# Automorphisms of 3-Dimensional Handlebodies

Ulrich Oertel*

Revised May, 2000

Rutgers University, Newark

**Abstract**

This paper gives a classification up to isotopy of automorphisms (self-homeomorphisms) of 3-dimensional handlebodies and compression bodies, analogous to the Nielsen-Thurston classification of automorphisms of surfaces. Indecomposable automorphisms analogous to pseudo-Anosov automorphisms are identified and called generic. The first steps are taken towards understanding generic automorphisms using invariant laminations. An automorphism $f : M \to M$ of an arbitrary compact, connected, orientable, irreducible 3-manifold $M$ with non-empty boundary can be understood by decomposing the 3-manifold into $f$-invariant submanifolds including the Jaco-Shalen-Johannson characteristic manifold and Bonahon's characteristic compression body.

*Keywords:* handlebody, compression body, automorphism, homeomorphism, mapping class group, 3-manifold.

## 1 Introduction

The mapping class group of a handlebody has been studied extensively, for example by S. Suzuki, [17], who gives generators, and by B. Wajnryb, [19], who gives a presentation. D. McCullough and A. Miller studied the mapping class group of compression bodies and 3-manifolds with compressible boundary, see [15]. In related work, McCullough studied the cohomological dimension of the mapping class groups of 3-manifolds, see [14]. H. Masur described the action of the mapping class group of a handlebody on the projective lamination space of the boundary surface, see [13].

This paper is concerned with individual automorphisms (self-diffeomorphisms or self-homeomorphisms) of 3-handlebodies and compression bodies, rather

*Research supported by the National Science Foundation and by the Max Planck Institute for Mathematics, Bonn

than the mapping class group. The goal is to describe a theory of automorphisms of handlebodies and compression bodies analogous to the Nielsen-Thurston classification up to isotopy of surface automorphisms, [18].

We shall define a class of indecomposable automorphisms of handlebodies and compression bodies which are analogous to pseudo-Anosov automorphisms of surfaces. We call these automorphisms "generic." This answers a question raised by K. Johannson in [11] whether such an analogue exists.

A *handlebody* $H$ of dimension 3 and genus $g$ is a 3-ball with some number $g \geq 1$ of 3-dimensional 1-handles attached. We will follow the convention that a disjoint union of handlebodies is also called a handlebody. Of course, an automorphism $f : H \to H$, where $H$ is a handlebody, induces a surface automorphism $\partial f : \partial H \to \partial H$. One can apply Thurston's result to $\partial f$, and it is reasonable to expect that the topology and dynamics of $f$ are in some way related to those of $\partial f$. This is indeed the case.

We begin by reminding the reader of the statement of Thurston's theorem on the classification of surface automorphisms, but we will not give the definition of a *pseudo-Anosov* automorphism. The reader is referred to [18], [5], and [3] for this definition. A *reducing curve system* for an automorphism $g : S \to S$ is a set of disjointly embedded essential simple closed curves in $S$ whose union is $g$-invariant up to isotopy. The automorphism $g$ is *reducible* if a reducing curve system exists for $g$.

**Theorem 1.1. (W. Thurston)** *Suppose $g : S \to S$ is a automorphism of a surface $S$, $\chi(S) \leq 0$. Then one of the following holds. The automorphism $g$ is:*
*1) reducible,*
*2) periodic, or*
*3) pseudo-Anosov.*

In the reducible case, there is a canonical decomposition of the automorphism and the surface:

**Theorem 1.2. (W. Thurston)** *Suppose $g : S \to S$ is a automorphism of a surface $S$, $\chi(S) \leq 0$. There is a canonical minimal reducing curve system for $g$. Cutting the surface $S$ open on this curve system gives a surface $\hat{S}$ with boundary, and $g$ induces an automorphism $\hat{g}$ of $\hat{S}$. The automorphism $\hat{g}$ may permute the components of $\hat{S}$, but each component is invariant under some power $\hat{g}^k$, and $\hat{g}^k$ is either periodic or pseudo-Anosov on that component (up to isotopy).*

We refer to the components of $\hat{S}$ in the decomposition of $S$ given in Theorem 1.2 as the *surface elements* of the decomposition on a minimal reducing curve system. We refer to "pseudo-Anosov" or "periodic" surface elements, meaning that some power $g^k$ is pseudo-Anosov or periodic on the surface element.

In this introduction, we will consider automorphisms $f : H \to H$ of handlebodies, leaving the problem of the classification of automorphisms of compression bodies for later sections. Throughout the paper we consider isotopy classes of automorphisms. When we refer to an $f$-invariant subspace, we mean

$f$-invariant up to isotopy; when we refer to a periodic automorphism, we mean periodic up to isotopy.

For an automorphism $f$ of a handlebody, we shall see that there may be $f$-invariant reducing surfaces analogous to the reducing curve systems of Thurston's theory. These can be incompressible or compressible. A partial definition will be given below after we define compression bodies; the complete definition will be given in Section 2. As one might expect, the reducing surfaces have boundaries which are reducing curve systems for the induced automorphism $\partial f$ on $\partial H$.

A *compression body* is a 3-manifold pair $(Q, U)$, $U \subset \partial Q$, constructed from the disjoint union of a product $U \times I = U \times [0, 1]$ and a collection of balls by attaching 1-handles to $U \times 1$ and the boundaries of the balls, see Figure 1. Usually, in the literature, a compression body is constructed from a closed surface $U$, but we allow $U$ with boundary. Each component of the surface $U$ is either a closed surface, not the sphere, or a surface with boundary, possibly including disc components. We identify $U$ with $U \times 0 \subset \partial Q$. We give $I$-bundles a structure analogous to that of compression bodies. We regard an $I$-bundle $p : Q \to S$ as an *$I$-bundle pair* $(Q, U)$ where $U$ is the total space of the associated $\partial I$-bundle. We say the commpression body or $I$-bundle pair is *spotless* if $U$ contains no disc components. We say a compression body is *trivial* or a *product compression body* if it has the form $(Q, U) = U \times I$ with $U = U \times 0$. In particular, a handlebody is a spotless compression body. We let $\partial_i Q$ denote the surface $U$, also called the *interior boundary* of $Q$; we let $\partial_e Q$ denote the surface $W = \partial Q - \mathring{U}$, also called the *exterior boundary* of $Q$, even when $(Q, U)$ is an $I$-bundle pair.

Another useful model of a compression body is the following. Let $F$ be a surface possibly including sphere components and let $U$ be obtained from $F$ by removing sphere components. Then the compression body $(Q, U)$ is obtained from $F \times I$ by attaching 1-handles to $F \times 1$ and by capping all sphere components of $F \times 0$ by balls.

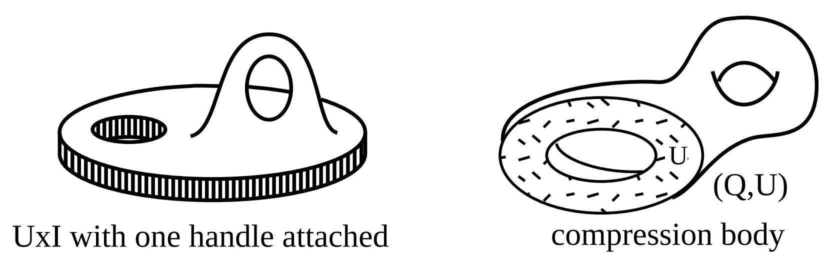



**Figure 1**

We follow the convention that compression bodies, $I$-bundle pairs, and handlebodies need not be connected. Handlebodies and compression bodies are irreducible. When analyzing automorphisms of a handlebody $H$, we will work with compression bodies and $I$-bundle pairs $Q$ embedded in $H$ as submanifolds with $\partial_e Q \subset \partial H$ and with $\text{int}(\partial_i Q) \subset \text{int}(H)$.

A handlebody is itself a compression body. If $(H, V)$ denotes a spotless compression body, then this is a handlebody if $V = \emptyset$. The following is a partial definition of reducing surfaces for automorphisms of handlebodies and compression bodies; it does not deal with arbitrary automorphisms of compression bodies. The complete definition will be given in Section 2.

**Definition 1.3.** Suppose $f : H \to H$ is an automorphism of a connected spotless handlebody which restricts to a pseudo-Anosov automorphism on $\partial H$. Suppose $Q$ is an $f$-invariant compression body in $H$ with $\partial_e Q = \partial H$. Then $\partial_i Q$ is a *reducing surface*. We say the automorphism $f : H \to H$ is *reducible* if it has a reducing surface, and it is *irreducible* if it does not have a reducing surface.

We can now define generic automorphisms of handlebodies. These will play the role of pseudo-Anosov automorphisms in the Nielsen-Thurston theory of automorphisms of surfaces.

**Definition 1.4.** If $H$ is a connected handlebody, a *generic* automorphism $f : H \to H$ is an automorphism which restricts to a pseudo-Anosov automorphism on $\partial H$ and which is irreducible.

In the definition of a generic automorphism of a handlebody, we note that if $f : H \to H$ restricts to a pseudo-Anosov automorphism on $\partial H$, then any reducing surface must be closed. Any reducing surface must therefore also be compressible.

Suppose $H$ is a handlebody and $f : H \to H$ is an automorphism. If $f$ preserves a structure of a spotless compression body $(H, V)$ for $H$, then we shall see from the complete definition of reducing surfaces, to be given in Section 2, that $f$ has a reducing surface and is reducible.

There is another important type of handlebody automorphism; it is essentially the same as a surface automorphism. Suppose $p : H \to S$ is an $I$-bundle over a surface $S$. We can regard the bundle as an $I$-bundle pair $(H, U)$. An automorphism $g : S \to S$ induces an automorphism $f : (H, U) \to (H, U)$ called the *lift* of $g$ which satisfies $p \circ f = g \circ p$. We are particularly interested in lifts of pseudo-Anosov automorphisms of surfaces. These automorphisms were described in Johannson's book, [11]. Suppose $H$ is a handlebody and $f : H \to H$ is an automorphism which preserves a structure of an $I$-bundle pair $(H, V)$ for $H$, then again we shall see that $f$ has a reducing surface and is reducible. It is not difficult to verify that any automorphism of an $I$-bundle pair is actually a lift of an automorphism of the base surface of the $I$-bundle.

Other reducing surfaces for automorphisms of handlebodies are more like the reducing curve systems which appear in the theory of automorphisms of surfaces.

The following theorem should be compared to Thurston's theorem on the classification of automorphisms of surfaces, see Theorem 1.1.

**Theorem 1.5.** *Suppose $f : H \to H$ is an automorphism of a connected handlebody. Then the automorphism is:*
*1) reducible,*
*2) periodic, or*
*3) generic on the handlebody.*

In Section 3, we shall prove a similar theorem classifying automorphisms of compression bodies, even compression bodies with closed interior boundary whose underlying topological space is not a handlebody. Using the following established theorems, it is then possible to obtain a classification of automorphisms of arbitrary compact, connected, irreducible manifolds with non-empty boundary, assuming that such a classification is possible for Seifert fibered 3-manifolds. The results indicate that the most interesting automorphisms of irreducible 3-manifolds with non-empty boundary are supported in the characteristic compression body and in the characteristic manifold. Away from these, automorphisms must be periodic.

**Theorem 1.6. (F. Bonahon, [2].)** *Suppose $M$ is a compact, irreducible, orientable 3-manifold. Suppose $f : M \to M$ is an automorphism. Then $f$ preserves, up to isotopy, the characteristic compression body $Q_{\partial M}$ (associated to the surface $\partial M$).*

We will describe the characteristic compression body in Section 2.

**Theorem 1.7. (W. Jaco and P. Shalen, [9]; K. Johannson, [10].)** *Suppose $M$ is a compact, irreducible, $\partial$-irreducible, orientable 3-manifold. Suppose $f : M \to M$ is an automorphism. Then $f$ preserves, up to isotopy, the characteristic manifold in $M$.*

**Theorem 1.8.** *(***K. Johannson, [10].)** *Suppose $M$ is an orientable, $\partial$-irreducible Haken 3-manifold with empty characteristic submanifold. Suppose $f : M \to M$ is an automorphism. Then, up to isotopy, $f$ is periodic.*

Clearly, an automorphism of an irreducible $\partial$-irreducible manifold with non-empty boundary yields an induced automorphism of the characteristic manifold, which can be subdivided into submanifolds which are either Seifert fibered or $I$-bundles. Automorphisms of a Seifert fibered manifold are considered to be well understood. The Seifert fiber structure is usually unique, and automorphisms preserve it up to isotopy. Similar comments apply to $I$-bundles of the kind that appear in the characteristic manifold. Therefore, we can regard the induced automorphisms of characteristic manifolds as being fairly well understood.

We need a better understanding of generic automorphisms of handlebodies. As in the Nielsen-Thurston theory, we shall investigate generic automorphisms using invariant laminations. Here, in the introduction, we state a theorem which applies to automorphisms of handlebodies; later, see Propostion 7.1, we state a result which applies to arbitrary compression bodies. In the following theorem $H_0$ is a concentric copy of $H$ embedded in the interior of $H$ such that $H - \mathring{H}_0$ has the structure of a product $\partial H \times I$.

**Theorem 1.9.** *Suppose $f : H \to H$ is a generic automorphism of a 3-dimensional handlebody. Then there is a 2-dimensional measured lamination $\Lambda \hookrightarrow \mathring{H}$ with transverse measure $\mu$ such that, up to isotopy, $f((\Lambda, \mu)) = (\Lambda, \lambda\mu)$ for some $\lambda > 1$. The lamination has the following properties:*

*1) Each leaf $\ell$ of $\Lambda$ is an open 2-dimensional disc.*

*2) The lamination $\Lambda$ fills $H_0$, in the sense that the components of $H_0 - \Lambda$ are contractible.*

*3) For each leaf $\ell$ of $\Lambda$, $\ell - \mathring{H}_0$ is incompressible in $H - \mathring{H}_0$.*

*4) $\Lambda \cup \partial H$ is a closed subset of $H$.*

*There is also a 1-dimensional lamination $\Omega$, transverse to $\Lambda$, with transverse measure $\nu$ and a map $\omega : \Omega \to \mathring{H}_0$ such that $f(\omega(\Omega, \nu)) = \omega(\Omega, \nu/\lambda)$. The map $f$ is an embedding on $f^{-1}(N(\Lambda))$ for some neighborhood $N(\Lambda)$. The statement that $f(\omega(\Omega, \nu)) = \omega(\Omega, \nu/\lambda)$ should be interpreted to mean that there is an isomorphism $h : (\Omega, \nu) \to (\Omega, \nu/\lambda)$ such that $f \circ \omega = \omega \circ h$.*

We note that the laminations in the above statement are "essential" only in a rather weak sense. For example, lifts of leaves of $\Lambda$ to the universal cover of $H$ are not necessarily properly embedded. The map $\omega$ need not be proper either: If we regarded $w : \Omega \to H$ as a homotopy class, in some examples there would be much unravelling. The fact that the invariant laminations are somewhat badly behaved should not be regarded as a flaw. Rather, the strange properties of the laminations shed light on the nature of automorphisms of 3-dimensional handlebodies.

**Remark 1.10.** There is an obvious question we have not addressed yet: Do generic automorphisms of handlebodies exist? Leonardo Navarro de Carvalho has explicitly constructed such examples, in work which is not yet published. Navarro has also constructed examples of generic automorphisms which induce the trivial automorphism of the fundamental group.

The following theorem was proved in response to a question asked by Feng Luo.

**Theorem 1.11.** *Suppose $H$ is a handlebody, and $f : H \to H$ is an automorphism with the property that $\partial f$ is a composition of Dehn twists on disjoint curves. Then $f$ is a composition of Dehn twists on a disjoint collection of discs and annuli.*

In order to avoid technicalities in the introduction, we have relegated the statements of some important results to later sections. For example, we will state and prove a theorem analogous to Theorem 1.2, showing how to decompose an arbitrary automorphism of a handlebody or compression body, see Theorem 4.1. We list open questions and problems in Section 9.

# 2 Reducing Surfaces

We will delay giving the complete definition of "reducing surface." Many reducing surfaces will be obtained using a version of Francis Bonahon's characteristic compression body, see [2]. Our version is more general than Bonahon's. Let $W$ be a compact essential subsurface of $\partial M$, where $M$ is an irreducible, orientable 3-manifold. If $\mathcal{D} = \{D_1, D_2, \ldots, D_q\}$ is a maximal collection of non-isotopic disjoint compressing discs of $W$ in $M$, a *characteristic compression body $Q$ associated to $W$* is defined to be a regular neighborhood $N = N(W \cup \mathcal{D})$, with boundary spheres capped off by the balls they bound in $M$. Here we abuse notation by using $\mathcal{D}$ also to denote $\cup_i D_i$. The characteristic compression body is a pair $(Q, U)$, where $U = \partial Q - \mathring{W} = \partial_i Q$, and $W = \partial_e Q$. We shall be using characteristic compression bodies only in a special class of irreducible 3-manifolds, namely in compression bodies.

The characteristic compression body $(Q, U)$ described above is, of course, a compression body according to our previous definition. To see this, note that in $N = N(W \cup \mathcal{D})$ the complement of $N(\mathcal{D})$ is a product $F \times I$, where $F = \partial N - \mathring{W}$. Each disc $D_i \in \mathcal{D}$ is dual to a 1-handle, so $N$ is obtained by attaching 1-handles to $F \times 1$. Thus $(N, F)$ has the structure of a compression body, except that $F$ may contain sphere components. We cap the sphere components of $F \subset \partial N$ with balls to obtain $(Q, U)$, where $U$ is obtained from $F$ by removing sphere components. Alternatively, we see that $(Q, U)$ is obtained by attaching 1-handles to $U \times 1 \subset U \times I$ and to the boundaries of a finite collection of balls.

The following proposition (a slight generalization of a result of F. Bonahon, see [2]) gives the essential properties of characteristic compression bodies.

**Proposition 2.1.** *Let $M$ be an irreducible, orientable 3-manifold, and $W$ an essential surface in $\partial M$.*

*(i) If $Q$ is a characteristic compression body associated to $W$, then $\partial_i Q$ is incompressible in $M$, but possibly with disc components.*

*(ii) The characteristic compression body $Q$ associated to $W$ is uniquely determined up to isotopy.*

*Proof.* (i) We have defined $Q$ as $\bar{N}(W \cup \mathcal{D})$, where $\mathcal{D}$ is a maximal collection of mutually non-isotopic compressing discs of $W$, and where the bar indicates that we augment $N = N(W \cup \mathcal{D})$ by capping sphere boundary components with the balls they bound in $M$. Initially, we will work with $N$ rather than $Q$. The submanifold $N$ is a $F \times I$ with a 1-handle attached to $F \times 1$ for every $D_i \in \mathcal{D}$, the disc $D_i$ being a cocore of the 1-handle. Suppose now that $\partial_i Q$ is compressible, then a non-sphere component of $F = \partial_i N = \partial N - \mathring{W}$ is compressible. It cannot be compressible in $N$, so it is compressible in $M - \mathring{N}$. Suppose $D$ is a compressing disc in $M - \mathring{N}$, with $\partial D \subset F$. We extend $D$ through the product by attaching $\partial D \times I \subset F \times I$ to $\partial D \subset F \times 0$. The attached collar can be can be perturbed so that it becomes disjoint from the handle attaching discs in $F \times 1$, so the extended $D$ satisfies $\partial D \subset W$. If $D$ were parallel to one of the $D_i$'s, then the original $D$ could not be a compressing disc of $\partial_i N$. Therefore, we have a contradiction; $\mathcal{D}$ is not a maximal collection of non-isotopic compressing discs.

(ii) To prove uniqueness, we suppose $Q' = \bar{N}(\mathcal{D}')$ where $\mathcal{D}'$ is another maximal collection of compressing discs of $W$. Using standard methods, intersections of $\mathcal{D}'$ with $\partial_i Q$ can be removed by isotopy, since $\partial_i Q$ is incompressible.

Therefore $N(W \cup \mathcal{D}') \subset Q$. Any boundary spheres of $N(W \cup \mathcal{D}')$ bound balls in $Q$, so $\bar{N}(W \cup \mathcal{D}') = Q' \subset Q$. Now by (i) $\partial_i Q' = U'$ is incompressible, hence it can be made disjoint from $\mathcal{D}$, so it is an incompressible surface in the product $U \times I$ with $\partial U'$ isotopic to $\partial W$ or $\partial U \times 1$. It is well-known that such a surface $U'$ must be isotopic to $U \times 0$, or $\partial_i Q$, hence $(Q', U')$ is isotopic to $(Q, U)$. $\square$

Some examples of characteristic compression bodies in handlebodies are given in Figures 2 and 3. As another example, again in a handlebody $H$, consider the characteristic compression body associated to a compressible annulus $W$ in $\partial H$. It has the form $(Q, U)$, where $Q$ is a ball and $U$ is the disjoint union of two discs in $\partial Q$.

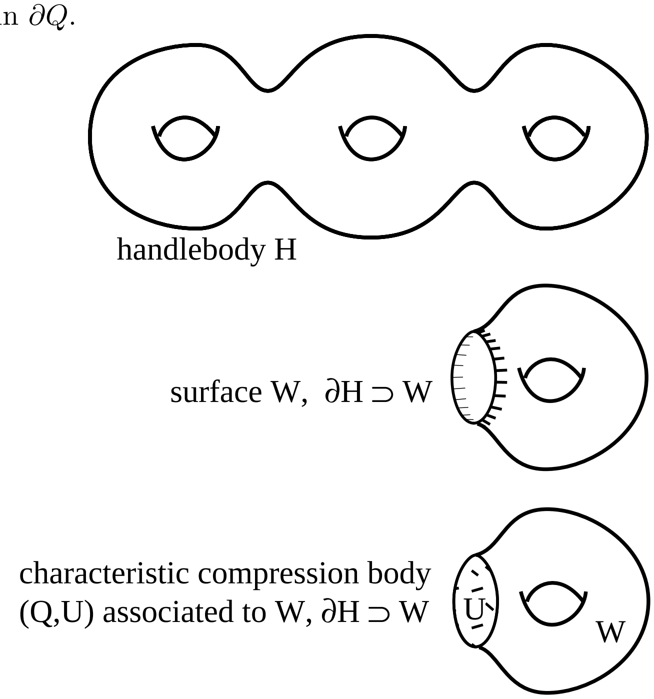



**Figure 2**

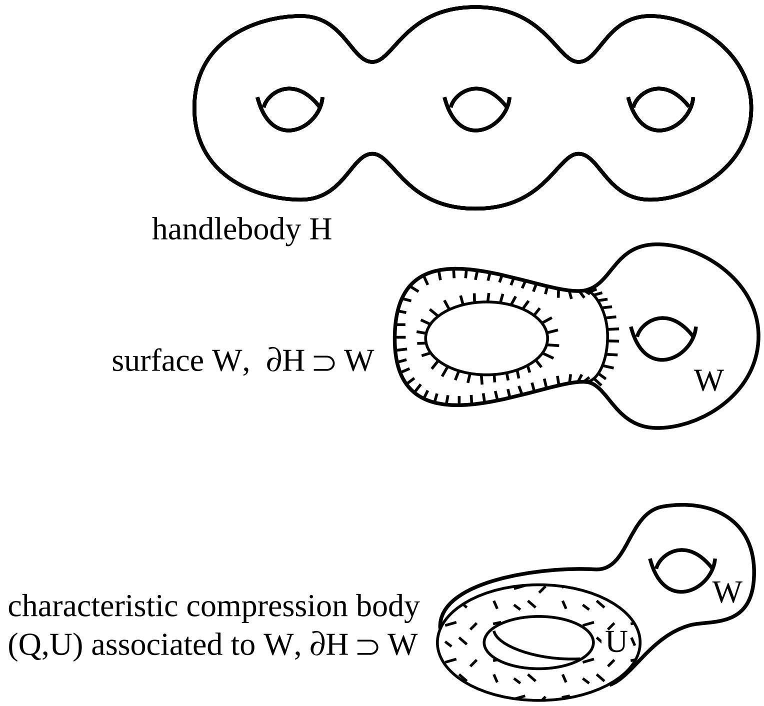



**Figure 3**

Our main interest in the characteristic compression body arises from the following corollary.

**Corollary 2.2.** *Suppose $f : M \to M$ is an automorphism of an irreducible 3-manifold $M$ and $W \hookrightarrow \partial M$ is a compact surface essential in $\partial M$. Suppose $W$ is $f$-invariant up to isotopy, then the associated characteristic compression body $Q$ is $f$-invariant up to isotopy.*

In the introduction we stated Theorem 1.5, which classifies automorphisms of handlebodies. We will go further in two directions. First, we will deal with arbitrary compression bodies whose underlying spaces are not handlebodies. Second, for an arbitrary compression body $(H, U)$, we will show in Section 4 how to decompose an arbitrary automorphism $f : (H, U) \to (H, U)$. After decomposing on a reducing surface and passing to a finite power of $f$, we may assume that the induced automorphism acts on a connected "piece." The induced automorphisms on these pieces, in turn, must be analyzed by decomposing on further reducing surfaces, etc. The following definitions describe the types of pieces we will have to deal with.

**Definitions 2.3.** We have already defined a *spotless compression body* $(H, V)$: The space $H$ is obtained from a disjoint union of balls and a product $V \times I$ by attaching 1-handles to the boundaries of the balls and to $V \times 1$; the surface $V \hookrightarrow \partial H$ is the same as $V \times 0$. We require that $V$ contain no disc or sphere components. When $H$ is connected and $V = \emptyset$, $(H, V)$ is a handlebody or ball. If $(H, V)$ has the form $H = V \times I$ with $V \times 0 = V$, then $(H, V)$ is called a *spotless product compression body* or a *trivial compression body.* As usual, $\partial_e H = \partial H - \mathring{V}$ is the *exterior boundary*, even if $V = \emptyset$.

A *spotted compression body* is a triple $(H, V, \Delta)$ where $(H, V)$ is a spotless compression body and $\Delta \neq \emptyset$ denotes a union of discs or "spots" embedded in $\partial_e H = \partial H - \mathring{V}$.

A *spotted product* is a spotted compression body $(H, V, \Delta)$ of the form $H = V \times I$ with $V = V \times 0$ and with $\Delta \neq \emptyset$ a union of discs embedded in $V \times 1$.

A *spotted ball* is a spotted compression body whose underlying space is a ball. It has the form $(B, \Delta)$ where $B$ is a ball and $\Delta \neq \emptyset$ is a disjoint union of discs in $\partial B = \partial_e B$.

An *$I$-bundle pair* is a pair $(H, V)$ where $H$ is the total space of an $I$-bundle $p : H \to S$ over a surface $S$ and $V$ is $\partial_i H$, the total space of the associated $\partial I$-bundle.

**Remark 2.4.** Given a compression body $(Q, U)$, say a characteristic compression body associated to a surface $W$ in $\partial H$, the surface $U = \partial_i(Q, U)$ may contain disc components. We shall often view $(Q, U)$ instead as a spotted compression body $(Q, V, \Delta)$, where $\Delta$ is the union of discs in $U$ and $V = U - \Delta$. For example, a compression body $(Q, U)$ can become a spotted ball $(B, \Delta)$ or $(B, \emptyset, \Delta)$. We consistently use $U$ for the interior boundary of a compression body possibly including discs, and $V$ for the interior boundary not including discs.

A spotless compression body or $I$-bundle pair is an example of a *Haken pair*, i.e. a pair $(M, F)$ where $M$ is an irreducible 3-manifold and $F \subset \partial M$ is incompressible in $M$.

As in the introduction, if $(H, V)$ is an $I$-bundle pair $p : H \to S$, we say the *exterior boundary*, $\partial_e J$, is $p^{-1}(\partial S)$ and the *interior boundary* is the total space of the $\partial I$-bundle associated to the bundle $p : H \to S$.

**Definition 2.5.** Given an automorphism $f : (H, U) \to (H, U)$ of a connected compression body $(H, U)$ with $\partial_e(H, U) = W$, a *reducing surface* for $f$ is an $f$-invariant surface of one of the following types:

i) Suppose $(Q, T)$ is a non-product $f$-invariant compression body in $H$ with $\partial_e Q \subset W = \partial_e(H, U)$. Suppose it is not the case that $(Q, \partial_e Q)$ is isotopic via an isotopy of pairs in $(H, W)$ to all of $(H, W)$, so that $T = U$ after the isotopy. Then the union of the non-$U$-parallel components of $T$ is a *compressional reducing surface.*

ii) Suppose $(Q, T)$ is an $f$-invariant $I$-bundle pair embedded in $H$ with $\partial_e(Q, T) \subset W$ and with $Q$ isotopic rel $\partial_e(Q, T)$ to $H$, then $T$ is a reducing surface.

iii) If $f$ is any automorphism of any spotless compression body $(H, U) = (H, V)$, and $G$ is an $f$-invariant union of essential (incompressible and non-$\partial$-parallel) annuli, with $\partial G \subset (\mathring{V} \cup \mathring{W})$, then $G$ is an *annular reducing surface.*

An automorphism $f$ of a (spotted) compression body is *reducible* if there is a reducing surface for $f$, and is *irreducible* otherwise.

In (i), suppose $(Q,T)$ is a spotted product $(Q,\hat{T},\Delta)$, with every component of $\Delta$ isotopic to a disc in $U$. Then $\hat{T}$ is a *spot-removing reducing surface* for $f$.

In (i) or (ii), suppose $(Q,T)$ is an invariant compression body or $I$-bundle pair with $(Q,\partial_e Q)$ isotopic in $(H,W)$ to all of $H$, and suppose it is not the case that $(Q,\partial_e Q)$ is isotopic via an isotopy of pairs in $(H,W)$ to all of $(H,W)$ such that $T=U$ after the isotopy. Then $T$ is a *peripheral reducing surface*.

For completeness, if $(H,V)$ is a spotless $I$-bundle pair, we define reducing surfaces for automorphisms $f:(H,V)\to(H,V)$. Any $f$-invariant annular non-$\partial$-parallel surface with boundary in $V$ is such an *annular reducing surface*, and these are the only reducing surfaces.

Now it is quite easy to check that Definition 1.3 given in the introduction agrees with the definition given here. We will show later, see Corollary 2.18, that if an automorphism $f:(H,V)\to(H,V)$ of a compression body is reducible with a non-empty canonical minimal reducing curve system for $\partial_e f$, then there is a reducing surface for $f$ without closed components. Thus one need only consider reducing surfaces without closed components, except in special cases like the case where $\partial_e f$ is pseudo-Anosov on a closed surface.

Peripheral reducing surfaces exist for an automorphism $f:(H,V)\to(H,V)$ whenever $H$ can be given another structure as an $f$-invariant compression body or $I$-bundle pair with strictly larger (up to isotopy) $V$. Thus, for example, any automorphism of a trivial product compression body is reducible, since there is a peripheral reducing surface arising from an $I$-bundle pair, as in (ii) of the definition.

**Example 2.6.** Suppose that $f:H\to H$ is an automorphism of a connected handlebody and that $F$ is a pseudo-Anosov surface element for $\partial f$. Suppose that $F, f(F), f^2(F)$ are distinct surface elements but $f^3(F)=F$. Let $S=F\cup f(F)\cup f^2(F)$. Suppose $F$ is compressible. Then, if $Q$ is the characteristic compression body associated to $S$, $\partial_i Q$ is an incompressible reducing surface.

**Example 2.7.** Let $g:(Q,U)\to(Q,U)$ be any automorphism of a a non-trivial compression body $(Q,U)$, where $U$ is connected with $\partial U\neq\emptyset$. Doubling $Q$ on $U$ to get $H$ and doubling the automorphism $g$ we obtain an automorphism $f$ of the handlebody $H$. The doubling surface is an incompressible reducing surface.

**Example 2.8.** Let $g$ be a reducible automorphism of a surface $S$ with boundary. For specificity, let us suppose $S$ is divided by a simple closed curve $\alpha$ into two surface elements $S_1$ and $S_2$ on which $g$ induces pseudo-Anosov automorphisms $g_1$ and $g_2$, see Figure 4. Suppose $S$ and $S_1$ have one boundary component each, and that $S_2$ has two boundary components. Let $f$ be the lift of $g$ to the product $H=S\times I$, then $f$ is an automorphism of the handlebody $H$. Now $\alpha\times I$ is a compressional reducing surface, since $S_2\times I$ is an invariant compression body $Q$ , with $\partial_i Q$ the annulus $\alpha\times I$. Of course $\alpha\times I$ is also a annular reducing surface.

The same automorphism $f$ is also reducible via a peripheral reducing surface. Clearly there is an invariant (product) $I$-bundle pair $(Q,T)$ with $Q$ of the form $S \times I$ with $\partial_e(Q,T) \subset \partial H$ and with $Q$ isotopic rel $\partial_e(Q,T)$ to $H$.

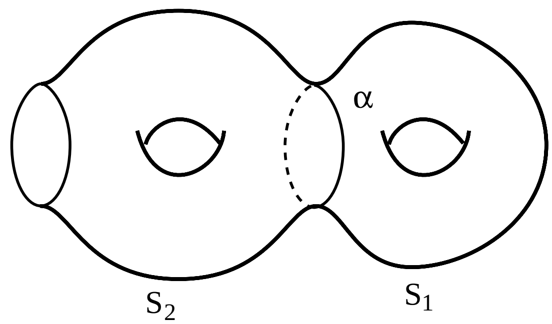



**Figure 4**

**Example 2.9.** Let $f : (H,V) \to (H,V)$ be any automorphism of a compression body. Now let $R \subsetneq V)$ be an $f$-invariant surface. Then $f : (H,R) \to (H,R)$ is reducible; the components of $V$ not isotopic to a components of $R$ form a reducing surface when they are isotoped to be properly embedded in $H$. This example illustrates the role of peripheral reducing surfaces: They detect a choice of compression body or $I$-bundle pair structure for $H$ which is not optimal.

Using the characteristic compression body, we can often detect reducing surfaces using the following corollary of Proposition 2.1.

**Corollary 2.10.** *Suppose $f : (H,V) \to (H,V)$ is an automorphism of a connected spotless compression body. If a surface $S \subset \partial_e(H,V)$ is $\partial f$-invariant and compressible, then the characteristic compression body $(Q,T)$ associated to $S$ yields an incompressible compressional reducing surface or $(Q,S)$ is isotopic in $(H,W)$ to $(H,W)$, where $W = \partial_e H$. This applies also when $V = \emptyset$.*

*Proof.* Let $Q$ be the characteristic compression body associated to $S$. We may discard components of $S$ whose characteristic compression bodies are trivial. If $\partial_i Q = T$ contains components which are not $V$-parallel, their union forms a reducing surface for $f$. □

We eliminate the nuisance of spots using spot-removing reducing surfaces. Suppose $f : (H,V,\Delta) \to (H,V,\Delta)$ is an automorphism of a connected spotted compression body, $\Delta \neq \emptyset$. Notice that $W = \partial H - \mathring{V}$ is a connected surface. Let $N = N(W)$ be a product neighborhood of $W$ and let $T$ be the closure in $\partial N$ of $\partial N - \partial H$. If $T$ is isotopic to $V$, then $H$ is a spotted product, otherwise $T$ is a spot-removing reducing surface, or $T$ is a sphere bounding a ball in $H$. We have proved:

**Proposition 2.11.** *Let $f : (H,V,\Delta) \to (H,V,\Delta)$ be an automorphism of a connected spotted compression body with $\Delta \neq \emptyset$. Then there is a spot-removing* *reducing surface, which cuts a spotted product from* $(H, V, \Delta)$*, or* $(H, V, \Delta)$ *is a spotted product or a spotted ball.*

Using decomposition on these spot-removing reducing surfaces, we can always decompose an automorphism of a spotted compression body to obtain automorphisms of spotted products, spotted balls, and spotless compression bodies. Notice that an automorphism of a spotted product may itself have a spot-removing reducing surface.

We regard automorphisms of spotted products $(H, V, \Delta)$, where $H = V \times I$ and $\Delta \subset V \times 1$, as being understandable. Such an automorphism is obtained from a lift $g \times \mathrm{id}$ of an automorphism $g : V \to V$ of the base surface by composing with an automorphism "stirring" the spots in $V \times 1$. The stirring automorphism is probably generated by Dehn twists and fractional Dehn twists, in $(V \times 1)$-parallel discs and annuli separating some spots of $\Delta$ from the product.

In view of Proposition 2.11, we can now restrict our attention to automorphisms of spotless compression bodies.

Given an automorphism $f$ of a connected spotless compression body $(H, V)$, to find incompressible compressional reducing surfaces for $f$, we use the canonical minimal reducing curve system for $\partial_e f$ and the surface elements obtained by cutting $\partial_e H$ on this system. If a surface element $F$ is compressible in $H$, then the union of iterates of $F$ gives an $f$-invariant surface $S$. By Corollary 2.10 the associated characteristic compression body $Q$ yields an incompressible reducing surface contained in $\partial_i Q$ if $(Q, \partial_e Q)$ is not isotopic to all of $(H, \partial_e H)$. More generally, if all surface elements in $\partial_e H$ are incompressible in $H$, an $f$-invariant, compressible, union of surface elements can also yield a reducing surface in the same way.

In some situations we must detect annular reducing surfaces. In particular, when surface elements in $\partial_e H$ are all incompressible, the most obvious strategy for finding an incompressible compressional reducing surface fails. That is, the characteristic compression body associated to a union of iterates of a surface element does not give a compressional reducing surface, since the characteristic compression body is trivial. To deal with this situation we prove the following sequence of lemmas and propositions.

We shall need the following notion of $\partial$-incompressibility. Let $M$ be a 3-manifold and $R \hookrightarrow \partial M$ an essential subsurface of the boundary. Let $S \hookrightarrow M$ be properly embedded in $M$ with $\partial S$ intersecting $\partial R$ transversely. Then $S$ is *$\partial$-incompressible with respect to* $R$ if for every half-disc $(K, \alpha, \beta) \hookrightarrow (M, S, R)$ (where $\alpha$ and $\beta$ are complementary arcs in the boundary of the disc $K$ meeting only at endpoints), there is a half-disc $(K', \alpha, \beta') \hookrightarrow (S, S, \partial S \cap R)$.

**Lemma 2.12.** *1) Suppose* $H$ *is a handlebody and* $R \hookrightarrow \partial H$ *is a surface embedded in* $\partial H$ *incompressible in* $H$*. Then there is an essential disc* $E \hookrightarrow H$ *with the property that* $\partial E$ *intersects* $\partial R$ *transversely and minimally among representatives of its isotopy class, and that* $E$ *is* $\partial$*-incompressible with respect to* $R$*. If* $R_0$ *is a component of* $R$*,* $E$ *can be chosen so the intersection with* $R_0$ *is non-empty.*

*2) Suppose $(H,V)$ is a spotless compression body and $R \hookrightarrow \partial_e(H,V)$ is an incompressible surface embedded in $W = \partial H - \mathring{V}$, with $R_0$ any component of $R$. Then one of the following is true.*

*i) There is an essential disc $(E, \partial E) \hookrightarrow (H, W)$ with the property that $\partial E$ is isotoped to intersect $\partial R$ minimally and transversely, that $E$ is $\partial$-incompressible with respect to $R$, and that $E \cap R_0 \neq \emptyset$.*

*ii) There is an essential annulus $A$ in $(H,V)$ with one boundary component in $R_0$ and the other boundary component in $V$.*

*Proof.* The first statement is a special case of the second, so it is enough to prove the second. Let $\mathcal{E}$ denote a complete system of compressing discs for $W$. This means that cutting on $\mathcal{E}$ yields balls and a space homeomorphic to the product $V \times I$. If there is a complete system $\mathcal{E}$ for which no $E_i \in \mathcal{E}$ intersects $R_0$, clearly there is an essential annulus $A$ with boundaries in $R_0$ and $V$. Otherwise we assign a lexicographic complexity $(a,b)$ to $\mathcal{E}$ where $b$ is the number of arcs $|\mathcal{E} \cap R|$, while $a$ is the minimum of $|E_i \cap R|$ over all $E_i \in \mathcal{E}$ which intersect $R_0$. (We use $|X|$ to denote the number of components of a space $X$.)

We choose a complete system $\mathcal{E}$ of minimal complexity. Let $E_0 \in \mathcal{E}$ realize $a$ in the complexity. We claim $E_0$ satisfies the $\partial$-incompressibility condition with respect to $R$. If not, let $K$ be a $\partial$-compressing disc. Surgery of $E_0$ using $K$ yields two discs $E_0'$ and $E_0''$, and replacing $E_0$ by these discs in $\mathcal{E}$ yields a complete system $\mathcal{E}'$ of smaller complexity as follows. Either after surgery $\partial \mathcal{E}'$ already intersects $\partial R$ minimally, and the entry $a$ is reduced, or $\partial \mathcal{E}'$ can be isotoped to reduce the total number of arcs of intersection with $R$, so the entry $b$ can be reduced without increasing $a$. This is a contradiction, so we can take the disc $E$ of the statement to be $E_0$. ☐

In the following statement $J$ denotes the total space of an $I$-bundle $p : J \to S$, which we regard as an $I$-bundle pair.

**Lemma 2.13.** *Let $f : H \to H$ be an automorphism of a connected handlebody $H$. Suppose the canonical minimal reducing curve system is $\mathcal{R}$ and $R = \partial H - \mathring{N}(\mathcal{R})$ is incompressible. If $\partial f$ is pseudo-Anosov on a subsurface of $R$, then there is a submanifold $J \hookrightarrow H$, the total space of an $I$-bundle, with $\partial_i J \hookrightarrow \partial H$, and with (up to isotopy) the restriction of $f$ to $J$ the lift of a pseudo-Anosov automorphism of a surface.*

*Proof.* After replacing $f$ by a finite power of itself, we may suppose $f$ is pseudo-Anosov on a component $R_0$ of $R$. We enlarge $R$ by adding $N(\mathcal{R})$ and enlarge $\mathcal{R}$ by replacing it by $\partial N(\mathcal{R})$, where $N(\mathcal{R})$ always means a neighborhood in $\partial H$. We have added an annular component to $R$ for every curve of the canonical minimal reducing curve system. Now we may assume that $\partial f$ is the identity on $\partial H - \mathring{R}$, since Dehn twists are now supported in the annular components of $R$. Thus we may as well assume that the components of $R$ abut on curves of $\mathcal{R}$.

By Lemma 2.12 there is a disc $E$ which is $\partial$-incompressible with respect to $R$ and intersects $R_0$. The disc $E$ should be thought of as a polygon with vertices corresponding to intersections of $\partial E$ with $\mathcal{R}$.

For some large $j$ we consider $E \cap f^j(E)$. Let $\alpha_0$ denote $E \cap R_0$, so $\alpha_0$ is a union of sides of $E$. Letting $R_*$ be the union of non-annular components of $R$, we may assume that the sides of $f^j(E)$ in $R_*$ are isotoped as pairs to minimize intersections with the sides of $E$. Thus, if $s$ is a side of $E$ in $R_*$, $(f^j(s), f^j(\partial s))$ is isotoped in $(R_*, \partial R_*)$ to minimize intersections with $\partial E$. The isotopy just performed ensures that there is never an arc of $\partial E \cap R_*$ which together with an arc of $f^j(E) \cap R_*$ cuts a triangular disc from $R_*$, with one side of the triangle on $\partial R_*$. In annular components of $R$ we minimize intersections rel $\partial$, allowing triangular regions next to the boundary. We remove closed curves of $f^j(E) \cap E$ by isotopy. Since $f$ is pseudo-Anosov on $R_0 \subset R_*$, $|f^j(\alpha_0) \cap \alpha_0| \to \infty$ as $j \to \infty$, where we count the minimum number of intersections up to isotopy of $f^j(\alpha_0)$. If an arc $\gamma$ of $E \cap f^j(E)$ in $E$ joins two points of $\partial E$, one of which is in $\alpha_0$, then there are two possibilities. Either the arc is inessential in $E$, joining a side in $\alpha_0$ to itself, or it is essential, joining a side in $\alpha_0$ to a distinct side. Assuming there are inessential arcs, we suppose $\gamma$ is innermost in $E$, cutting a half-disc $K$ from $E$. The $\partial$-incompressibility condition satisfied by $E$ and $f^j(E)$ ensures that in this case the arc of intersection could be removed by isotopy, contradicting the fact that we have already isotoped the sides of $E$ and $f^j(E)$ in $R_0$ to minimize intersections. Similar arguments apply in other components of $R$, so we can assume that the pattern of arcs of $E \cap f^j(E)$ in $E$ or $f^j(E)$ contains no inessential arcs.

It follows that for sufficiently large $j$, we obtain arbitrarily many essential arcs of intersection in $E$ with one end in $\alpha_0 \cap f^j(\alpha_0)$. Viewed in $E$, these arcs have one end in a side in $\alpha_0$ and the other end in a distinct side of $E$; viewed in $f^j(E)$, one end is in a side in $f^j(\alpha_0)$ and the other end in a distinct side of $f^j(E)$. Thus we obtain arbitrarily many rectangular regions subtended from $f^j(E)$ by consecutive adjacent arcs of intersection with one end in $f^j(\alpha_0)$. Since the number of essential arc types (isotopy classes of arcs joining distinct sides) in $E$ is bounded, for sufficiently large $j$ we can find a sequence of consecutive adjacent rectangles in $f^j(E)$ which can be joined to form an annulus $A$ or Möbius band with boundary in $\mathring{R}$. A Möbius band can be replaced by the boundary of its regular neighborhood, yielding an annulus $A$. In any case, the annulus $A$ has at least one boundary component in $R_0$. The annulus is essential in the Haken pair $(H, \partial H - \mathring{N}(\mathcal{R})$: If it were $\partial$-compressible to $R_0$, this would yield a $\partial$-compression for $f^j(E)$, a contradiction. Clearly $\partial A \cap R_0$ is not null-homotopic in $R_0$, otherwise $|\alpha_0 \cap f^j(\alpha_0)|$ is not minimized by isotopy.

It is also easy to verify that $A \cap R_0$ does not contain a closed curve $\partial$-parallel in $R_0$, otherwise a component of $f^j(\alpha_0)$ would spiral towards a component of $\partial R_0$, and one would have triangle adjacent to $\partial R_0$ in the pattern formed in $R_0$ by $\alpha_0$ and $f^j(\alpha_0)$.

Applying $f$ iteratively to $A$, we obtain a sequence of essential annuli, each intersecting $R_0$ in at least one boundary circle. Combining two of these, we shall construct an invariant essential $I$-bundle $p : J \to S$ for the pair $(H, \partial H - \mathring{N}(\mathcal{R}))$, with at least one component of $\partial_i J$ contained in $R_0$. One can show that in a Haken pair $(M, F)$ ($F \subset \partial M$ incompressible) there exists an essential I-bundle $p : K \to S$ with $\partial_i K \subset F$ maximal subject to the condition that it contain no component which is a bundle over an annulus or Möbius band. This I-bundle is unique up to isotopy of $(K, \partial_i K)$ in $(M, F)$. Using two iterates of $A$ which intersect essentially in arcs, we see that there is an $I$-bundle $L$ with $\partial_i L$ intersecting $R_0$ in an essential non-annular surface, hence $K$ is non-empty. In fact, $\partial_i K$ must contain $R_0$ up to isotopy since $K$ is invariant, and $R_0$ has no proper non-annular subsurfaces, invariant up to isotopy.

The surface $R_0$ could have been any pseudo-Anosov surface element, so we know that $\partial_i K$ contains all pseudo-Anosov surface elements in $R$. Taking $(M, F) = (H, \partial H - \mathring{N}(\mathcal{R}))$, we let $J$ be the subbundle of $K$ consisting of the fibers of $K$ intersecting pseudo-Anosov surface elements in $R$. Since $J$ is $f$-invariant, we conclude that up to isotopy $f|_J$ is the lift of a pseudo-Anosov automorphism of a surface. ☐

**Lemma 2.14.** *Let $H$ be a connected handlebody and let $\mathcal{R} \subset \partial H$ be an embedded curve system with the property that no closed curve in $\mathcal{R}$ bounds a disc in $H$. Let $f : H \to H$ be an automorphism of the connected handlebody $H$ whose restriction $\partial f$ to $\partial H$ is a composition of non-trivial Dehn twists on the finitely many disjoint curves $C_i$ of $\mathcal{R}$.*

*i) Then for every $C_i$, there is an invariant essential annulus $A_i$ with $C_i \subset \partial A_i$.*

*ii) There is also an invariant surface $G$ with $\partial G \subset N(\mathcal{R}) \subset \partial H$ consisting of essential annuli, with every $C_i$ isotopic to some component of $\partial G$.*

*Proof.* As in the previous lemma, $\mathcal{R}$ is the canonical minimal reducing curve system. As before, we replace each curve $C_i$ of $\mathcal{R}$ by $\partial N(C_i)$. Then $\partial H - \mathring{N}(\mathcal{R})$ contains an annulus neighborhood of each $C_i$. Unless $\partial H$ is a torus, we can assume all components $F_i$ of $\partial H - \mathring{N}(\mathcal{R})$ not supporting Dehn twists are not annuli, otherwise Dehn twists in adjacent annuli of $R$ separated by an annulus can be combined. If $\partial H$ is a torus, the only automorphisms $f : H \to H$ restricting to a compositions of Dehn twists on disjoint curves are Dehn twists on meridians, contrary to our assumption that the $C_i$'s do not bound discs in $H$. We let $R$ be the union of disjoint annuli $N(C_i)$, and we let $R_0$ denote one of these, so $\partial f$ restricts to a non-trivial Dehn twist supported in $R_0$. Applying Lemma 2.12, we choose an essential disc $E$ in $H$ which is $\partial$-incompressible with respect to $R$ and intersecting $R_0$. We let $\alpha_0$ denote $E \cap R_0$.

As before, $(\partial f)^j$ stretches $\alpha_0$, making $f^j(\alpha_0)$ arbitrarily long as measured by intersections with $\alpha_0$ for sufficiently large $j$. For large $j$, we isotope the sides $f^j(s)$ of $f^j(E) \cap R$ rel boundary in $R$ to minimize intersections with sides $s$ of $E$. Then $E$ and $f^j(E)$ do not intersect in $\partial H - R$. Notice that in this setting $\partial E - R \neq \emptyset$ and the "sides" of $E$ are components of $E \cap R$. There are arcs of $\partial E$ which are mapped to $\partial H - \mathring{R}$.

We consider intersections of $E$ and $f^j(E)$. The $\partial$-incompressibility condition guarantees that there are no inessential arcs of intersection in $E$ ($f^j(E)$) joining an arc of $\partial E \cap R$ ($f^j(\partial E) \cap R$) to itself. There are only essential arcs joining distinct sides of $E$ or $f^j(E)$. For sufficiently large $j$, we obtain arbitrarily many rectangular regions subtended from $f^j(E)$ by consecutive adjacent essential arcs of intersection with one end in $f^j(\alpha_0)$. For sufficiently large $j$, these rectangular regions can be joined to form an embedded annulus or Möbius band. As before, we replace a Möbius band by the boundary of its regular neighborhood to obtain an annulus $A$. If this annulus were $\partial$-parallel, with both boundaries in $R_0$ and isotopic to an annulus $A'$ in $R_0$ we would obtain a $\partial$-compression with respect to $R$ for $f^j(E)$, a contradiction. There is also a possibility that $A$ might be isotopic to an annulus $A' \subset \partial H$, but with $A'$ not contained in an $R_i$. However, this would imply that there is an annulus in $\partial H - R$, which has already been ruled out. We conclude that $A$ is essential, and clearly $\partial A \subset R$.

We have proved statement i). To prove statement ii), observe that the characteristic manifold $N$, see [11], for the Haken manifold pair $(H, R)$ must intersect each annulus $N(C_i)$. Then $\text{fr}(N)$ gives an invariant embedded surface $G$ with annulus components as required. ☐

**Proposition 2.15.** *Let $f : H \to H$ be an automorphism of a connected handlebody $H$. Suppose $\mathcal{R}$ is the canonical minimal reducing curve system for the induced automorphism $\partial f$ of $\partial H$. Suppose $R = \partial H - \mathring{N}(\mathcal{R})$ is incompressible. Then $f$ is reducible.*

*In particular, either*

*1) $H$ can be given an $f$ invariant structure as an $I$-bundle pair $(H, V)$, and $f$ therefore has a peripheral reducing surface, or*

*2) $f$ has an annular reducing surface $G$, with every component of $\partial G$ isotopic in $\partial H$ to a curve of $\mathcal{R}$.*

**Remark 2.16.** K. Johannson's Proposition 3.11 of [11] is related, in that it detects incompressible $f$-invariant non-$\partial$-parallel surfaces, including unions of annuli. With our assumptions, we are always guaranteed reducing surfaces. Other incompressible, $f$-invariant, non-$\partial$-parallel surfaces obtained from Johannson's proposition are not necessarily reducing surfaces according to our definition. With some additional work, it might be possible to use Johannson's proposition to prove Proposition 2.15.

*Proof. (Proposition 2.15.)* Since $R$ is incompressible, clearly $\mathcal{R} \neq \emptyset$, so either there is a pseudo-Anosov surface element, or all surface elements are periodic, and $\partial f$ is periodic on $R = \partial H - \mathring{N}(\mathcal{R})$, with Dehn twists supported in neighborhoods of curves of $\mathcal{R}$.

In the first case, Lemma 2.13 gives a submanifold $(J, \partial_i J) \hookrightarrow (H, \partial H)$ such that up to isotopy the restriction of $f$ to $J$ is a lift of a pseudo-Anosov automorphism of a surface. If $J$ is isotopic to $H$, then $f$ is itself a lift. Otherwise $\partial_e J$ gives an annular reducing surface.

If $\partial f$ is a composition of Dehn twists, Lemma 2.14 yields an $f$-invariant surface $G$ consisting of essential annuli in the frontier of the characteristic manifold for the Haken pair $(H, N(\mathcal{R}))$, where, as usual, $N(\mathcal{R})$ denotes a regular neighborhood in $\partial H$. If $f$ is periodic in the complement of annuli supporting Dehn twists, and $f$ possibly permutes the annuli supporting Dehn twists, then by passing to a finite power of $f$, we may suppose that $\partial f$ is the identity on $R$ with Dehn twists supported in $N(\mathcal{R})$. Again by Lemma 2.14 the characteristic manifold for the pair $(H, N(\mathcal{R}))$ is non-empty and the frontier yields an $f$-invariant annular surface for the original $f$. (Actually, one can show that if $\partial f$ is a composition of Dehn twists on the disjoint closed curves of $\mathcal{R}$, then $\partial H - \mathcal{R}$ must be compressible, see the proof of Theorem 1.11 at the end of Section 4, so this case is vacuous.) ∎

We need a version of Proposition 2.15 which applies to connected compression bodies. The most interesting new case to consider is a compression body $(H, V)$ of the form $V \times I$ with 1-handles attached to $V \times 1$, where $V$ is a closed surface. An automorphism $f : (H, V) \to (H, V)$ induces an automorphism $\partial_e f$ on $\partial_e H$. We will work with surface elements for $\partial_e f$, the automorphism induced on $\partial_e(H, V)$ by $f$.

In compression bodies, it will be convenient to distinguish different types of annular reducing surfaces: Those annuli with both boundary components in $W = \partial H - \mathring{V}$, and those with one boundary component in $V$ and the other in $W$. (An annulus $A$ with $\partial A \subset V$ must be $\partial$-parallel.) We call the annuli of the former type *horizontal* and the annuli of the latter type *vertical*. If an annular reducing surface $G$ consists of horizontal annuli it is called *a horizontal annular reducing surface*; if it consists of vertical annuli, it is called a *vertical annular reducing surface* for $f : (H, V) \to (H, V)$. If $f : (H, V) \to (H, V)$ is an automorphism of an $I$-bundle pair, then any (annular) reducing surface is *vertical*.

**Proposition 2.17.** *Let $f : (H, V) \to (H, V)$ be an automorphism of a connected spotless compression body $(H, V)$, where $V \neq \emptyset$. Suppose $\mathcal{R}$ is the canonical minimal reducing curve system for the induced automorphism $\partial_e f$ of $W = \partial_e(H, V)$, and suppose $R = \partial_e(H, V) - \mathring{N}(\mathcal{R})$ is incompressible. Then $f$ is reducible.*

*In particular, either*

*1) $(H, V)$ is a product $V \times I$, with $V = V \times 0$ and $f$ is the lift of a pseudo-Anosov or periodic automorphism to the product, so $f$ has a peripheral reducing surface, or*

*2) $H$ is an $I$-bundle $p : H \to S$ with $R$ the total space of the associated $\partial I$-bundle and with $V \subset p^{-1}(\partial S) = \partial_e H$, so $f$ has a peripheral reducing surface, or*

*3) $f$ has an annular reducing surface $G$.*

*Proof.* Our assumption says that surface elements in $R$ are incompressible. If $\mathcal{R} = \emptyset$, then $W = R$ is incompressible, and $(H, V)$ is a product compression body. The automorphism $\partial_e f$ must be periodic or pseudo-Anosov, hence $f$ is the lift of a periodic or pseudo-Anosov automorphism to the product compression body $V \times I$. By definition, the automorphism of the product compression body is reducible, since the product compression body can be replaced by an invariant $I$-bundle pair.

If $\mathcal{R} \neq \emptyset$, there must be at least one pseudo-Anosov surface element in $R$ or $\partial_e f$ is a composition of Dehn twists on curves of $\mathcal{R}$, after replacing $f$ by a power of itself. In the former case we proceed as in the proof of Lemma 2.13, in the latter as in the proof of Lemma 2.14.

In the former case, when $\partial_e f$ has at least one pseudo-Anosov surface element $R_0$, we let $\mathcal{R}$ denote the canonical minimal reducing curve system for $\partial_e f$, we replace $\mathcal{R}$ by $\partial N(\mathcal{R})$, then let $R$ denote the surface obtained by cutting $W$ on the modified $\mathcal{R}$. Thus $R$ contains an annulus corresponding to every curve of the canonical minimal reducing curve system. Applying Lemma 2.12 we either obtain a disc $E$ $\partial$-incompressible with respect to a $R$ and intersecting $R_0$, or there is a vertical essential annulus with one boundary component in each of $R_0$ and $V$. If there is an essential vertical annulus, two essentially intersecting iterates give an essential $I$-bundle $J$ which is not a bundle over an annulus or Möbius band and with $\partial_i J \subset V \cup R$. There is then a unique maximal such $I$-bundle $J$ associated to the Haken manifold pair $(H, V \cup R)$, subject to the condition that no component of $J$ is an $I$-bundle over an annulus or Möbius band. Then $\partial_e J$ yields a vertical annular reducing surface. Otherwise we have an essential disc $E$ which is $\partial$-incompressible with respect to $R$, and we can finish the argument in Lemma 2.13 to obtain a submanifold $J$, which is the total space of an $I$-bundle, with $\partial_i J \subset R$. In this case, if $J$ is isotopic to all of $H$, then $V \subset \partial_e J$ and we get 2) of the statement. Otherwise $\partial_e J$ yields a horizontal annular reducing surface.

In the remaining case, $\partial_e f$ has no pseudo-Anosov surface elements, and after replacing $f$ by a power of itself we have $\partial_e f$ a composition of Dehn twists on disjoint curves. As in the proof of Lemma 2.14, we let $R$ denote a union of annuli supporting the Dehn twists, and let $R_0$ be one of these. We may assume that no component of $\partial H - \mathring{R}$ is an annulus, otherwise either $\partial H$ is a torus, $V$ is an annulus, and $(H, V)$ is a product compression body; or $W$ and $V$ are closed, $W$ is a torus, and $(H, V)$ is a product compression body; or $H$ is a solid torus with $V = \emptyset$, contrary to hypothesis. The proposition can be verified in the special cases which have not been excluded.

By Lemma 2.12, either we obtain an essential disc $(E, \partial E) \hookrightarrow (H, W)$ $\partial$-incompressible with respect to $R$ and intersecting $R_0$, or we obtain an essential vertical annulus with one boundary component in each of $R_0$ and $V$. In the latter case, the union of the iterates of such a vertical essential annulus $A$ give an annular reducing surface for the original $f$, since clearly $\cup_i f^i(A) \cap W$ is invariant up to isotopy and it is then straightforward to show that $\cup_i f^i(A)$ is invariant up to isotopy. In the former case, where we obtain the disc $E$ which is $\partial$-incompressible to $R$, we construct a horizontal annulus $A$ as in Lemma 2.14, with $\partial A$ essential in $R$, and as before we check that $A$ is $\partial$-incompressible with respect to $R$. Hence $A$ is not $\partial$-parallel in $\partial H$, since components of $\partial H - R$ are not annuli. The frontier of the characteristic manifold for the Haken manifold pair $(H, R)$ then gives an annular reducing surface for the original $f$. ☐

**Corollary 2.18.** *Suppose $f : (H, V) \to (H, V)$ is an automorphism of a compression body, and suppose the canonical minimal reducing curve system $\mathcal{R}$ for $\partial_e f$ is non-empty. Then there is a reducing surface without closed components.*

*Proof.* If a surface element $F$ for $\partial_e f$ is compressible, then the union of iterates of $F$ yields a compressible invariant surface $S \subset W = \partial_e(H, V)$, and the associated characteristic compression body yields a compressional reducing surface. Otherwise all surface elements are incompressible and Propositions 2.15 and 2.17 show that there are reducing surfaces without closed components. ☐

# 3 Classification

In this section we prove a classification theorem for automorphisms of compression bodies. This will imply the classification theorem for automorphisms of handlebodies, Theorem 1.5, given in the introduction.

The following lemma will be needed in the proof of the classification theorem.

**Lemma 3.1.** *1) Suppose $H$ is a handlebody and $f : H \to H$ is an automorphism. If $\partial f : \partial H \to \partial H$ is periodic, then so is $f$. If $\partial f$ is the identity, so is $f$ (up to isotopy).*

*2) Suppose $f : (H, V) \to (H, V)$ is an automorphism of a spotless compression body. If $f|_{\partial_e(H,V)} = \partial_e f$ is periodic, then $f$ is periodic on $H$.*

*Proof.* 1) This statement is well known, see for example [11].

2) For the second statement, since $\partial_e f$ is periodic, there exists an integer $n$ such that $h = f^n$ is isotopic to the identity on $\partial_e(H, V)$. Applying the following Lemma 3.2, we conclude $h = f^n$ is isotopic to the identity. ☐

**Lemma 3.2.** *Suppose $(H, V)$ is a spotless compression body, and $f : (H, V) \to (H, V)$ is an automorphism with the property that $\partial_e f$ is isotopic to the identity. Then $f$ is isotopic to the identity on $H$.*

*Proof.* Without loss of generality $H$ is connected. We have already dealt with the case $V = \emptyset$, so we also assume $V \neq \emptyset$. Let $A$ be a vertical essential annulus or rectangle in $H$. This means that if $H$ is a product $V \times I$ with handles attached to $V \times 1$, then $A$ is of the form $\gamma \times I$ where $\gamma$ is an essential curve in $V$. (We choose $\gamma$ so that $\gamma \times 1$ is disjoint from attaching discs for 1-handles.)

After a slight isotopy, $f(A) \cap A \cap W = \emptyset$, where $W = \partial H - \mathring{V}$, and we can assume that $f(A) \cap A$ consists of closed curves and arcs with ends on $V$. Trivial closed curves can be removed by isotopy of $f$. Using the incompressibility of $V$, arcs can also be removed by isotopy. If $A$ is an annulus and $f(A) \cap A$ contains a closed curve essential in $A$ and $f(A)$, then cut-and-paste on a curve in $A$ nearest $\partial A \cap V$ yields a new embedded annulus with both boundaries in $V$. Such an annulus in a product must be $\partial$-parallel, which implies that the curve of intersection can be removed by isotopy of $f$. Thus we may assume that $f(A) \cap A = \emptyset$, with $A \cap W$ isotopic to $f(A) \cap W$ in $W$. Identifying the isotopic curves $A \cap W$ and $f(A) \cap W$, $A \cup f(A)$ yields a new annulus $A'$, with $\partial A' \subset V$. Again, $A'$ must be isotopic to an annulus in $V$, since $V$ is incompressible. This also shows that $A$ is isotopic to $f(A)$. After further isotopy, we may assume $f(A) = A$. Now we can decompose $H$ on $A = f(A)$ to obtain a new compression body $(H, V)$ with a new induced automorphism $f$ which is the identity on $\partial_e H$. After finitely many such decompositions, we obtain a new $f : H \to H$ where $\partial_i H$ is a collection of discs. It is then clear that since $f$ is the identity on $\partial_e H$, it is isotopic to the identity on $\partial Q$, hence on the disconnected handlebody $Q$. $\square$

Now we state the classification theorem for compression bodies. We need a notion of generic automorphisms for compression bodies:

**Definition 3.3.** Suppose $f : (H, V) \to (H, V)$ is an automorphism of a connected compression body which is neither a ball nor a product compression body. If $f$ has no reducing surfaces, then $f$ is *generic.*

**Theorem 3.4.** *Suppose $f : (H, V) \to (H, V)$ is an automorphism of a connected compression body. Then the automorphism is*
*1) reducible,*
*2) periodic, or*
*3) a generic automomorphism $f : (H, V) \to (H, V)$.*

*Proof.* Suppose $f : (H, V) \to (H, V)$ is an automorphism of a connected compression body, and suppose that $\partial_e f$ has a canonical minimal reducing curve system $\mathcal{R}$ cutting $W = \partial_e(H, V)$ into finitely many surfaces $W_i$. We may suppose that $\cup W_i = \partial H - \mathring{N}(\mathcal{R})$. For each $W_i$, there exists a $k \geq 1$ such that $(\partial f)^k$ is either pseudo-Anosov or periodic on $W_i$. We replace $f$ by a finite power of itself as necessary, so that $\partial_e f$ maps each $W_i$ to itself.

Here is a list of obviously exhaustive and mutually exclusive cases:

A) $\mathcal{R} = \emptyset$, and there is only one $W_i$, say $W_0 = W$ , and on $W$ the automorphism $\partial_e f$ is either

(i) periodic or

(ii) pseudo-Anosov.

B) $\mathcal{R} \neq \emptyset$ (no $W_i = W$) and either

(i) some $W_i$ is compressible, or

(ii) all $W_i$ are incompressible.

In case A(i), by Lemma 3.1 $f$ is periodic on $H$.

In case A(ii), if $W$ is compressible, then by definition, $f$ is generic or there is a compressible reducing surface. If $W$ is incompressible, then $(H, V)$ is a product compression body and $f$ has a peripheral reducing surface coming from an invariant $I$-bundle pair.

In case B(i), if $W_0$ is compressible, then $W_0$ together with its iterates gives an $f$-invariant surface $S$, and by Corollary 2.10 the associated characteristic compression body $(Q,T)$ yields an incompressible reducing surface.

In case B(ii) the $W_i$'s are all incompressible, and we apply Proposition 2.17, to show $f$ is reducible. ☐

# 4 Decomposition

The purpose of this section is to describe a decomposition of a reducible automorphism of a compression body into automorphisms of sub- compression bodies of the compression body, and automorphisms of $I$-bundle pairs. We state our analogue of Theorem 1.2 immediately.

**Theorem 4.1.** *Suppose $f:(H,V)\to(H,V)$ is an automorphism of a connected spotless compression body. Then the compression body can be decomposed hierarchically on reducing surfaces into connected compression bodies and $I$-bundle pairs, with induced diffeomorhisms of the following kinds. For each element of the decomposition, some power of $f$ induces an automorphism $\hat f$ as described:*

*1) A spotless handlebody $H$ on which $\hat f$ is a generic automorphism.*

*2) A spotless compression body $(Q,V)$, with induced automorphism $\hat f$ periodic.*

*3) A spotless compression body $(Q,V)$, $V\neq\emptyset$, with induced automorphism $\hat f$ generic on the compression body.*

*4) An $I$-bundle pair $(Q,V)$ associated to the $I$-bundle $p:Q\to S$ with induced automorphism $\hat f$ the lift of a pseudo-Anosov or periodic automorphism on the base surface $S$.*

*5) A spotted product or spotted ball with some induced automorphism $\hat f$.*

**Remark 4.2.** In (5) of the statement, automorphisms of spotted balls and products can be further decomposed. If $(P,V,\Delta)$ is a product and $f:(P,V,\Delta)\to(P,V,\Delta)$ is pseudo-Anosov on $\partial P-(\mathring{V}\cup\Delta)$, then, for example, no further non-trivial decompositions are possible. We choose not to examine the details of the decomposition of automorphisms of spotted products and spotted balls.

The idea of the proof of Theorem 4.1 should be clear: If $f:(H,V)\to(H,V)$ has a reducing surface, we use the reducing surface to decompose the automorphism and consider induced automorphisms of resulting compression bodies and $I$-bundle pairs, inductively decomposing as often as possible.

We will need the following fact.

**Lemma 4.3.** *a) Suppose $H$ is a handlebody, and $W\subset\partial H$ is an essential subsurface of $\partial H$. Suppose $Q\hookrightarrow H$ is a compression body in $H$ (not necessarily characteristic) with $\partial_e Q=W$. Then $cl(H-Q)$ is a union of handlebodies and balls.*

*b) Suppose $(H, V)$ is a spotless compression body, and $W \subset \partial_e(H, V)$ is an essential subsurface of $\partial_e(H, V)$. Suppose $Q \hookrightarrow H$ is a compression body in $(H, V)$ (not necessarily characteristic) with $\partial_e Q = W$. Then $(cl(H - Q), V)$ is a spotless compression body.*

*Proof.* To prove a), we consider $Q$, which is constructed as $F \times I$ with handles attached to $F \times 1$ and with balls capping sphere components of $F \times 0$. The surface $U$ is the surface $F$ with sphere components removed. Let $\mathcal{E}$ be the set of compressing discs in $Q$ of $\partial_e Q$ dual to the handles attached to $F \times 1$. Cutting $Q$ on the discs of $\mathcal{E}$ yields the product $F \times I$ with balls capping sphere components of $F \times 0$, so cutting $H$ on the discs of $\mathcal{E}$ yields a space homeomorphic to $H' = \text{cl}(H - Q)$. On the other hand, a handlebody has the property that cutting on an arbitrary set of compressing discs yields handlebodies and balls. This shows that $H'$ is a spotless compression body consisting of handlebodies and balls.

The idea for the proof of b) is the same. □

In the following proposition, we explain the relationship between annular reducing surfaces and compressional reducing surfaces.

**Proposition 4.4.** *1) Suppose $f : (H, V) \to (H, V)$ is an automorphism of a spotless compression body and suppose $G$ is a horizontal ($\partial G \subset W = \partial H - \mathring{V}$) annular reducing surface for $f$. Then $f$ has a compressional reducing surface disjoint from $G$.*

*2) If $G$ is vertical annular reducing surface for $f$, then decomposing on $G$ yields an automorphism $f'$ of a new compression body $(H', V')$, where $H' = H - \mathring{N}(G)$ and $V' = V - \mathring{N}(G)$. If the underlying space $H$ is a handlebody, then the underlying space $H'$ is a handlebody.*

*3) If there is is an annular reducing surface for an automorphism $f : H \to H$ of a handlebody $H$, then there is a compressional reducing surface for $f$.*

*Proof.* 1) Let $W = \partial H - \mathring{V}$. We observe that $H$ is not a trivial compression body, a product $V \times I$, with $V = V \times 0$, since then there could not be any horizontal essential annuli, so $W$ is compressible. Let $Q$ be the compression body associated to $S = W - \mathring{N}(\partial G)$. We claim this is not a product compression body. For let $E$ be any compressing disc of $W$. Consider $E \cap G$. Eliminate trivial closed curves of intersection by isotopy. If there are any arcs of intersection $\partial$-parallel in $G$, choose an innermost arc bounding a half-disc $K$. Surgering $E$ on $K$ gives a new compressing disc (which we now call $E$) intersecting $G$ in fewer arcs. Repeating the argument, we finally arrive at a situation where all arcs of $E \cap G$ are essential in $G$. If $E$ is disjoint from $G$, we have proved our claim. Otherwise, choose an arc of intersection $\alpha$ innermost on $E$, cutting a half-disc $K$ from $E$, which is essential in an annulus $A$ of $G$. Then two copies of $K$ together with the rectangle obtained by cutting $A$ on $\alpha$ give a compressing disc for $S$. If this were not a compressing disc, $A$ would be $\partial$-parallel. Now we know the characteristic compression body $Q$ associated to $S$ is not a product, and is disjoint from $G$, hence it yields a reducing surface.

2) This statement should be clear.

3) This is the special case of 1) where $V = \emptyset$.

□

In the proof of the decomposition theorem, Theorem 4.1, we shall need a measure of the complexity of a spotted compression body or $I$-bundle pair which ensures that the process of decomposition ends with finitely many indecomposable compression bodies and $I$-bundle pairs.

**Definition 4.5.** If $H$ is a connected handlebody, we define the *genus*, genus$(H)$, as the number of 1-handles one attaches to a 0-handle to obtain $H$. If $(H, U)$, $U \neq \emptyset$, is a connected compression body, we define the *genus* as genus$(H, U) = g$ where $g$ is the number of 1-handles one attaches to $U \times 1 \subset U \times I$ to obtain $(H, U)$ from $U \times I$. For a connected $I$-bundle pair $(H, V)$ we define genus$(H, V) = 0$. For a disconnected compression body $(H, U)$ we define genus$(H, U)$ as the sum of the genera of its components, and similarly for disjoint unions of connected compression bodies and $I$-bundle pairs.

**Lemma 4.6.** *Let $(H, U)$ be a connected compression body with $U = V \cup \Delta$, where $\Delta$ is the union of discs in $U$.*

*1) If $U \neq \emptyset$, genus$(H, U) = (\chi(U) - \chi(W))/2$; if $U = \emptyset$, genus$(H, \emptyset) = (2 - \chi(W))/2$.*

*2) genus$(H, U) \geq 0$ and genus$(H, U) = 0$ if and only if*

*a) $V = \emptyset$, $H$ is a ball, and $\Delta$ consists of 0 or 1 disc, or*

*b) $V \neq \emptyset$, $\Delta = \emptyset$ and $(H, V)$ is a trivial product compression body.*

*3) genus$(H, V) \leq$ genus$(H, U)$. If $\Delta \neq \emptyset$ and $V \neq \emptyset$, then genus$(H, V) <$ genus$(H, U)$. Also, if $V = \emptyset$ and $\Delta$ contains at least two discs, then genus$(H, V) <$ genus$(H, U)$.*

*Proof.* 1) If $U \neq \emptyset$, then $(H, U)$ can be constructed by attaching 1-handles to $\partial P - U$ in the product compression body $(P, U) = (U \times I, U \times 0)$. If $g$ 1-handles are attached to $\partial P - U$, then $W = \partial_e(H, U)$ has Euler characteristic $\chi(W) = \chi(U) - 2g$. This gives the formula for $g =$ genus$(H, U)$. If $U = \emptyset$, then $H$ is a handlebody or a ball, and $H$ is obtained from a ball by attaching $g$ 1-handles. We calculate that $\chi(W) = \chi(\partial H) = 2 - 2g$, whence we obtain the formula for $g =$ genus$(H)$.

2) From the definition, clearly genus$(H, U) \geq 0$. Suppose genus$(H, U) = 0$. If $U \neq \emptyset$, $H$ can be obtained from $(U \times I, U \times 0)$ by attaching $0 =$ genus$(H, U)$ 1-handles to $U \times 1$. Hence $(H, U)$ is a product compression body. If $U$ is a disc, then $(H, U)$ is a ball with one spot. If $U = \emptyset$, $H$ is obtained by attaching $0 =$ genus$(H, \emptyset)$ 1-handles to the boundary of a ball, hence $H$ is a ball.

3) If $V \neq \emptyset$, then we have genus$(H, U) = (\chi(U) - \chi(W))/2$, genus$(H, V) = (\chi(V) - \chi(W'))/2$ where $W' = \partial_e(H, V)$. If $U$ has $k$ discs, $\chi(W') = \chi(W) + k$, $\chi(V) = \chi(U) - k$, hence genus$(H, V) =$ genus$(H, U) - k$. If $V = \emptyset$, and $U$ consists of $k$ discs, then genus$(H, U) = (\chi(U) - \chi(W))/2 = (k - \chi(W))/2$ while genus$(H, V) =$ genus$(H, \emptyset) = (2 - \chi(W'))/2 = (2 - (\chi(W) + k))/2 = (2 - k - \chi(W))/2$, hence genus$(H, V) <$ genus$(H, U)$ provided $k > 1$.

□

*Proof. (Theorem 4.1.)* Given an automorphism $f : (H, V) \to (H, V)$ which is reducible, we will decompose repeatedly to end with induced automorphisms of sub- compression bodies and $I$-bundle pairs of the types listed in the statement of the theorem. We shall use genus as a complexity to show that the decomposition must end after finitely many steps.

We decompose on reducing surfaces, and assign a definite structure as a compression body or $I$-bundle pair to each component of the cut-apart compression body. The process of decomposition will be inductive; every decomposition will replace $f : (H, V) \to (H, V)$ by several automorphisms of compression bodies and $I$-bundle pairs. We will then use $f : (H, V) \to (H, V)$ to denote an automorphism of one of these new compression bodies or $I$-bundle pairs. Whenever the decomposition yields an automorphism of a spotted product or spotted ball, we pay no further attention to that element of the decomposition, since we have chosen not to examine these further. In the special case that $f : (H, V) \to (H, V)$ has peripheral reducing surfaces, we will not actually decompose the compression body; instead we will enlarge $V$.

*Case I. The automorphism $f : (H, V) \to (H, V)$ of a compression body has a compressional reducing surface and $(H, V)$ is a compression body.* There is an invariant non-product compression body $Q$ with $\partial_e Q = S \subset \partial_e(H, V)$, either characteristic or not. Whenever possible, we use reducing surfaces without closed components. We cut off $Q$ on $T = \partial_i Q$, and it retains its structure as the compression body $(Q, T)$. The compression body $H' = \text{cl}(H - Q)$ retains the same $\partial_i H' = V$ as $H$, and the remainder of $\partial H'$ (including the copy of $T$) becomes $\partial_e H'$. With this convention, we easily check that $\text{genus}(H, V) = \text{genus}(H', V) + \text{genus}(Q, T)$. We know $\text{genus}(Q, T) > 0$.

*Case IA. genus$(H', V) > 0$.* In this case each of $(Q, T)$ and $(H', V)$ have smaller genus. If only decompositions of this type occurred, the inductive decomposition would stop after finitely many steps. If the reducing surface coming from $(Q, T)$ is peripheral, we can discard the automorphism of $(H', V)$. (If we do not discard it, it will become an automorphism of a product $I$-bundle pair after applying the reduction in Case III.)

*Case IB. genus$(H', V) = 0$.* By Lemma 4.6 and the fact that $V$ contains no discs, this occurs when $(H', V)$ is a union of spotless balls and spotless product compression bodies. We shall show later that automorphisms of spotless product compression bodies can be decomposed completely in finitely many steps.

*Case IB1. The compression body $Q$ is disconnected.* Without loss of generality we can assume that every component of $(Q, T)$ is a non-product compression body. It follows that each component has smaller genus than $(H, V)$ and we can continue the decomposition of the components of $(Q, T)$ inductively.

*Case IB2. The compression body $Q$ is connected.*

*Case IB2(i). $(H',V)$ contains a spotless ball.* Then $T$ must contain discs, and we change the structure of $(Q,T)$ so that the discs become spots in $\partial_e Q$, removing them from $T$ to obtain $S$. Thus, we view $(Q,T)$ as the spotted compression body $(Q,S,\Delta_Q)$, $\Delta_Q \neq \emptyset$. Using a spot-removing reduction, we obtain a spotless compression body $(Q,S)$. By Lemma 4.6, genus$(Q,S) <$ genus$(Q,T)$ unless $S=\emptyset$ and $T=\Delta_Q$ is single disc, attached to the spotless ball $(H',\emptyset)$. But then $T$ clearly is not a reducing surface. We are left with an automorphism of $(Q,S)$, where genus$(Q,S) <$ genus$(H,V)$, as well as automorphisms of spotted products, spotless balls, and product compression bodies. We will deal with automorphisms of product compression bodies later.

*Case IB2(ii). $(H',V)$ is a spotless product.* Writing $(Q,T) = (Q,S,\Delta_Q)$ again, where $S$ contains no discs, we consider two further subcases.

*Case IB2(ii)a. $S=\emptyset$.* In this case $(Q,S)$ is a connected handlebody or ball, $\Delta_Q \neq \emptyset$ and unless $\Delta_Q$ is a single disc and $Q$ is a ball, again by by Lemma 4.6, we can decompose $(Q,S,\Delta_Q) = (Q,\emptyset,\Delta_Q$ to obtain automorphisms of $(Q,S)=(Q,\emptyset)$ and a spotted product, with genus$(Q,S) <$ genus$(H,V)$. If $Q$ is a ball and $\Delta_Q$ is a single disc, then $T$ is not a reducing surface.

*Case IB2(ii)b. $S \neq \emptyset$.* If also $\Delta_Q \neq \emptyset$, we can again decompose $(Q,S,\Delta_Q)$ to obtain automorphisms of $(Q,S)$ and a spotted product, with genus$(Q,S) <$ genus$(H,V)$. If $\Delta_Q=\emptyset$, then the invariant curve system $\partial T$ in $W=\partial_e(H,V)$ yields a vertical annular reducing surface in $(H,V)$, and we use the argument Case IIB.

*Case II. The automorphism $f:(H,V)\to(H,V)$ of a compression body has an annular reducing surface.*

*Case IIA. The automorphism $f:(H,V)\to(H,V)$ has a horizontal annular reducing surface.* By Proposition 4.4, it also has a compressional reducing surface, and we can argue as in Case I.

*Case IIB. The automorphism $f:(H,V)\to(H,V)$ has a vertical annular reducing surface.* Decompositions on vertical annular reducing surfaces yield decompositions on reducing curve systems of $\partial_e f$, so decomposition on vertical annular surfaces is a finite process. The decompositions yield automorphisms of compression bodies, possibly including product compression bodies, with genus no larger than genus$(H,V)$. If $\partial_e f$ is periodic, it is not necessary to perform further decompositions on vertical annular reducing surfaces.

*Case III. The automorphism $f:(H,V)\to(H,V)$ of a compression body has a peripheral reducing surface coming from an invariant $I$-bundle pair $(Q,T)$.* In this case, we replace $(H,V)$ by $(Q,T)$. Either the genus is reduced, or we replace a product compression body by a product $I$-bundle pair, which cannot be decomposed further except by vertical annular reducing surfaces.

*Case IV. The automorphism $f : (H, V) \to (H, V)$ is an automorphism of an $I$-bundle pair $(H, V)$.* Since $f$ is a lift of an automorphism of a surface, and decomposition on annular surfaces corresponds to decomposition of the surface automorphism on reducing curve systems, a finite decomposition yields lifts of pseudo-Anosov or periodic surface automorphisms. □

**Remark 4.7.** The reader may wonder whether the decomposition of Theorem 4.1 is unique when the decomposition of spotted products and spotted balls is completed. There is certainly some non-uniqueness in the choice of the hierarchy. It is however possible that there is nevertheless some kind of uniqueness.

Finally, we give the proof of a theorem stated in the introduction, Theorem 1.11.

*Proof. (Theorem 1.11.)* We may assume that the canonical reducing curve system $\mathcal{R}$ is non-empty; $N(\mathcal{R})$ supports the Dehn twists of $\partial f$. The surface $\partial H - \mathcal{R}$ may or may not be incompressible. Suppose there are curves of $\mathcal{R}$ which bound discs in $H$. Letting $S$ be a disjoint union of annular neighborhoods of these curves, the associated characteristic compression body $Q$ is a union of balls, with $\partial_i Q$ containing two discs for each component. For the induced automorphism of the remaining handlebody $H' = \text{cl}(H - Q)$ there are again, in general, curves supporting Dehn twists in $\partial H'$, but these do not bound discs in $H$.

Now we consider $f' : H' \to H'$ with $\partial f'$ a composition of Dehn twists on disjoint curves in the curve system $\mathcal{R}'$, none of which bound discs in $H'$. We let $R = N(\mathcal{R}')$, a union of incompressible annuli in $\partial H'$. We are in the setting of Lemma 2.14 which guarantees the existence of an essential annulus with boundary in $R$ and with one boundary component in any given annulus of $R$. This means the characteristic manifold $N$ for the Haken pair $(H', R)$ consists of solid tori with frontier $G$ say, $G$ a union of annuli with $\partial G = \partial R$. Let $E$ be a compressing disc for $\partial H'$, isotoped to minimize $|\partial E \cap R|$. The disc $E$ may be disjoint from $G$, in which case $\partial H' - \mathring{N}$ is compressible. If $E$ intersects $G$, then we simplify intersections until only arcs essential in $G$ remain. Now an innermost arc $\gamma$ in $E$ cuts a disc $K$ from $E$, and the disc $K$ cannot lie in $N$. Two copies of $K$ and the rectangle obtained by cutting an annulus of $G$ on $\gamma$ yield a compressing disc for $\partial H' - \mathring{N}$. We have shown that $S = \partial H' - \mathring{N} = \partial H - \mathring{R}$ is compressible, so there is an associated invariant non-product characteristic compression body $Q$. The automorphism on $\partial_e Q$ is the identity, so the induced automorphism on $Q$ is the identity. In fact, this shows that $Q$ must be $H' - \mathring{N}$ up to isotopy, otherwise a repetition of our argument with $H'$ replaced by the handlebody obtained by removing $Q$ would show that there are still compressing discs of $\partial H' - \mathring{R}$.

Now we can reconstruct the automorphism $f' : H' \to H'$ and the automorphism $f : H \to H$. Regluing the annuli of $\partial_i Q$ to the solid tori of $N$, we can introduce only Dehn twists on annuli. Regluing discs in $\partial H'$ to the balls obtained in the first step of the decomposition, one can introduce only Dehn twists on discs. $\square$

# 5 Invariant 2-Dimensional Laminations

The next task is to begin the construction of invariant 2-dimensional laminations for generic automorphisms of handlebodies or compression bodies. The idea for the construction of 2-dimensional laminations is the same as in [12]. Throughout this section and the following sections, we will deal with a connected, spotless compression body $(H, V)$. The surface $\partial H - \overset{\circ}{V}$ will always be denoted $W$. Also, throughout the section, we assume genus$(H, V) > 0$. To extract the main ideas, it is useful to think of the case of a handlebody.

Let $(H_0, V_0) \subset \overset{\circ}{H}$ be a "concentric compression body" with a product structure on its complement. If the compression body has the form $V \times [0, 1]$ with handles attached to $V \times 1$, $H_0 = V \times [0, 1/2]$ with handles attached to $V \times 1/2$ in such a way that $H - H_0 \cup \text{fr}(H_0)$ has the structure of a product $W \times [0, 1]$. Here $V = V \times 0$ and $W = W \times 1$.

The goal of this section is to prove the following preliminary result; it guarantees the existence of an invariant lamination. Later, we shall modify the invariant laminations in order further to improve their properties.

**Proposition 5.1.** *Suppose $f : (H, V) \to (H, V)$ is generic automorphism of an 3-dimensional compression body. Then there is a 2-dimensional measured lamination $\Lambda \subset \overset{\circ}{H}$ with transverse measure $\mu$, such that $f(\Lambda, \mu) = (\Lambda, \lambda\mu)$, up to isotopy, for some stretch factor $\lambda > 1$. The leaves of $\Lambda$ are planes. Further, $\Lambda$ "fills" $H_0$, in the sense that each component of $H_0 - \Lambda$ is either contractible or deformation retracts to $V$.*

The statement of Proposition 5.1 will be slightly expanded later, to yield Proposition 5.6.

An automorphism $f : (H, V) \to (H, V)$ is called *outward expanding* with respect to $(H_0, V)$ if $f|_{W \times I} = f|_{W \times 0} \times h$, where $h : I \to I$ is a homeomorphism moving every point towards the fixed point 1, so that $h(1) = 1$ and $h(t) > t$ for all $t < 1$. We define $H_t = f^t(H_0)$ for all integers $t \geq 0$ and reparametrize the interval $[0, 1)$ in the product $W \times [0, 1]$ such that $\partial H_t = W \times t = W_t$, and the parameter $t$ now takes values in $[0, \infty)$. Finally, for any $t \geq 0$ we define $(H_t, V)$ to be the compression body cut from $(H, V)$ by $W \times t = W_t$.

Since any automorphism $f : (H, V) \to (H, V)$ of a compression body, after a suitable isotopy, agrees within a collar $W \times I$ of $W \subset \partial H$ with a product homeomorphism, a further "vertical" isotopy of $f$ within this collar gives:

**Lemma 5.2.** *Every automorphism of a compression body is isotopic to an outward expanding automorphism.*

Henceforth, we shall always assume that automorphisms $f$ of compression bodies have been isotoped such that they are outward expanding.

Let $\mathcal{E} = \{E_i, i = 1 \dots q\}$. be a collection of discs essential in $(H_0, V)$, with $\partial E_i \subset W_0$, cutting $H_0$ into a product of the form $V \times I$, possibly together with one or more balls. Such a collection of discs is called a *complete* collection of discs. (When $V = \emptyset$, $\mathcal{E}$ cuts $H$ into one or more balls.) We abuse notation by also using $\mathcal{E}$ to denote the union of the discs in $\mathcal{E}$. We further abuse notation by often regarding $\mathcal{E}$ as a collection of discs properly embedded in $H$ rather than in $H_0$, using the obvious identification of $H_0$ with $H$. Thus, for example, we shall speak of $W$-parallel discs in $\mathcal{E}$, meaning discs isotopic to discs in $W_0$. Corresponding to the choice of a complete $\mathcal{E}$, there is a dual object $\Gamma$ consisting of the surface $V$ with a graph attached to $\text{int}(V)$ at finitely many points, which are vertices of the graph. The edges correspond to discs of $\mathcal{E}$; the vertices, except those on $V$, correspond to the complementary balls; and the surface $V$ corresponds to complementary product. If $V = \emptyset$, and $H$ is a handlebody, then $\Gamma$ is a graph. Now $H_0$ can be regarded as a regular neighborhood of $\Gamma$, when $\Gamma$ is embedded in $H_1$ naturally, with $V = V \times 0$. By isotopy of $f$ we can arrange that $f(\mathcal{E})$ is transverse to the edges of $\Gamma$, see Figure 5, and meets $H_0 = N(\Gamma)$ in discs, each isotopic to a disc of $\mathcal{E}$. Any collection $\mathcal{E}$ of discs properly embedded in $H_0$ (or $H$) with $\partial\mathcal{E} \subset W_0$ (or $\partial\mathcal{E} \subset W$), not necessarily complete and not necessarily containing only essential discs, is called *admissible* if every component of $f(\mathcal{E}) \cap H_0$ is a disc isotopic to a disc of $\mathcal{E}$. We have shown:

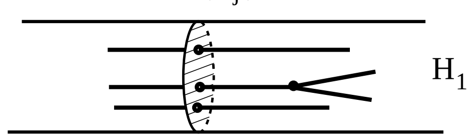



**Figure 5**

**Lemma 5.3.** *After a suitable isotopy every outward expanding automorphism $f : (H, V) \to (H, V)$ admits a complete admissible collection $\mathcal{E} \subset H_0$ as above, where every $E_i \in \mathcal{E}$ is a compressing disc of $W$.*

We shall refer to admissible collections $\mathcal{E}$ of discs $(E_i, \partial E_i) \hookrightarrow (H, W)$ as *systems* of discs. Sometimes, we shall retain the adjective "admissible" for emphasis, speaking of "admissible systems." A system may contain discs which are not compressing discs of $W$. Also, even if a system contains these $W$-parallel discs, the definition of completeness of the system remains the same. We will say a system is *$W$-parallel* if every disc in the system is $W$-parallel.

Let $P_j$ denote the punctured disc $f(E_j) - H_0$. Let $m_{ij}$ denote the number of parallel copies of $E_i$ in $f(E_j)$, and let $M = M(\mathcal{E})$ denote the matrix $(m_{ij})$, which will be called the *incidence matrix* for $\mathcal{E}$ with respect to $f$. A system $\mathcal{E}$ is *irreducible* if the incidence matrix is irreducible, see Section 7. In terms of the discs $E_i$, the system $\mathcal{E}$ is irreducible if for each $i, j$ there exists a $k \geq 1$ with

$f^k(E_j) \cap H_0$ containing at least one disc isotopic to $E_i$. It is a standard fact that a matrix $M$ with non-negative integer entries has an eigenvector $x$ with non-negative entries and that the corresponding eigenvalue $\lambda = \lambda(\mathcal{E})$ satisfies $\lambda \geq 1$. If the matrix is irreducible, the eigenvector is unique and its entries are positive.

It turns out that the lack of reducing surfaces for $f$ in $(H, V)$ is related to the existence of irreducible complete systems.

**Lemma 5.4.** *Suppose $f : (H, V) \to (H, V)$ is an automorphism of a compression body, and suppose that there is an (admissible) system which is not complete and not $W$-parallel. Then there is a reducing surface for $f$.*

*Proof.* Suppose $\mathcal{E}$ is properly embedded in $H$ (not $H_0$), and suppose it is admissible, non-complete, and not $W$-parallel. Letting $Q = \bar{N}(\mathcal{E} \cup W)$, it is easy to check that $f(Q) \subset Q$ up to isotopy. Since $f(Q)$ is a characteristic compression body for $W$ in $Q$, we conclude $f(Q) = Q$ up to isotopy. Since $\mathcal{E}$ is not $W$-parallel, $Q$ is not a product. Since $\mathcal{E}$ is not complete, $\partial_i Q$ is not isotopic to $V$. It follows that $\partial_i Q$ is a reducing surface. □

**Proposition 5.5.** *If the automorphism $f : (H, V) \to (H, V)$ is generic, then there is a complete irreducible system $\mathcal{E}$ for $f$. Also, any non-$W$-parallel complete system $\mathcal{E}$ has a complete irreducible subsystem $\mathcal{E}'$ with no $W$-parallel discs. Further, $\lambda(\mathcal{E}') \leq \lambda(\mathcal{E})$, and $\lambda(\mathcal{E}') < \lambda(\mathcal{E})$ if $\mathcal{E}$ contains $W$-parallel discs.*

*Proof.* Choose any non-$W$-parallel, admissible system $\mathcal{E}$, which must be complete by Lemma 5.4. If it contains any $W$-parallel discs, these can be eliminated as follows. Where images of other discs intersect $H_0$ in these $W$-parallel discs, $f$ can be isotoped to eliminate the intersections. This strictly reduces some entries in the incidence matrix, hence the eigenvalue $\lambda(\mathcal{E})$ is reduced, see Section 7. If the new $\mathcal{E}$ is not irreducible, the incidence matrix is reducible, which implies, that after relabelling, the incidence matrix can be made upper block diagonal, with diagonal blocks irreducible, and with more than one block on the diagonal. Then discarding all $E_i$'s except those with indices corresponding to the last diagonal block, we obtain an irreducible system with eigenvalue no larger. By Lemma 5.4, if this is not complete, then there is a reducing surface, which is a contradiction. We have actually proved the second (stronger) statement of the proposition. □

We always assume, henceforth, that $f$ is generic, so there are no reducing surfaces. Assuming $\mathcal{E}$ is *any* (complete) irreducible system of discs for $f$ in the compression body $(H, V)$, we shall construct a branched surface $B = B(\mathcal{E})$ in $\mathring{H}$. (Some background on branched surfaces will be given in Section 6; more information can be found in the references quoted there.) First we construct $B_1 = B \cap H_1$: it is obtained from $f(\mathcal{E})$ by identifying all isotopic discs of $f(\mathcal{E}) \cap H_0$, as shown schematically in Figure 6. To complete the construction of $B$ we note that $f(B_1 - \mathring{H}_0)$ can be attached to $\partial B_1$ to obtain $B_2$ and inductively, $f^i(B_1 - \mathring{H}_0)$ can be attached to $B_i$ to construct a branched surface

$B_{i+1}$. Alternatively, $B_i$ is obtained by identifying all isotopic discs of $f^i(\mathcal{E})\cap H_j$ successively for $j = i-1, \ldots, 0$. Up to isotopy, $B_i \cap H_r = B_r,\ r < i$, so we define $B = \cup_i B_i$. The branched surface $B$ is a non-compact branched surface with infinitely many sectors. (Sectors are completions of components of the complement of the branch locus.) Note that the branched surface $B$ does not have boundary on $W_i$; this follows from the irreducibility of $\mathcal{E}$. If $\mathcal{E}$ is merely admissible, the same construction works, but $B$ may have boundary on $W_i,\ i \geq 0$.

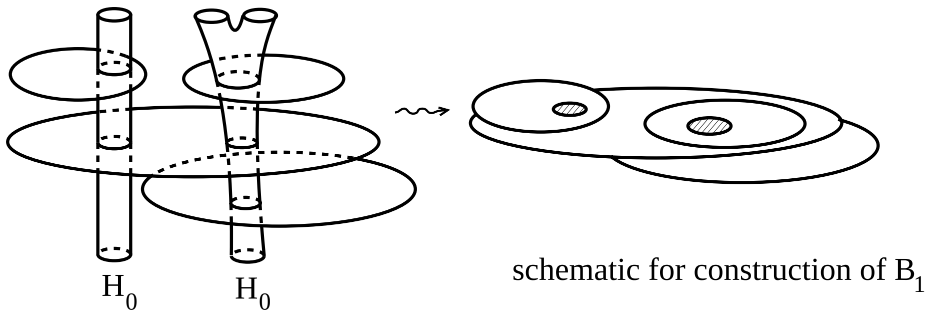



**Figure 6**

If $x$ is an eigenvector corresponding to the irreducible system $\mathcal{E}$, each component $x_i$ of the eigenvector $x$ can be regarded as a weight on the disc $E_i \in \mathcal{E}$. The eigenvector $x$ now yields an infinite weight function $w$ which assigns a positive weight to each sector:

$$w(E_i) = x_i \text{ and } w(f^t(P_i)) = x_i/\lambda^{t+1} \text{ for } t \geq 1.$$

Recall that $P_i$ is a planar surface, $P_i = f(E_i) - \mathring{H}_0$. As is well known, a weight vector defines a measured lamination $(\Lambda, \mu)$ carried by $B$ if the entries of the weight vector satisfy the *switch conditions*. This means that the weight on the sector in $H_t - H_{t-1}$ adjacent to a branch circle in $W_t$ equals the sum of weights on sectors in $H_{t+1} - H_t$ adjacent to the same branch circle, with appropriate multiplicity if a sector abuts the branch circle more than once. It is not difficult to check that our weight function satisfies the switch conditions, using the fact that it is obtained from an eigenvector for $M(\mathcal{E})$.

At this point, we have constructed a measured lamination $(\Lambda, \mu)$ which is $f$-invariant up to isotopy and fully carried by the branched surface $B$. Applying $f$, by construction we have $f(\Lambda, \mu) = (\Lambda, \lambda\mu)$. We summarize the results of our construction in the following statement, which emphasizes the fact that the lamination depends on the choice of non-$W$-parallel system:

**Proposition 5.6.** *Suppose $f : (H, V) \to (H, V)$ is a generic automorphism of a 3dimensional compression body. Given an irreducible system $\mathcal{E}$ which is not $W$-parallel, there exists a lamination $(\Lambda, \mu)$, carried by $B(\mathcal{E})$, which is uniquely determined, up to isotopy of $\Lambda$ and up to scalar multiplication of $\mu$.*

*The lamination $\Lambda$ "fills" $H_0$, in the sense that each component of the complement is either contractible or deformation retracts to $V$. Also, $\Lambda \cup W$ is closed.*

It will be important to choose the best possible systems $\mathcal{E}$ to construct our laminations. A fairly obvious requirement for a "good" system is given by the following definition. A system $\mathcal{E}$ is *efficient* if no two distinct $E_i$ are parallel and no $E_i$ is $W$-parallel.

We conclude with a lemma which will be crucial step in the proof in the existence of an invariant 2-dimensional with the properties described in Theorem 1.9, as it was a crucial step in the proof of the main theorem of [12]. The proof depends on a lemma in linear algebra whose statement was confirmed and proved by Michael Boyle, Morris Newmann, Robert C. Thompson, and Robert Williams, see Lemma 10.7 .

**Lemma 5.7.** *Suppose we are given a system $\mathcal{E}$ which is irreducible and not $W$-parallel, but not efficient. Then there exists an irreducible, efficient subsystem $\mathcal{E}'$ with $\lambda(\mathcal{E}') \leq \lambda(\mathcal{E})$. If $\mathcal{E}$ contains $W$-parallel discs, then $\lambda(\mathcal{E}') < \lambda(\mathcal{E})$.*

*Proof.* If there are parallel discs $E_a$ and $E_b$ in $\mathcal{E}$, then we can divert all occurrences of an $E_a$ in any $f(E_j)$ to $E_b$, or vice versa, by isotopy of $f$ and discard $E_a$ from $\mathcal{E}$. Each diversion affects the incidence matrix as in the hypotheses of Lemma 10.7. This lemma then guarantees that $\lambda(\mathcal{E}'') \leq \lambda(\mathcal{E})$. After finitely many such modifications we obtain a subsystem $\mathcal{E}''$, with $\lambda(\mathcal{E}'') \leq \lambda(\mathcal{E})$, which does not contain distinct discs which are parallel .

Now, using Proposition 5.5, we can replace $\mathcal{E}''$ by a subsystem $\mathcal{E}'$ of itself to obtain an irreducible subsystem without $W$-parallel discs. Since $\mathcal{E}' \subset \mathcal{E}''$ also does not contain parallel discs, it is efficient. Clearly one has $\lambda(\mathcal{E}') \leq \lambda(\mathcal{E}'')$.

Notice that, if some of the original discs in $\mathcal{E}$ are $W$-parallel, then, as a first step, we can isotope the $f(E_j)$'s to eliminate those. Since entries of the incidence matrix relative to $\mathcal{E}$ are strictly reduced, this strictly reduces the Perron-Frobenius eigenvalue $\lambda$, see Lemma 10.5 (b). ☐

# 6 Branched Surfaces

In this section we shall present some notation related to branched surfaces. Branched manifolds were first defined by R. Williams in [20]; branched 1-manifolds or train tracks were used to study laminations in surfaces by W. Thurston in various settings, e.g. in [18]; branched surfaces were used to study incompressible surfaces in 3-manifolds in [6], and were then used to study laminations in 3-manifolds in [16], [7], and in many later papers.

Suppose $B$ is the branched surface as constructed in Section 5, embedded in a compression body $H$, then a *fibered neighborhood* $N(B)$ of $B$ in $H$ is a closed regular neighborhood of $B$ foliated by interval fibers, as shown in Figure 7, with the frontier of $N(B)$ being the union of the *horizontal boundary*, $\partial_h N(B)$, and the *vertical boundary* $\partial_v N(B)$ as shown. If $B$ is locally constructed near a point on a branch circle by identifying $r$ discs along half-disc subsets of each, then the *branching order* of the branch sphere is $r$.

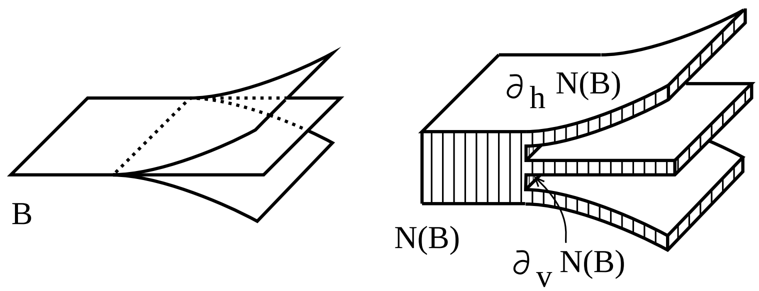


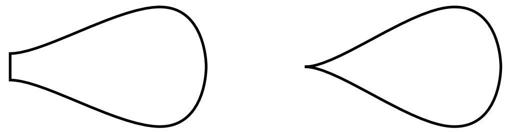

**Figure 7**

The vertical boundary $\partial_v N(B)$ is a trivial $I$-bundle over a 1-manifold as shown, where the 1-manifold is called the *maw manifold* whose components are *maw circles*. The maw circles correspond to the spaces between branches in the left side of Figure 7. There is a *projection map* $\pi : \mathring{H} \to \mathring{H}/\sim$ which maps $\mathring{H}$ to a quotient space in which fibers of $N(B)$ are contracted to points of $\mathring{H}/\sim$. Notice that $\mathring{H}/\sim$ can be identified with $\mathring{H}$, and $N(B)$ is collapsed to a branched surface which can be identified with $B$. The maw circles appear as cusps in $\mathring{H}_B$, the manifold obtained by cutting $\mathring{H}$ on $B$. More precisely, $\mathring{H}_B$ and $\mathring{H}-\mathring{N}(B)$ have homeomorphic interiors, and $\partial\mathring{H}_B$ can be obtained as a quotient of $\partial(\mathring{H}-\mathring{N}(B))$ as follows: Every component of $\partial_v(N(B)) \subset \partial(\mathring{H}-\mathring{N}(B))$, a product of type $S^1 \times I$, is collapsed along $I$-fibers to give a cusp, homeomorphic to $S^1$, in $\partial\mathring{H}_B$.

A lamination is *carried* by $B$ if it can be embedded in $N(B)$ transverse to interval fibers; it is *fully carried* if every interval fiber of $N(B)$ is intersected by the lamination.

A *0-gon for* $B$ is a disc $D$ embedded in $H - \mathring{N}(B)$ with $D \cap N(B) \subset \partial_h N(B) = \partial D$. The 0-gon is *essential* if $\partial D$ does not bound a disc in $\partial_h N(B)$. Applying the projection map $\pi$ to $H - \mathring{N}(B)$, we obtain a disc properly embedded in $H_B$. Furthermore, $\partial\pi(D)$ is disjoint from the cusp locus $\pi(\partial_v N(B)$ in $\partial H_B$, so the $\partial\pi(D)$ is smooth. We also call the disc $\pi(D)$ a *0-gon*.

A *monogon* for $B$ is an embedding $\kappa : (K,\beta,\alpha) \to (\mathring{H}-\mathring{N}(B), \partial_h N(B), \partial_v N(B))$ of a disc $K$, see Figure 8a, where $\alpha$ and $\beta$ are complementary closed arcs in $\partial K$. The arc $\alpha \subset \partial_v N(B)$ is called the *tip* of the monogon. A triple $(K,\beta,\alpha)$ as above is called a *half-disc*.



(a) (b)

**Figure 8**

Applying the projection map $\pi$ to $H - \mathring{N}(B)$, we get another point of view, see Figure 8b: if $K$ is a monogon for $B$, then $\pi(K)$, which we also denote by $K$, is a disc embedded in $H_B$ with the property that $\partial K$ intersects the cusp manifold $\pi(\partial_v N(B)) \subset \partial H_B$ transversely at a single point $\pi(\alpha)$, called the *tip*.

To construct the invariant 2-dimensional lamination of Theorem 1.9, we shall modify $\mathcal{E}$ and $B = B(\mathcal{E})$ in later sections, so that the following proposition applies.

**Proposition 6.1.** *Suppose $B = B(\mathcal{E})$ is a branched surface obtained from an admissible system, and suppose for simplicity that $\partial B(\mathcal{E}) = \emptyset$ as when $\mathcal{E}$ is irreducible. Suppose $B - \mathring{H}_0$ has no monogons or essential 0-gons in $H - \mathring{H}_0$. If $\Lambda$ is a lamination fully carried by $B$, then the leaves of $\Lambda - \mathring{H}_0$ are incompressible in $H - \mathring{H}_0$.*

Note: The branched surface $\hat{B} = B - \mathring{H}_0$ has branch locus at its boundary in $W_0$. If this causes any confusion, one can replace $\hat{B} = B - H_0$ by $B - H_{-\epsilon}$, where $H_{-\epsilon}$ is a concentric compression body slightly smaller than $H_0$.

*Proof.* Suppose $D$ is a compressing disc in $H - H_0$ for a leaf $\ell$ of $\Lambda - H_0$. The compressing disc is contained in $H_i$ for some $i$, and $\Lambda \cap H_i$ has leaves belonging to one of finitely many parallel families of discs. We form a surface $S$ in $H_i$ having one copy of a disc from each parallel family. This surface $S$ is fully carried by $B \cap H_i$. We let $\hat{S} = S - \mathring{H}_0$, and we let $\hat{B}$ denote $B - \mathring{H}_0$. Then $\hat{S}$ is fully carried by $\hat{B} \subset H_i - \mathring{H}_0$. By assumption $\hat{B}$ has no monogons or essential 0-gons. Also, it is easy to check that $\hat{B}$ has no discs of contact; therefore, using Theorem 2 of [6], we conclude that $\hat{S}$ is incompressible in $H_i - H_0$, which contradicts our assumption that $D$ was a compressing disc of $\ell$, since every component of $\ell \cap H_i$ is isotopic to a component of $\hat{S}$. □

In our setting, the "no monogons" condition is superfluous, as the following lemma shows.

**Lemma 6.2.** *Suppose $B = B(\mathcal{E})$ is a branched surface obtained from an admissible system, and suppose for simplicity that $\partial B(\mathcal{E}) = \emptyset$ as when $\mathcal{E}$ is irreducible. Suppose $B - \mathring{H}_0$ has no essential 0-gons in $H - \mathring{H}_0$. Then $B - \mathring{H}_0$ has no monogons in $H - \mathring{H}_0$.*

*Proof.* Suppose $K$ is a monogon with tip on an annulus $A$ of $\partial_v N(B - \mathring{H}_0)$. We may suppose that the annulus $A$ lies in $W_i$. The tip $\alpha \subset \partial K$ cuts $A$ to yield a rectangle $R = A - \mathring{N}(\alpha)$. Glueing to $R$ two parallel copies of $K$, i.e. components of $\text{fr}(N(K))$ where the frontier is in $H - (\mathring{N}(B) \cup H_0)$, one obtains a 0-gon $D$ for $B - \mathring{H}_0$. Since $D$ cannot be essential, it is isotopic to a disc $D'$ in $\partial_h N(B)$. This shows that there is an annulus component $A'$ of $\partial_h N(B)$, with $\partial A = \partial A'$. By construction, if $A \subset W_i$ then components of $\partial A'$ bound discs $E$ and $E'$ isotopic to discs of $f^i(\mathcal{E})$ in $H_i$ and carried by $B_i = B \cap H_i$, hence $B$ carries the sphere $A' \cup E \cup E'$. This is impossible; any connected surface carried by $B_k$, $k \geq 0$ must be a disc, a fact which can be proved by induction on $k$. □

# 7 Modification of 2-Dimensional Laminations

Given a generic automorphism $f : (H, V) \to (H, V)$, from the results of Section 5, we know that for any admissible system $\mathcal{E}$, we can construct a branched surface $B = B(\mathcal{E})$ carrying a measured invariant lamination $(\Lambda, \mu)$ with incidence matrix $M = M(\mathcal{E})$ having Perron-Frobenius eigenvalue $\lambda = \lambda(\mathcal{E}) > 1$. The lamination $\Lambda$ (as well as the measure $\mu$) is also determined by $\mathcal{E}$, so we sometimes write $\Lambda = \Lambda(\mathcal{E})$. We wish to construct an invariant lamination with better properties. We will prove the following proposition, which implies the part of Theorem 1.9 referring to 2-dimensional invariant laminations. The part of the theorem dealing with 1-dimensional invariant laminations will be proved in Section 8.

**Proposition 7.1.** *Suppose $f : (H, V) \to (H, V)$ is a generic automorphism of a compression body. Then there is a 2-dimensional measured lamination $\Lambda \subset \mathring{H}$ with transverse measure $\mu$ such that, up to isotopy, $f((\Lambda, \mu)) = (\Lambda, \lambda\mu)$ for some $\lambda > 1$. The lamination has the following properties:*

*1) Each leaf $\ell$ of $\Lambda$ is an open 2-dimensional disc.*

*2) The lamination $\Lambda$ fills $H_0$, in the sense that each component of $H_0 - \Lambda$ is contractible or deformation retracts to $V$*

*3) For each leaf $\ell$ of $\Lambda$, $\ell - \mathring{H}_0$ is incompressible in $H - (\mathring{H}_0 \cup \mathring{V})$.*

*4) $\Lambda \cup (\partial H - \mathring{V})$ is closed in $H$.*

**Remark 7.2.** In more detail, what we will prove is that there is a complete, irreducible, efficient system $\mathcal{E}$ such that the branched surface $B = B(\mathcal{E})$ has no essential 0-gons in $H - \mathring{H}_0$ and no monogons in $H - \mathring{H}_0$. It follows that $\Lambda(\mathcal{E})$ has the property that leaves of $\Lambda(\mathcal{E}) - \mathring{H}_0$ are incompressible in $\Lambda - \mathring{H}_0$. Any lamination $\Lambda(\mathcal{E})$ constructed from a complete, admissible, irreducible system $\mathcal{E}$ with minimal $\lambda(\mathcal{E})$ has this incompressibility property.

To prove Proposition 7.1 we will need to show that the lamination $\Lambda$ constructed in Section 5, starting from an automorphism $f$ of a compression body

$H$, can be replaced by one having the incompressibility property, that each leaf $\ell - H_0$ be incompressible in $H - H_0$. The proof is done by identifying good properties of $\mathcal{E}$ and showing that these are preserved by a sequence of modifications designed to eliminate compressing discs. We will show that each modification reduces or leaves unchanged the eigenvalue $\lambda(\mathcal{E})$. The eigenvalue will serve as a complexity which can be reduced by performing a sequence of modifications whenever $B(\mathcal{E})$ has an essential 0-gon in the complement of $H_0$. With care, one returns to an efficient system $\mathcal{E}$, and since eigenvalues of the corresponding bounded integer Perron-Frobenious matrices are well-ordered (Lemma 10.6), the complexity shows that the sequence of modifications will lead to a branched surface without essential 0-gons in the complement of $H_0$. Alternatively, one could initially choose a complete, admissible, irreducible system $\mathcal{E}$ with minimal $\lambda(\mathcal{E})$. Assuming that the incompressibility property does not hold, one would then obtain a contradiction. Some, but not all, of the modifications described here are the same as those described in [12].

We begin by summarizing the results of Section 5.

**Proposition 7.3.** *Suppose $f : (H, V) \to (H, V)$ is generic. There exists a non-$W$-parallel irreducible system $\mathcal{E}$, which must be complete. A non-$W$-parallel, irreducible system $\mathcal{E}$ determines a lamination $\Lambda(\mathcal{E}) = \Lambda$ fully carried by $B(\mathcal{E})$, with a transverse measure $\mu$, where the lamination is uniquely determined by $\mathcal{E}$ up to isotopy and the measure is uniquely determined up to multiplication by a scalar. The lamination is $f$-invariant, $f(\Lambda, \mu) = (\Lambda, \lambda\mu)$, with stretch factor $\lambda > 1$ equal to the Perron-Frobenius eigenvalue $\lambda(\mathcal{E})$. The measured lamination $(\Lambda, \mu)$ is fully carried by $B(\mathcal{E})$.*

Henceforth, when we say $(\Lambda, \mu)$ is unique we will always mean that $\Lambda$ is unique up to isotopy and that $\mu$ is unique up to multiplication by a scalar.

We shall show that the property of irreducibility is preserved by each of our modifications of $\mathcal{E}$; this implies that we will always be working with systems $\mathcal{E}$ such that $B(\mathcal{E})$ has no boundary. We will also show that each modification preserves the property of non-$W$-parallelity. This implies by Lemma 5.4 that each modification results in a new complete, irreducible system.

Some of our modifications will leave not only the eigenvalue $\lambda$ unchanged, but will also leave the lamination $(\Lambda, \mu)$ unchanged up to isotopy. Other more drastic modifications will change both $\lambda$ and $\Lambda$. The first modification of our system $\mathcal{E}$, which we shall use throughout this section, is called splitting.

**Definition 7.4.** We choose an $E_v$ in the system $\mathcal{E}$ for an automorphism $f$, and we suppose that the branch circle of $B = B(\mathcal{E})$ at $\partial E_v$ has branching order $r \geq 2$. At the branch circle, there are $r$ punctured discs attached in a specified order, say $P_{j_1} \dots P_{j_r}$, each $P_{j_m}$ being attached on an inner boundary $\sigma_{j_m}$. We do not rule out repetitions among the $P_{j_m}$'s, but there are no repetitions among the $\sigma_{j_m}$'s. We form a new system $\mathcal{E}'$ by replacing $E_v$ by two parallel discs $E_{v_1}$ and $E_{v_2}$. We choose an integer $s, 1 \leq s \leq r - 1$, and we attach $P_{j_1}, \dots , P_{j_s}$ to $E_{v_1}$ while attaching $P_{j_{s+1}}, \dots , P_{j_r}$ to $E_{v_2}$, see Figure 9. To preserve equivariance, we split $P_v$ and $f^t(P_v), t \geq 1$ in the same way to get the new branched surface $B' = B(\mathcal{E}')$. Clearly, the choice of splitting is actually determined by the choice of a maw circle at $\partial E_v$, namely the maw between $P_{j_s}$ and $P_{j_{s+1}}$. We say that the new system $\mathcal{E}'$ or the new branched surface $B'$ is obtained from the original system or branched surface by a *simple splitting* (at the chosen maw). There is a projection map $p : \mathcal{E}' \to \mathcal{E}$ or $p : B' \to B$, which maps both $E_{v_1}$ and $E_{v_2}$ to $E_v$ and which commutes with $f$.

A composition of a finite sequence of simple splittings is called a *splitting*.

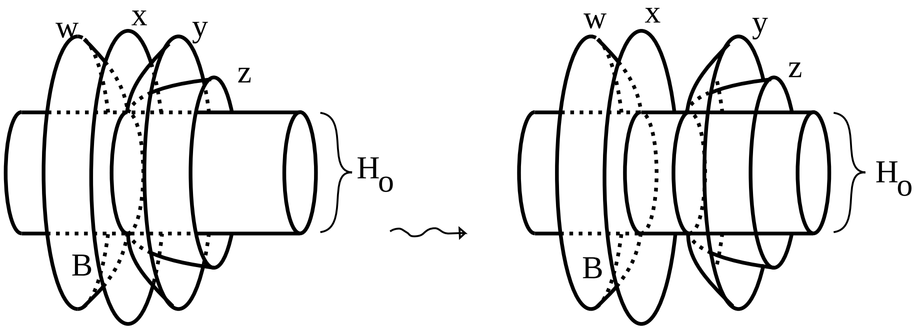



**Figure 9**

**Lemma 7.5.** *Suppose a system $\mathcal{E}$ is irreducible and not $W$-parallel. Then splitting yields a system $\mathcal{E}'$ such that*

*1) $\mathcal{E}'$ is irreducible and not $W$-parallel,*

*2) $\Lambda(\mathcal{E}')$ is isotopic to $\Lambda(\mathcal{E})$, and*

*3) $\lambda(\mathcal{E}') = \lambda(\mathcal{E})$.*

*Proof.* For a simple splitting, $\mathcal{E}'$ differs from $\mathcal{E} = \{E_1, \ldots, E_k\}$ only in that one $E_i$, say $E_1$ is r copies $E_0'$ and $E_1'$.

The fact that $\mathcal{E}'$ is non-$W$-parallel is immediate.

Let $P_{i_0}'$ and $P_{i_1}'$ be punctured discs with inner boundary on $E_0'$ and $E_1'$ respectively. Then $P_{i_0} = p(P_{i_0}')$ and $P_{i_1} = p(P_{i_1}')$ are punctured discs attached to $E_1$. To prove irreducibility, we must show that given any $r, s$, $f^t(E_r')$ contains a copy of $E_s'$ for some $t \geq 1$. We know there exists $k$ with $f^k(p(E_r'))$ containing $p(E_s')$. This proves the lemma if $s \neq 0, 1$. If $s = 0$, then we know that for some $k$, $f^k(E_r)$ contains an $E_{i_0}$, hence $f^{k+1}(E_r')$ contains an $E_s'$. Similarly if $s = 1$.

To prove statements (2) and (3), we examine eigenvectors before and after the splitting. An eigenvector $x'$ with eigenvalue $\lambda'$ for the incidence matrix $M(\mathcal{E}')$ defines a weight function $w'$ on $B' = B(\mathcal{E}')$ which satisfies the switch conditions. Hence it determines an $f$-invariant measured lamination with stretch factor $\lambda'$, and this lamination agrees with $(\Lambda', \mu')$, by Proposition 7.3. Furthermore one has $\lambda' = \lambda(\mathcal{E}')$.

But via the projection map $p : B' \to B$ the function $w'$ induces a weight function $w$ on $B = B(\mathcal{E})$ which also satisfies the switch conditions. Hence, it comes from an eigenvector $x$ of $M(\mathcal{E})$, and the corresponding eigenvalue is equal to $\lambda'$. By Proposition 7.3 the lamination determined by $w$ is equal to $(\Lambda, \mu)$, and $\lambda' = \lambda(\mathcal{E})$. The isotopy moving $(\Lambda', \mu')$ to $(\Lambda, \mu)$ corresponds to the projection of $B'$ to $B$. □

As we have mentioned, some of the modifications we will perform preserve $\lambda$ and preserve $\Lambda(\mathcal{E})$ up to isotopy. The following modification does not preserve $\lambda$ or $\Lambda(\mathcal{E})$.

**Definition 7.6.** Suppose we are given a system $\mathcal{E}$ such that there is an embedded essential 0-gon $D$ for $B$ contained in $H_1 - \mathring{H}_0$. The disc $D$ yields a compressing disc for $f(E_i) - H_0$ for some $i$, and $\partial D$ bounds a disc $D'$ in $f(E_i)$. By the irreducibility of $H$, we can then isotope $f$ to move $D' \subset f(E_i)$ to $D$, thus eliminating intersections of $f(E_i)$ with $H_0$, see Figure 10. We say that the resulting system $\mathcal{E}''$ is obtained from $\mathcal{E}$ by a *diversion*.

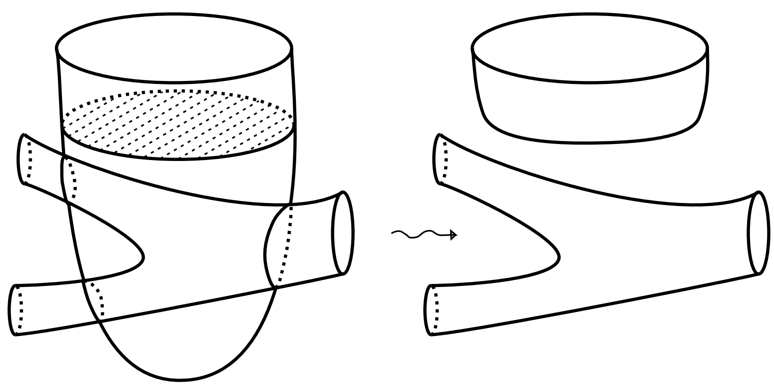


**Figure 10**

Notice that though $\mathcal{E}'$ is isotopic to $\mathcal{E}$, the incidence matrix will have changed and in general the new lamination will not be isotopic to the old one.

**Lemma 7.7.** *Suppose a system $\mathcal{E}$ is irreducible and not $W$-parallel. Then a diversion yields a system $\mathcal{E}''$ with a subsystem called $\mathcal{E}'$ such that*
*1) $\mathcal{E}'$ is irreducible and not $W$-parallel, and*
*2) $\lambda(\mathcal{E}') < \lambda(\mathcal{E})$.*

*Proof.* After performing a diversion we get a system $\mathcal{E}''$ which is isotopic to $\mathcal{E}$. In the diversion we do not eliminate any disc $E_i$, even if after the diversion the branching order at $\partial E_i$ is zero. In particular, since $\mathcal{E}$ is not $W$-parallel, neither is $\mathcal{E}''$. The incidence matrix has some entries in one column reduced. It follows from Lemma 10.5 that the new incidence matrix has strictly smaller Perron-Frobenius eigenvalue. We can then pass to an irreducible and non-$W$-parallel subsystem by Proposition 5.5 without increasing $\lambda$. □

Suppose $D$ is an (essential) 0-gon for $B$ in $H - H_0$. We can always replace our system $\mathcal{E}$ with one which has no $W$-parallel discs, without increasing $\lambda$. We may then remove trivial curves of the intersection $D \cap (\cup W_i)$ by isotopy of $D$. The *complexity* of a 0-gon is the number of arcs of intersection of $D \cap (\cup W_i)$. Our next two modifications are designed to reduce the complexity of a compressing disc. The half-disc in each definition will be a half-disc cut from $D$ by an innermost arc of intersection with $\cup W_i$.

**Definition 7.8.** Suppose $K$ is an embedded half-disc in $H_1 - \mathring{H}_0$, with $u$ and $v$ complementary closed arcs in $\partial K$ having endpoints in common, and with $v \subset B, u \subset W_0$, see Figure 11. Suppose that the points of $\partial u = \partial v$ lie on $\partial E_a$ and $\partial E_b$, and suppose $B$ has branching order 1 at $\partial E_a$ and at $\partial E_b$. Suppose finally that $E_a$ and $E_b$ are distinct. We define a *down-move on the half-disc* $K$ changing $\mathcal{E}$ to $\mathcal{E}'$. We isotope the arc $v$ to $u$ and a little beyond through $K$, performing the isotopy equivariantly and extending the isotopy to $\mathcal{E}$. This has the effect of performing an ambient boundary connect sum (often called a *band sum*) joining $E_a$ and $E_b$ to yield $E_a \# E_b$, while also joining $P_a$ to $P_b$, and $f^t(P_a)$ to $f^t(P_b)$. We discard $E_a$ and $E_b$ from $\mathcal{E}$ and add $E'_c = E_a \# E_b$ to obtain $\mathcal{E}'$.

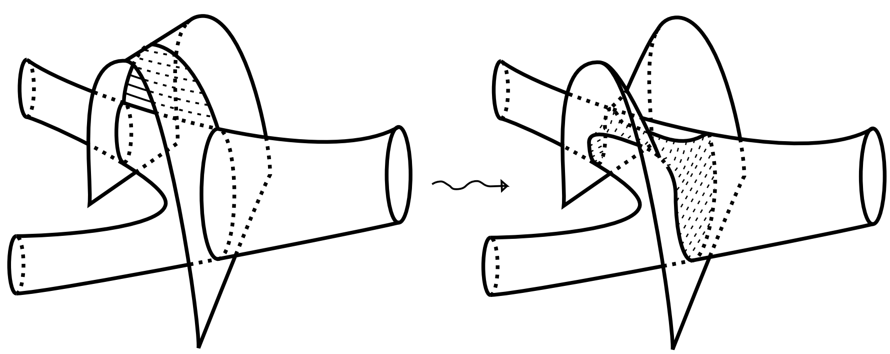


**Figure 11**

**Lemma 7.9.** *Suppose a system $\mathcal{E}$ is irreducible and not $W$-parallel. Then a down-move on a half-disc $K$ yields a system $\mathcal{E}'$ such that*

*1) $\mathcal{E}'$ is irreducible and not $W$-parallel,*
*2) $\Lambda(\mathcal{E}')$ is isotopic to $\Lambda(\mathcal{E})$, and*
*3) $\lambda(\mathcal{E}') = \lambda(\mathcal{E})$.*

*Proof.* We begin with the proof of (1). To prove the irreducibility of the new system $\mathcal{E}'$, we must show that, given any indices $r, s$, $r \neq s$, for discs in $\mathcal{E}'$, $f^t(E'_r)$ contains a copy of $E'_s$ for some $t \geq 1$. If neither $E'_r$ nor $E'_s$ is $E'_c$, then this follows from the irreducibility of $\mathcal{E}$. If $E'_r = E'_c$, then since we know that $f^t(E_a)$ contains $E_s$ for some $t$, it follows that $f^t(E'_c)$ contains $E'_s = E_s$. If $E'_s = E'_c$, then we consider the punctured disc $P_j$ having $E_a$ and $E_b$ attached to its inner boundaries, and we suppose initially that $j \neq a, j \neq b$. We know that for some $t$, $f^t(E_r)$ contains $E_j$, hence $f^{t+1}(E'_r)$ contains $E'_c$. If $E_j = E_a$, say, then since $f^t(E_r)$ contains an $E_j = E_a$ it also contains an $E_b$. This is because $\partial E_a$ has branching order 1 and $f(E_j)$ is the only disc of $f(\mathcal{E})$ which contains $E_a$. Hence $f^t(E_r)$ contains $f(E_j)$ and thus $E_b$. Hence $f^t(E'_r)$ contains $E'_c$.

Next, we prove that $\mathcal{E}'$ is not $W$-parallel. Notice that, if $\mathcal{E}'$ is $W$-parallel, then in particular $f(E'_c) = f(E_a)\#f(E_b)$ is $W$-parallel, so that $f(E_a)$ and $f(E_b)$ are isotopic: Hence either both $E_a$ and $E_b$ are $W$-parallel, or neither is. The system $\mathcal{E}'$ contains discs isotopic to all discs of $\mathcal{E}$ except $E_a$ and $E_b$. Thus we need only worry about the possibility that every disc of $\mathcal{E}$ except $E_a$ and $E_b$ is $W$-parallel. Since $\mathcal{E}$ is not $W$-parallel, we know that $E_a$ and $E_b$ are not both $W$-parallel, hence neither is $W$-parallel. Because $\partial E_a$ and $\partial E_b$ have branching order one, all discs of $f(\mathcal{E})$ except possibly $f(E_j)$ meet $H_0$ only in $W$-parallel discs. Thus there is at most one disc, $f(E_j)$, of $f(\mathcal{E})$ which is not $W$-parallel: All other discs of $f(\mathcal{E})$ meet $H_0$ only in $W$-parallel discs. Hence we have shown that if there are two discs in $\mathcal{E}$ (or $f(\mathcal{E})$) which are not $W$-parallel, then there is only one, a contradiction.

The proofs of statements (2) and (3) are similar to those of (2) and (3) in Lemma 7.5. We replace the projection $p$ in this proof by the $f$-equivariant isotopy from $B(\mathcal{E})$ to $B(\mathcal{E}')$ which is given by the isotopy on $\mathcal{E}$ described in Definition 7.8. □

**Definition 7.10.** Suppose $K$ is an embedded half-disc in $H_1 - \mathring{H}_0$, with $u$ and $v$ complementary closed arcs in $\partial K$ having endpoints in common, and with $v \subset B, u \subset W_1$, see Figure 12. Suppose that $v \subset P_c \subset H_1 - \mathring{H}_0$ is an essential arc in $P_c$ and suppose $B$ has branching order 1 at $\partial P_c \cap W_1$. We will now define an *up-move on the half-disc* $K$ changing $\mathcal{E}$ to $\mathcal{E}'$. We isotope the arc $v$ to $u$ and a little beyond through $K$, performing the isotopy equivariantly and extending the isotopy to $\mathcal{E}$. This has the effect of cutting $E_c$ on a properly embedded arc to yield discs $E'_a$ and $E'_b$, while also cutting $P_c$, and $f^t(P_c)$. We discard $E_c$ from $\mathcal{E}$ and add $E'_a$ and $E'_b$ to obtain $\mathcal{E}'$.

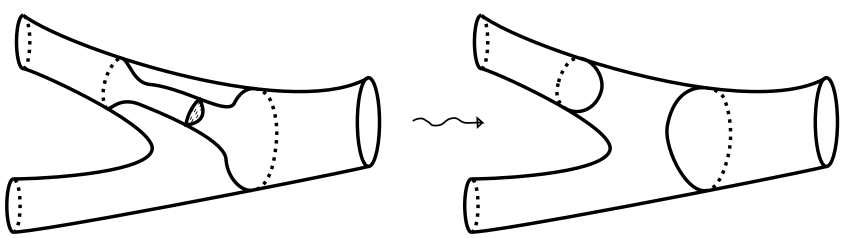


**Figure 12**

**Lemma 7.11.** *Suppose a system $\mathcal{E}$ is irreducible and not $W$-parallel. Then an up-move on a half-disc $K$ yields a system $\mathcal{E}'$ such that*

*1) $\mathcal{E}'$ is irreducible and non-$W$-parallel,*

*2) $\Lambda(\mathcal{E}')$ is isotopic to $\Lambda(\mathcal{E})$, and*

*3) $\lambda(\mathcal{E}') = \lambda(\mathcal{E})$.*

*Proof.* We begin with the proof of (1). To prove the irreducibility of the new system $\mathcal{E}'$, we must show that, given any indices $r, s$, $r \neq s$ for discs in $\mathcal{E}'$, $f^t(E'_r)$ contains a copy of $E'_s$ for some $t > 1$. If neither $E'_r$ nor $E'_s$ is $E'_a$ or $E'_b$, then this follows from the irreducibility of $\mathcal{E}$.

If $E'_r = E'_a$, but $E'_s \neq E'_a$, $E'_s \neq E'_b$, then we consider how the up-move splits $P_c$ on an essential arc to yield $P'_a$ and $P'_b$. Regarding $P'_a$ as a subset of $P_c$, it must have some $E_d$ attached to an inner boundary. Since $f^t(E_d)$ contains $E_s = E'_s$ for some $t$, we conclude that $f^{t+1}(E'_r) = f^{t+1}(E'_a)$ contains $E'_s$. The argument is similar if $E'_r = E'_b$.

If $E'_s = E'_a$, but $E'_r \neq E'_a$, $E'_r \neq E'_b$, then since $f^t(E_r)$ contains $E_c$ for some $t$, $f^t(E'_r)$ contains $E'_a$. Similarly if $E'_s = E'_b$, but $E'_s \neq E'_a$, $E'_s \neq E'_b$.

Finally, if, say, $E'_r = E'_a$ and $E'_s = E'_b$, then regarding $P'_a$ as a subset of $P_c$, it must have some $E_d$ attached to its inner boundary. Since $E_c \subset f^t(E_d)$ for some $t$, $E'_b \subset f^{t+1}(E'_a)$.

This completes the proof of irreducibility.

Next, we prove that $\mathcal{E}'$ is not $W$-parallel. Notice that, if $\mathcal{E}'$ is $W$-parallel, then in particular $E'_a$ and $E'_b$ are $W$-parallel, which implies that $E_c$ was $W$-parallel. It follows that some disc other than $E_c$ in $\mathcal{E}$ was not $W$-parallel. Since this other disc of $\mathcal{E}$ becomes a disc of $\mathcal{E}'$, we conclude $\mathcal{E}'$ is not $W$-parallel.

The proofs of statements (2) and (3) are similar to those of (2) and (3) in Lemma 7.5. We replace the projection $p$ in this proof by the $f$-equivariant isotopy from $B(\mathcal{E})$ to $B(\mathcal{E}')$ which is given by the isotopy on $\mathcal{E}$ described in Definition 7.10. □

We are now ready to prove Proposition 7.1.

*Proof. (Proposition 7.1.)* We begin with an irreducible and non-$W$-parallel system $\mathcal{E}$, obtained from Proposition 7.3 with eigenvalue $\lambda$. Initially, the discs of $\mathcal{E}$ are all essential discs. From $\mathcal{E}$, we construct the branched surface $B = B(\mathcal{E})$.

From Lemma 6.2, we know that if there are no 0-gons for $B - H_0$, then there are also no monogons. Assume now there is an essential 0-gon $D$ for $B - H_0$. We will do modifications on $\mathcal{E}$ with the goal of obtaining an essential 0-gon or monogon for $B$ in $H_1 - H_0$. Once this goal has been achieved, we do a diversion move, which reduces the eigenvalue $\lambda$. Throughout our argument, we can always assume that the punctured discs $P_i$ in $H_1 - \mathring{H}_0$ are incompressible in $H_1 - \mathring{H}_0$, otherwise we have reached our goal: to reduce $\lambda$ using a diversion move. Similarly, we can always assume that there is no monogon in $H_1 - \mathring{H}_0$, since a monogon yields an essential 0-gon.

Since $D$ is disjoint from $H_0$, using the fact that the $P_i$'s are incompressible, we can isotope $D$ to eliminate closed curves of intersection with $\cup_i W_i$.

Let $K$ be a half-disc cut from $D$ by an innermost arc of intersection. Suppose $\partial K = u \cup v$ where $v \subset \partial D \subset B$ and the complementary closed arc $u$ lies in $W_i$ for some $i$.

There are two cases to consider. If $K \subset H_i$, then in Case 1, $u \subset W_i$ and in Case 2, $u \subset W_{i-1}$.

*Case 1.* The arc $v$ lies in some sector $P = f^t(P_j)$, with both ends of $v$ on the outer boundary of $P$,which is a punctured disc. By applying a negative power of $f$, we may assume that $K$ lies in $H_1 - H_0$.

*Case 1a.* If $v$ is inessential in $P$, it is clear that one can isotope $v$ within $B$ out of $H_1$ and a little beyond. We extend the isotopy to all of $D$, then we eliminate closed curves of intersection of $D$ with $\cup W_i$ which may have appeared. Thus we have reduced the complexity of $D$, i.e. the number of arcs of intersection $D \cap (\cup_{i \geq 0} W_i)$. This isotopy involves only $D$, not the system $\mathcal{E}$, and we do not need to worry about equivariance.

*Case 1b.* If $v$ is essential in its punctured disc $P$, we will use the half-disc $K$ to perform an up-move. Before doing the up-move, however, we must ensure that the branching order at $\partial v$ is 1. This is done by performing a splitting as shown in Figure 13. The essential 0-gon $D$ remains essentially unchanged after the splitting. Lemma 7.5 shows that the new system $\mathcal{E}'$ which results from splitting remains irreducible and not $\partial$-parallel, with the same eigenvalue $\lambda$ and with $B(\mathcal{E}')$ carrying the same lamination $\Lambda$ up to isotopy. We rename the new system $\mathcal{E}$, for simplicity. Next we perform the up-move, using the half-disc $K$. Once again, now by Lemma 7.11, the resulting system $\mathcal{E}'$ is irreducible, with the same eigenvalue and with $B(\mathcal{E}')$ carrying the same lamination $\Lambda$ up to isotopy. Again, we have reduced the complexity of $D$.

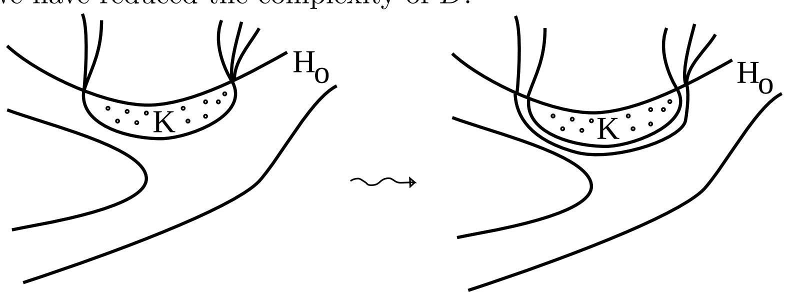



**Figure 13**

*Case 2.* The half-disc $K$ satisfies $u \subset W_{i-1}$. Applying an $f^t$ for appropriate negative $t$, we may assume that $K$ and $P$ lie in $H_1$ with $u \subset W_0$. The two points of $\partial v = \partial u$ lie in $\partial E_a$ and $\partial E_b$, say, where possibly $E_a = E_b$.

*Case 2a.* If $v$ is an inessential arc in $P$, joining an inner boundary to itself, an isotopy of $v$ in $P$ extended to $D$ eliminates an arc of intersection, as in Case 1a.

*Case 2b* In this case, $v$ is an essential arc joining inner boundary components of $P$, and the endpoints of $v$ are attached to $E_a$ and $E_b$.

*Case 2b(i).* If the essential arc $v$ joins distinct inner boundary components of $P$ which are identified with $\partial E_a \neq \partial E_b$, then we do simple splittings at the maws adjacent to the arc $v$ as shown in Figure 14, to ensure that the branching order at $\partial E_a$ and at $\partial E_b$ becomes 1. This gives a new system $\mathcal{E}'$, but Lemma 7.5 shows that $\mathcal{E}'$ is irreducible and not $\partial$-parallel, with the same eigenvalue $\lambda$, and $B(\mathcal{E}')$ carries the same lamination up to isotopy. Now we are ready to do a down-move on $K$. By Lemma 7.9, this yields a new irreducible and non-$W$-parallel system $\mathcal{E}$ with the same Perron-Frobenius eigenvalue, which gives rise to an isotopic lamination. While doing the down-move, we isotope $D$ to eliminate the arc of intersection $u$, thus we obtain a 0-gon of smaller complexity.

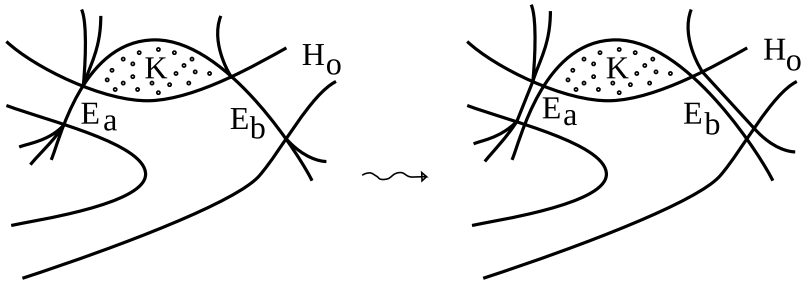



**Figure 14**

*Case 2b(ii).* If the essential arc $v$ joins distinct inner boundary components of $P$ which are identified with $\partial E_a = \partial E_b$ then we do a simple splitting at the maw(s) adjacent to the arc $v$, to ensure that $E_a$ and $E_b$ become distinct, and that the branching order at $\partial E_a$ and at $\partial E_b$ becomes 1. This splitting can be done, since, if the given $E_a$ and $E_b$ are equal, then the branching order at $\partial E_a = \partial E_b$ is necessarily bigger than 1. As usual, by Lemma 7.5 we obtain a new $\mathcal{E}$ which is irreducible and not $\partial$-parallel, with the same eigenvalue $\lambda$, and yielding the same measured lamination up to isotopy.

Now we are ready to do a down-move on $K$ exactly as in Case 2b(i) to obtain a 0-gon of smaller complexity and a new system $\mathcal{E}$ with the usual properties.

*Case 2b(iii).* In this case, the essential arc $v$ joins an inner boundary component of $P$ to itself, and the inner boundary component is identified with $\partial E_a$. If the branching order at $\partial E_a$ is greater than 1, then as before, we do a simple splitting at the adjacent maw to ensure that the branching order becomes 1. Notice that if we attempt to isotope the arc $v$ in $P$ down to $W_0$ and a little beyond, as in a down-move, we produce a component of $B \cap H_0$ which is an annulus, not a disc. According to our definition, this is not a down-move. However, we will show that it is possible to do a sequence of genuine down-moves, which removes an arc of intersection with $D$. To see this, we temporarily perform the isotopy of $v$ down to $\partial H_0$ and slightly beyond. This cuts the incompressible $P$ into two incompressible punctured discs $T$ and $R$. One of these, $T$ say, has an outer boundary, in $W_1$; the other punctured disc, $R$, has all boundary components in $W_0$. It is a standard fact that an incompressible surface such as $R$ in a product of the form $F \times I$, with all of its boundary in $F \times 0$, must be $\partial$-parallel. Regarding $R$ as a subset of the original $P$, this shows that it is possible to do a finite sequence of genuine down-moves to coalesce punctures of $P$, which correspond to punctures of $R$, with the puncture of $P$ containing $\partial v$. After performing these down-moves, the arc $v$ becomes inessential in $P$, and we are reduced to Case 1a.

This completes the case analysis. In every case, it is possible to reduce the complexity of the essential 0-gon $D$. Repeating these operations on the simpler $D$, we must ultimately obtain a 0-gon with complexity 0, i.e a $D$ contained in

$H_1 - H_0$. Now a diversion move gives a new irreducible and non-$W$-parallel system $\mathcal{E}'$ with strictly smaller eigenvalue by Lemma 7.7. Lemma 5.7 applies, and we obtain a system $\mathcal{E}_1$ and a lamination $(\Lambda_1, \mu_1)$ with $\lambda(\mathcal{E}_1) \leq \lambda(\mathcal{E}') < \lambda(\mathcal{E})$ and with $\mathcal{E}_1$ irreducible and efficient.

If the new system $\mathcal{E}_1$ is such that there is an essential 0-gon for $B(\mathcal{E}_1)$, we repeat everything we have done in the proof up to now to obtain another irreducible and efficient system $\mathcal{E}_2$ with $\lambda(\mathcal{E}_2) < \lambda(\mathcal{E}_1)$. We cannot reduce $\lambda(\mathcal{E}_i)$ infinitely often, always obtaining efficient and irreducible $\mathcal{E}$, because efficient systems have bounded cardinality, and the set of eigenvalues of non-negative integer irreducible matrices of bounded size is well-ordered, see Lemma 10.6. Thus we must finally obtain a system with $B$ having no 0-gons. $\square$

# 8 Invariant 1-Dimensional Laminations

In this section, we will construct invariant 1-dimensional laminations for generic automorphisms of handlebodies, completing the proof of Theorem 1.9. We will also attempt to give a description of generic automorphisms of handlebodies and compression bodies using the 2-dimensional invariant measured laminations $(\Lambda, \mu)$ constructed earlier. We shall restrict our attention, initially, to the case of handlebodies.

To understand to what extent the invariant lamination $(\Lambda, \mu)$ for a generic automorphism $f$ determines the automorphism $f$, we make an obvious observation. Suppose $(\Lambda, \mu)$ is the invariant lamination constructed earlier for a generic automorphism $f : H \to H$ of a handlebody. We always assume in this section that this lamination is a preferred one, with minimal $\lambda$ and with the property that leaves $\ell - \mathring{H}_0$ are incompressible in $H - \mathring{H}_0$. Let $\Lambda_j = \Lambda \cap H_j$, a measured lamination consisting of parallel families of discs of $f^j(\mathcal{E})$, the family of discs parallel to $f^j(E_i)$ having measure $x_i/\lambda^j$, where $x = (x_i)$ is the eigenvector constructed in the previous sections, and $\lambda$ is the eigenvalue. The obvious observation is that $f$ takes $\Lambda_0 \subset H_0$ to $\Lambda_1 \subset H_1$, dividing measure by $\lambda$. Identifying $H_j,\ j = 0, 1$ with $H$, we see that the automorphism is almost determined by $\Lambda$, $H_0$, and $H_1$, since $f$ must map the discs of $\Lambda_0$ to the discs of $\Lambda_1$, and these discs form complete systems $\mathcal{E}$ and $f(\mathcal{E})$ in $H_0$ and $H_1$ respectively. In fact, the automorphism is determined up to precomposing with an automorphism of $H$ which (up to isotopy) takes $\Lambda_0$ to itself (when $H$ is identified with $H_0$) and induces an automorphism of the measured lamination $\Lambda_0$. Thus, if we understand the transition from $\Lambda_0$ to $\Lambda_1$, then, to a large degree, we have understood the automorphism. We shall understand the transition using Morse theory. As a height function we use the natural height in the product $H - \mathring{H}_0$; namely, recall that we had a 1-parameter family $H_t,\ t \geq 0$ of concentric handlebodies. The parameter $t$ gives a height function assigning height $t$ to $\partial H_t$, $0 \leq t < \infty$.

**Proposition 8.1.** *(i) The branched surface $B \cap H - \mathring{H}_0$ can be isotoped to a Morse position without centers relative to the height function $t$ described above. It follows that the lamination $\Lambda$ carried by $B$ can be put into Morse position* *without centers. (ii) For every $t \geq 0$ which is not a critical value for the height function $t$ on $\Lambda$, $\Lambda_t = \Lambda \cap H_t$ is a measured lamination $(\Lambda_t, \mu_t)$ each of whose leaves is isotopic to a disc of a complete system $\mathcal{E}_t$ of essential discs in $H_t$, with every disc of $\mathcal{E}_t$ represented in $\Lambda_t$. If the measure of discs isotopic to $E_{ti} \in \mathcal{E}_t$ is $x_{ti}$, then the vector $x_t = (x_{ti})$, is an eigenvector with the same eigenvalue $\lambda$ as the eigenvector $x$.*

*Proof.* (i) The branched surface $B$ intersects $H_1 - \mathring{H}_0$ in a collection of incompressible punctured discs $P_i$, each with one boundary component in $\partial H_1$ and at least one boundary in $\partial H_0$. Letting $P$ denote the union of these $P_i$'s, we observe that $P$ is incompressible by our construction of $\Lambda$. A standard result says that such an incompressible surface in a product of the form $F \times I$, $F$ a surface, can be put in Morse position without centers (provided $P$ has no $\partial$-parallel components). We know that $P$ cannot have $\partial$-parallel components since it has boundary components on each boundary component of the product. Performing the isotopy of $P$ equivariantly, so that all images $f^i(P)$ are isomorphic, we obtain $B$ in the required position. It is obvious that any lamination $\Lambda$ carried by $B$ can then be put in Morse position with no centers; a saddle in $B$ gives a 1-parameter family of saddles of $\Lambda$.

(ii) We now consider $\Lambda \cap H_t$, assuming $t$ is not a critical value on $\Lambda$. A component of intersection of a leaf with $H_t$ must be a planar surface. If such a component $C$ is a disc and is $\partial$-parallel, then it cannot intersect $H_0$, since this would contradict incompressibility of $\Lambda - \mathring{H}_0$. Otherwise the disc is contained in $H_t - H_0$, so it would have to have a minimum, contradicting Morse position without centers.

Next we suppose that there is a component $C$ which is planar but not a disc. In this case, at least one boundary component of $C$ must bound a disc $D$ in its leaf, with $\mathring{D}$ disjoint from $C$. The disc $D$ has a collar of its boundary in $H - H_t$, with boundary at $\partial H_t$, hence it must have a maximum in $H - H_t$, a contradiction. We conclude that every leaf of $\Lambda_t = \Lambda \cap H_t$ is an essential disc.

There is a finite collection $\mathcal{E}_t$ of essential discs, each carried by $B_t = B \cap H_t$, with every leaf of $\Lambda_t$ having the same carrying map as some $E_{ti} \in \mathcal{E}_t$. Now we must show that $\mathcal{E}_t$ is complete. If $j > t$ is an integer, we know that $\mathcal{E}_j$ is complete; its complement in $H_j$ has only ball components. Further $\mathcal{E}_j \cap H_t$ is a union of discs, each a disc of $\mathcal{E}_t$, with each disc of $\mathcal{E}_t$ represented. It follows that if $H_t - \mathcal{E}_t$ contained a loop $\gamma$ homotopically non-trivial in $H_t - \mathcal{E}_t$, then $H_t - \mathcal{E}_j$ would also contain such a loop $\gamma$. It would be null-homotopic in $H_j - \mathcal{E}_j$. But it is easy to verify that for each component $X$ of $H_t - \mathcal{E}_t$ the homomorphism $\pi_1(X) \to \pi_1(H_t)$ induced by inclusion is injective, while the homomorphism $\pi_1(H_t) \to \pi_1(H_j)$ is a bijection. Hence the loop $\gamma$ must be homotopically trivial in $H_t$, a contradiction.

The truth of the statement about eigenvectors and eigenvalues should now be clear. □

We note that the proof of Proposition 8.1 does not use the full force of the incompressibility of leaves of $\Lambda - \mathring{H}_0$; it only uses the incompressibility of leaves of $\Lambda \cap (H_1 - \mathring{H}_0)$. We shall make more use of the incompressibility condition later.

**Remark 8.2.** Using Proposition 8.1 one can show that the invariant lamination $(\Lambda, \mu)$ for a generic automorphism $f$ of a handlebody $H$, as described in Theorem 1.9, corresponds to an half-infinite path in the *disc space* for $H$. For a definition of disc space, see for example [14].

Now, with $\Lambda$ in Morse position without centers, we examine the transition from $H_0$ to $H_1$. For simplicity, we can suppose that saddles of $B$ are at distinct levels. A saddle in $B$ yields a 1-parameter family of saddles of $\Lambda$. Figure 15 shows a typical move corresponding to increasing from $H_s$ to $H_t$, with with a 1-parameter family of saddles of $\Lambda$ corresponding to a saddle of $B$ lying between the levels $s$ and $t$. The move describes the relationship between $(H_s, \Lambda_s, \mu_s)$ and $(H_t, \Lambda_t, \mu_t)$. Corresponding to the lamination $\Lambda_t$, we have a dual metric graph $\Gamma_t$ embedded in $H$ such that $N(\Gamma_t)$ is a concentric handlebody in $H$. In fact, there is a map $p_t : H_t \to \Gamma_t$ collapsing non-boundary leaves and complementary regions of $\Lambda_t$ to points of $\Gamma_t$. The transverse measure on $\Lambda_t$ yields a metric on $\Gamma_t$. For $t > s$ as above, with one saddle in between, $\Gamma_t$ is obtained from $\Gamma_s$ by an *ambient interval pinch.* The ambient pinch is a homotopy $h : \Gamma_s \times I \to H$ through embeddings $h_u, 0 \leq u < 1$. For $u = 1$, $h_u$ is no longer an embedding, and the pinch is achieved by identifying two embedded intervals in $\Gamma_s$, initially sharing an endpoint, with interiors disjoint from vertices, which are mapped by $h_1$ to the same interval in $H$. The result of identification is the embedded graph $\Gamma_t$, see Figure 15. The metric is respected in the pinching.

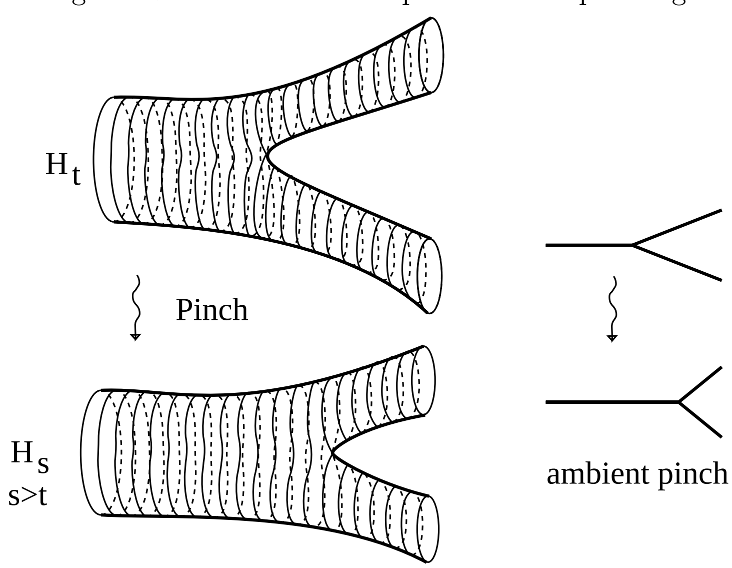



**Figure 15**

If we look at two ambient pinches in succession, more complex behavior is possible, see Figures 16 and 17.

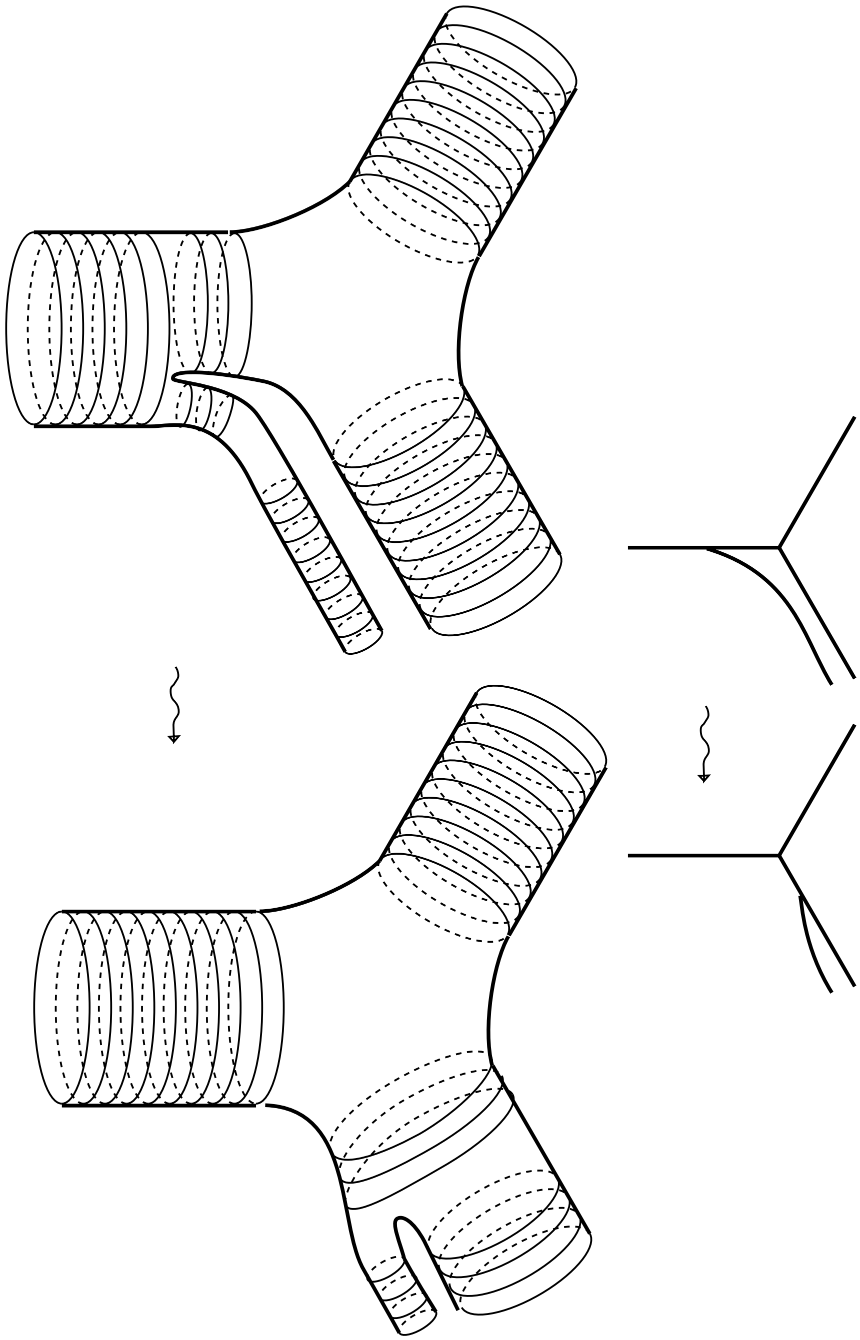


**Figure 16**

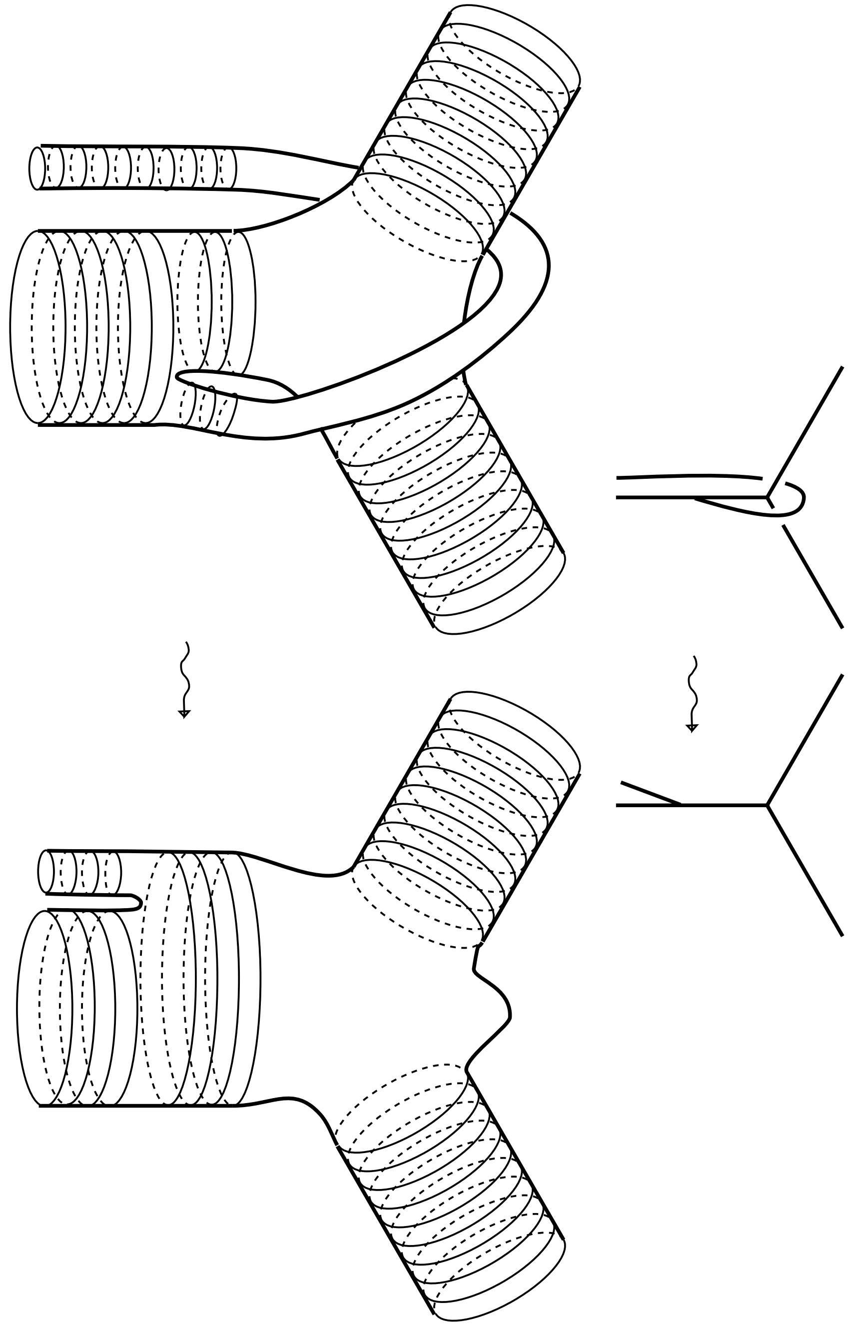


**Figure 17**

We note that the homotopy equivalence from $\Gamma_0$ to $\Gamma_1$ corresponding to the transition from $t = 0$ to $t = 1$ is, in general, far from being train track map as defined in [1].

We can define graphs $\Gamma_r$ for negative $r$ as well, with homotopy equivalences

$h_r : \Gamma_0 \to \Gamma_r$ and we can regard $\Gamma_r$ as being embedded in $H_0$ when $r < 0$. Roughly speaking, the 1-dimensional lamination $\Omega$ of Theorem 1.9 is the inverse limit as $r \to -\infty$ of the $\Gamma_r$'s.

*Proof. Conclusion, Theorem 1.9.* We construct the invariant 1-lamination $\Omega$ for a generic diffeomorphism $f$ of a handlebody using the measured 2-dimensional lamination $(\Lambda, \mu)$ constructed earlier. Recall from the statement of the theorem that $\Omega$ is not an embedded lamination. However, the lamination $\Omega$ is almost the same as the inverse limit as $r \to -\infty$ of $\Gamma_r$, or $\cap_r H_r$, again with $r \leq 0$, where the $\Gamma_r$'s are obtained from $(\Lambda, \mu)$ as in the discussion. It is clear that $\Theta = \cap_r H_r$ is $f$-invariant and closed. If $E_i \in \mathcal{E}$ we readily see that $E_i \cap (\cap_r H_r) = E_i \cap \Theta$ is a Cantor set in the disc. Thus $\Theta$ intersects a 1-handle of $H_0$ dual to $E_i$ in a 1-dimensional lamination. In the complementary 0-handles, $\Theta$ is more complex, due to the fact that handles of $H_r$ with smaller $r$ can link in the 0-handles of $H_0$.

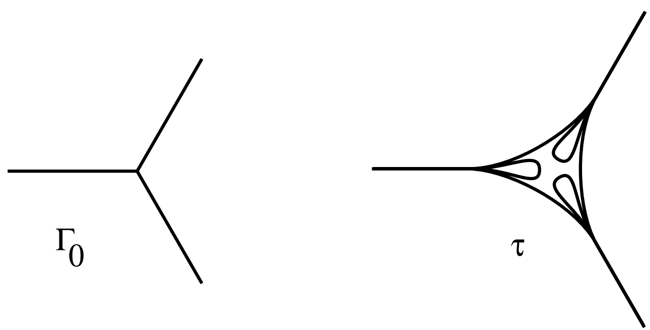



**Figure 18**

We will avoid dealing with the pathological topology of $\Theta$ by building an abstract lamination $\Omega$, and a map $\omega : \Omega \to H_0$ with $\omega(\Omega) = \Theta$. First we construct a "universal" train track $\tau$ for $H_0$ or $\Gamma_0$. The train track contains a segment $\epsilon_i$ for each edge $e_i$ of the graph $\Gamma_0$. Corresponding to each vertex $v$ of $\Gamma_0$, we have a collection of segments of $\tau$, one associated to each pair of ends of edges $e_i$ incident at the vertex. These segments are assembled as shown in Figure 18 to yield $\tau$. We have a projection $p_0 : H_0 \to \Gamma_0$ which maps 1-handles dual to discs $E_i$ of $\mathcal{E}$ to edges of $\Gamma_0$, and maps complementary 0-handles to vertices of $\Gamma_0$. Also, we have a natural embedding $\psi : \tau \to H_0$, such that $p_0 \circ \psi(\epsilon_i) = e_i$ , and $p_0 \circ \psi$ collapses all the other segments of $\tau$ to appropriate vertices. Embedding $\tau$ in $H_0$ via $\psi$, $f^{-1}(\tau)$ is carried by $\tau$ provided we allow ourselves to *homotop* $f^{-1}(\tau)$ in $H_0$ into $N(\tau)$ using a homotopy which is supported in the 0-handles together with an isotopy supported in the 1-handles. We will let $g : \tau \to \tau$ be the carrying map obtained in this way by homotoping $f^{-1}(\tau)$ into $N(\tau)$ and then composing with the projection $\pi : N(\tau) \to \tau$. We may assume that the fibers of the projection $N(\tau) \to \tau$ are discs foliating $N(\tau)$; to say that $f^{-1}(\tau)$ is carried by $\tau$ means that $f^{-1}(\tau)$ (after homotopy) is mapped into $N(\tau)$ transverse to the disc fibers. There is an incidence matrix associated to $g$; the number of times $g(\epsilon_i)$ passes over the segment $\epsilon_j$ is exactly the same as the corresponding entry $m_{ij}$ for the incidence matrix of $f$ with respect to the system $\mathcal{E}$ which was used to construct $(\Lambda, \mu)$. The eigenvector $x$ for the matrix $(m_{ij})$ assigns weights $x_i$ to the segments $\epsilon_i$ of $\tau$. We have the same eigenvalue $\lambda$ as the eigenvalue $\lambda$ associated to $(\Lambda, \mu)$.

The weights $x_i/(\lambda^s)$ on the segments $\epsilon_i$ of $\tau$ induce a weight vector $v_s$ on $\tau$ by pushing forward via $g^s : \tau \to \tau$. Regarding $x$ and $v_s$ as a functions on $\tau$, piecewise constant on edges of $\tau$, with $x = 0$ on edges of $\tau$ other than the edges $\epsilon_i$, this means that for any $y \in \tau$, $v_s(y) = \sum x(z)$, where the sum is over points $z \in g^{-s}(y)$. The weight vector $v_s$ does not satisfy the switch conditions on $\tau$, but

$$w = \lim_{s\to\infty} v_s$$

is a weight vector satisfying the switch conditions. Further, on $\epsilon_i$ the weights $w_i$ are equal to $x_i$.

The inverse system

$$\cdots \xrightarrow{g} (\tau, \frac{w}{\lambda^2}) \xrightarrow{g} (\tau, \frac{w}{\lambda}) \xrightarrow{g} (\tau, w)$$

has the property that the weights $w/(\lambda^i)$ on $\tau$ are induced by the weights $w/(\lambda^{i+1})$ via the map $g$. This means that if $y \in \tau$, the weight at $y$ is the sum of weights assigned by the weight vector $w/(\lambda^{i+1})$ to the finitely many points of $g^{-1}(y)$. Such an inverse system of train tracks (or branched manifolds) with invariant weight vectors compatible under carrying maps $g$ determines a measured lamination $(\Omega, \nu)$, where $\Omega$ is the inverse limit of the train tracks. We see that $g^{-1}$ induces an isomorphism $h : (\Omega, \nu) \to (\Omega, \nu/\lambda)$ as in the statement of Theorem 1.9.

Next, we define the map $\omega : \Omega \to H$ as follows. A point in $\Omega$ is an inverse sequence $\{y_s\} = y,\ y_s \in \tau,\ g(y_s) = y_{s-1},\ s \geq 0$. Recall that the topology for the inverse limit $\Omega$ is the smallest one such that the maps $\phi_k : \Omega \to \tau_k$ given by $\phi_k(y) = y_k$ are continuous. Let $\psi_k : \tau \to H_{-k}$ denote $f^{-k} \circ \psi$ for $k \geq 1$. We define $\omega(y) = \lim_k \psi_k \circ \phi_k$, but we must explain why the limit makes sense. We choose a metric for $H$ such that the diameters of the sets $f^{-k}(p_0^{-1}(z)) \subset H_{-k}$, $z \in \Gamma_0$, become arbitrarily small as $k \to \infty$, i.e.

$$\lim_{k\to\infty} \sup_{z\in\Gamma_0} (\mathrm{diam} f^{-k}(p_0^{-1}(z)) = 0.$$

With this assumption, by the Uniform Limit Theorem, the sequence $\psi_k \circ \phi_k$ converges uniformly, and we obtain a continuous map $\omega : \Omega \to H$. □

# 9 Questions

There is a large number of questions and problems which remain unanswered or unresolved at the time of this writing. The author may pursue some of these in the immediate or more distant future. Leonardo Navarro de Carvalho is engaged in a further study of generic automorphisms. Here are some of the questions and problems:

**Question 9.1.** *Is $(\Lambda,\mu)$ in Theorems 1.9 and Proposition 7.1 unique up to isotopy? Is the stretch factor $\lambda$ unique? The first question is important. The laminations are probably not unique up to isotopy, but perhaps it will be possible to prove uniqueness up to some equivalence relation. The possible factors $\lambda$ for laminations obtained from the construction are well-ordered. One can restrict attention to laminations with minimal $\lambda$, so the second question is less important.*

**Question 9.2.** *If $f : H \to H$ is a generic automorphism of a handlebody, what is the relationship between the minimum stretch factor $\lambda$ associated to an invariant lamination $(\Lambda,\mu)$ and the stretch factor $\lambda'$ associated to the pseudo-Anosov automorphism $\partial f$?*

**Question 9.3.** *What is the relationship between $(\Lambda,\mu)$ and the corresponding lamination for $f^{-1}$?*

**Problem 9.4.** *Complete a description of automorphisms of 3-manifolds, again in the spirit of the Nielsen-Thurston theory.*

**Problem 9.5.** *A special case of Problem 9.4 is to describe automorphisms of connected sums of $S^2 \times S^1$'s, defining reducing surfaces and generic automorphisms, then describing invariant laminations for generic automorphisms. The author has already give some thought to this problem.*

**Problem 9.6.** *Improve the description of generic automorphisms, especially generic automorphisms of compression bodies, and describe their properties.*

**Question 9.7.** *In the case of an automorphism $f : (H,V) \to (H,V)$ of a compression body, with $V \neq \emptyset$, it is fairly clear that there is a 1-dimensional measured lamination $(\Omega,\nu)$ in $\mathring{H}$ with $\Omega \cup V$ closed. What is the behavior of $\Omega$ near $V$? How does this depend on the restriction $\partial_i f$ of $f$ to $V$?*

**Problem 9.8.** *Develop methods for constructing examples of generic automorphisms of handlebodies and compression bodies. Compare these automorphisms with the automorphisms of free groups they induce. Compare irreducibility and eigenvalues. Study examples of generic automorphisms of handlebodies of genus $g$ whose induced automorphisms of the free group $F_g$ are not irreducible.*

**Problem 9.9.** *Study generic automorphisms in the context of dynamical systems, choosing suitable representatives of their isotopy classes.*

**Problem 9.10.** *Develop the Nielsen-Thurston theory of automorphisms of surfaces (with non-empty boundary), using the constructions of invariant laminations described in [12] and here.*

**Question 9.11.** *Is the decomposition of Theorem 4.1 unique in some sense?*

# 10 Linear Algebra

This section is essentially the same as a brief introduction to the theory of Perron-Frobenius matrices which was included in [12]. Lemma 10.7 was developed, with the help of experts, for use in that paper. The remaining statements are standard.

The theory of non-negative matrices is well-developed and by now classic; several expository books are available. We will use [8] as standard reference; we also recommend the short treatment in [4].

A matrix $A = (a_{ij})$ is called *non-negative* if all entries are real numbers which satisfy $a_{ij} \geq 0$. Such a matrix is called *reducible* if after a suitable permutation of indices, the matrix takes the shape of an upper triangular block matrix with more than one block. Otherwise it is called *irreducible.* For a reducible matrix, by refining the block structure if necessary, we can always assume that the diagonal blocks are irreducible.

The most basic fact in the theory is that for a non-negative matrix $A$ is that there is always a real eigenvalue $\lambda = \lambda_A \geq 0$, called the *Perron-Froebenius eigenvalue*, which realizes the spectral radius of $A$, and that there is always a corresponding eigenvector $x$, called a *Perron-Frobenius eigenvector*, which has non-negative entries. If $A$ is irreducible, then among eigenvectors $x$ with non-negative entries, $x$ is uniquely determined up to scalar multiplication.

We would like to remind the reader of the following characterization of the Perron-Frobenius eigenvalue for any (possibly reducible) non-negative matrix $A$. If $A$ is irreducible this is shown, for example, in [8], §13 (40); otherwise, use the fact that every reducible matrix is the limit of irreducible non-negative matrices:

**10.1.** *(a)* $\lambda_A = \inf\{\max\{(Aw)_k/w_k\} \mid w = (w_k)$ *a vector with positive entries*$\}$
*(b)* $\lambda_A = \sup\{\min\{(Aw)_k/w_k\} \mid w = (w_k)$ *a vector with positive entries*$\}$

The following fact follows directly from the definition of irreducibility, see [8].

**Lemma 10.2.** *Each of the following properties of the matrix $A \neq 0$ is equivalent to the irreducibility of $A$.*

*1) For every index pair $(i, j)$ there exists a $t \geq 1$ such that the $(i, j)$-th entry of $A^t$ is positive.*

*2) For any two entries $a_{i_1 j_1}$, $a_{i' j'}$ there exists a sequence of index pairs $\{(i_k, j_k)\}$, $k = 2, \dots, t$, such that $i_t = i', j_t = j'$ and for all $k \geq 1$*

*(a) $i_k = j_{k-1}$, and*

*(b) $a_{i_k j_k} > 0$.*

From 10.1 (b), considering $w = (1, 1, \dots, 1)$, and from the definition of $\lambda_A$ as spectral radius, one deduces (compare also [4], Proposition II.1.10):

**Lemma 10.3.** *Let $A = (a_{ij})$ be a non-zero, irreducible matrix with non-negative integer entries. Then the Perron-Frobenius eigenvalue satisfies $\lambda \geq 1$. If $\lambda = 1$, then $A$ is a transitive permutation matrix.*

**Lemma 10.4.** *Every non-negative matrix $A$ (possibly reducible) with $\lambda_A = 0$ satisfies $A^k = 0$ for some $k \geq 1$.*

**Lemma 10.5.** *Let $A = (a_{ij})$ and $A' = (a'_{ij})$ be two non-negative matrices of same size, with Perron-Frobenius eigenvalues $\lambda_A$ and $\lambda_{A'}$.*

*(a) If $a_{ij} \geq a'_{ij}$ holds for all indices $i, j$, then it follows $\lambda_A \geq \lambda_{A'}$.*

*(b) Assume furthermore that $A$ is irreducible and that there is at least one pair of indices with strict inequality $a_{ij} > a'_{ij}$, then it follows $\lambda_A > \lambda_{A'}$.*

*Proof.* (a) This follows directly from the definition of the Perron-Frobenius eigenvalue as spectral radius.

(b) If $A'$ is irreducible, this is proved in [8], §13.2 , Satz 6. If $A'$ is reducible, then let $A'_*$ be the irreducible diagonal block submatrix of $A'$ with $\lambda_{A'}$ as eigenvalue, and let $A_*$ be the corresponding submatrix of $A$. For $0 < \mu < 1$ let $A_\mu$ be the matrix which agrees with $A$ on $A_*$, and with $\mu A$ everywhere else. Thus $A_\mu$ is irreducible, and from the case $A'$ irreducible, considered previously, we obtain the inequalities $\lambda_{A_\mu} < \lambda_{A_{\mu'}} < \lambda_A$ if $0 < \mu < \mu' < 1$. Since $\lambda_{A'} = \lambda_{A'_*} \leq \lambda_{A_*} \leq \lambda_{A_\mu}$, this proves our claim. □

**Lemma 10.6.** *For any infinite set of non-negative irreducible integer matrices $A(n)$, all of the same size, the set of Perron-Frobenius eigenvalues $\lambda_{A(n)}$ is unbounded. In other words, the set of eigenvalues of non-negative irreducible integer matrices of a given size is well-ordered.*

*Proof.* Since $\{A(n) | n \in \mathbb{N}\}$ is infinite, there must be at least one index pair $(i, j)$ such that the entry $a_{i,j}(n)$ of $A(n)$ is unbounded for $n \in \mathbb{N}$. Let $A'(n)$ be obtained from A(n) by replacing every non-zero entry distinct from $a_{i,j}(n)$ by 1. Then $A'(n)$ is irreducible, and Lemma 10.5(a) implies $\lambda_{A(n)} \geq \lambda_{A'(n)}$. By further replacing some of the entries equal to 1 by 0's we can change each $A'(n)$ into a matrix $A''(n)$ which is obtained from a permutation matrix by changing the $(i, j)$-entry from 1 to $a''_{i,j}(n) = a_{i,j}(n)$. It satisfies $\lambda_{A'(n)} \geq \lambda_{A''(n)}$. A suitable power $A''(n)^k$ with $k \leq n!$ is a diagonal matrix with at least one diagonal entry bigger or equal to $a_{i,j}(n)$. As these are unbounded, it follows that the set $\lambda_{A''(n)}$ and hence the set $\lambda_{A(n)}$ is unbounded. □

We consulted Michael Boyle, Morris Newmann, Robert C. Thompson, and Robert Williams about the following lemma. They confirmed that the statement was correct and offered two proofs.

**Lemma 10.7.** *Let $A$ be an irreducible non-negative matrix, and, for indices $i \neq j$, let $A_i$ be obtained from $A$ by replacing the $i$-th column by the sum of the $i$-th and the $j$-th column and then erasing the $j$-th column and row. Similarly, let $A_j$ be obtained from $A$ by replacing the $j$-th column by the sum of the $i$-th and the $j$-th column, and erasing the $i$-th column and row. Then one has*

$$\lambda_{A_i} \leq \lambda_A \qquad \text{or} \qquad \lambda_{A_j} \leq \lambda_A .$$

*Proof.* (essentially Boyle's version) Let $A'_i$ be obtained from $A$ by replacing the $j$-th row by an identical copy of the $i$-th row. Let $N$ differ from the identity matrix only in that the entry $(j, i)$ is equal to 1. Then $N^{-1}A'_iN$ has as $i$-th column the sum of the $i$-th and the $j$-th column of $A$, except that all entries of its $j$-th row are equal to 0. Everywhere else it agrees with $A$. We have $\lambda_{A_i} = \lambda_{A'_i}$.

As a consequence it suffices to prove the above inequalities for $A'_i$ and $A'_j$ (defined analogously) rather than $A_i$ and $A_j$: Let the column $v$ be a Perron-Frobenius eigenvector of $A$, and suppose without loss of generality that its entries satisfy $v_i \leq v_j$. Let $v'$ be obtained from $v$ by replacing the entry $v_j$ with $v_i$. This gives $A'_iv = \lambda_A v'$. As the entries of $v'$ and $v$ satisfy $v'_k \leq v_k$, it follows $(A'_iv')_k \leq (A'_iv)_k = \lambda_A v'_k$. Since $A$ is irreducible, the eigenvector $v$ and hence the vector $v'$ has positive entries, and thus 10.1(a) directly gives $\lambda_{A'_i} \leq \lambda_A$. □

*Acknowledgements:* I thank Martin Lustig, who worked with me on a precursor of this paper [12] dealing with invariant laminations for automorphisms of higher dimensional handlebodies, and who has therefore had a considerable influence on this paper as well. In particular, Section 10, a summary of linear algebra needed in both papers, was written by Lustig for the previous paper [12]. I am grateful to Feng Luo and Leonardo Navarro de Carvalho for suggesting improvements and corrections to the manuscript. I thank Michael Boyle, Morris Newmann, Robert C. Thompson, and Robert Williams, who resolved a question about linear algebra for Lustig and myself, see Section 10. I thank the Max Planck Institute for Mathematics in Bonn for support during the year 1995-96. I also gratefully acknowledge the support of the National Science Foundation.

# References

[1] M. Bestvina and M. Handel, *Train tracks and automorphisms of free groups*, Ann. Math. **135** (1992), 1–51.

[2] F. Bonahon, *Cobordism of automorphisms of surfaces*, Ann. Sci. École Norm. Sup. (4) **16** (1983), no. 2, 237–270.

[3] A. Casson and S. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*, Student Texts, vol. 9, London Math. Soc., 1988.

[4] W. Dicks and E. Ventura, *The group fixed by a family of injective endomorphisms of a free group*, Contemporary Math. **195** (1996).

[5] A. Fathi, F. Laudenbach, and V. Poenaru, *Travaux de Thurston sur les surfaces*, vol. 66, Astérisque, 1979.

[6] W. Floyd and U. Oertel, *Incompressible surfaces via branched surfaces*, Topology **23** (1984), 117–125.

[7] D. Gabai and U. Oertel, *Essential laminations in 3-manifolds*, Ann. of Math. **130** (1989), 41–73.

[8] F. Gantmacher, *Matrizentheorie*, Springer Verlag, 1986.

[9] W. Jaco and P. Shalen, *Seifert fibered spaces in 3-manifolds*, Mem. Amer. Math. Soc (1979), no. 220, 1–192.

[10] K. Johannson, *Homotopy equivalences of 3-manifolds with boundaries*, LNM, no. 761, Springer Verlag, 1979.

[11] ______, *Topology and combinatorics of 3-manifolds*, Springer Verlag, 1995.

[12] M. Lustig and U. Oertel, *Invariant laminations for diffeomorphisms of handlebodies*, to appear, Oxford Quarterly J. of Math.

[13] H. Masur, *Measured foliations and handlebodies*, Ergod. Th. & Dynam. Sys. **6** (1986), 99–116.

[14] D. McCullough, *Virtually geometrically finite mapping class groups of 3-manifolds*, J. Diff. Geom. **33** (1991), 1–65.

[15] D. McCullough and A. Miller, *Homeomorphisms of 3-manifolds with compressible boundary*, Mem. Amer. Math. Soc. (1986), no. 344, 1–100.

[16] U. Oertel, *Measured laminations in 3-manifolds*, Trans. Am. Math. Soc. **305** (1988), no. 2, 531–573.

[17] S. Suzuki, *On homeomorphisms of a 3-dimensional handlebody*, Canadian J. Math. **29** (1977), no. 1, 111–124.

[18] W. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. A.M.S. **19** (1988), 417–431.

[19] B. Wajnryb, *Mapping class group of a handlebody*, Fundamenta Mathematicae **158** (1998), 195–228.

[20] R. Williams, *Expanding attractors*, Pub. Math. I.H.E.S. **43** (1974), 169–203.