TRAIN-TRACKS FOR SURFACE HOMEOMORPHISMS

M. BESTVINA and M. HANDEL

(Received 15 February 1992; in revised form 12 October 1993)

0. INTRODUCTION

0.1. Overview

The purpose of this paper is to present an algorithmic proof of Thurston’s classification theorem [8, 3, 1, 4, 6, 7] for surface homeomorphisms. It is a modification of the proof ([2, Section 1]) of the analogue of Thurston’s theorem for irreducible automorphisms of free groups. This proof is constructive (it neither requires passing to the universal cover, nor does it use limiting constructions of measured foliations), and one can program a computer to find an invariant train-track for a homeomorphism given, say, as a composition of Dehn twists. In particular, one can effectively decide whether a mapping class is pseudo-Anosov. The given homeomorphism is described by its action on a spine of the surface. We formalize this in Section 1, where we introduce the notion of a fibered surface carrying the homeomorphism. Section 2 discusses moves which improve the efficiency of a fibered surface. The efficiency is measured by the growth rate of the transition matrix (for the induced map on the spine of the surface). Section 3 describes the algorithm that attempts to find the most efficient fibered surface carrying a homeomorphism isotopic to the given one, in the case that the surface has one puncture, and fails only after discovering a reduction. The modifications in the general case are described in Sections 4 (several punctures) and 5 (no punctures). We also explain how to construct an invariant train-track (in the sense of Thurston), the stable and unstable foliations, and a Markov partition for a homeomorphism isotopic to one carried by an efficient fibered surface (again, the construction fails only after discovering a reduction). The proof relies on elementary Perron–Frobenius theory, which we briefly review below. We have chosen to deal with punctured surfaces, rather than compact surfaces with boundary which would require a straightforward variation in our exposition. Section 6 contains 3 examples which illustrate the algorithm.

0.2. Thurston’s theorem

Let $S$ be a closed surface with finitely many (maybe no) punctures (= distinguished points) such that $S_0 = S - \{\text{punctures}\}$ has negative Euler characteristic, and let $f: S \to S$ be a homeomorphism permuting the punctures. Recall that $f$ is periodic if $f^n = \text{identity}$ for some $n > 0$, and it is reducible if there is an $f$-invariant closed 1-manifold $J \subset S_0$ whose complementary components in $S_0$ have negative Euler characteristic or else are Möbius bands. We refer to $J$ as a reduction of $f$. Finally, $f$ is pseudo-Anosov if there is a number $\lambda > 1$ and a pair $\mathcal{F}^s, \mathcal{F}^u$ of transverse measured foliations with singularities modelled on $k$-prongs, $k = 1, 2, \ldots$ (Fig. 1) such that $f(\mathcal{F}^s) = (1/\lambda)\mathcal{F}^s$ and $f(\mathcal{F}^u) = \lambda\mathcal{F}^u$. Furthermore, the one-prong singularities of these foliations are allowed to occur only at the punctures.
Remark 0.2.1. In practice, we will detect reducibility by finding an $f$-invariant compact subsurface $S'$ of $S_0$ so that

1. no component of $S_0 - S'$ is a disk,
2. for no component $S''$ of $S'$ is the inclusion $S'' \subset S_0$ a homotopy equivalence, and
3. $S'$ contains a loop that is not homotopic within $S_0$ into a small neighborhood of a puncture.

Indeed, it easily follows from (1)–(3) that not all components of the invariant 1-manifold $\partial S'$ bound a disk or a once-punctured disk. Remove the ones that do, the remaining 1-manifold is still invariant. Finally, replace each parallel family of curves by a curve in the family, maintaining invariance, to obtain the desired reduction of $f$. By abuse of terminology, we also refer to $S'$ as a reduction for $f$.

Theorem 0.2.2 (Thurston [9]). Every homeomorphism $f: S \to S$ is isotopic rel punctures to one that is either periodic, or reducible, or pseudo-Anosov.

If $f$ is reducible, we can cut $S$ along the invariant 1-manifold. We can apply Theorem 0.2.2 to all components and their first return maps (where we crush all boundary components to punctures). The canonical representative for $f$ can be obtained by piecing these together. This yields a more detailed statement of the classification theorem.

Theorem 0.2.3 (Thurston [9]). Every homeomorphism $f: S \to S$ is isotopic rel punctures to a homeomorphism $f'$ with the following property. There is a collection $\mathcal{C}$ (maybe empty) of pairwise disjoint, non-parallel, non-peripheral simple closed curves such that $f'$ leaves invariant the union of disjoint regular neighborhoods of curves in $\mathcal{C}$, and such that the first return map on each complementary component is either periodic, or pseudo-Anosov.

0.3. Perron–Frobenius theory [5]

Let $M$ be a square matrix with non-negative integer entries. $M$ is reducible if there is a permutation of the index set such that it assumes a triangular block form

$$
\begin{pmatrix}
A & B \\
0 & C
\end{pmatrix}
$$

Otherwise $M$ is irreducible. A non-negative irreducible matrix $M$ has a unique positive eigenvector (up to scale), and the associated eigenvalue $\lambda$ equals the spectral radius of $M$, we say that $\lambda$ is the growth rate of $M$. The reason for this terminology is that the entries of $M^k$
grow like \((\text{const}) \times \lambda^k\) as \(k \to \infty\) (if a power of \(M\) is reducible, this is true only after averaging appropriately). If \(\lambda = 1\), then \(M\) is a (cyclic) permutation matrix, i.e. it has the form
\[
\begin{pmatrix}
0 & 0 & \cdots & 0 & 1 \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 \\
\end{pmatrix}
\]
after permuting the index set if necessary. Otherwise \(\lambda > 1\) and for every \(i, j\) there is \(n > 0\) such that the \(ij\)-entry of \(M^n\) is arbitrarily large. Any reducible matrix can be put in the canonical block form \((M_{ij})\) where \(M_{ij} = 0\) for \(i > j\) and \(M_{ii}\) is irreducible or a zero matrix. The growth rate of \(M\) is the largest growth rate of the \(M_{ii}\)'s (where a zero matrix has growth rate 0).

1. FIBERED SURFACES

Definition 1.1. A fibered surface is a compact surface \(F\) with boundary which is decomposed into arcs and into polygons that are modelled on \(k\)-junctions, \(k = 1, 2, 3, \ldots\). The components of the subsurface fibered by arcs are strips (Fig. 2).

Shrinking the decomposition elements to points produces a graph \(G\), where vertices (of valence \(k\)) correspond to \((k\)-junctions and strips to edges. We can think of \(G\) as being embedded in \(F\), representing the spine of \(F\). Let \(S\) be a closed surface with non-empty set of punctures, and let \(f: S \to S\) be a homeomorphism permuting the punctures.

Definition 1.2. A fibered surface \(F = S_0(- S - \{\text{punctures}\})\) carries \(f\) provided \(F \cong S_0\) is a homotopy equivalence, \(f\) maps each decomposition element of \(F\) into a decomposition element, and each junction into a junction; in particular, \(f\) maps \(F\) into \(F\).

If \(F\) carries \(f\), then \(f\) induces a map \(g: G \to G\) which sends vertices to vertices, and each edge to an edge-path. Form the transition matrix \(M\) whose entry \(m_{ij}\) equals the number of times that the edge-path \(g\) (\(j\)th edge) crosses the \(i\)th edge in either direction. The growth rate of \(g\) (and of \(f: F \to F\)) is the growth rate of \(M\), and we denote it \(\lambda(F, f)\).

Definition 1.3. Assuming that \(g\) does not collapse any edges, there is an induced map \(Dg = \text{derivative of } g\) defined on the disjoint union \(\mathcal{L} = \bigcup \{Lk(v, G) | v \text{ is a vertex of } G\}\) of links of vertices in \(G\) (which can also be thought of as the set of oriented edges in \(G\)) as follows: if \(d \in Lk(v, G)\) is the point corresponding to an edge \(a\) emanating from \(v\), let \(Dg(d) \in Lk(g(v), G)\) be the point determined by the path \(g(a)\) emanating from \(g(v)\); i.e. if \(e\) is an edge then the first edge crossed by \(g(e)\) is \(Dg(e)\). We say that two elements of \(\mathcal{L}\) are

\[\begin{array}{c}
\text{1-junction} \\
\text{2-junction} \\
\text{3-junction}
\end{array}\]

Fig. 2.
equivalent if they come from the same vertex in $G$ and get identified by some power of $D_g$. The equivalence classes are gates.

Definition 1.4. We refer to a collection of points in the link $Lk(v, G)$ of a vertex $v$ in $G$ as connected, provided that in the natural cyclic order (in the non-orientable case it should be called the dihedral order) $l_1, l_2, \ldots, l_k$ of the link, if $l_i$ and $l_j$ both belong to the collection and $i < j$, then either $l_{i+1}, \ldots, l_{j-1}$ are in the collection or $l_{j+1}, \ldots, l_k, l_1, \ldots, l_{i-1}$ are in the collection.

For example, it is easy to see that gates are connected.

2. THE MOVES

Fibered surfaces are not unique. In this section we describe several ways to modify a given fibered surface carrying a homeomorphism $f$. Some of the moves require an isotopy of $f$. These moves are all analyzed in detail in [2, Section 1]. Moves 1, 5, and 6 do not affect the growth rate. Moves 2 and 4 never increase growth rate and often decrease it. Move 3 can increase or decrease growth rate.

2.1. Collapsing an invariant forest

Suppose the induced map $g: G \to G$ has an invariant forest, i.e. a subgraph each of whose components is contractible. The homeomorphism $f' = f$ and the fibered surface $F' = F$ are unchanged, and the decomposition of $F'$ is obtained from the one on $F$ by declaring the subsurface corresponding to each tree in the forest a new junction. The graph $G'$ associated with $F'$ is obtained from $G$ by collapsing each component of the invariant forest to a point. Each edge $E$ of $G'$ can be thought of as an edge of $G$ and the edge path $g'(E)$ is obtained from the edge path $g(E)$ by removing all occurrences of edges in the invariant forest.

2.2. Valence 1 isotopy

Suppose that $G$ has a valence 1 vertex. Change $f$ to $f'$ by an isotopy with support contained in the union of the strip $E$ corresponding to the edge incident to the valence one vertex with the two adjacent junctions so that $f'(F)$ misses the strip $E$. Finally, let $F'$ be $F$ with the strip and the 1-junction removed. The graph $G'$ associated with $F'$ is obtained from $G$ by removing the valence one vertex and the edge that is incident to it. Each edge $E$ of $G'$ can be thought of as an edge of $G$ and the edge path $g'(E)$ is obtained from the edge path $g(E)$ by removing all occurrences of the edge incident to the valence one vertex.

2.3. Valence 2 isotopy

Suppose that $G$ has a valence 2 vertex with adjacent edges $E_i$ and $E_j$. Choose one of these two edges, say $E_i$, and postcompose $f$ by an isotopy which is supported in the union of the strip corresponding to $E_i$ and the two junctions incident to $E_i$, and which pushes the 2-junction through the strip into the other adjacent junction. Note that after this isotopy no junction of $F$ maps into the 2-junction. Now foliate the 2-junction into arcs. The graph $G'$ associated with $F'$ is obtained from $G$ by removing the valence two vertex and amalgamating $E_i$ and $E_j$ into a single edge labelled $E'_j$. Each edge $E$ of $G'$ other than $E'_j$ can be thought of as an edge in $G$ and the edge path $g'(E)$ is obtained from the edge path $g(E)$ by removing all occurrences of $E_i$ and changing each $E_j$ to $E'_j$. The edge path $g'(E_j)$ is obtained from the concatenation $g(E_i) \cdot g(E_j)$ by removing all occurrences of $E_i$ and changing each $E_j$ to $E'_j$. 
2.4. Pulling tight

There are two ways to pull tight. First, if \( v \) is a vertex and if \( Dg(E_i) \) is independent of the choice of edge \( E_i \) initiating at \( v \), then postcompose \( f \) by an isotopy with support contained in the strip corresponding to \( Dg(E_i) \) and in the two junctions adjacent to this strip to remove the initial occurrence of \( Dg(E_i) \) in each edge path \( g(E_i) \). After performing this operation a finite number of times, working on "innermost" vertices first, we may assume that \( Dg \) is not constant at any vertex.

Second, suppose that \( g \) sends an edge \( E \) of \( G \) to an edge-path that backtracks; i.e. that the edge path \( g(E) \) has a subpath of the form \( E_iE_i \). Postcompose \( f \) by an isotopy (Fig. 3) to shorten the length of \( g(E) \) by removing the subpath \( E_iE_i \). After performing this operation a finite number of times, working on "innermost" edges first, we may assume that \( g \) sends every edge of \( G \) to an edge-path that does not backtrack. Note that if we perform these tightening operations as much as possible, then the resulting map on \( G \) is independent of the order in which we do the tightening and we can ignore considerations of which edges and vertices are innermost.

2.5. Folding

Stallings [8] introduced folding and used it successfully to obtain several results about free groups and their homomorphisms.

Suppose we have a connected collection of points in the link of a vertex in \( G \) such that the corresponding collection of edges maps by \( g \) to the same edge-path which does not backtrack. Thus, \( f \) sends the corresponding strips to "parallel" bands. Assume that \( f(F) \) does not come between these bands (we can always arrange this by an isotopy of \( f \)). We leave \( f' = f \) unchanged, while naturally enlarging \( F \) to \( F' \) so that all strips in \( F \) corresponding to the edges under consideration are covered by a single strip in \( F' \). The new graph \( G' \) is obtained from \( G \) by identifying all edges in the collection, say \( E_{i1}, \ldots, E_{ik} \), to a single edge \( E^* \). This operation on \( G \) is called folding. Each edge \( E \) of \( G' \) (including \( E^* \)) can be thought of as an edge of \( G \) and the edge path \( g'(E) \) is obtained from the edge path \( g(E) \) by replacing each occurrence of \( E_{ij} \) with \( E^* \).

2.6. Subdivision

If \( v \) is a point in the interior of an edge \( E_i \) such that \( g(v) \) is a vertex, we can declare \( v \) to be a new vertex, and introduce a 2-junction in the corresponding strip of \( F \). The graph \( G' \) is obtained from \( G \) by replacing \( E_i \) by a pair of edges \( E'_i \) or \( E''_i \) with \( v \) as a common vertex. Each edge \( E \) of \( G' \) other than \( E'_i \) or \( E''_i \) can be thought of as an edge in \( G \) and the edge path \( g'(E) \) is

![Fig. 3.](image-url)
obtained from the edge path $g(E)$ by replacing each occurrence of $E_i$ with the concatenation $E_i \cdot E'_i$. Replacing each occurrence of $E_i$ in $g(E_i)$ with $E_i E'_i$ yields $g'(E'_i)$. Similarly, if $v_1, v_2, \ldots, v_k$ is an $f$-invariant collection of points in $G$, we can make them all vertices of valence 2.

3. ONE PUNCTURE

3.1. Definitions and statements

Throughout this section we assume that $S$ has one puncture. Let $f: S \to S$ be an orientation-preserving homeomorphism carried by a fibered surface $F$, let $g: G \to G$ be the induced map on the graph $G$, and let $M$ be the transition matrix.

Definition 3.1.1. Maps $g: G \to G$ and $f: F \to F$ are irreducible if $M$ is irreducible and $G$ has no valence 1 or 2 vertices.

Note that if $g$ is irreducible then $G$ has no $g$-invariant subgraphs and the size of $M$ is bounded by $3 \times \text{rank } H_1(S_0) - 3$.

Lemma 3.1.2. Suppose that $g: G \to G$ and $f: F \to F$ are irreducible. Then the following are equivalent:

1. Every iterate of $g$ sends each edge to an edge-path that does not backtrack.
2. For every edge $E$ in $G$ the edge-path $g(E) = E_1 E_2 \cdots E_k$ does not backtrack and it determines distinct gates each time it passes through a vertex, i.e. for $i = 1, 2, \ldots, k - 1$ the points in $\mathcal{L}$ corresponding to the terminal point of $E_i$ and the initial point of $E_{i+1}$ are not equivalent.

Proof of Lemma 3.1.2. The key observation is that $g$ sends each edge to an edge-path that does not backtrack if and only if $g$ is locally one-to-one, except perhaps at vertices. If $g$ satisfies this property but $g^k$ does not, then there must be a point $x$ in the interior of an edge $E$ such that $g^i(x) = v$ is a vertex for some $1 \leq i \leq k - 1$ and such that $g^{k-i}$ is not one-to-one on any neighborhood of $v$ in the edge path $g^i(E)$. Replacing $x$ by $g^{k-i}(x)$, we may assume that $i = 1$. Thus, (1) and (2) are equivalent.

We say that $g: G \to G$ is efficient if it satisfies conditions (1) and (2) of Lemma 3.1.2. Note that by condition (2), efficiency can be checked in a predictable number of steps.

Efficient maps on graphs are called train-track maps in [2]. Since this notion is different from Thurston's in the context of surface homeomorphisms, we choose a different name, and discuss the relationship with (Thurston's) train-tracks in Section 3.3. We prove the following two statements that together imply Theorem 0.2.2 for once punctured surfaces.

Theorem 3.1.3. Every homeomorphism of a once punctured surface $S$ is isotopic rel puncture to one which is either reducible or carried by an efficient fibered surface.

Theorem 3.1.4. If a homeomorphism $f: S \to S$ of a once punctured surface is carried by an efficient fibered surface, then $f$ is isotopic rel puncture to a homeomorphism which is either periodic, or reducible, or pseudo-Anosov.

The proof of Theorem 3.1.3 is in Section 3.2, and the proof of Theorem 3.1.4 is in Sections 3.3 and 3.4.
3.2. The algorithm

This entire section is devoted to a proof of Theorem 3.1.3. The algorithm we describe is applied in Example 6.1. Start with any fibered surface $F$ that carries the given homeomorphism $f: S \to S$. For example, $F$ can be taken to be a regular neighborhood of a spine $G$ of $S$, with the natural fibered surface structure. Then isotop $f$ so that the vertices of $G$ map into junctions of $F$, and so that the edges of $G$ map into $F$ transversely to the arc fibers. Finally, by a further isotopy, arrange that decomposition elements of $F$ map to decomposition elements.

We denote the growth rate of the induced map $g: G \to G$ by $\lambda = \lambda(F,f)$. We wish to modify $f$ and $F$ so that the new fibered surface carrying the new homeomorphism is efficient. Our measure of how close we are to an efficient fibered surface is the growth rate: the smaller the growth rate, the more efficient the carrier. Steps (1)-(5) are designed to modify $f$ and $F$ to find a reduction or an irreducible fibered surface carrying $f$. During these steps, $\lambda(F,f)$ does not increase [2, Section 1].

1. If $g$ is not tight, we can pull tight until this is rectified.
2. If $g$ has an invariant, non-trivial forest, collapse it, and repeat until there are no non-trivial invariant forests. If the new map is not tight, go back to (1).
3. If $G$ has valence 1 vertices, perform valence 1 isotopies to remove all of them. If the map on the new graph is not tight, or has a non-trivial invariant forest, go back to (1) and (2).
4. If $G$ contains a non-trivial invariant subgraph $G_0$, we argue that $f$ is isotopic to a reducible homeomorphism. In this case the algorithm stops. Note that $f$ maps the corresponding subsurface $F_0$ into itself. No component of $F_0$ is a disk. Since there are no valence one vertices, $G_0$ is not homotopy equivalent to $G$ and hence $F_0$ is not homotopy equivalent to $F$. The boundary curve of $S_0$ is represented in $G$ by a circuit that must go through every edge of $G$ (this is where we use the fact that $S$ has only one puncture). Consequently, $F_0$ is not an annulus parallel to the puncture. Since $f(F_0) = F_0$ is $\pi_1$-injective, by an isotopy we can arrange $f(F_0) = F_0$, and we obtain a reduction of $f$ (see Remark 0.2.1).
5. Suppose $G$ has a valence 2 vertex $v$ with incident edges $e_1$ and $e_2$. The transition matrix for $g$ is irreducible (or else $G$ would have a non-trivial invariant subgraph) and therefore a positive eigenvector assigns a weight $w(e)$ to each edge of $G$, equalling the $e$-coordinate of the eigenvector. If $w(e_1) \geq w(e_2)$ we isotop $f$ across the strip corresponding to $e_1$ and then foliate the 2-junction (valence 2 isotopy). This removes $v$ from $G$, and the growth rate does not increase (it decreases if $w(e_1) > w(e_2)$, see [2, Lemma 1.13]). If the induced map on the new graph is not tight, or if there are invariant subgraphs, go through the appropriate steps above again. Since each of these moves decreases the “complexity” measured as the pair (number of edges in the graph, $\sum_{e \in G} (\text{combinatorial length of } g(e))$), we are guaranteed to stop, either by discovering a reduction of the given homeomorphism as in (4), or by obtaining an irreducible fibered surface $F$ carrying (a homeomorphism isotopic to) $f$, and the growth rate of the induced map does not exceed $\lambda$.
6. If the fibered surface $F$ is efficient, the algorithm stops. Otherwise, we will construct a new fibered surface $F'$ and a homeomorphism $f'$ isotopic to $f$ with $\lambda(F',f') < \lambda(F,f)$ [2, p. 16–18]. Applying steps (1)-(5) if necessary, we may assume that $F'$ is irreducible. Since the set of growth rates of non-negative integer matrices of bounded size is a discrete subset of $[1, \infty)$, the above process is guaranteed to stop in finitely many steps, after discovering either a reduction, or an efficient fibered surface carrying a homeomorphism isotopic to $f$, thus completing the proof of Theorem 3.1.3.
By definition, if $F$ fails to be efficient, there exists $k > 0$, an edge $E$ in $G$ and a point $p$ in the interior of $E$ such that after $k$ iterations by $g$ the edge $E$ folds at $p$, i.e. $g^k$ fails to be injective in a neighborhood of $p$.

If $k$ is chosen as small as possible, then the iterates $v_i = g^i(p)$ are vertices ($i = 1, 2, \ldots$). Subdivide so that $p$ is now a valence 2 vertex. The two points $a, b$ in the link of $p$ determine (by iterating the derivative $Dg$) pairs of points $a_i, b_i$ in the link of $v_i$ ($i = 1, 2, \ldots, k - 1$), and a single point $c$ in the link of $v_k$. We now want to fold, first at $v_{k-1}$, then at $v_{k-2}$, etc. (Fig. 4) and finally at $v_1$ to arrive at a graph map that is not tight. Tightening then reduces the growth rate. To set the stage, we first subdivide a few times.

Consider all points $d_1, d_2, \ldots, d_n$ in the link of $v_{k-1}$ that map to $c$ under $Dg$. They form a connected collection in the cyclic order induced by the surface. Subdivide the corresponding edges $D_1, D_2, \ldots, D_n$ of $G$ if necessary so that each maps to an edge-path $C$ of length 1 (determined by $c$ and therefore equal for all these edges). If $p$ is not an endpoint of any of these edges, fold them to reduce $k$. If $p$ is an endpoint of an edge in this collection, folding would increase the valence of $p$, and we have to avoid that. To fix this, subdivide some more: Iterate $C$ until it maps to an edge-path of length $> 1$ (if that does not happen, $g$ is a homeomorphism, and hence the fibered surface $F$ is efficient with $\lambda - 1$). Then subdivide $C$ and its iterates. Finally, subdivide $D_1, \ldots, D_n$ and then fold as above. The resulting transition matrix is still irreducible and $\lambda$ has not changed.

After performing the above operation $k - 1$ times, we will have a new fibered surface $F'$ that carries $f$ with $\lambda(F', f) = \lambda(F, f)$, but the induced graph $G'$ has an edge $E'$ such that $g'|E'$ is not tight. Tightening now reduces the growth rate.

3.3. From an efficient fibered surface to a train-track

In this section and the next we assume that $F$ is an efficient fibered surface carrying $f$. If the growth rate $\lambda(F, f) = 1$, the homeomorphism carried by $F$ is easily seen to be isotopic to a periodic one. In the remaining part of the discussion we assume that $\lambda(F, f) > 1$. We construct an invariant train track $\tau$ for $f$, or else discover a reduction.

Inside each junction $J$ draw a small circle $C_J$ (Fig. 5), and to each gate $\gamma$ at $J$ assign a point $p(\gamma)$ on the circle, in the correct cyclic order. The train track $\tau$ will have two types of edges: the ones outside the $C_J$'s (real edges) and the ones inside (infinitesimal edges). The reason for this terminology is explained below.

There is a real edge for each edge of $G$; each endpoint is the appropriate $p(\gamma)$. These edges intersect the $C_J$'s only at the endpoints, and these intersections are orthogonal. We do not distinguish between real edges of $\tau$ and the edges of $G$.
Join \( p(y_1) \) and \( p(y_2) \) by an infinitesimal edge in \( \tau \) if and only if for some edge \( e \) of \( G \) and some \( m > 0 \) the edge-path \( g^m(e) \) enters \( J \) through \( y_1 \) and then exits through \( y_2 \). This edge is contained in the disk bounded by \( C_J \); it intersects \( C_J \) and the other edges only at the endpoints (Fig. 5), and the intersections with \( C_J \) are orthogonal. The map \( g \) induces a map on \( \tau \): each real edge \( e \) is sent to the edge-path \( g(e) \) where between every two edges we insert the appropriate infinitesimal edge of \( \tau \). Each infinitesimal edge is sent to another infinitesimal edge, determined by \( g \) (since distinct gates are mapped to distinct gates). The transition matrix for this map has the form

\[
M' = \begin{pmatrix} N & A \\ 0 & M \end{pmatrix}
\]

where \( N \) is the transition matrix restricted to the infinitesimal edges, and \( M \) is the transition matrix on the real edges (which equals the transition matrix on \( G \)). Notice that each column of \( N \) and of any positive power of \( N \) has one non-zero entry, and it equals one. In particular, no eigenvalue of \( N \) is greater than 1 (they are all roots of unity or 0), and the matrix \( I - (1/\lambda)N \) is invertible, with inverse \( I + (1/\lambda)N + (1/\lambda^2)N^2 + \ldots \).

**Proposition 3.3.1.** The train-track \( \tau \) constructed above does not change if we replace the homeomorphism \( f \) (and the induced map \( g \)) by a power. Furthermore, the infinitesimal edges do not cross, and \( \tau \) can be viewed as being embedded in the surface.

**Proof.** It is clear that the gates do not change after taking a power. Let \( e \) be an infinitesimal edge. For some \( m > 0 \) and some real edge \( e \) the edge-path \( g^m(e) \) crosses \( e \). Let \( k > 0 \) and select a real edge \( e' \) such that the edge-path \( g^k(e') \) crosses \( e \). Then \( g^k(e') \) crosses \( e \), so \( e \) is an infinitesimal edge for \( g^k \). Now suppose that two infinitesimal edges \( e_1 \) and \( e_2 \) cross. Let \( e_i (i = 1, 2) \) be edges of \( G \) such that the edge path \( g^{m_i}(e_i) \) crosses \( e_i \) for some \( m_i > 0 \) \((i = 1, 2)\). Find edges \( e'_i \) of \( G \) so that \( g^{m_i}(e'_i) \) crosses \( e_1 \) and \( g^{m_i}(e'_i) \) crosses \( e_2 \). However, then the paths \( f^{m_i m_2}(e'_1) \) and \( f^{m_1 m_2}(e'_2) \) would not have disjoint interiors, contradicting the fact that \( f \) is an embedding (if \( e'_1 = e'_2 \) the conclusion is that \( f^{m_i m_2} \) is not an embedding when restricted to the interior of \( e'_1 = e'_2 \)).

**Proposition 3.3.2.** If there is a component of \( S - \tau \) that is not a disk (or equivalently if the union of the infinitesimal edges inside some \( C_J \) fails to connect all the gates at \( J \)), then \( f \) is isotopic to a reducible homeomorphism.
Proof. In this case the regular neighborhood \( N \) of \( \tau \) has at least one essential non-peripheral boundary component. By an isotopy we can arrange that \( f(N) = N \) giving us a reduction.

On the other hand, if \( S - \tau \) is a union of disks (one of which contains the puncture, and the others are “infinitesimal polygons”, i.e. each is contained inside some \( C_J \)), \( \tau \) is what Thurston calls an invariant train-track for \( f \). The fact that \( S - \tau \) consists of disks corresponds to the statement that \( \tau \) “fills” the surface. By construction, any two edges incident to the same vertex of \( \tau \) are tangent, i.e. they form an angle of either 0 or \( \pi \). Therefore, each point of \( \tau \) has a naturally defined 1-dimensional tangent space, and the induced map \( \tau \to \tau \) (which we also call \( g \)) is an immersion, since the efficiency of \( g: G \to G \) (see Definition 3.1.2(1)) guarantees that \( \pi \)-angles are mapped to \( \pi \)-angles (care must be taken to make the map smooth, and in particular the map can no longer be thought of as linear on each edge). This is now a Thurston train-track for \( f \). In the following section we argue that \( f \) is (isotopic to) a pseudo-Anosov homeomorphism.

We conclude this section by analyzing the structure of the set of the infinitesimal edges. In Propositions 3.3.3–3.3.5 we assume that \( F \) is an irreducible fibered surface carrying a homeomorphism \( f \) with \( \lambda(F,f) > 1 \), and that each component of \( S - \tau \) is a disk.

**Proposition 3.3.3.** There are at least two gates at each junction in \( F \). Furthermore, \( f \) induces a permutation of the set of junctions in \( F \) with at least three gates. More precisely, if \( f(J) \in J' \) and there are at least three gates at \( J \), then \( J \) and \( J' \) have the same number of gates; \( f \) induces a 1–1 correspondence between the infinitesimal edges in \( C_J \) and \( C_{J'} \), and no other junction with at least three gates maps into \( J' \).

Proof. A junction could not have a single gate, for otherwise a path of the form \( f^n(e) \) that enters the junction through this gate would have no gates to exit the junction. Now suppose \( f(J) \subseteq J' \), \( J \) has \( k \geq 3 \) gates, and no junction has more than \( k \) gates. In particular, no junction with more than \( k \) gates maps to \( J' \). Since distinct gates map to distinct gates, it follows that \( J' \) must have \( k \) gates as well. Note that if \( f(J_0) \subseteq J' \), then either \( J_0 = J \) or \( J_0 \) has two gates, for else there would not be enough room (see also the proof of Proposition 3.3.4) in \( J' \) for \( f/F \) to be an embedding. This can be proved by considering the components of \( f(F) \) (disk bounded by \( C \) that contains \( f(J) \)). In particular, \( f \) induces a permutation of junctions with \( k \) gates. Clearly, \( f \) induces an injective map from the set of infinitesimal edges of \( J \) to those of \( J' \). Since a power of \( f \) fixes the gates of \( J \), it follows that \( J \) and \( J' \) have the same number of infinitesimal edges, and hence \( f \) induces a 1–1 correspondence between the infinitesimal edges in \( C_J \) and \( C_{J'} \).

We now repeat this argument for the junctions with \( k_1 > 3 \) edges, where \( k_1 < k \) is the largest possible integer, etc.

**Proposition 3.3.4.** Let \( J \) be a junction of \( F \) with \( k \) gates. Then one of the following holds (see Fig. 6):

1. \( k = 2 \), there is a unique infinitesimal edge in \( C_J \), and it joins the two gates.
2. \( k > 2 \) and there is one infinitesimal edge joining each pair of adjacent gates, so that the infinitesimal edges in \( C_J \) form a \( k \)-gon.
3. \( k > 2 \) and there is one infinitesimal edge joining each pair of adjacent gates with one exception, so that the infinitesimal edges in \( C_J \) form a \( k \)-gon with one side missing.

Proof. Suppose \( k > 2 \). Replacing \( g \) by a power if necessary, we can assume that \( f \) fixes all infinitesimal edges in \( C_J \) and that it reembeds the train-track \( \tau \) in a small regular neighbor-
Fig. 6.

hood of \( \tau \). If three infinitesimal edges were incident with the same gate, no edge path \( g^m(e) \) that crosses the middle one would be able to exit through this gate. It follows that at most two infinitesimal edges can be incident to each gate. Similarly, we argue that the "N" pattern is impossible, i.e. four gates \( \gamma_i, i = 1, 2, 3, 4 \), arranged cyclically on \( C \), so that the infinitesimal gates connect \( \gamma_1 \) with \( \gamma_2 \) and \( \gamma_3 \) with \( \gamma_4 \): an edge-path \( g^m(e) \) that crosses the infinitesimal edge \( e \) connecting \( \gamma_4 \) and \( \gamma_5 \) can go through the gate \( \gamma_1 \) only if the path \( f^m(e) \) passes on the side of \( e \) determined by \( \gamma_4 \), and it can go through the gate \( \gamma_3 \) only if the path \( f^m(e) \) passes on the side of \( e \) determined by \( \gamma_2 \), so this pattern is impossible. Since the number of gates at \( v \) is at least 2, and any two gates are joined by an infinitesimal edge-path, the claim follows.

**Proposition 3.3.5.** Each component of \( S - \tau \) is either a disk with at least three points where the boundary forms angle 0, or else it is a punctured disk with at least one such point.

**Proof.** Each infinitesimal component of \( S - \tau \) is a disk with at least three 0 angle points. by Proposition 3.3.4. The boundary curve \( \gamma \) of the component of \( S - \tau \) containing the puncture cannot be smooth, since all smooth curves increase in length after iteration by \( g \), while \( \gamma \) is fixed. \( \square \)

### 3.4. The Markov partition and the invariant measured foliations

Here we assume that \( f \) is carried by an efficient fibered surface \( F \), that \( \lambda(F, f) > 1 \) and that the complementary components in \( S \) of the invariant train-track \( \tau \) constructed in the previous section are all disks. We construct a Markov partition and the invariant measured foliations for (a map isotopic to) \( f \), thus showing that \( f \) is pseudo-Anosov.

To each edge \( e \) in \( G \) assign a positive number \( w(e) \) (the width) such that for some \( \lambda \) we have that \( \lambda w(e) \) equals the sum of widths (with multiplicities) of all strips that map across \( R_e \). In other words, the vector \( W = [w(e)]_e \), that has a component (equal to \( w(e) \)) for every non-oriented edge \( e \in G \), is a positive eigenvector of the transition matrix \( M \), and is therefore unique up to scale, and \( \lambda = \lambda(F, f) \). We can now uniquely define the widths of infinitesimal edges \( e \) of \( \tau \) by solving for \( X \) the equation

\[
\begin{pmatrix}
N & A \\
0 & M
\end{pmatrix}
\begin{pmatrix}
X \\
W
\end{pmatrix}
= \lambda
\begin{pmatrix}
X \\
W
\end{pmatrix}.
\]

Since \( \lambda > 1 \) is not an eigenvalue of the permutation matrix \( N \), there is a unique solution; it is positive and equals \((1/\lambda)(I + (1/\lambda)N + (1/\lambda^2)N^2 + \ldots)A W \). Furthermore, the switch equations hold, namely at every gate the sum of the widths of the real edges (the outside width)
equals the sum of the widths of the infinitesimal edges (the inside width). This can be seen as follows. Fix a large integer $k$. Then for any edge $e$ (real or infinitesimal)

$$w(e) = \frac{1}{\lambda^k} \sum_{\text{edges } e' \text{ of } \tau} w(e') \left( \# \text{ of times } g^k(e') \text{ crosses } e \right).$$

Since every edge-path $g^k(e')$ enters and exits every gate the same number of times except for the endpoints, the contribution of the sum to the outside width equals the contribution of the sum to the inside width at any gate, except for a bounded amount. Letting $k \to \infty$ yields the result.

Similarly, we can assign a positive number $l(e)$ (the length) to each edge $e$ of $G$ (i.e. real edge of $\tau$) by solving the equation $LM = \mu L$ for $L$. The transpose of $L$ is a positive eigenvector for the transpose of $M$. It follows that the lengths are unique up to scale, and $\mu = \lambda$. To extend this solution to the infinitesimal edges, we are forced to make their lengths equal to 0 (and hence the name).

With each edge $e$ of $G$ we associate a rectangle $R_e = S_e \times U_e$ endowed with the stable foliation whose leaves are $S_e \times \{y\}$, $y \in U_e$, and the unstable foliation whose leaves are $\{x\} \times U_e$, $x \in S_e$ (so the unstable leaves will get stretched by $f$). Each foliation is equipped with a transverse measure. The transverse measure on the (un)stable foliation is induced by the Lebesgue measure on $U_e$ (which is viewed as a segment of length $l(e)$ ($w(e)$)).

Place these rectangles in the surface so that (Fig. 7 picturing a neighborhood of the vertex corresponding to the middle picture in Fig. 6)

1. each component of the stable part $S_e \times \partial U_e$ of the boundary of $R_e$ is contained in the appropriate circle $C_j$,
2. each edge $e$ is isotopic into $R_e$ by an isotopy which keeps the endpoints of $e$ contained in the $C_j$'s,
3. if $e, e'$ are adjacent edges in a gate at $v$ corresponding to a junction $J$, then $R_e \cap C_j$ and $R_{e'} \cap C_j$ intersect in a point, and
4. if $e \neq e'$, then $R_e$ and $R_{e'}$ are disjoint except for the intersection points forced by (3).

Each gate $v$ at $J$ gives rise to a segment $L_v$ (contained in a stable leaf) in $C_J$ whose transverse measure equals $\sum_{e \in v} w(e)$. Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r$ be the infinitesimal edges incident to

![Fig. 7.](image-url)
this gate (by Proposition 3.3.4. we know that \( r = 1 \) or 2 but we will not use this). The switch equation states that \( \sum_{i=1}^r w(e_i) = \sum_{i} w(e_i) \), which together with a natural linear order of the \( e_i \)'s (given by an orientation of \( C_j \)) yields a partition of \( L_n \) into segments \( L_1', L_2', \ldots, L_r' \) with transverse measure of \( L_i' \) equal to \( w(e_i) \). Therefore, an infinitesimal edge \( e \) incident to gates \( y \) and \( y' \) gives rise to segments \( L_i' \) and \( L_j' \) with the same total measure. Identify such segments in a measure-preserving, orientation-reversing fashion (Fig. 7).

The resulting space \( \mathcal{R} \) (union of the rectangles \( R_n \) with identifications described above) is still viewed as a subset of \( S \) and can be thought of as a kind of a thickening of the invariant train-track \( \tau \), with all infinitesimal complementary components filled in, or equivalently as a thickening of \( G = S \). (However, note that \( \mathcal{R} \) is not in general a regular neighborhood of \( G \). For example, if the measure of each segment is 1, then \( \mathcal{R} \) is not a surface. However, generically, \( \mathcal{R} \) is a regular neighborhood of \( G \).) By our assumptions \( S - \mathcal{R} \) is a (punctured) disk.

We now define a map \( \Phi: \mathcal{R} \to \mathcal{R} \). It will map each rectangle by an affine map stretching the unstable foliation by \( \lambda \) and shrinking the stable foliation by \( \lambda \).

Shrinking the stable leaves in \( \mathcal{R} \) to points produces a map \( \pi: \mathcal{R} \to G \). View each edge \( e \) of \( G \) as having length \( l(e) \) so that \( \pi \) is "measure preserving" with respect to the Lebesgue measure in \( G \) and the measure associated to the stable foliation in \( \mathcal{R} \). Also view \( g: G \to G \) as being linear on each edge. If \( g(\epsilon) = \epsilon_1 \epsilon_2 \cdots \epsilon_k \), then \( l(e) = l(\epsilon_1) + l(\epsilon_2) + \cdots + l(\epsilon_k) \), so \( g \) has slope \( \lambda \) on each edge. We now lift \( g \) to \( \Phi: \mathcal{R} \to \mathcal{R} \). A stable leaf \( \pi^{-1}(g(x)) \) is sent into the stable leaf \( \pi^{-1}(g(x)) \). To determine where to place \( \Phi(\pi^{-1}(x)) \), which is an interval of width \( (1/\lambda) \) (width of \( \pi^{-1}(x) \)), as a subinterval of \( \pi^{-1}(g(x)) \), consider all points \( x = x_1, x_2, \ldots, x_m \in G \) that map to \( g(x) \) under \( g \). The fact that \( [w(e_i)]_u \) is an eigenvector for the transition matrix with eigenvalue \( \lambda \) implies that the widths of the \( \Phi(\pi^{-1}(x_i)) \)'s add up to the width of \( \pi^{-1}(g(x)) \). We then stack up the \( \Phi(\pi^{-1}(x_i)) \)'s on \( \pi^{-1}(g(x)) \) so that the adjacent arcs intersect in the common endpoint. The map \( f \) on the fibered surface determines the order in which the intervals \( \Phi(\pi^{-1}(x_1)), \Phi(\pi^{-1}(x_2)), \ldots, \Phi(\pi^{-1}(x_m)) \) are placed in \( \pi^{-1}(g(x)) \). Note that \( \Phi \) is surjective and only \( \partial \mathcal{R} \) maps to \( \partial \mathcal{R} \).

The peripheral curve in \( S_0 \) is represented as \( \partial \mathcal{R} \), and has a structure of a smooth polygon with, say, \( k \) sides. Since \( f \) preserves or flips the peripheral curve (up to isotopy), we see that \( \Phi^{2k} \) (the power is taken to ensure that \( \partial \mathcal{R} \) is not flipped or rotated) maps each side of \( \partial \mathcal{R} \) over itself (Fig. 8), and therefore each side has a unique fixed point. Each vertex of the

Fig. 8.
polygon $\partial R$ is equidistant (in terms of the transverse measure on the stable foliation) from the fixed points on adjacent sides. (Say the distances from a vertex $v$ to the fixed points $x_1, x_2$ on adjacent sides are $d_1, d_2$. The image of the segment $[v, x_i]$ of the unstable leaf under $\Phi^{2k}$ is the segment $[w, x_i] \supset [v, x_i]$ of length $\lambda^{2k}d_i = \text{length}[w, v] + d_i$. It follows that $d_i = \text{length}[w, v]/(\lambda^{2k} - 1)$ and in particular $d_1 = d_2$.) Identify adjacent sides in a length-preserving fashion up through these points. As a result, all of the fixed points get identified to a single point $p$, and the resulting space can be identified with $S$ with $p$ corresponding to the puncture. The (un)stable foliation $\mathcal{F}^u(\mathcal{F}^s)$ on $S$ is obtained by combining the (un)stable foliations in the rectangles. Now $\Phi$ induces a homeomorphism of $S$ isotopic to $f$ (one way to see this is to note that this homeomorphism is homotopic to $g : G \to G \subset S_0$; the homotopy being given essentially by collapsing the unstable leaves), and it maps each rectangle by an affine map stretching the unstable foliation by $\lambda$ and shrinking the stable foliation by $1/\lambda$. The rectangles cover the surface and form a Markov partition (or a rectangle decomposition) for the homeomorphism.

4. SEVERAL PUNCTURES

4.1. Definitions and statements

Throughout this section $S$ will be a closed surface with at least one puncture. If the punctures are cyclically permuted by the homeomorphism, then the argument of Section 3 is easily modified. (Straightforward changes in Algorithm 3.2(4) and Section 3.4 are sufficient.) More generally, when there is more than one orbit of punctures, we cannot expect the transition matrix $M$ to be irreducible: there may be a proper $g$-invariant subgraph $G_0 \subset G$ that has non-contractible components but does not determine a reduction for $f$ because each component of the associated subsurface is an annulus surrounding a puncture. To isolate this problem, we restrict our fibered surfaces as follows.

Suppose that there are $n \geq 1$ orbits of punctures. Select $n - 1$ of them. All fibered surfaces in this section will be required to satisfy the following property.

(10) The graph $G$ contains a $g$-invariant subgraph $P$ (= peripheral subgraph) whose components are circles, representing the peripheral curves in $S_0$ corresponding to the selected punctures. The map $g$ restricts to a (simplicial) homeomorphism of $P$.

We define a pair of associated subgraphs as follows. Let $\text{pre}-P$ be the subgraph of $G$ consisting of edges that are eventually mapped into $P$ (i.e. $E \subset \text{pre}-P$ if and only if $f^k(E) \subset P$ for some $k > 0$) and let $H$ be the subgraph consisting of edges not in $P$ or $\text{pre}-P$. Then $G = P \cup \text{pre}-P \cup H$ and the transition matrix $M$ has the form

$$
M = \begin{pmatrix}
N & A & B \\
0 & C & D \\
0 & 0 & M_H
\end{pmatrix}
$$

where $N$ (respectively, $C, M_H$) is the transition matrix for $P$ (respectively $\text{pre}-P, H$). Since $N$ is a permutation matrix and some iterate of $C$ is the zero matrix, the growth rate $\lambda(F, f)$ of $F$ is the growth rate of $M_H$.

Remark. One can think of the edges of $P$ and of $\text{pre}-P$ as being infinitesimal. When lengths are assigned to the edges of $G$ in order to construct stable and unstable foliations (see Sections 3.4 and 4.4.), the edges of $P \cup \text{pre}-P$ will all be given length zero.
There are two places in the argument of Section 3 that need modification. The first has
to do with making $M_n$ irreducible. In the once punctured case (see Algorithm 3.2(4)), after
performing valence one homotopies and collapsing invariant forests, we either find a reduc-
tion of $f$ or $M$ is irreducible. In the multipunctured case, we must also take into account
invariant subgraphs that deformation retract onto $P$. We remove these by a method (see
Section 4.2.) that is analogous to collapsing an invariant forest. After that, we either find
a reduction for $f$ or $M_H$ is irreducible.

The second modification is more subtle. In the very last step of the algorithm for the
once punctured surface (see Algorithm 3.2(6)), we folded until the graph was not tight, and
then reduced the growth rate by tightening. In the multipunctured case, we must insure that
the edge we are tightening lies in $H$ and that tightening reduces an entry of $M_H$ and not just
of the submatrices $B$ or $D$. This accounts for condition (12) in the definition of irreducibility
given below and for steps 4.3(3) and (6) of the algorithm.

**Definition 4.1.1.** We say that a fibered surface $F$ carrying $f$ with the induced map
$g: G \to G$ is **irreducible** if, in addition to (I0), properties (I1) and (I2) below are satisfied:

(I1) The transition matrix $M_H$ for $H$ is irreducible and has at most $3 \times H_1(S_0) - 3$ rows
and columns.

(I2) If $E_0$ and $E_1$ are distinct-oriented edges with the same initial vertex and if
$Dg(E_0) = Dg(E_1)$, then $Dg(E_0) = Dg(E_1) \subset H$.

**Remark.** If $G$ has no valence one or two vertices, then $M$ has at most $3 \times H_1(S_0) - 3$
rows and columns. Since we are allowing $G$ to have valence two vertices (see Algorithm
4.3(7)), we have added this bound on $M_H$ as part of the definition of irreducibility.

**Remark.** Condition (I2) implies that $g|_{\text{pre-P}}$ is locally injective and hence that each
component of $\text{pre-P}$ is an arc. Moreover, $P$ and $\text{pre-P}$ are disjoint. (It follows from the
definitions that they have no edges in common; the point here is that they have no common
vertices.)

We define **gates** the same way as before. They are the equivalence classes of elements of
$L = \bigcup \{ Lk(v, G) \mid v \text{ is a vertex of } G \}$ where two are equivalent if they come from the same
vertex of $G$ and get identified by some power of $Dg$.

We define efficiency as before: see Lemma 3.1.2.

We again state two theorems that imply Theorem 0.2.2 for the case that $S$ has at least
one puncture. The proof of the first is in Section 4.3 and of the second in Section 4.4.

**Theorem 4.1.3.** Every homeomorphism of a surface $S$ with at least one puncture is isotopic
rel punctures to one which is either reducible, or carried by an efficient fibered surface.

**Theorem 4.1.4.** If a homeomorphism $f: S \to S$ of a surface with at least one puncture is
carried by an efficient fibered surface, then $f$ is isotopic rel punctures to a homeomorphism
which is either periodic, or reducible, or pseudo-Anosov.

### 4.2. Absorbing into $P$

The algorithm for the multipunctured case requires one move that was not described in
Section 2. The following lemma lists the properties of our new move. We say that $F', f'$ and
$g': G' \to G'$ are obtained from $F, f$ and $g: G \to G$ by **absorbing into $P$**. This move is analogous
to the "core subdivision" of [2, p. 42] and to the collapse of an invariant forest.
If \( G_0 \) is any subgraph of \( G \), then we define \( Lk(G_0, G) \) to be the set of oriented edges of \( G \setminus G_0 \) that have initial endpoint in \( G_0 \).

**Lemma 4.2.1.** Given any fibered surface \( F \) and \( g : G \to G \) there exists a homeomorphism \( f' \) that is isotopic to \( f \) and a fibered surface \( F' \) so that the induced map \( g' : G' \to G' \) satisfies the following:

1. The only \( g' \)-invariant subgraph that deformation retracts onto \( P' \) ( = peripheral subgraph of \( G' \)) is \( P' \) itself.
2. \( Dg' \) maps \( Lk(P', G') \) into itself.
3. \( P' \) has no vertices of valence 2 in \( G' \).
4. \( \lambda' \leq \lambda \).
5. The number of edges in \((G' \setminus P')\) does not exceed the number of edges in \((G \setminus P)\).
6. If \( G \) has no valence one vertices, then \( G' \) has no valence one vertices.

**Proof.** Let \( G_0 \) be the maximal \( g \)-invariant subgraph that deformation retracts onto \( P \). The corresponding subsurface \( A \) of \( F \) is abstractly a union of annuli \( A_i \). Each \( A_i \) has an inner (adjacent to a puncture) boundary component and an outer boundary component. The frontier \( Fr A \) of \( A \) consists of arcs in outer boundary components that divide junctions inside \( A \) and strips in \( F \) outside \( A \). In other words, \( Fr A \) is the set of arcs along which strips not in \( A \) are attached to \( A \). There is a bijection between the components of \( Fr A \) and \( Lk(G_0, G) \).

Suppose that \( E \) is an oriented edge in \( Lk(G_0, G) \). Let \( E_0 \) be the maximal initial segment of \( E \) whose \( g \)-image is entirely contained in \( G_0 \). Since \( G_0 \) is assumed to be maximal, \( E_0 \) is a proper, possibly trivial, subpath of \( E \). Let \( U \) be the subsurface of \( F \) determined by the union of \( G_0 \) with all of the \( E_0 \)'s. Thus, \( U \) is a union of components of \( f^{-1}(A) \cap F \) and is obtained from \( A \) by adding proper subsets of each strip of \( F \setminus A \) that is attached to \( A \). Choose an isotopy \( \phi_t \) that expands \( A \) onto \( U \). Postcomposing with \( f \) gives an isotopy \( f \phi_t \), where \( f \phi_t \) maps \( A \) into \( A \) and maps each component of \( Fr A \) into a component of \( Fr A \). Moreover, \( f \phi_t \) maps an initial substrip of \( E \) into \( F \setminus A \) and \( f(E) \) crosses the same strips of \( F \setminus A \) in the same order as does \( f(E) \).

Identify the components of \( Fr A \) with \( Lk(G_0, G) \). Then \( f \phi_1 \) induces a map \( Dg_1 : Lk(G_0, G) \to Lk(G_0, G) \). If \( E \) is an oriented edge in \( Lk(G_0, G) \), then \( Dg_1(E) \) is the first edge of \( G \setminus G_0 \) in the edge path \( g(E) \).

We pause in our proof of Lemma 4.2.1. to state and prove the following lemma.

**Lemma 4.2.2.** There are unique disjoint arcs \( I_{ij} \) in the outer boundary component of \( A_i \) such that

1. Each component of \( Fr A_i \) is contained in some \( I_{ij} \); two components of \( Fr A_i \) are contained in the same \( I_{ij} \) if and only if they have the same image under some iterate of \( Dg_1 \).
2. The endpoints of \( I_{ij} \) lie in \( Fr A_i \).
3. Each \( f_1(I_{ij}) \) is isotopic rel \( Fr A \) into some \( I_{ij} \).

**Proof.** Partition \( Lk(G_0, G) \) into equivalence classes whose elements have the same image under some iterate of \( Dg_1 \). Each equivalence class is connected (in the sense that the corresponding components of \( Fr A \) are adjacent to each other in the outer boundary component of some \( A_i \)).

If there are at least two equivalence classes in \( Fr A_i \) or if there is only one component of \( Fr A_i \), then conditions (1) and (2) determine a unique collection of arcs in \( Fr A_i \). Condition
(3) follows from the fact that the $f_i$-image of the outer boundary component of $A_i$ is an essential (i.e. non-contractible) simple closed curve in some $A_i$.

Suppose then that all of the components $X_1, \ldots, X_r$ of $\text{Fr } A_i$ belong to the same equivalence class. There exists $l > 0$ so that $f_j^l(X_1), \ldots, f_j^l(X_r)$ all lie in the same component of $\text{Fr } A_i$. Let $x_1, \ldots, x_r$ be the complementary components of $X_1 \cup \cdots \cup X_r$ in the outer boundary component of $A_i$. Since the $f_i$-image of the outer boundary component of $A_i$ is an essential simple closed curve in $A_i$, exactly one $f_j^l(x_k)$ is not homotopic rel endpoints into $\text{Fr } A_i$. We choose $I_{ij}$ to be the complement of the interior of $x_k$. The reader will easily check that condition (3) is satisfied.

**Proof of Lemma 4.2.1.** (conclusion). We define $F'$ to be $F$ with a new fibered structure on $A$. The new fibered structure on $A_i$ is a product with the decomposition of the outer boundary component given by the $I_{ij}$'s. More precisely, the $j$th junction of $A_i$ intersects the outer boundary component of $A_i$ in $I_{ij}$ and intersects the inner boundary component in a single interval. The regions in between these junctions are fibered strips. The graph associated with $F'$ is called $G'$ and the subgraph $P'$ associated to $A$ is a union of circles. The vertices of $P'$ correspond to equivalence classes in $Lk(G_0, G)$; if the equivalence class has $k$ elements, then the vertex has valence $k + 2$. In particular, there are no valence two vertices in $P'$. Choose an isotopy $\psi$, with support in $A$ so that $f' = \psi f_j$ induces a map $g' : G' \to G'$.

Lemma 4.2.2 and the fact that the $f_j$-image of the outer boundary component of $A_i$ is an essential simple closed curve in $A$ implies that we may choose $f'$ so that $g'|P'$ is a simplicial homeomorphism.

Property (1) follows from the maximality of $G_0$. Property (2) follows from the fact that $f_j$ maps an initial substrip of $E$ into $F \setminus A$. We have already verified (3). Property (4) follows from the fact that the transition matrix for $(G' \setminus P')$ is a submatrix of the transition matrix for $(G' \setminus G_0)$. Finally, (5) and (6) follow from the fact that $(G' \setminus P')$ is combinatorially equivalent to $(G' \setminus G_0)$.

**Example 4.2.3.** Consider the situation in Fig. 9 in which we draw only the relevant part of $G$. The circle $x \cup \beta$ encloses a puncture. The map $g$ is given by $x = \alpha, \beta = \beta, x = y \ldots, y = \beta \ldots, \beta = \beta \ldots, y = x$ In this case $G_0 = P$. The map $Dg_j$ satisfies $x \mapsto y, y \mapsto z$ and $z \mapsto y$. Thus, $x$ and $z$ form one equivalence class in $Lk(G_0, G)$ and $y$ forms another. The graph $P'$ is a circle with two vertices that is rotated by $\pi$ (because the equivalence classes are permuted). The new edge paths are $g'(x') = \beta', g'(\beta') = x', g'(x') = y' \ldots, g'(x') = y' \ldots$ and $g'(y') = z' \ldots$

Further examples can be found in Section 6.2.

**4.3. The algorithm**

The algorithm for finding an efficient fibered surface for a given mapping class (or else discovering a reduction) is a modification of the one from Section 3.2. Start with any fibered

---

**Fig. 9.**
surface $F$ satisfying (10) and carrying a given homeomorphism $f$. The purpose of steps (1)–(6) below is to replace $f$ by an isotopic homeomorphism $f'$ and $F$ by an irreducible fibered surface $F'$, without increasing the growth rate.

**Remark.** Steps (1)–(5) below are performed in order. If at any stage, a previously achieved property is lost, return to the appropriate step in the algorithm and continue from there. This is guaranteed to stop since the "complexity" pair ($\sum_{e \in G-P}$ (combinatorial length of the edge-path $g(e)$ after removing from $g(e)$ edges in $P$) decreases after each move, except in (3) when it does not increase, but need not decrease. If step (3) is carried out and the complexity pair does not decrease, then step (3) will not be repeated until some other step (which does decrease the complexity pair) is performed.

(1) If $G$ has an invariant forest that does not intersect $P$, collapse it. Tighten if necessary.

(2) If $G$ has valence one vertices, perform valence one isotopies to remove them all. Tighten if necessary.

(3) If $G$ does not satisfy the conclusions of Lemma 4.2.1, absorb into $P$. Tighten if necessary. (Note that if tightening does not decrease the complexity pair, then the conclusion of Lemma 4.2.1 still holds so that (3) is not immediately repeated.)

(4) If $M_H$ is not irreducible, then we argue that $f$ is isotopic to a reducible homeomorphism and so the algorithm stops. There is an invariant subsurface $F_0$ whose components are not contractible and do not deformation retract onto components of $P$. The union of the peripheral curves surrounding the orbit of punctures that is not represented by $P$, can not be represented on any proper subgraph of $G$. Thus, $F_0$ is not a union of annuli parallel to punctures, and we obtain a reduction of $f$ as in Remark 0.2.1.

For all succeeding steps we assume that $M_H$ is irreducible.

(5) If $G$ has valence two vertices, then perform valence two isotopies to remove them all according to the following rules. By Lemma 4.2.1(3), the edges incident to a valence two vertex $v$ do not lie in $P$. If at least one of the edges adjacent to $v$ lies in pre-$P$, then perform the isotopy across the strip corresponding to such an edge. This latter move has no effect on $M_H$ and so does not change $\lambda$. If both edges incident to $v$ lie in $H$ and $\lambda > 1$, then follow the recipe in Algorithm 3.2(5). If $\lambda = 1$, then simply erase all valence two vertices with both incident edges in $H$; the resulting map is still simplicial and $\lambda$ is unchanged. Tighten if necessary.

Note that at this stage $G$ has no valence one or two vertices. Thus, $H$ has at most $3 \times H_0(S_0) - 3$ edges and (11) is satisfied.

(6) Lemma 4.2.1(2) implies that each component of pre-$P$ is disjoint from $P$. Let $P_0 = P$ and inductively define $P_i$ to be the union of the components of pre-$P$ that satisfy $g(P_i) \subseteq P_{i-1}$.

If (12) is not satisfied, then there is some smallest $i \geq 0$, and oriented edges $E_0, E_1$ with the same initial vertex such that $Dg(E_0) = Dg(E_1) \subseteq P_i$. By Lemma 4.2.1(2), $F_0, F_1 \subseteq P_i$. Fold $E_0$ and $E_1$ as much as possible. This reduces the total number of times (counted with multiplicity) that the $g$-image of edges in $G \setminus P_i$ cross edges in $P_i$. After finitely many such folds, $i$ increases. Since $P_i = \emptyset$ for all sufficiently large $i$, we eventually achieve (12).

Since these folds have no effect on $M_H$, (11) still holds and we have produced an irreducible fibered surface.
(7) If an irreducible fibered surface $F$ is not efficient, we construct a new fibered surface $F'$ and a homeomorphism $f'$ isotopic to $f$ with $\lambda(F', f') < \lambda(F, f)$. As in the once punctured case, the entire process can be repeated until one eventually arrives at an efficient fibered surface. Follow step (6) of Algorithm 3.2. Condition (12) implies that the edge $E$ is in $H$ and that when $g'|E'$ is tightened, one of the entries of $M_{H'}$ decreases. Thus, $\lambda' < \lambda$.

4.4. The invariant foliations

To obtain an invariant train-track, invariant measured foliations, and the Markov partition from an efficient fibered surface carrying $f$, we follow the analysis of the once punctured case. Draw the train-track by inserting infinitesimal edges which connect gates with the property that some power of $g$ sends some edge in $G$ to an edge path that enters and then immediately exits a vertex through that pair of gates. If not all complementary components of the train-track are (once punctured) disks, we find a reduction; otherwise, proceed to construct the invariant measured foliations. First assign widths and lengths to the edges in $H$ by finding a Perron–Frobenius eigenvector $\lambda$ for $M_H$ and its transpose. Then extend the solution to the other edges by solving

$$
\begin{bmatrix}
N & A & B \\
0 & C & D \\
0 & 0 & M_H
\end{bmatrix}
\begin{bmatrix}
X \\
W \\
V
\end{bmatrix}
= \lambda
\begin{bmatrix}
X \\
W \\
V
\end{bmatrix}
$$

for $X$ and $W$. Note that there is a unique solution since $\lambda > 1$ and the growth rates of $N$ and $C$ are one and zero, respectively. Of course, solving the similar equation for lengths assigns 0 to the edges of $P \cup \text{pre-}P$, so that they can also be considered infinitesimal. Finally, assign widths to the infinitesimal edges by solving the appropriate equation. The construction of the Markov partition and the invariant measured foliations now follows the once punctured case. Propositions 3.3.1.–3.3.5. hold without any changes. We leave it to the reader to supply the proof.

As an example, consider the homeomorphism of the sphere $(= \mathbb{R}^2 \cup \{ \infty \})$ with 6 punctures (one of which is $\infty$), defined by an efficient fibered surface in Fig. 10, where all peripheral loops (labelled by greek letters) are oriented counterclockwise and $\tilde{a}, \tilde{g}, \ldots$ denote $a, \gamma, \ldots$ with opposite orientation: $g(a) = \tilde{a}\tilde{d}, g(b) = \tilde{b}\tilde{d}\tilde{e}\tilde{e}\tilde{c}\tilde{e}, g(c) = \gamma\tilde{e}\tilde{c}\tilde{e}\tilde{b}\tilde{a},
g(d) = \tilde{b}\tilde{e}\tilde{e}\tilde{c}\tilde{e}\tilde{e}\tilde{c}\tilde{e}\tilde{b}\tilde{a}\tilde{d}, g(e) = \tilde{c}\tilde{e}\tilde{e}, g(\alpha) = \delta, g(\beta) = \epsilon, g(\gamma) = \alpha, g(\delta) = \beta, g(\varepsilon) = \gamma.$
The only nontrivial gate consists of the terminal endpoints of $a$ and $d$. However, $\infty$ is a 3-prong singularity, as the valence 4 vertex "opens up" and contributes two more prongs. The other punctures are 1-prong singularities.
5. NO PUNCTURES

We finish by proving Theorem 0.2.2 for surfaces with no punctures. Let $S$ be a closed surface with negative Euler characteristic, and let $f: S \to S$ be a homeomorphism. First assume that the Lefschetz number $L(f)$ is negative. Choose a fixed point $p$ for $f$. Now apply the algorithm from Section 3 to $f$ rel $p$. Any reduction we discover of $f$ on the punctured surface yields a reduction of $f$ on $S$. Suppose that the algorithm yields an efficient fibered surface $F$ that carries (a homeomorphism isotopic to) $f$ on $S$ rel $p$ and the associated pair of measured foliations $\mathcal{F}^s, \mathcal{F}^u$. If the puncture $p$ does not correspond to a 1-prong singularity of $\mathcal{F}^s$, the algorithm stops; $f$ is a pseudo-Anosov homeomorphism projectively fixing $\mathcal{F}^s$ and $\mathcal{F}^u$; otherwise, proceed as follows. The puncture $p$ has Lefschetz index 0, so there is another fixed point $q$ for $f$ with negative Lefschetz index. (This fixed point can be found by looking at the graph $G$ associated with $F$.) In particular, $f$ fixes at least two prongs of $\mathcal{F}^s$ at $q$. (Recall that the index of $q$ equals $1 - \#$ of fixed prongs of $\mathcal{F}^s$.) Draw a large segment $L$ of the unstable leaf in $F$ through $q$ determined by the two fixed directions. (This corresponds to iterating a path in $G$ through $q$ which maps over itself.) We can assume that the endpoints of $L$ are in junctions and that $L$ is so long that it passes through every pair of strips that determine an infinitesimal edge of the train-track associated with $F$ (see Section 3.3). Let $F'$ be $F$ cut open along $L$; it can be given the structure of a fibered surface (Fig. 11).

$L$ was chosen to be so long that the component of $S - F$ containing $p$ intersects only one junction with valence $> 2$, i.e. this component looks like a completely standard monogon (Fig. 12).

Now define a new fibered surface $F''$ for $f$ rel $q$, by "filling in" this monogon. On the level of graphs, the monogon contains a natural fixed point, which is "antipodal" to the vertex, and it is Nielsen equivalent to $p$. Identify the two halves of the monogon (Fig. 13). We obtain a fibered surface $F''$ that carries (a map isotopic to) $f$ on $S - q$ whose growth rate is $\lambda$. However, this fibered surface is not efficient, since the new junction has valence 1. Proceed with the algorithm from Section 3, after renaming $q$ to $p$. Repeat this procedure until discovering either a reduction or a pair of invariant measured foliations with no 1-prong singularities. This is guaranteed to stop, since the sequence of growth rates is strictly decreasing, and belongs to a discrete subset of $[1, \infty)$.

Even if $L(f) \geq 0$, it follows from linear algebra that there exists an integer $k > 0$ such that $L(f^k) < 0$. Instead of using fixed points of $f$ as above, we use orbits of fixed points of $f^k$ (see the beginning of Section 4.1). The details are left to the reader.
6. EXAMPLES

We illustrate the above algorithm on three examples. Edges of all graphs are oriented (and their orientation is specified most of the time by arrows in diagrams). We write \( \bar{a} \) for edge \( a \) with the opposite orientation, etc. Fibered surfaces are defined by their spines (\( G \) in the preceding sections), the reader can easily supply the decomposition into arcs and polygons. All homeomorphisms are assumed to be carried by the current fibered surface, and only the action on \( G \) is given. The most often used move is folding, and we suppress the preparatory subdivision. Thus, when we say fold \( a \) and \( b \) we mean if necessary subdivide \( a \) and \( b \) so that their initial segments map to the same edge-path, and then fold these initial segments. We also use \( a \) and \( \bar{a} \) to describe the points in the union of the links of vertices corresponding to the initial and the terminal endpoint of the edge \( a \).

6.1. Genus 2 surface

Consider the homeomorphism \( f \) of the once punctured genus 2 orientable surface given as the composition \( D_4 D_2 D_6 \) of Dehn twists in indicated curves. The Dehn twists are recorded by their action on the spine of the surface (Fig. 14), which is pictured as an octagon with side identifications, and the puncture corresponds to the vertex. The curve \( e \) is in the free homotopy class of \( bd \).
$D_g(b) = ab$, $D_g(a) = a\bar{b}$, $D_g(d) = cd$, $D_g(a) = adb$, $D_g(b) = \bar{b}adb$, $D_g(c) = cdb$, $D_g(d) = bdb$, where the edges not mentioned are fixed (e.g. $D_g(a) = a$, etc.). Finally, $g$ is given by $a \rightarrow ab\bar{b}dca$, $b \rightarrow acd\bar{b}acb$, $c \rightarrow ccd\bar{b}$, $d \rightarrow bc\bar{b}$.

In what follows, we sometimes simplify the picture of the fibered surface in the octagon by isotoping it to minimize the intersection with the 1-skeleton (the sides of the octagon).

Move 1. The fibered surface is not efficient, since $g(b) = \ldots ab \ldots$ and $Dg(a) = Dg(b) = a$, so there is a point $p$ in the interior of $b$ at which $g^2$ is not locally injective. The transition matrix is

$$
\begin{pmatrix}
3 & 2 & 0 & 0 \\
2 & 3 & 1 & 2 \\
1 & 2 & 2 & 1 \\
1 & 2 & 1 & 1
\end{pmatrix}
$$

whose growth rate is $\lambda = 6.268 \ldots$. We follow Algorithm 3.2(6) with $E = b$ and $k = 2$ to
lower the growth rate. Both $b$ and $\bar{a}$ map to $acd\bar{b}a\cdots$. Fold them (Fig. 15; here is a place where we suppress subdivision; both $b$ and $\bar{a}$ are first subdivided once to create edges that map to $acd\bar{b}a$).

The old edges are expressed in terms of the new ones as $a = a'x'$, $b = \bar{x}'b'$, $c = c'$, and $d = d'$ (where $x'$ is the edge arising by identifying parts of $a$ and $b$). The new map can be expressed as $a' \rightarrow (a'x')$, $b' \rightarrow c'd'(b'x')$, $c' \rightarrow c'c'd'(b'x')$, $x' \rightarrow (b'x')(\bar{x}'a')(\bar{x}'b')\bar{a}'(\bar{x}'a')$. The new map can be expressed as $a' \rightarrow (a'x')$, $b' \rightarrow c'd'(b'x')$, $c' \rightarrow c'c'd'(b'x')$, $x' \rightarrow (b'x')(\bar{x}'a')(\bar{x}'b')\bar{a}'(\bar{x}'a')$.

We now pull tight (notice that cancellation occurs in the image of $x$ where $p$ lands after folding) and drop the primes to get $a \rightarrow ax$, $b \rightarrow cdx$, $c \rightarrow ccdx$, $d \rightarrow b$, $x \rightarrow \bar{b}\bar{a}\bar{d}\bar{e}\bar{x}\bar{a}$.

The growth rate has dropped to $\lambda = 5.353\ldots$.

Move 2. We again follow Algorithm 3.2(6) with $k = 2$ and $E = e$ since $g(e) = \ldots\bar{d}\ldots$ and $Dg(b) = Dg(b) = \bar{x}$. Fold $\bar{b}$ and the initial segment of $d$ that maps to $\bar{xb}\bar{d}\bar{e}$. The new graph is pictured in Fig. 16 (the fold is easier to perform graphically if we first isotop the graph so that the edge $b$ is "short"). The new edges are expressed in terms of the new ones as $d = d'b'$, $a' = a$, $b' = b$, $c' = c$, $x' = x$.

The map is $a' \rightarrow a'x'$, $b' \rightarrow c'(d'b')\bar{b}'x'$, $c' \rightarrow c'c'(d'b')\bar{b}'x'$, $d' \rightarrow \bar{x}'b'$, $x' \rightarrow \bar{b}'\bar{a}'\bar{x}'b'(\bar{b}'\bar{d}')\bar{e}'\bar{x}'\bar{a}'$. Pull tight and drop the primes to obtain $a \rightarrow ax$, $b \rightarrow cdx$, $c \rightarrow ccdx$, $d \rightarrow b$, $x \rightarrow \bar{b}\bar{a}\bar{d}\bar{e}\bar{x}\bar{a}$, $y \rightarrow \bar{x}$ and the growth rate is $\lambda = 4.125\ldots$.

Move 3. Next take $E = b$, $g(b) = cd\ldots$, and $Dg(c) = Dg(d) = \bar{x}$. Fold $\bar{d}$ and the initial segment of $e$ that maps to $\bar{x}$. Here $c = c'y'$ and $d = y'd'$ (Fig. 17).

We leave it to the reader to write down the new map and then pull tight and drop the primes. The result is $a \rightarrow ax$, $b \rightarrow cdx$, $c \rightarrow c'y'\bar{c}d$, $d \rightarrow b$, $x \rightarrow \bar{b}\bar{a}\bar{d}\bar{e}\bar{x}\bar{a}$, $y \rightarrow \bar{x}$ and the growth rate is $\lambda = 3.537\ldots$.

Move 4. Take $E = x$, $g(x) = \ldots\bar{x}\bar{d}\ldots$ and $Dg(\bar{x}) = Dg(\bar{d}) = \bar{b}$. Fold $\bar{d}$ and the initial segment of $x$ that maps to $b$. We have $x = d'\bar{x}'$ (see Fig. 18; we again isotop the graph so that $d$ becomes "short") and after pulling tight and dropping the primes we obtain $a \rightarrow adx$, $b \rightarrow c\bar{x}$, $c \rightarrow c'y'\bar{c}d$, $d \rightarrow b$, $x \rightarrow \bar{a}\bar{x}\bar{c}\bar{d}\bar{x}\bar{a}$, $y \rightarrow \bar{x}d$. The new growth rate is $\lambda = 3.378\ldots$. 

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig16.png}
\caption{Fig. 16.}
\end{figure}
Move 5. Take $E = c, g(c) = c\ldots$, and $Dg(\overline{c}) = Dg(\overline{y}) = \overline{d}$. Fold the initial segments of $\overline{c}$ and $\overline{y}$ that map to $\overline{d}$. Again, $c = c\zeta'$ and $y = y\zeta'$ and we obtain (see Fig. 19)

$$a \rightarrow a\overline{d}x, b \rightarrow czx, c \rightarrow c\overline{yc}z, d \rightarrow b, x \rightarrow a\overline{xc}zda, y \rightarrow x, z \rightarrow d$$

with $\lambda = 3.346\ldots$

Move 6. $E = x, g(x) = ax\ldots, Dg(a) = Dg(x) = a$. Fold $a$ and the initial segment of $\overline{x}$ that maps to $a\overline{d}x (x = x'\overline{a}')$ and get (Fig. 20, isotoping to make $a$ short)

$$a \rightarrow a\overline{dx}a, b \rightarrow cxza, c \rightarrow c\overline{yc}z, d \rightarrow b, x \rightarrow x\overline{zc}, y \rightarrow ax, z \rightarrow d$$

with $\lambda = 2.807\ldots$

Move 7. Here we have to take $k = 3$ (since $g^2$ is locally 1–1 in the interior of every edge), $E = a, g(a) = \ldots ax\ldots, g^2(a) = \ldots \overline{bx}\ldots$, and $Dg(b) = Dg(\overline{x}) = c$, so fold $\overline{x}$ and the initial segment of $b$ that maps to $cza$. The map is (Fig. 21).

$$a \rightarrow a\overline{dx}a, b \rightarrow \overline{a}, c \rightarrow c\overline{yc}z, d \rightarrow \overline{xb}, x \rightarrow \overline{xc}, y \rightarrow ax, z \rightarrow d.$$  

There is no cancellation yet, so the growth rate remains unchanged.

Fig. 20.

Fig. 21.
To complete this move, fold the initial segments of \(d\) and \(x\) that map to \(\tilde{x}\) (Fig. 22):
\[
\begin{align*}
u &\to u\tilde{u}x\tilde{u}, \\
b &\to \tilde{a}, \\
c &\to c\tilde{c}z, \\
d &\to b, \\
x &\to \tilde{z}c, \\
y &\to a\tilde{x}\tilde{w}, \\
z &\to wd, \\
w &\to \tilde{x}\tilde{w},
\end{align*}
\]
with growth rate \(\lambda = 2.680\ldots\).

**Move 8.** We have to take \(k = 3\) again, with \(E = a, g(a) = a\tilde{a} \ldots , g^2(a) = \ldots \tilde{a}\tilde{b} \ldots \), and \(Dg(a) = Dg(\tilde{b}) = a\). We would like to fold \(\tilde{b}\) and the initial segment of \(a\) that maps to \(a\), but this would increase the valence (from 2 to 4) of the point where \(g^3\) fails to be locally injective. Instead, we first subdivide \(a\) into \(a = a_1a_2\), where \(a_1 \to a_1a_2\tilde{a}, a_2 \to x\tilde{a}_2\tilde{a}_1\). Now fold the initial segments of \(a_1\) and \(b\) that map to \(a_1\) (Fig. 23):
\[
\begin{align*}
\text{al} &\to \text{al}_1\text{al}_2, \\
\text{al}_2 &\to x\tilde{a}_2\tilde{a}_1\tilde{u}, \\
b &\to \tilde{a}_2, \\
c &\to c\tilde{c}z, \\
d &\to b\tilde{u}, \\
x &\to \tilde{z}c, \\
y &\to ua_1a_2\tilde{x}\tilde{w}, \\
z &\to wd, \\
w &\to x\tilde{w}, \\
u &\to ua_1.
\end{align*}
\]

For the second part of the move, fold the initial segments of \(\tilde{a}_2\) and \(\tilde{d}\) which map to \(u\) and pull tight to obtain (Fig. 24)
\[
\begin{align*}
\text{a}_1 &\to \text{a}_2\tilde{d}, \\
\text{a}_2 &\to x e\tilde{a}_2\tilde{a}_1, \\
b &\to e\tilde{a}_2, \\
c &\to c\tilde{c}z, \\
d &\to b, \\
x &\to \tilde{z}e, \\
y &\to ua_1a_2\tilde{e}\tilde{w}, \\
z &\to wd\tilde{e}, \\
w &\to \tilde{x}\tilde{w}, \\
u &\to ua_1, \\
e &\to u.
\end{align*}
\]
The growth rate is \(\lambda = 2.653\ldots\).

**Move 9.** Now remove two valence 2 vertices. A little computation shows that a positive eigenvector assigns larger weight to \(e\) than to \(x\) (\(w(e)/w(x) = 1.117\ldots\)), and larger weight to \(a_2\) than to \(a_1\) (\(w(a_2)/w(a_1) = 1.107\ldots\)). We note that \(a_1a_2 \to a_2\tilde{d}xe\tilde{a}_2\tilde{a}_1\) and \(xe \to \tilde{z}e\).
Homotoping across $e$ and $a_2$ amounts to dropping all occurrences of $e$ and $a_2$ in the image edge-paths, as well as replacing the image of $x$ (and $a_1$) by the image of $xe$ (and $a_1a_2$). We also drop the subscript from $u_2$ and obtain (Fig. 25)

$$a \to dxa, b \to *, c \to cycz, d \to b, x \to \tilde{z}cu, y \to uaxw, z \to wd, w \to \tilde{x}w, u \to ua.$$ 

Now edges $b$ and $d$ form an invariant forest. Collapse it to get (Fig. 26):

$$a \to dxa, c \to cycz, x \to \tilde{z}cu, y \to uaxw, z \to w, w \to \tilde{x}w, u \to ua.$$ 

The growth rate is $\lambda = 2.61803$. . . . It is easy to see that the only non-trivial gates are given by \{u, y\} and \{y, z\}, and hence we have an efficient fibered surface.

Following the recipe of Section 3.3, we produce an invariant train-track for the homeomorphism (Fig. 27).

To construct the invariant measured foliations, follow Section 3.4. Both infinitesimal quadrilaterals give rise to 4 prong singularities, and they are permuted by the homeomorphism. The square of the homeomorphism rotates these singularities by $\pi/2$. The puncture is contained in a bigon, so it becomes a non-singular point of the invariant foliations. (This is
therefore also an example of a pseudo-Anosov homeomorphism of the closed genus 2 surface.)

6.2. Four times punctured sphere

We regard the 2-sphere as $\mathbb{R}^2 \cup \infty$, and $\infty$ is one of the punctures. The three loops around the other three punctures are oriented counterclockwise (Fig. 28), and form the invariant subgraph $P$ from the text. The remaining edges are in $H$. The left Dehn twists in the curves 1 and 2 are given by $D_1(a) = a, D_1(c) = ax\beta c$ and $D_2(a) = \beta c\gamma e, D_2(c) = c$ (The three loops are fixed under both maps). Consider the homeomorphism $f$ given by $f = D_2D_1^2$, i.e. by $g(a) = \beta c\gamma e a, g(c) = \beta c\gamma e a x c \gamma e b c\gamma e a x c \gamma e$. All subsequent modifications of $g$ are going to fix $\alpha, \beta, \gamma$. 

Fig. 26.

Fig. 27.
Move 1 (see Section 4.2).

\[ M_{II} = \begin{pmatrix} 1 & 4 \\ 2 & 7 \end{pmatrix} \]

and the growth rate \( \lambda = 4 + \sqrt{17} = 8.123 \ldots \). The representative is not irreducible, as it fails to satisfy (I2). We follow the procedure from 4.2 and use the notation of the proof of Lemma 4.2.1. \( \text{Fr } A_x \) and \( \text{Fr } A_y \) have only one component so no change is made in the graph at these components of \( P \). \( \text{Fr } A_y \) has two components, one for \( a \) and one for \( c \). The map \( Dg_1 \) satisfies \( a \mapsto c \) and \( c \mapsto c \) so there is a single equivalence class of elements of \( \text{Fr } A_y \). The arc \( I_{a1} \) is as shown in Fig. 29. Note that \( I_{a1} \) can be thought of as the terminal segment of \( a \) that maps to \( \bar{\beta} \) followed by the initial segment of \( c \) that maps to \( \beta \); thus, \( g_1(I_{a1}) = \bar{\beta} \beta \) is contractible rel endpoints. There is no change in the graph \( G \). The new map is given by \( a \to cyca, c \to cyca \bar{c} \bar{a} \bar{c} \bar{c} \bar{y} \bar{c} \bar{d} \bar{a} \bar{c} \bar{a} \bar{c} \). The growth rate has not changed.

Move 2 (see Algorithm 4.3(7)). The above irreducible fibered surface is not efficient. We can take \( E = a \), since \( g(a) \) enters and exits through the same gate \( \{a, c\} \). Fold \( a \) and the initial segment of \( c \) that maps to \( cyca \) to get (notice that \( c = a'c', \) see Fig. 30) \( a' \to a'c'y\bar{c}\bar{a}'a' \) and \( c' \to a\bar{a}'a'c'y\bar{c}\bar{a}'a'c'y\bar{c}\bar{a}'a' \) or after pulling tight and dropping the primes
The transition matrix on the non-peripheral edges $a, c$ is now
\[
\begin{pmatrix}
1 & 2 \\
2 & 5
\end{pmatrix}
\]
and the growth rate has dropped to $\lambda = 3 + 2\sqrt{2} = 5.828 \ldots$

*Move 3* (see Section 4.2). The fibered surface is not irreducible ($\mathcal{I}_2$ fails again). The frontiers of $A_1, A_2, A_3$ have, respectively, two, one and one component. The arc $I_{a_1}$ is shown in Fig. 30. Note $I_{a_1}$ can be thought of as $\alpha$ followed by an initial segment of $c$ that maps to $a$; thus, $g_1(I_{a_1}) = \bar{a}a$ is contractible rel endpoints. The resulting graph is shown in Fig. 31. The map is given by $a \to aacy\alpha, c \to ac\gamma bacy\gamma c$. The growth rate is unchanged.

This is an efficient fibered surface, the only non-trivial gate is $\{a, c\}$. The corresponding train-track is depicted in Fig. 32, the invariant foliations have 1-prong singularities at the punctures.

### 6.3. Sphere with 5 punctures

We view the 2-sphere as $\mathbb{R}^2 \cup \{\infty\}$, and $\infty$ is one of the punctures (Fig. 33). The map is given by $a \to b, b \to c, c \to d, d \to d\bar{e}d\bar{e}b$, so the punctures are cyclically permuted and the algorithm from Section 3 applies. The fibered surface is not efficient since $g^2$ maps $d$ to an edge path that backtracks. Fold $a$ and the last quarter of $d$ (Fig. 33) to obtain $a \to b, b \to c,$
c → dā, d → a(ād)c =  āc. The new fibered surface is efficient. The gates are {c, c}, {ā, a, d}, and {b, b, d}. The corresponding train-track is pictured in Fig. 33. All punctures have a 1-prong singularity, and there is another 3-prong singularity which rotates by 2π/3.

Acknowledgements—We thank Joe Christy, who read carefully the preliminary version of the manuscript and corrected some of our mistakes. We also thank the NSF, the Alfred P. Sloan Foundation, and the Institute for Advanced Study for their support while the research leading to this paper was conducted. In addition, the first author acknowledges the support of the Presidential Young Investigator Award and the I.H.E.S.

REFERENCES


*Department of Mathematics*
*University of Utah*
*Salt Lake City, Utah 84112, U.S.A.*

*Cuny Lehman College*