STATISTICAL HYPERBOLICITY IN TEICHMÜLLER SPACE

SPENCER DOWDALL, MOON DUCHIN, AND HOWARD MASUR

ABSTRACT. In this paper we explore the idea that Teichmüller space is hyperbolic "on average." Our approach focuses on studying the geometry of geodesics which spend a definite proportion of time in some thick part of Teichmüller space. We consider several different measures on Teichmüller space and find that this behavior for geodesics is indeed typical. With respect to each of these measures, we show that the average distance between points in a ball of radius r is asymptotic to $2r$, which is as large as possible. Our techniques also lead to a statement quantifying the expected thinness of random triangles in Teichmüller space, showing that "most triangles are mostly thin."

1. INTRODUCTION

Let S be a closed surface of genus $g > 1$. In this paper we continue the study of metric properties of Teichmüller space $\mathcal{T}(S)$, which is the parameter space for several types of geometric structures on S. Equipped with the Teichmüller metric $d_{\mathcal{T}}$, it is a complete metric space homeomorphic to $\mathbb{R}^{6g-6}$. It is not δ–hyperbolic [18], and several kinds of obstructions to hyperbolicity are known: for instance, pairs of geodesic rays through the same point may fellow-travel arbitrarily far apart [12], and there are large "thin parts" of the space which, up to bounded additive error, are isometric to product spaces equipped with sup metrics (and therefore rule out hyperbolicity in the space as a whole) [19]. However, these exceptions to negative curvature seem to come from rare occurrences, while a long list of properties associated with hyperbolicity do hold globally or in specialized situations. Thus one may expect such properties to hold generically or "on average." This paper aims to show that this is indeed the case.

Our motivating goal is to understand the generic geometry of Teichmüller space, particularly with regard to negative-curvature phenomena. For example, geodesics that stay in the thick part of $\mathcal{T}(S)$ are well understood and exhibit many properties characteristic of hyperbolicity. Geodesics lying completely in the thin part are also well understood, and exhibit no negative-curvature characteristics. However, much more typical is for a geodesic to spend time in both the thick and thin parts of $\mathcal{T}(S)$—indeed, a generic geodesic ray will switch between these parts infinitely often. In this paper we develop tools to study these types of geodesics, and we discover that certain negative-curvature phenomena do hold in this setting. For example, we obtain the following variant of the thin triangle property.

Theorem A. *For any $\epsilon > 0$ and $0 < \theta \leq 1$, there exist constants C, L such that if $I \subset [x, y] \subset \mathcal{T}(S)$ is a geodesic subinterval of length at least L and at least*

proportion θ of I is ϵ–thick, then for all $z \in \mathcal{T}(S)$, we have

$$I \cap \mathrm{Nbhd}_C([x, z] \cup [y, z]) \neq \emptyset.$$

This is a generalization of a result of Rafi, Theorem 3.13 below, which gives the same conclusion under the stronger hypothesis that the entire interval I is thick. As a consequence (Corollary 3.15), we can assert for instance that if the three sides of a triangle each spend more than half their time in the ϵ–thick part, then half of each side is within a uniformly bounded distance of the union of the other two sides.

Since our goal is to study the geometry of generic rays, Theorem A motivates a consideration of whether randomly sampled geodesics are, with high probability, likely to spend a definite fraction of time in a given thick part. To address this, we investigate a number of a priori different measures on Teichmüller space, which are natural from various points of view, and show that having a definite proportion in the thick part is in fact typical for all of them.

For instance, as a metric space $\mathcal{T}(S)$ carries a $(6g - 6)$–dimensional Hausdorff measure η. This allows us to fix x and choose $y \in \mathcal{B}_r(x)$ at random from the r–ball as a way of sampling geodesics $[x, y]$. Other measures come from the Finsler structure (Busemann measure μ_{B} and Holmes–Thompson measure μ_{HT}), from the holonomy coordinates on the cotangent bundle (holonomy, or Masur–Veech, measure $\mathbf{m}$), and from the symplectic structure. In §4, we find that all of these measures are mutually absolutely continuous and in fact are related by explicit inequalities. Further interesting measures are provided by the identification of the metric r–sphere $\mathcal{S}_r(x)$ with the unit sphere $\mathcal{Q}^1(x)$ in the vector space of quadratic differentials on x via the Teichmüller map. The latter has various natural measures, and corresponding measures on $\mathcal{S}_r(x)$ will be called *visual measures*; we will pay special attention to two standard visual measures, denoted $\mathrm{Vis}_r(\nu_x)$ and $\mathrm{Vis}_r(s_x)$. We will also use the term visual measures for the induced measures on $\mathcal{T}(S)$, denoted $\mathrm{Vis}(\nu_x)$ and $\mathrm{Vis}(s_x)$, obtained by integrating radially.

As one application of our statistical approach, we compute a statistic built by combining a metric and a measure to quantify how fast a space "spreads out." Suppose we are given a family of probability measures μ_r on the spheres $\mathcal{S}_r(x)$ of a metric space (X, d). Then let $E(X) = E(X, x, d, \{\mu_r\})$ be the average normalized distance between points on large spheres:

$$E(X) := \lim_{r \to \infty} \frac{1}{r} \int_{\mathcal{S}_r(x) \times \mathcal{S}_r(x)} d(y, z) \ d\mu_r(y) d\mu_r(z),$$

if the limit exists. This creates a numerical index varying from 0 (least spread out) to 2 (most spread out). It is shown in [6] that non-elementary hyperbolic groups all have $E(G, S) = 2$ for any finite generating set S; this is also the case in the hyperbolic space $\mathbb{H}^n$ of any dimension endowed with the natural measure on spheres. By contrast, it is shown that $E(\mathbb{R}^n) < \sqrt{2}$ for all n, and that $E(\mathbb{Z}^n, S) < 2$ for all n and S, with nontrivial dependence on S. (See [6] for more examples.) Motivated by these findings, we may regard a measured metric space with $E = 2$ as being "statistically hyperbolic."

We note that finding that $E = 2$ for hyperbolic groups makes use of homogeneity. In contrast, it is easy to build (highly non-regular) locally finite trees, equipped with counting measure on spheres, for which E obtains any value from 0 to 2; see [6, p.4]. Thus neither δ–hyperbolicity nor exponential growth is sufficient to ensure

$E = 2$. Indeed, since the measures are normalized, the growth rate of the space has no direct effect on E. As an illustration, note that the Euclidean plane could be endowed with a visual measure, constructed just like the ones we study below in §4.3, which would give it exponential volume growth while leaving $E = 4/\pi$ unchanged. On the other hand, other measures on $\mathbb{R}^2$ would give different values of E; the statistic is quite sensitive to the choice of measure.

The following theorem concerns the average distance between points in the ball $\mathcal{B}_r(x)$ of radius r centered at x. We show that this average distance is asymptotic to $2r$, which, in light of the triangle inequality, is the maximum possible distance.

Theorem B. *Let μ denote the Hausdorff measure η, holonomy measure $\mathbf{m}$, or either standard visual measure $\mathrm{Vis}(\nu_x)$ or $\mathrm{Vis}(s_x)$. Then for every point $x \in \mathcal{T}(S)$,*

$$\lim_{r \to \infty} \frac{1}{r} \frac{1}{\mu(\mathcal{B}_r(x))^2} \int_{\mathcal{B}_r(x) \times \mathcal{B}_r(x)} d_{\mathcal{T}}(y, z) \ d\mu(y) d\mu(z) = 2.$$

Of course, by the remarks above, this also holds for all the other measures discussed in the paper. Indeed, we will work with properties of measures on $\mathcal{T}(S)$ that suffice to guarantee this conclusion: a *thickness property* (P1) defined in §5.2 guaranteeing that typical rays spend a definite proportion of their time in the thick part, and a *separation property* (P2) defined in §6 asserting that typical pairs of rays will exceed any definite amount of separation. In some places we use a stronger separation property (P3) which is a quantified version with an exponential bound.

With respect to the standard visual measures, the same methods yield:

Theorem C. *For every point $x \in \mathcal{T}(S)$ and either family $\{\mu_r\}$ of standard visual measures $\mu_r = \mathrm{Vis}_r(\nu_x)$ or $\mathrm{Vis}_r(s_x)$ on the spheres $\mathcal{S}_r(x)$, we have*

$$E(\mathcal{T}(S), x, d_{\mathcal{T}}, \{\mu_r\}) = 2.$$

As a second application of our approach, we promote Theorem A to a quantitative statement about the expected thinness of typical triangles. This is expressed in the following theorem, which shows that "most triangles are mostly thin." For a fixed δ, let $0 \leq \Theta_\delta(\triangle) \leq 1$ denote the proportion of the perimeter of a geodesic triangle $\triangle$ that lies within δ of the other two sides. Then let $\Theta_\delta(X) = \Theta_\delta(X, x, d, \mu)$ be the limiting average of this value:

$$\Theta_\delta(X) := \liminf_{r \to \infty} \frac{1}{\mu(\mathcal{B}_r(x))^2} \int_{\mathcal{B}_r(x) \times \mathcal{B}_r(x)} \Theta_\delta(\triangle(x, y, z)) \ d\mu(y) d\mu(z).$$

Theorem D. *Let μ denote either the Hausdorff measure η or the holonomy measure $\mathbf{m}$. Then for all $x \in \mathcal{T}(S)$ and $\sigma > 0$ there exists $\delta > 0$ such that*

$$\Theta_\delta(\mathcal{T}(S), x, d_{\mathcal{T}}, \mu) \geq 1 - \sigma.$$

In other words, the proportion of a triangle's perimeter that is close to the other two sides can be made arbitrarily close to 1 in expectation. By contrast, note that $\Theta_\delta(\mathbb{R}^n) = 0$ for all δ, whereas δ–hyperbolic spaces X automatically satisfy $\Theta_\delta(X) = 1$ by definition.

We sketch here the main ideas in the proofs of the theorems. Theorem A is put together with distance estimates coming from subsurface projections, using reverse triangle inequalities (following Masur–Minsky and Rafi), and antichain bounds (Rafi and Schleimer). A crucial ingredient is to show that geodesics spending a definite proportion of time in the thick part have shadows that make definite progress in the curve complex (Theorem 3.10).

The idea for Theorem B is that the separation property (P2) ensures that most pairs of geodesics will have stopped fellow-traveling in the Teichmüller metric by a threshold time. Then one would hope that, as in a hyperbolic space, the geodesic joining their endpoints would follow the first geodesic back to approximately where they separate before following the other so that its length is roughly the sum of the lengths of the two geodesics, as on the left in Figure 1.
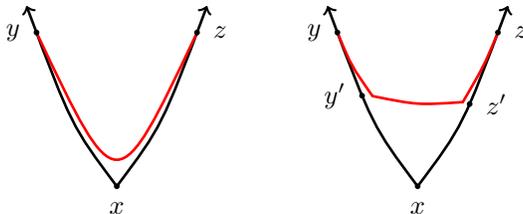

FIGURE 1. We will show that the geodesic between points on generic rays "dips back" near the basepoint. While Minsky's product regions theorem says that the connecting geodesic can instead take a "shortcut" when $[x, y']$ and $[x, z']$ go through thin parts corresponding to disjoint subsurfaces, we show this effect is rare.

One obstruction to this hyperbolic-like behavior is that the pair of geodesics can enter thin parts corresponding to disjoint subsurfaces, in which case Minsky's product region theorem [19] allows the length of the third side to be smaller than the sum, as on the right in Figure 1. The thickness property (P1) and Theorem A together rule out this shortcut behavior. Theorem C follows from this and exponential growth of the metric.

Theorem D uses Theorem A and a strengthening of the thickness property (P1) to find thick points in various locations around a typical triangle. The proof then concludes by applying Rafi's fellow-traveling theorem [22] to deduce that most of each side lies close to other sides.

In establishing the needed thickness and separation properties for the above results, we use a variety of recently developed tools such as volume asymptotics in Teichmüller space (Athreya–Bufetov–Eskin–Mirzakhani [3]) and the random walk model for discretized Teichmüller geodesics (Eskin–Mirzakhani [7]). A detailed treatment of the latter is included in Appendix B. We also make use of a simplified version of Rafi's distance formula [21], which is derived in Appendix A.

1.1. **Acknowledgments.** We would like to thank Benson Farb, Curtis McMullen, and especially Alex Eskin and Kasra Rafi for numerous helpful comments and explanations. In particular, we are indebted to Kasra Rafi for encouraging us to investigate expected thinness (Theorem D) and for suggesting the idea of the proof for Theorem 5.10. We are also grateful for the suggestions of the anonymous referees, which led to major improvements in the paper.

2. BACKGROUND

2.1. **Teichmüller space and quadratic differentials.** Recall that Teichmüller space $\mathcal{T}(S)$ is the space of marked Riemann surfaces X that are homeomorphic to the topological surface S. More precisely, it consists of pairs (X, f), where $f \colon S \to$

X is a homeomorphism, up to the equivalence relation that $(X_1, f_1) \sim (X_2, f_2)$ when there exists a conformal map $F \colon X_1 \to X_2$ such that $F \circ f_1$ is isotopic to f_2. Alternately, we may define $\mathcal{T}(S)$ as the space of marked hyperbolic surfaces (ρ, f); namely, the markings are maps $f \colon S \to \rho$ with $(\rho_1, f_1) \sim (\rho_2, f_2)$ when there exists an isometry $F \colon \rho_1 \to \rho_2$ such that $F \circ f_1$ is isotopic to f_2.

The space $\mathcal{T}(S)$ is homeomorphic to the ball $\mathbb{R}^{6g-6}$, and from now on we will use $h = 6g - 6$ to designate this dimension. In this paper, we will typically denote a point of $\mathcal{T}(S)$ by x, regarding it either as a Riemann surface or a hyperbolic surface, and suppressing the marking f.

Using the first definition of $\mathcal{T}(S)$, the Teichmüller distance is given by

$$d_{\mathcal{T}}((X_1, f_1), (X_2, f_2)) := \inf_{F \sim f_2 \circ f_1^{-1}} \frac{1}{2} \log K(F),$$

where the minimum is taken over all quasiconformal maps F and $K(F)$ is the maximal dilatation of F. Equipped with this metric, Teichmüller space becomes a unique geodesic metric space. For $x, y \in \mathcal{T}(S)$, the Teichmüller geodesic segment joining x to y will usually be denoted $[x, y]$. We will also write y_t for the time–t point on the ray based at x and going through y.

A *quadratic differential* on a Riemann surface X is a holomorphic 2–tensor $q = \phi(z)dz^2$ on X. The space of all quadratic differentials on all Riemann surfaces homeomorphic to S is denoted $\mathcal{Q}(S)$. A point of $\mathcal{Q}(S)$ will be denoted q, with the underlying complex structure implicit in the notation. The real dimension of $\mathcal{Q}(S)$ is $12g - 12 = 2h$. Reading off the Riemann surface, we obtain a projection to the Teichmüller space $\pi \colon \mathcal{Q}(S) \to \mathcal{T}(S)$. Under this projection, $\mathcal{Q}(S)$ forms vector bundle over $\mathcal{T}(S)$ which is canonically identified with the cotangent bundle of $\mathcal{T}(S)$. Each fiber $\mathcal{Q}(X)$ is equipped with a norm given by the total area of q; namely $\|q\| = \int_X |\phi(z)dz^2|$. Recall that $d_{\mathcal{T}}$ is not a Riemannian metric on $\mathcal{T}(S)$, but rather a Finsler metric; it comes from dualizing the norm on $\mathcal{Q}$ to give a norm on each tangent space of $\mathcal{T}(S)$ that is not induced by any inner product.

It is the famous theorem of Teichmüller that the infimum in the definition of $d_{\mathcal{T}}$ is realized uniquely by a *Teichmüller map* from X_1 to X_2. A Teichmüller map is determined by an initial quadratic differential $q = \phi(z)dz^2$ on X_1 and the number K. The Teichmüller map expands along the horizontal trajectories of q by a factor of $K^{1/2}$ and contracts along the vertical trajectories by the same factor to obtain a terminal quadratic differential q' on the image surface X_2. If we fix q and let $K = e^{2t}$ vary over $t \in [0, \infty)$ we get a Teichmüller geodesic ray. We will denote by ν_+ the horizontal foliation of q and by ν_- the vertical foliation.

Recall that the *mapping class group* of S, defined by

$$\mathrm{Mod}(S) := \mathrm{Diff}^+(S)/\mathrm{Diff}_0(S),$$

is the discrete group of orientation-preserving diffeomorphisms of S, up to isotopy. This group acts isometrically on $\mathcal{T}(S)$ by changing the marking: $\phi \cdot (X, f) = (X, f \circ \phi^{-1})$. In fact, by a result of Royden [24], $\mathrm{Mod}(S)$ is the full group of (orientation-preserving) isometries of $(T(S), d_{\mathcal{T}})$.

2.2. **Curve complex.** When we speak of a *curve* on S, this will mean an isotopy class of essential simple closed curves. Given $x \in \mathcal{T}(S)$, the *length* $l_x(\alpha)$ of a curve α is the length of the geodesic in the isotopy class in the hyperbolic metric x.

We recall the definition of the *curve complex* (or curve graph) $\mathcal{C}(S)$ of S. The vertices of $\mathcal{C}(S)$ are the curves on S. Two vertices are joined by an edge if the corresponding curves can be realized disjointly. Assigning edges to have length 1 we have a metric graph. Properly speaking, $\mathcal{C}(S)$ is the flag complex associated to this curve graph, but since we are working coarsely, we can identify $\mathcal{C}(S)$ with the graph.

It is known that the curve graph is hyperbolic [17]. That is, there exists a constant $\delta > 0$ such that every geodesic triangle in $\mathcal{C}(S)$ is δ–thin: each side of the triangle is contained in the union of the δ–neighborhoods of the other two sides. Furthermore, in any δ–hyperbolic metric space and for any quasi-isometry constants (K, C), there exists a constant τ, depending only on δ, K, C, such that any two (K, C)–quasi-geodesic segments with the same endpoints remain within τ of each other. Since actual geodesics are $(1, 0)$–quasi-geodesics, this implies that every (K, C)–quasi-geodesic triangle is $(\delta + 2\tau)$–thin.

2.3. **Thick parts and subsurface projections.** For any given ϵ, we say a curve is ϵ–short if its hyperbolic length is less than ϵ. Then define the ϵ–*thick part* of Teichmüller space to be the subset $\mathcal{T}_\epsilon \subset \mathcal{T}(S)$ corresponding to those hyperbolic surfaces on which no curve is ϵ–short. Its complement is called the ϵ–*thin part* or, when ϵ is understood, simply the thin part.

For each $x \in \mathcal{T}(S)$ there is associated a *Bers marking* μ_x. To construct μ_x, greedily choose a shortest *pants decomposition* of the surface (a collection of $3g - 3$ disjoint simple geodesics). Then for each pants curve β, choose a shortest geodesic crossing β minimally (either once or twice depending on the topology) that is disjoint from all other pants curves. The total collection of $6g - 6$ curves is called a Bers marking and is defined up to finitely many choices. Notice that the curves comprising μ_x form a diameter–2 subset of $\mathcal{C}(S)$.

Recall that there exists a universal *Margulis constant* such that any two curves with hyperbolic length (on any surface $x \in \mathcal{T}(S)$) less than this value are disjoint. When discussing the ϵ–thick part $\mathcal{T}_\epsilon$, we always assume ϵ is less than the Margulis constant. In particular, this ensures that for $x \in \mathcal{T}(S) \setminus \mathcal{T}_\epsilon$, the Bers marking μ_x contains every curve α with $l_x(\alpha) \leq \epsilon$.

Throughout, a *proper subsurface of S* will mean a compact, properly embedded subsurface $V \subset S$ which is not equal to S and for which the induced map on fundamental groups is injective. Subsurfaces which are isotopic to each other will not be considered distinct. The proper subsurfaces of S fall into two categories, annuli and non-annuli, which behave somewhat differently. Nevertheless, we will strive to develop intuitive notation under which these two possibilities may be dealt with on equal footing.

Every proper subsurface V has a nonempty boundary ∂V consisting of a disjoint union of curves on S. We say that two subsurfaces V and W *transversely intersect*, denoted $V \pitchfork W$, if they are neither (isotopically) disjoint nor nested. In this case, ∂V necessarily intersects W, and ∂W intersects V.

Consider a non-annular subsurface V, possibly equal to S. The *subsurface projection* $\pi_V(\beta)$ of a simple closed curve $\beta \subset S$ to V is defined as follows: Realize β and ∂V as geodesics (in any hyperbolic metric on S). If $\beta \subset V$, then $\pi_V(\beta)$ is defined to be β. If β is disjoint from V, then $\pi_V(\beta)$ is undefined. Otherwise, $\beta \cap V$ is a disjoint union of finitely many homotopy classes of arcs with endpoints on ∂V, and we obtain $\pi_V(\beta)$ by choosing any arc and performing a surgery along ∂V to

create a simple closed curve contained in V. The subsurface projection of a point $x \in \mathcal{T}(S)$ is then defined to be the collection

$$\pi_V(x) := \{\pi_V(\beta)\}_{\beta \in \mu_x}$$

of curves obtained by varying β in the Bers marking at x. This is a non-empty subset of the curve complex $\mathcal{C}(V)$ with uniformly bounded diameter.

Definition 2.1 (Non-annular projection distance)**.** For a non-annular subsurface $V \subseteq S$, the *projection distance in V* of a pair of points $x, y \in \mathcal{T}(S)$ is defined to be

$$d_V(x, y) := \operatorname{diam}_{\mathcal{C}(V)}(\pi_V(x) \cup \pi_V(y)).$$

In particular, $d_S(x, y)$ denotes the curve complex distance. When convenient, we will also denote this distance by $d_{\mathcal{C}(V)} := d_V$.

For an annular subsurface $A \subset S$ with *core curve* $\alpha = \partial A$, there are two kinds of projection distances: one that measures twisting about α and is analogous to the definition above, and a second which also incorporates the length of α. Any simple closed curve β that crosses α may be realized by a geodesic and then lifted to a geodesic $\tilde{\beta}$ in the annular cover $\tilde{A}$, that is, the quotient of $\mathbb{H}^2$ by the deck transformation corresponding to α, with the Gromov compactification. For a pair β, γ of such curves, we may then consider the intersection number $\operatorname{i}(\tilde{\beta}, \tilde{\gamma})$ in $\tilde{A}$. The *twisting distance* in A of a pair of points $x, y \in \mathcal{T}(S)$ is then defined as

$$d_{\mathcal{C}(A)}(x, y) := \sup_{\beta \in \mu_x, \gamma \in \mu_y} \operatorname{i}_{\tilde{A}}(\tilde{\beta}, \tilde{\gamma}).$$

We additionally define a hyperbolic projection distance as follows.

Definition 2.2 (Annular projection distance)**.** For an annular subsurface $A \subset S$ with core curve $\alpha = \partial A$, we let $\mathbb{H}_\alpha$ denote a copy of the standard horoball $\{\operatorname{Im}(z) \geq 1\} \subset \mathbb{H}^2$. Given $x, y \in \mathcal{T}(S)$, we consider the points $(0, 1/l_x(\alpha))$ and $(d_{\mathcal{C}(A)}(x, y), 1/l_y(\alpha)) \in \mathbb{H}^2$ and denote their closest point projections to the horoball $\mathbb{H}_\alpha$ by

$$\pi_\alpha(x) = \left(0, \max\left\{1, \frac{1}{l_x(\alpha)}\right\}\right), \quad \pi_\alpha(y) = \left(d_{\mathcal{C}(A)}(x, y), \max\left\{1, \frac{1}{l_y(\alpha)}\right\}\right).$$

The *projection distance* in A (or *hyperbolic distance* $d_{\mathbb{H}_\alpha}$) between x and y is then defined to be

$$d_A(x, y) := d_{\mathbb{H}^2}\left(\pi_\alpha(x), \pi_\alpha(y)\right).$$

2.4. **Notation.** Following Rafi [21, §2.4], we fix a parameter $\epsilon_0 > 0$ for the entirety of this paper which is smaller than the Margulis constant and small enough for a few other fundamental results to hold (Minsky's product regions theorem and Rafi's distance estimates described in the following section). Note that the definition of ϵ_0 depends only on the topology of the surface S, and we therefore view ϵ_0 as a global constant.

Our analysis involves many inequalities that have controlled multiplicative and additive error. To streamline the the presentation, we will often avoid explicitly writing the constants involved and will instead rely on the following notation: For real-valued expressions A and B, we use the notation

$$A \dot{\prec} B$$

to mean that there exists a universal constant $c \geq 1$, depending only on the topology of the surface S, such that $A \leq cB$. We will use $A \overset{\cdot}{\asymp} B$ to mean that $A \overset{\cdot}{\prec} B$ and $A \overset{\cdot}{\succ} B$ both hold. (The dot in the symbols indicates that the error is only multiplicative.) When allowing for multiplicative and additive error we will instead use symbols $\prec$, $\succ$, and $\asymp$. Thus $A \asymp B$ means that there exists a universal constant $c \geq 1$ so that $A \leq cB + c$ and $B \leq cA + c$.

When the implied constant depends on additional parameters we will list these as subscripts of the binary relation. For example, $A \overset{\cdot}{\prec}_{\epsilon, \theta} B$ means that there exists a constant c depending only on ϵ, θ, and the topology of S such that $A \leq cB$.

2.5. **Distance formula.** The following distance formula due to Rafi relates the Teichmüller distance between two points x and y to the combinatorics of the corresponding Bers markings μ_x and μ_y. Recall the global constant $\epsilon_0 > 0$ introduced in §2.4 above.

Theorem 2.3 (Distance formula, Rafi [21]). *Given any sufficiently large threshold M_0, for all $x, y \in \mathcal{T}(S)$ we have*

$$d_{\mathcal{T}}(x, y) \; \asymp_{M_0} \; d_S(x, y) + \sum_V [d_V(x, y)]_{M_0} + \max_{\alpha \in \Gamma_{xy}} d_{\mathbb{H}_\alpha}(x, y)$$

$$+ \sum_{A \,:\, \partial A \notin \Gamma_{xy}} \log_+ \big[d_{\mathcal{C}(A)}(x, y) \big]_{M_0} + \max_{\alpha \in \Gamma_x} \log_+ \left(\frac{1}{l_x(\alpha)} \right) + \max_{\alpha \in \Gamma_y} \log_+ \left(\frac{1}{l_y(\alpha)} \right),$$

where the first sum is over all non-annular proper subsurfaces $V \subsetneq S$, where Γ_{xy} is the set of ϵ_0–short curves in both x and y, Γ_x is the set of curves that are ϵ_0–short in x but not in y, and Γ_y is defined similarly. Here and throughout, $\log_+$ is a modified logarithm so that $\log_+ a = 0$ for $a \in [0, 1]$; and $[\cdot]_{M_0}$ is a threshold function for which $[N]_{M_0} := N$ when $N \geq M_0$ and $[N]_{M_0} := 0$ otherwise.

By instead making all annular measurements with the hyperbolic distance on $\mathbb{H}_\alpha$ we will obtain a particularly simple restatement of this formula.

Proposition 2.4 (Repackaged distance formula). *Given any sufficiently large threshold M_0, for all $x, y \in \mathcal{T}(S)$ we have:*

$$(2.5) \qquad\qquad d_{\mathcal{T}}(x, y) \; \asymp_{M_0} \; d_S(x, y) + \sum_Y [d_Y(x, y)]_{M_0}$$

Here, the sum is over all (annular and non-annular) proper subsurfaces.

Remark 2.6. Our definition of $d_{\mathbb{H}_\alpha} = d_A$ is technically different than that used by Rafi in [21]; however, the two definitions agree up to bounded additive error.

In calling it "repackaged" we mean to say that the content of (2.5) is essentially contained in Rafi [21]. We include a detailed proof here in Appendix A.

2.6. **Thin intervals.** We will use some results from Rafi's work combinatorializing the Teichmüller metric. Specifically, Corollary 3.4 and Proposition 3.7 of [21] show that for every Teichmüller geodesic and every proper subsurface V, there is a (possibly empty) interval along the geodesic where ∂V is short. Outside of this interval, the projections d_V move by at most a bounded amount. In the form that we will use below: for each positive $\epsilon \leq \epsilon_0$ there are positive constants M_ϵ and $\epsilon' \leq \epsilon$ such that for any pair of points $x, y \in \mathcal{T}(S)$ there is a possibly empty (and not uniquely defined) connected interval I_V^ϵ along the geodesic segment $[x, y]$ such that

- for $a \in \mathsf{I}_V^\epsilon$, the length each component of ∂V on a is at most ϵ;
- for $a \in [x, y] \setminus \mathsf{I}_V^\epsilon$, some component β of ∂V has $l_a(\beta) \geq \epsilon'$;
- for a, b in the same component of $[x, y] \setminus \mathsf{I}_V^\epsilon$, we have $d_V(a, b) < \mathsf{M}_\epsilon$; and
- if $V \pitchfork W$ then $\mathsf{I}_V^\epsilon \cap \mathsf{I}_W^\epsilon = \emptyset$.

This I_V^ϵ is called the ϵ–*thin interval* for V, or just the *thin interval* when ϵ is understood. While $[x, y]$ is suppressed in the notation, the geodesic with respect to which the interval I_V^ϵ is defined should be clear from context.

Remark 2.7. Note that for us I_V^ϵ is a segment in Teichmüller space, whereas Rafi works with the corresponding time interval $I \subset \mathbb{R}$. We also caution that I_V is not necessarily the same as the "active interval" for V considered by Rafi in [22], where it is additionally required that the restriction of $[x, y]$ to V behaves like a unit-speed Teichmüller geodesic in $\mathcal{T}(V)$.

If $\mathsf{I}_V^\epsilon \neq \emptyset$ we will say that V *becomes thin* along $[x, y]$. In particular if $d_V(x, y) \geq \mathsf{M}^\epsilon$, then $\mathsf{I}_V^\epsilon \neq \emptyset$ and so V becomes thin along $[x, y]$. Note that the second condition above says that the complement of the union of thin intervals (for all proper subsurfaces) lies in the ϵ'–thick part of $\mathcal{T}(S)$.

We always assume that M_ϵ is chosen large enough to be a valid threshold in in the distance formula (2.5). In the case $\epsilon = \epsilon_0$ we will omit the parameter and simply write M and I_V for M_{ϵ_0} and $\mathsf{I}_V^{\epsilon_0}$. Thus M is a global constant that depends only on the topology of S.

2.7. **Reverse triangle inequality.** We will repeatedly use the fact that the projection of a Teichmüller geodesic to the curve complex of any subsurface other than an annulus forms an unparameterized quasi-geodesic that, in particular, does not backtrack. This phenomenon is captured by the following "reverse triangle inequality," which was proved first in the case of the curve complex of the whole surface by Masur–Minsky [17] and then for general subsurfaces by Rafi [22, Thm B].

Lemma 2.8 (Reverse triangle inequality). *There exists a global constant* $\mathsf{B} > 0$ *such that for any non-annular subsurface V (including S itself) and for any geodesic interval $[x, y] \subset \mathcal{T}(S)$ and any point $a \in [x, y]$ we have*

$$(2.9) \qquad d_V(x, a) + d_V(a, y) \leq d_V(x, y) + \mathsf{B}.$$

In the exceptional annulus case, Rafi [22] shows that the reverse triangle inequality for $d_{\mathcal{C}(A)}$ does hold when the twisting distance is measured with respect to the quadratic differential defining the geodesic. However, it is unknown whether the reverse triangle inequality holds with twisting defined in terms of the hyperbolic metric, as it is in this paper. So, to deal with this exceptional case we instead appeal to the following result, still following Rafi [20], (c.f., Theorem 5.5 of [22]): though projection to the annulus may not be large between the points we consider, we find a subsurface that does register a large projection distance.

Lemma 2.10 (R.T.I. exception). *For any sufficiently large M', there exists $\epsilon' > 0$ with the following property. Suppose that a simple curve α on S satisfies $l_a(\alpha) \leq \epsilon'$ for some point $a \in [x, y]$ with $d_S(a, x), d_S(a, y) \geq 6 + \mathsf{B}$. Then there exists a subsurface $Z \subsetneq S$ disjoint from α (possibly the annulus with core curve α) for which $d_Z(x, y) > M'$.*

Proof. Let $\nu^\pm$ denote horizontal and vertical foliations for the Teichmüller geodesic $[x, y]$. By choosing ϵ' small, thus forcing $1/l_a(\alpha)$ to be large, Theorem 5.6 of [20]

ensures that we can find a component Y of $S \setminus \alpha$ (possibly the annulus with core curve α) for which the intersection number

$$i_Y(\nu_Y^+, \nu_Y^-)$$

is as large as we like (see §2 of [20] for the definitions of i_Y and of the projections $\nu_Y^\pm$ to the "arc and curve complex" of Y). As is well known, the two arc systems $\nu_Y^\pm$ can only have large intersection number if the projection $d_Z(\nu_Y^+, \nu_Y^-)$ to some subsurface $Z \subset Y$ is large. Thus, by choosing ϵ' sufficiently small, we may assume that there is a subsurface $Z \subset S$ disjoint from α (possibly the annulus with core curve α) for which $d_Z(\nu^+, \nu^-) > M' + 2\mathsf{M}$. It follows that Z determines a nonempty thin interval I_Z along the bi-infinite geodesic $[\nu^-, \nu^+]$. Moreover, since $d_S(\alpha, \partial Z) \leq 1$, the reverse triangle inequality implies that for any $t \in \mathrm{I}_Z$ and any $b \in [t, a]$ we have

$$d_S(a, b) \leq d_S(a, b) + d_S(b, t) \leq d_S(a, t) + \mathsf{B} \leq 5 + \mathsf{B}.$$

Thus it must be the case that $\mathrm{I}_Z \subset [x, y]$, for otherwise either $d_S(x, a)$ or $d_S(y, a)$ would be smaller than $5 + \mathsf{B}$, which is not the case. Therefore the projection to Z changes by at most M outside of $[x, y]$, and so we conclude that $d_Z(x, y) > M'$ as desired. $\qquad\square$

Going forward, for each $\epsilon \leq \epsilon_0$ we additionally assume that M_ϵ is chosen large enough to satisfy $\mathsf{M}_\epsilon \geq \mathsf{B}$ and so that Theorem 2.10 above applies with $M' = \mathsf{M}_\epsilon$.

3. The geometry of statistically thick geodesics

In the study of Teichmüller geometry, one finds that the thick part $\mathcal{T}_\epsilon$ behaves very much like a negatively-curved space. For example in Theorems 4.4 and 7.6 of [11] Kent and Leininger show that geodesic triangles contained entirely within $\mathcal{T}_\epsilon$ are δ–thin for some δ depending on ϵ, and that the projection of any geodesic $\gamma \subset \mathcal{T}_\epsilon$ to the curve complex $\mathcal{C}(S)$ is an honest *parametrized* quasi-geodesic which, in particular, must progress at a linear rate. These facts can also be deduced from the distance formula (2.5) together with properties of thin intervals I_V^ϵ (§2.6 above).

All of these negative-curvature properties are lost when geodesics are allowed to enter the thin part. For example, Minsky's product region theorem [19] shows that geodesic triangles in $\mathcal{T}(S) \setminus \mathcal{T}_\epsilon$ need not be δ–thin for any δ, and it is easy to construct arbitrarily long geodesics in $\mathcal{T}(S) \setminus \mathcal{T}_\epsilon$ that project to uniformly bounded diameter sets in $\mathcal{C}(S)$.

However, each of these extremes—living entirely in $\mathcal{T}_\epsilon$ or entirely in $T(S) \setminus \mathcal{T}_\epsilon$—is quite rare, as a typical Teichmüller geodesic will spend part of its time in $\mathcal{T}_\epsilon$ and part of its time in $\mathcal{T}(S) \setminus \mathcal{T}_\epsilon$. In this section we develop tools to study geodesics with exactly this behavior (later on, in §5 we will show that this behavior is in fact generic in a certain quantifiable sense).

Our techniques rely on controlling the fraction of time a Teichmüller geodesic spends in a given thick part $\mathcal{T}_\epsilon$. We call this quantity the *thick-stat*; for a nondegenerate geodesic segment $[x, y] \subset \mathcal{T}(S)$ it is denoted by

$$\mathsf{Thk}_\epsilon^\%[x, y] = \frac{\left| \{0 \leq s \leq d_\mathcal{T}(x, y) : y_s \in \mathcal{T}_\epsilon\} \right|}{d_\mathcal{T}(x, y)},$$

where y_s is the time–s point on the geodesic ray from x through y. Thus $[x, y] \subset \mathcal{T}_\epsilon$ is equivalent to $\mathsf{Thk}_\epsilon^\%[x, y] = 1$. In the following subsections, we will show that

negative-curvature properties similar to those mentioned above for thick geodesics also hold when $\mathsf{Thk}_\epsilon^\%$ is merely bounded away from zero.

3.1. **Progress in the curve complex.** The goal of this subsection is to prove Theorem 3.10, which says that geodesics that spend a definite fraction of their time in the thick part $\mathcal{T}_\epsilon$ must move at a definite linear rate in the curve complex.

The idea is that long subintervals contained in $\mathcal{T}_\epsilon$ contribute to progress in $\mathcal{C}(S)$; alternately, one could consider intervals in the complement of all the ϵ–thin intervals I_V^ϵ for proper subsurfaces $V \subsetneq S$. For this analysis, we would like to bound the number of connected components of $\bigcup_V \mathrm{I}_V^\epsilon$ in terms of $d_S(x,y)$. One bound is given by the number of nonempty thin intervals. While there may be arbitrarily many such I_V^ϵ, some of these will be redundant in the sense that $\mathrm{I}_V^\epsilon \subset \mathrm{I}_W^\epsilon$ for some other subsurface W.

Definition 3.1 (Thin-significance)**.** Fix $0 < \epsilon \le \epsilon_0$. A proper subsurface $V \subsetneq S$ is said to be ϵ–*thin-significant* (or simply *thin-significant*) for the geodesic segment $[x,y]$ if $d_{\mathcal{C}(V)}(x,y) \ge 3\mathsf{M}_\epsilon$ and for every other proper subsurface $Z \subsetneq S$ with $d_{\mathcal{C}(Z)}(x,y) \ge 3\mathsf{M}_\epsilon$ we have $\mathrm{I}_V^\epsilon \not\subset \mathrm{I}_Z^\epsilon$.

Remark. In this subsection we will focus on the curve complex distance $d_{\mathcal{C}(V)}$ for a subsurface V. Recall that this agrees with the usual projection distance d_V in the case that V is non-annular, but that $d_{\mathcal{C}(A)}$ and d_A differ for annuli. We will take care to handle exceptional annuli carefully.

Our first goal is to bound the number of thin-significant subsurfaces along an arbitrary geodesic. For this, we will use the work of Rafi–Schleimer [23] bounding the size of an antichain in the poset of subsurfaces of S.

Definition 3.2 (Antichain)**.** Given a subsurface $\Sigma \subset S$ a pair of points $x, y \in \mathcal{T}(S)$ and constants $T_1 \ge T_0 > 0$, a collection Ω of proper subsurfaces of Σ is an *antichain* for (Σ, x, y, T_0, T_1) if the following hold:

- if $Y, Y' \in \Omega$, then Y is not a proper subsurface of Y';
- if $Y \in \Omega$, then $d_{\mathcal{C}(Y)}(x,y) \ge T_0$; and
- if $Z \subsetneq \Sigma$ and $d_{\mathcal{C}(Z)}(x,y) \ge T_1$, then $Z \subset Y$ for some $Y \in \Omega$.

Lemma 3.3 (Antichain bound [23, Lem 7.1])**.** *For every $\Sigma \subset S$ and sufficiently large $T_1 \ge T_0 > 0$, there is a constant $A = A(\Sigma, T_0, T_1)$ so that if Ω is an antichain for (Σ, x, y, T_0, T_1) then*
$$|\Omega| \le A \cdot d_{\mathcal{C}(\Sigma)}(x,y).$$

We now prove a proposition showing that if there are a large enough number of thin-significant subsurfaces along a geodesic, then the image of the geodesic makes definite progress in the curve complex. The following notation will be used in the proof.

Definition 3.4. Consider a geodesic segment $[x,y] \subset \mathcal{T}(S)$ and a collection Ω of proper subsurfaces of S. We will consider three partial orders on the set Ω:

(1) $V \le_1 W \iff V \subset W$,
(2) $V \le_2 W \iff \mathrm{I}_V^\epsilon \subset \mathrm{I}_W^\epsilon$, and
(3) $V \le_3 W \iff V \subset W$ and $\mathrm{I}_V \subset \mathrm{I}_W$.

The subcollection of Ω consisting of maximal elements with respect to $\le_*$ will be denoted $(\Omega)_*$; notice that these sets are related by $(\Omega)_1 \subset (\Omega)_3 \supset (\Omega)_2$. Elements of $(\Omega)_1$ are said to be *topologically maximal* with respect to Ω.

Proposition 3.5 (Progress from thin-significant subsurfaces). *For any positive $\epsilon \leq \epsilon_0$ and any t_0, there is a constant N such that if $d_{\mathcal{C}(S)}(x, y) \leq t_0$, then the number of ϵ–thin-significant subsurfaces along $[x, y]$ is at most N.*

Proof. Let $\Omega = \{V \subsetneq S : d_{\mathcal{C}(V)}(x, y) \geq 3\mathsf{M}_\epsilon\}$ be the collection of proper subsurfaces which have a large projection. By definition, the set of ϵ–thin-significant subsurfaces is exactly given by $(\Omega)_2$. On the other hand, the subcollection $(\Omega)_1$ of topologically maximal subsurfaces clearly forms an antichain for $(S, x, y, 3\mathsf{M}_\epsilon, 3\mathsf{M}_\epsilon)$. By Lemma 3.3, we therefore have $|(\Omega)_1| \leq At_0$ for some constant A. We will extend this to a bound on the cardinality of the larger set $(\Omega)_3$; this will imply the proposition because $(\Omega)_2 \subset (\Omega)_3$.

Fix a proper subsurface $W \in \Omega$ and consider the set $\mathcal{U}_W = \{V \in (\Omega)_3 : V \subsetneq W\}$. We claim that $|\mathcal{U}_W|$ is bounded by a constant depending only on the complexity of W. By the above, this will suffice because each $V \in (\Omega)_3$ is either equal to or properly contained in some topologically maximal proper subsurface $W \in (\Omega)_1$.

First consider those $V \in \mathcal{U}_W$ for which $\mathrm{I}_V^\epsilon \cap \mathrm{I}_W^\epsilon \neq \emptyset$. The definition of $\leq_3$ implies that $\mathrm{I}_V^\epsilon \not\subset \mathrm{I}_W^\epsilon$; therefore I_V^ϵ must overlap with at least one endpoint of I_W^ϵ. If $\mathrm{I}_{V_1}^\epsilon$ and $\mathrm{I}_{V_2}^\epsilon$ both contain the initial endpoint of I_W^ϵ, then $\mathrm{I}_{V_1}^\epsilon \cap \mathrm{I}_{V_2}^\epsilon \neq \emptyset$ and so we cannot have $V_1 \pitchfork V_2$. Since there is a universal bound on the number of subsurfaces such that no two intersect transversely, this bounds the number of $V \in \mathcal{U}_W$ for which $\mathrm{I}_V^\epsilon \cap \mathrm{I}_W^\epsilon \neq \emptyset$.

It remains to bound the number of $V \in \mathcal{U}_W$ for which $\mathrm{I}_V^\epsilon \cap \mathrm{I}_W^\epsilon = \emptyset$; we will only focus on the case that I_V^ϵ occurs before I_W^ϵ when traveling from x to y. Suppose that $\mathrm{I}_W^\epsilon = [a, b] \subset [x, y]$ and consider the set

$$\Omega' = \{V \in \Omega : V \subsetneq W \text{ and } d_{\mathcal{C}(V)}(x, a) \geq 2\mathsf{M}_\epsilon\}$$
$$\cup \{A \text{ an annulus} : A \subsetneq W, \text{ and } d_{\mathcal{C}(A)}(x, a) \geq 4\mathsf{M}_\epsilon\}.$$

We claim that the subcollection $(\Omega')_1$ forms an antichain for $(W, x, a, 2\mathsf{M}_\epsilon, 4\mathsf{M}_\epsilon)$: The only issue is to check that every $Y \subsetneq W$ with $d_{\mathcal{C}(Y)}(x, a) \geq 4\mathsf{M}_\epsilon$ is contained in an element of $(\Omega')_1$. If Y satisfies the reverse triangle inequality then $d_{\mathcal{C}(Y)}(x, y) \geq d_{\mathcal{C}(Y)}(x, a) - \mathsf{B} \geq 3\mathsf{M}_\epsilon$ and therefore $Y \in \Omega'$. If Y does not satisfy the reverse triangle inequality it is an annulus and automatically $Y \in \Omega'$. Since $d_{\mathcal{C}(W)}(x, a) \leq \mathsf{M}_\epsilon$, Lemma 3.3 now gives a bound on $|(\Omega')_1|$.

Finally, notice that for each $V \in \mathcal{U}_W$ with I_V^ϵ occurring before I_W^ϵ along $[x, y]$, the triangle inequality gives $d_{\mathcal{C}(V)}(x, a) \geq d_{\mathcal{C}(V)}(x, y) - \mathsf{M}_\epsilon \geq 2\mathsf{M}_\epsilon$ and so ensures that $V \in \Omega'$. Therefore each such V is contained in some topologically maximal $Z \in \Omega'$; that is to say, each $V \in \mathcal{U}_W$ with I_V^ϵ occurring before I_W^ϵ along $[x, y]$ is contained in $\mathcal{U}_Z$ for some $Z \in (\Omega')_1$. The bound on $|\mathcal{U}_W|$ now follows by the bound on $|(\Omega')_1|$ and induction on the complexity of the subsurface W. $\square$

Proposition 3.5 gives control on the union of the ϵ–thin intervals I_V^ϵ for all subsurfaces V with $d_{\mathcal{C}(V)}(x, y)$ large. However, the potential failure of the reverse triangle inequality enables $[x, y]$ to contain many thin intervals that are not accounted for by Proposition 3.5. The following lemma allows us to control these as well by extending the thin intervals so as to have certain large projections that will help with our bookkeeping.

Lemma 3.6 (Extended thin interval). *For any $\epsilon \leq \epsilon_0$ and $t_0 > 0$, there exists $M' > 0$ with the following property. If a subsurface $W \subsetneq S$ satisfies $d_W(x', y') > M'$*

*for some pair of points $x', y' \in [x, y]$, then there is a connected interval $[a, b] \subset [x, y]$
containing I_W^ϵ and contained entirely in the ϵ–thin part of $\mathcal{T}(S)$ such that either*

- *$d_S(a, b) \geq t_0 + 3\mathsf{B}$ (and thus also $d_S(x, y) > t_0$ by Lemma 2.8), or*
- *$[a, b]$ nontrivially intersects the ϵ–thin interval $\mathrm{I}_V^\epsilon \subset [x, y]$ of some subsurface $V \subsetneq S$ satisfying $d_V(x, y) \geq 3\mathsf{M}_\epsilon$.*

We call such an interval $J = [a, b]$ an extended ϵ–thin interval for W.

Proof. Let K be the implied constant in the distance formula (2.5) corresponding
to the threshold $5\mathsf{M}_\epsilon$. For this K and the given t_0, set

$$M' = K^2(t_0 + 3\mathsf{B}) + K^2 + K + 2\mathsf{M}_\epsilon.$$

Now suppose that W determines a thin interval $\mathrm{I}_W^\epsilon \subset [x, y]$ of length $L \geq 0$ (where
we allow the possibility that $\mathrm{I}_W^\epsilon = \emptyset$ and $L = 0$). Since the projection to W can
change by at most M_ϵ outside of I_W^ϵ, the distance formula implies that

$$d_W(x', y') \leq 2\mathsf{M}_\epsilon + KL + K$$

for any $x', y' \in [x, y]$. Therefore, the hypothesis $d_W(x', y') \geq M'$ on W ensures that

$$L \geq (M' - 2\mathsf{M}_\epsilon - K)/K \geq K(t_0 + 3\mathsf{B}) + K.$$

In particular $\mathrm{I}_W^\epsilon \neq \emptyset$. Thus we have shown that there exists a nonempty interval
$J = [a, b] \subset [x, y]$ (for example, I_W^ϵ itself) that

(1) contains I_W^ϵ, and
(2) is the union of finitely many nonempty thin intervals $\mathrm{I}_V^\epsilon \subset [x, y]$ for proper
    subsurfaces $V \subsetneq S$ (and so is entirely contained in the ϵ–thin part of $\mathcal{T}(S)$).

We claim that for any such interval J that fails to satisfy the conclusion of the
lemma, there exists a *strictly* larger subinterval $J' \supsetneq J$ that again satisfies (1)–(2).
As there are only finitely many subintervals satisfying (2) (since at most finitely
many curves become shorter than ϵ along the compact segment $[x, y]$), repeated
applications of the claim will eventually produce the desired subinterval.

To prove the claim, we may suppose that $J = [a, b]$ satisfies (1)–(2) above and
fails to meet the conclusion of the lemma. In that case, there necessarily exists a
nonempty collection Ω of subsurfaces $V \subsetneq S$ for which $d_V(a, b) \geq 5\mathsf{M}_\epsilon$, for otherwise
the distance formula (applied with threshold $5\mathsf{M}_\epsilon$) would give

$$d_{\mathcal{T}}(a, b) \leq Kd_S(a, b) + K < K(t_0 + 3\mathsf{B}) + K \leq L,$$

contradicting the assumption $[a, b] \supset \mathrm{I}_W^\epsilon$ (recall that L is the length of I_W^ϵ).

Note that the property $d_V(a, b) \geq 5\mathsf{M}_\epsilon$ implies that the thin interval $\mathrm{I}_V^\epsilon \subset [x, y]$
of each $V \in \Omega$ nontrivially intersects $[a, b]$. If any such interval I_V^ϵ were completely
contained within $[a, b]$ then, since the projection to V can move at most M_ϵ outside
of I_V^ϵ, the triangle inequality would give $d_V(x, y) \geq 3\mathsf{M}_\epsilon$. As this is evidently not
the case (since J fails to satisfy the conclusion of the lemma), it must be that the
thin interval I_V^ϵ of each $V \in \Omega$ nontrivially intersects the complement of $[a, b]$ as
well. Therefore,

$$J' := [a, b] \cup \bigcup_{V \in \Omega} \mathrm{I}_V^\epsilon$$

is a connected subinterval of $[x, y]$ that properly contains J and again satisfies
(1)–(2). Thus the claim holds and so the lemma is verified. $\square$

By the distance formula (2.5), any long interval disjoint from all thin intervals along $[x, y]$ must travel a large distance in the curve complex $\mathcal{C}(S)$ of the whole surface. The following lemma says that each such subinterval contributes to the curve complex distance along the total geodesic.

Lemma 3.7 (Cumulative contribution of subintervals)**.** *There exist constants $0 < \rho_1 < 1$ and $D_1 > 0$ such that for all $d > D_1$, if $[x, y]$ is a Teichmüller geodesic that contains n subintervals $[x_i, y_i]$ with disjoint interiors whose endpoints satisfy $d_S(x_i, y_i) \geq d$, then*

$$d_S(x, y) \geq \rho_1 n d.$$

Proof. Applying the reverse triangle inequality (2.9) to the points x_i and y_i we have $d_S(x, x_i) + d_S(x_i, y_i) + d_S(y_i, y) \leq d_S(x, y) + 2\mathsf{B}$. By recursively applying this observation to $[x, x_i]$ and $[y_i, y]$ and then throwing out the complementary intervals, we find that

$$d_S(x, y) \geq \sum d_S(x_i, y_i) - 2n\mathsf{B} \geq nd - 2n\mathsf{B}.$$

Choose $D_1 > 4\mathsf{B}$ and $\rho_1 = 1/2$. Then for $d \geq D_1$ the quantity on the right side is at least $\rho_1 n d$. $\qquad\square$

We now fix once and for all a "definite progress" constant $\mathsf{D} > 0$, sufficiently large so that $\rho_1 \mathsf{D} > D_1$ (and thus $\mathsf{D} > D_1$ as well), and make the following definition.

Definition 3.8. For any $\epsilon \leq \epsilon_0$, set $t_0 = \rho_1 \mathsf{D}$ and let $M' = M'(\epsilon)$ be the corresponding constant provided by Lemma 3.6. Then define the *primary ϵ–thin portion* $\mathcal{W}_\epsilon$ of a geodesic segment $[x, y]$ to be the union of ϵ–thin intervals $\mathrm{I}_V^\epsilon \subset [x, y]$ for all proper subsurfaces with $d_{\mathcal{C}(V)}(x, y) \geq 3\mathsf{M}_\epsilon$ together with an extended ϵ–thin interval $J_W^\epsilon \subset [x, y]$ for any proper subsurface W satisfying $d_W(x', y') \geq M'$ for some pair of points $x', y' \in [x, y]$.

Lemma 3.9 (Primary thin portion)**.** *The primary ϵ–thin portion $\mathcal{W}_\epsilon$ is completely contained in the ϵ–thin part of $\mathcal{T}$. Furthermore, if $d_S(x, y) \leq \rho_1 \mathsf{D}$, then the number of connected components of $\mathcal{W}_\epsilon$ is bounded by a constant N' depending only on ϵ.*

Proof. The first assertion is immediate since $\mathcal{W}_\epsilon$ is a union of ϵ–thin intervals. For the second assertion, note that the union $\mathcal{W}'$ of ϵ–thin intervals I_V^ϵ for all proper subsurfaces with $d_{\mathcal{C}(V)}(x, y) \geq 3\mathsf{M}_\epsilon$ has a bounded number of connected components by Proposition 3.5 (since passing to thin-significant subsurfaces does not change the union). Consider now an extended ϵ–thin interval J_W^ϵ contributing to $\mathcal{W}_\epsilon$. Since $d_S(x, y) \leq \rho_1 \mathsf{D} = t_0$, Lemma 3.6 implies that J_W^ϵ intersects I_V^ϵ for some subsurface V with $d_V(x, y) \geq 3\mathsf{M}_\epsilon$. We claim that either $d_{\mathcal{C}(V)}(x, y) \geq 3\mathsf{M}_\epsilon$, so that $\mathrm{I}_V^\epsilon \subset \mathcal{W}'$ and thus $J_W^\epsilon \cap \mathcal{W}' \neq \emptyset$, or else V is an annulus with $\min(l_x(\partial V), l_y(\partial V)) < \epsilon_0$. Since there can be at most $6g - 6$ such annuli and $\mathcal{W}'$ has a bounded number of components, this will suffice.

If V is non-annular, then $d_{\mathcal{C}(V)} = d_V$ and the claim is immediate. Otherwise V is an annulus with $d_V(x, y) \geq 3\mathsf{M}_\epsilon$. Since M_ϵ may be assumed large enough to satisfy the universal condition $\mathsf{M}_\epsilon \geq 36 \log_+(1/\epsilon_0) + 6$, Lemma A.3 implies that either $\min(l_x(\partial V), l_y(\partial V)) < \epsilon_0$ or else $d_{\mathcal{C}(V)}(x, y) \geq e^{\mathsf{M}_\epsilon/2} \geq 3\mathsf{M}_\epsilon$, as claimed. $\quad\square$

For any interval $[a, b] \subset [x, y] \setminus \mathcal{W}_\epsilon$ in the complement of the primary ϵ–thin portion, the construction of $\mathcal{W}_\epsilon$ ensures that $d_W(a, b) \leq M'$ for all proper subsurfaces W of S. Applying the distance formula (2.5) with $M' = M'(\epsilon)$ as the threshold, we now see that $d_{\mathcal{T}}(a, b) \prec_\epsilon d_S(a, b)$ for any connected interval $[a, b] \subset [x, y] \setminus \mathcal{W}_\epsilon$.

This gives rise to a fixed value L_ϵ such that any interval $[a, b]$ of length at least L_ϵ that lies entirely in $[x, y] \setminus \mathcal{W}_\epsilon$ satisfies $d_S(a, b) \geq \mathsf{D}$. Thus according to Lemma 3.7, if I is any interval along a geodesic that contains a subinterval of length L_ϵ that is disjoint from $\mathcal{W}_\epsilon$, then the distance in the curve complex between the endpoints of I is at least $\rho_1 \mathsf{D}$. We now come to the main result of this subsection.

Theorem 3.10 (Definite progress). *For every $\epsilon > 0$ and $0 < \theta < 1$, there exists a constant $R_1 > 0$ such that*

$$d_S(x, y) \succ_{\epsilon, \theta} d_{\mathcal{T}}(x, y)$$

for every Teichmüller geodesics $[x, y]$ satisfying $d_{\mathcal{T}}(x, y) \geq R_1$ and $\mathsf{Thk}_\epsilon^\%[x, y] \geq \theta$.

Proof. Since shrinking ϵ preserves the hypothesis $\mathsf{Thk}_\epsilon^\%[x, y] \geq \theta$, we may assume $\epsilon \leq \epsilon_0$. Let $N' = N'(\epsilon)$ denote the constant obtained from Lemma 3.9. Choose n so that $n\theta > 1$ and make the following definitions:

$$\theta' = \frac{n\theta - 1}{n - 1}, \quad T_0 \geq \frac{\mathsf{L}_\epsilon(N' + 1)}{\theta'}, \quad R_1 = 2T_0, \quad \rho = \frac{\rho_1^2 \mathsf{D}}{2nT_0}.$$

Let $[x, y]$ be a Teichmüller geodesic of length $r \geq R_1$ satisfying $\mathsf{Thk}_\epsilon^\%[x, y] \geq \theta$. Set $m = \lfloor r/T_0 \rfloor$ and divide $[x, y]$ into m subsegments of length $r/m \geq T_0$. Let us say that a subsegment $[a, b] \subset [x, y]$ is *stalled* if $d_S(a, b) < \rho_1 \mathsf{D}$ and *progressing* if $d_S(a, b) \geq \rho_1 \mathsf{D}$. Suppose that m_1 of the subsegments are stalled, and thus $m_2 = m - m_1$ are progressing.

Given a stalled segment $[a, b]$, we decompose it into its primary ϵ–thin portion $\mathcal{W}_\epsilon$ and note that, since it is stalled, Lemma 3.9 ensures $\mathcal{W}_\epsilon$ has at most N' connected components. Therefore we conclude that $\mathcal{W}_\epsilon$ has at most $N' + 1$ complementary subintervals in $[a, b]$. Furthermore, each complementary subinterval has length at most L_ϵ, for otherwise we would have $d_S(a, b) \geq \rho_1 \mathsf{D}$ by the paragraph preceeding Theorem 3.10. Since $\mathcal{W}_\epsilon$ is contained in the ϵ–thin part, we see that the total amount of time that this interval $[a, b]$ spends in the thick part is at most

$$(N' + 1)\mathsf{L}_\epsilon \leq \theta' T_0 \leq \theta' r/m.$$

Therefore the total amount of time that the full interval $[x, y]$ spends in the thick part is at most

$$\left(\theta' \frac{r}{m}\right) m_1 + \left(\frac{r}{m}\right) m_2 = \frac{r}{m}(\theta' m_1 + m_2).$$

We claim that $m_2 \geq m/n$. If this were not the case, then we necessarily have $m_1 > (n - 1)m/n$. Since $\theta' < 1$, it follows that

$$\theta' \cdot m_1 + 1 \cdot m_2 < \theta' \cdot m \frac{n - 1}{n} + 1 \cdot m \frac{1}{n},$$

where the inequality is valid by the elementary fact that for any constants $a, b, c, d, \alpha, \beta$ such that $a + b = c + d$ and $0 < \alpha < \beta$ we have

$$(3.11) \qquad\qquad \alpha \cdot a + \beta \cdot b < \alpha \cdot c + \beta \cdot d \iff a > c.$$

But then the amount of time that $[x, y]$ is thick is less than

$$\frac{r}{m}\left(\theta' m \frac{n - 1}{n} + m \frac{1}{n}\right) = r\left(\frac{n\theta - 1}{n - 1} \cdot \frac{n - 1}{n} + \frac{1}{n}\right) = r\theta,$$

which contradicts the assumption on $[x, y]$. Therefore $m_2 \geq m/n$, as claimed.

On each of the m_2 progressing intervals, the curve complex distance between end-points is at least $\rho_1 \mathsf{D}$. Therefore, cumulative contribution of subintervals (Lemma 3.7) implies that

$$d_S(x,y) \geq \rho_1 m_2(\rho_1 \mathsf{D}) \geq \rho_1^2 \mathsf{D} \frac{m}{n} \geq \frac{\rho_1^2 \mathsf{D}}{n} \left(\frac{r}{T_0} - 1 \right) \geq \frac{\rho_1^2 \mathsf{D}}{2nT_0} r = \rho r. \qquad \square$$

Remark 3.12. After developing our proof of Theorem 3.10 we learned of an independent yet closely related result of Hamenstädt's, namely Proposition 2.1 of [9], which under the same hypotheses provides a lower bound on $d_S(x,y)$ that is constant rather than linear in $d_\mathcal{T}(x,y)$. In fact, the linear bound in Theorem 3.10 may be deduced from Hamenstädt's result by breaking $[x,y]$ into subintervals, applying [9, Proposition 2.1] to those with large thick-stat, and adding the resulting contributions using Lemma 3.7, much as we have done above. With this approach [9, Proposition 2.1] would effectively replace the use of Proposition 3.5 and Lemma 3.6 in our argument. However, we have decided to retain our original argument using Proposition 3.5 and Lemma 3.6 as we believe these to be of independent interest.

3.2. **A statistical thin triangles statement.** In this subsection we prove Theorem A and obtain thinness results for geodesic triangles whose sides satisfy various thick-stat conditions. Recall that given $\epsilon > 0$ there is a $\delta > 0$ such that every geodesic triangle whose sides live entirely in $\mathcal{T}_\epsilon$ is δ–thin. This fact can be deduced from the following theorem of Rafi, which gives specific information under much more general conditions.

Theorem 3.13 (Rafi [22, Theorem 8.1]). *For every $\epsilon > 0$ there exist constants C_1, L_1 such that if $I \subset [x,y] \subset \mathcal{T}(S)$ is a geodesic subinterval of length at least L_1 lying entirely in the ϵ–thick part, then for all $z \in \mathcal{T}(S)$, we have*

$$I \cap \mathrm{Nbhd}_{C_1}([x,z] \cup [y,z]) \neq \emptyset.$$

We weaken the hypothesis to only require a definite thick-stat.

Theorem A. *For any $\epsilon > 0$ and $0 < \theta \leq 1$, there exist constants C, L such that if $I \subset [x,y] \subset \mathcal{T}(S)$ is a geodesic subinterval of length at least L and at least proportion θ of I is ϵ–thick, then for all $z \in \mathcal{T}(S)$, we have*

$$I \cap \mathrm{Nbhd}_C([x,z] \cup [y,z]) \neq \emptyset.$$

Before proving this result, we discuss two consequences. Firstly we observe that there is not merely one point in the subinterval I which is close to $[x,z] \cup [y,z]$, but in fact this conclusion holds for a large fraction of the interval I.

Proposition 3.14. *For any $\epsilon > 0$ and $0 < \theta' < \theta \leq 1$, there are constants L', C' so that if a side $[x,y]$ of a geodesic triangle $\triangle(x,y,z) \subset \mathcal{T}(S)$ contains a subinterval $I \subset [x,y]$ of length at least L' with $\mathsf{Thk}_\epsilon^\%(I) \geq \theta$, then at least proportion θ' of I is within distance C' of $[x,z] \cup [y,z]$. That is, if* length *denotes Lebesgue measure along a geodesic segment,*

$$\mathrm{length}\left(\{ I \cap \mathrm{Nbhd}_{C'}([x,z] \cup [y,z]) \} \right) \geq \theta' \cdot \mathrm{length}(I).$$

Proof. Let $\rho = \frac{\theta - \theta'}{1 - \theta'}$. Apply Theorem A to $\mathcal{T}_\epsilon$ with the fraction ρ and let $L' = L$ and C be the corresponding constants. Given a subinterval $I \subset [x,y]$ satisfying the hypotheses of the theorem, divide I into $n = \lfloor \mathrm{length}(I)/L \rfloor$ subintervals of equal length (the length will be between L and $2L$). Let a denote the fraction of

these subintervals that have $\mathsf{Thk}_\epsilon^{\%} \geq \rho$ (so $a = \frac{k}{n}$ for some $k \in \{0, \ldots, n\}$). Each of these na subintervals can spend at most all of their time in $\mathcal{T}_\epsilon$ and the other $n(1-a)$ subintervals spend less than proportion ρ of their time in $\mathcal{T}_\epsilon$. Therefore the maximum amount of time the whole interval I can spend in $\mathcal{T}_\epsilon$ is less than

$$1 \cdot na \cdot \frac{\text{length}(I)}{n} + \rho \cdot n(1-a) \cdot \frac{\text{length}(I)}{n} = \text{length}(I)(a + \rho - \rho a).$$

Since we have $\mathsf{Thk}_\epsilon^{\%}(I) \geq \theta$ by hypotheses, this implies $\theta < a + \rho - a\rho$. That is,

$$a > \frac{\theta - \rho}{1 - \rho} = \theta'.$$

Now, Theorem A implies that each subinterval with $\mathsf{Thk}_\epsilon^{\%} \geq \rho$ contains a point within distance C of $[x, z] \cup [y, z]$. Therefore, each of the na subintervals satisfying $\mathsf{Thk}_\epsilon^{\%} \geq \rho$ is contained entirely within the $C' = C + 2L$ neighborhood of $[x, z] \cup [y, z]$. As the union of these na subintervals comprise proportion $a > \theta'$ of the interval I, the statement follows. $\qquad\square$

From this we obtain the following immediate corollary.

Corollary 3.15 (Statistically thin triangles). *For all $\epsilon > 0$ and $0 < \theta' < \theta \leq 1$ there exists a constant δ with the following property. For any geodesic triangle in $\mathcal{T}(S)$ whose three sides have $\mathsf{Thk}_\epsilon^{\%} \geq \theta$, at least proportion θ' of each side of the triangle is contained within δ of the union of the other two sides.*

We now give the proof of Theorem A.

Proof of Theorem A. We will find L so that the conclusion of the theorem applies to any subinterval I with $L \leq \text{length}(I) \leq 2L$ and $\mathsf{Thk}_\epsilon^{\%}(I) \geq \theta$. This will suffice because any long interval with $\mathsf{Thk}_\epsilon^{\%} \geq \theta$ can be partitioned into subintervals satisfying this length condition, one of which must have $\mathsf{Thk}_\epsilon^{\%} \geq \theta$.

Recall that the (coarsely defined) projection $\pi_S \colon \mathcal{T}(S) \to \mathcal{C}(S)$ sends a point $w \in \mathcal{T}(S)$ to the set of simple closed curve in the Bers marking μ_w (which is a set of diameter 2 in the curve complex). The work of Masur–Minsky [17] shows that there are universal constants K, C so that Teichmüller geodesics project to (unparametrized) (K, C)–quasi-geodesics under π_S. By the hyperbolicity of $\mathcal{C}(S)$ (see §2.2), each quasi-geodesic fellow travels any geodesic with the same endpoints, and so there is a constant $\tau > 0$ so that every (K, C)–quasi-triangle in $\mathcal{C}(S)$ is τ–thin. We may furthermore assume that $\tau \geq \mathsf{B}$.

By Theorem 3.10 the interval I moves a definite amount in $\mathcal{C}(S)$; that is, by making L large, we can arrange for I to project to an arbitrarily long subsegment of the (unparametrized) quasi-geodesic $\pi_S([x, y])$. In particular, by choosing L sufficiently large, we can ensure that there is a point $w \in I$ so that either

$$d_S(\mu_w, \pi_S([y, z])) \geq 2\tau + 6 \qquad \text{or} \qquad d_S(\mu_w, \pi_S([x, z])) \geq 2\tau + 6.$$

By choosing such w with μ_w near the center of $\pi_S(I)$, we can moreover ensure that

$$d_S(w, t) \geq 2\tau + 6$$

for all points $t \in [x, y]$ outside of I, and in particular that $d_S(w, x), d_S(w, y) \geq 2\tau + 6$.

Assuming without loss of generality that

$$d_S(\mu_w, \pi_S([y, z])) \geq 2\tau + 6 > \tau,$$

hyperbolicity implies that there is a point $u \in [x, z]$ so that $d_S(w, u) \leq \tau$. We will show that $d_V(w, u) \prec_{\epsilon, \theta} 1$ for all proper subsurfaces $V \subset S$. The result will then follow from the distance formula (2.5). We first establish the following

Claim 3.16. *There is a constant M_0 (depending only on ϵ and θ) so that for any proper subsurface $V \subset S$ satisfying $d_S(\partial V, \mu_w) \leq 2\tau + 3$ we have*

$$d_V(x, y), d_V(y, z), d_V(x, z) \leq M_0.$$

To see this, first observe that for any such V the triangle inequality implies $d_S(\partial V, \pi_S([y, z])) \geq 3$. Therefore V does not become thin along $[y, z]$ and so we may conclude $d_V(y, z) \leq \mathsf{M}$. If V does not become thin along $[x, y]$, then we have the same bound on $d_V(x, y)$. However, V may become thin along $[x, y]$ in which case there is a point $t \in [x, y]$ at which the length of ∂V is smaller than ϵ_0. Therefore μ_t contains ∂V, which implies

$$d_S(w, t) \leq 2\tau + 5 < 2\tau + 6$$

and consequently that $t \in I \subset [x, y]$. Thus the entire thin interval I_V for V is contained within I and in particular has length at most $\mathrm{length}(I) \leq 2L$. Since the projection π_V to $\mathcal{C}(V)$ is a Lipschitz map and, up to an additive error, the projection of $[x, y]$ to $\mathcal{C}(V)$ can only change in the thin interval I_V (see §2.6), we conclude that $d_V(x, y)$ is bounded in terms of L (and L depends only on ϵ and θ). Finally, the triangle inequality and the above bounds on $d_V(y, z)$ and $d_V(x, y)$ together provide a uniform bound on $d_V(x, z)$. This completes the proof of Claim 3.16.

We now show that $d_V(w, u)$ is uniformly bounded for all proper subsurfaces. Consider any V with $d_V(w, u) \geq \mathsf{M}$. Then V becomes thin along $[w, u]$ and so ∂V lies within distance $\tau + 2$ of the $\mathcal{C}(S)$–geodesic from μ_w to μ_u. In particular

$$d_V(\partial V, \mu_w) \leq \tau + 2 + d_S(w, u) \leq 2\tau + 2$$

and so Claim 3.16 implies that $d_V(x, y)$ and $d_V(x, z)$ are at most M_0.

If V is a non-annular surface, then the reverse triangle inequality, applied to $[x, y]$ and $[x, z]$, yields bounds on $d_V(x, w)$ and $d_V(x, u)$ so that we may bound $d_V(w, u)$ by the triangle inequality.

If V is an annulus whose core curve $\alpha = \partial V$ satisfies $l_w(\alpha) > \epsilon_0$, then w cannot be contained in the (possibly empty) thin interval $\mathrm{I}_V \subset [x, y]$. Thus at least one of the intervals $[x, w]$ or $[w, y]$ is disjoint from I_V and consequently has d_V–projection at most M. Thus we may conclude $d_V(x, w) \leq \mathsf{M} + M_0$ by the triangle inequality. If $l_w(\alpha) \leq \epsilon_0$, then μ_w necessarily contains α and we instead appeal to Lemma 2.10. According to that theorem applied to $2M_0$, there is a constant ϵ' such that if $l_w(\alpha) \leq \epsilon'$, then there exists a subsurface Z with $d_S(\partial Z, \alpha) \leq 1$ so that $d_Z(x, y) \geq 2M_0$. Since this contradicts Claim 3.16, we must in fact have $l_w(\alpha) > \epsilon'$. As above, it follows that w cannot be contained in the ϵ'–thin interval $\mathrm{I}_V^{\epsilon'} \subset [x, y]$ and thus that $d_V(x, w) \leq M_0 + \mathsf{M}_{\epsilon'}$ by the triangle inequality and the theory of thin intervals.

Since $d_S(x, w), d_S(w, z) \geq 2\tau + 6$ and $d_S(w, u) \leq \tau$ by assumption, we also have $d_S(x, u), d_S(u, z) \geq \tau + 6 \geq \mathsf{B} + 6$ by the triangle inequality. Therefore we may apply the same argument, using Lemma 2.10 as needed, to obtain a bound on $d_V(x, u)$ as well. The triangle inequality then gives the desired bound on $d_V(w, u)$. $\square$

4. Comparing measures

To address genericity and averaging questions, one of course needs to consider a measure. In the present context of metric geometry, it is perhaps most natural to consider Hausdorff measure of the appropriate dimension.

Definition 4.1 (Hausdorff measure). The n–dimensional Hausdorff measure on a metric space will be denoted by η. It is defined by

$$\eta(E) := \lim_{\delta \to 0} \left[\inf \sum \operatorname{diam}(U_i)^n \right],$$

where the infimum is over countable covers $\{U_i\}$ of E with $\operatorname{diam} U_i < \delta \ \forall i$.

For the Teichmüller metric, there is a nontrivial h–dimensional Hausdorff measure (recalling that $h = 6g - 6$). As we shall see, in order to understand average distances with respect to this measure, it will be necessary to compare with other measures, defined below, which are also natural to consider in their own right.

4.1. **Measures on Finsler manifolds.** The Teichmüller space carries several natural volume forms coming from its structure as a Finsler manifold. Let us discuss these general constructions first before returning to the case of $M = \mathcal{T}(S)$. The treatment closely follows the survey by Álvarez and Thompson [1].

Recall that a Finsler metric on an n–dimensional Finsler manifold M is a continuous function $F \colon T(M) \to \mathbb{R}$ that restricts to a norm on each tangent space $T_x(M)$. There is a dual norm on each cotangent space $T_x^*(M)$. For a point $x \in M$, let $B_x \subset T_x(M)$ and $B_x^* \subset T_x^*(M)$ denote the unit balls for these two norms. A local coordinate system $(x_1, \ldots, x_n)$ on M induces a pair of isomorphisms

$$(4.2) \qquad \phi \colon T_x(M) \to \mathbb{R}^n \qquad \text{and} \qquad \psi \colon T_x^*(M) \to \mathbb{R}^n$$

defined by writing vectors and covectors with respect to the dual bases $\{\partial_{x_1}, \ldots, \partial_{x_n}\}$ and $\{dx_1, \ldots, dx_n\}$. By definition of the dual norm, the pairing $T_x(M) \times T_x^*(M) \to \mathbb{R}$ is sent to the standard inner product on $\mathbb{R}^n$ under these isomorphisms. In the local coordinate chart we may now define two functions

$$f(x) = \frac{\varepsilon_n}{\lambda\left(\phi(B_x)\right)} \qquad \text{and} \qquad g(x) = \frac{\lambda\left(\psi(B_x^*)\right)}{\varepsilon_n},$$

where λ is Lebesgue measure and $\varepsilon_n := \lambda(\mathrm{Ball}^n)$ is the Lebesgue measure of the standard unit ball in $\mathbb{R}^n$. While these functions clearly depend on the choice of coordinates $(x_1, \ldots, x_n)$, one may easily check that the n–forms

$$f(x)\, dx_1 \wedge \cdots \wedge dx_n \qquad \text{and} \qquad g(x)\, dx_1 \wedge \cdots \wedge dx_n$$

are independent of the coordinate system and therefore define global volume forms on M. The former is called the *Busemann volume* on the Finsler manifold and the latter is the *Holmes–Thompson volume*; see [1] for more details. These both define measures on M.

A third measure to consider is the one induced by the canonical symplectic form ω on the cotangent bundle, defined as follows. Consider local coordinates $(x_1, \ldots, x_n)$ defined in a neighborhood $U \subset M$. The 1–forms $dx_1, \ldots, dx_n$ then give a trivialization of $T^*(M)$ over U, and we have a local coordinate system on $T^*(M)$ given by

$$(4.3) \qquad (x_1, y_1, \ldots, x_n, y_n) \mapsto \left((x_1, \ldots, x_n), \sum_{i=1}^n y_i\, dx_i \right).$$

In these coordinates the canonical symplectic form may be written simply as $\omega = \sum dx_i \wedge dy_i$. Taking exterior powers then yields a volume form $\mu_{\text{sp}} = \omega^n/n!$ on $T^*(M)$. By restricting to the unit disk bundle $T^{*,\leq 1}(M)$ and pushing forward by the projection $\pi\colon T^*(M) \to M$, we obtain a *symplectic measure* $\mathbf{n}$ on M.

Finally, a Finsler metric on a smooth manifold M^n induces a path metric d in the usual way, and this in turn gives rise to a Hausdorff measure in any dimension.

Recall that a centrally symmetric convex body $\Omega \subset \mathbb{R}^n$ determines a *polar body* $\Omega^\circ \subset (\mathbb{R}^n)^* = \mathbb{R}^n$ via

$$\Omega^\circ := \{\xi \in \mathbb{R}^n \mid \xi \cdot v \leq 1 \ \forall v \in \Omega\}.$$

The *Mahler volume* of Ω is then defined to be the product $M(\Omega) := \lambda(\Omega) \cdot \lambda(\Omega^\circ)$ of the Lebesgue volumes of Ω and Ω°. For any centrally symmetric convex body Ω, it is known that

$$(4.4) \qquad \frac{\varepsilon_n^2}{n^{n/2}} \leq M(\Omega) \leq \varepsilon_n^2 = M(\text{Ball}^n).$$

The first inequality was established by John [10], and the latter, which gives an equality if and only if the norm is Euclidean, is known as the Blaschke–Santaló inequality [4].

Theorem 4.5 (Assembling facts on Finsler measures)**.** *Suppose that M^n is a continuous Finsler manifold. Then*

- *the Busemann measure μ_{B} and the n–dimensional Hausdorff measure η are equal;*
- *the Holmes–Thompson measure μ_{HT} and the symplectic measure $\mathbf{n}$ are scalar multiples: $\mu_{\text{HT}} = \frac{1}{\varepsilon_n}\mathbf{n}$;*
- *$\mu_{\text{HT}} \leq \mu_{\text{B}} \leq (n^{n/2})\mu_{\text{HT}}$, with equality of measures if and only if the metric is Riemannian.*

Note that it is still possible for μ_{HT} and μ_{B} to be scalar multiples of each other in the non-Riemannian case, for instance on a vector space with a Finsler norm.

Proof. The first statement was originally shown by Busemann in the 1940s in [5] and is stated in modern language in [1, Thm 3.23].

The second statement is straightforward and we include a proof for completeness. Working in the local coordinates and applying the Fubini theorem, we see that the Holmes–Thompson volume of a subset $E \subset M$ is given by:

$$\int_E g(x)\, dx_1 \wedge \cdots \wedge dx_n = \int_E \left(\int_{\psi(B_x^*)} \frac{1}{\varepsilon_n} d\lambda \right) dx_1 \wedge \cdots \wedge dx_n$$

$$= \frac{1}{\varepsilon_n} \int_{\pi^{-1}(E) \cap T^{*,\leq 1}(M)} dy_1 \wedge \cdots \wedge dy_n \wedge dx_1 \wedge \cdots \wedge dx_n$$

$$= \frac{1}{\varepsilon_n} \mathbf{n}(E).$$

For the third statement, recall that the measures are defined by

$$\mu_{\text{B}}(E) = \int_E f(x)\, dx_1 \wedge \cdots \wedge dx_n \quad \text{and} \quad \mu_{\text{HT}}(E) = \int_E g(x)\, dx_1 \wedge \cdots \wedge dx_n.$$

For each $x \in M$, the unit ball $B_x \subset T_x(M)$ is sent to a centrally symmetric convex body $\phi(B_x) \subset \mathbb{R}^n$ under the isomorphism ϕ defined in (4.2). The polar

body is exactly given by $\phi(B_x)^\circ = \psi(B_x^*)$. Therefore, the Mahler volume of $\phi(B_x)$ is

$$M(\phi(B_x)) = \lambda(\phi(B_x)) \cdot \lambda(\psi(B_x^*)) = \varepsilon_n^2 \frac{g(x)}{f(x)}.$$

Combining with (4.4) now implies that $n^{-n/2} f(x) \leq g(x) \leq f(x)$ for all $x \in M$. We conclude that $\mu_{\mathrm{HT}}(E) \leq \mu_{\mathrm{B}}(E) \leq n^{n/2} \mu_{\mathrm{HT}}(E)$ for all $E \subset M$. Finally, since Blaschke–Santaló can only give equality for a Euclidean norm, it follows that μ_{B} and μ_{HT} can only be equal for a Riemannian metric. $\qquad\square$

4.2. **Measures coming from quadratic differentials.** Recall that quadratic differential space $\mathcal{Q}(S)$ is naturally identified with the cotangent bundle $T^*(\mathcal{T}(S))$ of Teichmüller space, and that each quadratic differential $q \in \mathcal{Q}(S)$ has a norm $\|q\|$ given by the area of the flat structure on S induced by q. The unit disk bundle for this norm will be denoted by

$$\mathcal{Q}^{\leq 1}(S) = \{q \in \mathcal{Q}(S) : \|q\| \leq 1\}.$$

Using this disk bundle, the natural symplectic measure μ_{sp} on $\mathcal{Q}(S)$ descends to a measure $\mathbf{n}$ on $\mathcal{T}(S)$ exactly as above. We note that ω and therefore μ_{sp} and $\mathbf{n}$ are invariant under the action of the mapping class group.

The space $\mathcal{Q}(S)$ also carries a natural $\mathrm{Mod}(S)$–invariant measure μ_{hol} that is defined in terms of holonomy coordinates and which we will refer to as *holonomy measure*; it is also sometimes called Masur–Veech measure in the literature (see [14] for details). This measure has been studied extensively, for instance to establish ergodicity results for the geodesic flow. The measure μ_{hol} is also related to the "Thurston measure" μ_{TH} on the space of measured foliations $\mathcal{MF}$ induced by the piecewise-linear structure of $\mathcal{MF}$ [8]. Indeed, as seen in [14], μ_{hol} is equal to the pullback of $\mu_{\mathrm{TH}} \times \mu_{\mathrm{TH}}$ under the $\mathrm{Mod}(S)$–invariant map $\mathcal{Q}(S) \to \mathcal{MF} \times \mathcal{MF}$ that sends a quadratic differential to its vertical and horizontal foliations.

Just as μ_{sp} induces $\mathbf{n}$, the holonomy measure μ_{hol} descends to a measure $\mathbf{m}$ on $\mathcal{T}(S)$. Explicitly, the $\mathbf{m}$–measure of a set $E \subset \mathcal{T}(S)$ is given by

$$\mathbf{m}(E) := \mu_{\mathrm{hol}}\left(\pi^{-1}(E) \cap \mathcal{Q}^{\leq 1}(S)\right).$$

This measure $\mathbf{m}$ has been studied previously in [3] and [7].

Proposition 4.6. [16, p.3746] *There is a scalar $k > 0$ such that $\mu_{\mathrm{sp}} = k \cdot \mu_{\mathrm{hol}}$.*

We recall the outline of the argument here. In [16], it was shown that the Teichmüller geodesic flow on $\mathcal{Q}(S)$ is a Hamiltonian flow for the function

$$H(q) = \frac{\|q\|^2}{2}.$$

As such, the Teichmüller flow preserves the symplectic form ω and the corresponding measure μ_{sp}. The measures μ_{sp} and μ_{hol} both descend to the quotient space $\mathcal{Q}(S)/\mathrm{Mod}(S)$; furthermore, the latter defines an ergodic measure for the Teichmüller flow on $\mathcal{Q}(S)/\mathrm{Mod}(S)$ [14]. Since μ_{sp} is absolutely continuous with respect to μ_{hol}, the proposition follows.

We therefore also have $\mathbf{n} = k\mathbf{m}$, and combining Proposition 4.6 with Theorem 4.5 we get:

Corollary 4.7. *There are scalars $k_2 > k_1 > 0$ such that*

$$k_1 \mathbf{m} \leq \eta \leq k_2 \mathbf{m}.$$

4.3. **Visual measures.** The unit sphere subbundle of $\mathcal{Q}(S)$ will be denoted by

$$\mathcal{Q}^1(S) = \{q \in \mathcal{Q}(S) \colon \|q\| = 1\}.$$

For each $x \in \mathcal{T}(S)$, the fiber $\mathcal{Q}^1(x)$ is identified with the "space of directions" at x, and the Teichmüller geodesic flow $\varphi_t \colon \mathcal{Q}(S) \to \mathcal{Q}(S)$ gives rise to a homeomorphism

$$\Psi_x \colon \quad \begin{aligned} \mathcal{Q}^1(x) \times (0, \infty) &\quad \to \quad \mathcal{T}(S) \setminus \{x\} \\ (q, r) &\quad \mapsto \quad \pi(\varphi_r(q)) \, , \end{aligned}$$

which serves as "polar coordinates" centered at x. Furthermore, this conjugates φ_t to a *radial flow* based at x given by

$$\hat{\varphi}_t(\pi(\varphi_r(q))) := \pi(\varphi_{r+t}(q)).$$

We will consider measures on $\mathcal{T}(S)$ that are compatible with these polar coordinates and with the radial flow.

Definition 4.8 (Visual measure)**.** Given any measure κ_x on the unit sphere $\mathcal{Q}^1(x) \cong \mathbb{S}^{h-1}$, we define the corresponding *visual measures* on $\mathcal{S}_r(x)$ and $\mathcal{T}(S)$ as follows. Firstly, the visual measure $\mathrm{Vis}_r(\kappa_x)$ on the sphere $\mathcal{S}_r(x)$ of radius r is just the push-forward of $e^{hr}\kappa_x$ under the homeomorphism $\mathcal{Q}^1(x) \times \{r\} \cong \mathcal{S}_r(x)$. Integrating these over $(0, \infty)$ then gives a visual measure on $\mathcal{T}(S)$ defined by

$$\mathrm{Vis}(\kappa_x)(E) := \int_{(q,r) \in E \subset \mathcal{Q}^1(S) \times (0,\infty)} e^{hr} d\kappa_x(q) d\lambda(r).$$

Said differently, $\mathrm{Vis}(\kappa_x)$ is equal to the push-forward of $\kappa_x \times \lambda_0$ under the homeomorphism Ψ_x, where λ_0 is the weighted Lebesgue measure on $(0, \infty)$ given by $\lambda_0([a, b]) = \int_a^b e^{hr} d\lambda(r) = (e^{hb} - e^{ha})/h$. (We have scaled things in this way so that the visual measure of the ball of radius R grows like e^{hR}.)

The essential feature of visual measures is that they enjoy the following "normalized invariance" under the radial flow: For any $t \geq 0$ and measurable $E \subset \mathcal{S}_r(x)$ we have

$$\frac{\mathrm{Vis}_{r+t}(\kappa_x)(\hat{\varphi}_t(E))}{\mathrm{Vis}_{r+t}(\kappa_x)(\mathcal{S}_{r+t}(x))} = \frac{\mathrm{Vis}_r(\kappa_x)(E)}{\mathrm{Vis}_r(\kappa_x)(\mathcal{S}_r(x))}.$$

The same invariance holds for $\mathrm{Vis}(\kappa_x)$ when we normalize with respect to annular shells $\mathcal{B}_b(x) \setminus \mathcal{B}_a(x)$ instead of spheres.

There are two visual measures that specifically interest us. Firstly, the normed vector space $\mathcal{Q}(x)$ carries a unique translation-invariant measure ν_x normalized so that $\nu_x(B_x^*) = 1$; recall that the unit ball B_x^* is just the intersection $\mathcal{Q}^{\leq 1}(S) \cap \mathcal{Q}(x)$. This induces a measure (also denoted ν_x) on the unit sphere $\mathcal{Q}^1(x)$ via the usual method of coning off: $\nu_x(E) := \nu_x([0, 1] \times E)$ for $E \subset \mathcal{Q}^1(x)$.

Secondly, since $\mathcal{Q}(S)$ has the structure of a fiber bundle over $\mathcal{T}(S)$, we can define a conditional measure s_x on $\mathcal{Q}(x)$ by disintegration from μ_{hol}. More precisely, s_x is the unique measure on $\mathcal{Q}(x)$ such that the μ_{hol}–measure of $E \subset \mathcal{Q}(S)$ is given by

$$\mu_{\mathrm{hol}}(E) = \int_{\mathcal{T}(S)} s_x(E \cap \mathcal{Q}(x)) \, d\mathbf{m}(x).$$

Via the process of coning off, we again think of s_x as a measure on $\mathcal{Q}^1(x)$.

The space $\mathcal{Q}(S)$ of quadratic differentials is a complex vector bundle; as such, there is a natural circle action $S^1 \curvearrowright \mathcal{Q}(S)$ that preserves each fiber $\mathcal{Q}(x)$ and unit sphere $\mathcal{Q}^1(x)$. We say that a visual measure $\mathrm{Vis}(\kappa_x)$ is *rotation-invariant* if the corresponding measure κ_x on $\mathcal{Q}^1(x)$ is invariant under this action of S^1. The

visual measure $\mathrm{Vis}(\nu_x)$ is rotation-invariant because S^1 preserves the unit ball B_x. Similarly, $\mathrm{Vis}(s_x)$ is rotation-invariant because S^1 preserves μ_{hol}.

4.4. **Summary.** The measures on $\mathcal{T}(S)$ considered above are $\mathbf{n}$ and $\mathbf{m}$ (induced by the symplectic and holonomy measures on $\mathcal{Q}(S)$, respectively, via the covering map), Hausdorff measure η, the visual measures $\mathrm{Vis}(\kappa_x)$ created by radially flowing measures on the sphere of directions $\mathcal{Q}^1(x)$, and the measures μ_{B} and μ_{HT} coming from the Finsler structure.

We found that $\mathbf{n}$, $\mathbf{m}$, and μ_{HT} are scalar multiples of each other, Hausdorff measure and Busemann measure coincide, and all five of these are mutually comparable in the sense of being bounded above and below by scalar multiples of each other. In the following section we will establish results about the structure of generic geodesic rays with respect to these measures and the visual measures.

5. Thickness statistics for geodesic rays

In §3 we studied the behavior of Teichmüller geodesics that spend a definite fraction of their time in some thick part $\mathcal{T}_\epsilon$. In this section we will show that most Teichmüller geodesics in fact satisfy this property. Therefore the tools developed in §3 apply generically and we may use them in studying averaging questions such as Theorems B, C, and D.

In all of what follows, if x is a fixed basepoint and $y \in \mathcal{B}_r(x)$ is a point in the ball centered at x, then we will write y_t to denote the time–t point on the geodesic ray based at x and traveling through y.

5.1. **Volume estimates.** We begin by recalling some estimates on the volume of Teichmüller balls and using these to reduce to the case of annular shells.

Athreya, Bufetov, Eskin, and Mirzakhani [3] have found the following asymptotic estimate for the $\mathbf{m}$–volume of a ball of radius r.

Theorem 5.1 (Volume asymptotics [3, Theorem 1.3])**.** *There is a (bounded) function $f \colon \mathcal{T}(S) \to (0, \infty)$ such that for each $x \in \mathcal{T}(S)$*

$$\lim_{r \to \infty} \frac{\mathbf{m}(\mathcal{B}_r(x)))}{e^{hr}} = f(x).$$

Corollary 5.2 (Definite exponential growth)**.** *Let μ denote Hausdorff measure η, holonomy measure $\mathbf{m}$, or any visual measure $\mu_x = \mathrm{Vis}(\kappa_x)$. For each $x \in \mathcal{T}(S)$, there exist constants $C_1 \leq C_2$ such that for all sufficiently large r (depending on x) we have*

$$C_1 e^{hr} \leq \mu(\mathcal{B}_r(x)) \leq C_2 e^{hr}.$$

Proof. This is built into the definition of the visual measure $\mathrm{Vis}(\kappa_x)$. For the holonomy measure $\mathbf{m}$, this follows from Theorem 5.1 above. The same estimate then holds for η by Proposition 4.6 and Corollary 4.7. $\qquad\square$

Of course, this holds for the other measures discussed in this paper as well by the comparisons in the last section.

For any $r > k > 0$, let $\mathcal{A}_r^k(x) = \mathcal{B}_r(x) \setminus \mathcal{B}_{r-k}(x)$ denote the annular shell between radii r and $r - k$. The fact that the volume of a ball grows exponentially in the radius means that we can focus our attention on annuli rather than on balls.

Lemma 5.3 (Reduction to annuli). *Fix $x \in \mathcal{T}(S)$ and let μ be any measure with definite exponential growth (i.e., satisfying the conclusion of Corollary 5.2). Suppose that for all $k > 0$ we have*

$$\lim_{r \to \infty} \frac{1}{r} \frac{1}{\mu(\mathcal{A}_r^k(x))^2} \int_{\mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x)} d_{\mathcal{T}}(y,z) \ d\mu(y)d\mu(z) = 2.$$

Then the same holds when $\mathcal{A}_r^k(x)$ is replaced by $\mathcal{B}_r(x)$.

Proof. Let C_1, C_2 be as in Corollary 5.2 above. For each k sufficiently large (satisfying $\frac{C_2}{C_1} e^{-hk} < 1$) and all sufficiently large r we have

$$\begin{aligned}
2 &\geq \frac{1}{r} \frac{1}{\mu(\mathcal{B}_r(x))^2} \int_{\mathcal{B}_r(x) \times \mathcal{B}_r(x)} d_{\mathcal{T}}(y,z) \ d\mu(y)d\mu(z) \\
&\geq \left(\frac{\mu(\mathcal{B}_r(x))}{\mu(\mathcal{B}_r(x))} - \frac{\mu(\mathcal{B}_{r-k}(x))}{\mu(\mathcal{B}_r(x))} \right)^2 \frac{1}{r} \frac{1}{\mu(\mathcal{A}_r^k(x))^2} \int_{\mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x)} d_{\mathcal{T}}(y,z) \ d\mu(y)d\mu(z) \\
&\geq \left(1 - \frac{C_2}{C_1} e^{-hk} \right)^2 \frac{1}{r} \frac{1}{\mu(\mathcal{A}_r^k(x))^2} \int_{\mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x)} d_{\mathcal{T}}(y,z) \ d\mu(y)d\mu(z).
\end{aligned}$$

The claim now follows since, by assumption, the latter becomes arbitrarily close to 2 when r and k are sufficiently large. $\qquad\square$

5.2. The thickness property.

Recall from §3 that the thick-stat of a nondegenerate geodesic $[x,y] \subset \mathcal{T}(S)$ is defined by

$$\mathsf{Thk}_\epsilon^\%[x,y] = \frac{\left| \{ 0 \leq s \leq d_{\mathcal{T}}(x,y) : y_s \in \mathcal{T}_\epsilon \} \right|}{d_{\mathcal{T}}(x,y)},$$

where y_s denotes the time–s point on the geodesic ray from x through y. The goal of this section is to show that, for all measures of interest, the thick-stat $\mathsf{Thk}_\epsilon^\%[x,y]$ is uniformly bounded below for most $y \in \mathcal{A}_r^k(x)$. More precisely, we will show that these measures satisfy the following property.

Definition 5.4 (Property P1). We say a measure μ on $\mathcal{T}(S)$ has the *thickness property* (P1) if for all $0 < \theta, \sigma < 1$, there exists $\epsilon > 0$ such that

$$\lim_{r \to \infty} \frac{\mu \left(\{ y \in \mathcal{A}_r^k(x) : \mathsf{Thk}_\epsilon^\%[x,y_t] \geq \theta \text{ for all } \sigma r \leq t \leq r \} \right)}{\mu(\mathcal{A}_r^k(x))} = 1$$

holds for all $x \in \mathcal{T}(S)$ and $k > 0$.

We first observe that, for visual measures, the thickness property follows from the ergodicity of the Teichmüller geodesic flow φ_t.

Proposition 5.5 (Thickness statistics for visual measures). *Let κ_x denote either of the visual measures s_x or ν_x on $\mathcal{Q}^1(x)$. For all $0 < \theta < 1$ there exists $\epsilon > 0$ such that for all $x \in \mathcal{T}(S)$ we have*

$$\lim_{R_0 \to \infty} \kappa_x \left(\{ q \in \mathcal{Q}^1(x) : \mathsf{Thk}_\epsilon^\%[x, \pi(\varphi_r(q))] \geq \theta \text{ for all } r > R_0 \} \right) = 1.$$

Proof. Choose $\epsilon > 0$ sufficiently small so that the proportion of the **m**–volume of moduli space that is ϵ–thick is larger than θ; that is, so that

$$\mathbf{m}\big(\mathcal{T}_\epsilon / \operatorname{Mod}(S)\big) > \theta \cdot \mathbf{m}\big(\mathcal{T}(S) / \operatorname{Mod}(S)\big).$$

By the ergodicity of the geodesic flow [14], it follows that the geodesic ray determined by μ_{hol}–almost every $q \in \mathcal{Q}^1(S)$ spends more than proportion θ of its time in $\mathcal{T}_\epsilon$, asymptotically. The vertical foliation of each such q is uniquely ergodic [15]. If two quadratic differentials have the same vertical uniquely ergodic measured foliation then they are forwards asymptotic [13]. We conclude that almost every measured foliation $F \in \mathcal{MF}$ (with respect to Thurston measure μ_{TH}) has the property that any quadratic differential q with vertical foliation F satisfies $\lim_{r \to \infty} \mathsf{Thk}_\epsilon^\%[\pi(q), \pi(\varphi_r(q))] \geq \theta$.

The map $\mathcal{Q}(x) \to \mathcal{MF}$ which assigns to q its vertical foliation is a smooth map off the multiple zero locus, so it is smooth on a set of full measure. Thus it is absolutely continuous with respect to the measures κ_x and μ_{TH}. Thus the property of asymptotically spending at least proportion θ of the time in $\mathcal{T}_\epsilon$ holds for κ_x–almost every $q \in \mathcal{Q}(x)$. This means that for each $x \in \mathcal{T}(S)$ the quantity

$$\kappa_x \left(\left\{ q \in \mathcal{Q}^1(x) : \mathsf{Thk}_\epsilon^\%[x, \pi(\varphi_r(q))] \geq \theta \text{ for all } r > R_0 \right\} \right)$$

increases to 1 as the threshold R_0 tends to infinity. $\qquad \square$

Since visual measures on $\mathcal{T}(S)$ are obtained by integrating the above measures on spheres, we immediately obtain the thickness property for visual measures.

Corollary 5.6. *The thickness property* (P1) *holds for* $\mathrm{Vis}(s_x)$ *and* $\mathrm{Vis}(\nu_x)$.

5.3. **Random walks.** We next verify (P1) for $\mathbf{m}$, which is considerably more involved. The proof uses ideas of Eskin and Mirzakhani on discretizing geodesics into sample paths of a random walk. We begin the setup by combining some results on the volume of balls from Athreya–Bufetov–Eskin–Mirzakhani [3] and Eskin–Mirzakhani [7].

Lemma 5.7 (Volume of balls [3, Theorem 1.2], [7, Lemma 3.1]). *There exists a constant* $\mathsf{c} > 0$ *(depending only on the topology of S) such that* $\mathbf{m}(\mathcal{B}_\mathsf{c}(y)) \stackrel{\cdot}{\asymp} 1$ *for all* $y \in \mathcal{T}(S)$. *Additionally, given* $\epsilon > 0$ *we have* $\mathbf{m}(\mathcal{B}_r(y)) \stackrel{\cdot}{\prec}_\epsilon e^{hr}$ *for all* $y \in \mathcal{T}_\epsilon$.

Proof. The first claim is exactly Lemma 3.1 of [7]. For the second claim, choose a point $x \in \mathcal{T}_\epsilon$. Note that our choice of x depends only on ϵ. By the volume asymptotics (Theorem 5.1), there exists constant R_0 such that

$$\mathbf{m}(\mathcal{B}_r(x)) \stackrel{\cdot}{\prec}_\epsilon e^{hr}$$

for all $r \geq R_0$. Furthermore, by increasing R_0 if necessary, we may assume that the $\mathrm{Mod}(S)$–translates of $\mathcal{B}_{R_0}(x)$ cover $\mathcal{T}_\epsilon$. It follows that for any $y \in \mathcal{T}_\epsilon$ and any $r \geq 0$ we have

$$\mathbf{m}(\mathcal{B}_r(y)) \leq \mathbf{m}(\mathcal{B}_{r+R_0}(x')) = \mathbf{m}(\mathcal{B}_{r+R_0}(x)) \stackrel{\cdot}{\prec}_\epsilon e^{hR_0} e^{hr}$$

for some $\mathrm{Mod}(S)$–translate x' of x. This establishes the second claim. $\qquad \square$

To define a random walk on $\mathcal{T}(S)$ with basepoint x, first choose a net $\mathcal{N}$ of points in $\mathcal{T}(S)$, choosing so that $x \in \mathcal{N}$ and such that the net points are c–separated and $(2\mathsf{c})$–dense (i.e., the distances between net points are at least c but the $(2\mathsf{c})$–balls about net points cover Teichmüller space). Here c is the constant from Lemma 5.7, depending only on the topology of S.

Given a parameter τ, a *sample path of length s* (starting at x) is a map

$$\lambda \colon \{0, \ldots, \lfloor s/\tau \rfloor\} \to \mathcal{N}$$

such that $\lambda_0 = x$ and for each index, $d_{\mathcal{T}}(\lambda_k, \lambda_{k+1}) \leq \tau$. Let $\mathsf{P}_\tau^x(s)$ be the set of sample paths λ starting at x of length at most s, and let P_τ^x be the set of all sample paths of any length. By (36) of [7], for any $\delta > 0$ and sufficiently large τ (depending on δ) we have,

$$|\mathsf{P}_\tau^x(s)| \leq e^{s(h+\delta)}$$

for all $s \geq 0$. (Note that the constant C_2 in (36) of [7], coming from [7, Proposition 4.5], can be taken to equal 1.)

Now given τ we define a map $F_\tau \colon \mathcal{T}(S) \to \mathsf{P}_\tau^x$ which takes a point y and "discretizes" the geodesic $[x, y]$ to a sample path. For any $[x, y]$ we mark off points along the geodesic starting at x and spaced by time $\tau - 2\mathsf{c}$. For each such point along the geodesic we choose a nearest point in $\mathcal{N}$. This is the sample path $F_\tau(y)$ associated to y. For any $\epsilon_1 > 0$, we can choose τ sufficiently large so that the image under F_τ of the ball $\mathcal{B}_r(x)$ is contained in $\mathsf{P}_\tau^x(r(1 + \epsilon_1))$. Furthermore, after increasing τ if necessary, the above estimate on the cardinality of $\mathsf{P}_\tau^x(s)$ shows that we additionally have

$$|\mathsf{P}_\tau^x(r(1 + \epsilon_1))| \leq e^{hr(1+2\epsilon_1)}$$

for all r. We are now ready to show that the measure of points determining a ray with a too-large thick-stat decays exponentially in r.

Theorem 5.8 (Thickness estimate for holonomy measure). *For all $0 < \theta, \sigma < 1$, $x \in \mathcal{T}(S)$, and $k > 0$, there exist $\epsilon = \epsilon(\theta, \sigma) > 0$ and $\alpha > 0$ such that*

$$\frac{\mathbf{m}\left(\left\{y \in \mathcal{A}_r^k(x) : \mathsf{Thk}_\epsilon^{\%}[x, y_t] < \theta \text{ for some } t \in [\sigma r, r]\right\}\right)}{\mathbf{m}(\mathcal{A}_r^k(x))} < e^{-\alpha r}$$

for all sufficiently large r, where y_t is the time–t point on the geodesic ray from x through y.

Proof. We start with a geodesic $[x, y]$ for $y \in \mathcal{A}_r^k(x)$. As in the above discussion, for any $\epsilon_1 < 1$, we can choose $\tau \geq \frac{4\mathsf{c}}{\epsilon_1}$ so that the geodesic determines a sample path λ in $\mathsf{P}_\tau^x(r(1 + \epsilon_1))$. Let $F_\tau \colon \mathcal{A}_r^k(x) \to P_\tau$ be the map carrying y to the sample path associated to $[x, y]$. In Appendix B we give a self-contained proof of the following statement (Theorem B.6) which is indicated in the proof of Theorem 5.1 of [7]: given θ, there exists $\delta' > 0$ such that for all large τ there is an ϵ' so that for each $\tau \leq t \leq r$, the estimate

$$\frac{1}{\lfloor t/\tau \rfloor} \left|\left\{1 \leq i \leq \lfloor t/\tau \rfloor : \lambda_i \in \mathcal{T}_{\epsilon'}\right\}\right| \geq \theta$$

holds for all but at most

$$e^{-\delta' t} e^{hr(1+2\epsilon_1)}$$

of the sample paths $\lambda \in \mathsf{P}_\tau^x(r(1 + \epsilon_1))$. This is the crucial ingredient that lets us control our thickness statistic. For the given σ, we now choose ϵ_1 small enough (forcing τ to be large) so that $\kappa := \frac{\delta' \sigma}{h} - 2\epsilon_1 > 0$. In particular, note that τ and δ' depend only on θ and σ. If we let $\Omega \subset \mathsf{P}_\tau^x(r(1 + \epsilon_1))$ denote the union of these exceptional sample paths corresponding to any $\sigma r \leq t \leq r$, it follows that

$$|\Omega| \leq \sum_{k=0}^{\infty} e^{-\delta'(\sigma r + k\tau)} e^{hr(1+2\epsilon_1)} \mathrel{\dot{\prec}}_{\theta, \sigma} e^{-\delta' \sigma r} e^{hr(1+2\epsilon_1)} = e^{hr(1-\kappa)}.$$

We know that the ball of radius c centered at any point has $\mathbf{m}$–measure $O(1)$ by Lemma 5.7. Consequently, when r is sufficiently large we have

$$\mathbf{m}(F_\tau^{-1}(\Omega)) \mathrel{\dot{\prec}} |\Omega| \mathrel{\dot{\prec}}_{\theta,\sigma} e^{hr}e^{-\kappa hr} \le \mathbf{m}(\mathcal{A}_r^k(x)) \cdot e^{-\alpha r},$$

for a suitable choice of α. We conclude that for all $y \in \mathcal{A}_r^k(x)$ except for a set of at most this measure, the sample path associated to the geodesic from x has to y has $\mathsf{Thk}_{\epsilon'}^{\%} \ge \theta$ for ending times $t \ge \sigma r$. Now every point on the geodesic is within distance τ of a point on the sample path. Let $\epsilon = \epsilon' e^{-\tau}$. A point at distance at most τ from a point in $\mathcal{T}_{\epsilon'}$ lies in $\mathcal{T}_\epsilon$. This concludes the proof. □

This says that, except for set of endpoints y of exponentially small measure, geodesics have the property that they eventually have spent a definite fraction of their time in the thick part. In particular, this implies the following.

Corollary 5.9. *The measures* $\mathbf{m}$ *and* η *satisfy the thickness property* (P1).

We will again use random walks to now show that a typical geodesic has a long interval where it stays in the thick part. Specifically we say that a geodesic segment $[x,y]$ contains an ϵ–*thick interval* I if there is a geodesic subsegment I along $[x,y]$ such that $I \subset \mathcal{T}_\epsilon$.

Theorem 5.10 (Thick intervals). *For all* $0 < \sigma < 1$, $M > 0$, *and sufficiently small* ϵ, *there exists* $\beta > 0$ *such that for all sufficiently large* r,

$$\frac{\mathbf{m}\left(\left\{y \in \mathcal{A}_r^k(x) : [y_{\sigma r}, y_{2\sigma r}] \text{ contains no } \epsilon\text{–thick interval of length } M\right\}\right)}{\mathbf{m}(\mathcal{A}_r^k(x))} < e^{-\beta r},$$

where y_t *is the time–t point on the geodesic ray from x through y.*

Proof. For small ϵ_1 define τ as in Theorem 5.8. Fix some $0 < \theta < 1$ by Theorem 5.8, for sufficiently small ϵ', except for an exponentially small set of paths the interval $[y_{\sigma r}, y_{2\sigma r}]$ contains a proportion θ of ϵ'–thick points. There exists $\rho > 0$ such that for any net point $y \in \mathcal{T}_{\epsilon'}$, the probability that the next step in the random walk starting at y remains in $\mathcal{T}_{\epsilon'}$ is at least ρ.

Let $\kappa := 1 - \rho^{M/\tau} < 1$. Then given a point $y \in \mathcal{T}_{\epsilon'}$, the probability that in the next $\lfloor M/\tau \rfloor$ steps in the random walk (that is, the sample path of length M) starting at y at least one of the points is ϵ'–thin is at most κ.

Consider a sample path of length $L = \sigma r$ (so having $\lfloor L/\tau \rfloor$ points) for which at least θ proportion of its points are in $\mathcal{T}_{\epsilon'}$. Divide the path into subpaths with $\lfloor M/\tau \rfloor$ points, called *pieces*, so that the number of pieces is additively close to L/M if $M \gg \tau$. Let J_1 be the collection of odd-index pieces and J_2 the collection of even-index pieces. Suppose without loss of generality that J_1 contains at least as many pieces with an ϵ'–thick point as J_2 does. The number of J_1 pieces containing a point in $\mathcal{T}_{\epsilon'}$ is then at least $N := \frac{L\theta}{2M}$. The probability that the sample path of length M starting at an ϵ' thick point enters the ϵ'–thin part is at most κ, as we have seen. If two thick starting points are in different J_1 pieces, then these events are independent, because between the two points is a J_2 piece of length M. Thus the probability that the random path does not have any pieces of length M that lie entirely in $\mathcal{T}_{\epsilon'}$ is at most κ^N.

Now as in the discussion in the previous theorem, we have the map $F_\tau \colon \mathcal{A}_r^k(x) \to \mathsf{P}_\tau^x(r(1+\epsilon_1))$. Let $y \in \mathcal{A}_r^k(x)$. We consider the segment $[y_{\sigma r}, y_{2\sigma r}] \subset [x,y]$. Its image under F_τ is a path of length $L = \sigma r$. The probability that this random path fails

to have the desired segment of length M in the ϵ'–thick part is at most $\kappa^{\sigma r \theta / 2M}$, by the statement at the end of the last paragraph. Since κ is a fixed number smaller than 1, we can find an upper bound for this proportion of the form $e^{-\beta r}$ for some $\beta > 0$. As we saw in the proof of the last theorem, since this property holds for an exponentially small proportion of sample paths, the corresponding property holds for an exponentially small proportion of endpoints $y \in \mathcal{A}_r^k(x)$.

Let $\epsilon = e^{-\tau} \epsilon'$. Note again that if a point on the random path is ϵ'–thick, then the corresponding point on the geodesic $[x, y]$ is ϵ–thick. We conclude, as in the last theorem, that the measure of the set of points $y \in \mathcal{A}_r^k(x)$ such that the geodesic $[x, y]$ does not have an ϵ–thick interval of length M in $[\sigma r, 2\sigma r]$ is at most $e^{-\beta r}$. $\square$

6. SEPARATION STATISTICS FOR PAIRS OF RAYS

We need one final ingredient before proving Theorems B and C. Namely, in order to apply Theorem A to show that the geodesic $[y, z]$ connecting two generic points $y, z \in \mathcal{B}_r(x)$ must "dip back" towards x, we must first know that $[x, y]$ and $[x, z]$ become C–separated, where C is the constant from Theorem A. Thus we need an estimate for the probability that two geodesic rays based at x fellow-travel past a given radius. The appropriate sort of control is ensured by the following property.

Definition 6.1 (Property P2). We say a measure μ on $\mathcal{T}(S)$ satisfies the *separation property* (P2) if for all $M_0, k > 0$, $0 < \sigma < 1$, and $x \in \mathcal{T}(S)$, we have

$$\lim_{r \to \infty} \frac{\mu \times \mu \left(\{ (y, z) \in \mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x) : d_{\mathcal{T}}(y_t, z_t) \geq M_0 \text{ for all } t \in [\sigma r, r] \} \right)}{\mu \times \mu \left(\mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x) \right)} = 1.$$

We will also consider the following time-specific version this separation property:

Definition 6.2 (Property P3). We say that a measure μ on $\mathcal{T}(S)$ satisfies the *strong separation property* (P3), or has exponential decay of fellow travelers, if for all $x \in \mathcal{T}(S)$, $M_0, k > 0$, $0 < \sigma < 1$, there exist $\alpha, R_0 > 0$ such that

$$\frac{\mu \times \mu \left(\{ (y, z) \in \mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x) : d_{\mathcal{T}}(y_t, z_t) < M_0 \} \right)}{\mu \times \mu \left(\mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x) \right)} \leq e^{-\alpha t},$$

whenever $R_0 \leq \sigma r \leq t \leq r$.

Theorem 6.3 (Strong separation for visual measures). *All rotation-invariant visual measures $\mu_x = \mathrm{Vis}(\kappa_x)$ on $\mathcal{T}(S)$, and in particular $\mathrm{Vis}(\nu_x)$ and $\mathrm{Vis}(s_x)$, have the strong separation property* (P3).

Proof. Choose $\sigma r \leq t \leq r$, fix a point $y \in \mathcal{A}_r^k(x)$, and let $E = \{ z \in \mathcal{A}_r^k(x) : d_{\mathcal{T}}(y_t, z_t) < M_0 \}$. Looking instead in the sphere $\mathcal{S}_t(x)$, we have the set $E_t = \{ z \in \mathcal{S}_t(x) : d_{\mathcal{T}}(y_t, z) < M_0 \}$. Notice that, by definition,

$$E = \bigcup_{s \in [r-t-k, r-t]} \hat{\varphi}_s(E_t).$$

(Recall that $\hat{\varphi}_s$ denotes the radial geodesic flow based at x.) Therefore, by the normalized invariance, we have

$$
\begin{aligned}
\mu(E) &= \int_{r-k}^{r} \mathrm{Vis}_s(\kappa_x)(\hat{\varphi}_{s-t}(E_t)) \, d\lambda(s) \\
&= \int_{r-k}^{r} \mathrm{Vis}_t(\kappa_x)(E_t) \frac{\mathrm{Vis}_s(\kappa_x)(\mathcal{S}_s(x))}{\mathrm{Vis}_t(\kappa_x)(\mathcal{S}_t(x))} \, d\lambda(s) \\
&= \frac{\kappa_x(E_t)}{\kappa_x(\mathcal{Q}^1(x))} \mu_x(\mathcal{A}_r^k(x)),
\end{aligned}
$$

where, in the last line, we have identified E_t with its image in $\mathcal{Q}^1(x) \cong \mathcal{S}_t(x)$.

It remains to find R_0 (independent of y) such that $\kappa_x(E_t)/\kappa_x(\mathcal{Q}^1(x)) \stackrel{\cdot}{\prec} e^{-t}$ when $t \geq R_0$. Recall that S^1 acts freely on $\mathcal{Q}^1(x)$ by rotations. Choosing orbit representatives, we may realize $\mathcal{Q}^1(x)$ as a setwise product $(\mathcal{Q}^1(x)/S^1) \times S^1$. The measure κ_x pushes forward to a measure on $\mathcal{Q}^1(x)/S^1$. By disintegration, we then obtain a measure on each fiber S^1 which, by the rotation-invariance of κ_x, must agree with Lebesgue measure up to a scalar. For any two points $z, z' \in E_t$, the triangle inequality gives $d_{\mathcal{T}}(z, z') \leq 2M_0$. Now suppose that z and z' lie in the same Teichmüller disk, meaning that the unit quadratic differentials associated to the geodesics $[x, z]$ and $[x, z']$ lie in the same S^1–orbit. Each Teichmüller disk is an isometrically embedded copy of the hyperbolic plane. Thus, when t is large compared to M_0, hyperbolic geometry implies that the fraction of each S^1–orbit contained in E_t is $\stackrel{\cdot}{\prec} e^{-t}$. Using the product structure and integrating over the $\mathcal{Q}^1(x)/S^1$ factor, Fubini's theorem then implies that $\kappa_x(E_t)/\kappa_x(\mathcal{Q}^1(x)) \stackrel{\cdot}{\prec} e^{-t}$ as well. $\qquad\square$

Theorem 6.4 (Strong separation for holonomy measure)**.** *The measure* $\mathbf{m}$ *satisfies the strong separation property* (P3)*.*

Proof. For a given r, t, let

$$
D_{r,t} = \{(y, z) : d_{\mathcal{T}}(y_t, z_t) < M_0\} \subset \mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x)
$$

denote the the set in question. Set $\theta = {}^1\!/_2$ and choose ϵ and $\alpha > 0$ as in Theorem 5.8 so that the $\mathbf{m}$–measure of the set

$$
E_r = \left\{ y : \mathsf{Thk}_\epsilon^\%[x, y_t] < \theta \text{ for some } t \in [\sigma r, r] \right\} \subset \mathcal{A}_r^k(x)
$$

is at most $\mathbf{m}(\mathcal{A}_r^k(x)) e^{-\alpha r}$ for all large r. We may now write $D_{r,t} = D'_{r,t} \cup D''_{r,t}$, where

$$
D'_{r,t} = \{(y, z) \in D_{r,t} : y \in E_r\} \quad \text{and} \quad D''_{r,t} = \{(y, z) \in D_{r,t} : y \notin E_r\}.
$$

By the above, we have that

(6.5)
$$
\frac{\mathbf{m} \times \mathbf{m}(D'_{r,t})}{\mathbf{m}(\mathcal{A}_r^k(x))^2} \leq \frac{\mathbf{m}(E_r)\mathbf{m}(\mathcal{A}_r^k(x))}{\mathbf{m}(\mathcal{A}_r^k(x))^2} \leq e^{-\alpha t}
$$

for all large r and all $t \leq r$.

Choose any point $y \in \mathcal{A}_r^k(x) \setminus E_r$ and fix some $t \in [\sigma r, r]$. Then any point $z \in \mathcal{A}_r^k(x)$ satisfying $d_{\mathcal{T}}(y_t, z_t) < M_0$ must be contained in the ball of radius $r - t + M_0$ about y_t. Since $t \geq \sigma r$ and $y \notin E_r$; we know that $\mathsf{Thk}_\epsilon^\%[x, y_t] \geq \theta$. Choosing any $0 < \delta < \theta$, there must exist a time $t' \in [\delta t, t]$ for which $y_{t'} \in \mathcal{T}_\epsilon$.

Furthermore, each such z lies within the ball of radius $(1-\delta)t + M_0 + r - t$ about $y_{t'}$. Applying Lemma 5.7, we see that

$$(6.6) \qquad \frac{\mathbf{m} \times \mathbf{m}(D''_{r,t})}{\mathbf{m}(\mathcal{A}^k_r(x))^2} \leq \frac{\mathbf{m}(\mathcal{B}_{r-\delta t + M_0}(y_{t'}))}{\mathbf{m}(\mathcal{A}^k_r(x))} \dot{\prec}_\epsilon \frac{e^{hM_0}e^{hr}e^{-h\delta t}}{\mathbf{m}(\mathcal{A}^k_r(x))}.$$

The Theorem now follows from (6.5), (6.6) and the fact that, by Corollary 5.2, we have $\mathbf{m}(\mathcal{A}^k_r(x)) \dot{\succ}_{x,k} e^{hr}$ for all large r. $\qquad\square$

Finally, we see that for any measure enjoying exponential decay of fellow travelers, most pairs of geodesic rays are in fact *never* near each other beyond some threshold.

Proposition 6.7. *The strong separation property* (P3) *implies the separation property* (P2).

Proof. If $(y,z) \in \mathcal{A}^k_r(x) \times \mathcal{A}^k_r(x)$ does not lie in the set

$$E_r = \big\{(y,z) \in \mathcal{A}^k_r(x) \times \mathcal{A}^k_r(x) : d_\mathcal{T}(y_t, z_t) \geq M_0 \text{ for all } t \in [\sigma r, r]\big\},$$

then there is some $n \in \mathbb{N}$, $n \leq (1-\sigma)r$, such that $d_\mathcal{T}(y_{\sigma r + n}, z_{\sigma r + n}) < M_0 + 2$. Thus all such points are contained in the union of exceptional sets corresponding to the radii $\sigma r, \sigma r + 1, \ldots, \sigma r + \lfloor r - \sigma r \rfloor$. Using the exponential bound provided by property (P3), we see that for large r the complement of E_r has measure at most

$$\mu(\mathcal{A}^k_r(x))^2 \left(e^{-\alpha\sigma r} + e^{-\alpha(\sigma r + 1)} + \cdots + e^{-\alpha(\sigma r + \lfloor r - \sigma r \rfloor)} \right)$$

$$\leq \mu(\mathcal{A}^k_r(x))^2 \left(\frac{1}{1 - e^{-\alpha}} \right) e^{-\alpha\sigma r}. \qquad\square$$

Corollary 6.8. *The Hausdorff measure η, holonomy measure $\mathbf{m}$, and the visual measures* $\mathrm{Vis}(\nu_x)$ *and* $\mathrm{Vis}(s_x)$ *all satisfy property* (P2).

Thus we can conclude that after throwing out a subset of $\mathcal{A}^k_r(x) \times \mathcal{A}^k_r(x)$ of an arbitrarily small proportional measure, all pairs of geodesics stay separated by an arbitrarily chosen distance in Teichmüller space after a threshold time σr has elapsed.

7. STATISTICAL HYPERBOLICITY

We can now assemble our results to prove Theorems B and C.

Theorem 7.1 (Annulus version of statistical hyperbolicity)**.** *Let μ be any measure on $\mathcal{T}(S)$ satisfying the thickness property* (P1) *and the separation property* (P2)*. Fix a basepoint $x \in \mathcal{T}(S)$ and an arbitrary $k > 0$. Then*

$$\lim_{r\to\infty} \frac{1}{r} \frac{1}{\mu(\mathcal{A}^k_r(x))^2} \int_{\mathcal{A}^k_r(x) \times \mathcal{A}^k_r(x)} d_\mathcal{T}(y,z) \; d\mu(y)d\mu(z) = 2.$$

We have shown that these hypotheses are satisfied by the standard visual measures $\mathrm{Vis}(\nu_x)$ and $\mathrm{Vis}(s_x)$, the holonomy measure $\mathbf{m}$, and the Hausdorff measure η (Corollaries 5.6, 5.9, 6.8), and thus also all the other measures considered in this paper. Therefore, combining Theorem 7.1 with Lemma 5.3 (Reduction to annuli) we immediately obtain Theorem B:

Theorem B. *Let μ denote the Hausdorff measure η, holonomy measure $\mathbf{m}$, or either standard visual measure $\mathrm{Vis}(\nu_x)$ or $\mathrm{Vis}(s_x)$. Then for every point $x \in \mathcal{T}(S)$,*

$$\lim_{r \to \infty} \frac{1}{r} \frac{1}{\mu(\mathcal{B}_r(x))^2} \int_{\mathcal{B}_r(x) \times \mathcal{B}_r(x)} d_{\mathcal{T}}(y,z) \; d\mu(y)d\mu(z) = 2.$$

Proof of Theorem 7.1. Set $\theta = \sfrac{3}{4}$. For any $0 < \delta, \sigma < \sfrac{1}{3}$, let $\epsilon = \epsilon(\theta, \sigma) > 0$ be the corresponding thickness parameter guaranteed by Property (P1). For this ϵ and $\theta' = \sfrac{1}{2}$, let C and L be the corresponding constants provided by Theorem A. Properties (P1) and (P2) together imply that for all large r, we may restrict to a subset $E_r \subset \mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x)$ whose complement has proportional μ–measure at most δ and such that all pairs $(y, z) \in E_r$ satisfy

$$\mathsf{Thk}_{\epsilon}^{\%}[x, y_t], \mathsf{Thk}_{\epsilon}^{\%}[x, z_t] \geq \theta \qquad \text{and} \qquad d_{\mathcal{T}}(y_t, z_t) \geq 3C$$

for all $t \in [\sigma r, r]$, where y_t and z_t are the time–t points on the geodesic rays from x through y and z, respectively. Notice that, in this case, the point y_t cannot be within C of any point on $[x, z]$ (by the triangle inequality, any point on $[x, z]$ within C of y_t must lie in $[z_{t-C}, z_{t+C}]$).

We now let $t = 2\sigma r$. Since $\mathsf{Thk}_{\epsilon}^{\%}[x, y_{2\sigma r}] \geq \theta = \sfrac{3}{4}$, it follows that the interval $I_r = [y_{\sigma r}, y_{2\sigma r}]$ satisfies $\mathsf{Thk}_{\epsilon}^{\%} I_r \geq \sfrac{1}{2} = \theta'$ for all large r. We also have length$(I_r) = \sigma r \geq L$ when r is large, and $\sigma < \sfrac{1}{3}$ ensures $I_r \subset [x, y]$. Since $I_r \cap \mathrm{Nbhd}_C([x, z]) = \emptyset$, as noted above, Theorem A now implies that $I_r \cap \mathrm{Nbhd}_C([y, z]) \neq \emptyset$. Therefore $[y, z]$ contains a point in the ball $\mathcal{B}_{2\sigma r + C}(x)$, and so we conclude that

$$d_{\mathcal{T}}(y, z) \geq 2(r - k - (2\sigma r + C))$$

for all $(y, z) \in E_r$. Putting the above estimates together, we find that

$$\liminf_{r \to \infty} \frac{1}{r} \frac{1}{\mu(\mathcal{A}_r^k(x))^2} \int_{\mathcal{A}_r^k(x) \times \mathcal{A}_r^k(x)} d_{\mathcal{T}}(y, z) \, d\mu(y)d\mu(z)$$

$$\geq \liminf_{r \to \infty} \frac{1}{r}(1 - \delta)(2r - 2k - 4\sigma r - 2C) = (1 - \delta)(2 - 4\sigma).$$

Since δ and σ can be chosen arbitrarily small, the result follows. $\qquad\square$

Remark 7.2. One could give an alternate proof of Theorem B that does not rely on Theorem A, but rather combines Rafi's Theorem 3.13 with our Theorem 5.10, which strengthens property (P1) to show that with high probability a sufficiently long geodesic has a totally thick subinterval of definite length. (This holds despite the fact that the probability of such a subinterval occurring at any specified time is small).

We similarly obtain Theorem C:

Theorem C. *For every point $x \in \mathcal{T}(S)$ and either family $\{\mu_r\}$ of standard visual measures $\mu_r = \mathrm{Vis}_r(\nu_x)$ or $\mathrm{Vis}_r(s_x)$ on the spheres $\mathcal{S}_r(x)$, we have*

$$E(\mathcal{T}(S), x, d_{\mathcal{T}}, \{\mu_r\}) = 2.$$

Proof. Visual measures on $\mathcal{T}(S)$ are constructed by radially integrating these visual measures on spheres. In fact, Proposition 5.5 and Theorem 6.3 were proved for annuli by first verifying them for spheres, and so analogous formulations of properties (P1) and (P2) also hold for the visual measures $\{\mu_r\}$ on spheres. The result thus follows by the same argument used to prove Theorem 7.1 above. $\qquad\square$

8. Thin triangles

For $y, z \in \mathcal{B}_r(x)$, we can form the geodesic triangle $\triangle = \triangle(x, y, z)$ whose sides are $[x, y], [x, z]$ and $[y, z]$. Since with positive probability the points y and z are in the thin part of $\mathcal{T}(S)$, we can not expect that almost every triangle is thin as $r \to \infty$. However, it is true that almost every triangle is mostly thin: For a fixed δ, let $0 \leq \Theta_\delta(\triangle) \leq 1$ denote the proportion of the three sides of $\triangle$ consisting of points that lie in a δ–neighborhood of the union of the other two sides.

Theorem D. *Let μ denote either the Hausdorff measure η or the holonomy measure $\mathbf{m}$. Then for all $x \in \mathcal{T}(S)$ and $\sigma > 0$ there exists $\delta > 0$ such that*

$$\liminf_{r \to \infty} \frac{1}{\mu(\mathcal{B}_r(x))^2} \int_{\mathcal{B}_r(x) \times \mathcal{B}_r(x)} \Theta_\delta(\triangle(x, y, z)) \; d\mu(y) d\mu(z) \geq 1 - \sigma.$$

Before embarking on the proof, we use the results in the appendix to establish the following analog of Theorem 5.8.

Lemma 8.1 (Thick-stat near the end)**.** *For all $0 < \theta, \sigma < 1$, $x \in \mathcal{T}(S)$ there exist $\epsilon'' = \epsilon''(\theta, \sigma) > 0$ and $\alpha > 0$ such that for all $k > 0$ and all sufficiently large r*

$$\frac{\mathbf{m}\left(\left\{ y \in \mathcal{A}_r^k(x) : \mathsf{Thk}_{\epsilon''}^{\%}[y_{(1-\sigma)r}, y] < \theta \right\}\right)}{\mathbf{m}(\mathcal{A}_r^k(x))} < e^{-\alpha r},$$

where y_t denotes the time–t point along the geodesic ray from x through y.

Proof. For the given θ, we let $\rho = 1 - \theta$ and take $\delta' = \rho/8$. For this δ' and the given σ, we choose $\epsilon_1 \ll \sigma$ and $\tau \geq \frac{4\mathsf{c}}{\epsilon_1}$ accordingly so that $\kappa := \frac{\delta' \sigma}{h} - 2\epsilon_1 > 0$. Recall that every geodesic $[x, y]$ with $y \in \mathcal{A}_r^k(x)$ then determines a sample path in the set $\mathsf{P}_\tau^x(r(1 + \epsilon_1))$ of random walks of at most $r(1 + \epsilon_1)/\tau$ steps starting at x. As in the Appendix, we let μ_τ^x denote the probability measure on $\mathsf{P}_\tau^x(r(1 + \epsilon_1))$, and by Lemma B.2 we moreover assume that τ is sufficiently large so that the cardinality of any subset $A \subset \mathsf{P}_\tau^x(r(1 + \epsilon_1))$ is bounded by $\mu_\tau^x(A)e^{hr(1+2\epsilon_1)}$.

Fixing τ and ρ as above and choosing ϵ so that $x \in \mathcal{T}_\epsilon$, we now apply Theorem B.5 with a constant γ satisfying $c_\tau^\rho < \gamma < c_\tau^{\rho/2}$. This provides ϵ' and a cocompact set $\mathcal{K} \supset \mathcal{T}_\epsilon$ for which the conclusion of that theorem holds. Since τ is assumed to be large, by Proposition B.3 we may suppose $c_\tau < e^{-\tau/2}$ so that $\gamma < e^{-\rho\tau/4}$. Roughly, the idea is to now apply Theorem B.5 to the set of sample paths starting at $y_{(1-2\sigma)r}$.

To this end, we first argue that most random walks $\lambda \in \mathsf{P}_\tau^x(r(1 + \epsilon_1))$ land in $\mathcal{K}$ for some step in the interval $[2\sigma r/\tau, \sigma r/\tau]$ of steps from the end. The subset $\Omega \subset \mathsf{P}_\tau^x(r(1 + \epsilon_1))$ of exceptional sample paths that avoid $\mathcal{K}$ between steps $2\sigma r/\tau$ and $\sigma r/\tau$ from the end is the union over $j \geq 0$ of sets Ω_j consisting of paths that lie in $\mathcal{K}$ at step $2\sigma r/\tau + j$ from the end and stay outside $\mathcal{K}$ for the next $\sigma r/\tau + j$ steps. Now (B.7) in the Appendix says that for each j the probability of Ω_j satisfies

$$\mu_\tau^x(\Omega_j) \leq M \left(e^{-\tau/4} \right)^{\frac{\sigma r}{\tau} + j} = M e^{-\sigma r/4} e^{-\tau j/4}.$$

for some constant $M = M(\tau)$ (note that the constant γ_0 appearing in (B.7) satisfies $\gamma_0 < \gamma^{1/\rho} < e^{-\tau/4}$). Assuming that r is large, if we sum over j we see that the probability of a random walk lying in Ω satisfies

$$\mu_\tau^x(\Omega) \leq e^{-\sigma r/8} < e^{-\delta' \sigma r}.$$

Now consider the paths that do land in $\mathcal{K}$ for some step in the interval $[2\sigma r/\tau, \sigma r/\tau]$ of steps from the end, namely the set

$$\Sigma := \mathsf{P}_\tau^x(r(1+\epsilon_1)) \setminus \Omega.$$

This can be partitioned into sets Σ_j for $j = 0, 1, \ldots, \sigma r/\tau$, where a path is in Σ_j if its first $\mathcal{K}$–point in the interval of $[2\sigma r/\tau, \sigma r/\tau]$ steps from the end appears at $n_j = \sigma r/\tau + j$ steps from the end.

We claim that most paths in Σ_j will spend at least θ proportion of their final n_j steps in $\mathcal{T}_{\epsilon'}$. Indeed, by Theorem B.5, the probability that a random walk in Σ_j fails to have at least θ proportion of its final n_j steps $\mathcal{T}_{\epsilon'}$ is bounded by

$$\gamma^{n_j} < e^{-n_j\tau\rho/4} = e^{-2\delta'n_j\tau} = e^{-2(\delta'\sigma r + \delta'\tau j)} < e^{-\delta'\sigma r}.$$

Except for these paths, every path in Σ_j spends at least θ proportion of its last $n_j = \sigma r/\tau + j$ steps in $\mathcal{T}_{\epsilon'}$, and therefore (since $j \leq \sigma r/\tau$) at least $\theta/2$ proportion of its final $2\sigma r/\tau$ steps in $T_{\epsilon'}$.

Since the above bound of $e^{-\delta'\sigma r}$ holds for each set Σ_j, it follows that the probability that a random walk in Σ fails to spend at least $\theta/2$ proportion of its last $2\sigma r/\tau$ steps in $\mathcal{T}_{\epsilon'}$ is at most $e^{-\delta'\sigma r}$. Letting $\Sigma' \subset \Sigma$ denote these exceptional paths, we conclude that

$$\mu_\tau^x(\Sigma') \leq e^{-\delta'\sigma r}\mu_\tau^x(\Sigma) \leq e^{-\delta'\sigma r}.$$

Adding this to our previous estimate on $\mu_\tau^x(\Omega)$, we find that except for the subset $\Omega \cup \Sigma' \subset \mathsf{P}_\tau^x(r(1+\epsilon_1))$ of μ_τ^x–measure at most $2e^{-\delta'\sigma r}$, every sample path in $\mathsf{P}_\tau^x(r(1+\epsilon_1))$ spends at least $\theta/2$ proportion of its final $2\sigma r/\tau$ steps in $\mathcal{T}_{\epsilon'}$. In particular, by Lemma B.2 the number of these exceptional sample paths is at most

$$\mu_\tau^x(\Omega \cup \Sigma')e^{hr(1+2\epsilon_1)} \leq 2e^{-\delta\sigma r}e^{hr(1+2\epsilon_1)} = 2e^{hr(1-\kappa)}.$$

(Recall that $\epsilon_1 \ll \sigma$ was chosen so that $\kappa = \frac{\delta'\sigma}{h} - 2\epsilon_1 > 0$). Since the ball of radius c centered at any point has $\mathbf{m}$–measure $\stackrel{.}{\prec} 1$ by Lemma 5.7, it follows that for all large r, the set of points $y \in \mathcal{A}_r^k(x)$ for which the sample path associated to the geodesic $[x, y]$ lies in $\Omega \cup \Sigma'$ has $\mathbf{m}$–measure at most on the order of

$$2e^{hr(1-\kappa)} \stackrel{.}{\prec} \mathbf{m}(\mathcal{A}_r^k(x))e^{-\kappa r} \leq \mathbf{m}(\mathcal{A}_r^k(x))e^{-\kappa r/2}.$$

We conclude that for all $y \in \mathcal{A}_r^k(x)$ except for a set of at most this measure, the sample path associated to the geodesic $[x, y]$ spends at least $\theta/2$ proportion of its final $2\sigma r/\tau$ steps in $\mathcal{T}_{\epsilon'}$. Since every point of such a geodesic is within distance τ of a point on the sample path, by setting $\epsilon'' = \epsilon'e^{-\tau}$ it follows that $\mathsf{Thk}_{\epsilon''}^{\%}[y_{(1-2\sigma)r}, y] \geq \theta/2$ for every such $y \in \mathcal{A}_r^k(x)$. Therefore we have proved the claim for $\theta/2$ and 2σ. As this can be done for any $0 < \theta, \sigma < 1$, the theorem follows. $\square$

Proof of Theorem D. By Corollary 4.7, it suffices to assume $\mu = \mathbf{m}$, and by Lemma 5.3 it suffices to estimate the expected value over $\mathcal{A}_r^k(x)^2$ rather than $\mathcal{B}_r(x)^2$. As in the proof of 7.1, we set $\theta = 3/4$ and let $\epsilon = \epsilon(\theta, \sigma)$ be the smaller of the constants provided by Property (P1) and Lemma 8.1. For this ϵ and $\theta' = 1/2$, we then let C and L be the corresponding constants provided by Theorem A.

Properties (P1)–(P2) and Lemma 8.1 show that for all large r we may restrict to a subset $E_r \subset \mathcal{A}_r^k(x)^2$ – with complement having arbitrarily small proportional $\mathbf{m}$–measure – such that for all $(y, z) \in E_r$ the four intervals

$$[y_{\sigma r}, y_{2\sigma r}], \quad [z_{\sigma r}, z_{2\sigma r}], \quad [y_{(1-2\sigma)r}, y_{(1-\sigma)r}], \quad \text{and} \quad [z_{(1-2\sigma)r}, z_{(1-\sigma)r}]$$

all have $\mathsf{Thk}_\epsilon^{\%} \geq \theta' = 1/2$ and such that $d_{\mathcal{T}}(y_t, z_t) \geq 3C$ for all $t \geq \sigma r$. By Theorem A, it follows that these intervals respectively contain points y', z', y'', z'' that each lie within distance C of the geodesic $[y, z]$. Thus by Rafi's fellow traveling result [22, Theorem C], there is a constant $C' > C$ such that the intervals $[y', y'']$ and $[z', z'']$ are entirely contained within C' of $[y, z]$.

Now, $[y', y'']$ and $[z', z'']$ both have length at least $(1 - 4\sigma)r$, and so the subintervals of $[y, z]$ that they fellow travel must each have length at least $(1 - 4\sigma)r - 2C'$. Thus we have identified subintervals of $\triangle(x, y, z)$ whose lengths total at least $4(1 - 4\sigma)r - 4C'$ and which lie within C' of the union of the other two sides of $\triangle(x, y, z)$. Since the total perimeter of $\triangle(x, y, z)$ is at most $4r$, it follows that

$$\Theta_{C'}(\triangle(x, y, z)) \geq 1 - 4\sigma - \frac{C'}{r}$$

for all $(y, z) \in E_r$. Since the complement of E_r has arbitrarily small proportional **m**–measure, the result follows. $\square$

Remark 8.2. The above proof may be easily adapted to show that the conclusion of Theorem D also holds for the expected value of $\Theta_\delta(y_1, y_2, y_3)$ over all triples of points $y_1, y_2, y_3 \in \mathcal{B}_r(x)$.

APPENDIX A. REPACKAGED DISTANCE FORMULA

The purpose of this appendix is to repackage Rafi's distance formula (Theorem 2.3) in a way that treats annular and non-annular subsurfaces on equal footing:

Proposition A.1 (Repackaged distance formula)**.** *Given any sufficiently large threshold M_0, for all $x, y \in \mathcal{T}(S)$ we have:*

$$d_{\mathcal{T}}(x, y) \ \asymp_{M_0} \ d_S(x, y) + \sum_Y [d_Y(x, y)]_{M_0}$$

Here, the sum is over all (annular and non-annular) proper subsurfaces.

For simplicity and without loss of generality, below we suppose that ϵ_0 has been chosen small enough that $\log_+(1/\epsilon_0) \geq 100$, say. We begin with a straightforward reformulation.

Lemma A.2. *For any sufficiently large threshold M_0, for all $x, y \in \mathcal{T}(S)$ we have:*

$$d_{\mathcal{T}}(x, y) \ \asymp_{M_0} \ d_S(x, y) + \sum_V [d_V(x, y)]_{M_0} + \sum_{A\,:\,\partial A \in \Gamma_{xy}} [d_A(x, y)]_{M_0} \ +$$

$$\sum_{A\,:\,\partial A \notin \Gamma_{xy}} \left[\max \left\{ \log_+ \left(d_{\mathcal{C}(A)}(x, y) \right), \log_+ \left(\frac{1}{l_x(\partial A)} \right), \log_+ \left(\frac{1}{l_y(\partial A)} \right) \right\} \right]_{\log M_0}$$

Proof. Since Γ_{xy}, Γ_x and Γ_y each contain at most $3g - 3$ curves, each max over these sets is within bounded multiplicative error of the corresponding sum, and applying a threshold only creates bounded additive error, so the first three terms of the lemma are established. By the definition of Γ_x we have

$$\sum_{\alpha \in \Gamma_x} \log_+ \left(\frac{1}{l_x(\alpha)} \right) = \sum_{\alpha \notin \Gamma_{xy}} \log_+ \left[\frac{1}{l_x(\alpha)} \right]_{1/\epsilon_0} .$$

Since this is a sum with at most $3g - 3$ nonzero terms, we can increase the threshold to any number $M_0 \geq 1/\epsilon_0$ with bounded additive error. Finally, for functions f, g, h,

we have (in fact with the implied multiplicative constant equal to 3)

$$\log_+[f]_{M_0} + \log_+[g]_{M_0} + \log_+[h]_{M_0} \; \asymp \; \big[\max\{\log_+ f, \log_+ g, \log_+ h\}\big]_{\log M_0}. \qquad \square$$

We now show that each term in the last summand is bilipschitz equivalent to the corresponding hyperbolic distance $d_A(x, y)$.

Lemma A.3. *Consider an annular subsurface $A \subset S$ with core curve $\partial A = \alpha$. For each pair of points $x, y \in \mathcal{T}(S)$, set*

$$H_A(x, y) := \max \left\{ \log_+\big(d_{\mathcal{C}(A)}(x, y)\big), \log_+\left(\frac{1}{l_x(\alpha)}\right), \log_+\left(\frac{1}{l_y(\alpha)}\right) \right\}.$$

If $x, y \in \mathcal{T}(S)$ are such that $\alpha \notin \Gamma_{xy}$ and either $d_A(x, y)$ or $H_A(x, y)$ is greater than $36 \log_+(1/\epsilon_0)$, then $6^{-1} d_A(x, y) \le H_A(x, y) \le 6 d_A(x, y)$.

Proof. Choose points $x, y \in \mathcal{T}(S)$ that satisfy the hypotheses. To fix notation, set $\pi'_\alpha(x) = (0, 1)$ and $\pi'_\alpha(y) = (d_{\mathcal{C}(A)}(x, y), 1)$. These are the closest-point projections of $\pi_\alpha(x)$ and $\pi_\alpha(y)$ to the horocycle bounding $\mathbb{H}_\alpha$, and their distances from these points are exactly given by $\log_+(1/l_x(\alpha))$ and $\log_+(1/l_y(\alpha))$. Let

$$B = d_{\mathbb{H}^2}(\pi'_\alpha(x), \pi'_\alpha(y)) = \operatorname{arccosh}\left(1 + \frac{d_{\mathcal{C}(A)}(x, y)^2}{2}\right)$$

denote the hyperbolic distance between these projections. Using this formula, one may easily check that the inequalities

$$(\text{A.4}) \qquad\qquad \log_+ d_{\mathcal{C}(A)}(x, y) \le B \le 4 \log_+ d_{\mathcal{C}(A)}(x, y)$$

hold provided that either $B \ge 3$ or $d_{\mathcal{C}(A)}(x, y) \ge 3$.

Applying the triangle inequality with the points $\pi'_\alpha(x)$ and $\pi'_\alpha(y)$ implies that

$$(\text{A.5}) \qquad\qquad d_A(x, y) \le \log_+\left(\frac{1}{l_x(\alpha)}\right) + B + \log_+\left(\frac{1}{l_y(\alpha)}\right).$$

Then (A.4), (A.5), and the definition of H_A imply that $d_A(x, y) \le 6 H_A(x, y)$ in the case that $B \ge 3$. If $B < 3$, we claim that the hypotheses of the Lemma ensure that B cannot be the largest term on the right-hand side and therefore that $d_A(x, y) \le 3L \le 3 H_A(x, y)$, where L denotes the larger of the other two terms. Indeed, if B were the largest term and $B < 3$, then (A.5) would imply $d_A(x, y) < 9$, and (A.4) would necessitate $\log_+ d_{\mathcal{C}(A)}(x, y) < 3$ so that $H_A(x, y) < 3$. But then both d_A and H_A are less than 9, contradicting the hypothesis.

By the above, the assumption $d_A(x, y) \ge 36 \log_+(1/\epsilon_0)$ implies that $H_A(x, y) \ge 6 \log_+(1/\epsilon_0)$; therefore all cases will be covered by proving that this in turn implies $H_A(x, y) \le 6 d_A(x, y)$. Without loss of generality, we may assume that $l_x(\alpha) \le l_y(\alpha)$; since $\alpha \notin \Gamma_{xy}$ this guarantees $l_y(\alpha) \ge \epsilon_0$. First suppose that $\log_+ d_{\mathcal{C}(A)}(x, y) \ge 3 \log_+(1/l_x(\alpha))$, in which case we have $\log_+ d_{\mathcal{C}(A)}(x, y) = H_A(x, y) \ge 6 \log_+(1/\epsilon_0)$. In particular we certainly have $d_{\mathcal{C}(A)}(x, y) \ge 3$; thus (A.4) and the triangle inequality give

$$\log_+ d_{\mathcal{C}(A)}(x, y) \le B \le \log_+\left(\frac{1}{l_x(\alpha)}\right) + d_A(x, y) + \log_+\left(\frac{1}{l_y(\alpha)}\right).$$

Therefore $H_A(x, y) = \log_+ d_{\mathcal{C}(A)}(x, y) \le 3 d_A(x, y)$ in this case. The remaining possibility $\log_+ d_{\mathcal{C}(A)}(x, y) \le 3 \log_+(1/l_x(\alpha))$ necessitates $3 \log_+(1/l_x(\alpha)) \ge H_A(x, y)$.

Recall that $\pi'_\alpha(x)$ is the *closest* point projection of $\pi_\alpha(x)$ to the horocycle bounding $\mathbb{H}_\alpha$; since $\pi'_\alpha(y)$ is also on this horocycle we have

$$\log_+\left(\frac{1}{l_x(\alpha)}\right) \leq d_{\mathbb{H}^2}(\pi_\alpha(x), \pi'_\alpha(y)) \leq d_A(x,y) + \log_+\left(\frac{1}{l_y(\alpha)}\right).$$

The assumptions $3\log_+(1/l_x(\alpha)) \geq H_A(x,y) \geq 6\log_+(1/\epsilon_0)$ and $l_y(\alpha) \geq \epsilon_0$ now ensure that $H_A(x,y) \leq 6d_A(x,y)$. $\qquad\square$

Corollary A.6. *Let $H_A(x,y)$ be defined as in Lemma A.3. Then for any threshold $M_0 \geq 36\log_+(1/\epsilon_0)$ and any $x, y \in \mathcal{T}(S)$ we have*

$$\sum_{\partial A \notin \Gamma_{xy}} 6^{-1}\left[d_A(x,y)\right]_{6M_0} \leq \sum_{\partial A \notin \Gamma_{xy}} \left[H_A(x,y)\right]_{M_0} \leq \sum_{\partial A \notin \Gamma_{xy}} 6\left[d_A(x,y)\right]_{M_0/6}$$

With these estimates, we can derive the simplified distance formula.

Proof of Repackaged Distance Formula. Choose any sufficiently large threshold M_0 such that Lemma A.2 holds for both e^{6M_0} and $M_0/6$ and such that $M_0/6 \geq 36\log_+(1/\epsilon_0)$. Notice that, in any sum of the form $\sum[f]_M$, raising the threshold can only decrease the value of the sum, and lowering the threshold can only increase its value. Therefore, combining Lemma A.2 and Corollary A.6 we find that for any $x, y \in \mathcal{T}(S)$ the various distances satisfy

$$d_\mathcal{T} \prec_{\mathsf{M}_0} d_S + \sum_V \left[d_V\right]_{e^{6M_0}} + \sum_{\partial A \in \Gamma_{xy}} \left[d_A\right]_{e^{6M_0}} + \sum_{\partial A \notin \Gamma_{xy}} \left[H_A\right]_{6M_0}$$

$$\leq 6\left(d_S + \sum_V \left[d_V\right]_{M_0} + \sum_{\partial A \in \Gamma_{xy}} \left[d_A\right]_{M_0} + \sum_{\partial A \notin \Gamma_{xy}} \left[d_A\right]_{M_0}\right),$$

where we have suppressed the x and y in the notation. The lower bound on $d_\mathcal{T}(x,y)$ is similar. $\qquad\square$

Appendix B. Thickness statistics for random walks

The purpose of this Appendix is to give a detailed proof of Theorem B.5 and its consequence Theorem B.6. A version of the later statement, as well as a sketch of the proof, was indicated in [7]. However, as our application requires more control of the constants than the statement in [7] provides, we include a precise formulation. Eskin–Mirzakhani in [7, §4.1] define for each $\tau > 0$ a $\mathrm{Mod}(S)$–invariant function u_τ which descends to a proper function on the quotient moduli space. The actual definition of u_τ is somewhat complicated, but $u_\tau(x)$ is (up to multiplicative constants depending on τ) comparable to $1/(\text{length of the shortest curve on } x)$. Sets of the form $\{x : u_\tau(x) \leq C\}$ are cocompact subsets of $\mathcal{T}$. There is a constant M_τ such that if $d_\mathcal{T}(x,y) \leq \tau$ then

$$(\text{B.1}) \qquad\qquad \frac{u_\tau(x)}{u_\tau(y)} \leq M_\tau.$$

Recall from §5.3 that $\mathcal{N}$ denotes a fixed net in $\mathcal{T}$ and that, given a parameter τ, P_τ^x denotes the set of all sample paths $\lambda\colon [0, 1, \dots] \to \mathcal{N}$ (of any length) starting at the basepoint x and satisfying $d_\mathcal{T}(\lambda_i, \lambda_{i+1}) \leq \tau$ for all i. If random sample paths $\lambda = (\lambda_0, \lambda_1, \dots)$ are constructed via the Markov process in which the net point λ_{i+1} is selected uniformly at random among all net points in the ball $\mathcal{B}_\tau(\lambda_i)$, then this Markov process determines a probability measure μ_τ^x on the set P_τ^x of all sample

paths (i.e., random walks) starting at x. Recall that $\mathsf{P}_\tau^x(r)$ denotes the set of sample paths of at most $\lfloor r/\tau \rfloor$ steps, so that the distance from start to end is at most r. For our applications, we will be concerned about the *number* of sample paths satisfying a certain property, rather than the μ_τ^x–measure of such paths. These quantities are related in the following manner, which was implicit in [7].

Lemma B.2. *For any $\delta > 0$, there exists τ_0 such that for all $\tau > \tau_0$ and all r, the cardinality of any subset $A \subset \mathsf{P}_\tau^x(r)$ satisfies*

$$|A| \leq \mu_\tau^x(A) \cdot e^{(h+\delta)r}.$$

Proof. Since A is a finite set, $\mu_\tau^x(A)$ is simply the sum of the measures of the individual elements of A. By definition of the Markov process, the μ_τ^x measure of an element $\lambda \in A$ is the reciprocal of the product of the number of choices for the first step of λ times the number of choices for the second step, and so on. Thus

$$\mu_\tau^x(\lambda) = \left(|\mathcal{B}_\tau(\lambda_0) \cap \mathcal{N}| \cdot |\mathcal{B}_\tau(\lambda_1) \cap \mathcal{N}| \cdots |\mathcal{B}_\tau(\lambda_{\lfloor r/\tau \rfloor - 1}) \cap \mathcal{N}| \right)^{-1}$$

Combining Proposition 4.5 and equation (17) of [7], it follows that for the given δ there exists τ_0 such that for all all $\tau \geq \tau_0$ one has $|\mathcal{B}_\tau(y) \cap \mathcal{N}| \leq e^{(h+\delta)\tau}$ for all $y \in \mathcal{T}(S)$. Therefore, for each $\lambda \in A$ we find that

$$\mu_\tau^x(\lambda) \geq \left(e^{(h+\delta)\tau} \right)^{-\lfloor r/\tau \rfloor} \geq e^{-(h+\delta)r}.$$

Adding these estimates for all $\lambda \in A$ yields the claimed inequality. $\square$

We now define an averaging operator A_τ which, given a function $f \colon \mathcal{N} \to \mathbb{R}_+$, produces the new function $A_\tau f \colon \mathcal{N} \to \mathbb{R}_+$ defined as

$$A_\tau f(x) := \int_{\lambda \in \mathsf{P}_\tau^x} f(\lambda_1) \, d\mu_\tau^x(\lambda).$$

Thus $A_\tau f(x)$ is the expected value of $f(\lambda_1)$ among all random walks starting at x.

In [7], Eskin and Mirzakhani defined the averaging operator slightly differently. For a function $f \colon \mathcal{T} \to \mathbb{R}_+$ they defined

$$A_\tau f(x) := \frac{1}{\mathbf{m}(B_\mathcal{T}(x,\tau))} \int_{B_\mathcal{T}(x,\tau)} f(y) d\mathbf{m}(y)$$

and then established the following estimates regarding their function u_τ.

Proposition B.3 (Theorem 4.1 of [7]). *For all sufficiently large $\tau \geq 0$, there are constants c_τ and b_τ such that*

- $(A_\tau u_\tau)(y) \leq c_\tau u_\tau(y) + b_\tau$ *for all $y \in \mathcal{T}$, and*
- $c_\tau \leq C' e^{-\tau}$ *for a universal constant C' depending only on genus.*

The same estimates hold for our averaging operator over the discrete random walks on the net. This follows from (17) of [7] which says that the number of net points in a ball of radius τ is comparable to the volume of the ball, and the fact that on a ball of radius $2\mathsf{c}$ the value of u_τ is (up to multiplicative constants depending on c) the value at the center.

We now state a general proposition about Markov chains proved by Athreya in [2]. Denote a state space by $\mathcal{S}$ and let P_s be the probability measure on all random walks in $\mathcal{S}$ starting at s. For a subset $C \subset \mathcal{S}$ and a random walk $X = (X_0, X_1, \dots)$ starting at $s = X_0$, denote by $\tau_C(X) := \inf\{n \geq 0 \mid X_n \in C\}$ the first step at which

X enters C. Proposition 3.1 of [2] then gives the following in terms of an averaging operator A defined over steps of the random walk, as above.

Proposition B.4. *Suppose there exist constants $0 < c < 1$ and $b \geq 0$ and a function $V \colon \mathcal{S} \to \mathbb{R}_+$ defined on the state space of a random walk that satisfies $(AV)(s) \leq cV(s) + b$ for all $s \in \mathcal{S}$. Then for all $l \geq 0$, all $s \notin C_l := \{y \in \mathcal{S} \mid V(y) \leq l\}$, and all $n \geq 0$ we have*

$$P_s\big(\tau_{C_l}(X) > n\big) \leq \frac{V(s)}{l}\left(c + \frac{b}{l}\right)^n.$$

Following Athreya's proof of Theorem 2.3 in [2, §6], we now use Proposition B.3 and Proposition B.4 to get an *exponential* bound on the probability that a random walk λ in $\mathcal{T}$ spends a large fraction of its time outside of large compact sets.

Theorem B.5 (Fraction in the thick part; c.f. [2, Theorem 2.3]). *For all $0 < \rho < 1$, $\epsilon, \tau > 0$, and γ such that $c_\tau^\rho < \gamma < 1$, there exists $\epsilon' > 0$ and a cocompact set $\mathcal{K}$ containing $\mathcal{T}_\epsilon$ such that for all $x \in \mathcal{N} \cap \mathcal{K}$ and all $n \geq 1$ we have*

$$\mu_\tau^x\left(\left\{\lambda \in \mathsf{P}_\tau^x : \frac{|\{1 \leq k \leq n : \lambda_k \notin \mathcal{T}_{\epsilon'}\}|}{n} > \rho\right\}\right) < \gamma^n.$$

As an almost immediate consequence we will have the following (c.f. [7]).

Theorem B.6. *Given $0 < \theta < 1$, there exist $\delta' > 0$ such that for all $\epsilon, \epsilon_1 > 0$ and all large τ there exists ϵ' so that for all $x \in \mathcal{T}_\epsilon \cap \mathcal{N}$ and each $\tau \leq t \leq r$, the estimate*

$$\frac{1}{\lfloor t/\tau \rfloor}\left|\left\{1 \leq i \leq \lfloor t/\tau \rfloor : \lambda_i \in \mathcal{T}_{\epsilon'}\right\}\right| \geq \theta$$

holds for all but at most

$$e^{-\delta' t} e^{hr(1+2\epsilon_1)}$$

of the sample paths in $\mathsf{P}_\tau^x(r(1+\epsilon_1))$.

Proof. We assume τ is large enough so that, by Lemma B.2, the cardinality of any subset $A \subset \mathsf{P}_\tau^x(r(1+\epsilon_1))$ satisfies $|A| \leq \mu_\tau^x(A)e^{hr(1+2\epsilon_1)}$. Let $\rho = 1 - \theta$ and set $\delta' = \rho/4$. For large τ, we have $c_\tau < e^{-\tau/2}$, and we choose γ so that $c_\tau^\rho < \gamma < c_\tau^{\rho/2} < 1$. By Theorem B.5 there is a corresponding ϵ' for these $\epsilon, \tau, \rho, \gamma$ such that if $x \in \mathcal{T}_\epsilon$, then the μ_τ^x–measure of the set of paths in P_τ^x that in their first $n = \lfloor t/\tau \rfloor$ steps spend more than ρ proportion of time outside $T_{\epsilon'}$ is at most $\gamma^n < c_\tau^{n\rho/2} < e^{-n\rho\tau/4}$. By Lemma B.2 the number of paths that spend more than ρ proportion of time outside the $\mathcal{T}_{\epsilon'}$ (equivalently spending at most θ proportion inside $\mathcal{T}_{\epsilon'}$) is thus bounded by

$$\gamma^n e^{hr(1+2\epsilon_1)} < e^{-\rho t/4} e^{hr(1+2\epsilon_1)} = e^{-\delta' t} e^{hr(1+2\epsilon_1)}. \qquad \square$$

We end with the proof of Theorem B.5.

Proof. For the given γ, we have $c_\tau < \gamma^{1/\rho} < 1$. Therefore we may choose $l > 0$ sufficiently large so that $\gamma_0 = (c_\tau + \frac{b}{l}) < \gamma^{1/\rho}$. We also take l large enough to satisfy $\mathcal{T}_\epsilon \subset C_l = \{y \in \mathcal{T} : u_\tau(y) \leq l\}$ for the given ϵ. Define $\mathcal{K} := C_l$, which is cocompact, and fix any point $x \in \mathcal{N} \cap \mathcal{K}$. Note that $u_\tau(x) \leq l$.

For integers $i \geq 0$ define functions $t_i \colon \mathsf{P}_\tau^x \to \mathbb{N}$ as follows:

- $t_0 = \inf\{k \geq 0 \mid \lambda_k \in C_l\} = 0$,
- $t_{2i-1} = \inf\{k > t_{2i-2} \mid \lambda_k \notin C_l\}$ for $i \geq 1$, and

- $t_{2i} = \inf\{k > t_{2i-1} \mid \lambda_k \in C_l\}$ for $i \geq 1$.

That is, evaluated on sample path $\lambda \in \mathsf{P}_\tau^x$, t_0 is the first time that λ is in C_l (which is always the first step since we assume $\lambda_0 = x \in C_l$), t_1 is the first time that λ steps outside of C_l, t_2 is the next time that λ steps into C_l, t_3 is the next time that λ steps outside of C_l, and so on. Now define $s_i = t_i - t_{i-1}$ for $i \geq 1$. So:

- s_{2i} is the number of steps that $\lambda \in \mathsf{P}_\tau^x$ takes *outside* of C_l on its ith sojourn outside, and
- s_{2i-1} is the number of steps that $\lambda \in \mathsf{P}_\tau^x$ takes *inside* of C_l on its ith sojourn inside.

Now, for each $\lambda \in \mathsf{P}_\tau^x$, we define a function $F_\lambda \colon \mathbb{N} \to \{0,1\}$ as an indicator for C_l, as follows:

$$F_\lambda(k) = \begin{cases} 1, & t_{2i-1} \leq k < t_{2i} \text{ for some } i; \\ 0, & t_{2i} \leq k < t_{2i+1} \text{ for some } i. \end{cases}$$

Then

$$\frac{1}{n}\big|\{1 \leq k \leq n : u_\tau(\lambda_k) > l\}\big| = \frac{1}{n}\sum_{k=1}^{n} F_\lambda(k).$$

By definition, we have $|t_{2k} - t_{2k-2}| \geq 2$ for all k, which in turn implies that $k \leq t_{2k}$ for all $k \geq 1$. It now follows that

$$\sum_{k=1}^{n} F_\lambda(k) \leq \sum_{k=1}^{t_{2n}} F_\lambda(k) = \sum_{k=1}^{n} s_{2k},$$

We therefore conclude that

$$\mu_\tau^x\left(\frac{1}{n}\big|\{1 \leq k \leq n : u_\tau(\lambda_k) > l\}\big| > \rho\right) = \mu_\tau^x\left(\sum_{k=1}^{n} F_\lambda(k) > n\rho\right) \leq \mu_\tau^x\left(\sum_{i=1}^{n} s_{2i} > n\rho\right).$$

Now, for each k, Proposition B.4 applied to the function u_τ implies that

(B.7) $$\mu_\tau^x(s_{2i} > k) \leq M_\tau \gamma_0^k.$$

We now employ a trick to *exclude short sojourns*: Let $C' \geq 1$ denote some large threshold that will be determined later. For each $i \geq 1$ we now define

$$s_{2i}' = \begin{cases} 0, & s_{2i} \leq C' \\ s_{2i}, & \text{else} \end{cases}$$

By (B.7) above, we have $\mu_\tau^x(s_{2i} = k) \leq \mu_\tau^x(s_{2i} > k - 1) \leq M_\tau \gamma_0^{k-1}$ for each $i \geq 1$ and $k \geq 1$. This implies that for all $i \geq 1$ we have

$$\mu_\tau^x(s_{2i}' = k) \leq \begin{cases} 0, & 1 \leq k \leq C' \\ M_\tau \gamma_0^{k-1}, & k > C'. \end{cases}$$

The point here is that if $s_{2i} \leq C'$ then λ takes at most C' steps outside of C_l between t_{2i-1} and t_{2i}. By (B.1), this implies that $u_\tau(\lambda_k) \leq l M_\tau^{C'}$ for each of these steps (i.e., for $t_{2i-1} \leq k < t_{2i}$). Therefore, by increasing l to $l' := l M_\tau^{C'}$, the number of steps that λ takes outside of $C_{l'}$ (in its first n steps) is bounded above by $\sum_{i=1}^{n} s_{2i}'$. Therefore, it suffices to find an exponentially small upper bound on $\mu_\tau^x(\sum_{i=1}^{n} s_{2i}' > \rho n)$.

Our choice of $\gamma_0 < \gamma^{1/\rho}$ ensures that we may choose $\theta > 0$ satisfying the inequalities $\ln(\gamma_0) < -\theta < \ln(\gamma)/\rho$. In particular, we have $e^\theta \gamma_0 < 1$. Let $g : \mathsf{P}_\tau^x \to \mathbb{R}$ be given by $g = g_\theta = \exp(\theta \cdot \sum_{i=1}^n s'_{2i})$. Observe that

$$E(g) = \int_{\mathsf{P}_\tau^x} g(\lambda) d\mu_\tau^x(\lambda) \geq \int_{\{g > e^{\theta\rho n}\}} g(\lambda) d\mu_\tau^x(\lambda) \geq \mu_\tau^x(g > e^{\theta\rho n}) \cdot e^{\theta\rho n}.$$

Therefore

$$\mu_\tau^x\left(\sum_{i=1}^n s'_{2i} > \rho n\right) = \mu_\tau^x(g > e^{\theta\rho n}) \leq e^{-\theta\rho n} E(g)$$

$$= e^{-\theta\rho n} \int_{\mathsf{P}_\tau^x} \left(\prod_{i=1}^n e^{\theta s'_{2i}(\lambda)}\right) d\mu_\tau^x(\lambda) = e^{-\theta\rho n} \prod_{i=1}^n \int_{\mathsf{P}_\tau^x} e^{\theta s'_{2i}(\lambda)} d\mu_\tau^x(\lambda),$$

where in the last line we have used the fact that the random variables $s'_1, \ldots, s'_{2n}$ are independent of each other (so that the expected value of the product is equal to the product of the expected values). We now estimate each expected value in the product. For each $i \geq 1$ we have

$$\int_{\mathsf{P}_\tau^x} e^{\theta s'_{2i}(\lambda)} d\mu_\tau^x(\lambda) = \sum_{k=0}^\infty e^{\theta k} \mu_\tau^x(s'_{2i} = k) = e^{\theta \cdot 0} \mu_\tau^x(s'_{2i} = 0) + \sum_{k=\lceil C'\rceil}^\infty e^{\theta k} \mu_\tau^x(s'_{2i} = k)$$

$$\leq 1 + \sum_{k=\lceil C'\rceil}^\infty e^{\theta k} M_\tau \gamma_0^{k-1} \leq 1 + \tfrac{M_\tau}{\gamma_0(1-e^\theta \gamma_0)}(e^\theta \gamma_0)^{C'}.$$

Note that the assumptions on θ, γ, ρ imply $e^{\theta\rho}\gamma > 1$. Therefore, since $e^\theta \gamma_0 < 1$, we may now choose C' sufficiently large so that

$$\Gamma := \left(1 + \tfrac{M_\tau}{\gamma_0(1-e^\theta \gamma_0)}(e^\theta \gamma_0)^{C'}\right) < e^{\theta\rho}\gamma.$$

By the above calculations, we may now conclude that

$$\mu_\tau^x\left(\sum_{i=1}^n s_{2i} > \rho n\right) \leq e^{-\theta\rho n} \prod_{i=1}^n \int_{\mathsf{P}_\tau^x} e^{\theta s'_{2i}(\lambda)} d\mu_\tau^x(\lambda)$$

$$\leq e^{-\theta\rho n}\left(1 + \tfrac{M}{\gamma_0(1-e^\theta \gamma_0)}(e^\theta \gamma_0)^{C'}\right)^n = \left(e^{-\theta\rho}\Gamma\right)^n < \gamma^n.$$

This shows that for any x satisfying $u_\tau(x) \leq l$, the probability that a random walk $\lambda \in \mathsf{P}_\tau^x$ spends more than proportion ρ of its time outside of the cocompact set $C_{l'} = \{y \in \mathcal{T} : u_\tau(y) \leq l M_\tau^{C'}\}$ during its first n steps is at most γ^n. Taking $\epsilon' > 0$ sufficiently small so that $u_\tau(y) \leq l M_\tau^{C'} \implies y \in \mathcal{T}_{\epsilon'}$ completes the proof. $\qquad\square$

References

[1] J.C. Álvarez Paiva, A. Thompson, *Volumes on normed spaces and Finsler spaces*, MSRI Publications 50 (2004).

[2] J.S. Athreya, Quantitative recurrence and large deviations for Teichmüller geodesic flow. Geom. Dedicata 119 (2006), 121–140.

[3] J. Athreya, A. Bufetov, A. Eskin, and M. Mirzakhani, *Lattice point asymptotics and volume growth on Teichmüller space*, Duke Math. J. 161 (2012), no. 6, 1055–1111.

[4] L. A. Santaló, *An affine invariant for convex bodies of n-dimensional space*, Portugaliae Math. 8 (1949), 155–161.

[5] H. Busemann, *Intrinsic area*, Ann. of Math. 48 (1947), 234-267.

[6] M. Duchin, S. Lelièvre, and C. Mooney, *Statistical hyperbolicity in groups*, Alg. and Geom. Top. 12 (2012), 1–18.

[7] A. Eskin and M. Mirzakhani, *Counting closed geodesics in moduli space*, J. Modern Dynamics 5 (2011), no. 1, 71–105

[8] A. Fathi, F. Laudenbach, P. Poenaru, *Travaux de Thurston sur les surfaces*, Asterisque 66-67 (1979).

[9] U. Hamenstädt, *Distance in the curve graph*, arxiv 1104.4001, to appear in Geom. Dedicata.

[10] F. John, *Extremum problems with inequalities as subsidiary conditions*, 187–204 in Studies and essays presented to R. Courant on his 60th birthday, edited by K. O. Friedrichs, O. Neuegebauer, and J. J. Stoker, Interscience, New York, 1948.

[11] R. Kent, C. Leininger, *Shadows of mapping class groups: capturing convex cocompactness* . Geom. Funct. Anal. 18 (2008), 1270-1325.

[12] H. Masur, *On a class of geodesics in Teichmüller space*, Ann. of Math. (2) 102 (1975), no. 2, 205–221.

[13] H. Masur, *Uniquely ergodic quadratic differentials*, Comment. Math. Helvetici 55 255–266 (1980).

[14] H. Masur, *Interval exchange transformations and measured foliations*, Annals of Math. 115 (1982) 169–200.

[15] H. Masur, *Hausdorff dimension of the set of nonergodic foliations of a quadratic differential*, Duke Math. Journal 66 387–442 (1992).

[16] H. Masur, *The Teichmüller flow is Hamiltonian*, Proceedings of the American Mathematical Society 123 (1995) **12** 3739–3747.

[17] H. Masur, and Y. Minsky, *Geometry of the complex of curves. I. Hyperbolicity*, Invent. Math. 138 (1999), no. 1, 103–149.

[18] H. Masur and M. Wolf, *Teichmüller space is not Gromov hyperbolic*, Ann. Acad. Sci. Fenn. Ser. A I Math., 20(2) (1995), 259–267.

[19] Y.N. Minsky, *Extremal length estimates and product regions in Teichmüller space*, Duke Math. J. 83:2(1996), 249–286.

[20] K. Rafi, *A characterization of short curves of a geodesic in Teichmller space* Geom. Topol 9: (2005), 179–202.

[21] K. Rafi, *A combinatorial model for the Teichmüller metric*, Geom. Funct. Anal. 17:3 (2007), 936–959.

[22] K. Rafi, *Hyperbolicity in Teichmüller space*, arxiv 1011.6004.

[23] K. Rafi and S. Schleimer, *Covers and the curve complex*, Geom. Topol. 13 (2009), no. 4, 2141–2162.

[24] H. L. Royden, *Automorphisms and isometries of Teichmüller space*, Advances in the Theory of Riemann Surfaces (Proc. Conf., Stony Brook, N.Y., 1969), Ann. of Math. Studies, No. 66. Princeton Univ. Press, Princeton, N.J., (1971), 369–383.

SPENCER DOWDALL, MATHEMATICS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801
E-mail address: `dowdall@math.uiuc.edu`

MOON DUCHIN, MATHEMATICS, TUFTS UNIVERSITY, MEDFORD, MA 02155
E-mail address: `moon.duchin@tufts.edu`

HOWARD MASUR, MATHEMATICS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637
E-mail address: `masur@math.uchicago.edu`