Hausdorff dimension of sets of nonergodic measured foliations

By Howard Masur* and John Smillie**

Introduction

Measured foliations on surfaces arise naturally in several areas of mathematics including the topology of surfaces, the study of quadratic differentials and the study of rational billiards. In the second and third examples, measured foliations occur naturally in one-parameter families $F_\theta$ parametrized by $\theta \in S^1$. These families can be described in terms of a geometric structure on the surface which can be defined using either complex analysis, as in the case of quadratic differentials, or Riemannian geometry. We refer to this structure as a flat structure.

An important topological property of measured foliations is the property of minimality. A measured foliation on $M$ is minimal if every closed set which is a union of leaves is either empty or all of $M$. A somewhat analogous measure-theoretic property of a measured foliation is ergodicity. A measured foliation on $M$ is ergodic if every measurable set which is a union of leaves has measure zero or full measure. Minimality is a topological irreducibility property. Ergodicity is a measure-theoretic irreducibility criterion. That these properties are different is established by examples of minimal nonergodic measured foliations constructed implicitly or explicitly by Veech [V1], Keynes and Newton [K-N], Satayev [Sa] and Keane [K2].

Let $q$ denote a flat structure. Let $NM(q)$ be the set of $\theta \in S^1$ for which $F_\theta$ is not minimal. Results of [Z-K], [B-K-M] and [S] show that $NM(q)$ is countable. Let $NE(q)$ be the set of $\theta \in S^1$ for which $F_\theta$ is not ergodic. The set $NE(q)$ is considerably harder to analyze than the set $NM(q)$.

In the paper [K-M-S] it is shown that for any $q$ the set $NE(q)$ has measure zero. In this paper we deal with the Hausdorff dimension of the set $NE(q)$. The only cases in which the Hausdorff dimension was known were cases in which it was zero. There are for example surfaces for which $NE(q)$ is equal to $NM(q)$ and hence is countable. The simplest example is the flat torus. The measured

---

*Partially supported by NSF Grant #DMS-8902270.
**Partially supported by NSF Grant #DMS-8803228.
foliations that arise for the torus are closely related to rotations on the circle. As is the case for rotations of the circle, minimality is equivalent to ergodicity. In this case \(\text{NE}(q) = \text{NM}(q)\) consists of “rational” directions. The same phenomenon occurs for flat structures on the sphere with four singular points.

The set of flat structures on a fixed surface with certain specified topological data forms a moduli space which we call a stratum. It is convenient to study the behavior of \(\text{NE}(q)\) when \(q\) is restricted to a particular stratum. We refer to the stratum of the flat tori and the stratum of spheres with four singular points as exceptional strata. For every \(q\) in an exceptional stratum \(\text{NE}(q) = \text{NM}(q)\); so, in particular, the Hausdorff dimension of \(\text{NE}(q)\) is zero.

Veech shows in [V4] that every stratum contains a dense set of flat structures \(q\) for which \(\text{NE}(q) = \text{NM}(q)\). We will show that in some sense this behavior is not typical. A stratum has the structure of a smooth manifold with singularities so that it has a natural measure class. When we say a property holds for “almost every \(q\)” in a given stratum we mean with respect to this natural measure class. In this paper we show that for all strata other than the exceptional strata the set \(\text{NE}(q)\) has positive Hausdorff dimension for almost every \(q\). We will state this more precisely. Our main theorem states:

**Main Theorem.** For each component \(C\) of each nonexceptional stratum there exists a \(\delta > 0\) so that for almost every \(q\) in \(C\) the Hausdorff dimension of the set of \(\theta\) for which \(F_{\theta}\) is not ergodic is \(\delta\).

Veech proved in [V2] that each stratum has finitely many components.

This result implies the existence of many minimal nonergodic measured foliations. Since \(\text{NM}(q)\) is countable, the Hausdorff dimension of \(\text{NE}(q)\) is equal to the Hausdorff dimension of \(\text{NE}(q) - \text{NM}(q)\) which by our main theorem must be positive in many cases. In particular the main theorem establishes the existence of minimal nonergodic foliations in all nonexceptional strata.

A measured foliation is uniquely ergodic if it has a unique transverse invariant measure. This is a topological property of the measured foliation; that is to say, it depends only on the foliation and not on the measure. If a measured foliation is uniquely ergodic then it is ergodic with respect to its transverse invariant measure. Inversely a nonergodic measured foliation is not uniquely ergodic. Thus the main theorem establishes the existence of many minimal nonuniquely ergodic measured foliations.

The main theorem allows us to estimate the dimensions of sets of nonergodic foliations in several different contexts.

**Corollary 1.** Let \(\text{NE}\) be the set of \(q \in C\) so that the vertical foliation \(F_{\theta}\) of \(q\) is not ergodic. Then the Hausdorff dimension of \(\text{NE}\) is at least \(\dim C - 1 + \delta\).
The set of measured foliations on a surface of a given genus is a manifold with a piecewise projective structure (cf. [F-L-P]).

**Corollary 2A.** Let $MF(g)$ be the space of measured foliations on a surface of genus $g > 1$ and let $NE$ be the subset of nonergodic measured foliations in $MF(g)$. Then $\dim NE > \dim MF - 1$.

Interval exchange maps (see [K1] or [K-N]) are closely related to measured foliations. Interval exchange maps with a fixed permutation can be parametrized by the lengths of the intervals; thus the set of interval exchange maps with a fixed permutation is a simplex.

**Corollary 2B.** Let $(\Lambda, \sigma)$ be the space of interval exchange maps on $n > 3$ intervals with a fixed irreducible permutation $\sigma$ and let $NE$ be the subset of nonergodic examples. Then $\dim NE > \dim \Lambda - 1$.

The following corollary establishes a topological property of the set of nonergodic directions for a typical flat structure.

**Corollary 3.** Let $\mathcal{M}$ be a nonexceptional stratum. For almost every $q \in \mathcal{M}$ the set of $\theta$ such that $F_\theta$ is nonergodic and minimal is dense in the circle.

We give an outline of the ideas involved in the proof.

Let $C$ be a component of a stratum. The Teichmüller flow is a flow on $C$. There is a smooth measure $\mu$ on $C$ which is invariant under the Teichmüller flow. The Teichmüller flow is ergodic with respect to the measure $\mu$. This was proved first for principle strata (genus $\geq 2$, cone angles $= 3\pi$) by Masur [M1, 2]. The general case was proved by Veech [V3]. The Hausdorff dimension of the set of nonergodic directions is a measurable function on $C$ invariant under the Teichmüller flow. Ergodicity of the flow implies that this function is constant almost everywhere. Now let $C$ be a component of a nonexceptional stratum. We will prove the main theorem by finding a set $W_\infty \subset C$ of positive measure so that for $q \in W_\infty$ the set of nonergodic directions on $q$ has positive Hausdorff dimension.

The following definition will play a central role.

**Definition.** A metric cylinder is a surface isometric to a product of $S^1$ and the interval. If $M$ is a surface with flat structure $q$ then a metric cylinder in $M$ is an embedding of a metric cylinder which is an isometry with respect to the underlying Riemannian structure. The direction of the cylinder is the direction of a waist circle.
Note that we do not specify the relative lengths of the circle or the interval. In many cases we will be dealing with cylinders that are quite degenerate. If the interval factor is long and the circle factor is short we say the cylinder is pinched.

In order to construct a positive Hausdorff dimension, Cantor set of nonergodic directions on a surface $M$ with flat structure $q$, we will construct an infinite set of metric cylinders in $M$. It is convenient to think of these cylinders as forming a “family tree”. There will be a single cylinder at the top of this tree. It will have a finite number of offspring. Each offspring in turn will have a finite number of offspring. We will construct this tree of metric cylinders so that if we follow any infinite sequence of offspring $A_0, A_1, \ldots$ then the corresponding sequence of directions of waist curves $\theta_0, \theta_1, \ldots$ will converge to a direction $\theta_\infty$ and the foliation $F_{\theta_\infty}$ will not be ergodic. We denote the set of limiting directions by $\Lambda_q$. The set $\Lambda_q$ will be a Cantor set of positive Hausdorff dimension.

We construct the tree of cylinders inductively. The inductive step is the following: given a cylinder $A$ on $M$ find a collection of cylinders $B_i$ on $M$ which will be the offspring of $A$. The offspring of $A$ will be chosen to be disjoint from $A$. The number, directions and lengths of these offspring will be chosen to satisfy certain a priori bounds. These bounds will be used to show that the foliations in the limiting directions are not ergodic and that the Cantor set of limit directions has positive Hausdorff dimension. In Section 2 we derive a criterion which we will use to show that the limiting directions are not ergodic. The key property is showing that the area of $M - (A \cup B_1)$ is sufficiently small. In Section 3 we will derive a criterion which will be used to show that the Cantor set of limit directions has positive Hausdorff dimension. It will be important to show that there are sufficiently many offspring and that their directions are relatively evenly spaced on the circle.

We will make use of a natural action of $\mathrm{PSL}(2, \mathbb{R})$ on $C$. The action of $\mathrm{PSL}(2, \mathbb{R})$ on $q$ does not leave the metric on $q$ or the vertical direction invariant but it does leave the underlying affine structure and area unchanged. The Teichmüller flow corresponds to a one-parameter subgroup of $\mathrm{PSL}(2, \mathbb{R})$. It acts by contracting the vertical direction and expanding the horizontal direction.

Let $A$ be a cylinder in $M$ with flat structure $q$. We describe the procedure through which we construct offspring of $A$. This construction will only be possible on a subset of $C$. After we describe the procedure we will describe how to estimate the measure of the set on which this procedure is successful. The important point is that it be possible to construct a complete tree on a set of flat structures which has positive measure in $C$. We begin with a flat structure $q$ with diameter bounded above by a constant given a priori. The first step is to deform the flat structure $q$ by using an element of $\mathrm{PSL}(2, \mathbb{R})$ so that the metric
cylinder is pinched. We do this by contracting in the direction of the waist curve of the cylinder and expanding in the perpendicular direction. We assume that the surface that results from removing this cylinder from the deformed surface has a diameter less than some constant, given a priori. We then attempt to simplify the deformed surface by successively shrinking curves. We find a segment in the surface which is not too long and with the property that if we shrink the curve using an element of $\text{PSL}(2, \mathbb{R})$ then the resulting surface is connected and has diameter bounded by an a priori constant. We use the results of Section 7 to insure that such a curve exists. We continue this process of shrinking curves. Our constants are chosen so that once a curve is shortened, further deformations of the surface do not make it become long. Thus when the initial curves are first shortened, they are made much shorter than they will be at the end of the process. We continue the shortening operation until we reach a stage at which there are no remaining curves to shrink. At this stage we use the results of Section 8 to characterize the resulting surface as a connected sum of a torus, and a surface of small diameter with the cylinder $A$ sewn into the surface of small diameter. (If $M$ is a sphere, the four times punctured sphere will play the role of the torus. This exceptional case will be described completely at the end of Section 1.) We will use the results of Section 9 to find metric cylinders on the torus which miss the region of small diameter and satisfy the spacing requirements. These cylinders are constructed with respect to a flat structure which is affinely equivalent to the original flat structure. However the property of being a metric cylinder depends only on the underlying affine structure so that the metric cylinders that we have constructed give metric cylinders in the original surface.

The following relation of strata will be important in making measure estimates. We say that a stratum $\mathcal{M}'$ is a face of a stratum $\mathcal{M}$ in one of two situations. In the first case the surfaces in $\mathcal{M}'$ result from coalescing two distinct singularities of the surfaces in $\mathcal{M}$. In the second case the surfaces in $\mathcal{M}'$ result from pinching a simple closed curve in the surfaces in $\mathcal{M}$ and desingularizing the resulting surface. This second operation lowers the genus of the surface. (We will describe this relationship more completely in Section 6.) Associated to each stratum $\mathcal{M}$ there is a corresponding space of marked flat structures $\mathcal{D}$. The space $\mathcal{M}$ is a quotient of $\mathcal{D}$ under the action of an appropriate group of homotopy equivalences. In Section 6 we will construct coordinate charts for $\mathcal{D}$ and $\mathcal{D}'$ so that the images of both charts lie in a vector space $V$. The image of the chart for $\mathcal{D}$ will be an open subset $U$ of $V$. The image of the chart for $\mathcal{D}'$ will be an open subset $U'$ of a linear subspace $V' \subset V$ so that $U'$ is contained in the closure of $U$. The points of $U$ which are near $U'$ correspond to flat structures with some very short curves. We will construct a
linear retraction of $U$ into $U'$. In a sense which is made precise in Section 6, a surface $q$ in $U$ is geometrically close to its image surface $q'$ under the retraction.

We estimate the $\mu$ measure of the set of surfaces for which the shrinking operation fails at some point. We do this by induction on the dimension of the strata. Let $\mathcal{D}$ be a stratum. Say that for all faces of $\mathcal{M} \mathcal{D}$ we have estimates on the measures of the sets of flat structures for which the shortening process works. We discard from $\mathcal{M} \mathcal{D}$ the set of surfaces with large diameter. This is a set of small measure. (The required estimates are proved in Section 10.) In the set that remains we can shorten the lengths of the shortest shrinkable curves in each flat structure. This puts us in a neighborhood of the face strata. Flat structures which are geometrically close to flat structures for which the shrinking procedure is successful also have the property that the shrinking procedure is successful. We use the coordinate charts described in Section 6 to analyze the measure of the set of surfaces which after shrinking can be "degenerated" to a flat structure in a face stratum for which the shrinking is successful.

In Section 4 we construct a Delaunay triangulation for flat structures. This will be used in Sections 5 and 6. In Section 5 we find explicit estimates which show that flat surfaces with large diameter contain long metric cylinders. These estimates are used in Section 10. In Section 11 we prove the main theorem. In Section 12 we prove the corollaries.

The authors would like to thank the referee for suggesting a simplification of the original argument used in the proof of Theorem 5.3.

1. Preliminaries

In this section we review some standard material and establish notation. Much of this information is contained in [V2] though our terminology is not the same.

If $P$ is a polygon in the plane with angles that are rational multiples of $2\pi$ then the billiard flow for $P$ can be studied by constructing a surface $M$ obtained by gluing together copies of $P$. This surface has a flat Riemannian metric away from a finite set of points $\Sigma$ corresponding to vertices. A Riemannian metric on a surface is flat if it has Gaussian curvature zero or equivalently if the surface is locally isometric to $\mathbb{R}^2$. We say that a Riemannian metric on $M - \Sigma$ has cone type singularities if in a neighborhood of a point in $\Sigma$ this metric can be written as

$$ds^2 = dr^2 + (cr\,d\theta)^2$$

where $c$ is a positive real number. We say that the metric has a cone-type
singularity with cone angle $2\pi c$. The surface coming from the polygon also possesses a parallel line field $v$. We can define line fields $v_\theta$ by rotating the line field $v$ by an angle of $\theta$. A foliation $F_\theta$ is obtained by integrating the line field $v_\theta$.

If $M$ is a compact Riemann surface and $q$ is a meromorphic quadratic differential with finite norm then $q$ also gives rise to a family of measured foliations. For each $\theta$ in the unit circle let $F_\theta$ be the vertical foliation of the quadratic differential $e^{i\theta}q$ defined by $\text{Re}(e^{i\theta}q)^{1/2}dz = 0$. Every family $F_\theta$ arising from the billiard construction is equivalent to a family $F_\theta$ arising from a quadratic differential. Not every family $F_\theta$ arising from a quadratic differential is equivalent to a family arising from a billiard. But every quadratic differential is equivalent to some flat Riemannian metric with cone-type singularities and a parallel line field.

We will explain how both quadratic differentials and flat Riemannian metrics with parallel vector fields give rise to the same geometric structure. The family of measured foliations $F_\theta$ for either flat metrics or quadratic differentials can be defined in terms of this underlying geometric structure. One method of defining geometric structures on manifolds is by giving an atlas of charts where the change of coordinate maps is restricted. This description in terms of charts will also be useful when we define the $\text{PSL}(2,\mathbb{R})$ action on the space of structures.

Let $q$ be a meromorphic quadratic differential on a surface $M$. A canonical chart for $q$ is a coordinate chart in which the quadratic differential has the form $dz^2$. Away from the zeros and poles of $q$, $M$ can be covered by canonical charts. The change-of-coordinate functions between canonical charts have the form $v \mapsto v + c$ or $v \mapsto -v + c$.

Let $M$ be a surface with a flat Riemannian metric with cone-type singularities. Assume that $M$ possesses a parallel line field, that is to say, a line field which is invariant under parallel translation. Note that the existence of such a line field implies that the cone angles are multiples of $\pi$. Pick one such line field and call it the vertical line field. We will deal only with orientable surfaces and we assume an orientation has been chosen. In a neighborhood of each point of $M - \Sigma$ we can construct a chart which is an orientation preserving isometry and takes the vertical line field to the vertical line field on $\mathbb{R}^2$. The change of coordinate functions between these charts have the form $v \mapsto v + c$ or $v \mapsto -v + c$.

We will use the terminology “flat structure” to refer to a structure with cone angle singularities and a parallel line field. More general flat structures with cone angles that are not multiples of $\pi$ and without parallel line fields have been considered by Veech [V5] and Thurston [T].
If $q$ is a quadratic differential on a compact surface $M$ then the construction above defines a flat Riemannian metric on $M$. A zero of $q$ of order $k$ gives a cone angle $\pi(k + 2)$ singularity and a simple pole defines a cone-angle $\pi$ singularity. The norm of $q$ is the area of $M$ with respect to the area form of the Riemannian metric. If $q$ has finite norm then the poles of $q$ must be simple poles. There is no restriction on the zeros of $q$. We will use the letter $q$ to denote a particular flat structure on the surface $M$. Our terminology for this geometric structure is similar to that of Veech who uses the term “F-structure”.

It is often convenient to describe cone-type singularities by giving the curvature at the singularity rather than the cone angle. At a cone-type singularity the curvature $\kappa$ and the cone angle $c$ are related by the formula $\kappa = 2\pi - c$. With this definition of curvature the Gauss-Bonnet theorem is valid for surfaces with cone-type singularities.

The flat structure $q$ determines certain topological data. Let $\Sigma$ denote the set of singularities of $q$. A flat structure $q$ determines a function $\sigma: \Sigma \to \mathbb{Z}$ where $\pi \sigma(p)$ is the cone angle at $p$. A flat structure is orientable if we can assign a continuous orientation to the leaves of $F_q$. A flat structure is orientable if the corresponding quadratic differential is the square of an Abelian differential.

A flat structure on $M$ defines a homomorphism $\eta$ called the linear holonomy homomorphism from $\pi_1(M - \Sigma)$ to $\mathbb{Z}/2\mathbb{Z}$. Let $\gamma$ be a loop based at a nonsingular point $p$. Let $v$ be a tangent vector based at $p$. If we parallel-translate $v$ along $\gamma$ we get a new vector $w$ based at $p$. The restrictions on the change of coordinate functions imply that $w = \pm v$. We define $\eta(\gamma)$ to be zero if $w = v$ and $\eta(\gamma)$ to be 1 if $w = -v$. A flat structure is orientable if and only if the homomorphism $\eta$ is trivial.

The group $h(M, \Sigma)$ of orientation-preserving homeomorphisms of $M$ fixing $\Sigma$ acts on the space of flat structures which have cone angle $\pi \sigma(p)$ at $p \in \Sigma$, and for which $\eta \neq 0$. The quotient by the group $h_0(M, \Sigma)$ of homeomorphisms homotopic to the identity is the stratum $\mathcal{D}(M, \Sigma, \sigma, -)$. The group $h(M, \Sigma)$ also acts on the space of (orientable) flat structures for which $\eta = 0$. The quotient by $h_0(M, \Sigma)$ is denoted $\mathcal{D}(M, \Sigma, \sigma, +)$.

The Gauss-Bonnet theorem implies that a necessary condition for $\mathcal{D}(M, \Sigma, \sigma, \pm)$ to be nonempty is that the sum of the curvature at the singularities be $2\pi$ times the Euler characteristic of $M$. For $\mathcal{D}(M, \Sigma, \sigma, +)$ to be nonempty, an obvious additional necessary condition is that the cone angles must all be multiples of $2\pi$. Except for two cases the necessary conditions are also sufficient. If $M$ has genus 2, $\Sigma$ consists of a single point $x$, and $\sigma(x) = 6$, then $\mathcal{D}(M, \Sigma, \sigma, -) = \emptyset$. That is, there are no flat structures in genus 2 with nontrivial linear holonomy and a single singularity. Again in genus 2, if $\Sigma =$
\{(x_1, x_2), \sigma(x_1) = 5, \text{ and } \sigma(x_2) = 3, \text{ then again } \mathcal{D}(M, \Sigma, \sigma, -) = \emptyset. \text{ The authors are grateful to Irwin Kra for pointing out these examples. A forthcoming note will show that other than these two exceptions, the necessary conditions are also sufficient for a stratum to be nonempty. For the purposes of this paper we can simply assume that there are a wealth of nonempty strata.}

Let \( \tau(M) \) denote the tangent bundle of \( M - \Sigma \). A flat structure \( q \) on \( M \) gives a canonical local trivialization of this bundle which is determined (as a flat bundle) by the homomorphism \( \eta \). We also use the notation \( \mathbb{R}^2_\eta \). We can use this bundle as a coefficient system for cohomology. Let \( H^1(M, \Sigma; \tau(M)) \) be the first relative cohomology group of the pair \( (M, \Sigma) \) with coefficients in the bundle \( \tau(M) \). When \( q \) is orientable the bundle \( \tau(M) \) is canonically a product. In this case \( H^1(M, \Sigma; \tau(M)) \) can be identified with \( H^1(M, \Sigma; \mathbb{R}^2) \).

When \( q \) is not orientable there is also a method of interpreting \( H^1(M, \Sigma; \mathbb{R}^2_\eta) \) in terms of cohomology with constant coefficients. We first form the orientation cover of \( M \). Let \( \tilde{M} - \tilde{\Sigma} \) be the double cover of \( M - \Sigma \) corresponding to the kernel of \( \eta \). We can lift the flat metric from \( M - \Sigma \) to \( \tilde{M} - \tilde{\Sigma} \) and complete the metric by adding a finite set of points. Call the resulting surface \( \tilde{M} \) and the resulting flat structure \( \tilde{q} \). Let \( M' \) be the singular surface formed by identifying pairs of points that lie over the same orientable singularity. Let \( \Sigma' \) be the set of identified points. The covering map \( \tilde{M} - \tilde{\Sigma} \to M - \Sigma \) extends to a branched covering map from \( M' \) to \( M \). The deck transformation of \( \tilde{M} - \tilde{\Sigma} \) extends to a deck transformation of \( M' \). The branched covering map induces a homomorphism \( j \):

\[
H^1(M, \Sigma; \tau(M)) \to H^1(M', \Sigma'; \tau(M')) = H^1(M', \Sigma'; \mathbb{R}^2).
\]

The deck transformation on \( M' \) gives an involution \( \iota \) of \( H^1(M', \Sigma'; \mathbb{R}^2) \). We call the \( +1 \) eigenspace of \( \iota \) the even cohomology and we call the \( -1 \) eigenspace the odd cohomology. The map \( j \) is an injection and its image is the odd cohomology.

Let \( \tilde{M} \) denote the universal cover of \( M - \Sigma \). There is an isometry of \( \tilde{M} \) into \( \mathbb{R}^2 \) which is an isomorphism of flat structures. This map is called the developing map.

A flat structure \( q \) defines a cohomology class in \( H^1(M, \Sigma; \tau(M)) \) which we call \( \text{dev}(q) \). If \( \gamma \) is a one-simplex in \( M \) the developing map maps a lift of \( \gamma \) to a curve in \( \mathbb{R}^2 \). We define \( h(\gamma) = \text{dev}(q)(\gamma) \) to be the difference of the endpoints of the image of \( \gamma \) and call it the holonomy vector of \( \gamma \). We will use the map \( \text{dev} \) to define a geometric structure on \( \mathcal{D} \).

Let \( \tau \) be a triangulation of \( M \) with \( \Sigma \) as the set of vertices of \( \tau \).
Definition. Let $N(\tau, \eta)$ be the set of flat structures in $\mathcal{D} = \mathcal{D}(M, \Sigma, \sigma, \pm)$ for which the edges of $\tau$ are homotopic to geodesic segments, the triangles are non-degenerate and have the standard orientation and $\eta$ is the linear holonomy.

The holonomy homomorphism defines a map $\text{dev}: N(\tau, \eta) \to H^1(M, \Sigma; \mathbb{R}^2_\eta)$. We use these maps to give charts for $\mathcal{D}$.

**Lemma 1.1.** The map $\text{dev}$ is injective and locally onto when restricted to $N(\tau, \eta)$.

**Proof.** Geodesic triangles with the same coordinates are isometric. If $M_1$ and $M_2$ are surfaces in $N(\tau, \eta)$ with the same image under $\text{dev}$ we construct an isometry between them by identifying corresponding triangles which, by the first remark, are isometric. Each class near $\text{dev}(q_0)$ is $\text{dev}(q)$ for $q$ near $q_0$. Form the flat structure $q$ by gluing neighboring triangles isometrically along common edges.

It is a standard fact that every flat structure has a triangulation as above. We will see in Proposition 4.1 that every flat structure has a useful Delaunay triangulation. In any case, the existence of these triangulations shows that the system of charts described above covers the space of marked flat structures. It follows that there is an affine structure on $\mathcal{D}$ for which the holonomy homomorphism gives local charts.

**Proposition 1.2.** If $\mathcal{D}(M, \Sigma, \sigma, +)$ is nonempty, then its dimension is $2(2g - 1 + \text{card } \Sigma)$. If $\mathcal{D}(M, \Sigma, \sigma, -)$ is nonempty, then its dimension is $2(2g - 2 + \text{card } \Sigma)$.

**Proof.** Because the map $\text{dev}$ gives local charts we need to compute the dimension of the cohomology space which has the same dimension as $H_1(M, \Sigma; \mathbb{R}^2)$. If $\eta$ is trivial, the orientable case, this is just twice the dimension of $H_1(M, \Sigma; \mathbb{R})$ as claimed. Thus we may assume we are in the nonorientable case. Let $n_1$ and $n_2$ denote the number of nonorientable and orientable singularities respectively. The space $H_1(M', \Sigma'; \mathbb{R}^2)$ can be written as a direct sum of the subspaces that are even and odd with respect to the involution $i$. We need to compute the dimension of the odd homology $H_1(M', \Sigma'; \mathbb{R}^2)^{-}$.

Let $g'$ be the genus of $M'$. An Euler characteristic computation gives $2 - 2g' = 2(2 - 2g) - n_1$. The dimension of $H_1(M', \Sigma'; \mathbb{R}^2)$ is therefore $2(2g' + n_2) = 2(4g - 2 + \text{card } \Sigma)$. On the other hand we claim that $j$ defines an isomorphism from $H_1(M; \mathbb{R}^2)$ to the even homology $H_1(M', \Sigma'; \mathbb{R}^2)^{+}$ and thus the latter space has dimension $4g$. To prove the claim, for each homology class in $H_1(M; \mathbb{R}^2)$, choose a representative $\gamma$ that misses $\Sigma$. Its lift $p^{-1}(\gamma)$ to $M'$ is an element of the even homology and if it is trivial so is $\gamma$. Therefore $j$ is an injection. On the other hand any element of the even homology has a representative that misses the set $\Sigma'$ and thus is the lift of a $\gamma$. This proves the claim.
Subtracting the dimension of the even homology from the dimension of the homology gives the result.

The mapping class group \( h(M, \Sigma)/h_0(M, \Sigma) = \text{Map}(M, \Sigma) \) acts on \( \mathcal{D} \). It also acts on the union of the spaces \( H^1(M, \Sigma; \mathbb{R}^2_\lambda) \) as \( \lambda \) varies over the finite set of homomorphisms from \( \pi_1(M - \Sigma) \). The development map is equivariant. Thus \( \text{Map}(M, \Sigma) \) preserves the affine structure on \( \mathcal{D} \). The moduli space of flat structures \( \mathcal{M} \) is the quotient of \( \mathcal{D} \) by the action of \( \text{Map}(\Sigma, \sigma) \). Since the action of \( \text{Map}(M, \Sigma) \) on \( \mathcal{D} \) is properly discontinuous, the quotient space inherits the structure of an affine orbifold.

The action of \( \text{Map}(M, \Sigma) \) on the union of the \( H^1(M, \Sigma; \mathbb{R}^2_\lambda) \) is volume-preserving: \( \text{Map}(M, \Sigma) \) preserves the integer lattice \( H^1(M, \Sigma; \mathbb{Z}^2) \) so that each element has determinant \( \pm 1 \). There is a translation invariant measure on \( H^1(M, \Sigma; \mathbb{R}^2) \) with respect to which the lattice is unimodular. Let \( \mu \) be the corresponding measure on \( \mathcal{M} \).

The vector space \( H^1(M, \Sigma; \mathbb{R}^2_\lambda) \) contains the lattice \( H^1(M, \Sigma; \mathbb{Z}^2) \). We can choose a volume form that determines a multiple of Lebesgue measure on \( H^1(M, \Sigma; \mathbb{R}^2_\lambda) \) for which the lattice is unimodular. The action of \( \text{Map}(M, \Sigma) \) on the union of spaces \( H^1(M, \Sigma; \mathbb{R}^2_\lambda) \) preserves this natural volume form since it preserves the lattices \( H^1(M, \Sigma; \mathbb{Z}^2) \). Let \( \mu \) denote the measure on \( \mathcal{M} \) obtained from the pull-back, the natural volume form by the developing map.

For the rest of the paper \( \mathcal{M} \) will refer to flat structures with area less than or equal to one.

Let \( \mathcal{D}^1 \subset \mathcal{D} \) be the subspace of flat structures with area one. Let \( \mathcal{M} \mathcal{D}^1 \) be the subspace of \( \mathcal{M} \) consisting of flat structures with area one.

We will define smooth measures on \( \mathcal{D}^1 \) and \( \mathcal{M} \mathcal{D}^1 \). We can define a scalar multiplication on \( \mathcal{D} \). If \( M \) is an element of \( \mathcal{D} \) then let \( rM \) be the structure in which all lengths are multiplied by \( r \). Since this operation commutes with the action of \( \text{Map}(M, \Sigma) \) it is well-defined on \( \mathcal{M} \). We define a measure \( \mu^1 \) on \( \mathcal{M} \mathcal{D}^1 \). If \( X \subset \mathcal{M} \mathcal{D}^1 \) then \( \mu^1(X) = \mu\{tx: x \in X, 0 \leq t \leq 1\} \).

A basic result proved in Section 10 is that the measure of \( \mathcal{M} \mathcal{D}^1 \) is finite. This was first proved in the case of the principal stratum and the strata of orientable flat structures in [M1, 2]. It was proved for all strata in [V2].

The group \( \text{GL}(2, \mathbb{R}) \) acts on each \( \mathcal{D} \). Let \( a \) be an element of \( \text{GL}(2, \mathbb{R}) \) and let \( \{\phi_i\} \) be an atlas for the flat structure \( q \). We can define a new atlas by \( \{a\phi_i\} \).

The change of coordinate functions for the new atlas also have the form \( v \mapsto \pm v + c \). In fact the new atlas also gives a flat structure which we call \( a(q) \). Positive scalar multiples of the identity act by rescaling the flat structure. In particular they change the area. The matrix which is the negative of the identity acts trivially on \( \mathcal{D} \). The action of \( \text{SL}(2, \mathbb{R}) \) induces an action of \( \text{PSL}(2, \mathbb{R}) = \text{SL}(2, \mathbb{R})/\pm I \).
The following subgroups of PSL(2, R) will play an important role in this paper.

\[ g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, \text{ and} \]

\[ r_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \]

We refer to \( g \) as the Teichmüller geodesic flow. In Teichmüller theory the image of \( q \) under \( g_t \) is called the terminal quadratic differential of the Teichmüller map determined by \( q \) with maximal dilatation \( e^t \). The action of \( r_\theta \) on \( q \) is the same as multiplying \( q \) by \( e^{2i\theta} \). The PSL(2, R) orbit of \( q \) is called the Teichmüller disc through \( q \). The PSL(2, R) action commutes with the action of Map(\( M, \Sigma \)) so that it descends to an action on the quotient \( \mathcal{M}/\mathcal{D} \).

The measure \( \mu \) on \( \mathcal{M}/\mathcal{D} \) is PSL(2, R)-invariant.

We describe more fully the special case of a stratum of spheres. We will make a normalization. There must be at least four points of \( \Sigma \) with cone angle \( \pi \) singularities. We distinguish three points of \( \Sigma \) and require each to have a cone angle \( \pi \) singularity. We will also require in Section 6 that none of the distinguished points be the endpoint of any curve that we shrink. The results of Section 8 will imply that after all curves are shrunk, the resulting sphere is a connected sum of a sphere with three cone angle \( \pi \) singularities and a sphere with small diameter with the cylinder \( A \) sewn into the sphere with small diameter.

2. A criterion for nonergodicity

Let \( q \) be a fixed flat structure on \( M \). Let \( E_0, E_1, \ldots \) be a sequence of metric cylinders in \( M \) so that \( E_n \) is disjoint from \( E_{n+1} \). Assume that the directions of the cylinders \( \theta_0, \theta_1, \ldots \) converge to \( \theta_\infty \). In this section we establish a criterion for showing that \( F_{\theta_\infty} \) is not ergodic.

Define a sequence of partitions \( P_n = \{ A_n, B_n \} \) of \( M \) as follows. If \( n \) is even then \( A_n = E_n \) and \( B_n = E_n^c \). If \( n \) is odd then \( A_n = E_n^c \) and \( B_n = E_n \). We can rotate the coordinate system so that \( \theta_\infty \) is vertical direction. Let \( h_n \) be the sum of the horizontal components of the boundary curves separating \( A_n \) and \( B_n \).

**Theorem 2.1.** Suppose that

i) \( \lim_{n \to \infty} h_n = 0 \),

ii) \( 0 < c \leq \mu(A_n) \leq c' < 1 \) for some \( c, c' \),

iii) \( \sum_{n=1}^{\infty} \mu(A_n \Delta A_{n+1}) < \infty \).

Then the vertical foliation is nonergodic.
Proof. By iii) and the Borel-Cantelli lemma the set of \( x \)'s which are in infinitely many sets \( A_n \Delta A_{n+1} \) has \( \mu \) measure 0. Accordingly, let
\[
A_\omega = \{ x : \exists N \text{ such that for } n \geq N, x \in A_n \},
\]
\[
B_\omega = \{ x : \exists N \text{ such that for } n \geq N, x \in B_n \}.
\]
Then \( A_\omega \) and \( B_\omega \) satisfy
1) \( \mu(M - (A_\omega \cup B_\omega)) = 0 \),
2) \( A_\omega \cap B_\omega = \emptyset \),
3) \( \mu(A_\omega \Delta A_n) \to 0 \) as \( n \to \infty \),
4) \( 0 < \mu(A_\omega) < 1 \).

To see 3) note that
\[
A_\omega \Delta A_n \subset \bigcup_{i=-n}^{\infty} A_i \Delta A_{i+1}
\]
so that
\[
\mu(A_\omega \Delta A_n) \leq \sum_{i=-n}^{\infty} \mu(A_i \Delta A_{i+1}).
\]

Hypothesis iii) implies that the right-hand side goes to zero as \( n \) goes to infinity. Item 4) follows from item 3).

We may assume that the vertical foliation of \( q \) is orientable. If it is not we may replace \( M \) by the orientation cover and replace \( A_\omega \) and \( B_\omega \) by their lifts to the orientation cover. Let \( f_t \) be the flow along vertical leaves.

Claim. For any \( t \), \( \mu(f_t(A_\omega) \Delta A_\omega) = 0 \).

Suppose on the contrary the measure is \( \delta > 0 \) for some \( t_0 \). Choose \( n \) large enough so that
\[
\mu(A_\omega \Delta A_n) < \delta/8 \quad \text{and} \quad t_0 h_n < \delta/8.
\]
This is possible by 3) and hypothesis i). Then
\[
\mu(f_{t_0}(A_n) \Delta A_n) \geq \delta - 2 \cdot \delta/8 = 3\delta/4
\]
by (1.1), the assumption, and the fact that \( f_{t_0} \) preserves \( \mu \). Thus in time \( t_0 \), \( 3\delta/8 \) of the measure of \( E_{2n+1} \) flows to its complement. However, in time \( t_0 \) at most \( t_0 h_n < \delta/8 \) of measure can cross the boundary of \( A_n \). This is a contradiction proving the claim.

We would like to conclude therefore that \( A_\omega \) is an invariant set under the flow which by (4) would prove the theorem. The claim only proves \( A_\omega \) a.e. invariant for every \( t \). The theorem is a consequence of the following lemma.
Lemma 2.2. Let $f_t$ be a flow on $X$ preserving a probability measure $\mu$. Suppose for every $t$, $\mu(f_t(A) \Delta A) = 0$. Then there is a set $A'$ invariant under $f_t$ so that $\mu(A \Delta A') = 0$.

Proof. Let $\lambda$ be Lebesgue measure on $\mathbb{R}$. Let $A' = \{x: f_t(x) \in A \text{ for a.e. } t\}$. Clearly $A'$ is flow invariant. We will show $\mu(A \Delta A') = 0$. Let

$$A_0 = \{(x, t): x \in A\} \text{ and } A_1 = \{(x, t): f_t(x) \in A\}.$$  

For every $t$, $\mu(x: (x, t) \in A_0 \Delta A_1) = \mu(f_t(A) \Delta A) = 0$. This implies $(\mu \times \lambda)(A_0 \Delta A_1) = 0$. By Fubini's theorem there is a set $X'$ of full $\mu$-measure so that for $x_0 \in X'$,

$$\lambda(t: (x_0, t) \in A_0 \Delta A_1) = 0.$$  

If $x_0 \in X' \cap A$ then $(x_0, t) \in A_0$ so that $\lambda(t: (x_0, t) \not\in A_1) = 0$. Therefore $\{t: f_t(x_0) \in A\}$ has full $\lambda$-measure and so $x_0 \in A'$. If $x_0 \in X' - A$ then $\{t: (x_0, t) \not\in A_1\}$ has full $\lambda$-measure so that $x_0 \not\in A'$. Thus $A \Delta A'$ is contained in the complement of $X'$ and so $\mu(A \Delta A') = 0$.

3. Hausdorff dimension

In this section we give a criterion which we will use to establish that the set of limit directions has positive Hausdorff dimension.

Suppose $C \subset [0, 1]$ is a Cantor set constructed as a decreasing intersection of closed sets $C_n$ each of which is a disjoint union of closed intervals $\Delta_n$ called the level $n$ sets. We suppose there are constants $k_2 > k_1 + 1 > 0$ such that:

1) Each level $n$ set $\Delta_n$ contains at least $|\Delta_n|^{-k_1}$ level $n + 1$ subsets.
2) The sizes of the level $n + 1$ subsets are at least $|\Delta_n|^{k_2}$.
3) The gap between successive level $n + 1$ subsets is at least as big as the largest level $n + 1$ subset.

Theorem 3.1. The Hausdorff dimension of the set $C$ constructed above is at least $\rho$ where

$$\rho = \frac{k_1}{2(k_2 - k_1)k_2}.$$  

Note that $\rho \leq 1/2$. We first prove

Lemma 3.2. Suppose $U_1, \ldots, U_\rho$ is a cover of $\Delta_n \cap C$ for some level $n$ set $\Delta_n$ such that at least half of the level $n + 1$ subsets of $\Delta_n$ have the property that
both it and a neighbor are intersected by the same $U_j$. Then for any $0 < \sigma < 1$

$$\sum_{i=1}^{p} |U_i|^\sigma \geq (1/4)^\sigma |\Delta_n|^\sigma (k_2 - k_1).$$

Proof. If a $U_j$ intersects two or more level $n + 1$ subsets, its linear measure is at least as big as the sum of the gaps between them which by 3) is at least $|\Delta_n|^{k_2}$ multiplied by the number of gaps covered. The number of gaps covered is at least $1/2$ the number of sets with the property that both it and a neighbor are intersected by the same $U_j$. This is in turn at least $1/2$ the total number of sets by hypothesis. This gives $\sum_{i=1}^{p} |U_i| \geq (1/4) |\Delta_n|^{k_2 - k_1}$. The lemma follows since the function $f(x) = x^\sigma$ is concave for $\sigma < 1$.

Proof of Theorem 3.1. We will show that the Hausdorff measure in dimension $\rho$, $H^\rho(C)$, is positive. Assume $H^\rho(C) = 0$. Let $\delta_m$ be the maximum linear size of any level $m$ set $\Delta_m$. By (2) $\lim_{m \to \infty} \delta_m = 0$. Choose $K$ large enough so that for $m \geq K - 1$,

$$(1/2)(1/4)^{\rho} \delta_m^{\rho(k_2 - k_1)k_2 - k_1} \geq 1.$$ 

This is possible since $\rho(k_2 - k_1)k_2 - k_1 = -k_1/2 < 0$. Now if $H^\rho(C) = 0$, for any $0 < \epsilon < 1$, there is a finite cover $U_1, \ldots, U_p$ of $C$ with

a) $|U_i| < \min_{\Delta_k} |\Delta_k|$

b) $\sum |U_i|^\rho < \epsilon$.

Now 3) and a) imply that each $U_i$ intersects at most one level $K$ set. Let $n_i \geq K$ be the largest integer such that $U_i$ intersects only one level $n_i$ set. Since $\lim_{m \to \infty} \delta_m = 0$, this number is well-defined. Let $N = \max n_i$. Let

$$F_n = \{\Delta_n; \text{ such that if } U_i \cap \Delta_n \neq \emptyset \text{ then } U_i \cap \Delta_n = \emptyset \text{ for all } \Delta_n \neq \Delta_n\}.$$ 

Then $F_{N+1} = \emptyset$ and all level $K$ sets are in $F_K$. Let $J$ be the largest number, $K \leq J \leq N$, such that for some $\Delta_{J-1}$, at least half of its level $J$ subsets are in $F_j$. Let $\Delta_{J-1}$ be a level $J - 1$ set with the above property. Let $\delta = |\Delta_{J-1}|$. Then by 1) and the definition, $\Delta_{J-1}$ contains at least $1/2 \delta^{-k_1}$ level $J$ subsets $\Delta_J$ which are in $F_j$. For each such $\Delta_J \in F_j$ at least half of its level $J + 1$ subsets are not in $F_{J+1}$ since $J$ was maximal. Associate to each $\Delta_J \in F_j$, the set of $U_i$ which intersect it. By the definition of $F_j$ the $U_i$ associated to different $\Delta_J \in F_j$ are distinct. For any such $\Delta_J \in F_j$ the lemma gives $\sum |U_i|^\rho \geq 1/4 |\Delta_J|^{\rho(k_2 - k_1)}$, where the sum is taken over the $U_i$ associated to $\Delta_J$. Since the $U_i$ associated to different $\Delta_J$ are distinct,

$$\sum_{i=1}^{p} |U_i|^\rho \geq \sum_{\Delta_J \in F_j} (1/4)^{\rho} |\Delta_J|^{\rho(k_2 - k_1)}.$$
By 2), \(|\Delta_j| \geq \delta^{k_z}\). Since there are at least \(1/2\delta^{−k_1}\) such \(\Delta_j\),
\[
\sum_{i=1}^{p} |U_i|^p \geq \frac{1}{2} (1/4)^p \delta^{\rho_{j}(k_z−k_1)} \geq \frac{1}{2} (1/4)^p \delta^{\rho_{j−1}(k_z−k_1)−k_1},
\]
the last inequality due to the fact that \(\delta_{j−1} \geq \delta\) and the fact that the exponent is negative. But the last quantity is at least 1 and so \(\sum_{i=1}^{p} |U_i|^p \geq 1\), a contradiction.

4. The Delaunay triangulation

In this section we construct the Delaunay triangulation of \(M\) and prove some properties that we will use later. The Delaunay triangulation is useful for two reasons. For a generic surface \(M\) the Delaunay triangulation is canonically determined by the geometry of \(M\). Secondly, we can bound the degeneracy of the triangles in the triangulation in terms of straightforward geometric invariants of \(M\). The construction of a Delaunay triangulation is sketched in [T] in a slightly different context. The construction is modelled on a construction in \(\mathbb{R}^2\) which is described in [E]. In this section we assume only that \(M\) is a compact manifold with a flat metric with cone-type singularities. We make no assumption on the holonomy of \(M\). In particular our results apply in the context of [T].

If \(\phi\) is a path in \(M\) from \(p\) to \(q\) then we write \(d_{\phi}(p, q)\) for the length of the shortest path from \(p\) to \(q\) in the relative homotopy class of \(\phi\). We use \(d(p, q)\) for the distance between \(p\) and \(q\) which can also be described as the minimum of \(d_{\phi}(p, q)\) over all paths \(\phi\). We write \(d(p, \Sigma)\) for the minimum of \(d(p, s)\) where \(s \in \Sigma\). A length-minimizing path from \(p\) to \(\Sigma\) is a path \(\phi\) from \(p\) to \(s \in \Sigma\) such that \(d_{\phi}(p, s) \leq d_{\phi'}(p, s')\) for any path \(\phi'\) from \(p\) to \(s' \in \Sigma\). Note that any length-minimizing path from \(p\) to \(\Sigma\) is homotopic relative to its endpoints to a geodesic segment.

Definition. We define \(d(M)\) to be the maximum value of \(d(x, \Sigma)\) for \(x \in M\). We refer to \(d(M)\) as the diameter of \(M\). The quantity \(d(M)\) is related to the standard notion of diameter of a metric space but is more useful for our purposes.

We begin by describing the decomposition of \(M\) into Voronoi cells. The Delaunay decomposition will be dual to this decomposition in the sense that Delaunay \(i\)-cells will correspond to Voronoi \((2 - i)\)-cells. The two-dimensional facets of the Voronoi decomposition are the path components of the set of points which have unique length-minimizing paths to \(\Sigma\). The one-dimensional facets are the path components of the set of points which have exactly two length-minimizing paths to \(\Sigma\). The vertices of this decomposition are points which have three or more length-minimizing paths to \(\Sigma\).
Let \( p \) be a point in \( M - \Sigma \). We write \( p' \) for the origin in \( T_p \), the tangent space to \( M \) at \( p \). For \( q' \in T_p \), we write \( D(q', r) \) for the disk of radius \( r \) centered at \( q \). Write \( s = s_p \) for \( d(p, \Sigma) \). Because \( M \) has singularities the exponential map cannot be defined on all of \( T_p \). We can however define an exponential map from the interior of \( D(p', s) \) to \( M \) which maps the center of the disk to \( p \). The image of this map avoids the singularities of \( M \). Since \( M - \Sigma \) has a metric of zero curvature the exponential map is a locally isometric immersion. Let \( \iota_p : D(p', s) \to M \) denote the continuous extension of this map to the closed disk. For each point \( \sigma' \in \partial D(p', s) \) that maps to a singularity, the image of the radius \( p'\sigma' \) under \( \iota_p \) is a length-minimizing path to \( \Sigma \). Furthermore, every length-minimizing path from \( p \) to \( \Sigma \) has this form.

In the following proof and the remainder of this section it will be useful to define “exponential maps” on subsets of \( T_p \) more general than disks. We give the construction here. Let \( \overline{M - \Sigma} \) denote the universal cover of \( M - \Sigma \). Fix a point \( \tilde{p} \) in \( \overline{M - \Sigma} \) that maps to \( p \) under the covering projection \( \pi \). The developing map \( \overline{M - \Sigma} \) to \( T_p \) is defined to be the unique local isometry that takes \( \tilde{p} \) to \( p' \in T_p \) and has the identity as its derivative at \( \tilde{p} \). If \( X \) is any connected subset of \( \overline{M - \Sigma} \) on which the developing map is injective and \( X' \) is the image of \( X \) we can define a map \( \iota : X' \to M - \Sigma \) by \( \iota = \pi \circ \text{dev}^{-1} \). We can extend this map by continuity to a map from the closure of \( X' \) to \( M \). We say that a subset \( X \) of \( \overline{M - \Sigma} \) is star-like with respect to one of its points \( q \) if each point of \( X \) is connected to \( q \) by a unique geodesic segment. If \( X \) is star-like with respect to some point then the developing map is injective on \( X \) and the above procedure can be carried out. This criterion will suffice to justify the extension of the exponential map in all of our applications.

**Proposition 4.1.** The set of facets for the Voronoi decomposition is finite. The 0-facets are points. The 1-facets are geodesic one-manifolds.

**Proof.** Let \( p \in M - \Sigma \). The Voronoi decomposition of \( M \) induces a decomposition of \( D(p', s_p/3) \). We define a facet in \( D(p', s_p/3) \) to be a component of the inverse image of a facet in \( M \). We will prove that for each \( p \in M - \Sigma \) the induced decomposition of \( D(p', s_p/3) \) is finite. This fact together with the compactness of \( M \) will establish the proposition.

Let \( q' \) be an element of \( D(p', s_p/3) \). Write \( q \) for \( \iota_p(q') \). We can use the exponential function at \( q \) to extend the domain of the map \( \iota_p \) from \( D(p', s_p) \) to \( D(p', s_p) \cup D(q', s_q) \). In fact there is a unique extension so that the new map is still a local isometry on the interior of its domain. There is some ambiguity in our notation. The notation \( D(q', s_q) \) can represent a subset of \( T_p \) or \( T_q \). In fact
the map $\iota_q$ defined on $D(q', s_q) \subset T_q$ agrees with the extension of the map $\iota_p$ defined on $D(p', s_p) \subset T_p$, so perhaps this ambiguity will cause no confusion.

Write $E = \bigcup D(q', s_q)$ where the union is taken over all $q' \in D(p', s_p/3)$. Let $\nu$ denote the extension of $\iota_p$ to $E$. We will show that if $q' \in D(p', s_p/3)$ then $D(q', s_q) \supset D(p', s_p/3)$. We have $s_p = d(p, \Sigma) \leq d(p, q) + d(q, \Sigma)$. By construction $d(p, q) \leq s_p/3$ so that $s_p \leq s_p/3 + s_q$ and $s_q \geq 2s_p/3$. If $x' \in D(p', s_p/3)$ then

$$d(x', q') \leq d(x', p') + d(p', q') \leq d(x', p') + d(p, q) \leq s_p/3 + s_p/3 = 2s_p/3.$$  

(The second inequality holds because the exponential map does not increase distances.) This gives $x \in D(q', 2s_p/3) \subset D(q', s_q)$.

Let $S'$ denote the set of points $\sigma' \in \partial E$ which are mapped to elements of $\Sigma$ by $\nu$. For any $\sigma' \in S'$ and $q' \in D(p', s_p/3)$ the line segment $\sigma'q'$ lies in $E$. This is because $\sigma' \in \partial D(x', s_x)$ for some $x' \in D(p', s_p/3)$ and by the previous assertion $q' \in D(x', s_x)$. So $\sigma'$ and $q'$ are in the convex set $D(x', s_x)$.

Any length-minimizing path from any point $\iota(q')$ to $\Sigma$ for $q' \in D(p', s_p/3)$ is the image of some straight line segment from some $\sigma' \in S'$ to $q'$.

There is an upper bound to the length of line segments $\sigma'q'$. Any point in $E$ is in some set $D(q', s_q)$ where $d(p, q) \leq s_p/3$. So the distance to $p$ is at most $s_p/3 + s_q$. We have

$$s_q = d(q, \Sigma) \leq d(p, q) + d(p, \Sigma) \leq s_p/3 + s_p = 4s_p/3.$$  

Now every point in $E$ is within distance $5s_p/3$ from $p$. The set of relative homotopy classes of curves from $p$ to $\Sigma$ with length at most $C$ is finite. We conclude that $S'$ is a finite set.

We can construct a Voronoi decomposition of $T_p$ corresponding to the points $\sigma'$. The two-dimensional facets of the Voronoi decomposition are the path components of the set of points which have unique length-minimizing paths to $S'$. The one-dimensional facets are the path components of the set of points which have exactly two length-minimizing paths to $S'$. The vertices of this decomposition are points which have three or more length-minimizing paths to $S'$ (see [E, p. 293–298] for more details). The number of facets in this decomposition is bounded by the number of subsets of $S'$ and is therefore finite. The one-dimensional facets are connected subsets of the perpendicular bisector of a pair of points and are therefore geodesic one-manifolds.

For any point $q'$ in $D(p, s_p/3)$ the length-minimizing paths from $\iota(q')$ to $\Sigma$ are precisely the images of length-minimizing paths from $q'$ to $S'$. It follows that $\nu$ defines an isomorphism of decompositions between the decomposition of
$D(p, s_p/3)$ induced by the Voronoi decomposition of $T_p$ and the decomposition of $\mathcal{D}(D(p, s_p/3))$ induced by the Voronoi decomposition of $M$. In particular there are only finitely many Voronoi cells in $\mathcal{C}(D(p', s_p/3))$. The 1-facets in the Voronoi decomposition of $T_p$ are subsets of a line of points equidistant between two elements of $S'$. It follows that the 1-facets of the Voronoi decomposition of $\mathcal{D}(D(p, s_p/3))$ are geodesic one-dimensional manifolds.

Q.E.D.

For any $p \in M$ we can define a Delaunay facet. Let $S'$ be the set of points in $\partial D(p', s_p)$ that map to points in $P$. Let $H_p \subset T_p$ be the convex hull of $S'$. If $p$ is a Voronoi 0-facet then $H_p$ is two dimensional and the corresponding facet is is the image under $\mathcal{C}(D(p', s_p/3))$ of the interior of $H_p$. If $p$ is in a Voronoi 1-facet then $H_p$ is a segment and the corresponding Delaunay facet is the image of the segment minus its endpoints. It is not hard to show that the Delaunay facet corresponding to $p$ depends only on the Voronoi facet containing $p$ and not on $p$. If $p$ is in a Voronoi 2-cell then $S'$ has only one element $\sigma'$ and the corresponding Delaunay 0-facet is $\mathcal{C}(\Sigma')$.

**Lemma 4.2.** The map $\mathcal{C}(D(p))$ is injective and for distinct points $p$ and $q$ the sets $\mathcal{C}(\mathcal{D}(p))$ and $\mathcal{C}(\mathcal{D}(q))$ are disjoint.

**Proof.** Given a segment in $M - \Sigma$ with $p$ as one endpoint we define its lift to $T_p$ by lifting the segment to $M - \Sigma$ and then applying the developing map. We will use this construction several times in the following proof.

If $\mathcal{C}(D(p))$ fails to be injective then we can assume that there are points $y_p \neq x_p$ with $\mathcal{C}(y_p) = \mathcal{C}(x_p) = x$. The line segment $x_p x'$ is contained in $T_p$. The image of this segment $\mathcal{C}(x_p x')$ determines a geodesic path from $p$ to $x$. We can isometrically lift this path to $T_x$ so that $x$ lifts to $x'$. Let $p_x \in T_x$ denote the lift of the other endpoint of this path. The segment $\mathcal{C}(y_p x')$ determines a distinct geodesic path from $p$ to $x$. Let $q_x \in T_x$ denote the lift of the other endpoint of this path. Since the paths $\mathcal{C}(x_p x')$ and $\mathcal{C}(y_p x')$ are not equal we conclude that $p_x \neq q_x$.

If $p \neq q$ but $\mathcal{C}(\mathcal{D}(p))$ and $\mathcal{C}(\mathcal{D}(q))$ intersect then let $x_p$ be a point in $H_p$ and let $y_q$ be a point in $H_q$ so that $\mathcal{C}(x_p) = \mathcal{C}(y_q) = x$. As before we can lift the path $\mathcal{C}(x_p x')$ to $T_x$ so that $x$ lifts to $x'$. Let $p_x$ be the lift of the other endpoint of this path. We can also lift $\mathcal{C}(y_q x')$ to $T_x$ so that $x$ lifts to $x'$. Let $q_x$ be the other endpoint of this lift.

In either case, if $\mathcal{C}(D(p))$ fails to be injective or if $p \neq q$ but $\mathcal{C}(\mathcal{D}(p))$ and $\mathcal{C}(\mathcal{D}(q))$ intersect, the exponential map $\mathcal{C}(D(p))$ can be extended to the set $D(p_x, s_p) \cup D(q_x, s_q) \subset T_x$. In both cases $p_x \neq q_x$. The remainder of this argument applies to both cases.
Let \( \sigma'_i \) be the points on the boundary of \( D(p_x, s_p) \) that map to singularities under \( \iota \). Write \( H'_p \) for the convex hull of the set \( \{\sigma'_i\} \). Let \( \tau'_j \) be the points on the boundary of \( D(q_x, s_q) \) that map to singularities under \( \iota \). Write \( H'_q \) for the convex hull of the points \( \tau'_j \). We have \( \iota(H'_p) = \iota_p(H_p) \) and \( \iota(H'_q) = \iota_q(H_q) \).

The points \( \sigma'_i \) lie in \( (\partial D(p_x, s_p)) - D(q_x, s_q) \). In particular the set \( (\partial D(p_x, s_p)) - D(q_x, s_q) \) is not empty. We conclude that \( D(p_x, s_p) \) is not strictly contained in \( D(q_x, s_q) \). The points \( \tau'_i \) lie in \( (\partial D(q_x, s_q)) - D(p_x, s_p) \). This implies that \( D(q_x, s_q) \) is not strictly contained in \( D(p_x, s_p) \). The disks are not disjoint because they both contain \( x' \). The disks are not equal because they have distinct centers. We conclude that the circles \( \partial D(p_x, s_p) \) and \( \partial D(q_x, s_q) \) intersect in two points. The line between these points separates all points \( \sigma'_i \) from all points \( \tau'_j \). Thus the interiors of the convex hulls \( H'_p \) and \( H'_q \) are disjoint. This contradicts the assumption that \( x' \) is a point in common.

We wish to show that the closures of the Delaunay 2-cells cover \( M \). The proof requires two lemmas.

**Lemma 4.3.** For each Delaunay 1-cell \( C_1 \) there are Delaunay 2-cells \( C_2 \) and \( C'_2 \) such that \( C_1 \cup C_2 \cup C'_2 \) contains a neighborhood of the interior of \( C_1 \).

**Proof.** Let \( p \) be an interior point of the Voronoi 1-cell \( V_1 \) dual to \( C_1 \). The map \( \iota_p \) gives an immersion from \( D_p \) to \( M \). There are points \( \sigma'_0 \) and \( \sigma'_1 \) in \( \partial D_p \) so that \( \sigma'_0 \sigma'_1 \) maps to \( C_1 \). Let \( \phi: [a, b] \rightarrow M \) parametrize the maximal geodesic through \( p \) in the direction perpendicular to the direction of \( C_1 \). (Generically we will have \( a = -\infty \) and \( b = +\infty \).) Assume that \( \phi(0) = p \). Let \( \phi': [a, b] \rightarrow T_p \) parametrize the perpendicular bisector to the segment between \( \sigma'_0 \) and \( \sigma'_1 \). We can extend the definition of \( \iota \) so that \( \iota \) is defined on the image of \( \phi' \) and \( \iota \circ \phi' = \phi \). The Voronoi cell \( V_1 \) is contained in the image of \( \phi \). In fact \( V_1 \) is the image of the maximal interval \( (a', b') \) containing 0 and consisting of points with exactly two length-minimizing paths to \( \Sigma \). A priori either \( a' \) or \( b' \) could be infinite. We will show that both are finite. The image of the segments \( \sigma'_0 \phi'(t) \) and \( \sigma'_1 \phi'(t) \) under \( \iota \) are the length-minimizing paths from \( \phi'(t) \) to \( \Sigma \). We have \( d(\phi(t), \Sigma) = \sqrt{t^2 + (c/2)^2} \) where \( c = d(\sigma'_0, \sigma'_1) \). Since \( M \) is compact the function \( d(x, \Sigma) \) is bounded on \( M \). We conclude that \( b' \neq \infty \). The same argument shows that \( a' \neq -\infty \). The points \( \phi(a') \) and \( \phi(b') \) have at least one additional length-minimizing path to \( \Sigma \) and so are Voronoi 0-cells. Write \( \alpha \) for \( \phi'(a') \) and \( \beta \) for \( \phi'(b') \). The polygons \( H_\alpha \) and \( H_\beta \) can be thought of as subsets of \( T_p \). Both these sets contain the segment \( \sigma'_0 \sigma'_1 \) but they lie on opposite sides of the line determined by this segment. It follows that \( H_\alpha \cup H_\beta \) contains a neighborhood of the interior of \( \sigma'_0 \sigma'_1 \). Let \( C_2 \) and \( C'_2 \) be the corresponding
Delaunay 2-cells. Since \( \iota \) is a local homeomorphism the conclusion of the lemma follows.

**Theorem 4.4.** Every point in \( M \) is contained in a unique Delaunay cell. Every Delaunay cell is isometric to a polygon inscribed in a disk of radius less than or equal to \( d(M) \).

**Proof.** The geometric description of Delaunay cells follows immediately from their construction. The uniqueness of the Delaunay cell containing a given point has already been proved. We consider existence. Consider the subset of points in \( M - \Sigma \) which are contained in Delaunay 1-cells and 2-cells. By Lemma 2 this set is open in \( M - \Sigma \). By Proposition 1 the union of a 2-cell and its boundary 1-cells is closed and there are only finitely many cells; thus this set is closed in \( M - \Sigma \). We conclude that the set is all of \( M - \Sigma \). The points in \( \Sigma \) are Delaunay 0-cells. Thus all points of \( M \) are accounted for. Q.E.D.

The cells of the Delaunay decomposition are isometric to convex polygons inscribed in disks. We can subdivide these polygons into triangles without adding any new vertices. We call the result of any such subdivision a Delaunay triangulation. This terminology agrees with that of [T] in that we have decomposed \( M \) into triangles but the decomposition is not a triangulation in the combinatorial sense in that triangles are not determined by their vertices.

## 5. Surfaces with large diameter

The purpose of this section is to prove that surfaces of large diameter must contain metric cylinders. Similar results have been proved before using information about compactifications of Teichmüller space. Our approach to this question here is by way of differential geometry. This approach allows us to get the explicit constants in Theorem 5.3. In this section we reinstate the hypothesis that the holonomy of \( M \) is contained in \( \{ \pm 1 \} \).

We say that \( M \) contains a disk of radius \( r \) at \( p \) if there is an isometric embedding of the open disk of radius \( r \) in \( \mathbb{R}^2 \) into \( M \) which takes 0 to \( p \). We define the injectivity radius at \( p \) to be the positive real number \( r_p \) which is the supremum of real numbers \( r \) for which \( M \) contains a disk of radius \( r \) at \( p \).

As before let \( s_p \) denote the distance from \( p \) to the closest point of \( \Sigma \).

**Lemma 5.1.** Let \( p \) be a point of \( M \) with \( s_p > r_p \); then there is a closed geodesic of length \( 2r_p \) passing through \( p \).

**Proof.** We can use the exponential map to construct an immersion of the open disk of radius \( s_p \) into \( M \). Since \( r_p < s_p \) there is an immersion \( \iota \) of the closed disk of radius \( r_p \) into \( M \) with center \( p \). This immersion cannot be an
embedding (for then we could find larger embedded open disks centered at p); so there must be two points q and q' on the circle of radius r in \( \mathbb{R}^2 \) which map to the same point in M. We construct a map \( \sigma \) from the segment \([0, 2r_p]\) to M. The map sends the interval \([0, r_p]\) isometrically to the image of the radius from 0 to q so that 0 maps to p and \( r_p \) maps to \( \iota(q) \). The map sends \([r_p, 2r_p]\) isometrically to the image of the radius from 0 to \( q' \) so that \( r_p \) maps to \( \iota(q') = \iota(q) \) and \( 2r_p \) maps to \( p \). The map \( \sigma \) is continuous. It sends both endpoints of \([0, 2r_p]\) to \( p \) and so it defines a loop. The loop is a geodesic except perhaps at the points 0 and \( r_p \). In fact we will show that the loop is smooth at 0 and \( r_p \) and hence a closed geodesic. To verify smoothness at \( r_p \) we need to verify that the outward pointing normal vector at \( q \) and the inward pointing normal vector at \( q' \) have the same image under \( D\sigma \). Choose neighborhoods \( U \) of \( q \) and \( U' \) of \( q' \) in the closed disk of radius \( r_p \). The interiors of these disks map disjointly. It follows that the outward pointing normal vector at \( q \) maps to the inward pointing normal vector at \( q' \).

To verify smoothness at 0 we will calculate the holonomy around the loop \( \sigma \). Let \( \theta \) be the measure of the angle \( \angle q0q' \). Let \( v \) and \( v' \) be the unit vectors at 0 tangent to the radii 0q and 0q' respectively. If we parallel translate the vector \( Du(v) \) along the path \( \sigma \) it returns to \( Du(-v') \). The holonomy along \( \sigma \) is a rotation. The angle of the rotation is the angle between \( v \) and \( -v' \) which is \( \theta - \pi \). Since \( M \) is a flat surface the holonomy along any path in M must be rotation by 0 or \( \pi \) by assumption so that \( \theta \) must be 0 or \( \pi \). Since \( q \) and \( q' \) are distinct we conclude that \( \theta = \angle q0q' \neq 0 \) so that \( \theta = \pi \). This shows that the loop \( \sigma \) is actually a closed geodesic of length \( 2r_p \).

The following lemma is well-known in the theory of quadratic differentials. We include a proof for the reader who may not be familiar with this literature.

**Lemma 5.2.** Let \( \sigma \) be a closed geodesic. Assume that \( M \) contains at least one singularity. Assume that for every point \( p \) on \( \sigma \) the quantity \( s_p \) is greater than a fixed quantity \( s \). Then \( \sigma \) is a waist curve in an embedded annulus with height at least \( 2s \).

**Proof.** Since \( M \) is oriented, the normal bundle to \( \sigma \) is oriented. We can identify it with \( \sigma \times \mathbb{R} \). There is an exponential map from the normal bundle into \( M \) which is defined on \( \sigma \times (-s, s) \). If this map fails to be injective we can find an \( r \) with \( 0 < r < s \) and so that the exponential map is injective on \( \sigma \times (-r, r) \) but not on \( \sigma \times (-r, r) \). Let \( p \) and \( q \) be points in \( \sigma \times (-r, r) \) which map to the same point in M. The curves \( \sigma \times \{r\} \) and \( \sigma \times \{-r\} \) map to geodesic curves. Using the fact that the exponential map is injective on \( \sigma \times (-r, r) \) we see that the tangent line to the curve through \( p \) maps to the image of the tangent line through \( q \). Since a geodesic through a point is determined by its direction we
conclude that the entire geodesic containing \( p \) is identified with the entire geodesic containing \( q \). If there are points \( p \) and \( q \) on opposite boundaries which are identified then the image of the exponential map has no boundary and since \( M \) is connected it must be all of \( M \). On the other hand the image consists of points within distance \( s \) of \( \sigma \) and hence contains no singularities. We have assumed that \( M \) does contain a singularity. This yields a contradiction. The remaining possibility is that there are points \( p \) and \( q \) on the same boundary component that are identified and the other boundary component is mapped injectively. In this case the image of the exponential map must be a cross cap. The fact that it contains a parallel line field consisting of lines parallel to \( \sigma \) shows that it has Euler characteristic zero. The only connected surface with one boundary component and Euler characteristic zero is the cross cap. Since \( M \) is orientable it cannot contain a subsurface homeomorphic to the cross cap. This contradiction proves the lemma.

**Definition.** Let \( s = \sqrt{2/\pi} \). Let \( B_s \) be the set of points in \( p \in M \) with \( s_p \leq s \).

**Theorem 5.3.** Let \( M \) be a flat manifold with \( \text{area}(M) \leq 1 \). The complement of \( B_s \) is contained in a union of disjoint metric cylinders.

**Proof.** We say a metric cylinder is maximal if it is not contained in any larger metric cylinder. Consider the set of maximal metric cylinders with the property that their circumference is less than their height. Two distinct cylinders in this set are disjoint. In fact if \( C \) is such a cylinder and \( p \in C \) then the unique shortest curve through \( p \) is a waist curve of \( C \); any other geodesic through \( p \) must cross both boundaries of \( C \) and hence have length greater than or equal to the height of \( C \). If \( p \) is contained in some other cylinder \( C' \) in this collection then \( C \) and \( C' \) have a waist curve in common. Two maximal cylinders sharing the same waist curve must be equal and so \( C = C' \). To prove the theorem it suffices to show that every point outside of \( B_s \) is contained in such a cylinder.

Let \( p \) be a point outside of \( B_s \); thus \( s_p > s \) by definition. Let \( r_p \) be the injectivity radius at \( p \). The surface \( M \) contains an embedded disk of radius \( r \) at \( p \). By comparing the area of the disk to the area of \( M \) we conclude that \( \pi r_p^2 < 1 \) so that \( r_p < 1/\sqrt{\pi} \). Since \( s_p > s > 1/\sqrt{\pi} > r_p \), Lemma 5.1 implies that there is a geodesic \( \sigma \) of length \( L \) through \( p \) with \( L < 2/\sqrt{\pi} \). Let \( S \) be the distance from \( \sigma \) to the nearest singularity. By Lemma 5.2 there is an embedded metric cylinder around \( \sigma \) with height at least \( 2S \). The boundary of this cylinder must contain a singularity \( q \). The maximum distance between \( p \) and a point on the boundary of \( S \) is given by the Pythagorean formula as \( \sqrt{1/\pi + S^2} \). On the other hand, since \( p \) is outside \( B_s \), the distance from \( p \) to \( q \) is greater than \( s \). This
gives \( \sqrt{1/\pi + S^2} \geq s_p > s \) which implies \( S > 1/(\sqrt{\pi}) \). Thus the height of the cylinder is at least \( 2S = 2/\sqrt{\pi} \). Therefore the height is greater than the circumference.

**Proposition 5.4.** Let \( s \) be as above. The Delaunay triangulation of \( M \) consists of edges which either have length less than or equal to \( 2s \) or which cross a cylinder \( C \) whose height \( h \) is greater than its circumference \( c \). If an edge crosses \( C \) then its length is bounded below by \( h \) and bounded above by \( \sqrt{h^2 + c^2} \).

**Proof.** Given a Delaunay edge let \( p \) be a point in the dual Voronoi 1-cell. We consider two cases. In the first case \( p \) is in \( B_s \). The disk \( D_p \) has radius \( d(p, \Sigma) \leq s \). The Delaunay edge is the image under \( \iota_p \) of a secant in \( D_p \). The length of this secant is bounded above by the diameter of \( D_p \) which is bounded above by \( 2s \).

We now consider the case when \( p \in B_s \) where we may assume, by Theorem 5.3, that \( p \in C \) for some maximal cylinder \( C \). Let \( r \) be the circumference of the cylinder \( C \). Let \( R \) be a rectangle inscribed in \( D_p \), with vertices on the boundary of \( D_p \) so that one pair of opposite sides of \( R \) is parallel to the waist curve and one pair of opposite sides is perpendicular to the waist curve and such that the sides parallel to the waist curve have length \( r \). To fix notation assume that the waist curve is vertical. Since \( r \) is less than the radius of \( D_p \) this construction can be made. Any vertical segment in \( R \) maps to a waist curve in \( C \). In particular its opposite endpoints are identified. Thus the two horizontal sides of \( R \) are mapped to the same segment in \( M \). The image of \( R \) is a cylinder. In \( T_p \), extend the sides of \( R \) which are parallel to the waist curve to lines. Let \( R' \) be the set of points in \( T_p \) which are between these lines. The points in \( D_p \cap R' \) map to the image of \( R \). In particular no points in the interior of this set map to elements of \( \Sigma \) under \( \iota_p \). The secant that maps to the Delaunay edge connects point \( s'_0 \) and \( s'_1 \) that map to elements of \( \Sigma \). If these points are contained in the same component of \( D_p - R' \) then the distance between them is at most the height of \( R \) which is less than \( 2/\sqrt{\pi} \). If these points are in different components of \( D_p - R' \) then the secant curve crosses \( R \) and its image under \( \iota_p \) crosses \( C \). The length of this edge is bounded below by the height of \( C \). The length is bounded above by the diameter of \( D_p \). The diameter of \( D_p \) is equal to the diameter of the inscribed rectangle \( R \). One dimension of \( R \) is equal to the circumference of \( C \); the other is bounded above by the height of \( C \). This gives the upper bound of \( \sqrt{h^2 + c^2} \).

**Corollary 5.5.** Let \( M \) be a flat surface with diameter \( d \). There exists a constant \( C \) so that if \( d \geq s = \sqrt{2}/\pi \) then \( M \) contains a cylinder of height at least \( Cd \).
Proof. Our definition of $d$ implies that there is a point in $p \in M$ with $s_p = d$. Since $d \geq s > 1/\sqrt{\pi}$, the point $p$ is contained in a cylinder $C$. Let $2S$ be the height of the cylinder. As in the proof of Theorem 5.3 we have $\sqrt{1/\pi + S^2} \geq s_p = d$. This gives $S > \sqrt{d^2 - 1/\pi}$. Let $g(x) = \sqrt{x^2 - 1/\pi}$. The function $g(x)/x$ is increasing. If $C = 2g(s)/s > 0$, we have $2S \geq 2g(d) \geq Cd$.

**Corollary 5.6.** There is a constant $C_1$ depending only on the stratum, such that when the diameter $d$ of a flat structure $M$ is greater than $s = \sqrt{2/\pi}$, the circumference $c$ of the shortest metric cylinder on $M$ satisfies $d \leq C_1/c$.

Proof. Let $k$ denote the maximum number of geodesic segments, disjoint except for endpoints. It is easy to prove ([K-M-S]) that the number $k$ depends only on the stratum, not on the flat structure $M$ in the stratum. Each boundary component of a metric cylinder is made up of at most $k$ geodesic segments. Theorem 5.3 says that under the hypothesis of the corollary, there must be a metric cylinder $A$ and a point $p \in A$ such that the distance of $p$ to the nearest singularity is $d$. Let $d'$ be the distance of $p$ to the nearest singularity on the boundary of $A$. Then $d \leq d'$. The boundary of $A$ consists of at most $k$ segments each of whose length is at most $2s$, by Proposition 5.4. The circumference $c_0$ of $A$ therefore satisfies

$$c \leq c_0 \leq 2ks.$$ 

Let $h$ be the height of $A$. The Pythagorean theorem gives

$$d^2 \leq d'^2 \leq h^2 + c_0^2.$$ 

The area of any cylinder is at most 1. Therefore $hc_0 \leq 1$. This gives

$$d^2 \leq c_0^2 + \frac{1}{c_0^2} \leq \frac{1}{c_0^2} \left(1 + (2ks)^4\right) \leq \frac{1}{c^2} \left(1 + \frac{144k^4}{\pi^2}\right).$$

Take $C_1 = (1 + 144k^4/\pi^2)^{1/2}$. This gives the corollary.

**Corollary 5.7.** Let $M$ be a flat surface with diameter $d$. Let $\beta$ be any geodesic segment joining singularities whose length is minimal among all geodesic segments that join singularities. Assume that $\beta$ is not a simple closed curve which bounds a metric cylinder. Let $\delta$ be the length of the shortest geodesic segment $\gamma$ joining singularities not equal to $\beta$. Then there is a constant $C_5$ such that $d \leq C_5/\delta$. 


Proof. It is possible \( \gamma \) also has minimal length. Assume first that \( d \geq s = \sqrt{2} / \pi \). Corollary 5.6 says

\[
(1) \quad d \leq C_1 / c
\]

where \( c \) is the circumference of the shortest metric cylinder \( A \). If the boundary of \( A \) is a single segment of length \( c \), then by assumption \( \beta \) is not the boundary of \( A \). Since \( \gamma \) is the shortest segment not equal to \( \beta \), the boundary is at least as long as \( \gamma \) and we have \( c \geq \delta \). Therefore (1) says \( d \leq C_1 / \delta \). If the boundary of \( A \) is made up of more than one segment, one of those segments is not equal to \( \beta \) and therefore has length at least \( \delta \). This again gives \( c \geq \delta \) and by (1) again, \( d \leq C_1 / \delta \).

We can therefore assume that \( d \leq s \). The Delaunay triangulation of the flat surface is by edges with length at most \( 2d \). Since \( d \leq s \), we have a triangulation by edges of length at most \( 2s \) and \( \delta \leq 2s \). Thus \( d \leq s \leq s^2 / \delta \). With \( C_5 = s^2 \), this proves the corollary in this case as well.

6. Degenerating surfaces

The bundle of flat structures for a fixed surface \( M \) is a union of strata of different dimensions. In this section we define what it means for one stratum to be a face of another stratum. This relationship is motivated by the description of how strata fit together in the bundle of flat structures but it is more general. If a stratum \( \mathcal{M}' \) is obtained from a stratum \( \mathcal{M} \) by coalescing two singularities then \( \mathcal{M}' \) is in the closure of \( \mathcal{M} \) in the bundle of flat structures. If a stratum \( \mathcal{M}' \) is obtained from a stratum \( \mathcal{M} \) by pinching a curve and then desingularizing the resulting surface, then the two strata lie in different bundles of flat structures and we cannot interpret the relationship between strata in the same way.

The purpose of this section is to make this operation of collapsing along a segment precise. In Theorem 6.1 we show the following. Given a stratum \( \mathcal{M} \) there is a subset of large measure so that for \( q \) in this subset there is a well-defined operation that collapses the shortest segment in \( q \) to give a new flat structure \( q' \) in a stratum \( \mathcal{M}' \) which is a face of \( \mathcal{M} \). Moreover with respect to the affine structures on \( \mathcal{M} \) and \( \mathcal{M}' \) the collapsing map is an affine projection. We will also show that there is a map \( g \) from \( q \) to \( q' \) which has small distortion away from the curve which is collapsed.

In defining the collapsing operations we will restrict ourselves to those collapsing operations which we actually use in our proof. In particular we will only consider collapses in which the set of nonsingular points in the resulting surface is connected.
Now suppose $\mathcal{D}(M, \Sigma, \sigma, \pm)$ is a stratum. As in Section 1, let $\tau$ be a triangulation of $\mathcal{M}$ with vertices at points of $\Sigma$. Recall the definition of $N(\tau, \eta)$ as the set of flat structures in $\mathcal{M}\mathcal{D}$ with holonomy homomorphism $\eta$ for which the edges of $\tau$ are geodesic segments.

**Definition.** Let $M(\tau, \eta)$ be the set of $q \in \mathcal{D}$ such that $\tau$ is the Delaunay triangulation of $q$.

The sets $M(\tau, \eta)$ form coordinate charts for $\mathcal{D}$. Notice that $M(\tau, \eta) \subset N(\tau, \eta)$. Let $\prec$ represent an ordering of the homology classes of edges of the triangulation $\tau$. We can partition $M(\tau, \eta)$ up to a set of measure zero into disjoint subsets $M(\tau, \eta, \prec)$ each with the property that for any two nonhomologous edges $e_1, e_2$ and all flat structures in the subset, $|e_1| < |e_2|$ if the homology class of $e_1$ is less than the homology class of $e_2$ in the ordering $\prec$.

The action of any nontrivial element of $\text{Map}(M, \Sigma)$ on $\mathcal{D}$ maps each set $M(\tau, \eta, \prec)$ to a disjoint set. Therefore we may consider the set $M(\tau, \eta, \prec)$ as a subset of $\mathcal{M}\mathcal{D}$. Except for a set of zero measure, a finite number of $M(\tau, \eta, \prec)$ cover $\mathcal{M}\mathcal{D}$. We call these Delaunay triangulation pieces. Our map to lower dimensional strata will be defined on certain subsets of these $M(\tau, \eta, \prec)$.

Suppose $\kappa$ and $\delta$ are a pair of numbers such that $\kappa < \delta$.

**Definition.** Let $M^\delta_\kappa(\tau, \eta, \prec)$ be the set of flat structures in $M(\tau, \eta, \prec)$ such that:

a) The shortest edge has length at most $\kappa$ and every nonhomologous edge has length at least $\delta$.

b) The shortest edge does not divide the surface and its union with a homologous edge does not divide. In particular, the shortest edge is not a boundary component of a metric annulus.

c) If genus $M = 1$, the shortest edge or pair of homologous edges does not form a simple closed curve.

d) If genus $M = 0$, the shortest edge does not have any of the 3 distinguished points as an endpoint.

We now define what it means for a stratum to be a face of $M^\delta_\kappa(\tau, \eta, \prec)$ and we define a retraction from $M^\delta_\kappa(\tau, \eta, \prec)$ to that face. There are two basic types of faces depending on whether the shortest segment $\beta$ (resp. pair of shortest homologous segments $\beta_1, \beta_2$) on the flat surfaces in $M^\delta_\kappa(\tau, \eta, \prec)$ forms a simple closed curve or not. If it does not, the face will be called *coalescing* and we will say we have coalesced the surface along the segment $\beta$ (resp. $\beta_1, \beta_2$). If it does, the face will be called *pinching* and we will say we have pinched along the simple closed curve. In each case we will define the one-segment case first, indicating the differences for the two-segment case.
**Definition.** \( \mathcal{D}' = \mathcal{D}(M', \Sigma', \sigma', \pm ) \) is a coalescing face of \( M^g_\ast(\tau, \eta, <) \) along the segment \( \beta \) if:

a) genus \( M' = \text{genus } M \).

b) There is a map \( g: M \to M' \) which collapses \( \beta \) to a point \( x' \in \Sigma' \) and which is a bijection from \( \Sigma - x_1 - x_2 \) to \( \Sigma' - x' \), where \( x_1, x_2 \) are the endpoints of \( \beta \). If genus \( M = 0 \), \( g \) takes the three distinguished points of \( \Sigma \) to the three distinguished points of \( \Sigma' \).

c) \( \sigma'(x') = \sigma(x_1) + \sigma(x_2) - 2 \) and \( \sigma'(g(x)) = \sigma(x) \) for all \( x \in \Sigma - x_1 - x_2 \).

Condition c) implies that the sum of the cone angles at points of \( \Sigma' \) is \( 2\pi \) less than the sum of the cone angles of points of \( \Sigma \).

If there are two shortest homologous segments \( \beta_1, \beta_2 \) whose union is not closed then we require:

b) There is a map \( g: M \to M' \) which collapses \( \beta_1 \cup \beta_2 \) to a pair of points \( x', x'' \in \Sigma' \) if \( \beta_1 \) and \( \beta_2 \) have no points in common, and collapses the union to a single point if they have one endpoint in common.

c) The sum of the cone angles is reduced by \( 4\pi \) instead of \( 2\pi \) as in the one-segment case.

**Remark.** If the genus of \( M = 1 \) then condition c) in the definition of \( M^g_\ast(\tau, \eta, <) \) forces the face to be coalescing so that \( \mathcal{D}' \) is a stratum of tori as well.

**Definition.** \( \mathcal{D}' = \mathcal{D}(M', \Sigma', \sigma', \pm ) \) is a pinching face of \( M^g_\ast(\tau, \eta, <) \) along the simple closed geodesic \( \beta \) if:

a) genus \( M' = \text{genus } M - 1 \).

b) There is a map \( g: M - \beta \to M' \) which collapses the two boundary components of \( M - \beta \) to two points \( x', x'' \in \Sigma' \) and which is a bijection from \( \Sigma - x_1 \) to \( \Sigma' - x' - x'' \), where \( x_1 \) is the singularity on \( \beta \).

c) \( \sigma'(x') + \sigma'(x'') = \sigma(x_1) - 2 \) and \( \sigma'(g(x)) = \sigma(x) \) for all \( x \in \Sigma - x_1 \).

If the two shortest segments \( \beta_1, \beta_2 \) are homologous and their union is a simple closed geodesic then we require:

b) The map \( g \) collapses \( M - \beta_1 \cup \beta_2 \) to a pair of points.

c) The total cone angle is reduced by \( 4\pi \).

**Remark.** It is easy to see by Proposition 1.2 that in all cases

\[ \dim \mathcal{D}' = \dim \mathcal{D} - 2. \]

The following definition will be used in Theorem 6.1.

**Definition.** Let \( f: M \to N \) be a map between flat surfaces. Let \( \gamma \) be a curve in \( M \). Let \( h(\gamma) \) be the holonomy vector associated to \( \gamma \). We define the
distortion of \( \gamma \) under \( f \) to be \( \| h(\gamma) - h(f(\gamma)) \|/	ext{length}(\gamma) \). Note that if \( \gamma \) is a geodesic then \( \text{length}(\gamma) = \| h(\gamma) \| \).

**Theorem 6.1.** There are constants \( c_1, c_2 \) such that if \( \kappa < c_1 \delta^3 \), then each \( M^g_\kappa(\tau, \eta, <) \) has a face \( \mathcal{MD}' \) and there is a map \( f: M^g_\kappa(\tau, \eta, <) \to \mathcal{MD}' \). In addition to (a), (b), and (c) in the definition of face the following properties are satisfied:

i) There is a triangulation \( \tau' \) of \( M' \) with vertices in \( \Sigma' \) and a homomorphism \( \eta': \pi_1(M' - \Sigma') \to \mathbb{Z}/2\mathbb{Z} \) such that \( f(M^g_\kappa(\tau, \eta, <)) \subset N(\tau', \eta') \).

ii) For each \( q \in M^g_\kappa(\tau, \eta, <) \) the map \( g \) can be realized as a map that collapses the triangles containing \( \beta \) (resp. \( \beta_1, \beta_2 \) in the two-segment case) to edges in \( \tau' \) of the flat structure \( f(q) \) and maps all other triangles of \( \tau \) affinely to triangles of \( \tau' \).

iii) For each \( \gamma \in \pi_1(M - \beta) \) (resp. \( M - (\beta_1 \cup \beta_2) \)), \( \eta'(g(\gamma)) = \eta(\gamma) \). In particular this means \( H^1(M', \Sigma'; \mathbb{R}_\eta^2) \) is naturally a subspace of \( H^1(M, \Sigma; \mathbb{R}_\eta^2) \).

iv) The holonomy vectors \( v \) and \( v' \) of edges \( e \in \tau \) and \( e' = g(e) \in \tau' \) satisfy the relation \( v' = v + \epsilon v_0 \) where \( \epsilon = \epsilon(v) \) belongs to \( \{-1, 0, 1\} \) and \( v_0 \) is the holonomy vector of the edge \( \beta \) (resp. \( \beta_1, \beta_2 \) in the two-segment case).

v) For curves other than \( \beta \) (resp. \( \beta_1, \beta_2 \) in the two-segment case) the distortion of \( g \) is less than \( c_2 \kappa / \delta^3 \).

vi) There are constants \( c \) and \( p \) such that for each vector \( v_0 \), and \( q' \in \mathcal{MD}' \), there are at most \( c / \delta^p \) surfaces \( q \in M^g_\kappa(\tau, \eta, <) \) such that \( f(q) = q' \) and the holonomy vector of \( \beta \) (resp. \( \beta_1, \beta_2 \) in the two-segment case) is \( v_0 \).

**Remark.** We will call \( g \) the collapsing map and \( f \) the retraction to the face. Conclusion iv) says, in holonomy coordinates, that the map \( f \) is locally given by an elementary matrix followed by a projection. Conclusion vi) says restricted to each “level” \( v_0 = \) constant, the map \( f \) is globally finite to one with the given bound.

The proof of the theorem will occupy the rest of this section. The first step in the proof is Proposition 6.4 which constructs a certain cohomology class.

**Lemma 6.2.** Let \( c \) be the \( \mathbb{Z}/2\mathbb{Z} \) cohomology class dual to a geodesic segment or curve \( \gamma \). Let \( \eta \) be the tangent cohomology class. Then \( c \neq \eta \).

**Proof.** Consider the parallel foliation of \( M - \gamma \) with leaves perpendicular to \( \gamma \). All but finitely many leaves have both endpoints in \( \gamma \). Let \( \alpha \) be a leaf with both endpoints in \( \gamma \). Let \( p \) and \( q \) be the endpoints of \( \gamma \). Let \( \beta \) be the segment contained in \( \gamma \) with endpoints \( p \) and \( q \). We claim that the two cohomology classes above differ on the loop \( \alpha \beta \). There are two cases. If \( \alpha \) leaves and returns to \( \gamma \) on the same side then the tangent vectors to \( \alpha \) at \( p \) and \( q \) point on
supported on edges crossing \( \sigma \). Since \( \eta(\sigma) \) is zero the coefficient system \( R_\lambda \) can be trivialized over \( \sigma \). This allows us to associate \( c' \) with a class \( c \in H^1(M, \Sigma; R_\lambda) \).

Our next objective is Lemma 6.7 which gives estimates to show that a small perturbation of a nondegenerate triangle is still a nondegenerate triangle. We will use this to deform our flat structures. Lemmas 6.6 and 6.8 will be needed to estimate distortions.

Let \( \delta > 0 \) be given. Let \( T_1 \) be the space of ordered triples of points in \( \mathbb{R}^2 \) modulo the action of the group of translations \( \mathcal{A} \). Let \( T_2 \subset T_1 \) be the set of triples with positive determinant. Let \( \Delta(\delta, d) \subset T_2 \) be the set of isometry classes of triangles, with all edges of length greater than or equal to \( \delta \), which can be inscribed in circles of radius less than or equal to \( d \). If \( \alpha \) and \( \beta \) are ordered triples of points in \( \mathbb{R}^2 \) then there is a unique affine map, \( A(\beta, \alpha) \), which takes the points of \( \alpha \) to the corresponding points of \( \beta \). Let \( D \) be an equilateral triangle with each side of length one.

**Lemma 6.5.** Let \( \alpha \in \Delta(\delta, d) \). The norm of the affine map \( A(D, \alpha) \) is bounded by \( 2d/\delta^2 \). The norm of the affine map \( A(\alpha, D) \) is bounded by \( 4d/\sqrt{3} \).

**Proof.** We begin by proving the first statement. Let \( l \) be a line in the direction of maximal expansion for \( A(D, \alpha) \). There is a vertex \( a \) of \( \alpha \) so that the translate of \( l \) through \( a \) passes through the side of \( \alpha \) opposite to \( a \) at some point \( a' \). Let \( s_1 \) be the segment with endpoints \( a \) and \( a' \). Let \( s_2 \) be the image of \( s_1 \) under the map \( A \). Now \( ||A|| = \text{length}(s_2)/\text{length}(s_1) \). Let \( m \) be the minimum distance between \( a \) and the opposite side. Then \( \text{length}(s_1) \geq m \) and \( \text{length}(s_2) \leq 1 \) so that \( ||A|| \leq 1/m \). Now consider \( \alpha \) inscribed in a circle of radius \( d \). The value of \( m \) decreases as vertices are moved towards \( a \). Thus \( m \) is minimized (and \( 1/m \) is maximized) over triangles in \( \Delta(\kappa, d) \) by the triangle for which the sides adjacent to \( a \) both have length \( \delta \).

We calculate this value of \( m \). Assume that \( \alpha \) is inscribed in the circle of radius \( d \) tangent to the \( x \)-axis at the origin. Assume that \( a \) is at the origin. The other two points of \( \alpha \) lie at \( (r, m) \) and \( (-r, m) \) for some \( r \). We have the relations \( r^2 + m^2 = \delta^2 \) and \( r^2 + (d - m)^2 = d^2 \). Substituting the value for \( r \) obtained in the first relation into the second relation gives \( \delta^2 - m^2 + (d - m)^2 = d^2 \). This simplifies to \( m = \delta^2/2d \). This is the minimum value for \( m \) and so the maximum value for \( ||A|| \) is \( 2d/\delta^2 \).

We now prove the second statement. As before there are segments \( s_1 \) in \( D \) and \( s_2 \) in \( \alpha \) so that the expansion of \( A(\alpha, D) \) is \( \text{length}(s_2)/\text{length}(s_1) \). Now \( \text{length}(s_1) \geq \sqrt{3}/2 \). Since \( s_2 \) is contained in a circle of radius \( d \), \( \text{length}(s_2) \leq 2d \). We conclude that \( ||A|| \leq 4d/\sqrt{3} \).
Lemma 6.6. The smallest angle $\theta$ of any triangle in $A(\delta, d)$ satisfies $\sin(\theta) \geq \delta^3/2d^3$.

Proof. For any triangle $\Delta$ and angle $\theta$, $\sin(\theta) = 2 \text{area}(\Delta)/l_1l_2$ where $l_1, l_2$ are the lengths of the adjacent sides. We have $l_i \leq d$. The minimum length of the base of any triangle is $\delta$ and as we saw in the proof of the previous lemma, the minimum height is $\delta^2/2d$. Putting these inequalities together proves the lemma.

Let $p_1$, $p_2$ and $p_3$ be the vertices of $D$. Fix a constant $C_3$ so that if $q_1$, $q_2$ and $q_3$ are points in $\mathbb{R}^2$ and $\|p_i - q_i\| \leq C_3$ then $q_1$, $q_2$ and $q_3$ are vertices of a non-degenerate triangle with the same orientation as $D$.

Lemma 6.7. Let $\alpha \in A(\delta, d)$. Let $\alpha'$ be a triangle such that each vertex of $\alpha'$ differs from the corresponding vertex of $\alpha$ by at most $C_3\delta^2/2d$. Then $\alpha'$ is a non-degenerate triangle with the same orientation as $\alpha$.

Proof. We apply the map $A = A(D, \alpha)$ to $\alpha'$. By Lemma 6.5, $D' = A(\alpha')$ differs from $D$ by at most $C_3$ and thus it is non-degenerate and has the same orientation as $D$. These properties are preserved by $A$. Thus $\alpha'$ is non-degenerate and has the same orientation as $\alpha$.

Lemma 6.8. Let $\alpha \in A(\delta, d)$. Let $\alpha'$ be a triangle such that each vertex of $\alpha'$ differs from the corresponding vertex of $\alpha$ by at most $\epsilon < C_3\delta^2/2d$. Then the norm of $A(\alpha, \alpha') - I$ is bounded by $C_4d^3\epsilon/\delta^4$ for some $C_4$.

Proof. We want to estimate the norm of $A(\alpha', \alpha) - I$. Let $D' = A(D, \alpha)\alpha'$. Thus $A(D', \alpha') = A(D, \alpha)$. Now

$$A(\alpha', \alpha) = A(\alpha, D)A(D', D)A(D, \alpha).$$

Thus

$$A(\alpha', \alpha) - I = A(\alpha, D)(A(D', D) - I)A(D, \alpha).$$

The norm of the term on the left is bounded by the product of the norms of the terms on the right and so by Lemma 6.5,

$$\|A(\alpha', \alpha) - I\| \leq (4d/\sqrt{3})\|A(D', D) - I\|(2d/\delta^2).$$

Let $M = A(D', D) - I$. The norm of $M$ is the square root of the largest eigenvalue of $MM^T$. The coefficients of the characteristic polynomial, $P$, of $MM^T$ are polynomials in the coefficients of $M$. The term $\det(MM^T)$ has fourth order dependence on the coefficients of $M$. The term $\text{tr}(MM^T)$ has second order dependence on the coefficients of $M$. The discriminant of the characteristic polynomial has fourth order dependence. The roots of $P$ have second order dependence. Thus the norm of $M$ has first order dependence on coefficients of
The coefficients of $M$ have first order dependence on how much $D'$ differs from $D$ which is at most $\|A(D, \alpha)\|e \leq 2de/\delta^2$. It follows that there is a constant $C$ for which $\|A(D', D) - I\| \leq C2de/\delta^2$. This finishes the proof of the lemma.

**Proof of Theorem 6.1.** Let $d$ be the diameter of the flat surface. Corollary 5.7 says there is a constant $C_5$ such that $d \leq C_5/\delta$. Let $c_1 = C_5/2C_5$. Let $M$ be a flat surface satisfying the hypotheses of the theorem. Then $\kappa < c_1\delta^3 \leq C_3\delta^2/2d$.

Let $\tau$ be the Delaunay triangulation of $M$ as constructed in Section 4. Denote by $\beta$ the shortest segment, or pair of segments if there are two homologous shortest segments in $M$. Since $\beta$ is shortest it is an edge of $\tau$. Every edge of $\tau$ other than $\beta$ has length at least $\delta$. The triangles not containing $\beta$ are contained in the set $\Delta(\delta, d)$. Let $c \in H^1(M, \Sigma; \mathbb{R}^2)$ be the class giving the flat structure on $M$. Let $c'$ be the cohomology class given by the Lemma 6.4.

We may think of $c'$ as an element of $H^1(M, \Sigma; \mathbb{R}^2)$ by including $\mathbb{R}^2$ into $\mathbb{R}^2$ as first coordinate. By scaling and rotating the class $c'$ we may assume that $c'(\beta) = c(\beta)$ and $\|c'(\alpha)\| \leq \|c'(\beta)\| \leq \kappa$ on each edge $\alpha$ of the triangulation.

We will construct a flat structure on $M'$ so that it will correspond to the class $c - c'$.

We can expand the definition of flat surfaces with cone points to include surfaces with boundaries. We require that for each boundary point, $p$, there is a chart mapping a neighborhood of $p$ to a neighborhood of the origin in the upper half plane so that with respect to this chart the metric has the form

$$ds^2 = dr^2 + (crd\theta)^2.$$  

The cone angle at the boundary point is then $c\pi$. Boundary points with cone angle $\pi$ are called regular boundary points. Boundary points with cone angle other than $\pi$ are called singular boundary points. Note that the boundary curve has a length parameter. Polygons in the plane are examples of surfaces with such structures.

We observe that if two flat surfaces with boundary are “glued together” along a segment of the boundaries so that the length parameters are preserved then the result is a flat surface.

For example we can define a flat structure on a triangulated surface by describing the flat structure on each triangle of the triangulation. In particular this allows us to define a flat structure on $M'$ as we will now explain. On each triangle not containing $\beta$ the class $c - c'$ defines an element of $T_1$. Since the element determined by $c$ lies in $\Delta(\delta, d)$ and $|c'| \leq \kappa \leq C_3\delta^2/2d$, the element corresponding to $c - c'$ lies in $T_2$ and thus corresponds to a non-degenerate
triangle by Lemma 6.7. The class $c - c'$ assigns the value zero to $\beta$. The triangles containing $\beta$ collapse to intervals. We construct the flat structure on $M'$ from the non-degenerate triangles by gluing together the edges on either side of the degenerate triangles. The triangles on either side of these intervals are nondegenerate.

In the coalescing case the new flat structure is nondegenerate. We realize the map $g: M \to M'$ by mapping each triangle of $M$ affinely to the corresponding triangle or edge in $M'$ and collapsing $\beta$ to a point. All other points of $\Sigma$ are in one-to-one correspondence with points of $M'$.

In the pinching case, collapsing the simple closed curve $\beta$ gives rise to a singular surface. We desingularize the surface, which gives rise to the two points of $M'$. The collapsing map $g$ maps the two components of $M - \beta$ to these two points.

We wish to show that the classes $\eta$ for the surfaces $M$ and $M'$ are related as described in iii), and the functions $\sigma$ for the surfaces $M$ and $M'$ are related as described in c) of the definition of coalescing face. It will be convenient to define a family of surfaces $M_t$ for $0 \leq t \leq 1$ where $M_t$ corresponds to the class $c - tc'$.

It will be useful to recall some facts about vector fields over curves. Let $M$ be any flat manifold with set of singularities $\Sigma$. Let $\rho: [a, b] \to M - \Sigma$ be a loop. Let $v$ be a unit vector field along $\rho$; that is, $v(t)$ is a tangent vector at $\rho(t)$. In a coordinate chart we can think of $v$ as a map into the unit circle in $\mathbb{R}^2$. The vector $dv/dt$ is a multiple of the vector $v^\perp$ perpendicular to $v$ where $(v, v^\perp)$ defines the standard orientation. Write $dv/dt = w(t)v^\perp$.

The quantity $w(t)$ is well-defined independent of the choice of chart. Define the total rotation of $v$ to be

$$tr(v) = \int_{[a, b]} w(t) \, dt.$$ 

We note that $tr(v)$ depends only on the homotopy class of the vector field $v$.

If $\rho$ is a smooth curve parametrized by arclength then $tr(\rho)$ is the total curvature of $\rho$. When $M$ is triangulated and $\rho$ is a simple closed curve we have a formula for $tr(\rho)$ in terms of the angles in the triangulation. Say $\rho$ crosses edges $e_1, e_2, \ldots, e_k$ at points $p_1, p_2, \ldots, p_k$. Edges $e_i$ and $e_{i+1}$ meet at a vertex which we call $v_i$. Let $\alpha_i$ be the angle between $e_i$ and $e_{i+1}$ at $v_i$. Let $\epsilon_i = 1$ if the vector towards $v_i$ and the tangent vector to $\rho$ at $p_i$ give the standard orientation on $M$ and let $\epsilon_i = -1$ otherwise. Then

$$tr(\rho) = \sum_i \epsilon_i \alpha_i.$$
triangle by Lemma 6.7. The class $c - c'$ assigns the value zero to $\beta$. The triangles containing $\beta$ collapse to intervals. We construct the flat structure on $M'$ from the non-degenerate triangles by gluing together the edges on either side of the degenerate triangles. The triangles on either side of these intervals are nondegenerate.

In the coalescing case the new flat structure is nondegenerate. We realize the map $g: M \to M'$ by mapping each triangle of $M$ affinely to the corresponding triangle or edge in $M'$ and collapsing $\beta$ to a point. All other points of $\Sigma$ are in one-to-one correspondence with points of $M'$.

In the pinching case, collapsing the simple closed curve $\beta$ gives rise to a singular surface. We desingularize the surface, which gives rise to the two points of $M'$. The collapsing map $g$ maps the two components of $M - \beta$ to these two points.

We wish to show that the classes $\eta$ for the surfaces $M$ and $M'$ are related as described in iii), and the functions $\sigma$ for the surfaces $M$ and $M'$ are related as described in c) of the definition of coalescing face. It will be convenient to define a family of surfaces $M_t$ for $0 \leq t \leq 1$ where $M_t$ corresponds to the class $c - tc'$.

It will be useful to recall some facts about vector fields over curves. Let $M$ be any flat manifold with set of singularities $\Sigma$. Let $\rho: [a, b] \to M - \Sigma$ be a loop. Let $v$ be a unit vector field along $\rho$; that is, $v(t)$ is a tangent vector at $\rho(t)$. In a coordinate chart we can think of $v$ as a map into the unit circle in $\mathbb{R}^2$. The vector $dv/dt$ is a multiple of the vector $v^\perp$ perpendicular to $v$ where $(v, v^\perp)$ defines the standard orientation. Write $dv/dt = w(t)v^\perp$.

The quantity $w(t)$ is well-defined independent of the choice of chart. Define the total rotation of $v$ to be

$$\text{tr}(v) = \int_{[a, b]} w(t) \, dt.$$  

We note that $\text{tr}(v)$ depends only on the homotopy class of the vector field $v$.

If $\rho$ is a smooth curve parametrized by arclength then $\text{tr}(\rho)$ is the total curvature of $\rho$. When $M$ is triangulated and $\rho$ is a simple closed curve we have a formula for $\text{tr}(\rho)$ in terms of the angles in the triangulation. Say $\rho$ crosses edges $e_1, e_2, \ldots, e_k$ at points $p_1, p_2, \ldots, p_k$. Edges $e_i$ and $e_{i+1}$ meet at a vertex which we call $v_i$. Let $\alpha_i$ be the angle between $e_i$ and $e_{i+1}$ at $v_i$. Let $\epsilon_i$ be 1 if the vector towards $v_i$ and the tangent vector to $\rho$ at $p_i$ give the standard orientation on $M$ and let $\epsilon_i$ be $-1$ otherwise. Then

$$\text{tr}(\rho) = \sum_i \epsilon_i \alpha_i.$$
This formula is easily proved. Deform the vector field \( \dot{\rho} \) so that it is perpendicular to \( v_i \) at \( p_i \) for each \( i \). This will not change the total rotation. Now sum the contribution from each triangle.

Let \( \rho \) be a curve in \( M_0 \) which does not cross any edge \( \gamma \) which will be collapsed in \( M_1 \). We can construct a path \( \rho_t \) in each \( M_t \) which passes through the same sequence of edges as \( \rho \). We claim that \( \operatorname{tr}(\dot{\rho}_t) \) depends continuously on \( t \). By the above formula this is clear for \( t < 1 \). At \( t = 1 \) the triangulation changes so that some triangles become edges and thus the angles \( \alpha_i \) opposite degenerating edges do not contribute to the sum. But note that as \( t \) approaches 1 these angles \( \alpha_i \) go to zero. Thus the sum changes continuously.

In a flat manifold with holonomy in \( \{ \pm 1 \} \) the quantity \( \operatorname{tr}(v) \) can take on only integral multiples of \( \pi \). Thus \( \operatorname{tr}(\dot{\rho}_t) \) is independent of \( t \). Changing the homotopy class of a vector field \( v \) changes \( \operatorname{tr}(v) \) by a multiple of \( 2\pi \). Thus \( \operatorname{tr}(v) = \operatorname{tr}(\dot{\rho}) \mod 2\pi \). Since the latter is constant, \( \operatorname{tr}(v) \) is constant \( \mod 2\pi \). By taking \( v \) to be a parallel vector field it is easy to see that \( \operatorname{tr}(v)/\pi \mod 2 = \eta(\rho) \). In particular \( \eta(\rho_0) = \eta(\rho_1) \) as was to be shown.

Assume that points \( p \) and \( q \) with cone angles \( c_p \) and \( c_q \) in \( \Sigma \) become coalesced to a point \( r \) in \( M' \) with cone angle \( c_r \). Let \( \rho_0 \) be a path around \( p \) and \( q \). Using the interpretation of \( \operatorname{tr} \) as the integral of curvature, the Gauss-Bonnet formula gives \( \operatorname{tr}(\dot{\rho}_0) = c_p + c_q - 2\pi \). As above we have a family of paths \( \rho_t \) where \( \rho_1 \) surrounds the cone point \( r \). Now \( \operatorname{tr}(\dot{\rho}_0) = \operatorname{tr}(\dot{\rho}_1) \) and again by Gauss-Bonnet, \( \operatorname{tr}(\dot{\rho}_1) = c_r \). Thus \( c_r = c_p + c_q - 2\pi \). The argument for a curve shrinking degeneration is similar.

We have shown i)--iii); iv) follows from the fact that \( c' = \varepsilon c(\beta) \) on each edge of \( \tau \).

We need next to estimate distortions and start with the next lemma. Recall the constant \( C_5 \) in Corollary 5.7 and assume \( C_5 > 1 \).

**Lemma 6.9.** Let \( \alpha \) be an edge of a Delaunay triangulation for a flat structure in the set \( M_\kappa^\delta(\tau, \eta, <) \). For \( \delta \) sufficiently small, any geodesic segment \( \sigma \) starting and ending on \( \alpha \) which does not cross the edge that collapses to a point has length greater than or equal to \( \delta^9 / 8 C_5^4 \).

**Proof.** The proofs for the coalescing and pinching degenerations are slightly different. We give the proof for the coalescing case and then point out the differences for the pinching case. Corollary 5.7 says that \( d \leq C_5 / \delta \). We need to estimate the lengths of certain subsegments of \( \sigma \) as they cross triangles.

**Case I.** Suppose \( p \) is a point on an edge \( \alpha \) of a nondegenerating triangle \( \Delta \). Suppose that the distance \( s \) from \( p \) to either vertex of \( \alpha \) is at least \( \delta^3 / 4 C_5 \). Let \( \sigma_1 \) be a segment from \( p \) that crosses \( \Delta \) to an edge \( \alpha' \) cutting off a vertex angle \( \theta \).
If $\theta$ is acute then $\sigma_1$ is at least as long as the segment $\sigma_2$ from $p$ to $\alpha'$ which is perpendicular to $\alpha'$. If $\theta$ is obtuse then $\sigma_1$ is at least as long as the segment along $\alpha$ from $p$ to the vertex. In the acute case, Lemma 6.6 gives

$$|\sigma_1| \geq |\sigma_2| = \sin(\theta)s \geq \sin(\theta)\delta^3/4C_5 \geq \delta^3 \frac{\delta^3}{8C_5 d^3} \geq \frac{\delta^9}{8C_5^4}.$$ \hfill (6.1)

In the obtuse case by hypothesis

$$|\sigma_1| \geq s \geq \delta^3/4C_5 \geq \frac{\delta^9}{8C_5^4},$$ \hfill (6.2)

for $\delta$ sufficiently small.

Case II. Again suppose the triangle $\Delta$ is nondegenerating but the point $p$ is within $\delta^3/4C_5$ of a vertex and $\sigma_1$ is a segment crossing $\Delta$ from $p$ to the side opposite the vertex.

We saw in Lemma 6.5 that for nondegenerating triangles the minimum distance from a vertex to the opposite side of the triangle is at least $\delta^3/2d \geq \delta^3/2C_5$. Therefore

$$|\sigma_1| \geq \delta^3/2C_5 - \delta^3/4C_5 \geq \frac{\delta^9}{8C_5^4}$$ \hfill (6.3)

for $\delta$ sufficiently small.

Case III. Suppose $\Delta$ is a degenerating triangle, $p$ is on a noncollapsing edge of $\Delta$, but is within $\delta^3/4C_5$ of a vertex of the collapsing edge and $\sigma_1$ crosses $\Delta$ to within $\delta^3/4C_5$ of the vertex opposite the edge that collapses.

Since each noncollapsing edge has length at least $\delta$, for $\delta$ sufficiently small,

$$|\sigma_1| \geq \delta - 2\delta^3/4C_5 \geq \frac{\delta^9}{8C_4^4}.$$

(6.4)

Now for each singularity that is not an endpoint of the edge that collapses to a point, take a $\delta^3/4C_5$ neighborhood of that singularity. Because the distance between such singularities is at least $\delta$, these neighborhoods are disjoint and simply connected. Moreover they do not intersect edges other than those emanating from that singularity. Let $V_1$ be the union of all these neighborhoods. Let $V_2$ be a simply connected $\delta^3/4C_5$ neighborhood of the edge that collapses to a point. It is disjoint from $U$ and is disjoint from any edge that does not emanate from the endpoints of the edge that collapses to a point.
We first show that the lemma will follow from (6.1–6.4) if \( \sigma \) is not contained in \( V_1 \cup V_2 \). Assume that is the case. If a subsegment \( \sigma_1 \) crosses a nondegenerating triangle and both of its endpoints are in \( V_1 \cup V_2 \), but \( \sigma_1 \) leaves \( V_1 \cup V_2 \), we use (6.3) for the lower bound on the length. If \( \sigma_1 \) crosses a nondegenerating triangle and an endpoint is not in \( V_1 \cup V_2 \), then we use either (6.1) or (6.2). If \( \sigma_1 \) crosses a degenerating triangle with one endpoint in \( V_1 \) and the other in \( V_2 \), then we use (6.4). If \( \sigma_1 \) crosses a degenerating triangle and one endpoint is not in \( V_1 \cup V_2 \), then \( \sigma \) crosses an adjacent nondegenerating triangle with an endpoint not in \( V_1 \cup V_2 \) and we use (6.1) or (6.2).

The fact that \( \sigma \) must leave \( V_1 \) follows from some standard facts about quadratic differentials with finite area. Let \( U \) be a simply connected neighborhood of a singularity \( x \) which is disjoint from all other singularities, \( \alpha \) a segment in \( U \), and \( \beta \) a second segment that leaves and returns to \( \alpha \), remaining in \( U \). Then an application of the argument principle ([G], p. 43) or the Gauss Bonnet formula implies that the homotopically trivial simple closed curve which is the union of a piece of \( \alpha \) and \( \beta \) must contain a singularity with cone angle \( \pi \) (simple pole) in its interior and in that case neither \( \alpha \) nor \( \beta \) emanates from the singularity. In the case at hand we have \( \sigma \) leaving and returning to an edge that does emanate from the singularity. Thus it cannot remain in \( V_1 \).

We need to show that \( \sigma \) cannot remain in \( V_2 \). Call the collapsing edge \( e \), joining vertices \( x_1, x_2 \). Suppose \( \sigma \) leaves and returns to \( \alpha \), remaining in \( V_2 \). Both segments are disjoint from \( e \). Then the argument principle shows the trivial simple closed curve, which is the union of \( \alpha \) and \( \sigma \), must contain two points with cone angle \( \pi \) in its interior. That is, the endpoints of \( e \) must have cone angle \( \pi \). But this is ruled out by assumption. This completes the proof of the lemma in the one-segment coalescing case. The two-segment coalescing case is similar.

The only difference in the pinching case is that the \( \delta^3/4C_3 \) neighborhood of the pinching curve \( \beta \) is not simply connected. Rather it is an annulus homotopic to the loop being pinched. Suppose a geodesic segment \( \sigma \) left and returned to a segment \( \alpha \) while remaining in the neighborhood. Then \( \sigma \) and a piece of \( \alpha \) would form a simple loop \( \gamma \) homotopic to \( \beta \). Let \( a_1 \) and \( a_2 \) be the two points on \( \gamma \) where \( \alpha \) and \( \sigma \) meet. The sum of the exterior angles of \( \gamma \) at \( a_i \) is zero, while each exterior angle on \( \beta \) is negative. The two curves bound an annulus with no interior cone points. The annulus has Euler characteristic zero which contradicts Gauss-Bonnet.

\textbf{Remark.} Let \( C = 8C_3^4 \). We conclude from Lemma 6.9 that a geodesic of length \( L \) can cross any edge \( \alpha \) at most \( CL/\delta^6 \) times: we may divide the geodesic into subsegments each of which has both endpoints on \( \alpha \). The length of
each subsegment is at least $\delta^9/C$ by the lemma. There can be at most $CL/\delta^9$ of them.

We analyze the distortion of lengths of curves under $g$. It suffices to bound the distortion for geodesic segments because any curve is homotopic to a union of geodesic segments. Let $\sigma$ be a geodesic in $M$. We can write $\sigma$ as the union of subsegments $\sigma_i$ where each $\sigma_i$ is a component of the intersection of $\sigma$ with a triangle. For nondegenerate triangles we have the estimate, by Lemma 6.8,

$$\|h(\sigma_i) - h(g(\sigma_i))\| / \|h(\sigma_i)\| \leq C_4 kd^3/\delta^4 \leq C_4C_3^3\kappa/\delta^7.$$ 

For degenerate triangles,

$$\|h(\sigma_i) - h(g(\sigma_i))\| \leq \kappa.$$

Let $N$ be the set of indices corresponding to nondegenerate triangles. Let $D$ be the set of indices corresponding to degenerate triangles. Let $\|\sigma\|$ denote the length of $\sigma$. Then

$$\|h(\sigma) - h(g(\sigma))\| / \|\sigma\| = \left\| \sum h(\sigma_i) - \sum h(g(\sigma_i)) \right\| / \|\sigma\|$$

$$\leq \sum_{i \in N} \frac{\|h(\sigma_i) - h(g(\sigma_i))\|}{\|\sigma\|}$$

$$= \sum_{i \in N} \frac{\|h(\sigma_i)\|}{\|\sigma\|} \frac{\|h(\sigma_i) - h(g(\sigma_i))\|}{\|h(\sigma_i)\|}$$

$$+ \sum_{i \in D} \frac{\|h(\sigma_i) - h(g(\sigma_i))\|}{\|\sigma\|}$$

$$\leq \sum_{i \in N} \frac{\|h(\sigma_i)\|}{\|\sigma\|} C_4C_3^3\kappa/\delta^7 + \sum_{i \in D} \frac{\kappa}{\|\sigma\|}$$

$$\leq C_4C_3^3\kappa/\delta^7 + \text{card}(D) \frac{\kappa}{\|\sigma\|}$$

$$\leq C_4C_3^3\kappa/\delta^7 + C\kappa/\delta^9$$

$$\leq \kappa C_2/\delta^9$$

for some $C_2$, for $\delta$ sufficiently small. This proves $v$).

Finally we prove statement $vi$). The point here is that although the flat structure in the domain of $f$ is given its Delaunay triangulation $\tau$, the triangulation $\tau'$ induced on the image flat structure by the collapsing map may not be Delaunay. We lift the map $f$ to a map from $M^g_\kappa(\tau, \eta, \prec) \subset \mathcal{D}$ to $\mathcal{D}'$. The issue is that distinct points may have images that are equivalent under the
mapping class group of $\mathcal{D}'$. We need to compute how many equivalent points there can be.

Any $q \in M^\delta_\kappa(\tau, \eta, <)$ has diameter at most $C_5/\delta$. Therefore its image $q' = f(q)$ under the retraction map has diameter at most $c'/\delta$ for a fixed constant $c'$. Give $q'$ its Delaunay triangulation. The edges of this triangulation have lengths at most $2c'/\delta$. Now if two equivalent points in the image have the same Delaunay triangulation they are the same. A triangulation is determined by a triple of edges. Since we are merely trying to show that the number of equivalence classes grows polynomially in $1/\delta$, it suffices to show that the number of homotopy classes of geodesic segments of length at most $2c'/\delta$ on all possible flat structures $q'$ in the image of any $M^\delta_\kappa(\tau, \eta, <)$ grows polynomially in $1/\delta$.

For each such geodesic, its preimage under the collapsing map $g$ is a piecewise geodesic curve with respect to the Delaunay triangulation of some flat structure $q$. The curve has no self-intersections although it may pass through a vertex more than once. Its length with respect to the $q$ metric is at most $c/\delta$ for a new constant $c$. Thus we need to show that the number of homotopy classes of piecewise geodesics of length at most $c/\delta$ on flat structures $q \in M^\delta_\kappa(\tau, \eta, <)$ grows polynomially with $1/\delta$. By the remark that follows the proof of Lemma 6.9, for any flat structure in $M^\delta_\kappa(\tau, \eta, <)$, any curve of length $1/\delta$ intersects each edge in the Delaunay triangulation of $q$ at most $c/\delta^{10}$ times for a new constant $c$. The conclusion vi) then follows from the next lemma.

**Lemma 6.10.** Suppose $\tau$ is a triangulation of a flat surface, $\gamma_1, \gamma_2$ are two piecewise geodesics each without self-intersections except possibly for multiple intersections at a vertex and suppose they each intersect each edge of $\tau$ the same number of times. Then $\gamma_1$ and $\gamma_2$ are homotopic relative to the vertices.

**Proof.** We need to show that the homotopy class is determined by the number of intersections with the edges. Fix any triangle $\Delta$ with edges $e_1, e_2, e_3$ and let $x_1, x_2, x_3$ denote the number of intersections of a geodesic with $e_1, e_2, e_3$. There are two cases.

Case I. One of the $x_i$ is larger than the sum of the other two. There are three possibilities. Assume say $x_1 \geq x_2 + x_3$. Let $v_1$ be the vertex opposite $x_i$. Then $x_2$ segments must pass from $e_1$ to $e_2$, $x_3$ segments from $e_1$ to $e_3$ and $x_1 - (x_2 + x_3)$ segments join $e_1$ to $v_1$. Moreover because the curve is assumed to be simple the way points of $e_1$ are joined to points of $e_2, e_3$ and $v_1$ is determined. That is, if $x_2 \geq 2$, the segment that crosses $e_1$ closest to $v_3$ must cross $e_2$ closest to $v_3$; if $x_2 \geq 2$ the segment that crosses $e_1$ next closest to $v_3$ crosses $e_2$ next closest to $v_3$, etc. The same holds for segments passing from $e_1$,
to \( e_3 \). The segments passing from \( e_1 \) to \( v_1 \), if any, separate those passing from \( e_1 \) to \( e_2 \) from those passing from \( e_1 \) to \( e_3 \).

**Case II.** Each \( x_i \) is less than the sum of the other two. This implies that the number of segments passing from \( e_i \) to \( e_j = x_i + x_j - x_k/2 \), where \( k \neq i, j \). Moreover the order in which these segments join points on the edges is completely determined since the geodesic has no self-intersections just as in Case I.

We have shown that for each triangle the numbers \( x_i \) determine the number of segments that pass from one edge to another edge and the number of segments that join an edge to the vertex opposite the edge. Suppose now that \( \gamma_1, \gamma_2 \) have the same number of intersections with each edge. Then in each triangle, \( \gamma_1 \) and \( \gamma_2 \) have the same number of segments joining each pair of sides and the same number of segments joining each edge to the opposite vertex. Therefore for each triangle there is a homeomorphism of the triangle which maps the segments of \( \gamma_1 \) to the segments of \( \gamma_2 \). It takes each edge to itself, fixing the vertices. These homeomorphisms fit together to give a homeomorphism of the surface, homotopic to the identity relative to the vertices, which takes \( \gamma_1 \) to \( \gamma_2 \).

### 7. Finding shrinkable segments

One step of the proof of the main theorem is finding a geodesic segment or curve that we can shrink by the action of \( \text{PSL}(2, \mathbb{R}) \) so that the surface resulting from removing the segment is connected. There are two situations which need to be avoided. The first occurs if the segment we are shrinking divides the surface and the second occurs if the segment we are shrinking is homologous to a second segment and together the two segments divide the surface. We call a segment shrinkable if it does not divide the surface and together with a homologous segment it does not divide the surface. In this section we will show that we can find a shrinkable segment. Even after finding a shrinkable segment, it may be that after shrinking it, the surface is degenerate, in the sense that the retraction map to a face (Section 6) is not defined. For example it may happen that there is a second segment parallel to the shrinkable segment of about the same length. When we shrink the shrinkable curve, the parallel curve is also short, and the deformed flat structure does not lie in a set \( M_\delta^s(\tau, \eta, \prec) \) so that the retraction map is not defined. In Section 10 we will show that except for a set of small measure we can find a shrinkable segment that is not too long and that the retraction map is defined on the surface found by shrinking the segment.
We adopt the following terminology in this section. A geodesic curve is either a closed or nonclosed geodesic, not necessarily connected, which has singularities as endpoints and may have singularities as interior points. As always a segment refers to a geodesic path with singularities at the endpoints and no singularities in the interior. (We include the possibility of a closed curve but we will exclude the case of a closed curve with no singularities at all.) A parallel geodesic refers to a union of parallel segments. Since all our surfaces are orientable we use the terminology orientable flat surface to mean that the corresponding foliations are orientable ($\eta = 0$). In this section we will use $d$ to denote the diameter of $M$ as a metric space.

As indicated in the introduction to this section we want to find a geodesic segment $\sigma$ along which we can collapse without separating $M$. Specifically we want to avoid the situation where either $\sigma$ is disconnecting or where $\sigma$ is homologous to a curve $\sigma'$ in $H_1$ so that $\sigma \cup \sigma'$ disconnects $M$. We will also add an extra restriction in the case when the surface is a torus or sphere, since there were extra hypotheses in the statement of Theorem 6.1.

**Definition.** A segment $\sigma$ is shrinkable if none of the following hold:

(a) The curve $\sigma$ is a dividing curve.

(b) A second segment $\sigma'$ exists so that $\sigma \cup \sigma'$ bounds an orientable subsurface (this implies that $\sigma'$ is parallel to $\sigma$ and has the same length as $\sigma$).

(c) The curve $\sigma$ has nonorientable singularities as endpoints and there is a second parallel curve $\sigma'$ so that $\sigma'$ bounds a submanifold $M_1$ containing $\sigma$ and $M_1 - \sigma$ is orientable. (This implies that on the orientable double cover the two lifts of $\sigma$ and a lift of $\sigma'$ bound a subsurface and this implies the length of $\sigma'$ is twice the length of $\sigma$.)

If the surface is a torus not in the exceptional stratum we require:

(d) The curve $\sigma$ not be a simple closed curve.

If the surface is a sphere not in the exceptional stratum we distinguish three points of $\Sigma$ for which the cone angle is $\pi$ and require:

(e) The curve $\sigma$ not contain any of the distinguished points.

As noted above, a shrinkable segment can have a segment parallel to it and together they do not form a boundary in $H_1$.

**Proposition 7.1.** Let $M$ be a flat surface not in an exceptional stratum. Let $d$ be the diameter of $M$ (in the standard sense). There is a shrinkable geodesic segment of length less than $16d$ consisting of one segment.

**Proof.** The proof is much easier in the cases of the torus and sphere and so we give these cases first. Assume the surface $M$ is a torus. Recall that $\pi_1\sigma(s)$ is the cone angle at $s$. For $s$ on the boundary of a domain $U$, define $\pi_1\sigma U(s)$ to be the boundary cone angle. For any domain $U \subset M$, define $n(U) =$
\[ \sum \pi(\sigma(s) - 2), \] the sum over \( \Sigma \cap U \). Now \( n(M) = 0 \). If there are no cone angle \( \pi \) singularities then by Gauss-Bonnet all points of \( \Sigma \) must have cone angle \( 2\pi \), and \( M \) is a flat torus. Since the stratum is not terminal, \( \Sigma \) consists of at least two points and at least one has a cone angle \( \pi \) singularity. We will show that there is a segment joining a cone angle \( \pi \) singularity to a singularity which does not have cone angle \( \pi \) of length less than or equal to \( 4d \). To see that, begin by letting \( \gamma'_1 \) be the shortest segment of length less than or equal to \( d \) joining a cone angle \( \pi \) singularity to any other singularity. If the other endpoint is not a cone angle \( \pi \) singularity, we are done. Suppose the other endpoint also has cone angle \( \pi \). Then there is a geodesic \( \gamma \) parallel to \( \gamma'_1 \) not passing through any singularity such that \( \gamma \) bounds a disc. The disc contains \( \gamma'_1 \) in its interior. As a result there is an embedding of a metric cylinder in \( \mathbb{R}^3 \) into the surface such that the waist curves are homotopic to \( \gamma \). One boundary component of the cylinder in \( \mathbb{R}^3 \) maps two-to-one onto \( \gamma'_1 \). The image of the other boundary component must contain a singularity which does not have cone angle \( \pi \). Otherwise, the other boundary component would also map two-to-one onto a segment joining two cone angle \( \pi \) singularities. The closure of the image of the cylinder would be the entire torus. The torus would contain exactly four cone angle \( \pi \) singularities which is impossible. Therefore, let \( \gamma_1 \) be a segment crossing the cylinder joining a cone angle \( \pi \) singularity to one which is not cone angle \( \pi \). The circumference of the cylinder is at most \( 2d \). The height of the cylinder is at most \( 2d \). The length of \( \gamma_1 \) is at most \( 4d \). That proves the claim.

Since one endpoint of \( \gamma_1 \) is a cone angle \( \pi \) singularity it clearly satisfies conditions (a), (b), and (d). Suppose it does not satisfy (c) so that there is a parallel dividing \( \omega \) of length at most \( 8d \) passing through a single singularity \( x \). Let \( M_1 \) be the component of the complement of \( \omega \) containing \( \gamma_1 \) and let \( M_2 \) be the complementary component.

\( M_1 \) is either a torus minus a disc or a disc. It cannot be a disc since by assumption the only cone angle \( \pi \) singularity it contains is an endpoint of \( \gamma_1 \) and a disc with a geodesic boundary must contain more than one cone angle \( \pi \) singularity in its interior by Gauss-Bonnet. Therefore \( M_1 \) is a torus minus a disc and \( M_2 \) is a disc.

We claim \( n(M_2) \leq -3\pi \). To prove the claim we have

\[
0 = n(M) = n(M_1) + n(M_2) + \pi(\sigma(x) - 2) = n(M_1) + n(M_2) + \pi(\sigma_{M_1}(x) - 2) + \pi(\sigma_{M_2}(x) - 2) + 2\pi.
\]

Since \( M_1 \) is a torus minus a disc, with boundary \( \omega \),

\[
n(M_1) + \pi(\sigma_{M_1}(x) - 2) = \pi.
\]
Together with (1) this implies
\begin{equation}
-3\pi = n(M_2) + \pi(\sigma_{M_3}(x) - 2).
\end{equation}
However
\[
\sigma_{M_3}(x) - 2 \geq 0.
\]
Together with (2) this proves the claim.

Choose the shortest segment $\gamma$ contained in $M_2$ joining a cone angle $\pi$ singularity in $M_2$ to a singularity with cone angle greater than $\pi$. It is clear that $\gamma$ has length less than or equal to $16d$ since $\omega$ has length at most $8d$. We claim $\gamma$ must satisfy (c) as well as (a), (b), and (d). Otherwise $\gamma$ would be contained in a torus $M_3$ with boundary consisting of a parallel segment $\delta$ passing through a single point $y$.

$M_3$ divides $M_2$ into components each of which is homeomorphic to a disc, two of which are bounded by a segment of $\delta$ and a segment of $\omega$ and the rest of which are parallelograms bounded by two segments of $\delta$ and two segments of $\omega$.

For the parallelograms, $n = 0$. For the discs bounded by two segments, $n = -\pi$. Thus these domains only contribute $-2\pi$ to $n(M_2)$. This is a contradiction.

Suppose next that $M$ is a sphere and $x_1, x_2, x_3 \in \Sigma$ are three distinguished points with cone angle $\pi$. Since the sphere is not terminal, there are at least two other points with cone angle $\pi$ and at least one point with cone angle at least $\pi$. Let $\gamma$ be the shortest segment joining a cone angle $\pi$ singularity to a singularity that is not cone angle $\pi$. Exactly as in the torus case the length of $\gamma$ must be at most $4d$. Automatically $\gamma$ must satisfy (c) for now a separating $\gamma_1$ would bound a disc with exactly one cone angle $\pi$ singularity which is impossible.

We now begin the proof of the general case.

Definition. Let $\gamma$ be a geodesic curve. If $\gamma$ is a closed curve then let $\lambda(\gamma) = \frac{1}{2}\text{length}(\gamma)$. If $\gamma$ is connected but not a closed curve then let $\lambda(\gamma) = \text{length}(\gamma)$. If $\gamma$ is not connected then let $\lambda(\gamma)$ be the maximum value of $\lambda(\gamma_i)$ where the $\gamma_i$ are the components of $\gamma$.

We will derive the proposition from the following claims.

Claim 1. Let $M$ be an orientable flat surface which is not a torus. Let $d$ be the diameter of $M$. There is a segment $\sigma$ on $M$ with $\lambda(\sigma) \leq d$ such that:
1) $\sigma$ is not dividing;
2) For no segment $\sigma'$ does $\sigma \cup \sigma'$ bound a subsurface.
Claim 2. Let $M$ be an orientable flat surface with an involution $\iota$ which reverses the orientation of the foliations. There is a segment $\sigma$, satisfying $\lambda(\sigma) \leq d$, 1), and 2) above, which also satisfies:

3) There is no single-segment closed curve $\sigma'$ such that it together with $\sigma \cup \iota(\sigma)$ bounds a subsurface.

We first show that the proposition follows from the claims. If $M$ is orientable then the proposition follows from Claim 1. Assume that $M$ is nonorientable. Let $\tilde{M}$ be the orientation cover of $M$. Let $\iota$ be the orientation-reversing involution on $\tilde{M}$ and let $p$ be the projection from $\tilde{M}$ to $M$. Let $d'$ be the diameter of $\tilde{M}$. Then $d' \leq 2d$. Application of Claim 2 to the orientable manifold $\tilde{M}$ gives a segment $\tilde{\sigma}$ in $\tilde{M}$ of length less than or equal to $2d' \leq 4d$. Let $\sigma$ be $p(\tilde{\sigma})$. There are two possibilities. If $\iota(\tilde{\sigma})$ is disjoint from $\tilde{\sigma}$ except possibly for endpoints, then $\sigma$ is a geodesic segment of the same length as $\tilde{\sigma}$ (the endpoints of the segment may be identified). If $\iota(\tilde{\sigma})$ intersects $\tilde{\sigma}$ at interior points then $\iota$ acts on $\sigma$ by reflecting $\tilde{\sigma}$ through its midpoint. In this case $\sigma$ is a geodesic segment half the length of $\tilde{\sigma}$ and with distinct endpoints. The midpoint projects to a point with cone angle $\pi$. We will show that $\sigma$ satisfies the conclusion of the proposition. If $\sigma$ were to divide $\tilde{M}$ then $p^{-1}(\sigma)$ would be a curve composed of $\tilde{\sigma}$ and $\iota(\tilde{\sigma})$ which together would divide $\tilde{M}$, contrary to assumption.

If $\sigma$ together with some curve $\sigma'$ bounds an orientable subsurface $M_1$ then $p^{-1}(M_1)$ consists of two components one of which contains $\tilde{\sigma}$ in its boundary. Let $\tilde{M}_1$ denote this component. The boundary of $\tilde{M}_1$ consists of two segments. Thus $\tilde{\sigma}$, together with a second segment, bounds a submanifold, contrary to assumption.

Suppose the curve $\sigma$ has nonorientable singularities as endpoints, there is a second curve $\sigma'$ such that $\sigma'$ bounds a submanifold $M_1$ containing $\sigma$, and $M_1 - \sigma$ is orientable. Then a lift of $\sigma'$, $\tilde{\sigma}'$, and $\iota(\tilde{\sigma})$ gives the situation ruled out by case 3 of the claim.

Definition. A connected curve is stable if it is simple and it minimizes the distance between any two of its points. A disconnected curve is stable if each of its components is stable.

Lemma 7.2. If $\sigma$ is a stable geodesic then $\lambda(\sigma) \leq d$.

Proof. Note that $\lambda = \lambda(\sigma)$ is just the diameter of $\sigma$, thought of as a metric space with the path metric. We can find points $p$ and $q$ in $\sigma$ which cannot be connected by a path in $\sigma$ of length less than $\lambda$. The stability assumption implies that $p$ and $q$ cannot be connected by a shorter path in $M$. This shows that the diameter of $M$ is at least $\lambda$. 
Proof of claims. The proof of Claim 1 involves considering a subset of the cases considered in the proof of Claim 2. We will only prove Claim 2. We prove the claim by contradiction assuming that every stable segment violates one of the conditions of the claim. We will show that this assumption allows us to construct a sequence of submanifolds $M_0 \supset M_1 \supset \ldots$ where the boundary of $M_i$ is a stable curve $\gamma_i$. This construction will yield an infinite set of distinct geodesic segments with disjoint interiors. On the other hand the number of such segments is bounded by a constant depending only on the genus of $M$. This contradiction will establish the claim.

Our hypothesis is that every stable curve violates one of the conditions of the claim. Condition 1 is already satisfied by every geodesic curve consisting of one segment. If such a curve $\sigma$ divided $M$ into submanifolds $N_1$ and $N_2$, then since $M$ is orientable, $\sigma$ would be homologically trivial. In particular, the image of $\sigma$ under the development map would be a closed curve. On the other hand, since $\sigma$ is a single segment, the image under the development map is a nontrivial straight-line segment. This is a contradiction. Thus assume every stable curve violates conditions 2) or 3) of the claim. We will call a segment “good” if it violates condition 2) and “bad” if it violates condition 3).

We now describe inductively how to find the sequence $M_i$. We begin by finding $M_0$. The first step is to choose a curve which minimizes $\lambda$. The following lemma will be used to show that a curve which minimizes $\lambda$ is stable.

Lemma 7.3. Let $p$ and $q$ be points in $M$. Let $\sigma$ and $\tau$ be distinct locally length-minimizing paths connecting $p$ and $q$. Then the curve $\sigma \cup \tau$ is homotopically nontrivial. (Homotopies passing through the singularities of $M$ are allowed.)

Proof. Let us assume that $\sigma \cup \tau$ is homotopically trivial. It is a fact about curves in surfaces that we can find subcurves $\sigma' \subset \sigma$ and $\tau' \subset \tau$ meeting at points $p'$ and $q'$ so that $\sigma' \cup \tau'$ bounds an embedded disk $D$. Let $\alpha_{p'}$ and $\alpha_{q'}$ be the external angles at $p'$ and $q'$. Since the curve is embedded we have $\alpha_{p'} < \pi$ and $\alpha_{q'} < \pi$. Let $\beta_1, \ldots, \beta_k$ be the external angles at interior points of $\sigma'$ and $\tau'$. Since these curves are locally length-minimizing we have $\beta_i < 0$. Let $\gamma_i$ be the cone angles in the interior of $D$. Since $M$ is orientable $\gamma_i > 2\pi$.

Gauss-Bonnet gives:

$$\alpha_p + \alpha_q + \sum \beta_i - \sum (\gamma_i - 2\pi) = 2\pi \chi(D) = 2\pi.$$  

But this yields a contradiction because the left hand side of the equation is strictly less than $2\pi$.

We will now show that a curve minimizing $\lambda$ is stable. Let $\gamma$ be a curve minimizing $\lambda$. If $\gamma$ is not stable then there are points $p$ and $q$ in $\gamma$ and a curve $\alpha$ between $p$ and $q$ so that $\alpha$ is shorter than the curve $\gamma$, which is the shortest
curve in $\gamma$ connecting $p$ and $q$. The curve $\alpha \cup \gamma_1$ has a smaller $\lambda$ value than $\gamma$ and it is homotopically nontrivial by the lemma. This contradiction shows that $\lambda$ is stable.

We will now find a stable curve $\gamma_0$ such that $\lambda(\gamma_0) = \lambda(\gamma)$, and submanifold $M_0$ such that $\gamma_0$ is the boundary of $M_0$. We can assume $\gamma$ consists of a single segment. If $\gamma$ is composed of two or more segments, then one of these segments has length less than or equal to half the length of $\gamma$. A subsegment of a stable curve is stable if its length is less than or equal to half the length of the curve. Thus choose some such subsegment and call it $\gamma$. The segment $\gamma$ is stable and so, by hypothesis, one of two situations can occur. If $\gamma$ is good then there is a segment $\gamma'$ so that $\gamma$ and $\gamma'$ together bound a subsurface. In this case let $\gamma_0 = \gamma \cup \gamma'$ and let $M_0$ be either component of the complement. If $\gamma$ is bad then $\gamma$ and $\gamma$ together with some closed geodesic $\gamma'$ bound a subsurface $N$. We can choose $\gamma'$ so that $N$ is disjoint from $\varnothing(N)$. In this case let $\gamma_0$ be the union of these three segments and let $M_0 = N$. In either case $\gamma_0$ is stable.

Now suppose we have constructed $M_i$ inductively. Our inductive hypothesis includes the following. The boundary, $\gamma_i$, is stable and consists of two or three parallel segments. If $\gamma_i$ consists of three segments then we require it to have a special form. In this case $\gamma_i$ has two components and $\varnothing(M_i)$ and $M_i$ have disjoint interiors.

We now show how to construct $M_{i+1}$. Since $\gamma_i$ is a union of parallel segments and there are no cone angle $\pi$ singularities, an application of Gauss-Bonnet shows that $M_i$ cannot be a disk. If $M_i$ has positive genus we can find a nonperipheral curve in $M_i$. If $M_i$ is an annulus we can find a path connecting singularities on the boundaries. In either of these cases we can construct a minimal-length path in the given homotopy class. This path is a geodesic curve. At least one of its segments is nonperipheral. Let $\alpha$ be a nonperipheral segment in $M_i$, not contained in $\gamma_i$, that minimizes the quantity $\lambda$ when restricted to $M_i$.

We will show that $\alpha$ is stable. If $\alpha$ is not stable then there are points $p$ and $q$ on $\alpha$ and a segment $\beta$ which is shorter than the segment $\alpha_1 \subset \alpha$ connecting $p$ and $q$. If $\beta$ lies in $M_i$ then, arguing as before, we can construct a segment between $p$ and $q$ lying in $M_i$ with smaller $\lambda$ value, contrary to assumption. If $\beta$ does not lie in $M_i$ then let $p'$ be the point $\beta \cap \gamma_i$ closest to $p$ along $\beta$. Let $q'$ be the point on $\beta \cap \gamma$ closest to $q$ along $\beta$. Let $\beta_1$ be the subcurve of $\beta$ not containing $p$ or $q$ connecting $p'$ and $q'$. Let $\gamma_1$ be the shorter subcurve of $\gamma_i$ connecting $p'$ and $q'$. Since $\gamma_i$ is stable, $\text{length} \gamma_1 \leq \text{length} \beta_1$. Thus we can replace $\beta_1$ by $\gamma_1$ without increasing the length of $\beta$. The new curve $\beta$ lies in $M_i$ and we have ruled out this possibility.

We now consider the two possibilities that $\alpha$ is good or bad. We begin with the easier case of $\alpha$ good. There is a second curve $\alpha'$ so that $\alpha \cup \alpha'$ is dividing.
This implies that $\alpha'$ is parallel to and has the same length as $\alpha$. We will show that the interior of $\alpha'$ is disjoint from $\gamma_i$. We consider two cases.

**Case 1: $\alpha'$ has distinct endpoints.** If $\alpha'$ intersects a component $\gamma_1$ of $\gamma_i$ at a nonsingular point $p$, then since $\alpha \cup \alpha'$ is homologically trivial and $\alpha$ is disjoint from $\gamma_i$, it must be the case that $\alpha'$ intersects $\gamma_1$ in at least two points. Since $\gamma_1$ is stable, the subsegment of $\gamma_1$ between these two points is no longer than the corresponding segment in $\alpha'$ in the complement of $M_i$. Replacing this piece of $\alpha'$ by $\gamma_1$ does not increase the length. The new curve is not a geodesic since it is not locally length-minimizing at the nonsingular point $p$. Thus we may shorten $\alpha'$. Repeating this process we may assume that $\alpha'$ is contained in $M_i$. Since the length of the original $\alpha'$ is the same as that of $\alpha$ we have produced a closed curve in $M_i$ with smaller $\lambda$ value than $\alpha$, contradicting the assumption that $\alpha$ has minimal $\lambda$.

**Case 2: $\alpha'$ is a loop.** As before if $\alpha'$ intersects a component $\gamma_1$ of $\gamma_i$ it intersects in at least two points. We can find a subsegment of $\alpha'$ contained in $M_i$. The endpoints of this subsegment are connected by a subsegment of $\gamma_1$. Joining the two segments creates a closed curve in $M_i$ shorter than $\alpha$. The lemma implies that this new curve is homotopically nontrivial which is a contradiction.

These two cases show that the interior of $\alpha'$ is disjoint from $\gamma_i$. This means that either $\alpha'$ is in $M_i$ or in the complement of $M_i$. We will construct $\gamma_{i+1}$. If $\alpha'$ is in $M_i$ then we can take $\gamma_{i+1}$ to be $\alpha' \cup \alpha$. If $\alpha'$ is not in $M_i$ then $\alpha$ and $\alpha'$ bound a submanifold $N$. The boundary of $N \cap M_i$ consists of $\alpha$ and a subset of $\gamma_i$. We take $\gamma_{i+1}$ to be $\alpha$ together with this subset. If $\gamma_{i+1}$ contains three segments, two of them are segments of $\gamma_i$. These two segments are parallel by the induction hypothesis which forces $\alpha$ to be parallel to both. Thus $\gamma_{i+1}$ contains three parallel segments.

We take $M_{i+1}$ to be the component of $M - \gamma_{i+1}$ which does not contain $\gamma_i$ in its interior. If $\gamma_{i+1}$ contains three segments then so does $\gamma_i$. The induction hypothesis that $\iota(M_i)$ and $M_i$ have disjoint interiors implies that $\iota(M_{i+1})$ and $M_{i+1}$ have disjoint interiors. This completes the induction step in the case that $\alpha$ is good.

It remains to consider the case when $\alpha$ is bad. There are several cases with subcases.

**Case A.** The segment $\iota(\alpha)$ intersects $\gamma_i$ at an interior point of $\iota(\alpha)$.

If $\gamma_i$ had three segments then the assumption that $M_i$ and $\iota(M_i)$ have disjoint interiors would rule out the possibility that $\iota(\alpha)$ intersects $\gamma_i$ at an interior point. We conclude that $\gamma_i$ has two segments.
Subcase A1. The curve $\gamma_i$ has two components.

Let $\gamma$ be a segment of $\gamma_i$ that intersects $\iota(\alpha)$. Since $\iota$ is an involution, $\alpha$ intersects $\iota(\gamma)$. The segment $\iota(\gamma)$ is parallel to $\gamma_i$. Two parallel segments are equal, disjoint or intersect in an endpoint. Since $\alpha$ does not intersect $\gamma_i$ at an interior point, $\iota(\gamma)$ and $\gamma_i$ are not equal. We conclude that they intersect in at most an endpoint and $\iota(\gamma) \subset M_i$. By construction, the segment $\gamma$ minimizes $\lambda$ in $M_{i-1}$. Since $\lambda(\iota(\gamma)) = \lambda(\gamma)$ and $M_i \subset M_{i-1}$ we conclude that $\iota(\gamma)$ also minimizes $\lambda$ in $M_i$. It follows that $\iota(\gamma)$ is stable. Let $M_{i+1} = M_i \cap \iota(M_i)$ and let $\gamma_{i+1}$ be the boundary of $M_{i+1}$. Then $\gamma_{i+1}$ consists of $\iota(\gamma)$ and a component of $\gamma_i$.

Subcase A2. The curve $\gamma_i$ has one component.

In this case $\iota(\alpha)$ meets $\gamma_i$ in a homologically trivial way. We can use the curve shortening argument as in Cases 1 and 2 to derive a contradiction.

Case B. The interior of the segment $\iota(\alpha)$ is disjoint from $\gamma_i$ and contained in $M_i$.

Let $\alpha'$ be the curve homologous to $\alpha \cup \iota(\alpha)$ chosen so that the surface $N$ they bound has disjoint interior from $\iota(N)$. In this case $\alpha'$ links either component of $\gamma_i$ in a homologically trivial way. If $\alpha'$ intersects $\gamma_i$ in its interior, we could use a curve-shortening argument as in Case 1 to derive a contradiction. If $\alpha'$ is inside $M_i$ we can take as $M_{i+1}$ the submanifold $N$ of $M_i$ bounded by $\alpha'$ and $\alpha \cup \iota(\alpha)$. If $\alpha'$ is outside $M_i$ we can take as $M_{i+1}$ the intersection of $N$ with $M_i$. The boundary $\gamma_{i+1}$ has three segments, one from each of $\gamma_i$ and $\alpha$ and $\iota(\alpha)$. Since $\alpha$ and $\iota(\alpha)$ are parallel, they are parallel to the segment of $\gamma_i$.

The condition that $\iota(M_{i+1})$ should have disjoint interior from $M_{i+1}$ is satisfied by construction.

Case C. The interior of the segment $\iota(\alpha)$ is disjoint from $\gamma_i$ but not contained in $M_i$. This means $\iota(\alpha)$ lies in the complement of $M_i$. It also implies that the endpoints of $\alpha$ lie on $\gamma_i$.

Subcase C1. The curve $\gamma_i$ consists of two components and $\alpha$ has endpoints in both components.

If $\gamma_i$ has three segments then $M_i$ and $\iota(M_i)$ have disjoint interiors by assumption. If $\gamma_i$ has two segments and if the interiors of $M_i$ and $\iota(M_i)$ intersect, then $\iota(\gamma) \subset M_i$ for some segment $\gamma$ of $\gamma_i$. We could then use the method of Case A1 to construct $M_{i+1}$. Thus we may assume that $M_i$ and $\iota(M_i)$ have disjoint interiors. Each component $C$ of $\gamma_i$ contains an endpoint of $\alpha$ which is a fixed point of $\iota$. Thus each $C$ intersects $\iota(C)$ in at least one point. If $C \neq \iota(C)$ then the intersection consists of singular points. If $C = \iota(C)$ then $\iota$
has two fixed points on \( C \) and these are singular points. We conclude that if \( C \) contains only one singular point, then \( C \cap \iota(C) \) consists of only one point. If \( \gamma_i \) has two segments each component of \( \gamma_i \) contains one singularity. If \( \gamma_i \) has three segments exactly one component of \( \gamma_i \) contains one singularity. In either case there is a component which intersects its image exactly once. Furthermore one component cannot intersect the image of the other component without producing an excessive number of singular points. Let \( \alpha' \) be a curve homologous to \( \alpha \cup \iota(\alpha) \) and let \( N \) be the surface bounded by these curves chosen so that \( N \) and \( \iota(N) \) have disjoint interiors. Let \( R = N \cap M_i \). The boundary of \( R \) consists of \( \alpha, I_1, \alpha' \cap M_i \) and \( I_2 \) where \( I_1 \) and \( I_2 \) are subintervals of \( \gamma_i \) with one interval in each component. If \( \alpha \) and \( \alpha' \) meet some component in a single point then the corresponding interval \( I \) reduces to a point. When \( \alpha \) and \( \alpha' \) do intersect they intersect at a singular point. Since \( \alpha' \) is a one-segment curve it has only one singular point. It follows that \( I_1 \) and \( I_2 \) cannot both be trivial.

The subsurface \( R \) is oriented. Therefore its boundary is homologically trivial and the developing image of \( \partial R \) is a closed curve. The image of each side of \( \partial R \) is a line segment. Each pair of opposite sides is parallel. In particular if one interval \( I \) reduces to a point then the other must also. But this possibility is ruled out above. We conclude that both intervals are nontrivial and that the image of \( \partial R \) is a parallelogram. In a parallelogram opposite pairs of sides have the same length. In particular \( \alpha \) and \( \alpha' \cap M_i \) have the same length. By the same argument \( \iota(\alpha) \) has the same length as \( \alpha' \cap \iota(M_i) \). But \( \iota(\alpha) \) has the same length as \( \alpha \). Since the length of \( \alpha' \) is exactly twice the length of \( \alpha \) we conclude that the interior of \( \alpha' \) is contained in \( M_i \cup \iota(M_i) \). Thus when \( \alpha' \) crosses from \( M_i \) to \( \iota(M_i) \) it does so at a point of intersection of \( \gamma_i \) and \( \iota(\gamma_i) \). These intersection points are distinct from the endpoints of \( \alpha \). This implies that both components of \( \gamma_i \) meet their images under \( \iota \) in two points. But this contradicts our previous argument which showed that there is one component that meets its image in at most one point.

Subcase C2. The curve \( \gamma_i \) is connected.

Let \( \alpha' \) be a curve homologous to \( \alpha \cup \iota(\alpha) \), again chosen so that the surface \( N \) they bound has disjoint interior from \( \iota(N) \). If the curve \( \alpha' \) does not lie on one side of the homologically trivial curve \( \gamma_i \) then it must intersect \( \gamma_i \) in at least two points. Since \( \alpha' \) has only one singularity, one of the intersection points must be a nonsingular point. Replacing a piece of \( \alpha' \) with a piece of \( \gamma_i \) as in Case 2, we can use the curve-shortening argument to derive a contradiction. If \( \alpha' \) lies inside \( M_i \), let \( M_{i+1} \) be \( M_i \cap N \). Let \( \gamma_{i+1} \) be the boundary of \( M_{i+1} \). Then \( \gamma_{i+1} \) consists of the closed curve \( \alpha' \) and a curve composed of \( \alpha \) and a segment of \( \gamma_i \). Once again the induction hypotheses are satisfied for \( M_{i+1} \) and \( \gamma_{i+1} \). If \( \alpha' \) lies outside
of $M_i$ then $N \cap M_i$ has a boundary consisting of a segment of $\gamma_i$ and $\alpha$. Again let $M_{i+1}$ be $M_i \cap N$.

Subcase C3. The curve $\gamma_i$ has two components and $\alpha$ has both endpoints in the same component.

Since the endpoints of $\alpha$ are distinct, one component of $\gamma$ must have two segments. So $\gamma$ must have three segments. Let $\sigma$ denote the one-segment component. The curves $\alpha$ and $\iota(\alpha)$ are disjoint from $\sigma$ since $\iota(M_i)$ and $M_i$ have disjoint interiors. Therefore $\alpha'$ must intersect $\sigma$ in a homologically trivial way. If there are some interior intersection points we could use the curve-shortening argument to derive a contradiction. Since $\alpha'$ has only one singularity, we conclude that $\alpha'$ does not cross $\sigma$. A similar argument shows that $\alpha'$ does not cross the other component of $\gamma_i$. Let $M_{i+1}$ be $N \cap M_i$ as in Subcase C2.

This completes the construction of $M_{i+1}$ and completes the proof of the proposition.

8. Flat structures on the torus minus a disk and the sphere minus a disk

In this section we develop a criterion which allows us to recognize a flat surface as the connected sum of a torus (or sphere with cone angle $\pi$ singularities) with a surface of small diameter. In the proof of the main theorem we apply this result to the surfaces which result after we have shrunk all shrinkable curves. The fact that these surfaces are close to being tori (or spheres with cone angle $\pi$ singularities) is used in finding cylinders on these surfaces.

Theorem 8.1. Let $M$ be a flat surface without singularities and with piecewise smooth boundary homeomorphic to a torus minus a disk. Assume that the length of the boundary is less than $l$ and assume that every closed curve of length less than $l/2$ is homotopic to a multiple of the boundary curve. Then $M$ contains a subsurface $N$ so that

1) area$(M - N) < l^2/\pi$.

2) The length of the boundary of $N$ is less than $l$.

3) There are a flat torus $T$ and a convex disk $D \subset T$ so that $N$ is isometric to $T - D$.

Remark. Flat structures on surfaces with nonempty boundaries can be quite wild even if they have no singularities. As an example, any homomorphism from $\pi_1(M)$ to the group of Euclidean motions $\mathcal{X}$ can be realized as the holonomy of such a structure. The preceding theorem shows that if the length of the boundary is short the possible flat structures are quite restricted.
Proof. We begin by finding a curve which will be the boundary of \( N \). We can find a smooth simple curve \( \beta \) in the interior of \( M \) isotopic to \( \partial M \) with length less than \( l \) so that the area between \( \partial M \) and \( \beta \) is small. Thus we may assume that \( \partial M \) is smooth and \( M \) is contained in the interior of a flat manifold \( M' \).

Let \( \Gamma \) be the set of curves in \( M \) which are freely homotopic to nonzero multiples of the boundary of \( M \). The set \( \Gamma \) contains \( \partial M \) and so is nonempty. Let \( \gamma \) be an element of \( \Gamma \) of minimal length. Clearly \( \text{length}(\gamma) \leq \text{length}(\partial M) < l \).

Claim. \( \gamma \) is simple. If \( \gamma \) is not simple there are points \( a \) and \( b \) in the circle such that \( \gamma(a) = \gamma(b) = p \). So \( \gamma \) is the product (in \( \pi_1(M, p) \)) of two curves \( \gamma_1 \) and \( \gamma_2 \) based at \( p \). One of these curves, say \( \gamma_1 \), must have length less than or equal to \( l/2 \). By the hypothesis of the theorem this curve is homotopic to a multiple of the boundary curve. But since \( \gamma_1 \) is shorter than \( \gamma \) the minimality of \( \gamma \) implies that \( \gamma_1 \) must not be a nonzero multiple of \( \partial M \). We conclude that \( \gamma_1 \) is trivial. Since \( \gamma = \gamma_1\gamma_2 \) we have \( \gamma = \gamma_2 \). But \( \gamma_2 \) is shorter than \( \gamma \), contradicting the minimality of \( \gamma \).

The only multiples of \( \partial M \) represented by a simple curve are \( \pm 1 \). Thus the previous claim implies that \( \gamma \) is homotopic to \( \partial M \).

The curve \( \gamma \) divides \( M' \) into two components. Since \( \gamma \) is peripheral, one of these components is homeomorphic to an annulus. Let \( A \) be the annulus component and let \( N \) be the component of \( M' - \gamma \) not homeomorphic to an annulus. The set \( M - N \) is contained in the annulus \( A \). Perturbing \( \partial M \) if necessary we may assume that \( M - N \) has only finitely many components. The boundary of each component \( C_i \) is the union of a segment \( \gamma_i \) of \( \gamma \) and a segment \( m_i \) of \( \partial M \). It is a topological consequence of the fact that \( \partial M \) is a simple curve of degree one in the annulus \( A \) that the curves \( \gamma_i \) and \( m_i \) are homotopic. This implies that \( C_i \) is a disk. The sum of the lengths of the boundaries of these disks is less than or equal to the sum of the lengths of \( \gamma \) and \( \partial M \) which is less than or equal to \( 2l \). By the isoperimetric inequality (for immersed disks) the sum of the areas of these disks is not greater than the area of a disk with circumference \( 2l \). The area of such a disk is \( l^2/\pi \).

Claim. The geodesic curvature of \( \gamma \) is nonpositive. Choose an orientation on \( M \). This orientation determines a compatible orientation on \( \partial M \). The orientation on \( M \) determines one on \( N \) and on \( \gamma \). Let \( n \) be the normal to \( \gamma \) pointing into \( N \). Assume at some point \( p \) of \( \gamma \) that the geodesic curvature \( k_{\gamma}(p) \) is not zero. The mean curvature vector is \( k_{\gamma}n \). Moving \( \gamma \) in the direction of \( k_{\gamma}n \) decreases length. Specifically we could find a family of curves \( \gamma' \) with \( \gamma'^0 = \gamma \), \( \gamma' \subset M \) and \( \text{length}(\gamma') < \text{length}(\gamma) \). The existence of such curves contradicts
the minimal length property of $\gamma$. We conclude that $k_{g_n} = 0$ where $\gamma \subset \text{int}(M)$ and that $k_{g_n} \mathbf{n}$ does not point into $M$ where $\gamma \subset \partial M$. This implies that $k_{g} \leq 0$.

Claim. The total geodesic curvature of $\gamma$ is $-2\pi$. The curve $\gamma$ is composed of finitely many segments each of which is either a segment in $\partial M$ or a straight line. At points where these segments meet, the tangent vectors coincide. Thus $\gamma$ is a $C^1$ curve which is piecewise $C^\infty$. For the sake of applying the Gauss-Bonnet formula we may think of $\gamma$ as a piecewise smooth curve with all external angles equal to zero. We have

$$\int \int_N K + \int_{\partial N} k_{g} = 2\pi \chi(N).$$

The integral of scalar curvature in $N$ is zero. The Euler characteristic of $N$ is $-1$. Thus $\int_{\partial N} k_{g} = -2\pi$.

Since the linear holonomy group $\mathcal{L} = \{ \pm 1 \}$ is abelian the linear holonomy around the commutator $\gamma$ is trivial. There is a consistent notion of direction on $\gamma$. We can define a Gauss map from $\gamma$ to the unit circle $S^1$ mapping a point to the direction of its tangent vector.

Claim. The inverse image of a point under the Gauss map is connected. The geodesic curvature is the derivative of the Gauss map and the total curvature on $\gamma$ is the winding number of the Gauss map multiplied by $2\pi$. Assume that $p$ and $q$ map to the same point under the Gauss map. These points divide $\gamma$ into two intervals. The winding number on each interval is an integer and is nonpositive. Since the total winding number is $-1$ the winding number on one interval is zero. Since the derivative of the Gauss map is nonpositive the Gauss map must be constant on this interval. Thus $p$ and $q$ lie in a connected set on which the Gauss map is constant.

Our next objective is to show that the linear holonomy of $N$ is trivial. Since the linear holonomy is trivial on $\gamma = \partial N$ there is a well-defined homomorphism from $H_1(N, \partial N)$ to the $\mathcal{L}$. It suffices to show that this homomorphism is trivial.

Claim. The linear holonomy along any path from $\partial N$ to itself is trivial. Let $\alpha$ be a minimal-length path from $p \in \partial N$ to $q \in \partial N$ with nontrivial holonomy. If $\alpha$ were to intersect $\partial N$ in its interior then it would be the sum of two shorter curves rel $\partial N$. One of these curves would have nontrivial linear holonomy, contradicting the choice of $\alpha$. We conclude that $\alpha$ is a geodesic segment. Since the linear holonomy along $\partial N$ is trivial there is a consistent way of comparing directions at two different points. Choose an orientation along $\alpha$. If the linear holonomy along $\alpha$ is trivial then the vectors at either end of $\alpha$ pointing into $\alpha$
have opposite directions. If the linear holonomy along $\alpha$ is non-trivial then the vectors at either end of $\alpha$ pointing into $\alpha$ have the same direction (with respect to the consistent sense of direction on $\partial N$).

Consider the case that $p = q$. Now $\alpha$ is a map from an interval $I$ to $N$. The map $\alpha$ is not injective; nevertheless we claim that the set $\alpha(I)$ is a one-dimensional submanifold of $N$. Say $\alpha(x_1) = \alpha(x_2) = y$. Let $D$ be a small metric disk around $y$. Let $I_1$ and $I_2$ be the intervals around $x_1$ and $x_2$ which map into $D$. Since $\alpha$ is a geodesic, $\alpha(I_1)$ and $\alpha(I_2)$ are straight-line segments through $y$. Also, since $\alpha$ is a connected geodesic which does not meet the boundary of $M$, the direction of the tangent line to $\alpha$ is constant. Thus $\alpha(I_1)$ and $\alpha(I_2)$ are segments through the same point in the same direction. We conclude that they are equal. This argument shows that the intersection of $\alpha(I)$ with the interior of $M$ is a manifold. A similar argument shows that $p = q$ is an endpoint of this manifold. We have produced a connected one-dimensional manifold with boundary which has $p = q$ as its only endpoint. But a connected one-dimensional manifold must have zero or two endpoints. This shows that the case $p = q$ is impossible.

Consider the case $p \neq q$. At the points $p$ and $q$ the tangent vectors to $\alpha$ point in the same direction. These vectors are normal to $\gamma$; thus, the tangent vectors to $\gamma$ at $p$ and $q$ point in the same direction. Let $J$ be the segment of $\partial N$ with endpoints $p$ and $q$ along which the Gauss map is constant. At $p$ we have a unit tangent vector pointing into $J$. If we parallel translate this vector along $\alpha$ we get a Jacobi field. At $q$ the parallel translate points into $J$ (the direction has reversed). Translating $\alpha$ along this Jacobi field gives a family of geodesics. The lengths of these geodesics do not change because the tangent vectors to these geodesics remain perpendicular to the direction of $J$. The translation is possible as long as no geodesic hits $\partial N$ in its interior. If some geodesic did hit $\partial N$ in its interior then arguing as before we could break it up into two shorter pieces, one of which must have nontrivial linear holonomy, contradicting the choice of $\alpha$. Thus we can move the points $p$ and $q$ together until they coincide. But this contradicts the previous argument.

The affine holonomy around the curve $\gamma$ is trivial: Since the linear holonomy homomorphism is trivial the image of the affine holonomy is contained in the pure translation subgroup. This group is abelian and the homotopy class represented by $\gamma$ is a commutator. Hence the image of the class represented by $\gamma$ is trivial.

Since the affine holonomy around $\partial N$ is trivial the developing map is well defined on $\partial N$. The image of $\partial N$ is an immersed curve with total curvature $-2\pi$ and curvature of constant sign. We claim that any such curve is embedded.
and is the boundary of a convex disk. This is an elementary fact about curves in the plane which we will not prove.

The image of $\partial N$ under the developing map bounds a convex disk $D$ in the plane. The flat structure on the plane gives a flat structure on $D$. The developing map gives an identification of $\partial D$ with $\partial N$. Let $T$ be the torus obtained by gluing $\partial D$ to $\partial N$ by means of this identification. The flat structures on $D$ and $N$ patch together to give a flat structure on $T$. This follows because the developing maps patch together to give a consistent developing map. The surface $T$ is the flat torus promised in assertion (3) of the theorem. This completes the proof of the theorem.

**Theorem 8.2.** Let $M$ be a flat surface with three cone angle $\pi$ singularities and with piecewise smooth boundary homeomorphic to a disk. Let $l$ be the length of the boundary of $M$ and assume that every curve of length less than $l$ is homotopic to a multiple of the boundary curve and that no singularity is within distance $l/2$ of the boundary. Then $M$ contains a subsurface $N$ so that

1. $\text{area}(M - N) \leq 2l^2/\pi$.
2. The length of the boundary of $N$ is less than or equal to $l$.
3. There are a sphere with four cone angle $\pi$ singularities $S$ and a convex disk $D \subset S$ so that $N$ is isometric to $S - D$.

**Proof.** Let $M'$ be $M$ with the singularities removed. The linear holonomy homomorphism maps $\pi_1(M')$ to $\{\pm 1\}$. A small loop around any of the singularities maps nontrivially. Since the boundary is homologous to a sum of an odd number of singularities, the boundary maps nontrivially. We can construct a linear holonomy cover over $M'$ and add the singularities to the holonomy cover. The flat metric extends across these points to give a flat metric without singularities. We call this surface the branched holonomy cover. Since the holonomy around the boundary was nontrivial the boundary is double covered by a circle in the boundary of the holonomy cover. Thus the branched holonomy cover has a single boundary component. We calculate the Euler characteristic of the branched holonomy cover to be twice the Euler characteristic of $M'$ plus 3 for the three singular points. This gives an Euler characteristic of $-1$. The branched holonomy cover is therefore a torus minus a disk. We now apply Theorem 8.1 to the branched holonomy cover. This produces a submanifold $N$ with boundary $\gamma$ and a flat torus $T$ containing $N$. Note that in taking the cover the length of the boundary doubles, and the area doubles. The covering involution on the holonomy cover extends to an isometry of the flat torus. From the proof of Theorem 8.1 we see that $\gamma$ is the unique minimal length curve homotopic to $\partial N$. In particular $\gamma$ is invariant under the covering involution. The flat torus modulo this involution is the sphere $S$ sought.
9. Lattice points

In the proof of the main theorem we reduce the problem of finding cylinders on a flat surface, which are disjoint from a given cylinder, to a problem of finding cylinders on a torus. In this section we prove a result which describes the spacing, on the circle, of the directions of these cylinders on the torus. We will use this result to obtain information about the spacing of intervals used to define the Cantor set of nonergodic directions. This information will be used to estimate the Hausdorff dimension of the Cantor set.

With a slight abuse of language we will define the angle between two points in \( \mathbb{R}^2 \), neither of which is the origin, to be the smaller angle between the lines joining those two points to the origin.

**Theorem 9.1.** Let \( \Lambda \) be a unimodular lattice in \( \mathbb{R}^2 \) and let \( \delta \) be the length of the shortest nonzero element of \( \Lambda \). For \( \delta \) sufficiently small, in any sector of radius \( r \geq 64\pi/\delta \) and angle \( \pi/2 \), there is a subsector of radius \( r \) and angle \( \pi/8 \) such that there are at least \( \pi r/100 \) primitive lattice points of distance at least \( r/2 \) from the origin and such that the angle \( \theta \) between any two satisfies

\[
2/r \leq \theta \leq 10/r.
\]

**Lemma 9.2.** Suppose \( \Delta \) is a triangle with one vertex at the origin and the other two at lattice points \( \nu_1, \nu_2 \). Suppose there are no other lattice points on the sides or the interior of \( \Delta \). Then \( \Delta \) has area 1/2.

**Proof.** Let \( x = (\nu_1 + \nu_2)/2 \) be the midpoint of \( \nu_1 \nu_2 \). Then rotation by \( \pi \) about \( x \) preserves \( \Lambda \) since it is rotation by \( \pi \) about the origin followed by translation by \( 2x = \nu_1 + \nu_2 \). Each of these motions preserves \( \Lambda \). Thus if there are no other lattice points in \( \Delta \), there are no lattice points in the interior or sides of the parallelogram \( P \) with vertices 0, \( \nu_1, \nu_2, \nu_1 + \nu_2 \). The parallelogram \( P \) is a fundamental domain for \( \Lambda \) so that, since \( \Lambda \) is unimodular, the area of \( P \) is one.

**Proof of Theorem 9.1.** Suppose \( \nu_1, \nu_2 \) are primitive lattice points in the circle of radius \( r \). Let \( \theta \) be the angle between them. Suppose there are no other lattice points inside or on the boundary of the sector determined by the two lines through the origin and the points and the circle of radius \( r \). Under these circumstances we say \( \nu_1 \) and \( \nu_2 \) are a pair of consecutive lattice points and the sector lies between them. Lemma 9.1 says the triangle with vertices \( \nu_1, \nu_2 \) and the origin has area 1/2. This gives \( 1/2 = (1/2)|\nu_1| |\nu_2| \sin \theta \). Moreover if \( |\nu_1| \leq r/2 \), then \( |\nu_2| \geq r/2 \), for otherwise \( \nu_1 + \nu_2 \) would be a lattice point in the sector between \( \nu_1, \nu_2 \), contradicting the assumption. Thus if \( |\nu_1| \geq 1 \), for
\( i = 1, 2, \) then

\[
\sin \theta = \frac{1}{|\nu_1||\nu_2|} \leq 2/r.
\]

By rotating our coordinate system, we may assume that the shortest lattice elements are \((\pm \delta, 0)\). If \( \delta \) is small, \((\pm \delta, 0)\) are the only primitive lattice elements within distance 1 of the origin. There are two primitive lattice points which form a consecutive pair with \((\delta, 0)\) and two primitive lattice points which form a consecutive pair with \((-\delta, 0)\). For each of these four pairs, the angle between the two lattice points may not satisfy (1). They do satisfy

\[
\theta \leq 2 \sin \theta \leq \frac{2}{\delta \cdot r/2} = \frac{4}{\delta r} \leq \pi/16.
\]

Call the four sectors between these four pairs of lattice points, the large sectors. Then (2) says that in any sector of angle \( \pi/2 \) there is a subsector of angle \( \pi/8 \) disjoint from these four large sectors. Since all primitive lattice points other than \((\pm \delta, 0)\) are distance at least 1 from the origin, inequality (1) says that for any consecutive pair of lattice points in the subsector of angle \( \pi/8 \), the angle \( \theta \) between them satisfies

\[
\theta \leq 2 \sin \theta \leq 4/r.
\]

Fix one line on the boundary of the sector of angle \( \pi/8 \). The line through each primitive lattice point and the origin makes an angle with this fixed line. Order the primitive lattice points in this sector so that this angle is increasing. We will choose a set of primitive lattice points in this sector satisfying the conclusions of Theorem 9.1. Pick the first lattice point \( \nu_1 \) farther than \( r/2 \) from the origin. Since consecutive lattice points cannot both be within \( r/2 \) of the origin, and the angle between consecutive lattice points is at most \( 4/r \), the angle \( \nu_1 \) makes with the fixed line is at most \( 8/r \). Consider the first lattice point \( \nu_2 \) after \( \nu_1 \) such that the angle between \( \nu_1 \) and \( \nu_2 \) is at least \( 2/r \). Since \( \nu_2 \) is the first such lattice point following \( \nu_1 \), by (1), the angle between \( \nu_1 \) and \( \nu_2 \) is at most \( 6/r \). If \( \nu_2 \) is farther than \( r/2 \) from the origin, pick it as the second lattice point in the set. If it is not farther than \( r/2 \) from the origin, the lattice point \( \nu_2' \) following \( \nu_2 \), which forms a consecutive pair with it, is farther than \( r/2 \) from the origin. Moreover the angle between \( \nu_2' \) and \( \nu_2 \) is at most \( 4/r \). Therefore the angle between \( \nu_2' \) and \( \nu_1 \) is between \( 2/r \) and \( 10/r \). In this case pick \( \nu_2' \) as the second one in the set. We now start the choosing process with the second lattice point (either \( \nu_2 \) or \( \nu_2' \)) in the set. We look for the next lattice that makes an angle of at least \( 2/r \) with the second lattice point in the set. If it is farther than \( r/2 \) from the origin, choose it as the third point in the set. If it is not farther than \( r/2 \)
from the origin, the lattice point immediately following it is farther than $r/2$
from the origin and is chosen as the third element in the set. We continue in this
fashion until we reach the other boundary line of the sector of angle $\pi/8$. The
number of elements in the set is at least $(\pi/8 - 16/r)/(10/r) \geq \pi r/100$, for
$\delta$ sufficiently small ($r$ sufficiently large).

10. Computation of measures

In this section we will make the following basic computations of measures.
We will start with a quotient stratum $\mathcal{M}$ and the union of strata $\mathcal{M}'$ which
are the faces of $\mathcal{M}$. We are given $E' \subset \bigcup \mathcal{M}'$, a set of small measure. We
think of $\bigcup \mathcal{M}' - E'$ as the subset of $\mathcal{M}'$ on which the shrinking process is
successful. We will define $E \subset \mathcal{M}$ such that for each $q \in \mathcal{M} - E$ the
following inductive shrinking process is successful. On the flat structure of $q$ we
will be able to find a shrinkable segment $\beta$ whose length is less than some
number given a priori. We will have a compact subset of $\text{SL}(2, \mathbb{R})$, given a priori,
so that if we deform $q$ via an element of this compact subset, $\beta$ becomes short.
We will then require that the deformed flat structure lie in a set $M^\delta(\tau, \eta, \prec)$ so
that the retraction map which either coalesces or pinches along $\beta$ is defined on
the deformed flat structure. We then further require that the image under the
retraction map lie in $\bigcup \mathcal{M}' - E'$ so that further shrinking is possible. We will
find estimates for the measure of $E$ in terms of the given data of lengths, the
compact subset of $\text{SL}(2, \mathbb{R})$, and the measure of $E'$.

Along the way we will reprove the basic result of Veech that the measure of
the space of normalized flat structures in each quotient stratum is finite. Our
inductive procedure on strata treats the first stratum differently from succeeding
strata. We refer to the first stratum as the initial stratum. Recall from the
introduction that for the initial stratum we need to collapse a metric cylinder to a
simple closed curve, shrink the simple closed curve so that it is short and then
pinch the curve giving a flat structure in another stratum. For a noninitial
stratum, we shrink a segment which may or may not be closed but in any case
does not bound a cylinder, and then either pinch or coalesce along the segment.
We will consider completely the case of a noninitial stratum and then point out
the modifications necessary in the initial stratum case. Assume $\mathcal{M}$ is not the
exceptional strata of flat tori or spheres with four points and, to begin, assume
also that $\mathcal{M}$ is not the initial stratum.

For the rest of this section $(\kappa', \delta')$ will refer to a pair of numbers so that

* \[ \kappa' \leq c_1 \delta'^3 \]
and

\[ \frac{c_2 \kappa'}{\delta'^0} \leq \kappa'^{1/2}, \]

where \( c_1, c_2 \) are the constants given by Theorem 6.1.

Now suppose \( E' \) is a subset of the union of the \( \mathcal{MD}' \), that is to say the faces of the sets \( M_\tau^\delta(\tau, \eta, <) \subset \mathcal{MD} \). We will think of \( E' \) as a set of small measure in the union of the \( \mathcal{MD}' \) on which the shrinking process in \( \mathcal{MD}' \) fails.

We first define two sets in \( \mathcal{MD} \) “near” the faces for which some aspect of the shrinking process will fail on the “higher” stratum \( \mathcal{MD} \). On the first the retraction map is not defined. On the second the retraction map is defined but its image lies in the set \( E' \).

**Definition.** Let \( V_1(\kappa', \delta') \) be the set of \( q \in \mathcal{MD} \) such that the shortest shrinkable geodesic segment \( \beta \) has length at most \( \kappa' \) and some nonhomologous geodesic segment \( \gamma \) has length at most \( \delta' \) or two nonhomologous geodesic segments have the same length.

**Definition.** Let \( V_2(E', \kappa', \delta') \) be the set of \( q \in \mathcal{MD}' \) such that the shortest shrinkable geodesic segment \( \beta \) has length at most \( \kappa' \), \( q \notin V_1(\kappa', \delta') \), but the retraction map \( f \) applied to \( q \) which collapses \( \beta \) has image in \( E' \).

Note that if the shortest shrinkable saddle connection has length at most \( \kappa' \) and \( q \notin V_1(\kappa', \delta') \), then by definition \( q \in M_\tau^{\delta}(\tau, \eta, <) \), so that the retraction map is defined and so the definition of \( V_2 \) makes sense. Now for each \( L \), let

\[ \mathcal{MD}(L) = \{ q \in \mathcal{MD} : q \) has a shrinkable geodesic segment of length at most \( L \} \].

Suppose \( \kappa' \) is given. We want to associate to each \( q \in \mathcal{MD}(L) \) an element of \( \text{SL}(2, \mathbb{R}) \) so that if that element of \( \text{SL}(2, \mathbb{R}) \) is applied to \( q \), the result is a flat surface with a shrinkable segment of length at most \( \kappa' \). To do this, we establish the following notation. Divide the circle of directions into disjoint half-open intervals \( I_j \), all of the same size, with size as close as possible to, but less than \( \kappa'^2/2L^2 \). For \( q \in \mathcal{MD}(L) \) let \( \phi(q) \) be the angle such that the holonomy of the shortest shrinkable geodesic segment saddle has zero horizontal coordinate with respect to the flat structure \( r_{\phi(q)} q \). Then \( \phi(q) \in I_j \) for some unique \( j \). Define \( \theta = \theta(q) \) to be the left endpoint of \( I_j \). Let

\[ \mathcal{MD}_j(L) = \{ q \in \mathcal{MD}(L) : \phi(q) \in I_j \} \].

Define \( t \) by

\[ e^{t/2} = \sqrt{2} L / \kappa' \].
Lemma 10.1. For $q \in \mathcal{MD}(L)$, $g_t r_{\theta(q)} q$ has a shrinkable segment of length at most $\kappa'$.

Proof. We measure the holonomy coordinates $(h, v)$ of the shortest shrinkable segment $\beta$ with respect to the flat structure $r_{\theta(q)} q$ and then compute them after flowing time $t$. Since $\beta$ has length at most $L$, automatically $v \leq L$. Flowing by $t$ contracts vertical coordinates by $e^{t/2} = \sqrt{2} L / \kappa'$. The vertical coordinate of $\beta$ with respect to $g_t r_{\theta(q)} q$ is therefore at most $\kappa' / \sqrt{2}$. On the other hand since both $\phi(q)$ and $\theta(q)$ belong to $I_j$,

$$h/|\beta| = |\sin(\phi(q) - \theta(q))| \leq |\phi(q) - \theta(q)| \leq \frac{\kappa'^2}{2L^2}.$$ 

Since $|\beta| \leq L$, and flowing by $t$ increases horizontal coordinates by $e^{t/2}$, the horizontal coordinate of $\beta$ with respect to $g_t r_{\theta(q)} q$ is also at most $\kappa' / \sqrt{2}$. The lemma follows.

Definition. Let $E$ be the set of $q$ in $\mathcal{MD}$ such that either $q \in \mathcal{MD} - \mathcal{MD}(L)$ or if $q \in \mathcal{MD}(L)$ then $g_t r_{\theta} q \in V_1(\kappa', \delta') \cup V_2(E', \kappa', \delta')$.

The next proposition whose proof is almost immediate says that on the complement of $E$, the shrinking process as described in the introduction to this section is possible.

Proposition 10.2. For $q \notin E$, $g_t r_{\theta(q)} q$ has a shrinkable geodesic segment $\beta$ of length at most $\kappa'$. The retraction map $f$ is defined on the flat surface of $g_t r_{\theta(q)} q$ and has image in the complement of $E'$. The collapsing map $h$ is a $\sqrt{\kappa'}$ distortion on the complement of $\beta$.

Proof. By definition, $q \in \mathcal{MD}(L)$. By Lemma 10.1, $g_t r_{\theta(q)} q$ has a shrinkable segment $\beta$ of length at most $\kappa'$. Since $g_t r_{\theta(q)} q \notin V_1(\kappa', \delta') \cup V_2(E', \kappa', \delta')$, the image under the retraction map does not lie in $E'$. The distortion estimate comes from Theorem 6.1 v) and **.

The main purpose of this section is to prove Theorems 10.3, 10.4, and 10.5. Recall that $p$ is the constant in the conclusion of Theorem 6.1. Let $\mu$ and $\mu'$ refer to the measures in $\mathcal{MD}$ and $\mathcal{MD}'$ respectively.

Theorem 10.3. There is an absolute constant $c$ such that if $L$ is sufficiently large, $\kappa'$, $\delta'$ and $\mu'(E') = \epsilon'$ are sufficiently small, then $\mu(E) < c(L^2 \epsilon' / \delta' p + 1/L^2 + L^2 / \delta'^2)$.

Before proving the theorem we state a corresponding but easier result for the exceptional strata.
Definition. Let $\mathcal{D}$ be the exceptional torus stratum. For each $a > 0$ let $E(a)$ be the set of $q \in \mathcal{D}$ such that the shortest simple closed curve has length less than or equal to $a$.

Definition. For $\mathcal{D}$ the exceptional sphere stratum and $a > 0$, let $E(a)$ be the set of $q \in \mathcal{D}$ such that the shortest segment joining two singularities has length less than or equal to $a$.

Theorem 10.4. There is a constant $c$ such that for either exceptional stratum, $\mu(E(a)) \leq ca^2$.

Proof of Theorem 10.3. We will identify three sets whose union contains $E$ and show that the measure of each set is less than a term in the sum.

Recall from Sections 4 and 5 the definition of diameter of a surface $q$, denoted $d(q)$ as the maximum distance of any point to a singularity. Let $d_s(q)$ denote the diameter in the standard sense. We need a simple relationship between the two definitions.

Claim. There is a constant $c'$ depending only on the stratum such that $d(q) \leq d_s(q) \leq c'd(q)$ for all $q$.

Proof of claim. The left-hand inequality is obvious. For the right-hand inequality, take the Delaunay triangulation of $q$. The edges have length at most $2d(q)$ and the maximum distance of any point to a vertex is at most $d(q)$. The number of edges $k$ in the triangulation depends only on the stratum. The maximum distance between any two vertices is $2kd(q)$. The maximum distance between any two points is therefore $(2k + 2)d(q)$, proving the claim.

Definition. $E_1$ is the set of $q \in \mathcal{D}$ such that $d_s(q) \geq L/16$.

Definition. $E_2$ is the set of $q \in \mathcal{D}(L)$ such that $\varphi_t r_{\theta(q)}q \in V_1(\kappa', \delta')$.

Definition. $E_3$ is the set of $q \in \mathcal{D}(L)$ such that $\varphi_t r_{\theta(q)}q \in V_2(E', \kappa', \delta')$.

Suppose now $q \in E$. We show first that $q \in E_1 \cup E_2 \cup E_3$. If $q \in (\mathcal{D} \setminus \mathcal{D}(L)) \cap E$, it does not have a shrinkable segment of length less than or equal to $L$. By Proposition 7.1, $d_s(q) \geq L/16$ so that $q \in E_1$. If $q \in \mathcal{D}(L) \cap E$ then $\varphi_t r_{\theta(q)}q \in V_1(\kappa', \delta') \cup V_2(E', \kappa', \delta')$ by definition, and so $q \in E_2 \cup E_3$.

The main step is to estimate the measure of each $E_i$. We will do this by estimating the measure of the intersection of $E_i$ with each canonical triangulation piece $M(\tau, \eta, \prec)$. This is sufficient since a bounded number of $M(\tau, \eta, \prec)$ cover the quotient space $\mathcal{D}$, the bound depending only on the stratum. Now on each $M(\tau, \eta, \prec)$ the developing map defines an injection into the vector
space $H^1$. We will choose a set of edges which form a homology basis. Their
holonomy vectors then form global coordinates for $M(\tau, \eta, <)$. We may then
identify $H^1$ with Euclidean space and the measure $\mu$ with a product Lebesgue
measure. Let $\lambda_1$ denote $l$ dimensional Lebesgue measure. We will estimate
measures of various sets by iterated integration.

We begin with $E_1$ and will show that

$$\mu(E_1) \leq \frac{c}{L^2}$$

for some constant $c$. For any $d$ consider the set $H(d)$ of flat structures
$q \in M(\tau, \eta, <)$ such that $d(q) \geq d$. Here we are using diameter in the sense of
Sections 4 and 5. By the claim, it suffices to show that

$$(1) \quad \mu(H(d)) \leq \frac{c}{d^2}$$

for some constant $c$ that does not depend on $d$. We first consider $M(\tau, \eta, <) - H(s)$ where as in Section 5, $s = \sqrt{2/\pi}$. The holonomy vectors of the edges of
the flat structures in $M(\tau, \eta, <) - H(s)$ have length at most $2s$ and therefore
the image of $M(\tau, \eta, <) - H(s)$ under the developing map is a bounded set in
Euclidean space. Therefore for some $c'$ we have

$$\mu(M(\tau, \eta, <) - H(s)) \leq c' = \frac{c's^2}{s^2}.$$ 

Now if $d \leq s$, the above inequality implies

$$(2) \quad \mu(H(d) - H(s)) \leq \mu(M(\tau, \eta, <) - H(s)) \leq \frac{c's^2}{d^2}.$$ 

Therefore (2) implies that we need only prove (1) in the case that $d \geq s$.

Therefore assume $d \geq s$. By Theorem 5.3, each $q \in H(d)$ contains a
disjoint set of metric cylinders $A_1(q), \ldots, A_n(q)$ whose heights are larger than
their circumference. We say those metric cylinders belong to $q$. By Corollary 5.5
there is a constant $C$ such that at least one of the cylinders belonging to $q$ has
height $h \geq Cd$. Since the area of any metric cylinder is at most 1, the
circumference $c$ of that cylinder satisfies

$$c \leq \frac{1}{h} \leq \frac{1}{Cd}.$$ 

By Proposition 5.4, the Delaunay triangulation of $q$ consists of edges on the
boundary of the cylinders $A_i$, other edges of length at most $2s$, and edges that
cross the cylinders $A_i$. Since the triangulation and hence the edges of $\tau$ are
fixed, there are a bounded number of homotopy classes of curves that can be the
waist curve of a metric cylinder belonging to \( q \in H(d) \). The bound depends only on the number of edges in the triangulation.

Fix then any collection of disjoint metric cylinders \( A_1, \ldots, A_n \) that belong to some \( q_0 \in H(d) \). Since the number of homotopy classes of such collections is bounded, it is enough to give the required estimate of the measure in \( H(d) \) of those \( q \) such that the metric cylinders that belong to \( q \) are in the classes of \( A_1, \ldots, A_n \). Restrict \( q \) to lie in this set. Then each edge of the triangulation is one of three types, independent of the \( q \) in the set. Namely, it is either on the boundary of a metric cylinder in the homotopy class of an \( A_i \) or it crosses one of these cylinders or it neither crosses a metric cylinder or is on the boundary of one.

Choose a set \( e_1, \ldots, e_j \) of edges that are on the boundary of the metric cylinders homotopic to the \( A_i \), that are linearly independent in homology and such that any other edge on the boundary of the cylinders is a linear combination of the \( e_i \). Let \( e_{j+1}, \ldots, e_k \) be a further set of edges so that \( e_1, \ldots, e_k \) is a homology basis. For each flat structure the holonomy vector \( \text{hol}(e_j) \) of each edge has two components. Our strategy is to fix a \( j \)-tuple of vectors \( v = (\text{hol}(e_1), \ldots, \text{hol}(e_j)) \) and consider the slice

\[
H_v(d) = \left\{ (\text{hol}(e_{j+1}), \ldots, \text{hol}(e_k)) \right\} \subset \mathbb{R}^{2k-2j}
\]

such that \( (\text{hol}(e_1), \ldots, \text{hol}(e_j), \text{hol}(e_{j+1}), \ldots, \text{hol}(e_k)) \in H(d) \). We will first show that \( m(v) = \lambda_{2k-2j}(H_v(d)) \) is bounded independently of \( v \) where \( \lambda_{2k-2j}(H_v(d)) \) is the \((2k - 2)\)-dimensional Lebesgue measure of this slice. Then by iterated integration we will compute

\[
\mu(H(d)) = \int m(v) \, d\lambda_{2j}(v),
\]

where \( \lambda_{2j} \) is \( 2j \)-dimensional Lebesgue measure.

To show \( \lambda_{2k-2j}(H_v(d)) \) is bounded, it suffices to show \( H_v(d) \subset X_{j+1} \times \cdots \times X_k \) where each \( X_i \) is a subset of \( \mathbb{R}^2 \) of bounded 2-dimensional Lebesgue measure.

Let \( e_i, i > j \), be any edge. There are two possibilities. Either \( e_i \) does not cross a metric cylinder or it does. In the first case \(|e_i| = |\text{hol}(e_i)| \leq 2s \) for all flat structures in the set. Therefore the set of vectors \( \text{hol}(e_i) \) lie in a rectangle \( X_i \) with sides \( 4s \) which clearly is a set of bounded 2-dimensional measure.

Next suppose \( e_i, i > j \), crosses one of the metric cylinders \( A_i \) homotopic to one of the \( \{A_j\} \). We will show \( \text{hol}(e_i) \) lies in a set \( X_i \subset \mathbb{R}^2 \) of bounded Lebesgue measure even though \( X_i \) itself need not be a bounded set. Let

\[
c(A) = (c_1(A), c_2(A))
\]
be the coordinates of the holonomy vector of the boundary of $A$. The circumference of $A$ is $|c(A)|$. Let $|h(A)|$ be the height of $A$. The boundary of $A$ is a linear combination of the maximal linearly independent set of edges $e_1, \ldots, e_j$ on the boundaries of the $A_i$. Since the holonomy vectors of these edges are assumed to be fixed, the vector $c(A)$ is fixed. Let

$$w = (w_1(A), w_2(A))$$

be the coordinates of $\text{hol}(e_i)$. Let $X_i \subset \mathbb{R}^2$ consist of the set of vectors $w$ that occur as we vary over the set of flat structures $q$ such that the metric cylinders that belong to $q$ are in the homotopy classes of the $A_i$. We wish to estimate the Lebesgue measure $\lambda_2$ of $X_i$. This is done by iterated integration.

The edge $e_i$ crosses $A$ and intersects any segment that crosses $A$ and is perpendicular to the boundary, at most once. The area of $A$ is at most 1. The circumference is less than the height. We have

$$|\text{hol}(e_i)| = |w|^2 \leq |h(A)|^2 + |c(A)|^2 \leq 2|h(A)|^2 \leq \frac{2}{|c(A)|^2}.$$  

Without loss of generality we may assume $|c_2(A)| \geq |c_1(A)|$. Otherwise we reverse the roles of $c_1$ and $c_2$ in the arguments below.

**Claim.** For a fixed value of $w_1(A)$, and the above assumption, the values of $w_2(A)$ lie in an interval $I_{w_1}(A)$ of size at most $4|c(A)|/\sqrt{2}$.

**Proof of claim.** Recall that the vector $c(A)$ is assumed fixed as well as $w_1$. Suppose first we are in the special case that $c_1(A) = 0$ so that the boundary of $A$ is a vertical vector. Therefore $e_1$ crosses each horizontal segment crossing $A$ at most once. This implies

$$|w_2(A)| \leq |c_1(A)| = |c(A)|.$$  

Next suppose $0 < |c_1(A)| \leq |c_2(A)|$. Choose $-\pi/4 \leq \theta \leq \pi/4$ so that

$$(\cos \theta)c_1(A) + (\sin \theta)c_2(A) = 0.$$  

Then the holonomy in the horizontal direction of the boundary of $A$ is 0 with respect to the rotated flat structure $r_{\theta}q$. The special case (4) says that the vertical holonomy $w_2'(A)$ of $e_i$ with respect to $r_{\theta}q$ satisfies

$$-|c(A)| \leq w_2'(A) \leq |c(A)|.$$  

Since

$$w_2'(A) = (-\sin \theta)w_1(A) + (\cos \theta)w_2(A),$$
we have
\[ |c(A)| - (\sin \theta)w_1(A) \leq (\cos \theta)w_2(A) \leq |c(A)| - (\sin \theta)w_1(A). \]
Now \( w_1(A) \) is assumed to be fixed and \( |\cos \theta| \geq \sqrt{2}/2 \). Then (5) implies that the possible values of \( w_2(A) \) lie in an interval \( I_{w_1(A)} \) of size
\[ \frac{4|c(A)|}{\sqrt{2}}. \]
The claim follows.

Now by the claim, (3), and Fubini’s Theorem,
\[ \lambda_2(X_i) \leq \int_0^{\sqrt{2}/|c(A)|} \left( \int_{I_{w_1(A)}} dw_2(A) \right) dw_1(A) \leq 4. \]
This shows that the two dimensional Lebesgue measure of this set of vectors is bounded and concludes the proof that
\[ m(v) = \lambda_{2k-2j}(H_v(d)) \]
is bounded. Now
\[ \mu(H(d)) = \int m(v) \, d\lambda_{2j}(v). \]
The holonomy vector of each \( e_i, i \leq j \), is bounded by \( 2s \) and at least one of the vectors has length at most \( 1/Cd \). Since each vector has two components, the multivector \( v \) lies in a rectangle in \( \mathbb{R}^{2j} \) all of whose sides have length at most \( 4s \) and two of whose sides have length at most \( 2/Cd \). When we integrate the bounded function \( m(v) \) over this rectangle we find the measure is at most
\[ c/d^2 \]
for some constant \( c \). This completes the computation of the measure of \( E_1 \).

We now show
\[ \mu(E_2) \leq cL^2\delta^{12}. \]
Again we need only consider a fixed \( M(\tau, \eta, \prec) \). We claim
\[ \mu(V_1(\kappa', \delta')) \leq c\kappa'^2\delta'^2 \]
for a universal constant \( c \). To prove the claim we may again restrict ourselves to the intersection of \( V_1(\kappa', \delta') \) with the flat structures having a given Delaunay triangulation. Via the developing map we consider \( V_1 \) as a subset of \( \mathbb{R}^{2k} \). First note that the set of flat structures with two nonhomologous edges with the same length have measure zero; so we may disregard that set. Now if \( \beta \) and \( \gamma \) are not both edges of the Delaunay triangulation, there are a pair of nonhomologous
edges in the Delaunay triangulation whose lengths are not larger than $\beta$ and $\gamma$ respectively. Thus from the point of view of finding upper bounds for the measure we may as well assume that $\beta$ and $\gamma$ are edges. We choose $\beta$ and $\gamma$ as two elements $e_1$ and $e_2$ of a homology basis and complete to a basis $e_1, e_2, \ldots, e_k$. Our strategy is similar to the case of $E_1$. We fix a pair of holonomy vectors $v = (\text{hol}(e_1), \text{hol}(e_2))$ and consider the slice

$$V_1^v(\kappa', \delta') = \{(\text{hol}(e_3), \ldots, \text{hol}(e_k)) \in \mathbb{R}^{2k-4}\}$$

such that

$$\text{hol}(e_1), \text{hol}(e_2), \text{hol}(e_3), \ldots, \text{hol}(e_k) \in V_1(\kappa', \delta').$$

We first show that

$$m(v) = \lambda_{2k-4}(V_1^v(\kappa', \delta'))$$

is bounded independently of $v$ where $\lambda_{2k-4}$ is $(2k - 4)$-dimensional Lebesgue measure. The proof is identical to the previous proof that $H_{\ast}(d)$ had bounded Lebesgue measure in the considerations of $E_1$. Namely, we divide the holonomy vectors of $e_i, i \geq 3$, into the two cases. Either $|e_i| \leq 2s$ or $e_i$ crosses a metric cylinder whose height is larger than its circumference. In the first case the vectors $\text{hol}(e_i)$ lie in a bounded rectangle in $\mathbb{R}^2$. In the second case, by (6) again, the vectors $\text{hol}(e_i)$ lie in a set $X_i \subset \mathbb{R}^2$ of bounded 2-dimensional Lebesgue measure.

Now

$$\mu(V_1(\kappa', \delta')) = \int m(v) \, d\lambda_4(v)$$

where $\lambda_4$ is 4-dimensional Lebesgue measure. Since $|e_1| = |\text{hol}(e_1)| \leq \kappa'$ and $|e_2| = |\text{hol}(e_2)| \leq \delta'$, the 4-tuple $v$ lies in rectangle with two sides of length at most $2\kappa'$ and two sides of length at most $2\delta'$. Since $m(v)$ is bounded, we find

$$\mu(V_1(\kappa', \delta')) \leq c\kappa'^2\delta'^2,$$

proving Claim (7).

Now consider the sets $\mathcal{M}_j(L)$ defined in the paragraph before Lemma 10.1. By definition $\theta = \theta(q)$ is constant on $\mathcal{M}_j(L)$ and $g_tr_\theta q \in V_1(\kappa', \delta')$ for all $q \in \mathcal{M}_j(L) \cap E_2$. The flow $g_t$ and the rotation $r_\theta$ are measure-preserving. These facts and (7) imply

$$\mu(\mathcal{M}_j(L) \cap E_2) \leq c\kappa'^2\delta'^2$$

for each $j$. There are $3\pi L^2/\kappa'^2$ such indices $j$. This gives

$$\mu(E_2) \leq c(L^2/\delta'^2)\kappa'^2\delta'^2 = cL^2\delta'^2.$$
We will now show that

$$\mu(E_3) \leq \frac{cL^2\varepsilon'}{\delta''}. $$

We again restrict ourselves to a canonical triangulation piece $M(\tau, \eta, \prec)$. Let $f$ be the retraction map defined on $M_\kappa^\delta(\tau, \eta, \prec)$. Then

$$f^{-1}(E) = V_2(E', \kappa', \delta').$$

We first show

$$\mu(f^{-1}(E)) \leq \frac{c\kappa'^2\varepsilon'}{\delta''}. $$

Suppose $e_1$ is any edge. It is enough to show (8) for $f$ restricted to the flat structures in $V_2(E', \kappa', \delta')$ for which $e_1$ is the shortest edge. Let $e_2, \ldots, e_k$ be edges, so that together with $e_1$ they form a homology basis. Let $v_i$ denote the holonomy vector of $e_i$. For each flat structure $q$ the image of $\tau$ under the collapsing map $g$ is a geodesic triangulation $\tau'$ of $f(q)$ which is not necessarily Delaunay. Denote by $e'_2, \ldots, e'_k$ the images of $e_2, \ldots, e_k$ under $g$. They form a homology basis for $\tau'$. Let $v'_i$ be the holonomy vector of $e'_i$. Fix a vector $v_1$ with $|v_1| \leq \kappa'$. According to Theorem 6.1 vi), the map $f$ restricted to the set of $q$ such that the shortest segment $e_1$ has holonomy vector $v_1$ is at most $c/\delta''$ to 1. Using Theorem 6.1 iv) we may therefore decompose the domain of $f$ into a disjoint union of at most $c/\delta''$ sets $\Omega$ such that on each $\Omega$, the map $f$ has the form

$$ (v'_2, \ldots, v'_k) = (v_2 - \varepsilon_2 v_1, \ldots, v_k - \varepsilon_k v_1) $$

where $\varepsilon_i$ is among $-1, 1, 0$. This map can be written as a composition of a map $f_1$ which is

$$ (v_1, v_2, \ldots, v_k) \rightarrow (v_1, v_2 - \varepsilon_2 v_1, \ldots, v_k - \varepsilon_k v_1) $$

followed by the projection $f_2$ which is

$$ (v_1, v_2, \ldots, v_k) \rightarrow (v_2, \ldots, v_k). $$

Now $f$ has the form $f_2 \circ f_1$. We will consider $f$ restricted to each $\Omega$.

Consider the inverse image of $E'$ under the projection $f_2$ restricted to $\Omega$. The measure $\mu'$ on $\mathcal{M} \Omega'$ is defined by the holonomy vectors $v'_2, \ldots, v'_k$ because $f(q) \in N(\tau', \eta')$. Via the projection, the measure $\mu$ can be considered as the product measure of $\mu'$ together with Lebesgue measure in $\mathbb{R}^2$ defined by the holonomy vector $v_1$. Since $v_1$ has two coordinates, $|v_1| \leq \kappa'$ and $\mu(E') = \varepsilon'$, we
find
\[ \mu\left(f_2^{-1}(E') \cap \Omega\right) \leq 4\kappa'^2\varepsilon'. \]
The map \( f_1 \) is a nonsingular linear map so that it multiplies the measure \( \mu \) by a constant factor. Therefore
\[ \mu\left(f^{-1}(E') \cap \Omega\right) \leq c\kappa'^2\varepsilon'^2. \]
The estimate on the number of \( \Omega \) implies
\[ \mu(f^{-1}(E')) = \mu(V_2(E', \kappa', \delta')) \leq \frac{c\kappa'^2\varepsilon'}{\delta''}. \]
which is (8). The rest of the argument follows just as in the discussion of \( E_2 \).
Now we have by definition
\[ g_tr_\theta(HD_j(L) \cap E_3) \subset V_2(E', \kappa', \delta'). \]
Since \( g_t \) and \( r_\theta \) preserve the measure \( \mu \),
\[ \mu(HD_j(L) \cap E_3) \leq \frac{c\kappa'^2\varepsilon'}{\delta''}. \]
Since there are \( 3\pi L^2/\kappa'^2 \) indices \( j \),
\[ \mu(E_3) \leq \frac{cL^2\varepsilon'}{\delta''}, \]
for a new constant \( c \). This concludes the proof of Theorem 10.3.

**Proof of Theorem 10.4.** In the case of the flat torus the shortest simple closed curve \( e_1 \) has length at most \( a \) and in the case of the sphere the shortest geodesic \( e_1 \) joining two cone angle \( \pi \) singularities has length at most \( e \). In either case, choose a second geodesic \( e_2 \) crossing the metric cylinder once so that \( e_1 \) and \( e_2 \) are a basis for homology. Fix \( v = \text{hol}(e_1) \), and let \( E_v(a) \) be the slice \( \{ \text{hol}(e_2) : (\text{hol}(e_1), \text{hol}(e_2)) \in E(a) \} \). The estimate (6) again shows \( m(v) = \lambda_2(E_v(a)) \) is bounded by some \( c \) independently of \( v \). Since \( |e_1| \leq a \),
\[ \mu(E(a)) = \int m(v) \, d\lambda_2(v) \leq ca^2. \]
This proves the theorem.

We now indicate the differences for the case of the initial stratum. Now we will have a homotopy class of simple closed curves \( \beta \) which are nondividing.

**Definition.** For any \( L \) let \( KD(\beta, L) \) be the set of \( q \in KD \) such that \( q \) has a metric cylinder in the homotopy class of \( \beta \) of length less than or equal to \( L \).
Definition. We say a metric cylinder is shrinkable if each boundary component contains exactly one singularity with cone angle \( k\pi, \; k \geq 3 \), or one boundary component contains exactly one singularity with cone angle \( k\pi, \; k \geq 3 \), and the other contains two cone angle \( \pi \) singularities.

For each shrinkable metric cylinder \( C \) we indicate a canonical way of collapsing it to either a simple closed curve or a segment, producing in either case a flat structure in a different stratum.

1) Each boundary component of \( C \) is a simple closed curve joining a point of \( \Sigma \) to itself. Collapse \( C \) to a simple closed curve and collapse the two points of \( \Sigma \) on the boundary components to a single point. The dimension of the stratum of the new flat structures is two less than the dimension of the stratum we start with.

II) One boundary component of \( C \) is a segment joining two cone angle \( \pi \) singularities. The other component is a simple closed curve \( \beta \) joining a point \( x \) of \( \Sigma \), with cone angle \( k\pi \), to itself. Then \( \beta \) bounds a disc on the surface. Remove the disc. Identify points on \( \beta \) which are of equal distance on \( \beta \) from \( x \). The result is a point with cone angle \( (k - 1)\pi \) corresponding to \( x \), a point with cone angle \( \pi \) corresponding to the point antipodal to \( x \) and a geodesic segment joining them. This gives a flat structure in a new stratum and the dimension of the stratum is two less, again.

We give definitions of the two sets \( V_1 \) and \( V_2 \) similar to the definitions given before.

Definition. \( V_1(\beta, \kappa', \delta') \) consists of the set of flat structures in \( \mathcal{M}(\beta, \kappa') \) such that some saddle connection not homologous to \( \beta \) has length at most \( \delta' \), or two nonhomologous segments have the same length.

Definition. Let \( V_2(E', \beta, \kappa', \delta') \) be the set of flat structures that are in \( \mathcal{M}(\beta, \kappa') - V_1(\beta, \kappa', \delta') \), but the map which is the composition of first shrinking the cylinder and then applying the retraction map has image in \( E' \).

Remark. For \( q \in \mathcal{M}(\beta, \kappa') - V_1(\beta, \kappa', \delta') \), the cylinder is shrinkable. For if it were not shrinkable there would be two or more edges on the boundary not homologous to \( \beta \), each of length at most \( \kappa' \). Further, after collapsing the cylinder, the resulting flat structure lies in a set \( M_{\delta'}^\kappa(\tau, \eta, <) \) so that the retraction map is defined and the definition of \( V_2 \) makes sense.

Define the intervals \( I_j \) exactly as before. For each \( \kappa' \), we give \( t \) the same definition. In this case, the Main Theorem states the following:

Theorem 10.5. Suppose \( E' \in \mathcal{M}(\kappa', \delta') \), and \( L \) are given and \( \mu'(E') = \epsilon' \). Define \( E = \{ q \in \mathcal{M}(L, \beta); \; g_t r_q \in V_1(\beta, \kappa', \delta') \cup V_2(E', \beta, \kappa', \delta') \} \). Then
there is a constant $c$ such that for $L$ sufficiently large, and $\varepsilon', \delta', \kappa'$ sufficiently small, $\mu(E) < c(L^2 \varepsilon'/\delta'^p + L^2 \delta'^q)$.

We note the changes in the proof. Since we are already working with the set $\mathcal{M}(\beta, L)$ of flat structures with cylinders of length at most $L$ we do not need to consider the set $E_1$. The sets $E_2, E_3$ are defined as before.

The measure of $V_1(\beta, \kappa', \delta')$ is computed as before. We note that the holonomy vector of a segment crossing the metric cylinder defined by $\beta$ may be large, but exactly as in the computation (6) of the Lebesgue measure in $\mathbb{R}^2$ of the set $X$, it provides a bounded contribution to the product measure.

The computation of the measure of $V_2(E', \beta, \kappa', \delta')$ is slightly different because of the presence of an edge crossing the cylinder of $\beta$ in the Delaunay triangulation. Let $\gamma$ be that edge. Let $w$ be its holonomy vector; $v_1, \ldots, v_k$ together with $w$ are the holonomy vectors for a basis, where $v_1$ is the holonomy vector of $\beta$. The holonomy map

$$(w, v_1, \ldots, v_k) : M(\tau, \eta, <) \to \mathbb{R}^{2k+4}$$

is an injection and defines the measure $\mu$. The image of any edge other than $\beta$ and $\gamma$ under the collapsing map is an edge with holonomy vector $v'_i$. Again these holonomy vectors define the measure $\mu'$. Then the map which is the composition of first collapsing the cylinder and then applying the retraction map $f$ has the form

$$(v'_2, \ldots, v'_k) = (v_2 - \varepsilon_2 v_1, \ldots, v_k - \varepsilon_k v_1)$$

followed by projection on the last $k - 1$ vectors. We have again $|\beta| = |v_1| \leq \kappa'$ and the same estimate, $c/\delta'^p$, on the number of inverse images of any $q' \in E'$ with given $v_1$ and $w$. The same estimate as given in Theorem 10.4 based on (6) shows that the measure in $\mathbb{R}^4$ defined by the holonomy vectors of $v_1$ and $w$ is at most $c \kappa'^2$. The form of the map $f$ again shows that the inverse image of $E'$ still has measure $\leq c \kappa'^2 \varepsilon'/\delta'^p$.

Finally we note:

**Theorem 10.6 (V3).** $\mu(\mathcal{M}) < \infty$.

**Proof.** As in the proof of Theorems 10.3, 10.4, and 10.5, we need only consider edges which cross metric cylinders. The estimate (6) shows as before that they contribute a bounded amount to the iterated product measure.

11. Construction of Cantor sets and the proof of the Main Theorem

In this section we prove the Main Theorem. We first need a lemma which gives an estimate for the number of simple closed curves on a surface.
Lemma 11.1. There is a constant $\omega$ depending only on $\mathcal{H}$ such that, for each flat structure of area one, there is a constant $c$ such that the number of simple closed curves of length less than or equal to $T$ is bounded above by $cT^{\omega}$.

Proof. By the results of [F-L-P], if the genus of the surface is at least two, there is a constant $c$ such that with respect to the conformally equivalent Poincaré metric $\rho$, the number of simple closed geodesics of length at most $T$ is at most $cT^{\frac{5}{2}g - 6}$. The lemma follows from the next claim.

Claim. For any flat metric $q$, the ratio of the length of the geodesic in a homotopy class to the $\rho$ length of the geodesic in the same homotopy class is bounded above, and below away from zero.

Proof of claim. For any tangent vector $v$ to the surface, let $\rho(v)$ and $q(v)$ denote the lengths with respect to the metrics $\rho$ and $q$. Let $||l||_\rho$ and $||l||_q$ denote the lengths of a curve $l$. Let $V$ be an open set which is a disjoint union of small open neighborhoods of the singularities of $q$.

Since the complement of $V$ is compact, there are positive constants $c_1$ and $c_2$ so that for any tangent vector $v$ based at a point in the complement of $V$,

$$c_1 < \frac{\rho(v)}{q(v)} < c_2.$$

Now let $\beta$ be the Poincaré geodesic for its homotopy class. Replace each component $l$ of $\beta \cap V$, by an arc $l'$ on the boundary of $V$ with the same endpoints. For some constant $c$ we have $||l'||_\rho \leq c||l||_\rho$. There is a simple closed curve $\beta'$ homotopic to $\beta$ such that $||\beta'||_\rho \leq c||\beta||_\rho$. Since $\beta'$ is contained in the complement of $V$,

$$||\beta'||_q \leq c_2||\beta'||_\rho \leq cc_2||\beta||_\rho.$$

Because the geodesic with respect to the metric $q$ in the homotopy class of $\beta$ is shorter than $||\beta'||_q$, we have the inequality in one direction.

The proof of the inequality in the other direction is similar. Take the geodesic $\beta$ for the metric $q$ and find a homotopic $\beta'$ in the complement of $V$ such that

$$||\beta'||_q \leq c||\beta||_q.$$

The Poincaré length of $\beta'$ has length bounded in term of the length with respect to $q$ and the geodesic in the class of $\beta$ is even shorter. This proves the claim.

If the surface has genus 0 or 1, we may pass to the double cover, ramified over the nonorientable singularities. Each simple curve either lifts to a simple closed curve, or the simple curve traversed twice lifts to a simple closed curve.
The double cover has genus at least two so that we can apply the previous argument.

Before proving the Main Theorem we state the theorem that allows us to form the family of trees of metric cylinders. This is the inductive step of the construction. First we establish some notation. Recall the constant $p$ from Theorem 6.1 vi) and appearing in the statements of Theorems 10.3 and 10.5. Let

$$\tau = \max(2, p, \omega).$$

For $\alpha$ a homotopy class, let $\mathcal{MD}^1(\alpha, T)$ be the set of (area-1) flat structures $q \in \mathcal{MD}^1$ which have a metric cylinder $A_{\alpha}$ in the homotopy class of $\alpha$ of length at most $T$. Recall also that $\mu^1$ refers to the measure on the area-1 flat structures in a stratum, while $\mu$ simply refers to measure on the stratum.

**Definition.** If $q \in \mathcal{MD}(\alpha, T)$, denote by $\theta_\alpha$ the direction of the waist curve of the metric cylinder in the homotopy class of $\alpha$, the direction with respect to the flat structure $q$.

**Theorem 11.2.** There are positive constants $c_1, c_2, p_1, p_2, s, k$, such that $p_2 > p_1 > \max(s, 2)$ and $k > 1$, with the following properties. For any natural number $n$, homotopy class $\alpha$, and number $T_0$ sufficiently large, for all $T \geq (400/T_0)^{p_1^s-1}$, there is a subset $F(\alpha) \subset \mathcal{MD}^1(\alpha, T)$ and for each $q \in F(\alpha)$, there is a set $\Gamma_q(\alpha)$ of homotopy classes of curves such that for each $\gamma \in \Gamma_q(\alpha)$ the surface $q$ contains a metric cylinder $A_\gamma$ in the homotopy class $\gamma$ such that $A_\alpha \cap A_\gamma = \emptyset$. In addition,

i) $\mu^1(\mathcal{MD}^1(\alpha, T) - F(\alpha)) = O((1/T)^{1/2})$.

ii) The cardinality of $\Gamma_q(\alpha)$ is at least $T^s$.

iii) area$(M - (A_\alpha \cup A_\gamma)) < 1/10 n^2$.

iv) For each $q \in F(\alpha)$ there is a $p'$ satisfying $p_1 < p' < p_2$ such that for each $\gamma \in \Gamma_q(\alpha)$, $T^{p_1} \leq c_1 T^{p'} \leq |\gamma|_q \leq c_2 T^{p'} \leq T^{p_2}$, length measured with respect to the metric defined by $q$.

v) For each $q \in F(\alpha)$ there is an interval $\Delta$ in the circle of directions of width $1/T^k$ centered at $\theta_\alpha$ containing all $\theta_\gamma$ for $\gamma \in F_q(\alpha)$.

vi) For $\gamma, \gamma' \in \Gamma_q(\alpha)$, $|\theta_\gamma - \theta_{\gamma'}| \geq 4(c_2 T^{p'})^{-k}$.

**Remarks.** The theorem says that each $q \in F(\alpha)$ has metric cylinders $A_\gamma$ disjoint from $A_\alpha$. Condition iii) about the areas is needed to apply Theorem 2.1. That theorem allows us to conclude that if we have a flat structure $q$, and if by repeatedly applying Theorem 11.2, we can find a sequence $A_{\gamma_n}$ of metric cylinders on $q$, then the limit $\theta$ of the $\theta_{\gamma_n}$ is a nonergodic direction for $q$. Conditions ii), iv), v), and vi) are used to control the building of the Cantor set which will consist of all such limiting $\theta$ so that Theorem 3.1 can be applied to
guarantee positive Hausdorff dimension. In particular ii) will tell us that a "level" $n$ interval defined by $\alpha$ contains a specified number of level $n + 1$ subintervals defined by $\gamma \in \Gamma_q(\alpha)$. Condition v) gives the size of the interval determined by each $\alpha$ in terms of the length of $\alpha$. Condition iv) controls the length of the level $n + 1$ curves $\gamma$ in terms of the length of the level $n$ curve $\alpha$. Together with iv) this controls the size of the intervals. Condition vi) controls the size of the gap between intervals. Finally condition i) says that the construction of subintervals of each interval is successful if we remove a set of flat structures with small measure. We will use this last estimate to show that we can make the infinite construction of a Cantor set by removing countably many sets of flat structures whose union still has small measure. The construction will then be possible on the set of positive measure that remains.

Before proving Theorem 11.2 we use it to prove the Main Theorem.

Proof of Main Theorem. Fix a component $C$ of $\mathcal{M}^1$. The main part of the proof will consist of finding a positive number $\delta$ and a set $W_\infty \subset \mathcal{M}^1$ of positive measure such that for each $q \in W_\infty$ there is a Cantor set $\Lambda_q \subset [0, \pi]$ of Hausdorff dimension greater than $\delta$ so that for each $\theta \in \Lambda_q$ the vertical foliation of $r_\theta q$ is nonergodic.

For a set of flat structures $q$ we will construct the Cantor set $\Lambda_q$ inductively as a decreasing intersection of sets, each of which is a union of "level" $n$ intervals. Each level $n$ interval will be defined by the waist curve $\alpha$ of a metric cylinder of $q$. To find the level $n + 1$ intervals, we will apply Theorem 11.2 which guarantees the existence of metric cylinders disjoint from the given one, if we throw out a set of flat structures of small measure. Thus the infinite construction of the Cantor set $\Lambda_q$ will be possible after removing countably many sets of flat structures. We will show that a set of positive measure remains.

To begin, we apply the theorem of [M2] which asserts that any flat structure in the stratum contains a metric cylinder. Let $q_0$ be any flat structure of area 1 and $\beta$ the homotopy class of the cylinder $C$. By either sewing a cylinder into $C$ or possibly removing a subcylinder of $C$ and then rescaling the metric, we can assume the cylinder has area 5/12 and the flat structure still has area 1. By applying an element of $\text{SL}(2,\mathbb{R})$ to $q_0$, we can assume the cylinder has circumference 800. On taking a compact neighborhood of this flat structure, we can assume we have a compact set $W_0$ of positive measure $\delta_0$ such that each $q \in W_0$ has a metric cylinder homotopic to $\beta$ with area between $\frac{1}{2}$ and $\frac{1}{3}$ and length at least 400.

By Lemma 11.1 and the fact that $W_0$ is compact, there is a $c_1$ such that for any $q \in W_0$, the number of geodesics of length at most $T$ is bounded above by $c_1 T^\gamma$. 
Fix $c > 1$ to be any constant. Let
\[
c_2 = \sum_{j=1}^{\infty} \frac{c^{\tau(j+1)-2^j}}{400^{2^j}}.
\]
For $T_0$ sufficiently large, and $T \geq (400/T_0)T_0^{p_1^{-1}}$, by (i) of Theorem 11.2 there is a constant $c'$ such that for any $\alpha$,
\[
(2) \quad \mu^1(\mathcal{M}(\alpha, T) - F(\alpha)) \leq c' \left( \frac{1}{T} \right)^{2^\tau}.
\]
Since $p_1 > 1$ and $\tau - 2^\tau < 0$, the series
\[
c_1 c_2 c' \sum_{n=1}^{\infty} T_0^{2^\tau n p_1^{-1}} T_0^{(\tau-2^\tau)p_1^{-1}}
\]
converges to a function $f(T_0)$ and the function $f(T_0)$ converges to zero as $T_0 \to \infty$. Choose $T_0$ greater than or equal to 400 and large enough so that
* Theorem 11.2 holds,
** $|\beta|_q \leq T_0$ for all $q \in W_0$,
*** $f(T_0) < \delta_0 / 2$.

We now describe the inductive construction of a Cantor set based on Theorem 11.2. Set $S_0 = \{\beta\}$, $W_0(\beta) = W_0$, and for each $q \in W_0$ define $\Delta_1 = \Delta_1(\beta, q)$ to be the (level 1) interval of width $1/T_0^k$ about $\theta_\beta$. Assume inductively that for each $1 \leq j \leq n$, we have:
1) Sets $S_{j-1}$ of simple closed curves,
2) For each $\alpha \in S_{j-1}$ a set $W_{j-1}(\alpha)$ such that each $q \in W_{j-1}(\alpha)$ has a metric cylinder $A_\alpha$ in the homotopy class of $\alpha$,
3) $\frac{400}{T_0} T_0^{p_1^{-1}} \leq \min |\alpha|_q \leq \max |\alpha|_q = T(\alpha) \leq T_0^{p_1^{-1}}$

where the minimum and maximum are taken over $q \in W_{j-1}(\alpha)$.
4) For each $q \in W_{j-1}(\alpha)$, there is a level $j$ interval $\Delta_j = \Delta_j(\alpha, q)$ of width $1/T(\alpha)^k$ centered at $\theta_\alpha$ with the following property. For $j > 1$, $\Delta_j$ is contained in some level $j-1$ interval $\Delta_{j-1} = \Delta_{j-1}(\omega, q)$ where $\omega \in S_{j-2}$ and $q \in W_{j-2}(\omega)$. Moreover any level $j-1$ interval and its level $j$ subintervals satisfy the hypotheses of Theorem 3.1.

It is easy to see that 1)–4) hold when $j = 1$. We now make the required constructions for $n + 1$, of sets of simple closed curves, sets of flat structures associated to those simple closed curves, and level $n + 1$ intervals. For each $\alpha \in S_{n-1}$ and $q \in W_{n-1}(\alpha)$, let $\Delta = \Delta_n(\alpha, q)$ be the level $n$ interval centered at $\theta_\alpha$. Let $F(\alpha)$ be the subset of $\mathcal{M}(\alpha, T(\alpha))$ produced by Theorem 11.2. By
Theorem 11.2,

a) For all \( q \in F(\alpha) \cap W_{n-1}(\alpha) \), \( \Delta \) contains at least \( T(\alpha)^{s} \) angles \( \theta_{\gamma} \) defined by metric cylinders \( A_{\gamma} \) on \( q \), where \( A_{\alpha} \cap A_{\gamma} = \emptyset \).

b) \( T(\alpha)^{p_1} \leq c_1 T(\alpha)^{p'} \leq |\gamma|_{q} \leq c_2 T(\alpha)^{p''} \leq T(\alpha)^{p_2} \) for some \( p' \), where \( p_1 < p' < p_2 \).

c) The gap between successive angles \( \theta_{\gamma} \) divided by \((c_2 T(\alpha)^{-k})\) is at least 4.

Call such a \( \gamma \) as that produced by Theorem 11.2 an offspring of \( \alpha \) and \( q \). Let \( S \) be the set consisting of all \( \gamma \) that are the offspring of \( \alpha \in S_{n-1} \) and some \( q \in W_{n-1}(\alpha) \). Let \( W_{n}(\gamma) \) consist of those \( q \) for which \( \gamma \) is an offspring of some \( \alpha \in S_{n-1} \) and \( q \in W_{n-1}(\alpha) \). Then 1) and 2) in the induction step are satisfied for \( n + 1 \). Define

\[
T(\gamma) = \max |\gamma|_{q} \leq c_2 T(\alpha)^{p''},
\]

the maximum over \( q \in W_{n}(\gamma) \). Now b) and the induction hypothesis that 3) holds for \( j = n \) imply 3) holds with \( \gamma \) in place of \( \alpha \) and \( n + 1 \) in place of \( n \).

For each \( q \in W_{n}(\gamma) \), take an interval \( \Delta_{n+1}(\gamma, q) \) of width \( T(\gamma)^{-k} \) about each \( \theta_{\gamma} \). These are the level \( n + 1 \) subintervals of \( \Delta \). Now suppose \( \gamma \neq \gamma' \), both curves are in \( S \) and \( q \in W_{n}(\gamma) \cap W_{n}(\gamma') \). By c) above, the gap between \( \theta_{\gamma} \) and \( \theta_{\gamma'} \) is at least \((4c_2 T(\alpha)^{-k})\) which by (4) and the definition of \( \Delta_{n+1}(\gamma, q) \) is at least four times as large as any level \( n + 1 \) interval. Thus the gap between the level \( n + 1 \) intervals is at least as large as the size of any level \( n + 1 \) interval. Set \( k_1 = s/k \) and \( k_2 = p_2 \). Then a) and b) say that the interval \( \Delta = \Delta_{n}(\alpha, q) \) with size

\[
|\Delta| = \frac{1}{T(\alpha)^{k}}
\]

has

\[
|\Delta|^{-k_1}
\]

level \( n + 1 \) subintervals \( \Delta_{n+1}(\gamma, q) \) each of size at least

\[
\left( \frac{1}{T(\alpha)^{p_2}} \right)^{k} = |\Delta|^{k_2}.
\]

Moreover as we have seen, the gap between the intervals is larger than the intervals. The hypotheses of Theorem 3.1 are satisfied and so 4) holds. The inductive step is complete.

For each \( q \in W_{n-1}(\alpha) \cap F(\alpha) \), we have shown how to construct level \( n + 1 \) subintervals of the level \( n \) interval \( \Delta_{n}(\alpha, q) \). The construction of level \( n + 1 \) subintervals of \( \Delta_{n}(\alpha, q) \) fails on the complement of \( F(\alpha) \) in \( W_{n-1}(\alpha) \). By
(2), an upper bound for the measure of this set is given by the estimate

\begin{equation}
\mu^1(W_{n-1}(\alpha) - W_{n-1}(\alpha) \cap F(\alpha)) \leq c'/T(\alpha)^{2^r}.
\end{equation}

Let

\[ U_n = \bigcup_{\alpha \in S_{n-1}} (W_{n-1}(\alpha) - W_{n-1}(\alpha) \cap F(\alpha)). \]

This is the set of all flat structures for which the construction of level \( n \) intervals \( \Delta_n(\alpha, q) \) is possible, but for which the construction of level \( n + 1 \) intervals from the level \( n \) intervals fails for some level \( n \) interval. Thus on the complement of \( U_n \), the construction succeeds for all level \( n \) intervals. To find the measure of \( U_n \), we have to sum (5) over \( \alpha \in S_{n-1} \). Divide the interval

\[ \left[ \frac{400}{T_0} T_0^{p_1^{n-1}}, \frac{400}{T_0} T_0^{p_2^{n-1}} \right] \]

of possible values of \( T(\alpha) \) given by (3) into a union of intervals of the form

\[ I_j = \left[ c^j \left( \frac{400}{T_0} T_0^{p_1^{n-1}}, c^{j+1} \left( \frac{400}{T_0} T_0^{p_1^{n-1}} \right) \right. \right]. \]

By the choice of \( c_1 \), for each \( j \), there are at most

\begin{equation}
c_1 c^{\tau(j+1)} T_0^{\tau p_1^{n-1}}
\end{equation}

\( \alpha \)'s with \( T(\alpha) \in I_j \). Then (5), (6), and the definition of \( c_2 \) imply

\begin{equation}
\mu^1(U_n) \leq \sum_{\alpha} \frac{c'}{T(\alpha)^{2^r}} \leq \sum_j c_1 c' \frac{c^{\tau(j+1)} T_0^{\tau p_1^{n-1}}}{\left( \frac{400}{T_0} \right)^{2^r} T_0^{2^r p_1^{n-1}}} = c_1 c' c_2 T_0^{(\tau - 2^r) p_1^{n-1} + 2^r}. \end{equation}

Now we let

\[ W_\infty = W_0 - \bigcup_{n=0}^\infty U_n. \]

By removal of the union of the \( U_n \) from \( W_0 \), the sets of flat structures for which the construction of level \( n + 1 \) subintervals from level \( n \) intervals fails for some level \( n \) interval, we have guaranteed that for all \( q \in W_\infty \), the infinite construction of a Cantor set \( \Lambda_q \) is possible. We compute \( \mu^1(W_\infty) \).
Now we know by *** that
\[ f(T_0) = c_1 c' c_2 \sum_{n=1}^{\infty} T_0^{(\gamma - 2)^{(n-1)} + 2} \leq \delta_0/2. \]
This together with (7) implies
\[ \sum_{n=1}^{\infty} \mu^1(U_n) \leq f(T_0) \leq \delta_0/2. \]
We have
\[ \mu^1(W_\infty) \geq \mu^1(W_0) - \sum \mu^1(U_n) \geq \delta_0 - \delta_0/2 = \delta_0/2. \]
Thus \( W_\infty \) is a set of measure greater than \( \delta_0/2 \) such that for each \( q \in W_\infty \) we can form the Cantor set \( \Lambda = \Lambda_q \). Each \( \Lambda \) has positive Hausdorff dimension since all hypotheses of Theorem 3.1 are satisfied.

Next we check the criteria i)--iii) of Theorem 2.1 for \( q \in W_\infty \) and \( \theta \in \Lambda_q \) for \( r_\theta q \) to be nonergodic. In the construction of \( \Lambda \), consider level \( n \) and \( n + 1 \) sets \( \Delta_n \supset \Delta_{n+1} \) defined by disjoint metric cylinders \( A_n = A_\alpha \) and \( A_{n+1} = A_\gamma \). Thus successive partitions of the surface are defined by the boundaries of \( A_n \) and \( A_{n+1} \) and the components are \( A_n \) and \( A_{n+1} \). Now Theorem 11.2 iii) gives
\[ \text{area}(M - (A_\alpha \cup A_\gamma)) < 1/10n^2 \]
so that \( \text{area}(A_{n+1}^c - A_n) \leq 1/10n^2 \). Thus
\[ \sum \text{area}(A_{n+1}^c - A_n) < \infty \]
and iii) of Theorem 2.1 is satisfied.

Furthermore
\[ \text{area}(A_n) = 1 - \text{area}(A_{n-1}) \leq \text{area}(A_{n-2}) + \frac{1}{10(n-2)^2}. \]
Thus if \( n = 2m + 1 \),
\[ \text{area}(A_n) \leq \text{area}(A_1) + \frac{1}{10} \sum_{j=1}^{m} \frac{1}{(2j-1)^2} \leq \frac{1}{2} + \frac{1}{10} \frac{\pi^2}{6} < 1. \]
If \( n \) is even, then \( \text{area}(A_n) \leq \text{area}(A_2) + 1/10 \pi^2/6 \). But \( \text{area}(A_2) \leq 1 - \text{area}(A_1) \leq 2/3 \) so that \( \text{area}(A_n) \leq 2/3 + 1/10 \pi^2/6 < 1 \) in this case as well, proving ii) of Theorem 2.1.

Finally, we check criterion i) of Theorem 2.1. Suppose \( |\gamma| = T \). Then \( |\theta - \theta_\gamma| \leq 1/T^k \). For a geodesic segment let \( v_\theta(\alpha) \) denote the holonomy vector of \( \alpha \) in the direction \( \theta \) and \( h_\theta(\alpha) \) denote the component in the perpendicular direction. Then \( h_\theta(\gamma) = T |\sin(\theta - \theta_\gamma)| \leq O(T/T^k) \to 0 \) as \( T \to \infty \) since \( k > 1 \).
We now finish the proof of the Main Theorem. Let $HD(q)$ be the Hausdorff dimension of $NE(q)$. The function $HD$ is measurable as a function of $q$. The sets $NE(q)$ and $NE(g_i,q)$ are related by a linear fractional transformation of $S^1$. A linear fractional transformation is a diffeomorphism and hence preserves the Hausdorff dimension of the set of nonergodic directions. Thus $HD$ is invariant under the Teichmüller flow $g_i$. The flow $g_i$ is ergodic on each component of the stratum by [V2]. Ergodicity implies that $HD$ is constant almost everywhere on the component $C$. For $q \in W_n$ the set $NE(q)$ contains $\Lambda_q$. In particular, for $q \in W_n$ the function $HD(q)$ is positive. Since $W_n$ has positive measure, $HD$ must be equal to a positive constant almost everywhere.

**Proof of Theorem 11.2.** We are given the initial stratum $M^1 \subset M^1(\alpha, T)$ and the set $M^1(\alpha, T) \subset M^1$ of flat structures with area 1 and the subset $M^1(\alpha, T) \subset M^1$ of the same dimension. $\Gamma_{i+1}$ will be called a successor of $\Gamma_i$. The $\Gamma_i$ satisfy the following properties.

i) If the genus of surfaces in $M^1(\alpha, T) \subset M^1$ is at least 1, $\Gamma_N = M^1(\alpha, T) \subset M^1$ is the exceptional stratum of flat structures on a torus. If the genus is 0, $\Gamma_N = M^1(\alpha, T) \subset M^1$ is the exceptional stratum of spheres.

ii) For $2 \leq i \leq N$, each stratum in the union $\Gamma_i$ is a face of some $M^1(\alpha, T) \subset M^1(\alpha, T)$.

iii) $\Gamma_1$ is the union of the strata of flat structures obtained from collapsing shrinkable cylinders in the class of $\alpha$ and then collapsing the resulting simple closed curve or segment.

Now we know $\dim \Gamma_{i+1} = \dim \Gamma_i - 2$ except when $i = 0$, in which case $\dim \Gamma_1 = \dim \Gamma_0 - 4$. This implies that the length $N + 1$ of the sequence depends only on the dimension of the stratum $\Gamma_0$. In fact, since the dimension of either exceptional stratum is 4,

$$N = \frac{1}{2} (\dim \Gamma_0 - 6).$$

We first consider the case where the flat structures in $\Gamma_0$ have genus greater than or equal to 1 so that the exceptional stratum that ends the sequence is the stratum of flat tori.

The first step in the proof will be to define inductively subsets $E_i$ of $\Gamma_i$. We will start with a subset $E_N$ of small measure in the quotient flat torus stratum $\Gamma_N$. $E_N$ will consist of those flat tori which have a very short closed geodesic. This will then determine $E_{N-1} \subset \Gamma_{N-1}$. $E_{N-1}$ will consist of those flat structures for which some aspect of the shrinking process fails: either every shrinkable curve is very long, or there is a shrinkable curve which is not too long, but
after shrinking it, the retraction map is not defined on the deformed flat structure, or the retraction map is defined, but the image of the flat structure under the retraction map lies in $E_N$. Using the results of Section 10, we will show the measure of $E_{N-1}$ is small as well. We similarly use $E_{N-1}$ to define $E_{N-2}$ and so forth. In this way we determine a subset $E(\alpha)$ of $\mathcal{M}(\alpha, T)$ which will have small measure. Our desired subset $F(\alpha)$ will be the complement of $E(\alpha)$. The estimate i) in the conclusion of the theorem will come from the definition of $E(\alpha)$ and the results of Section 10.

The second step will be to show that each $q \in F(\alpha)$ has metric cylinders $A_\gamma$ that satisfy conclusions ii)–vi) of the theorem. This will involve estimating various lengths and distortions, and applying the results of Sections 6, 8, and 9.

Before defining the $E_i$ we will choose the constants $s, p_1, p_2, k$ appearing in the statement of the theorem. The choice of $c_1$ and $c_2$ will appear in the body of the proof. First recall the constant $p$ appearing in Theorems 10.3 and 10.5, and the constant $\tau$ defined by (1). Let

$$s = p2^{\tau N}.$$

We will now choose a sequence of numbers with certain properties and define $p_1, p_2$ and $k$ in terms of these numbers. Choose a sequence $l_1 > l_2 > \cdots > l_{N-1} > l_N$ such that conditions (9), and (10), and (13) hold. In the paragraph after formulae (18) we will give an explanation for requiring these inequalities as well as the significance of the other constants chosen. Choose $l_j$ so that

$$l_j \geq (10p)2^{\tau j},$$

$$l_j/2 \geq 3l_{j+1}/2 + 2p2^{\tau j}.$$  

Set

$$m_1 = l_1 + 1 - \sum_{i=2}^{N} (2p\tau(i - 1) + l_i)$$

and

$$m_2 = l_1 + 1 + \sum_{i=2}^{N} (2p\tau(i - 1) + l_i).$$

Then we also require

$$2 + \frac{m_2}{1 + s + m_1} \leq m_1 + l_N.$$

Notice this last condition can be achieved by making $l_1$ large compared to the
\( l_i, \ i \geq 2. \) Let
\[
(14) \quad k = m_1 + l_N,
\]
\[
(15) \quad p_1 = s + m_1,
\]
and
\[
(16) \quad p_2 = s + 2 + m_2.
\]
We now suppose \( T_0 \) and \( T \geq (400/T_0)T_0^{n^{n-1}} \) are given. We next define other quantities in terms of \( T \) and the previous constants. The sets \( E_i \) in turn will be defined in terms of these latter quantities. Let
\[
(17) \quad \delta_0 = \frac{1}{T}
\]
and for each \( 1 \leq j \leq N \) define \( \delta_j \) and \( \kappa_j \) by
\[
(18) \quad \delta_j = \frac{1}{T^{2^{j+1}}} , \quad \kappa_j = \frac{1}{T^2}.
\]
The point of requiring inequality (9) is that we can and will choose \( T_0 \) sufficiently large, so that the inequality \( 10p \geq 18 \), and (9) imply that for each \( j \), \( \kappa_j \) and \( \delta_j \) are a pair of numbers that satisfy the inequalities * and ** in Section 10. On the union of strata \( \mathcal{Y}_j \) we will apply the Teichmüller flow \( g_t \), where \( t \) is defined below by (20), shrinking a segment so that its length is at most \( \kappa_j \) while requiring all nonhomologous segments to have length at least \( \delta_j \). Because * and ** are satisfied, we will be able to apply Theorem 10.3 which makes estimates of measures. There are two reasons for requiring (10); \( l_j \) is large compared to \( l_{j+1} \), which means \( \kappa_j \) is small compared to \( \kappa_{j+1} \). Moreover the time (20) that it takes to shrink segments in the strata \( \mathcal{Y}_{j+1} \) to length \( \kappa_{j+1} \) is also small compared to \( \kappa_j \). This will imply that once we shrink a segment to length at most \( \kappa_j \), after a further deformation of the flat structure that shrinks another curve to length at most \( \kappa_{j+1} \), the first curve will remain shorter than \( \kappa_{j+1} \). In other words, once a segment is made short, it remains short (although not as short) under further deformations. The second point of (10) is the following. When we deform a flat structure in \( \mathcal{Y}_j \) so that a segment has length at most \( \kappa_j \), the flat structure is geometrically close to a flat structure on the face strata \( \mathcal{Y}_{j+1} \). The closeness is measured by the distortion of the collapsing map. The action of \( SL(2, \mathbb{R}) \) changes distortions. Inequality (10) will imply that the flat structures in \( \mathcal{Y}_j \) and \( \mathcal{Y}_{j+1} \) are still close after further deformations. In particular this will imply that when we shrink a segment on the surface in \( \mathcal{Y}_{j+1} \), there is a corresponding segment which is short on the surface in \( \mathcal{Y}_j \). Combined with the previous explanation for (10), this will mean there are two short segments on the flat surfaces in \( \mathcal{Y}_j \). Inductively we will be able to shorten \( N \) segments in the initial stratum.
The composition of the various elements of $\text{SL}(2, \mathbb{R})$ used to shrink segments will give an element of $\text{SL}(2, \mathbb{R})$ which will depend on the flat structure in the initial stratum. The numbers $m_1$ and $m_2$ are defined so as to give uniform lower and upper bounds for the amount of the diagonal (Teichmüller flow) action (inequality (25)). After the shrinking process is finished, the flat structure in the initial stratum will be close to a flat torus. We pick closed orbits on the flat torus of length about $T^\circ$. Flowing back will give metric cylinders $A_\gamma$ on the beginning flat structure whose length will be bounded in terms of $s$, $m_1$, $m_2$, and therefore $p_1$ and $p_2$. On our initial flat structure, we will be able to compute an upper bound for the angle the waist curve of each $A_\gamma$ makes with the waist curve of the cylinder $A_a$. The number $k$ is defined in terms of this upper bound. Finally inequality (13) is used to show that the angle between the waist curves of different $A_\gamma$ is larger than the interval of angles they define at the next level of the construction. This is needed to guarantee that the gaps between intervals are larger than the size of the intervals. To finish the preparation for the proof we define the set of times that it will take to contract segments to lengths $\kappa_i$. Define $t_1$ by

$$e^{t_1/2} = \frac{1}{\delta_0 \kappa_1}.$$  

and for $2 \leq i \leq N$ define $t_i$ by

$$e^{t_i/2} = \frac{1}{\delta_{i-1} \kappa_i}.$$  

Finally for $1 \leq i \leq N$, define $L_i$ by

$$L_i = \frac{1}{\delta_i^{\nu}}.$$  

Now we are ready to define the sets $E_i$ promised in the introduction. For $q \in \mathcal{MD}(L)$, recall from Section 10 the definition of $\theta = \theta(q)$ as the endpoint of an interval $I_j$ of angles. Now let $E_N$ be the set of $q \in \mathcal{MD}_N = \mathcal{V}_N$ such that the shortest closed curve on the flat surface has length at most $\delta_N^\nu$. By Theorem 10.4, for some constant $c$,

$$\mu(E_N) < \frac{c}{\delta_N^{2\nu}}.$$
Now suppose inductively we have chosen, for \( i + 1 \leq j \leq N \), a set \( E_j \subset Y_j \) such that:

a) For \( j < N \), and \( q \in E_j \), either \( q \in Y_j - \bigcup M_i(L_j) \), or if \( q \in \bigcup M_i(L_j) \), then \( g_{t_j}r_\varrho q \in V_1(\kappa_{j+1}, \delta_{j+1}) \cup V_2(E_{j+1}, \kappa_{j+1}, \delta_{j+1}) \).

b) \( \mu(E_j) \leq c\delta_j^{2p} \) for a universal constant \( c \).

The definition a) is chosen so that as in Proposition 10.2, for \( q \in E_j \), some step in the shrinking procedure fails. We wish to define \( E_i \) and satisfy a) and b). Because of a) the definition is obvious.

**Definition.** For \( i \geq 1 \), let \( E_i \) be the set of \( q \in V_i \) such that \( q \notin \bigcup M_i(L_i) \) or \( q \in \bigcup M_i(L_i) \) and \( g_{r_i}r_\varrho q \in V_i(\kappa_{i+1}, \delta_{i+1}) \cup V_2(E_{i+1}, \kappa_{i+1}, \delta_{i+1}) \).

**Definition.** Let \( E_0 = E(\alpha) \) be the set of \( q \in M_0(\alpha, T) \) such that

\[
g_{r_0}r_\varrho q \in V_1(\alpha, \kappa_1, \delta_1) \cup V_2(E_1, \alpha, \kappa_1, \delta_1).
\]

By construction, a) is satisfied. We check b). Since \( \kappa_j \) and \( \delta_j \) satisfy * and ** of Section 10 we can apply Theorem 10.3 with \( \epsilon' = \delta_i^{2p} \), \( L = 1/\delta_i^p \), \( \kappa' = \kappa_{i+1} \), and \( \delta' = \delta_{i+1} \). We have for \( i \geq 1 \),

\[
\mu(E_i) < \frac{c\delta_{i+1}^{2p}}{\delta^{2p}_i \delta_{i+1}^{2p}} + \frac{c\delta_{i+1}^2}{\delta_i^{2p}} + c\delta_i^{2p}.
\]

The definition of \( \tau \) in terms of \( p \) in (1), the definition (18) of \( \delta_i \), and a simple computation show that each term in the sum is less than or equal to \( c\delta_i^{2p} \). Thus

\[
\mu(E_i) < c\delta_i^{2p}.
\]

Similarly, Theorem 10.5 applied with \( \epsilon' = \delta_1^{2p}, L = 1/T, \kappa' = \kappa_1 \) and \( \delta' = \delta_1 \) yields the inequality

\[
\mu^1(E(\alpha)) \leq c \frac{1}{T^{2^r}}.
\]

Thus the induction step is complete. We have constructed a set \( E(\alpha) \). Let

\[
F(\alpha) = M_0(\alpha, T) - E(\alpha).
\]

Then \( F(\alpha) \) satisfies i) in the conclusion of the theorem. This completes the first step.

The second step is to check the other conclusions. Suppose \( q \in F(\alpha) \) so that \( q \notin E(\alpha) \). Then by Proposition 10.2, the length of \( \alpha \) on the flat surface of \( g_{r_0}r_\varrho(q) \) is at most \( \kappa_1 \). Moreover, there is a collapsing map \( h_1 \) which maps the flat surface defined by \( g_{r_0}r_\varrho(q)q \) to the flat surface of some \( q_1 \notin E_1 \). It collapses
\(\alpha = \alpha_1\) to a point or pair of points and has distortion at most \(c\kappa_1^{1/2}\) on the complement of \(\alpha_1\).

Now since \(q_1 \not\in E_1\), \(q_1\) has a shrinkable segment \(\alpha_2\) of length less than or equal to

\[
L_1 = \frac{1}{\delta_l^p}
\]

so that \(q_1 \in \cup_s D_1(L_s)\). In order to shrink \(\alpha_2\), we will need to apply \(g_{t_2}r_{\theta(q_1)}\) to \(q_1\). We want to keep track of the flat structure \(q\) in the initial stratum \(Y_0\) and the lengths of segments of \(q\) under these successive deformations. Therefore we apply \(g_{t_2}r_{\theta(q_1)}\) to \(g_{t_1}r_{\theta(q)}q\) as well as to \(q_1\). We want to check the effect of rotating and flowing time \(t_2\) on the length of \(\alpha_1\) and on the distortion of the collapsing map \(h_1\). Since flowing by \(t_2\) increases lengths by a factor of at most \(e^{t_2/2}\), the length of \(\alpha_1\) with respect to the flat structure of \(g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q\) is at most

\[e^{t_2/2}\kappa_1\]

From this, the definitions (20) and (18), and then the inequality (10), we find that the \(g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q\) length of \(\alpha_1\) is at most

\[
T^{2^{2^l}+l_2-l_1} \leq \frac{1}{T^{l_2}} = \kappa_2.
\]

Next we estimate the distortion of the collapsing map \(h_1\) defined on the flat structure \(g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q\). Its distortion has changed from its distortion on the flat structure \(g_{t_1}r_{\theta(q)}q\), since we have applied the element \(g_{t_2}r_{\theta(q_1)}\) of \(SL(2,\mathbb{R})\) to the domain \(g_{t_1}r_{\theta(q)}q\) and image \(q_1\). The collapsing map \(h_1\) is a simplicial map on each geodesic triangle. Rotating does not increase distortions. After flowing time \(t_2\), we still have a collapsing map \(h_1\) which takes geodesic triangles to geodesic triangles and the distortion on each triangle is increased by a factor of at most \(e^{t_2}\). Since the original map \(h_1\) has distortion at most \(c\kappa_1^{1/2}\), the new map \(h_1\) has distortion at most

\[ce^{t_2}\kappa_1^{1/2}\]

on the complement of \(\alpha_1\) which by (20) and (18) is equal to

\[
\frac{T^{(2p)^{2^l}T^{2l_2}}}{c T^{l_2/2}}.
\]

By (10) again, this quantity is at most

\[
\frac{c}{T^{l_2/2}} = c\kappa_2^{1/2}.
\]
Now we shrink $\alpha_2$ by applying $g_{t_2}r_{\theta(q_1)}$ to $q_1$. Since $q_1 \notin E_1$, $\alpha_2$ has length at most $\kappa_2$ on the flat surface of $g_{t_2}r_{\theta(q_1)}q_1$. By the distortion estimate (22) on $h_1$, the length of the segment $\alpha_2$ with respect to $g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q$ is at most $2\kappa_2$. Thus we have concluded that $\alpha_1$ has length at most $\kappa_2$ and $\alpha_2$ length at most $2\kappa_2$ on the flat surface of $g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q$. Again since $q_1 \notin E_1$,

$$g_{t_2}r_{\theta(q_1)}q_1 \notin V_1(\kappa_2, \delta_2) \cup V_2(E_2, \kappa_2, \delta_2).$$

Therefore there is a map $h_2$, defined on the flat surface of $g_{t_2}r_{\theta(q_1)}q_1$, to the flat surface of some $q_2 \notin E_2$ which collapses $\alpha_2$. The distortion of $h_2$ on the complement of $\alpha_2$ is at most

$$c\kappa_2^{1/2}.$$ 

The composition $h_2 \circ h_1$ gives a map from the flat surface of $g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q$ to $q_2$ which collapses both $\alpha_1$ and $\alpha_2$. The distortion of $h_2 \circ h_1$ on the complement of $\alpha_1$ and $\alpha_2$ is at most a constant times the sum of the distortion of $h_1$ and the distortion of $h_2$. By (22) the distortion of $h_2 \circ h_1$ is at most

$$c\kappa_2^{1/2}$$

for a new constant $c$.

Since $q_2 \notin E_2$, $q_2$ has a shrinkable segment $\alpha_3$ of length at most $L_2$. We can apply $g_{t_3}r_{\theta(q_2)}$ to both $q_2$ and $g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q$.

Repeating the above arguments and estimates using the definitions (18), (20), and the inequality (10), we find that the $g_{t_3}r_{\theta(q_2)}g_{t_2}r_{\theta(q_1)}g_{t_1}r_{\theta(q)}q$ length of $\alpha_1$, $\alpha_2$, and $\alpha_3$ are all at most $2\kappa_3$, and there is a map $h_3 \circ h_2 \circ h_1$ from this flat surface to some $q_3 \notin E_3$ which collapses all three segments. The distortion of $h_3 \circ h_2 \circ h_1$ is at most $c\kappa_3^{1/2}$.

Continuing in this manner, we find

1) geodesic segments $\alpha_1, \ldots, \alpha_N$,
2) flat structures $q_1, \ldots, q_N$, where $q_i \notin \mathcal{Y}_i - E_i$,
3) angles $\theta(q), \theta(q_1), \ldots, \theta(q_{N-1})$, and
4) a map $h_0 = h_N \circ h_{N-1} \circ \cdots \circ h_1$ from the flat surface

$$M_N = q_1' = g_{t_N}r_{\theta(q_{N-1})}g_{t_{N-1}} \cdots g_{t_1}r_{\theta(q)}q$$

onto the flat torus $M_N$ of $q_N$.

The map $h_0$ collapses the union of the $\alpha_i$ to the distinguished point of $M_N$. The definitions (18) and (20) of $\kappa_i$, $\delta_i$, and $e^t_i$, and inequality (10) show the map $h_0$ is a

$$c\kappa_N^{1/2}$$

distortion on the complement of $\alpha_1, \ldots, \alpha_N$ and the length of each $\alpha_i$ with
respect to the flat structure of \( q_N \) is at most

\[
2\kappa_N = \frac{2}{T^{l_N}}.
\]

Moreover \( M'_N - A_{a_1} - \bigcup_{j=2}^N \alpha_j \) is homeomorphic to a torus minus a disc and the boundary of the disc has length at most \( 2\kappa_N \).

Since \( q_N \notin E_N \), \( M_N \) does not have a closed curve shorter than \( \delta_N^p = 1/T^{p2r_N} \). If \( T_0 \), hence \( T \), is large enough, then the definition of \( \kappa_N \) in (18) shows that the upper bound \( c\kappa_N^{1/2} \) for the distortion of \( h_0 \) satisfies

\[
c\kappa_N^{1/2} \leq 1/2.
\]

This together with the fact that \( M_N \) does not have a simple closed curve shorter than \( \delta_N^p \) implies that \( M'_N - A_{a_1} - \bigcup_{j=2}^N \alpha_j \) does not have a simple closed curve shorter than

\[
\frac{\delta_N^p}{2} = \frac{1}{2T^{p2r_N}},
\]

not homotopic to the boundary. However by (9), the upper bound \( 2\kappa_N = 2/T^{l_N} \) for the boundary length is smaller than \( 1/2T^{p2r_N} \), the lower bound for the length of the shortest closed curve not homotopic to the boundary. We can apply Theorem 8.1. There is \( \Omega \subset M'_N - A_{a_1} - \bigcup_{i=2}^N \alpha_i \) such that

a) area\( (M'_N - A_{a_1} - \Omega) < 4\kappa_N^2/\pi \leq 1/20n^2 \);

b) Length boundary \( \Omega \leq \kappa_N \);

c) There are a flat torus \( M \) and convex disc \( D \subset M \) so that \( \Omega \) is isometric to \( M - D \).

We now use the closed orbits on the flat torus \( M \) to find metric cylinders on \( M' \) disjoint from \( A_{\alpha} \). To begin, we notice from the definition (8) of \( s \) and (9) that

\[
p2r_N + 2 < s + 3 < l_N.
\]

By the isometry of \( \Omega \) and \( M - D \), the shortest simple closed curve on \( M \) has length at least

\[
\frac{1}{2T^{p2r_N}} \geq \frac{1}{T^{s+1}}.
\]

Now

\[
T^{s+1} \geq 128\pi T^s = \frac{64\pi}{1/2T^s}.
\]

Since \( r = T^{s+1} \) and \( \delta = 1/2T^s \), the hypotheses of Theorem 9.1 hold. By
Theorem 9.1 there are at least

\[ \pi T^{s+1}/100 \]

homotopy classes \( \gamma \) of closed orbits on the flat torus \( M \) with lengths between \( T^{s+1}/2 \) and \( T^{s+1} \). Furthermore we can choose the \( \gamma \) so that their vertical angles lie in a sector of size at least \( \pi/8 \) which is bounded away from the vertical direction of \( r_{\pi/2}q_N \). Now

\[ \frac{\pi T^{s+1}}{100} \geq T^s \]

since \( T \geq (400/T_0)T_0^{p_1^{-1}} \). Let \( \Gamma_q(\alpha) \) be the set of all such \( \gamma \). Fix such a homotopy class of orbits \( \gamma \) of length at most \( T^{s+1} \). Since length \( D \leq \kappa_N = 1/T^{2N} \),

\[ \text{area}\{x: \text{orbit through } x \text{ intersects } D\} \leq \frac{T^{s+1}}{T^{2N}} \leq \frac{1}{T^2} \leq \frac{T_0^2}{(400)^2 T_0^2 p_1^{-1}} \leq \frac{1}{20n^2}, \]

the last inequality since \( p_1 \geq 2 \). The complement of this orbit set is a metric cylinder on \( M \) which by the isometry of \( \Omega \) and \( M - D \) gives a metric cylinder \( A_\gamma \) on \( M_N' \), disjoint from \( A_\alpha \). Since the flat structure \( M_N' \) of \( q_N' \) is the result of applying an element of \( \text{SL}(2, \mathbb{R}) \) to the original flat structure \( q \), we find we have at least \( T^s \) metric cylinders \( A_\gamma \) on \( q \) disjoint from \( A_\alpha \). This gives conclusion ii). The area of the complement of \( A_\gamma \) in \( \Omega \) is at most \( 1/(20n^2) \). This together with a) above implies that

\[ \text{area}\{M_N' - (A_\gamma \cup A_\alpha)\} \leq \frac{1}{20n^2} + \frac{1}{20n^2} = \frac{1}{10n^2}. \]

Therefore conclusion (iii) of the theorem holds.

We need to check iv), v), and vi). These will come from an analysis of the time needed for the deformations. For some \( t, \theta \), we have \( q_N' = g_\theta r_\theta q \), where

\[ t_1 - \sum_{i=2}^{N} t_i \leq t \leq \sum_{i=1}^{N} t_i. \]

The definitions of \( e^{\cdot/t}, \kappa_t, \delta_t \), in (18) and (20), and the definitions of \( m_1 \) and \( m_2 \) given in (11) and (12) imply that

(25) \[ T^{m_1} \leq e^{t/2} \leq T^{m_2}. \]

Define \( p' \) so that

\[ T^{p'} = T^{s+1} e^{t/2}. \]

With (25) this gives

(26) \[ T^{m_1+s+1} \leq T^{p'} \leq T^{m_2+s+1}. \]
We consider $g, r_{\pi/2} q'_{N} = -r_{\theta} q$. That is, we flow by the Teichmüller flow back to $-r_{\theta} q$. The vertical angles of the $\gamma \in \Gamma_{q}(\alpha)$ are bounded away from the direction of the flow. This means that in flowing time $t$ back to $-r_{\theta} q$, the lengths of the $\gamma$ increase by an amount $\Omega(e^{t/2})$. Here $\Omega(L)$ refers to any quantity such that $\Omega(L)/L$ is bounded above, and below away from zero, by constants. We conclude from this, the fact that on $M_{N}$ the lengths of the $\gamma$ are $\Omega(T^{s+1})$, the definition of $p', \gamma_{q}$, and (26), that for some $c_{1}$ and $c_{2}$,

\begin{equation}
(27) \quad c_{1} T^{m_{1}+s+1} \leq c_{1} T^{'p} \leq |\gamma|_{q} = \Omega(T^{'p}) \leq \Omega(T^{s+1} e^{t/2}) \leq c_{2} T^{p} \leq c_{2} T^{m_{2}+s+1}.
\end{equation}

Now Definitions (15) and (16) of $p_{1}$ and $p_{2}$ imply that for sufficiently large $T_{0}$, hence sufficiently large $T$,

\[ T^{p_{1}} \leq c_{1} T^{m_{1}+s+1} \]

and

\[ c_{2} T^{m_{2}+s+1} \leq T^{p_{2}}. \]

Together with (27) this gives conclusion iv).

We now check v). Flowing by $g$, contracts vertical lengths by $e^{t/2}$. Therefore the maximum vertical length of any $\gamma \in \Gamma_{q}(\alpha)$ with respect to $-r_{\theta} q$ is $O(e^{-t/2} T^{s+1})$. This is the maximum horizontal length $h_{\theta}(\gamma)$ of $\gamma$ with respect to $r_{\theta} q$. This together with (26) implies

\begin{equation}
(28) \quad |\sin(\theta_{\gamma} - \theta)| = \frac{h_{\theta}(\gamma)}{|\gamma|} = O\left( e^{-t/2} \frac{T^{s+1}}{T^{s+1} e^{t/2}} \right) = O\left( e^{-t} \right).
\end{equation}

On the other hand $\alpha = \alpha_{1}$ has been shortened to length at most $2\kappa_{N} = 2/T_{N}$ in time $t$, so that $h_{\theta}(\alpha) e^{t/2} \leq 2/T_{N}$. Since $|\alpha|_{q} \geq 400$,

\begin{equation}
(29) \quad |\sin(\theta_{\alpha} - \theta)| = \frac{h_{\theta}(\alpha)}{|\alpha|_{q}} \leq \frac{2 e^{-t/2}}{400 T^{l_{N}}}. \quad (29)
\end{equation}

Then (25), (28), and (29) imply

\[ |\sin(\theta_{\alpha} - \theta_{\gamma})| \leq O(e^{-t}) + \frac{e^{-t/2}}{200 T^{l_{N}}} \leq O\left( \frac{1}{T^{m_{1}}} \right) + O\left( \frac{1}{T^{m_{1}+l_{N}}} \right) \]

\[ \leq \frac{1}{T^{2m_{1}+1}} + \frac{1}{T^{m_{1}+l_{N}+1}} \]

for $T$ sufficiently large. By the definition (14) of $k$, for $T_{0}$, hence $T$ sufficiently large, this quantity is at most

\begin{equation}
(30) \quad \frac{2}{T^{m_{1}+l_{N}+1}} < \frac{1}{2T^{k}}.
\end{equation}
Thus all $\theta_\gamma$ lie in an interval of size \[ \frac{1}{T^k} \]
centered at $\theta_{\alpha}$. This gives v).

Finally we check vi). The gaps between the vertical angles of the $\gamma$ on $M'_N$ are $\Omega(T^{-s-1})$. Since the vertical angles are bounded away from the direction of the flow $g_t$, the gaps decrease by a factor of $\Omega(e^{-t})$ as we flow $t$. Using (25) and the definition of $p'$, we find the gaps between the $\theta_\gamma$ (on the flat surface $q$) are
\[ \Omega(e^{-t}T^{-s-1}) \geq \Omega(T^{-p'})T^{-m_2} = \Omega(T^{p'})^{-1-m_2/p'} . \]

By (26) and (15), the last quantity is at least
\[ \Omega(T^{p'})^{-1-m_2/(p_1+1)} \]
and by inequality (13), and the definitions (14), and (15), this is at least
\[ \Omega(T^{p'})^{-1-k} . \]

Then the gap size which is at least $\Omega(T^{p'})^{-1-k}$ divided by $c_2(T^{p'})^{-k}$ goes to infinity as $T \to \infty$. Thus by taking $T_0$ large enough, we can make the ratio at least $4$. This proves vi) and concludes the proof of Theorem 11.2 in the case that the initial stratum surfaces have genus greater than or equal to 1.

The proof for the sphere is almost identical. Recall for each stratum that there are three distinguished points with cone angle $\pi$ which are not the endpoints of any edge to be collapsed. The three distinguished points are mapped by the collapsing map to the three distinguished points of the face stratum. This implies that the three distinguished points in the initial stratum are not endpoints of any of the edges $\alpha_1, \ldots, \alpha_N$ that will be shortened. Let
\[ E_N = \{ q \in \mathcal{P}_N : \text{shortest saddle connection} \]
\[ \text{joining points of } \Sigma \text{ has length } \leq \delta_N \} . \]

The estimate of the measure of $E_N$ is exactly the same as the corresponding set in the torus case. The argument and estimates now proceed exactly as before. We now have that the complement of $A_{\alpha_1}$ and $\alpha_2, \ldots, \alpha_N$ in $M'_N$ is a flat surface which has three cone angle $\pi$ singularities, the three distinguished points, and is homeomorphic to a disk. We can then apply Theorem 8.2 to conclude there is a subsurface isometric to a sphere with four cone angle $\pi$ singularities minus a convex disk. Passing to the double cover gives a flat torus minus a convex disk to which we may apply Theorem 9.1 to conclude that there are metric cylinders. Projecting back to the sphere gives metric cylinders on the sphere. This concludes the proof of Theorem 11.2.
12. Proof of corollaries

*Proof of Corollary 1.* We can locally write $\mathcal{D}$ as $[0, 2\pi) \times I$ for $I$ an interval of dimension $n - 1 = \dim \mathcal{D} - 1$ and such that for each $y \in I$, $((\theta, y): \theta \in [0, 2\pi)) = \{e^{i\theta}q: \theta \in [0, 2\pi)\}$ for some $q$. The Main Theorem and Fubini’s theorem say for a.e. $y \in I$, $C_y = (\theta: (\theta, y)$ is nonergodic) has Hausdorff dimension $\delta$. Then for $\delta_1 < \delta$, $H^{\delta_1}(C_y) = \infty$ for a.e. $y$. Theorem 5.8 of [Fa] says $H^{n-1+\delta}(0, 2\pi) \times I) = \infty$ so that dim $NE \geq n - 1 + \delta_1$ for each $\delta_1 < \delta$.

*Proof of Corollary 2A.* The principal stratum $\mathcal{D}$ is defined to be the set of flat structures such that all singularities have cone angle $3\pi$. The principal stratum can locally be written as $J \times I$ where $J$ is an open set in $MF$, $I$ is an interval of the same dimension and where for each $x \in J$ all $q = (x, y)$, $y \in I$, have the same vertical foliation. By Corollary 1, $(q = (x, y): q$ is nonergodic) has Hausdorff dimension $> 2n - 1$. By the above remark this set is $NE \times I$. Thus the Hausdorff dimension of $NE > n - 1$.

*Proof of Corollary 2B.* For each $(\Lambda, \sigma)$, $\Lambda \times I$ for an appropriate interval $I$ locally describes a stratum where again for each $x \in \Lambda$ all $(x, y)$ have as vertical foliation the suspension of the interval exchange $x$. Just as above, Hausdorff dim $NE > \dim \Lambda - 1$.

*Proof of Corollary 3.* For any ergodic $q$ let $L(q)$ denote the length of the largest interval $I \subset [-\pi, \pi]$ about 0 such that $r_\theta q$ is not minimal and nonergodic for $\theta \in I$. We wish to show $L(q) = 0$ for a.e. $q$.

**Claim.** If $L(q) \neq 0$, $2\pi$ and $t > 0$, then $L(g_t(q)) > L(q)$.

**Proof of claim.** The interval $I$ corresponds to an interval on the circle at $\infty$, namely a set of directions. Since $g_t(q)$ moves $q$ toward that interval at $\infty$ the angle of directions viewed from $g_t(q)$ is greater than the angle viewed from $q$, proving the claim.

Thus for any $c \geq 0$ let $E_c = \{x: L(x) \geq c\}$. The claim says $g_t E_c \subseteq E_c$ for any $t$. Since $g_t$ is ergodic this implies $\mu(E_c) = 0$ or its complement has measure 0 for each $c$. Here $\mu$ is a measure on $Q$. This in turn implies $L$ is constant a.e. The constant must be 0 or $2\pi$ since $L(q)$ increases under the flow when $L(q) \neq 0$, $2\pi$. Now by our theorem $L(q) \neq 2\pi$ for a.e. $q$. Thus $L(q) = 0$ for a.e. $q$. 

*University of Illinois at Chicago, Chicago, IL*  
*Cornell University, Ithaca, NY*
References


(Received February 12, 1990)
(Revised November 21, 1990)