

HARMONIC MAPS, LENGTH, AND ENERGY IN TEICHMÜLLER SPACE

YAIR N. MINSKY

Abstract

The limiting behavior of high-energy harmonic maps between closed hyperbolic surfaces is analyzed. In general a measured foliation on the domain is shown to be mapped very nearly (exponentially in the energy) to its geodesic representative in the range. This foliation is in fact the horizontal foliation Φ_h of the Hopf differential Φ of the harmonic map. Φ_h is also characterized as nearly maximizing (up to an additive constant) the ratio of squared hyperbolic length in the range to extremal length in the domain, among all simple closed curves in the domain. The same ratio gives the energy of the map up to an additive constant. This can be viewed as an analogy to other canonical maps between surfaces, for which different optimization problems are characterized by corresponding length-ratio maximizations.

In addition, the asymptotics of a family of harmonic maps obtained when the domain surface is varied along a classical Teichmüller ray are studied. As expected, the limiting Hopf foliation and the foliation determining the ray are equivalent as topological (not measured) foliations.

1. Introduction

The Teichmüller space of a surface can be viewed as the space of complex (or conformal) structures on the surface, or alternatively as the space of hyperbolic structures on it. A key feature in the study of Teichmüller spaces is the construction of comparison maps between surfaces which are "optimal" in some sense appropriate to the point of view. This paper studies *harmonic maps*, which are obtained by minimizing *energy* over a homotopy class of maps between surfaces. The energy of a map between Riemannian manifolds is the integral over the domain of the squared derivative of the map (see §3 for a precise definition). In particular, we analyze the approximate behavior of such maps (in the case of closed surfaces) when their energy is high or, equivalently, when the domain and range are far apart as points in Teichmüller space. We obtain, in Theorems 7.1 and 7.2, analogies between the behavior of harmonic maps and

that of other “optimal” maps, namely Teichmüller maps and Thurston’s stretch maps.

Summary of results. Central to the discussion are the notions of *measured foliations* (and equivalently, measured laminations), which are generalizations of simple closed curves on a surface, and which we can think of, for now, as “infinite simple curves” (we are avoiding here any explanation of the meaning of “measure”—see §2). Let γ denote such an object, and let $M = (S, \sigma)$ be a closed surface S with a hyperbolic metric σ . Then the geodesic representative γ^* can be defined as can its length, denoted $l_\sigma(\gamma)$. The *extremal length* $E_\sigma(\gamma)$ can also be defined in a way that generalizes the extremal length of a simple closed curve (this being the reciprocal of the maximal conformal modulus of an embedded annulus in S whose core is homotopic to γ). We note that $E_\sigma(\gamma)$ is the same for any metric in the conformal class of σ , and thus is properly a *conformal* rather than a hyperbolic invariant of γ .

Let $N = (S, \rho)$ denote a second hyperbolic structure on the surface. The homotopy class of the identity, viewed as a map from M to N , contains a unique harmonic representative f whose energy $\mathcal{E}(f)$ is minimal in this class (§3). We shall prove

Theorem 7.2. *There is a constant C depending only on $\chi(S)$, and a measured foliation φ on S such that*

$$\frac{1}{2} \frac{l_\rho^2(\varphi)}{E_\sigma(\varphi)} \leq \mathcal{E}(f) \leq \frac{1}{2} \frac{l_\rho^2(\varphi)}{E_\sigma(\varphi)} + C$$

Actually, the inequality $\frac{1}{2} l_\rho^2(\gamma) E_\sigma(\gamma) \leq \mathcal{E}(f)$ is true for any measured foliation γ (Proposition 3.1), so it follows that φ gives, up to the constant C , the maximal value for the ratio l_ρ^2/E_σ over all measured foliations on S . (Clearly this result is meaningful only for large values of $\mathcal{E}(f)$.)

This situation, in which the solution to a length-ratio maximization problem for foliations gives an estimate for the solution of an optimization problem for maps, echoes what happens in the other two examples of optimal maps, as we now describe.

A Teichmüller map is obtained by considering the conformal rather than the metric properties of M and N . The extent to which a map distorts conformal structure is called its *quasi-conformal dilatation*, and Teichmüller showed (see [10]) that there is a unique map, minimizing this dilatation in the homotopy class of (say) the identity. Furthermore, there is a measured foliation ψ in S for which the ratio $K = E_\rho(\psi)/E_\sigma(\psi)$ is maximal among all measured foliations, and this K is precisely the dilatation of the Teichmüller map (see [17]).

Thurston's stretch maps, on the other hand, are obtained by considering the hyperbolic metrics of both domain and range, and attempting to minimize the *Lipschitz constant* of maps in the homotopy class of the identity. A minimizing map is found (not quite uniquely, however) and again there is a measured foliation λ such that the ratio $l_\rho(\lambda)/l_\sigma(\lambda)$ is maximal, and is equal to the minimal Lipschitz constant.

Harmonic maps, it turns out, compare the hyperbolic structure of N to the *conformal* structure of M since the energy of a map with a two-dimensional domain depends not on the domain metric itself but on its conformal class. Thus the appropriate comparison of lengths should involve the hyperbolic invariant l_ρ and the conformal invariant E_σ —as borne out in Theorem 7.2 (why l_ρ^2 appears, instead of l_ρ , will be made clear in §2.)

Theorem 7.1, from which Theorem 7.2 follows, gives a geometric characterization of a harmonic map in terms of this "maximal stretch foliation," in further analogy with the other two types of maps. To discuss this result we need briefly to describe the two canonical concrete realizations of a measured foliation on a Riemann surface.

The first, relevant to the conformal structure, is as the horizontal foliation of a holomorphic quadratic differential. In brief, this means there is a Euclidean metric with singularities which is conformally equivalent to the hyperbolic metric, and the leaves of our foliation are Euclidean straight lines which meet in prongs of three or more at the singularities.

In the hyperbolic point of view, we do not obtain a foliation of the entire surface. Instead, we straighten each leaf to its hyperbolic geodesic representative (if there are closed leaves then a whole family of homotopic leaves will collapse to one geodesic), obtaining a closed set of geodesics whose complement consists of surfaces bounded by closed or infinite geodesics (the simplest example is an ideal triangle in the hyperbolic plane). This is called a *geodesic lamination*. Each complementary region of the lamination corresponds to a configuration of singularities in the quadratic differential picture above.

A Teichmüller map is obtained, geometrically, as follows. The measured foliation ψ maximizing E_ρ/E_σ is realized via a holomorphic quadratic differential ψ_σ in the domain, and via ψ_ρ in the range. The Teichmüller map takes the leaves of ψ_σ to those of ψ_ρ , in the process expanding the Euclidean metric by exactly K along the leaves.

Similarly, let λ maximize the ratio l_ρ/l_σ . A Thurston stretch map is obtained by mapping the leaves of the geodesic representative of λ in

M to those of its representative in N , and filling in the complementary regions (in a nonunique way).

The analogous result for harmonic maps is that the leaves of the quadratic differential in M are mapped, *approximately*, to the leaves of the geodesic representative in N . The full statement of the theorem is somewhat technical, and we summarize it here. The distance $d(\cdot, \cdot)$ is the Euclidean metric of the quadratic differential, and \mathcal{P}_{R_0} refers to a region of the surface M containing the singularities of the metric and corresponding roughly to the complementary regions of the geodesic representative of the foliation.

Theorem 7.1. *The harmonic map $f: M \rightarrow N$ takes the leaves of φ ϵ -close in a C^1 sense to the geodesic representative of their images in N , where ϵ at each point $p \in M$ is given by $A \exp(-Bd(p, \mathcal{P}_{R_0}))$, A and B being a-priori positive constants depending only on the topological type of S .*

We note that the diameter of M in the singular Euclidean metric is at least proportional to $\mathcal{E}(f)^{1/2}$, whereas the diameter of \mathcal{P}_{R_0} is bounded, so that "most" of the leaves of φ are in fact mapped quite close to their geodesic representatives.

In §8, we consider the family of harmonic maps obtained when the target N is fixed and the structure of the domain M is allowed to vary (as a point in Teichmüller space). Fixing a surface M_0 , Teichmüller space can be parametrized by *Teichmüller rays* based at M_0 ; these are one-parameter families of deformations of M_0 , each ray obtained by squeezing the conformal structure (via a Teichmüller map) along a fixed foliation. The foliation φ associated to the harmonic map changes as we progress along such a ray, and its limiting values are described by

Theorem 8.1. *A limit point of the foliations φ along a Teichmüller ray is equivalent (as an unmeasured foliation) to the foliation determining that ray.*

We note that there is a subtlety involving the measures which will be briefly discussed in §8. Not having defined the notion of measure for a foliation, we avoid this issue for now.

In §4 we develop, as tools for the proof of the main theorems, some estimates of conformal moduli that may be of independent interest. In particular, we convert geometric information about annuli in a singular Euclidean metric (arising from a holomorphic quadratic differential) to information about moduli. We define the notion of a *regular annulus* (which generalizes standard examples like round annuli in the plane) and

compute a simple geometric invariant μ which estimates the modulus via:

Theorem 4.5 and 4.6. *Fix a holomorphic quadratic differential metric on M . For a regular annulus $A \subset M$,*

$$\mu(A) \leq \text{Mod}(A) \leq \mu(A) + \min\left(c_1\mu(A), c_2\sqrt{\mu(A)}\right).$$

If A is any homotopically nontrivial annulus in M and $\text{Mod}(A) \leq m_0$, then A contains a regular annulus B such that

$$\mu(B) \leq c_3 \text{Mod}(A) - c_4,$$

where the constants m_0 and $c_1 \dots, c_4$ depend only on $\chi(M)$.

The author is very grateful to Mike Wolf for his considerable help and encouragement through various stages of this project. It is worth noting here that in [31]–[33] Wolf has already analyzed in detail some special cases of the phenomena we describe in §§7 and 8—in particular, he has considered the case where M is fixed and N degenerates, and also a particular type of degeneration of M , when a finite number of simple closed curves is pinched.

This paper has developed from a portion of the author's Ph.D. thesis [20], written at Princeton University. He is grateful to his advisor, Bill Thurston, for help and inspiration, and to Dick Canary, for years of talking and listening. The referee deserves thanks as well, for his relentless comments and suggestions, and for his willingness to read the paper carefully.

A generalization of these results to the case where the image is a hyperbolic 3-manifold will appear in [21]. This requires a revision of the analytical estimates in §3, as well as a somewhat careful use of the properties of pleated surfaces.

Method of proof. We conclude the introduction with a brief discussion of the techniques used in proving Theorem 7.1. Let us fix the harmonic diffeomorphism $f : M \rightarrow N$. The first piece of information, namely the measured foliation appearing in Theorems 7.1 and 7.2, is obtained as follows. To a harmonic map of a surface is associated a holomorphic quadratic differential on the domain, known as the *Hopf differential* (see §3.2 or [16], [25], [31]), and written as Φ . It defines a Euclidean metric with singularities, as hinted above (and described in detail in §2), which we write as $|\Phi|$, and a foliation by straight lines which we call Φ_h . This foliation is characterized by the condition that, at each point on the surface, the direction in which df has the greatest expansion is tangent to a leaf of Φ_h . The metric $|\Phi|$ has the additional characterization that its

area measure is, away from the singularities, an approximation to energy density for the map (inequality (3.5)).

The Jacobian of f obeys an elliptic second-order partial differential equation (3.2) from which follows an estimate (§3.3) that implies that the images of Φ_h are tightly squeezed together and are nearly geodesic in the hyperbolic metric ρ , provided the leaves are sufficiently far from the singularities, of Φ , as measured in the $|\Phi|$ metric. These estimates fail at the singularities and in regions of M where the injectivity radius of $|\Phi|$ is small, but improve exponentially with $|\Phi|$ -distance from these "bad" regions. Accordingly, in §5, we show how to construct a family of regions $\mathcal{P}_R \subset M$ ($R > 0$) such that \mathcal{P}_R includes an R -neighborhood of the bad regions, as measured in $|\Phi|$, while maintaining control over the size (area, boundary length, etc.) and shape of \mathcal{P}_R .

§6 gives the main body of the proof, in which this local information about the curvature of images of the foliation leaves outside \mathcal{P}_R (for sufficiently large R) is converted to the global fact that these are mapped C^1 -close to their geodesic representatives. This is not a trivial assertion, because the lack of control on curvature inside \mathcal{P}_R means, a priori, that the image leaves could be anywhere on the surface.

The main idea is to control these leaves by using Thurston's *train-tracks*. A train-track may be visualized as a slight thickening of a 1-complex whose edges meet tangentially at its vertices. The thickened edges are foliated by short "ties" transverse to the edges, and we define a *train route* as a path that is transverse to the ties. If the edges of the 1-complex have sufficiently low curvature (in a hyperbolic surface), then any closed train route is (correspondingly) close to its geodesic representative.

The regions \mathcal{P}_R are shaped in such a way that their complements have a natural train-track structure with respect to which the leaves of Φ_h describe train routes. For large enough R , the image of this in the ρ -metric is a train-track with nearly straight edges, and it is with this train-track that we obtain control over the geodesic representative of Φ_h . Controlling the leaves in the complement of this train track is the main concern of §6, in which the fact that the complement is a (nearly) convex subsurface in the hyperbolic metric ρ , and the existence of bounds on the lengths of the runaway leaves, are used to "straighten" them without disturbing the well-behaved leaves.

2. Structures on surfaces

Hyperbolic constructions. References for the material in this section can be found in [4]–[6], [9], and [28].

Let $M = (S, \sigma)$ denote a closed surface S of genus $g > 1$ with a hyperbolic metric σ . A *geodesic lamination* on M is a closed subset of M which is a disjoint union of complete geodesic leaves. Such a set has zero area (in fact Hausdorff dimension 1). Denote by $GL(M)$ the set of geodesic laminations on M .

Let $\mathcal{ML}(M)$ be the set of geodesic laminations equipped with *transverse measures*. A transverse measure assigns Borel measures to arcs transverse to the lamination, which are invariant under translation along the lamination. The mass of the measure assigned to α by $\lambda \in \mathcal{ML}(M)$ is denoted $i(\alpha, \lambda)$, and arcs disjoint from the lamination have zero measure. $\mathcal{ML}(M)$ has a natural topology which should be thought of as the geometric topology, weighted by the measures. To be more precise, a neighborhood of a lamination λ in this topology is determined by a finite number of (say) smooth arcs transverse to λ and a number $\epsilon > 0$, and contains laminations that are transverse to these arcs and deposit a measure on them differing from that of λ by at most ϵ . $\mathcal{PML}(M)$ is the projectivization of $\mathcal{ML}(M)$, identifying measures that are multiples of each other. $\mathcal{PML}(M)$ is a sphere of dimension $6g - 7$ (see [28], [9]), and in particular is compact.

Any simple closed geodesic γ admits the *counting measure*, which assigns to a transverse arc α the mass $i(\alpha, \gamma) = \#(\alpha \cap \gamma)$. The simple closed geodesics endowed with positive multiples of the counting measures are dense in $\mathcal{ML}(M)$. The notion of hyperbolic length $l_M(\gamma)$ for a simple closed geodesic generalizes to a continuous function on $\mathcal{ML}(M)$ which scales homogeneously:

$$l_M(c\lambda) = cl_M(\lambda).$$

It is computed as the mass of the measure defined locally as the product of the transverse measure with the regular length measure along the geodesic leaves.

The geometric intersection number of simple closed geodesics similarly extends to a continuous nonnegative function

$$i: \mathcal{ML}(M) \times \mathcal{ML}(M) \rightarrow \mathbf{R},$$

which is homogeneous in both arguments. It is given by the mass of the measure on the surface defined locally as the product of the transverse measures of the two laminations. We note also that, as for closed curves, $i(\alpha, \beta)$ can be considered to be defined on a homotopy class of laminations, as it gives the minimal intersection number over all homotopic representatives of α and β (see [18] for a discussion).

Conformal constructions. If we consider only the conformal structure induced by σ , then M becomes a Riemann surface, and a different set of constructions is natural. A holomorphic quadratic differential Φ is a differential expressed as $\Phi(z)dz^2$ in a local holomorphic coordinate system, such that $\Phi(z)$ is holomorphic in z (see [27]). The space of such differentials on M is denoted $QD(M)$, and is a $6g-6$ (real)dimensional vector space. Removing 0 and taking a quotient by the positive real numbers, we obtain a sphere, $PQD(M)$.

Any $\Phi \in QD(M)$ determines a pair of *measured singular foliations* on M : away from the (isolated) zeros of Φ choose a branch of the square root and define a local holomorphic coordinate $\zeta(z)$ by integrating $d\zeta = \sqrt{\Phi(z)}dz$. This defines ζ up to sign and translations, and in the ζ coordinate system $\Phi = d\zeta^2$. The lines of constant $\text{Im}(\zeta)$ form a well-defined foliation on M , called Φ_h or the *horizontal foliation*, and similarly the lines of constant $\text{Re}(\zeta)$ form the vertical foliation Φ_v . At the zeros of Φ , these foliations have $(k+2)$ -prong “saddle” singularities, where k is the degree of the zero. To visualize this, draw the integral curves of the line field $v = \pm z^{-k/2}$ around $z = 0$ (or see Figure 5 in §5).

The foliation Φ_h inherits a natural transverse measure from Φ —the measure of a (short) transverse arc is its vertical height in the ζ plane (and vice versa for Φ_v).

Φ also determines a (singular) metric written as $|\Phi|$ or $|\Phi(z)||dz|^2$, which is just the Euclidean metric in the ζ plane. At a zero the $|\Phi|$ metric has concentrated negative curvature which is an integral multiple of π . We will explore the global properties of such metrics in more detail in §4.

A *Teichmüller map* is a homeomorphism between two Riemann surfaces that achieves the lowest possible quasiconformal distortion in its homotopy class. Such a map is always given by a quadratic differential Ψ on the domain surface M and a number $K \geq 1$; the conformal structure of the range is obtained by dilations in the horizontal and vertical directions of Ψ whose ratio is K . For example we can define a metric on the range by contracting the $|\Psi|$ metric on the domain by K along the leaves of Ψ_h and leaving it unchanged in the Ψ_v direction. The conformal class of this metric gives a Riemann surface $M_{\Psi, K}$, and the quadratic differential Ψ^K defined by the foliations of Ψ and that metric has a horizontal foliation in the same measure class as Ψ_h . Note that this is the reverse of the usual definition, where the vertical foliation is contracted with respect to the horizontal. This will seem more reasonable in §8, and at any rate the difference amounts to replacing Ψ by $-\Psi$.

We will call $\{M_{\Psi, K} : K \geq 1\}$ the *Teichmüller ray* along Ψ based at M .

Topological constructions. $QD(M)$ and $\mathcal{ML}(M)$ are related to a purely topological construction, and hence to each other. Let $\mathcal{MF}(S)$ denote the space of measured foliations on S with only saddle singularities, modulo isotopy and Whitehead moves (which collapse singularities that are connected by a leaf of the foliation).

The map

$$\mathcal{H} : QD(M) \rightarrow \mathcal{MF}(S)$$

sending Φ to $[\Phi_h]$ is a homeomorphism (see [17], [15]). Similarly, there is a map

$$\mathcal{L} : \mathcal{MF}(S) \rightarrow \mathcal{ML}(M),$$

which produces a lamination from a foliation by “straightening” all the leaves to geodesics, in the process creating spaces between the leaves (see [19] for a detailed discussion). This map, too, is a homeomorphism. All these maps descend to homeomorphisms between the associated projectivized spaces (see also [14]).

A lamination is *minimal* if it contains no proper sublamination. Every measured lamination decomposes into a union of finitely many minimal components (see [28], [6]). Note that this is false for a general geodesic lamination λ , which may contain a proper sublamination whose complement in λ is not closed. We briefly discuss this decomposition, and the way it appears from the point of view of foliations. This will become relevant in §8.

Say that a closed subsurface $R \subset M$ is a *supporting subsurface* for a lamination $\lambda \subset R$ if R is *incompressible* (i.e., $\pi_1(R)$ injects into $\pi_1(M)$), and if any essential closed curve $\alpha \subset R$ which is disjoint from λ may be isotoped through $R - \lambda$ into ∂R . It is clear that every lamination has a supporting subsurface: begin with M , and cut away neighborhoods of essential nonperipheral curves disjoint from the lamination a finite number of times (by an Euler characteristic argument.) Such a subsurface is also unique up to isotopy: for any supporting subsurface R of λ , find a neighborhood of λ which is contained in R , and adjoin to it the disks bounded by any compressible curves in its boundary. The resulting surface N has boundary curves which (by the properties of R) are isotopic outside N into ∂R , and is therefore isotopic to R . Since any two supporting subsurfaces R, R' must have such a neighborhood in common, they are isotopic to each other.

It is clear, also, that a measured lamination is minimal if and only if its supporting subsurface is connected. We can now give the correspondence between minimal sublaminations and subfoliations. Let \mathcal{F} denote a measured foliation, and $\lambda = \mathcal{L}(\mathcal{F})$ the corresponding measured lamination. Let Σ denote the union of leaves of \mathcal{F} that meet singularities. The compact leaves in Σ (those which meet singularities at two ends) form a compact 1-complex, possibly empty. Let Σ_c denote the union of noncontractible components of this 1-complex. We have:

Lemma 2.1 (Minimal components of a foliation). *The components of the complement of an open regular neighborhood of Σ_c are supporting subsurfaces for the minimal components of λ .*

Proof. In Levitt's construction [19], the leaves of λ are obtained as follows: each nonsingular leaf of \mathcal{F} lifts to the universal cover \widetilde{M} of M to a union of leaves that have well-defined geodesic representatives (geodesics with the same endpoints at infinity). A component of the lift of Σ is a tree, a regular neighborhood of which has boundaries with well-defined geodesic representatives. All these leaves project to λ , and the regular neighborhoods of the trees have projections which are isotopic to the complement of λ (one must take some care here because Σ may be dense in M).

Now let $\gamma \subset M$ be an essential simple closed curve in the complement of λ . Then a lift of γ to \widetilde{M} may be isotoped to one of the regular neighborhoods of a component of Σ . This component must have a noncontractible projection in M , which as an abstract surface can be retracted into a component of Σ_c .

In other words γ may be isotoped into a regular neighborhood of Σ_c , and it follows immediately that the complement of this neighborhood is a supporting subsurface for λ , whose components are supporting subsurfaces for the minimal components of λ .

Modulus and extremal length. For a family of curves (closed curves or segments) Γ in a Riemann surface M extremal length is defined as

$$(2.1) \quad E_M(\Gamma) = \sup_{\rho} \frac{l_{\rho}^2(\Gamma)}{\text{Area}(\rho)},$$

where the supremum is taken over all choices of metric ρ consistent with the conformal structure of M , and $l_{\rho}(\Gamma)$ is the infimum of ρ -length over curves in Γ (see [2]). Generally Γ is a free homotopy class or a homotopy class with a constraint on endpoints.

In particular, if M is an annulus and Γ is the class of paths connecting the boundaries of M , then $E_M(\Gamma)$ is the *modulus* $\text{Mod}(M)$, which is defined uniquely as the ratio of the height to the circumference of any Euclidean annulus conformally equivalent to M . Further, if Γ is the free homotopy class of a circumferential curve (core) of M , then $E_M(\Gamma) = 1/\text{Mod}(M)$.

For a general M , for any simple closed curve $\gamma \subset M$ we take $E_M(\gamma)$ to be the extremal length of the free homotopy class of γ in M . The fact analogous to the previous paragraph is that

$$(2.2) \quad E_M(\gamma) = \inf_A \frac{1}{\text{Mod}(A)},$$

where the infimum is taken over all annuli $A \subset M$ whose core is homotopic to γ . In [17] this is called the *geometric* definition of extremal length, and (2.1) is called the *analytic* definition.

The infimum and supremum problems have the same solution. The quadratic differential Ψ such that $\Psi_h \equiv \gamma$ (as measured laminations) defines a cylinder—the union of leaves of Ψ_h which do not meet the singularities—which maximizes modulus among cylinders homotopic to γ . On the other hand the $|\Psi|$ metric maximizes the ratio of (2.1).

Extremal length can be extended to arbitrary measured foliations. First, from definition (2.1) it is fairly clear that the right way for extremal length to behave under scaling of measures is

$$E(c\gamma) = c^2 E(\gamma).$$

In [17], Kerckhoff showed that with this property extremal length extends to all of $\mathcal{MF}(M)$ continuously, and the extremal metric is always given by the appropriate quadratic differential. It is also easy to see that $E_M(\Psi_h) = \text{Area}(|\Psi|)$. We note that this is also true for noncompact surfaces of finite topological type.

Since extremal length scales quadratically with measure and hyperbolic length scales linearly, the ratio

$$\frac{l_\rho^2(\gamma)}{E_\sigma(\gamma)}$$

is the natural comparison between hyperbolic and conformal invariants of γ on surfaces (S, ρ) and (S, σ) . This ratio is invariant under scaling and so defines a continuous function on the compact set $\mathcal{PMF}(S)$ which must therefore realize a maximum. (The factor of $\frac{1}{2}$ that appears later is due to the $\frac{1}{2}$ in the traditional definition of energy.)

3. Harmonic maps

3.1 Basic properties. A harmonic map $f : M \rightarrow N$ between Riemannian manifolds is a stationary point of the energy functional on C^1 maps

$$\mathcal{E}(f) = \int_M e \, dV_M,$$

where e is the pointwise energy, $\frac{1}{2}|df|^2$, and dV_M is the volume form on M . One may think of e as a multiple of the average squared stretch of the map at a point: if $\{\varepsilon_\alpha\}$ are the elements of an orthonormal frame in M , then

$$e = \frac{1}{2} \sum_\alpha \|df(\varepsilon_\alpha)\|_N^2.$$

In direct analogy with harmonic functions, a map is harmonic if and only if it satisfies an Euler-Lagrange equation $\tau f = 0$, where τ is a second-order elliptic partial differential operator (see [7] for a more complete discussion).

The basic existence result was obtained by Eells and Sampson [8], who proved the existence of a harmonic map in any homotopy class of maps in the case where N is compact and has nonpositive sectional curvatures. Hartman [13] showed that the map is unique if N is negatively curved, and if the image is not contractible to a point or a geodesic.

In such cases, it also follows that the unique harmonic map varies smoothly under smooth deformation of either the domain or range metric (see Sampson [23]).

3.2. Maps of surfaces. We restrict now to the case where M is two-dimensional. The situation is greatly simplified here by the immediate observation that energy, and therefore harmonicity, are unaffected by conformal changes in the domain metric. This is because multiplying the metric of M by a positive function divides e and multiplies dV by the square of the function. The equation $\tau f = 0$ is also unaffected by conformal changes.

The only relevant properties of M in the context of harmonic maps are therefore its properties as a Riemann surface, or its conformal properties. Our first example of this is a simple inequality relating energy, length of curves in the image, and extremal length in the domain. Suppose R is a conformal rectangle in M with modulus m . If we realize R as the Euclidean rectangle $\{0 \leq x \leq L\} \times \{0 \leq y \leq H\}$, so that $m = H/L$, then

for any map $f: M \rightarrow N$ we can bound the average of the lengths in N of the images of the horizontal arcs $\gamma_y = \{(x, y): 0 \leq x \leq L\}$:

$$\begin{aligned}
 \text{Avg}\{l_N(f(\gamma_y))\} &= \frac{1}{H} \int_0^H \int_0^L \left| \frac{\partial f}{\partial x} \right|_N dx dy \\
 (3.1) \qquad \qquad \qquad &\leq \frac{1}{H} \left[\text{Area}(R) \cdot \int_R \left| \frac{\partial f}{\partial x} \right|_N^2 \right]^{1/2} \leq \left[\frac{L}{H} \cdot 2\mathcal{E}(f) \right]^{1/2}.
 \end{aligned}$$

This argument works as well for a conformal cylinder whose horizontal arcs are all homotopic to some curve γ in M . Since the extremal length of γ is the infimum of $1/m$ over all such cylinders, we obtain

Proposition 3.1 (Energy lower bound). *For any simple closed curve $\gamma \subset M$ and any map $f: M \rightarrow N$, the energy of f is bounded below by*

$$\mathcal{E}(f) \geq \frac{1}{2} \frac{l_N^2(f(\gamma))}{E_M(\gamma)},$$

where E_M is extremal length in M , and $l_N(f(\gamma))$ is the infimum of lengths in N of representatives of $f(\gamma)$.

By continuity, this inequality extends to arbitrary measured laminations (or foliations).

Hopf differentials. Let $f: M^2 \rightarrow N$ be a harmonic map from a Riemann surface. The Hopf differential of f is the quadratic differential

$$\Phi = (f^*\rho)^{2,0},$$

where ρ is the metric of N (in other words, it is the dz^2 part of the pullback of the metric tensor ρ).

One can show (see [16], [24], [23]) that the harmonicity of f implies that Φ is holomorphic. Thus, Φ defines a singular Euclidean metric $|\Phi|$ on M and measured foliations Φ_h and Φ_v as in §2.

The leaves of Φ_h are tangent to the maximal stretch directions of the map. In fact, choose a local holomorphic coordinate $z = x + iy$ such that the leaves of Φ_h are parallel to the x -axis (and Φ_v to the y -axis). Let $\sigma(z)|dz|^2$ be a metric on M , and e the pointwise energy with respect to this metric. Then

$$f^*(\rho) = \Phi dz^2 + \sigma e dz d\bar{z} + \bar{\Phi} d\bar{z}^2$$

or, since in these coordinates $\Phi(z) = |\Phi(z)|$,

$$f^*(\rho) = (\sigma e + 2|\Phi|)dx^2 + (\sigma e - 2|\Phi|)dy^2.$$

Two-dimensional ranges. Restrict further, now, to the case where N is a surface. There is a collection of useful notation, introduced in [23] and

used in [31]-[33]. Let w be a local complex coordinate on N , and let $\rho(w)|dw|^2$ be the local form of the metric of N .

Let \mathcal{J} be the Jacobian of f relative to the σ metric, and let ν be the Beltrami differential of f , defined by

$$\nu = \frac{w_z d\bar{z}}{w_{\bar{z}} dz}.$$

The Beltrami differential measures the failure of a map to be conformal, and in particular its norm $|\nu|$, which is a well-defined function on M (unless $df = 0$, which can only happen at the zeros of Φ), is 0 if f is conformal, 1 if df collapses one tangent direction, and ∞ if f is anticonformal.

It will be convenient to define $\mathcal{E} = \log(1/|\nu|)$. The heart of the analytical preliminaries to our proof is the fact that high energy maps tend to have \mathcal{E} very near zero on most of M , which implies that f nearly collapses the vertical direction of Φ .

A simple computation shows

$$\cosh \mathcal{E} = \frac{\sigma e}{2|\Phi|}, \quad \sinh \mathcal{E} = \frac{\sigma \mathcal{J}}{2|\Phi|}$$

away from the zeros of Φ . The behavior of \mathcal{E} is controlled by the following equation (see [23] or [32]):

$$(3.2) \quad \Delta \mathcal{E} = -2K(\rho)\mathcal{J} = -4K(\rho)\frac{|\Phi|}{\sigma} \sinh \mathcal{E},$$

where $K(\rho)$ is the Gaussian curvature of the $\rho|dw|^2$ metric, and Δ is the Laplacian in the $\sigma|dz|^2$ metric. This equation holds whenever f is nonsingular, and, as we shall see shortly, it forces \mathcal{E} to decay to 0 "exponentially fast," relative to the $|\Phi||dz|^2$ metric.

The signs of \mathcal{J} and \mathcal{E} depend on the choice of orientation for the target. This being fixed they are smooth functions (with singularities for \mathcal{E} at the zeros of Φ), and in the case of a degree 1 map are everywhere positive ([23], [25]).

Let us also introduce the following notation:

- $dA(\sigma) = \sigma dx \wedge dy$: the area form for $\sigma|dz|^2$,
- $\mathcal{E}(U) = \int_U e dA(\sigma)$: the energy of f associated to a subsurface $U \subset M$,
- $\|\Phi\|_U = \int_U dA(|\Phi|)$: the $|\Phi|$ -mass of U , or its area in the $|\Phi|$ metric. We also write $\|\Phi\| = \|\Phi\|_M$.

In coordinates where $\Phi \equiv 1$ we have

$$(3.3) \quad f_\rho^* = 2(\cosh \mathcal{E} + 1)dx^2 + 2(\cosh \mathcal{E} - 1)dy^2.$$

Thus when \mathcal{E} is very small, the map tends to squeeze together the horizontal leaves of Φ , and its derivative along the leaves is approximately 2.

With a little algebra one can see that

$$2|\Phi| \leq \sigma e \leq 2|\Phi| + \sigma \mathcal{E},$$

so that, integrating over any subsurface $U \subset M$ (see also [31]),

$$(3.4) \quad 2\|\Phi\|_U \leq \mathcal{E}(U) \leq 2\|\Phi\|_U + \text{Area}(f(M)).$$

3.3. Estimates on \mathcal{E} . Assume now that N is a closed hyperbolic surface, and that $f: M \rightarrow N$ is a harmonic map homotopic to a diffeomorphism. From [23], [25] we know that f is itself a diffeomorphism, and in particular $\mathcal{E} > 0$. Since, by the Gauss-Bonnet theorem, the area of a hyperbolic surface is given by

$$\text{Area}(f(M)) = -2\pi\chi(M),$$

inequality (3.4) becomes

$$(3.5) \quad 2\|\Phi\|_U \leq \mathcal{E}(U) \leq 2\|\Phi\|_U + 2\pi|\chi(M)|.$$

The formula for $\text{Area}(f(M))$ gives us an easy bound on \mathcal{E} :

Lemma 3.2 (Rough bound). *Let $p \in M$ be a point with a neighborhood U such that U contains no zeros of Φ , and in the $|\Phi|$ -metric is a round disk of radius r centered on p . Then there is a bound*

$$\mathcal{E}(p) \leq \sinh^{-1}(|\chi(M)|/r^2).$$

Proof. The simplest consequence of (3.2) is that $\Delta\mathcal{E} \geq 0$, and thus \mathcal{E} is subharmonic in U . It is sufficient, therefore, to bound the average of \mathcal{E} on U (in the $|\Phi|$ metric).

Recall from above that

$$\sinh \mathcal{E} = \frac{1}{2} \frac{\sigma}{|\Phi|} \mathcal{E}.$$

Using the concavity of \sinh^{-1} on the positive real axis and the formula for $\text{Area}(f(M))$, we obtain

$$\begin{aligned} |\Phi| \cdot \text{Avg}_U(\mathcal{E}) &= |\Phi| \cdot \text{Avg}_U \left(\sinh^{-1} \frac{1}{2} \frac{\sigma}{|\Phi|} \mathcal{E} \right) \\ &\leq \sinh^{-1} \left(|\Phi| \cdot \text{Avg}_U \left(\frac{1}{2} \frac{\sigma}{|\Phi|} \mathcal{E} \right) \right) \\ &= \sinh^{-1} \left(\frac{1}{2\pi r^2} \int_U \frac{\sigma}{|\Phi|} \mathcal{E} \, dA(|\Phi|) \right) \\ &\leq \sinh^{-1}(|\chi(M)|/r^2). \quad \text{q.e.d.} \end{aligned}$$

This lemma can be used to obtain a preliminary bound in parts of the surface where the injectivity radius in the $|\Phi|$ metric is large enough. In § 5 it will become clearer what to do about other parts. Meanwhile, we utilize the full power of (3.2) and state an estimate from [32].

Lemma 3.3 (Exponential estimate). *Let $p \in M$ be at a $|\Phi|$ -distance $> d > 0$ from any zeros of Φ , and let \mathcal{G} be bounded above by B in a neighborhood of $|\Phi|$ -radius d around p . Then*

$$\mathcal{G}(p) < B / \cosh d.$$

Proof. If we choose $\sigma = |\Phi|$ in (3.2), we obtain

$$\Delta \mathcal{G} = 4 \sinh \mathcal{G},$$

where Δ is now the Laplacian for the $|\Phi|$ metric, and we recall that the sectional curvature of the image surface is -1 . This equation holds in a Euclidean disk of radius d about p , which is the lift to the universal cover of a d -neighborhood of p in the $|\Phi|$ metric (*Note:* this disk does not have to be embedded in S ; it can wrap around many times, since we are not going to use the area bound on the image. The conditions of the lemma ensure that the disk contains no zeros of Φ , so it is Euclidean).

If (x, y) are Euclidean coordinates based at p , we define a comparison function F on the disk by

$$F(x, y) = \frac{B}{\cosh d} \cosh \sqrt{2}x \cosh \sqrt{2}y.$$

Then $F \geq B$ on the boundary, $\Delta F = 4F$ everywhere, and $F(p) = B / \cosh d$. A maximum principle argument shows that $F \geq \mathcal{G}$ everywhere:

$$\Delta(F - \mathcal{G}) = 4F - 4 \sinh \mathcal{G} \leq 4(F - \mathcal{G}),$$

and $\Delta(F - \mathcal{G}) \geq 0$ when $F - \mathcal{G}$ achieves its minimum. Thus $F - \mathcal{G} \geq 0$.

Notation. We will frequently use $\varepsilon(r)$ to denote any function of the form ae^{-br} , where a and b are positive constants depending, unless otherwise stated, on nothing but the topological type of the surface S . Thus, the conclusion of the above lemma could have been written $\mathcal{G}(p) \leq B\varepsilon(d)$.

Curvature estimates. The estimates on \mathcal{G} have two important geometric consequences. First, as follows immediately from (3.3), vertical arcs of Φ_h are very short for small \mathcal{G} :

$$\|df(v)\| = O(\mathcal{G})$$

for a unit vector v tangent to Φ_v . Further, horizontal arcs are nearly straight: if $k_h(p)$ denotes the geodesic curvature in N of the leaf of Φ_h

through p and $\mathcal{G} \leq \epsilon$ on a neighborhood of p of size 1 in the $|\Phi|$ metric, then

$$k_h(p) = O(\epsilon).$$

This follows from standard elliptic estimates which give bounds on the gradient of \mathcal{G} (see [32] for details).

4. The geometry of quadratic differentials

Having seen that estimates on \mathcal{G} depend to a great extent on the geometry of the Hopf differential Φ (in particular, on placement with respect to the zeros of Φ and injectivity radius in the $|\Phi|$ metric), we proceed to examine this geometry more carefully. Our main goal is to prepare the ground for Theorem 5.1 (polygonal region). We will also obtain some estimates of extremal lengths which will be useful in § 8.

4.1. Definitions. The distinguishing characteristics of the $|\Phi|$ metric are the following:

(A) It is Euclidean everywhere except on a set \mathcal{Z} of isolated singularities. A singularity p has concentrated negative curvature $-n\pi$, or a cone angle of $(n+2)\pi$, where $n = \deg(p)$ is a positive integer.

(B) It admits a foliation by straight lines with $(n+2)$ -prong singularities at degree n singular points of the metric.

Let $d(\cdot, \cdot)$ denote a *complete* metric on a surface S with properties (A) and (B). Special cases in which we will be interested will be when S is (i) the closed surface M , (ii) an open disk Δ (usually the universal cover of M), and (iii) an open cylinder Γ (usually the cover of M corresponding to some simple closed curve $\gamma \subset M$).

Condition (A) alone is enough to give the metric d many properties which are analogous to those of smooth metrics of nonpositive curvature. We discuss them here without rigorous proof (see [2]). A geodesic is defined as a path that is locally shortest, and from this one can conclude that geodesics consist of Euclidean straight lines away from the singularities, which might meet at a singularity making an angle no less than π on either side. Between any two points in a complete disk Δ there is a unique geodesic segment (see, e.g., [27]); the uniqueness follows from an application of the Gauss-Bonnet theorem.

Similarly, for any nontrivial free homotopy class of simple closed paths on a surface S there is a geodesic representative. The representative might not be simple because it may have self-tangencies along subarcs. (Consider, for example, the complex plane slit along two congruent real intervals, the

resulting holes then sewn together to make a handle. The curve running around both slits in the original plane has a geodesic representative with such a self-tangency.) However it will not have self-crossings by the Gauss-Bonnet theorem, and its lift in the cylindrical covering corresponding to the homotopy class will be simple.

The geodesic representative is unique, except for the case where it is one of a continuous family of closed Euclidean geodesics in a Euclidean annulus (obtained by gluing a pair of opposite sides of a Euclidean rectangle). These annuli, which we also call *flat cylinders*, will play an important role in our analysis. We note here that a *maximal* flat cylinder \mathcal{F} is one that contains all its geodesic representatives, and that such a cylinder must necessarily contain singularities on both of its boundaries.

We will regard as *annuli* some 2-complexes that are not true annuli, but rather closed regions between two concentric curves that might touch. In this case we call them “pinched” annuli if the distinction is important (in particular a single closed curve can be considered an annulus with no interior).

The curvature of a piecewise-smooth path γ in S is well defined (as a measure, with atoms at the corners) up to choice of sign if γ does not meet any singularities. If γ is the boundary of some set C , we choose the sign to be positive when the acceleration vector points into C (or, at a corner, when the interior angle is at most π). In such a case we say that γ is *curved inwards* (or positively) with respect to C . If γ passes through a singularity, we can use the same definition, but we must keep in mind that γ might be curved outwards with respect to both C and its complement. If γ is curved positively (or negatively) with respect to C at every point, we say it is *monotonically curved* with respect to C .

Because d admits a foliation by straight lines (property (B)), the curvature of a path γ is just the rate of turning of the foliation leaves relative to γ . The total curvature $\kappa(\gamma)$ of any closed curve γ that contains no singularities is therefore an integral multiple of π (note that this is just the argument principle in disguise). It is not hard to show, using the Gauss-Bonnet theorem, that there is also an upper bound $2\pi|\chi(M)|$ to the magnitude of $\kappa(\gamma)$ for a separating simple closed curve γ . This is false for a nonseparating curve. However, if we assume that γ is monotonically curved, then it and its geodesic representative bound an embedded (possibly pinched) annulus (see the discussion of convexity), and another Gauss-Bonnet argument yields a bound of $\kappa(\gamma) \leq 4\pi|\chi(M)|$.

Controlling neighborhoods. The dimensions of R -neighborhoods in a compact surface are controlled with a standard application of the Gauss-

Bonnet theorem. Let $\mathcal{N}_R(C)$ denote the R -neighborhood of a set C . We have:

Lemma 4.1 (Size of R -neighborhood). *For any connected 2-complex C with piecewise smooth boundary in a compact surface M with a quadratic differential metric,*

$$l(\partial\mathcal{N}_R(C)) \leq l(\partial C) + K_1R,$$

$$\text{Area}(\mathcal{N}_R(C)) \leq \text{Area}(C) + l(\partial C)R + K_2R^2,$$

where K_1 and K_2 depend only on $\chi(M)$.

Proof of Lemma 4.1. The piecewise-smooth level curves $\gamma_r = \partial\mathcal{N}_R(C)$ move at unit speed (with respect to r) orthogonally to themselves, so at all but the finitely many values of r where γ_r passes through a singularity,

$$\frac{d}{dr}l(\gamma_r) = \kappa(\gamma_r).$$

By the Gauss-Bonnet theorem,

$$\begin{aligned} \kappa(\gamma_r) &= 2\pi\chi(\mathcal{N}_r(C)) - K(\mathcal{N}_r(C)) \\ &\leq 2\pi(1 + |\chi(M)|), \end{aligned}$$

where $K(X)$ is the total Gaussian curvature of X , obtained in this case by summing over the singularities in X .

Integrating, we have

$$l(\gamma_r) \leq l(\gamma_0) + 2\pi(1 + |\chi(M)|)R,$$

and integrating again gives the bound on area.

4.2. Convexity. As for any smooth negatively curved metric, if S is simply-connected, then the distance function $d(\cdot, \cdot)$ is *convex* in the sense that for every pair of geodesics γ, β parametrized at constant speeds, $d(\gamma(t), \beta(t))$ is a convex function of t . (It is easy to construct a proof of this without appeal to the general machinery by using elementary Euclidean geometry and considering the cases that occur when geodesics pass through singularities.) We will call a set C in S *convex* if any path connecting two points in C can be deformed to a geodesic lying in C .

We list without proof some standard consequences of convexity:

Lemma 4.2 (Convexity facts). *Let d be a complete convex distance function on an open ball B^m . Then the following hold:*

(1) *If $f: B^m \rightarrow \mathbf{R}$ is a convex function, then $L_R(f) = \{x \in B^m: f(x) \leq R\}$ is a convex set.*

(2) *If $C \subset B^m$ is convex, then the function $x \mapsto d(x, C)$ is convex.*

(3) For any convex $C \subset B^m$ there is a well-defined projection $\pi: B^m - C \rightarrow \partial C$ that takes each point of $B^m - C$ to its closest point in C . The projection is distance-nonincreasing.

An immediate consequence of (1) and (2) is that R -balls in a complete disk Δ are convex, as are R -neighborhoods of geodesics. Similarly, for a complete cylinder Γ we can see that an R -neighborhood of a core geodesic is convex, by lifting to the universal cover.

Nonpositive curvature supplies the following *isoperimetric inequality*:

$$(4.1) \quad \text{Area}(\mathcal{D}) \leq \frac{1}{4\pi} l^2(\partial\mathcal{D})$$

for any topological disk $\mathcal{D} \subset \Delta$ with rectifiable boundary. This follows immediately from the results in [22] by considering smooth nonpositively curved approximations to our singular metric. Similarly, let \mathcal{A} denote a closed annulus carrying a quadratic differential metric. (\mathcal{A} could be a subannulus of Δ , or of a complete cylinder Γ .) If L is a bound for the lengths of the boundaries of \mathcal{A} , and D is the minimal distance between them, then

$$(4.2) \quad \text{Area}(\mathcal{A}) \leq \frac{1}{\pi} (D + L)^2$$

is obtained by cutting along the shortest curve connecting the boundaries and applying (4.1).

The *convex hull* $\mathcal{H}(U)$ of a set $U \subset S$ is the smallest convex subset of S containing U . It follows from Lemma 4.2 that bounded sets in Δ or in a complete cylinder Γ have bounded convex hulls (the latter obtained by lifting to the universal cover and considering neighborhoods of a lift of a geodesic representative of the core curve). We can obtain more precise information:

Lemma 4.3 (Bounds on convex hulls). *Let Δ and Γ be a disk and a cylinder with complete metrics satisfying property (A). Let $U \subset \Delta$ be a closed disk, and γ a closed curve isotopic to the core of Γ . Then $\mathcal{H}(U)$ is a closed disk, $\mathcal{H}_{(\gamma)}$ is a closed cylinder retracting to γ , and:*

- (1) $l(\partial\mathcal{H}(U)) \leq l(\partial U)$,
- (2) $\mathcal{H}(U) \subset \mathcal{N}_{l(\partial U)}(U)$,
- (3) $l(\alpha) \leq l(\gamma)$ for each component α of $\partial\mathcal{H}(\gamma)$.

If, in addition, the metric on Γ is foliated (i.e., satisfies property (B)), then:

- (4) $\mathcal{H}(\gamma) \subset \mathcal{N}_{l(\gamma)}(\gamma)$,
- (5) $\text{Area}(\mathcal{H}(\gamma)) \leq \frac{4}{\pi} l^2(\gamma)$.

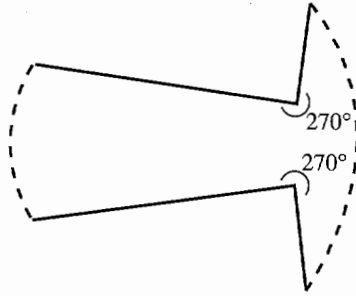


FIGURE 1. THE DOUBLE OF THIS FIGURE ALONG ITS UPPER AND LOWER (SOLID) BOUNDARIES HAS A METRIC WITH PROPERTY (A), BUT DOES NOT ADMIT A SINGULAR FOLIATION BY STRAIGHT LINES. PARTS (4) AND (5) OF LEMMA 4.3 CAN FAIL ARBITRARILY BADLY HERE

Proof. The boundary of a convex set cannot contain homotopically trivial components unless the set is a disk, since it is clear that the convex hull of the boundary of a disk contains the disk. Therefore $\mathcal{CH}(U)$ is a closed disk, and $\mathcal{CH}(\gamma)$ is a closed cylinder.

Let α be an arc of $\partial\mathcal{CH}(U)$ which does not touch U . Then α must be a geodesic, because otherwise we could replace a small piece of α by a shortcut into the interior. Thus, if $E = \partial\mathcal{CH}(U) \cap U$ is the set of extreme points of U , E is closed and every arc of $\partial\mathcal{CH}(U) - E$ (note that E must be nonempty since there are no closed geodesics in Δ) is shorter than the corresponding arc of ∂U , and (1) follows. The same reasoning applies to (3), except that the boundary of $\mathcal{CH}(\gamma)$ has two components, and it is conceivable that one (or both) of them misses E , if it is a geodesic.

Part (2) follows immediately from part (1) of Lemma 4.2 (convexity facts): because any point in U is contained in a geodesic segment with endpoints on ∂U , and geodesics in Δ are unique, U lies in the $(l(\partial U)/2)$ -neighborhood of any point in ∂U , which is a convex set by the lemma. Thus $\mathcal{CH}(U)$ lies in the same set.

Parts (4) and (5) are slightly trickier because they depend strongly on the foliation. In fact they are false for a metric satisfying only property (A), as evidenced by the example in Figure 1.

If γ intersects a geodesic core γ^* of Γ , then again as a consequence of Lemma 4.2(1) the convex hull of γ is contained in an $(l(\gamma)/2)$ -neighborhood of γ^* . (4) follows immediately, and (5) is a consequence of inequality (4.2). Assume therefore that γ lies on one side (say the right side) of γ^* , where we now take γ^* to be the rightmost geodesic if Γ has a

flat part. We know then that $\mathcal{H}(\gamma)$ lies to the right of γ^* (in fact γ^* is the left boundary of $\mathcal{H}(\gamma)$). Let γ_r be the right boundary of the r -neighborhood of γ^* . As in the previous section, the integral of curvature around γ_r is a positive multiple of π , except for isolated values of r for which γ_r contains a singularity of the metric, and where the integral is not defined. Therefore,

$$\frac{d}{dr}l(\gamma_r) = \int_{\gamma_r} k = n\pi, \quad n \geq 1,$$

so $l(\gamma_r) \geq \pi r$. Let r_0 be the first r such that γ_r touches γ . Then $l(\gamma_{r_0}) \leq l(\gamma)$ because the closest-point projection to a convex curve (like γ_r) is distance-decreasing. This gives us

$$d(\gamma^*, \gamma) = r_0 \leq \frac{1}{\pi}l(\gamma).$$

Since (again because of Lemma 4.2) $\mathcal{H}(\gamma)$ is contained in the annulus bounded by γ^* and $\gamma_{r_0+l(\gamma)/2}$, (4) follows immediately, and inequality (4.2) again implies (5).

Boundary-convex hulls. In a closed surface M it is impractical to take convex hulls, since by our definition the only convex sets are the empty set and M itself. Instead the useful notion to work with is that of *boundary-convex* sets. A set C is boundary-convex if its boundary is everywhere curved outward with respect to $M - C$ (this is not quite the same as being curved inward with respect to C ; see §4.1). It is easy to see that C is boundary-convex if and only if for every component U of C , its lift to the cover of M determined by the image of $\pi_1(U)$ in $\pi_1(M)$ is convex. Another way to say this is that any geodesic arc which is deformable rel endpoints into C in fact lies in C . It is also clear that each component of a boundary-convex set is *incompressible*, i.e., the map $i_* : \pi_1(U) \rightarrow \pi_1(M)$ is injective.

We say that two components γ_1 and γ_2 of ∂C (where possibly $\gamma_1 = \gamma_2$) are *r-separated* if every arc α in S with endpoints on γ_1 and γ_2 that cannot be deformed (rel endpoints) into C has length greater than $r \max(l(\gamma_1), l(\gamma_2))$. If all of C 's boundaries are *r-separated* from each other (and from themselves) we say that C is *r-separated*. In other words, disjoint embedded collars of radius $rl(\gamma)/2$ can be appended to C at each boundary component γ .

Theorem 4.4 (Boundary-convex hull). *Let M be a closed surface carrying a quadratic differential metric $d(\cdot, \cdot)$, and let $A \subset M$ be any closed 2-complex with piecewise-smooth boundary. For any $s \leq 0$ there is a set containing A , which we call $\mathcal{B} = \mathcal{BCH}(A; s)$, such that*

1. \mathcal{B} is boundary-convex and s -separated,
2. $l(\partial\mathcal{B}) \leq k_1 l(\partial A)$,
3. $\text{Area}(\mathcal{B}) \leq \text{Area}(A) + k_2 l^2(\partial A)$, and
4. $\mathcal{B} \subset \mathcal{N}_{k_3 l(\partial A)}(A)$, where k_1, k_2 , and k_3 are constants depending only on the topological type of M , and on s .

Proof. Let U be a component of A , and M_U the cover of M corresponding to the inclusion of $\pi_1(U)$ in $\pi_1(M)$. The basic step of the construction is to lift U to its homeomorphic cover $\tilde{U} \subset M_U$, build the convex hull $\mathcal{H}(\tilde{U})$ in M_U , and hope that its projection back down to M is an embedding.

Assume that U is a subsurface, thickening it very slightly if necessary. This allows us to assume that ∂U is a union of simple closed curves. The added area may be made as small as we like and we shall ignore it.

We can control the size of $\mathcal{H}(\tilde{U})$ by observing that it is composed of the union of \tilde{U} with a disk for every boundary component that is compressible in $M_U - \tilde{U}$, and an annulus (probably pinched) for every other boundary component. If γ is a boundary component of the first kind, then the added disk has area bounded by $\frac{1}{4\pi} l^2(\gamma)$, by (4.1). If γ is of the second kind, then either γ is compressible in M_U or it is not. If it is, M_U is a disk Δ and $\mathcal{H}(\tilde{U}) = \mathcal{H}(\gamma)$, whose area is again bounded by $\frac{1}{4\pi} l^2(\gamma)$, and whose boundary has length at most $l(\gamma)$ by Lemma 4.3 (bounds on convex hulls). If γ is nontrivial, we can lift to the cylindrical cover Γ determined by γ . The component of $M_U - \tilde{U}$ bounded by γ lifts homeomorphically to Γ , and the added annulus is just the corresponding component in Γ of $\mathcal{H}(\gamma) - \gamma$ (see Figure 2). Again by Lemma 4.3, the added area is no more than $\frac{4}{\pi} l^2(\gamma)$ and the new boundary has length at most $l(\gamma)$. Part (4) of Lemma 4.3 implies that $\mathcal{H}(\tilde{U})$ lies in the $l(\gamma)$ -neighborhood of \tilde{U} .

Let \mathcal{E} denote the disjoint union $\coprod_{U_i \subset A} \mathcal{H}(\tilde{U}_i)$ of convexified components of A , and $\mathbf{p} : \mathcal{E} \rightarrow M$ the union of the projection maps $\mathbf{p}_i : \mathcal{H}(\tilde{U}_i) \rightarrow M$. If \mathbf{p} is an embedding and the resulting boundary-convex set $\mathbf{p}(\mathcal{E})$ is s -separated, we are done, by setting $\mathcal{B}\mathcal{H}(A; s) = \mathbf{p}(\mathcal{E})$.

If \mathbf{p} is not an embedding, define $A' = \mathbf{p}(\mathcal{E})$. If \mathbf{p} is an embedding but $\mathbf{p}(\mathcal{E})$ is not s -separated, define $A' = \mathbf{p}(\mathcal{E}) \cup \alpha$, where α is a (slightly thickened) arc of length no more than $sl(\partial\mathcal{E}) \leq sl(\partial A)$ whose endpoints are on $\partial\mathbf{p}(\mathcal{E})$ and which is not deformable into $\mathbf{p}(\mathcal{E})$. We then have

$$l(\partial A') \leq l(\partial A)(1 + 2s),$$

(note that α counts twice in the boundary length of A' , once for each

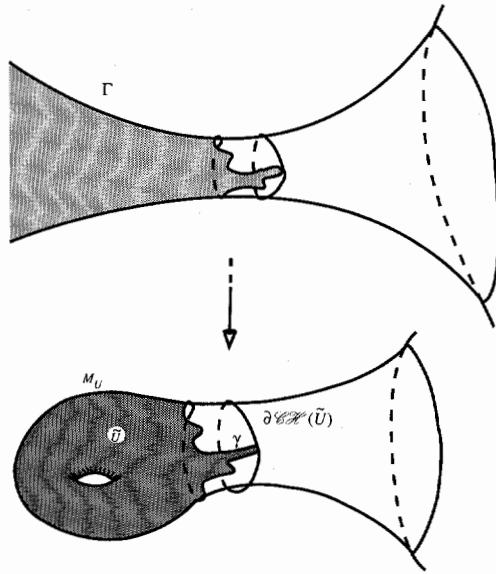


FIGURE 2

side), and

$$\text{Area}(A') \leq \text{Area}(A) + 2l^2(\partial A).$$

We can now repeat the process of taking components and constructing boundary-convex hulls. It remains to show that this can only happen a bounded number of times.

For a subsurface X of M , denote by \hat{X} the surface obtained by appending to X any disk components of $M - X$. We will show that $\chi(M - \hat{A})$, which is nonpositive, grows each time we repeat the above process, and this will bound the number of steps. (The reason we can ignore disk components of $M - A'$ is that they will get filled in anyway in the next step.)

In the case where \mathbf{p} is an embedding and an arc α was added, it is immediate that A' is not retractable into \hat{A} . If the images of two components $\mathbf{p}(\mathcal{H}(\tilde{U}_1))$ and $\mathbf{p}(\mathcal{H}(\tilde{U}_2))$ intersect, the same is true. Suppose finally that \mathbf{p} fails to be an embedding on a single component $\mathcal{H}(U)$. If the projection can be retracted to \hat{U} , then every boundary component γ of the projection can be retracted into $\partial\hat{U}$, and is therefore either homotopically trivial or isotopic to a boundary component of \hat{U} . In the latter case γ lifts homeomorphically to M_U , and therefore must comprise the image of an entire boundary component of $\mathcal{H}(\tilde{U})$. Since the projection

is injective on fundamental groups, all the boundaries of $\mathcal{H}(\tilde{U})$ must project this way, and the former case (γ is homotopically trivial) cannot occur. The projection, being already the restriction of a covering map, is therefore an embedding.

We conclude that, unless we are at the last step, A' (and therefore \hat{A}') does not retract into \hat{A} . The rest is an elementary computation. Recall that we are slightly thickening A' if necessary, so that is a subsurface rather than a general 2-complex. Let $X = M - \hat{A}$, $Y = M - \hat{A}'$, and $W = \text{closure}(X - Y)$. Then $\chi(Y) = \chi(X) - \chi(W) + \chi(W \cap Y)$. For any component Z of W , $\chi(Z \cap Y) - \chi(Z)$ is nonnegative, and zero only if Z is a disk or annulus that retracts to ∂X . Since W does not retract, there must be at least one component that makes a positive contribution to the sum. Therefore, $\chi(Y) > \chi(X)$, and we are done.

4.3 Modulus and extremal length. In a hyperbolic surface there is a direct correspondence between the extremal length of a curve and its hyperbolic length, when these are small enough. To a curve of small extremal length E there always corresponds a long thin hyperbolic cylinder whose core length is approximately $2\pi E$. In a quadratic differential metric a more complicated, but still quantifiable, relationship exists between the geometry of cylinders and the extremal length of their cores.

Let A be a closed annulus in M with boundaries ∂_0 and ∂_1 . Denote their (signed) total curvatures by $\kappa(\partial_0)$ and $\kappa(\partial_1)$, defined with respect to A as in §4.1. Suppose that both boundaries are monotonically curved with respect to A (recall that this means that the curvature vector of a boundary consistently points into (resp. out of) A at smooth points, and internal angles are at most (resp. at least) π at corners). Further, suppose the boundaries are equidistant from each other, and that $\kappa(\partial_0) \leq 0$. We then call A a *regular annulus*, and if $\kappa(\partial_0) < 0$ we call A *expanding* and say that ∂_0 is the inner boundary and ∂_1 is the outer boundary. If the interior of A contains no zeros we say A is a *primitive annulus*, and write $\kappa(A) = -\kappa(\partial_0) = \kappa(\partial_1)$. When $\kappa(A) = 0$, it is a *flat annulus* and is foliated by closed Euclidean geodesics homotopic to the boundaries.

The modulus of a flat annulus is clearly $\text{Mod}(A) = d(\partial_0, \partial_1)/l(\partial_0)$. The other well-known example is a round annulus between two concentric circles in the plane, for which $\kappa(A) = 2\pi$ and $\text{Mod}(A) = \frac{1}{2\pi} \log l(\partial_1)/l(\partial_0)$. Therefore we define

$$\mu(A) = \begin{cases} \frac{1}{\kappa(A)} \log(l(\partial_1)/l(\partial_0)) & \text{if } \kappa(A) > 0, \\ d(\partial_0, \partial_1)/l(\partial_0) & \text{if } \kappa(A) = 0 \end{cases}$$

for a primitive regular annulus A .

A regular annulus is foliated by level curves $\gamma_r = \{p \in A : d(p, \partial_0) = r\}$ which by Lemma 4.2 (convexity facts) are themselves monotonically curved. Therefore by cutting along the curves γ_{r_i} which contain singularities we decompose A canonically into a union $A = \bigcup A_i$ of most $2|\chi(M)| + 1$ primitive annuli, where $\partial A_i = \gamma_{r_i} \cup \gamma_{r_{i+1}}$. We then define $\mu(A) = \sum \mu(A_i)$. We justify this notation with:

Theorem 4.5 (Modulus of regular annulus). *For any regular annulus $A \subset M$,*

$$\mu(A) \leq \text{Mod}(A) \leq \mu(A) + \min(c_1 \mu(A), c_2 \sqrt{\mu(A)}),$$

where c_1 and c_2 depend only on $\chi(M)$.

Proof. Define a scaling function $\rho(r) = l(\gamma_0)/l(\gamma_r)$ and let d_ρ denote the distance function obtained by the conformal change of metric $|\Phi(p)dz^2| \mapsto \rho^2(d(p, \partial_0))|\Phi(p)dz^2|$ where $|\Phi dz^2|$ is the original metric of M . In this metric each γ_r has length $l_\rho(\gamma_r) = l(\gamma_0) = l(\partial_0) \equiv L$.

Let r_1, \dots, r_n be the values of r for which γ_{r_i} pass through singularities, set $r_0 = 0$ and $r_{n+1} = R = d(\partial_0, \partial_1)$, and let A_0, \dots, A_n be the corresponding primitive annuli. The total curvature of γ_r (with respect to the annulus it bounds together with ∂_0) is $\kappa(\gamma_r) = \kappa(A_i)$ if $r_i < r < r_{i+1}$ and $\frac{d}{dr}l(\gamma_r) = \kappa(\gamma_r)$, so we can obtain a lower bound H for the ρ -length of any path from ∂_0 to ∂_1 by computing

$$\begin{aligned} H &= \int_0^R \rho(r) dr = \sum_{i=0}^n \int_{r_i}^{r_{i+1}} \frac{l(\partial_0)}{l(\gamma_r)} dr \\ &= \sum_{i=0}^n \frac{l(\partial_0)}{\kappa(A_i)} \log \frac{l(\gamma_{r_{i+1}})}{l(\gamma_{r_i})} = L\mu(A). \end{aligned}$$

By a similar computation we can obtain

$$\text{Area}_\rho(A) = \int_0^R \rho^2(r)l(\gamma_r) dr = L^2\mu(A).$$

Therefore (see §2),

$$\text{Mod}(A) \geq \frac{H^2}{\text{Area}_\rho(A)} = \mu(A).$$

The bound in the other direction is trickier. In particular, the ρ -length of the circumference of A is not bounded below by L . Denote by A_ϵ the

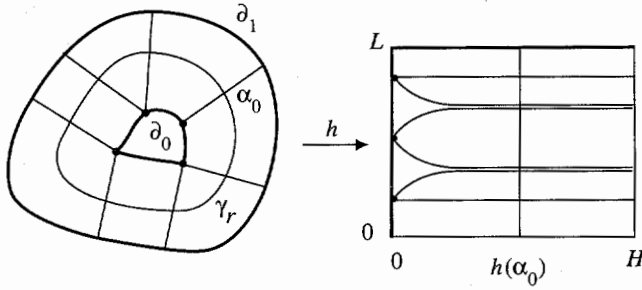


FIGURE 3

Euclidean annulus of circumference L and height H , parametrized as a rectangle $[0, H] \times [0, L]$ with edges $[0, H] \times \{0\}$ and $[0, H] \times \{L\}$ identified. We will construct a map $h: (A, d_\rho) \rightarrow A_e$ that is approximately an isometry, and then use a simple inequality involving the energy of h to get an estimate of the modulus.

The complement of the singularities in the interior of A is foliated by arcs orthogonal to the level curves γ_r , which represent shortest paths from points in A to ∂_0 (see Figure 3). Let α_0 be one such "radial" arc. We define h by the requirement that h maps α_0 isometrically to $[0, h] \times \{0\}$ and each level set γ_r is isometrically to the circle corresponding to $\{\tilde{r}\} \times [0, L]$, where $\tilde{r}(r) = \int_0^r \rho(s) ds$ is the ρ -distance of γ_r from γ_0 .

We proceed to estimate the energy of h . Fix another radial arc α , parametrized with respect to arclength by r , and with respect to ρ -arclength by \tilde{r} . After choosing one of the two rectangles bounded by α_0 and α , let $y(r)$ denote $l(\beta_r)$, where β_r is the segment of γ_r between α_0 and α in the rectangle. Let $\tilde{y}(\tilde{r})$ denote $l_\rho(\beta_{r(\tilde{r})})$. For $r_i < r < r_{i+1}$, we have $\frac{dy}{dr} = k(A_i)$, where $k(A_i) \in [0, \kappa(A_i)]$ is the total curvature of β_r . We also have, by the definition of ρ ,

$$\frac{d\rho}{dr} = \frac{l(\gamma_0)}{l^2(\gamma_r)} \frac{d(l\gamma_r)}{dr} = -\rho^2(r) \frac{\kappa(A_i)}{L}.$$

Since $\tilde{y}(\tilde{r}(r)) = \rho(r)y(r)$ we have

$$\frac{d\tilde{y}}{dr} = \rho \frac{dy}{dr} + y \frac{d\rho}{dr} = \rho k(A_i) - \rho \frac{\kappa(A_i)}{L} \tilde{y}.$$

Now applying the chain rule using $\frac{d\tilde{r}}{dr} = \rho(r)$, we have

$$(4.3) \quad \frac{d\tilde{y}}{d\tilde{r}} = k(A_i) - \frac{\kappa(A_i)}{L} \tilde{y}, \quad r_i < r < r_{i+1}.$$

By discussion in §4.1, $\kappa(A_i) \leq N_0\pi$ for a constant N_0 depending on $\chi(M)$, so

$$\left| \frac{d\tilde{y}}{d\tilde{r}} \right| \leq N_0\pi.$$

Since $h(\alpha)$ is given by the path $(\tilde{r}, \tilde{y}(\tilde{r}))$ in A_e , this implies

$$(4.4) \quad |dh|^2 \leq 2 + N_0^2\pi^2$$

in the interiors of the A_i . This alone bounds the energy of h by

$$(4.5) \quad \mathcal{E}_h(A) \leq \frac{1}{2}(2 + N_0^2\pi^2)L^2\mu(A).$$

We can do a little better, however. In A_i , the solution of (4.3) is

$$\tilde{y}(\tilde{r}) = [\tilde{y}(\tilde{r}_i) - \frac{L}{\kappa(A_i)}k(A_i)] \exp\left(-\frac{\kappa(A_i)}{L}(\tilde{r} - \tilde{r}_i)\right) + \frac{L}{\kappa(A_i)}k(A_i),$$

so

$$\begin{aligned} \left| \frac{d\tilde{y}}{d\tilde{r}} \right| &= \left| \frac{\kappa(A_i)}{L}\tilde{y}(\tilde{r}_i) - \kappa(A_i) \right| \exp\left(-\frac{\kappa(A_i)}{L}(\tilde{r} - \tilde{r}_i)\right) \\ &\leq \kappa(A_i) \exp\left(-\frac{\kappa(A_i)}{L}(\tilde{r} - \tilde{r}_i)\right). \end{aligned}$$

Therefore the energy of h restricted to A_i is bounded by

$$\begin{aligned} \mathcal{E}_h(A_i) &\leq \frac{1}{2} \int_{\tilde{r}_i}^{\tilde{r}_{i+1}} \left[2 + \kappa^2(A_i) \exp\left(-2\frac{\kappa(A_i)}{L}(\tilde{r} - \tilde{r}_i)\right) \right] L d\tilde{r} \\ &= L(\tilde{r}_{i+1} - \tilde{r}_i) + \frac{\kappa(A_i)L^2}{4} \left[1 - \exp\left(-2\frac{\kappa(A_i)}{L}(\tilde{r}_{i+1} - \tilde{r}_i)\right) \right] \\ &\leq L^2 \left(\mu(A_i) + \frac{1}{4}N_0\pi \right), \end{aligned}$$

Summing over the (at most $N_0 + 1$) cylinders we obtain

$$(4.6) \quad \mathcal{E}_h(A) \leq L^2(\mu(A) + N_1),$$

where N_1 again depends only on $\chi(M)$. This energy gives us bounds on modulus by the following argument. If A_1 and A_2 are two Euclidean annuli, then by an argument like that of Proposition 3.1 (energy lower bound), we can show that the lowest energy for a map from $(A_1, \partial A_1)$ to $(A_2, \partial A_2)$ is

$$\mathcal{E}(A_1, A_2) = \frac{1}{2} \left[\frac{\text{Mod}(A_1)}{\text{Mod}(A_2)} + \frac{\text{Mod}(A_2)}{\text{Mod}(A_1)} \right] \text{Area}(A_2),$$

(and is in fact realized by a linear map). Since energy depends on the conformal rather than the metric structure of the domain, this gives a lower bound for the energy of any map to A_2 from an annulus conformally equivalent to A_1 . Applying this to our situation, where $\text{Mod}(A_e) = \mu(A) = L^2/\text{Area}(A_e)$, we have

$$\frac{\mathcal{E}_h(A)}{\text{Area}(A_e)} \geq \frac{1}{2} \left[\frac{\text{Mod}(A)}{\mu(A)} + \frac{\mu(A)}{\text{Mod}(A)} \right]$$

or, by applying (4.5) and (4.6),

$$\frac{1}{2} \left[\frac{\text{Mod}(A)}{\mu(A)} + \frac{\mu(A)}{\text{Mod}(A)} \right] \leq \min\left(1 + \frac{N_1}{\mu(A)}, 1 + N_2\right),$$

where $N_2 = 1 + N_0^2\pi^2/2$. The statement of the theorem follows easily. q.e.d.

We note briefly some geometric facts that follow from the above discussion. For an expanding annulus A , $l(\gamma_r) = l(\gamma_0) + k(r)r$, where $k(r)$ is bounded below by π and above by $N_0\pi$. Thus, if $r(A) = d(\partial_0, \partial_1)$ is the radius of A , then $r(A)/l(\partial_0)$ is bounded above and below by two exponential functions of $\mu(A)$. Integrating, we see that (in the original metric) the area of A is bounded above and below by $l^2(\partial_0)$ times exponential functions of $\mu(A)$. For a flat annulus, of course, $\text{Area}(A) = l^2(\partial_0)\mu(A)$.

What we really want in this section is a statement relating arbitrary high modulus annuli in M to the quadratic differential metric. The following theorem supplies this, at the expense of the sharpness of the bounds.

Theorem 4.6 (Modulus of any annulus). *If $A \subset M$ is any homotopically nontrivial annulus with $\text{Mod}(A) \leq m_0$, then A contains a regular annulus B such that*

$$\mu(B) \geq c_3 \text{Mod}(A) - c_4,$$

where m_0, c_3 , and c_4 depend only on $\chi(M)$.

Proof. The technique is essentially that of the previous theorem, except that we first have to find B .

If the total curvatures of both boundaries of A are positive with respect to A , then we can find a circumferential curve $\gamma \subset A$, dividing A into two annuli A_0 and A_1 , whose total curvature is nonpositive with respect to both annuli, and such that $\text{Mod}(A_0) + \text{Mod}(A_1) = \text{Mod}(A)$. Thus there is a subannulus $A' \subseteq A$ with $\text{Mod}(A') \leq \frac{1}{2} \text{Mod}(A)$ such that at least one of its boundaries, say ∂'_0 , has nonpositive total curvature with respect to A' .

Lift A' homeomorphically to the cylindrical cover Γ corresponding to its core. Let γ_0 be the boundary of $\mathcal{H}(\partial'_0)$ which lies on the same side of



FIGURE 4

∂_0' as A' (see Figure 4). γ_0 must touch ∂_0' because otherwise γ_0 would be a geodesic core of Γ , and ∂_0' would have positive total curvature with respect to A' , by the Gauss-Bonnet Theorem. If γ_r is the r -equidistant curve from γ_0 on the same side, let R denote the first value of r for which γ_r touches ∂_1' . Let B be the regular annulus bounded by γ_0 and γ_r (if $R = 0$, then B is a single curve, which may not be entirely contained in A').

The energy argument used to bound the modulus from above in the previous theorem encounters technical problems here, because we would need to define the map h on $A' - B$, whose shape could be very wild. Rather than involve ourselves in this messy problem, we will use a simpler estimate and be satisfied with bounds which are not as tight.

Define $\rho(r)$ as in the previous theorem. Let $r: \Gamma \rightarrow \mathbf{R}$ be defined as signed distance from γ_0 so that $r < 0$ on the side of γ_0 containing ∂_0' . In particular $r(\gamma_r) = r$. Define a function $\rho': \mathbf{R} \rightarrow \mathbf{R}$ as follows:

$$\rho'(r) = \begin{cases} 1 & \text{if } -l(\gamma_0)/2 < r < 0, \\ \rho(r) & \text{if } 0 \leq r \leq R, \\ \rho(R) & \text{if } R < r < R + l(\gamma_R)/2, \\ 0 & \text{otherwise,} \end{cases}$$

and let d_ρ denote the metric d scaled by $(\rho'(r(p)))^2$. Let us compute the circumference $C_\rho(A')$, i.e., the minimal d_ρ -length of a curve in A' isotopic to the boundary.

We first claim the minimal length is achieved by a curve lying entirely to the right of γ_0 (in other words the subannulus of Γ where $r \geq 0$). If γ is a length minimizing curve, it must at least enter the right side of γ_0 since γ_0 touches ∂_0' . Let α be an arc of γ with endpoints on γ_0 and interior on the left side of γ_0 . If α leaves the $(l(\gamma_0)/2)$ -neighborhood of

γ_0 then its $d_{\rho'}$ -length is at least $l(\gamma_0)$. If it does not, then $l_{\rho'}(\alpha) = l(\alpha)$. In either case the geodesic segment of γ_0 homotopic to α rel endpoints is shorter than α in the ρ' metric.

On the other hand, if an arc B of γ exits on the other side, then it can be deformed to γ_R by the distance-decreasing projection given in Lemma 4.2 (convexity facts). We conclude that $C_{\rho'}(A') \geq C_{\rho'}(B)$. Note that this holds true even in the case where $R = 0$ and B is not contained in A' .

The $d_{\rho'}$ metric restricted to B is exactly the d_{ρ} metric of the previous theorem, so we can construct a map $h: B \rightarrow B_e$ where B_e is the Euclidean annulus of circumference $L = l(\gamma_0)$ and height $H = L\mu(B)$. We can again bound the derivative of h by using (4.4), so

$$C_{\rho'}(B) \geq \frac{1}{\sup |dh|} C_{\text{Euc}}(B_e) = \frac{L}{\sqrt{2 + N_0^2 \pi^2}}.$$

We obtain an upper bound on the $d_{\rho'}$ -area of A' by observing that the area of $A' \cap r^{-1}[-l(\gamma_0)/2, 0]$ is bounded by $N_3 l^2(\gamma_0)$, and that the area of $A' \cap r^{-1}[R, R + l(\gamma_R)/2]$ is bounded by $N_3 l^2(\gamma_R)$ where N_3 depends only on $\chi(M)$ (Lemma 4.1 (size of R -neighborhood)). Thus,

$$\text{Area}_{\rho'}(A') \leq L^2(2N_3 + \mu(B)).$$

By the definitions in §2, this implies

$$\text{Mod}(A') \leq \frac{\text{Area}_{\rho'}(A')}{C_{\rho'}^2(A')} \leq N_4 + N_5 \mu(B),$$

where N_4 and N_5 are combinations of the previous constants. Note in particular that if $\text{Mod}(A) > 2N_4$, then B is actually contained in A' . The theorem follows, with $m_0 = 2N_4$, $c_3 = 1/2N_5$, and $c_4 = N_4/N_5$.

4.4. Injectivity radius. The injectivity radius in a quadratic differential metric, unlike in a hyperbolic surface, has no intrinsic conformal meaning because there is no natural scale at which to view such a metric. We can expand or contract it by any constant and still have a quadratic differential metric, and indeed conformally equivalent subsurfaces can have injectivity radii that are widely different. Taking this into account, we can still produce a geometric decomposition of the surface analogous to, and bearing direct correspondence with, the thick-thin decomposition of a hyperbolic surface.

We start with a rough characterization of injectivity radius in terms of distance from the set \mathcal{Z} of singularities of the metric.

Lemma 4.7 (Injectivity away from zeros). *The injectivity radius of a point $p \in M - \mathcal{N}_R(\mathcal{Z})$ is at least $2R$, unless p lies in a flat cylinder of M . Conversely, let U be a component of $\mathcal{B}_R = \mathcal{BCH}(\mathcal{N}_R(\mathcal{Z}); 0)$. If U is not a disk, then the injectivity radius anywhere in U is at most KR , where K depends only on $\chi(M)$.*

Recall that $\mathcal{BCH}(X; 0)$ denotes the boundary-convex hull of X (with no separation requirement on the boundaries).

Proof of Lemma 4.7. Let γ be any geodesic path with both endpoints on $p \in M - \mathcal{N}_R(\mathcal{Z})$. If γ passes through a singularity of the metric, then $l(\gamma) \geq 2R$ since $d(p, \mathcal{Z}) \geq R$. If, on the other hand, γ contains no singularities, then it is a Euclidean geodesic, and makes a constant angle with the foliation. Thus in fact the two ends of γ meet at p making an angle of π , and γ is a closed geodesic core of a flat cylinder. This proves the first statement.

To prove the converse, let β be a boundary component of U . Since U is not a disk, β is nontrivial, and its length is at most some K_0R by Lemma 4.1 (size of R -neighborhood) and Theorem 4.4 (boundary-convex hull). Thus the injectivity radius anywhere on the boundary is at most $K_0R/2$. By part (4) of Theorem 4.4, $U \subset \mathcal{N}_{c_1R}(\mathcal{Z})$ with $c_1 = c_1(\chi(M))$, so any point in U can be connected to ∂U by a path of length at most c_1R . The injectivity radius is therefore bounded above by $(2c_1 + K_0)R$. q.e.d.

We can now form something analogous to a thick-thin decomposition. Choose $\varepsilon \ll \min(\varepsilon_0/2\pi, 1/m_0)$, where ε_0 is the Margulis constant, and m_0 is the modulus from Theorem 4.6 (modulus of any annulus). Let $\{\gamma_1, \dots, \gamma_n\}$ be the simple closed curves of M whose extremal lengths, $\{E(\gamma_i)\}$, are less than ε . These curves must then be disjoint, and we can find a collection of disjoint annuli $\{A_i\}$ with moduli $\text{Mod}(A_i) \geq 1/E(\gamma_i) - m_1$, where m_1 is a constant depending only on $\chi(M)$; this is a standard consequence of the thick-thin decomposition—just take the ε_0 Margulis tubes around the γ_i .

By Theorem 4.6 (modulus of any annulus), each A_i contains a regular annulus B_i with $\mu(B_i) \geq c_3 \text{Mod}(A_i) - c_4$. Let C_i denote the maximal union of regular annuli contained in A_i . These C_i will be our “thin” part, and $M_T = M - \cup C_i$ will be our “thick” part. Note that because of the slack in Theorem 4.6, M_T might actually contain some of the hyperbolic thin part, but it certainly contains the entire hyperbolic thick part, and on the other hand the C_i account for a definite fraction of every component of the thin part.

5. Building \mathcal{P}_R

From here on we assume that $M = (S, \sigma)$ and $N = (S, \rho)$ represent different hyperbolic structures on an underlying closed surface S , and that $f: M \rightarrow N$ is the unique harmonic map in the homotopy class of the identity.

Lemma 3.2 (rough bound) provides a preliminary bound on \mathcal{G} whenever the injectivity radius of the $|\Phi|$ metric is not too small. Lemma 3.3 (exponential estimate) provides an exponential improvement in this bound as $|\Phi|$ -distance from the zeros increases. Lemma 4.7 (injectivity away from zeros) shows that distance from zeros implies large injectivity radius, except on thin flat cylinders. The situation on such cylinders is not hard to control, either. Let \mathcal{F} be a maximal flat cylinder in the $|\Phi|$ metric with circumference $W \leq 1$ (recall from §4.1). By Lemma 3.2 (rough bound) and Lemma 3.3 (exponential estimate) we have

$$\mathcal{G}(p) \leq \frac{\sinh^{-1}(4|\chi(M)|/W^2)}{\cosh(r - W/2)}$$

at a point $p \in \mathcal{F}$ which is a distance $r > W/2$ from $\partial\mathcal{F}$. Thus, for

$$r \geq r_W = \frac{1}{2}W + \log \frac{1}{2} \sinh^{-1}(4|\chi(M)|/W^2)$$

we have $\mathcal{G}(p) \leq 1$, and the pair of subcylinders $\mathcal{F}_0 = \{p \in \mathcal{F} : d(p, \partial\mathcal{F}) \leq r_W\}$ has $|\Phi|$ -area bounded by a constant A_0 independent of W . In $\mathcal{F} - \mathcal{F}_0$ we have $\mathcal{G}(p) \leq \varepsilon(d(p, \mathcal{F}_0))$ (recall the notation of $\varepsilon(\cdot)$ from §3). If $W > 1$ we can set $r_W = 1$ and still have the exponential estimate.

Thus in general the “bad” parts of M are neighborhoods of zeros and the subcylinders \mathcal{F}_0 of the flat cylinders, and estimates on \mathcal{G} improve exponentially with distance from them. Our goal in this section is to construct a growing family of surfaces \mathcal{P}_R ($R > 0$) which include R -neighborhoods of the “bad” parts and whose shape and dimensions are well-suited to the train-track arguments of the next section, by which we will show that the leaves of Φ_h outside \mathcal{P}_R are mapped by the harmonic map to curves that are $\varepsilon(R)$ -nearly tangent to the corresponding leaves of the geodesic lamination $f(\Phi_h)^*$.

We should think of \mathcal{P}_R as a generalization of the following two examples, which were analyzed in [31]-[33]. First, consider a fixed M and a fixed quadratic differential Φ_0 , and the ray of hyperbolic surfaces $\{N_r : r > 0\}$ leaving every compact set in Teichmüller space for which the harmonic map $f_r: M \rightarrow N_r$ has the Hopf differential $\Phi_r = r\Phi_0$. For large enough r , we can build polygons with alternating horizontal and vertical

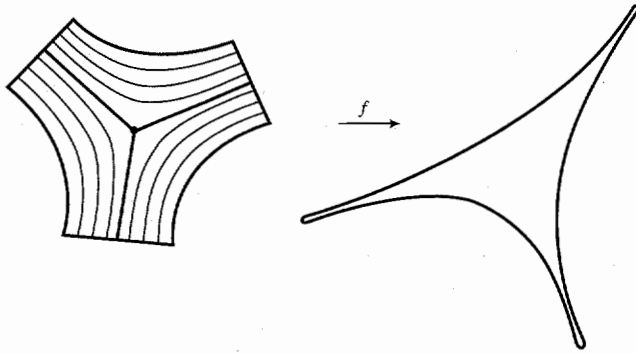


FIGURE 5

sides around the zeros, which take up only a small percentage of the total $|\Phi_r|$ -area of M , and which map approximately to ideal polygons in N_r (see Figure 5).

In the second example we take a fixed surface N and a family of surfaces M_r in which a particular curve γ is shrinking. In that case the Hopf differential Φ_r of f_r develops a long flat cylinder homotopic to γ , and the rest of the surface contributes a bounded amount of Φ_r -area. \mathcal{P}_R in this case (for a bounded value of R) is the complement of the flat cylinder, and the boundary of \mathcal{P}_R maps nearly geodesically to the geodesic representative of γ in N (see Figure 6).

In the general case a boundary component γ of \mathcal{P}_R will be of one of the following two types:

1. An alternating sequence of horizontal and vertical arcs of the quadratic differential, meeting at angles of $\pi/2$ with respect to \mathcal{P}_R ; in this case we call γ a *polygonal* boundary component.
2. A geodesic core of a flat cylinder—in which case we call γ a *straight* boundary component.

The polygonal type of component is a generalization of the first example, and we expect its image under the harmonic map to approximate an ideal polygon whose edges are leaves of the image lamination. The flat type is a generalization of the second example, and we expect the images of leaves of Φ_h in the corresponding cylinder to bunch or spiral around a geodesic in the image, or to approximate a portion of the image lamination in the thin part (depending on the angle of the leaves in the cylinder).

We will also require, for our later convenience, that straight boundary components always occur in pairs bounding flat cylinders. In such a case \mathcal{P}_R will always include at least an $(r_W + R)$ -neighborhood of the boundary

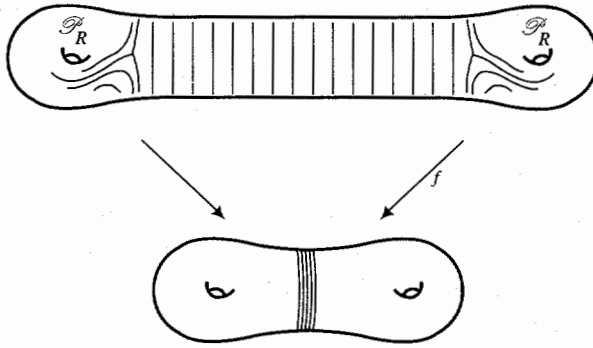


FIGURE 6. PINCHING CYLINDER

of the corresponding maximal cylinder, so that $\varepsilon(R)$ estimates on \mathcal{G} hold in the complement of \mathcal{P}_R , as discussed in the beginning of the section. As R grows, such a flat cylinder will gradually be engulfed until it is contained in \mathcal{P}_R and the straight boundaries disappear. We note that the number of distinct maximal flat cylinders whose subcylinders can occur as components of $M - \mathcal{P}_R$ as R grows from 0 to ∞ is bounded by $3|\chi(M)|/2$, since they represent homotopically distinct, disjoint simple closed curves in M . Accordingly we will number them $\mathcal{F}_1, \dots, \mathcal{F}_n$ ($n \leq 3|\chi(M)|/2$), in the order in which they occur.

In the following theorem we show how to construct \mathcal{P}_R and control its dimensions in various senses.

Theorem 5.1 (polygonal region). *Let $s_1 > 0$ and $c_1, \dots, c_{3|\chi(M)|/2} > 0$ be chosen constants, where M is a closed surface carrying a quadratic differential metric with singularity set \mathcal{Z} . For any $R > 0$ there exists a boundary-convex set $\mathcal{P}_R \subset M$ with the following properties:*

- (1) $\mathcal{N}_R(\mathcal{Z}) \subset \mathcal{P}_R$.
- (2) Every component of $\partial\mathcal{P}_R$ is either polygonal or straight (as above), and the straight components occur in pairs bounding homotopically distinct flat cylinders.
- (3) If \mathcal{F}_k is the k th maximal flat cylinder whose subcylinders occur as components of $M - \mathcal{P}_r$ for $r \leq R$, and \mathcal{F}_k is partially contained in \mathcal{P}_R , then $\mathcal{F}_k \cap \mathcal{P}_R$ is a pair of flat cylinders with length at least $r_W + R + c_k R^2 / W$, where $W = W(\mathcal{F}_k)$ is the circumference of \mathcal{F}_k .
- (4) $l(\partial\mathcal{P}_R) \leq K_1 R$.
- (5) $\text{Area}(\mathcal{P}_R) \leq A_1 + (K_2 + 2 \sum_{i=1}^k c_i) R^2$, where k is the number of flat cylinder components of $M - \mathcal{P}_r$ that have occurred for $r \leq R$.

(6) Each edge of a polygonal boundary component has length at least $K_3 R$.

(7) The polygonal components of $\partial\mathcal{P}_R$ are s_1 -separated.

In the above, A_1 , K_1 , K_2 , and K_3 are constants depending only on s_1 and $\chi(M)$.

Proof. The basic strategy is to form the R -neighborhood of the singularities, boundary-convexify it and “square out the corners” to satisfy property (2). We must take some care to control the size of the resulting set, and also treat the flat cylinders separately.

Let \mathcal{B} denote $\mathcal{BCH}(\mathcal{N}_R(\mathcal{L}); s_2)$, where $s_2 = 1 + \sqrt{2}s_1$. $l(\partial\mathcal{B})$ and $\text{Area}(\mathcal{B})$ are bounded by $k_1 R$ and $k_2 R^2$, respectively, by virtue of Lemma 4.1 (size of R -neighborhood) and Theorem 4.4 (boundary-convex hull), where $k_i = k_i(s_1, \chi(M))$.

Any component of $\partial\mathcal{B}$ which meets a flat cylinder and is isotopic to its core is automatically a core of that cylinder, by the boundary-convexity of \mathcal{B} . Let γ_0 be such a component in a flat cylinder \mathcal{F} , and suppose that the opposite side of \mathcal{F} is not bounded by a corresponding curve. Then there is some core curve γ_1 of \mathcal{F} which touches a component of $\partial\mathcal{B}$ different from γ_0 . Appending γ_1 to \mathcal{B} we obtain a set whose boundary length is greater by no more than $2l(\gamma_1) \leq 2l(\partial\mathcal{B}) \leq 2k_1 R$, so the boundary-convex hull of the new set, $\mathcal{BCH}(\mathcal{B} \cup \gamma_1; s_2)$, still has boundary length and area bounded proportionally to R and R^2 , respectively. We repeat this process as long as necessary. As in Theorem 4.4 (boundary-convex hull), the process must terminate after a bounded number of steps because the topological type of the complement of \mathcal{B} decreases each time. The resulting set, which we call \mathcal{B}' , satisfies the part of property (2) relating to straight boundary components, as well as properties (1), (4), and (5) (with appropriate constants) and is s_2 -separated.

To satisfy (3) we simply extend \mathcal{B}' into each flat cylinder \mathcal{F}_k that it adjoins, until the intersection $\mathcal{B}' \cap \mathcal{F}_k$ is composed of cylinders of the desired length, $r_w + R + c_k R^2 / W$. Note that this can only increase the area of \mathcal{B}' by some $(K_2 + 2c_k)R^2$. If the height of \mathcal{F}_k is less than the desired length of the cylinders being added, we just include the whole cylinder in \mathcal{B}' and eliminate those boundary components. Call the resulting set \mathcal{C} . Note that the polygonal boundaries of \mathcal{C} are still s_2 -separated, and that property (5) is still satisfied.

The next step is to add “corners” to each nonstraight component of \mathcal{C} so that property (2) will be satisfied. Let γ be such a component. The horizontal and vertical foliations turn monotonically around γ , and we

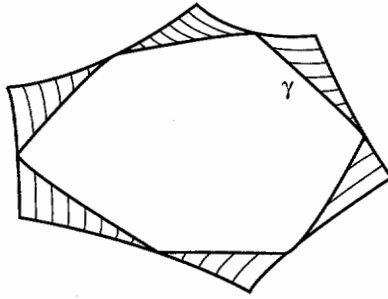


FIGURE 7



FIGURE 8



FIGURE 9

want to add approximately triangular corners with one vertical side, one horizontal side, and one side on γ , as in Figure 7.

Consider an arc α where the line field turns through 90 degrees, from being tangent to α at one endpoint to being perpendicular at the other. The developing image of α into \mathbf{R}^2 , and thus of a neighborhood of it, must be an embedding (Figure 8).

Because the polygonal boundaries of \mathcal{E} are s_2 -separated, there are no other portions of $\partial\mathcal{E}$, and in particular no singularities, within $l(\partial\mathcal{E})$ of α , so we can continue the developing image for that distance. It is an easy fact of Euclidean geometry that a curve which turns through 90° monotonically in the plane is no more than $l/2$ away from its corresponding "corner" (i.e., the path consisting of two orthogonal segments tangent to α at its endpoints), where l is the length of the curve, so we can adjoin a corner (Figure 9) to each such arc, and the resulting surface, which is the desired \mathcal{P}_R , will still be embedded.

The polygonal boundary components of \mathcal{P}_R are in fact still separated by arcs of length at least $(s_2 - 1)l(\partial\mathcal{E})$. Again as a simple consequence of Euclidean geometry, a corner attached to an arc α as above has length at most $\sqrt{2}l(\alpha)$, so $l(\partial\mathcal{P}_R) \leq \sqrt{2}l(\partial\mathcal{E})$ and the polygonal boundaries of \mathcal{P}_R are s_1 -separated by our choice of s_2 . This gives us property (7), as well as (4) and (5) with revised constants.

Property (6) follows from the fact that the boundary of \mathcal{Z} has curvature at most $1/R$ at any point—this is true for the convex parts of the boundary of $\mathcal{N}_R(\mathcal{Z})$, and the boundary of the convex hull can only have less curvature.

6. Building train-tracks

Let α be a (not necessarily continuous) path in a hyperbolic surface. We say that α is (l, ϵ) -*nearly-straight* if it is piecewise C^2 with each C^2 piece having length at least l , its curvature is at most ϵ where defined, and it makes jumps of at most ϵ at its discontinuities (measured in the lift to the unit tangent bundle—so the turning angle is bounded too). Standard hyperbolic geometry shows that, if ϵ is sufficiently small, α is in a $c\epsilon$ -neighborhood of its geodesic representative α^* (rel endpoints, if it is not a closed path), where $c = c(l)$. Further, the closest-point projection $\pi: \alpha \rightarrow \alpha^*$ is a diffeomorphism on each C^2 piece, with $1 - c\epsilon \leq |d\pi| \leq 1$.

This fact together with the estimates from the previous sections on the curvatures (in N) of leaves of Φ_h outside \mathcal{P}_R are, sadly, not sufficient to allow us to conclude that Φ_h is near its geodesic representative in N . The trouble is that we have no control over the leaves of Φ_h inside \mathcal{P}_R , and in general even one very sharp corner is enough to leave an otherwise straight curve far from its geodesic representative.

The saving grace in our situation is that Φ_h in $M - \mathcal{P}_R$ forms a *train-track* (see [12] for a detailed treatment) in the N metric, which greatly limits the amount of damage caused by the lack of control inside \mathcal{P}_R .

We extend the standard definition of train-track to an (l, ϵ) -*nearly straight train-track*, which is composed of branches that are (l, ϵ) -nearly straight paths, meeting in switches where jumps of at most ϵ in the unit tangent bundle can occur between incoming and outgoing branches. It is then clear that any lamination carried on such a track is in a $c\epsilon$ -neighborhood of its geodesic representative.

One can also realize the branches of a track as long thin rectangles, foliated in one direction by long nearly straight arcs, and in the other by very short arcs, called “ties.” Our construction will actually be a hybrid of these two types of train-tracks.

Note. By reparametrizing N we may assume the harmonic map $f: M \rightarrow N$ is precisely the identity on S . Where confusion is unlikely we will make this assumption implicitly.

6.1. The train-track structure of $M - \mathcal{P}_R$. Denote by T_0 the union of components of $M - \mathcal{P}_R$ which are not flat cylinders (we will return to the

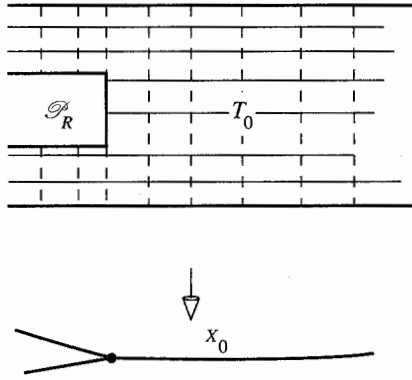


FIGURE 10. A VERTICAL ARC IN ∂T_0 PROJECTS TO A SWITCH OF X_0 . Φ_h ARCS ARE DRAWN SOLID, AND Φ_v ARCS DOTTED

flat cylinders later). Our first attempt at a train-track will be the 1-complex X_0 obtained by identifying points in T_0 which are connected (in T_0) by a vertical arc of Φ . Vertices of X_0 are images of vertical boundary arcs of T_0 , and inherit a switch structure in the obvious way; the edges of X_0 coming from one side of a vertical arc in ∂T_0 are considered to come from one side of the switch, as in Figure 10.

The preimage of a branch of X_0 is either a Euclidean rectangle in the $|\Phi|$ metric whose boundary lies along horizontal and vertical arcs of Φ , or it is a portion of a flat cylinder, as shown in Figure 11 (next page) (note that although T_0 has no components which are flat cylinders it may well contain portions of flat cylinders in its interior). In case (b) of the figure, the interior of the branch has a rectangular preimage, and the branch forms a closed loop.

The $|\Phi|$ metric on T_0 projects to a metric on X_0 which gives each branch the same length as any of the horizontal arcs in the preimage. Recall from (3.3) that this length closely reflects the length in N of the image leaves. We want ultimately to obtain a (l, ϵ) -nearly straight train-track in N , so we need to ensure that the branches of X_0 have lengths at least 1. This will be done with to operations on T_0 :

1. *Splitting* (compare [12],[3]), which consists of enlarging \mathcal{P}_R by moving forward a vertical boundary, thus moving the corresponding switch in X_0 . This is illustrated in Figure 12. Note that there needs to be enough room in $M - \mathcal{P}_R$ for this operation to be possible. The next proposition shows that only a bounded amount of splitting is necessary.

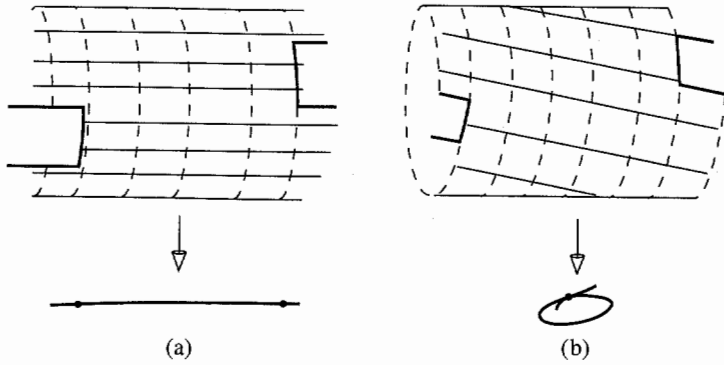


FIGURE 11. IN CYLINDER (B), THE DOTTED ARCS REPRESENT A SINGLE VERTICAL LEAF, WHOSE IMAGE IS THE SWITCH

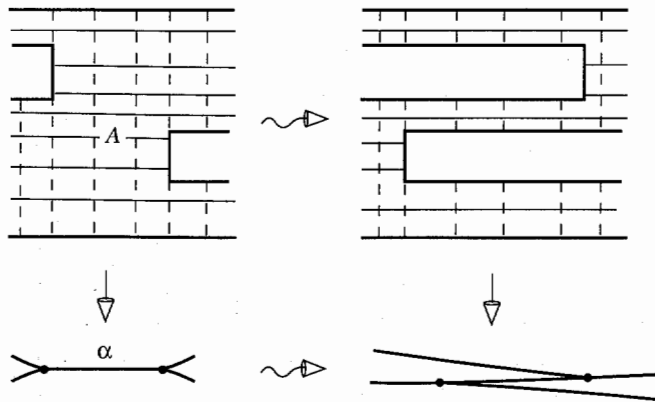


FIGURE 12. SPLITTING ONE SWITCH PAST ANOTHER, IN T_0 AND IN X_0

2. *Slicing* some of the flat cylinders in T_0 along horizontal leaves (Figure 13). Such slices can be long, but they do not affect the area of T_0 . Instead, they produce switches which have only one incoming and one outgoing edge.

Proposition 6.1 (Adjust train-track). *There is a choice of $R_0 > 0$ and a separation constant s_1 for the construction of \mathcal{P}_R such that for $R > R_0$, we can split through a bounded distance and slice through some flat cylinders to obtain a new set T_1 whose train-track X_1 has branches of length at least 1 with interiors that have rectangular preimages in T_1 .*

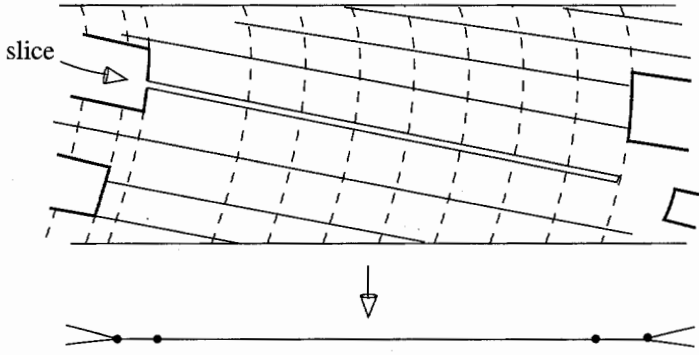


FIGURE 13

Proof. Call a branch of X_0 *short* if its length is at most 3. Let $\alpha \subset X_0$ be a short branch whose preimage in T_0 is a rectangle A . We say that α has an incoming switch if some vertical arc of ∂T_0 is properly contained in ∂A . We can split through α in such a case, by pushing all its incoming switches forward by the length of α , as in Figure 12. The result is to break α into several pieces, some of which may merge with adjacent branches and some of which remain short. All of these short branches no longer have incoming switches, but new incoming switches may have been added to adjacent branches.

Consider now the operation of simultaneously splitting through all the short branches which have incoming switches. After such an operation, if there is still a short branch with incoming switches, then those switches must have been produced by splitting through an adjacent short branch.

Since there is a bounded number of branches (depending on $\chi(M)$) and not all of them are short (provided R_0 is large enough), after some bounded number of repetitions of this operation one of the following two situations must hold: Either all short branches have no incoming switches, or there is a cycle of short branches such that starting at one of them and making successive splits we arrive back at the original branch.

Such a cycle must correspond to a flat cylinder which we can slice, because this splitting is done by pushing a vertical boundary to T_0 across a sequence of rectangles only to meet the first one again, which must therefore wrap around some cylinder (Figure 14, next page). The core of this cylinder is represented by a curve consisting of a segment of vertical boundary of the original rectangle and a horizontal segment that runs along the new portion of $\partial \mathcal{P}_R$ added during the splitting. Two such cylinders must therefore have disjoint cores, so there is a bounded number of them that can occur.

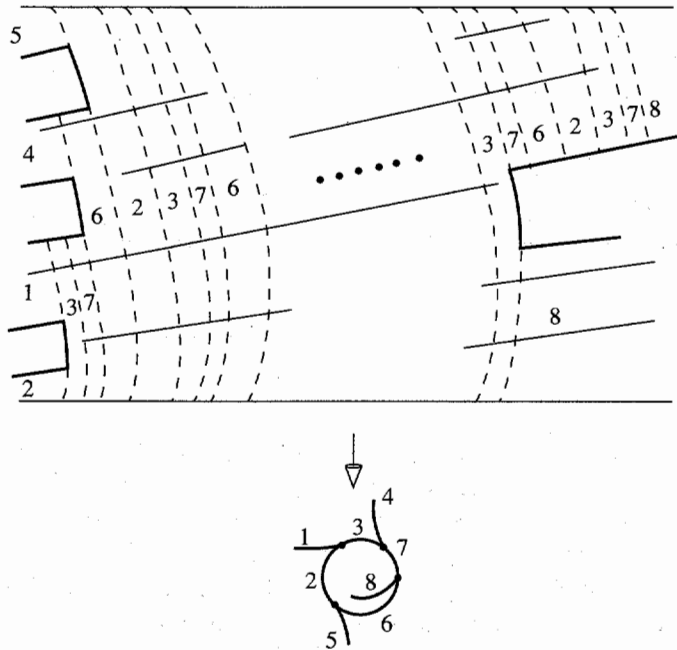


FIGURE 14. A CYCLE OF SHORT BRANCHES ARISING FROM A FLAT CYLINDER. SLICING THROUGH THIS CYLINDER WILL PRODUCE A LONG BRANCH SPIRALING AROUND THE FORMER CYCLE

If we now alter our previous procedure by first slicing through all such cylinders (including cylinders whose cores are vertical leaves) we know that the result of the bounded splitting procedure must be that no short branches have incoming switches.

Let α be a branch of length at most 1 after the above procedure is done. α can have no incoming switches, so it must have an endpoint at which the switches to either side of α are incoming to an adjacent branch, β . Thus $l(\beta) \geq 3$. Split all of β 's incoming switches a distance 1 into β . This leaves $l(\beta) \geq 1$, and now all the branches adjacent to it have length at least 1 also. No new incoming switches have been generated, so we can repeat this step until all branches have length at least 1.

Since we have shown that the total distance through which we need to split is bounded, this is easily made possible by a suitable choice of constant s_1 in Theorem 5.1 (polygonal region).

Note. We will assume from now on that the splitting adjustments of the previous proposition are done as part of the construction of \mathcal{P}_R . Thus T_1 differs from $M - \mathcal{P}_R$ only by the slicing adjustments.

We can now construct a $(1, \varepsilon(R))$ -nearly straight train-track in N corresponding to the image of T_1 . For each branch of T_1 choose one horizontal leaf and let its image in N be a branch of the train-track, which we will call τ_1 . We already know from §3.3 that these branches are of length at least $2 - \varepsilon(R)$, and have curvature bounded above by $\varepsilon(R)$. The next lemma insures that all the horizontal arcs from a given branch are $\varepsilon(R)$ -nearly tangent. This implies immediately that τ_1 is a $(1, \varepsilon(R))$ -nearly straight track, and that our choice of leaves for branches was not important.

Lemma 6.2 (narrow rectangles). *Let $\beta \subset T_1$ be a vertical arc which projects to a point in X_1 . Then $l_N(f(\beta)) \leq \varepsilon(R)$.*

Proof. By (3.3), the length of $f(\beta)$ can be computed by

$$l_N(f(\beta)) = \int_{\beta} \sqrt{2(\cosh \mathcal{E} - 1)} ds,$$

where ds is arclength along β in the $|\Phi|$ metric. By our estimates on \mathcal{E} this is

$$l_N(f(\beta)) \leq \int_{\beta} \varepsilon_1(R + d(p, \mathcal{P}_R)) ds(p),$$

where ε_1 is an inversely exponential function.

Let $A_n = \mathcal{N}_n(\mathcal{P}_R) - \mathcal{N}_{n-1}(\mathcal{P}_R)$, where $n = 1, 2, 3, \dots$, and \mathcal{N}_r denotes an r -neighborhood in the $|\Phi|$ metric. Lemma 4.1 and Theorem 5.1 (4) give

$$l(\partial \mathcal{N}_{n-1}(\mathcal{P}_R)) \leq C_1 R + C_2(n - 1),$$

where the C_i depend, as usual, only on $\chi(M)$. Lemma 4.1 applied to $\mathcal{N}_n(\mathcal{P}_R) = \mathcal{N}_1(\mathcal{N}_{n-1}(\mathcal{P}_R))$ also yields

$$\text{Area}(\mathcal{N}_n(\mathcal{P}_R)) \leq \text{Area}(\mathcal{N}_{n-1}(\mathcal{P}_R)) + l(\partial \mathcal{N}_{n-1}(\mathcal{P}_R)) + C_3$$

or

$$\text{Area}(A_n) \leq C_4(R + n).$$

If we also define $A_0 = \mathcal{P}_R \cap \mathcal{N}_1(\partial \mathcal{P}_R)$, we easily obtain the same inequality for $n = 0$ (provided R is not too small).

Since β projects to a point in a branch of X_1 of length at least 1, β must be contained in a band B of width 1, which is embedded in M . For every point $x \in \beta \cap A_n$ ($n \geq 1$) it is immediate from the definitions that $\mathcal{N}_1(x) \subseteq A_{n+1} \cup A_n \cup A_{n-1}$. Thus,

$$\begin{aligned} L_{|\Phi|}(\beta \cap A_n) &\leq \text{Area}(B \cap (A_{n+1} \cup A_n \cup A_{n-1})) \\ &\leq 3C_5(R + n). \end{aligned}$$

We may therefore write

$$\begin{aligned}
 l_N(f(\beta)) &\leq \sum_{n=1}^{\infty} \int_{\beta \cap A_n} \varepsilon_1(R + d(p, \mathcal{P}_R)) ds(p) \\
 &\leq \sum_{n=1}^{\infty} 3C_5(R + n)\varepsilon_1(R + n - 1 \leq \varepsilon_2(R),
 \end{aligned}$$

where ε_2 is another inversely exponential function.

6.2. Enlarging the train-track. Although τ_1 is a perfectly good train-track, Φ_h itself is not in general carried on it. Missing are, first of all, branches to carry the leaves inside \mathcal{P}_R . These account for a small proportion of the total mass of Φ , but they are still not to be ignored in view of the discussion at the beginning of the section. They will be dealt with in the next lemma. The real problem, however, is with the flat cylinder components of $M - \mathcal{P}_R$ which we have ignored so far. These may contain much or most of the mass of Φ . Indeed, there are cases where $M - \mathcal{P}_R$ consists entirely of flat cylinders, and in those cases we have done nothing so far.

These may contain much or most of the mass of Φ . Indeed, there are cases where $M - \mathcal{P}_R$ consists entirely of flat cylinders, and in those cases we have done nothing so far.

The leaves of Φ_h in these cylinders are already $(1, \varepsilon(R))$ -nearly straight, but they are disconnected from τ_1 . Our task will be to join them to τ_1 with geodesic arcs that are homotopic to leaves of Φ_h and meet the cylinder arcs and τ_1 with $\varepsilon(R)$ -small angles at their endpoints.

In preparation, we first take care of the easiest part, the leaves of Φ_h that enter and exit \mathcal{P}_R through a vertical arc of one of the polygonal boundaries.

Lemma 6.3 (Straighten by convexity). *if α is an arc of Φ_h contained in \mathcal{P}_R with endpoints on vertical segments of polygonal boundaries of \mathcal{P}_R , then the geodesic representative $f(\alpha)^*$ of $f(\alpha)$ in N meets $f(\alpha)$ at its endpoints at angles bounded by $\varepsilon(R)$.*

Proof. Let \mathcal{F} be the component of \mathcal{P}_R containing α . Lift \mathcal{F} homeomorphically to the cover $N_{\mathcal{F}}$ of N corresponding to $\pi_1(\mathcal{F}) \subset \pi_1(N)$, and let \mathcal{F}^* denote the embedded subsurface of $N_{\mathcal{F}}$ obtained by “straightening” \mathcal{F} as follows: replace each straight cylindrical boundary component by its geodesic representative (in N ’s hyperbolic metric), and for each polygonal boundary component replace each horizontal and vertical arc by a geodesic (very short ones, for the vertical arcs), adjusting the corners a small $(\varepsilon(R))$ amount if necessary, so that the resulting polygon

is embedded. The resulting surface is convex, and each end of $f(\alpha)$ is within $\varepsilon(R)$ of a short vertical boundary arc. This arc adjoins two long horizontal boundary arcs that are, therefore, $\varepsilon(R)$ -nearly tangent. The end of $f(\alpha)$, being $\varepsilon(R)$ -nearly straight and contained in \mathcal{F} , is $\varepsilon(R)$ -nearly tangent to these arcs, and therefore to the corresponding geodesic arcs of $\partial\mathcal{F}^*$.

Since \mathcal{F}^* is convex, $f(\alpha)^*$ is contained within it. Thus, each end of $f(\alpha)^*$ is trapped between the two geodesics of $\partial\mathcal{F}^*$ incident to the same vertical arc, and by elementary hyperbolic geometry, must be $\varepsilon(R)$ -nearly tangent to them and therefore to the corresponding end of $f(\alpha)$.

Adding cylinder branches. The next lemma allows us to attach leaves from the flat cylinders as new branches of the train-track.

Lemma 6.4 (Guide wires for flat cylinders). *There is a choice of $R_0 > 0$ and of constants in Theorem 5.1 (polygonal region) which depends only on $\chi(M)$ such that the following holds: If $R > R_0$, the train-track τ_1 can be enlarged to a train-track τ_n which contains a branch (possibly closed) for every flat cylinder \mathcal{F} of $M - \mathcal{P}_R$, which is isotopic (mod endpoints) to a leaf of Φ_h passing through $\mathcal{F} \cdot \tau_n$ is realized in N as a $(1, \varepsilon(R))$ -nearly straight broken train-track, and images of the leaves of Φ_h in \mathcal{F} are $\varepsilon(R)$ -near to the corresponding branch of τ_n .*

Proof. This lemma is the most technical step in the proof of Theorem 7.1, and we begin therefore with an outline of the strategy of the proof.

Each flat cylinder \mathcal{F} contains leaves whose images in N are already $\varepsilon(R)$ -nearly straight, spiraling tightly around a geodesic in N or cutting through a long thin part. We will join each end of such a leaf to the rest of the train-track with a leaf segment of Φ_h which passes through \mathcal{P}_R . The trouble, of course, is that such a leaf is not nearly straight, and when we replace it with its geodesic representative we may get a branch that does not meet the existing train-track in small angles, or in which the portion in \mathcal{F} has been “unwound.”

The strategy will be to control the construction in such a way that the added arc in \mathcal{P}_R is short compared to the nearly straight arcs in \mathcal{F} . The length of the arcs in \mathcal{F} is bounded from below by a judicious choice of the constants c_k in the construction of \mathcal{P}_R . The length of the arc in \mathcal{P}_R is bounded from above via an estimate of the total energy of f in \mathcal{P}_R . Lemma 6.5 provides the hyperbolic geometry that translates these estimates into control of the geodesic representatives of the chosen leaves.

For any flat cylinder \mathcal{F} of $M - \mathcal{P}_R$, let $W(\mathcal{F})$ denote its circumference. The leaves of Φ_h spiral around F making some angle ζ with any geodesic core of the cylinder (if $\zeta = 0$ the leaves are closed, and this is the

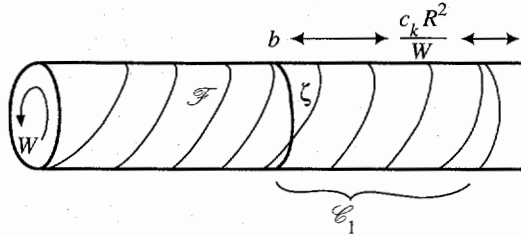


FIGURE 15

simplest case to deal with). The vertical distance between successive turns of any leaf is $V(\mathcal{F}) = W \sin \zeta$, which gives the total transverse measure of the leaves of Φ_h entering \mathcal{F} (Figure 15).

We now order the cylinders according to decreasing values of V and successively adjoin each one to the track-train. Let τ_i be the train-track constructed at the i th stage, and let \mathcal{F} be the next cylinder to be adjoined. For each boundary component b of \mathcal{F} we will find an arc α of Φ_h in \mathcal{P}_R that connects b to the rest of $M - \mathcal{P}_R$, and whose image is $\varepsilon(R)$ -nearly tangent to its geodesic representative at the endpoints. These arcs, together with the image of a leaf of Φ_h in \mathcal{F} , will form a $(1, \varepsilon(R))$ -nearly straight broken path which we add to τ_i to make a new train-track τ_{i+1} .

Let us get an easy special case out of the way first. If $\zeta(\mathcal{F}) = 0$ then, by the estimates of §3.3 and the construction of \mathcal{P}_R , the images of closed trajectories of Φ_h in \mathcal{F} are $\varepsilon(R)$ -nearly straight and $\varepsilon(R)$ -near to each other. Thus any of them will do as the added (closed) branch to τ_i . The $\varepsilon(R)$ -nearby geodesic curve in this homotopy class will be the added leaf to λ_i .

If $\zeta(\mathcal{F}) \neq 0$ there is more work to do. Each end of each leaf of Φ_h exits through $\partial\mathcal{F}$ into \mathcal{P}_R . We claim that all but finitely many of these ends must also eventually leave \mathcal{P}_R . Call a leaf of Φ_h in \mathcal{P}_R *critical* if it meets a critical point of Φ_h before exiting \mathcal{P}_R . Then there are only finitely many such leaves, and therefore only finitely many intersecting $\partial\mathcal{F}$. If a leaf α exiting \mathcal{F} is not critical, it must be in some band of width $\delta > 0$ of noncritical leaves, which must be embedded in \mathcal{P}_R . Since \mathcal{P}_R has finite $|\Phi|$ -area, this band, and therefore α , must have finite length.

Fix a boundary component b of \mathcal{F} and let \mathcal{F}_k denote the maximal flat cylinder containing \mathcal{F} (with index k determined as in Theorem 5.1). The portion of $\mathcal{F}_k - \mathcal{F}$ adjacent to b has length at least $r_W + R + c_k R^2 / W$, so we let \mathcal{E}_1 denote the subcylinder of length $c_k R^2 / W$ adjacent to b . For any arc α of Φ_h exiting \mathcal{F} through b , let α_1 denote its intersection with

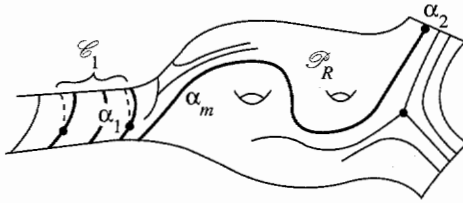


FIGURE 16

\mathcal{E}_1 . If the opposite end of α is in another as yet unadjoined cylinder $\mathcal{F}' \subset \mathcal{F}_{k'}$, let \mathcal{E}_2 denote the corresponding cylinder of length $c_{k'}R^2/W(\mathcal{F}')$ and α_2 the intersection of α with \mathcal{E}_2 (conceivably $\mathcal{F}' = \mathcal{F}_k$ but then by index considerations α is not deformable rel endpoints into \mathcal{F} , so in some cover the cylinders are different). If α does not terminate in such a cylinder let α_2 denote just the far endpoint of α . Let α_m denote the closure of $\alpha - \alpha_1 - \alpha_2$ (see Figure 16). We note that $f(\alpha)$ is $\varepsilon(R)$ -nearly straight at $f(\alpha_1)$ and $f(\alpha_2)$.

Let \mathcal{P}'_R denote \mathcal{P}_R minus the portions of flat cylinders bounded by straight components of $\partial\mathcal{P}_R$. Then $\|\Phi\|_{\mathcal{P}'_R} \leq (K_2 + \sum_{j < k, k'} 2c_j)R^2$. The set of leaves α exiting b has $|\Phi|$ -height $V(\mathcal{F})$. By the averaging argument of Proposition 3.1 (energy lower bound), the average length of the images in N of the middle arcs α_m is bounded by

$$\begin{aligned} \text{Avg}_{\mathcal{E}_\alpha} l_N(\alpha_m) &\leq \frac{1}{V} [\|\Phi\|_{\mathcal{P}'_R} \mathcal{E}(\mathcal{P}'_R)]^{1/2} \\ &\leq \frac{1}{V} [\|\Phi\|_{\mathcal{P}'_R} (2\pi|\chi(M)| + 2\|\Phi\|_{\mathcal{P}'_R})]^{1/2} \end{aligned}$$

by inequality (3.5). (Note that we are abbreviating $l_N(f(\alpha))$ to $l_N(\alpha)$, and so on.) If we choose R_0 large enough and $\mathcal{E}(M)$ is sufficiently large, this gives

$$\text{Avg}_{\mathcal{E}_\alpha} l_N(\alpha_m) \leq \frac{(K_2 + \sum_{j < k, k'} 2c_j)R^2}{V}.$$

We conclude that there must be at least one arc, which we continue to label α , for which $l_N(\alpha_m)$ is no more than this average. On the other hand, $l_{|\Phi|}(\alpha_1)$ is $1/\sin \zeta$ times the length of \mathcal{E}_1 . By the estimates on \mathcal{E} in \mathcal{E}_1 , we obtain

$$l_N(\alpha_1) \geq (2 - \varepsilon(R))l_{|\Phi|}(\alpha_1) \geq \frac{(2 - \varepsilon(R))c_k R^2}{V}.$$

Choosing c_k large enough relative to $K_2 + \sum_{j < k, k'} 2c_j$, we can have $l_N(\alpha_1)/l_N(\alpha_m) \leq Q_1$ for any constant Q_1 which we will determine shortly.

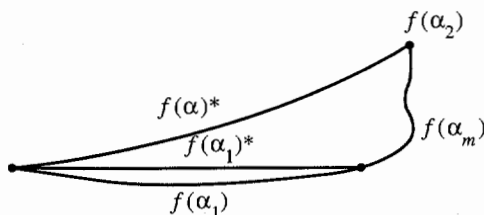


FIGURE 17. GOOD TRIANGLE

If α_2 is also in an unadjoined cylinder \mathcal{F}' , we can repeat the above computation to obtain

$$l_N(\alpha_2) \geq \frac{(2 - \varepsilon(R))c_k'R^2}{V(\mathcal{F}')} \geq \frac{(2 - \varepsilon(R))c_k'R^2}{V(\mathcal{F})},$$

since, by the ordering of the construction, $V(\mathcal{F}') \leq V(\mathcal{F})$. Thus we can again have a ratio $l_N(\alpha_2)/l_N(\alpha_m) \geq Q_1$. Additionally, since $V \leq W \leq K_1R$, we can choose c_k (or c_k') so that $l_N(\alpha_i) \geq Q_2R$ for any constant Q_2 .

With these estimates on the lengths of $f(\alpha_i)$ and $f(\alpha_m)$, we proceed to show that the geodesic representative $f(\alpha^*)$ meets $f(\alpha)$ with $\varepsilon(R)$ -small angles at the endpoints.

Consider first the case where α_2 is a single point, i.e., the far end of α exits \mathcal{P}_R either at a polygonal boundary or through a cylinder that has already been adjoined to the train-track.

We have a configuration as in Figure 17 in the universal cover. If we choose the ratio Q_1 between $l_N(\alpha_1)$ and $l_N(\alpha_m)$ large enough (ahead of time), hyperbolic trigonometry assures us that $f(\alpha)^*$ and $f(\alpha)$ are $\varepsilon(R)$ -nearly tangent at the α_1 endpoint.

To prove this for the other endpoint, we can easily adapt the argument of Lemma 6.3 (straighten by convexity). Take first the case where α exits \mathcal{P}_R through a vertical component of \mathcal{P}_R . Construct \mathcal{F}^* as in that lemma, and note that the α_1 endpoint of $f(\alpha)$ is either contained in \mathcal{F}^* or, if it is not then $f(\alpha)$ enters \mathcal{F}^* through the geodesic boundary corresponding to \mathcal{F} , and thus has intersection number 1 with it. In either case, the α_2 end of $f(\alpha)^*$ must be contained in \mathcal{F}^* and by the arguments of the lemma, must be $\varepsilon(R)$ -nearly tangent to $f(\alpha)$ at $f(\alpha_2)$.

Finally, there is the case where α exits through a cylinder \mathcal{F}'' that has already been adjoined. The same argument works, but first we have to cut \mathcal{P}_R along the arc β which was added to the train-track and contains one of the leaves of Φ_h through the cylinder \mathcal{F}'' . A convex surface can

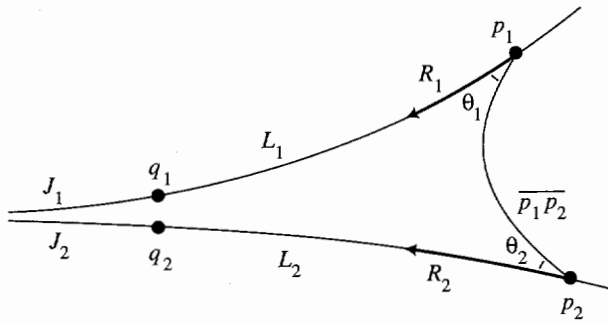


FIGURE 18

be constructed again with two copies of the geodesic representative of β in its boundary, and again $f(\alpha)^*$ is trapped between these two curves. The distance between them is bounded by $\varepsilon(R)$, because in \mathcal{F}'' there are vertical arcs with both endpoints on a horizontal leaf, whose $|\Phi|$ -length is bounded by K_1R and which are therefore contracted in the N metric to arcs of length $\varepsilon(R)$. Thus again $f(\alpha)^*$ is constrained to be $\varepsilon(R)$ -nearly tangent to $f(\alpha)$ at $f(\alpha_2)$.

It remains to deal with the case where both α_1 and α_2 lie in unadjoined cylinders. We first require some observations about curves in hyperbolic space and their geodesic representatives.

Let L_1 and L_2 be geodesics in \mathbf{H}^2 (or \mathbf{H}^3 for that matter), and fix $s_0 > 0$. Define the *juncture* (or s_0 -*juncture*) of L_1 and L_2 to be $J = J_1 \cup J_2$, where

$$J_i = \{x \in L_i : d(x, L_j) \leq \max(s_0, d(L_1, L_2))\}$$

for $(i, j) = (1, 2)$ or $(2, 1)$, so that J_i is a line segment or a ray when $d(L_1, L_2) < s_0$, and a single point when $d(L_1, L_2) \geq s_0$.

Let R_1 and R_2 denote rays on L_1 and L_2 with basepoints p_1 and p_2 , respectively. Let θ_i denote the angle made by the ray R_i and the geodesic $\overline{p_1 p_2}$ as in Figure 18. Suppose that R_1 , say, contains J_1 . Then elementary hyperbolic trigonometry shows that

$$(6.1) \quad \theta_1 \leq \varepsilon(d(p_1, J_1)),$$

where the function ε depends on the choice of s_0 . Let q_1 be the endpoint of J_1 closest to p_1 , and q_2 the endpoint of J_2 closest to q_1 . If, now, $q_2 \in R_2$ then it is also easy to see that

$$(6.2) \quad \theta_2 \leq \varepsilon(d(p_2, q_2)).$$

Now let β be a path in \mathbf{H}^2 which is composed of a sequence of three segments β_1, β_m , and β_2 such that β_1 and β_2 are geodesics. Let L_i ,

for $i = 1, 2$, be the infinite geodesic containing β_i . Denote by b_i the endpoint common to β_i and β_m , and by p_i the other endpoint of β_i . Let R_i be the ray of L_i based at p_i and containing β_i . Then θ_i as defined above is also the angle at p_i between β and its geodesic representative β^* . We have:

Lemma 6.5 (Straighten corners). *There is an inversely exponential function ε and a constant A , depending only on s_0 , such that the following hold:*

(a) *If $l(\beta_i) > l(\beta_m) + l(J_i) + A$, ($i = 1, 2$), then*

$$\theta_i \leq \varepsilon(l(\beta_i) - l(\beta_m) - l(J_i) - A).$$

(b) *If $l(\beta_2) > l(\beta_m) + A$ and $J_1 \subset R_1 - \beta_1$, then*

$$\theta_1 \leq \varepsilon(l(\beta_1)), \quad \theta_2 \leq \varepsilon(l(\beta_2) - l(\beta_m) - A).$$

Proof (Sketch). We first observe that there is a constant $A > 0$ such that

$$d(b_i, J_i) \leq l(\beta_m) + A.$$

This follows from the fact that $l(\beta_m) \geq d(b_i, L_{3-i})$, and from either elementary hyperbolic trigonometry or a compactness argument on the space of configurations.

Condition (a) on $l(\beta_i)$ insures, therefore, that R_i contains J_i , and the $d(p_i, J_i) \geq l(\beta_i) - l(\beta_m) - l(J_i) - A$. An application of (6.1) concludes the proof of case (a).

If the conditions of case (b) hold, then R_1 already contains J_1 and $d(p_1, J_1) \geq l(\beta_1)$. Define q_1 and q_2 as before. If $d(L_1, L_2) \geq s_0$, then, again, $d(b_2, q_2) \leq l(\beta_m) + A$. If $d(L_1, L_2) < s_0$ and $q_2 \notin R_2 - \beta_2$, then it is easy to see that $d(b_2, q_2) \leq d(b_2, q_1) + d(q_1, q_2) \leq d(b_2, b_1) + s_0 \leq l(\beta_m)k + s_0$. Thus in any case $q_2 \in R_2$ and $d(p_2, q_2) \leq l(\beta_2) - l(\beta_m) - A$. The bounds on θ_i follow from (6.1) and (6.2). *q.e.d.*

We return now to the proof of Lemma 6.4. We wish to apply Lemma 6.5 to the path β obtained from $f(\alpha)$ by replacing $f(\alpha_1)$ and $f(\alpha_2)$ with their nearby geodesic representatives. In order to do this we require, in addition to the estimates on $l_N(\alpha_i)/l_N(\alpha_m)$, some bounds on the junctures J_i . Here we need to make strong use of properties of finite-area hyperbolic surfaces.

Let \mathcal{E} denote $\mathcal{F} \cup \mathcal{E}_1$ or $\mathcal{F}' \cup \mathcal{E}_2$. A vertical arc in \mathcal{E} with both endpoints on a given leaf of Φ_h has $|\Phi|$ -length $V \leq W \leq K_1 R$ and therefore its image has N -length $v \leq \varepsilon(R)$. The $|\Phi|$ -length of the segment of Φ_h between the endpoints of the vertical arc is $W \cos \eta$, so its N -length is $l \leq 2W \cos \zeta$ by (3.3).

Let $\epsilon > 0$ be small enough to give a thick-thin decomposition for hyperbolic surfaces. Further, assume ϵ is small enough so that any simple geodesic in N which enters a 1-neighborhood of the ϵ -thin part is either a core of some component of the thin part or intersects the core of some component. (This choice of ϵ does not depend on N .) We separate the situation into two possibilities:

(a) $l \leq \epsilon/2$. Then assuming R is large enough depending only on ϵ so that $v \leq \epsilon/2$, the image of the entire cylinder lies in the ϵ -thin part of N corresponding to its core curve.

(b) $l > \epsilon/2$. Then the horizontal segment makes an $(\epsilon/2, \epsilon(R))$ -nearly straight broken circuit. We conclude that for large enough R , again depending only on ϵ , the leaves of Φ_h in \mathcal{E} are $\epsilon(R)$ -near to the geodesic γ in N corresponding to the core curve of \mathcal{E} .

Note that in case (b), $l_N(\gamma) \leq 2W(1 + \epsilon(R)) \leq K_4R$, an estimate we can easily obtain from (3.3).

Lift α , \mathcal{E}_1 , and \mathcal{E}_2 to $\tilde{\alpha} = \tilde{\alpha}_1 \cup \tilde{\alpha}_m \cup \tilde{\alpha}_2$, $\tilde{\mathcal{E}}_1$, and $\tilde{\mathcal{E}}_2$ in $\tilde{N} = \mathbf{H}^2$ (so that $\tilde{\alpha}_i \subset \tilde{\mathcal{E}}_i$). If both \mathcal{E}_1 and \mathcal{E}_2 fall under case (b), let $L_i = \tilde{\gamma}_i$ be the lift of γ_i corresponding to $\tilde{\mathcal{E}}_i$, and let β be obtained from $\tilde{\alpha}$ by projecting $\tilde{\alpha}_i$ to L_i (recall from the above that this projection moves points a distance $\epsilon(R)$ and distorts lengths by a factor of $1 - \epsilon(R)$). Defining J_i as in Lemma 6.5, we observe that there is a universal choice of s_0 such that $l(J_i) \leq \min(l_N(\gamma_1), l_N(\gamma_2))$; this follows directly from the fact that all the translates of $\tilde{\gamma}_2$, say, by the translation of $l_N(\gamma_1)$ along $\tilde{\gamma}_1$ are disjoint. Hence $l(J_i) \leq K_4R$. Since $l_N(\alpha_i) \geq Q_2R$ and also $l_N(\alpha_i) \geq Q_1l_N(\alpha_m)$, we can choose, say, $Q_1 = 2$, $Q_2 = 4K_4$, and $R > R_0 = 2A/K_4$ and apply Lemma 6.5 to conclude that

$$\theta_i \leq \epsilon(K_4R - A) = \epsilon(R),$$

where θ_i are the angles between β and β^* . The same estimate then holds for α and α^* .

Consider next the case that \mathcal{E}_1 (say) is in case (a) and \mathcal{E}_2 is in case (b). Let T_1 be the lift of the corresponding thin part which contains $\tilde{\mathcal{E}}_1$, and let L_1 denote the infinite geodesic containing the endpoints of $\tilde{\alpha}_1$. Define β as above, and recall the rest of the notation of Lemma 6.5. If $J_1 \subset R_1 - \beta_1$, given a choice of constants as in the previous paragraph we can use part (a) of that lemma to conclude that $\theta_i \leq \epsilon(R)$. If J_1 is not contained in $R_1 - \beta_1$, since $d(b_1, J_1) \leq l_N(\beta_m) + A$ and $l_N(\beta_1) \geq Q_1l_N(\beta_m)$ we conclude that J_1 must meet β_1 provided we choose $Q_1 > 1$. By the choice of ϵ , $\tilde{\gamma}_2$ must be outside a 1-neighborhood of T_1 . Thus, provided

$s_0 < 1$, we must have $l(J_1) = 0$. In this case we can use part (b) of Lemma 6.5.

The last possibility, that both \mathcal{E}_i are in case (b), is handled in much the same way. If $J_i \subset R_i - \beta_i$ for either $i = 1$ or 2 , we may apply Lemma 6.5(b). If not, then we can again argue that the junctures meet β_1 and β_2 , so that $l(J_i) = 0$ and Lemma 6.5(a) applies. q.e.d.

We summarize the results of this section in

Lemma 6.6 (Final train-track). *The original train-track τ_1 may be enlarged to a $(1, \varepsilon(R))$ -nearly straight train-track τ which carries $f(\Phi_h)$, and which contains branches $\varepsilon(R)$ -nearly tangent to all the images of leaves in $M - \mathcal{P}_R$.*

Proof. In Lemma 6.3 we showed how leaves in \mathcal{P}_R with both endpoints on a vertical boundary may be added to τ_1 . In Lemma 6.4 we added branches that are nearly tangent to the flat cylinder leaves and contain segments homotopic to segments of leaves in \mathcal{P}_R .

The only branches left to add to the train-track are those that carry leaves contained in \mathcal{P}_R with one or both of its endpoints on a straight cylinder boundary, and branches to carry leaves that never leave \mathcal{P}_R at all. We note that, as was shown in the course of proving Lemma 6.4, a nonsingular leaf segment contained in \mathcal{P}_R must either be doubly infinite, or compact.

Now that we have added the cylinders to the train-track, the first type of branch can be added using the same sort of argument as in Lemma 6.3 (straighten by convexity)—the role of the nearly-convex surface \mathcal{F} is now played by the complement of τ_n .

Finally, the leaves of Φ_h that are entirely contained in \mathcal{P}_R form a subfoliation μ_R . The corresponding sublamination may be approximated in N by a nearly straight train-track which lies in a supporting subsurface for the sublamination. Again by the near-convexity of the complement of τ_n , this train-track may be made to be disjoint from τ_n and can therefore be adjoined as a separate component; we note that this part of the train-track is inessential to the rest of our arguments, since it anyway only carries leaves in \mathcal{P}_R , which have bounded total length and therefore have no effect on the estimates that follow.

7. The main theorems

Using the train-track constructed in Lemma 6.6 we can give a final geometric statement, and prove the length-energy inequality. Let $f(\Phi_h)^*$ denote the geodesic representative of $f(\Phi_h)$ in N .

Theorem 7.1 (Map foliation near lamination). *Let $f: M \rightarrow N$ be a diffeomorphic harmonic map between closed hyperbolic surfaces. There are choices of constants $s_1 > 0$ and $c_1, \dots, c_{3|\chi(M)|/2} > 0$ for the construction of \mathcal{P}_R and an $R_0 > 0$, all depending only on $\chi(M)$, such that in the complement of \mathcal{P}_{R_0} there is a map π from the leaves of Φ_h to the lamination $f(\Phi_h)^*$ that factors through f , and is a local diffeomorphism on each leaf of Φ_h , mapping it to the corresponding geodesic representative of its image. For any point p on a leaf in $M - \mathcal{P}_{R_0}$,*

$$d_N(f(p), \pi(p)) < \varepsilon(d_{|\Phi|}(p, \mathcal{P}_{R_0})),$$

and the derivative of π along leaves, with respect to the $|\Phi|$ metric, satisfies

$$||d\pi| - 2| \leq \varepsilon(d_{|\Phi|}(p, \mathcal{P}_{R_0})),$$

where the factor of 2 comes from the derivative of f , which is approximately 2 along the horizontal leaves. The constants determining the inversely exponential function ε also depend only on $\chi(M)$.

Note that the theorem is trivial for maps of energy less than $K_2 R_0^2$ because then $\mathcal{P}_{R_0} = M$.

Proof. The idea is to lift each leaf to the universal cover and take the nearest-point projection to its geodesic representative.

Choose constants as prescribed by the proofs in the previous sections. Let μ be a leaf of Φ_h , which is not entirely contained in \mathcal{P}_{R_0} , and let p be a point on μ , which is outside \mathcal{P}_R for $R > R_0$. By the results of the previous section we can employ the train-track τ corresponding to $M - \mathcal{P}_R$ to alter $f(\mu)$ only at the images of $\mu \cap \mathcal{P}_R$ so that the resulting broken path $f(\mu)'$ is $(1, \varepsilon(R))$ -straight. Lifting to the universal cover, we conclude that the closest-point projection from $\tilde{f}(\mu)'$ to the geodesic $\tilde{f}(\mu)^*$ moves points by at most $\varepsilon(R)$, has derivative within $\varepsilon(R)$ of 1, and is diffeomorphic at the interior points of segments (and in particular at $\tilde{f}(p)$).

This projection is equivariant with respect to $\pi_1(M)$, so we can project it back to N and compose with f to get π .

Theorem 7.2 (Energy is length-ratio). *There is a constant C depending only on $\chi(M)$ such that in the situation described by the previous theorem,*

$$\frac{1}{2} \frac{l_N^2(\Phi_h)}{E_M(\Phi_h)} \leq \mathcal{E}(f) \leq \frac{1}{2} \frac{l_N^2(\Phi_h)}{E_M(\Phi_h)} + C.$$

Note that we abbreviate $l_N(f(\Phi_h)^*)$ to $l_N(\Phi_h)$.

In view of Lemma 3.1 (energy lower bound), this theorem implies that Φ_h comes within C of maximizing the ratio $l_N^2/2E_M$ over all of $\mathcal{MF}(S)$.

Proof. We use Theorem 7.1 (map foliation near lamination) to obtain a lower bound on $l_N(f(\Phi_h)^*)$. First we restrict to the complement of \mathcal{P}_{R_0} :

$$l_N(f(\Phi_h)^*) \geq l_N(\pi(\Phi_h \cap (M - \mathcal{P}_{R_0}))).$$

By the bounds on $|d\pi|$ and recalling that transverse measure on Φ_h is just vertical length in the $|\Phi|$ metric, we have

$$\begin{aligned} l_N(f(\Phi_h)^*) &\geq \int_{M - \mathcal{P}_{R_0}} 2 - \varepsilon(d(\cdot, \mathcal{P}_{R_0})) dA(|\Phi|) \\ &\geq 2\|\Phi\|_{M - \mathcal{P}_{R_0}} - \int_0^\infty \varepsilon(r) l_{|\Phi|}(\partial\mathcal{N}_r(\mathcal{P}_{R_0})) dr. \end{aligned}$$

We recall (Lemma 4.1 (size of R -neighborhood)) that $l_{|\Phi|}(\partial\mathcal{N}_r(\mathcal{P}_{R_0})) \leq K_1 R_0 + N_1 r$ where N_1 depends on $\chi(M)$. Also, $\|\Phi\|_{\mathcal{P}_{R_0}}$ is bounded as in Theorem 5.1 (polygonal region), so that

$$\begin{aligned} l_N(f(\Phi_h)^*) &\geq 2\|\Phi\|_{M - \mathcal{P}_{R_0}} - \int_0^\infty \varepsilon(r)(K_1 R_0 + N_1 r) dr \\ &\geq 2\|\Phi\|_M - C_1, \end{aligned}$$

where C_1 depends on the other constants. Since $E_M(\Phi_h) = \|\Phi\|_M$, we have

$$\frac{1}{2} \frac{l_N^2(f(\Phi_h)^*)}{E_M(\Phi_h)} \geq 2\|\Phi\| - C_1 + \frac{1}{2} \frac{C_1^2}{\|\Phi\|} \geq \mathcal{E}(f) - C_1 - 2\pi|\chi(M)|$$

by (3.5), and hence the proof is complete.

8. Limits and compactifications

Consider now the situation where $N = (S, \rho)$ is fixed and the conformal structure σ of $M = (S, \sigma)$ is allowed to vary. There is a map

$$\Phi_h: \mathcal{T}(S) \rightarrow \mathcal{MF}(S)$$

which associates to each choice of M the measured horizontal foliation of the Hopf differential of the harmonic map $f: M \rightarrow N$, homotopic to the identity. This map is continuous by general considerations; see §3.

It is also proper. We first show that the energy $\mathcal{E}(f)$ is a proper function of σ . This was shown in [26] in a broader context, but we give here a short proof using our terminology. Let $M_i = (S, \sigma_i)$ be a sequence of Riemann surfaces leaving every compact set of Teichmüller space. The quasiconformal distortion between M_i and N then goes to

infinity, and therefore the maximal ratio $E_\rho(\gamma)/E_{\sigma_i}(\gamma)$ over all foliations $\gamma \in \mathcal{MF}(S)$ goes to infinity as well. Since (see (8.1) later in this section) $l_\rho^2(\gamma)/E_\rho(\gamma)$ is bounded above and below by positive constants, it follows that $\max_\gamma (l_\rho^2(\gamma)/E_{\sigma_i}(\gamma)) \rightarrow \infty$. By (3.1) $\mathcal{E}(f_{\sigma_i}) \rightarrow \infty$, so the energy is a proper function. Now by Theorem 7.2 (energy is length-ratio) and inequality (3.5) we have $l_\rho(\mathcal{L}(\Phi_h^i)) \rightarrow \infty$, and in particular Φ_h^i must leave every compact set in $\mathcal{MF}(S)$.

$\mathcal{MF}(S)$ is naturally compactified by the sphere $\mathcal{PMF}(S)$, and $\mathcal{T}(S)$ can be compactified by $\mathcal{PMF}(S)$ in several different ways, so it is natural to ask if and how the map Φ_h extends to these compactifications.

We note here that a related problem—in which M is fixed and N varies—has been analyzed quite carefully by Wolf ([31],[32]). In that case there is a proper diffeomorphism $\Phi: \mathcal{T}(S) \rightarrow QD(M)$ where we recall $QD(M) \cong \mathcal{MF}(S)$, which extends at infinity to give the Thurston compactification of $\mathcal{T}(S)$.

Such a pleasant eventuality is less likely in our situation, for the following general reason. Thurston’s compactification is naturally related to the asymptotic behavior of *hyperbolic* invariants of a surface, namely lengths of curves, and indeed, the harmonic map “detects” these properties in the range N . On the other hand, the important invariants for the domain of a harmonic map are *conformal*, namely extremal lengths of curves, as we have seen in the previous sections. The following example illustrates the difference between the two at infinity.

Consider a base (hyperbolic) surface M_0 on which two distinct disjoint simple closed geodesics γ_1 and γ_2 have been fixed, and suppose their hyperbolic lengths are 1. Let $\{M_n\}$ denote a sequence of hyperbolic structures on S which are obtained from the hyperbolic metric on M_0 by (a) shrinking the length of γ_1 to ϵ_n , and (b) Dehn-twisting n times around γ_2 (see Figure 19).

We note a few facts without much proof, since they are only intended to give a plausibility argument. For a fixed homotopy class α in S , it is not hard to see that

$$l_{M_n}(\alpha) \approx c_1 + c_2 |\log \epsilon_n| i(\alpha, \gamma_1) + c_3 n i(\alpha, \gamma_2).$$

(Recall that $i(\alpha, \gamma)$ is the minimal intersection number of curves homotopic to α and γ .) The second term counts the number of times that α must cross the collar around γ_1 , and the third term counts the number of times α must wind around γ_2 , n for each intersection with γ_2 . Choose $\epsilon_n = 1/n^2$. The third term eventually dominates, so in Thurston’s com-

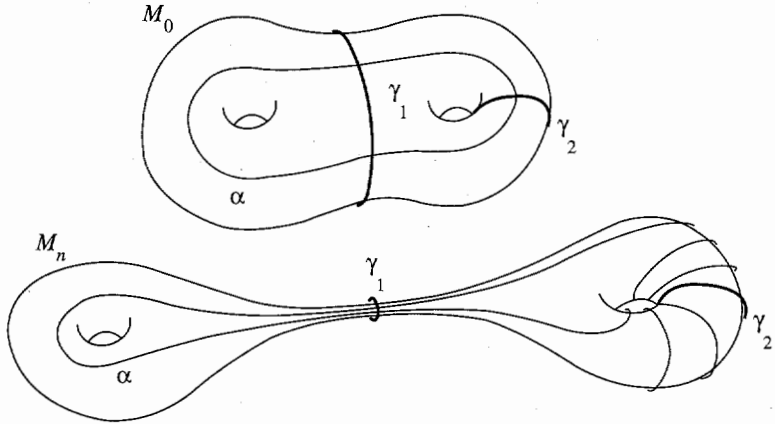


FIGURE 19. IN THE HYPERBOLIC STRUCTURE ON M_n , n DEHN TWISTS HAVE BEEN PERFORMED AROUND γ_2 WHILE γ_1 HAS BEEN SHRUNK TO LENGTH ϵ_n

pactification, which is based on modeling of projective length structure by intersection number, the limit of the sequence is the curve γ_2 . On the other hand, the extremal lengths of γ_1 and γ_2 are

$$E_{M_n}(\gamma_1) \approx \frac{\epsilon_n}{2\pi} = \frac{1}{2\pi n^2} \quad \text{and} \quad E_{M_n}(\gamma_2) \geq \frac{c_4}{n},$$

so in view of Theorem 7.2 (energy is length-ratio) we expect the limiting foliation Φ_h to be γ_1 rather than γ_2 .

We will fare a little better if we consider instead the Teichmüller compactification, which is obtained from the ray structure imposed on $\mathcal{T}(S)$ by the Teichmüller rays $\{M_{K,\Psi}\}$ (see §2). We will examine next the asymptotic behavior of harmonic maps as M goes to infinity along such a ray, but we emphasize that we do *not* obtain a proof that Φ_h extends continuously to the Teichmüller boundary, because it is not clear if the convergence is uniform over the set of rays.

Limits of Teichmüller rays. Fix a holomorphic quadratic differential Ψ on M , and for $K > 1$ let $M_K = M_{K,\Psi}$. Let $f_K: M_K \rightarrow N$ be the harmonic map in the homotopy class of the identity, and $\Phi^K \in QD(M_K)$ its Hopf differential. Recall also the natural identification $\mathcal{L}: \mathcal{MF}(S) \rightarrow \mathcal{ML}(N)$ (§2). We will prove:

Theorem 8.1 (Teichmüller ray limit). *If $[\lambda] \in \mathcal{PML}(N)$ is a limit point of the projective classes $[\mathcal{L}(\Phi_h^K)]$, then the underlying laminations of $\mathcal{L}(\Psi_h)$ and $[\lambda]$ are identical.*

Note. We do not prove that the transverse measures on $\mathcal{L}(\Psi_h)$ and λ are projectively equivalent; in fact is this most likely false. At the very least, we must contend with the fact that the Teichmüller compactification is not independent of choice of basepoint—in particular, the same sequence of surfaces can be “viewed” from different base surfaces as endpoints of Teichmüller rays along foliations whose limiting projective measure classes are different (see [17]). Since we have not specified M , we must expect this to be a problem.

However, even if we make the natural choice of $M = N$, we cannot expect to obtain the right projective measure class in the limit, since Ψ_h optimizes ratios of extremal lengths and Φ_h optimizes ratios of hyperbolic to extremal lengths. Thus, if $\mathcal{L}(\Psi_h)$ has several components, the theorem tells us that λ will have the same components, but the choice of relative weights on the components which maximizes the extremal length ratio is likely to be different than the choice which maximizes the hyperbolic-to-extremal ratio.

For $\alpha, \beta \in \mathcal{MF}(S)$ define

$$I(\alpha, \beta) = \frac{i(\alpha, \beta)}{l_N(\alpha)l_N(\beta)},$$

where $l_N(\alpha)$ is understood to mean $l_N(\mathcal{L}(\alpha))$. This clearly only depends on the projective classes of α and β , and therefore determines a pairing on $\mathcal{PMF}(S)$. We then have the following easy corollary of Theorem 7.2 (energy is length-ratio):

Lemma 8.2 (Intersections vanish). *For M_K, Ψ , and Φ^K as above we have*

$$I(\Psi_h, \Phi_h^K) \leq \frac{c_1}{K}$$

for $K \geq c_2$, where c_1 and c_2 depend only on the choice of M and N .

Proof. We note first that there are constants $c_3, c_4 > 0$ depending strongly on M and N such that

$$(8.1) \quad \frac{1}{c_3} l_N^2(\gamma) \leq E_M(\gamma) \leq c_4 l_N^2(\gamma)$$

for any measured foliation $\gamma \in \mathcal{MF}(S)$. A simple proof is as follows: The functions l_N^2 and E_M are continuous and positive on $\mathcal{MF}(S) - \{0\}$,

and because they both scale quadratically (see §2), l_N^2/E_M is a continuous positive function on the compact set $\mathcal{PMF}(S)$. Thus it takes on a maximum and a minimum, and (8.1) follows.

Fix a large K , and denote by $\Psi^K \in QD(M_K)$ the image of the differential Ψ in M_K , obtained by contracting the $|\Psi|$ metric by K in the horizontal direction. Thus $\|\Psi^K\| = \|\Psi\|/K$, and Ψ_h^K and Ψ_h represent the same element of $\mathcal{MF}(S)$. In particular,

$$E_{M_K}(\Psi_h) = E_{M_K}(\Psi_h^K) = \|\Psi^K\| = E_M(\Psi_h)/K.$$

By the analytic characterization of extremal length, for any simple closed curve $\gamma \subset S$,

$$E_{M_K}(\gamma) \geq \frac{l_{|\Psi^K|}^2(\gamma)}{\text{Area}(|\Psi^K|)} \geq \frac{i(\gamma, \Psi_h^K)^2}{\|\Psi\|^2/K} \geq \frac{K}{E_M(\Psi_h)} i(\gamma, \Psi_h)^2.$$

By the continuity of $i(\cdot, \cdot)$ and E_{M_K} over $\mathcal{MF}(S)$, we conclude

$$E_{M_K}(\Phi_h^K) \geq \frac{K}{E_M(\Psi_h)} i(\Phi_h^K, \Psi_h)^2.$$

Thus,

$$I^2(\Phi_h^K, \Psi_h) \leq \frac{E_M(\Psi_h)}{K} \frac{E_{M_K}(\Phi_h^K)}{l_N^2(\Phi_h^K) l_N^2(\Psi_h)}.$$

Applying (8.1) to Ψ_h and Theorem 7.2 (energy is length-ratio) yields

$$I^2(\Phi_h^K, \Psi_h) \leq \frac{c_3}{K} \frac{1}{2(\mathcal{E}(f_K) - C)},$$

and, since

$$\mathcal{E}(f_K) \geq \frac{1}{2} \frac{l_N^2(\Psi_h)}{E_{M_K}(\Psi_h)} \geq \frac{1}{2c_3} \frac{E_M(\Psi_h)}{E_{M_K}(\Psi_h)} \geq \frac{K}{2c_3},$$

we have

$$I^2(\Phi_h^K, \Psi_h) \leq \frac{c_3^2}{K(K - 2Cc_3)}.$$

Hence the lemma follows for, say, $K \geq 4Cc_3$. q.e.d.

Now consider a subsequence $\{K_i\}$ for which $[\mathcal{L}(\Phi_h^{K_i})] \rightarrow [\lambda]$ in $\mathcal{PML}(N)$. Lemma 8.2 tells us that $i(\lambda, \mathcal{L}(\Psi_h)) = 0$, but not yet that the two laminations are the same. We proceed with the proof.

Proof of Theorem 8.1. Two minimal geodesic measured laminations with zero intersection number are either identical as sets of geodesics, or are disjoint—any nontransverse intersection is a sublamination (see §2

for a discussion of minimal laminations). Therefore it suffices to show that each minimal component of λ is contained in $\mathcal{L}(\Psi_h)$ (forgetting the measures), and vice versa.

Let ν be a minimal component of λ , and suppose first that ν is not a simple closed curve. If S_ν is a supporting subsurface for ν (see §2) in N , then S_ν must contain a simple closed curve γ not isotopic into ∂S_ν , since ν is a limit of such curves in S_ν and is not itself isotopic to ∂S_ν (in other words, S_ν cannot be an annulus or a pair of pants). Thus, $i(\gamma, \nu) > 0$ by the definition of S_ν , but $i(\gamma, \nu') = 0$ for any other minimal component ν' of λ or of $\mathcal{L}(\Psi_h)$.

The conclusion that the underlying lamination of ν is a component of $\mathcal{L}(\Psi_h)$ now follows from the following two assertions for any simple closed curve $\gamma \subset S$:

1. If $i(\gamma, \lambda) > 0$, then $\lim_{i \rightarrow \infty} E_i(\gamma) = \infty$.
2. If $i(\gamma, \Psi_h) = 0$, then $E_i(\gamma)$ remains bounded, (Here E_i denotes extremal length in M_{K_i} .)

Proof of (1). Denoting $\Phi^i = \Phi^{K_i}$, by the analytic definition of extremal length we have:

$$E_i(\gamma) \geq \frac{l_{|\Phi^i|}^2(\gamma)}{\|\Phi^i\|} \geq \frac{i(\gamma, \Phi_h^i)^2}{\|\Phi^i\|} = \left[\frac{i(\gamma, \Phi_h^i)}{l_N(\Phi_h^i)} \right]^2 \frac{l_N^2(\Phi_h^i)}{E_i(\Phi_h^i)} \\ \geq [l_N(\gamma)I(\gamma, \Phi_h^i)]^2 \cdot 2(\mathcal{E}(f_i) - C),$$

where the last inequality (and the constant C) comes from Theorem 7.2 (energy is length-ratio). Since $I(\gamma, \Phi_h^i) \rightarrow I(\gamma, \lambda) > 0$ and $\mathcal{E}(f_i) \rightarrow \infty$ as $i \rightarrow \infty$, it follows that $E_i(\gamma) \rightarrow \infty$ as well.

Proof of (2). We will in fact prove a little more. Let Σ denote the set of critical leaves of Ψ_h , and $\Sigma_c \subset \Sigma$ the noncontractible components of the subset of compact leaves, as in §2. Then the neighborhood in the $|\Psi|$ metric $\mathcal{N} = \mathcal{N}_\epsilon(\Sigma_c)$ (for ϵ small enough that the neighborhood is a thickening of Σ_c) supports any curve γ with $i(\gamma, \Psi_h) = 0$, by Lemma 2.1. We show

Lemma 8.3 (Psi complement bounded). *For any $\gamma \subset \mathcal{N}$, $E_{M_K}(\gamma)$ is bounded as $K \rightarrow \infty$, and, for $\gamma \subset \partial \mathcal{N}$, $E_{M_K}(\gamma) \rightarrow 0$.*

Proof. The $K^2|\Psi^K|$ metric on M_K is obtained by stretching the $|\Psi|$ metric in the vertical direction by a factor of K , so there is an isometric copy of \mathcal{N} in M_K for every K . Thus

$$E_{M_K}(\gamma) \leq E_{\mathcal{N}}(\gamma)$$

by the geometric definition of extremal length.

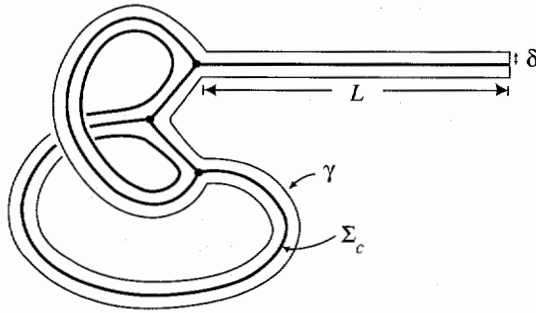


FIGURE 20

If γ is a component of $\partial\mathcal{N}$, we will construct regular annuli in M_K representing γ whose moduli will grow without bound. If γ is a horizontal trajectory of Ψ , then this is easy—the flat annulus bounded by γ and Σ_c has modulus $\epsilon K/l_{|\Psi|}(\gamma)$ in M_K . More likely, however, γ will intersect some noncompact leaves of Σ emanating from a singularity in Σ_c . In this case, choose a (large) $L > 0$ and append to \mathcal{N} an initial segment of length L from each such leaf, along with a δ -neighborhood of it, where $\delta(L)$ is small enough that this neighborhood is embedded (see Figure 20). The boundary curve representing γ in this new surface bounds (together with Σ_c) an annulus which, for $K > L/\delta(L)$, has radius at least L in the $K^2|\Psi^K|$ metric. Thus its modulus is logarithmically related to L (see §4.3) which was arbitrarily large. q.e.d.

Note that the dependence of δ on L is hard to control, so we do not know how fast $E_{M_K}(\gamma)$ approaches zero.

This concludes the proof that any minimal component of λ , which is not a simple closed curve, appears in $\mathcal{L}(\Psi_h)$. Consider now the case where $\nu \subset \lambda$ is a simple closed curve. Let $\hat{\nu} = m\nu$ denote the multiple of ν which has the counting measure. We shall show:

3. $E_i(\hat{\nu}) < c/K_i$, where c is independent of K_i .

4. This implies that ν is isotopic to the core of a flat annulus of $|\Psi|$. Since $i(\nu, \Psi_h) = 0$, it follows that the geodesic cores of the flat annulus are leaves of Ψ_h , and thus ν is a component of $\mathcal{L}(\Psi_h)$.

Proof of (3). Let γ denote any simple closed curve on N . The computation in the proof of assertion (1) above yields:

$$E_i(\gamma) \geq 2(\mathcal{E}_i - C)[i(\gamma, \Phi_h^i)/l_N(\Phi_h^i)]^2.$$

As in Lemma 8.2, $\mathcal{E}_i \geq c_1 K_i$ for some $c_1 > 0$. Further,

$$\lim i(\gamma, \Phi_h^i)/l_N(\Phi_h^i) = i(\gamma, \lambda)/l_N(\lambda).$$

Finally, $i(\gamma, \lambda) \geq i(\gamma, \hat{\nu}) = i(\gamma, \hat{\nu})/m$, so there is some $c_2 > 0$ depending on everything but K_i such that

$$E_i(\gamma) \geq c_2 K_i i(\gamma, \hat{\nu}).$$

On the other hand, the following is an easy consequence of the thick-thin decomposition for hyperbolic surfaces: there exist $\epsilon_0, c_3 > 0$ such that for any simple closed curve α on any closed Riemann surface M of bounded genus, if $E_M(\alpha) \leq \epsilon_0$ then there is a curve $\beta \subset M$ with $i(\alpha, \beta) = 2$ and

$$E_M(\beta) \leq \frac{c_3}{E_M(\alpha)}.$$

(β is obtained from two strands that cross through the thin part of α , together with arcs of bounded hyperbolic length in the thick part.)

Setting $\alpha = \hat{\nu}$ and $\gamma = \beta$, we obtain

$$E_i(\hat{\nu}) \leq \frac{c_3/(2c_2)}{K_i}.$$

Proof of (4). Given (3), for large enough i there exists an annulus $A \subset M_i$ with core $\hat{\nu}$ such that $\text{Mod}(A) \geq K_i/c$. By Theorem 4.6, A contains an annulus B which is regular in the $|\Psi^{K_i}|$ metric with $\mu(B) \geq K_i/c_4$. Let L_K be the minimal $|\Psi^K|$ -length of a curve representing $\hat{\nu}$. Then (see § 4.3) the $|\Psi^{K_i}|$ -area of B is bounded below by

$$\|\Psi^{K_i}\|_B \geq L_K^2 G(\mu(B)),$$

where $G(\mu) = \mu$ if B is flat, and is bounded between two fixed exponential functions if B is expanding. Since $|\Psi^K|$ is obtained from $|\Psi|$ by contracting the horizontal leaves by K , we have $L_K \geq L_1/K$ and $\|\Psi^{K_i}\|_B \leq \|\Psi^{K_i}\|_{M_i} \leq c_5/K$. Thus

$$L_1^2 \leq \frac{c_5 K_i}{G(K_i/c_4)}.$$

For large enough K_i this produces a contradiction if B is expanding, and thus eventually B must be flat. This of course means that there must be a flat annulus isotopic to B in the original $|\Psi|$ metric, and the intersection number argument, as above, implies that the cores of this annulus are leaves of Ψ_h .

Now, by Proposition 3.1 (energy lower bound), we have

$$\mathcal{E}_K(S_\nu^K) \geq \frac{1}{2} \frac{l_N^2(\nu)}{E_{S_\nu^K}(\nu)} \geq c_9 K$$

and, by inequality (3.4),

$$(8.3) \quad \|\Phi^K\|_{S_\nu^K} \geq c_{10} K$$

for large enough K .

The next step is to show that the large $|\Phi^K|$ -area of S_ν^K implies that a train-track approximation can be built with a component of definite length supported by S_ν .

For a boundary component γ of S_ν let γ^K denote a geodesic representative in the $|\Phi^K|$ metric on M_K . Let $L_K = \max_{\gamma \in \partial S_\nu} (l_{|\Phi^K|}(\gamma^K))$. We claim:

$$(8.4) \quad \lim_{K \rightarrow \infty} L_K^2/K = 0.$$

We can assume, possibly by restricting to a subsequence, that there is some fixed $\gamma \in \partial S_\nu$, which is longest in the $|\Phi^K|$ metric. Since $E_K(\gamma) \rightarrow 0$, we can find for each K a regular annulus in the $|\Phi^K|$ metric representing γ whose modulus $m(K)$ approaches infinity; its $|\Phi^K|$ -area is at least $L_K^2 G(m(K))$, where G is a function at least linear in its argument, again from §4.3. Since $\|\Phi^K\| \leq cK$, it follows that $L_K^2/K \leq c/G(m(K)) \rightarrow 0$.

Let T_ν^K be the 2-complex homotopic to S_ν which is bounded by the $|\Phi^K|$ -geodesics $\{\gamma^K\}$ (we say 2-complex rather than surface because some γ^K might not be embedded, but in that case T_ν^K is still isotopic to a deformation-retract of S_ν^K). If B_γ^K is expanding “into” S_ν^K , and the boundary it shares with S_ν^K is outwardly curved with respect to S_ν^K , then γ^K is outside the interior of S_ν^K . If B_γ^K is expanding “out” of S_ν^K , then γ^K is contained in S_ν^K , but the annulus between γ^K and the outer boundary of B_γ^K has $|\Phi^K|$ -area at most $c_{11} L_K^2$, since by our construction the modulus of this annulus is bounded by m_1 . In any case, the $|\Phi^K|$ -mass of the set $S_\nu^K - T_\nu^K$ is a shrinking proportion of K , so by (8.4) and (8.3), we have

$$(8.5) \quad \|\Phi^K\|_{T_\nu^K} \geq c_{12} K.$$

Choose $\epsilon > 0$ small enough that a $(1, \epsilon)$ -nearly straight train-track gives a $c\epsilon$ -approximation to any lamination it carries (see §6), and choose

R_0 large enough that the train-track construction of the previous section on $M - \mathcal{P}_R$ for any M yields a $(1, \epsilon)$ -straight train-track for any $R > R_0$. If we now choose $R_K = \max(R_0, L_K/2)$, then \mathcal{P}_{R_K} in M_K separates T_ν^K from the rest of M_K . By (8.4) the mass of the components of $M_K - \mathcal{P}_{R_K}$ in T_ν^K is at least $c_{13}K$.

The corresponding portion of the train-track thus lies in a subsurface isotopic to S_ν , and its total length is a definite proportion of the total length of $\mathcal{L}(\Phi_h^K)$. The limiting lamination must therefore meet this subsurface.

It remains to consider the case where ν is a simple closed curve. In this case S_ν is just the geodesic representative of ν , and S_ν^0 is the flat cylinder consisting of closed Ψ_h trajectories homotopic to ν . As before (but easier), $E_K(\nu) \leq c_{15}/K$. We deduce from this that ν is represented by a long flat cylinder in the $|\Phi^K|$ metric, using a similar argument to that used in the first half of the proof. Let $L = l_{|\Phi^K|}(\nu^K)$, where ν^K is a $|\Phi^K|$ -geodesic representing ν . If A is any regular annulus representing ν in M_K , then $\|\Phi^K\|_A = L^2 G(\mu(A))$, where $G(\mu) = \mu$ if A is flat, and is bounded between two exponential functions of μ if A is expanding.

A lower bound on L can now be obtained. By Theorem 4.6 (modulus of any annulus) there is a regular annulus A such that $\mu(A) \geq c_{16}K$, and we can assume that one of its boundaries is ν^K . Take a subannulus $B \subset A$, also regular and bounded by ν^K , for which $\mu(B) = 1 + 4\pi|\chi(M)|/l_N^2(\nu)$. By (3.5) and Proposition 3.1 (energy lower bound), we obtain

$$\frac{1}{2} \frac{l_N^2(\nu)}{E_B(\nu)} \leq \mathcal{E}(B) \leq 2\pi|\chi(M)| + 2\|\Phi^K\|_B,$$

or, since $1/E_B(\nu) = \text{Mod}(B) \geq \mu(B)$, and using the above formula,

$$l_N^2(\nu)\mu(B) \leq 4\pi|\chi(M)| + 4L^2 G(\mu(B)),$$

which implies

$$L^2 \geq L_0^2 \equiv l_N^2(\nu)/4G(\mu(B)).$$

Now we can use the fact that $\|\Phi^K\|_A \leq \|\Phi^K\| \leq c_{17}K$ to argue that

$$L_0^2 G(\mu(A)) \leq c_{17}K,$$

or, since $\mu(A) \geq c_{16}K$ and G is increasing,

$$L_0^2 \leq \frac{c_{17}K}{G(c_{16}K)}.$$

If A is expanding, this immediately produces a contradiction for large enough K , so in fact the regular annulus of modulus proportional to K must be flat.

Finally, using the estimates of §3.3, we can see that for large enough K the images of the horizontal trajectories of Φ^K in A spiral closely around the geodesic representative of ν in N , and their length contribution (using the train-track construction) is proportional to $\|\Phi^K\|_A \geq c_{16} L_0^2 K$. Therefore, the geodesic ν must be contained in the limiting lamination λ .

References

- [1] L. Ahlfors, *Conformal invariants—Topics in geometric function theory*, McGraw-Hill, New York, 1973.
- [2] W. Ballmann, M. Gromov & V. Schroeder, *Manifolds of nonpositive curvature*, Birkhäuser, Boston, 1985.
- [3] F. Bonahon, *Bouts des variétés hyperboliques de dimension 3*, Ann. of Math. (2) **124** (1986) 71–158.
- [4] R. D. Canary, D. B. A. Epstein & P. Green, *Notes on notes of Thurston*, Analytical and Geometric Aspects of Hyperbolic Space, London Math. Soc. Lecture Notes Series, No. 111, Cambridge University Press, 1987, 3–92.
- [5] A. J. Casson, *Automorphisms of surface after Nielsen and Thurston*, Notes by S. A. Bleiler, University of Texas, Austin, 1982.
- [6] A. J. Casson & S. A. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*, Cambridge University Press, 1988.
- [7] J. Eells, Jr. & L. Lemaire, *Selected topics in harmonic maps*, CBMS Regional Conf. Ser., No. 50, Amer. Math. Soc., Providence, RI, 1980.
- [8] J. Eells, Jr. & J. H. Sampson, *Harmonic mappings of Riemannian manifolds*, Amer. J. Math. **86** (1964), 109–160.
- [9] A. Fathi, F. Laudenbach, & V. Poenaru, *Travaux de Thurston sur les surfaces*, Asterisque **66–67** (1979).
- [10] F. P. Gardiner, *Teichmüller theory and quadratic differentials*, Wiley-Interscience, New York, 1987.
- [11] D. Gilbarg & N. S. Trudinger, *Elliptic partial differential equations of second order*, Springer, Berlin, 1983.
- [12] J. Harer & R. Penner, *Combinatorics of train tracks*, preprint.
- [13] P. Hartman, On homotopic harmonic maps., Canad. J. Math. **19** (1967) 673–687.
- [14] A. E. Hatcher, *Measured lamination spaces for surfaces, from the topological viewpoint*, Topology Appl. **30** (1988) 63–88.
- [15] J. Hubbard & H. Masur, *Quadratic differentials and foliations*, Acta Math. **142** (1979) 221–274.
- [16] J. Jost, *Harmonic maps between surfaces*, Lecture Notes in Math., Vol. 1062, Springer, Berlin, 1984.
- [17] S. Kerckhoff, *The asymptotic geometry of Teichmüller space*, Topology **19** (1980) 23–41.
- [18] —, *The Nielsen realization problem*, Ann. of Math. (2) **117** (1983) 235–265.
- [19] G. Levitt, *Foliations and laminations on hyperbolic surfaces*, Topology **22** (1983) 119–135.
- [20] Y. Minsky, *Harmonic maps and hyperbolic geometry*, Ph.D. thesis, Princeton University, 1989.
- [21] —, *Harmonic maps into hyperbolic 3-manifolds*, Trans. Amer. Math. Soc., to appear.

- [22] R. Osserman, *The isoperimetric inequality*, Bull. Amer. Math. Soc. **84** (1978) 1182–1238.
- [23] J. H. Sampson, *Some properties and applications of harmonic mappings*, Ann. Sci. École Norm. Sup. (4) **11** (1978) 211–228.
- [24] R. Schoen, *Analytic aspects of the harmonic map problem*, preprint.
- [25] R. Schoen & S.-T. Yau, *On univalent harmonic maps between surfaces*, Invent. Math. **44** (1978) 265–278.
- [26] —, *Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature*, Ann. of Math. (2) **110** (1979) 127–142.
- [27] K. Strebel, *Quadratic differentials*, Springer, Berlin, 1984.
- [28] W. Thurston, *The geometry and topology of 3-manifolds*, Princeton University Lecture Notes, 1982.
- [29] —, *Hyperbolic structures on 3-manifolds. I: Deformation of acylindrical manifolds*, Ann. of Math. (2) **124** (1986) 203–246.
- [30] —, *Minimal stretch maps between hyperbolic surfaces*, unpublished manuscript, Princeton University.
- [31] M. Wolf, *The Teichmüller theory of harmonic maps*, J. Differential Geometry **29** (1989) 449–479.
- [32] —, *High energy degeneration of harmonic maps between surfaces*, Topology, to appear.
- [33] —, *Infinite energy harmonic maps and degeneration of hyperbolic surfaces in moduli space*, J. Differential Geometry, to appear.

YALE UNIVERSITY