GEOMETRY OF THE COMPLEX OF CURVES I: HYPERBOLICITY

HOWARD A. MASUR AND YAIR N. MINSKY

1. Introduction

In topology, geometry and complex analysis, one attaches a number of interesting mathematical objects to a surface S. The Teichmüller space $\mathcal{T}(S)$ is the parameter space of conformal (or hyperbolic) structures on S, up to isomorphism isotopic to the identity. The Mapping Class Group $\mathrm{Mod}(S)$ is the group of auto-homeomorphisms of S, up to isotopy. The geometric and group-theoretic properties of these objects are tied to each other via the intrinsic combinatorial topology of S.

In [18], Harvey associated to a surface S a finite-dimensional simplicial complex $\mathcal{C}(S)$, called the *complex of curves*, which was intended to capture some of this combinatorial structure, and in particular to encode the asymptotic geometry of Teichmüller space in analogy with Tits buildings for symmetric spaces. The vertices of Harvey's complex are homotopy classes of simple closed curves in S, and the simplices are collections of curves that can be realized disjointly. This complex was then considered by Harer [16, 17] from a cohomological point of view, and by Ivanov [21, 20, 22] with applications to the structure of $\mathrm{Mod}(S)$ (in particular a new proof of Royden's theorem).

In this paper we begin a study of the intrinsic geometry of $\mathcal{C}(S)$, which can be made into a complete geodesic metric space in a natural way by making each simplex a regular Euclidean simplex of sidelength 1 (see Bridson [6]). Our main result is the following:

Theorem 1.1. (Hyperbolicity) *Let S be an oriented surface of finite type. The curve complex $\mathcal{C}(S)$ is a δ-hyperbolic metric space, where δ depends on S. Except when S is a sphere with 3 or fewer punctures, $\mathcal{C}(S)$ has infinite diameter.*

(See §2.1 for a definition of δ-hyperbolicity.)

We remark that in a few sporadic cases our definition of $\mathcal{C}(S)$ varies slightly from the original; see §2.2. Note also that we can just as well consider the 1-skeleton $\mathcal{C}_1(S)$ rather than the whole complex: δ-hyperbolicity is a quasi-isometry invariant, and $\mathcal{C}(S)$ is evidently quasi-isometric to its 1-skeleton.

Harer showed [16], in the non-sporadic cases, that $\mathcal{C}(S)$ is homotopy equivalent to a wedge of spheres of dimension greater than 1, and in particular is simply-connected but not contractible. It follows that $\mathcal{C}(S)$ cannot be given a $\mathrm{CAT}(\kappa)$ metric for any $\kappa \le 0$ (see e.g. Ballmann [3, §I.4]). This rules out the most simple way to prove δ-hyperbolicity by a local argument. One might still ask if $\mathcal{C}(S)$ can be embedded quasi-isometrically in a $\mathrm{CAT}(\kappa)$ space for $\kappa \le 0$. If S has boundary then $\mathcal{C}(S)$ embeds in a related *arc complex*, whose vertices are allowed to be arcs with endpoints on the boundary, and which Harer proved in [16] is contractible. It is thus an interesting question whether this complex admits a $\mathrm{CAT}(\kappa)$ metric for $\kappa \le 0$.

Date: January 2, 1998.

The first author was partially supported by NSF grant #DMS 9201321. The second author was partially supported by an NSF postdoctoral fellowship and a fellowship from the Alfred P. Sloan Foundation.

Stony Brook IMS Preprint #1996/11
October 1996
Revised version: January 1998

Theorem 1.1 is motivated in part by the need to understand the extent of an important but incomplete analogy between the geometry of the Teichmüller space and that of complete, negatively curved manifolds, or more generally of δ-hyperbolic spaces. There are many senses in which this analogy holds, and it was exploited, for example, by Bers [4], Kerckhoff [25], and Wolpert [41]. On the other hand, Masur [27] showed that (except for the simplest cases) the Teichmüller metric on $\mathcal{T}(S)$ cannot be negatively curved in a local sense, and more recently Masur-Wolf [31] showed that it is not δ-hyperbolic. The Weil-Petersson metric on $\mathcal{T}(S)$ has negative sectional curvatures, however they are not bounded away from zero [28, 42].

The failure of δ-hyperbolicity in $\mathcal{T}(S)$ is closely related to the presence of infinite diameter regions where the metric on $\mathcal{T}(S)$ is nearly a product (let us consider from now on only the Teichmüller metric on $\mathcal{T}(S)$). Fixing a small $\epsilon_0 > 0$, let

$$H_\alpha = \{x \in \mathcal{T}(S) : Ext_x(\alpha) \leq \epsilon_0\}$$

denote the region in $\mathcal{T}(S)$ where a simple closed curve α has small extremal length (see Section 2 for definitions). Then (see Minsky [34]) the Teichmüller metric in this region is approximated by a product of infinite-diameter metric spaces, and so cannot be δ-hyperbolic.

As a consequence of the Collar Lemma (see e.g. [23, 7]), when ϵ_0 is sufficiently small the intersection pattern of the family $\{H_\alpha\}$ is exactly encoded by the complex $\mathcal{C}(S)$ (it is the *nerve* of this family). Thus, one interpretation of our main theorem is that the regions $\{H_\alpha\}$ are the only obstructions to hyperbolicity, and once their internal structure is ignored, the way in which they fit together is hyperbolic.

This can be made precise by Farb's notion of *relative hyperbolicity* [11], and in Section 7 we will prove:

Theorem 1.2. (Relative Hyperbolicity 1) *The Teichmüller space $\mathcal{T}(S)$ is relatively hyperbolic with respect to the family of regions $\{H_\alpha\}$.*

A similar discussion can be carried out for the mapping class group. A group is *word hyperbolic* if its Cayley graph is δ-hyperbolic. It is known that a group acting by isometries on a δ-hyperbolic space with finite point stabilizers and compact quotient must itself be word-hyperbolic, and it is plain that $\mathrm{Mod}(S)$ acts isometrically on $\mathcal{C}(S)$, with compact quotient. Nevertheless, $\mathrm{Mod}(S)$ is known not to be word-hyperbolic for all but the simplest cases, because it contains high-rank abelian subgroups. This is not a contradiction, because the action on $\mathcal{C}(S)$ has infinite point stabilizers.

Indeed, abelian subgroups in $\mathrm{Mod}(S)$ are generated by elements that stabilize disjoint subsurfaces, and in particular their boundary curves, and hence are "invisible" from the point of view of coarse geometry of the complex. One can formalize this intuition as we did with Teichmüller space by considering subgroups of $\mathrm{Mod}(S)$ fixing certain curves, and their cosets. In Section 7 we will carry this out and prove:

Theorem 1.3. (Relative Hyperbolicity 2) *The group $\mathrm{Mod}(S)$ is relatively hyperbolic with respect to left-cosets of a finite collection of stabilizers of curves.*

Farb shows in [11] that relative-hyperbolicity results such as these are useful in converting information about subgroups (such as automaticity) to information about the full groups. Although his work does not apply directly to our situation, it is nonetheless possible to use the results of this paper as the first step in an inductive analysis of the structure of the Mapping Class Group. Such an analysis will be carried out in [30].

We remark finally that although our main theorem has essentially a topological statement, the proof we have found uses Teichmüller geometry in an essential way. It would be very interesting to find a purely combinatorial proof. In particular, it would be nice to have an effective estimate of the constant δ, which our proof does not provide since it depends on bounds obtained from a compactness argument in the Moduli space.

2. Outline of the Proof

In this section, after describing some necessary background, we will give an outline of the proof of the Hyperbolicity Theorem 1.1, which reduces it to a number of assertions. These assertions will then be proven in sections 3 through 6.

2.1. **Hyperbolicity.** A geodesic metric space X is a path-connected metric space in which any two points x, y are connected by an isometric image of an interval in the real line, called a geodesic and denoted $[xy]$ (we use this notation although $[xy]$ is not required to be unique).

We say that X satisfies the *thin triangles condition* if there exists some $\delta \geq 0$ such that, for any x, y and $z \in X$ the geodesic $[xz]$ is contained in a δ-neighborhood of $[xy] \cup [yz]$. This is one of several equivalent conditions for X to be δ-*hyperbolic* in the sense of Gromov, or *negatively curved* in the sense of Cannon. (We remark that there are formulations of hyperbolicity that do not require X to be a geodesic space, but we will not be concerned with them here. See Cannon [8], Gromov [15] and also [5, 10, 14].)

Important examples of hyperbolic spaces are the classical hyperbolic space $\mathbf{H}^n$, all simplicial trees (here $\delta = 0$), and Cayley graphs of fundamental groups of closed negatively curved manifolds.

We note also that every finite-diameter space is trivially δ-hyperbolic with δ equal to the diameter, which is the reason we must check that the complex of curves has infinite diameter.

2.2. **The complex of curves.** Let S be a closed surface of genus g with p punctures. Except in the sporadic cases mentioned below, define a complex $\mathcal{C}(S)$ as follows: k-simplices of $\mathcal{C}(S)$ are $(k+1)$-tuples $\{\gamma_0, \gamma_1, \ldots, \gamma_k\}$ of distinct non-trivial homotopy classes of simple, non-peripheral closed curves, which can be realized disjointly. This complex is obviously finite-dimensional by an Euler characteristic argument, and is typically locally infinite.

Sporadic cases. In a number of cases $\mathcal{C}(S)$ (and hence our main theorem) is either trivial or already well-understood. When S is a sphere ($g = 0$) with $p \leq 3$ punctures, the complex is empty. In this case we can say Theorem 1.1 holds vacuously. When $g = 0$ and $p = 4$, or $g = 1$ and $p \leq 1$, Harvey's complex has no edges, and is just an infinite set of vertices. In these cases it is useful to alter the definition slightly, so that edges are placed between vertices corresponding to curves of smallest possible intersection number (1 for the tori, 2 for the sphere). When this is done, we obtain the familiar *Farey graph*, for which Theorem 1.1 is fairly easy to prove. See [35] for an exposition of this case.

For the remainder of the paper we exclude the surfaces with $g = 0, p \leq 4$ and $g = 1, p \leq 1$, which we call *sporadic*.

In all other cases, the dimension of the complex is easily computed to be $3g - 4 + p$, which in particular is at least 1. Letting $\mathcal{C}_k$ denote the k-skeleton of $\mathcal{C}$, we focus on the graph $\mathcal{C}_1$. We turn $\mathcal{C}_1$ into a metric space by specifying that each edge has length 1, and we denote by $d_\mathcal{C}$ the distance function obtained by taking shortest paths. Note also that $\mathcal{C}_1$ is a geodesic metric space.

For $\alpha, \beta \in \mathcal{C}_0(S)$, let $i(\alpha, \beta)$ denote the geometric intersection number of α with β on S, which is equal to the number of transverse intersections of their geodesic representatives in a hyperbolic metric on S.

Lemma 2.1. *If S is not sporadic, $\mathcal{C}_1$ is connected. Moreover for any two curves α, β, $d_{\mathcal{C}}(\alpha,\beta) \le 2i(\alpha,\beta)+1$.*

Remark. In fact for large $d_{\mathcal{C}}$ a better estimate is that $i(\alpha,\beta)$ is at least exponential in $d_{\mathcal{C}}$, as we shall see in Section 3.

Proof. Assume that α and β are realized with minimal intersection number If $i(\alpha,\beta)=1$ then a regular neighborhood of $\alpha\cup\beta$ is a punctured torus whose boundary γ must be nontrivial and nonperipheral since the torus and punctured torus are excluded. Since γ is disjoint from both α and β, $d(\alpha,\beta)=2$.

For $i(\alpha,\beta)\ge 2$, fixing two points of $\alpha\cap\beta$ adjacent in α there are two distinct ways to do surgery on these points, replacing a segment of β with the segment of α between them, producing homotopically nontrivial simple curves β_1,β_2 such that $i(\alpha,\beta_j)\le i(\alpha,\beta)-1$. If the two intersections agree in orientation then $i(\beta,\beta_j)=1$, and neither β_j can be peripheral (if it bounds a punctured disk then α must enter it and has a non-essential intersection with β). Thus $d(\alpha,\beta)\le 2+d(\alpha,\beta_j)$ and we are done by induction.

If the two intersections have opposite signs then actually $i(\alpha,\beta_j)\le i(\alpha,\beta)-2$, and $i(\beta,\beta_j)=0$ for $j=1,2$. Thus if at least one β_j is nonperipheral, we again apply induction (and get a better estimate than above). If both β_1,β_2 are peripheral then β must bound a twice punctured disk on the side containing the α segment between the intersections. Thus consider a segment of α between intersections, which is adjacent to this one. If it also falls into the last category then β bounds a twice punctured disk on its other side too, and S must be a 4-times punctured sphere, which has been excluded. □

2.3. **Teichmüller space.** An analytically finite conformal structure on S is an identification of S with a closed Riemann surface minus a finite number of points. Let $\mathcal{T}(S)$ denote the Teichmüller space of analytically finite conformal structures on S, modulo conformal isomorphism isotopic to the identity.

Given an element $x\in\mathcal{T}(S)$ and a simple closed curve α in S, we recall that the *extremal length* $Ext_x(\alpha)$ is the reciprocal of the largest conformal modulus of an embedded annulus in S homotopic to α. We remark also that an alternate definition is $Ext_x(\alpha)=\sup_\sigma |\alpha^*|_\sigma^2$ where σ ranges over conformal metrics of area 1 on (S,x), and $|\alpha^*|_\sigma$ denotes σ-length of a shortest representative of α. (See e.g. Ahlfors [1].)

The Teichmüller metric $d_{\mathcal{T}}$ on $\mathcal{T}(S)$ can be defined in terms of maps with minimal quasiconformal dilatation, but for us it will be useful to note Kerckhoff's characterization [24]:

$$d_{\mathcal{T}}(x,y)=\tfrac{1}{2}\log\sup_{\alpha\in\mathcal{C}_0(S)}\frac{Ext_y(\alpha)}{Ext_x(\alpha)}. \tag{2.1}$$

A holomorphic quadratic differential q on a Riemann surface is a tensor of the form $\varphi(z)dz^2$ in local coordinates, with φ holomorphic. Away from zeroes, a coordinate ζ can be chosen so that $q=d\zeta^2$, which determines a Euclidean metric $|d\zeta^2|$ together with a pair of orthogonal foliations parallel to the real and imaginary axes in the ζ plane. These are well-defined globally and are called the *horizontal* and *vertical* foliations, respectively. (See Gardiner [13] or Strebel [39].)

Geodesics in $\mathcal{T}(S)$ are determined by quadratic differentials. Given q holomorphic for some $x\in\mathcal{T}(S)$, for any $t\in\mathbf{R}$ we consider the conformal structures obtained by scaling the horizontal foliation of q by a factor of e^t, and the vertical by e^{-t}. The resulting family, which we write $L_q(t)$, is a geodesic parametrized by arclength.

For a closed curve or arc α in S, denote by $|\alpha|_q$ its length in the q metric. Let $|\alpha|_{q,h}$ and $|\alpha|_{q,v}$ denote its horizontal and vertical lengths, respectively, by which we mean the total lengths of the (locally defined) projections of α to the horizontal and vertical directions of q.

Finally we note that the variation of horizontal and vertical lengths is given by

$$|\alpha|_{q_t,h} = |\alpha|_{q_0,h}e^t \tag{2.2}$$

and

$$|\beta|_{q_t,v} = |\beta|_{q_0,v}e^{-t}. \tag{2.3}$$

2.4. **The proof of the Hyperbolicity Theorem.** One way to prove hyperbolicity is to find a class of paths with the following contraction property:

Definition 2.2. *Let X be a metric space. Say that a path $\gamma : I \to X$ (where $I \subset \mathbf{R}$ is some interval, possibly infinite) has the* contraction property *if there exists $\pi : X \to I$ and constants $a, b, c > 0$ such that:*

1. *For any $t \in I$,* $\mathrm{diam}(\gamma([t, \pi(\gamma(t))])) \le c$
2. *If $d(x, y) \le 1$ then* $\mathrm{diam}\,\gamma([\pi(x), \pi(y)]) \le c$.
3. *If $d(x, \gamma(\pi(x))) \ge a$ and $d(x, y) \le bd(x, \gamma(\pi(x)))$, then*

$$\mathrm{diam}\,\gamma[\pi(x), \pi(y)] \le c.$$

(Here for $s, t \in \mathbf{R}$ we take $[s, t]$ to mean the interval with endpoints s, t regardless of order.)

One should think of this property in analogy with closest-point projection to a geodesic in $\mathbf{H}^n$. Condition (1) is a coarsening of the requirement that points in $\gamma(I)$ be fixed. Condition (2) states that the projection is coarsely Lipschitz. Condition (3) is the most important, stating that the map is, in the large, strongly contracting for points far away from their images in $\gamma(I)$. Note that this holds in $\mathbf{H}^n$ for $b = 1$.

Note also that we give π as a map to the parameter interval I rather than its image, in order to avoid requiring anything about the speed of parametrization of γ: for example γ is allowed to be constant for long intervals, and on the other hand it need not be continuous.

We say that a family Γ of paths has the contraction property if every $\gamma \in \Gamma$ has the contraction property, with respect to a uniform $a, b, c > 0$.

Call a family of paths *coarsely transitive* if there exists $D \ge 0$ such that for any x and y with $d(x, y) \ge D$ there is a path in the family joining x to y. In section 6 we will prove the following theorem, which is probably well-known.

Theorem 2.3. *If a geodesic metric space X has a coarsely transitive path family Γ with the contraction property then X is hyperbolic. Furthermore, the paths in Γ are uniformly quasi-geodesic.*

(See §6 for the definition of quasi-geodesic in this context).

Our family of paths will be constructed using Teichmüller geodesics, in the following manner. There is a natural map Φ from $\mathcal{T}(S)$ to finite subsets of $\mathcal{C}(S)$, assigning to any $x \in \mathcal{T}(S)$ the set of curves of shortest Ext_x (extremal length is convenient for us, though hyperbolic will do as well). A geodesic in $\mathcal{T}(S)$ traces out, via Φ, a path in $\mathcal{C}(S)$ up to some bounded ambiguity.

That is, let q be a quadratic differential on a Riemann surface x and let $L_q : \mathbf{R} \to \mathcal{T}(S)$ be the corresponding Teichmüller geodesic (parametrized by arclength). Let a map

$$F \equiv F_q : \mathbf{R} \to \mathcal{C}(S)$$

be defined by assigning to t one of the curves of $\Phi(L_q(t))$. The actual choices will not matter, as $\Phi(x)$ has uniformly bounded diameter:

Lemma 2.4. *There exists $c = c(S)$ such that $\operatorname{diam}_{\mathcal{C}} \Phi(x) \leq c$ for all $x \in \mathcal{T}(S)$.*

Proof. There exists $e_0(S)$ such that the shortest nonperipheral curve on (S, x) has extremal length at most e_0. Thus Lemma 2.5 below immediately bounds the distance between any two shortest curves, by $2e_0 + 1$.

Note in fact that there exists ϵ_0 such that if x has a curve α of extremal length at most ϵ_0 then any curve intersecting α has extremal length greater than ϵ_0. In this case the diameter of $\Phi(x)$ is at most 1. □

Lemma 2.5. *For $\alpha, \beta \in \mathcal{C}_0(S)$, if $Ext_x(\alpha) \leq E$ and $Ext_x(\beta) \leq E$ for some conformal structure x on S, then $d_{\mathcal{C}}(\alpha, \beta) \leq 2E + 1$.*

Proof. It is an elementary fact (see e.g. [33]) that $Ext_x(\alpha) Ext_x(\beta) \geq i(\alpha, \beta)^2$. Thus the assumption of the lemma gives $i(\alpha, \beta) \leq E$. Now by Lemma 2.1, $d_{\mathcal{C}}(\alpha, \beta) \leq 2E + 1$. □

If q has a closed vertical leaf then there is a collection of (up to homotopy) disjoint vertical curves whose extremal lengths go to 0 as $t \to \infty$. In this case choose a fixed one of these to be the value of F as $t \to \infty$, and let this also be denoted by $F(\infty)$. Similarly define $F(-\infty)$ if there are horizontal curves.

The projection for F will be defined using the notion of *balance*. Recalling the notation of §2.3, we say that β is balanced with respect to q if $|\beta^*|_{q,h} = |\beta^*|_{q,v}$, where β^* is a q-geodesic representative (it may be necessary for β^* to go through punctures – see §4.1).

We note that β^* is also geodesic with respect to any q_t. Since $|\cdot|_{q_t,h}$ and $|\cdot|_{q_t,v}$ vary like e^t and e^{-t} (by (2.2) and (2.3)), if β^* is not entirely vertical or horizontal with respect to q there is a unique t for which β is balanced, and this is also the minimum of the quantity $|\beta^*|_{q_t,h} + |\beta^*|_{q_t,v}$. We observe also that, since the q-length of β^* is estimated by

$$\frac{1}{\sqrt{2}}(|\beta^*|_{q,h} + |\beta^*|_{q,v}) \leq |\beta^*|_q \leq |\beta^*|_{q,h} + |\beta^*|_{q,v},$$

the minimum of $|\beta^*|_{q_t}$ also occurs within bounded distance (in fact $\frac{1}{2}\cosh^{-1}\sqrt{2}$) of the balance point. (Compare with the projection used in [36]).

Let $\mathcal{C}_b = \mathcal{C}_b(q)$ denote the set of simple closed curves that are not entirely horizontal or vertical for q. We define $\pi = \pi_q : \mathcal{C}_0 \to \mathbf{R}$ as follows: for $\beta \in \mathcal{C}_b$ let $\pi(\beta)$ be the unique t for which β is balanced for q_t. For $\beta \in \mathcal{C} \setminus \mathcal{C}_b$ let $\pi(\beta)$ be $+\infty$ if β is vertical, and $-\infty$ if β is horizontal. (As above, in this case $F(\pm\infty)$ makes sense).

Suppose now $d(\alpha, \beta) \geq 3$. Then α and β *fill* S, in that there is no γ disjoint from both. There is therefore a quadratic differential q whose nonsingular vertical leaves are homotopic to α and whose nonsingular horizontal leaves are homotopic to β (see [12, Exposé 13]). Then $F_q(+\infty) = \alpha$ and $F_q(-\infty) = \beta$. This shows that the family $\{F_q\}$ is coarsely transitive.

Hyperbolicity will therefore be a consequence of Theorem 2.3 and the following:

Theorem 2.6. (Projection Theorem) *The path family $\{F_q\}$ satisfies the contraction property with the projections π_q defined above.*

The proof of this theorem will be given in section 5.

We will begin in Section 3 by developing tools for controlling distances between curves in $\mathcal{C}_0(S)$. Using Thurston's train-track coordinates, we will analyze a covering of $\mathcal{C}_0(S)$ by a family of polyhedra which have the property that a point contained in a deeply nested sequence of polyhedra will be a definite distance from any point outside the outermost polyhedron (Lemma 3.2). A partial converse to this will be the Nesting Lemma 3.7, which given two distant curves will allow us to construct a nested sequence of polyhedra separating them. We will apply these tools to prove Lemma 3.12, which relates intersection numbers to distance in $\mathcal{C}(S)$ in a way which can be directly applied in Section 5.

Proposition 3.6 in Section 3.3 will establish the infinite-diameter claim in the Hyperbolicity Theorem 1.1.

3. The nested train-track argument

3.1. **Train-tracks.** We refer to Penner-Harer [38] for a thorough treatment of train-tracks, recalling here some of the terminology. A train-track on a surface S is an embedded 1-complex τ satisfying the following properties. Each edge (called a branch) is a smooth path with well-defined tangent vectors at the endpoints, and at any vertex (called a switch) the incident edges are mutually tangent. The tangent vector at the switch pointing toward the interior of an edge can have two possible directions, and this divides the ends of edges at the switch into two sets, neither of which is permitted to be empty. Call them "incoming" and "outgoing". The valence of each switch is at least 3, except possibly for one bivalent switch in a closed curve component. Finally, we require that the components of $S \setminus \tau$ have negative generalized Euler characteristic, in this sense: For a surface R whose boundary consists of smooth arcs meeting at cusps, define $\chi'(R)$ to be the Euler characteristic $\chi(R)$ minus $1/2$ for every outward-pointing cusp (internal angle 0), plus $1/2$ for every inward-pointing cusp (internal angle 2π). For the train track complementary regions all cusps are outward, so that the condition $\chi'(R) < 0$ excludes annuli, once-punctured disks with smooth boundary, or unpunctured disks with 0, 1 or 2 cusps at the boundary. We will usually consider isotopic train-tracks to be the same.

A *train route* is a non-degenerate smooth path in τ; in particular it traverses a switch only by passing from incoming to outgoing edge (or vice versa). A *transverse measure* on τ is a non-negative function μ on the branches satisfying the switch condition: For any switch the sums of μ over incoming and outgoing branches are equal. A closed train-route induces the counting measure on τ.

A train-track is *recurrent* if every branch is contained in a closed train route, or equivalently if there is a transverse measure which is positive on every branch.

Fixing a reference hyperbolic metric on S, a *geodesic lamination* in S is a closed set foliated by geodesics (see e.g. [25, 19]). A geodesic lamination is *measured* if it supports a measure on arcs transverse to its leaves, which is invariant under isotopies preserving the leaves. The space of all compactly supported measured geodesic laminations on S, with suitable topology, is known as $\mathcal{ML}(S)$, and we note that different choices of reference metric on S yield equivalent spaces. A geodesic lamination λ is *carried* on τ if there is a homotopy of S taking λ to a set of train routes. In such a case λ induces a transverse measure on τ, which in turn uniquely determines λ. The set of measures on τ gives local coordinates on $\mathcal{ML}(S)$, and in fact $\mathcal{ML}(S)$ is a manifold (homeomorphic to a Euclidean space).

For a recurrent train-track τ, let $P(\tau)$ denote the polyhedron of measures supported on τ. We will blur the distinction between $P(\tau)$ as a subset of $\mathcal{ML}(S)$, and as a subset of the space $\mathbf{R}_+^{\mathcal{B}}$ of non-negative functions on the branch set $\mathcal{B}$ of τ.

We note that $P(\tau)$ is preserved by scaling, so it is a cone on a compact polyhedron in projective space. However we will need to consider actual measures in $P(\tau)$ rather than projective classes.

By $int(P(\sigma))$ we will mean the set of weights on σ which are positive on every branch. Note that unless the dimension $\dim(P(\sigma))$ is maximal, this is different from the interior of $P(\sigma)$ as a subset of $\mathcal{ML}(S)$, which is empty.

We write $\sigma < \tau$ if σ is a *subtrack* of τ; that is, σ is a train track which is a subset of τ. We also say that τ is an *extension* of σ in this case. We write $\sigma \prec \tau$ if σ is *carried* on τ, by which we mean that there is a homotopy of S such that every train route on σ is taken to a train route on τ. It is easy to see that $\sigma < \tau$ is equivalent to $P(\sigma)$ being a subface of $P(\tau)$, and $\sigma \prec \tau$ is equivalent to $P(\sigma) \subseteq P(\tau)$.

Say that σ *fills* τ if $\sigma \prec \tau$ and $int(P(\sigma)) \subseteq int(P(\tau))$. When both tracks are recurrent this is equivalent to saying that every branch of τ is traversed by some branch of σ. Similarly, a curve α fills τ if $\alpha \prec \tau$ and traverses every branch of τ.

Call a train-track τ *large* if all the components of $S \setminus \tau$ are polygons or once-punctured polygons. We will also say that $P(\tau)$ is large in such a situation.

We say that σ is *maximal* if it is not a proper subtrack of any other track. This is equivalent to saying that all complementary regions of σ are triangles or punctured monogons. Also, except in the case of the punctured torus, it is equivalent to $\dim(P(\tau)) = \dim(\mathcal{ML}(S))$. In any case, maximal implies large.

A *vertex cycle* of τ is a non-negative measure on τ which is an extreme point of $P(\tau)$. That is, its projective class is a vertex of the projectivized polyhedron. A vertex μ is always rational, i.e. $\mu(b)/\mu(b') \in \mathbf{Q}$ for $b, b' \in \mathcal{B}$. This is because an irrational μ can be approximated by a rational μ' with the same support, and then the decomposition $\mu = \epsilon\mu' + (\mu - \epsilon\mu')$, for ϵ sufficiently small, contradicts the assumption that μ is an extreme point. Furthermore, up to scaling, a vertex cycle can always be realized by the counting measure on a single, simple closed curve: the alternative is a union of nonparallel curves, and this again gives a nontrivial decomposition of the measure. We will always assume that a vertex is of this form.

A track τ is *transversely recurrent* if every branch of τ is crossed by some simple curve α intersecting τ transversely and *efficiently* – that is, so that $\alpha \cup \tau$ has no bigon complementary components. For us it will only be important to note the following equivalent geometric property, which is established in Theorem 1.4.3 of [38]: τ is transversely recurrent if and only if for any L (large) and ϵ (small) there is a complete finite-area hyperbolic metric on S in which τ can be realized so that all edges have length at least L and curvatures at most ϵ (including at the switches).

Furthermore, we record (Lemma 1.3.3 of [38]) that if τ is transversely recurrent and τ' is a subtrack of τ or is carried in τ, then τ' is also transversely recurrent.

We call a track *birecurrent* if it is both recurrent and transversely recurrent.

Splitting. Let τ be a generic train-track (all switches are trivalent). A *splitting* move is one of the three elementary moves on a local configuration, as shown in figure 1. The three splits are called a left split, a collision, and a right split, and the resulting tracks in each case are carried by τ. If we place a positive measure μ on τ and use the labelling as in the figure, we note that a positive measure is induced on the right split track if $\mu(a) > \mu(c)$, on the left split track if $\mu(a) < \mu(c)$, and on the collision track if $\mu(a) = \mu(c)$. We call a a *winner* of the splitting operation if $\mu(a) > \mu(c)$ (note that d is then also a winner).

Any measured lamination β carried on τ determines a sequence of possible splittings by the rule in the previous paragraph, and all the resulting train-tracks carry β. Note that if β fills τ then it will continue to fill the split tracks, and in particular they will all be recurrent. This process can continue as long as the split track is not a simple closed curve. (We must check that for any recurrent track that is not a simple curve there is a "splittable" edge, that is one which is in the configuration of figure 1: simply consider any transverse measure whose support is all of τ and take an edge of maximal weight. See [26]).

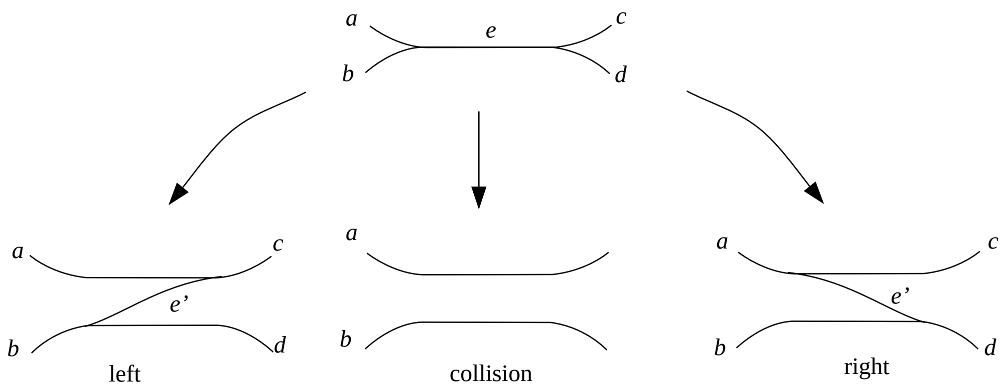

FIGURE 1. The three ways to split through an edge.

When β is a simple closed curve we will usually terminate the sequence as soon as we reach a track for which β is a vertex.

Note also that if σ is a right or left splitting of τ then $P(\sigma)$ and $P(\tau)$ must share at least one vertex. To see this, note that $P(\sigma)$ is one of the pieces obtained by cutting $P(\tau)$ by a hyperplane. Such a subset always contains a vertex of the original polyhedron. In the case of a collision splitting, we at least see that σ is a subtrack of a track that shares a vertex with τ.

Finally, we note that when τ is not generic, each switch can be slightly perturbed ("combing" in [38]) to yield a generic track carrying τ which carries the same set of laminations.

A basic observation. Although it is relatively easy to understand geometrically when pairs of curves are a distance at most 2 in $\mathcal{C}_1$, larger distances are more subtle to detect. This entire section is motivated by the observation that, if α and β are disjoint curves (distance 1 in $\mathcal{C}_1$) and α is carried on a maximal train-track σ in such a way that it passes through every branch, then β is also carried on σ. In other words

$$\mathcal{N}_1(int(P(\sigma))) \subset P(\sigma), \tag{3.1}$$

where $\mathcal{N}_1$ denotes a radius 1 neighborhood in $\mathcal{C}_1$. This implies inductively that if $\tau_j, j = 0, \ldots, n$ is a sequence of maximal tracks such that $P(\tau_j) \subset int(P(\tau_{j-1}))$ and $\beta_j, j \geq 1$ is a sequence in $\mathcal{C}_1$ such that β_j is in $int(P(\tau_{j-1}))$ but not in $P(\tau_j)$, then

$$d_{\mathcal{C}_1}(\beta_1, \beta_j) \geq j.$$

The issue is more subtle when the τ_j are not maximal, or equivalently if they are maximal but the β_j are carried on a face. This leads to a discussion of diagonal extensions and Lemma 3.4 which generalizes the above inequality. A partial converse is given in Lemma 3.7. These two Lemmas are then applied to prove Lemma 3.12.

Diagonal extensions. Let σ be a large track. A *diagonal extension* of σ is a track κ such that $\sigma < \kappa$ and every branch of $\kappa \setminus \sigma$ is a *diagonal* of σ: that is, its endpoints terminate in corner of a complementary region of σ. It is easy to see that if σ is transversely recurrent then so is any diagonal extension – after realizing σ with long edges with nearly zero curvature, its complementary regions are nearly convex and we can make the diagonals nearly geodesic too.

Let $E(\sigma)$ denote the set of all recurrent diagonal extensions of σ. Note that it is a finite set, and let $PE(\sigma)$ denote $\bigcup_{\kappa\in E(\sigma)} P(\kappa)$.

Further, let us define $N(\tau)$ to be the union of $E(\sigma)$ over all large recurrent subtracks $\sigma < \tau$. Define $PN(\tau) = \bigcup_{\kappa \in N(\tau)} P(\kappa)$. In some sense this should be thought of as a "neighborhood" of $P(\tau)$; compare Lemma 3.4 and (3.2).

Let $int(PE(\sigma))$ denote the set of measures $\mu \in PE(\sigma)$ which are positive on every branch of σ. We also define $int(PN(\tau)) = \bigcup_\kappa int(PE(\kappa))$, where κ varies over the large recurrent subtracks of τ.

If a lamination $\mu \in \mathcal{ML}(S)$ has complementary regions which are ideal polygons or once-punctured ideal polygons, then an ϵ-neighborhood of μ, for ϵ sufficiently small, gives rise to a train track σ (see Penner-Harer [38]) which must be large, and μ puts positive weight on every branch of σ. One of the properties of the measure topology on $\mathcal{ML}(S)$ is then that in a small enough neighborhood of μ all laminations are carried on some diagonal extension of σ, and in particular in $int(PE(\sigma))$. A more quantitative form of this idea is the following sufficient condition for containment in $int(PE(\sigma))$.

Lemma 3.1. *There exists $\delta > 0$ (depending only on S) for which the following holds. Let $\sigma < \tau$ where σ is a large track. If $\mu \in P(\tau)$ and, for every branch b of $\tau \setminus \sigma$ and b' of σ, $\mu(b) < \delta\mu(b')$, then σ is recurrent and $\mu \in int(PE(\sigma))$.*

Proof. Whenever there are branches of $\tau \setminus \sigma$ which meet an edge e of the boundary of a complementary region of σ at other than a corner point, the situation can be simplified by either a *split* or a *shift*. First, there may be a splitting move involving these branches which replaces σ by an equivalent track (also called σ). Such a move either reduces the number of edges of $\tau \setminus \sigma$ incident to σ, or moves one of them closer to a corner (see figure 2). These are the only two possibilities.

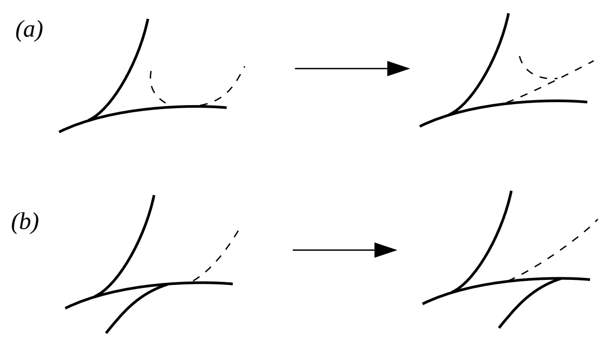

FIGURE 2. Solid edges are in σ, dotted edges in $\tau \setminus \sigma$. In (a), the splitting reduces the number of dotted edges incident to σ. In (b), the splitting brings a dotted edge closer to a corner.

If a branch of $\tau \setminus \sigma$ is facing "toward" a corner of a complementary region, and separated from it only by non-splittable edges, we can perform a *shift move* (see figure 3) which takes this branch to the corner without affecting the set of measures carried on the track. Thus any sequence of such splitting and shifting moves must terminate after a bounded number of steps in a track $\tau' \in E(\sigma)$.

The measure μ will be carried on such a τ' provided it is consistent with the sequence of moves. That is, whenever a splitting is determined by a comparison between a branch b of $\tau \setminus \sigma$ and a branch c of σ, the branch of σ must win (that is, $\mu(b) < \mu(c)$). After such a splitting, there is a branch of σ with measure $\mu(c) - \mu(b)$. Thus, if k is the bound on the number of splittings that take place, a sufficient condition that all the splittings are consistent with μ is $\min_\sigma \mu > (k+1)\max_{\tau\setminus\sigma} \mu$.

Therefore, setting $\delta = 1/(k+1)$, we guarantee that $\mu \in PE(\sigma)$, and furthermore that μ puts positive measure on each branch of σ after the splitting, so that $\mu \in int(PE(\sigma))$.

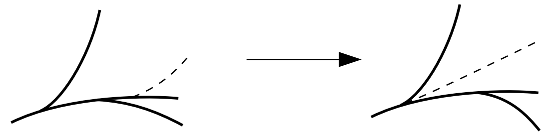

FIGURE 3. Shifting an edge of $\tau \setminus \sigma$ into a corner

Finally, we must show that σ is recurrent. It is an easy fact of linear algebra that, if there is an assignment of positive weights on the branches of σ such that at every switch the difference of incoming and outgoing weights is less than a fixed constant δ_1 times the minimum weight (where δ_1 depends just on the surface S), then these weights can be perturbed to positive weights satisfying the switch conditions, and hence σ is recurrent. It follows that, with sufficiently small δ, the measure μ restricted to σ gives such a set of weights. □

The next two lemmas show that, when tracks are nested, their diagonal extensions are nested in a suitable sense, and the way in which the diagonal branches cover each other is controlled.

Lemma 3.2. *Let σ and τ be large recurrent tracks, and suppose $\sigma \prec \tau$. If σ fills τ, then $PE(\sigma) \subseteq PE(\tau)$. Even if σ does not fill τ, we have $PN(\sigma) \subseteq PN(\tau)$.*

Proof. We may thicken τ slightly to get a regular neighborhood τ_ϵ, which can be foliated by short arcs called "ties" transverse to τ. Then σ can be embedded in τ_ϵ so that it is transverse to the ties.

The assumption that σ fills τ implies that every edge of τ is traversed by some edge of σ, and thus σ crosses every tie. Any component D of $S \setminus \sigma$, which is a polygon or a once-punctured polygon, must have some subset F foliated by ties. F consists of a neighborhood of the boundary and bands joining different boundary edges. Each component of $D \setminus F$ is isotopic to a component of $S \setminus \tau$, and the quotient of D obtained by identifying each tie to a point can be identified with some union of complementary regions of τ, joined by train routes in τ. Any diagonal edge e in D joining two corners of D may therefore be put in minimal position with respect to the ties (so that e meets ties transversely, and no disks are bounded by a segment of e and a tie), and hence gives rise to a train route through the union of τ with some diagonal edges.

It follows that any diagonal extension of σ can be carried by a diagonal extension of τ, and hence $PE(\sigma) \subset PE(\tau)$.

Now in general, if κ is a large recurrent subtrack of σ, let ρ be the smallest subtrack of τ carrying κ. Note that ρ is necessarily large and recurrent, and κ fills ρ. Thus the same argument applies to the faces, and we can conclude $PN(\sigma) \subseteq PN(\tau)$ □

Lemma 3.3. *Let $\sigma \prec \tau$ where σ is a large recurrent track, and let $\sigma' \in E(\sigma)$, $\tau' \in E(\tau)$ such that $\sigma' \prec \tau'$. Then any branch b of $\tau' \setminus \tau$ is traversed with bounded degree m_0 by branches of σ'. The number m_0 depends only on S.*

Proof. It will suffice to show that no branch of σ' passes through a branch of $\tau' \setminus \tau$ more than twice; we then obtain m_0 from a topological bound on the number of branches of any train track in S.

As in the previous lemma, let τ'_ϵ be a regular neighborhood of τ' foliated by ties and isotope σ' so that it is contained in τ'_ϵ and is transverse to the ties. Because each component of $S \setminus \tau'$ is either a polygon with $d \geq 3$ corners or a once-punctured polygon with $p \geq 1$ corners, we may extend the ties to a foliation $\mathcal{F}$ of S with one index $1 - d/2 \leq -1/2$ singularity in each d-gon and an index $1 - p/2 \leq 1/2$ singularity at the puncture of each punctured p-gon. (See figure 4.)

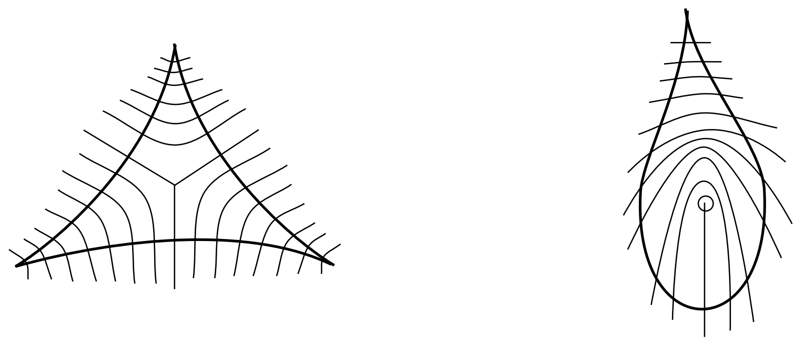

FIGURE 4. The foliation $\mathcal{F}$ in a triangular component, and in a punctured monogon.

Fix a tie t that crosses the regular neighborhood of a branch b of $\tau' \setminus \tau$. Since $\sigma \prec \tau$, no branch of σ crosses t. Suppose a branch e of $\sigma' \setminus \sigma$ crosses t twice and let t' be a segment in t between two successive crossings. Then t' and an interval $e' \subset e$ form an embedded loop in a complementary region U of σ, which is either a disk or once-punctured disk since σ is large.

Since the foliation is transverse to e and parallel to t, we can see that the index of the foliation around $t' \cup e'$ is therefore $+1$ if the ends of e' meet t on opposite sides, and $+1/2$ if they meet t on the same side. The first case cannot occur since the disk bounded by $t' \cup e'$ can contain at most one singularity of $\mathcal{F}$, of index at most $1/2$. The second case can occur if $t' \cup e'$ surrounds a puncture. However, in this case we can see that there are no other points of $t \cap e$. For if there were we could combine two such loops to find a loop in U with index 1, again a contradiction. □

3.2. **Nesting and $\mathcal{C}$-distance.** Equation (3.1) is a special case of the following more general fact:

Lemma 3.4. *If σ is a large birecurrent train-track and $\alpha \in int(PE(\sigma))$ then*

$$d_{\mathcal{C}}(\alpha, \beta) \leq 1 \implies \beta \in PE(\sigma).$$

In other words,

$$\mathcal{N}_1(int(PE(\sigma))) \subset PE(\sigma),$$

where $\mathcal{N}_1$ denotes a radius 1 neighborhood in $\mathcal{C}_1$.

Proof. If $\beta \notin PE(\sigma)$ then, by the more quantitative Lemma 3.5 below, $i(\alpha, \beta)$ is no less than the minimum weight α puts on any branch of σ. This is positive since $\alpha \in int(PE(\sigma))$, but then it follows $d_{\mathcal{C}}(\alpha, \beta) > 1$. □

We note the following immediate consequence:

$$\mathcal{N}_1(int(PN(\sigma)) \subset PN(\sigma) \tag{3.2}$$

which is obtained by applying Lemma 3.4 to the large subtracks of σ.

It remains to prove the following lemma, which will also be used at the end of this section.

Lemma 3.5. *Let τ be a large birecurrent track, let α be carried on a diagonal extension $\tau' \in E(\tau)$, and let β be a curve not carried on any diagonal extension of τ. Then $i(\alpha, \beta) \geq \min_b \alpha(b)$ where the right hand side denotes the minimum weight α puts on all branches b of τ.*

Proof. Consider first the case that $\tau' = \tau$, so that α is actually carried in τ.

Since τ is transversely recurrent, for any $\epsilon > 0$ there is a hyperbolic metric on S for which all train routes through τ have curvature at most ϵ. Fixing such a metric for $\epsilon < 1$, and lifting τ to a train-track $\widetilde{\tau}$ in the universal cover $\mathbf{H}^2$, we have that each train route r of $\widetilde{\tau}$ is uniformly quasi-geodesic, and in particular has two distinct endpoints $\partial r = \{r_+, r_-\}$ on the circle $\partial\mathbf{H}^2$ and stays in a uniform neighborhood of the geodesic r^* connecting them.

Choose a component $\widetilde{\beta}$ of the lift of β to $\mathbf{H}^2$, and note it is also quasi-geodesic. Let T_β be a generator of the subgroup of $\pi_1(S)$ preserving $\widetilde{\beta}$. We say that an edge $\widetilde{e}$ of $\widetilde{\tau}$ *separates* $\widetilde{\beta}$ *consistently* if, for any train route r passing through $\widetilde{e}$, its endpoints $r_\pm$ separate $\widetilde{\beta}_\pm$. If this occurs for some $\widetilde{e}$ then, letting e be its projection to S, we deduce immediately that $i(\beta, \alpha) \geq \alpha(e)$, since α lifts to a collection of train routes with $\alpha(e)$ of them passing through $\widetilde{e}$, and through each translate $T_\beta^m(\widetilde{e})$.

Thus, let us now prove that, if no edge of $\widetilde{\tau}$ separates $\widetilde{\beta}$ consistently, then $\widetilde{\beta}$ is carried on a diagonal extension of $\widetilde{\tau}$ (at the end we will check that this projects down to an extension of τ).

Each train route r separates $\mathbf{H}^2$ into two open disks, which we call halfplanes. Note that each is contained in a uniformly bounded neighborhood of a geodesic halfplane bounded by r^*. For a halfplane H, let H' denote $\bar{H} \setminus \bar{r}$ (where the bar denotes closure in the closed disk), i.e. the union of H with an open arc on the boundary circle.

Let J_+ and J_- be the components of $\partial\mathbf{H}^2 \setminus \partial\widetilde{\beta}$. Let $\mathcal{H}_+$ denote the union of all halfplanes H (bounded by train routes) that meet the boundary entirely in J_+, and define $\mathcal{H}_-$ similarly.

Any halfplane H_+ from $\mathcal{H}_+$ must be disjoint from any halfplane H_- from $\mathcal{H}_-$. Suppose otherwise: their intersection would be an open set which is either bounded or meets infinity only at $\partial\widetilde{\beta}$. In the first case it must be a bigon bounded by arcs of the train route boundaries of H_+ and H_-. A bigon has generalized Euler characteristic (as defined in §3.1) $\chi' = 0$. But since, as is easily checked, χ' is additive for unions of closures of complementary regions of $\widetilde{\tau}$ (see also Casson-Bleiler [9]), this contradicts the condition $\chi' < 0$ for complementary regions of the train-track. If the intersection is unbounded it is a "generalized bigon", i.e. a region R between two train routes r_+ and r_-, which are either biinfinite and hence parallel to $\widetilde{\beta}$, or are infinite rays emanating from a common point and asymptotic at infinity. Since they are quasigeodesics they remain a bounded distance apart. Either case again contradicts the Euler characteristic condition: For any $M > 0$, the intersection of R with a ball of radius M contains a union R_M of complementary regions with a *bounded* number of cusps, only depending on the distance between r_+ and r_-. Thus $|\chi'(R_M)|$ is bounded. On the other hand, by additivity, $-\chi'(R_M)$ grows linearly with M. This is a contradiction.

We conclude that $\mathcal{H}_+$ and $\mathcal{H}_-$ are disjoint. Now let $\mathcal{K} = \mathbf{H}^2 \setminus (\mathcal{H}_+ \cup \mathcal{H}_-)$. This is a closed set, nonempty since $\mathbf{H}^2$ is connected, and by construction must be a union of vertices, arcs and complementary regions of $\widetilde{\tau}$.

We further claim that $\mathcal{K}$ is connected, and in fact ϵ'*-convex* for some small $\epsilon' < 1$, meaning that any two points in $\mathcal{K}$ are connected by a path in $\mathcal{K}$ whose curvature is bounded by ϵ' at every point.

To see this, note that $\mathcal{K}$ is the intersection of a sequence of *closed* halfplanes $\{C_n\}$, each the complement of an open halfplane from $\mathcal{H}_\pm$. Letting $\mathcal{K}_n = \cap_{i<n} C_i$, we show the claims for each K_n. First consider the local structure of $\partial\mathcal{K}_n$.

Suppose that two branches a and b of $\widetilde{\tau}$, which are in $\mathcal{K}_n$, meet at a switch s. If they are on the same side of the switch, the cusp region between them must be contained in $\mathcal{K}_n$. For otherwise there is a halfplane outside $\mathcal{K}_n$, whose boundary passes through s and separates a from b – so one of them is in the halfplane and not in $\mathcal{K}_n$.

We conclude from this that the boundary of $\mathcal{K}_n$ is comprised of train routes meeting at *outward pointing* cusps. It follows that each component X of $\mathcal{K}_n$ is ϵ'-convex for some ϵ' which can be made arbitrarily small by suitable choice of ϵ. (For example, use the fact that a small radial neighborhood of X is convex, to deform into X any geodesic with endpoints in X).

Furthermore, we can see by induction that $\mathcal{K}_n$ is connected: $\mathcal{K}_2 = C_1$, and $\mathcal{K}_{n+1} = \mathcal{K}_n \cap C_n$. If $\mathcal{K}_n$ is connected, we note that the train route boundary of C_n must intersect $\mathcal{K}_n$ in a connected (possibly infinite) arc, for otherwise we obtain an arc in $\mathbf{H}^2 \setminus \mathcal{K}_n$ with endpoints on $\mathcal{K}_n$, which together with a piece of $\partial\mathcal{K}_n$ must bound a region with $\chi' \geq 0$, again a contradiction. It follows that $\mathcal{K}_{n+1}$ is connected.

The intersection of a nested sequence of closed ϵ'-convex sets is itself ϵ'-convex, and in particular connected. Thus our conclusions follow for $\mathcal{K}$.

We now claim that $int(\mathcal{K})$ is disjoint from $\widetilde{\tau}$, and hence is a union of complementary components. For, if $\widetilde{e}$ is any branch of $\widetilde{\tau}$, by our assumption $\widetilde{e}$ does not consistently separate $\widetilde{\beta}$, and hence lies on a train route both of whose endpoints are in either $\bar{J}_+$ or $\bar{J}_-$. Thus, $\widetilde{e}$ is on the boundary of one of the halfplanes comprising either $\mathcal{H}_+$ or $\mathcal{H}_-$, and hence in $\bar{\mathcal{H}}_\pm$. (Note, this is the only place where we use the assumption).

Since $\mathcal{K}$ is invariant by T_β, its closure in $\bar{\mathbf{H}}^2$ contains $\partial\widetilde{\beta}$. This together with ϵ'-convexity implies that $\mathcal{K}$ contains an infinite path p of curvature bounded by ϵ', connecting $\widetilde{\beta}_+$ with $\widetilde{\beta}_-$. Using the above description of the local structure of $\mathcal{K}$, we see that p may be taken to be a union of train routes and paths through diagonals of complementary regions of $\widetilde{\tau}$. That is, p is carried by a diagonal extension of $\widetilde{\tau}$. Both p and the extension may be assumed invariant by T_β, so p projects to a closed curve in S homotopic to β, and carried on τ together with a number of diagonal branches.

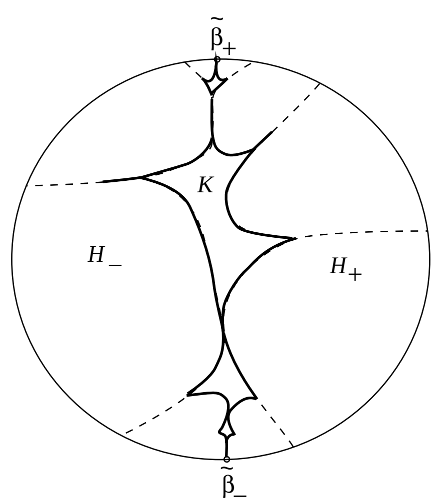

FIGURE 5. Part of the region $\mathcal{K}$, outlined in solid lines. The dotted lines indicate a few of the halfplanes comprising $\mathcal{H}_+$ and $\mathcal{H}_-$.

It remains just to check that the projected extension is still a train track, i.e. that none of the diagonals cross each other. But this is the same as checking that, if we translate the construction upstairs by $\pi_1(S)$, no region is traversed by two crossing diagonals. If this were to happen we clearly would obtain two translates of $\widetilde{\beta}$ whose endpoints separate each other, which would contradict the assumption that β is simple.

This concludes the proof in the case that $\tau' = \tau$. Now in general, note that we have so far proved the following: If β is not carried in any extension of τ, then there is an edge e of τ whose lift $\widetilde{e}$ has the property that all train routes of $\widetilde{\tau}$ through $\widetilde{e}$ separate $\partial\widetilde{\beta}$. Now consider a train route r' of the lift $\widetilde{\tau}'$ of τ' which passes through $\widetilde{e}$. Since $\widetilde{\tau}'$ is a diagonal extension of $\widetilde{\tau}$, r' must be sandwiched in between two routes of $\widetilde{\tau}$ that pass through $\widetilde{e}$. It follows that r' also separates $\partial\widetilde{\beta}$. Hence when α is carried on τ', it lifts to $\alpha(e)$ routes of $\widetilde{\tau}'$ through $\widetilde{e}$, and as before we obtain $i(\alpha,\beta) \geq \alpha(e)$. ☐

3.3. **Infinite diameter and action of pseudo-Anosovs.** We now have sufficient tools to prove the following result on the action of $\mathrm{Mod}(S)$ on $\mathcal{C}(S)$:

Proposition 3.6. *For a non-sporadic surface S there exists $c > 0$ such that, for any pseudo-Anosov $h \in \mathrm{Mod}(S)$, any $\gamma \in \mathcal{C}_0$ and any $n \in \mathbf{Z}$,*

$$d_{\mathcal{C}}(h^n(\gamma), \gamma) \geq c|n|.$$

As an immediate corollary we have

$$\mathrm{diam}(\mathcal{C}(S)) = \infty,$$

which gives part of the conclusion of Theorem 1.1. (In fact F. Luo has pointed out an easier proof that the diameter is infinite, which we sketch here: Let μ be a maximal geodesic lamination and let γ_i be any sequence of closed geodesics converging geometrically to μ. Then if $d_{\mathcal{C}}(\gamma_0, \gamma_n)$ remains bounded, after restricting to a subsequence we may assume $d_{\mathcal{C}}(\gamma_0, \gamma_n) = N$ for all $n > 0$. For each γ_n we may then find β_n with $d(\beta_n, \gamma_n) = 1$ and $d(\gamma_0, \beta_n) = N - 1$. But $\gamma_n \to \mu$ and μ maximal implies $\beta_n \to \mu$ as well, since γ_n and β_n are disjoint in S. Proceeding inductively we arrive at the case $N = 1$, and the conclusion is that $\beta_n \to \mu$ and $\beta_n = \gamma_0$, a contradiction.)

Proposition 3.6 should be compared to a property of the action of a word-hyperbolic group G on its Cayley graph Γ: namely, for a fixed $c(G) > 0$, if $h \in G$ has infinite order then $d(h^n\gamma, \gamma) \geq c|n|$ for all $n \in \mathbf{Z}$, $\gamma \in \Gamma$ (see [15, 14, 2]).

In Gromov's terminology [15], Proposition 3.6 says that the action of a pseudo-Anosov h on $\mathcal{C}(S)$ is *hyperbolic*. In general there are two more types of isometries of a δ-hyperbolic space: *elliptic*, for which orbits are bounded, and *parabolic*, for which orbits are unbounded but $\inf d(\gamma, h^n(\gamma))/|n| = 0$. When h is not pseudo-Anosov, it must be reducible or finite-order. In either case some vertex of $\mathcal{C}(S)$ is fixed by a finite power of h, and hence h is elliptic. Thus it follows from Proposition 3.6 that the action of $\mathrm{Mod}(S)$ on $\mathcal{C}(S)$ has no parabolics.

Proof of Proposition 3.6. A pseudo-Anosov map $h : S \to S$ determines measured laminations $\mu, \nu \in \mathcal{ML}(S)$, called stable and unstable laminations, with the following properties (for more about pseudo-Anosov homeomorphisms, see e.g. [12, 40, 4]). They are transverse to each other, and the complementary regions of each are ideal polygons or once punctured ideal polygons. Both projective classes $[\mu]$ and $[\nu]$ in $\mathcal{PML}(S) = (\mathcal{ML}(S) \setminus \{0\})/\mathbf{R}_+$ are fixed points for h, such that $[\mu]$ is attracting in $\mathcal{PML}(S) \setminus [\nu]$, and $[\nu]$ is repelling in $\mathcal{PML}(S) \setminus [\mu]$. In particular, since μ and ν cannot have closed-curve components, every vertex of $\mathcal{C}(S)$ approaches $[\mu]$ under forward iteration of h, and $[\nu]$ under backward iteration.

Let τ_0 be a generic train-track formed from a regular ϵ neighborhood of μ. If ϵ is sufficiently small, the complementary domains of τ_0 are in one-to-one correspondence with those of μ, and τ_0 is birecurrent (see Penner-Harer [38]).

One can homotope h to a standard form in which it permutes the complementary regions of μ, is expanding on the leaves of μ, and contracting in the transverse direction near μ. Thus, the image train-track $h(\tau_0)$ must be carried in τ_0, and fills it.

If $\tau \in E(\tau_0)$ is a diagonal extension, then $h(\tau)$ is carried in some $\tau' \in E(\tau_0)$ by Lemma 3.2. Since the number of tracks in $E(\tau_0)$ is bounded in terms of the topology of S, there is some $k_0(S)$ such that, for some $k \le k_0$ the power $h' = h^k$ takes τ to a track carried by τ. Let $\mathcal{B}$ denote the branch set of τ, and $\mathcal{B}_0 \subset \mathcal{B}$ the branch set of τ_0. In the coordinates of $\mathbf{R}^{\mathcal{B}}$ we may represent h' as an integer matrix M, with a submatrix M_0 giving the restriction to $\mathbf{R}^{\mathcal{B}_0}$ (see Penner [37]). Penner shows in [37] that M_0^n has all positive entries where n is the dimension $|\mathcal{B}_0|$, and in fact $|M_0^n(x_0)| \ge 2|x_0|$ for any vector x_0 representing a measure on τ_0. Indeed M_0 has a unique eigenspace in the positive cone of $\mathbf{R}^{\mathcal{B}_0}$, which corresponds to $[\mu]$. On the other hand, for a diagonal branch $b \in \mathcal{B} \setminus \mathcal{B}_0$ we have, by Lemma 3.3, that $|M^i(x)(b)| \le m_0|x|$ for all $x \in \mathbf{R}^{\mathcal{B}}$ and all powers $i > 0$. Now, any transverse measure x on τ must put some positive measure on a branch of $\mathcal{B}_0$, since τ_0 is generic. It follows immediately that given any $\delta > 0$ there exists m_1, depending only on δ and S, such that for some $m \le m_1$ we have $\max_{b\in\mathcal{B}\setminus\mathcal{B}_0} h^m(x)(b) \le \delta \min_{b\in\mathcal{B}_0} h^m(x)(b)$, for any $x \in P(\tau)$. Applying this to each $\tau \in E(\tau_0)$, and invoking Lemma 3.1, we conclude that, for suitable choice of δ,

$$h^m(PE(\tau_0)) \subset int(PE(\tau_0)).$$

Thus letting $\tau_j = h^{mj}(\tau_0)$ we find by induction that $PE(\tau_{j+1}) \subset int(PE(\tau_j))$. Now if $\beta \in \mathcal{C}_0(S)$ satisfies $\beta \notin PE(\tau_0)$ but $h^m(\beta) \in PE(\tau_0)$, we have that $h^{km}(\beta) \in PE(\tau_{k-1})$ for $k \ge 1$. But then Lemma 3.4 applied inductively shows that $d_{\mathcal{C}}(h^{km}(\beta), \beta) \ge k$.

For arbitrary $n \in \mathbf{Z}$ we note, since h is an isometry on $\mathcal{C}$, that $|n| \le d_{\mathcal{C}}(h^{nm}(\beta), \beta) \le m d_{\mathcal{C}}(h^n(\beta), \beta)$, and conclude that $d_{\mathcal{C}}(h^n(\beta), \beta) \ge |n|/m$.

Finally, for arbitrary $\gamma \in \mathcal{C}_0(S)$ we note that $[h^n(\gamma)] \to [\mu]$ as $n \to \infty$, and (see the discussion before Lemma 3.1) some neighborhood of μ is contained in $int(PE(\tau_0))$. Thus eventually $h^n(\gamma) \in int(PE(\tau_0))$. On the other hand $[h^{-n}(\gamma)] \to [\nu]$ as $n \to \infty$, and ν is not contained in $PE(\tau_0)$ since by the previous discussion all of $[PE(\tau_0)]$ converges to $[\mu]$ under iterations of h. Since $[PE(\tau_0)]$ is closed it misses a neighborhood of $[\nu]$, and we conclude that there is some $p \in \mathbf{Z}$ for which $h^p(\gamma) \notin PE(\tau_0)$ but $h^{m+p}(\gamma) \in PE(\tau_0)$. Applying the previous two paragraphs to $\beta = h^p(\gamma)$, we obtain the desired bound for γ as well.

We thus have our conclusion for $c = 1/m_1$, which by construction is independent of h or γ. □

3.4. **The nesting lemma.** We will need the following notation. If $\alpha \in \mathcal{C}_0(S)$ and σ, τ are train tracks, let $d_{\mathcal{C}}(\alpha, \sigma)$ denote $\min_v d_{\mathcal{C}}(\alpha, v)$ and $d_{\mathcal{C}}(\sigma, \tau) = \min_{v,w} d_{\mathcal{C}}(v, w)$, where v ranges over the vertices of σ and w ranges over the vertices of τ.

The goal of this section is the following lemma, whose proof will appear at the end of it.

Lemma 3.7. (Nesting Lemma) *There exists a $D_2 > 0$ such that, whenever ω and τ are large recurrent generic tracks and $\omega \prec \tau$, if $d_{\mathcal{C}}(\omega, \tau) \ge D_2$, we have*

$$PN(\omega) \subset int(PN(\tau)).$$

Given a train-track τ and a measure $\mu \in P(\tau)$ we can define a *combinatorial length* $\ell_\tau(\mu)$ as $\sum_b \mu(b)$, where the sum is over the branches b of τ. Similarly if $\mu \in PN(\tau)$ we can define $\ell_{N(\tau)}(\mu)$ as the minimum of combinatorial lengths in the tracks of $N(\tau)$ that carry μ.

There are some easy consequences of there being only finitely many combinatorial types of train-tracks on S. For example, if $\lambda \in P(\tau)$ and one writes λ as a combination $\sum_i a_i \alpha_i$ (not necessarily unique) of the vertices α_i of τ with nonnegative coefficients, then

$$\max_i a_i \le \ell_\tau(\lambda) \le C_1 \max_i a_i \tag{3.3}$$

where C_1 depends on a bound on the number of vertices of τ, and a bound C_0 for $\ell_\kappa(\omega)$ over all train tracks κ and vertices ω.

Another consequence of finiteness is that there is a constant B depending only on S, such that any two vertices of a train-track are $\mathcal{C}$-distance at most B apart. (We conjecture that $B = 2$).

Furthermore we have:

Lemma 3.8. *Given $L > 0$ there exists $D_0(L)$ so that, if $\alpha \in P(\tau)$ and $\ell_\tau(\alpha) < L$ then $d_\mathcal{C}(\alpha, \tau) < D_0$.*

Proof. Fixing L and τ, only finitely many curves α are carried by τ with $\ell_\tau(\alpha) \le L$. Thus there is an upper bound on their distance from the vertices of τ. Taking a maximum over all combinatorial types of train-tracks in S, we have the desired statement. □

We also observe:

Lemma 3.9. *If $\alpha \in P(\tau)$ and $d_\mathcal{C}(\alpha, \tau) \ge 3$ then α fills a large subtrack of τ.*

Proof. Suppose that α is carried in $\kappa < \tau$ which is not large. Then $S \setminus \kappa$ contains a nontrivial, nonperipheral curve β, so that $d_\mathcal{C}(\beta, \alpha) \le 1$ and $d_\mathcal{C}(\beta, v) \le 1$ for any vertex v of κ. By the triangle inequality $d_\mathcal{C}(\alpha, v) \le 2$, and since v is also a vertex of τ, $d_\mathcal{C}(\alpha, \tau) \le 2$. □

The next lemma addresses the following issue. A closed curve carried on an extension of a track σ does not necessarily trace through any complete cycle on σ. However, if σ is sufficiently deeply nested in τ, then any curve on an extension of σ is forced to run through a cycle of τ, and in fact must put a definite amount of weight on that cycle.

Lemma 3.10. *There exists M_0, and for any L there exists $D_1(L)$ such that if σ is large, $\sigma \prec \tau$, and $d(\sigma, \tau) \ge D_1(L)$ then the following holds. Suppose $\sigma' \in E(\sigma)$ and $\tau' \in E(\tau)$, and $\sigma' \prec \tau'$. Then any curve β carried on σ' can be expressed in $P(\tau')$ as $\beta_\tau + \beta'_\tau$, where $\beta_\tau \in P(\tau)$, and*

$$\ell_{\tau'}(\beta'_\tau) \le M_0 \ell_{\sigma'}(\beta), \tag{3.4}$$

$$\ell_\tau(\beta_\tau) \ge L \ell_{\sigma'}(\beta). \tag{3.5}$$

Proof. It suffices to prove the lemma when β is a vertex v of σ'. For the general case, express β as a combination of vertices and use (3.3).

Let W_0 be a bound (by finiteness) on the weights that v puts on any branch of σ', so that by Lemma 3.3 v puts at most $m_0 W_0$ on the branches of $\tau' \setminus \tau$. Write the vertices of τ' as $\{\alpha_i\} \cup \{\gamma_j\}$, where α_i are the ones supported in τ. Then in the coordinates of $P(\tau')$ we may write $v = v_\tau + v'_\tau$ where $v_\tau = \sum a_i \alpha_i$ and $v'_\tau = \sum_j c_j \gamma_j$, with $a_i, c_j \ge 0$.

For each branch b of $\tau' \setminus \tau$ we have $v(b) = \sum c_j \gamma_j(b) \le m_0 W_0$. Since for each j some b has $\gamma_j(b) \ge 1$, we have $c_j \le m_0 W_0$.

We have shown

$$\ell_{\tau'}(v'_\tau) \le m_0 W_0 C_0,$$

where recall C_0 is a bound for the combinatorial length of all vertices of a train track. Letting $M_0 = m_0 W_0 C_0$, and noting that $\ell_{\sigma'}(v) \ge 1$, we have the first desired inequality (3.4). Let D_0 be the distance bound provided by Lemma 3.8 for a length bound of $C_0 L + M$. Then let $D_1 = 2B + D_0$.

Since the distance between a vertex of σ' and any vertex of σ is at most B, and the same for τ' and τ, we conclude from the assumption $d_\mathcal{C}(\sigma, \tau) \ge D_1$ that we have $d_\mathcal{C}(v, \tau') \ge D_0$, and by Lemma 3.8, we have $\ell_{\tau'}(v) \geq C_0 L + M_0$. Now $\ell_\tau(v_\tau) = \ell_{\tau'}(v) - \ell_{\tau'}(v'_\tau) \geq C_0 L$, and since $\ell_{\sigma'}(v) \leq C_0$ we have the second inequality (3.5). □

Proof of Lemma 3.7 (Nesting Lemma). Let $\omega \prec \tau$ with $d_{\mathcal{C}}(\omega, \tau) \geq D_2$, where D_2 will be determined shortly. Let σ be any large subtrack of ω. We will prove that $PE(\sigma) \subset int(PE(\kappa))$ for some large subtrack κ of τ. Thus by definition we will have $PN(\omega) \subset int(PN(\tau))$, which is the desired statement.

Let $\tau = \tau_0$ and $\cdots \tau_2 \prec \tau_1 \prec \tau_0$ be a sequence of tracks obtained from τ_0 by splitting, so that $\sigma \prec \tau_j$ for each j. Let ρ be the first τ_j for which $d_{\mathcal{C}}(\tau_j, \tau) > 2$. Note that we may assume σ fills ρ: in a splitting move determined by σ, if an edge becomes empty we use a collision move and erase the edge. Since σ is large, ρ must be as well.

By the properties of splitting sequences (see Section 3.1), τ_j either shares a vertex with τ_{j-1} or is a subtrack of a track that shares a vertex with it. Thus for any vertex v of $\rho = \tau_j$, $d_{\mathcal{C}}(v, \tau_{j-1}) \leq B$, and it follows that $d_{\mathcal{C}}(v, \tau) \leq 2 + 2B$. Therefore $d_{\mathcal{C}}(\sigma, \rho) \geq D_2 - 2 - 2B$.

Fix now β carried by $\sigma' \in E(\sigma)$, and let us show that $\beta \in int(PE(\kappa))$ for some large subtrack $\kappa < \tau$. The idea will be that, by Lemma 3.10, β will place definite weight on some cycle of ρ, and by Lemma 3.9 this cycle will fill a large subtrack κ_0 of τ. On the other hand β will place relatively little weight on any extension branches outside τ, and we will be able to reach our conclusion for some κ containing κ_0.

Since σ fills ρ, by Lemma 3.2 there is some $\rho' \in E(\rho)$ carrying β. Fix L_1 (to be determined shortly), and let C_1 be the constant in inequality (3.3). Lemma 3.10 implies that for sufficiently large D_2 (depending on $L_1 C_1$), we can write $\beta = \beta_\rho + \beta'_\rho$ where $\ell_\rho(\beta_\rho) \geq L_1 C_1 \ell_{\sigma'}(\beta)$. Inequality (3.3) then implies that we can write $\beta_\rho = \sum_i a_i \alpha_i$ where α_i are vertices of ρ, such that $a_1 \geq L_1 \ell_{\sigma'}(\beta)$. Now applying Lemma 3.9, since $d_{\mathcal{C}}(\alpha_1, \tau) > 2$ there is a large subtrack κ_0 of τ such that $\alpha_1(b) \geq 1$ for each branch b of κ_0.

Therefore $\beta(b) \geq L_1 \ell_{\sigma'}(\beta)$ for every branch b of κ_0. However, we don't know if $\beta \in PE(\kappa_0)$. The trouble is that the enlargement of κ_0 that supports β may not be a diagonal extension. Thus we will find an intermediate track between τ and κ_0 by adding branches to κ_0 that have too much weight to be pushed to the corner, and show that this process terminates with the desired track.

Let $\tau' \in E(\tau)$ be a track carrying β (by Lemma 3.2). By Lemma 3.3 we know that $\beta(b) \leq m_0 \ell_{\sigma'}(\beta)$ for any branch b of $\tau' \setminus \tau$.

If for all branches c of $\tau' \setminus \kappa_0$ which meet κ_0 we have $\beta(c) < \delta L_1 \ell_{\sigma'}(\beta)$ then by Lemma 3.1, $\beta \in int(PE(\kappa_0))$ and we are done. If not, let c violate this inequality, and define an extension κ_1 of κ_0 containing c as follows: let $c_\pm$ be the ends of c where $c_- \in \kappa_0$. If $c_+ \in \kappa_0$ then $\kappa_1 = \kappa_0 \cup c$ is a train track. If not, then c_+ is incoming to some switch with at most m_1 branches outgoing ($m_1 = m_1(S)$). At least one of those, c_1, a branch of τ', has measure $\beta(c_1) \geq \frac{1}{m_1}\beta(c) \geq \frac{\delta}{m_1} L_1 \ell_{\sigma'}(\beta)$. Add this branch, and continue adding branches of τ' until we find one which touches κ_0 again. Let κ_1 denote κ_0 together with this chain of branches, and note that for all branches b of κ_1, $\beta(b) \geq \frac{\delta}{m_2} L_1 \ell_{\sigma'}(\beta)$, where $m_2 = m_2(S)$.

If now there is no edge of $\tau' \setminus \kappa_1$ adjacent to κ_1 with measure at least $\delta \frac{\delta}{m_2} L_1 \ell_{\sigma'}(\beta)$, we are done. Otherwise, we can repeat this process, obtaining a sequence κ_i of extensions, each of which is a subtrack of τ' and which must terminate after at most m_3 steps, $m_3 = m_3(S)$. Thus, $\beta(b)$ for any branch b of κ_i is always at least $\left(\frac{\delta}{m_2}\right)^{m_3} L_1 \ell_{\sigma'}(\beta)$. If we have chosen L_1 sufficiently large that $m_0 < \left(\frac{\delta}{m_2}\right)^{m_3} L_1$, this process must terminate *without* appending to κ_j any branches of $\tau' \setminus \tau$, since as above all such branches have β-measure at most $m_0\ell_{\sigma'}(\beta)$. Therefore we must end with some $\kappa_j < \tau$ for which $\beta \in int(PE(\kappa_j))$, and we are done. $\square$

We note that a corollary of the proof is the following quantitative version of the Nesting Lemma, obtained by taking the constant L_1 sufficiently large:

Lemma 3.11. *The constant D_2 in the Nesting Lemma may be chosen so that, if $\omega \prec \tau$ and $d_{\mathcal{C}}(\omega,\tau) \geq D_2$, then for any $\beta \in PN(\omega)$ there is a subtrack $\kappa < \tau$ such that $\beta \in PE(\kappa)$ and, for any branch b of κ,*

$$\beta(b) \geq 2\ell_{N(\omega)}(\beta).$$

3.5. **Growth of intersection numbers.** Lemma 3.11 implies, in particular, that the combinatorial length of a curve carried on a train track grows at least exponentially with its distance from a fixed point, say a vertex of the track. As a consequence (see also Lemma 3.5), its intersection number with any fixed curve not carried on the track should grow exponentially.

A finer analysis shows that, if two curves are both far from a fixed one and relatively near each other, then both are deeply nested in diagonal extensions of the same track, and as a consequence their intersection numbers with any fixed curve are very large compared to their intersection number with each other. In the closing argument of the Projection Theorem 2.6, in Section 5, we will use the following quantitative version of this observation.

Lemma 3.12. *Given $Q, k > 0$ there exist D_3, ν such that the following holds. If α, β and γ in $\mathcal{C}_0(S)$ are such that*

$$d_{\mathcal{C}}(\beta,\alpha) \geq D_3$$

and

$$d_{\mathcal{C}}(\gamma,\beta) \leq \nu d_{\mathcal{C}}(\beta,\alpha)$$

then

$$\min_{\alpha'} i(\beta,\alpha') \cdot \min_{\alpha'} i(\gamma,\alpha') \geq Q i(\beta,\gamma),$$

where α' varies over the k-neighborhood of α in $\mathcal{C}$.

Proof. Extend α to a pair of pants decomposition of S. Such a decomposition determines a family of *standard train-tracks*, obtained by choosing one of a finite number of configurations in each pair of pants and in a connecting collar between any adjacent pairs of pants. (See Penner-Harer [38]). Each such track is generic and birecurrent, and The family of tracks has the property that *any* simple closed curve on S is carried by one of them. Thus, let τ_0 denote a standard train-track carrying β. Depending on the choice of local picture in an annulus neighborhood of α, α is either a vertex cycle of τ_0, or is distance at most 2 from a vertex cycle. It follows that $d_{\mathcal{C}}(\beta,\tau_0) \geq d_{\mathcal{C}}(\beta,\alpha) - (2+B)$.

Since β is assumed far from α, by Lemma 3.9 it fills a large subtrack of τ_0, which we will continue to call τ_0.

Now we will show that we can find a sequence $\tau_n \prec \cdots \prec \tau_0$ of train-tracks, each carrying β, so that $d_{\mathcal{C}}(\tau_{j+1},\tau_j) > D_2$ and the length of the sequence is $n \geq d_{\mathcal{C}}(\beta,\tau)/(D_2+2B)$. This is done by splitting. Perform a sequence of splittings of τ_0 determined by the weights of β, and terminating with a track that has β as a vertex. Inductively define τ_{j+1} to be the first track in the sequence such that $d_{\mathcal{C}}(\tau_{j+1},\tau_j) > D_2$. Since τ_{j+1} is the first, we may conclude as in the proof of Lemma 3.7 that $d_{\mathcal{C}}(v,\tau_j) \leq D_2 + 2B$ for any vertex v of τ_{j+1}. It follows that $d_{\mathcal{C}}(\beta,\tau_{j+1}) \geq d(\beta,\tau_j) - (D_2+2B)$, and we may continue.

Lemma 3.7 now guarantees that $PN(\tau_{j+1}) \subset int(PN(\tau_j))$.

If $d_{\mathcal{C}}(\alpha,\alpha') \le k$ then $d_{\mathcal{C}}(\alpha',\tau_0) \le k+2$, and by applying Lemma 3.4 inductively we see that α' cannot be in $PN(\tau_{k+3})$. (Note that to apply Lemma 3.4 we need each τ_j to be transversely recurrent, but this follows from the fact that τ_0 is.)

Another application of Lemma 3.4 shows that if $d_{\mathcal{C}}(\beta,\gamma) \le m < n$, then $\gamma \in PN(\tau_{n-m})$.

Thus, assuming n is sufficiently large compared to m, we may conclude that both β and γ are contained in $PN(\tau_{k+3})$. Furthermore, applying Lemma 3.11 repeatedly, we also have that for a large subtrack κ of τ_{k+3}, γ puts weight at least $2^{n-m-k-3}\ell_{N(\tau_{n-m})}(\gamma)$ on every branch of κ. The same holds for β and a large subtrack κ' of τ_{k+3}. Applying Lemma 3.5, we find that

$$i(\alpha',\gamma) \ge 2^{n-m-k-3}\ell_{N(\tau_{n-m})}(\gamma),$$

and similarly for β.

On the other hand it is easy to see that

$$i(\beta,\gamma) \le C_2 \ell_{N(\tau_{n-m})}(\beta)\ell_{N(\tau_{n-m})}(\gamma).$$

where C_2 depends only on the topological type of S. This is because in every branch of τ_{n-m} a strand of β and one of γ can only have one essential intersection, and a strand in a diagonal branch of $N(\tau_{n-m})$ can only hit strands in diagonal branches, and two diagonal branches can intersect at most once if the complementary domain is a polygon, or twice if it is a punctured polygon.

Putting these inequalities together, if $n-m-k$ is sufficiently high we have the desired inequality. □

4. Geometry of quadratic differentials

Let q be a holomorphic quadratic differential of area 1 with respect to some conformal structure x on S. In this section we will study the geometry imposed by q, with particular regard to the way nearly horizontal and nearly vertical geodesics are arranged, and how they intersect each other. Our main goals are the Vertical Domain Lemma 4.6, which gives a particular "thickening" of a nearly vertical curve with some useful properties, and the Intersection Number Lemma 4.8, which gives conditions for a nearly vertical and a nearly horizontal curve to have large intersection number.

4.1. **Basic properties and uniform estimates.** A *straight segment* with respect to q is a path containing no singularities in its interior, and which is geodesic in the locally Euclidean metric of q. If S has no punctures, a *geodesic segment* is composed of straight segments which meet at singularities making an angle of at least π on either side. A straight segment connecting two singularities is also called a *saddle connection.*

A *metric cylinder* in q is an annulus which is isometric to the product of a circle and a line segment.

When S has no punctures, each nontrivial homotopy class has a geodesic representative. However when there are punctures the metric of q is incomplete and we must slightly generalize the notion. From now on by "geodesic representative" of a closed curve α we mean a curve α^* in the compactified surface $\hat{S}$ (adding the punctures) such that $\alpha^*\cap S$ is composed of geodesic arcs, and there is a homotopy from α to α^* which until the last moment is contained in S. One can formalize this notion by first excising from S open r-neighborhoods of the punctures, considering geodesic representatives in the resulting compact surface, and then letting r tend to 0. It is not hard to see that any non-peripheral homotopy class has such a geodesic representative, which has minimal length, and the representative is unique unless there is a metric cylinder foliated by curves in the homotopy class (however the homotopy class is not uniquely determined by the representative). The same discussion works for the geodesic representative of a homotopy class of paths rel endpoints.

If we start with a homotopy class of simple curves then the geodesic representative does not have to be simple: even in the absence of punctures, it may have self-tangencies along saddle connections, because the metric is not smooth at the zeros of q. Furthermore, when a geodesic representative passes through a puncture, since the total angle around the puncture is π, it is easy to see that in fact the path approaches the puncture along a straight segment, and then retraces the same segment in the opposite direction.

Topological constants. For later reference, $n_1, \ldots, n_5$ will denote the following bounds, which may easily be computed in terms of the genus and number of punctures of S. Let n_1 bound the number of singularities of q, including punctures. Let n_2 bound the number of disjoint saddle connections which may appear simultaneously in S. Let n_3 be an isoperimetric constant, such that $\text{Area}_q(X) \leq n_3 \,\text{diam}_q(X)^2$ for any subset X of S. Let n_4 bound the size of a sequence $X_1 \subset \cdots \subset X_{n_4} \subset S$ for which $i_*\pi_1(X_j)$ is a proper subgroup of $i_*\pi_1(X_{j+1})$, where i_* is the map induced on π_1 by inclusion into S. Let n_5 bound $1/\pi$ times the sum of cone angles over all singularities of q.

Definite collars. Let the *width* of an annulus A in a metric q denote the minimal distance between boundaries, and the *circumference* the minimal length of a curve going once around A. A compactness argument using the moduli space of Riemann surfaces yields the following:

Lemma 4.1. *There exists $W > 0$ depending only on the topology of S such that, for each unit area quadratic differential q there exists a nonperipheral annulus of width W. Furthermore, given μ there exists $L > 0$ so that the annulus can be chosen either to be a metric cylinder of modulus at least μ, or to have circumference at least L.*

Proof. If the statement is false, then there is a sequence of conformal structures x_i on S, unit-area holomorphic quadratic differentials q_i, and $L_i \to 0$, $W_i \to 0$ such that there is no annulus in (S, q_i) of width at least W_i which either has circumference at least L_i or is a metric cylinder of modulus μ.

We can now apply a compactification argument whose details may be found in Masur [29]. We may take a subsequence so that (S, x_i) converge in a compactified moduli space to a noded Riemann surface (S', x), where S' may be taken as the complement in S of a collection of disjoint curves, and q_i converge on compact sets of S' to some q. Given $\mu > 0$ there is a $K(\mu) > 0$ (depending on the topological type of S) such that the following alternative holds: if $\text{diam}(q_i) \geq K$ for all sufficiently high i then eventually (S, x_i, q_i) contains a metric cylinder of width at least W, and modulus at least μ. In this case we have contradicted the choice of sequence, hence we are done. If $\text{diam}(q) \leq K$ for all sufficiently high i then the limiting q is non-zero on at least one component R of S'. (The two possibilities are not mutually exclusive). We also note that q has at most simple pole singularities at the punctures.

Since R supports a non-zero holomorphic quadratic differential of finite area, it cannot be a sphere with less than 4 punctures. It follows that there is some simple, nontrivial, nonperipheral curve in R, so let A be any collar for this curve. If W and L are the width and circumference of A, then in the approximating metrics of q_i for high enough i we obtain annuli of nearly these width and circumferences, again contradicting the choice of sequence. □

Definite boxes. We will need the following notion, where a *rectangle* denotes an embedded Euclidean rectangle with respect to q, in particular containing no singularities in its interior.

Definition 4.2. *Let ω denote a q-geodesic segment or closed curve. If $N, \delta > 0$, an (N, δ) box for ω is a rectangle containing at least N parallel strands of ω (counting multiplicity) of equal length δ,* *parallel to two of the sides of the rectangle. The endpoints of the strands are on the orthogonal sides of the rectangle. The lengths of the orthogonal sides are at most δ.*

Remark: Note that if $N < 1$, an (N, δ) box means a $(1, \delta)$-box.

As a consequence of Lemma 4.1 we can prove the following:

Lemma 4.3. *Let q be a unit-area holomorphic quadratic differential on (S, x), and suppose that there are no q-metric cylinders of modulus greater than 2 in S. Let A denote the nonperipheral annulus of width W and length L provided by Lemma 4.1. There exist $\delta, r > 0$, depending only on the topology of S, such that for any closed q-geodesic γ which has intersection number $N > 0$ with the core of A, there is a (rN, δ)-box for γ in A. Furthermore, the q-injectivity radius at the center of the box is at least δ.*

Proof. On a smaller annulus $A' \subset A$ of width $W/2$ the q-injectivity radius is at least $\delta_1 = \min(W/4, L/2)$. There are N segments (with multiplicity) of γ of length $W/2$ passing through A'. Centered on any nonsingular point of a segment σ of $\gamma \cap A'$ there is a geodesic segment orthogonal to σ of length $2\delta_1$, and so (recalling n_1 from above) there must be a segment on σ of length at least $W/2n_1$ for which these orthogonal segments meet no singularities, and therefore make a $(1, W/2n_1)$ box for σ, with center on σ

Consider all such boxes in A. There are N (with multiplicity) and we must check that there is sufficient overlap. For each box consider a box of half the size with the same center. Since each box has definite area and the area of q is 1, we find that there must be a point simultaneously in $\max(rN, 1)$ half-boxes for a fixed $r > 0$, and hence a box containing $\max(rN, 1)$ centers of boxes. This is the desired box. ∎

4.2. **Vertical and horizontal.** From now on, let us suppose that two constants $\theta, \epsilon > 0$ have been fixed satisfying a short list of constraints which will appear in the course of the proof. For now assume $\theta < \min(1/2, \epsilon^2)$.

Definition 4.4. *We say a straight segment is* almost vertical *(respectively* almost horizontal*) with respect to q if its direction is within θ of the vertical (resp. horizontal) direction of q. We say a geodesic segment or closed curve is almost vertical (resp. almost horizontal) if it is composed of straight segments each of which is almost vertical (resp. almost horizontal) or has length at most ϵ.*

Note that a (weak) consequence of the condition $\theta < 1/2$ is that

$$|\alpha|_{q,v} > \tfrac{1}{2}|\alpha|_q.$$

We now define a certain type of thickening, which we call a *vertical* (or *horizontal*) *domain*, that will be useful in several places.

Definition 4.5. *Let ω be an almost vertical geodesic segment or closed curve. The* vertical domain *$\Omega_\epsilon(\omega)$ is constructed as follows. For any point $p \in \omega$ let σ_p be the maximal open q-horizontal segment about p which contains no singularities or punctures, and such that each component of $\sigma_p - \{p\}$ has length at most ϵ. Let $\Omega_\epsilon(\omega)$ be the closure of $\cup_{p\in\omega}\sigma_p$.*

We similarly define a horizontal domain *$\Psi_\epsilon(\omega)$ if ω is almost horizontal, where the σ_p are vertical segments.*

Let us record some useful properties of this construction.

Lemma 4.6. *There exist positive L_0 and a_0, a_1, a_2 such that the following holds. Let q be a unit-area quadratic differential on S and let ω be an almost vertical geodesic representative of a simple*

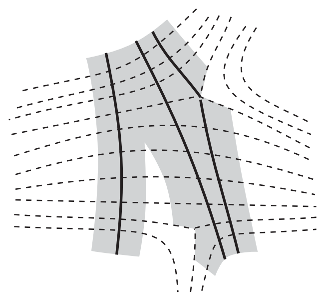

FIGURE 6. An example of a vertical domain. The horizontal foliation is dotted, ω is solid and $\Omega_\epsilon(\omega)$ is in grey

closed curve or segment (rel endpoints). If ω is a segment, assume it has length at least L_0. Let τ be an almost-horizontal straight segment of diameter $\mathrm{diam}_q(\tau) \geq a_0\epsilon$, which is disjoint from ω. Let $\Omega = \Omega_\epsilon(\omega)$. Then we have:

1. *The map $\pi_1(\Omega) \to \pi_1(S)$ induced by inclusion has non-trivial, non-peripheral image.*
2. *There is a subsegment of τ of diameter at least $a_1 \mathrm{diam}_q(\tau)$ which is disjoint from Ω.*
3. *The boundary of Ω has q-length at most $a_2\epsilon + 1/\epsilon$*
4. *The boundary of Ω has horizontal length at most $a_2\epsilon$.*

Proof. Let n_1, n_2 and n_5 be the topological constants described in §4.1, and choose $L_0 = 2n_2\epsilon + 6/\epsilon$. We will see that in fact the a_i can be written explicitly in terms of the n_i.

We first prove Part (1). If ω passes through a puncture then it follows from the definition that Ω contains a neighborhood of the puncture. Thus if ω is the (generalized) geodesic representative of a closed curve, or even if it passes through punctures more than once, part (1) is obviously satisfied. Therefore, possibly restricting to a subsegment, we may assume that ω is a segment of length at least $L_0/2$, passing through no punctures in its interior.

For all non-singular $p \in \omega$ let σ_p be the horizontal arcs of Definition 4.5. For all $y \in \Omega$ let $f(y) = \#\{p : y \in \sigma_p\}$ (where the number is counted with multiplicity if ω is not embedded). Then Ω can be described as the closure of a finite union of open parallelograms with horizontal sides of length 2ϵ, whose heights sum to $|\omega|_{q,v}$, and f gives the degree of overlap of these parallelograms. It follows that $\int_\Omega f(y) = 2\epsilon|\omega|_{q,v}$ (where the integral is with respect to q-area). On the other hand $\int_\Omega f(y) \leq \max(f)\,\mathrm{Area}(\Omega)$. Since the area of q is 1 we conclude

$$\max f \geq 2\epsilon|\omega|_{q,v}.$$

If ω' is the union of almost vertical straight segments of ω and ω'' is the rest, then $|\omega''|_q \leq n_2\epsilon$, so $|\omega'|_q \geq L_0 - n_2\epsilon$. Since $|\omega'|_{q,v} \geq \frac{1}{2}|\omega'|_q$, the choice of L_0 guarantees that $\max f \geq 3$.

We conclude that there is a point y contained in a horizontal segment $\sigma \subset \Omega$ which cuts ω in three places (counted with multiplicity if ω is not embedded). If there are two consecutive intersection points on σ with the same orientation, then a segment of ω together with an interval of σ make a simple curve in Ω, geodesic except at the intersection points where the total turning angle (measured in the q metric) is at most $2\theta < \pi$, and hence it cannot be trivial or peripheral. (A curve bounding a disk has total turning angle at least 2π, and a curve bounding a puncture has total angle at least π). If every two consecutive intersections have opposite orientations, we can take three consecutive points so that the orientation matches on the outer two, produce a curve passing through a segment of σ and ω with one self-intersection point, and do surgery to get a simple curve with total turning angle at most $4\theta < \pi$. Again it must be non-trivial and non-peripheral. This proves part (1).

We now consider parts (3) and (4). For any $y \in \partial\Omega$, there is some $z \in \omega$ joined to y by a horizontal arc σ which meets ω only in z. Note that σ may pass through a singularity or a puncture (possibly z itself). If so, then a portion of σ may lie on $\partial\Omega$, and contributes at most ϵ to $|\partial\Omega|_{q,h}$. The number of horizontal arcs issuing from a singularity is $1/\pi$ times its cone angle, so this contribution to $|\partial\Omega|_{q,h}$ is bounded by $n_5\epsilon$.

If σ meets no singularities then y is contained in a segment of $\partial\Omega$ parallel to a segment of ω containing z. The total length of such portions of $\partial\Omega$ which are not almost vertical is therefore bounded by $2n_2\epsilon$: there are at most n_2 such segments in ω, they can be approached from either side, and each has length at most ϵ.

Finally for any segment κ of the portion of $\partial\Omega$ which is almost vertical, we note that the segments σ form an embedded parallelogram of width ϵ and height $|\kappa|_{q,v}$. Since q is unit area, we conclude that the vertical length of this portion of the boundary is at most $1/\epsilon$. Its horizontal length is bounded by $(1/\epsilon)\tan\theta < 2\theta/\epsilon$ (assuming $\theta < 1/2$), which is bounded by 2ϵ since $\theta < \epsilon^2$. Putting these together we have a bound $|\partial\Omega|_{q,h} \le (n_5 + 2n_2 + 2)\epsilon$, and $|\partial\Omega|_q \le (n_5 + 2n_2 + 2)\epsilon + 1/\epsilon$, which proves parts (3) and (4).

Finally it remains to prove (2). For $y \in \tau \cap \Omega$, let σ_y be the horizontal segment of length at most ϵ joining y to ω. Suppose that y is at least 2ϵ away from any singularity of q, endpoint of ω or τ, or segment of ω that is not almost vertical. Then (with the assumption $\theta < 1/2$) a segment of the almost-horizontal τ of length 2ϵ must intersect the almost-vertical segment of ω passing through the endpoint of σ, but we have assumed $\tau \cap \omega = \emptyset$. Thus, $\tau \cap \Omega$ is contained in a 2ϵ-neighborhood of the singularities, endpoints and non-almost-vertical segments of ω. There are at most $k = n_1 + n_2 + 2$ of these, each of diameter at most ϵ. Let d be the largest diameter of a component of $\tau \setminus \Omega$. Then the diameter of τ is bounded by $5\epsilon k + (k+1)d$ (the 5ϵ bounds the diameter of a 2ϵ neighborhood of a segment of diameter ϵ). For $\mathrm{diam}_q(\tau) \ge 10\epsilon k$, say, we find that $d \ge \mathrm{diam}_q(\tau)/2(k+1)$, which gives part (2). ☐

Let us also note the following observation which will be used in the proofs of Lemmas 4.8 and 5.5.

Lemma 4.7. *If q is a holomorphic quadratic differential on S and τ is an embedded straight segment in a disk D in S, then $|\partial D|_{q,h} \ge 2|\tau|_{q,h}$. If τ is in a once-punctured disk D' and q has at most a simple pole at the puncture then $|\partial D'|_{q,h} \ge |\tau|_{q,h}$.*

Proof. Consider first a disk D. For any $p \in \tau$ extend a vertical segment σ in both directions until it hits ∂D. This will happen since D is simply connected. It follows immediately that the horizontal length of the portion of ∂D cut off by these segments is equal to twice the horizontal length of D. For a punctured disk D', note that a segment σ could hit τ at both ends, on the same side of τ, if it goes around the puncture. However since the puncture is at most a pole this can only happen on one side of τ, and the vertical segments extended from the other side still give the desired bound. ☐

4.3. **The intersection number lemma.** Our first application of the vertical domain and box lemmas is the following lemma, which states that an almost horizontal and an almost vertical curve which intersect every bounded curve a definite amount also intersect each other proportionally. Compare this fact with Lemma 3.12; the two will be applied together to yield a contradiction.

Lemma 4.8. *Suppose that $\epsilon < \min(\delta/4, a_1^{n_4}\delta/2a_0, a_1^{n_4}\delta/2n_4a_2)$, in addition to previous constraints. There exist $M, h > 0$ such that, given $x \in \mathcal{T}(S)$ and unit area quadratic differential q on (S,x), if β is* *an almost horizontal closed geodesic and γ is an almost vertical closed geodesic with respect to q, and for every non-peripheral simple closed α of q-length at most M,*

$$i(\beta,\alpha) \geq B \text{ and } i(\gamma,\alpha) \geq C, \tag{4.1}$$

then

$$i(\beta,\gamma) \geq hBC$$

Proof. Apply lemma 4.1 to get constants $L, W > 0$ such that either q has a flat cylinder of modulus 2, or an annulus of circumference at least L and radius W. Let $M = \max(1/\sqrt{2}, 1/W, n_4(a_2\epsilon + 1/\epsilon))$. Consider now both possible cases.

Case A. If q has a flat cylinder of modulus 2, let α be the core of this cylinder. Then α has q-length at most $1/\sqrt{2}$, so that (4.1) gives B and C strands of β and γ, respectively, crossing the annulus. It is easy to see that any two nearly orthogonal segments cutting through the annulus must intersect at least once. It follows that $i(\beta,\gamma) \geq BC$, so with $h \leq 1$, we are done.

Case B. If q has an annulus A with circumference at least L and width W, note that A has modulus at least W^2 (since its area is at most 1), and hence $Ext_x(\alpha) \leq 1/W^2$, where α is the core of A. In particular $|\alpha^*|_q \leq 1/W$ where α^* is the q-geodesic representative (see §2.3). Thus γ contains at least C segments (with multiplicity) crossing A and hence of length at least W, and similarly β contains at least B such segments.

Lemma 4.3 guarantees an almost horizontal (cB, δ)-box H for β, with injectivity radius at least δ at its center. Let τ_0 denote the segment of length $\delta/2$ centered on the center of H, parallel to the direction of β. It has the property (recalling $\theta < 1/2$) that any almost-vertical segment that meets τ_0 must cut through all the β strands in H.

Since γ must have length at least CW, we may divide it into at least $CW/L_0 - 1$ pieces, each of which has length at least L_0. (If $CW/L_0 < 2$ we may instead take the whole closed curve γ, and prove the theorem for $C = 1$. The discrepancy is absorbed in the constants.) Each of these is almost vertical, though they may traverse saddle connections of length at most ϵ which are not almost vertical. However, these short segments cannot meet τ_0, since H contains no singularities (here we are using the assumption $\epsilon < \delta/4$). Thus, for each segment that meets τ_0 we obtain rB essential intersections with β. If all of them do meet τ_0, then we are done.

Thus suppose that one segment ω_1 is disjoint from τ_0. Let $X_1 = \Omega_\epsilon(\omega_1)$. Lemma 4.6 guarantees that X_1 generates a nontrivial, nonperipheral subgroup of $\pi_1(S)$. Note that the diameter of τ_0 is $\delta/2$, by the injectivity radius lower bound in H. Thus part (2) of Lemma 4.6 guarantees that a subarc τ_1 of τ_0, with diameter $a_1\delta/2$, is disjoint from X_1 (to apply the Lemma we need $\delta/2 \geq a_0\epsilon$, which is implied by the conditions on ϵ).

Because the q-length of any nontrivial nonperipheral component of ∂X_1 is bounded by $a_2\epsilon + 1/\epsilon \leq M$, γ intersects it essentially, C times. Thus for any component Y of $S \backslash X_1$ which is not a disk or punctured disk, there are C arcs of γ (with multiplicity) passing through Y with both endpoints on ∂Y, which are not deformable back into X_1.

Apply this where Y is the component of $S \setminus X_1$ containing τ_1. This cannot be a disk or punctured disk, because the horizontal length of ∂Y, which is at most $a_2\epsilon$ by Lemma 4.6, is smaller than the length $a_1\delta/2$ of τ_1 by our assumptions on ϵ, and we may apply Lemma 4.7.

Thus, if all C arcs in Y meet τ_1 then as before we have our required intersections between β and γ, and we are done.

If one arc ω_2 is disjoint from τ_1 then define $X_2 = X_1 \cup \Omega_\epsilon(\omega_2)$. We may apply Lemma 4.6 and the same arguments as before to find a subarc τ_2 of τ_1, of diameter $a_1^2\delta/2$, which is disjoint from $\Omega_\epsilon(\omega_2)$, and hence from X_2.

We may continue by induction, generating a sequence $X_1 \subset \cdots X_j \subset X_{j+1}$ and subarcs $\tau_j \subset \tau_1$ of length $a_1^j\delta/2$ disjoint from X_j. At each step, $|\partial X_j|_q$ is incremented by at most $a_2\epsilon + 1/\epsilon$, and $|\partial X_j|_{q,h}$ goes up by at most $a_2\epsilon$. Since X_{j+1} cannot be deformed into X_j the process must terminate within n_4 steps. By our assumption on ϵ we can apply Lemma 4.7 each time so that the component of $S - X_j$ containing τ_j is never a disk or punctured disk. It follows that the only way the process can terminate is by giving C intersections of γ with τ_j for some $j \leq n_4$, which concludes the proof. ☐

5. Proof of the projection theorem

In this section let q be a quadratic differential of area 1 with respect to a conformal structure x on S. let L_q denote the corresponding Teichmüller geodesic. We will denote the Riemann surfaces along L_q by $x_t = L_q(t)$ where t is arclength, and the quadratic differentials by q_t.

Recall that the geodesic gives rise to a map $F_q : \mathbf{R} \to \mathcal{C}$, and a projection $\pi = \pi_q : \mathcal{C} \to \mathbf{R}$ defined as in section 2. To prove the Projection Theorem 2.6 we must show that this projection satisfies the contraction property (Definition 2.2).

We will also assume that our constants ϵ, θ satisfy the assumptions of the Intersection Number Lemma 4.8.

5.1. Bounded adjustments.

We will first need to examine transitions along L_q from mostly vertical to balanced to mostly horizontal curves. As measured by the Teichmüller length parameter, a nearly vertical curve can take a very long time to become balanced. However we find that in a number of crucial situations the transition takes bounded time (independent of q) *as viewed in the curve complex* (that is, when considering quantities such as $\operatorname{diam}_{\mathcal{C}}(F[s,t])$ instead of $|s-t|$).

The relevant insight is illustrated by this sketch of the proof of Lemma 5.5: Consider a very long nearly vertical segment with respect to q_0, which does not fill the whole surface (say it avoids a definite-length horizontal segment). Then if for $t > 0$ the segment is still long and nearly vertical, it fills up some proper subsurface of S which can only shrink as t increases. The boundaries of the resulting sequence of surfaces form a bounded-length sequence in $\mathcal{C}(S)$. This is made precise using the Vertical Domain construction.

Our first observation about the map F is that it is, on a large scale, Lipschitz:

Lemma 5.1. (Lipschitz) *There exist $C, D > 0$ such that for any q, t_1 and t_2 we have*

$$d_{\mathcal{C}}(F_q(t_1), F_q(t_2)) \leq C|t_2 - t_1| + D.$$

Proof. As in Lemma 2.4, let $e_0(S)$ be such that for any conformal structure on S there is a curve with extremal length at most e_0. Suppose that $|t_2 - t_1| \leq 1$. Let α_i be a curve of shortest extremal length for x_{t_i}, for $i = 1, 2$. Then $Ext_{x_{t_1}}(\alpha_2) \leq e^2 e_0$ (by (2.1)). A bound on $d_{\mathcal{C}}(\alpha_1, \alpha_2)$ follows from Lemma 2.5.

The case where $|t_2 - t_1| > 1$ follows by subdividing. ☐

In the exceptional cases of projecting curves that are entirely horizontal or vertical, we observe the following:

Proposition 5.2. *If $\beta \in \mathcal{C} \setminus \mathcal{C}_b(q)$ then $d_{\mathcal{C}}(\beta, F_q(\pi_q(\beta))) \leq 1$.*

Proof. Assume that β is vertical. Let Σ denote the union of compact singular leaves of the vertical foliation of q, and let Σ_ϵ denote a regular neighborhood of Σ. Then it is not hard to see (e.g. [32]) that for any non-peripheral curve γ in S the extremal length $Ext_{x_t}(\gamma)$ remains bounded as $t \to +\infty$ if and only if γ can be deformed into Σ, and $Ext_{x_t}(\gamma) \to 0$ as $t \to +\infty$ if and only if γ is homotopic to a boundary component of Σ_ϵ.

It follows that $F_q(+\infty) = F_q(\pi_q(\beta))$ is one of these boundary components, and since β is in Σ, we obtain $d_{\mathcal{C}}(\beta, F_q(\pi_q(\beta))) \leq 1$. If β is horizontal we make a similar argument, reversing the t-direction. ☐

In Lemmas 5.3-5.7 we will find a series of constants d_1-d_5 that depend only on the topology of S and not on the particular q. The next lemma shows that the image under the map F of the set of t where a curve α is close to its minima has bounded diameter in $\mathcal{C}(S)$. In particular, once the q_t-length of α is sufficiently short, we only need to wait a bounded amount until it starts to grow again.

Lemma 5.3. *There exist $\epsilon_1, d_1 > 0$, depending only on the topology of S, with the following property. If α is a closed q-geodesic homotopic to a simple curve, let*

$$J = \{t : |\alpha|_{q_t} \leq \epsilon_1\}.$$

Then $\operatorname{diam}_{\mathcal{C}}(F(J)) \leq d_1$.

(Note that J is a bounded interval in $\mathbf{R}$ *unless α is completely vertical or horizontal.)*

Proof. Let $\epsilon_1 = 2W$. By Lemma 4.1, for each t there is a nonperipheral curve β_t with a collar neighborhood of q_t-width W. Since for $t \in J$, $|\alpha|_{q_t} \leq 2W$, we may conclude that $i(\alpha, \beta_t) = 0$. Hence for any $t, s \in J$, $d_{\mathcal{C}}(\beta_t, \beta_s) \leq 2$.

The existence of the collar implies $Ext_{x_t}(\beta_t) \leq 1/W^2$. Hence by Lemma 2.5, we conclude $d_{\mathcal{C}}(\beta_t, F(t)) \leq 1/W^2 + 1$ for $t \in J$.

It follows that $d_{\mathcal{C}}(F(s), f(t)) \leq 2/W^2 + 4$, and we set d_1 accordingly. ☐

The following lemma will allow us to convert a long almost-horizontal arc which has small diameter (i.e. winds around tightly) to one which has a definite diameter, within bounded distance in $\mathcal{C}$.

Lemma 5.4. *There exist constants $\epsilon_3 > \epsilon_2 > 0$ and $d_2 > 0$, depending on the topology of S and the initial choice of ϵ, θ, so that the following holds. Let τ be an almost horizontal straight segment with respect to q, of length $|\tau|_q \geq \epsilon_3$. Let J be the interval*

$$J = \{t \geq 0 : \operatorname{diam}_{q_t}(\tau) < \epsilon_2\}$$

and suppose $0 \in J$. Then $\operatorname{diam}_{\mathcal{C}}(F(J)) \leq d_2$.

Proof. For any t let β_t be the homotopy class of the core of the annulus of width W given by Lemma 4.1. Let $\epsilon_2 = W - 2\epsilon$.

Let Ψ^t denote the horizontal domain $\Psi_\epsilon(\tau)$ with respect to q_t (see Definition 4.5). Then $\operatorname{diam}_{q_t}(\Psi^t) \leq \operatorname{diam}_{q_t}(\tau) + 2\epsilon < W$ for $t \in J$. Thus any closed curve in Ψ^t has 0 intersection number with β_t. We will show that if $|\tau|_q$ is sufficiently large, there exists a nontrivial, nonperipheral curve κ which is contained in Ψ^t for all $t \in J$.

We use an argument similar to the proof of Lemma 4.6 part (1). Recall the vertical intervals σ_x of radius ϵ around nonsingular $x \in \tau$ from the definition of Ψ^0. For any $y \in S$ let $f(y) = \#\{x \in \tau : y \in \sigma_x\}$. Then $\int_{\Psi^0} f(y)$ with respect to q-area is $2\epsilon|\tau|_{q,h}$. On the other hand the integral is at most $\operatorname{Area}_q(\Psi^0) \max f$, so that

$$\max f \geq \frac{2\epsilon}{\operatorname{Area}_q(\Psi^0)}|\tau|_{q,h} \geq \frac{\epsilon}{\operatorname{Area}_q(\Psi^0)}|\tau|_q,$$

where the second inequality is due to τ being almost horizontal. We also have $\mathrm{Area}_q(\Psi^0) \le n_3 \,\mathrm{diam}_q(\Psi^0)^2 \le n_3 W^2$ where n_3 was defined in §4.1. Thus we have $\max f \ge \epsilon|\tau|_q/(n_3 W^2)$. Set $\epsilon_3 = 3n_3W^2/\epsilon$, and now $|\tau|_q \ge \epsilon_3$ implies $\max f \ge 3$. As in Lemma 4.6 we conclude that Ψ^0 contains a nontrivial nonperipheral curve κ.

Since lengths in the vertical direction shrink as t increases, for all $t \ge 0$ we have $\Psi^0 \subset \Psi^t$. Thus κ is in all the Ψ^t, and hence must have 0 intersection number (hence $\mathcal{C}$-distance 1) with all β_t for $t \in J$, as above. As in Lemma 5.3, $d_{\mathcal{C}}(\beta_t, F(t)) \le 1/W^2 + 1$, so it follows that $F(t)$ is within bounded $\mathcal{C}$-distance of κ for all $t \in J$. □

The next lemma shows that if an almost vertical straight segment misses an almost horizontal segment of definite length, then after a bounded wait as measured in the curve complex, it will either be very short, or almost horizontal.

Lemma 5.5. *In addition to our previous assumptions suppose we also have $\epsilon < \min(\epsilon_2/a_0, \epsilon_2 a_1/a_2)$. There is a number $d_3 = d_3(\epsilon, \theta)$ with the following property. Suppose α is a straight segment of q disjoint from an almost horizontal straight segment τ of length ϵ, let*

$$J = \{t \ge 0 : |\alpha|_{q_t} > \epsilon_1 \text{ and } \alpha \text{ not almost horizontal with respect to } q_t\}.$$

Then $\mathrm{diam}_{\mathcal{C}}(F(J)) \le d_3$.

Remark. Since ϵ_2 was given as $W - 2\epsilon$ in Lemma 5.4, it is evident that our added conditions of the form $\epsilon < C\epsilon_2$ are satisfied for ϵ sufficiently small.

Proof. If α is not almost vertical, then for a bounded T (depending on ϵ, θ) it will be almost horizontal with respect to q_T. By Lemma 5.1, $F([0,T])$ has bounded diameter in $\mathcal{C}$.

Thus we may assume α is almost vertical to begin. Suppose its length is at most L_0. Then for $T = \log 2L_0/\epsilon_1$, its q_T-vertical length is at most $\epsilon_1/2$; thus it either has q_T-length less than ϵ_1 or it is not almost vertical. In the first case we have satisfied the conclusion of the Lemma, again bounding $\mathrm{diam}_{\mathcal{C}}(F([0,T]))$ by Lemma 5.1. In the second case we are also done by the argument in the first paragraph.

Thus finally assume α is almost vertical and has length greater than L_0. Since $|\tau|_q = \epsilon$ and τ is almost horizontal, for $t_1 = \log 2\epsilon_3/\epsilon$ we have $|\tau|_{q_{t_1}} \ge \epsilon_3$. Lemma 5.4 implies that either $\mathrm{diam}_{\mathcal{C}}(F([t_1,\infty))) \le d_2$ in which case we are done, or there is a $t_2 > t_1$, with $\mathrm{diam}_{\mathcal{C}}(F([t_1,t_2])) \le d_2$, so that $\mathrm{diam}_{q_{t_2}}(\tau) \ge \epsilon_2$.

Now if α is not almost vertical or has length at most L_0 with respect to q_{t_2}, we are done by the above cases. Otherwise, we construct the vertical domain $\Omega_1 = \Omega_\epsilon(\alpha)$ with respect to q_{t_2}. Since α is disjoint from τ and (by assumption) $\epsilon_2 > a_0\epsilon$, Lemma 4.6 gives a subarc τ_1 of τ of length $a_1\epsilon_2$, disjoint from Ω_1. The total horizontal length of $\partial\Omega$ is bounded by $a_2\epsilon$ by part (4) of Lemma 4.6. Thus, since $a_2\epsilon < a_1\epsilon_2$ and applying Lemma 4.7, we conclude that the component Y of $S \setminus \Omega$ containing τ_1 cannot be a disk or once-punctured disk. Thus $\partial\Omega_1$ has nontrivial and nonperipheral components. For each such component σ, we have $|\sigma^*|_{q_{t_2}} \le \ell_0 = a_2\epsilon + 1/\epsilon$ by part (3) of Lemma 4.6, where σ^* is the geodesic representative. This bound means that within bounded Teichmüller distance either $|\sigma^*|_{q_t}$ reaches ϵ_1, or it starts to increase. Applying Lemma 5.3, we conclude that either the remaining $\mathrm{diam}_{\mathcal{C}}(F([t_2,\infty)))$ is bounded, in which case we are done, or there is t_3 with bounded $\mathrm{diam}_{\mathcal{C}}(F([t_2,t_3]))$, such that for $t > t_3$, $|\sigma^*|_{q_t} > \epsilon_1$. It follows that for an additional t_4 with $t_4 - t_3$ bounded, $|\sigma^*|_{q_{t_4}} \ge 2\ell_0$.

We can now repeat the argument: There is a t_5 such that $\mathrm{diam}_{\mathcal{C}}(F([t_4,t_5]))$ is bounded, so that with respect to q_{t_5} we either have the desired condition for α, or α is almost vertical, of length at least L_0, and τ_1 now has diameter at least ϵ_2. Thus Lemma 4.6 again gives a vertical domain $\Omega_2 = \Omega_\epsilon(\alpha)$ with respect to q_{t_5} whose boundary components have q_{t_5}-length at most ℓ_0. Since our previous boundary components σ now have $|\sigma^*|_{q_{t_5}} \geq 2\ell_0$, no nontrivial nonperipheral component of $\partial\Omega_2$ is homotopic to σ. The vertical domains decrease monotonically as t increases, so we conclude that $i_*(\pi_1(\Omega_2))$ is a proper subgroup of $i_*(\pi_1(\Omega_1))$.

We may repeat this procedure, obtaining a sequence $\Omega_{j+1} \subset \Omega_j$ which terminates in at most n_4 steps, at which point α has length less than ϵ_1 or is almost horizontal, as desired, or we find that a remaining interval $[t, \infty)$ has bounded-diameter image. □

From now on let us assume that ϵ, θ satisfy the conditions of Lemma 5.5 as well as the previous conditions.

The following lemma shows that if a curve is balanced at $L_q(0)$, then in the forward direction it will become almost horizontal after an interval of bounded size in the curve complex.

Lemma 5.6. (Almost Horizontal) *There exists $d_4 = d_4(\epsilon, \theta)$ so that if β is balanced with respect to q and*

$$J = \{t \geq 0 : \beta \text{ is not almost horizontal with respect to } q_t\},$$

then $\operatorname{diam}_{\mathcal{C}}(F(J)) \leq d_4$.

Proof. Let n_3 be the bound for the number of mutually disjoint saddle connections one can have in S. Thus, β runs through at most n_3 saddle connections, although some may be traversed arbitrarily many times.

Since $|\beta|_{q,h} = |\beta|_{q,v}$, for $t > t_1 = \frac{1}{2}\log 1/\theta$ we have $|\beta|_{q_t,h} > \frac{1}{\theta}|\beta|_{q_t,v}$. It follows that if $\beta_{0,t}$ is the subset of β which traverses almost-horizontal arcs with respect to q_t, we have $|\beta_{0,t}|_{q_t} \geq |\beta \setminus \beta_{0,t}|_{q_t}$. Now by Lemma 5.3, we have t_2 with bounded $\operatorname{diam}_{\mathcal{C}}(F([t_1, t_2]))$ such that $|\beta|_{t_2} \geq \epsilon_1$, and therefore $|\beta_{0,t_2}|_{q_{t_2}} \geq \epsilon_1/2$. It may still be that this length is obtained by traversing many times a very short almost horizontal curve σ, but applying Lemma 5.3 again we obtain t_3 with bounded $\operatorname{diam}_{\mathcal{C}}(F([t_2, t_3]))$ such that $|\sigma|_{q_{t_3}} \geq \epsilon_1$.

In particular β contains an embedded straight segment τ which is almost-horizontal with respect to q_{t_3} and of length at least ϵ.

Now suppose β is not almost horizontal for q_{t_3}, so that there is a segment β_1 of β which has length at least ϵ and is not almost horizontal. Since β has no self intersections, β_1 is disjoint from τ. Applying Lemma 5.5, there exists t_4 with $\operatorname{diam}_{\mathcal{C}}(F([t_3, t_4])) \leq d_3$, such that if $t > t_4$ then either β_1 is almost horizontal with respect to q_t, or $|\beta_1|_{q_t} \leq \epsilon_1$. In the latter case, set $t_5 = t_4 + \log 2\epsilon_1/\epsilon\theta$, and note that for $t > t_5$, β_1 will either have length less than ϵ or be almost horizontal. Apply this to all of the saddle connections of β. □

The next lemma shows that, unless a curve is almost vertical in $L_q(0)$, it can be balanced in the forward direction after an interval of bounded size in the curve complex.

Lemma 5.7. (Almost Vertical) *There exists $d_5 = d_5(\epsilon, \theta)$ such that if γ is not almost vertical with respect to q then for*

$$J = \{t \geq 0 : |\gamma|_{q_t,v} > |\gamma|_{q_t,h}\}$$

we have $\operatorname{diam}_{\mathcal{C}}(F(J)) \leq d_5$.

Proof. Since γ is not almost vertical with respect to q, it contains a segment τ that is not almost vertical and has length at least ϵ. For $t_1 = \log 2/\theta$, τ will be almost horizontal and have length at least ϵ with respect to q_{t_1}. We can assume that there is a set of almost vertical saddle connections $\omega \subset \gamma$ that carry at least 1/2 of the length of γ, for otherwise $|\gamma|_{q_t,h}$ would dominate for $t > t_2$ for a bounded t_2. Obviously each ω is disjoint from τ, since γ does not have self intersections.

By Lemma 5.5, there is $t_3 > t_1$ with $\text{diam}_{\mathcal{C}}(F([t_1, t_3])) \le d_3$ such that either ω is almost horizontal or has length at most ϵ_1 with respect to q_{t_3}. If all the ω's are in the former case we are done since then γ would have been balanced for $t \le t_3$.

Suppose, then, that some ω has length smaller than ϵ_1. If we follow any strand of γ starting at ω until it returns with the same orientation, we may find in the set of saddle connections it traverses a geodesic loop homotopic to a simple loop. There is a bound (in terms of the number of possible saddle connections n_2) on the number of such loops γ', and hence there must be some γ' each of whose saddle connections are traversed at least a definite fraction of the number of times ω is traversed.

Either γ' contains an almost horizontal saddle connection, or it has length at most $n_2\epsilon_1$. In that case, within bounded $t > t_3$ it will either have length ϵ_1 or begin to grow, and we may apply Lemma 5.3 to get a t_4 with bounded $\text{diam}_{\mathcal{C}}(F([t_3, t_4]))$, such that for $t > t_4$ $|\gamma'|_{q_t} \ge \epsilon_1$ and $|\gamma'|_{q_t}$ is growing, which implies that after an additional bounded t_5, it will be balanced.

Thus for $t > t_5$, the contribution of ω to γ is either itself horizontal or offset (to within a bounded factor) by the other segments in γ', and after another bounded interval $[t_5, t_6]$ we have balance. □

5.2. **Completion of proof.** We must show that all three conditions of definition 2.2 hold for our projection π_q, with suitable constants a, b and c (independent of the geodesic L_q).

In what follows, fix q and $x = L_q(0)$. Let $\alpha = F_q(0)$ be a shortest curve on x. Assume $\pi_q(\beta) = 0$ so that β is balanced at 0. We assume that all curves are q-geodesics (hence q_t-geodesics for any t), and furthermore that ϵ, θ satisfy the conditions in Lemma 4.8 and Lemma 5.5.

Let us restate the conditions in our current terminology, and prove them.

Condition (1): $\text{diam}_{\mathcal{C}}(F_q([0, \pi_q(F_q(0))])) \le c$.

We may assume $|\alpha|_{q,v} > |\alpha|_{q,h}$, or equivalently that the balance point $\pi_q(\alpha)$ is positive. Since α has minimal extremal length with respect to x, we have a bound $|\alpha|_q \le Ext_x(\alpha)^{1/2} \le \sqrt{e_0}$. Thus for bounded t_1, the vertical length $|\alpha|_{q_{t_1},v}$ becomes ϵ_1, so either α is balanced for $t \le t_1$, in which case we are done, or $|\alpha|_{q_{t_1}} \le \epsilon_1$. In the latter case we apply Lemma 5.3 to see that either α is vertical, in which case $\text{diam}_{\mathcal{C}}(F([0,\infty)))$ is bounded and $\pi_q(\alpha) = +\infty$, so we are done, or there is t_2 with $\text{diam}_{\mathcal{C}}(F([t_1, t_2]))$ bounded, and $|\alpha|_t$ is increasing after t_2, so it is balanced for some $t < t_2 + \frac{1}{2}\cosh^{-1}\sqrt{2}$ and again we are done.

Condition (2): If $d_{\mathcal{C}}(\beta, \gamma) \le 1$ then $\text{diam}_{\mathcal{C}}(F_q([\pi_q(\beta), \pi_q(\gamma)])) \le c$.

Recall $\pi_q(\beta) = 0$. Assume $\pi_q(\gamma) > 0$, so $|\gamma|_{q,v} > |\gamma|_{q,h}$. Since β is balanced at 0, Lemma 5.6 gives $t_1 > 0$ with $\text{diam}_{\mathcal{C}}(F_q([0, t_1]))$ bounded, such that β is almost horizontal with respect to q_{t_1}. Lemma 5.3 then gives t_2 with $\text{diam}_{\mathcal{C}}(F_q([t_1, t_2]))$ bounded, so that $|\beta|_{q_{t_2}} \ge \epsilon_1$. Thus there is a $t_3 \ge t_2$ with $F_q[t_2, t_3]$ bounded, so that β has an almost horizontal segment of length ϵ with respect to q_{t_3}.

Now, either γ is already balanced for $t \le t_3$, in which case we are done, or it is still mostly vertical with respect to q_{t_3}. In this case, since γ is disjoint from β it misses the horizontal segment of length ϵ and Lemma 5.5 gives t_4 with $\text{diam}_{\mathcal{C}}(F_q([t_3, t_4]))$ bounded, so that every saddle connection of γ is either almost horizontal or has length at most ϵ_1 with respect to q_{t_4}. Thus for t_5 with bounded $t_5 - t_4$, either the length of γ shrinks to ϵ_1, or it begins to increase so γ is balanced for $t \le t_5$. In the former case, Lemma 5.3 says that the segment J of $t > t_5$ where $|\gamma|_{q_t} \le \epsilon_1$ has bounded-diameter image in $\mathcal{C}$. This includes the case where γ is completely vertical and $\pi_q(\gamma) = +\infty$. In all other cases, γ will be balanced for some $t \in J$ and again we are done.

Condition (3): If $d_{\mathcal{C}}(\beta, F_q(\pi_q(\beta))) \ge a$ and $d_{\mathcal{C}}(\beta, \gamma) \le b d_{\mathcal{C}}(\beta, F_q(\pi_q(\beta)))$ then $\text{diam}_{\mathcal{C}}\, F_q([\pi_q(\beta), \pi_q(\gamma)]) \le c$.

Recall that $F_q(\pi_q(\beta)) = F_q(0) = \alpha$. Assume without loss of generality that γ is more vertical than horizontal at q_0. By Proposition 5.2, we can assume that $\beta \in \mathcal{C}_b$, for otherwise its distance from the image of F is at most 1. By Lemma 5.6 (Almost Horizontal), there is some $t_1 > 0$ with $\mathrm{diam}_{\mathcal{C}}(F_q[0,t_1])$ bounded such that β is almost horizontal at q_{t_1}.

We next show that γ cannot be almost vertical at q_{t_1}. Let M, h be the constants given by lemma 4.8, and suppose by contradiction that γ is almost vertical. If α' is any curve of q_{t_1}-length at most M, there is a bound $d(M)$ on $d_{\mathcal{C}}(\alpha', \alpha)$ by the following: Lemma 4.1 gives a nonperipheral annulus A of width W so that the intersection of α' with its core σ is at most M/W. Lemma 2.1 then bounds $d_{\mathcal{C}}(\alpha', \sigma)$. Lemma 2.5 in turn bounds $d_{\mathcal{C}}(\sigma, F_q(t_1))$ since both have bounded extremal length. Finally $d_{\mathcal{C}}(F_q(t_1), F_q(0)) = d_{\mathcal{C}}(F_q(t_1), \alpha)$ is bounded by choice of t_1.

Now applying Lemma 3.12 with $Q = 2/h$, and $k = d(M)$, we obtain, provided $d_{\mathcal{C}}(\alpha, \beta) \geq D_3$ and $d_{\mathcal{C}}(\gamma, \beta) \leq \nu d_{\mathcal{C}}(\alpha, \beta)$, that

$$\min_{\alpha'} i(\beta, \alpha') \min_{\alpha'} i(\gamma, \alpha') \geq Q i(\beta, \gamma)$$

where α' varies over all curves of q_{t_1}-length at most M. On the other hand, Lemma 4.8 gives the opposite inequality

$$i(\gamma, \beta) \geq h \min_{\alpha'} i(\beta, \alpha') \min_{\alpha'} i(\gamma, \alpha').$$

This is a contradiction since we have chosen $Q > 1/h$, and we conclude that γ cannot be almost vertical at t_1. Thus by Lemma 5.7 (Almost Vertical), there exists t_2 with $\mathrm{diam}_{\mathcal{C}}(F_q([t_1, t_2]))$ bounded such that γ is balanced at t_2. This concludes the proof of Theorem 2.6.

6. Contraction property and hyperbolicity

To complete the proof of Theorem 1.1, it remains to prove Theorem 2.3, that if a geodesic metric space X has a coarsely transitive path family Γ with the contraction property then X is hyperbolic.

For our purposes a path $\gamma : I \to X$ is a *quasi-geodesic* if the following inequality holds for any $x, y \in I$:

$$\mathrm{length}_s(\gamma[x,y]) \leq K d_X(\gamma(x), \gamma(y)) + \delta$$

where $K \geq 1$ and $\delta, s \geq 0$ are fixed constants, and length_s for $s > 0$ is "arclength on the scale s", which is defined as follows: $\mathrm{length}_s(\gamma[x,y]) = sn$ where n is the smallest number for which $[x,y]$ can be subdivided into n closed subintervals $J_1, \ldots, J_n$ with $\mathrm{diam}_X(\gamma(J_i)) \leq s$. (This definition circumvents the need for checking the behavior of the parametrization at small scale; we let length_0 denote normal length). Note also that the opposite inequality $d_X(\gamma(x), \gamma(y)) \leq \mathrm{length}_s(\gamma[x,y])$ holds automatically.

The proof is in two steps. We say that X has *stability of quasi-geodesics* if for all $K \geq 1, \delta, s \geq 0$ there exists $R > 0$ such that any (K, δ, s)-quasi-geodesic $\alpha : I \to X$ with endpoints x, y remains in an R-neighborhood of any geodesic $[xy]$.

Lemma 6.1. *If X has a coarsely transitive path family Γ with the contraction property then X has stability of quasi-geodesics. In addition, the paths of Γ themselves are uniform quasi-geodesics.*

Lemma 6.2. *Stability of quasi-geodesics implies hyperbolicity.*

Proof of Lemma 6.1. We may assume that the path family Γ is transitive, since for paths of length bounded by a fixed D it is easy to define a contracting projection, simply by mapping all of X to one endpoint.

Consider $\gamma : [0, M] \to X$ in Γ, and let $\alpha : [0, L] \to X$ be a (K, δ, s)-quasi-geodesic such that $\alpha(0) = \gamma(0)$ and $\alpha(L) = \gamma(M)$. Note that a $(K, \delta, 0)$-quasi-geodesic is also a $(K, \delta + s, s)$-quasi-geodesic for any $s > 0$ since $\mathrm{length}_s \leq \mathrm{length}_0 + s$. Thus from now on we may assume $s > 0$.

We show that α remains in a $R(K,\delta,s)$-neighborhood of γ. The proof is somewhat complicated by the fact that we do not assume continuity of α, γ or π, but the idea is simple and well-known: large excursions of α away from γ can, using the contraction property, be circumvented by short cuts that travel along the projection to γ.

Let $r(u) = d(\alpha(u), \gamma(\pi(\alpha(u))))$. We will bound $r(u)$ uniformly in terms of K, δ, s, and the constants a, b and c of the contraction property (Definition 2.2).

Divide $[0, L]$ into closed intervals $J_1, \ldots, J_n$ such that $ns = \operatorname{length}_s(\alpha[0, L])$ and $\operatorname{diam} \alpha(J_i) \leq s$. Then by part (2) of definition 2.2, $\operatorname{diam} \gamma(\pi(\alpha(J_i))) \leq s'$, where $s' = c$ if $s \leq 1$ and $s' = 1 + cs$ if $s > 1$.

Fix $R_0 > 0$, to be determined shortly. For any $u \in [0, L]$, if $r(u) \geq R_0 + s'$ then u is contained in some interval $J = [u_0, u_1]$, a union of J_i, such that $r \geq R_0$ in J and $r \leq R_0 + s'$ at u_0 and u_1. Subdivide J into intervals $K_1, \ldots, K_m$, each a union of at most bR_0/s of the J_i, so that for each j $\operatorname{diam} \alpha(K_j) \leq bR_0$, and the number m is at most $1 + \operatorname{length}_s(\alpha(J))/bR_0$. Now assuming $R_0 \geq a$ and applying the contraction property (part 3) to each of these we obtain

$$\operatorname{diam} \gamma([\pi(\alpha(u_0)), \pi(\alpha(u_1))]) \leq mc$$

and by the triangle inequality

$$d(\alpha(u_0), \alpha(u_1)) \leq 2(R_0 + s') + \left(1 + \frac{\operatorname{length}_s(\alpha(J))}{bR_0}\right) c. \tag{6.1}$$

Since α is a (K, δ, s)-quasi-geodesic, $\operatorname{length}_s(\alpha(J)) \leq K d(\alpha(u_0), \alpha(u_1)) + \delta$. Combining with (6.1), we get

$$\operatorname{length}_s(\alpha(J)) \leq \frac{Kc}{bR_0} \operatorname{length}_s(\alpha(J)) + 2Kc(R_0 + s') + \delta. \tag{6.2}$$

Make the (a priori) choice of R_0 sufficiently large that $Kc/bR_0 < 1/2$. Then (6.2) gives an upper bound R on $\operatorname{length}_s(\alpha(J))$ depending only on the initial constants.

Thus $d(\alpha(u), \{\alpha(u_0), \alpha(u_1)\})$ is at most $R/2$, and in particular, no point in $\alpha(J)$ can be further than $R_0 + R/2$ from $\gamma([0, M])$. Furthermore by applying part (2) of the contraction property, it follows that $r(u)$ is bounded uniformly.

This implies that we can project from γ back to α, in the following sense: For any $t \in [0, M]$ we can find $u \in [0, L]$ such that $d(\gamma(t), \gamma(\pi(\alpha(u))))$ is bounded by a uniform constant, just by chopping α into bounded-length pieces and applying parts (1) and (2) of the contraction property. Now by the bound on $r(u)$ we can bound $d(\gamma(t), \alpha(u))$ uniformly.

Apply this to an actual geodesic α and a quasi-geodesic β with the same endpoints. Letting $\gamma \in \Gamma$ be a path with the same endpoints, we project from β to γ and then from γ to α as above. Both steps move a bounded distance, so we conclude that β lies in a bounded neighborhood of α. Hence, we have stability of quasi-geodesics. □

Proof of Lemma 6.2. To prove hyperbolicity it suffices to establish the thin triangle condition. Let x, y, z be three points in X. We must show that $[xy]$ lies in a δ-neighborhood of $[xz] \cup [yz]$, for uniform δ.

Let $z' \in [xy]$ be a point that minimizes distance from z to $[xy]$. We claim that the broken geodesic $[xz'] \cup [z'z]$ is a $(3, 0, 0)$-quasi-geodesic. If $z' = x$ this is obvious, so assume $z' \neq x$. Let u lie in $[xz']$ and v lie in $[z'z]$.

It follows from the choice of z' that it also minimizes distance from $[v]$ to $[xy]$ (via the triangle inequality). Thus $d(u, v) \geq d(z', v)$.

By the triangle inequality, $d(u,v) \geq d(u,z') - d(z',v)$. Thus adding this to twice the previous inequality we get $3d(u,v) \geq d(z',v) + d(u,z')$. This is exactly the fact that length$([uz'] \cup [z'v])$ estimates $d(u,v)$, so we conclude $[xz'] \cup [z'z]$ is a $(3,0,0)$-quasi-geodesic.

Now by stability of quasi-geodesics, we have that $[xz'] \cup [z'z]$ is in a uniform δ-neighborhood of $[xz]$, and in particular $[xz']$ is. Applying the same argument for y replacing x, we see that all of $[xy]$ is in a δ-neighborhood of $[xz] \cup [yz]$. This concludes the proof. □

7. Relative Hyperbolicity

In this final section we establish Theorems 1.2 and 1.3, which provide an interpretation of our hyperbolicity theorem in terms of the geometry of Teichmüller space, and the structure of the Mapping Class Group.

The following terminology is due to Farb [11]: If X is any geodesic metric space and $\mathcal{H}$ is a family of regions in X, let the *electric distance* d_e on X be the path metric imposed by shrinking each $H \in \mathcal{H}$ to diameter 1, in the following way: For each $H \in \mathcal{H}$ create a new point c_H and an interval of length 1/2 from c_H to every point in H. The new metric is induced by shortest paths in this enlarged space $\hat{X}$ (called the *electric space*). We say X is *relatively hyperbolic* with respect to $\mathcal{H}$ if $(\hat{X}, d_e)$ is δ-hyperbolic for some δ.

7.1. In Teichmüller space.

Fixing $\epsilon_0 > 0$ sufficiently small that the Collar Lemma holds for ϵ_0, let $\mathcal{H}_C = \{H_\alpha\}_{\alpha \in \mathcal{C}_0(S)}$ denote the family of regions in $\mathcal{T}(S)$ defined as in the introduction:

$$H_\alpha = \{x \in \mathcal{T}(S) : Ext_x(\alpha) < \epsilon_0\}.$$

Then it is easy to see that a set of points $\alpha_1, \ldots, \alpha_k$ is a simplex in $\mathcal{C}(S)$ if and only if $H_{\alpha_1} \cap \cdots \cap H_{\alpha_k}$ is non-empty. In other words, $\mathcal{C}(S)$ is the *nerve* of the family $\mathcal{H}_C$.

The statement of Theorem 1.2 is a direct consequence of Theorem 1.1 and the following:

Lemma 7.1. *The electric space $(\hat{\mathcal{T}}(S), d_e)$ defined with respect to the family $\mathcal{H}_C$ is quasi-isometric to $\mathcal{C}_1(S)$.*

Proof. There is a natural map $\varphi : \mathcal{C}_0(S) \to \hat{\mathcal{T}}(S)$ taking each α to the new point $c_\alpha \equiv c_{H_\alpha}$. The set $\mathcal{C}_0(S)$ is clearly 1/2-dense in $\mathcal{C}_1(S)$. Let us check that its image $\{c_\alpha\}$ is d_0-dense in $\hat{\mathcal{T}}(S)$, for some $d_0 < \infty$.

Recall that for any conformal structure x on S there is a curve $\alpha \in \mathcal{C}_0(S)$ with $Ext_x(\alpha) \leq e_0$. Then for this α, we see that x is a bounded Teichmüller distance (in fact $\frac{1}{2}\log(e_0/\epsilon_0)$) from H_α: we may apply to x a Teichmüller map whose vertical foliation consists of leaves homotopic to α. It follows that $\{c_\alpha\}$ is $(\frac{1}{2} + \frac{1}{2}\log(e_0/\epsilon_0))$-dense in $\hat{\mathcal{T}}(S)$.

Now we need only show that for any $\alpha, \beta \in \mathcal{C}_0(S)$

$$\frac{1}{K} d_{\mathcal{C}}(\alpha,\beta) - a \leq d_e(c_\alpha, c_\beta) \leq K d_{\mathcal{C}}(\alpha,\beta) + a \tag{7.1}$$

with fixed $K, a > 0$, to show that φ induces a quasi-isometry. One direction is easy: if $d_{\mathcal{C}}(\alpha,\beta) = 1$ then $H_\alpha \cap H_\beta$ is nonempty, and any point x in this set is connected to each of c_α and c_β by a segment of length 1/2. Hence φ is 1-Lipschitz.

To obtain the other direction, consider for any $x \in \mathcal{T}(S)$ the set $\Phi(x)$ of elements in $\mathcal{C}_0(S)$ of minimal Ext_x. This set has diameter at most $2e_0 + 1$ by Lemma 2.4. Now if $d_{\mathcal{T}}(S)(x,y) \leq 1$ we see also that $\Phi(x) \cup \Phi(y)$ has bounded diameter, by Lemma 5.1. Note also that if $x \in H_\alpha$ then $d_{\mathcal{C}}(\alpha, \Phi(x)) \leq 1$.

Thus, any map that associates to $x \in \mathcal{T}(S)$ some (any) element of $\Phi(x)$ and to c_α associates α will expand distances by a bounded multiplicative and additive amount, and serve as an inverse to φ. It follows that φ is a quasi-isometry. □

7.2. **In the Mapping Class Group.** To carry out a similar analysis for $\mathrm{Mod}(S)$, recall first that for any group G with a fixed finite generating set Γ, the Cayley graph $\mathcal{G} = \mathcal{G}_{G,\Gamma}$ is a 1-complex whose vertex set is G and whose edges are all pairs $(g, g\gamma)$ with $\gamma \in \Gamma$. Giving all edges length 1, we obtain a complete locally finite geodesic metric space.

Now for $G = \mathrm{Mod}(S)$, we single out a number of subgroups as follows. Up to the action of $\mathrm{Mod}(S)$, there are only a finite number of distinct non-trivial non-peripheral homotopy classes of simple curves in S (distinguished by the topological type of their complement). Let $\{\alpha_1, \ldots, \alpha_N\}$ be a fixed list of representatives of these $\mathrm{Mod}(S)$-orbits. Let $Fix(\alpha_j)$ be the subgroup of $\mathrm{Mod}(S)$ fixing α_j.

Given any $\beta \in \mathcal{C}_0(S)$, let α_j be the unique representative of β in the list, and let G_β be the left-coset of $Fix(\alpha_j)$ defined by $G_\beta = \{g \in \mathrm{Mod}(S) : g(\alpha_j) = \beta\}$.

Now we may form the electric space $\hat{\mathcal{G}}$ of $\mathcal{G}$ relative to the family of cosets $\{G_\beta\}$, and its electric distance d_e. The analogue to Lemma 7.1 is:

Lemma 7.2. *Fixing a choice of generating set Γ and representatives $\{\alpha_1, \ldots, \alpha_N\}$ of $\mathrm{Mod}(S)$-orbits in $\mathcal{C}_0(S)$, the electric space $(\hat{\mathcal{G}}, d_e)$ is quasi-isometric to $\mathcal{C}_1(S)$.*

Again, this together with Theorem 1.1 proves Theorem 1.3, where the relative hyperbolicity of $\mathrm{Mod}(S)$ is with respect to this family of cosets $\{G_\beta\}$.

Proof. Let c_β denote the new point added to G_β in the construction of $\hat{\mathcal{G}}$. The natural map $\varphi : \mathcal{C}_0(S) \to \hat{\mathcal{G}}$ is again $\varphi(\beta) = c_\beta$. In this case it is clear that $\{c_\beta\}$ is 1/2-dense in $\hat{\mathcal{G}}$ since every $g \in \mathrm{Mod}(S)$ is in the coset $gFix(\alpha_j) = G_{g(\alpha_j)}$ for each $j \leq N$. It remains to check that the inequalities (7.1) hold.

Up to the action of $\mathrm{Mod}(S)$ there are only finitely many pairs (β, β') of disjoint curves in $\mathcal{C}_0(S)$ (i.e. edges in $\mathcal{C}_1(S)$). Let $\{(\beta_i, \beta'_i)\}_{i=1}^{L}$ be an enumeration of representatives of $\mathrm{Mod}(S)$-orbits. For each α_j there is some (in fact several) β_i equivalent to it under $\mathrm{Mod}(S)$, so let w_{ij} be a fixed group element such that $w_{ij}(\alpha_j) = \beta_i$. Define w'_{ij} similarly. Since this is a finite list, there is some upper bound B on their lengths as words in the generating set Γ.

Now let $\beta, \beta' \in \mathcal{C}_0(S)$ be any two curves of distance 1. Hence there exists $g \in \mathrm{Mod}(S)$ and $i \leq L$ such that $g(\beta_i) = \beta$ and $g(\beta'_i) = \beta'$. There also exist $j, k \leq N$ such that $w_{ij}(\alpha_j) = \beta_i$ and $w'_{ik}(\beta_k) = \beta'_i$.

Thus $gw_{ij} \in G_\beta$ and $gw'_{ik} \in G_{\beta'}$, and these two elements are separated by a path in $\mathcal{G}$ of distance at most $2B$. We conclude that $d_e(c_\beta, c_{\beta'}) \leq 2B + 1$ and hence the map φ is $(2B+1)$-Lipschitz.

To obtain a bound in the other direction, note that for any $g \in \mathrm{Mod}(S)$ we may associate the set $A_g = \{g(\alpha_j)\}_{j \leq N}$ in $\mathcal{C}_0(S)$, and that the diameter of this set in $\mathcal{C}(S)$ is equal to the diameter of $A_{id} = \{\alpha_j\}_{j \leq N}$, which is some fixed D (with appropriate choice of α_J we can easily get $D = 2$). Now given g and $g\gamma$ where $\gamma \in \Gamma$ is a generator, the distance between the sets A_g and $A_{g\gamma}$ is equal to that between A_{id} and A_γ, which is again bounded. Note finally that if $g \in G_\beta$ then $\beta \in A_g$. Thus we can map the vertices of $\hat{\mathcal{G}}$ back to $\mathcal{C}_0$, taking each c_β to β, and each g to some (any) element of A_g, and the resulting map is Lipschitz, and inverts φ. This proves that φ is a quasi-isometry. □

References

[1] L. Ahlfors, *Conformal invariants: topics in geometric function theory*, McGraw-Hill, 1973.

[2] Alonso, Brady, Cooper, Ferlini, Lustig, Mihalik, Shapiro, and Short, *Notes on word hyperbolic groups*, Group Theory from a Geometrical Viewpoint, ICTP Trieste 1990 (E. Ghys, A. Haefliger, and A. Verjovsky, eds.), World Scientific, 1991, pp. 3–63.

[3] W. Ballmann, *Lectures on spaces of nonpositive curvature*, Birkhäuser, 1995.
[4] L. Bers, *An extremal problem for quasiconformal mappings and a theorem by Thurston*, Acta Math. **141** (1978), 73–98.
[5] B. Bowditch, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, Group theory from a geometrical viewpoint (Trieste, 1990), World Scientific Publishing, 1991, pp. 64–167.
[6] M. R. Bridson, *Geodesics and curvature in metric simplicial complexes*, Group Theory from a Geometrical Viewpoint, ICTP Trieste 1990 (E. Ghys, A. Haefliger, and A. Verjovsky, eds.), World Scientific, 1991, pp. 373–463.
[7] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*, Birkhäuser, 1992.
[8] J. Cannon, *The theory of negatively curved spaces and groups*, Ergodic theory, symbolic dynamics, and hyperbolic spaces (Trieste, 1989), Oxford Univ. Press, 1991, pp. 315–369.
[9] A. J. Casson, *Automorphisms of surfaces after Nielsen and Thurston*, Notes by S. A. Bleiler, U. T. Austin, 1982.
[10] M. Coornaert, T. Delzant, and A. Papadopoulos, *Géométrie et theorie de groupes: les groups hyperboliques de Gromov*, Springer-Verlag, 1990.
[11] B. Farb, *Relatively hyperbolic and automatic groups with applications to negatively curved manifolds*, Ph.D. thesis, Princeton University, 1994.
[12] A. Fathi, F. Laudenbach, and V. Poenaru, *Travaux de Thurston sur les surfaces*, vol. 66-67, Asterisque, 1979.
[13] F. Gardiner, *Teichmüller theory and quadratic differentials*, Wiley Interscience, 1987.
[14] E. Ghys and P. de la Harpe, *Sur les groupes hyperboliques d'aprés Mikhael Gromov*, Birkhäuser, 1990.
[15] M. Gromov, *Hyperbolic groups*, Essays in Group Theory (S. M. Gersten, editor), MSRI Publications no. 8, Springer-Verlag, 1987.
[16] J. Harer, *Stability of the homology of the mapping class group of an orientable surface*, Ann. of Math. **121** (1985), 215–249.
[17] ———, *The virtual cohomological dimension of the mapping class group of an orientable surface*, Invent. Math. **84** (1986), 157–176.
[18] W. J. Harvey, *Boundary structure of the modular group*, Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference (I. Kra and B. Maskit, eds.), Ann. of Math. Stud. 97, Princeton, 1981.
[19] A. E. Hatcher, *Measured lamination spaces for surfaces, from the topological viewpoint*, Topology Appl. **30** (1988), 63–88.
[20] N. V. Ivanov, *Automorphisms of complexes of curves and of Teichmüller spaces*, Preprint.
[21] ———, *Complexes of curves and the Teichmüller modular group*, Uspekhi Mat. Nauk **42** (1987), 55–107.
[22] ———, *Complexes of curves and Teichmüller spaces*, Math. Notes **49** (1991), 479–484.
[23] L. Keen, *Collars on Riemann surfaces*, Discontinuous groups and Riemann surfaces (Proc. Conf., Univ. Maryland 1973), Ann. of Math. Studies 79, Princeton, 1974, pp. 263–268.
[24] S. Kerckhoff, *The asymptotic geometry of Teichmüller space*, Topology **19** (1980), 23–41.
[25] ———, *The Nielsen realization problem*, Ann. of Math. **117** (1983), 235–265.
[26] ———, *Simplicial systems for interval exchange maps and measured foliations*, Ergodic Theory and Dynamical Systems **5** (1985), 257–271.
[27] H. A. Masur, *On a class of geodesics in Teichmüller space*, Ann. of Math. **102** (1975), 205–221.
[28] ———, *The extension of the Weil-Petersson metric to the boundary of Teichmüller space*, Duke Math. J. **43** (1976), 623–635.
[29] ———, *Interval exchange transformations and measured foliations*, Ann. of Math. **115** (1982), 169–200.
[30] H. A. Masur and Y. Minsky, *Geometry of the complex of curves II*, in preparation.
[31] H. A. Masur and M. Wolf, *Teichmüller space is not Gromov hyperbolic*, MSRI preprint No. 011-94, 1994.
[32] Y. Minsky, *Harmonic maps, length and energy in Teichmüller space*, J. of Diff. Geom. **35** (1992), 151–217.
[33] ———, *Teichmüller geodesics and ends of hyperbolic 3-manifolds*, Topology **32** (1993), 625–647.
[34] ———, *Extremal length estimates and product regions in Teichmüller space*, Duke Math J. **83** (1996), 249–286.
[35] ———, *A geometric approach to the complex of curves*, Proceedings of the 37th Taniguchi Symposium on Topology and Teichmüller Spaces (S. Kojima et. al., ed.), World Scientific, 1996, pp. 149–158.
[36] ———, *Quasi-projections in Teichmüller space*, J. Reine Angew. Math. **473** (1996), 121–136.
[37] R. Penner, *Bounds on least dilatations*, Proc. Amer. Math. Soc. **113** (1991), 443–450.
[38] R. Penner and J. Harer, *Combinatorics of train tracks*, Annals of Math. Studies no. 125, Princeton University Press, 1992.
[39] K. Strebel, *Quadratic differentials*, Springer-Verlag, 1984.

[40] W. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. **19** (1988), 417–431.
[41] S. A. Wolpert, *Geodesic length functions and the Nielsen problem*, J. Differential Geom. **25** (1987), 275–296.
[42] ______, *The hyperbolic metric and the geometry of the universal curve*, J. Differential Geom. **31** (1990), 417–472.

University of Illinois at Chicago

SUNY Stony Brook