Invent. math. 132, 607–632 (1998)

Inventiones mathematicae

Simple geodesics and a series constant over Teichmuller space

Greg McShane

Laboratoire de Mathématique, Université Paul Sabatier, F-31062 Toulouse, France

Oblatum IV-1995 & 5-VI-1997

Abstract We investigate the Birman Series set in a neighborhood of a cusp on a punctured surface, showing that it is homeomorphic to a Cantor set union countably many isolated points cross a line. The local topology of the Cantor set is shown to be related in a simple way to the global behavior of simple geodesics. From this we deduce that a certain series is constant across the Teichmuller space.

1. Introduction

In [11], by analysing the pointset of all simple geodesics on the torus, the following striking identity for the lengths of closed simple geodesics was obtained:

Theorem 1. *Let M be a once punctured torus then*

$$\sum_{\gamma} \frac{1}{1+\exp|\gamma|} = \frac{1}{2} \quad ,$$

where the sum is over all closed simple geodesics γ on M.

Remark 1.1. By an argument taking a quotient of the Z_2 homology cover of the punctured torus [8] it is possible to deduce from the above that if M is a 4 punctured sphere then

Substantial parts of this work appeared in the authors thesis, this was funded by an SERC studentship.The author was supported by an EC grant no. ERBCHBGCT930251 while preparing this work.

$$\sum_{\gamma} \frac{1}{1+\exp\frac{1}{2}|\gamma|} = \frac{1}{2} \ ,$$

where the sum is over all closed simple geodesics γ on M.

This identity differs from previous series in that the sum runs not over all closed geodesic, as is the case with the Selberg trace formula and its relatives, but just over the simple closed geodesics; the closed simple geodesics are a small subset of all closed geodesics. In this article we show how the techniques of [11] can be applied to a more general surface $\mathbf{M}$ (convex finite area with a cusp and without boundary) to give a description of the Birman Series set near a cusp. From this we calculate a series constant across the Teichmuller space of $\mathbf{M}$ which generalises the above identity.

Theorem 2. *For the surface* $\mathbf{M}$,

$$\sum \frac{1}{1+\exp\frac{1}{2}(|\alpha|+|\beta|)} = \frac{1}{2} \ ,$$

where the sum is over all pairs of closed simple geodesics α, β *which bound an embedded pair of pants containing the cusp point, and* $|\alpha|$ (*resp.* $|\beta|$) *is the length of* α (*resp.* β).

Remark 1.2. One recovers the identity for the torus by noting that for such embedded pants the two boundary geodesics are equal. For the 4 punctured sphere all such embedded pants have 2 punctures and only one geodesic boundary component; one can think of this as always having a bounding geodesic in the pair of length 0.

We begin, therefore, by reviewing some definitions and generalities concerning surfaces and geodesic laminations. Much of this appears in some form in [14] but a better exposition of the material is given by [5].

Throughout $\mathbf{M}$ denotes a Riemann surface equipped with a complete hyperbolic structure of finite area, this means that there is a faithful representation of $\pi_1(\mathbf{M})$, $\Gamma < \mathrm{Isom}(\mathrm{H}^2)$ such that Γ is a discrete subgroup and the quotient H^2/Γ is isometric to $\mathbf{M}$. The main feature of our surface is a chosen cusp region: a *cusp region* is a portion of the surface isometric to $\{z : \mathrm{Im}\, z \geq 1\}/z \mapsto z+p$. A simple calculation shows that the area of the cusp region is p. We insist that the cusp region is embedded in $\mathbf{M}$ and not merely immersed. In what follows we will specify a cusp region by its area; by a *small cusp region* we mean one of sufficiently small area. By *near the cusp* we mean in a sufficiently small cusp region. Following Lemma 2 we will make a quantitative definition of what we mean by a small cusp region.

A *geodesic* is a curve which is locally length minimising on $\mathbf{M}$. One says that a geodesic is *complete* iff it is not strictly contained in any other geodesic. A geodesic or any curve for that matter is *simple* iff it contains no self intersections. After Thurston one defines a *geodesic lamination* to be any collection of disjoint complete simple geodesics which is closed as a pointset; each complete geodesic contained in the lamination is called a *leaf*. A lamination is *minimal* iff no proper subset is a lamination. An example of a minimal lamination is a single closed geodesic; it has a single leaf. There are, however, minimal laminations which have more than one leaf (see [6] for examples); necessarily such a lamination must have uncountably many leaves [5]. We shall call such a lamination (minimal, compact, more than one leaf) *non-trivial*.

We say a geodesic γ *spirals* to a lamination $\Omega(\gamma)$ if $\Omega(\gamma)$ is in the closure of γ on $\mathbf{M}$. We adopt this notation following the standard definition of the *ω-limit set* of an orbit in a dynamical system, see [9]. We note that $\Omega(\gamma)$ is necessarily minimal in the sense above. On the other hand a geodesic has an *end up the cusp* iff that end is contained in a small cusp region on $\mathbf{M}$. With these definitions:

(classification of geodesics) *any complete simple geodesic falls into exactly one of the following three classes*:

1. *It is a leaf of a compact lamination.*
2. *It has a single end spiraling into a compact lamination and the other end up a cusp.*
3. *It has both ends up a cusp.*

More usually ([5] Theorem 4.2.8) one thinks of a lamination as decomposing as a finite number of minimal sublaminations union a finite number of geodesics which spiral to these. Thus a geodesic is either a leaf of a minimal lamination (this includes the case of a geodesic with both ends up the cusp) or has an end spiraling to a minimal sublamination. However, we will find it more convenient to classify the leaves as above as this emphasises the behavior of the geodesic with respect to the cusp.

Henceforth, we shall say a geodesic is a *cusp geodesic* if it has both its ends up the cusp.

The definitions above are concerned with collections of geodesics which are leaves of individual laminations, we now discuss properties of collections of laminations.

One of the most important tools in the study of geodesic laminations is the fact that the set of closed simple geodesics is dense in the set of compact minimal laminations ([5] Theorem 4.2.15). This fact is proved using an approximation technique due to Thurston; a large part of our work will be to improve on this method (see for example Theorem 13).

A fundamental question one can ask about the set of all laminations on a surface is how big is it as a subset of $\mathbf{M}$? It is well known that the pointset of a geodesic lamination on a hyperbolic surface has measure zero. For a sufficiently complicated surface there are uncountably many laminations so a naive measure theory argument does not show anything. For a number of years the question of the size of the union of all laminations was an open question, finally it was settled by the following remarkable Theorem 3:

Theorem 3 (Birman Series). *Let G be the set of all simple geodesics on a hyperbolic surface. The set S of points which lie on a geodesic $\gamma \in G$ has Hausdorff dimension* 1.

Henceforth we call the pointset of the union of all complete geodesics the *Birman Series set* and we abreviate this to BS set. Near a generic point of the BS set the topological/combinatorial structure of the set is very complicated [12]. However, in a sufficiently small neighborhood of the cusp (cusp region) the BS set has a beautifully simple description which we give in the following paragraphs.

Let $\mathscr{H} \subset \mathbf{M}$ be some fixed small cusp region. We shall see later Lemma 2 that we can choose $\mathscr{H}$ so that if an end of a simple geodesic enters this region then it will never leave, that is the end goes up the cusp. Now define E to be the set of all points of $\mathscr{H}$ which lie on some complete simple geodesic or other. In broad terms $\mathscr{H}$ is homeomorphic to a cylinder, that is to the product of a line and the circle; we will identify the circle factor in a natural way with $\partial\mathscr{H}$. This product structure restricts to a product structure on E, the set $\partial\mathscr{H} \cap E$ is a *cross section* of the set of ends in that E is homeomorphic to $(\partial\mathscr{H} \cap E) \times \mathrm{R}^+$. We show that $\partial\mathscr{H} \cap E$ is a (non-empty) Cantor set union countably many isolated points. Characterisations will be given of the isolated points and the boundary points of the Cantor set in terms of the global topological behavior of complete geodesics. We shall call a maximal component of $\mathscr{H} \backslash E$ (resp. $\partial\mathscr{H} \backslash E$) a *gap*.

We recall the definitions of pointset topology which we use in the sequel, Let Y be a topological space and $X \subset Y$ then $x \in X$ is *isolated* if there is an open set $U, x \in U$ such that $U \cap X = \emptyset$. On the other hand x is a *boundary point* if x is the limit of points in X *and* there is a connected open set U contained in the complement $Y \backslash X$ such that x is in the closure of U. Note this second condition is somewhat stronger the more usual assumption that any neighborhood of x meets $Y \backslash X$; if we adopt that definition then for a Cantor set in R *every* point would be a boundary point.

By definition each point $x \in E$ lies on a complete simple geodesic γ_x. The point x lies on a portion of this geodesic which contains an end of γ_x which lies up the cusp. There is another end of γ_x, we will call this in what follows simply *the other end for x*. By considering the behavior of this other end one can decide what the set E looks like in a neighborhood of x. We shall show:

Theorem 4. *Let* $\mathbf{M}$ *be a punctured hyperbolic surface without boundary and $x \in E \cap \partial\mathscr{H}$, with E as above, then*

1. *The point x is isolated iff the other end goes up a cusp.*

2. *The point x is a boundary point iff the other end spirals to a closed simple geodesic.*

3. *Each gap is bounded by two points $x, y \in \partial\mathscr{H} \cap E$, exactly one point, x say, lies on a geodesic with both ends up the chosen cusp. The point y lies either on a geodesic with an end spiralling to a closed simple geodesic or on a*

geodesic with its other end up some other cusp; both these cases occur, however if the surface has only one cusp only the first case occurs.

Remark 1.3. It is perhaps convenient to think of the other cusp in part 3 as a 0 length geodesic and a geodesic with an end up it as somehow spiralling to this.

We have immediately.

Corollary 5. *The intersection of a suitably small horocycle with the set of simple geodesics with an end up the cusp but which are not cusp geodesics is a Cantor set.*

Proof. Recall that a Cantor set is a closed nowhere dense set with no isolated points. Let $\partial\mathscr{H}$ be a suitably small horocycle. Since the union of all simple geodesics is closed as a pointset $E \cap \partial\mathscr{H}$ is indeed closed and by the arguments of Birman-Series [3] also nowhere dense. One checks that on removing the isolated points from $E \cap \partial\mathscr{H}$ these 2 properties are preserved. By the theorem what remains is the set of points of E not lying on any cusp geodesic. $\square$

Our structure theorem (Theorem 4) together with the Birman Series result imply the following:

Corollary 6. *The area of a cusp region is equal to the area of the union of all gaps.*

We deduce 2 from this by a simple computation of the area of a cusp region.

The paper is organised as follows. We begin by discussing in an expository manner simple geodesics on pants (Section 2) and the behavior of simple geodesics with respect to small cusp regions (Section 3). From this we deduce that ends of cusp geodesics give rise to isolated points in E (Section 4, Theorem 9). By using Dehn twists (Section 5), we show that geodesics which spiral to simple closed geodesics are boundary points of E (Theorem 11). Finally we give an argument (Section 6) that shows that a geodesic which spirals to a non-trivial lamination is neither isolated nor a boundary point (Theorem 13). Given these 3 results we demonstrate Theorem 4 (Section 8) and Theorem 2 (Section 9).

Generalisations etc. In [13] the author, by using a geometric coding scheme for geodesics, gives a description of the BS set on the punctured torus near a *Weierstrass point* and in this way calculates 3 other identities similar to 1.

Brian Bowditch [4] has given a different proof of 1 by a means of the Fricke trace relations for $SL(2, \mathbf{C})$ and the Markoff binary tree.

Acknowledgements. I should like to thank T. Jorgensen, A. Marden and D. Epstein for their encouragements and B.H. Bowditch, C.M. Series, S.J. Patterson, S.E. Cross for useful conversations and various assistance whilst preparing this article. I am most grateful to the Ecole Normale Superieure de Lyon for their hospitality and in particular to P. Iglesias for technical assistance. The author is indebted to the referee for his patience.

2. Pairs of pants

The least complicated of all hyperbolic surfaces with a non-trivial convex core is a pair of pants. A *pair of pants*, or *triply connected domain*, is a surface homeomorphic to a sphere minus 3 points. Such a surface admits many convex hyperbolic structures though there is only a single complete finite volume structure; the space of structures is 3 dimensional and parameterised by the lengths of the so called *boundary geodesic* (see [2]). We shall insist that all our pants have a single cusp, that is one of the boundary geodesics has length 0. What concerns us here is that there are only finitely many simple geodesics on pants and these are easy to describe in terms of their ends. This description will prove to be key at various stages in our subsequent investigations.

Any simple loop on a pair of pants which does not bound a disc is homotopic to the boundary. Thus the set of simple closed geodesics is finite, since simple closed geodesics are dense in compact minimal laminations all such laminations must be simple geodesics; one easily sees from the classification of geodesics that there are only finitely many complete simple geodesics on a pair of pants. One of these geodesics is the unique geodesic with both ends up the cusp γ. There are 4 other geodesics each with one end up the cusp the other end spiraling to the boundary. In addition to these there are 4 other simple geodesics each with both ends spiraling to different boundary components; these are investigated in [13] but will not concern us here. Finally there are 4 simple geodesics which each have both ends spiraling to the same boundary geodesic. In summary there are 15 simple geodesics on our pair of pants of which 2 are closed (boundary geodesics), one has both ends up the cusp, 4 others have an end up the cusp and their other end spiraling to one of the boundary geodesics, another 8 have both ends spiraling to boundary geodesics in various ways. Since we are principally concerned with ends of geodesics near the cusp these last eight will play no role in what follows.

Now let one of these boundary geodesics shrink to 0 length so that we now have 2 cusps and only one boundary geodesic. This surface has a symmetry (order 2 isometry) that swaps the cusp points. Choose a cusp point. There are 4 simple geodesics with ends up this cusp: one geodesic with both ends up the cusp, two with ends spiraling to the boundary geodesic and a fourth with an end up each cusp.

The following Lemma will help us reduce the study of the BS set near cusp geodesics to an analysis of pants.

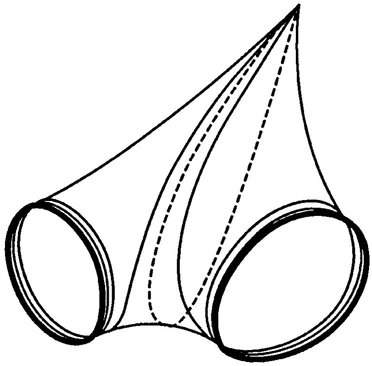

Fig. 1. The figure shows a pair of pants with some of geodesics drawn on it. The two closed simple are shown as heavy black curves, the dotted curve represents the unique simple geodesic with both ends up the cusp; also shown are two of the 4 geodesics possessing one end up the cusp and the other spiraling to the closed geodesics

Proposition 1. *Let* **M** *be a surface with at least one cusp and let γ be a geodesic with both ends up the cusp then γ is contained in a unique embedded pair of pants.*

Proof. Suppose that **M** has just one cusp.

Let $H \subset \mathbf{M}$ be a small horocycle that meets γ in exactly two points. One cuts and pastes γ and H to make two curves α, β (see figure). A small perturbation of this pair of curves gives a pair of simple disjoint loops disjoint from γ; one straightens these to get a pair of simple disjoint geodesics disjoint from γ. When one cuts along these geodesics one component of the cut surface is a pair of pants containing γ. This is the embedded pants P.

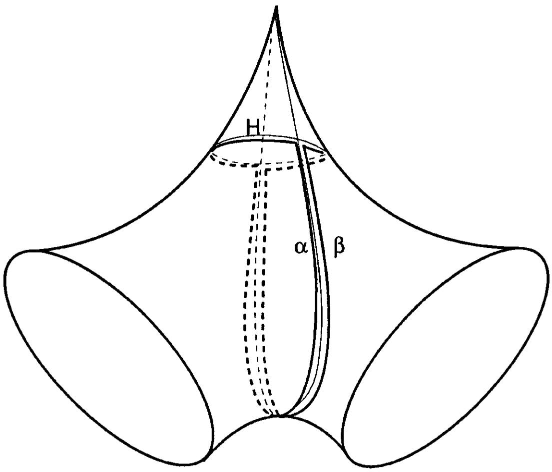

Fig. 2. The construction of the curves α, β

To see uniqueness suppose that P' is some other pair of pants containing γ. Let α be one of the boundary geodesics of P'. There is a curve α' homotopic to α and contained in P; by convexity this straightens to a (simple closed) geodesic in P and so must be one boundary geodesics of P. It follows $P' = P$.

These arguments are easily modified for the case that **M** has more than one cusp. $\square$

3. Cusp regions

We provide an exposition of the relationship between simple geodesics and small cusp regions; most of this has appeared elsewhere [5] but we provide new sharpness results. In particular we show leaves of compact laminations cannot enter small cusp regions and that a simple geodesic must behave in a very nice way near the cusp (recall that by *near the cusp* we just mean in a sufficiently small cusp region). To do this we introduce the horocyclic foliation of the cusp region which, in suitably small cusp regions, a simple geodesic must always intersect orthogonally.

Let C be a cusp region; identify it with the quotient $\{z : \operatorname{Im} z \geq 1\}/z \mapsto z + p$ where p is its area. The upper half plane is foliated by horizontal lines each of which one thinks of as a horocycle based at ∞. This foliation is invariant under the translation $z \mapsto z + p$ and so descends to a foliation on the quotient i.e. on the cusp region C. This is the *horocyclic foliation*; each leaf is a concentric horocyclic curve homotopic to the cusp point. If a cusp region is included in a larger cusp region then each leaf of its horocyclic foliation is a leaf of the horocyclic foliation on the larger one. The orthogonal foliation to the horocyclic foliation consists of leaves each of which is a geodesic segment with an end up the cusp. This gives the cusp region a natural, in the sense that it comes from the geometry, product structure $S^1 \times \mathrm{R}$, where the circle factor is identified with a horocycle and the line factor with a leaf of the orthogonal foliation. This will be important in what follows.

Lemma 2. *Let M be a surface with a cusp; such a surface has a cusp region of area 2. The portion of a complete simple geodesic which lies in a cusp region of area less than 2 always meets the horocyclic foliation perpendicularly. The number 2 is the least such number with this property.*

Remark 3.1. This lemma appears in [5] without the sharpness statement.

We can now say quantitatively what we mean by a small cusp region – it is just one of area less than 2.

Proof. To show that every surface with a cusp has a cusp region of area 2 one proceeds as follows. Take a geodesic on the surface with both ends up the cusp; by the preceding result such a geodesic is contained in a unique pair of pants P. It is then a simple excercise in hyperbolic trigonometry (left to the reader) to verify that all pants with a cusp have a cusp region of area 2.

Suppose first that we lift to H^2 and conjugate Γ so that the cusp is at ∞ and the cusp region of area 2 is covered by $\{z : \operatorname{Im} z > 0\}/z \mapsto z + 2$.

Let γ be a simple geodesic. A lift $\gamma' \subset H^2$ is either a semicircle with center on the real line or a vertical line. If γ' is a semicircle then it must have radius less than 1. This is because the translate of γ' by $z \mapsto z + 2$ will also be a lift and if the radius is bigger than 1 then the two will intersect; the intersection would project to a self intersection of γ on the surface.

If γ enters the cusp region on $\mathbf{M}$ then there is a lift γ' which enters the cusp region in $\mathbb{H}^2$. By the above argument this cannot be a semicircle and so has to be a vertical line this meets the horocyclic foliation perpendicularly. The result follows on passing to the quotient.

To see that 2 is sharp consider a pair of pants with bounding geodesics α and β. One constructs a sequence of surfaces for which the length of α tends to 0 while the length of β goes to ∞. The shortest distance between β and the cusp region goes to 0 along these surfaces. By doubling/gluing to another surface one sees that 2 is sharp for finite volume and even single cusp surfaces. $\square$

An immediate consequence of this lemma is the following:

Corollary 7. Let $\epsilon > 0$. *Any geodesic which is the leaf of a compact lamination on a punctured surface does not intersect the cusp region whose bounding curve has length* $2 - \epsilon$. *Moreover this bound is sharp.*

Proof. By the lemma a complete simple geodesic which intersects the cusp region whose bounding curve has length $2 - \epsilon$ must have an end up the cusp. Thus it cannot be contained in any compact subset of the surface.

The sharpness at 2 goes by the same argument as in the theorem above.

The following tells us that E has a product structure in the cusp. $\square$

Corollary 8. *Let γ be a complete simple geodesic on a punctured surface. The geodesic γ intersects no other complete simple geodesic in any cusp region whose bounding curve has length less than* 2.

Proof. Let $X \subset \gamma$ be the portion of such a geodesic in the cusp region $\mathscr{H}$; by the above this is the leaf of the orthogonal foliation. Firstly X cannot meet the leaf of a compact lamination since no such geodesic can enter the cusp region. Secondly if $\alpha(\neq \gamma)$ is some other cusp geodesic then $\alpha \cap \mathscr{H}$ consists of a pair of leaves of the foliation orthogonal to the horocyclic foliation and so this set is disjoint from X.

The case where α only has one end up the cusp goes likewise. $\square$

4. Isolated points

In this section we use a simple geometric argument to establish that geodesics with both ends up cusps give rise to isolated points of E.

Theorem 9. *Let* **M** *be a surface with at least* 1 *cusp. If* $x \in E$ *is a point which lies on a geodesic with its other end up a cusp then* x *is an isolated point.*

Proof. Suppose that M has only one cusp and let γ be a geodesic with both ends up the cusp. By Lemma 1 there is a unique embedded pair of pants containing γ; let α, β be the bounding geodesics for these pants. There are, by our analysis of geodesics on pants, precisely four simple geodesics on this surface having the property that each has a single end up the cusp and another end spiraling to the boundary (either α or β). These divide $\mathscr{H}$ into 4 regions, 2 of which contain ends of γ; we now show that there are no other points of E in these 2 regions.

We work in the upper half plane and conjugate the covering group, Γ, for M so that ∞ is a cusp point. Let Λ denote the maximal subset of H^2 which projects to $\alpha \cup \beta$ on M. Now let μ be a vertical line and suppose that its image on M is not equal to one of the 4 geodesics above. It is easy to see that this image curve must at some point leave the pair of pants bounded by $\alpha \cup \beta$, and so μ must meet Λ. The set Λ consists of a disjoint union of semicircles. Suppose, for notational convenience, that the first such semicircle μ meets is a lift of α; set μ^-, μ^+ to be the pair of vertical lines through the endpoints of this semicircle. Both μ^-, μ^+ project to simple geodesics on the surface. One see this by constructing curve which has the same endpoints as μ^- (resp. μ^+) and which projects to a simple curve on the surface: for example, the piecewise geodesic in the hyperbolic plane consisting of a portion of μ and a portion of α, having a single "corner" where these portions meet, and which has the same endpoints as μ^- (resp. μ^+). This piecewise geodesic is simple since both μ and α are simple and the only intersections between the portion of μ and the portion of α is the "corner"

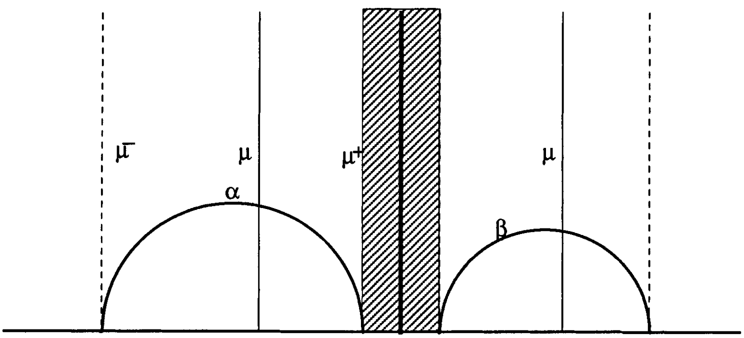

Fig. 3. The diagram of lifts in H^2 used in the proof of Theorem 9

(another intersection between these two portions would contradict the fact that α was chosen to be of maximal diameter). Now since both μ^+, μ^- are asymptotic to a lift of α, their images on M both spiral to α. Finally no vertical line which projects to γ on M can lie between μ^-, μ^+ because it would meet Λ; for then γ would meet $\alpha \cup \beta$ on the surface, contradicting our hypothesis.

One can easily modify this argument for the case when **M** has more than 1 cusp. The only difference is that, in the case where one of the components of $\mathbf{M} - \gamma$ is a punctured monogon, the pair of pants containing γ has two punctures. This means that one of the lines μ^+, μ^- has both its endpoints at cusp points; on the pants two of the geodesics which spiraled to the boundary, which has "shrunk to a cusp", have been merged together. However, the conclusion that the strip between μ^+, μ^- meets no simple geodesics is still valid by the same argument as given above.

Finally, if **M** has more than one cusp and x is a point on γ whose other end is up a different cusp, one takes the unique embedded pair of pants with 2 cusps and one geodesic boundary which contains γ. This pair of pants has a unique simple geodesic with both ends up our chosen cusp and, in addition, this geodesic is disjoint from γ. By the arguments of the previous paragraph the punctured monogon bounded by this geodesic meets no simple geodesic other than γ. $\square$

The following is an immediate consequence of the proof of the above Theorem:

Corollary 10. *Let K be a gap with a point x in its boundary which lies on a geodesic γ with both ends up the cusp.*

1. *If γ bounds a punctured monogon in the surface and K is contained in this monogon, then the other point in the boundary of K lies on a geodesic which goes up the other cusp on the monogon.*

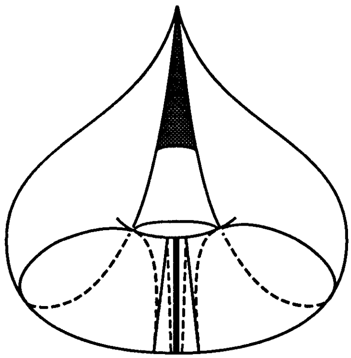

Fig. 4. The figure shows the location of a gap in the Birman series set (the shaded region at the top) associated to a certain closed simple geodesic (the heavy black line at front) on a punctured torus

2. *If γ does not bound a punctured monogon then the other point in ∂K lies on a geodesic which spirals to a closed geodesic α and α, γ bound an embedded cylinder in* $\mathbf{M}$.

Remark 4.1. In the spirit of the discussion of pants with two cusps these cases can be viewed as being more or less the same, that is one thinks of the other cusp as being a zero length geodesic and the geodesics with an end up that cusp as "spiraling to this geodesic".

We can in fact compute the width of the gap.

Proposition 3. *Let x be an isolated point which lies on a geodesic γ with both ends up the same cusp, x is in the closure of 2 gaps. The area of the union of these gaps abutting on x is*

$$\frac{1}{1+\exp\frac{1}{2}(|\alpha|+|\beta|)} \times \textit{area of cusp region} \ ,$$

where $|\alpha|$ and $|\beta|$ are the lengths of the boundaries of the pants containing γ.

Remark 4.2. If the pants containing γ contain another cusp then we simply set one of the boundary lengths to 0 and the formula remains valid.

Proof. In what follows a strip is a portion of a surface isometric to $\operatorname{Im} z > p, 0 < \operatorname{Re} z < 1$, the area of a strip is $1/p$.

One cuts the pants into 4 quadrilaterals (cut along the geodesic with both ends up the cusp, along the unique simple geodesic segment perpendicular to the boundaries, along the pair of simple geodesics which have an end up the cusp and meet the boundary geodesics perpendicularly). Each of the quadrilaterals has 3 right angles and a zero angle (at the cusp); there are two sides of infinite length (abutting the cusp) and two sides of finite length – one of these finite sides has length $|\gamma|/2$ where γ is one or other of the boundary geodesics.

Consider a quadrilateral in the upper half plane which is congruent to one of the above. We normalise so that the two infinite sides lift to $\operatorname{Re} z = 0, \operatorname{Re} z = 1$ respectively. The finite sides lift to arcs of semicircles (marked C,D in the diagram), one takes the arc, marked C say, corresponding to the boundary geodesic and extends it to meet the boundary at a point marked x in the diagram. The vertical line which has x as its endpoint represents the cusp geodesic which spirals to the boundary on the pants. This vertical divides the strip into two parts (in the diagram shaded, unshaded – the shaded part corresponds to a gap), we calculate the ratio of the areas of these parts. From this we will calculate the area of the gap, by Corollary 6, in a small cusp region.

Let a, b be the Euclidean lengths of the radii shown in the diagram; note that because of our normalisation and right angles $a^2 + b^2 = 1$. The

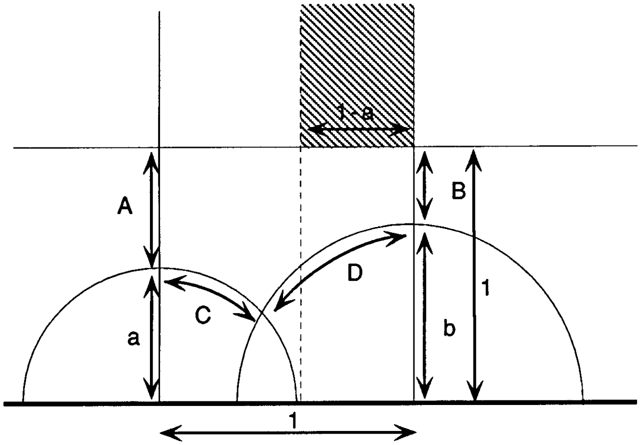

Fig. 5. The figure shows a quadrilateral used to calculate the area of a gap associated to a particular closed simple geodesic. The capital letters (A, B, C, D) are lengths measured in the Poincare metric on the upper half plane; the other two lengths (a, b – the radii of the circles) are measured in the usual Euclidean metric

shaded area is $(1 - a)$ times the total area. A simple computation shows that

$$C = \log\frac{(1+a)}{b} = \frac{1}{2}\cdot\log\frac{1+a}{1-a} = \frac{1}{2}\mathrm{cotanh}(a/2) \ .$$

One rearranges this to see $a = \tanh(\frac{1}{2}$ length boundary geodesic γ); it follows from standard trigonometric identities that $b = \mathrm{sech}(\frac{1}{2}$length boundary geodesic γ).

We now take 4 of these quadrilaterals, 2 abutting on α and two abutting on β and identify edges to recover the original pants. Consider the cusp region of area $2\,\mathrm{sech}(\frac{1}{2}|\alpha|) + 2\,\mathrm{sech}(\frac{1}{2}|\beta|)$. This meets each quadrilateral abutting on α (resp. β) in a strip of area $2(1 - \tanh(\frac{1}{2}|\alpha|)) \cdot \mathrm{sech}(\frac{1}{2}|\beta|)$ (resp. $2(1 - \tanh(\frac{1}{2}|\beta|) \cdot \mathrm{sech}(\frac{1}{2}|\alpha|))$ (one sees this by a calculation based on the observation that the strips considered in the paragraph above will not in general glue together to make a cusp region and one must weight them appropriately.) Thus the ratio of the total area of the gaps to that of the cusp region is

$$\frac{(1 - \tanh(\frac{1}{2}|\alpha|)) \cdot \cosh(\frac{1}{2}|\alpha|) + (1 - \tanh(\frac{1}{2}|\beta|)) \cdot \cosh(\frac{1}{2}|\beta|)}{\cosh(\frac{1}{2}|\alpha|) + \cosh(\frac{1}{2}|\beta|)} \ ,$$

and the proposition follows on rearrangement of this. □

5. Limits of iterated Dehn twists

Throughout **M** denotes a surface with a distinguished cusp which we call the *chosen cusp*.

We next show that if x lies on a geodesic γ spiralling to a closed geodesic $\Omega(\gamma)$ then x is the limit point of a sequence in E. This implies that x is a boundary point – since one can find an embedded pair of pants on M such that $\Omega(\gamma)$ is one of the boundary geodesics of this pants (with a suitable interpretation of this in the case of the punctured torus), and such that γ is one of the four simple geodesics spiralling to the boundary. By Theorem 9 there is a gap in the complement of E which has x in its closure.

The technique we employ is to use *Dehn twists* to construct a sequence of geodesics γ_n with limit a lamination containing the geodesic γ. If the closed geodesic $\Omega(\gamma)$ is separating then the sequence we construct consists of geodesics each with exactly one end up the cusp. If the closed geodesic $\Omega(\gamma)$ is non-separating then each geodesic in the sequence has both ends up the cusp. In either case, and an important feature of our construction, for each $n > 0$, γ_n meets $\Omega(\gamma)$ *exactly once*; this avoids unnecessary technicalities arising from multiple intersections (see [7]). Background material relating to Dehn twists can be found in [6]. More detailed accounts are given in [10] or [7].

We begin by constructing a nice geodesic, the γ_1, of our sequence.

Lemma 4. *Let M be a punctured surface without boundary. Let γ be a simple geodesic with one end up the cusp and the other end spiralling to a closed geodesic $\Omega(\gamma)$. There is a complete simple geodesic, β, such that*

1. *it meets $\Omega(\gamma)$ in a single point,*
2. *for any lift γ' of γ to H^2 there is a lift, $\Omega(\gamma)'$, of $\Omega(\gamma)$ and a lift, β', of β such that $\gamma', \beta', \Omega(\gamma)'$ form a triangle in H^2 (see diagram).*

Remark 5.1. The first condition is not strictly necessary but averts subsequent complications, due to multiple intersections with the twist locus, in calculating the image of β under iterated Dehn twists round $\Omega(\gamma)$. The second condition ensures that under such iterated twists the limit of β will contain γ.

Proof. We shall give a construction of β which is different in the case when $\Omega(\gamma)$ is separating from when it is non-separating. The construction in each case consists of two parts; the first of these takes place in the upper half space and is common to both cases (separating, non-separating).

Suppose $\Omega(\gamma)$ is non-separating. Lift to H^2 so that the cusp is at ∞ as usual. There is a vertical line γ' which is a lift of γ and this is asymptotic to a semicircle, which we denote $\mathscr{D}$, which projects to $\Omega(\gamma)$ on **M**. By following γ' down from ∞ and then taking a short arc across to $\mathscr{D}$ one obtains a curve which projects to a simple curve on **M** (since $\gamma, \Omega(\gamma)$ are disjoint). Now take a vertical line starting at ∞, β_1', which has the same endpoint on $\mathscr{D}$ as this

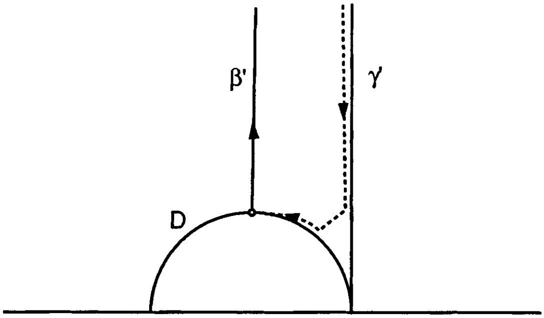

Fig. 6. The diagram shows the construction of a lift β' from a lift of γ, denoted γ', and a lift of a leaf of μ, denoted D. The dotted line is a lift of a simple curve with the same endpoints as β, so β' projects to a simple curve on the surface

curve. The image, β_1, of β_1' on $\mathbf{M}$ is simple since it is a geodesic fixed endpoint homotopic to a simple curve. Moreover, since ∞ is a cusp point, β_1 has an end up the cusp and an endpoint on $\Omega(\gamma)$.

Cut $\mathbf{M}$ along $\Omega(\gamma)$; the resulting surface has 2 boundary components and the image of β_1 has an endpoint on one of these. Choose β_2 to be any simple arc disjoint from β_1, with an end up the cusp and an endpoint on the other boundary component of the cut surface. When we reglue the cut surface to get $\mathbf{M}$, β_1 and β_2 are a pair of disjoint curves each with an endpoint on $\Omega(\gamma)$. Take the curve β to be the union of β_1, β_2 and a segment of $\Omega(\gamma)$ joining the two endpoints of these curves on $\Omega(\gamma)$. Straighten β to get a geodesic satisfying the conditions of the lemma.

Now suppose that $\Omega(\gamma)$ is separating and that $\mathbf{M}$ is not a four punctured sphere (we deal with this case below). Construct the curve β_1 in the upper half plane and continue to denote by β_1 its projection to $\mathbf{M}$ just as before. The component of the cut surface which does not contain the chosen cusp has at least one closed geodesic. There is a simple geodesic β_2 with an endpoint on $\Omega(\gamma)$ and a single end spiralling into some closed geodesic in this component. The images of β_1, β_2 on the surface are a pair of disjoint curves each with an endpoint on $\Omega(\gamma)$. The curve β is the union of β_1, β_2 and a segment of $\Omega(\gamma)$ joining the two endpoints. Straighten β we get a geodesic satisfying the conditions of the lemma.

Suppose now that $\mathbf{M}$ is the four punctured sphere. Each geodesic on this surface is separating and the cut surface consists of two components each of which is a pair of pants. In this case we can take β_1 to be any simple arc running from the fixed cusp to the boundary and β_2 to be any simple arc on the other component of the cut surface with an end up the cusp and an endpoint on the boundary. We then proceed as before. □

We now define a Dehn twist; an account of the properties of these maps is given in [1]. Let γ be a closed simple geodesic on a surface. There is a regular neighborhood of γ which is homeomorphic to an annulus. Parameterize this neighborhood as $\{[r, \theta] : 1 \leq r \leq 2\}$. The *Dehn twist* round γ is the homeomorphism which is the identity of this annulus and the map $[r, \theta] \mapsto [r, \theta + 2\pi r]$ on the annulus.

The image of a geodesic β which intersects γ, under a Dehn twist round γ, is not a geodesic. However, it is a piecewise smooth curve and there is a unique geodesic determined by straightening. If the geodesic β is simple then the geodesic determined by its image is simple. If the geodesic β intersects γ exactly n times then the geodesic determined by the image of β under a Dehn twist round γ intersects γ exactly n times.

Theorem 11. *Let M be a closed punctured surface. Let γ be a simple geodesic with one end up the cusp and the other end spiralling to a closed geodesic $\Omega(\gamma)$. Then there is a sequence of points of E converging to γ. Further these points can be taken to lie on a sequence of simple geodesics which are given by the images of a fixed geodesic under iterated Dehn twists round $\Omega(\gamma)$. The choice of the fixed geodesic depends on γ.*

Proof. We give a proof in the case where $\Omega(\gamma)$ is a non-separating geodesic; it is left to the reader to check that the method extends, with minor modifications, to the case where $\Omega(\gamma)$ is separating. Let β be the geodesic constructed in Lemma 4; we will consider the sequence of geodesics $\mathrm{T}^n(\beta), n \in \mathrm{N}$, where T denotes the Dehn twist round $\Omega(\gamma)$. It will be convenient to work in the upper half plane.

We lift to H^2 in the same manner as in Lemma 4 (see diagram). Conjugate the group of covering transformations for M, Γ, so that ∞ is a cusp point. Choose a vertical line γ' which is a lift of γ, and let $\mathscr{D}$ be the largest semi-circle which is tangent to γ' and is a lift of $\Omega(\gamma)$. By the proof of Lemma 4 there is a lift of β, β' say, which meets $\mathscr{D}$. Let $A \in \Gamma$ be the hyperbolic transformation which has axis $\mathscr{D}$, covers $\Omega(\gamma)$, and such that the attracting fixed point of A is the endpoint of γ'.

We now describe a procedure to calculate a sequence of geodesics, which we denote by β'_n, such that β'_n covers $\mathrm{T}^n\beta$ for each n. Let $\beta'_0 = \beta'$, as constructed above, and let β'_n, for $n \geq 1$, be the vertical line with endpoint the image of the finite endpoint of β' under A^n. One shows that β'_n covers $\mathrm{T}^n\beta$ as follows. In the surface, one can construct a piecewise geodesic curve in the same homotopy class as $\mathrm{T}^n\beta$ by following β down from the cusp till it meets $\Omega(\gamma)$, then going n times round this closed geodesic (in the right direction) and finally returning to the cusp via β. In the universal cover H^2 this construction corresponds to following β' down from ∞, walking along $\mathscr{D}$ until we come to the nth lift of β that we meet, and then travelling along this lift of β to its endpoint on the other side of $\mathscr{D}$ from infinity. It is easy to see that the nth lift of β that we meet will be precisely $A^n\beta'$ and the endpoint concerned is the image of the finite endpoint of β' under A^n. Thus $A^n\beta'$ has the same endpoints as a lift of a curve in the same homotopy class as $\mathrm{T}^n\beta$; since, in addition, β' is geodesic it must cover $\mathrm{T}^n\beta$. Note that this construction was facilitated very much by the choice of the initial curve β.

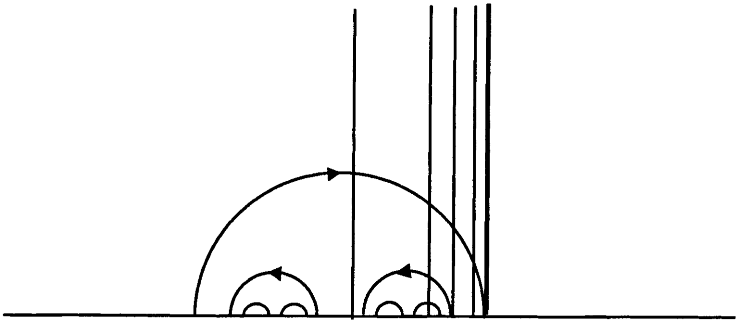

Fig. 7. The figure shows the sequence $T^n\beta$ in the upper half plane. The semi-circles represent lifts of the geodesic in $\Omega(\gamma)$. The bold line on the left is the limit of the $T^n\beta$

The sequence of (finite) endpoints of the β_n' converges to the attracting fixed point of A which is none other than the (finite) endpoint of γ'. Thus β_n' converges to γ' and so, when we take the projection to the quotient surface, $T^n\beta$ converges to a lamination containing γ. □

6. Approximating generic ends

We have now described the set E near points x for which the other end of the geodesic is up a cusp or spirals to a closed simple geodesic. However, these two cases can only account for countably many components of E and we know that the set of components of E is uncountably infinite. Thus we think of the remaining case as being that of the *generic point* of E.

Let x be a point of E which lies on the geodesic γ whose other end spirals to a non-trivial lamination $\Omega(\gamma)$. Lift to H^2 so that the cusp is at ∞ and such that γ lifts to a vertical line. We want to show that this vertical line does not bound a vertical strip which, in the cusp region at infinity, avoids all simple geodesics; this will mean that x is not isolated and moreover cannot be a boundary point. One can use the standard machinery for approximating laminations in [5] to show that γ is not isolated, however, this approach as it stands falls short of showing that it cannot be a boundary point.

Our strategy is as follows. We begin as in the previous section by constructing a "nice" arc. Using this arc we construct two families of simple geodesics, the *joins* and the *double joins*. The technique used to construct these is based fundamentally on the method in [5], however, we introduce a new method to ensure that the sequences we construct are always simple geodesics with both ends up the cusp. We show that, in the cusp region, a subsequence of the joins converges to γ. Finally, by using a contradiction argument involving both the joins and the double joins, we eliminate the possibility that γ represents a boundary point.

Proposition 5. *Let $\gamma, \Omega(\gamma) \subset \mathbf{M}$ be as above. Conjugate Γ so that the vertical line $\gamma' \subset \mathrm{H}^2$ is a lift of γ, and let $\Omega(\gamma)'$ denote the largest subset of H^2 which projects to $\Omega(\gamma)$ on the surface. Then there is an arc β' in H^2 so that $\gamma', \beta', \Omega(\gamma)'$ bound a triangle with 2 ideal vertices as in the diagram. The image, β, of β' on the surface is a simple arc with one end up the cusp and $\beta' \cap \Omega(\gamma)'$ is a Cantor set, and so uncountably infinite. The arc β' can be chosen so it crosses the first leaf that it meets of $\Omega(\gamma)'$ perpendicularly.*

Remark 6.1. The orthogonality condition at the intersection of β' and the leaf of $\Omega(\gamma)$ will greatly facilitate calculations in the sequel.

Proof. The line γ' is asymptotic to a collection of semicircles which are leaves of $\Omega(\gamma)'$. One takes the largest of these semicircles on the left of γ' and define β' to be the vertical line which meets this semicircle at its highest point (as in the proof of Lemma 4). The image of β' on the surface is a simple curve (by an argument similar to that in Lemma 4) and meets $\Omega(\gamma)$ in a single point. Since the injectivity radius of $\mathbf{M}$ is bounded away from 0 at the endpoint of β one can extend it slightly so that it remains a simple arc. Now β meets $\Omega(\gamma)$ transversely in more than one point and, because $\Omega(\gamma)$ is minimal and not a simple geodesic, their intersection is a Cantor set [5]. □

There is an obvious method, using β, γ, for constructing geodesics with both ends up the cusp which we outline now. Since $\Omega(\gamma)$ is in the closure of γ and $\Omega(\gamma)$ meets β uncountably often, γ must meet β (countably) infinitely often. There is a natural ordering on these intersection points inherited from the arclength parameterisation on γ, so label them $y_i, i > 0$. To make a curve with both ends up the cusp one follows γ round till it meets β at y_i then go back up the cusp via β; pull tight to obtain a geodesic – we call this the *join at y_i*. Another method of constructing a curve would be to follow β down to some y_j, then take γ to y_i and then return up the cusp via β; we call this the *double join between y_i and y_j*. Unfortunately, many of the geodesics constructed in these two ways are not simple.

However, one can ensure that a geodesic constructed by the methods above is simple as follows. Choose a point $p \in \beta$ which lies in a suitably

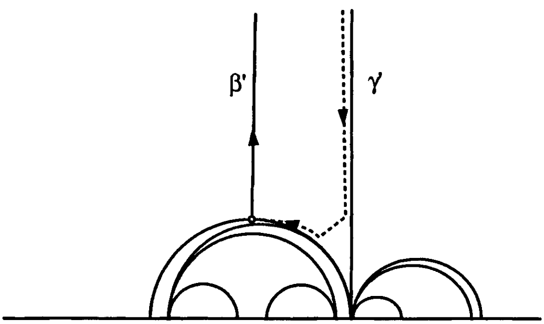

Fig. 8 The diagram shows the lift of the nice arc β we construct in Proposition 5

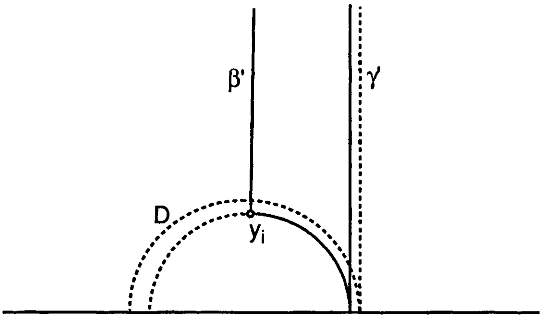

Fig. 9. The lift of a join to H^2 (the heavy black line on the right). The piecewise geodesic curve used in the construction of the join is also shown (the solid curve with the same endpoints on the boundary of H^2)

small cusp region and set b to be the nearest point of $\Omega(\gamma) \cap \beta$ to p, measuring distance in arclength along β. Define a function $\Psi : \gamma \cap \beta \to R^+$ taking as its value the distance between y_i and b measured along β. In this way Ψ is a well defined bounded function and, since γ is simple, it is injective. We say y_j is a *closest approach* iff $\Psi(y_j)$ is less than $\Psi(y_i), i < j$; two closest approaches $y_i, y_j, i < j$ are *consecutive* iff there is no closest approach among the $y_k, i < k < j$.

Lemma 6. *With the notation above*:

1. *If y_i is a closest approach then the join at y_i is a simple geodesic with both ends up the cusp.*

2. *If y_i, y_j are consecutive closest approaches then the double join between them is a simple geodesic with both ends up the cusp.*

Proof. We give a proof only for a join, the double join is dealt with in a similar fashion.

Let y_i be a closest approach. One first sees that the join is always a non-trivial geodesic, that is when one pulls tight the piecewise smooth curve, it does not just vanish up the cusp. Work in H^2 with $\gamma', \beta', \Omega(\gamma)'$ as in Lemma 5. There is a point, y_i', on β' which is a lift for y_i. There is exactly one semicircle, γ_i' say, which passes through y_i' and is a lift of γ. Exactly one endpoint of the arc γ_i' is a cusp point and the vertical line through this point projects to the join through y_i. Thus the join has a lift which is a geodesic in H^2 and so is a non-trivial curve in the surface.

To check that it is simple one proceeds as follows. The piecewise smooth curve used above to define the join consists of a segment of γ and a segment of β; these segments are both simple arcs. Moreover these segments can meet only in y_i – for, if they met in another point, y_j say, then $j < i$. But this point would lie between the cusp and y_i on β; this contradicts the hypothesis that y_i was a closest approach. Thus this piecewise smooth curve is simple and so, being in the same homotopy class, the join is simple too. □

Notation: Henceforth, y_i^* will be the ith closest approach; $\{\beta.\gamma\}_i$ will denote the join of β, γ at y_i^* and $\{\beta.\gamma.\beta\}_i$ will denote the double join between y_i^* and y_{i+1}^*.

We now describe some of the properties of the sequence of closest approaches y_i^* which we will use in the proof that γ is neither isolated nor a boundary point in E.

Lemma 7. *Let y_i^* be the sequence of closest approaches then $y_i^* \to b$ as $i \to \infty$ and the sequence tends to b monotone in the sense that Ψ is monotone decreasing.*

Proof. The limit of the y_i^* exists, essentially since $\Psi(y_i^*)$ is monotone decreasing; let a denote this limit and suppose, for a contradiction, that $a \neq b$. Now by definition b must lie on β between a and the cusp. Since $\Omega(\gamma) \cap \beta$ has no isolated points there is a point of this set between b and a. Thus, by the density of γ in $\Omega(\gamma)$, there is a point of γ between a and b. This is a contradiction. □

We work in the upper half plane and adopt the notation used in Proposition 5. The lift β' meets the semicircle $\mathscr{D}$ (see diagram) to the right of γ'; note that Proposition 5 allows us to take β' perpendicular to $\mathscr{D}$. In this picture the point b lifts to the intersection of β' and $\mathscr{D}$. For each point y of β' the pencil of semicircles through y which avoid $\mathscr{D}$ always contains the semicircle perpendicular to β, and the pencil decreases in size as y approaches b'. Since γ is disjoint from $\Omega(\gamma)$ on $\mathbf{M}$, the lift of γ passing through y_i, γ_i, avoids $\mathscr{D}$ – so γ_i is always in this pencil. We have:

Corollary 12. *Let $\theta(y_i)$ be the angle between β and γ then $\theta(y_i) \to \pi/2$ as $i \to \infty$.*

We are now in a position to show that γ is not isolated in the set of ends. The proof is by construction of a subsequence of joins with limit containing γ. This, however, does not show that the end of γ is not a boundary point; the proof of this comes in Theorem 13.

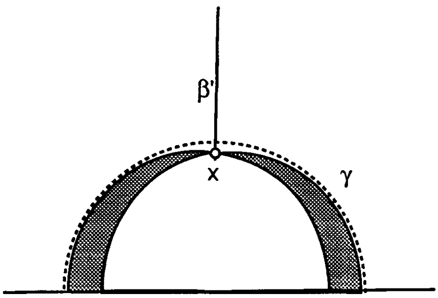

Fig. 10. The pencil of rays disjoint from $\mathscr{D}$

Lemma 8. *The sequence of geodesics which consists of joins through the y_i has a subsequence which limits to a lamination containing γ.*

Proof. We work in $\mathbb{H}^2$ with ∞ a cusp point, choose a vertical line, γ', which is a lift of γ and let y_i' be the point of γ' which covers y_i. The collection y_i' cannot remain within a compact set, since this would imply that a compact geodesic subarc of γ would pass close, say less than $1/4$ the injectivity radius of the thick part of the surface, to itself infinitely often; using this one could construct a set of essential closed loops with length going to 0. Also the y_i' cannot approach ∞, since it is a cusp point and they all lie on a compact lamination. Thus the y_i' converge to the endpoint of γ'. There is a (unique) lift of β, β_i', through y_i'; the vertical line, denoted $\{\beta.\gamma\}_i'$, through the cusp endpoint of β_i' is a lift of the join through y_i. One sees from Corollary 12 that, for sufficiently large i, β_i' is an arc of a semicircle approximately perpendicular to γ'. Since $\beta_i' \cap \gamma' = y_i'$ tends to the finite endpoint of γ' the diameter of the semicircle containing β_i' goes to 0; so the sequence $\{\beta.\gamma\}_i'$ converges to γ'.

On passing to the surface **M** one sees that there is a subsequence of joins with limit a lamination containing γ. □

Finally, we shall need the following:

Lemma 9. *Let $d_\gamma(y_{i+1}, y_i)$ denote the distance between y_i and y_{i+1} measured along γ. Then $d_\gamma(y_{i+1}, y_i) \to \infty$ as $i \to \infty$.*

Proof. Suppose there is a subsequence such that $d_\gamma(y_{i+1}, y_i) < K$ for some $K > 0$. Since $y_i \to b$ the distance between y_i and y_{i+1} tends to 0. One can use the standard techniques discussed in [5] to construct from the portion of γ between y_i and y_{i+1}, a sequence of closed simple geodesics, α_i, converging to $\Omega(\gamma)$. The length of each α_i is bounded by K plus the distance between y_i and y_{i+1} in the surface, and so is bounded independent of i. Thus, since for all $t > 0$ the set of geodesics in **M** of length less than t is finite, the limit of the α_i, i.e. $\Omega(\gamma)$, is a closed geodesic. This contradicts the hypothesis on $\Omega(\gamma)$.

□

Theorem 13. *Let γ be a simple geodesic with a single end up the cusp and its other end spiralling into a minimal lamination which is not a closed geodesic. Then the geodesic γ does not represent a boundary point in E.*

Proof. Let y_i^* be the sequence of closest approaches. We work in H^2 and assume that ∞ is a cusp point for Γ – the covering group of **M**. Conjugate Γ if necessary and so that we have the same picture as in Lemma 8: the lift, γ', of γ is a vertical line and the line β' and the semicircle $\mathscr{D}$ are as before. Let $\{\beta.\gamma\}_i'$ be the vertical line covering the join through y_i^*.

Suppose, for a contradiction, that γ does represent a boundary point in E. This means that there exists an N such that for all $i \geq N$, the vertical lines $\{\beta.\gamma\}_i'$ lie on one side, say the right, of γ'. If we assume that they are all to the

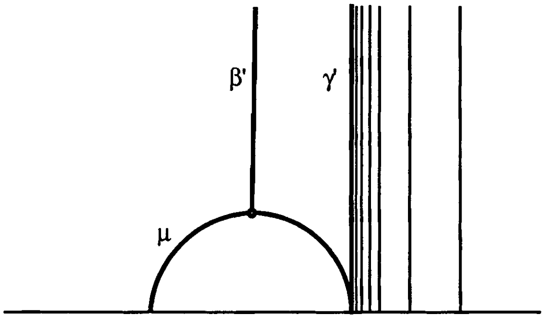

Fig. 11. The figure illustrates the running hypothesis in the proof of Theorem 13: all the lifts of the simple cusp geodesic *eventually* lie to the right of the line γ'

left then it is necessary to go back to Lemma 5 and take β' to the right of γ' and then work through from there.

Consider then the sequence of double joins between y_i^*, y_{i+1}^*. We construct a lift, $\{\beta.\gamma.\beta\}_i'$, of $\{\beta.\gamma.\beta\}_i$ as follows: there is a unique lift of y_i^* on β' and through this point a unique lift of γ which we denote γ_i', we follow this lift round till we encounter the (unique) lift of y_{i+1}^* lying on it; take the (unique) lift of β, β_{i+1}', through this point to a cusp point on the boundary. The vertical line, $\{\beta.\gamma.\beta\}_i'$, with this cusp point as endpoint is a lift of $\{\beta.\gamma.\beta\}_i$.

Claim: the lifts $\{\beta.\gamma.\beta\}_i'$ converge to γ'.

The proof consists of analysing the construction of $\{\beta.\gamma.\beta\}_i'$ a little more carefully. We followed β' to the point y_i^*, this meets the lift of γ, γ_i'; this lift has one endpoint at a cusp point and the other at some non-cusp point in the ideal boundary. Recall that β and γ meet almost at a right angle in y_{i+1}^* so β_{i+1}' is nearly perpendicular to γ_i'. Thus the semicircles which contain β_{i+1}' can be approximated by a sequence of semicircles, P_i, perpendicular to γ_i' and such that the intersection point remains the same i.e. $P_i \cap \gamma_i' = \beta_{i+1}' \cap \gamma_i'$ (by approximated we mean β_{i+1}' is within an ϵ neighborhood of P_i in the Euclidean metric on the upper half plane). Since the distance $d_\gamma(y_i^*, y_{i+1}^*)$ tends to ∞ as $i \to \infty$ (Lemma 9), the lift of y_{i+1}^* on γ_i' (i.e. $P_i \cap \gamma_i'$) approaches the boundary along γ_i' heading in the direction of the non-cusp endpoint. Since, in addition, the diameter of γ_i is bounded below, one can conclude from this that the diameter of P_i tends to 0 as $i \to \infty$. It follows that, given $\epsilon > 0$ the non-cusp endpoint of γ_i' and the cusp endpoint of β_{i+1}' are within 3ϵ (in the Euclidean metric) for i sufficiently large. Note that the latter is none other than the (finite) endpoint of $\{\beta.\gamma.\beta\}_i'$. Thus, to show that $\{\beta.\gamma.\beta\}_i'$ converges to γ', it suffices to show that the sequence of non-cusp endpoints of the γ_i' converges to the endpoint of γ'.

Recall that γ_i' is the lift of γ through $y_i^{*\prime}$; note that there is a unique element $A_i \in \Gamma$ such that $\gamma_i' = A_i\gamma'$. Consider the piecewise geodesic curve, c_i, obtained as follows: start at infinity, travel along γ' to the lift of y_i^*, then along the lift of β to its cusp endpoint. Note that c_i has the same endpoints as $\{\beta.\gamma\}_i'$ and, in particular, both its end points are cusp points in the boundary. Applying A_i, c_i gets mapped to a curve containing a portion of β' and the segment of γ_i' between y_i^* and the cusp point of γ_i'. By examining c_i and $A_i c_i$ one sees that, for i sufficiently large, the cusp point of γ_i' is indeed its endpoint on the left (recall that by hypothesis, for i sufficiently large, the $\{\beta.\gamma\}_i'$, and so c_i, has its finite endpoint to the right of the endpoint of γ'). Since the cusp endpoint is on the left, the non-cusp endpoint is the endpoint of γ_i' on the right. As $i \to \infty$, γ_i' converges to $\mathscr{D}$ and so the sequence of non-cusp endpoints of the γ_i' converge to the endpoint of $\mathscr{D}$ on the right – which is none other than the endpoint of γ'.

Claim: the endpoints of $\{\beta.\gamma.\beta\}_i'$ must eventually lie to the left of γ'; that is there exists N such that for $i > N$ the endpoint of $\{\beta.\gamma.\beta\}_i'$ is to the left of γ'.

Suppose for a contradiction that, for all i sufficiently large, the endpoint of $\{\beta.\gamma.\beta\}_i'$ lies to the right. Thus, in particular, the piecewise smooth lift of $\{\beta.\gamma.\beta\}_i$ introduced above contains an arc β_{i+1}' which has its endpoint to the right of γ'. Clearly $\{\beta.\gamma.\beta\}_i'$ crosses γ'. In fact β_{i+1}' must meet γ' for, since γ is simple in the surface, the portion of γ_i' in $\{\beta.\gamma.\beta\}_i'$ cannot meet γ'. The point $\beta_{i+1}' \cap \gamma'$, projects to one of the y_i on $\mathbf{M}$, say $y_k, k > 0$; note that $\Psi(y_k) < \Psi(y_{i+1}^*)$ since y_k lies between y_{i+1}^* and the cusp along β. We will show below that if $y_{i+1}^* = y_j$ then $k < j$ (recall that we "renamed" the closest intersection points y_n^* and writing $y_{i+1}^* = y_j$ simply amounts to calling this point by its old name). This, of course, leads to a contradiction since y_{i+1}^* is a closest approach.

We will deduce that $k < j$ by estimating how "far away" from the cusp region the two points y_{i+1}^*, y_k are; we make this notion precise now. From the discussion of cusp regions we know that there is a region $\mathscr{H} \subset \mathbf{M}$ which is disjoint from $\Omega(\gamma)$. Fix one such region and choose a point p lying on γ in its interior. For a point $y \in \gamma$ let $d(p, y)$ be the distance from p to y along γ. It is clear, from the definition of y_i given earlier, that $i < j$ iff $d(p, y_i) < d(p, y_j)$.

Let $p_i', y_k', y_i^{*\prime}, y_{i+1}^{*\prime}$ be the lifts on γ_i' of p, y_k, y_i^* and y_{i+1}^* respectively. We estimate the quantity $d(p_i, y_{i+1}^*)$ in a simple way. Leaving p_i' (see diagram)

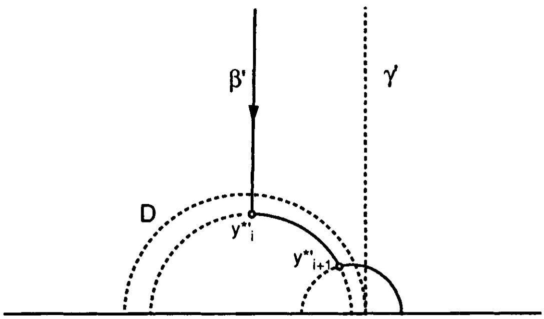

Fig. 12. The figure shows a lift of the piecewise geodesic curve used to construct the doublejoin between y_i and y_{i+1}

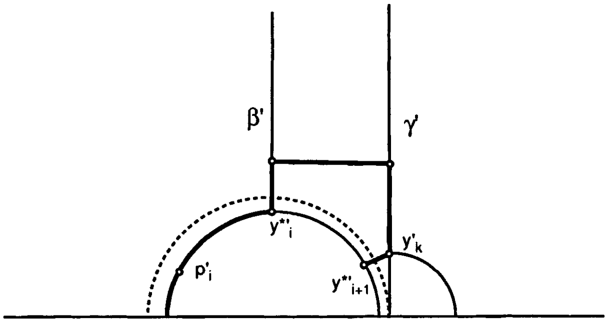

Fig. 13. The figure shows the path used to estimate the quantity $d(p, y^*_{i+1})$

one travels along γ'_i to $y^{*\prime}_i$, then up along β' till we reach the horizontal line tangent to the top of $\mathscr{D}$, we travel along this horizontal, then down γ' to y'_k and finally by β'_{i+1} to $y^{*\prime}_{i+1}$. We show this tortuous path is not much longer than just travelling direct along γ'_i.

Firstly, one has:

$$d(p, y^*_{i+1}) \geq (\text{length of tortuous path}) - K - 2\epsilon \ ,$$

where K is the length of the horocyclic segment (= horizontal line in diagram) and $\epsilon = \Psi(y^*_i)$. this bounds the distance travelled along of the subarcs of β. Secondly, there exists a constant $C > 0$ such that the distance travelled along γ' is not less than $d(p, y_k) - C$; one can take C to be the distance from p' to the horocyclic segment plus the length of the horocyclic segment. Note that this quantity C is independent of i. Thus,

$$d(p, y^*_{i+1}) \geq d(p, y^*_i) + d(p, y_k) - C - K - 2\epsilon \ ,$$

where C, K, ϵ are bounded above, independent of i. Since (see proof of Lemma 8) the y'_i go to the boundary along γ', one has that $d(p, y^*_i) \to \infty$ as $i \to \infty$. Thus the inequality above yields that $d(p, y_k) < d(p, y^*_{i+1}) = d(p, y_j)$ for sufficiently large i. This implies that, for sufficiently large i, $k < j$. Now, since y_k is between y^*_i and the cusp, $\Psi(y^*_{i+1}) = \Psi(y_j) > \Psi(y_k)_i$; so y_k is a closer approach than y^*_{i+1}. This is a contradiction: a closest approach must satisfy $\Psi(y_j) < \Psi(y_n)$ for all $n < j$. $\square$

7. Proof of the structure theorem

Having assembled the 3 theorems 9,11,13 we are in a position to demonstrate the proof of 4. The proofs of the various assertions proceed by eliminating cases using our 3 theorems.

Proof. 1. By Theorem 9 such points are isolated in E, the constructions of Theorem 11 and Theorem 13 eliminate the possibility that geodesics spiraling to simple closed geodesics respectively a nontrivial lamination are isolated.

2. By part 1 above a point of a cusp geodesic cannot be a limit point of E and by Theorem 13 a point on a geodesic spiraling to a nontrivial lamination cannot be in the closure of an open interval lying wholly in the complement of E. Let x lie on a geodesic spiraling to a closed simple geodesic; the construction of Theorem 11 shows that it is a limit point of E. On the other hand given a geodesic which spirals to a closed simple geodesic one can easily find an embedded pair of pants in the surface which contains it. By Theorem 9 this provides the neighborhood U in the complement of E.

3. Let K be a gap and consider the 2 points of E in its closure x, y say. By Theorem 13 neither of x, y lies on a geodesic spiralling to a non-trivial lamination. Also, by Theorem 11 at most one of x, y can spiral to a closed geodesic. Thus, by elimination, we are in the situation of the corollary to Theorem 9.

8. A constant series

With these observations the proof of Theorem 2 is a formality.

Proof. We choose a cusp region and compute its area as the sum of areas of gaps as follows.

The complement of set E in the cusp region is a countable union of gaps. One has associated to each simple cusp geodesic γ a set of 4 gaps; the sum of the areas of these gaps is $2(1+\exp\frac{1}{2}(|\alpha|+|\beta|))^{-1}\times$ (area of the cusp region), where α, β are the geodesics bounding the pants containing the cusp geodesic (proposition Theorem 3). Moreover, by Theorem 4 the union of all such gaps constitutes the complement of E. Since the measure of E is 0 (the Birman Series theorem) the sum of the areas of all gaps is exactly the area of the cusp region. We have on dividing this equation through by the area of the cusp region,

$$\sum_{\gamma} \frac{2}{1+\exp\frac{1}{2}(|\alpha|+|\beta|)} = 1 \ .$$

$\square$

References

[1] W. Abikoff. The real analytic theory of Teichmuller space, Lecture notes in mathematics, 820. Springer, New York, 1980

[2] A.F. Beardon. The geometry of discrete groups. Springer, New York, 1983

[3] J. Birman, C. M. Series. Geodesics with bounded intersection are sparse. Topology, **24**, 217–225 (1985)
[4] B.H. Bowditch. A proof of McShane's identity via Markoff triples. to appear in Bull. London Math. Soc.
[5] R.D. Canary, D.B.A. Epstein, P. Green. Notes on notes of Thurston. In Analytical and geometric aspects of hyperbolic space, LMS Lecture Notes Series 111, pages 3–92. Cambridge University Press, Cambridge, 1987
[6] A.J. Casson, S. A. Bleiler. Automorphisms of Surfaces after Nielsen and Thurston, LMS Student Texts 9. Cambridge University Press, Cambridge, 1988
[7] O. A. Goodman. Doctoral thesis. Ph.D. thesis, University of Warwick, 1989
[8] A. Haas. Diophantine approximation on hyperbolic Riemann surfaces. Acta Mathematica, **156**, 33–82, (1986)
[9] W. de Melo, J. Palis. Geometric theory of dynamical systems. Springer, 1982
[10] S. P. Kerckhoff. The Nielsen realisation problem. Annals of Math, **117**, 235–265, (1983)
[11] G. McShane. Doctoral thesis. Ph.D. thesis, University of Warwick, 1991
[12] G. McShane. Homeomorphisms which preserve simple geodesics. preprint, University of Warwick, 1993
[13] G. McShane. Weierstrass points and simple geodesics. preprint, Ecole Normale Superieure de Lyon, 1995
[14] W. P. Thurston The Geometry and Topology of 3-Manifolds. To Appear