

Spaces which are not negatively curved

LEE MOSHER AND ULRICH OERTEL ¹

Introduction.

The discovery by M. Gromov of negatively curved metric spaces and groups [Gro87] has strongly influenced geometric topology and combinatorial group theory. Spaces which are negatively curved generalize hyperbolic manifolds, and they share many of the geometric features of hyperbolic manifolds. For example, in a compact, negatively curved space with non-elementary fundamental group, geodesics behave much as they do in a hyperbolic manifold: the universal cover has a space at infinity corresponding to endpoints of geodesics, and there is an ergodic geodesic flow. For compact spaces which are not negatively curved, it seems desirable to find some common geometric features, and it would be even better to find a geometric feature which characterizes lack of negative curvature. This characterization should be similar to W. Thurston's Hyperbolization Conjecture for 3-Manifolds, which says that a closed 3-manifold M with infinite, freely indecomposable fundamental group is hyperbolic if and only if M has no immersed, incompressible torus. This conjecture can be weakened as follows [Mos95]:

Weak Hyperbolization Conjecture. *Given a closed 3-manifold M , exactly one of the following happens:*

1. M is negatively curved in the sense of Gromov
2. M has an immersed incompressible torus.

We present a geometric characterization for compact spaces which are not negatively curved, in terms of their 2-skeleta. The reader will recognize, even without precise definitions of the terminology, that our characterization parallels the above statement.

¹The first author is supported in part by NSF grant DMS 95-04946.

Theorem A. *Given a finite 2-complex X , exactly one of the following happens:*

1. X is negatively curved in the sense of Gromov.
2. There is compact, 2-dimensional measured lamination Λ of Euler characteristic zero, and a “nice” map $f: \Lambda \rightarrow X$ which is strongly least area on leaves.

Here is a brief discussion of negative curvature and laminations, sufficient to understand the statement of Theorem A; §1 contains a more formal treatment.

Negative curvature for a finite, simplicial 2-complex X is defined in terms of a linear isoperimetric inequality. Following [BF92], we say that X is negatively curved if there is a constant K such that for every homotopically trivial closed curve γ , there exists a disc map $d: D \rightarrow X$ spanning γ such that $\text{Area}(d) \leq K \cdot \text{Length}(\gamma)$. Length and area are measured by using some piecewise simplicial metric on X , for example a metric that makes the 2-simplices into equilateral Euclidean triangles of some constant side length. Gromov has proved that negative curvature of X is equivalent to the universal cover \tilde{X} being a *hyperbolic metric space* (see [Gro87], [CDP90], or appendix 2 of [GS90])—there exists $\delta \geq 0$ such that for each geodesic triangle Δ in \tilde{X} and for each point p in one side of Δ , the distance from p to the union of the other two sides is at most δ .

The property of negative curvature can be formulated for any finite simplicial complex Y . Negative curvature is independent of the actual metric chosen on Y . Indeed, it depends only on the fundamental group [Gro87]: Y is negatively curved if and only if, for any finite simplicial complex X with $\pi_1(X) \approx \pi_1(Y)$, X is also negatively curved. It follows that if you want to study negative curvature of Y , you are free to replace Y with any X having the same fundamental group. In order to simplify the lamination theory, we pick X to be a finite 2-complex.

We should point out that the term “negatively curved space” is sometimes used in the literature to refer to a space which is “locally $\text{CAT}(K)$ with $K < 0$ ”; this is a local condition on X or on \tilde{X} , whereas our concept of negatively curved space is an asymptotic condition on \tilde{X} .

We use the term *lamination* for what has been called a “foliated space” in the literature. Our laminations are always 2-dimensional, so a lamination has the local structure of the product of a 2-disc with a locally compact Hausdorff space T called the *local transversal*. A *measured lamination* is a lamination Λ together with an assignment of a finite, positive Borel measure

to each local transversal, so that a compatibility condition holds as the local transversal moves around. The simplest example of a measured lamination is a compact surface, whose local transversals are points which are given the Dirac measure.

A “nice” map from a measured lamination Λ to a 2-complex X is a continuous function $f: \Lambda \rightarrow X$, with certain restrictions on the local behavior of f on leaves, in order to rule out pathologies like constant maps from Λ to X . The map f is *normal* if for each leaf L of Λ , there is a cell-decomposition of L such that each k -cell maps homeomorphically to a k -cell of X , and for any 2-cells $\sigma \neq \tau$ in L sharing a common vertex, we have $f(\sigma) \neq f(\tau)$. If we relax this definition by allowing a 2-cell of L to map as an n -fold cyclic cover of a 2-cell in X , and if n is bounded above by some constant N , we obtain the definition of an N -*normal* map. In theorem A, “nice” means “3-normal”. This cannot be strengthened very much: there are examples of 2-complexes X which are not negatively curved, and for which the second condition of theorem A fails if “nice” means “normal”.

The *Euler characteristic* of a measured lamination Λ generalizes the Euler characteristic of a compact surface. We use a definition which generalizes Euler’s formula $\chi = F - E + V$ for a surface with F faces, E edges, and V vertices. This formula generalizes most easily when there is a map $f: \Lambda \rightarrow X$ which is a local embedding on leaves. In that case, F represents the total transverse measure of Λ over the barycenters of 2-cells of X , E is the total transverse measure over the barycenters of 1-cells, V is the total transverse measure over the 0-cells, and we have $\chi(\Lambda) = F - E + V$.

A map $f: \Lambda \rightarrow X$ is *strongly least area* on leaves if for each leaf $L \subset \Lambda$ with universal cover \tilde{L} , and for each embedded disc $D \subset \tilde{L}$, the induced disc map $D \rightarrow X$ has least area among all disc maps with the same boundary. Notice that when Λ is a 2-torus, this condition is quite unlike the usual condition that $f: \Lambda \rightarrow X$ have least area in its homotopy class.

The laminations produced by Theorem A are called *limiting laminations*, for reasons which will be explained below. The strongly least area property is necessary in order for the lamination to carry any geometric information about X . It is instructive to consider the case where X has a limiting lamination which is a closed, connected oriented surface, that is, a torus T^2 . If there is a strongly least area map $f: T^2 \rightarrow X$, then f is π_1 -injective, so $\pi_1(X)$ has a \mathbf{Z}^2 subgroup. To see why, consider the induced map $\tilde{f}: \mathbf{R}^2 \rightarrow X$, and let $\Gamma \subset \mathbf{R}^2$ be the lattice realizing the group of deck transformations. If f is not π_1 -injective, there is a line segment $E \subset \mathbf{R}^2$ connecting two lattice points such that $\gamma = \tilde{f}(E)$ is a homotopically trivial curve in X . It follows that γ extends to a disc map $d: D \rightarrow X$. Let A be the area of this disc

map. Now construct a large parallelogram P in \mathbf{R}^2 with E as one side, and with vertices in Γ , having area larger than $2A$. This parallelogram P , when mapped into X , is assumed to be least area. The boundary of P maps to a path of the form $\gamma\delta\gamma^{-1}\delta^{-1}$ in X , for some closed curve δ . This path also bounds a disc map whose domain is obtained by joining two copies of D by a long, thin strip of infinitesimal area. This disc map has area $2A$, contradicting that f is strongly least area; therefore f is π_1 -injective.

The above argument raises the hope of proving the Weak Hyperbolization Conjecture for a 3-manifold M by showing that if M is not negatively curved, then a limiting lamination $\Lambda \rightarrow M$ can be approximated by an incompressible torus. Using the approximation methods of Morgan & Shalen [MS88], this is true if the map $\Lambda \rightarrow M$ is an embedding; but this approach gives no information for nonembedded limiting laminations. Moreover, there exist spaces with limiting laminations which cannot be approximated by an incompressible torus. For instance, when $\pi_1(X)$ is the Baumslag-Solitar group $BS(1, 2) = \langle a, b \mid bab^{-1} = a^2 \rangle$, the space X is not negatively curved, but $\pi_1(X)$ has no \mathbf{Z}^2 subgroup. In §5 we give an explicit description of X and its limiting lamination Λ . This lamination does not have a Euclidean structure, in fact it has leaves which are quasi-isometric to the hyperbolic plane. We also describe an example of a space X with a limiting lamination that does have a Euclidean structure, yet this lamination cannot be approximated by a compact surface. These examples emphasize that any hope of proving the Weak Hyperbolization Conjecture for 3-manifolds must rest on special methods of 3-manifold topology.

Here is an outline of the proof of Theorem A. Suppose X is not negatively curved. It follows that there is a sequence of least area disc maps $d_j: D_j \rightarrow X$, such that

$$\text{Area}(D_j)/\text{Length}(\partial D_j) \rightarrow \infty \quad \text{as } j \rightarrow \infty$$

Such a sequence is called a *fat disc sequence* for X , and lack of negative curvature is characterized by the existence of a fat disc sequence. Regard these disc maps as a sequence of measured laminations with boundary, by putting a transverse Dirac measure on each disc D_j . Now normalize the Dirac measure, dividing it by $\text{Area}(D_j)$. Show that the fat disc sequence has a subsequence converging to a lamination map $f: \Lambda \rightarrow X$ such that Λ has no boundary and $\chi(\Lambda) = 0$. The fact that Λ has no boundary comes from an argument of Plante [Pla73]: in a fat disc sequence, the normalized boundary length is equal to $\text{Length}(\partial D_j)/\text{Area}(D_j)$ which vanishes in the limit. The fact that the $\chi(\Lambda) = 0$ comes from an argument we learned from Morgan-Shalen [MS88]: after passing to a subsequence, we have $D_j/\text{Area}(D_j) \rightarrow \Lambda$,

so $\chi(D_j)/\text{Area}(D_j) \rightarrow \chi(\Lambda)$. But $\chi(D_j) = 1$ and $\text{Area}(D_j) \rightarrow \infty$, so $\chi(D_j)/\text{Area}(D_j) \rightarrow 0$ and it follows that $\chi(\Lambda) = 0$. These arguments are carried out in §2.

The hardest part of this outline is obtaining convergence. We do this by constructing a compact universal lamination with boundary Ω , equipped with a map $f: \Omega \rightarrow X$, so that the leaves of Ω consist of all nice (i.e. 3-normal) maps of surfaces to X . Then we prove the Normal Disc Theorem which says that any fat disc sequence can be replaced by one which is 3-normal, hence the discs D_j in the sequence may be regarded as leaves of Ω . The space of normalized transverse measures on Ω is compact, hence the normalized Dirac measures on the D_j have a subsequence converging to a transverse measure on Ω , whose support is the desired measured lamination Λ .

The proof that $f: \Lambda \rightarrow X$ is strongly least area on leaves is carried out in §3. The main tool is the Strong Shadowing Theorem, which shows that for each leaf L of Λ with universal cover \tilde{L} , the induced map $\tilde{f}: \tilde{L} \rightarrow X$ is a Gromov-Hausdorff limit of a subsequence of the fat disc sequence $d_j: D_j \rightarrow X$. In particular, for each embedded disc $D \subset \tilde{L}$ there is an embedding $D \hookrightarrow D_j$ for some j such that the composed map $D \hookrightarrow D_j \rightarrow X$ is identical to $\tilde{f}|_D$, and hence the latter map is least area.

The converse is proved in §4 by quoting results from analysis; our thanks to E. Ghys for suggesting this method. Given a strongly least area measured lamination $f: \Lambda \rightarrow X$ with $\chi(\Lambda) = 0$, it is shown in §1 that Λ has no sphere leaves. Then we use a theorem of Ghys to conclude that almost every leaf L of Λ is conformally Euclidean. Next we use the Ahlfors Lemma which says, in modern language, that when L is conformally Euclidean it cannot be negatively curved; in fact, the universal cover \tilde{L} has an embedded fat disc sequence. Finally, using the fact that the induced map $\tilde{L} \rightarrow X$ is least area, it follows that X is not negatively curved.

Theorem A is applied to give new proofs, in the restricted setting of 2-complexes, for two geometric results: the Flat Plane Theorem (see §5), and the Conformal Minimal Plane Theorem (see end of §4). The Flat Plane Theorem says that if X is nonpositively curved in the sense of CAT(0) geometry, and if X is not negatively curved in the sense of Gromov, then the universal cover \tilde{X} contains an isometrically embedded copy of the Euclidean plane; this is proved in [Bri] for any nonpositively curved space. The Conformal Minimal Plane Theorem says that if X is not negatively curved in the sense of Gromov then there is nonconstant, conformal least area map of the Euclidean plane to X ; this is stated in [Gro87] for any compact space.

It might be interesting to prove an analogue of Theorem A in the category of closed Riemannian manifolds, where “least area” is interpreted in the classical sense of a solution to the Plateau problem. This should yield new proofs of the Flat Plane Theorem and the Conformal Minimal Plane Theorem in their full generality. The main difficulty in obtaining a Riemannian version of Theorem A will be the local lamination theory (we chose to restrict attention to 2-complexes in part because the local lamination theory is so simple). In particular, one would have to prove a Riemannian version of the Normal Disc Theorem, that any least area disc map into a closed Riemannian manifold may be replaced by one with no smaller isoperimetric ratio, and with bounded local geometry in the sense of [Sch]: uniform curvature bounds are satisfied at each point of the disc, up to and including points on the boundary.

Acknowledgements. We would like to thank Kay Tatsuoka and Mark Feighn for a seminar introducing us to Gromov’s theory of negatively curved groups and spaces. Our results were inspired by the methods of Plante, as developed by Morgan and Shalen. We would like to thank Benjamin Weiss for comments about aperiodic tilings, which led to the discovery of a Euclidean limiting lamination which cannot be approximated by a torus. And we thank Dave Gabai for reading a preliminary version of the paper and rooting out some errors.

1. Preliminaries.

1.1. Negatively curved spaces.

We shall work primarily in the simplicial category, so we will use the unadorned word *complex* to denote a simplicial complex; occasionally we also speak of cell complexes. The k -skeleton of a complex X is denoted $X^{(k)}$. The n -th barycentric subdivision of X is denoted $\Sigma^n X$. The link of a vertex $v \in X$ is denoted $\text{Link}(v)$. We shall use the notions of regular neighborhoods, simplicial maps, PL subcomplexes, PL maps, and general position of PL maps. A *triangulation* of a topological space X is a simplicial complex with underlying space X . By convention, if X is a complex and Y is a topological space, a map $f: Y \rightarrow X$ is called simplicial, PL, etc., if there is a triangulation of Y so that f is simplicial, PL, etc.

Given topological spaces X, Y, Z and continuous maps $f: Y \rightarrow X$ and $g: Z \rightarrow X$, the maps f and g are *equivalent* if there is a homeomorphism $h: Y \rightarrow Z$ such that $f = g \circ h$.

Let X be a finite 2-complex. If each edge of X is incident to no more than three 2-simplices then we say that X is *generic*. It is easy to change an arbitrary complex X so that it is generic, without changing $\pi_1(X)$, or indeed without changing the simple homotopy type of X . From now on, we fix a generic 2-complex X .

Each 2-simplex of X is given a metric making it isometric to an equilateral Euclidean triangle of side length 1. We will normalize the area measure so that each 2-simplex has total area 1. By integrating pull-back measures we therefore obtain a well-defined notion of length for each PL map of a 1-manifold into X , and of area for each PL map of a 2-manifold. For simplicial maps the measurement of length and area is more elementary, e.g. for a simplicial curve just count the number of edges in the domain mapping homeomorphically to edges of X . Having fixed these notions of length and area, we use them from now on without comment.

A *disc map* is a simplicial map from a disc to X . For emphasis we sometimes call this a *simplicial disc map*. We will also need the more general category of PL disc maps, which are just PL maps from a disc to X . The unadorned term “disc map” is reserved for simplicial maps.

A finite complex X is *negatively curved* if there is a constant K such that for every homotopically trivial closed simplicial curve γ in X , there is a disc map $d: D \rightarrow X$ extending γ (i.e. ∂d is equivalent to γ), so that the following *linear isoperimetric inequality* is satisfied:

$$\text{Area}(d) \leq K \cdot \text{Length}(\gamma)$$

This definition is equivalent to the usual “thin triangles” definition for either the 1-skeleton of the universal cover of X , or for any Cayley graph of $\pi_1(X)$ (see [Gro87], [CDP90], or appendix 2 of [GS90]).

Lack of negative curvature for X may be characterized as follows. The *isoperimetric ratio* of a disc map $d: D \rightarrow X$ is defined to be $\mathcal{I}(d) = \text{Area}(d)/\text{Length}(\partial d)$. A disc map $d: D \rightarrow X$ is said to be *least area* if, for every disc map $d': D' \rightarrow X$ such that ∂d and $\partial d'$ are equivalent, $\text{Area}(d) \leq \text{Area}(d')$. The complex X is not negatively curved if and only if there exists a sequence of least area disc maps $d_j: D_j \rightarrow X$ such that $\lim_{j \rightarrow \infty} \mathcal{I}(d_j) = \infty$. Such a sequence will be called a *fat disc sequence* for X , and when the entire sequence is understood, d_j will be called a *fat disc map* and D_j a *fat disc*.

We also use a PL notion of least area. A PL disc map $d: D \rightarrow X$ is PL least area if it has least area among all PL disc maps with the same boundary. Given a simplicial disc map, the property of PL least area is a priori stronger than simplicial least area. However, it is not hard to see these are equivalent; this follows from the first part of the proof of the Normal Disc Theorem.

1.2. Measured Laminations.

The term “lamination” is used in the theory of surfaces and 3-manifolds to mean a closed, foliated subspace; the local transversals of such spaces are often totally disconnected, instead of the k -disc transversal of a codimension- k foliation of a manifold. Our foliated spaces will not exist in any ambient manifold, and they will have totally disconnected transversals, hence we use the term “lamination”. The term “foliated space” has also been used in the literature. We begin by defining topological laminations, then we introduce additional structure to define measured laminations and simplicial laminations. The latter are useful for defining the Euler characteristic of a measured lamination. All laminations throughout this paper will be 2-dimensional.

Definition of a lamination. Let Λ be a compact topological space. A *lamination chart* for Λ consists of an open subset $V \subset \Lambda$, an open disc $D \subset \mathbf{R}^2$, a locally compact Hausdorff topological space T called the *local transversal*, and a homeomorphism $g: D \times T \rightarrow V$. A *plaque* of the chart is any set of the form $P = g(D \times t)$ for some $t \in T$. Informally, we shall often identify a plaque of the chart with D . Although the chart is formally a quadruple (V, D, T, g) , we shall often suppress the role of g and write $V = D \times T \subset \Lambda$; also, we shall sometimes replace D with some plaque P and write $V = P \times T \subset \Lambda$. We often confuse the local transversal T with any section of the projection $D \times T \rightarrow T$.

If we wish to deal with laminations having boundary, we must consider pairs of compact spaces $(\Lambda, \partial\Lambda)$. There will be two kinds of lamination charts. One kind was described above, mapped to an open set in $(\Lambda - \partial\Lambda)$. In addition we will have charts based on a half-disc model $(D, \partial D)$, where D is the intersection of the open unit disc in \mathbf{R}^2 with the closed upper half-plane $H \subset \mathbf{R}^2$, and $\partial D = D \cap \partial H$. The notation ∂D denotes the “half-boundary” of the half-disc D . Charts based on a half-disc are maps of pairs $(D \times T, \partial D \times T) \rightarrow (\Lambda, \partial\Lambda)$. For simplicity, we shall assume below that we are dealing with the closed case, $\partial\Lambda = \emptyset$.

A lamination chart (U, C, S, f) is a *sub-chart* of (V, D, T, g) if $U \subset V$, and the map

$$C \times S \xrightarrow{f} U \hookrightarrow V \xrightarrow{g^{-1}} D \times T$$

has the form $(x, s) \mapsto (h(x, s), p(s))$. The map $h_s: x \mapsto h(x, s)$ is an embedding of C into D for each $s \in S$, called a *leaf overlap map*. The map $p: S \rightarrow T$ is an embedding called the *transversal overlap map*.

A *lamination atlas* for Λ consists of a collection of lamination charts which cover Λ , such that for any two charts U, V and any $x \in U \cap V$, there is a common subchart W containing x . If Λ is equipped with a lamination atlas, then Λ is called a *lamination*. Any plaque of any chart of Λ is also called a plaque of Λ . The collection of all open subsets of all plaques of Λ forms a basis for the *leaf topology* on Λ , and any component of this topology is called a *leaf* of the lamination Λ . A local transversal is also called an *open transversal*, and any subset T' of an open transversal such that T' is a closed subset of Λ is called a *closed transversal* of Λ .

Next, a *transverse measure* ν on Λ assigns to each local transversal T a finite Borel measure, whose total mass is denoted $\nu(T)$, so that each transversal overlap map is measure preserving onto its image. The pair (Λ, ν) is called a *measured lamination* if $\nu(T) > 0$ for all local transversals T .

Cell structures and simplicial structures on measured laminations.

By making the leaves of a lamination Λ into cell complexes which fit together in the transverse direction, we obtain a *cell structure* on Λ , which is useful in defining the Euler characteristic of any transverse measure on Λ . When each 2-cell is a 2-simplex—a triangle—we obtain a *simplicial structure* on Λ . We shall not give the most general definitions; rather, our definitions are tailored to laminations having an atlas where each local transversal is both open and closed.

Let Λ be a lamination. If $D \times T \subset \Lambda$ is a lamination chart where T is an open and closed transversal, and if P is a polygon embedded in the open disc D , then $P \times T$ is called a *polygon chart* of Λ . We shall usually suppress the role of D , although the existence of the chart $D \times T$ is part of the definition of a polygon chart. A *cell structure* on Λ is a set of polygon charts \mathcal{P} covering Λ , such that if $P_1 \times T_1 \neq P_2 \times T_2 \in \mathcal{P}$ are not disjoint then the intersection, written in $P_1 \times T_1$ coordinates, has one of the following two forms:

- $E \times T$ where E is an edge of P_1 and T is an open and closed subset of T_1 .
- $V \times T$, where V is a vertex of P_1 and T is an open and closed subset of T_1 .

If $\mathcal{P} = \{P_1 \times T_1, \dots, P_N \times T_N\}$ is a finite cell structure on a compact lamination Λ , then each leaf L inherits the structure of a cell complex, where each 2-cell has the form $P_n \times t$ for some $n = 1, \dots, N$ and some $t \in T_n$. This collection of cell complexes, one for each leaf, has bounded local combinatorics: there is a uniform bound on the number of sides of each 2-cell and on the valence of each vertex. If each P_n is a 2-simplex then we obtain a *simplicial structure* on Λ .

If \mathcal{P} is a cell structure on Λ , the *barycentric subdivision* of \mathcal{P} is a simplicial structure $\Sigma^1\mathcal{P}$ defined in the obvious manner, so that the simplicial structure induced on a leaf L by $\Sigma^1\mathcal{P}$ is the barycentric subdivision of the cell structure on L induced by \mathcal{P} .

Laminations on 2-complexes. Consider a 2-complex X . In the crudest definition, a “lamination on X ” would just be a continuous map from a lamination to X . But we want a more refined definition with built-in restrictions on the local behavior on leaves, in order that the collection of laminations on X satisfy some kind of precompactness. We will achieve this by restricting the germs of maps at any given point of X .

If Λ is a lamination, a map $p: \Lambda \rightarrow X$ is called an *N -normal lamination map*, where N is a positive integer, if there exists a cell structure \mathcal{P} on Λ such that each leaf L of Λ satisfies the following:

- The map $p \mid L: L \rightarrow X$ is a nondegenerate simplicial map between barycentric subdivisions, that is, between the simplicial structure on L induced by $\Sigma^1\mathcal{P}$ and the simplicial complex $\Sigma^1 X$.
- For each 2-cell C of L , the map $p \mid C$ is a cyclic branched cover of degree $\leq N$ over some 2-simplex σ of X , branched over the barycenter of σ (see figure 1).
- If two distinct 2-cells in L have nonempty intersection, then they map to distinct 2-simplices in X .

If two distinct 1-cells in L have nonempty intersection, there is no requirement that they map to distinct 1-simplices in X , and so an N -normal lamination map may have leaves that are not locally embedded near vertices of

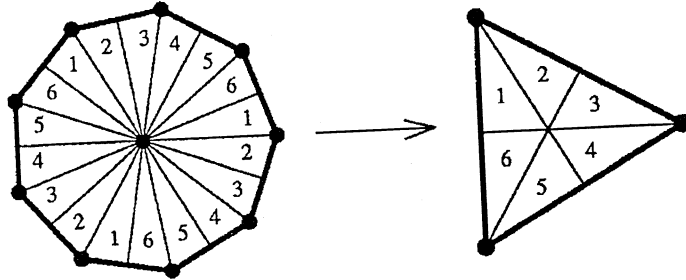


Figure 1: A 3-fold cyclic branched cover of a 2-simplex

X . Of course, there may also be leaves which are not locally embedded near barycenters of 2-simplices of X .

If $(\Lambda, \partial\Lambda)$ is a lamination with boundary, an N -normal lamination map is defined similarly as a map of pairs $p: (\Lambda, \partial\Lambda) \rightarrow (X, X^{(1)})$, with the additional requirement that leaves of $\partial\Lambda$ are locally embedded in $X^{(1)}$.

In the special case that the lamination Λ is just a surface, that is, all transversals are discrete, then an N -normal lamination map $\Lambda \rightarrow X$ is called an N -normal surface mapped to X .

1.3. Euler characteristic.

Suppose Λ is a measured lamination with a finite cell structure $\mathcal{P} = \{P_i \times T_i\}_{i=1}^I$. We shall define $\chi(\Lambda)$ in terms of \mathcal{P} . In principle, $\chi(\Lambda)$ should not depend on \mathcal{P} . When Λ has a Riemannian metric, this follows for all “smooth cell structures” from a Gauss-Bonnet type theorem which equates the polyhedral formula for $\chi(\Lambda)$ with a certain curvature integral; see §4.

To define the Euler characteristic, we need a way to count faces, edges, and vertices of \mathcal{P} , using the transverse measure ν . The *total weight of faces* is defined to be

$$F = \sum_{i=1}^I \nu(T_i)$$

We must also define the total weights of edges and vertices. First note that the “1-skeleton” $\bigcup \{\partial P_i \times T_i\}$ can be rewritten as the union of a finite collection of the form $\{E_j \times T'_j \mid j = 1, \dots, J\}$, so that for each j the set E_j is an edge of some P_i and $T'_j \subset T_i$, and so that if $E_{j_1} \times T'_{j_1}$ and $E_{j_2} \times T'_{j_2}$ are

not disjoint then their intersection has the form $V \times T''$ for some endpoint V of E_{j_1} and some $T'' \subset T'_{j_1}$. The *total weight of edges* is defined to be

$$E = \sum_{j=1}^J \nu(T'_j)$$

Similarly, the “0-skeleton” $\bigcup \{\text{Vertices}(P_i) \times T_i\}$ can be rewritten as a finite, pairwise disjoint collection of the form $\{V_k \times T''_k \mid k = 1, \dots, K\}$, so that for each k , the point V_k is a vertex of some P_i and $T''_k \subset T_i$. The *total weight of vertices* is defined to be

$$V = \sum_{k=1}^K \nu(T''_k)$$

Finally we define

$$\chi(\Lambda) = V - E + F$$

1.4. Statement of the theorem.

We need to define strongly least area maps. Suppose S is a connected surface and $p: S \rightarrow X$ is an N -normal map. The universal covering map $\pi: \tilde{S} \rightarrow S$ may be composed with p to give an N -normal map $p \circ \pi: \tilde{S} \rightarrow X$. We say that p is *strongly least area* if for every PL subcomplex $D \subset \tilde{S}$ homeomorphic to a 2-disc, the PL disc map $(p \circ \pi \mid D): D \rightarrow X$ is PL least area.

Theorem A. *Let X be a finite, generic 2-complex. Exactly one of the following is true:*

1. X is negatively curved.
2. There exists a compact measured lamination Λ with $\chi(\Lambda) = 0$, and a 3-normal map $f: \Lambda \rightarrow X$ which is strongly least area on each leaf.

One reason strongly least area laminations are nice is the following:

Lemma (No sphere leaves). *If $f: \Lambda \rightarrow X$ is N -normal and strongly least area on each leaf, then no leaf of Λ is a sphere.*

Proof. Suppose there is a sphere leaf L . By N -normality, the leaf L has positive area. Choose a PL closed curve $\gamma \subset L$ bounding a disc $D_1 \subset L$

of less than half the area of L . On the other side, γ bounds another disc D_2 of more than half the area. Both of the induced maps $D_1 \rightarrow X$ and $D_2 \rightarrow X$ are least area, but they have the same boundary and different areas, a contradiction. \square

2. Construction of limiting laminations.

Our method of constructing laminations uses branched surfaces, which are 2-dimensional analogues of train tracks. Just as 1-dimensional laminations are approximated by train tracks, so are 2-dimensional laminations approximated by branched surfaces. In fact, a 2-dimensional lamination can be completely specified by a finer and finer sequence of branched surfaces approximating the lamination. We shall show how a lamination can be constructed by taking the inverse limit of a certain sequence of branched surfaces, in much the same way that a 1-dimensional lamination is the intersection of a sequence of (neighborhoods of) train tracks.

In the next section we collect definitions concerning branched surfaces, and in the following sections we give a general construction for laminations as inverse limits of certain sequences of branched surfaces. The general construction can be applied to any finite 2-complex X , resulting in a lamination Ω called the “universal N -normal lamination” for X . Most of the leaves of the universal N -normal lamination are not useful for analyzing the geometry of X . However, if X is not negatively curved, we shall produce a transverse measure μ on Ω , and the support of μ will be a measured lamination Λ of Euler characteristic zero; in §3 its leaves are shown to be strongly least area.

2.1. Branched Surfaces.

Locally, a branched surface without boundary looks like a stack of pancakes which have been partially glued together. To formalize this, consider a finite, triangulated surface F . An F -stack is a finite 2-complex \mathcal{S} together with a PL map $\phi: \mathcal{S} \rightarrow F$ called the *stack projection*, such that each point of \mathcal{S} is in the image of a PL section of ϕ . We write $\mathcal{S} = F_1 \cup \cdots \cup F_n$ to indicate that F_1, \dots, F_n are all of the sections of ϕ . A *disc stack* is an F -stack where F is a disc.

Now we define a branched surface without boundary using an atlas of disc stacks. Consider a finite 2-complex B . A *branched surface atlas* for B is a finite collection of disc stacks \mathcal{A} satisfying the following properties:

1. For each $\phi: \mathcal{S} \rightarrow F$ in \mathcal{A} we have:
 - \mathcal{S} is a PL subcomplex of B .
 - $\text{int } \mathcal{S} = \phi^{-1}(\text{int } F)$ is an open subset of B .
 - The set $\text{Fr } \mathcal{S} = \mathcal{S} - \text{int } \mathcal{S}$ is the frontier of \mathcal{S} in B .
2. The collection $\{\text{int}(\mathcal{S}) \mid \mathcal{S} \in \mathcal{A}\}$ is an open cover of B .
3. If $\phi_i: \mathcal{S}_i \rightarrow F_i$ is in \mathcal{A} for $i = 1, 2$, and if $x \in \text{int}(\mathcal{S}_1) \cap \text{int}(\mathcal{S}_2)$, then there exists a disc stack $\phi: \mathcal{S} \rightarrow F$ satisfying item 1 and the following:
 - $x \in \text{int}(\mathcal{S})$.
 - $\mathcal{S} \subset \mathcal{S}_1 \cap \mathcal{S}_2$.
 - For each $i = 1, 2$ there exists a PL embedding $f_i: F \rightarrow F_i$ such that $f_i \circ \phi = \phi_i \mid \mathcal{S}$.

In item 3 the stack $\phi: \mathcal{S} \rightarrow F$ is not required to be part of the atlas \mathcal{A} .

A *branched surface without boundary* is defined to be a finite 2-complex B equipped with a branched surface atlas. When B is a branched surface without boundary, a “disc stack of B ” will always mean a disc stack \mathcal{S} that is compatible with \mathcal{A} in the sense that $\mathcal{A} \cup \{\mathcal{S}\}$ is also an atlas. Nevertheless, for certain technical reasons we will usually keep track of the original atlas. If $\mathcal{S} = D_1 \cup \cdots \cup D_n$ is a disc stack of B , each disc D_i is called a *smooth disc* of \mathcal{S} ; also, forgetting which stack the disc came from, D_i is called a smooth disc of B . The set of points where B is not locally homeomorphic to \mathbf{R}^2 is called the *branch locus* of B , denoted ΥB ; this is a subcomplex of the 1-skeleton $B^{(1)}$. The closure of a complementary component of ΥB is called a *sector* of B . Note that a sector is a union of 2-cells.

We shall also use branched surfaces with boundary. The definitions need to be adapted. One allows half-disc stacks as well as disc stacks. Our model F for a half-disc is the intersection of the closed unit disc with the closed upper half plane in \mathbf{R}^2 . The *half-boundary* ∂F is the intersection of F with the x -axis, the frontier $\text{Fr}(F)$ is the intersection with the unit circle, and $\text{int}(F) = F - \text{Fr}(F)$, while the *corners* of the half-disc F are the common endpoints of ∂F and $\text{Fr}(F)$. A *half-disc stack* is a PL map of pairs $(\mathcal{S}, \partial \mathcal{S}) \rightarrow (F, \partial F)$ so that each point of \mathcal{S} is contained in a PL section $F \rightarrow \mathcal{S}$; let $\text{int } \mathcal{S} = \phi^{-1}(\text{int } F)$ and $\text{Fr } \mathcal{S} = \phi^{-1}(\text{Fr}(F))$. A *branched surface with boundary* is a pair of complexes $(B, \partial B)$, with an atlas of disc stacks and half-disc stacks whose interiors form an open cover of B , whose frontiers are the topological frontiers in B , and with compatibility conditions

as above. We shall refer to branched surfaces with boundary simply as “branched surfaces”. Smooth discs and half-discs are defined as above. The branch locus is the set of points where $(B, \partial B)$ is not locally a surface with boundary. Sectors are defined as above.

Given a branched surface B and an arbitrary surface S , a *carrying map* is a PL map $f: (S, \partial S) \rightarrow (B, \partial B)$ such that each point of S is contained in the interior of a disc or half-disc D such that f embeds D as a smooth disc or half-disc in B . Given a compact lamination Λ , a continuous map $f: (\Lambda, \partial \Lambda) \rightarrow (B, \partial B)$ is called a *carrying map* (and Λ is *carried by* B) if the restriction of f to each leaf is a carrying map. If in addition f is surjective then f is a *full* carrying map, and Λ is *fully carried* by B .

Given a carrying map $f: \Lambda \rightarrow B$, if $\phi: \mathcal{S} \rightarrow F$ is a disc stack for B , then $f^{-1}(\mathcal{S})$ may be regarded as a lamination chart of the form $F \times T$, where $T = (\phi \circ f)^{-1}(y)$ for any $y \in F$. Given any $x \in B$, the disc stack and the point y may be chosen so that $x = \phi^{-1}(y)$, and we denote $T_x = T$. If σ is any cell of B , then we denote $T_\sigma = T_y$ for any $y \in \text{int}(\sigma)$. Note that for any σ , the transversal T_σ is both open and closed.

Now we consider the additional structure on a branched surface which is induced by the transverse measure from a measured lamination carried by that branched surface. An *invariant measure* on a branched surface B is a function $\mu: B \rightarrow \mathbf{R}_{\geq 0}$ such that for any disc stack $\phi: \mathcal{S} \rightarrow F$ of B , if we define a function $\mu_{\mathcal{S}}: \text{int}(F) \rightarrow \mathbf{R}_{\geq 0}$ by $\mu_{\mathcal{S}}(x) = \sum_{y \in \phi^{-1}(x)} \mu(y)$, then $\mu_{\mathcal{S}}$ is constant on $\text{int} F$.

Here is the standard method for constructing invariant measures. Start with a *weight vector* on B , which is a function μ assigning to each 2-cell $\sigma \subset B$ a number $\mu(\sigma) \geq 0$. If $x \in \text{int}(\sigma)$ then we may define $\mu(x) = \mu(\sigma)$. Given an edge $E \subset B^{(1)}$ not in ∂B , the 2-cells incident to E may be partitioned into two *sides* denoted $\sigma_1, \dots, \sigma_m$ and $\sigma'_1, \dots, \sigma'_n$, so that any $x \in \text{int}(E)$ is contained in a disc stack $\phi: \mathcal{S} \rightarrow D^2$ where $\mathcal{S} \cap (\sigma_1 \cup \dots \cup \sigma_m) = \phi^{-1}\{x \leq 0\}$ and $\mathcal{S} \cap (\sigma'_1 \cup \dots \cup \sigma'_n) = \phi^{-1}\{x \geq 0\}$. The weight vector μ determines an invariant measure if for any E with sides $\sigma_1, \dots, \sigma_m$ and $\sigma'_1, \dots, \sigma'_n$, we have $\mu(\sigma_1) + \dots + \mu(\sigma_m) = \mu(\sigma'_1) + \dots + \mu(\sigma'_n)$. For any $x \in \text{int}(E)$, we may then define $\mu(x)$ to be the common value of either side of this equation. It is straightforward to check that for any vertex x of B , and for any disc stack $\phi: \mathcal{S} \rightarrow F$ of B whose interior contains x and no other vertex, if $\phi^{-1}(\phi(x)) = \{x\}$ then the value of $\mu_{\mathcal{S}}(y)$ is constant for $y \in \text{int}(F) - \phi(x)$, and so we may define $\mu(x)$ to be this constant value.

Consider $f: \Lambda \rightarrow B$, a measured lamination carried on a branched surface, with transverse measure ν on Λ . There is a unique invariant measure μ on B that is *induced* by ν , in the following sense. Consider a disc stack

$\phi: \mathcal{S} \rightarrow D$ of B . The set $(\phi \circ f)^{-1}(\text{int } D)$ is a lamination chart of the form $\text{int } D \times T_{\mathcal{S}}$ for some open and closed transversal $T_{\mathcal{S}}$ of Λ . We say that μ is induced by ν if $\nu(T_{\mathcal{S}}) = \mu_{\mathcal{S}}(x)$, for any $x \in \text{int}(D)$ and any disc stack $\phi: \mathcal{S} \rightarrow D$ of B .

Given a measured lamination on a branched surface $f: \Lambda \rightarrow B$ with transverse measure ν , we may define a Borel measure on Λ as follows. Suppose that we have a positive Borel measure on B ; for instance, we might impose a measure on each 2-simplex σ of B by modelling it on some Euclidean triangle. In Λ , we have a coordinate chart $\text{int}(\sigma) \times T_{\sigma}$, and there is a Fubini product measure on this chart obtained as the product of the Borel measure on σ and the Borel measure ν on T_{σ} . This induces a Borel measure on all of Λ , called the *mass*, denoted $\text{Mass}(A)$ for a Borel set $A \subset \Lambda$. For any 2-simplex σ and any Borel subset $T' \subset T_{\sigma}$ we have $\text{Mass}(\sigma \times T') = \nu(T')$. Note that Mass is a positive Borel measure on Λ .

2.2. Laminations from refining sequences.

A PL map of branched surfaces $p: B \rightarrow B'$ is called a *submersion* if, for each $x \in B$ and each disc stack $\phi': \mathcal{S}' \rightarrow D'$ of B' with $p(x) \in \text{int}(\mathcal{S}')$, there exists a disc stack $\phi: \mathcal{S} \rightarrow D$ of B with $x \in \text{int}(\mathcal{S})$, and there exists an embedding $f: D \rightarrow D'$, such that $p(\mathcal{S}) \subset \mathcal{S}'$, and $f \circ \phi = \phi' \circ p$.

We shall construct laminations as inverse limits of branched surfaces. Let $p_k: B_k \rightarrow B_{k-1}$, $k = 1, 2, \dots$ be a sequence of submersions of branched surfaces. Let $p_{kj}: B_k \rightarrow B_j$ denote the composition $p_{j+1} \circ \dots \circ p_{k-1} \circ p_k$. The sequence of submersions is said a *refining sequence for the lamination* Λ if $\Lambda = \varprojlim B_k$, where we are taking an inverse limit of topological spaces. Recall the definition of inverse limits. An *inverse sequence* is any sequence $x = (x_k)$ with $x_k \in B_k$ such that $p_k(x_k) = x_{k-1}$. As a topological space, Λ consists of all inverse sequences, with the smallest topology so that the projection maps $f_k: \Lambda \rightarrow B_k$, given by $f_k(x) = x_k$, are continuous. We also require that the inverse sequence respect atlases: there exist atlases \mathcal{A}_k for B_k and a lamination atlas $\varprojlim \mathcal{A}_k$ for Λ such that $p_{kj}^{-1} \mathcal{A}_j = \mathcal{A}_k$ and $f_k^{-1}(\mathcal{A}_k) = \Lambda$. We shall use the abbreviations $B = B_0$ and $f = f_0$.

In general, any lamination carried by a branched surface B is the inverse limit of some refining sequence starting with B . While we shall not prove this general statement, we shall construct a “universal refining sequence”

$$\dots \xrightarrow{p_{k+1}} B_k \xrightarrow{p_k} \dots \xrightarrow{p_2} B_1 \xrightarrow{p_1} B_0 = B$$

whose inverse limit is a “universal lamination” for B . The universal lamination, roughly speaking, is a lamination having a leaf for every possible surface carried by B .

The construction of the universal refining sequence depends on a choice of an atlas $\mathcal{A} = \mathcal{A}_0$ for $B = B_0$. We require that \mathcal{A} is especially nice in that the boundaries of disc stacks are in general position with respect to each other. We choose an order on the disc stacks $\mathcal{A}_0 = \{\mathcal{S}_1, \dots, \mathcal{S}_R\}$. Since the inverse limit will respect this atlas, we can abuse notation by writing \mathcal{S}_r for the pull back stack $p_{k0}^{-1}(\mathcal{S}_r)$.

Now we construct the sequence B_k inductively, B_k being obtained from B_{k-1} by modifying the branched surface in a neighborhood of a disc stack \mathcal{S}_r of the atlas \mathcal{A}_{k-1} . One uses different disc stacks cyclically, so that for constructing B_k from B_{k-1} one uses \mathcal{S}_r where $r \equiv k \pmod{R}$. First we enlarge \mathcal{S}_r in B_{k-1} to obtain a disc stack \mathcal{S}'_r , ensuring that stack boundaries are still in general position when \mathcal{S}_r is replaced by its enlargement. The enlargement should have the property that each smooth disc of \mathcal{S}_r be contained in the interior of a smooth disc of \mathcal{S}'_r . An atlas of enlarged disc stacks $\mathcal{A}'_0 = \{\mathcal{S}'_1, \dots, \mathcal{S}'_R\}$ should be chosen once and for all in B_0 , and the required enlargements $\mathcal{A}'_{k-1} = \{\mathcal{S}'_1, \dots, \mathcal{S}'_R\}$ in B_{k-1} are obtained by pulling back enlargements from B_0 . In B_{k-1} , let $\mathcal{E} = \{E\}$ be the set of smooth discs of \mathcal{S}'_r . Each disc E is included into \mathcal{S}'_r , and hence also into B_{k-1} . We remove \mathcal{S}'_r from B_{k-1} and we replace it by the disjoint union of the discs E , attached to $B_{k-1} - \mathcal{S}'_r$ just as they were before the modification. The resulting branched surface is B_k . The map $p_{k,k-1}$ is the identity in the complement of \mathcal{S}'_r and for each attached disc E , the restriction of $p_{k,k-1}$ to E is induced by the inclusion map of E into B_{k-1} . We say that B_k is obtained from B_{k-1} by *resolving* the disc stack \mathcal{S}_r .

When \mathcal{S}_r is resolved, clearly any adjacent disc stack \mathcal{S}_i has also been modified. We would like to obtain an atlas \mathcal{A}_k for B_k with disc stacks $p_{k,k-1}^{-1}(\mathcal{S}_i)$. However, we have no assurance that these subsets of B_k are always disc stacks: the stack projection $\mathcal{S}_i \rightarrow D_i$ induces by composition a projection $p_{k,k-1}^{-1}(\mathcal{S}_i) \rightarrow D_i$ whose domain may not be covered by sections. We fix this problem by trimming away portions of B_k . The problem can be stated another way: the branched surface $p_{k,k-1}^{-1}(\mathcal{S}_i)$ may not fully carry any lamination, so that B_k also does not fully carry any lamination. Obviously, we are only interested in the maximal sub-branched surface of B_k which does fully carry at least one lamination. Therefore we replace B_k by this sub-branched surface. For a more algorithmic procedure, carry out the following inductive process for trimming away more and more of B_k . Letting i cycle through the index set $\{1, \dots, R\}$, repeat the following operation until it

stabilizes: remove any sector of B_k contained in $p_{k,k-1}^{-1}(\mathcal{S}_i)$ that does not lie in a section of the projection $p_{k,k-1}^{-1}(\mathcal{S}_i) \rightarrow D_i$. This process must stop since there are only finitely many sectors in B_k . Either way that B_k is trimmed, the pull back of each \mathcal{S}_i under the map $p_{k,k-1}$ yields a disc stack in B_k also denoted \mathcal{S}_i , and we obtain the atlas \mathcal{A}_k .

Theorem (Resolution of stacks yields refining sequence). *The sequence*

$$\dots \xrightarrow{p_{k+1}} B_k \xrightarrow{p_k} \dots \xrightarrow{p_2} B_1 \xrightarrow{p_1} B_0 = B$$

constructed above is a refining sequence, and thus $\Lambda = \varprojlim B_k$ is a compact, topological lamination carried by each B_k , and the natural projections $f_k: \Lambda \rightarrow B_k$ are carrying maps. Further one can choose atlases for the B_k 's and for Λ which are respected by the sequence.

Proof. Define $\Lambda = \varprojlim B_k$. At this point we do not know that Λ is a lamination, though we can say what the local transversals look like. Given a disc stack $\phi: \mathcal{S} \rightarrow D$ in the atlas \mathcal{A}_0 , denote $\phi_k = \phi \circ p_{k0}$, which is the stack projection of the disc stack $p_{k0}^{-1}(\mathcal{S})$ in the atlas \mathcal{A}_k . For any $y \in \text{int}(D)$, $\phi_k^{-1}(y)$ is a finite set T_{yk} of points in B_k all projecting to y , and the natural maps $T_{yk} \rightarrow T_{y,k-1}$ form an inverse system. The inverse limit $T_y = \varprojlim T_{yk}$ is the space whose points are inverse sequences $\{x_k \mid k \geq 1, x_k \in T_{yk}, \text{ and } p_{kj}(x_{ki}) = x_{ji}\}$. The generating sets for the topology of T_y are just the inverse images of points $x_k \in T_{yk}$ under the projection maps $T_y \rightarrow T_{yk}$. Ultimately, the space $T_y = (\phi \circ f)^{-1}(y)$ will be a local transversal for the lamination Λ . However, to prove Λ is a lamination we must get an expression for a local transversal which does not depend on the choice of $y \in \text{int}(D)$, but only on the disc stack \mathcal{S} .

To show $f: \Lambda \rightarrow B$ is a lamination carried by B , we use the atlases \mathcal{A}_j for the branched surfaces of the refining sequences. We shall construct a lamination atlas for Λ , with one lamination chart for every disc stack in \mathcal{A}_0 . Consider a stack $\mathcal{S} \in \mathcal{A}_0$, with stack projection $\phi: \mathcal{S} \rightarrow D$. We shall show that $f_0^{-1}(\text{int } \mathcal{S})$ is homeomorphic to $\text{int } D \times T$ for some compact topological space T . In what follows, we index disc stacks according to the branched surfaces they live in, so that $\mathcal{S}_k = p_{k0}^{-1}(\mathcal{S})$.

Let \mathcal{E}_k be the set of all smooth discs of \mathcal{S}_k . By construction, there is a subsequence B_{k_i} of the branched surfaces such that \mathcal{S}_{k_i} is the disjoint union of the discs of \mathcal{E}_{k_i} . More explicitly, if \mathcal{S} is the r -th disc stack of the enumeration of \mathcal{A} , then we can take $k_i = r + Ri$. The transversal $T_{y_{k_i}}$ is

constant throughout D , say $T_{y_{k_i}} = T_{k_i}$. It follows that the inverse limit transversal T_y is constant over D . Let us say $T_y = T$ for all y in D . This proves that the set $f_0^{-1}(\mathcal{S})$ in Λ has the form $D \times T$ as a set. In fact, by considering the subsequence \mathcal{S}_{k_i} , it is clear that $f_0^{-1}(\mathcal{S}) = \varprojlim \mathcal{S}_{k_i} = \varprojlim D \times T_{k_i} = D \times \varprojlim T_{k_i} = D \times T$. A similar argument shows that this defines an atlas for Λ . \square

Observe that the lamination constructed above can be given a simplicial structure. This follows from the fact that each B_n may be triangulated so that each submersion $p_n: B_n \rightarrow B_{n-1}$ is a simplicial map, not just a PL map. To see why this can be done, note that one may replace the original triangulation of B_n by an iterated barycentric subdivision, so that each disc stack in \mathcal{A}_0 and each enlarged disc stack in \mathcal{A}'_0 is a subcomplex. Once this is done, there is a triangulation on B_n defined by induction so that p_n maps each simplex of B_n homeomorphically to a simplex of B_{n-1} , hence p_n is simplicial for every n .

We call the refining sequence constructed above a *universal refining sequence* and we call the inverse limit a *universal lamination*. These names are justified by the following proposition:

Universality Proposition. *Let B_n be the universal refining sequence constructed from (B_0, \mathcal{A}_0) . Given a surface F and a carrying map $g_0: F \rightarrow B_0$, there exists a unique compatible sequence of carrying maps $g_n: F \rightarrow B_n$, in the sense that $p_{nm} \circ g_m = g_n$ for all $n > m$.*

Proof. Given n , suppose the sequence has been extended up to g_n . Let B_{n+1} be obtained from B_n by resolving discs of the disc stack \mathcal{S} . If D is a component of $(g_n)^{-1}(\mathcal{S})$, then $g_n(D)$ is a section of \mathcal{S} , so g_{n+1} is forced to take D to the unique smooth disc of B_{n+1} which projects to $g_n(D)$. The definition of g_{n+1} on the rest of F is also forced, since the map $B_{n+1} \rightarrow B_n$ is 1-1 except over the enlarged disc stack whose smooth discs are resolved in the modification of B_n to B_{n+1} . \square

It follows from this that the universal lamination constructed as an inverse limit is actually universal:

Corollary (Universality of the inverse limit). *If B is a branched surface and $f: \Omega \rightarrow B$ is a universal lamination on B , then for any lamination $g: \Lambda \rightarrow B$ carried on B , there exists a unique continuous map $q: \Lambda \rightarrow \Omega$*

respecting carrying maps, i.e. $f \circ q = g$. In particular, if Ω' is another universal lamination, then there is a unique homeomorphism $q: \Omega' \rightarrow \Omega$ such that $f \circ q = f'$

Proof. Let B_n be a universal refining sequence starting at $B = B_0$, for which $\Omega = \varprojlim B_n$, and let $f_n: \Omega \rightarrow B_n$ be the carrying maps. By the Universality Proposition, the carrying map $g = g_0: \Lambda \rightarrow B$ lifts to a compatible sequence of carrying maps $g_i: \Lambda \rightarrow B_n$, hence passing to the inverse limit we obtain a continuous map q respecting carrying maps. To prove uniqueness, if $q_1: \Lambda \rightarrow \Omega$ is any continuous map respecting carrying maps, then $f_n \circ q_1$ is a compatible sequence of carrying maps from Λ to B lifting g , hence $g_n = f_n \circ q_1$ so $q = q_1$.

If Ω and Ω' are universal, we obtain continuous maps in either direction respecting carrying maps, and by uniqueness the compositions must be identity maps. \square

Here are some remarks about the universal lamination.

It may happen that the resolution process applied to a branched surface B eventually produces an empty branched surface, i.e. the universal lamination is empty. This happens if and only if B carries no lamination. Because the resolution process can be carried out algorithmically, there is an algorithm that takes a branched surface as input and stops if and only if the branched surface does not carry any lamination. In other words, the set of branched surfaces carrying no lamination is recursively enumerable. However, this set is not recursive; see [Gab95].

There is one more comment worth making about the universal lamination. Given any lamination, one can modify it by adding a cover of a leaf, by adding parallel copies of a leaf, or by replacing a leaf by one of its own covers. When we speak of “covers of leaves” we mean covers compatible with carrying maps—for example if $f: F \rightarrow B$ and $g: G \rightarrow B$ are carrying maps from surfaces F and G to a branched surface B , then f covers g if there is a covering map $q: F \rightarrow G$ satisfying $g \circ q = f$. The universal lamination has the property that if it contains a certain leaf, then it does not contain a cover of that leaf, nor does it contain parallel copies of the leaf.

2.3. The splitting progress lemma.

The lemma described in this section will be used in proving that a limiting lamination is strongly least area. The proof is given here because it

follows naturally after the developments of the previous section.

Consider a universal refining sequence $\dots \rightarrow B_i \rightarrow \dots \rightarrow B_1 \rightarrow B_0$ with universal lamination $\Omega = \varprojlim B_i$. Consider also two leaves F, G of Ω , so we have two sequences of compatible carrying maps $f_i: F \rightarrow B_i$ and $g_i: G \rightarrow B_i$. We are assuming, as we may, that each submersion $B_i \rightarrow B_{i-1}$ is a simplicial map, hence F and G may be triangulated (tiled) so that the maps f_i, g_i are simultaneously simplicial. We wish to analyze how f_i and g_i can be parallel in B_i over respective subtilings of F and G , and how parallelism can fail on larger subtilings. Further, we wish to analyze how the “region of parallelism” shrinks as i increases and f_i is split apart from g_i . Our goal is to show that splitting progresses at a uniform rate.

We must say what types of “subtilings” will be needed. Suppose C is a *surface-like* 2-complex, with no more than two 2-cells incident to each edge; let ∂C be the union of edges incident to a unique 2-cell. We measure combinatorial distance between tiles of C as the minimum number of tiles a dual path from one tile to another must intersect, and we assume C has a connected dual 1-complex so any two tiles in C have finite combinatorial distance. Suppose $\sigma: C \rightarrow F, \nu: C \rightarrow G$ are simplicial local embeddings. By composition we obtain “carrying maps” $f_i \circ \sigma$ and $g_i \circ \nu$ from C to B_i . We say that σ and ν are *parallel* in B_i if $f_i \circ \sigma = g_i \circ \nu$. When this is so, let $\partial_d C$ be the portion of ∂C where the maps f_i and g_i “split apart” or “diverge”. That is, $\partial_d C$ is the union of 1-cells $E \subset \partial C$ across which the maps σ and ν cannot be extended without destroying parallelism in B_i : if $s \subset F - \sigma(C)$ and $t \subset G - \nu(C)$ are the 2-cells adjacent to $\sigma(E)$ and $\nu(E)$ respectively, then $f_i(s) \neq g_i(t)$.

The following lemma says that if each tile of C is within a bounded distance r of $\partial_d C$, then in some B_j with $j > i$, the sets $f_j(\sigma(C))$ and $g_j(\nu(C))$ are disjoint. Moreover, given $n = j - i$, disjointness in B_j is guaranteed for some r depending only on n .

Lemma (Splitting Progress). *Let B_i be a universal refining sequence. For each integer $n \geq 1$ there exists $r \geq 1$ with the following property. Suppose $F, G, \{f_i: F \rightarrow B_i\}, \{g_i: G \rightarrow B_i\}, C, \sigma: C \rightarrow F, \nu: C \rightarrow G$ are given as above. Given i , suppose that σ and ν are parallel in B_i . Suppose furthermore that each tile of C is within combinatorial distance r of $\partial_d C$. It follows that $f_{i+n}(\sigma(C))$ and $f_{i+n}(\nu(C))$ are disjoint in B_{i+n} .*

Proof. Suppose that the refining sequence is defined using an atlas \mathcal{A}_0 for B_0 containing R charts, yielding atlases \mathcal{A}_i for B_i . Any tile (2-simplex)

$\tau \subset C$ is contained in a disc D such that $\sigma(D)$ and $\nu(D)$ are mapped by f_i and g_i to the same smooth disc of \mathcal{A}_i . But if τ is adjacent to $\partial_d C$, then in B_{i+R} the discs $\sigma(D)$ and $\nu(D)$ are not mapped to the same smooth discs of atlas \mathcal{A}_{i+R} . It follows that in B_{i+2R} , the discs $\sigma(D)$ and $\nu(D)$ are mapped to disjoint smooth discs of \mathcal{A}_{i+2R} . By induction, any two tiles of C within combinatorial distance n of $\partial_d C$ are mapped disjointly into B_{i+2Rn} . \square

2.4. Normal Surfaces in a 2-complex.

We have shown how to pass from a single branched surface to a refining sequence and then to a lamination. To construct a lamination on a 2-complex X , we shall construct a canonical branched surface for X , by using an adaptation of the normal surface theory of Haken, see [Hak61], [Sch61]. Haken's normal surface theory was designed to analyze embedded surfaces in 3-manifolds. Our adaptation is designed to analyze surfaces mapped to a 2-complex X . For the moment, we do not assume that X is generic. That assumption will be made at the appropriate time.

Normal tiles. A *handle structure* on X is defined as follows. Consider the second barycentric subdivision $\Sigma^2 X$. For each 0-cell $v \in X$, the *0-handle with core v* is the regular neighborhood of v , i.e. the union of all cells of $\Sigma^2 X$ touching v . This is homeomorphic to the cone on the graph $\text{Link}(v)$. For each 1-cell $\alpha \subset X$, the *1-handle with core α* is the union of the closures of all cells of $\Sigma^2 X$ in $X - \{0\text{-handles}\}$ touching α . This is homeomorphic to $I \times (\text{cone on finitely many points})$. Finally, for each 2-cell $\sigma \subset X$, the *2-handle with core σ* is the closure of the 2-disc component of $X - \{0\text{-handles}\} \cup \{1\text{-handles}\}$ contained in $\text{int}(\sigma)$. Note that the core of a 1-handle or 2-handle is not a subset of the handle.

The frontier of a handle H is a graph K . Given an integer $N > 1$, we define an *N -normal cycle* in K to be a closed edge path γ in K , with the following restrictions. If H is a 0-handle, then γ passes through each edge of K at most once. If H is a 1-handle, then γ is embedded. And if H is a 2-handle, so K is a circle, then γ is an n -fold covering map of K for some $n \leq N$. The cone on γ to the barycenter of the core of H is a PL disc map, and any equivalent PL disc map is called an *N -normal tile* in H with frontier γ . Two equivalent N -normal tiles are said to belong to the same *tile type*. The most important fact about tile types is that for fixed N , there are only finitely many N -normal tile types.

We also define *half-tile types*. Given a handle H with frontier K , an edge path γ in K is called a *normal path* if it is not closed, and if the following conditions hold. If H is a 0-handle then γ passes over each edge of K at most once. If H is a 1-handle, then γ is embedded, and the endpoints of γ lie on the core of H . There are no normal paths in the frontier of a 2-handle. The cone on $(\gamma, \partial\gamma)$ gives a half-disc map $(D, \partial D) \rightarrow (X, X^{(1)})$ called a *normal half-tile* in H , with frontier γ . The endpoints $\partial\gamma$ are the image of the corners of the half-disc $(D, \partial D)$. There are only finitely many half-tile types, none in any 2-handle. When $D \rightarrow X$ is a normal tile, let ∂D denote the emptyset.

An N -normal surface is a PL map $f: (S, \partial S) \rightarrow (X, X^{(1)})$ for which the inverse image of each handle falls into components C , so that if $C \cap \partial S = \emptyset$ then $f|_C$ is an N -normal tile, and otherwise $f|_{(C, C \cap \partial S)}$ is a normal half-tile. Notice that this agrees with the earlier definition of N -normal laminations on X , specialized to a lamination with discrete transversals.

The unadorned term “normal” is a synonym for “1-normal”, and for any $N \geq 1$ the term “ N -normal half tile” is a synonym for “normal half-tile”.

The universal normal branched surface. Normal tiles and half-tiles are used to construct a branched surface B associated with each complex X , called the *universal normal branched surface* of X . Later we will construct an auxiliary complex \hat{X}_N and use it to define the *universal N -normal branched surface* of X .

We shall construct the universal normal branched surface B together with a map $b: (B, \partial B) \rightarrow (X, X^{(1)})$. To construct B , start with a disjoint union of 2-discs $\amalg D$, which are the domains of normal tiles and half-tiles mapped into X , one belonging to each tile type and half-tile type in each handle. The normal tile maps $d: D \rightarrow X$ yield a map $b: \amalg D \rightarrow X$. We shall declare certain frontier gluings among the collection $\amalg D$, and the resulting quotient space is B . The quotient of $\amalg \partial D$ will form ∂B . The gluings will respect the map $b: \amalg D \rightarrow X$, which therefore descends to a map $b: (B, \partial B) \rightarrow (X, X^{(1)})$.

Consider two normal tile or half-tile maps $d: D \rightarrow H$, $d': D' \rightarrow H'$, where H, H' are adjacent handles of distinct indices. Consider an embedded arc $\alpha \subset H \cap H'$, with the following restrictions:

1. $d^{-1}(\alpha)$ and $(d')^{-1}(\alpha)$ both map homeomorphically to α .
2. If one of H or H' is 2-handle then $\alpha = H \cap H'$.

3. If one is a 0-handle and the other a 1-handle, so $H \cap H'$ is a star centered on $\xi = H \cap H' \cap X^{(1)}$, then either α connects two endpoints of the star or α connects an endpoint to ξ ; furthermore, if $\xi \in \partial\alpha$ then $d^{-1}(\xi) \in \partial D$ and $(d')^{-1}(\xi) \in \partial D'$.

Then we declare that $d^{-1}(\alpha)$ and $(d')^{-1}(\alpha)$ are to be glued, respecting the homeomorphisms to α .

We give B the structure of a simplicial complex, by pulling back cells of $\Sigma^2 X$, so b is a simplicial map from B to $\Sigma^2 X$. To put a branched surface structure on B , we need an atlas. There will be one disc stack $\mathcal{S}(D)$ for each normal tile or half-tile D , and the stack projection will be the restriction of b to $\mathcal{S}(D)$. In order for disc stack boundaries to be in general position, the base of $\mathcal{S}(D)$ is a slightly modified version of D , as shown in Figure 2. The disc stacks will be components of inverse images under b of modified handles \hat{H} associated to handles H ; there will be one component of $b^{-1}(\hat{H})$ for each normal tile or half-tile in H . For $k = 2$, there is only one tile type in H^2 , hence we can let \hat{H}^2 be a regular neighborhood of H^2 in X , a slightly wider handle. The corresponding stack $b^{-1}(\hat{H}^2)$ is connected and generally has many smooth discs. At the other extreme, for $k = 0$, there are many tile types D , and the corresponding stacks are chosen to be the components of $b^{-1}(\hat{H}^0)$, where \hat{H}^0 is a slightly smaller handle than H^0 , obtained from H^0 by removing a very small regular neighborhood of ∂H^0 . Each component of $b^{-1}(\hat{H}^0)$ consists of a single smooth disc. Finally, if $k = 1$ then \hat{H}^1 is chosen to be “longer” and “narrower” than the 1-handle H^1 : for each 0-handle H^0 adjacent to H^1 , add to H^1 a regular neighborhood of the intersection with H^0 , and then for each adjacent 2-handle H^2 remove a very small regular neighborhood of the intersection with H^2 . This clearly defines a branched surface atlas with stack boundaries in general position.

We call B the *universal normal branched surface* for X , a terminology which is justified by the following obvious statement:

Proposition. *Given a surface F and a PL map $f: (F, \partial F) \rightarrow (X, X^{(1)})$, the map f is normal if and only if there exists a carrying map $\hat{f}: (F, \partial F) \rightarrow (B, \partial B)$ such that $b \circ \hat{f} = f$. Moreover, if \hat{f} exists then it is unique. \square*

If $p: \Lambda \rightarrow B$ is a universal lamination for B , then the composition $b \circ p: \Lambda \rightarrow X$ can therefore be regarded as a *universal normal lamination* on X . Any two universal normal laminations on X are homeomorphic by a homeomorphism which respects the maps to X .

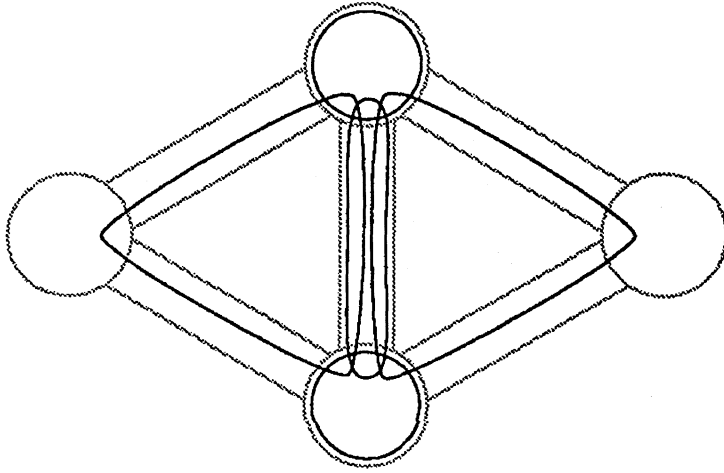


Figure 2: Base discs of a branched surface atlas

Next we define the universal N -normal branched surface B_N of X . Define a complex \hat{X}_N obtained from X as follows. For each 2-cell $\sigma \subset X$ and each integer n with $2 \leq n \leq N$, attach an additional 2-cell σ_n to X by using, as the attaching map, the n -fold cover of the attaching map of σ . There is a natural map $q: \hat{X}_N \rightarrow X$ which maps σ_n to σ by an n -fold cyclic covering map branched over the barycenter of σ . Pull back the handles of X to obtain a handle decomposition of \hat{X}_N , which is used to define normal tiles and half-tiles in \hat{X}_N . A normal tile or half-tile in \hat{X}_N is *good* if it projects to an N -normal tile or half-tile in X . Now construct a branched surface $\hat{b}: B_N \rightarrow \hat{X}_N$ following the procedure described above, but using only good normal tiles and half tiles. We therefore obtain a map

$$b: B \xrightarrow{\hat{b}} \hat{X}_N \xrightarrow{q} X$$

This is the universal N -normal branched surface of X , because:

Proposition. *Given a surface F and a PL map*

$$f: (F, \partial F) \rightarrow (X, X^{(1)}),$$

the map f is N -normal if and only if there exists a carrying map

$$\hat{f}: (F, \partial F) \rightarrow (B_N, \partial B_N)$$

such that $b \circ \hat{f} = f$. Moreover, if \hat{f} exists then it is unique. \square

2.5. Making disc maps normal.

We shall apply normal surface theory to study least area disc maps in X , in order to understand the absence of negative curvature. Starting with a fat disc sequence in X , we would like to lift it to a sequence of carrying maps in the universal N -normal branched surface B , for some uniform value of N , yielding a sequence of leaves in the universal N -normal lamination Ω . Then we can pass to a limit using compactness of Ω . However, a disc map $f: D \rightarrow X$ need not be N -normal, and so cannot be lifted to B in general. The point is that N -normality is a kind of “precompactness” condition on a set of disc maps, and there is no guarantee that a fat disc sequence is precompact (one way to make this precise is using Gromov-Hausdorff convergence: a sequence of fat discs is precompact in the Gromov-Hausdorff sense if and only if the combinatorics are locally bounded). The following result resolves this problem, by allowing an arbitrary fat disc sequence to be replaced by one which is “precompact”. Recall that for a PL disc map $f: D \rightarrow X$, the isoperimetric ratio is defined by $\mathcal{I}(f) = \text{Area}(f)/\text{Length}(\partial f)$.

Normal Disc Theorem. *Let X be a generic 2-complex. Given a least area disc map $f: D \rightarrow X$, there exists a 3-normal disc map $f': D' \rightarrow X$ such that f' is PL least area and $\mathcal{I}(f') \geq \mathcal{I}(f)$.*

Corollary. *If X is not negatively curved, then there exists a 3-normal, PL least area fat disc sequence $d_j: D_j \rightarrow X$.* \square

Remark. Simple examples show that the statement of the Normal Disc Theorem cannot be improved by requiring the resulting fat disc to be a local embedding. For the first example, build X from a torus T by taking an embedding $\alpha: [-1, 1] \rightarrow T$ and identifying $\alpha(t)$ with $\alpha(-t)$ for all $t \in [0, 1]$. The complex X is homotopy equivalent to a torus and hence is not negatively curved, but any fat disc sequence for X has discs which are not local embeddings over 0-handles. For another example, build X from a disc D and a one-holed torus T by gluing ∂T to ∂D via a 2-1 covering map $\partial T \rightarrow \partial D$. Again X is homotopy equivalent to a torus, but any fat disc sequence for X has discs which are not 1-normal. These also provide counterexamples to stronger versions of Theorem A which require the lamination to have locally embedded or normal leaves. On the other hand, we do not know

if the Normal Disc Theorem or Theorem A can be improved by replacing “3-normal” with “2-normal”.

Here is an outline of the proof of the Normal Disc Theorem. We shall perform alterations on f to make it more and more normal, without ever increasing area. In the N -Normal Disc Lemma we shall make f normal over 2-handles, then over 1-handles. Then we shall try to make f normal over 0-handles, but there is an obstruction. This obstruction may be pushed into the 2-handles, producing an N -normal disc map. Having proved the N -Normal Disc Lemma, we shall then show how N may be reduced to 3.

We begin the proof of the Normal Disc Theorem with a lemma which shows that PL and simplicial least area are closely related:

Simplicial least area lemma. *Given a PL disc map*

$$f: (D, \partial D) \rightarrow (X, X^{(1)})$$

with simplicial boundary, there exists a simplicial disc map

$$f': D' \rightarrow X$$

with equivalent boundary such that $\text{Area}(f') \leq \text{Area}(f)$.

Proof. We shall alter f until it is simplicial, without ever changing ∂f or increasing $\text{Area}(f)$.

Consider a 2-simplex $\sigma \subset X$. Let C be the closure of a component of $f^{-1}(\text{int } \sigma)$, and let $f_C = f|_C$. As $y \in \text{int}(\sigma)$ ranges over the regular values of f_C , let M be the minimum value of $|f_C^{-1}(y)|$, and let y be a regular value at which that minimum is achieved. Recalling that $\text{Area}(\sigma) = 1$, it follows that $\text{Area}(f_C) \geq M$. Choose a neighborhood U of y so that each component of $f_C^{-1}(U)$ maps homeomorphically to U . Now homotop f by stretching U over all of σ ; formally, compose with an ambient homotopy of X that is fixed outside σ and that crushes the annulus $\sigma - U$ onto $\partial\sigma$, taking U homeomorphically to σ . This procedure does not change f on $D - \text{int}(C)$, so in particular ∂f is unchanged, and the area is reduced by $\text{Area}(f_C) - M$. Doing this for each 2-simplex, if W is the union of all open 2-simplices of X then $\text{cl}(f^{-1}(W))$ is a disjoint union of closed 2-cells $\{C\}$, each disjoint from ∂D , each mapping homeomorphically to a 2-simplex; all further homotopies of f will be fixed on these 2-cells.

Now consider a 1-simplex $E \subset X$. Choose a regular value $z \in \text{int } E$ for the restriction of f to $\text{cl}(f^{-1}(\text{int}(E)))$. After the alterations of the previous

paragraph, no component of $f^{-1}(z)$ is a point, because the 2-cells $\{C\}$ are pairwise disjoint and disjoint from ∂D . Choose a neighborhood $V \subset E$ of z so that $f^{-1}(V)$ is a Cartesian product over V . Now homotop f by stretching this region out over all of E , so that $\text{cl}(f^{-1}(\text{int } E))$ is a Cartesian product over E , and make sure that the homotopy is stationary on ∂D . This procedure does not change $\text{Area}(f)$. The triangulation of ∂D may now be extended over all of D so that f is simplicial. \square

Corollary. *Any least area disc map is PL least area.* \square

Henceforth we use the unmodified term “least area” regardless of whether we are speaking about simplicial or PL disc maps.

***N*-Normal Disc Lemma.** *Given a constant C , consider the collection \mathcal{A}_C of all least area PL disc maps into X with simplicial boundary whose isoperimetric ratio is at least C . If \mathcal{A}_C is nonempty, then it has a member of minimal boundary length which is N -normal for some N .*

Proof. By the Simplicial Least Area Lemma, we may assume that \mathcal{A}_C has a member $f: D \rightarrow X$ with $\text{Length}(\partial f)$ minimal, so that f is simplicial. In particular f is normal over 2-handles, although this property will be destroyed later on.

Note first that ∂f cannot immediately retrace any edge of X . Otherwise, there is a segment of ∂D of the form $E_1 * J * E_2$, where E_1 and E_2 are mapped to the same edge E of X and J is mapped to an endpoint v of E . It follows that f may be homotoped to reduce $\text{Length}(\partial f)$ by 2, without changing $\text{Area}(f)$, by homotoping $E_1 * J * E_2$ to the endpoint of E opposite v . The homotoped map f' is least area, for if f' could be altered to reduce area without changing $\partial f'$, then $\partial f'$ could be pushed back out through E to give a map with the same boundary as f but with smaller area, contradicting that f is least area. Also, f' has larger isoperimetric ratio than f , and smaller boundary length. But now we have contradicted the choice of f .

Now we consider 1-handles. The 1-handle H corresponding to an edge E retracts to a subinterval of E , so that the retraction map is a fibration whose fiber is a star. The *cocore* of H is the fiber running through the center of E , denoted $\text{cocore}(H)$. Since f is already transverse to the endpoints of $\text{cocore}(H)$ and ∂f is simplicial, we can homotop f , using a homotopy with support in the 1-handle, to make f transverse to the cocore, without changing area or boundary length. Consider $f^{-1}(\text{cocore}(H))$. This is a

collection of closed curves and arcs. If there is a closed curve c , redefine f on the disc bounded by c to map that disc into $\text{cocore}(H)$, then homotop f slightly so the image of the disc is disjoint from $\text{cocore}(H)$, eliminating c as a component of $f^{-1}(\text{cocore}(H))$. Notice that before redefining f on the disc bounded by c , the area of that disc had to be 0, otherwise the new f would have smaller area than the old, a contradiction to least area.

At this point we know that $f^{-1}(\text{cocore}(H))$ contains only arcs. Among these arcs there might be an undesirable arc α with both endpoints mapped to the same vertex of the cocore. Either both ends of α are on ∂D and are mapped to the center of the cocore, or both ends of α are mapped to the same extreme vertex of the cocore. We show that neither case can happen.

Consider first the case where the ends of α are mapped to the same extreme vertex, lying on a 2-handle H' with core 2-cell σ . Let C_1, C_2 be the components of $f^{-1}(H')$ containing the endpoints of α . Now f may be homotoped to reduce area, using a homotopy supported on a neighborhood of $C_1 \cup C_2 \cup \alpha$. This contradicts that f is least area. Intuitively, f folds two 2-cells of D over the 2-simplex σ , with the fold occurring along E , so the fold may be pushed over the 2-simplex to reduce the area by 2.

Now consider the case where α is properly embedded in D and $\partial\alpha$ is mapped to the central point of $\text{cocore}(H)$. By a homotopy of f supported in a neighborhood of α , we may assume that f maps all of α to the central point of $\text{cocore}(H)$, without changing $\text{Area}(f)$ or ∂f . Cutting D along α , we obtain two PL disc maps $g_n: D_n \rightarrow X$, $n = 1, 2$, with surgery arcs $\alpha_n \subset \partial D_n$ that are identified with α . Now homotope g_n to push the surgery arc α_n off $\text{cocore}(H)$ and into an adjacent vertex, so that ∂g_n becomes simplicial. Clearly $\text{Length}(\partial g_1) + \text{Length}(\partial g_2) + 2 = \text{Length}(\partial f)$ and $\text{Area}(g_1) + \text{Area}(g_2) = \text{Area}(f)$. It follows that either $\mathcal{I}(g_1) \geq \mathcal{I}(f)$ or $\mathcal{I}(g_2) \geq \mathcal{I}(f)$. (This uses the fact that $(a_1 + a_2)/(b_1 + b_2)$ is between a_1/b_1 and a_2/b_2 if all the numbers are positive). Moreover, g_1, g_2 are least area, for if either area could be reduced preserving the boundary, then the new maps could be reglued along α , resulting in an area reduction of f preserving ∂f , a contradiction. It follows that one of the PL least area disc maps g_1 and g_2 has an isoperimetric ratio no smaller than that of f , and the length of its boundary is shorter than that of f . This contradicts our choice of f . It follows that no undesirable arcs can occur.

We have shown that every component of $f^{-1}(\text{cocore}(H))$ is an arc $\alpha \subset \text{int}(D)$ whose endpoints are mapped to distinct vertices of $\text{cocore}(H)$. To make f normal in H , homotop f to stretch a small neighborhood of $\text{cocore}(H)$ over the entire length of the 1-handle. The homotopy can be done relative to the boundary and the adjacent 2-handles. Because we have

excluded the undesirable arcs above, it is now easy to see that the components of $f^{-1}(H)$ give normal tiles and half-tiles mapped to H .

Now we consider 0-handles. Let H be a 0-handle with core vertex $v \in X$. Letting $K = \text{Fr } H$, note that f is already transverse to a neighborhood $N(K) \subset H$. If ∂f maps the endpoints of an arc of $(\partial f)^{-1}(H)$ to the same vertex of the graph K , then ∂f has an immediate retracing of an edge of X , which cannot happen. All arcs of $(\partial f)^{-1}(H)$ are therefore mapped to paths joining different vertices of K , and after a homotopy of f supported on $f^{-1}(H - N(K))$, the paths will all be embeddings. Now consider possibilities for $f^{-1}(H)$. Since f is already transverse to $N(K)$, each component U of $f^{-1}(H)$ is a surface with boundary. A priori several bad things could happen: U could be a disc with holes; or $U \cap \partial D$ could be disconnected. Also, by our construction so far the components of $U \cap f^{-1}(K)$ are mapped in relatively nicely, in that they do not have any immediate retracings of edges of the graph K ; nonetheless, they may traverse the same edge of K more than once.

Suppose now that one of the components, U , of $f^{-1}(H)$ is a disc with holes, having a circle boundary component α in the interior of D and bounding a disc E there. The disc E must include some normal tiles in 2-handles adjacent to H , and so E contributes a positive amount to the area of f . Redefining f on $\text{int}(E)$ so that it is mapped into H , we reduce $\text{Area}(f)$ without changing ∂f , contradicting that f is least area. We conclude that every component U of the surface $f^{-1}(H)$ has no holes.

Now suppose that there is a disc component U of $f^{-1}(H)$ having more than one arc of ∂D on its boundary. This means that U is cut from D by at least two arcs of $f^{-1}(\partial H)$. Suppose γ and δ are two arc components of $U \cap \partial D$. Choose an arc $\alpha \subset U$ joining points in the interiors of γ and δ , so that f maps α to the center of H ; such an arc α must exist, since $f(\gamma)$ and $f(\delta)$ each hit the center of H . Cutting along α produces two PL disc maps $g_n: D_n \rightarrow X$, $n = 1, 2$. Now we repeat the analysis used in 1-handles. We have $\text{Area}(f) = \text{Area}(g_1) + \text{Area}(g_2)$ and $\text{Length}(\partial f) = \text{Length}(\partial g_1) + \text{Length}(\partial g_2)$. It follows that $\mathcal{I}(g_1) \geq \mathcal{I}(f)$ or $\mathcal{I}(g_2) \geq \mathcal{I}(f)$. Also, since f is least area, so are g_1 and g_2 . Also, $\text{Length}(g_1)$ and $\text{Length}(g_2)$ are strictly less than $\text{Length}(f)$. But then we can replace f by g_1 or g_2 , whichever has the larger isoperimetric ratio, contradicting our original choice of f . We conclude that all components of $f^{-1}(H)$ must be discs in the interior of D , or half-discs intersecting ∂D in an arc.

The remaining problem is that for a disc or half-disc E of $f^{-1}(H)$, the set $\text{Fr } E = f^{-1}(K)$, the frontier of E relative to D , might traverse an edge of K more than once. First suppose that $\text{Fr } E$ traverses the edge $r \subset K$

two times in opposite directions, and choose an arc α in E joining two such preimages of r . In fact, we can choose α with endpoints in adjacent 2-handle normal tiles of D which are mapped to the same 2-handle H' of X , but with opposite orientation. Clearly there is a homotopy pushing α into H' , and then a further homotopy reducing area by 2, contradicting that f is least area. We conclude that $\text{Fr } E$ cannot traverse an edge of K twice in opposite directions.

However, it may be possible for $\text{Fr } E$ to traverse the same edge many times in the same direction. Before continuing, replace the map $f \mid E$ with the cone on the map $f \mid \text{Fr } E$, and so E contains a unique point c mapped to the core vertex v ; when E is a half-disc, we have $c \in \partial E$. Let r be an edge of K which is traversed by $m \geq 2$ edges of $\text{Fr } E$. The circle or arc $\text{Fr } E$ may now be written as $\alpha_0 * r_1 * \alpha_1 * r_2 * \alpha_2 * \cdots * r_m * \alpha_m$, where r_1, \dots, r_m are all the preimages of r , and $\alpha_0, \dots, \alpha_m$ are the complementary segments of $\text{Fr } E$ (see Diagram 3, which illustrates the case where E is a disc; when E is a half-disc, both α_0 and α_m may be non-empty, but when E is a disc we choose $\alpha_0 = \emptyset$). Now we show how to correct this problem, at the expense of increasing branching over the barycenters of 2-handles.

The edge $r \subset K$ is contained in a 2-cell σ of X . For each $i = 1, \dots, m$, let C_i be the 2-cell of D which contains r_i (see Diagram 3). A priori, it may happen that $C_{i_1} = C_{i_2}$ for some $i_1 \neq i_2 \in \{1, \dots, m\}$. However, this violates the fact that f is least area, as follows. Choose an embedded arc $\gamma \subset \sigma$ connecting the barycenter of σ to v , intersecting r at a single, transverse intersection point. Let $\gamma_i \subset C_i$ be an arc mapping homeomorphically to γ and connecting the center of C_i to c passing through r_i . It follows that $\gamma_{i_1} \cup \gamma_{i_2}$ is a closed curve in D , bounding a subdisc $D' \subset D$. Define a new disc D'' by removing $\text{int}(D')$ from D and gluing γ_{i_1} to γ_{i_2} respecting the map f . It follows that f induces a PL disc map $f'' : D'' \rightarrow X$ with the same boundary as f and with $\text{Area}(f'') < \text{Area}(f)$, contradicting that f is least area. This shows that the cells C_i are pairwise distinct, as illustrated in Figure 3.

We may express ∂C_i as an oriented closed curve β_i based at c , and $f \mid \beta_i$ induces the same orientation on $\partial \sigma$ for each i . Now define a new cell-decomposition of D by breaking c into m points c_1, \dots, c_m , one associated with each arc α_i , and joining the cells C_1, \dots, C_m into a single cell C with boundary $\beta_1 * \cdots * \beta_m$, as shown in Diagram 3 (when E is a half-disc, the point c_m lies on ∂D , and is associated to both α_0 and α_m). The cells of D outside $C_1 \cup \cdots \cup C_m$, and the maps from these cells to X , remain unaffected by this operation. The circle $\beta_1 * \cdots * \beta_m$ maps to $\partial \sigma$ by a covering map of degree m , which extends to a degree m cyclic cover of C over σ . This defines a

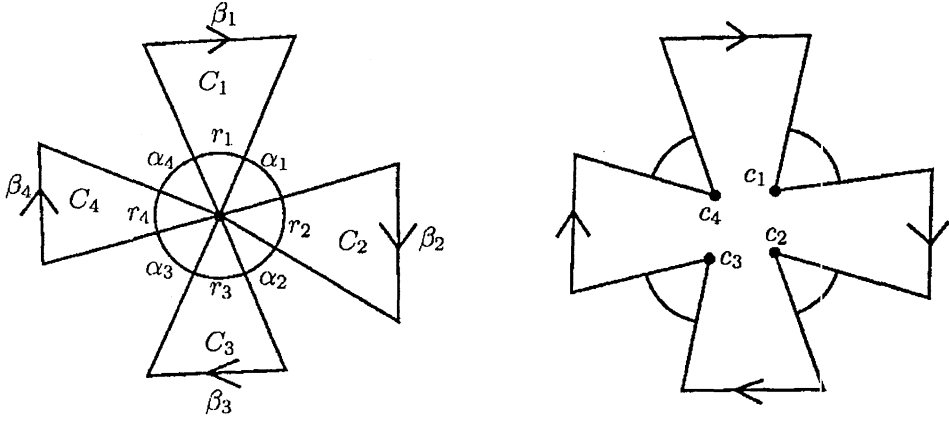


Figure 3: Disc maps can be made normal in 0-handles at the expense of introducing cyclic covers of 2-handles

PL disc map $f': D \rightarrow X$ with $\partial f = \partial f'$, and evidently $\text{Area}(f) = \text{Area}(f')$, because the contribution of $C_1 \cup \dots \cup C_k$ to $\text{Area}(f)$ equals the contribution of C to $\text{Area}(f')$, namely m . The map f' is N -normal over each 2-handle for some N , and f' is still normal over each 1-handle. Notice that E has been replaced by m distinct components of the preimage of H , and the frontiers of these components still do not traverse any edge of K twice in the opposite direction. In fact, these boundaries can be written as $\alpha_i * r'_i$ (or, in the case where E is a half-disc and $i = m$, as $\alpha_0 * r'_m * \alpha_m$) where $f' \upharpoonright \alpha_i$ is equivalent to $f \upharpoonright \alpha_i$, and r'_i maps homeomorphically to r .

The first time we do this operation, the map f starts out being 1-normal over 2-handles, and afterwards is N -normal over 2-handles for some N . Now we want to repeat the operation wherever there is a 0-handle H and a component E of $f^{-1}(H)$ such that $\text{Fr } E$ traverses some edge of ∂H more than once in the same direction. Since f is now only N -normal over 2-handles, some slight changes are needed in describing the operation. Consider the 2-cells C_1, \dots, C_m described above. Now we only know that C_i is a cyclic covering over σ of some degree N_i . After breaking c into m points c_1, \dots, c_m , the curve $\beta_1 * \dots * \beta_m$ maps to $\partial\sigma$ by a covering map of degree $N_1 + \dots + N_m$, so we may define a cyclic covering of C over σ of degree $N_1 + \dots + N_m$. Now the map is N -normal for some possibly larger value of N .

Each time the operation is repeated, the number of 2-cells of D decreases, so eventually the procedure must stop. Once this happens, we have produced a least area N -normal disc map $f: D \rightarrow X$, for some N . Moreover, we still

have that $f \in \mathcal{A}_C$, that is, $\mathcal{I}(f) \geq C$, and $\text{Length}(\partial f)$ is minimal among all PL disc maps with isoperimetric ratio $\geq C$. This finishes the proof of the N -Normal Disc Lemma. \square

Proof of Normal Disc Theorem. Consider a least area disc map $f_0: D \rightarrow X$. Applying the N -Normal Disc Lemma, we obtain a PL least area N -normal disc map $f: D \rightarrow X$ that has minimal boundary length among all least area PL disc maps with simplicial boundary of isoperimetric ratio $\geq \mathcal{I}(f_0)$.

Now we must show how to reduce N to 3, using the fact that X is generic. This will involve much cutting and pasting of the cell decomposition of D . These cut and paste operations are closely related to the “diamond moves” of combinatorial group theory (see e.g. [Ger87]). The intuition behind these operations is that branch points of f with high degree can be split up into several branch points of bounded degree. The difficulty is doing this so that different branch points of f live in different cells of D .

Consider a cell $C \subset D$ of degree > 3 . Let $W \subset X$ be the 2-simplex $f(C)$. Note that $C \neq D$, for otherwise by N -normality the map f is a cyclic branched cover of W , but then we may replace f by a map which is a homeomorphism onto W , reducing the boundary length of f without changing the isoperimetric ratio, contradicting the choice of f . Therefore C has at least one edge not lying on ∂D , which by N -normality must map onto an edge E of W that is adjacent to at least one other 2-simplex of X . Since X is generic, there exist either one or two other 2-simplices in X with the edge E , denoted B, R (ignore one of these letters if necessary). We shall think of R, W , and B as the colors red, white, and blue. Let x_R, x_W, x_B be the barycenters of R, W , and B , and let x_E be the barycenter of E . Let $\alpha_R, \alpha_W, \alpha_B$ be oriented, simple paths in R, W, B respectively, which go from x_R, x_W, x_B (resp.) to x_E .

Now define a region $\mathcal{I} \subset D$ by the following recursive procedure: start with $\mathcal{I} = C$, then add on to \mathcal{I} any cell of D whose intersection with \mathcal{I} contains a common edge mapped to E . Continue adding on cells of D in this manner until no more cells can be added. When the procedure halts, any edge of \mathcal{I} mapped to E is either interior to \mathcal{I} or lies on ∂D . We shall examine several features of \mathcal{I} in detail. Once this is done, we shall show how to change the map $f|_{\mathcal{I}}$ so that no 2-cells in \mathcal{I} have degree > 3 . The phrase “copy of E ” will refer to any edge of \mathcal{I} that is mapped to E .

Let $\mathcal{I}^* = \mathcal{I} \cap f^{-1}(\alpha_R \cup \alpha_W \cup \alpha_B)$ denote the “dual graph” of \mathcal{I} , with one vertex in each 2-cell of \mathcal{I} , and one vertex on each copy of E . Each 2-cell of \mathcal{I} is colored either R, W , or B , matching the color of the triangle of X

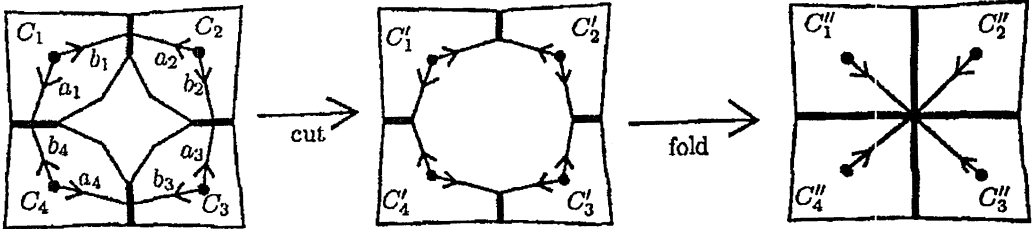


Figure 4: If \mathcal{I}^* is not a tree then $\text{Area}(f)$ can be reduced. In this and later diagrams, the inverse image of E is indicated by thicker arcs.

onto which that 2-cell is mapped. Each vertex of \mathcal{I}^* contained in a 2-cell of \mathcal{I} is colored either R , W , or B , matching the color of the containing 2-cell. Note that a 2-cell (or colored vertex) cannot be adjacent to another 2-cell (or colored vertex) of the same color.

Here is a sequence of facts about \mathcal{I} .

Fact 1. \mathcal{I}^* is a tree.

If not, we will contradict the fact that f is least area. If Fact 1 is false, then there is a simple loop in \mathcal{I}^* . This loop may be described as

$$\ell = a_1^{-1} * b_1 * a_2^{-1} * b_2 * \dots * a_n^{-1} * b_n,$$

for some $n \geq 1$, where a_i and b_i are located in some cell $C_i \subset \mathcal{I}$, and if C_i is labelled, say, W , then a_i and b_i are both preimages of α_W (see Figure 4). The integer n must be even, for as the loop is traversed, the copies of E that the loop crosses have orientations which alternate between pointing into and out of the loop. Now cut D open along ℓ , and throw away the disc bounded by ℓ . What remains is an annulus A , whose intersection with C_i is denoted C'_i . The annulus A will be sewn up to form another PL disc map with boundary ∂f , but with smaller area. This is done by making the following identifications on ∂A : for each $i = 1, \dots, n$, identify a_i with b_i . Since a_i and b_i share their tail endpoints, this identification can be described by saying that a_i and b_i are “folded” together. These identifications evidently make A into a disc D' , and the map $f|_A$ respects these identifications, since a_i and b_i both map to the same one of the arcs α_R , α_W , or α_B . We therefore obtain a new PL disc map $f': D' \rightarrow X$. Clearly $\partial f' = \partial f$, and $\text{Area}(f') < \text{Area}(f)$, contradicting that f is least area.

Fact 2. \mathcal{I} is a pinched topological disc.

In other words, we show that \mathcal{I} is a topological disc with possible identifications of boundary vertices. To see why, first note that any two 2-cells of \mathcal{I} can intersect only along a copy of E , or possibly along a common vertex which maps to a vertex of E ; this follows from the fact that any two of the triangles R, W, B intersect only along E , since X is a simplicial complex. Together with Fact 1, it follows that the interior of \mathcal{I} is an open topological disc, and it also follows that there are no identifications of edges along $\partial\mathcal{I}$. However, there may be vertex identifications along $\partial\mathcal{I}$, hence \mathcal{I} must be a pinched topological disc. In what follows, we shall assume that \mathcal{I} is actually a topological disc; the modifications needed to correct this assumption are minor.

Fact 3. There is at most one copy of E on $\partial\mathcal{I}$.

If not, we will reach a contradiction of the fact that f minimizes boundary length, among all least area PL disc maps with simplicial boundary and with isoperimetric ratio $\geq \mathcal{I}(f_0)$. Recall that any copy of E on $\partial\mathcal{I}$ necessarily lies on ∂D .

Supposing this fact is not true, there is a simple path in \mathcal{I}^* going from ∂D to ∂D . This path may be described as $\ell = a_1^{-1} * b_1 * a_2^{-1} * b_2 * \dots * a_n^{-1} * b_n$, as above. The endpoints of ℓ lie on two copies of E , namely $E_1, E_2 \subset \partial D$. Cutting open D along ℓ , we obtain two subdiscs D_k for $k = 1, 2$, each having a copy of ℓ on its boundary. On D_k , fold the copies of a_i and b_i for $1 \leq i \leq n$, yielding a new disc D'_k (see Figures 5 and 6). These identifications are consistent with $f \mid D_k$, hence we obtain a PL disc map $f_k: D'_k \rightarrow X$.

It is not immediately true that ∂f_1 and ∂f_2 are simplicial. Let $E_{ik} = E_i \cap D_k$, and let $x_k = E_{1k} \cap E_{2k}$ in D'_k . Given k , each of E_{1k} and E_{2k} maps to a half of E ; when n is odd they map to different halves of E ; and when n is even they map to the same half.

If n is odd, it follows that ∂f_1 and ∂f_2 are simplicial, as illustrated in Figure 5. Note that $\text{Length}(\partial f) = \text{Length}(\partial f_1) + \text{Length}(\partial f_2)$. Also, f_1 and f_2 are least area, for if there are PL disc maps f'_1 and f'_2 with the same boundaries and no larger area, one of which has smaller area, then f'_1 and f'_2 can be reglued into a disc map by stretching x_1 and x_2 into arcs and identifying; the resulting PL disc map has the same boundary as f but smaller area, a contradiction.

If n is even then ∂f_1 and ∂f_2 are not simplicial, as shown in Figure 6, because f_k folds the segment $E_{1k} \cup E_{2k}$ doubly over the same half of E . But f_k can then be homotoped to retract that segment to an endpoint of E , making ∂f_k simplicial. Having done this, it follows that $\text{Length}(\partial f) = \text{Length}(\partial f_1) + \text{Length}(\partial f_2) + 2$. The PL disc maps f_1 and f_2 are least area,

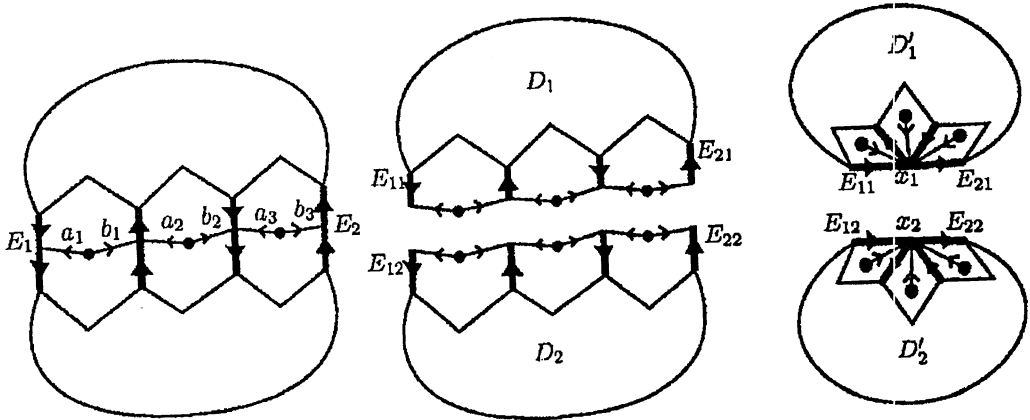


Figure 5: When $\partial\mathcal{I}$ has two copies of E : “odd” case

for if not then restoring their boundaries to the “unsimplicial” form, we can stretch and glue as above, obtaining a PL disc map with the same boundary as f and smaller area, a contradiction.

Whatever the parity of n , we have constructed two least area PL disc maps f_1 and f_2 , with

$$\begin{aligned} \text{Area}(f) &= \text{Area}(f_1) + \text{Area}(f_2) \quad \text{and} \\ \text{Length}(\partial f) &\geq \text{Length}(\partial f_1) + \text{Length}(\partial f_2) \end{aligned}$$

As we have seen before, at least one of these has isoperimetric ratio greater than or equal to $\mathcal{I}(f)$, contradicting the choice of f .

Fact 4. The map $f \mid \mathcal{I}$ preserves orientation.

By this we mean that any two cells of \mathcal{I} with the same color are mapped to the same triangle of X with the same orientation. Suppose this is false, say that two white cells W_1 and W_2 of \mathcal{I} are mapped to W with opposite orientation. Choose a simple edge path $\ell = a_1 * b_1^{-1} * a_2 * b_2^{-1} * \dots * a_n * b_n^{-1}$ in \mathcal{I}^* from the center of W_1 to the center of W_2 . Cutting open along ℓ leaves a hole whose boundary has two copies of each a_i and b_i , so as in Figure 7 we can label the boundary of this hole as

$$a_{1,1} * b_{1,1}^{-1} * \dots * a_{n,1} * b_{n,1}^{-1} * b_{n,2} * a_{n,2}^{-1} * \dots * b_{1,2} * a_{1,2}^{-1}$$

The result of cutting W_1 open along a_1 is denoted W'_1 , and the result of cutting W_2 open along b_n is denoted W'_2 . Now we make the following identifications, to define a new PL disc map. Identify $a_{1,1}$ with $b_{n,1}$, and identify

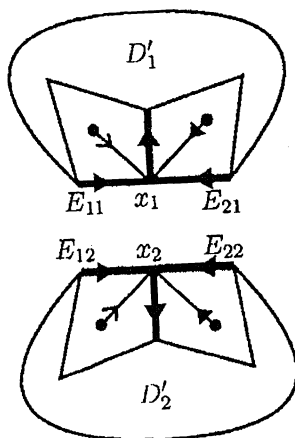


Figure 6: When $\partial\mathcal{I}$ has two copies of E : “even” case

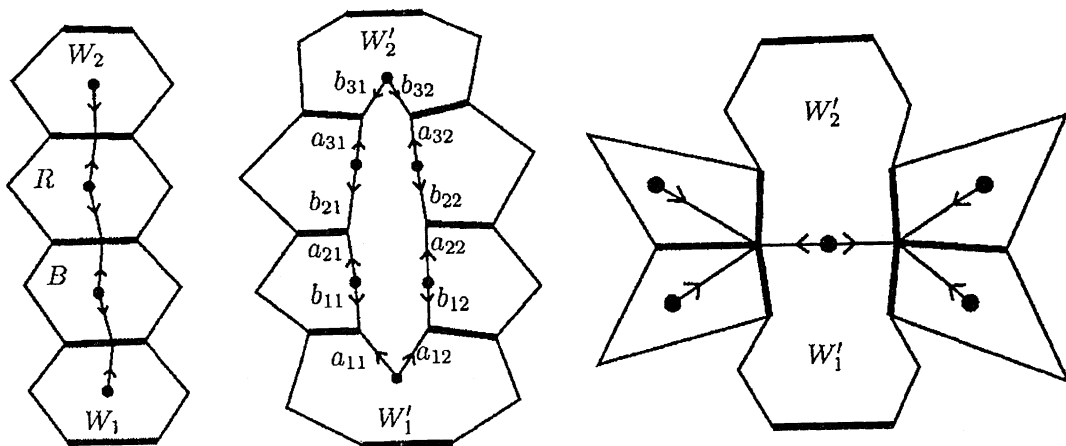


Figure 7: If W_1 and W_2 are mapped to W with opposite orientation, then after cut and paste, area can be reduced.

$a_{1,2}$ with $b_{n,2}$; note that these are *not* folds. Then, for $1 \leq i \leq n-1$, fold $b_{i,1}$ with $a_{i+1,1}$, and fold $b_{i,2}$ with $a_{i+1,2}$.

These identifications are consistent with the map f , hence we obtain a new PL disc map $f': D' \rightarrow X$, with $\partial f = \partial f'$. Now, however, we shall perturb f' to make its area strictly less than $\text{Area}(f)$. To do this, note that W'_1 has a boundary arc $a_{1,1}^{-1} * a_{1,2}$ and W'_2 has a boundary arc $b_{n,1}^{-1} * b_{n,2}$ which are then glued together to form a cell $W' = W'_1 \cup W'_2$. Now consider the boundary map $f' | \partial W': \partial W' \rightarrow \partial W$. Let δ' be the absolute value of the algebraic degree of this map,

$$\delta' = |\text{degree}(f' | \partial W')|$$

Similarly, let

$$\delta_1 = |\text{degree}(f | \partial W_1)|, \quad \delta_2 = |\text{degree}(f | \partial W_2)|$$

Because W_1 and W_2 are mapped to W with opposite orientation, we have $\delta' = |\delta_1 - \delta_2|$, and clearly this is strictly less than $\delta_1 + \delta_2$. The map $f' | W'$ has area $\delta_1 + \delta_2$, but it can be changed by a homotopy rel boundary so that the area is $|\delta_1 - \delta_2|$, which is strictly smaller. Since the boundary map is the same as ∂f , this contradicts the fact that f is least area.

The next two facts are combinatorial consequences of Fact 4. We measure the distance between colored vertices of \mathcal{I}^* by ignoring the points where \mathcal{I}^* intersects a copy of E .

Fact 5. The distance between any two vertices of the same color in \mathcal{I}^* is even.

Suppose not; choose, say, white vertices W_1, W_2 with an odd distance. The arc between these vertices determines a chain of cells $W_1 = C_0, C_1, \dots, C_n = W_2$ with n odd. Let o_i be the orientation on R, W , or B obtained by pushing the orientation on C_i forward via f . Let ϵ_i be the orientation on E obtained from o_i as the induced boundary orientation. Note that $\epsilon_i \neq \epsilon_{i+1}$ for all i , and so ϵ_i alternates values as i increases. Since n is odd it follows that $\epsilon_0 \neq \epsilon_n$, and so $o_0 \neq o_n$, contradicting Fact 4.

Before proceeding to the next fact, we say that a given color R, W , or B is a *central color* if there exists a vertex of \mathcal{I}^* with that color which is adjacent to vertices of both other colors.

Fact 6. There is at most one central color.

Suppose that, say, red and white are both central colors, so there is a path of vertices $W_0—R_0—B_0$ and another path $B_1—W_1—R_1$. Given a pair of vertices in a tree, if you make a new pair by moving one vertex of the original pair to an adjacent vertex, then the distance changes parity. By Fact 5, the distance between W_0 and W_1 is even, and it follows that the distance between B_0 and B_1 is odd, contradicting Fact 5.

Henceforth, by relabelling we may assume that the only central color is W , or that there are only two colors W and R , neither central.

Now we put these facts to work. The disc \mathcal{I} is decomposed into cells by copies of E . The degree of each cell is equal to the number of copies of E on its boundary. The strategy is to change this decomposition, by erasing the old copies of E in \mathcal{I} and drawing in new copies, so that each new cell has no more than three copies of E on its boundary, hence its degree is ≤ 3 . One could try to do this in a piecemeal fashion, doing cut and paste operations analogous to the operations we have been doing above. But it turns out to be easier to erase completely the old copies of E , ignore them henceforth, and then concentrate on the structure of $\partial\mathcal{I}$ as a guide for drawing in new copies of E .

So first we must say what Fact 6 reveals about the structure of $\partial\mathcal{I}$. The endpoints of each old copy of E give two points on $\partial\mathcal{I}$, which we call *dots*. The arc of $\partial\mathcal{I}$ between two adjacent dots will be called a *segment*. There is an even number of segments on \mathcal{I} . By Fact 3, at most one segment is a copy of E . Every other segment maps homeomorphically to one of $\partial R - E$, $\partial W - E$, or $\partial B - E$, hence segments may be colored R , W , or B , except for at most one which is a copy of E . The color of a segment is the same as the color of the adjacent 2-cell. Also, two adjacent colored segments of $\partial\mathcal{I}$ lie on adjacent 2-cells of \mathcal{I} , and hence they must have different colors. Since there is at most one central color, which we are assuming is W , it follows that every other segment of $\partial\mathcal{I}$ is colored W , and the remaining segments are colored either R or B , except that this pattern may be broken at most once at a segment which is a copy of E .

Now we ignore the old copies of E in the interior of \mathcal{I} , and we define a recursive procedure for drawing in new copies of E and carving \mathcal{I} up into new colored cells, each with at most three copies of E on its boundary. To begin with, paint the entire background of \mathcal{I} with gray, to indicate that we have not yet carved out any new cells from \mathcal{I} .

Step 0. To initialize the procedure there are two cases. If there is a copy of E on $\partial\mathcal{I}$, let that be the first new copy of E ; if the two adjacent segments

are colored R or B (they must be the same), connect the opposite endpoints of these segments with a new copy of E which cuts off a new cell of degree 2 colored R or B ; otherwise, if the two adjacent segments are colored W , go on to step 1. If there is no copy of E on $\partial\mathcal{I}$, choose an arbitrary R or B segment on $\partial\mathcal{I}$, and connect its endpoints with a new copy of E , which cuts off a new cell of degree 1 colored R or B .

Step 1. Now look at any new copy of E , call it E_1 , which is adjacent to a portion of \mathcal{I} that is still gray (if there is no such E_1 , then the procedure stops). Look at the segments of $\partial\mathcal{I}$ which are incident both to E_1 and to the gray part of \mathcal{I} . At this stage of the procedure, those segments are colored W . It could be that the segments are the same, making a single segment connecting the two endpoints of E_1 , in which case this segment together with E_1 cuts off a new cell of degree 1 colored W . Otherwise, there are two segments, both colored W ; we call these W_1 and W_2 . Now look at the next segments adjacent to W_1 and W_2 respectively, call them S_1 and S_2 . We have dots $x_i = W_i \cap S_i$ for $i = 1, 2$.

Step 1a. Suppose first that S_1 and S_2 have the same color, say R . Connect the dots x_1 and x_2 with a new copy E_2 of E , and label the resulting cell W , which has degree 2. If $S_1 = S_2$, then this segment together with E_2 cuts off another new cell, which is then colored R , and has degree 1. Otherwise, if $S_1 \neq S_2$, the endpoints of S_1 and S_2 opposite x_1 and x_2 are to be connected with another new copy of E , and the resulting cell is colored R , with degree 2. Now return to step 1.

Step 1b. Now suppose that S_1 is colored, say, R and S_2 is colored B . Going around the boundary of the gray portion of \mathcal{I} from S_1 to S_2 , it must happen that there are three adjacent segments colored, respectively, R , W , and B , which we label R' , W' , and B' (see Figure 8). Consider the dots $y = R' \cap W'$ and $z = W' \cap B'$. Now draw in two new copies of E connecting the pair x_1, y and the pair z, x_2 . This cuts off a new cell colored W with degree 3. The edge connecting x_1, y is then incident to segments labelled R , and we can proceed as in step 1a, drawing in another new copy of E (if $S_1 \neq R'$) and obtaining a new cell colored R with degree 1 or 2. Also, the edge connecting x_2, z is incident to segments labelled B , and proceeding as in step 1a we obtain a new cell colored B with degree 1 or 2. Now return to step 1.

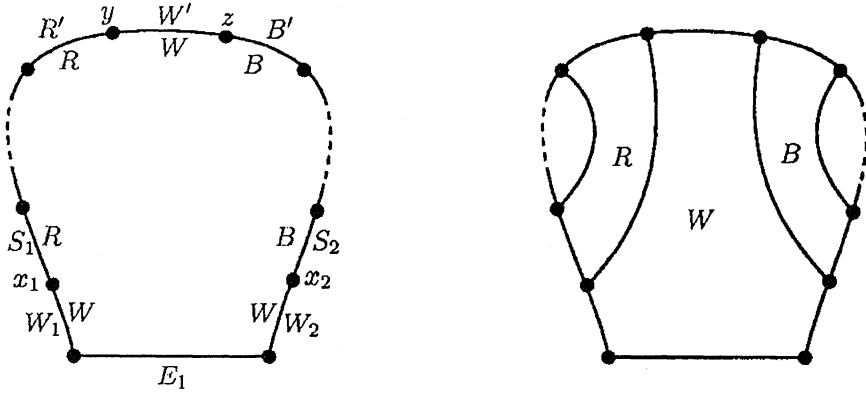


Figure 8: Cutting out new colored cells

This finishes the description of the procedure. As the procedure goes on, fewer and fewer segments of $\partial\mathcal{I}$ remain adjacent to gray parts of \mathcal{I} , while a larger and larger subdisc of \mathcal{I} is carved into colored cells. The procedure eventually halts with all of \mathcal{I} cut into new colored cells. When the procedure finishes executing, then we may use the new coloring to redefine $f|_{\mathcal{I}}$, and all branching numbers on \mathcal{I} are ≤ 3 . The new f is still an N -normal disc map for some N , with the same area and boundary as the old, and the new domain disc D has the same number of cells as the old, but since we started with a cell C having degree strictly greater than 3, the total number of cells in D with degree > 3 has been reduced. We may repeat the procedure until there are no cells with degree > 3 , at which time f is a least area 3-normal disc map with $\mathcal{I}(f) \geq \mathcal{I}(f_0)$. This finishes the proof of the Normal Disc Theorem. \square

Two interesting questions which arise are: Can 3-normal be replaced by 2-normal? Which nonhyperbolic 2-complexes X have a normal limiting lamination? (not all, by the counterexamples given earlier). In regard to this last question, in §5 we shall prove that if a 2-complex X is non-positively curved in a certain sense, then each limiting lamination has locally embedded leaves.

2.6. Construction of the lamination.

If X is not negatively curved, then by the corollary to the Normal Disc Theorem there is a 3-normal fat disc sequence $d_j: D_j \rightarrow X$.

Theorem. *Any N -normal fat disc sequence $d_j: D_j \rightarrow X$ has a subsequence which determines an N -normal measured lamination Λ on X , with $\partial\Lambda = \emptyset$ and $\chi(\Lambda) = 0$.*

Notation: After passing to the subsequence given by the theorem, we shall use the notation

$$\Lambda = \varprojlim \frac{d_j}{\text{Area}(d_j)}$$

and we call Λ a *limiting lamination* of the fat disc sequence d_j .

Proof. Let B_0 be the universal N -normal branched surface of X , and let B_i be the universal refining sequence of B_0 . Each d_j lifts to B_0 , and by the Universality Proposition we have lifts $d_j^i: D_j \rightarrow B_i$ of d_j . Regarding each D_j as a leaf of the universal N -normal lamination Ω , we may proceed by taking the transverse Dirac measure on each D_j , weighted to have total mass 1, and passing to a limit in the weak topology of transverse measures on Ω . We prefer to do this more explicitly, constructing the limiting measure in terms of the refining sequence as follows.

We will use the carrying map d_j^i to determine a normalized invariant measure v_{ij} on B_i , by counting and then dividing by $\text{Area}(d_j)$. Passing to a subsequence, v_{ij} will converge as $j \rightarrow \infty$ to an invariant measure u_i on B_i . Moreover, the u_i will form a *compatible family of invariant measures* on the refining sequence B_i , meaning that the pushforward of u_{i+1} through the submersion $B_{i+1} \rightarrow B_i$ yields u_i . It follows that the u_i induce a transverse measure ν on the universal lamination Ω . Taking Λ to be the support of ν , we obtain the measured lamination $(\Lambda, \nu) = \varprojlim (B_i, u_i)$. Using the fact that $R(d_j) = 1/\mathcal{I}(d_j) \rightarrow 0$ as $j \rightarrow \infty$, we will show that $\partial u_i = 0$, hence $\partial\Lambda = \emptyset$. Since $\chi(d_j)/\text{Area}(d_j) = 1/\text{Area}(d_j) \rightarrow 0$, it will also follow that $\chi(\Lambda) = 0$. Here are the details.

Since d_j^i is a carrying map for the lamination D_j on B_i , by pushing forward the Dirac transverse measure on D_j we obtain an invariant measure w_{ij} , with $w_{ij}(z) = |(d_j^i)^{-1}(z)|$ for each $z \in B_i$. Now define the *normalized counting measures*:

$$v_{ij} = \frac{w_{ij}}{\text{Area}(d_j)}$$

Each 2-cell of B_i has an area, obtained by pulling back the area measure from X . Recall that for any cell τ of B_i and any invariant measure w , $w(\tau)$ denotes $w(z)$ for any $z \in \text{int } \tau$. For each i, j it follows that $\text{Area}(d_j) = \sum_k w_{ij}(\sigma_k) \cdot \text{Area}(\sigma_k)$, where $\{\sigma_k\}$ is an enumeration of the 2-cells of B_i . Fixing i and letting j vary, we have a constant sum

$$\sum_k v_{ij}(\sigma_k) \cdot \text{Area}(\sigma_k) = 1$$

Therefore v_{ij} , or indeed any subsequence of v_{ij} , has a convergent subsequence as $j \rightarrow \infty$, and the limit is an invariant measure on B_i . This invariant measure has mass 1, hence is non-zero.

We do not yet have a compatible family of invariant measures for the refining sequence $\{B_i\}$. To get this, we use a diagonal argument. Starting on B_0 choose a subsequence $v_{0,j(0,n)}$ which converges to an invariant measure u_0 on B_0 . Since the disc maps d_j^0 in B_0 lift to the disc maps d_j^1 in B_1 , and since the counting weights of d_j^0 and d_j^1 are both normalized by dividing by $\text{Area}(d_j)$, the invariant measures $v_{1,j(0,n)}$ on B_1 induce the invariant measures $v_{0,j(0,n)}$ on B_0 . Now choose a subsequence $j(1,n)$ of $j(0,n)$, so that $v_{1,j(1,n)}$ converges to an invariant measure u_1 on B_1 . Continuing in this way, we get a nested sequence of subsequences $j(i,n)$ so that $v_{i,j(i,n)}$ converges to an invariant measure u_i on B_i . Consider the term of the diagonal sequence $v_{k,j(k,k)}$, which is an invariant measure on B_k . By construction, if $k \geq i$ then $v_{k,j(k,k)}$ induces an invariant measure u_{ik} on B_i . Fixing i , u_{ik} is a subsequence of $v_{i,j(i,n)}$, hence u_{ik} converges to u_i . The operation of inducement from invariant measures on B_{i+1} to invariant measures on B_i is continuous, and clearly $u_{i+1,k}$ induces u_{ik} , so u_{i+1} induces u_i . We have shown that the u_i form a compatible family of invariant measures on the refining sequence $\{B_i\}$.

The refining sequence with its compatible family of invariant measures defines a measured lamination $(\Lambda, \nu) = \varprojlim (B_i, u_i)$ carried by B , which is therefore an N -normal lamination on X .

Next we show that $\partial\Lambda = \emptyset$. If $\partial\Lambda$ were nonempty then by positivity of the transverse measure ν it would follow that $\partial\nu$ is a positive transverse measure on the 1-dimensional lamination $\partial\Lambda$. Now $\partial\nu$ induces the invariant measure ∂u_i on ∂B , and it would follow that $\partial u_i \neq 0$ for all i . It therefore suffices to show that $\partial u_i = 0$.

For any 1-cell $E \subset \partial B$, define $\text{Length}(E)$ by pulling back the length measure from $X^{(1)}$. For each i, j we have

$$\text{Length}(\partial d_j) = \sum_k \partial w_{ij}(E_k) \cdot \text{Length}(E_k)$$

where $\{E_k\}$ is an enumeration of the 1-cells of ∂B . Since

$$\text{Length}(\partial d_j) / \text{Area}(d_j) \rightarrow 0$$

as $j \rightarrow \infty$, it follows that

$$\sum_k \partial v_{ij}(E_k) \cdot \text{Length}(E_k) \rightarrow 0 \quad \text{as } j \rightarrow \infty$$

so

$$\sum_k \partial u_i(E_k) \cdot \text{Length}(E_k) = 0$$

Since $\text{Length}(E_k) > 0$ for all k , we have $\partial u_i = 0$.

To prepare for the proof that $\chi(\Lambda) = 0$, we first define the Euler characteristic of an invariant measure u on a branched surface B , as an alternating sum:

$$\chi(u) = \sum_{\tau_0} u(\tau_0) - \sum_{\tau_1} u(\tau_1) + \sum_{\tau_2} u(\tau_2)$$

where τ_i denotes an i -cell of B . Clearly if u is induced by a measured lamination carried by B , then $\chi(u)$ gives the Euler characteristic of that lamination. Also, $\chi(u)$ varies continuously with u .

Since D_j is a disc carried by B_i , we have $\chi(v_{ij}) = 1 / \text{Area}(d_j)$. Since $\mathcal{I}(d_j) \rightarrow \infty$, it follows that $\text{Area}(d_j) \rightarrow \infty$, hence $\chi(v_{ij}) \rightarrow 0$ as $j \rightarrow \infty$. It follows that $\chi(u_i) = 0$, so $\chi(\Lambda) = 0$. \square

3. Area minimization.

Existence of an N -normal measured lamination of Euler characteristic zero is not enough to characterize spaces which are not negatively curved. For example any 2-complex can be altered, without changing the fundamental group, to accept a normal map of a torus, indeed an embedded torus. The additional property needed is that the lamination is strongly least area on each leaf. Here we show that limiting laminations have this property:

Strong Least Area Theorem. *If $f: \Lambda \rightarrow X$ is an N -normal limiting lamination on X with $\Lambda = \lim_{j \rightarrow \infty} d_j / \text{Area}(d_j)$, for some N -normal fat disc sequence $d_j: D_j \rightarrow X$, then f is strongly least area on each leaf.*

The proof of this theorem may be cast in terms of Gromov-Hausdorff convergence. Let $g: P \rightarrow X$ be a PL map, where P is homeomorphic to the

plane, and let $p \in P$ be a base point. We say that g is a *Gromov-Hausdorff limit* of the fat disc sequence $d_j: D_j \rightarrow X$ if for each embedded disc $D \subset P$ containing p , and for all sufficiently large j , there exists a base point $x_j \in D_j$ and an embedding $p_j: (D, p) \rightarrow (D_j, x_j)$ so that the maps $d_j \circ p_j: D \rightarrow X$ converge uniformly to the map $g \upharpoonright D$. In the case where X is a 2-complex and the maps $g: P \rightarrow X$ and $d_j: D_j \rightarrow X$ are all N -normal, the sequence $d_j \circ p_j$ converges uniformly to $g \upharpoonright D$ if and only if $d_j \circ p_j$ is eventually equal to $g \upharpoonright D$. This is what we shall prove:

Strong Shadowing Theorem. *With the same hypotheses as the Strong Least Area Theorem, consider a leaf L of Λ with universal covering $\pi: \tilde{L} \rightarrow L$, and an embedded PL disc $D \subset \tilde{L}$. For all sufficiently large j there exists an embedding $s: D \rightarrow D_j$ such that $d_j \circ s = f \circ \pi \upharpoonright D$.*

Summarizing the preceding discussion, the Strong Shadowing Theorem is equivalent to the statement that the universal cover of each leaf of Λ is a Gromov-Hausdorff limit of the fat disc sequence.

We now recast the Strong Shadowing Theorem in more suggestive language. Any map $s: D \rightarrow D_j$ such that $d_j \circ s = f \circ \pi \upharpoonright D$ is called a *shadow* of D in D_j ; the definition does not require s to be an embedding. The theorem therefore says that every disc embedded in the universal cover of a leaf has an embedded shadow in the fat disc D_j , for sufficiently large j .

Proof that Strong Shadowing implies Strong Least Area. If D is a PL embedded disc in the universal cover of a leaf of Λ , let $s: D \rightarrow D_j$ be a shadow in some fat disc. Since $d_j: D_j \rightarrow X$ is PL least area, the restricted map $d_j \upharpoonright s(D)$ is also PL least area. Therefore, the equivalent map s is PL least area. □

Our proof of the Strong Shadowing Theorem will be a bootstrap argument. As a first step we prove:

Weak least area theorem. *With the same hypotheses as above, if L is a leaf of Λ , and if $D \subset L$ is a subcomplex homeomorphic to a disc, then $f \upharpoonright D$ is least area in X .*

This is an immediate consequence of:

Weak shadowing theorem. *If L is a leaf of Λ , and if $D \subset L$ is a subcomplex homeomorphic to a disc, there exists a fat disc $d_j: D_j \rightarrow X$ and an embedding $s: D \rightarrow D_j$ such that $d_j \circ s = f \upharpoonright D$.*

Remark. If all leaves of Λ were simply connected, the Weak Shadowing Theorem would be sufficient to prove the Strong Least Area Theorem. It seems likely to us that one could construct a limiting lamination with all leaves being simply connected. Nonetheless, we have chosen to prove the Strong Least Area Theorem for more general limiting laminations, in order to cover the fortuitous possibility that the limiting lamination has leaves which are not simply connected, perhaps even a torus leaf.

Remark. It is much easier to prove the following intermediate version of shadowing: each disc D embedded in the universal cover of a leaf is shadowed by some immersion $D \rightarrow D_j$ for some fat disc D_j . It is also easy to see that the map $D \rightarrow D_j$, being a disc immersion into another disc, is least area. Might it be true that the composition $D \rightarrow D_j \rightarrow X$ of least area maps is least area? This would give a much easier proof of the Strong Least Area Theorem. However Joel Hass, improving on the example given in [Has92], has shown us a 3-manifold M , a least area disc map $E \rightarrow M$, and a disc immersion $D \rightarrow E$, such that the composition $D \rightarrow E \rightarrow M$ is not least area. This example emphasizes how important it is to prove existence of embedded shadows $D \rightarrow D_j$.

We have two tasks remaining: proving the Weak Shadowing Theorem; and proving that the Weak Least Area Theorem implies the Strong Shadowing Theorem.

For the rest of this section, fix the universal N -normal refining sequence for X ,

$$\dots \xrightarrow{p_{i+1}} B_i \xrightarrow{p_i} B_{i-1} \xrightarrow{p_{i-1}} \dots \xrightarrow{p_1} B_0$$

The lamination $\Omega = \varprojlim B_i$ is the universal N -normal lamination of X , the maps $f_i: \Omega \rightarrow B_i$ are carrying maps, and the measured lamination $(\Lambda, \nu) = \varprojlim (B_i, \mu_i) = \lim d_j / \text{Area}(d_j)$ is a limiting lamination on X . Let $d_j^i: D_j \rightarrow B_i$ be the carrying map obtained from $d_j: D_j \rightarrow X$ as in the Universality Proposition.

Here is an outline of the proof of Weak Shadowing. First we prove the Stability Lemma, which says that the embedded disc $D \subset L \subset \Lambda$ may be thickened to a sublamination $D \times T$ in the universal lamination Ω , for some open, compact transversal T of Ω , so that D itself corresponds to $D \times t_0$ for some $t_0 \in T$; this is just a form of Reeb Stability. After a few technical lemmas, we next prove the Disc Stack Isolation Lemma, which says that for sufficiently large i , the sublamination $D \times T \subset \Omega$ can be “isolated” in the branched surface B_i in the following sense: the set $\mathcal{S}_i = f_i(D \times T)$ is a disc stack with a stack projection to D that agrees with the projection map

$D \times T \rightarrow D$, and $(f_i)^{-1}(S_i) = D \times T$. We may regard D_j as a leaf of the universal lamination Ω , and we want to know whether D_j intersects $D \times T$ nontrivially; if it does, any component of intersection is a shadow of D in D_j . Using the fact that the normalized Dirac measure on D_j converges to the transverse measure ν on Λ , and the fact that $D \times t_0$ is in the support of ν , it follows that $D_j \cap (D \times T) \neq \emptyset$ for all sufficiently large j , completing the proof.

3.1. Technical lemmas.

Each result about Ω contained in this section holds in a somewhat more general context. For example, the following version of “Reeb stability” holds for any lamination Ω carried on a branched surface.

Stability Lemma. *Let $f: \Omega \rightarrow B$ be a carrying map, let D be a 2-complex homeomorphic to the 2-disc, let L be a leaf of Ω , and let $g: D \rightarrow L$ be a simplicial local embedding. Then there exists a compact space T and a continuous map $G: D \times T \rightarrow \Omega$ with the following properties:*

1. *There exists $t_0 \in T$ such that $g = G(\cdot, t_0)$.*
2. *If $t \in T$ then G maps $D \times t$ into a leaf of Ω by a simplicial local embedding, and the map $D \rightarrow D \times t \xrightarrow{G} \Omega \xrightarrow{f} B$ is identical to the map $f \circ g: D \rightarrow B$.*
3. *For each $x \in D$, the map $G(x, \cdot)$ is a homeomorphism of T onto an open and closed transversal of Ω .*

Moreover, if we add the condition that $G(x \times T)$ be maximal with respect to (1)–(3), then G is unique up to an isomorphism of T .

To refer to this lemma, we will say that G is obtained by *thickening* g in B . Of course, the same disc immersion in Ω will produce different thickenings in different branched surfaces. In particular, if $B' \rightarrow B$ is a surjective submersion, the thickening of a disc immersion in B' will be thinner than the thickening in B .

Proof of Stability Lemma. We use the following classical fact from PL topology: any triangulation of a disc D has a *shelling*, a sequence of sub-complexes $E_0 \subset \cdots \subset E_R$ such that $D_0 = E_0$ is a 2-cell, and if $1 \leq r \leq R$ then $D_r = \text{cl}(E_r - E_{r-1})$ is a 2-cell intersecting E_{r-1} in an arc.

The Stability Lemma is proved by induction on R . If $R = 0$ then $\sigma = f(g(E_0))$ is a 2-cell of B . Let $T = T_\sigma$ and let G map $E_0 \times T$ isomorphically onto $\sigma \times T$. Since f is continuous we have that $\text{int}(E_0) \times T = f^{-1}(\text{int}(\sigma))$ is open, so T is open. Choosing $x \in \text{int}(\sigma)$, we may identify T with $x \times T = f^{-1}(x)$ so T is closed.

Assuming by induction that $G': E_{R-1} \times T' \rightarrow \Omega$ satisfies the conclusions of the lemma, consider $\sigma = f(g(D_R))$, a 2-cell of B . Let α be the arc $D_R \cap E_{R-1}$. Let $x \in B$ be any point in the set $f(g(\alpha))$, so $x \in \partial\sigma$. Consider the transversal $T_x = p^{-1}(x)$, which is both open and closed. Both of the transversals T' and T_σ may be identified with open and closed subsets of T_x . The set $T = T' \cap T_\sigma$ is therefore an open and closed subtransversal, and it is non-empty since it contains a point of $g(D)$. The required map $G: E_R \times T \rightarrow \Omega$ is obtained by pasting together $G' \mid E_{R-1} \times T$ and the isomorphism $D_R \times T \rightarrow \sigma \times T$.

Uniqueness follows by observing that any maximal G must be the same as the G constructed above by induction. \square

By a process similar to the proof of the Stability Lemma, we can define the holonomy of any lamination carried by a branched surface $p: \Omega \rightarrow B$. Let $\gamma: [0, 1] \rightarrow B$ be any dual edge path, choose edges $E_1, \dots, E_n \subset B$, numbers $0 = x_0 < x_1 < \dots < x_n = 1$, and 2-cells $\sigma_0, \sigma_1, \dots, \sigma_n \subset B$ such that $\gamma(x_i) \in \text{int}(\sigma_i)$ and $\gamma \mid [x_{i-1}, x_i]$ passes from σ_{i-1} to σ_i through the edge $E_i \subset \partial\sigma_{i-1} \cap \sigma_i$. For each i , $T_{\sigma_{i-1}}$ and T_{σ_i} both embed naturally as open and closed subtransversals of T_{E_i} . Define $T_0 = T_{\sigma_0}$. By induction, define $T_i \subset T_{\sigma_i}$ as the pull-back to T_{σ_i} of the intersection of the embedded images of T_{i-1} and T_{σ_i} in T_{E_i} ; this is an open and closed transversal of Ω , possibly empty. Via embedding in T_{E_i} we therefore obtain measure preserving embeddings $T_{i+1} \rightarrow T_i$. By composition, we obtain an embedding $T_n \rightarrow T_0$. The image of this embedding is denote $T_\gamma \subset T_0$, and taking the inverse we obtain a measure preserving homeomorphism $h_\gamma: T_\gamma \rightarrow T_n$ called the *holonomy map* of γ . The domain of h_γ is an open and closed subtransversal of T_0 , non-empty if and only if γ lifts to a path in some leaf of Ω .

The following lemma holds for any lamination Ω which is the inverse limit of a refining sequence $\dots \rightarrow B_i \rightarrow \dots \rightarrow B_0$.

Embedding Lemma. *If L is a leaf of Ω and $S \subset L$ is a finite subcomplex, then there exists N such that $f_N \mid S$ is an embedding in B_N . It follows that for all $n \geq N$, the map $f_n \mid S$ is an embedding in B_n .*

Proof. Suppose σ, τ are distinct cells of any dimension in S . Since $S \subset \Omega \subset \varprojlim(B_n)$, there must exist n such that $f_n(\sigma)$ and $f_n(\tau)$ have disjoint interiors. Since there are only finitely many pairs of cells, there exists N such that any two distinct cells of S are mapped into B_N with disjoint interiors, proving the lemma. \square

The next lemma applies to any lamination Ω which is the inverse limit of a refining sequence $\cdots \rightarrow B_n \rightarrow \cdots \rightarrow B_0$. In preparation for a later application of the Splitting Progress Lemma, we need to estimate distance to the “divergence locus”, denoted ∂_d in the Splitting Progress Lemma. Consider a disk D embedded in a leaf of Ω as a subcomplex of that leaf, with $D \cap \partial\Omega = \emptyset$. By the Embedding Lemma there exists m such that f_m embeds D in B_m ; let $\hat{D} = f_m(D)$. Consider a leaf $L \subset \Omega$ and a component C of $L \cap f_m^{-1}(\hat{D})$. Note that C is a surface-like 2-complex. One way that C can fail to map homeomorphically to \hat{D} is if there is an edge $E \subset \partial C$ such that $f_m(E) \subset \text{int}(\hat{D})$. Define $\partial_d C$ to be the set of all such edges. If $E \subset \partial_d C$, then there is a 2-cell σ of L adjacent to C along E , and $f_m(\sigma) \not\subset \hat{D}$. The set $\partial_d C$ is where C “diverges” from B .

The following lemma says that if C does not map homeomorphically to \hat{D} , then $\partial_d C$ is nonempty, and moreover every 2-cell in C is close to $\partial_d C$.

Lemma: Divergence is imminent. *Using the above notation, let ρ be the number of 2-cells in D . If C does not map homeomorphically to \hat{D} , then each 2-cell of C is within combinatorial distance ρ of $\partial_d C$.*

Proof. We use a holonomy argument. Let $\tau \subset C$ be any 2-cell. Choose a subcomplex $S \subset C$ which is maximal subject to the following conditions:

1. S is a connected subsurface.
2. $\tau \subset S$.
3. f_m is an injection on $\text{int}(S)$.

Define $\partial_d S = \partial S \cap \partial_d C$, and let $\partial_i S = \text{cl}(\partial S - \partial_d S)$. By maximality of S , the edges of $\partial_i S$ are identified in pairs under the map f_m .

We claim that if $\partial_i S \neq \emptyset$ then $\partial_d S \neq \emptyset$. If $\partial_i S$ is non-empty, choose a path $\gamma' \subset S$ connecting points $x \neq y$ in $\partial_i S$ which are identified in D . The projection of γ' to \hat{D} is a path γ such that under holonomy, $h_\gamma(x) = y \neq x$. Since σ is an embedding on $\text{int}(S)$, it follows that $\sigma(S) \neq D$, for otherwise,

by Reeb Stability, we reach the contradiction that $h_\gamma(x) = x$. By maximality of S it follows that $\partial_d S \neq \emptyset$, proving the claim.

Also, it cannot happen that $\partial_i S = \partial_d S = \emptyset$, for if so then $S = C$ is mapped by f_m homeomorphically to \hat{D} , contradicting the choice of C .

We conclude that $\partial_d S \neq \emptyset$. Now the area of the surface S is bounded above by the area of D , and any tile in S is therefore within combinatorial distance ρ of $\partial_d S \subset \partial_d C$. It follows that τ is within combinatorial distance ρ of $\partial_d C$. \square

The next lemma applies whenever Ω is the universal lamination of a given branched surface B_0 , i.e. the inverse limit of the universal refining sequence $\cdots \rightarrow B_i \rightarrow \cdots \rightarrow B_0$ constructed in §2.2. The lemma finds “isolated disc products” in Ω . Consider a product sublamination $F \times T \subset \Omega$, where F is a compact surface and T is a compact local transversal. We say that $F \times T$ is *isolated* in B_n if the following conditions hold:

1. The image $\mathcal{S} = f_n(F \times T)$ is an F -stack with stack projection $q: \mathcal{S} \rightarrow F$, so that for each $t \in T$, the map $F \rightarrow F \times t \xrightarrow{f_n} \mathcal{S} \xrightarrow{q} F$ is the identity.
2. $f_n^{-1}(\mathcal{S}) = F \times T$.

Given a sublamination Λ of Ω , if $F \times T \subset \Lambda$, and if the condition (2) is replaced by $f_n^{-1}(\mathcal{S}) \cap \Lambda = F \times T$, then we say $F \times T$ is *isolated in B_n relative to Λ* .

Disc stack isolation lemma. *Suppose D is a disc embedded in a leaf $L \subset \Omega$. There exists a product sublamination $D \times T \subset \Omega$, with $D = D \times t_0$ for some $t_0 \in T$, such that for all sufficiently large n , $D \times T$ is isolated in B_n .*

Proof. First apply the Embedding Lemma so that f_m embeds D in B_m . Then apply the Stability Lemma to thicken D in B_m , obtaining a product $D \times T$. Next, applying the lemma Divergence Is Imminent, if a component C of $f_m^{-1}(f_m(D))$ does not map homeomorphically to $f_m(D)$ then every 2-cell of C is within distance ρ of $\partial_d C$, where ρ is the number of 2-cells in D . Applying the Splitting Progress Lemma, there exists $n > m$, independent of C and depending only on ρ , so that $f_n(D \times T) \cap f_n(C) = \emptyset$. It follows that $D \times T$ is isolated in B_n . \square

3.2. Proofs of shadowing theorems.

Proof of Weak shadowing. Recall that Ω is the universal N -normal lamination for the complex X , and $(\Lambda, \nu) = \varprojlim (B_i, \mu_i)$ is the limit lamination as usual, with $\Lambda \subset \Omega$. We are given a disc D embedded as a subcomplex of a leaf L ; let $\delta: D \rightarrow X$ be the composition $D \hookrightarrow L \subset \Lambda \xrightarrow{f} X$. By the Disc Stack Isolation Lemma we may thicken D to a product sublamination $D \times T \subset \Omega$ so that the composition $D \times T \rightarrow X$ is the same as the composition $D \times T \rightarrow D \xrightarrow{\delta} X$, and so that if i is sufficiently large then $D \times T$ is isolated in B_i . Let $\mathcal{S}_i = f_i(D \times T)$, so \mathcal{S}_i is a D -stack and $f_i^{-1}(\mathcal{S}_i) = D \times T$.

The fat disc D_j may be regarded as a leaf of Ω , and the map $d_j^i: D_j \rightarrow B_i$ is the same as the restriction of f_i to D_j . Let $P_j^i = (d_j^i)^{-1}(\mathcal{S}_i)$, which we call the *picture* of $D \times T$ in D_j . We claim that if j is sufficiently large, then $P_j^i \neq \emptyset$. This follows from the fact that $\lim_{j \rightarrow \infty} \text{Mass}(P_j^i) = \text{Mass}(\mathcal{S}_i) = \text{Mass}(D \times T) \neq 0$, where $\text{Mass}(P_j^i) = \text{Area}(P_j^i) / \text{Area}(D_j)$, and $\text{Mass}(\mathcal{S}_i)$ is obtained by integrating the pullback of area measure in X against the function μ_i . In particular, for sufficiently large j we have $\text{Mass}(P_j^i) \neq 0$ so $P_j^i \neq \emptyset$.

Every component of the nonempty set P_j^i is of the form $D \times t$. Each such component gives an embedded shadow of D in D_j . □

The proof of the Strong Shadowing Theorem uses the Weak Least Area Theorem, and is technically more difficult. Here is an outline.

Recall that we start with an arbitrary PL embedded disc D in the universal cover \tilde{L} of a leaf L of the limit lamination Λ . Assume D is a subcomplex of \tilde{L} . In the Weak Shadowing Theorem, where D embeds in L , the strategy was to thicken D up to an isolated disc product $D \times T \subset \Omega$, and to understand how fat discs pass through $D \times T$. In general there is no hope of finding an isolated disc product, because D is embedded only in \tilde{L} , not in L itself. Instead, we look at the downstairs image $F \subset L$ of D , generically a subsurface of L . Since Λ may have holonomy around F , we cannot generally mimic the Disc Stack Isolation Lemma to produce an isolated product sublamination $F \times T$. But if there is holonomy, the identifications of the original disc map $D \rightarrow L$ can be partially simplified, replacing it by a disc map into a nearby leaf having fewer identifications. This replacement does not change the induced map $D \rightarrow X$. Simplifying the disc map as much as possible, the resulting surface F may be thickened to a product sublamination $F \times T \subset \Lambda$, which is isolated relative to Λ in some B_i . We also show

that if L_t is the leaf containing $F \times t$, then the induced map $D \rightarrow \tilde{L}_t$ is an embedding, for each t ; this is precisely where we use the Weak Least Area Theorem. This result is described in the lemma Simplifying Disc Maps.

The rest of the proof is a delicate study of how fat discs pass through the F -stack $\mathcal{S}_i = f_i(F \times T) \subset B_i$, with the goal of showing that for some fat disc D_j , some component of $D_j \cap f_i^{-1}(\mathcal{S}_i)$ is an embedded shadow of D . By using isolation of the stack \mathcal{S}_i together with a measure theoretic argument, we produce some shadows of D in fat discs. Finally, by using the fact that the maps $D \rightarrow \tilde{L}_t$ described above are embeddings, we show that there are embedded shadows of D in D_j , for sufficiently large j .

We start by taking care of a minor technical annoyance: if we project D to L , the image may not be a surface. Instead it may be a surface-like 2-complex having 0-cells where two or more 2-cells touch, so that a neighborhood of the 0-cell in the image of D is not a disc. To prevent this, we replace the disc D with a regular neighborhood: pulling back the second barycentric subdivision of X gives the second barycentric subdivision of the cell decomposition on \tilde{L} ; now attach to D all simplices of this subdivision which touch the original D . Having done this, the new D is still a disc in \tilde{L} , and its boundary maps to the union of handle boundaries in X . Any disc in \tilde{L} with this property is called a *standard disc*. Any standard disc in \tilde{L} , when projected to L , yields a subsurface $F \subset L$. Henceforth we assume D is standard.

The following lemma shows how to thicken F to a product $F \times T \subset \Lambda$, after possibly altering D to simplify its self-intersections in L . The lemma also provides crucial information about universal covers of leaves. The proof depends on the Weak Least Area Theorem.

Lemma: Simplifying disc maps. *Let L be a leaf of Λ with universal covering $\pi: \tilde{L} \rightarrow L$, and let $D \subset \tilde{L}$ be a standard disc. Let δ be the composition $D \xrightarrow{\pi} L \xrightarrow{f} X$. There exists a product sublamination $F \times T \subset \Lambda$, and a map $g: D \rightarrow F$, such that the map $g \times \text{id}: D \times T \rightarrow F \times T \subset \Lambda$ has the following properties. Let $g_t: D \rightarrow \Lambda$ denote the map $g_t(x) = (g \times \text{id})(x, t)$, let $F_t = F \times t = g_t(D)$, and let L_t be the leaf of Λ containing F_t . Then:*

1. For each $t \in T$ the composition $D \xrightarrow{g_t} \Lambda \rightarrow X$ is identical with δ .
2. Let L_t have universal cover \tilde{L}_t . Any lift $\tilde{g}_t: D \rightarrow \tilde{L}_t$ of the map g_t is an embedding.

3. There exists an integer I such that if $i \geq I$, then $F \times T$ is isolated relative to Λ in B_i .

Proof. Let $g = \pi|_D$. Now let $G_i: D \times T_i \rightarrow B_i$ be a maximal thickening of g in B_i . The sets T_i form a nested sequence of closed subsets of T_0 , whose intersection is $\{t_0\}$. We may identify the maps $G_i|_{D \times t}$ and g_t , and we may identify the maps $G_i|_{D \times t_0}$ and g . Let $F_t = g_t(D)$, a subsurface with boundary in L_t . Note that statement (1) is automatically true for any $t \in T_0$.

Let \hat{F}_t denote the surface obtained from F_t by attaching all caps in L_t , i.e. disc components of $L_t - \text{int}(F_t)$. It follows that \hat{F}_t is essential in L_t , meaning that the induced map $\pi_1(\hat{F}_t) \rightarrow \pi_1(L_t)$ is an injection. Let $F = F_{t_0}$, $\hat{F} = \hat{F}_{t_0}$, and $L = L_{t_0}$ (this will not be the final version of F). By the Embedding Lemma, we may take I' sufficiently large so that if $i \geq I'$ then \hat{F} is embedded by f_i in B_i . For the rest of the proof we fix I' to insure this is so, and we write $G = G_{I'}$, $f = f_{I'}$, and $B = B_{I'}$. We abuse notation by sometimes using \hat{F} to denote $f(\hat{F})$ in B . In particular, f induces a map $F_t \rightarrow F$ for each $t \in T_{I'}$.

We claim that if $I \geq I'$ is sufficiently large, then for all $i \geq I$ and $t \in T_i$, the map $f|_{F_t}: F_t \rightarrow F$ extends to a continuous map $\hat{F}_t \rightarrow \hat{F}$. From this claim, statement (2) follows for any $t \in T_I$, for if there is a lift $\tilde{g}_t: D \rightarrow \tilde{L}_t$ that is not an embedding, then there is a nonclosed curve $\gamma \subset D \subset \tilde{L}$ such that $\tilde{g}_t \circ \gamma$ is closed in \tilde{L}_t . The curve $g_t \circ \gamma$ is therefore homotopically trivial in L_t , and so also in \hat{F}_t since \hat{F}_t is essential in L_t . The map $(f|_{F_t}) \circ g_t \circ \gamma$ is therefore homotopically trivial in \hat{F} , and so also in L . But the latter curve is the same as the projection of γ to L under the universal covering map $\tilde{L} \rightarrow L$, which is homotopically nontrivial because $\gamma \subset \tilde{L}$ is not closed. This contradiction proves that \tilde{g}_t is an embedding.

The proof of the claim uses the Weak Least Area Theorem together with the Splitting Progress Lemma. Consider $t \in T_{I'}$, and suppose that for some cap $E \subset \hat{F}_t$ with $c = \partial E$, the map $f|_c$ does not extend to a map $E \rightarrow \hat{F}$. We want to show that $t \notin T_I$ for some I independent of t and E . Since \hat{F} is essential in L it follows that $f|_c$ does not extend to a map $E \rightarrow L$. Let $C \subset E$ be a maximal connected subcomplex of E containing c such that $f|_c$ extends to a map $C \rightarrow L$. Applying the Splitting Progress Lemma to C we conclude that the map $f|_c$ in L is not parallel to $c \subset L_t$ in the branched surface B_I , as long as I is sufficiently large; therefore D and $g_t(D)$ are not parallel in B_I , so $t \notin T_I$. The fact that I is independent of t and E follows from the Splitting Progress Lemma and the fact that

$\text{Area}(C) \leq \text{Area}(E)$, if we can show that $\text{Area}(E)$ is bounded independently of t and E . The number $\text{Length}(\partial F_t)$ is bounded by $\text{Length}(\partial D)$ for each t , so each component of ∂F_t has length bounded independently of t , hence by the Weak Least Area Theorem the area of each cap E of \hat{F}_t is bounded independently of t and E , establishing the claim.

Our family F_t does not necessarily define a product in Λ . To obtain a product, we choose $t_1 \in T_I$ such that g_{t_1} identifies a minimal number of cells (of any dimension); if M_t is the number of cells identified by g_t , then M_{t_1} is minimal. Since the fixed point set of the holonomy map is a closed set, we can choose a neighborhood $U \subset T_I$ of t_1 such that for any $t \in U$, if $g_t(x) = g_t(y)$ then $g_{t_1}(x) = g_{t_1}(y)$, for any $x, y \in D$. It follows that $M_t \leq M_{t_1}$ for $t \in U$. But since M_{t_1} is minimal we have $M_t = M_{t_1}$, and the identifications that g_t performs on D must be the same, so the surfaces F_t are all homeomorphic, for all $t \in U$.

Now let $\tau_i \subset T_i$ be the transversal for a maximal thickening $H_i: D \times \tau_i \rightarrow B_i$ of $h = h_{t_1} = g_{t_1}$, for $i \geq I$. By choosing i sufficiently large, we may assume by the Embedding Lemma that f_i embeds F_{t_1} in B_i , and we may also assume that $\tau_i \subset U$. Fix i so that this is so, and redefine I to be this value of i . Let $T = \tau_I$, redefine $F = F_{t_1}$, and we now have a product $F \times T \rightarrow \Lambda$ satisfying (1)–(2). As usual, we confuse F with its image in B_I .

The final step is to show that $F \times T$ can be isolated in B_i if $i \geq I$ is sufficiently large, using another application of the Splitting Progress Lemma. We use the fact that the maximal thickening of $D \times t_1$ relative to Λ in B_I has the form $h \times \text{id}: D \times T \rightarrow F \times T$ for some map $h: D \rightarrow F$. Consider a component C of $f_I^{-1}(F)$ which is not an F_t for any $t \in T$. If there were a map $g_C: D \rightarrow C$ such that $f_I \circ g_C = f_I \circ g_t$ for $t \in T_I$, then clearly our thickening would not be maximal. Therefore there is no such g_C . This means that if we start at an arbitrary tile of C and attempt to develop D onto C , then the leaf containing C must diverge from F before we have succeeded in developing all of D . It follows that every tile of C is within combinatorial distance $\text{Area}(D)$ of $\partial_d(C)$. Applying the Splitting Progress Lemma, there is an $i \geq I$ such that all components C are mapped disjointly from $f_i(F \times T)$ in B_i . \square

Next we analyze how discs of the fat disc sequence pass through the F -stack which is the image of the product lamination $F \times T$ described in the statement of the previous lemma.

As in the previous lemma, let F be a compact surface, and suppose $F \times T$ is embedded in the lamination Λ . Let $\mathcal{S}_i = f_i(F \times T) \subset B_i$. Further, suppose \mathcal{S}_i is an F -stack with stack projection $\Pi_i: \mathcal{S}_i \rightarrow F$, and suppose that for all

$t \in T$, the map $F \rightarrow F \times t \rightarrow \mathcal{S}_i \rightarrow F$ is the identity, and that $F \times T$ is isolated in B_i relative to Λ . Now define $P_j^i \subset D_j$ to be $P_j^i = (d_j^i)^{-1}(\mathcal{S}_i)$ —this set is called the *picture* of $F \times T$ in D_j .

A sector of B_i on which the invariant measure μ_i is zero is called a *ghost sector*. Define $\partial_g P_j^i$ as the union of edges of ∂P_j^i having an adjacent 2-tile mapped to a ghost sector of B_i . Define $\partial_\partial P_j^i$ to be the closure of $\partial P_j^i - \partial_g P_j^i$. Since $F \times T$ is isolated in B_i relative to Λ , it follows that $\partial_\partial P_j^i$ maps to ∂F under $\Pi_i \circ d_j^i$.

For any integer $\rho \geq 0$ define $P_j^i(\rho)$ to be the union of tiles in P_j^i whose combinatorial distance from $\partial_g P_j^i$ is greater than ρ .

The following lemma says that P_j^i is a “fat picture sequence” in the sense that, ignoring $\partial_\partial P_j^i$, the leftover boundary $\partial_g P_j^i$ becomes arbitrarily small compared to the area.

Fat picture lemma. *The picture is fat in the following sense:*

$$\lim_{j \rightarrow \infty} \frac{\text{Area}(P_j^i)}{\text{Length}(\partial_g P_j^i)} = \infty$$

Moreover, for any $\rho \geq 0$:

$$\lim_{j \rightarrow \infty} \frac{\text{Area}(P_j^i(\rho))}{\text{Area}(P_j^i)} = 1$$

Proof of Fat picture lemma. Recall that each carrying map $f_i: \Lambda \rightarrow B_i$ induces a positive Borel measure on Λ , by pulling back a standard Lebesgue measure of total mass 1 on each 2-cell of B_i , and taking the Fubini product with the transverse measure on Λ . Since each submersion $B_i \rightarrow B_0$ maps each 2-cell of B_i homeomorphically onto a 2-cell of B_0 , we may choose Lebesgue measure on a 2-cell of B_i by pulling back Lebesgue measure on the image 2-cell in B_0 , and so the Borel measure on Λ is independent of B_i . Given a Borel set $A \subset \Lambda$, let $\text{Mass}(A)$ denote its measure.

By the construction, the limit as $j \rightarrow \infty$ of the normalized counting measure of d_j^i is an invariant measure on B_i identical to that induced by the measured lamination Λ . We will use this fact to examine how d_j^i passes through \mathcal{S}_i . Note that the weights induced by Λ or by $F \times T$ on all sectors of \mathcal{S}_i are positive. We have isolated \mathcal{S}_i in order that all sectors of B_i adjacent to int \mathcal{S}_i but not contained in \mathcal{S}_i are ghost sectors, hence the weights induced by Λ on these sectors are all 0. This means that the normalized counting weights

induced by the disc maps d_j^i on these sectors approach 0 as $j \rightarrow \infty$. Each 2-cell in D_j mapped to a ghost sector of B_i touches a uniformly bounded number of edges of $\partial_g P_j^i$. Therefore,

$$\lim_{j \rightarrow \infty} \frac{\text{Length}(\partial_g P_j^i)}{\text{Area}(D_j)} = 0$$

On the other hand,

$$\lim_{j \rightarrow \infty} \frac{\text{Area}(P_j^i)}{\text{Area}(D_j)} = \frac{\text{Mass}(D \times T)}{\text{Mass}(\Lambda)} \neq 0$$

Combining these limits, it follows that

$$\lim_{j \rightarrow \infty} \frac{\text{Length}(\partial_g P_j^i)}{\text{Area}(P_j^i)} = 0$$

which proves the first statement. In particular, given any $\epsilon > 0$, if j is sufficiently large then

$$\frac{\text{Length}(\partial_g P_j^i)}{\text{Area}(P_j^i)} < \epsilon$$

To prove the second statement, let M be an upper bound for the area of a ball of radius ρ in any surface carried by B_i . Since every cell of $P_j^i - P_j^i(\rho)$ is within distance ρ of an edge of $\partial_g P_j^i$, we conclude that $\text{Area}(P_j^i - P_j^i(\rho))$ is bounded above by

$$M \cdot \text{Length}(\partial_g P_j^i) < M\epsilon \text{Area}(P_j^i)$$

for sufficiently large j . We conclude that $P_j^i - P_j^i(\rho)$ comprises an arbitrarily small proportion of $\text{Area}(P_j^i)$. This proves the second statement from the first. \square

Proof of the Strong Shadowing Theorem. Apply the lemma Simplifying Disc Maps to obtain the product $F \times T$, isolated in B_i relative to Λ for $i \geq I$. We have an F -stack $\mathcal{S}_i = f_i(F \times T)$ with $\Lambda \cap f_i^{-1}(\mathcal{S}_i) = F \times T$, and we have the picture $P_j^i = (d_j^i)^{-1}(\mathcal{S}_i)$. Let L_t be the leaf of Λ containing $F \times t$, and let $g_t: D \rightarrow L_t$ be the composition $D \rightarrow F \approx F \times t \subset L_t$.

Step 1: Constructing shadows. We use the Fat Picture Lemma to get lots of shadows $s: D \rightarrow D_j$ of $\delta: D \rightarrow X$. In fact we construct shadows

s which are in the picture P_j^i , in the sense that $s(D) \subset P_j^i$; this implies $d_j^i(s(D)) \subset \mathcal{S}_i$.

Choose a 2-cell $\Sigma \subset D$, and consider $g(\Sigma)$, a 2-cell in F . Denote $Q_j^i = \Pi_i \circ d_j^i : P_j^i \rightarrow F$, and consider the set $(Q_j^i)^{-1}(g(\Sigma)) \subset P_j^i$. Since $(\Lambda, \nu) = \varinjlim (B_i, \mu_i)$, it follows that for $i \geq I$ we have

$$\lim_{j \rightarrow \infty} \frac{\text{Area}((Q_j^i)^{-1}(g(\Sigma)))}{\text{Area}(P_j^i)} = \frac{\text{Area}(g(\Sigma))}{\text{Area}(F)} \neq 0. \tag{*}$$

Let ρ be the number of 2-cells in D . Applying the Fat Picture Lemma we obtain

$$\lim_{i \rightarrow \infty} \frac{\text{Area}(P_j^i(\rho))}{\text{Area}(P_j^i)} = 1. \tag{**}$$

and combining (*) and (**), it follows that for $i \geq I$, if j is sufficiently large then

$$(Q_j^i)^{-1}(g(\Sigma)) \cap P_j^i(\rho) \neq \emptyset. \tag{***}$$

Choose any 2-cell $\sigma \subset (Q_j^i)^{-1}(g(\Sigma)) \cap P_j^i(\rho)$. We construct a shadow $s : D \rightarrow D_j$ of δ in P_j^i such that $s(\Sigma) = \sigma$. As in the Stability Lemma, the shadow s is constructed by using a shelling $\Sigma = E_1 \subset \dots \subset E_\rho$ of D . Note that a partial shadow $s_1 : E_1 \rightarrow D_j$ is already defined, with image σ . Assume by induction that for $r < \rho$ we have constructed a partial shadow $s_r : E_r \rightarrow D_j$ extending s_1 . It follows that $s_r(E_r) \subset P_j^i$, because $s_1(E_1) \subset P_j^i(\rho)$. Let α be the arc $E_r \cap \text{cl}(E_{r+1} - E_r)$. We know that $s_r(\alpha) \not\subset \partial_g P_j^i$, because the combinatorial distance from $s_r(E_r)$ to $\partial_g P_j^i$ is at least $\rho - r$. We also know that $s_r(\alpha) \not\subset \partial_\partial P_j^i$ because $\alpha \not\subset \partial D$ and so $g(\alpha) \not\subset \partial F$. Since $s_r(\alpha) \not\subset \partial P_j^i$ it follows that s_r can be extended over E_{r+1} to yield a partial shadow $s_{r+1} : E_{r+1} \rightarrow P_j^i$. This completes the construction of a shadow of δ in P_j^i .

Step 2: Constructing embedded shadows. If any one of the shadows constructed in step 1 is embedded, the Strong Shadowing Theorem is proved. We study how shadows can fail to be embedded, and we use the Splitting Progress Lemma, together with the fact that D embeds in \tilde{L}_t for each $t \in T$, to show that for some $i \geq I$ and all j , any shadow of δ in P_j^i is embedded.

Consider a shadow $s : D \rightarrow D_j$ of δ in P_j^i which is not embedded. Let $\gamma \subset D$ be an embedded edge path of minimal length so that s identifies the endpoints of γ , hence $s \circ \gamma$ is a simple closed curve in D_j . Let E be the subdisc of D_j bounded by $s \circ \gamma$, and let $C = s(D) \cup E$. The number $\text{Area}(E)$ is bounded by a constant depending only on D and X , because $d_j \upharpoonright E$ is

a least area map into X and $\text{Length}(\partial E) = \text{Length}(\gamma)$ is bounded, being no larger than the number of edges in D . Since $\text{Area}(s(D))$ is bounded by $\text{Area}(D)$, it follows that $\text{Area}(C)$ is bounded. Fix $t \in T$. Since each lift $\tilde{g}_t: D \rightarrow \tilde{L}_t$ is an embedding, the curve $g_t(\gamma)$ is essential in L_t , so E cannot be parallel to a disc map in L_t , and it follows that C cannot be parallel to any map of the form $C \rightarrow L_t$. Applying the Splitting Progress Lemma it follows that if $i \geq I$ is sufficiently large then C is separated from each F_t in B_i , that is, $d_j^i(C) \cap f_i(F_t) = \emptyset$. It follows that s is not a shadow in P_j^i .

We have shown that there is an $i \geq I$ such that for any j , any non-embedded shadow of δ in P_j^I is not a shadow of δ in P_j^i . We know that for sufficiently large j there is a shadow s of δ in P_j^i ; moreover, since $P_j^i \subset P_j^I$ it follows that s is also a shadow in P_j^I , hence s must be embedded. This proves the Strong Shadowing Theorem, and finishes the proof of the Strong Least Area Theorem. \square

4. Proof of main theorems.

First we summarize the proof that if a finite, generic 2-complex X is not negatively curved, then there is a 3-normal measured lamination $f: \Lambda \rightarrow X$ with $\chi(\Lambda) = 0$ which is strongly least area on each leaf. By the corollary to the Normal Disc Theorem, X has a 3-normal fat disc sequence $d_j: D_j \rightarrow X$. By the theorem of §2.6, there is a measured limiting lamination $\Lambda = \lim d_j / \text{Area}(d_j)$ and a 3-normal map $f: \Lambda \rightarrow X$ such that $\chi(\Lambda) = 0$. By the Strong Least Area Theorem of §3, f is strongly least area on each leaf, completing the proof.

The proof of the converse rests on the following:

Theorem: Curvature of leaves. *If Λ is a compact, 2-dimensional measured lamination with $\chi(\Lambda) = 0$ such that almost every leaf is not a 2-sphere, then for almost every leaf L of Λ , the universal cover \tilde{L} is not negatively curved. In fact, \tilde{L} has an embedded PL fat disc sequence.*

Before giving the proof we first apply it:

Proof of theorem A. Given a finite 2-complex X , suppose that there is a n -normal measured lamination $f: \Lambda \rightarrow X$, such that $\chi(\Lambda) = 0$ and f is strongly least area on each leaf. From §1.4 it follows that Λ has no sphere leaves. By the theorem Curvature of Leaves it follows that Λ has a leaf L whose universal cover \tilde{L} has an embedded fat disc sequence. Composing

with the least area map $\tilde{L} \rightarrow X$ yields a PL fat disc sequence in X , which evidently implies the existence of a simplicial fat disc sequence in X , proving that X is not negatively curved. \square

The proof of the theorem Curvature of Leaves will use some analytic tools. First, we shall apply a result of Ghys [Ghy88] which shows that almost every leaf L of Λ is conformally Euclidean. Then we shall use the Ahlfors Lemma [Ahl35] p. 88, which provides an embedded fat disc sequence in \tilde{L} .

The results of [Ghy88] apply to orientable, Riemannian foliations of manifolds, but as Candel has pointed out [Can93], these results hold as well for oriented Riemannian laminations. We begin by making our lamination oriented and Riemannian, after which we can state and apply the results of [Ghy88].

A lamination is *oriented* if it is equipped with an atlas of charts of the form $D \times T$ where D is an oriented disc, such that leaf overlap maps preserve orientation. Without loss of generality we may assume that Λ is oriented, for otherwise the usual construction of orientation double covers for manifolds works to produce an orientation double covering map $\tilde{\Lambda} \rightarrow \Lambda$, and we may replace $f: \Lambda \rightarrow X$ with the map $\tilde{\Lambda} \rightarrow \Lambda \xrightarrow{f} X$; if Λ has an transverse measure of full support with Euler characteristic zero then the same is true of $\tilde{\Lambda}$, and if the map $\Lambda \rightarrow X$ is N -normal and strongly least area on each leaf then the same is true of the map $\tilde{\Lambda} \rightarrow X$; also if almost every leaf of $\tilde{\Lambda}$ has a universal cover with an embedded fat disc sequence then the same is true of Λ .

Riemannian structures on laminations are defined following Candel [Can93]. A lamination Λ is *smooth* if it is equipped with an atlas of charts $D \times T$ where D is a smooth disc, such that leaf overlap maps are smooth. When Λ is smooth, a *smooth Riemannian metric* on Λ is a choice of a positive definite quadratic form $Q(p)$ on the tangent plane to each point $p \in \Lambda$, such that for each chart $D \times T$ and each $t \in T$, the map Q_t which assigns to $x \in D$ the quadratic form $Q_t(x) = Q(x, t)$ is smooth as a function of x , and as t varies in T the map Q_t varies continuously in the C^∞ topology.

Branched surfaces may be used to impose Riemannian metrics on laminations that they carry. Any branched surface B can be embedded in some Euclidean space \mathbf{R}^n as a smooth subcomplex, so that each surface carried by B is locally a smooth submanifold of \mathbf{R}^n , and so that the given PL structure on B is compatible with the smooth structure. Each point of B is equipped with a unique tangent plane, and that plane inherits a Riemannian metric

from \mathbf{R}^n . Any carrying map $\Lambda \rightarrow B$ induces a smooth Riemannian metric on Λ , so that the smooth structure and the PL structure on Λ are compatible.

Given a carrying map $\Lambda \rightarrow B$, we now have two ways to measure length and area in Λ : PL measurements and Riemannian measurements. Given a transverse measure μ on Λ , this leads to two different definitions of $\chi(\Lambda)$. First there is the PL definition we have been using all along. The other definition is given in [Ghy88] and [Can93]: the Fubini product of μ with Lebesgue measure on leaves gives a positive Borel measure ν on Λ , and if $k: \Lambda \rightarrow \mathbf{R}$ denotes the Gaussian curvature function then $\chi(\Lambda) = \int k d\mu$. By applying the Gauss-Bonnet theorem to each sector of B it is easily proved that the two definitions of $\chi(\Lambda)$ agree.

Now we quote the theorems from analysis which will finish off our proof. First is a result of [Ghy88]:

Theorem: Conformal Type of Leaves. *If Λ is a compact, oriented Riemannian measured lamination such that almost every leaf is not a 2-sphere, and if $\chi(\Lambda) = 0$, then almost every leaf of Λ is conformally Euclidean, i.e. the universal cover is conformally equivalent to \mathbf{E}^2 .*

Since this theorem is not stated in [Ghy88], some comments are needed to give a full justification; our thanks to E. Ghys for supplying these comments, as well as for making us aware of the Ahlfors Lemma.

Proof. As in the proof of proposition 3.1 of [Ghy88], the measure μ can be written as a sum $\mu_- + \mu_0$ where μ_- is concentrated on the conformally hyperbolic leaves and μ_0 is concentrated on the conformally Euclidean leaves. Connes has proved that a 2-dimensional measured lamination with positive Euler characteristic must have a positive measure's worth of leaves homeomorphic to the 2-sphere [Con82], [MO97], and hence the integrals $\int k d\mu_-$ and $\int k d\mu_0$ are both non-positive. But since $\int k d\mu_- + \int k d\mu_0 = \int k d\mu = \chi(\Lambda) = 0$, it follows that $\int k d\mu_- = \int k d\mu_0 = 0$. Now, replacing μ by μ_- , it suffices to prove that if almost every leaf is conformally hyperbolic, then the integral of curvature is negative.

In the middle of p. 55 of [Ghy88] is the formula

$$\int k d\mu = \int^{\mathcal{F}} \Delta\phi d\mu - \int \exp(2\phi) d\mu \leq \int^{\mathcal{F}} \Delta\phi d\mu$$

where $\phi: \Lambda \rightarrow [-\infty, +\infty)$ is a certain function which takes value $-\infty$ on conformally Euclidean leaves and value $> -\infty$ on conformally hyperbolic leaves (in the latter case, $\exp(2\phi)$ is the "conformal multiplier function",

which the metric is multiplied by in order to make the curvature constant). However, since we are assuming that almost every leaf is conformally hyperbolic, it follows that $\phi > -\infty$ almost everywhere, hence the above inequality is a *strict* inequality, which yields

$$\int k \, d\mu < \int \mathcal{F} \Delta \phi \, d\mu.$$

Then in proposition 3.2 the integral on the right is proved to be 0, hence $\int k \, d\mu < 0$. \square

Finally, we need the following result, originally found in [Ahl35] p. 188; see also [Can93].

Ahlfors Lemma. *If S is a Riemannian surface which is conformally Euclidean, then the universal cover \tilde{S} has a smoothly embedded fat disc sequence.*

Remark: the method of proof in [Can93] shows that if $f: \mathbf{E}^2 \rightarrow \tilde{S}$ is a conformal isomorphism, and if D_r is the disc of radius r centered on the origin in the Euclidean plane, then there is a sequence r_n so that the isoperimetric ratio of $f(D_{r_n})$ goes to ∞ .

Proof of Theorem: Curvature of Leaves. By combining the theorem Conformal Type of Leaves with the Ahlfors Lemma it follows that for almost every leaf L of Λ , the universal cover \tilde{L} has a sequence of smoothly embedded discs $d_j: D_j \rightarrow \tilde{L}$ which is a fat disc sequence for some Riemannian metric on L . Letting Area_R and Length_R denote Riemannian area and length, this means $\text{Area}_R(d_j)/\text{Length}_R(\partial d_j) \rightarrow \infty$. We may assume that ∂d_j is PL, by slightly perturbing d_j so as to change $\text{Area}_R(d_j)$ and $\text{Length}_R(\partial d_j)$ by bounded factors. Letting $\text{Area}(d_j)$ and $\text{Length}(d_j)$ denote PL measurements, since Λ is compact there is a constant $K \geq 1$ such that

$$\frac{1}{K} \text{Area}(d_j) \leq \text{Area}_R(d_j) \leq K \text{Area}(d_j)$$

and

$$\frac{1}{K} \text{Length}(\partial d_j) \leq \text{Length}_R(\partial d_j) \leq K \text{Length}(\partial d_j)$$

and hence $\text{Area}(d_j)/\text{Length}(\partial d_j) \rightarrow \infty$. \square

Applying these results we also obtain the following special case of a statement of Gromov [Gro87], 6.8.S:

Conformal Minimal Plane Theorem. *Let X be a finite 2-complex. If X is not negatively curved then there is a nonconstant, conformal, least area map from the Euclidean plane to X .*

Proof. If X is not negatively curved, let $f: \Lambda \rightarrow X$ be a 3-normal limiting lamination. Applying the theorem Conformal Types of Leaves, there is a leaf $L \subset \Lambda$ whose universal cover \tilde{L} is conformally equivalent to the Euclidean plane. The induced map $\tilde{L} \rightarrow X$ is nonconstant, conformal, and least area. \square

5. Examples.

5.1. The Baumslag-Solitar group $BS(1, 2)$.

We shall give an explicit description of a normal measured lamination $\Lambda \rightarrow X$ with $\chi(\Lambda) = 0$ which is strongly least area on each leaf, where the fundamental group of the 2-complex X is the Baumslag-Solitar group $BS(1, 2) = \langle a, b \mid bab^{-1} = a^2 \rangle$. This lamination and analogous laminations for the groups $BS(1, n) = \langle a, b \mid bab^{-1} = a^n \rangle$ have played a role in studying quasi-isometric rigidity of these groups [MF97b], [MF97a].

The space X is the quotient of a cylinder $S^1 \times [0, 1]$ by the identification $(z, 1) \sim (z^2, 0)$. Another description of X starts with a rectangle H by inserting a fifth vertex in one of the horizontal sides, then identifying sides labelled a and b with orientations, as shown in Figure 9, to obtain X . We call H a *horobrick* for reasons which will become clear. We will refer to the horizontal side of H containing two a edges as the *long side*, and the other horizontal side will be the *short side*. The edge labelling determines a gluing pattern on the five edges of H , and the quotient complex is X . The universal cover \tilde{X} is tiled with copies of H . To describe \tilde{X} , let T be the unique connected, directed tree such that every vertex has one incoming edge and two outgoing edges. For each directed edge E of T , construct an infinite strip of copies of H glued along their b edges, and map the strip to E by a fibration over \mathbf{R} , so that each directed b edge maps to E preserving direction. At a vertex v of T , three of these strips are to be glued together, one strip $\bigcup_{i \in \mathbf{Z}} H_i$ corresponding to the incoming edge, and two strips $\bigcup_{i \in \mathbf{Z}} H'_i$ and $\bigcup_{i \in \mathbf{Z}} H''_i$ for each of the two outgoing edges. Now glue the long side of H'_i to the short sides of $H_{2i+1} \cup H_{2i+2}$, and glue the long side of H''_i to the short

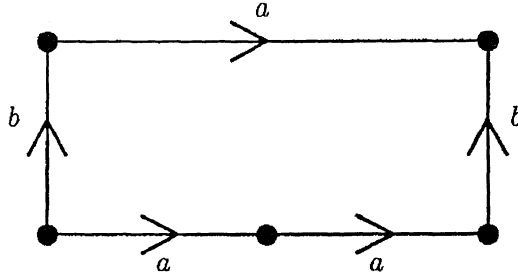


Figure 9: The horobrick: a fundamental domain for \tilde{X}

sides of $H_{2i} \cup H_{2i+1}$. The description given of \tilde{X} yields a fibration $\phi: \tilde{X} \rightarrow T$ with fiber \mathbf{R} , showing that \tilde{X} is contractible.

We describe some useful subsets of \tilde{X} . A *line* in T is any bi-infinite directed path $L = \dots * E_{i-1} * E_i * E_{i+1} * \dots$. The tiling on $\phi^{-1}(L)$ is combinatorially equivalent to a tiling of the upper half-space model $\mathbf{R} \times \mathbf{R}_+$ of the hyperbolic plane by isometric tiles, each a horobrick, so that for any edge $E_i \subset L$, the strip $\phi^{-1}(E_i)$ corresponds to the set $\mathbf{R} \times [2^i, 2^{i+1}]$. Each horobrick in this strip has the form $[x2^i, (x+2)2^i] \times [2^i, 2^{i+1}]$ for some x , i.e. a right angled “rectangle” with two geodesic sides of length $\ln(2)$, a long horocyclic side of length 2, and a short horocyclic side of length 1, hence the horobrick tiles are all isometric. The strips of horobricks are called *horostrips*. Imposing on H the above metric makes X and \tilde{X} into geodesic metric spaces. For each line $L \subset T$, $\phi^{-1}(L)$ is therefore an (isometric copy of the) hyperbolic plane. A *ray* in T is any positive half of any line. If R is a ray, then $\phi^{-1}(R)$ is a *horoball* in a hyperbolic plane. Let R, R' be any two rays in T which are disjoint except that they have the same endpoint v . The space $\phi^{-1}(R \cup R')$ is obtained by gluing two horoballs isometrically along their common boundary $\phi^{-1}(v)$; this is called a *doubled horoball*. The doubled horoball $\phi^{-1}(R \cup R')$ can be pictured by turning it upside down and thinking of sheet draped along the clothesline $\phi^{-1}(v)$; see Figure 10, which shows the sheet with a little air puffed between the two halves.

An explicit fat disc sequence for X can be described as follows. First note that since \tilde{X} is a contractible 2-complex, \tilde{X} has no non-trivial 2-cycles. Therefore, any simplicial embedded disc $d: D \subset \tilde{X}$ is area minimizing in \tilde{X} , for if $d': D' \rightarrow \tilde{X}$ is any other disc map with $\partial d \approx \partial d'$, then $d - d'$ is a 2-cycle, hence trivial, so d' must hit each cell of D , showing that $\text{Area}(d') \geq \text{Area}(d)$. Pick $n \in \mathbf{Z}_+$; we shall describe the n th fat disc in the sequence, with $n = 4$ being pictured in Diagram 11. The path $b^n a b^{-n} a^{-1} b^n a^{-1} b^{-n} a$

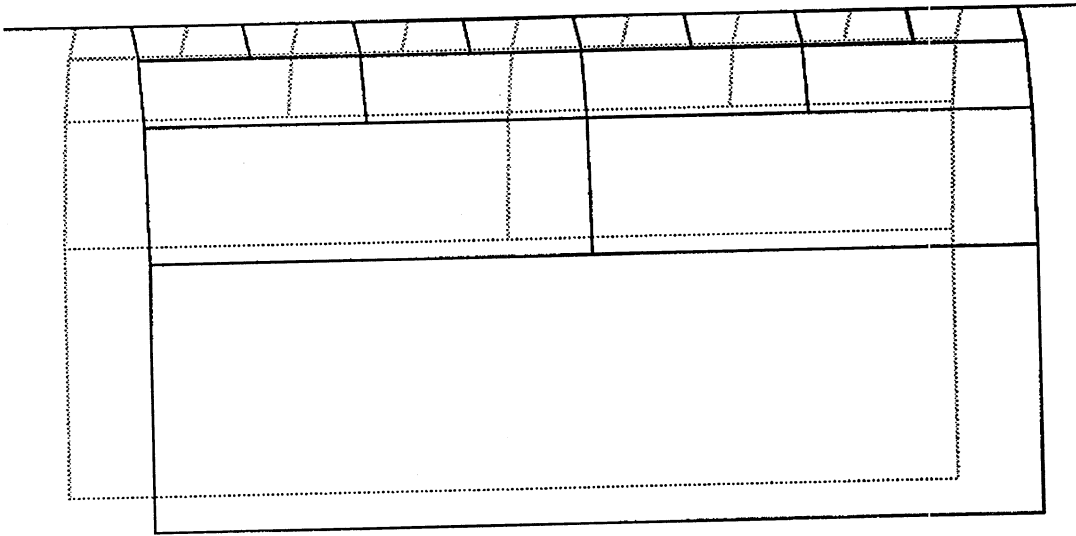


Figure 10: A portion of a doubled horoball

in X is null-homotopic, and it lifts to a simple closed path in \tilde{X} bounding an embedded disc D_n which is contained in a sheet $\phi^{-1}(R \cup R')$ draped along a clothesline $\phi^{-1}(v)$ as described above. By intersecting D_n with each half of the sheet, we obtain a decomposition $D_n = D \cup D'$ where $\partial D \cap \partial D'$ is a lift of the path $a^{2^{n-1}}$, and both of the simple closed curves ∂D and $\partial D'$ are lifts of the closed curve $b^n a b^{-n} a^{-2^n}$. Evidently we have $\text{Length}(\partial D_n) = 4 + 4n$, and $\text{Area}(D_n) = 2 * (2^{n-1} + \dots + 2 + 1) = 2^{n+1} - 2$, hence $\lim_{n \rightarrow \infty} \text{Length}(\partial D_n) / \text{Area}(D_n) = 0$. Projecting to X , we obtain a fat disc sequence $d_n: D_n \rightarrow X$.

The universal normal branched surface B for X can be described as follows. We use the description of X as the quotient of $S^1 \times [0, 1]$ under the identification $(z, 1) \sim (z^2, 0)$. Choose a small $\epsilon > 0$. One sector of B is an annulus mapping to $S^1 \times [\epsilon, 1 - \epsilon]$, which we call the *wide annulus* of B . Another sector of B , called the *narrow annulus*, is an annulus mapping to $S^1 \times [0, \epsilon] \cup S^1 \times [1 - \epsilon, 1]$; the core of the narrow annulus divides it into two sub-annuli, one double covering $S^1 \times [0, \epsilon]$ and the other single covering $S^1 \times [1 - \epsilon, 1]$. The remaining sector of B is a Möbius band, dividing down the middle into a single annulus whose interior singly covers $S^1 \times [1 - \epsilon, 1]$, with the core of the Möbius band doubly covering the circle $S^1 \times 0$ embedded in X . The branched surface B has a unique positive invariant measure up to multiplication by a positive constant. It can be checked directly that the

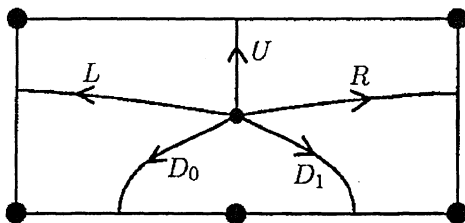


Figure 11: Dual 1-cells in a horobrick

normalized counting measures of the lifts of the fat discs d_n to B converge to an invariant measure on B which assigns a weight 1 to the wide annular sector, and weight 1/2 to the narrow annular sector and the Möbius band sector.

Instead of describing the canonical refining sequence of B , we shall describe the limiting lamination Λ directly, by constructing its local transversal over a base point in X , and by describing the holonomy maps around generating curves in $\pi_1(X)$. The holonomy maps are closely related to the “dyadic adding machine”, a well-known dynamical system [Rud90].

Choose a base point $x_0 \in \text{int}(H)$, viewed as a point of X . The five edges of H map to the two edges of X ; choose a point in the interior of each edge of X and pull back to a point in the interior of each edge of H , and connect x_0 to each such point with a path. These paths, called *dual 1-cells*, are labelled as in Figure 11: U for *up*, L for *left*, R for *right*, and D_0, D_1 for *down*. We shall give an explicit construction of the local transversal τ of Λ over x_0 , and explicit descriptions of the holonomy maps of Λ along the paths $R * L^{-1}$, $U * D_0^{-1}$, $U * D_1^{-1}$, and $D_0 * D_1^{-1}$.

To motivate the description of τ , let us imagine that we already have a lamination Λ so that each leaf $\ell \subset \Lambda$ has universal cover $\tilde{\ell}$ which is a hyperbolic plane or a doubled horoball, in either case tiled by horobricks. In each case, a leaf with a chosen base point can be thought of as a Gromov-Hausdorff limit of fat discs with chosen base points. In the case of a doubled horoball leaf, base points are chosen along the clothesline in each fat disc. In the case of a hyperbolic leaf, base points are chosen farther and farther out into a sheet of the doubled horoball. We also remark that each leaf ℓ is either simply connected or, when $\tilde{\ell}$ is a hyperbolic plane, either $\ell = \tilde{\ell}$ or ℓ is a cylinder obtained as the quotient of a hyperbolic plane by a homothety.

Associated to each horobrick in a leaf are certain data which describe how the given horobrick sits inside the leaf. Given a horobrick $H \subset \ell$, consider the point $x = \tau \cap H$. There is a unique infinite path in X which lifts to a

path in ℓ starting at x and which has the form $U * D_{n_1}^{-1} * U * D_{n_2}^{-1} * \dots * U * D_{n_k}^{-1} * \dots$. We can therefore associate to x the infinite sequence of 0's and 1's $(n_1, n_2, \dots, n_k, \dots)$. This data does not yet completely describe x . To get the next piece of data, consider a horostrip \mathcal{S} containing x (when ℓ is a cylinder, picture \mathcal{S} in the universal cover). If ℓ is a doubled horoball, \mathcal{S} may be adjacent to a clothesline, or there may be some finite number of horostrips below \mathcal{S} , intervening between \mathcal{S} and a clothesline (picture the sheet hanging upward from the clothesline); whereas if $\tilde{\ell}$ is a hyperbolic plane, there are infinitely many horostrips below \mathcal{S} . There is thus an element in the set $\mathbf{Z}_+ \cup \{\infty\}$ describing the number of horostrips below x , including \mathcal{S} . Finally, if ℓ is a doubled horoball, there is a binary piece of data describing which half of the doubled horoball contains x .

Now we are ready to define τ . Let S be a two point set $S = \{S_0, S_1\}$, for labelling halves of doubled horoballs. The transversal τ is defined to be the set $(\{0, 1\}^{\mathbf{Z}_+} \times \mathbf{Z}_+ \times S) \cup (\{0, 1\}^{\mathbf{Z}_+} \times \{\infty\})$. By projecting onto the first direct factor $\{0, 1\}^{\mathbf{Z}_+}$ we obtain functions $f_k: \tau \rightarrow \{0, 1\}$ for each $k \in \mathbf{Z}_+$. By projecting onto the second direct factor \mathbf{Z}_+ or $\{\infty\}$ we obtain a projection $g: \tau \rightarrow \mathbf{Z}_+ \cup \{\infty\}$. Finally, we have a projection $h: g^{-1}(\mathbf{Z}_+) \rightarrow S$. We impose on τ the topology generated by the functions f_k, g , and h , where $\{0, 1\}$ and S are given the discrete topology and $\mathbf{Z}_+ \cup \{\infty\}$ is topologized by taking the obvious bijection with the set $\{1/n \mid n \in \mathbf{Z}_+\} \cup \{0\}$. We make τ into a Borel space by choosing the measure of the set $f_k^{-1}(i) \cap g^{-1}(n) \cap h^{-1}(S_j)$ to be 2^{-n-2} for each $i, j \in \{0, 1\}$, each $k, n \in \mathbf{Z}_+$, and choosing the measure of $g^{-1}(\infty) = 0$. It follows that τ has total mass 1.

Now we describe holonomy. The holonomy maps will be denoted by the corresponding paths in X . The domain and image of the holonomy map $R * L^{-1}$ are both τ . To motivate this map, consider a horobrick B with center x in a tiling of a horoball, look at the path $U * D_{n_1}^{-1} * U * D_{n_2}^{-1} * \dots * U * D_{n_k}^{-1} * \dots$ described above, and consider how the subscripts are affected by replacing B with the next horobrick to the right. The action of $R * L^{-1}$ on a point $x \in \tau$ is obtained by interpreting the sequence $f(x) = (f_k(x))$ as a 2-adic integer and adding 1, that is:

$$f(R * L^{-1}(x)) = f(x) + 1$$

If $f(x)$ is a constant sequence of 1's, then adding 1 converts it to a constant sequence of 0's. The values of $g(x)$ and $h(x)$ are left unchanged by $R * L^{-1}$. Note that the action on $f(x)$ is identical to the so-called "dyadic adding machine" studied in dynamical systems [Rud90].

The domain of the holonomy map $U * D_0^{-1}$ is $f_1^{-1}(0)$. The action of $U * D_0^{-1}$ on the sequence $f(x)$ is a shift, removing the initial 0 from the sequence;

the action on $g(x)$ is to add 1; and $h(x)$ is left unchanged. The image of $U * D_0^{-1}$ is $\tau - g^{-1}(1)$. The action of $U * D_1^{-1}$ is similarly described. The domain and range of $D_0 * D_1^{-1}$ are both $g^{-1}(1)$; the effect of this holonomy map is motivated by noting that it moves between the two halves of a doubled horoball. The map leaves $g(x) = 1$ unchanged, and interchanges the values $h(x)$. The action of $D_0 * D_1^{-1}$ on $f(x)$ depends on which half of the doubled horoball x lies in, i.e. it depends on the value of $h(x)$. If $h(x) = S_0$, then $f(x)$ is left unchanged. If $h(x) = S_1$, then 1 is subtracted from $f(x)$, i.e. for $x \in g^{-1}(1) \cap h^{-1}(S_1)$,

$$f(D_0 * D_1^{-1}(x)) = f(x) - 1$$

It is straightforward that these definitions give a measure preserving pseudo-group acting on τ .

To prove that this pseudo-group defines a lamination Λ on X , there are three relators that must be checked, which come from the three normal tile types at the unique vertex of X . These tile types are in 1-1 correspondence with the three types of vertices that occur in a horobrick tiling of a hyperbolic plane or doubled horoball, as shown in Figure 12. Letting I denote the identity map on any space, the relators are:

- (a) $R * L^{-1} * U * D_1^{-1} * D_0 * U^{-1} = I$
- (b) $R * L^{-1} * U * D_0^{-1} * L * R^{-1} * D_1 * U^{-1} = I$
- (c) $D_0 * D_1^{-1} * R * L^{-1} * D_0 * D_1^{-1} = I$

There is one confusing point in understanding relator (c): recalling that X is the quotient of a horobrick H , note that the paths $R * L^{-1}$, $U * D_0^{-1}$ and $U * D_1^{-1}$ all preserve orientation on H , but $D_0 * D_1^{-1}$ reverses orientation on H . In Figure 12.c the horobrick in the upper part therefore has the orientation inherited from the page, whereas the horobricks in the lower part have the opposite orientation.

Once the relators are checked, Λ is constructed as the quotient of $H \times \tau$ by gluing sides: the right side of $H \times x$ is glued to the left side of $H \times (R * L^{-1}(x))$, etc. The relators show that Λ is indeed a lamination, and that projection to H induces a map $\Lambda \rightarrow X$ yielding a measured lamination on X . Each leaf of Λ is either a doubled horoball (when $g(x) \neq \infty$), or a hyperbolic plane (when $g(x) = \infty$ and $f(x)$ is not eventually periodic), or a hyperbolic cylinder (when $g(x) = \infty$ and $f(x)$ is eventually periodic), and the universal cover of each leaf embeds in \tilde{X} . The map $\Lambda \rightarrow X$ is therefore strongly least

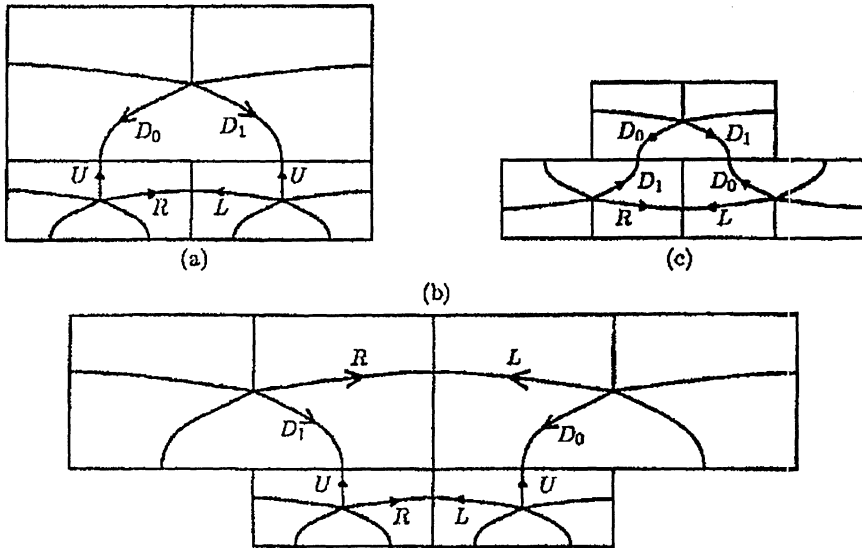


Figure 12: Relators

area on each leaf. It is straightforward to see that any measured lamination on X has Euler characteristic zero.

5.2. Euclidean laminations.

A lamination Λ is *Euclidean* if there is a lamination atlas where each chart has the form $D \times T$ for some open disc $D \subset \mathbb{E}^2$, so that leaf overlap maps are restrictions of Euclidean isometries. We shall describe two classes of spaces which are not negatively curved and that have Euclidean limiting laminations: non-positively curved 2-complexes, and Euclidean tiling spaces.

Non-positively curved 2-complexes. Consider a 2-complex X for which each 2-simplex is identified with some triangle in \mathbb{E}^2 , so that for each 1-cell E of X , any two 2-simplices incident to E are glued isometrically along E . Lengths of paths in X can be defined, and X can be made into a path metric space by defining $d(x, y)$ to be the length of the shortest path from x to y . Let D be a topological disc with a distinguished point $x_0 \in \text{int}(D)$. Given a 0-cell $x \in X$ and a local embedding $f: (D, x_0) \rightarrow (X, x)$, we can pull back the Euclidean metric on the 2-simplices to obtain a Euclidean metric on D with a cone singularity at x of some angle $\theta_f > 0$. We say that X has *non-positive curvature*, denoted $K \leq 0$, if for any f , $\theta_f \geq 2\pi$. Note that the universal cover \tilde{X} inherits a metric of non-positive curvature. We shall

prove:

Theorem. *Let X be a nonpositively curved 2-complex X , and suppose that X is not negatively curved. If $f: \Lambda \rightarrow X$ is an n -normal measured lamination on X with $\chi(\Lambda) = 0$, and if f is strongly least area on each leaf, then Λ is Euclidean, and each leaf maps into X by a local isometry. In particular, f is a local embedding on each leaf.*

Proof. For any $y \in \Lambda$, there is a natural way to put a Euclidean metric on a neighborhood of y by pulling back the metric from X , except when y maps to a 0-cell of X , or when y maps to the barycenter of a 2-cell of X . In each of the two exceptional cases, there is a cone angle $\theta_y \geq 2\pi$, and y has a neighborhood which is isometric to a Euclidean cone of cone angle θ_y . The Euler characteristic may be computed from cone angles as follows. Let $X^{(0)}$ denote the 0-skeleton of X as usual, and let $X_{(0)}$ denote the dual 0-skeleton, i.e. the union of the barycenters of all 2-cells. Noting that since $f^{-1}(X^{(0)} \cup X_{(0)})$ is a disjoint union of local transversals, one may integrate the angle defect $2\pi - \theta_y$ against the transverse measure, and the result is equal to $2\pi\chi(\Lambda) = 0$. Since θ_y is a continuous function of $y \in f^{-1}(X^{(0)} \cup X_{(0)})$, and since the transverse measure is a positive Borel measure on $f^{-1}(X^{(0)} \cup X_{(0)})$, and since $\theta_y \geq 2\pi$ for all y , it follows that $\theta_y = 2\pi$ for all y . We therefore obtain a Euclidean structure on Λ . The map f cannot have a zipped local embedding at any point $y \in f^{-1}(X^{(0)})$, and f cannot have a k -fold branched cover with $k \geq 2$ at any point $y \in f^{-1}(X_{(0)})$, because $\theta_y > 2\pi$ for such values of y . \square

Corollary: Flat Plane Theorem [Bri]. *Under the same hypotheses as above, there is an isometric embedding $\mathbf{E}^2 \rightarrow \tilde{X}$.*

Proof. If $f: \Lambda \rightarrow X$ is as in the previous theorem, and if L is a leaf with universal cover $\tilde{L} \approx \mathbf{E}^2$, then the induced map $\tilde{L} \rightarrow \tilde{X}$ is a local isometry. By [Gro87] (see also [Bal91]) this map is an isometric embedding. \square

Tiling groups. The theory of tilings of \mathbf{E}^2 has been pursued by many mathematicians. One is given a finite set of “tiles” in \mathbf{E}^2 , and the problem is to determine if \mathbf{E}^2 may be tiled by translated copies of the given tiles, with no overlap of the interiors. There are also restricted tiling problems

where one is given a particular Jordan curve γ in \mathbb{E}^2 , and the problem is to determine if the interior of γ can be tiled. The last problem has been analyzed by Conway [Thu90], who associates to each tiling set a group, which is the fundamental group of a certain complex associated to the tiling set. Conway shows that under certain conditions, γ has associated to it a word w in the tiling group, and a necessary condition for the interior of γ to be tiled is that w is trivial in the tiling group.

We shall make a brief study of the geometry of Conway's tiling groups, by proving that if the Euclidean plane can be tiled, then the tiling group is not negatively curved, and the associated 2-complex has a Euclidean limiting lamination. This phenomenon has been noticed in other contexts. For example, Benjamin Weiss pointed out to us that from Robinson's tilings of \mathbb{E}^2 [Gr87], where the tiles are all Euclidean squares with various protrusions to specify the matching rules, one obtains a measure preserving action of \mathbb{Z}^2 on a compact measure space, and in the context of Robinson's tilings this action is easily seen to be a manifestation of the Euclidean limiting lamination constructed below.

Our description of the general tiling problem is designed to lead to a quick definition of Conway's tiling group. Let \mathcal{P} be a finite collection of *tiles*, each of which is a topological polygon in \mathbb{E}^2 , i.e. a closed, embedded disc with a finite set of three or more distinguished points on the boundary called *vertices*; a closed arc between adjacent vertices on the boundary is called an *edge*. Suppose we are also given a finite set \mathcal{F} of translational isometries between certain pairs of edges in the collection \mathcal{P} ; these are the *matching rules* which tell how tiles may be joined. The tiling problem asks: given the *tiling data* $(\mathcal{P}, \mathcal{F})$, can \mathbb{E}^2 be tiled by translated copies of polygons in \mathcal{P} , with disjoint interiors, so that whenever two tiles meet, their intersection is either a vertex of each, or an edge of each whose identification is a consequence of one of the matching rules? More formally, for each tile T in the tiling, there must exist some $P_T \in \mathcal{P}$ and some Euclidean translation f_T such that $T = f_T(P_T)$. Moreover, given tiles T, T' whose intersection is more than a point, there must exist edges $E \subset P_T$ and $E' \subset P_{T'}$ such that $T \cap T' = f_T(E) = f_{T'}(E')$, and there must exist a matching rule $g: E \rightarrow E'$ in \mathcal{F} so that $f_T \upharpoonright E = f_{T'} \circ g \upharpoonright E$.

Associated to the tiling data $(\mathcal{P}, \mathcal{F})$ is the *tiling group* $\Gamma = \Gamma(\mathcal{P}, \mathcal{F})$, discovered and studied by Conway [Thu90]. To define this group, let $X = X(\mathcal{P}, \mathcal{F})$ be the 2-complex obtained from the disjoint union of \mathcal{P} , by identifying edges using \mathcal{F} . We make the additional hypothesis that X is connected. There may, of course, be other identifications which are forced by the given identifications in \mathcal{F} . Now define $\Gamma(\mathcal{P}, \mathcal{F}) = \pi_1(X(\mathcal{P}, \mathcal{F}))$. We

shall prove:

Theorem. *If \mathbf{E}^2 can be tiled using the tiling data $(\mathcal{P}, \mathcal{F})$, then $X = X(\mathcal{P}, \mathcal{F})$ is not negatively curved, and X has a Euclidean limiting lamination.*

Proof. The complex X has a path metric obtained from the Euclidean metric on each tile. We shall use this metric to measure area and length, and with respect to this metric the limiting lamination we construct will be strongly least area on each leaf. Without loss of generality, we may assume that the tile boundaries are rectifiable. This can always be achieved by replacing the original tiling data with new tiling data where tile boundaries are made piecewise straight.

First we construct a “developing map” $\delta: \tilde{X} \rightarrow \mathbf{E}^2$. From the construction, it will follow that δ has a section if and only if the plane can be tiled. Also, it will follow that any section is a least area map. This will allow us to construct a Euclidean limiting lamination.

To construct $\delta: \tilde{X} \rightarrow \mathbf{E}^2$, note that each edge of \tilde{X} can be thought of as a copy of a certain edge of the collection of polygons \mathcal{P} , well-defined up to translation. Choose a maximal tree T for the 1-skeleton of \tilde{X} . First we define $\delta \mid T$. Enumerate the edges of T as $\{E_n\}$, so that $E_1 \cup \dots \cup E_n$ is a tree for all n . Now define $\delta \mid E_n$ by induction, by letting $\delta(E_1)$ be an arbitrary translate of the associated edge of \mathcal{P} , then by induction define $\delta(E_n)$ to be the unique translate of the associated edge of \mathcal{P} so that $\delta \mid E_1 \cup \dots \cup E_n$ is continuous; in other words, the point $E_n \cap (E_1 \cup \dots \cup E_{n-1})$ must be mapped to the same point of \mathbf{E}^2 by $\delta \mid E_1 \cup \dots \cup E_{n-1}$ and by $\delta \mid E_n$.

Now consider an edge E of \tilde{X} which is not in T . The set $E \cap T = \partial E$ consists of two points, and we must show that there is a translate of the associated edge of E connecting the image of these two points under δ . To do this, let α be the unique embedded path in T connecting ∂E . Then $\alpha \cup E$ is a closed curve in \tilde{X} , so it bounds a cellular disc map $d: D \rightarrow \tilde{X}$. We show by induction on $\text{Area}(D)$ that there is a continuous map $g: D \rightarrow \mathbf{E}^2$ such that each tile of D is mapped to a translated copy of the associated tile of \mathcal{P} . By further translating the map g , it follows easily that $\delta \mid \alpha = g \mid \alpha$, and hence $\delta \mid \alpha$ can be defined as $g \mid E$. If $\text{Area}(D) = 1$, then D is a copy of a tile of \mathcal{P} which can be taken to be $g(D)$. If $\text{Area}(D) = n$, then choose a tile $P \subset D$ such that $P \cap \partial D$ is an arc; then $D - P$ is a disc of area $n - 1$ and so a map $D - P \rightarrow \mathbf{E}^2$ can be constructed as desired, and then the map may be extended to P by gluing across the common arc.

We have now defined δ on the 1-skeleton of \tilde{X} , and it follows that δ may be extended across each 2-cell. It follows immediately from the construction of δ that the plane can be tiled if and only if δ has a section.

Now suppose that $f: \mathbf{E}^2 \rightarrow \tilde{X}$ is a section of δ . We claim that f is a least area map. To prove this, let γ be any simple closed curve in \mathbf{E}^2 bounding a tiled disc of area A , and let $d: D \rightarrow \tilde{X}$ be any disc map bounded by $f \circ \gamma$. We have a disc map $\delta \circ d$ into \mathbf{E}^2 bounded by γ , and this disc map must have area at least A , but $\text{Area}(d) = \text{Area}(\delta \circ d) \geq A$, so f is least area.

A fat disc sequence for \tilde{X} is now constructed by taking an embedded fat disc sequence for \mathbf{E}^2 and mapping to \tilde{X} via f , as was done for non-positively curved complexes. The resulting limiting lamination is Euclidean. \square

A question one might be tempted to ask is: can a Euclidean limiting lamination be approximated by a torus? To make the question more precise, we formulate it as follows: if a branched surface B carries a Euclidean limiting lamination, does B also carry a torus? In this form, the answer is no, as can be seen from the tiling examples as follows. We may construct a “Euclidean normal” branched surface B mapping to the complex $X = X(\mathcal{P}, \mathcal{F})$, by using only those normal tiles which are local sections of $\delta: \tilde{X} \rightarrow \mathbf{E}^2$. It is evident that the fat disc sequence constructed in the previous proof is carried by B , hence the limiting lamination may be constructed so that it is carried by B . Note that B carries a torus if and only if there is a doubly periodic tiling of the plane. But it is well known that tile data $(\mathcal{P}, \mathcal{F})$ can be constructed such that the only tilings of the plane are aperiodic (see [Gr87]). For such tiling sets, the limiting lamination cannot be approximated by a torus. We were made aware of these examples by Benjamin Weiss.

Question: is there a tiling group $\Gamma(\mathcal{P}, \mathcal{F})$ which is not negatively curved and which does not contain a \mathbf{Z}^2 subgroup? Can this be done so that $X(\mathcal{P}, \mathcal{F})$ is non-positively curved?

References.

- [Ahl35] L. Ahlfors, *Zur Theorie der Überlagerungsflächen*, Acta Math. **65** (1935), 157–194.
- [Bal91] W. Ballman, *Singular spaces of non-positive curvature*, Sur les group hyperbolique d’après Mikhael Gromov, Progress in Math., vol. 83, Birkhauser, Boston, 1991.
- [BF92] M. Bestvina and M. Feighn, *A combination theorem for negatively curved groups*, J. Diff. Geom. **35** (1992), no. 1, 85–101.

- [Bri] M. Bridson, *On the existence of flat planes in spaces of nonpositive curvature*, Proc. AMS , 123 (1995), no.1, 223-235.
- [Can93] A. Candel, *Uniformization of surface laminations*, Ann. Scient. Éc. Norm. Sup. **26** (1993), 489–516.
- [CDP90] M. Coornaert, T. Delzant and A. Papadopoulos, *Géométrie et théorie des groupes*, Lec. Notes Math., vol. 1441, Springer-Verlag, 1990.
- [Con82] A. Connes, *A survey of foliations and operator algebras*, Operator algebras and Applications (R. V. Kadison, ed.), vol. 38, Proceedings of the symposium in pure mathematics, no. 1, Amer. Math. Soc., 1982, pp. 521–628.
- [Gab95] D. Gabai, *Eight problems in the geometric theory of foliations and laminations on 3-manifolds*, To appear, Proceedings of Georgia Topology Conference, 1995.
- [Ger87] S. M. Gersten, *Reducible diagrams and equations over groups*, Essays in Group Theory, MSRI Publications, vol. 8, Springer-Verlag, 1987, pp. 15–74.
- [Ghy88] E. Ghys, *Gauss-Bonnet theorem for 2-dimensional foliations*, J. Functional Analysis **77** (1988), 51–59.
- [Gr87] B. Grünbaum, *Tilings and patterns*, W. H. Freeman, 1987.
- [Gro87] M. Gromov, *Hyperbolic groups*, Essays in group theory (S. Gersten, ed.), MSRI Publications, vol. 8, Springer, 1987.
- [GS90] S. M. Gersten and H. Short, *Small cancellation theory and automatic groups*, Invent. Math. **102** (1990), 305–334.
- [Hak61] W. Haken, *Theorie der Normalflaechen*, Acta Math. **105** (1961), 245–375.
- [Has92] J. Hass, *Intersections of least area surfaces*, Pac. J. Math. **152** (1992), no. 1, 119–123.
- [MF97a] L. Mosher and B. Farb, *Quasi-isometric rigidity for the solvable Baumslag-Solitar groups, II*, preprint, 1997.
- [MF97b] L. Mosher and B. Farb, *A rigidity theorem for the solvable Baumslag-Solitar groups*, Invent. Math. (1998), to appear.
- [MO97] L. Mosher and U. Oertel, *Two-dimensional measured laminations of positive Euler characteristic*, preprint, 1997.

- [Mos95] L. Mosher, *Geometry of cubulated 3-manifolds*, *Topology* **34** (1995), 789–814.
- [MS88] J. Morgan and P. Shalen, *Degenerations of hyperbolic structures, II: Measured laminations in 3-manifolds*, *Ann. of Math.* **127** (1988), 403–456.
- [Pla73] J. Plante, *A generalization of the Poincaré-Bendixson theorem for foliations of codimension one*, *Topology* **12** (1973), 177–182.
- [Rud90] D. Rudolph, *Fundamentals of measurable dynamics*, Oxford University Press, 1990.
- [Sch] R. Schoen, *Estimates for stable minimal surfaces in three dimensional manifolds*, *Seminar on minimal submanifolds* (E. Bombieri, ed.), *Ann. of Math. Stud.*, vol. 103, Princeton Univ. Press, pp. 111–126.
- [Sch61] H. Schubert, *Bestimmung der Primfaktorzerlegung von Verkettungen*, *Math. Z.* **76** (1961), 116–148.
- [Thu90] W. P. Thurston, *Conway's tiling groups*, *Amer. Math. Monthly* **97** (1990), 757–773.

RUTGERS UNIVERSITY

NEWARK, NJ 07102

E-mail addresses: MOSHER@ANDROMEDA.RUTGERS.EDU,
OERTEL@ANDROMEDA.RUTGERS.EDU