$ we get a locally isometric cover $\mathcal{G}^\phi \to \Sigma^{\mathbf{vert}(X)}$. Since the domain of this cover is connected and simply connected, we get that it is nothing other than the universal cover of $\Sigma^{\mathbf{vert}(X)}$. This proves (4) and concludes the proof of the lemma. □

Remark. Although we will not need it here, let us comment briefly on the topology of the connected components of $\mathcal{G}^X$. The fundamental group of the connected component $\mathcal{G}^\phi$ containing a realization

$\phi \in \mathcal{G}^X$ is isomorphic to the centralizer $\mathcal{Z}_{\pi_1(\Sigma)}(\phi_*(\pi_1(X)))$ in $\pi_1(\Sigma)$ of the image of $\pi_1(X)$ under $\phi_* : \pi_1(X) \to \pi_1(\Sigma)$. It follows that $\mathcal{G}^\phi$ is isometric to the quotient under the diagonal action of $\mathcal{Z}_{\pi_1(\Sigma)}(\phi_*(\pi_1(X)))$ of $\mathbb{H}^2 \times \cdots \times \mathbb{H}^2$ of $\mathbf{vert}(X)$-copies of the hyperbolic plane. In particular, we have:

○ If ϕ is homotopically trivial, then $\mathcal{G}^\phi \simeq \pi_1(\Sigma)\backslash(\mathbb{H}^2 \times \cdots \times \mathbb{H}^2)$.
○ If $\phi_*(\pi_1(X)) \neq \mathrm{Id}_{\pi_1(\Sigma)}$ is abelian, then $\mathcal{G}^\phi \simeq \mathbb{Z}\backslash(\mathbb{H}^2 \times \cdots \times \mathbb{H}^2)$.
○ $\phi_*(\pi_1(X))$ is non-abelian, then $\mathcal{G}^\phi \simeq \mathbb{H}^2 \times \cdots \times \mathbb{H}^2$.

In the appendix, we will give a concrete description of the components of $\mathcal{G}^X$ for the case that X is a loop, that is, the graph with a single vertex and a single edge.

Lemma 2.2, with all its beauty, does not say anything about the existence of critical realizations. Our next goal is to prove that any realization that from far away kind of looks like a critical realization is actually homotopic to a critical realization.

Quasicritical realizations

Recall that a regular realization $\phi : X \to \Sigma$ of a trivalent graph X is *critical* if the angles between the images of any two half-edges incident to the same vertex are equal to $\frac{2\pi}{3}$. We will say that a regular realization is *quasicritical* if those angles are bounded from below by $\frac{1}{2}\pi$ and that the realization is ℓ_0*-long* if $\ell(\phi(e)) > \ell_0$ for all edges e of X. Recall that we measure all angles in $[0, \pi]$.

Our next goal here is to prove that every sufficiently long quasicritical realization is homotopic to a critical realization. This will follow easily from the following technical result:

Proposition 2.3. *For any trivalent graph X and constant $C \geq 0$, there exist constants $\ell_0, D > 0$ such that given an ℓ_0-long quasicritical realization $\phi : X \to \Sigma$, a trivalent graph Y and realization $\psi : Y \to \Sigma$ satisfying*

1. *$\ell(\psi) \leq \ell(\phi) + C$, and*
2. *there is a homotopy equivalence $\sigma : X \to Y$ with ϕ and $\sigma \circ \psi$ homotopic,*

then there exists a homeomorphism $F : Y \to X$ mapping each edge with constant speed such that $\sigma \circ F$ is homotopic to the identity and such that the geodesic homotopy $X \times [0, 1] \to \sigma$, $(x, t) \mapsto \phi_t(x)$ joining $\phi_0(\cdot) = \phi \circ F(\cdot)$ to $\phi_1(\cdot) = \psi(\cdot)$ has tracks bounded by D.

Since the proof of Proposition 2.3 is technical and pretty long, it may be a good idea to skip it in a first reading. Nevertheless, before proving it, we demonstrate that it might be useful:

Corollary 2.4. *Let X be a trivalent graph. There are positive constants ℓ_0 and D such that every component of $\mathcal{G}^X$ which contains an ℓ_0-long quasicritical realization $\phi : X \to \Sigma$ also contains a critical realization ψ, which moreover is unique and homotopic to ϕ by a homotopy whose tracks have length bounded by D.*

Proof. Let ℓ_0 and D be given by Proposition 2.3 for $C = 0$ and if needed increase ℓ_0 so that it is larger than $2D$. Let $\phi : X \to \Sigma$ be an ℓ_0-long quasicritical realization. Let $\psi : X \to \Sigma$ be the minimizer for the length function in the component $\mathcal{G}^\phi$ – it exists because the length function is proper. We claim that ψ is critical. In the light of Lemma 2.2, it suffices to prove that ψ is regular.

Being a minimizer we have that $\ell_\Sigma(\psi) \leqslant \ell_\Sigma(\phi)$. We thus get from Proposition 2.3 that there exists a homeomorphism $F : X \to X$ homotopic to the identity, mapping edges with constant speed and such that the geodesic homotopy from $\phi \circ F$ to ψ has tracks bounded by D. Now, since X is trivalent we get that the homeomorphism F, being homotopic to the identity and mapping edges with constant speed, is actually equal to the identity. What we thus have is a homotopy from ϕ and ψ with tracks bounded by D. Now, since each edge of $\phi(X)$ has at least length $\ell_0 > 2D$ and since the tracks of the homotopy are bounded by D, we get that ψ is regular and hence critical by Lemma 2.2, as we needed to prove. Lemma 2.2 also yields that ψ is unique. □

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Let us next prove the proposition:

Proof of Proposition 2.3. Let $\phi : X \to \Sigma$ be a quasicritical realization, and assume it is ℓ_0-long for an ℓ_0 large enough to satisfy some conditions we will give in the course of the proof. Let $H : X \times [0,1] \to \Sigma$ be the homotopy between ϕ and $\psi \circ \sigma$.

To be able to consistently choose lifts of ϕ, ψ and σ to the universal covers $\widetilde{X}, \widetilde{Y}$ and $\mathbb{H}^2$ of X, Y and Σ, let us start by picking base points. Fixing $x_0 \in X$, consider the base points $\sigma(x_0) \in Y$ and $\phi(x_0) \in \Sigma$, and pick lifts $\widetilde{x}_0 \in \widetilde{X}$, $\widetilde{\sigma(x_0)} \in \widetilde{Y}$ and $\widetilde{\phi(x_0)} \in \mathbb{H}^2$ of each one of those endpoints. Having chosen those base points we have uniquely determined lifts $\widetilde{\phi} : \widetilde{X} \to \mathbb{H}^2$ and $\widetilde{\sigma} : \widetilde{X} \to \widetilde{Y}$ of ϕ and σ satisfying $\widetilde{\phi}(\widetilde{x}_0) = \widetilde{\phi(x_0)}$ and $\widetilde{\sigma}(\widetilde{x}_0) = \widetilde{\sigma(x_0)}$. We can also lift to $\mathbb{H}^2$, starting at $\widetilde{\phi}(\widetilde{x}_0)$, the path in Σ given by $t \mapsto H(x_0, t)$. The endpoint of this path is a lift of $\psi \circ \sigma(x_0)$, and we take the lift $\widetilde{\psi} : \widetilde{Y} \to \mathbb{H}^2$ which maps $\widetilde{\sigma}(\widetilde{x}_0)$ to this point. All those lifts are related by the following equivariance property:

$$\widetilde{\phi}(g(x)) = \phi_*(g)(\widetilde{\phi}(x)) \text{ and } (\widetilde{\psi} \circ \widetilde{\sigma})(g(x)) = \phi_*(g)\Big((\widetilde{\psi} \circ \widetilde{\sigma})(x)\Big) \tag{2.4}$$

for all $x \in \widetilde{X}$ and for all $g \in \pi_1(X, x_0)$, where $\phi_* : \pi_1(X, x_0) \to \pi_1(\Sigma, \phi(x_0))$ is the homomorphism induced by ϕ and the chosen base points.

Note that, since ϕ is almost critical and ℓ_0-long we get, as long as ℓ_0 is large enough, that the lift $\widetilde{\phi} : \widetilde{X} \to \mathbb{H}^2$ is an injective quasi-isometric embedding. For ease of notation, denote by $T_X = \widetilde{\phi}(\widetilde{X}) \subset \mathbb{H}^2$ the image of $\widetilde{\phi}$ and let us rephrase what we just said: If ℓ_0 is sufficiently large, then T_X is a quasiconvex tree, with quasiconvexity constants only depending on a lower bound for ℓ_0. We denote by $\hat{\pi} : \mathbb{H}^2 \to T_X$ a nearest point retraction which is equivariant under $\phi_*(\pi_1(X, x_0))$. We would like to define

$$\pi : \widetilde{Y} \to T_X$$

as $\hat{\pi} \circ \widetilde{\psi}$, but since $\hat{\pi}$ is definitively not continuous, we have to be slightly careful. We define π as follows: First, set $\pi(v) = \hat{\pi}(\widetilde{\psi}(v))$ for all vertices $v \in \mathbf{vert}(\widetilde{Y})$ of $\widetilde{Y}$, and then extend (at constant speed) over the edges of Y. Note that Equation (2.4), together with the equivariance of π, implies that $\pi \circ \widetilde{\sigma} : \tilde{X} \to T_X$ satisfies

$$(\pi \circ \widetilde{\sigma})(g(x)) = \phi_*(g)(\pi \circ \widetilde{\sigma}(x))$$

for all $x \in \widetilde{X}$ and $g \in \pi_1(X, x_0)$.

Claim 1. *As long as ℓ_0 is large enough, we have that π maps each vertex of $\widetilde{Y}$ within* **const** *of one and only one vertex of T_X.*

Starting with the proof of the claim, note that there is a constant $A \geqslant 0$ depending only on the quasiconvexity constants for T_X with

$$d_{\mathbb{H}^2}(x, y) \geqslant -A + d_{\mathbb{H}^2}(\hat{\pi}(x), \hat{\pi}(y)) \tag{2.5}$$

for all $x, y \in \mathbb{H}^2$. In fact, there exists constants $K, A' > 0$ which once again depend only on the quasiconvexity constant of T_X such that whenever $d_{\mathbb{H}^2}(\hat{\pi}(x), \hat{\pi}(y)) > K$ we have the better bound

$$d_{\mathbb{H}^2}(x, y) \geqslant -A' + d_{\mathbb{H}^2}(x, \hat{\pi}(x)) + d_{\mathbb{H}^2}(\hat{\pi}(x), \hat{\pi}(y)) + d_{\mathbb{H}^2}(\hat{\pi}(y), y). \tag{2.6}$$

For $a, b \in T_X$, let $d_{T_X}(a, b)$ denote the interior distance in T_X between them. We will care about a modified version of this distance: Let B be the collection of balls in $\mathbb{H}^2$ of radius 1 centered at each vertex of T_X, and define $d_{T_X \operatorname{rel} B}(a, b)$ to be the length of the part of the path in T_X between them that

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

lies outside of B. For all ℓ_0 large enough (depending only on hyperbolicity of $\mathbb{H}^2$), we have from the choice of the radius[1] that

$$d_{T \operatorname{rel} B}(a, b) \leq d_{\mathbb{H}^2}(a, b) \tag{2.7}$$

for all $a, b \in T_X$.

Given two edges e, e' of $\widetilde{Y}$ adjacent to the same vertex, we define

$$\operatorname{Fold}(e, e') = \ell_{T \operatorname{rel} B}(\pi(e) \cap \pi(e'))$$

and let then

$$\operatorname{Fold}(\pi) = \max \operatorname{Fold}(e, e')$$

where the maximum is taken over all such pairs of edges. The reason to introduce this quantity is that we have

$$\sum_{[v,v'] \in \mathbf{edge}(Y)} d_{T_X \operatorname{rel} B}(\pi(v), \pi(v')) \geq \ell(\phi) + \operatorname{Fold}(\pi) - A'' \tag{2.8}$$

for some positive constant A'' depending only on that number of vertices (and the fact that we chose the balls B to have radius 1) – here, we have identified each edge of Y with one of its representatives in $\widetilde{Y}$.

Now, using Equations (2.5), (2.7) and (2.8), we get that

$$\begin{aligned}
\ell(\psi) &= \sum_{[v,v'] \in \mathbf{edge}(Y)} d_{\mathbb{H}^2}(\tilde{\psi}(v), \tilde{\psi}(v')) \\
&\geq -\mathbf{const} + \sum_{[v,v'] \in \mathbf{edge}(Y)} d_{\mathbb{H}^2}(\pi(v), \pi(v')) \\
&\geq -\mathbf{const} + \sum_{[v,v'] \in \mathbf{edge}(Y)} d_{T_X \operatorname{rel} B}(\pi(v), \pi(v')) \\
&\geq -\mathbf{const} + \ell(\phi) + \operatorname{Fold}(\pi),
\end{aligned}$$

where each **const** is a positive constant (but not necessarily the same) depending on the quasiconvexity constant and combinatorics of Y.

Since $\ell(\psi) \leq \ell(\phi) + C$, it follows that $\operatorname{Fold}(\pi) \leq \mathbf{const}$, where the constant depends on C and on the lower bound for ℓ_0 and on combinatorics of Y. In particular, for each $v \in \mathbf{vert}$ there is a vertex $v' \in T_X$ such that $d_{T_X \operatorname{rel} B}(\pi(v), v') \leq \mathbf{const}$ and we can choose ℓ_0 large such that v' is unique. We have proved Claim 1.

Armed with Claim 1, we can define a map $\widetilde{F} : \widetilde{Y} \to T_X$ by first mapping $v \in \mathbf{vert}(\widetilde{Y})$, to the unique vertex in T_X closest to $\pi(v)$, extending it so that it maps each $e \in \mathbf{edge}(\widetilde{Y})$ with constant speed. Note that equivariance of π implies that $\widetilde{F}$ satisfies $\widetilde{F}(\sigma_*(g)(y)) = g(\widetilde{F}(y))$ for all $y \in \widetilde{Y}$ and $g \in \pi_1(X, x_0)$, where $\sigma_* : \pi_1(X, x_0) \to \pi_1(Y, \sigma(x_0))$ is the isomorphism induced by σ and the chosen base points. It follows that $\widetilde{F}$ descends to a map $F : Y \to X$.

Claim 2. *$F : Y \to X$ is a homeomorphism with $F \circ \sigma$ homotopic to the identity.*

The fact that $F \circ \sigma$ is homotopic to the identity follows directly from the fact $\widetilde{F} \circ \widetilde{\sigma} : \widetilde{X} \to \widetilde{X}$ satisfies that $(\widetilde{F} \circ \sigma)(gx) = g((\widetilde{F} \circ \widetilde{\sigma})(x))$ for $g \in \pi_1(X, x_0)$. What we really have to prove is that F is a homeomorphism. To see that this is the case, note that since the maps $\widetilde{F}, \pi : \widetilde{Y} \to \widetilde{X}$ send points within **const** of each other, and since in our way to proving Claim 1 we proved that $\operatorname{Fold}(\pi) < \mathbf{const}$, we get that $\operatorname{Fold}(\widetilde{F}) < \mathbf{const}$. On the other hand, since $\widetilde{F}$ maps vertices to vertices and has constant speed on

[1] In fact, any radius greater than $\log \sqrt{2} \approx 0.3466$ would work for large ℓ_0.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

the edges we get that if $\mathrm{Fold}(\widetilde{F}) \neq 0$, then $\mathrm{Fold}(F)$ has to be at least as large as the shortest edge of X. This implies, using the fact that $\mathrm{Fold}(\widetilde{F})$ is uniformly bounded, that as long as ℓ_0 is large enough, then $\mathrm{Fold}(\widetilde{F}) = 0$. This implies in turn that $\widetilde{F}$ is locally injective. Local injectivity of $\widetilde{F}$, together with the fact that F, being a homotopy inverse to σ, is a homotopy equivalence, implies that F is a homeomorphism. We have proved Claim 2.

What is left is to bound the tracks of the geodesic homotopy from $\phi \circ F$ to ψ – the argument is similar to the proof of the existence of F. First, note that we know that for every vertex $v \in \widetilde{Y}$, the nearest point projection $\hat{\pi}$ maps $\widetilde{\psi}(v)$ close to the vertex $\tilde{\phi}(\widetilde{F}(v))$ of T_X. It follows that if we choose ℓ_0 large enough we can also assume that $d_{\mathbb{H}^2}(\hat{\pi}(\tilde{\psi}(v)), \hat{\pi}(\tilde{\psi}((v')))) > K$ for any two distinct vertices $v, v' \in \mathbf{vert}(\widetilde{Y})$. This means that Equation (2.6) applies, and hence that we have

$$\begin{aligned}
\ell(\psi(Y)) &= \sum_{[v,v']\in\mathbf{edge}} d_{\mathbb{H}^2}(\tilde{\psi}(v), \tilde{\psi}(v')) \\
&\geq -\mathbf{const} + \sum_{[v,v']\in\mathbf{edge}} d_{\mathbb{H}^2}(\hat{\pi}(\tilde{\psi}(v)), \hat{\pi}(\tilde{\psi}(v')))+ \\
&\quad + 3 \sum_{v\in\mathbf{vert}} d_{\mathbb{H}^2}(\tilde{\psi}(v), \hat{\pi}(\tilde{\psi}(v))) \\
&\geq -\mathbf{const} + \ell(\phi(X)) + 3 \sum_{v\in\mathbf{vert}} d_{\mathbb{H}^2}(\tilde{\psi}(v), \hat{\pi}(\tilde{\psi}(v))),
\end{aligned}$$

where as before each **const** is a positive constant, the first inequality follows by Equation (2.6), the second by Equations (2.7) and (2.8). Hence, again by assumption (1) in the lemma, we must have that $d_{\mathbb{H}^2}(\tilde{\psi}(v), \hat{\pi}(\tilde{\psi}(v))) < \mathbf{const}$ for all $v \in \mathbf{vert}(\widetilde{Y})$. Since $\widetilde{F}(v)$ and $\hat{\pi}(\tilde{\psi}(v))$ are near each other, we have that $d_{\mathbb{H}^2}(\tilde{\psi}(v), \widetilde{F}(v)) < \mathbf{const}$ for all $v \in \mathbf{vert}(\widetilde{Y})$. Convexity of the length function implies that the geodesic homotopy between $\tilde{\psi}$ and $\tilde{\phi} \circ \widetilde{F}$ has then tracks bounded by the same constant **const**, as we had claimed. □

ε-critical realizations

Corollary 2.4 asserts that whenever we have a realization with very long edges and which looks vaguely critical, then it is not far from a critical realization. Our next goal is to give a more precise description of the situation in the case that our realization looks even more as if it were critical. To be precise, suppose that X is a trivalent graph and ε positive and small. A regular realization $\phi : X \to \Sigma$ is *ε-critical* if

$$\angle(\phi(\vec{e}_1), \phi(\vec{e}_2)) \in \left[\frac{2\pi}{3} - \varepsilon, \frac{2\pi}{3} + \varepsilon\right]$$

for every two half-edges $\vec{e}_1, \vec{e}_2 \in \mathbf{half}_v$ incident to any vertex $v \in \mathbf{vert}(X)$.

The goal of this section is to determine how the set $\mathcal{G}^X_{\varepsilon-\mathrm{crit}}$ of ε-critical realizations of X looks. To do that, we need a bit of of notation and preparatory work. First, we understand a *tripod* to be a tuple $\tau = (p, \{v, v', v''\})$, where $p \in \mathbb{H}^2$ is a point and where $\{v, v', v''\} \subset T^1_p\mathbb{H}^2$ is an unordered collection consisting of three distinct unit vectors based at p. A tripod $\tau = (p, \{v, v', v''\})$ is *critical* if the three vectors $v, v', v,''$ have pairwise angle equal to $\frac{2\pi}{3}$. Any tripod τ determines three geodesic rays, that is, three points in $\partial_\infty\mathbb{H}^2$, that is an ideal triangle $T_\tau \subset \mathbb{H}^2$ (see the left-hand side of Figure 1). Conversely, every ideal triangle $T \subset \mathbb{H}^2$ determines a tripod $\tau^T_p = (p, \{v, v', v''\})$ for every $p \in \mathbb{H}^2$: The vectors $v, v', v,''$ are the unit vectors based at p and pointing to the (ideal) vertices of the ideal triangle T. Note that T consists of exactly those points p such that the vectors in the tripod τ^T_p have pairwise (always unoriented) angles adding up to 2π. For a given ε, we let

$$T(\varepsilon) = \left\{ p \in T \,\middle|\, \begin{array}{l} \text{the angles between the vectors} \\ \text{in } \tau^T_p \text{ lie in } [\frac{2\pi}{3} - \varepsilon, \frac{2\pi}{3} + \varepsilon] \end{array} \right\}$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

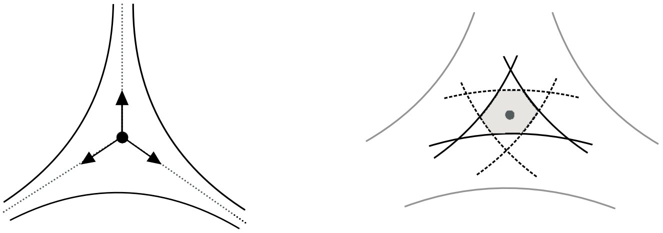

Figure 1. *Left: A critical tripod and (part of) the ideal triangle it determines. Right: The hexagon $T(\varepsilon)$. Each of the lines making up the boundary of the hexagon corresponds to points p having an angle between vectors in τ_P^T of measure $\frac{2\pi}{3} - \varepsilon$ (dotted lines) and $\frac{2\pi}{3} + \varepsilon$ (solid lines).*

be the set of points such that the angles of the tripod τ_p^T are within ε of $\frac{2\pi}{3}$. The set $T(\varepsilon)$ is, at least for $\varepsilon \in (0, \frac{\pi}{6})$, a hexagon centered at the center of the triangle – the sides of the hexagon are not geodesic but rather subsegments of curves of constant curvature; see Figure 1. However, if we scale everything by ε^{-1}, then $\varepsilon^{-1} \cdot T(\varepsilon)$ converges to the regular euclidean hexagon of side length $\frac{2}{3}$. This means in particular that

$$\operatorname{diam}(T(\varepsilon)) \sim \frac{4}{3}\varepsilon \text{ and } \operatorname{vol}(T(\varepsilon)) \sim \frac{2}{\sqrt{3}}\varepsilon^2 \text{ as } \varepsilon \to 0. \tag{2.9}$$

Remark. To get the shape of $T(\varepsilon)$, one can use elementary synthetic hyperbolic geometry. But one can also resort to hyperbolic trigonometry. For example, evoking formula 2.2.2 (iv) in the back of Buser's book [3] one gets

$$T(\varepsilon) = \left\{ p \in T \,\middle|\, \cot\left(\frac{\pi}{3} - \frac{\varepsilon}{2}\right) \le \sinh(d(p, \partial T)) \le \cot\left(\frac{\pi}{3} + \frac{\varepsilon}{2}\right) \right\}.$$

Suppose now that $\phi : X \longrightarrow \Sigma$ is a critical realization of a trivalent graph X in our closed hyperbolic surface, and recall that by Lemma 2.2 we have that the connected component $\mathcal{G}^\phi$ of geodesic realizations homotopic to ϕ is isometric to a product of hyperbolic planes $\mathbb{H}^2 \times \cdots \times \mathbb{H}^2$, one factor for each vertex of X. To each vertex x of X, we can associate first the tripod $\tau_x^\phi = (\phi(x), \{\phi(\vec{e})\}_{\vec{e} \in \mathbf{half}_x(X)})$ consisting of the image under ϕ of the vertex x and of the unit vectors $\phi(\vec{e})$ tangent to the images of the half-edges incident to x, and then $T_{\tau_x^\phi}$ the ideal triangle associated to the critical tripod τ_x^ϕ. The assumption that ϕ is critical implies that the point $\phi(x)$ is the center of $T_{\tau_x^\phi}$ for every vertex $x \in \mathbf{vert}$. We let

$$\mathcal{T}^\phi = \prod_{x \in \mathbf{vert}(X)} T_{\tau_x^\phi} \subset \prod_{x \in \mathbf{vert}(X)} \mathbb{H}^2 = \mathcal{G}^\phi$$

be the subset of $\mathcal{G}^\phi$ consisting of geodesic realizations homotopic to ϕ via a homotopy that maps each vertex x within $T_{\tau_x^\phi}$. Accordingly, we set

$$\mathcal{T}^\phi(\varepsilon) = \prod_{x \in \mathbf{vert}(X)} T_{\tau_x^\phi}(\varepsilon) \subset \mathcal{T}^\phi$$

and we note that

$$\operatorname{vol}(\mathcal{T}^\phi(\varepsilon)) \sim \left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^{|\mathbf{vert}(X)|}.$$

The reason we care about all these sets is that, asymptotically, $\mathcal{T}^\phi(\varepsilon)$ agrees with the set $\mathcal{G}^\phi_{\varepsilon\text{-crit}}$ of ε-critical realizations ψ homotopic to ϕ. Indeed, we get from Corollary 2.4 that the geodesic homotopy

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

from ψ to ϕ has lengths bounded by a uniform constant. This means that for all $e \in \mathbf{edge}(X)$ the geodesic segments $\phi(e)$ and $\psi(e)$ are very long but have endpoints relatively close to each other. This means that, when looked from (each) one of its vertices x, the segment $\psi(e)$ is very close to being asymptotic to $\phi(e)$. In particular, the angle between $\psi(e)$ and the geodesic ray starting at $\psi(x)$ and asymptotic to the ray starting in $\phi(x)$ in direction $\phi(e)$ is smaller than $\delta = \delta(\min_{e\in\mathbf{edge}(X)} \ell(\psi(e)))$ for some positive function with $\delta(t) \to 0$ as $t \to \infty$. With the rather clumsy notation, we find ourselves working with we have that δ bounds from above the angles between corresponding edges of the tripods τ_x^{ψ} and $\tau_{\psi(x)}^{T_x^{\phi}}$. Since by assumption the angles of τ_x^{ψ} belong to $[\frac{2\pi}{3} - \varepsilon, \frac{2\pi}{3} + \varepsilon]$, we get that the angles of $\tau_{\psi(x)}^{T_x^{\phi}}$ belong to $[\frac{2\pi}{3} - \varepsilon - 2\delta, \frac{2\pi}{3} + \varepsilon + 2\delta]$, meaning that $\psi(x) \in T_{\tau_x^{\phi}}(\varepsilon + 2\delta)$.

To summarize, what we have proved is that for all $\varepsilon_1 > \varepsilon$ there is ℓ_1 such that if $\phi : X \to \Sigma$ is critical with $\ell(\phi(e)) \geqslant \ell_1$ for all $e \in \mathbf{edge}$, then

$$\mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}} \subset \mathcal{T}^{\phi}(\varepsilon_1).$$

The same argument proves that for all $\varepsilon_2 < \varepsilon$ there is ℓ_2 such that if $\phi : X \to \Sigma$ is critical with $\ell(\phi(e)) \geqslant \ell_2$ for all $e \in \mathbf{edge}$, then

$$\mathcal{T}^{\phi}(\varepsilon_2) \subset \mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}}$$

We record these facts:

Lemma 2.5. *There are functions $h : (0, \frac{\pi}{6}) \to \mathbb{R}_{>0}$ and $\delta : (0, \varepsilon_0) \to \mathbb{R}_{>0}$ with $\lim_{t\to 0} h(t) = 0$ and $\lim_{t\to\infty} \delta(t) = 0$ such that for every critical realizations $\phi : X \to \Sigma$ we have*

$$\mathcal{T}^{\phi}(\varepsilon - r(\varepsilon, \phi)) \subset \mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}} \subset \mathcal{T}^{\phi}(\varepsilon + r(\varepsilon, \phi)),$$

where $r(\varepsilon, \phi) = h(\varepsilon) + \delta(\min_{e\in\mathbf{edge}(X)} \ell(\phi(e)))$.

To conclude this section, we collect what we will actually need about the set of ε-critical realizations in a single statement.

Proposition 2.6. *Let X be a trivalent graph. There are $\ell > 0$ and functions $\rho_0, \rho_1 : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ with $\lim_{t\to 0} \rho_0(t) = 0$ and $\lim_{t\to\infty} \rho_1(t) = 0$ and such that the following holds for all $\varepsilon \in [0, \frac{\pi}{6}]$:*

If $\phi \in \mathcal{G}_{\varepsilon-\mathrm{crit}}(X)$ is an ℓ-long ε-critical realization, then

$$\left| \mathrm{vol}(\mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}}) - \left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^V \right| \leqslant \rho_0(\varepsilon) \cdot \rho_1\left(\min_{e\in\mathbf{edge}} \ell(\phi(e)) \right) \cdot \varepsilon^{2V}. \tag{2.10}$$

Moreover, $\mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}}$ contains a unique critical realization ψ and we have

$$\max_{e\in\mathbf{edge}} |\ell_{\Sigma}(\phi(e)) - \ell_{\Sigma}(\psi(e))| \leqslant \rho_0(\varepsilon) + \rho_1\left(\min_{e\in\mathbf{edge}} \ell(\phi(e)) \right) \cdot \varepsilon. \tag{2.11}$$

Here, as before, V is the number of vertices of X and $\mathbf{edge} = \mathbf{edge}(X)$ *is its set of edges.*

Proof. Suppose to begin with that ℓ is at least as large as the ℓ_0 in Corollary 2.4. We thus get that $\mathcal{G}^{\phi}$ contains a unique critical realization ψ. Now, we get from Lemma 2.5 that

$$\mathcal{T}^{\phi}(\varepsilon - r(\varepsilon, \phi)) \subset \mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}} \subset \mathcal{T}^{\phi}(\varepsilon + r(\varepsilon, \phi)), \tag{2.12}$$

where $r(\varepsilon, \phi) = h(\varepsilon) + \delta(\min_{e\in\mathbf{edge}} \ell(\phi(e)))$ for some functions $h : (0, \frac{\pi}{6}) \to \mathbb{R}_{>0}$ with and $\delta : (0, \varepsilon_0) \to \mathbb{R}_{>0}$ with $\lim_{t\to\infty} h(t) = 0$ and $\lim_{t\to\infty} \delta(t) = 0$. The bounds (2.10) and (2.11) follow now from Equations (2.12) and (2.9). □

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

3. Variations of Delsarte's theorem

In this section, we establish the asymptotics of the number of realizations $\phi : X \to \Sigma$ satisfying some length condition, mapping each vertex to some predetermined point in Σ, and so that the tuple of directions of images of half-edges belongs to some given open set of such tuples – see Theorem 3.2 for details. Other than some bookkeeping, what is needed is a slight extension of Delsarte's lattice counting theorem [7], namely Theorem 3.1 below. This result is known to experts, and much more sophisticated versions than what we need here can be found in the literature – see, for example, [17, Théorème 4.1.1]. However, for the sake of completeness, we will explain how to derive Theorem 3.1 from the fact that the geodesic flow is mixing. We refer to the very nice books [2, 17] for the needed background.

We start by recalling a few well-known facts about dynamics of geodesic flows on hyperbolic surfaces. First, recall that we can identify $T^1\mathbb{H}^2$ with $\mathrm{PSL}_2\,\mathbb{R}$, and $T^1\Sigma = T^1(\Gamma\backslash\mathbb{H}^2)$ with $\Gamma\backslash\,\mathrm{PSL}_2\,\mathbb{R}$. More specifically, when using the identification

$$\mathrm{PSL}_2\,\mathbb{R} \to T^1\mathbb{H}^2 \quad g \mapsto \frac{d}{dt} g(e^t i)|_{t=0}$$

where the computation happens in the upper half-plane model and $i \in \mathbb{H}^2$ is the imaginary unit, then the geodesic and horocyclic flows amount to right multiplication by the matrices

$$\rho_t = \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix} \text{ and } h_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix},$$

respectively.

Note that $K = \mathrm{SO}_2 \subset \mathrm{PSL}_2\,\mathbb{R}$, the stabilizer of i under the action $\mathrm{PSL}_2\,\mathbb{R} \curvearrowright \mathbb{H}^2$, is a maximal compact subgroup – K corresponds under the identification $\mathrm{PSL}_2\,\mathbb{R} \simeq T^1\mathbb{H}^2$ to the unit tangent space $T_i^1\mathbb{H}^2$ of the base point $i \in \mathbb{H}^2$. The KAN decomposition (basically the output of the Gramm–Schmidt process from linear algebra) asserts that every element in $g \in \mathrm{PSL}_2\,\mathbb{R}$ can be written in a unique way as

$$g = k\rho_t h_s \tag{3.1}$$

for $k \in K$ and $t, s \in \mathbb{R}$. In those coordinates, the Haar measure of $\mathrm{PSL}_2\,\mathbb{R}$ is given by

$$\int f(g)\ \mathrm{vol}_{T^1\mathbb{H}^2}(g) = \iiint f(k\rho_t h_s) \cdot e^{-t}\, dkdtds$$

where dk stands for integrating over the arc length in $K \simeq T_i^1\mathbb{H}^2$, normalized to have total measure 2π.

The basic fact we will need is that the geodesic flow ρ_t on $T^1\Sigma$ is mixing [2, III.2.3], or rather one of its direct consequences, namely the fact that for each $x_0 \in \mathbb{H}^2$ the projection to Σ of the sphere $S(x_0, L)$ centered at x_0 and with radius L gets equidistributed in Σ when $L \to \infty$ [2, III.3.3]. To be more precise, note that we might assume without loss of generality that $x_0 = i$, meaning that we are identifying $T_{x_0}^1\mathbb{H}^2 = K$. The fact that the geodesic flow is mixing implies then that for every nondegenerate interval $I \subset K$ the spherical arcs $I\rho_t$ equidistribute in $\Gamma\backslash\,\mathrm{PSL}_2\,\mathbb{R}$ in the sense that for any continuous function $f \in C^0(T^1\Sigma) = C^0(\Gamma\backslash\,\mathrm{PSL}_2\,\mathbb{R})$ we have

$$\lim_{L\to\infty} \frac{1}{\ell(I)} \int_I f(\Gamma k\rho_L)\, dk = \frac{1}{\mathrm{vol}_{T^1\Sigma}(T^1\Sigma)} \int_{T^1\Sigma} f(\Gamma g) \cdot dg, \tag{3.2}$$

where $\ell(I)$ is the arc length of I and where the second integral is with respect to $\mathrm{vol}_{T^1\Sigma}$. Anyways, we care about equidistribution of the spheres for the following reason. If $f \in C^0(\Sigma)$ is a continuous function and if we let $\tilde{f}$ be the composition of f with the cover $\mathbb{H}^2 \to \Sigma$, then Equation (3.2) implies, with $I = K$, that

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

$$\lim_{L\to\infty}\frac{1}{\mathrm{vol}_{\mathbb{H}^2}(B(x_0,L))}\cdot\int_{B(x_0,L)}\tilde{f}(x)\,dx=\frac{1}{\mathrm{vol}(\Sigma)}\int_{\Sigma}f(x)dx.$$

Applying this to a nonnegative function $f_{y_0,\varepsilon}\in C^0(\Sigma)$ with total integral 1 and supported by $B(y_0,\varepsilon)$, we get that

$$\int_{B(x_0,L)}\tilde{f}_{y_0,\varepsilon}(x)\,dx\sim\frac{\mathrm{vol}_{\mathbb{H}^2}(B(x_0,L))}{\mathrm{vol}(\Sigma)}.$$

Now, from the properties of $f_{y_0,\varepsilon}$ we get that

$$\int_{B(x,L-\varepsilon)}\tilde{f}_{y_0,\varepsilon}(\cdot)d\,\mathrm{vol}_{\mathbb{H}^2}\leqslant|\Gamma\cdot y_0\cap B(x_0,L)|\leqslant\int_{B(x,L+\varepsilon)}\tilde{f}_{y_0,\varepsilon}(\cdot)d\,\mathrm{vol}_{\mathbb{H}^2}\,. \tag{3.3}$$

When taken together, the last two displayed equations imply that for all ε one has

$$\frac{\pi\cdot e^{L-\varepsilon}}{\mathrm{vol}(\Sigma)}\leqslant|\Gamma\cdot y_0\cap B(x_0,L)|\leqslant\frac{\pi\cdot e^{L+\varepsilon}}{\mathrm{vol}(\Sigma)}$$

and hence that

$$|\Gamma\cdot y_0\cap B(x_0,L)|\sim\frac{\pi\cdot e^{L}}{\mathrm{vol}(\Sigma)}\text{ as }L\to\infty. \tag{3.4}$$

This is Delsarte's lattice point counting theorem [7] – we refer to [2, III.3.5] for more details.

An observation: Note that counting elements of $\Gamma\cdot y_0$ contained in the ball $B(x_0,L)$ is exactly the same thing as counting geodesic arcs in Σ of length at most L going from x_0 to y_0, or to be precise, from the projection of x_0 to the projection of y_0. In this way, if we are given a segment $I\subset K=T^1_{x_0}\Sigma$ and we use the equidistribution of the spherical segments $I\rho_t$, then when we run the argument above we get the cardinality of the set $\mathbf{A}_{I,y_0}(L)$ of geodesic arcs of length at most L, starting in x_0 with initial speed in I, and ending in y_0. As expected, the result is that

$$|\mathbf{A}_{I,y_0}(L)|\sim\frac{\ell(I)}{2\pi}\cdot\frac{\pi\cdot e^{L}}{\mathrm{vol}(\Sigma)}\quad\text{when }L\to\infty.$$

Suppose now that we dial it up a bit giving ourselves a second sector $J\subset T^1_{y_0}\Sigma$ and care, for some fixed h, about the cardinality of the set $\mathbf{A}_{I,J}(L,h)$ of geodesic arcs with length in $[L,L+h]$, joining x_0 to y_0, and with initial and terminal velocities in I and J, respectively. We can obtain the asymptotics of $|\mathbf{A}_{I,J}(L,h)|$ following the same basic idea as in the proof of Delsarte's theorem above. We need, however, to replace the bump function $f_{y_0,\varepsilon}$ by something else.

Using the coordinates (3.1), consider for h and δ positive and small the set

$$\mathcal{J}(J,h,\delta)=\{J\rho_t h_s\text{ with }s\in[-\delta,\delta]\text{ and }t\in[0,h]\}\subset T^1\Sigma$$

and note that it has volume $\ell(J)\cdot 2\delta\cdot(1-e^{-h})$. Note also that the intersection of $\mathcal{J}(J,h,\delta)$ with the outer normal vector field of any horosphere consists of segments of the form $H_{v\rho_t}=\{v\rho_t h_s\text{ with }s\in[-\delta,\delta]\}$, where $v\in J$ and $t\in[0,h]$ and that each such segment has length 2δ. Finally, observe that each one of the sets $H_{v\rho_t}$ contains exactly one vector, namely $v\rho_t$, which lands in J when we geodesic flow it for some time in $[-h,0]$. It follows that for all intervals $\mathbb{I}\subset\mathbb{R}$ and $w\in T^1\Sigma$ we have

$$\left|\left|\left\{r\in\mathbb{I}\,\middle|\,\begin{array}{c}\exists s\in[-\delta,\delta],t\in[-h,0]\\ \text{with }wh_s\rho_t\in J\end{array}\right\}\right|-\frac{\ell(\{s\in\mathbb{I}\text{ with }wh_s\in\mathcal{J}\})}{2\delta}\right|\leqslant 2,$$

where $\mathcal{J} = \mathcal{J}(J, h, \delta)$ and where the number 2 is there to take into account possible over counting near the ends of the horospherical segment.

Using then the fact that when $L \to \infty$ the sphere $S(x_0, L + h) \subset \mathbb{H}^2$ looks more and more like a horosphere, one gets the following analogue of Equation (3.3): For any $\delta' < \delta < \delta''$ and $h' < h < h''$ and $J' \Subset J \Subset J''$, we have for all sufficiently large L

$$\int_{I\rho_{L+h}} \chi_{\mathcal{J}(J',h',\delta')} \leqslant 2\delta \cdot |\mathbf{A}_{I,J}(L,h)| \leqslant \int_{I\rho_{L+h}} \chi_{\mathcal{J}(J'',h'',\delta'')},$$

where $\chi_{\mathcal{J}}$ is the characteristic function of $\mathcal{J}$. From here, we get that

$$\begin{aligned}|\mathbf{A}_{I,J}(L,h)| &\sim \frac{1}{2\delta}\int_{I\rho_{L+h}} \chi_{\mathcal{J}(J,h,\delta)} \sim \frac{1}{2\delta}\cdot\frac{e^{L+h}}{2}\cdot\int_I \chi_{\mathcal{J}(J,h,\delta)}(k\rho_{L+h})\, dk \\ &\overset{(3.2)}{\sim} \frac{1}{2\delta}\cdot\frac{e^{L+h}}{2}\cdot\ell(I)\cdot\frac{\mathrm{vol}_{T^1\Sigma}(\mathcal{J}(J,h,\delta))}{\mathrm{vol}_{T^1\Sigma}(T^1\Sigma)} \\ &\sim \frac{\ell(I)\cdot\ell(J)}{4\pi}\cdot\frac{e^{L+h}-e^L}{\mathrm{vol}(\Sigma)}.\end{aligned}$$

We record this slightly generalized version of Delsarte's theorem for later reference:

Theorem 3.1 (Delarte's theorem for in-out sectors). *Let Σ be a closed hyperbolic surface, let h be positive and let $I \subset T_{x_0}\Sigma$ and $J \subset T_{y_0}\Sigma$ be nondegenerate segments. Let then $\mathbf{A}_{I,J}(L,h)$ be the set of geodesic arcs $\alpha : [0, r] \to \Sigma$ with length $r \in [L, L + h]$, with endpoints $\alpha(0) = x_0$ and $\alpha(r) = y_0$, and with initial and terminal speeds satisfying $\alpha'(0) \in I$ and $\alpha'(r) \in J$. Then we have*

$$|\mathbf{A}_{I,J}(L,h)| \sim \frac{\ell(I)\cdot\ell(J)}{4\pi}\cdot\frac{e^{L+h}-e^L}{\mathrm{vol}(\Sigma)}$$

when $L \to \infty$.

As we mentioned earlier, Theorem 3.1 is a very special case of the much more general [17, Théorème 4.1.1]. We would not be surprised if there were also other references we are not aware of.

Remark. Note also that the asymptotic behavior in Theorem 3.1 is uniform as long as I and J belong to a compact set of choices. For example, since the surface Σ is compact, we get that for all $\delta > 0$ the asymptotic behavior in Theorem 3.1 is uniform as long as I and J are intervals of length at least δ contained, respectively, in $T^1_{x_I}\Sigma$ and $T^1_{x_J}\Sigma$ for some x_I and x_J in Σ.

Now, let X be a trivalent graph with vertex set $\mathbf{vert}(X)$ and let $\vec{x} = (x_v)_{v\in\mathbf{vert}(X)} \in \Sigma^{\mathbf{vert}(X)}$ be a $\mathbf{vert}(X)$-tuple of points in the surface. Let also

$$U \subset \prod_{v\in\mathbf{vert}(X)}\left(\bigoplus_{\vec{e}\in\mathbf{half}_v(X)} T^1_{x_v}\Sigma\right) \overset{\text{def}}{=} \mathbb{T}_{\vec{x}}$$

be an open set where, as always, $\mathbf{half}_v(X)$ is the set of all half-edges of X starting at the vertex v.

Given for each edge $e \in \mathbf{edge}(X)$ a positive number L_e and writing $\vec{L} = (L_e)_{e\in\mathbf{edge}(X)}$, we are going to be interested in the set $\mathbf{G}^X_U(\vec{L}_e, h)$ of realizations $\phi : X \to \Sigma$

(a) mapping each vertex $v \in \mathbf{vert}(X)$ to the point x_v,
(b) with $(\phi(\vec{e}))_{\vec{e}\in\mathbf{half}(X)} \in U$ and
(c) with $\ell(e) \in [L_e, L_e + h]$ for all $e \in \mathbf{edge}(X)$.

In this setting, we have the following version of Delsarte's theorem relating the cardinality of $\mathbf{G}^X_U(\vec{L}_e, h)$ with the volume $\mathrm{vol}_{\vec{x}}(U)$ of U, where the volume is measured in the flat torus $\mathbb{T}_{\vec{x}}$.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Theorem 3.2 (Delarte's theorem for graph realizations). *Let Σ be a closed hyperbolic surface, let X be a finite graph, fix $\vec{x} \in \Sigma^{\mathbf{vert}(X)}$ and an open set $U_{\vec{x}} \subset \mathbb{T}_{\vec{x}}$. If $\mathrm{vol}_{\vec{x}}(\bar{U}_{\vec{x}} \setminus U_{\vec{x}}) = 0$, then for every $h > 0$ we have*

$$|\mathbf{G}^X_{U_{\vec{x}}}(\vec{L}, h)| \sim \frac{\mathrm{vol}_{\vec{x}}(U_{\vec{x}})}{(4\pi)^E} \cdot \frac{(e^h - 1)^E \cdot e^{\|\vec{L}\|}}{\mathrm{vol}(\Sigma)^E}$$

when $\min_{e \in \mathbf{edge}(X)} L_e \to \infty$. Here, $\bar{U}_{\vec{x}}$ is the closure of $U_{\vec{x}}$ in $\mathbb{T}_{\vec{x}}$ and $\mathrm{vol}_{\vec{x}}(\cdot)$ stands for the volume therein. Also, $E = |\,\mathbf{edge}(X)|$ is the number of edges in X, and $\|\vec{L}\| = \sum_{e \in \mathbf{edge}(X)} L_e$.

We stress that $\mathrm{vol}_{\vec{x}}$ is normalized in such a way that $\mathrm{vol}_{\vec{x}}(\mathbb{T}_{\vec{x}}) = (2\pi)^{2E}$. We also stress that this is consistent with the fact that we use the interval $[0, \pi]$ to measure unoriented angles.

Proof. Let us say that a closed set of the form

$$\prod_{v \in \mathbf{vert}(X)} \left(\prod_{\vec{e} \in \mathbf{half}_v(X)} I_{\vec{e}} \right) \subset \mathbb{T}_{\vec{x}} = \prod_{v \in \mathbf{vert}(X)} \left(\bigoplus_{\vec{e} \in \mathbf{half}_v(X)} T^1_{x_v} \Sigma \right),$$

where each $I_{\vec{e}}$ is a segment in $T^1_{x_v}\Sigma$ is *a cube*. We say that a closed subset of $\prod_{v \in \mathbf{vert}(X)} \left(\oplus_{\vec{e} \in \mathbf{half}_v(X)} T^1_{x_v}\Sigma \right)$ it *cubical* if it can be given as the union of finitely many cubes with disjoint interiors. The assumption that the open set $U = U_{\vec{x}}$ in the statement is such that $\bar{U} \setminus U$ is a null-set implies that U can be approximated from inside and outside by cubical sets $U' \subset U \subset U''$ with $\mathrm{vol}(U'') - \mathrm{vol}(U')$ as small as we want. It follows that it suffices to prove the theorem if U is (the interior of) a cubical set.

Note now that if $U = U_1 \cup U_2$ is the disjoint union of two sets U_1 and U_2 and if the statement of the theorem holds true for U_1 and U_2, then it also holds true for U. Since every cubical set is made out of finitely many cubes with disjoint interior, we deduce that it really suffices to prove the theorem for individual cubes

$$U = \prod_{v \in \mathbf{vert}(X)} \left(\prod_{\vec{e} \in \mathbf{half}_v(X)} I_{\vec{e}} \right).$$

Note that, up to shuffling the factors we can see U as

$$U = \prod_{e \in \mathbf{edge}(X)} (I_{e^+} \times I_{e^-}).$$

Here, we are denoting by e^+ and e^- the two half-edges of the edge $e \in \mathbf{edge}(X)$. Now, unpacking the notation one sees that a realization ϕ belongs to $\mathbf{G}_U(\vec{L}_e, h)$ if and only if for all $e \in \mathbf{edge}(X)$ the arc $\phi(e)$ belongs to $\mathbf{A}_{I_{e^+}, I_{e^-}}(L_e, h)$. It follows thus from Delsarte's theorem for in-out sectors that

$$\begin{aligned}
|\mathbf{G}^X_U(\vec{L}, h)| &= \prod_{e \in \mathbf{edge}(X)} \left| \mathbf{A}_{I_{e^+}, I_{e^-}}(L_e, h) \right| \\
&\sim \prod_{e \in \mathbf{edge}(X)} \left(\frac{\ell(I_{e^+}) \cdot \ell(I_{e^-})}{4\pi} \cdot \frac{(e^h - 1) \cdot e^{L_e}}{\mathrm{vol}(\Sigma)} \right) \\
&= \frac{\prod_{e \in \mathbf{edge}(X)} (\ell(I_{e^+}) \cdot \ell(I_{e^-}))}{(4\pi)^E} \cdot \frac{(e^h - 1)^E \cdot e^{\sum_{e \in \mathbf{edge}(X)} L_e}}{\mathrm{vol}(\Sigma)^E} \\
&= \frac{\mathrm{vol}(U)}{(4\pi)^E} \cdot \frac{(e^h - 1)^E \cdot e^{\|\vec{L}\|}}{\mathrm{vol}(\Sigma)^E},
\end{aligned}$$

where, as in the statement we have set $E = |\,\mathbf{edge}(X)|$. We are done. □

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Let us consider now the concrete case we will care mostly about. Suppose namely that X is a trivalent graph and that we want all the angles to be between $\frac{2}{3}\pi-\varepsilon$ and $\frac{2}{3}\pi+\varepsilon$. In other words, we are interested in the set of ε-critical realizations $\phi : X \to \Sigma$ which map the vertices to our prescribed tuple $\vec{x} \in \Sigma^{\mathbf{vert}(X)}$. Then we have to consider the set

$$U^X_{\vec{x},\varepsilon-\mathrm{crit}} \subset \prod_{v\in\mathbf{vert}(X)}\left(\bigoplus_{\vec{e}\in\mathbf{half}_v(X)} T^1_{x_v}\Sigma\right)$$

of those tuples $(v_{\vec{e}})_{\vec{e}\in\mathbf{half}(X)}$ with $\angle(v_{\vec{e}_1}, v_{\vec{e}_2}) \in [\frac{2\pi}{3}-\varepsilon, \frac{2\pi}{3}+\varepsilon]$ for all distinct $\vec{e}_1, \vec{e}_2 \in \mathbf{half}(X)$ incident to the same vertex.

Let us compute the volume of $U^X_{\vec{x},\varepsilon-\mathrm{crit}}$. Noting that the conditions on $U^X_{\vec{x},\varepsilon-\mathrm{crit}}$ associated to different vertices are independent, we get that it suffices to think vertex-by-vertex and then multiply all the numbers obtained for all vertices. For each vertex v, label arbitrarily the half-edges incident to v by $\vec{e}_1, \vec{e}_2$ and $\vec{e}_3$. We have no restriction for the position of $v_{\vec{e}_1} \in T_{x_v}\Sigma$. Once we have fixed $v_{\vec{e}_1} \in T^1_{x_v}(\Sigma)$, we get that $v_{\vec{e}_2}$ can belong to two segments, each one of length 2ε, in $T^1_{x_v}(\Sigma)$ – recall that all our angles are unoriented. Then, once we have fixed $v_{\vec{e}_1}$ and $v_{\vec{e}_2}$, we have to choose $v_{\vec{e}_3}$ in an interval of length $2\varepsilon - |\angle(v_{\vec{e}_1}, v_{\vec{e}_2}) - \frac{2\pi}{3}|$. This means that when we have chosen $v_{\vec{e}_1}$ and which segment $v_{\vec{e}_2}$ is in, then we have $3\varepsilon^2$ worth of choices of $(v_{\vec{e}_2}, v_{\vec{e}_3})$. This means that the set of possible choices in $T^1_{x_v}\Sigma\times T^1_{x_v}\Sigma\times T^1_{x_v}\Sigma$ has volume equal to $12\pi\varepsilon^2$. Since, as we already mentioned earlier, the conditions at all vertices of X are independent, we get that

$$\mathrm{vol}_{\vec{x}}(U^X_{\vec{x},\varepsilon-\mathrm{crit}}) = (12\pi\varepsilon^2)^{|\mathbf{vert}(X)|}.$$

From Theorem 3.2, we get thus that for all $h > 0$ and all $\vec{x} \in \Sigma^{\mathbf{vert}(X)}$ we have

$$|\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit}}(\vec{L}, h)| \sim \frac{(12\pi\varepsilon^2)^V}{(4\pi)^E}\cdot\frac{(e^h-1)^E\cdot e^{\|\vec{L}\|}}{\mathrm{vol}(\Sigma)^E} \text{ as } \min_{e\in\mathbf{edge}(X)} L_e \to \infty,$$

where we are writing V and E for the number of vertices and edges of X, and $\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit}}$ instead of $\mathbf{G}^X_{U_{\vec{x},\varepsilon-\mathrm{crit}}}$. Taking into account that X is trivalent and hence satisfies $V = -2\chi(X)$ and $E = -3\chi(X)$, we can clean this up to

$$|\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit}}(\vec{L}, h)| \sim \varepsilon^{4|\chi(X)|}\cdot\left(\frac{2}{3}\right)^{2\chi(X)}\cdot\pi^{\chi(X)}\cdot\frac{(e^h-1)^{-3\chi(X)}\cdot e^{\|\vec{L}\|}}{\mathrm{vol}(\Sigma)^{-3\chi(X)}}$$

as $\min_{e\in\mathbf{edge}(X)} L_e \to \infty$. Moreover, since the geometry of the set $U^X_{\vec{x}}(\varepsilon-\mathrm{crit})$ is independent of the point $\vec{x}$, we get that the speed of convergence to this asymptotic is independent of $\vec{x}$. Altogether, we have the following result:

Corollary 3.3. *Let Σ be a closed hyperbolic surface and X be a trivalent graph, fix $\varepsilon > 0$ and $h > 0$ and for $\vec{x} \in \Sigma^{\mathbf{vert}(X)}$, let $\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit}}(\vec{L}, h)$ be the set of ε-critical realizations $\phi : X \to \Sigma$ mapping the vertex v to the point $x_v = \phi(v)$. Then we have*

$$|\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit}}(\vec{L}, h)| \sim \varepsilon^{4|\chi(X)|}\cdot\left(\frac{2}{3}\right)^{2\chi(X)}\cdot\pi^{\chi(X)}\cdot\frac{(e^h-1)^{-3\chi(X)}\cdot e^{\|\vec{L}\|}}{\mathrm{vol}(\Sigma)^{-3\chi(X)}} \tag{3.5}$$

as $\min_{e\in\mathbf{edge}(X)} L_e \to \infty$.

Note that the set $U^X_{\vec{x},\varepsilon-\mathrm{crit}}$ needed to establish Corollary 3.3 is basically the same for all choices of $\vec{x}$. For example, if we take another $\vec{y} \in \Sigma^{\mathbf{vert}(X)}$ and we identity $\mathbb{T}_{\vec{x}}$ with $\mathbb{T}_{\vec{y}}$ isometrically by parallel transport (along any collection of curves whatsoever), then $U^X_{\vec{x},\varepsilon-\mathrm{crit}}$ is sent to $U^X_{\vec{y},\varepsilon-\mathrm{crit}}$. It follows that we can approximate $U^X_{\vec{x},\varepsilon-\mathrm{crit}}$ and $U^X_{\vec{y},\varepsilon-\mathrm{crit}}$ at the same time by cubical sets consisting of the same number of cubes with the same side lengths. It thus follows from the comment following Theorem 3.1 that the asymptotics in Corollary 3.3 is uniform in $\vec{x}$. We record this fact:

Addendum to Corollary 3.3. The asymptotics in Equation (3.5) is uniform in $\vec{x} \in \Sigma^{\mathbf{vert}(X)}$.

4. Counting critical realizations

In this section, we prove Theorem 1.3 from the introduction. Before restating the theorem, recall that if X is a trivalent graph then we denote by

$$\mathbf{G}^X(L) = \left\{ \begin{array}{c} \phi : X \to \Sigma \text{ critical realization} \\ \text{with length } \ell_\Sigma(\phi) \leqslant L \end{array} \right\} \tag{4.1}$$

the set of all critical realizations of length $\ell_\Sigma(\phi)$ at most L.

Theorem 1.3. *Let Σ be a closed, connected and oriented hyperbolic surface. For every connected trivalent graph X, we have*

$$|\mathbf{G}^X(L)| \sim \left(\frac{2}{3}\right)^{3\chi(X)} \cdot \frac{\mathrm{vol}(T^1\Sigma)^{\chi(X)}}{(-3\chi(X)-1)!} \cdot L^{-3\chi(X)-1} \cdot e^L$$

as $L \to \infty$.

Fixing for the remaining of this section the trivalent graph X, we will write $\mathbf{vert} = \mathbf{vert}(X)$ and $\mathbf{edge} = \mathbf{edge}(X)$ for the sets of vertices and edges and $V = |\,\mathbf{vert}\,|$ and $E = |\,\mathbf{edge}\,|$ for their cardinalities. Similarly, we will denote by $\mathcal{G} = \mathcal{G}^X$ the manifold of realizations of X in Σ, and by $\mathbf{G}(L) = \mathbf{G}^X(L)$ the set of critical realizations with length at most L.

The main step in the proof of Theorem 1.3 is to count critical realizations of X such that the corresponding vectors of lengths $(\ell(\phi(e)))_{e\in\mathbf{edge}}$ belong to a box of size $h > 0$. More concretely, we want to count how many elements there are in the set

$$\mathbf{G}(\vec{L}, h) = \left\{ \begin{array}{c} \phi : X \to \Sigma \text{ critical realization with} \\ \ell(\phi(e)) \in (L_e, L_e + h] \text{ for all } e \in \mathbf{edge} \end{array} \right\}, \tag{4.2}$$

where $\vec{L} = (L_e)_{e\in\mathbf{edge}}$ is a positive vector. We start by establishing some form of an upper bound for the number of homotopy classes of realizations when we bound the length of each individual edge – recall that by Lemma 2.2 any two homotopic critical realizations are identical.

Lemma 4.1. *For all $\vec{L} \in \mathbb{R}_+^{\mathbf{edge}(X)}$, there are at most* $\mathbf{const} \cdot e^{\|\vec{L}\|}$ *homotopy classes of realizations $\phi : X \to \Sigma$ with $\ell(\phi(e)) \leqslant L_e$ for all $e \in$* $\mathbf{edge}(X)$.

It is worth pointing out that Lemma 4.1 fails if Σ is allowed to have cusps – see Section 9.

Proof. Let us fix a point $x_0 \in \Sigma$, and note that every point in Σ – think of the images under a realization of the vertices of X – can be moved to x_0 along a path of length at most $\mathrm{diam}(\Sigma)$. It follows that every realization $\phi : X \to \Sigma$ is homotopic to a new realization $\psi : X \to \Sigma$ mapping all vertices of X to x_0 and with

$$\ell(\psi(e)) \leqslant \ell(\phi(e)) + 2 \cdot \mathrm{diam}(\Sigma) \leqslant L_e + 2 \cdot \mathrm{diam}(\Sigma) \tag{4.3}$$

for every edge $e \in \mathbf{edge}(X)$. Note that the homotopy class of ψ is determined by the homotopy classes of the loops $\psi(e)$ when e ranges over the edges of X. Now, Equation (4.3) implies that, up to homotopy, we have at most $\mathbf{const} \cdot e^{L_e}$ choices for the geodesic segment $\psi(e)$. This implies that there are at most $\mathbf{const} \cdot e^{\|\vec{L}\|}$ choices for the homotopy class of ψ, and hence for the homotopy class of ϕ. We are done. □

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Although it is evidently pretty coarse, Lemma 4.1 will play a key role in the proof of Theorem 1.3. However, the main tool in the proof of the theorem is the following:

Proposition 4.2. *For all $h > 0$, we have*

$$|\mathbf{G}(\vec{L}, h)| \sim \frac{2^{4\chi(X)}}{3^{3\chi(X)}} \cdot \pi^{\chi(X)} \cdot \frac{(e^h - 1)^{-3\chi(X)} \cdot e^{\|\vec{L}\|}}{\mathrm{vol}(\Sigma)^{-\chi(X)}}$$

as $\min_{e \in \mathbf{edge}} L_e \to \infty$. *Here,* $\mathbf{G}(\vec{L}, h)$ *is as in Equation (4.2).*

Proof. Denote by $\mathcal{G}(\vec{L}, h) \subset \mathcal{G}$ the set of all realizations $\phi : X \to \Sigma$ with $\ell_\Sigma(\phi) \in [L_e, L_e + h]$ for all $e \in \mathbf{edge}$ and then let

$$G_\varepsilon(\vec{L}, h) = \left\{\mathcal{G}^\phi \in \pi_0(\mathcal{G}) \text{ with } \mathcal{G}^\phi_{\varepsilon-\mathrm{crit}} \subset \mathcal{G}(\vec{L}, h)\right\}$$
$$\hat{G}_\varepsilon(\vec{L}, h) = \left\{\mathcal{G}^\phi \in \pi_0(\mathcal{G}) \text{ with } \mathcal{G}^\phi_{\varepsilon-\mathrm{crit}} \cap \mathcal{G}(\vec{L}, h) \neq \emptyset\right\}$$

be the sets of connected components of $\mathcal{G}$ whose set of ε-critical realizations is fully contained in (resp. which meet) $\mathcal{G}(\vec{L}, h)$. It follows from Corollary 2.4 that there is some ℓ_0 such that as long as $\vec{L}$ satisfies that $\min L_e \geqslant \ell_0$, then each component listed in $\hat{G}_\varepsilon(\vec{L}, h)$ contains exactly one critical realization of the graph X. Assuming from now on that we are in this situation, we get that

$$|G_\varepsilon(\vec{L}, h)| \leqslant |\mathbf{G}(\vec{L}, h)| \leqslant |\hat{G}_\varepsilon(\vec{L}, h)|.$$

Now, from Equation (2.10) in Proposition 2.6 we get that for all $\delta > 0$ there is $\ell_1 > \ell_0$ with

$$(1-\delta) \cdot \mathrm{vol}\left(\bigcup_{\mathcal{G}^\phi \in G_\varepsilon(\vec{L},h)} \mathcal{G}^\phi_{\varepsilon-\mathrm{crit}}\right) < \left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^V \cdot |G_\varepsilon(\vec{L}, h)|$$

$$(1+\delta) \cdot \mathrm{vol}\left(\bigcup_{\mathcal{G}^\phi \in \hat{G}_\varepsilon(\vec{L},h)} \mathcal{G}^\phi_{\varepsilon-\mathrm{crit}}\right) > \left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^V \cdot |\hat{G}_\varepsilon(\vec{L}, h)|$$

whenever ε is small enough and $\min L_e \geqslant \ell_1$. Altogether, we get that for all ε positive and small we have

$$(1-\delta) \cdot \left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^{-V} \cdot \mathrm{vol}\left(\bigcup_{\mathcal{G}^\phi \in G_\varepsilon(\vec{L},h)} \mathcal{G}^\phi_{\varepsilon-\mathrm{crit}}\right) < |\mathbf{G}(\vec{L}, h)|$$

$$(1+\delta) \cdot \left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^{-V} \cdot \mathrm{vol}\left(\bigcup_{\mathcal{G}^\phi \in \hat{G}_\varepsilon(\vec{L},h)} \mathcal{G}^\phi_{\varepsilon-\mathrm{crit}}\right) > |\mathbf{G}(\vec{L}, h)|$$

for all $\vec{L}$ with $\min L_e \geqslant \ell_1$.

We get now from Equation (2.11) in Proposition 2.6 that there is $\ell_2 > \ell_1$ such that, as long as ε is under some threshold, we have that whenever $\phi, \psi \in \mathcal{G}$ are homotopic ε-critical realizations with $\ell(\phi(e)), \ell(\psi(e)) \geqslant \ell_2$ for all $e \in \mathbf{edge}$, then we have that the lengths $\ell(\phi(e))$ and $\ell(\psi(e))$ differ by at most 2ε for each edge $e \in \mathbf{edge}$. This implies that for any such $\vec{L}$ and ε we have

$$\mathcal{G}_{\varepsilon-\mathrm{crit}}(\vec{L}+[2\varepsilon],h-4\varepsilon) \subset \bigcup_{\mathcal{G}^{\phi}\in\mathbf{G}_{\varepsilon}(\vec{L},h)} \mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}}$$

$$\mathcal{G}_{\varepsilon-\mathrm{crit}}(\vec{L}-[2\varepsilon],h+4\varepsilon) \supset \bigcup_{\mathcal{G}^{\phi}\in\hat{\mathbf{G}}_{\varepsilon}(\vec{L},h)} \mathcal{G}^{\phi}_{\varepsilon-\mathrm{crit}},$$

where $\vec{L}+[t]\in\mathbb{R}^{\mathbf{edge}}$ is the vector with entries $(\vec{L}+[t])_e=\vec{L}_e+t$.

Let us summarize what we have obtained so far:

$$(1-\delta)\cdot\left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^{-V}\cdot\operatorname{vol}\Big(\mathcal{G}_{\varepsilon-\mathrm{crit}}(\vec{L}+[2\varepsilon],h-4\varepsilon)\Big) < |\mathbf{G}(\vec{L},h)|$$

$$(1+\delta)\cdot\left(\frac{2}{\sqrt{3}}\varepsilon^2\right)^{-V}\cdot\operatorname{vol}\Big(\mathcal{G}_{\varepsilon-\mathrm{crit}}(\vec{L}-[2\varepsilon],h+4\varepsilon)\Big) > |\mathbf{G}(\vec{L},h)|.$$

Our next goal is to compute the volumes on the left. Using the cover (2.2), we can compute volumes $\operatorname{vol}(\mathcal{G}_{\varepsilon-\mathrm{crit}}(\vec{L},h))$ by integrating over $\Sigma^{\mathbf{vert}}$ the cardinality of the intersection

$$\mathbf{G}^{X}_{\vec{x},\varepsilon-\mathrm{crit}}(\vec{L},h)=\Pi^{-1}(\vec{x})\cap\mathcal{G}_{\varepsilon-\mathrm{crit}}(\vec{L},h)$$

of the fiber $\Pi^{-1}(\vec{x})$ with the set we care about. In light of Corollary 3.3, we get in this way that

$$\begin{aligned}
\operatorname{vol}(\mathcal{G}_{\varepsilon-\mathrm{crit}}(\vec{L},h)) &= \int_{\Sigma^{\mathbf{vert}}}\left|\mathbf{G}^{X}_{\vec{x},\varepsilon-\mathrm{crit}}(\vec{L},h)\right| d\vec{x} \\
&\overset{\text{Cor. 3.3}}{\sim} \int_{\Sigma^{\mathbf{vert}}} \varepsilon^{4|\chi(X)|}\cdot\left(\frac{2}{3}\right)^{2\chi(X)}\cdot\pi^{\chi(X)}\cdot\frac{(e^h-1)^{-3\chi(X)}\cdot e^{\|\vec{L}\|}}{\operatorname{vol}(\Sigma)^{-3\chi(X)}}\,d\vec{x} \\
&= \varepsilon^{4|\chi(X)|}\cdot\left(\frac{2}{3}\right)^{2\chi(X)}\cdot\pi^{\chi(X)}\cdot\frac{(e^h-1)^{-3\chi(X)}\cdot e^{\|\vec{L}\|}}{\operatorname{vol}(\Sigma)^{-3\chi(X)}}\operatorname{vol}(\Sigma)^{V} \\
&= \varepsilon^{4|\chi(X)|}\cdot\left(\frac{2}{3}\right)^{2\chi(X)}\cdot\pi^{\chi(X)}\cdot\frac{(e^h-1)^{-3\chi(X)}\cdot e^{\|\vec{L}\|}}{\operatorname{vol}(\Sigma)^{-\chi(X)}},
\end{aligned}$$

where we have used that $V=-2\chi(X)$ and where the asymptotics hold true when $\min_e L_e\to\infty$. This means that whenever $\min_e L_e$ is large enough we have

$$(1-\delta)\cdot\frac{2^{4\chi(X)}}{3^{3\chi(X)}}\cdot\pi^{\chi(X)}\cdot\frac{(e^{h-4\varepsilon}-1)^{-3\chi(X)}\cdot e^{\sum_{e\in\mathbf{edge}}(L_e+2\varepsilon)}}{\operatorname{vol}(\Sigma)^{-\chi(X)}} < |\mathbf{G}(\vec{L},h)|,$$

$$(1+\delta)\cdot\frac{2^{4\chi(X)}}{3^{3\chi(X)}}\cdot\pi^{\chi(X)}\cdot\frac{(e^{h+4\varepsilon}-1)^{-3\chi(X)}\cdot e^{\sum_{e\in\mathbf{edge}}(L_e-2\varepsilon)}}{\operatorname{vol}(\Sigma)^{-\chi(X)}} > |\mathbf{G}(\vec{L},h)|.$$

Since this is true for all $\varepsilon>0$, we get that

$$(1-\delta)\cdot\frac{2^{4\chi(X)}}{3^{3\chi(X)}}\cdot\pi^{\chi(X)}\cdot\frac{(e^{h}-1)^{-3\chi(X)}\cdot e^{\|\vec{L}\|}}{\operatorname{vol}(\Sigma)^{-\chi(X)}} \leqslant |\mathbf{G}(\vec{L},h)|$$

$$(1+\delta)\cdot\frac{2^{4\chi(X)}}{3^{3\chi(X)}}\cdot\pi^{\chi(X)}\cdot\frac{(e^{h}-1)^{-3\chi(X)}\cdot e^{\|\vec{L}\|}}{\operatorname{vol}(\Sigma)^{-\chi(X)}} \geqslant |\mathbf{G}(\vec{L},h)|$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

and hence, since for all $\delta > 0$ we can choose L large enough so that the above bounds hold, we have

$$\frac{2^{4\chi(X)}}{3^{3\chi(X)}} \cdot \pi^{\chi(X)} \cdot \frac{(e^h - 1)^{-3\chi(X)} \cdot e^{\|\vec{L}\|}}{\mathrm{vol}(\Sigma)^{-\chi(X)}} \sim |\mathbf{G}(\vec{L}, h)|$$

as we wanted to prove. □

Armed with Lemma 4.1 and Proposition 4.2, we can now prove the theorem:

Proof of Theorem 1.3. Let $h > 0$ be small, and for $\vec{n} \in \mathbb{N}^{\mathbf{edge}}$ consider, with the same notation as in Equation (4.2), the set $\mathbf{G}(h \cdot \vec{n}, h)$. Setting

$$\Delta(N) = \{\vec{n} \in \mathbb{N}^{\mathbf{edge}} \text{ with } \|\vec{n}\| \leqslant N\},$$

where $\|\vec{n}\| = \sum_e n_e$, note that

$$\sum_{\vec{n} \in \Delta(N-E)} |\mathbf{G}(h \cdot \vec{n}, h)| \leqslant |\mathbf{G}(h \cdot N)| \leqslant \sum_{\vec{n} \in \Delta(N)} |\mathbf{G}(h \cdot \vec{n}, h)|, \tag{4.4}$$

where $\mathbf{G}(h \cdot N) = \mathbf{G}^X(h \cdot N)$ is as in Equation (4.1) and where, once again, $E = |\,\mathbf{edge}\,|$ is the number of edges of the graph X. Finally, write

$$\kappa = \frac{2^{4\chi(X)}}{3^{3\chi(X)}} \cdot (\pi \cdot \mathrm{vol}(\Sigma))^{\chi(X)}$$

Proposition 4.2 now reads as

$$|\mathbf{G}(h \cdot \vec{n}, h)| \sim \kappa \cdot (e^h - 1)^{-3\chi(X)} \cdot e^{h \cdot \|\vec{n}\|},$$

where the asymptotic holds for fixed h when $\min \vec{n}_e \to \infty$. This means that for all h and δ there is $n(h, \delta)$ with

$$|\mathbf{G}(h \cdot \vec{n}, h)| > (\kappa - \delta) \cdot (e^h - 1)^{-3\chi(X)} \cdot e^{h \cdot \|\vec{n}\|}$$

and

$$|\mathbf{G}(h \cdot \vec{n}, h)| < (\kappa + \delta) \cdot (e^h - 1)^{-3\chi(X)} \cdot e^{h \cdot \|\vec{n}\|}$$

for all $\vec{n}$ with $\min \vec{n}_e \geqslant n(h, \delta)$. It follows thus from the left side of Equation (4.4) that

$$\begin{aligned} |\mathbf{G}(h \cdot N)| &\geqslant \sum_{\substack{\vec{n} \,\in\, \Delta(N-E), \\ \min \vec{n}_e \,\geqslant\, n(h,\delta)}} |\mathbf{G}(h \cdot \vec{n}, h)| \\ &> (\kappa - \delta)(e^h - 1)^{-3\chi(X)} \sum_{\substack{\vec{n} \,\in\, \Delta(N-E), \\ \min \vec{n}_e \,\geqslant\, n(h,\delta)}} e^{h \cdot \|\vec{n}\|} \\ &= (\kappa - \delta)(e^h - 1)^{-3\chi(X)} \sum_{K=0}^{N-E} P(K) \cdot e^{h \cdot K}, \end{aligned}$$

where $P(K)$ is the number of those $\vec{n} \in \mathbb{N}^{\mathbf{edge}}$ with $\|n\| = K$ and $\min \vec{n}_e \geqslant n(h, \delta)$. As K tends to ∞, we have $P(K) \sim \frac{1}{(E-1)!} K^{E-1}$. Taking into account that $E = -3\chi(X)$, we get that for all N large enough we have

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

$$\begin{aligned}
|\mathbf{G}(h\cdot N)| &\gtrsim \frac{\kappa-\delta}{(-3\chi(X)-1)!}(e^h-1)^{-3\chi(X)}\sum_{K=0}^{N-E} K^{-3\chi(X)-1}\cdot e^{h\cdot K}\\
&= \frac{\kappa-\delta}{(-3\chi(X)-1)!}\left(\frac{e^h-1}{h}\right)^{-3\chi(X)}\sum_{K=0}^{N-E}(hK)^{-3\chi(X)-1}\cdot e^{h\cdot K}\cdot h\\
&\gtrsim \frac{\kappa-\delta}{(-3\chi(X)-1)!}\left(\frac{e^h-1}{h}\right)^{-3\chi(X)}\int_0^{(N-E)h} x^{-3\chi(X)-1}e^x\,dx,
\end{aligned}$$

where the symbol $\gtrsim$ means that asymptotically the ratio between the left side and the right side is at least 1. When $N\longrightarrow\infty$ then the value of the integral is asymptotic to $((N-E)\cdot h)^{-3\chi(X)-1}\cdot e^{(N-E)h}$, and this means that for all N large enough we have

$$|\mathbf{G}(h\cdot N)| \gtrsim \frac{\kappa-\delta}{(-3\chi(X)-1)!}\left(\frac{e^h-1}{h}\right)^{-3\chi(X)}(Nh-Eh)^{-3\chi(X)-1}\cdot e^{Nh-Eh}.$$

This being true for all δ and all h and replacing Nh by L, we have

$$|\mathbf{G}(L)| \gtrsim \frac{\kappa}{(-3\chi(X)-1)!}L^{-3\chi(X)-1}\cdot e^L$$

as $L\longrightarrow\infty$. In other words, we have established the desired asymptotic lower bound.

Starting with the upper bound we get, again for h positive and small, from the right side in Equation (4.4) that

$$|\mathbf{G}(h\cdot N)| \leqslant \sum_{\substack{\vec{n}\,\in\,\Delta(N),\\ \min\vec{n}_e\,\geqslant\, n(h,\delta)}}|\mathbf{G}(h\cdot\vec{n},h)| + \sum_{\substack{\vec{n}\,\in\,\Delta(N),\\ \min\vec{n}_e\,\leqslant\, n(h,\delta)}}|\mathbf{G}(h\cdot\vec{n},h)|.$$

The same calculation as above yields that

$$\begin{aligned}
&\sum_{\substack{\vec{n}\,\in\,\Delta(N),\\ \min\vec{n}_e\,\geqslant\, n(h,\delta)}}|\mathbf{G}(h\cdot\vec{n},h)| \lesssim\\
&\qquad\lesssim \frac{\kappa+\delta}{(-3\chi(X)-1)!}\left(\frac{e^h-1}{h}\right)^{-3\chi(X)}(h(N+E))^{-3\chi(X)-1}\cdot e^{h(N+E)}
\end{aligned}\tag{4.5}$$

as $N\longrightarrow\infty$. On the other hand, we get from Lemma 4.1 that there is $C>0$ with $|\mathbf{G}(h\cdot\vec{n},h)|\leqslant C\cdot e^{h\cdot|\vec{n}|}$ for all $\vec{n}$. This means thus that

$$\sum_{\substack{\vec{n}\,\in\,\Delta(N),\\ \min\vec{n}_e\,\leqslant\, n(h,\delta)}}|\mathbf{G}(h\cdot\vec{n},h)| \leqslant C\cdot\sum_{K=1}^{N}Q(K)\cdot e^{h\cdot K},$$

where $Q(K)$ is the number of those $\vec{n}\in\mathbb{N}^{\mathbf{edge}}$ with $\min\vec{n}_e<n(h,\delta)$ and $|\vec{n}|=K$. When $K\longrightarrow\infty$ the function $Q(K)$ is asymptotic to $C'\cdot K^{E-2}$ for some positive constant C', meaning that we have

$$\sum_{\substack{\vec{n}\,\in\,\Delta(N),\\ \min\vec{n}_e\,\leqslant\, n(h,\delta)}}|\mathbf{G}(h\cdot\vec{n},h)| \leqslant \frac{(1+\delta)\cdot C\cdot C'}{h^{E-1}}\cdot\sum_{K=1}^{N}h\cdot(hK)^{E-2}\cdot e^{h\cdot K}$$

for all L large enough. A similar estimation as the one above yields thus that there is another positive constant C'' with

$$\sum_{\substack{\vec{n}\in\Delta(N),\\ \min\vec{n}_e\leqslant n(h,\delta)}} |\mathbf{G}(h\cdot\vec{n},h)| \leqslant C''\cdot(N\cdot h)^{-3\chi(X)-2}\cdot e^{N\cdot H}. \tag{4.6}$$

The quantity in the right-hand side of Equation (4.6) is negligible when compared to the right-hand side of Equation (4.5), and this means that we have

$$|\mathbf{G}(h\cdot N)| \preceq \frac{\kappa+2\delta}{(-3\chi(X)-1)!}\left(\frac{e^h-1}{h}\right)^{-3\chi(X)} (h(N+E))^{-3\chi(X)-1}\cdot e^{h(N+E)}$$

for all large N. Since this holds true for all δ and all h, replacing hN by L we deduce that

$$|\mathbf{G}(L)| \preceq \frac{\kappa}{(-3\chi(X)-1)!}L^{-3\chi(X)-1}\cdot e^L.$$

Having now also established the upper asymptotic bound, we are done with the proof of the theorem. □

Before moving on to other matters, we include an observation that we will use later on. The basic strategy of the proof of Theorem 1.3 was to decompose the problem of counting all critical realizations of at most some given length into the problem of counting those whose edge lengths are in a given box and then adding over all boxes. For most boxes, Proposition 4.2 gives a pretty precise estimation for the number of critical realizations in the box, and from Lemma 4.1 we get an upper bound for all boxes. We used these two to get the desired upper bound in the theorem, deducing that we could ignore the boxes where Proposition 4.2 does not apply. A very similar argument implies that the set of critical realizations where some edge is shorter than ℓ_0 is also negligible. We state this observation as a lemma:

Lemma 4.3. *For all ℓ, all but a negligible set of critical realizations are ℓ-long.*

It is probably clear from the context what negligible means here, but to be precise, we mean that the set is negligible inside the set of all critical realizations in the sense of Equation (7.2) below.

5. Fillings

Let S_g be a compact, connected and oriented surface of genus g and with one boundary component. Below, we will be interested in continuous maps

$$\beta : S_g \to \Sigma \tag{5.1}$$

which send ∂S_g to a closed geodesic $\gamma = \beta(S_g)$. We will refer to such a map as a *filling of genus g of γ*, a *genus g filling of γ*, or just simply as a *filling of γ* when the genus g is either undetermined or understood from the context. The genus of a curve $\gamma \subset \Sigma$ is the infimum of all g's for which there is a genus g filling Equation (5.1) with $\gamma = \beta(\partial S_g)$. Note that the genus of a curve is infinite unless γ is homologically trivial, that is, unless γ is represented by elements in the commutator subgroup of $\pi_1(\Sigma)$. Indeed, as an element of $\pi_1(S_g)$ the boundary ∂S_g is a product of g commutators. It follows that if a curve γ in Σ has genus g, then it is, when considered as an element in $\pi_1(\Sigma)$, a product of g commutators. Conversely, if γ is a product of g commutators, then there is a map as in Equation (5.1) with S_g of genus g. In a nutshell, what we have is that the genus and the commutator length of γ agree. We record this fact for ease of reference:

Lemma 5.1. *The genus of a curve agrees with its commutator length.*

Continuing with the same notation and terminology, suppose that a curve γ in Σ has genus g. We then refer to any $\beta : S \to \Sigma$ as in Equation (5.1) with S of genus g as a *minimal genus filling*. Minimal genus fillings have very nice topological properties. Indeed, suppose that $\beta : S_g \to \Sigma$ is a filling as in Equation (5.1) and suppose that there is an essential simple curve $m \subset S_g$ with $\beta(m)$ homotopically

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

trivial. Then, performing surgery on the surface S_g and the map β we get a smaller genus filling $\beta' : S_{g'} \to \Sigma$ with $\beta'(\partial S_{g'}) = \beta(\partial S_g)$ and $g' < g$. It follows that if $\beta : S_g \to \Sigma$ is a minimal genus filling for $\gamma = \beta(\partial S_g)$, then β is *geometrically incompressible* in the sense that there are no elements in $\ker(\beta_* : \pi_1(S_g) \to \pi_1(\Sigma))$ which are represented by simple curves. We record this fact:

Lemma 5.2. *Minimal genus fillings are geometrically incompressible.*

From now on, we will be working exclusively with minimal genus fillings and the reader can safely add the words ‘minimal genus’ every time they see the word ‘filling’.

We will not be that much interested in individual fillings, but rather in homotopy classes of fillings. In particular, we will allow ourselves to select particularly nice fillings. More precisely, we will be working with *hyperbolic fillings*, by which we will understand a particular kind of pleated surface. We remind the reader that according to Thurston [18] (see also [5]) a pleated surface is a map from a hyperbolic surface to a hyperbolic manifold with the property that every point in the domain is contained in a geodesic segment which gets mapped isometrically.

Definition. A filling $\beta : S \to \Sigma$ is *hyperbolic* if S is endowed with a hyperbolic metric with geodesic boundary and if the map β is such that every $x \in S$ is contained in the interior of a geodesic arc I such that β maps I isometrically to a geodesic arc in Σ.

An important observation is that if $\beta : S \to \Sigma$ is a hyperbolic filling, if $x \in \partial S$ and if $I \subset S$ is a geodesic segment with x in its interior, then $I \subset \partial S$. It follows that hyperbolic fillings map the boundary isometrically.

Lemma 5.3. *The restriction of any hyperbolic filling $\beta : S \to \Sigma$ to ∂S is geodesic.*

We should not delay making sure that hyperbolic fillings exist:

Proposition 5.4. *Every minimal genus filling is homotopic to a hyperbolic filling.*

To prove this proposition, we will first show that if $\beta : S \to \Sigma$ is any filling and if the surface S admits a triangulation with certain properties, then β is homotopic to a hyperbolic filling. For lack of a better name, we will say that a triangulation $\mathcal{T}$ of S is *useful* if it satisfies the following three conditions:

(i) $\mathcal{T}$ has exactly $g + 1$ vertices $v_0, \ldots, v_g$, with $v_0 \in \partial S$ and the others in the interior of S.
(ii) There is a collection of edges $I_0, \ldots, I_g$ of $\mathcal{T}$ such that both endpoints of I_i are attached to v_i for all $i = 0, \ldots, g$. Moreover, $I_0 = \partial S$.

These two conditions are evidently pretty soft. It is the condition we will state next what makes useful triangulations actually useful. We first need a bit of notation: If I is any edge of $\mathcal{T}$ other than $I_0, \ldots, I_g$, then let G_I be the connected component of $I \cup I_0 \cup \cdots \cup I_g$ containing I.

(iii) For any edge I other than $I_0, \ldots, I_g$, we have that the image under $\beta_* : \pi_1(G_I) \to \pi_1(\Sigma)$ is not abelian.

Note that $\pi_1(G_I)$ is a a free group of rank 2. This means that its image $\beta_*(\pi_1(G_I))$ has at most rank 2. Since we are assuming that it is not abelian, we actually get that it is a rank 2 free group. Free groups being Hopfian we deduce that $\beta_* : \pi_1(G_I) \to \pi_1(\Sigma)$ is an isomorphism onto its image. In other words, β_* is injective on $\pi_1(G_I)$.

The following result makes clear why we care about such triangulations:

Lemma 5.5. *If the domain S of a filling $\beta : S \to \Sigma$ admits a useful triangulation $\mathcal{T}$, then β is homotopic to a hyperbolic filling.*

Proof. As we just discussed, our conditions imply that β_* is injective on $\pi_1(G_I)$ for all I. This implies in particular that each one of the exceptional edges $I_0, \ldots, I_g$ of $\mathcal{T}$ closes up to a simple closed curve $\gamma_0, \ldots, \gamma_g$ in S which is mapped to a homotopically essential curve. This in turn means that the images

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

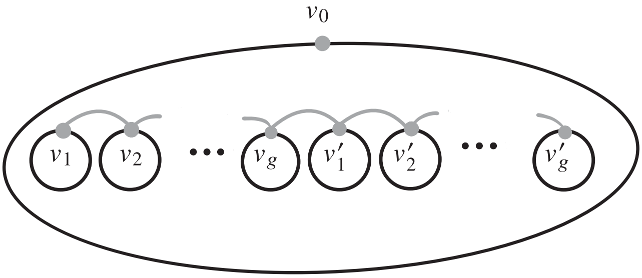

Figure 2. *Construction of the useful triangulation in Proposition 5.4: The $2g+1$ holed sphere obtained from S by cutting along $\gamma_1, \dots, \gamma_g$. The vertices $v_1, \dots, v_g$ and their copies $v'_1, \dots, v'_g$ are joined by the sequence of edges $[v_1, v_2], [v_2, v_3], \dots, [v_g, v'_1], [v'_1, v'_2], \dots, [v'_{g-1}, v'_g]$. To complete the triangulation, add as many edges as needed joining v_0 to the other vertices.*

of $\gamma_0, \dots, \gamma_g$ are homotopic to nontrivial closed geodesics. Since all these $g+1$ curves are mutually disjoint, we can then homotope β so that $\beta(\gamma_i)$ is a closed geodesic $\hat{\gamma}_i$ for all i.

Let now I be one of the remaining edges of $\mathcal{T}$, let v_i and v_j be its (possibly equal) endpoints and note that $G_I = \gamma_i \cup I \cup \gamma_j$. Since β_* is injective on $\pi_1(G_I, v_i)$, we know that the elements $\beta_*(\gamma_i)$ and $\beta_*(I * \gamma_j * I^{-1})$ do not commute and hence have distinct fixed points in $\partial_\infty \mathbb{H}^2$. This seemingly weak property is all we need to run the standard construction of pleated surfaces by spinning the edges of the triangulation over the geodesics $\hat{\gamma}_i$ – compare with the proof of Theorem I.5.3.6 in [5]. □

We can now prove Proposition 5.4.

Proof. Let $\beta : S \to \Sigma$ be a minimal genus filling of a geodesic $\gamma = \beta(\partial S)$, say of genus g. In light of Lemma 5.5, to prove that β is homotopic to a hyperbolic filling it suffices to show that S admits a useful triangulation. Let us start by taking g disjoint compact one-holed tori $T_1, \dots, T_g \subset S$. We claim that the restriction of β to each T_i is π_1-injective. Indeed, since we know that β is geometrically incompressible we deduce that ∂T_i is not in the kernel of the induced homomorphism at the fundamental group level. It follows that the image of $\beta_*(\pi_1(T_i))$ cannot be abelian. Now, this implies that $\beta_*(\pi_1(T_i))$ is free and evidently of rank 2. Hence, again since free groups are Hopfian, we get that the restriction of β_* to $\pi_1(T_i)$ is injective for all i.

Now, why do we care about that? Knowing that the restriction of β_* to $\pi_1(T_i)$ is injective for any i implies that when the images under β of the nonboundary parallel simple closed curves in T_i determine infinitely many conjugacy classes of maximal abelian subgroups of $\pi_1(\Sigma)$. We can thus choose for each $i = 1, \dots, g$ a nonboundary parallel simple closed curve $\gamma_i \subset T_i$ such that if we also set $\gamma_0 = \partial S$, then we have that

(*) no two of the the maximal abelian subgroups of $\pi_1(\Sigma)$ containing $\beta_*(\gamma_0), \dots, \beta_*(\gamma_g)$ are conjugate to each other.

Choosing now a vertex v_i in each one of the curves γ_i, we get from $(*)$ that if I is any simple path joining v_i to v_j for $i \neq j$, then the image of $\beta_*(\pi_1(\gamma_i \cup I \cup \gamma_j))$ is not abelian.

The upshot of all of this is that any triangulation $\mathcal{T}$ with

1. vertex set $v_0, \dots, v_g$,
2. such that there are edges $I_0, \dots, I_g$ incident on both ends to v_i and with image γ_i, and
3. such that all other edges connect distinct endpoints,

is useful. To see that such a triangulation exists, cut S along $\gamma_1, \dots, \gamma_g$. When doing this, we get a $2g+1$ holed sphere Δ and each vertex $v_1, \dots, v_g$ arises twice – denote the two copies of v_i by v_i and v'_i as in Figure 2. Now, any triangulation of Δ with vertex set $v_0, v_1, v'_1, v_2, v'_2, \dots, v_g, v'_g$ and which

- contains a sequence of $2g-1$ edges yielding a path

$$[v_1,v_2],[v_2,v_3],\dots,[v_g,v_1'],[v_1',v_2'],\dots,[v_{g-1}',v_g']$$

and
- such that all other edges are incident to v_0

yields a triangulation of S satisfying (1)–(3) above. Having proved that there is a useful triangulation, we get from Lemma 5.5 that β is homotopic to a hyperbolic filling, as we needed to prove. □

Being able to work with hyperbolic fillings is going to be key in the next section, where we bound the number of closed geodesics γ in Σ with length $\ell_\Sigma(\gamma)\leqslant L$ and which admit at least two homotopically distinct fillings.

Definition. Two fillings $\beta_1 : S_1 \to \Sigma$ and $\beta_2 : S_2 \to \Sigma$ are *homotopic* if there is a homeomorphism $F : S_1 \to S_2$ such that β_1 is homotopic to $\beta_2 \circ F$.

Evidently, to bound the number of pairs of nonhomotopic fillings with the same boundary we will need some criterion to decide when two fillings are homotopic. To be able to state it note that if $\beta_1 : S_1 \to \Sigma$ and $\beta_2 : S_2 \to \Sigma$ are homotopic hyperbolic fillings, then there is

> a closed hyperbolic surface $S = S_1 \cup_{\partial S_1=\partial S_2} S_2$ obtained by isometrically gluing S_1 and S_2 along the boundary, in such a way that there is a pleated surface $\Theta : S \to \Sigma$ with $\Theta|_{S_i} = \beta_i$.

For lack of a better name, we refer to $\Theta : S \to \Sigma$ as the *pseudo-double* associated to β_1 and β_2, and to the curve $\partial S_1 = \partial S_2 \subset S$ as the *crease* of the pseudo-double.

Our criterion to decide if the two hyperbolic fillings β_1 and β_2 of a geodesic γ are homotopic will be in terms of the structure of the ε_0 thin part of the domain of the associated pseudo-double, where we choose once and forever the constant

$$\varepsilon_0 = \frac{1}{10}\cdot\min\left\{\text{Margulis constant of } \mathbb{H}^2,\ \text{systole of } \Sigma\right\}. \tag{5.2}$$

The following is our criterion.

Lemma 5.6. *Suppose that $\beta_1 : S_1 \to \Sigma$ and $\beta_2 : S_2 \to \Sigma$ are hyperbolic minimal genus fillings of genus g of a geodesic γ, let $\Theta : S \to \Sigma$ be the associated pseudo-double and denote its crease $\partial S_1 = \partial S_2 \subset S$ also by γ. If*

1. *the ε_0-thin part of S has $6g-3$ connected components $U_1,\dots,U_{6g-3}$, and*
2. *γ traverses each U_i exactly twice,*

then β_1 and β_2 are homotopic.

Proof. Note that the surface S has genus $2g$. In particular, the assumption on the ε_0-thin part implies that there is a pants decomposition P in S consisting of closed geodesics of length at most $2\varepsilon_0$. Now, since Θ is evidently 1-Lipschitz and since $2\varepsilon_0$ is less than the systole of Σ we get that each one of the components of P is mapped to a homotopically trivial curve. The assumption that the crease γ, a simple curve, intersects each component of the pants decomposition exactly twice implies that γ cuts each pair of pants into two hexagons. Paint blue those in S_1 and yellow those in S_2.

We can now construct a homotopy relative to the crease as follows. Each component of P consists of a blue arc and a yellow arc whose juxtaposition is homotopically trivial. This implies that we can homotope all yellow arcs, relative to their endpoints to the corresponding blue arcs. Extend this homotopy to a homotopy fixing ∂S_2 and defined on the whole of S_2. Now, the boundary of each yellow hexagon is mapped to the boundary of each blue hexagon. Since Σ has trivial π_2, we deduce that those two hexagons are homotopic. Proceeding like this with each hexagon, we get a homotopy relative to the crease of the yellow parts of S to the blue part. This is what we wanted to get. □

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

6. Bounding the number of multifillings

The goal of this section is to prove Theorem 1.4:

Theorem 1.4. *For any g, there are at most* $\textbf{const} \cdot L^{6g-5} \cdot e^{\frac{L}{2}}$ *genus g closed geodesics γ in Σ with length $\ell(\gamma) \leqslant L$ and with two nonhomotopic fillings $\beta_1 : S_1 \to \Sigma$ and $\beta_2 : S_2 \to \Sigma$ of genus $\leqslant g$.*

The proof of Theorem 1.4 turns out to be kind of involved. We hope that the reader will not despair with all the weird objects and rather opaque statements they will find below.

Wired surfaces

Under a *wired surface*, we will understand a compact connected simplicial complex Δ obtained as follows: Start with a compact, possibly disconnected, triangulated surface $\mathrm{Surf}(\Delta)$ and an, evidently finite, subset P_Δ of the set of vertices of the triangulation of $\mathrm{Surf}(\Delta)$ such that every connected component of $\mathrm{Surf}(\Delta) \setminus P_\Delta$ has negative Euler characteristic. We think of the elements in P_Δ as *plugs*. Now, attach 1-simplices, to which we will refer as *wires*, by attaching both endpoints to plugs and do so in such a way that each plug arises exactly once as the end-point of a wire – we denote the set of wires by $\textbf{wire}(\Delta)$. A wired surface Δ without wires, that is, one with $\Delta = \mathrm{Surf}(\Delta)$, is said to be *degenerate*. Otherwise, it is *nondegenerate*.

How will wired surfaces arise? We will say that a pair $(\mathbb{F}, \mathbb{T})$ is a *decoration* of a surface S if

- ○ $\mathbb{F}$ is a partial foliation of S supported by the union of a finite collection of disjoint, essential and nonparallel, cylinders, and
- ○ $\mathbb{T}$ is a triangulation of the complement of the interior of those cylinders.

Now, if $(\mathbb{F}, \mathbb{T})$ is a decoration of S, then we get an associated wired surface $\Delta = S/\sim_{\mathbb{F}}$ by collapsing each leaf of $\mathbb{F}$ to a point and dividing each one of the arising bigons into two triangles by adding a vertex in the interior of the bigon. We will say that the quotient map $\pi : S \to \Delta$ is a *resolution* of Δ with *associated foliation* $\mathbb{F}$. Every wired surface admits an essentially unique resolution, unique in the sense that any two differ by a PL-homeomorphism mapping one of the foliations to the other one.

Suppose now that Δ is a wired surface. A *simple curve* on Δ is a map $\eta : \mathbb{S}^1 \to \Delta$ such that there are a resolution $\pi : S \to \Delta$ with associated foliation $\mathbb{F}$ and an essential simple curve $\eta' : \mathbb{S}^1 \to S$ which is transversal to $\mathbb{F}$ and with $\eta = \pi \circ \eta'$.

Note that transversality to the associated foliation implies that if η is a simple curve of a wired surface Δ then $\eta \cap \pi^{-1}(\Delta \setminus \mathrm{Surf}(\Delta))$ consists of a collection of segments, each one of them mapped homeomorphically to a wire. If I is such a wire, then we denote by $n_I(\eta)$ the *weight* of η in I, that is, the number of connected components of $\eta \cap \pi^{-1}(\Delta \setminus \mathrm{Surf}(\Delta))$ which are mapped homeomorphically to I, or in other words, the number of times that η crosses I. We refer to the vector

$$\vec{n}_\Delta(\eta) = (n_I(\eta))_{I \in \textbf{wire}(\Delta)} \tag{6.1}$$

as the *weight vector for η in Δ*. The intersection of the image of η with the surface part $\mathrm{Surf}(\Delta)$ is a simple arc system with endpoints in the set P_Δ. Note that up to homotopying η to another simple curve in Δ we might assume that all the components of the arc system $\eta \cap \mathrm{Surf}(\Delta)$ are essential. This is equivalent to asking that for each wire I we have $n_I(\eta) \leqslant n_I(\eta')$ for any other simple curve η' in Δ homotopic to η. We will suppose from now on, without further mention, that all simple curves in Δ satisfy these minimality requirements.

So far, wired surfaces are just topological objects. Let us change this. Under a *hyperbolic wired surface* we understand a wired surface Δ whose surface part $\mathrm{Surf}(\Delta)$ is endowed with a piece-wise hyperbolic metric, that is, one with respect to which the simplexes in the triangulation of $\mathrm{Surf}(\Delta)$ are isometric to hyperbolic triangles.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Let Δ be a hyperbolic wired surface, and as always let Σ be our fixed hyperbolic surface. We will say that a map $\Xi : \Delta \to \Sigma$ is *tight* if the following holds:

- ○ Ξ maps every wire to a geodesic segment, and
- ○ Ξ is an isometry when restricted to each one of the simplexes in the triangulation of $\mathrm{Surf}(\Delta)$.

We will be interested in counting pairs $(\Xi : \Delta \to \Sigma, \gamma)$ consisting of tight maps and simple curves. Evidently, without further restrictions, there could be infinitely many such pairs. What we are going to count is pairs were the curve has bounded length. We are, however, going to use a pretty strange notion of length. Consider namely for some given small but positive ε, the following quantity

$$\ell^{\varepsilon}_{\Xi}(\gamma) = \varepsilon \cdot \ell_{\mathrm{Surf}(\Delta)}(\gamma \cap \mathrm{Surf}(\Delta)) + \sum_{I \in \mathbf{wire}} n_I(\gamma) \cdot \max\left\{\ell_{\Sigma}(\Xi(I)) - \frac{1}{\varepsilon}, 0\right\}. \tag{6.2}$$

It is evident that this notion of length is exactly tailored to what we will need later on, but let us try to parse what Equation (6.2) actually means. What is the role of ε? If we think of the length as a measure of the cost of a journey, then the first ε just makes traveling along the surface part pretty cheap, meaning that for the same price we can cruise longer over there. Along the same lines, when traveling through the wires, we only pay when the wires are very long.

Lemma 6.1. *Let Δ be a nondegenerate hyperbolic wired surface with set of wires* $\mathbf{wire} = \mathbf{wire}(\Delta)$. *Fix a tight map $f : \Delta \to \Sigma$ and a positive integer vector $\vec{n} = (n_I)_{I \in \mathbf{wire}} \in \mathbb{N}_+^{\mathbf{wire}}$, and denote by*

$$\min = \min_{I \in \mathbf{wire}} n_I \geqslant 1 \text{ and } d = |\{I \in \mathbf{wire} \text{ with } n_I = \min\}|$$

the smallest entry of $\vec{n}$ and the number of times that this value is taken.

For any $\varepsilon > 0$, there are at most $\mathbf{const} \cdot L^{d-1} \cdot e^{\frac{L}{\min}}$ *homotopy classes of pairs $(\Xi : \Delta \to \Sigma, \gamma)$, where Ξ is a tight map with $\Xi|_{\mathrm{Surf}(\Delta)} = f|_{\mathrm{Surf}(\Delta)}$ and where γ is a simple multicurve in Δ with $n_I(\gamma) \geqslant n_I$ for every wire I and with $\ell^{\varepsilon}_{\Xi}(\gamma) \leqslant L$.*

Note that the fact that the obtained bound, or rather its rate of growth, does not depend on ε implies that actually the only way to get many homotopy classes is to play with the wires. In fact, since the given bound only depends on d and min, the only wires that matter are those which the curve crosses as little as possible.

Another comment before launching the proof. Namely, what happens if the wired surface Δ in Lemma 6.1 is degenerate? If there are no wires, then Δ is nothing other than a surface with a (piece-wise hyperbolic) metric. In such a surface, there are at most $\mathbf{const} \cdot L^{3 \cdot |\chi(\Delta)|}$ simple multicurves of length at most L – see, for example, [8]. This means that for a degenerate wired surface one actually gets a polynomial bound instead of an exponential one.

We are now ready to launch the proof of Lemma 6.1:

Proof. Note that, since the map Ξ is fixed on $\mathrm{Surf}(\Delta)$, we get that the homotopy type of the map, or even the map itself, is determined by what happens to the wires. In particular, as in the proof of Lemma 4.1 we get that if we give ourselves a positive vector $\vec{\lambda} = (\lambda_I)_{I \in \mathbf{wire}} \in \mathbb{R}_+^{\mathbf{wire}}$, then there are at most $\mathbf{const} \cdot e^{\|\vec{\lambda}\|}$ homotopy classes of tight maps $\Xi : \Delta \to \Sigma$ with $\Xi|_{\mathrm{Surf}(\Delta)} = f$ and such that

$$\lambda_I \leqslant \ell_{\Sigma}(\Xi(I)) \leqslant \lambda_I + 1 \text{ for all } I \in \mathbf{wire}\,. \tag{6.3}$$

As always, we have set $\|\vec{\lambda}\| = \lambda_1 + \cdots + \lambda_r$.

Now, we are not counting homotopy classes of maps but, rather, of pairs $(\Xi : \Delta \to \Sigma, \gamma)$, where the multicurve γ satisfies $\ell^{\varepsilon}_{\Xi}(\gamma) \leqslant L$. Note that, if Ξ satisfies Equation (6.3), then our given length bound $\ell^{\varepsilon}_{\Xi}(\gamma) \leqslant L$ implies that

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

$$\begin{aligned}\ell_{\mathrm{Surf}(\Delta)}(\gamma\cap\mathrm{Surf}(\Delta)) &= \frac{1}{\varepsilon}\left(\ell^{\varepsilon}_{\Xi}(\gamma)-\sum_{I\in\mathbf{wire}} n_I(\gamma)\cdot\max\left\{\ell_{\Sigma}(\Xi(I))-\frac{1}{\varepsilon},0\right\}\right)\\ &\leqslant \frac{1}{\varepsilon}\left(L-\sum_{I\in\mathbf{wire}} n_I\cdot\max\left\{\lambda_I-\frac{1}{\varepsilon},0\right\}\right)\\ &\leqslant \frac{1}{\varepsilon}\left(L-\langle\vec{n},\vec{\lambda}\rangle+\frac{1}{\varepsilon}\cdot\|\vec{n}\|\right).\end{aligned}$$

Now, since for any given length there are only polynomially many simple arc systems of bounded length we deduce that for each Ξ satisfying Equation (6.3) there are at most $\mathbf{const}\cdot(L-\langle\vec{n},\vec{\lambda}\rangle)^{\mathbf{const}}$ homotopy classes of simple multicurves γ in Δ with $\ell^{\varepsilon}_{\Xi}(\gamma)\leqslant L$ and satisfying $n_I(\gamma)\geqslant n_I$ for each wire I. Putting all of this together, we get the following:

Fact. There are at most $\mathbf{const}\cdot(L-\langle\vec{n},\vec{\lambda}\rangle)^{\mathbf{const}}\cdot e^{\|\lambda\|}$ homotopy classes of pairs $(\Xi:\Delta\to\Sigma,\gamma)$, where Ξ is a tight map with $\Xi|_{\mathrm{Surf}(\Delta)}=f$, satisfying Equation (6.3) and where γ is a simple curve in Δ with $n_I(\gamma)\geqslant n_I$ for every wire I and with $\ell^{\varepsilon}_{\Xi}(\gamma)\leqslant L$.

Now, we get that the quantity we want to bound, that is, the number of homotopy classes of pairs $(\Xi:\Delta\to\Sigma,\gamma)$, where Ξ is a tight map with $\Xi|_{\mathrm{Surf}(\Delta)}=f$ and where γ is a simple curve in Δ with $n_I(\gamma)\geqslant n_I$ for every wire I and with $\ell^{\varepsilon}_{\Xi}(\gamma)\leqslant L$ is bounded from above by

$$\sum_{\lambda\in\mathbb{N}^{\mathbf{wire}},\ \|\lambda\|\leqslant L}\mathbf{const}\cdot(L-\langle\vec{n},\vec{\lambda}\rangle)^{\mathbf{const}}\cdot e^{\|\lambda\|}.$$

This quantity is then bounded from above by the value of the integral

$$\mathbf{const}\int_{\{\vec{x}\in\mathbb{R}_+^{\mathbf{wire}},\ \langle\vec{n},\vec{x}\rangle\}\leqslant L}(L-\langle\vec{n},\vec{x}\rangle)^{\mathbf{const}}\cdot e^{\|\vec{x}\|}dx,$$

and now it is a calculus problem that we leave to the reader to check that this integral is bounded by $\mathbf{const}\cdot L^{d-1}\cdot e^{\frac{L}{n_{\min}}}$, where $n_{\min}=\min_I n_I\geqslant 1$ and where $d=|\{I \text{ wire with } n_I=n_{\min}\}|$. We are done. □

At this point, we know how to bound the number of homotopy classes of tight maps of wired surfaces. It is time to explain why we care about being able to do so.

Pseudo-doubles

Earlier, just before the statement of Lemma 5.6, we introduced the pseudo-double associated to two fillings. Let us extend that terminology a bit: Under a *pseudo-double*, we understand a pair $(\Theta:S\to\Sigma,\gamma)$ where

- $\Theta:S\to\Sigma$ is a pleated surface with S closed,
- $\gamma\subset S$, the *crease*, is a simple curve cutting S into two connected components and
- Θ maps γ to a geodesic in Σ and its restriction $\Theta|_{S\setminus\gamma}$ to the complement of γ is geometrically incompressible.

Two pseudo-doubles $(\Theta:S\to\Sigma,\gamma)$ and $(\Theta':S'\to\Sigma,\gamma')$ are *homotopic* if there is a homeomorphism $f:S\to S'$ with γ' homotopic to $f(\gamma)$ and with Θ homotopic to $\Theta'\circ f$.

Note that this terminology is consistent with the use of the word pseudo-double in the previous section.

Recall now that we fixed earlier some ε_0 satisfying Equation (5.2) and note that if $(\Theta:S\to\Sigma,\gamma)$ is a pseudo-double then the crease γ, being separating, crosses every component U of the thin part $S^{\leqslant\varepsilon_0}$ an even number of times $\iota(\gamma,U)\in 2\mathbb{N}$. In fact, since by the choice of ε_0 we get that Θ maps the core of every component of the thin part to a homotopically trivial curve and since the restriction of Θ to $S \setminus \gamma$ is geometrically incompressible we get that actually

$$\iota(\gamma, U) \geqslant 2 \text{ for all components } U \text{ of the thin part of } S, \tag{6.4}$$

by which we mean that γ traverses each such U at least twice.

Our next goal is to bound, for growing L, the number of homotopy classes of pseudo-doubles $(\Theta : S \to \Sigma, \gamma)$ where S has given topological type, where there are precisely d components U of the thin part with $\iota(\gamma, U) = 2$ and where $\ell_S(\gamma) \leqslant L$:

Proposition 6.2. *Let ε_0 be as in Equation (5.2), and suppose that S_0 is a closed orientable surface.*

1. *For every $d \geqslant 1$, there are at most $\mathbf{const} \cdot L^{d-1} \cdot e^{\frac{L}{2}}$ homotopy classes of pseudo-doubles $(\Theta : S \to \Sigma, \gamma)$, where S is homeomorphic to S_0 where the thin part $S^{\leqslant \varepsilon_0}$ of S has exactly d components U with $\iota(\gamma, U) = 2$ and where γ has length $\ell_S(\gamma) \leqslant L$.*
2. *There are at most $\mathbf{const} \cdot e^{\frac{L}{3}}$ homotopy classes of pseudo-doubles $(\Theta : S \to \Sigma, \gamma)$ where S is homeomorphic to S_0, where there is no component U of the thin part $S^{\leqslant \varepsilon_0}$ with $\iota(\gamma, U) = 2$, and where γ has length $\ell_S(\gamma) \leqslant L$.*

Recall that we declared a *decoration* $(\mathbb{F}, \mathbb{T})$ of a surface S to be a pair consisting of partial foliation $\mathbb{F}$ supported by the union of disjoint essential and nonparallel cylinders and of a triangulation $\mathbb{T}$ of the complement of the interior of those cylinders. To see where these decorations come from, assume that S is a hyperbolic surface.

- The ε_0-thick part of S is metrically bounded in the sense that it has a triangulation $\mathbb{T}$ whose vertices are $\frac{1}{10}\varepsilon_0$-separated and whose edges have length at most $\frac{1}{3}\varepsilon_0$ and are geodesic unless contained in the boundary $\partial(S^{\geqslant \varepsilon_0})$ of the ε_0-thick part – in that case, they have just constant curvature.
- The components of the ε_0-thin part $S^{\leqslant \varepsilon_0}$ are not metrically bounded but still have a very simple structure: They are cylinders foliated by constant curvature circles, namely the curves at constant distance from the geodesic at the core of the cylinder.

Putting these things together, that is, the triangulation $\mathbb{T}$ of the thick part and the foliation $\mathbb{F}$ of the thin part, we get what we will refer as a *thin-thick decoration* $(\mathbb{F}, \mathbb{T})$ *of* S – note that the triangulation $\mathbb{T}$ is not unique, and this is why we use an undetermined article. More importantly, note also that the number of components of the thin part is bounded just in terms of the topology of S and that the number of vertices in the triangulation $\mathbb{T}$ is bounded by some number depending on the chosen ε_0 and again the topological type of S. Since we have fixed ε_0, it follows that every compact surface S_0 admits finitely many decorations $(\mathbb{F}_1, \mathbb{T}_1), \ldots, (\mathbb{F}_r, \mathbb{T}_r)$ such that if S is any hyperbolic surface homeomorphic to S_0, then there is a homeomorphism $\sigma : S_0 \to S$ and some i such that $(\sigma(\mathbb{F}_i), \sigma(\mathbb{T}_i))$ is a ε_0-thin-thick decoration of S. We state this fact for later reference:

Lemma 6.3. *For every closed surface S_0, there are finitely many decorations $(\mathbb{F}_1, \mathbb{T}_1), \ldots, (\mathbb{F}_r, \mathbb{T}_r)$ such that for any hyperbolic surface S homeomorphic to S_0 there are $i \in \{1, \ldots, r\}$ and a homeomorphism $\sigma : S_0 \to S$ such that $\sigma(\mathbb{F}_i, \mathbb{T}_i)$ is a ε_0-thin-thick decoration of S.*

After these comments, we can finally launch the proof of Proposition 6.2:

Proof of Proposition 6.2. Starting with the proof of (1), note that from Lemma 6.3 we get that it suffices to prove for each fixed decoration $(\mathbb{F}, \mathbb{T})$ of S_0 that

(*) there are at most $\mathbf{const} \cdot L^{d-1} \cdot e^{\frac{L}{2}}$ homotopy classes of pseudo-doubles $(\Theta : S \to \Sigma, \gamma)$ where there is a homeomorphism $\sigma : S_0 \to S$ such that $(\sigma(\mathbb{F}), \sigma(\mathbb{T}))$ is a decoration of the ε_0-thin-thick decomposition of S, where the thin part $S^{\leqslant \varepsilon_0}$ of S has exactly d components U with $\iota(\gamma, U) = 2$ and where γ has length $\ell_S(\gamma) \leqslant L$.

Assuming from now on that we have an ε_0-thin-thick decoration $(\mathbb{F}, \mathbb{T})$, let $\Delta = S_0/\sim_{\mathbb{F}}$ be the wired surface obtained from S by collapsing each leaf of $\mathbb{F}$ to a point and let $\pi : S_0 \to \Delta$ be the corresponding quotient map.

The reason we consider Δ is that, as we already mentioned earlier, we have that by the choice of ε_0 all the leaves of the foliation $\sigma(\mathbb{F})$ are mapped by Θ to homotopically trivial curves. This implies that there is a map

$$\Xi : \Delta \to \Sigma$$

mapping the wires of Δ to geodesic segments and with $\Theta \circ \sigma$ homotopic to $\Xi \circ \pi$ by a homotopy whose tracks are bounded by **const**. In particular, the edges in $\mathbb{T}$ are mapped to paths homotopic to geodesic paths of at most length **const**. Pulling tight relative to the vertices, we can assume that Ξ maps two-dimensional simplices in the triangulation of Δ to hyperbolic triangles. Note that the bound on the lengths of the images of the edges of $\mathbb{T}$ imply that the restriction of Ξ to $\mathrm{Surf}(\Delta)$ is **const**-Lipschitz.

This uniform Lipschitz bound implies that $\Xi|_{\mathrm{Surf}(\Delta_0)}$ belongs to finitely many homotopy classes and that the tracks of the homotopy to any chosen representative of the correct homotopy class are bounded by **const**. Choosing the representatives to be tight maps with respect to some hyperbolic structure on Δ_0, we get:

Fact. There are finitely many hyperbolic structures $\Delta_1, \ldots, \Delta_r$ on Δ and finitely many tight maps $f_1, \ldots, f_r : \Delta_i \to \Sigma$ such that for any $\Theta : S \to \Sigma$ and $\sigma : S_0 \to S$ as in $(*)$ there are $i \in \{1, \ldots, r\}$ and a tight map $\Xi : \Delta_i \to \Sigma$ with $\Xi|_{\mathrm{Surf}(\Delta_i)} = f_i|_{\mathrm{Surf}(\Delta_i)}$ and such that $\Theta \circ \sigma : S_0 \to \Sigma$ is homotopic to $\Xi \circ \pi : S_0 \to \Sigma$ by a homotopy whose tracks have at most length **const**.

Continuing with the same notation, note that the bound on the tracks of the homotopy between $\Theta \circ \sigma$ and $\Xi \circ \pi$ means that when we compare the length of the geodesic γ in S with that of the curve $\eta = (\pi \circ \sigma^{-1})(\gamma)$ in the hyperbolic wired surface Δ_i then there is at most an increase by an additive amount every time γ crosses a simplex of the wired surface. It means that lengths

- increase at most by an additive amount every time we cross a component of the thin part, and
- increase by a multiplicative amount while we are in the thick part.

Said in other words: There is some R with

$$\begin{aligned} \ell_\Sigma(\Xi(\eta \cap \mathrm{Surf}(\Delta))) &\leqslant R \cdot \ell_S(\gamma \cap S^{\geqslant \varepsilon_0}) \\ \ell_\Sigma(\Xi(\eta \setminus \mathrm{Surf}(\Delta))) &\leqslant \sum_{\kappa \in \pi_0(\gamma \cap S^{\leqslant \varepsilon_0})} (\ell_S(\kappa) + R). \end{aligned}$$

Recalling that the wires I of Δ_0 correspond to the thin parts of $S^{\leqslant \varepsilon_0}$ and that the weight $n_I(\eta)$ is nothing other than the number of times that η crosses the wire I, we get from the last two inequalities that

$$\frac{1}{R} \cdot \ell_\Xi(\eta \cap \mathrm{Surf}(\Delta)) + \sum_{I \in \mathbf{wire}(\Delta_0)} n_I(\eta) \cdot \max\{\ell_\Xi(I) - R, 0\} \leqslant \ell_S(\gamma), \tag{6.5}$$

where the 'max' arises because a length is always nonnegative. Anyways, note that with the notation introduced in Equation (6.2) we can rewrite Equation (6.5) as

$$\ell_\Xi^{\frac{1}{R}}(\gamma) \leqslant \ell_S(\gamma).$$

Note also that from Equation (6.4) we get that $n_I(\eta) \geqslant 2$ for all $I \in \mathbf{wire}(\Delta_0)$. This means that using the notation of Lemma 6.1 we can restate the assumptions in Proposition 6.2 (1) as follows: min = 2 and this value is achieved $d \geqslant 1$ times. Lemma 6.1 implies thus that there are at most $\mathbf{const} \cdot L^{d-1} \cdot e^{\frac{L}{2}}$ homotopy classes of pairs (Ξ, η) arising from pseudo-doubles $(\Theta : S \to \Sigma, \gamma)$, where Θ is as in the fact

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

and where $\ell_S(\gamma) \leqslant L$. This implies a fortiori that we have at most that many choices for the homotopy class of Ξ. Since the homotopy class of Ξ determines that of Θ, we are done with the proof of (1). The proof of (2) is pretty much identical; the only difference is that now min $\geqslant 4$. Plugging this in the argument above, we get the bound $\mathbf{const} \cdot L^{k-1} \cdot e^{\frac{L}{4}}$ for some k. This is evidently a stronger bound that $\mathbf{const} \cdot e^{\frac{L}{3}}$, and we are done. □

The fruits of our labor

After all this work, we are now ready to prove Theorem 1.4.

Proof of Theorem 1.4. Suppose that γ is a closed geodesic with length $\ell_\Sigma(\gamma) \leqslant L$ and such that there are two nonhomotopic minimal genus fillings $\beta_1 : S_1 \to \Sigma$ and $\beta_2 : S_2 \to \Sigma$ with $\beta_i(\partial S_i) = \gamma$. From Proposition 5.4, we get that, without loss of generality, we might assume that both $\beta_1 : S_1 \to \Sigma$ and $\beta_2 : S_2 \to \Sigma$ are hyperbolic fillings.

Let then $S = S_i \cup_{\partial S_1 = \partial S_2} S_2$ be the hyperbolic surface obtained by gluing both surfaces S_1 and S_2 along the boundary in such a way that there is a pleated surface $\Theta : S \to \Sigma$ with $\Theta|_{S_i} = \beta_i$. Note once again that the map Θ maps the crease $\hat{\gamma} = \partial S_1 = \partial S_2$ geodesically to γ. Moreover, Lemma 5.2 implies that the restriction of Θ to $S \setminus \hat{\gamma}$ is geometrically incompressible. Taken together, all of this means that the pair $(\Theta : S \to \Sigma, \hat{\gamma})$ is a pseudo-double.

Now, Lemma 5.6 implies, together with the assumption that β_1 and β_2 are not homotopic, that the ε_0-thin part of S has at most $6g - 4$ connected components which are traversed twice by the crease $\hat{\gamma}$. We thus get from Proposition 6.2 that there are at most $\mathbf{const} \cdot L^{6g-5} \cdot e^{\frac{L}{2}}$ choices for the homotopy class of $(\Theta : S \to \Sigma, \hat{\gamma})$. Since the geodesic γ is determined by the homotopy class of $\Theta(\hat{\gamma})$, we have proved that there are $\mathbf{const} \cdot L^{6g-5} \cdot e^{\frac{L}{2}}$ choices for γ, as we had claimed. □

7. Proof of the main theorem

In this section, we prove Theorem 1.1 from the introduction.

Theorem 1.1. *Let Σ be a closed, connected and oriented hyperbolic surface and for $g \geqslant 1$ and $L > 0$ let $\mathbf{B}_g(L)$ be as in Equation (1.3). We have*

$$|\mathbf{B}_g(L)| \sim \frac{2}{12^g \cdot g! \cdot (3g-2)! \cdot \mathrm{vol}(T^1\Sigma)^{2g-1}} \cdot L^{6g-4} \cdot e^{\frac{L}{2}}$$

as $L \to \infty$.

Before we can even explain the idea of the proof of Theorem 1.1, we need to recall what fat graphs are and a few of their properties:

Fat graphs

A *fat graph* X is a graph endowed with a cyclic ordering of the set $\mathbf{half}_v$ for each v – fat graphs are also sometimes called *ribbon graphs*. Every fat graph is endowed with a canonically built thickening $\mathbf{neigh}(X)$, the *thickening of* X. For the sake of concreteness, let us discuss this in the particular case that X is trivalent. We start by taking an oriented filled-in hexagon G_v for every vertex v; see Figure 3. If we label by a, b, c the three elements in $\mathbf{half}_v$, given in the correct cyclic order, then we label the boundary components of G_v by a, ab, b, bc, c, ca, also given in the correct cyclic order. Now, for every edge $e \in \mathbf{edge}(X)$ let $v_1, v_2 \in \mathbf{vert}(X)$ be the two (possibly identical) vertices at which the two half-edges $\vec{e}_1 \in \mathbf{half}_{v_1}$ and $\vec{e}_2 \in \mathbf{half}_{v_2}$ corresponding to e are based, and identify in an orientation reversing way the $\vec{e}_1$-edge of ∂G_{v_1} with the $\vec{e}_2$-edge of ∂G_{v_2}. Proceeding like this for all edges, we end up with the *thickening* $\mathbf{neigh}(X)$ of X.

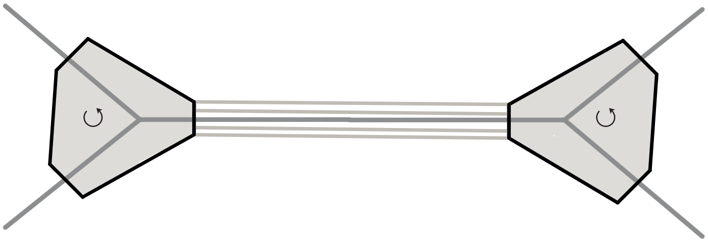

Figure 3. *Constructing the thickening of X. Pictured are two oriented hexagons corresponding to two vertices connected by an edge. The lighter lines indicate the gluing of those two hexagons.*

Definition. A trivalent fat graph X has *genus* g if $\mathbf{neigh}(X)$ is homeomorphic to a surface of genus g with one boundary component.

One of the quantities that appear in the right side of Theorem 1.1 is the number of genus g fat graphs or, rather, that number weighted by the number of automorphims of each such fat graph, where a map $F : X \to X'$ between two fat graphs is a *fat graph homeomorphism* if first it is a homeomorphism between two the underlying graphs and then if it sends one fat structure to the other one. In any case, what is really lucky for us is that Bacher and Vdovina [1] have computed that

$$\sum \frac{1}{|\operatorname{Aut}(X)|} = \frac{2}{12^g} \cdot \frac{(6g-5)!}{g! \cdot (3g-3)!},$$

where the sum takes place over all homeomorphism classes of genus g fat graphs.

Remark. Bacher and Vdovina's result is phrased in terms of one-vertex triangulations of the closed surface of genus g up to orientation preserving homeomorphism. Let us explain briefly how one goes from such triangulations to genus g fat graphs and back. The dual graph of a triangulation of a surface is a trivalent fat graph – its thickening is the surface minus the vertices of the triangulation. It follows that if the triangulation has a single vertex and the surface is closed of genus g, then the fat graph has genus g. Conversely, the thickening of a fat graph is equipped with a natural arc system (one arc dual to each edge). When collapsing each boundary component of the thickening to a point one gets a closed surface together with a triangulation with as many vertices as connected components of the boundary. It follows that if X is a genus g fat graph, then one gets a one-vertex triangulation of a genus g surface.

Let X now be a fat graph, and, for the sake of concreteness, assume that it has genus g. By construction, we have a canonical embedding of X into $\mathbf{neigh}(X)$ whose image is a spine. In particular, there is a retraction

$$\text{spine} : \mathbf{neigh}(X) \to X$$

such that the preimage of every vertex is a tripod and the preimage of every point in the interior of and edge in X is a segment. The image of $\partial\,\mathbf{neigh}(X)$ under spine runs twice over every edge of X. We will refer to this parametrized curve in X as ∂X.

Remark. Note that reversing the orientation of X, that is, reversing the cyclic order at each vertex, has the effect of reversing the orientation of ∂X.

The map

We will reduce the proof of Theorem 1.1 to the fact that we know how to count critical realizations of graphs, that is, to Theorem 1.3. Let us explain the basic idea. For given g, consider the set

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

$$\mathbf{X}_g = \left\{ (X,\phi) \left| \begin{array}{c} X \text{ is a fat graph of genus } g \text{ and} \\ \phi : X \to \Sigma \text{ is a critical realization} \\ \text{of the underlying graph} \end{array} \right. \right\} \Big/_{\text{equiv}}$$

of equivalence classes of realizations, where two realizations $\phi : X \to \Sigma$ and $\phi' : X' \to \Sigma$ are *equivalent* if there is a fat graph homeomorphism $\sigma : X \to X'$ such that $\phi = \phi' \circ \sigma$.

> **We stress something very important:** The critical realization ϕ for $(X,\phi) \in \mathbf{X}_g$ is explicitly not assumed to respect the fat structure. On the other hand, the homeomorphism σ in the definition of equivalence definitively has to preserve the fat structure.

Note that if (X,ϕ) is a equivalent to (X',ϕ'), then the curves $\phi(\partial X)$ and $\phi'(\partial X')$ are freely homotopic to each other. In particular, we have a well-defined map

$$\Lambda : \mathbf{X}_g \to \mathbf{C}, \quad (X,\phi) \mapsto \text{ geodesic homotopic to } \phi(\partial X), \tag{7.1}$$

where $\mathbf{C}$ is, as it was earlier, the collection of all oriented geodesics in Σ.

The basic idea of the proof of Theorem 1.1 is that the map (7.1) has the following informally stated properties:

1. Λ is basically injective with image basically contained in $\mathbf{B}_g$,
2. Λ is basically surjective onto $\mathbf{B}_g$, and
3. generically, the geodesic $\Lambda(X,\phi)$ has length almost exactly equal to $2 \cdot \ell(\phi) - C$ for some explicit constant C.

Let us start by clarifying the final point. Suppose that $(X,\phi) \in \mathbf{X}_g$ is such that ϕ is ℓ_0-long for some large ℓ_0. Since the curve ∂X runs exactly twice over each edge X, we get that it its image $\phi(\partial X)$ consists of $2\,\mathbf{edge}(X)$ geodesic segments of length at least ℓ_0 and with $3 \cdot \mathbf{vert}(X) = -6 \cdot \chi(X)$ corners where it makes an angle equal to $\frac{2\pi}{3}$. We get now from a standard hyperbolic geometry computation (or, if you so wish, from a limiting argument) that when ℓ_0 is large, then, up to an small error depending on ℓ_0, when we pull tight $\phi(\partial X)$ to get $\Lambda(X,\phi)$ we save $\log \frac{4}{3}$ at each one of those corners. Taking into account that $\chi(X) = 1 - 2g$ when X has genus g, this is what we have proved:

Lemma 7.1. *For every $\delta > 0$, there is ℓ_δ with*

$$\left| \ell_\Sigma(\Lambda(X,\phi)) - \left(2\ell_\Sigma(\phi) - 6 \cdot (2g-1) \cdot \log \frac{4}{3} \right) \right| \leqslant \delta$$

for every $(X,\phi) \in \mathbf{X}_g$ such that ϕ is ℓ_δ-long.

Remark. Where does $\log \frac{4}{3}$ come from? If $\Delta \subset \mathbb{H}^2$ is an ideal triangle with vertices θ_1, θ_2 and θ_3 and center p and if p' is the projection of p to the side say (θ_1,θ_2), then $\log \frac{2}{\sqrt{3}}$ is the difference between the values at p and p' of the Buseman function centered at θ_1, and every time we pass by a vertex we basically save twice that amount (see Figure 4).

Our next goal is to prove that the map Λ is basically bijective onto $\mathbf{B}_g$, but we should first formalize what we mean by ‘basically’. Suppose that we have a set Z consisting of either realizations, or curves, or of anything else consisting of elements $\alpha \in Z$ whose length $\ell_\Sigma(\alpha)$ can be measured, and set $Z(L) = \{\alpha \in Z \text{ with } \ell_\Sigma(\alpha) \leqslant L\}$. We will say that a subset

$$W \subset Z \text{ is } \textit{negligible} \text{ in } Z \text{ if } \limsup_{L\to\infty} \frac{|W(L)|}{|Z(L)|} = 0 \tag{7.2}$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

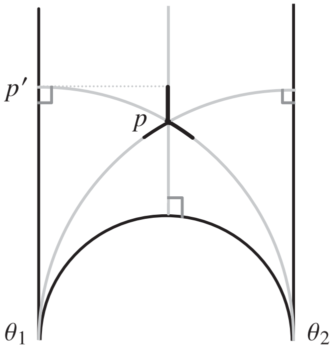

Figure 4. *The ideal triangle* Δ *with vertices* $\theta_1, \theta_2, \theta_3 = \infty$ *and center* p. *Each one of the legs of the bold printed tripod has length* $\log \frac{2}{\sqrt{3}}$.

If the ambient set Z is understood from the context, then we might just say that W is *negligible*. The complement of a negligible set is said to be *generic* and the elements of a generic set are themselves *generic*. For example, we get from Lemma 4.3 and Lemma 7.1 that for all δ we have that

$$\left|\ell_\Sigma(\Lambda(X,\phi)) - \left(2\ell_\Sigma(\phi) - 2\cdot(6g-3)\cdot\log\frac{4}{3}\right)\right| \leqslant \delta$$

for $(X,\phi) \in \mathbf{X}_g$ generic.

Basic bijectivity of Λ

Above we used the word ‘basically’ as meaning that something was true up to negligible sets. Let us start by proving that the map Λ is basically injective and that its image is contained in the set $\mathbf{B}_g$ of closed geodesics of genus g.

Lemma 7.2. *There is a generic subset* $W \subset \mathbf{X}_g$ *such that the restriction of* Λ *to* W *is injective and that its image is contained in* $\mathbf{B}_g$.

Proof. Let ℓ_0 and C be such that for every $(X,\phi) \in \mathbf{X}_g$ so that ϕ is ℓ_0-long we have

$$2\cdot\ell_\Sigma(\phi) - C \leqslant \ell_\Sigma(\Lambda(X,\phi)) \leqslant 2\ell_\Sigma(\phi),$$

and let $\mathbf{X}_{g,\ell_0}$ be the set of those pairs. We get from Lemma 4.3 that $\mathbf{X}_{g,\ell_0}$ is generic in $\mathbf{X}_g$. It follows hence from Theorem 1.3 that

$$|\{(X,\phi) \in \mathbf{X}_{g,\ell_0} \text{ with } \ell_\Sigma(\Lambda(X,\phi)) \leqslant L\}| \geqslant \mathbf{const}\cdot L^{6g-4}\cdot e^{\frac{L}{2}}. \tag{7.3}$$

Note now that each element (X,ϕ) of $\mathbf{X}_g$, and thus of $\mathbf{X}_{g,\ell_0}$, determines not only the curve $\Lambda(X,\phi)$ but also a homotopy class of fillings for this curve, namely $\phi\circ\text{spine} : \mathbf{neigh}(X) \to \Sigma$. Let now $Z \subset \mathbf{X}_{g,\ell_0}$ be the set of pairs (X,ϕ) so that the filling $\phi\circ\text{spine} : \mathbf{neigh}(X) \to \Sigma$ is not unique in the sense that the curve $\Lambda(X,\phi)$ admits another nonhomotopic genus g filling, and let

$$W = \mathbf{X}_{g,\ell_0} \setminus Z$$

be its complement. From Theorem 1.4, we get that Z consists of at most $\mathbf{const}\cdot L^{6g-5}\cdot e^{\frac{L}{2}}$ many elements and hence that W is generic.

Claim. If ℓ_0 is over some threshold, we have that $\Lambda(W) \subset \mathbf{B}_g$.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

Proof. First, we have by construction that the curve $\Lambda(X,\phi)$ admits the genus g filling $\phi \circ \mathrm{spine}_X : \mathbf{neigh}(X) \to \Sigma$. Suppose that it admits a smaller genus filling. Then, adding handles and mapping them to points we get that $\Lambda(X,\phi)$ admits a non-π_1-injective genus g filling. Since on the other hand the filling $\phi \circ \mathrm{spine}_X : \mathbf{neigh}(X) \to \Sigma$ is π_1-injective as long as ℓ_0 is over some threshold, we get that $\Lambda(X,\phi)$ admits two nonhomotopic genus g fillings, contradicting the assumption that $(X,\phi) \in W$. □

It remains to prove that the restriction of Λ to the generic set W is injective.

Suppose that we have $(X,\phi),(X',\phi') \in W$ with $\Lambda(X,\phi) = \Lambda(X',\phi')$. Since (X,ϕ) and (X',ϕ') belong to W, we know that the two fillings

$$\phi \circ \mathrm{spine}_X : \mathbf{neigh}(X) \to \Sigma \text{ and } \phi' \circ \mathrm{spine}_{X'} : \mathbf{neigh}(X') \to \Sigma$$

are homotopic. Recall that this means that there is a homeomorphism

$$\sigma : \mathbf{neigh}(X) \to \mathbf{neigh}(X')$$

with $\phi' \circ \mathrm{spine}_{X'} \circ \sigma$ homotopic to $\phi \circ \mathrm{spine}_X$. Since X and X' are spines of $\mathbf{neigh}(X)$ and $\mathbf{neigh}(X')$, we deduce that there is a homotopy equivalence $\bar\sigma : X \to X'$ such that $\phi' \circ \bar\sigma$ is homotopic to ϕ. Now, since the lengths of both $\phi(X)$ and $\phi'(X')$ are, up to a constant, basically half the length of $\Lambda(X,\phi) = \Lambda(X',\phi')$ we see that ϕ and ϕ' satisfy the conditions in Proposition 2.3. It thus follows that there is a homeomorphism $F : X' \to X$ mapping edges at constant velocity, with $F \circ \bar\sigma$ homotopic to the identity, and with $\phi \circ F$ homotopic to ϕ'. Now, both

$$\phi \circ F, \phi' : X \to \Sigma$$

are critical realizations and both are homotpic to each other. We get then from (1) in Lemma 2.2 that $\phi' = \phi \circ F$. To conclude, note that since F is a homotopy inverse of $\bar\sigma$ and since $\bar\sigma$ is induced by the homeomorphism $\sigma : \mathbf{neigh}(X') \to \mathbf{neigh}(X)$ we get that F is a fat graph homeomorphism. This proves that (X,ϕ) and (X',ϕ') are equivalent and hence that the restriction of Λ to W is injective. We are done with Lemma 7.2. □

Remark. Note that Equation (7.3) and Lemma 7.2 imply that

$$|\mathbf{B}_g(L)| > \mathbf{const} \cdot L^{6g-4} \cdot e^{\frac{L}{2}} \tag{7.4}$$

for all L large enough.

Our next goal is to show that the image of Λ contains a large subset of $\mathbf{B}_g$.

Lemma 7.3. *There is a generic subset of* $\mathbf{B}_g$ *which is contained in the image of* Λ.

Proof. Let $Z \subset \mathbf{B}_g$ be the set of those geodesics which admits a hyperbolic genus g filling $\beta : S \to \Sigma$ such that the ε_0-thin part of the double DS of S has at most $6g-4$ connected components U through with $\iota(U,\partial S) = 2$. It follows from Proposition 6.2 that Z has at most $\mathbf{const} \cdot L^{6g-5} \cdot e^{\frac{L}{2}}$ elements with length $\leqslant L$. It follows from Equation (7.4) that Z is negligible and hence that its complement $W = \mathbf{B}_g \setminus Z$ is generic. We claim that W is contained in the image of Λ, at least if ε_0 is chosen small enough.

Each $\gamma \in W$ admits a hyperbolic filling $\beta : S \to \Sigma$ such that the ε_0-thin part of the double DS has at least $6g-3$ connected components U with $\iota(\partial S, U) = 2$. Since DS has genus $2g$, we have that $6g-3$ is actually the maximal number of connected components that its thin part can have. It follows that the double DS of S admits a pants decomposition consisting of very short curves and that moreover ∂S cuts each one of them exactly twice. It follows that S has $6g-3$ orthogeodesics cutting S into a union of $4g-2$ right-angle hexagons. These hexagons have three alternating sides which are extremely short. It follows that each one of the hexagons contains a pretty large compact set which is almost isometric to a large neighborhood of the center of an ideal triangle in $\mathbb{H}^2$. Declare the center of the hexagon to be the image of the center of the ideal triangle by this almost isometric map. Now, we can represent the

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

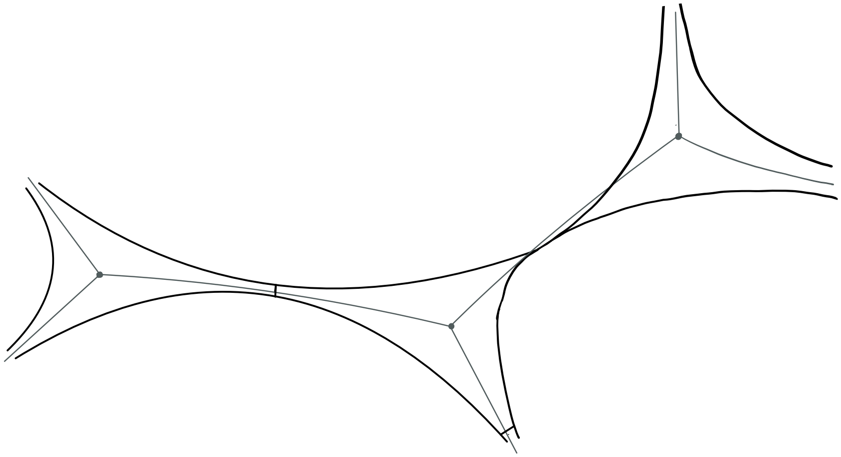

Figure 5. *Getting an almost critical realization out of a filling whose domain can be cut into hexagons by very short orthogeodesics.*

dual graph of the decomposition of S by our short orthogeodesic segments as a geodesic subgraph X of S with vertices in the centers of the hexagons; see Figure 5.

The graph X is a spine of S, and hence it inherits from S a fat graph structure with S as its neighborhood. Let now $\psi : X \to \Sigma$ be the realization obtained from the restriction of β to X by pulling the edges tight.

Claim. For all δ, there is ε such that if $\varepsilon_0 < \varepsilon$, then the realization $\psi : X \to \Sigma$ is δ-critical.

Proof. Suppose that the claim fails to be true. This means that for some δ there are a sequence of counter examples with $\varepsilon_0 \to 0$. Since $\varepsilon_0 \to 0$, we get that the images of the hexagons converge to a 1-Lipschitz map of a geodesic triangle into Σ which, however, maps the boundary geodesics to geodesics. Such a map is an isometric embedding of the triangle in question. Now, in a geodesic triangle, the geodesic rays starting in the center and pointing into the cusps make angle $\frac{2\pi}{3}$. This means that the angles in the approximates converge to $\frac{2\pi}{3}$ contradicting our assumption that some of them were at least δ off $\frac{2\pi}{3}$. We have proved the claim. □

It follows thus from the claim and Corollary 2.4 that, as long as ε_0 is under some threshold, the realization $\psi : X \to \Sigma$ is homotopic to a critical realization $\phi : X \to \Sigma$. This implies (X, ϕ) belongs to the domain of Λ and that $\gamma = \Lambda(X, \phi)$. We have proved that the generic set $W \subset \mathbf{B}_g$ is contained in the image of Λ. □

We are now ready to wrap all of this up.

Proof of Theorem 1.1

Combining Lemma 7.1, Lemma 7.2 and Lemma 7.3, we get that for all $\delta > 0$ there is a generic subset $W \subset \mathbf{X}_g$ which is mapped injectively under Λ to a generic subset of $\mathbf{B}_g$ in such a way that

$$|\ell_\Sigma(\Lambda(X, \phi)) - 2\ell_\Sigma(\phi) + 2\kappa| \leqslant \delta,$$

where

$$\kappa = -3\chi(X) \cdot \log \frac{4}{3} = \log\left(\left(\frac{4}{3}\right)^{6g-3}\right). \tag{7.5}$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

It follows that for all $\delta > 0$, our set $\mathbf{B}_g(L)$ has, for $L \longrightarrow \infty$, at least as many elements as $\mathbf{X}_g(\frac{L}{2} + \kappa - \delta)$ and at most as many as $\mathbf{X}_g(\frac{L}{2} + \kappa + \delta)$. In symbols, this just means that

$$\left|\mathbf{X}_g\left(\frac{L}{2} + \kappa - \delta\right)\right| \leq |\mathbf{B}_g(L)| \leq \left|\mathbf{X}_g\left(\frac{L}{2} + \kappa + \delta\right)\right| \tag{7.6}$$

for large L. It thus remains to estimate the cardinality of $\mathbf{X}_g(L)$.

Lemma 7.4. *We have*

$$|\mathbf{X}_g(L)| \sim \frac{1}{12^g} \cdot \left(\frac{3}{2}\right)^{6g-3} \cdot \frac{1}{g! \cdot (3g-2)! \cdot \mathrm{vol}(T^1\Sigma)^{2g-1}} \cdot L^{6g-4} \cdot e^L$$

as $L \longrightarrow \infty$.

Proof. From Theorem 1.3, we get that every trivalent graph X has

$$|\mathbf{G}^X(L)| \sim \left(\frac{2}{3}\right)^{3\chi(X)} \cdot \frac{\mathrm{vol}(T^1\Sigma)^{\chi(X)}}{(-3\chi(X)-1)!} \cdot L^{-3\chi(X)-1} \cdot e^L$$

critical realizations of length at most L in Σ. This implies that whenever X is fat graph of genus g then asymptotically there are

$$\frac{1}{|\operatorname{Aut}(X)|}\left(\frac{2}{3}\right)^{3\chi(X)} \cdot \frac{\mathrm{vol}(T^1\Sigma)^{\chi(X)}}{(-3\chi(X)-1)!} \cdot L^{-3\chi(X)-1} \cdot e^L$$

many elements in $\mathbf{X}_g(L)$ represented by (X, ϕ) for some critical $\phi : X \longrightarrow \Sigma$ of length $\ell(\phi) \leqslant L$. Adding over all possible types of genus g fat graphs, we get that

$$|\mathbf{X}_g(L)| \sim \sum_X \frac{1}{|\operatorname{Aut}(X)|}\left(\frac{2}{3}\right)^{3\chi(X)} \cdot \frac{\mathrm{vol}(T^1\Sigma)^{\chi(X)}}{(-3\chi(X)-1)!} \cdot L^{-3\chi(X)-1} \cdot e^L.$$

From the Bacher–Vdovina [1] result mentioned earlier and taking into consideration that $\chi(X) = 1-2g$, we get

$$|\mathbf{X}_g(L)| \sim \frac{2}{12^g} \cdot \frac{(6g-5)!}{g! \cdot (3g-3)!} \cdot \left(\frac{3}{2}\right)^{6g-3} \cdot \frac{\mathrm{vol}(T^1\Sigma)^{1-2g}}{(6g-4)!} \cdot L^{6g-4} \cdot e^L.$$

The claim follows now from elementary algebra. □

Now, from Lemma 7.4 we get that

$$\left|\mathbf{X}_g\left(\frac{L}{2}\right)\right| \sim \frac{1}{12^g} \cdot \left(\frac{3}{2}\right)^{6g-3} \cdot \frac{1}{g! \cdot (3g-2)! \cdot \mathrm{vol}(T^1\Sigma)^{2g-1}} \cdot \frac{L^{6g-4}}{2^{6g-4}} \cdot e^{\frac{L}{2}}.$$

Taking into account that

$$\left|\mathbf{X}_g\left(\frac{L}{2} + \kappa\right)\right| \sim \left|\mathbf{X}_g\left(\frac{L}{2}\right)\right| \cdot e^\kappa = \left|\mathbf{X}_g\left(\frac{L}{2}\right)\right| \cdot \left(\frac{4}{3}\right)^{6g-3},$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

we get that

$$\left|\mathbf{X}_g\left(\frac{L}{2}+\kappa\right)\right| \sim \frac{2}{12^g \cdot g! \cdot (3g-2)! \cdot \mathrm{vol}(T^1\Sigma)^{2g-1}} \cdot L^{6g-4} \cdot e^{\frac{L}{2}}.$$

It follows thus from Equation (7.6) that

$$|\mathbf{B}_g(L)| \sim \frac{2}{12^g \cdot g! \cdot (3g-2)! \cdot \mathrm{vol}(T^1\Sigma)^{2g-1}} \cdot L^{6g-4} \cdot e^{\frac{L}{2}}$$

as we wanted to show. This concludes the proof of Theorem 1.1.

8. Curves bounding immersed surfaces

We now turn our attention to Theorem 1.2:

Theorem 1.2. *Let Σ be a closed, connected and oriented hyperbolic surface and for $g \geqslant 1$ and $L > 0$ let $\mathbf{B}_g(L)$ be as in Equation (1.3). We have*

$$|\{\gamma \in \mathbf{B}_g(L) \textit{ bounds immersed surface of genus } g\}| \sim \frac{1}{2^{4g-2}}|\mathbf{B}_g(L)|$$

as $L \to \infty$.

Denote by

$$\mathbf{B}_g^{\mathrm{imm}}(L) = \{\gamma \in \mathbf{B}_g(L) \text{ bounds immersed surface of genus } g\}$$

the set we want to count. From the proof of Theorem 1.1, we get that

$$|\mathbf{B}_g^{\mathrm{imm}}(L)| \sim \left|\mathbf{X}_g^{\mathrm{imm}}\left(\frac{L}{2}+\kappa\right)\right|,$$

where κ is as in Equation (7.5) and where

$$\mathbf{X}_g^{\mathrm{imm}}(L) \overset{\mathrm{def}}{=} \left\{ (X,\phi) \in \mathbf{X}_g(L) \;\middle|\; \begin{array}{c} \text{the realization } \phi : X \to \Sigma \text{ extends} \\ \text{to an immersion of the thickening} \\ \mathbf{neigh}(X) \text{ of } X \end{array} \right\}.$$

Equivalently, $(X,\phi) \in \mathbf{X}_g$ belongs to $\mathbf{X}_g^{\mathrm{imm}}$ if the cyclic ordering at each vertex of X agrees with the one pulled back from Σ via ϕ, that is, the one coming from the orientation of Σ. We can refer to such realizations of a fat graph as *fat realizations*. With this language, we have that

$$\mathbf{X}_g^{\mathrm{imm}}(L) = \big\{(X,\phi) \in \mathbf{X}_g(L) \,\big|\, \phi \text{ is a fat realization of the fat graph } X\big\}.$$

For a given trivalent fat graph X, let

$$\mathbf{G}^X_{\mathrm{imm}}(L) = \{\phi \in \mathbf{G}^X(L) \mid \phi \text{ is a fat realization of } X\}$$

be the set of fat critical realizations of X of total length at most L. Note that

$$|\mathbf{X}_g^{\mathrm{imm}}(L)| = \sum_X \frac{1}{|\operatorname{Aut}(X)|}|\mathbf{G}^X_{\mathrm{imm}}(L)|.$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

What is still missing to be able to run the proof as in that of Theorem 1.1 is a version of Theorem 1.3 for $\mathbf{G}^X_{\text{imm}}$. Here it is:

Theorem 8.1. *Let Σ be a closed, connected and oriented hyperbolic surface. For every connected trivalent fat graph X, we have*

$$|\mathbf{G}^X_{\text{imm}}(L)| \sim 2^{2\chi(X)} \cdot \left(\frac{2}{3}\right)^{3\chi(X)} \cdot \frac{\text{vol}(T^1\Sigma)^{\chi(X)}}{(-3\chi(X)-1)!} \cdot L^{-3\chi(X)-1} \cdot e^L$$

as $L \to \infty$.

Assuming Theorem 8.1 for the moment, we get from $\chi(X) = 1-2g$ and from the Bacher–Vdovina theorem that

$$|\mathbf{X}^{\text{imm}}_g(L)| \sim \frac{1}{2^{4g-2}} \cdot \frac{2}{12^g} \cdot \frac{(6g-5)!}{g! \cdot (3g-3)!} \cdot \left(\frac{3}{2}\right)^{6g-3} \cdot \frac{\text{vol}(T^1\Sigma)^{1-2g}}{(6g-4)!} \cdot L^{6g-4} \cdot e^L$$

from where we get, as in the proof of Theorem 1.1, that

$$|\mathbf{B}^{\text{imm}}_g(L)| \sim \frac{1}{2^{4g-2}} \frac{2}{12^g \cdot g! \cdot (3g-2)! \cdot \text{vol}(T^1\Sigma)^{2g-1}} \cdot L^{6g-4} \cdot e^{\frac{L}{2}}.$$

The claim of Theorem 1.2 follows now from this statement combined with Theorem 1.1.

All that is left to do is to prove Theorem 8.1. Since it is basically identical to the proof of Theorem 1.3, we just point out the differences. The key is to obtain a fat graph version of Proposition 4.2. In a nutshell, the idea of the proof of this proposition was that

1. we could compute the volume $\text{vol}(\mathcal{G}^X_{\varepsilon\text{-crit}}(\vec{L}, h))$ of the set of ε-critical realizations of X whose edge lengths were in a box, and
2. we knew that every connected component contributes the same amount, and how much.

Let thus $\mathcal{G}^X_{\varepsilon\text{-crit}-\text{imm}}(\vec{L}, h) \subset \mathcal{G}^X_{\varepsilon\text{-crit}}(\vec{L}, h)$ be those ε-critical realizations in our box which preserve the fat structure. Recalling now that by Proposition 2.6 the connected components of $\mathcal{G}^X_{\varepsilon\text{-crit}}(\vec{L}, h)$ have small diameter, we deduce that the induced fat structure is constant over each such connected component. It follows that $\mathcal{G}^X_{\varepsilon\text{-crit}-\text{imm}}(\vec{L}, h)$ is a union of connected components of $\mathcal{G}^X_{\varepsilon\text{-crit}}(\vec{L}, h)$. In particular, to be able to obtain a fat graph version of Proposition 4.2 we just need to be able to compute $\text{vol}(\mathcal{G}^X_{\varepsilon\text{-crit}-\text{imm}}(\vec{L}, h))$.

Now, in the proof of Proposition 4.2, the key ingredient of the computation of the volume of $\mathcal{G}_{\varepsilon\text{-crit}}(\vec{L}, h)$ was Corollary 3.3 – we remind the reader that the statement of the said corollary was that for any $\vec{x} \in \Sigma^{\mathbf{vert}\,X}$ we have

$$|\mathbf{G}^X_{\vec{x},\varepsilon\text{-crit}}(\vec{L}, h)| \sim \varepsilon^{4|\chi(X)|} \cdot \left(\frac{2}{3}\right)^{2\chi(X)} \cdot \pi^{\chi(X)} \cdot \frac{(e^h-1)^{-3\chi(X)} \cdot e^{\|\vec{L}\|}}{\text{vol}(\Sigma)^{-3\chi(X)}}$$

as $\min_{e \in \mathbf{edge}(X)} L_e \to \infty$, where $\mathbf{G}^X_{\vec{x},\varepsilon\text{-crit}}(\vec{L}, h)$ is the set of ε-critical realizations $\phi : X \to \Sigma$ mapping the vertex v to the point $x_v = \phi(v)$. As we see, if we want to just copy line-by-line the computation of the volume of $\mathcal{G}_{\varepsilon\text{-crit}}(\vec{L}, h)$ to get the volume of $\mathcal{G}_{\varepsilon\text{-crit}-\text{imm}}(\vec{L}, h)$ what we need to know is the number of elements in the set $\mathbf{G}^X_{\vec{x},\varepsilon\text{-crit}-\text{imm}}(\vec{L}, h)$ of ε-critical realizations $\phi : X \to \Sigma$ mapping the vertex v to the point $x_v = \phi(v)$ and preserving the fat structure.

Now, to obtain the number of elements in $\mathbf{G}^X_{\vec{x},\varepsilon\text{-crit}}(\vec{L}, h)$ we invoked Theorem 3.2 and computed the volume of the set

$$U^X_{\vec{x},\varepsilon\text{-crit}} \subset \prod_{v \in \mathbf{vert}(X)} \left(\bigoplus_{\vec{e} \in \mathbf{half}_v(X)} T^1_{x_v}\Sigma \right)$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

of those tuples $(v_{\vec{e}})_{\vec{e}\in\mathbf{half}(X)}$ with $\angle(v_{\vec{e}_1}, v_{\vec{e}_2}) \in [\frac{2\pi}{3} - \varepsilon, \frac{2\pi}{3} + \varepsilon]$ for all distinct $\vec{e}_1, \vec{e}_2 \in \mathbf{half}(X)$ incident to the same vertex. Accordingly, to compute the number of elements in $\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit\,imm}}(\vec{L}, h)$ we need to compute the volume of the set

$$U^X_{\vec{x},\varepsilon-\mathrm{crit-imm}} \subset U^X_{\vec{x},\varepsilon-\mathrm{crit}}$$

consisting of tuples such that for any vertex $v \in \mathbf{vert}(X)$ the cyclic order of the half-edges incident to v agrees with the one of the corresponding unit tangent vectors. Following the computation of $\mathrm{vol}(U_{\vec{x},\varepsilon=\mathrm{crit}})$, we get that

$$\mathrm{vol}(U^X_{\vec{x},\varepsilon-\mathrm{crit-imm}}) = \frac{1}{2^{|\mathbf{vert}(X)|}} \mathrm{vol}(U^X_{\vec{x},\varepsilon-\mathrm{crit}}) = 2^{2\chi(X)} \cdot \mathrm{vol}(U^X_{\vec{x},\varepsilon-\mathrm{crit}}).$$

As we just discussed, this implies that

$$|\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit-imm}}(\vec{L}, h)| \sim 2^{2\chi(X)} \cdot |\mathbf{G}^X_{\vec{x},\varepsilon-\mathrm{crit}}(\vec{L}, h)|$$

and hence that

$$\mathrm{vol}(\mathcal{G}^X_{\varepsilon-\mathrm{crit-imm}}(\vec{L}, h)) \sim 2^{2\chi(X)} \cdot \mathrm{vol}(\mathcal{G}^X_{\varepsilon-\mathrm{crit}}(\vec{L}, h))$$

and thus that

$$|\mathbf{G}^X_{\mathrm{imm}}(L)| \sim 2^{2\chi(X)} \cdot |\mathbf{G}^X(L)|.$$

Theorem 8.1 follows now from Theorem 1.3.

9. Comments

In this section, we discuss where and how much we use the assumption that Σ is a closed orientable surface.

First, do the results here apply if we replace Σ by a compact two-dimensional orbifold $\mathcal{O} = \Gamma\backslash\mathbb{H}^2$?

The answer is yes for Theorem 1.3, with exactly the same proof. One should just do everything equivariantly. For example, a realization in $\mathcal{O}$ of a graph X should be a map $\tilde{\phi} : \tilde{X} \to \mathbb{H}^2$ from the universal cover of X to $\mathbb{H}^2$ which is equivariant under a homomorphism $\tilde{\phi}_* : \pi_1(X) \to \Gamma$ and where two such pairs $(\tilde{\phi}, \tilde{\phi}_*)$ and $(\tilde{\psi}, \tilde{\psi}_*)$ are identified if they differ by an element of Γ. Once we rephrase the situation in those terms, everything extends in the obvious way. For example, the space $\mathcal{G}^X$ of realizations of a graph in $\mathcal{O}$ is now an orbifold: Tthe map $\mathcal{G}^X \to \mathcal{O}^{\mathbf{vert}\,X}$ sending each realization to the images of the vertices is a covering in the category of orbifolds.

On the other hand, to prove Theorem 1.1 one needs to be a little bit careful because in Section 5 and Section 6 we used repeatedly that every sufficiently short curve in Σ is homotopically trivial, and this is no longer true if we are working in $\mathcal{O}$.

What about allowing Σ to have cusps?

Here, we again have problems with the discussion in Section 5 and Section 6, but this time it is much worse. In some sense, the results of Section 5 and Section 6 are just generalizations of Lemma 4.1, and this lemma fails in the presence of cusps: Indeed, suppose that Σ is a once punctured torus and X is a graph with two vertices x and x' (if you wish, you can make X trivalent while essentially keeping the same reasoning as we will present, but doing so might obscure things slightly) and with six edges f_1, f_2, e_1, e_2, e_3 and h such that

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

- ○ the edges f_1, f_2 are incident on both ends to x,
- ○ the edges e_1, e_2, e_3 are incidents on both ends to x' and
- ○ the edge h runs from x to x'.

Fix now a horospherical neighborhood of the cusps in Σ, fix a point $\mathbf{x}_0$ and let $\mathbf{x}_t$ be the point at distance t of $\mathbf{x}_0$ along the ray pointing directly into the cusp. Now, if we are given a vector $\vec{L} = (F_1, F_2, E_1, E_3, E_3, H)$ with positive real coefficients consider realizations $\phi : X \to \Sigma$ with $\phi(x) = \mathbf{x}_0$, with $\phi(x') = \mathbf{x}_H$, and with $\phi(h)$ equal to the segment of length H joining $\mathbf{x}_0$ and $\mathbf{x}_H$. Now, we have $\mathbf{const}\, e^{F_1+F_2}$ choices for the images of f_1 and f_2 subject to the restriction that $\phi(f_i)$ is for $i = 1, 2$ a geodesic segment of length at most F_i. Note also that the horospherical simple loop based at $\mathbf{x}_H$ has length $\mathbf{const} \cdot e^{-H}$ and that geodesic loops in the cusp, based at $\mathbf{x}_H$ and with length ℓ are homotopic to horospherical segments of length at most $\mathbf{const} \cdot e^{\frac{\ell}{2}}$. This implies that, if we want to map e_i to a loop in the cusp and of length at most E_i, then we have at least $\mathbf{const} \cdot e^{\frac{1}{2}E_i} \cdot e^H$ choices. Altogether we have at least

$$\mathbf{const} \cdot e^{F_1+F_2+\frac{1}{2}E_1+\frac{1}{2}E_2+\frac{1}{2}E_3+3H}$$

choices of (homotopy classes of) realizations of X into Σ with $\ell(\phi(f_1)) \leqslant F_1, \ell(\phi(f_2)) \leqslant F_2, \ell(\phi(e_1)) \leqslant E_1$, $\ell(\phi(e_2)) \leqslant E_2$ $\ell(\phi(e_3)) \leqslant E_3$, $\ell(\phi(H)) \leqslant H$. In particular, if we set

$$\vec{L} = (F_1, F_2, E_1, E_2, E_3, H) = \left(n, n, \frac{1}{2}n, \frac{1}{2}n, \frac{1}{2}n, \frac{5}{2}n\right)$$

we have at least $\mathbf{const} \cdot e^{\frac{61}{4}n}$ such realizations, and this is a much larger number than $\mathbf{const}\, e^{6n} = \mathbf{const}\, e^{\|L\|}$. This proves that the analogue of Lemma 4.1 fails if Σ is not compact.

In summary, if Σ has finite volume but is not compact, then we do not even know whether Theorem 1.3 holds. Although, we suspect that the answer is yes.

Can Σ be nonorientable?

If Σ is a compact hyperbolic surface which, however, is nonorientable, then we know that both Theorem 1.3 and Theorem 1.1 hold true: We never used orientability during their proofs. We did, however, in the proof of Theorem 1.2: Unless Σ is oriented, it makes little sense to speak about the induced fat graph structure. We do not know what happens with Theorem 1.2 when the ambient surface is not oriented.

And what about higher dimensions?

If we replace Σ by a closed hyperbolic manifold of other dimension than 2, then Theorem 1.3 and Theorem 1.1 should still hold and with proofs which, if not identical, keep the same spirit. It is, however, less clear whether there should be an interesting analogue of Theorem 1.2: at least if dim $\geqslant 5$, where every map of a surface can be deformed to an embedding.

A. The geometric prime number theorem

The argument we used to prove Theorem 1.3 can be used to recover Huber's geometric primer number theorem, and this is what we do here. Besides giving a simple proof of this theorem, it might help the reader understand the logic of the proof of Theorem 1.3.

The geometric prime number theorem (Huber). Let Σ be a closed, connected, orientable, hyperbolic surface, and let $\mathbf{C}(L)$ be the set of closed nontrivial oriented geodesics in Σ of length at most L. We have

$$|\mathbf{C}(L)| \sim \frac{e^L}{L}$$

as $L \to \infty$.

Remark. In Section 9, we discussed some of the difficulties one would face when extending the main results to the case that Σ is a finite volume surface or an orbifold. None of these problems really arise when proving the geometric prime number theorem, and indeed the proof we present here works with only minimal changes if Σ is replaced by an arbitrary finite area orbifold $\Gamma\backslash\mathbb{H}^2$. We decided to just deal with the compact case to avoid hiding the structure of the argument.

The proof of Huber's theorem would be cleaner and nicer if all closed geodesics were primitive. Luckily, this is almost true. Indeed, it follows, for example, from the work of Coornaert and Knieper [6] that there is some $C > 0$ with

$$\frac{1}{C}\cdot\frac{e^L}{L} \leqslant |\mathbf{C}(L)| \leqslant C\cdot e^L \tag{A.1}$$

for all $L > 0$. We stress that the Coornaert–Knieper argument is pretty coarse: What they actually prove is a statement about groups acting isometrically, discretely and cocompactly on Gromov-hyperbolic spaces.

The point for us is that Equation (A.1) implies that most geodesics of length at most L are primitive. Indeed, every nonprimitive geodesic of length at most L is a multiple of a primitive geodesic of length at most $\frac{1}{2}L$. On the other hand, if s_0 is the systole of Σ, then at most $\frac{L}{s_0}$ geodesics of length at most L arise as multiples of any given geodesic. These two observations, together with the right side of Equation (A.1), imply that there are at most $\frac{L}{s_0}\mathbf{C}(\frac{1}{2}L) \leqslant \frac{C}{s_0}\cdot L\cdot e^{L/2}$ nonprimitive geodesics of length at most L. Taking the left side of Equation (A.1) into consideration we deduce that, as we had claimed above, most geodesics are primitive.

Lemma A.1. *Fix h, and let $\mathbf{C}(L,h)$ and $\mathbf{P}(L,h)$ be, respectively, the sets of all geodesics and of all primitive geodesics of length in $[L, L+h]$. Then we have*

$$|\mathbf{C}(L,h)| \sim |\mathbf{P}(L,h)|$$

as $L \to \infty$.

After this preparatory comment, let us start the real business. Let $\mathcal{L}$ be the space of all geodesic loops in Σ, and consider the map $\Pi : \mathcal{L} \to \Sigma$ mapping each loop to its base point. As was the case for more general graphs, the map Π is a covering, meaning that when we pull back the hyperbolic metric using Π we can think of $\mathcal{L}$ as being a hyperbolic surface. Note also that the set of connected components of $\mathcal{L}$ agrees with the set of free homotopy classes of loops. It follows that for every closed geodesic γ we have a connected component $\mathcal{L}^\gamma$.

Now, given $\varepsilon > 0$ small let $\mathcal{L}_\varepsilon$ be the set of all geodesic loops with angle defect at most ε, that is, the set of geodesic loops whose initial and terminal velocity vectors meet with unoriented angle in $[0,\varepsilon]$. For $L > 0$, let $\mathcal{L}_\varepsilon(L, L+h)$ be the elements in $\mathcal{L}_\varepsilon$ with length between L and $L+h$. Accordingly, set $\mathcal{L}^\gamma_\varepsilon(L, L+h) = \mathcal{L}^\gamma \cap \mathcal{L}_\varepsilon(L, L+h)$.

Let us establish some basic properties of $\mathcal{L}_\varepsilon(L, L+h)$ and $\mathcal{L}^\gamma_\varepsilon(L, L+h)$:

Lemma A.2. *Fix $h > 0$ and $\varepsilon > 0$. We have that*

$$\mathrm{vol}(\mathcal{L}_\varepsilon(L, L+h)) \sim \varepsilon\cdot(e^{L+h} - e^L) \text{ as } L \to \infty.$$

Moreover, there is a function $C(\varepsilon)$ with $\lim_{\varepsilon\to 0} C(\varepsilon) = 1$ such that for every sufficiently long closed geodesic γ we have that

1. $\mathcal{L}^\gamma_\varepsilon(L, L+h) = \emptyset$ *unless* $\ell(\gamma) \in [L-\varepsilon, L+h]$,
2. $\mathrm{vol}(\mathcal{L}^\gamma_\varepsilon(L, L+h)) \leqslant C(\varepsilon)\cdot\varepsilon\cdot L$ *and*
3. $C(\varepsilon)^{-1} \leqslant \frac{1}{\varepsilon\cdot L}\mathrm{vol}(\mathcal{L}^\varepsilon_\gamma(L, L+h)) \leqslant C(\varepsilon)$ *if γ is primitive and* $\ell(\gamma) \in [L, L+h-\varepsilon]$.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

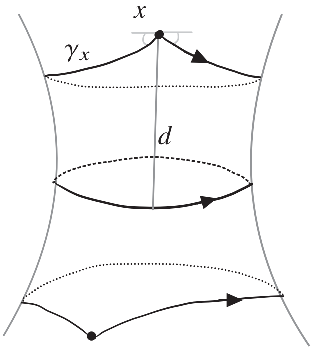

Figure 6. *The cylinder $\langle\gamma\rangle\backslash\mathbb{H}^2$. The marked angles each measures one half of the angle defect of γ_x.*

Proof. As in the proof of Proposition 4.2, we exploit the cover $\Pi : \mathcal{L} \to \Sigma$ to compute the volume of $\mathcal{L}_\varepsilon(L, L+h)$:

$$\mathrm{vol}(\mathcal{L}_\varepsilon(L, L+h)) = \int_\Sigma |\Pi^{-1}(x) \cap \mathcal{L}_\varepsilon(L, L+h)| dx.$$

Now, $\Pi^{-1}(x) \cap \mathcal{L}_\varepsilon(L, L+h)$ is nothing other than the number of geodesic arcs going from x to x with length within $[L, L+h]$ and such that the initial and terminal velocities make at most angle ε. Once we fix x, the set of admissive pairs of initial and terminal velocities is a subset $T^1_x\Sigma \times T^1_x\Sigma$ with volume $4\pi\varepsilon$. We thus get from Theorem 3.2 that

$$|\Pi^{-1}(x) \cap \mathcal{L}_\varepsilon(L, L+h)| \sim \varepsilon \cdot \frac{e^{L+h} - e^L}{\mathrm{vol}(\Sigma)}.$$

Since we are assuming that Σ is closed, this is uniform in x, meaning that we get

$$\mathrm{vol}(\mathcal{L}_\varepsilon(L, L+h)) \sim \int_\Sigma \varepsilon \cdot \frac{e^{L+h} - e^L}{\mathrm{vol}(\Sigma)} = \varepsilon \cdot (e^{L+h} - e^L).$$

We have proved the first claim.

To prove the rest, let us give a concrete description of $\mathcal{L}^\gamma$. Writing $\gamma = \eta^k$ for some η primitive and $k \geqslant 1$, consider the cover

$$\pi_\eta : \langle\eta\rangle\backslash\mathbb{H}^2 \to \Sigma$$

of Σ corresponding to η. Now, for each $x \in \langle\eta\rangle\backslash\mathbb{H}^2$ there is a unique hyperbolic loop γ_x based at x and freely homotopic to running k times over η. The basic observation is that the map

$$\langle\eta\rangle\backslash\mathbb{H}^2 \to \mathcal{L}^\gamma,\ x \mapsto \pi_\eta \circ \gamma_x$$

is an isometry between $\langle\eta\rangle\backslash\mathbb{H}^2$ and the connected component $\mathcal{L}^\gamma$.

Now, if one denotes by d the distance in $\langle\eta\rangle\backslash\mathbb{H}^2$ between x and the central geodesic (see Figure 6) one gets from formula 2.3.1(vi) in the final page in [3] that

$$\begin{aligned}\frac{1}{2} \cdot (\text{angle defect of } \gamma_x) &\sim \tan\left(\frac{1}{2} \cdot (\text{angle defect of } \gamma_x)\right)\\ &= \sinh(d) \cdot \tanh\left(\frac{\ell(\gamma)}{2}\right) \sim d,\end{aligned}$$

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press

where the asymptotics hold when $\ell(\gamma)$ is large and the angle defect is small. Now, claims (1), (2) and (3) follow from elementary considerations. □

Armed with these two lemmas, we are ready to conclude the proof of the geometric prime number theorem. Note that it suffices to prove that for h positive and fixed we have

$$|\mathbf{C}(L,h)| \sim \frac{e^{L+h}-e^{L}}{L} \tag{A.2}$$

as $L \longrightarrow \infty$. Here, $\mathbf{C}(L,h)$ is as in Lemma A.1.

Nevertheless, with notation as in both lemmas we have for ε positive and small and for L large that

$$\begin{aligned} |\mathbf{P}(L,h)| &\overset{\text{A.2(3)}}{\preceq} \frac{C(\varepsilon)}{\varepsilon\cdot L}\operatorname{vol}\big(\cup_{\gamma\in\mathbf{P}(L,h)}\mathcal{L}_\varepsilon^\gamma(L,L+h+\varepsilon)\big)\\ &\leqslant \frac{C(\varepsilon)}{\varepsilon\cdot L}\operatorname{vol}(\mathcal{L}_\varepsilon(L,L+h+\varepsilon))\\ &\overset{\text{A.2}}{\sim} C(\varepsilon)\cdot\frac{e^{L+h+\varepsilon}-e^{L}}{L}. \end{aligned}$$

On the other hand, we also have that

$$\begin{aligned} |\mathbf{C}(L,h)| &\overset{\text{A.2(2)}}{\geqslant} \frac{1}{C(\varepsilon)\cdot\varepsilon\cdot L}\operatorname{vol}\big(\cup_{\gamma\in\mathbf{C}(L,h)}\mathcal{L}_\varepsilon(L,h)\big)\\ &= \frac{1}{C(\varepsilon)\cdot\varepsilon\cdot L}\operatorname{vol}(\mathcal{L}_\varepsilon(L,h))\\ &\overset{\text{A.2}}{\sim} \frac{1}{C(\varepsilon)}\frac{e^{L+h}-e^{L}}{L}. \end{aligned}$$

Since this holds for all ε, and since $C(\varepsilon)$ tends to 1 when $\varepsilon \longrightarrow 0$ we have that

$$|\mathbf{P}(L,h)| \preceq \frac{e^{L+h}-e^{L}}{L} \preceq |\mathbf{C}(L,h)|.$$

Now, Lemma A.1 implies Equation (A.2), and Bob's your uncle.

Acknowledgements. We thank Vincent Delecroix, Lars Louder, Michael Magee and Peter Sarnak for their very helpful comments. The second author is also very grateful for the patience shown by Sébastien Gouëzel, Vincent Guirardel and François Maucourant with all his questions – it is nice to have one's office next to theirs.

Competing interests. The authors have no competing interest to declare.

Data availability statement. No data were used in this research.

Funding statement. The first author gratefully acknowledges support from EPSRC grant EP/T015926/1 and UiT Aurora Center for Mathematical Structures in Computations (MASCOT).

References

[1] R. Bacher and A. Vdovina, 'Counting 1-vertex triangulations of oriented surfaces', *Discrete Math.* **246** (2002), 13–27.
[2] B. Bekka and M. Mayer, *Ergodic Theory and Topological Dynamics of Group Actions on Homogeneous Spaces, London Mathematical Society Lecture Note Series*, vol. 269 (Cambridge University Press, Cambridge, 2000), x+200pp.
[3] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*, *Progress in Mathematics*, vol. 106 (Birkhäuser, 1992), xiv+454pp.
[4] D. Calegari, *scl*, MSJ Memoirs, vol. 20 (Mathematical Society of Japan, 2009), xii+209pp.
[5] R. Canary, D. Epstein and P. Green, 'Notes on notes of Thurston', in *Analytical and Geometric Aspects of Hyperbolic Space (Coventry/Durham, 1984)*, London Math. Soc. Lecture Note Ser., vol. 111 (Cambridge Univ. Press, Cambridge, 1987), 3–92.
[6] M. Coornaert and G. Knieper, 'Growth of conjugacy classes in Gromov hyperbolic groups', *GAFA* **12** (2002), 464–478.

[7] J. Delsarte, ‘Sur le gitter fuchsien’, *C. R. Acad. Sci. Paris* **214** (1942), 147–179.

[8] V. Erlandsson and J. Souto, *Mirzakhani's Curve Counting and Geodesic Currents*, *Progress in Mathematics*, vol. 345 (Birkhäuser, 2022), xii+226pp.

[9] D. Hejhal, *The Selberg Trace Formula for PSL* $(2, \mathbb{R})$, *Vol. II*, Lecture Notes in Mathematics, vol. 1001 (Springer-Verlag, Berlin-Heidelberg, 1983), viii+806pp.

[10] H. Huber, ‘Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen’, *Math. Ann.* **138** (1959), 1–26.

[11] A. Katsuda and T. Sunada, ‘Homology and closed geodesics in a compact Riemann surface’, *Amer. J. Math.* **110** (1988), 145–155.

[12] G. Margulis, *On Some Aspects of the Theory of Anosov Systems. With a Survey by Richard Sharp: Periodic Orbits of Hyperbolic Flows*, Springer Monographs in Mathematics (Springer-Verlag, 2004), vi+139pp.

[13] M. Mirzakhani, ‘Growth of the number of simple closed geodesics on hyperbolic surfaces’, *Ann. of Math.* (**2**) 168 (2008), 97–125.

[14] M. Mirzakhani, ‘Counting mapping class group orbits on hyperbolic surfaces’, Preprint, 2016, arXiv:1601.03342.

[15] P. Park, ‘Conjugacy growth of commutators’, *Journal of Algebra* **526** (2019), 423–458.

[16] R. Phillips and P. Sarnak, ‘Geodesics in homology classes’, *Duke Math. J.* **55** (1987), 287–297.

[17] T. Roblin, ‘Ergodicité et équidistribution en courbure négative’, *Mém. Soc. Math. Fr. No.* **95** (2003), vi+96pp.

[18] W. Thurston, ‘Geometry and topology of 3-manifolds, lecture notes (1980). URL: http://library.msri.org/books/gt3m/.

https://doi.org/10.1017/fms.2023.114 Published online by Cambridge University Press