Ergod. Th. & Dynam. Sys. (1997), **17**, 593–610
Printed in the United Kingdom © *1997 Cambridge University Press*

The forcing partial order on the three times punctured disk

MICHAEL HANDEL

Department of Mathematics and Science, Lehman College, CUNY,
Bedford Park Boulevard, Bronx, NY 10468-1589, USA

(*Received* 24 *February* 1995 *and revised* 29 *April* 1996)

Abstract. The two-dimensional analogue of the Sharkovski order on periods for maps of the interval restricts to a partial order on essential pseudo-Anosov conjugacy classes in the mapping class group of the n-times punctured disk. In this paper we give an explicit description of this restricted partial order in the case when $n = 3$.

Suppose that $f : D^2 \to D^2$ is an orientation-preserving homeomorphism of the disk and that P is a finite f-invariant set. Pseudo-Anosov theory tells us that the correct invariant for studying the implications of P on the global dynamics of f is the isotopy class of f relative to P. For example, if the relative isotopy class is pseudo-Anosov, then f has positive entropy and infinitely many periodic orbits. More precise information, such as a lower bound for the entropy of f and for the number of f-periodic points of a given period, can be obtained from a train track map ([**BH**], see also Expose 10 of [**FLP**]) for the relative isotopy class.

From the dynamics point of view, the actual location of P is irrelevant. To remove this data, we choose standard models D_n for the n-times punctured disk and a homeomorphism $h : (D^2 \setminus P) \to D_n$, where n is the cardinality of P. Conjugating f by h yields a homeomorphism $f' = hfh^{-1} : D_n \to D_n$ of the model space whose isotopy class captures all the dynamics of the isotopy class of f relative to P, but is independent of the location of P. The homeomorphism f' is only well defined up to conjugacy so we are assigning to P not an isotopy class but a conjugacy class of an isotopy class. Thus, the natural object to associate to P is a conjugacy class in the mapping class group $\mathrm{MCG}(D_n)$. Following Boyland [**Boy**] we refer to this conjugacy class as the *braid type* of P.

The crucial isotopy classes are the *essential* pseudo-Anosov ones (defined below), and we restrict our attention to these. Denote conjugacy class by $[\cdot]$. We say that $f : D^2 \to D^2$ *exhibits* $[mc]$ if there is a f-invariant set P whose braid type equals $[mc]$. The braid type of P determines a lower bound for the *number* of periodic orbits of f. It is natural to ask if the braid type of P determines a 'lower bound' for the *braid types* exhibited by the periodic orbits of f.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

We say that $[mc_1]$ *forces* $[mc_2]$, written $[mc_1] > [mc_2]$, if every homeomorphism that exhibits $[mc_1]$ also exhibits $[mc_2]$. In other words, if $[mc_1]$ is the braid type of P and if $[mc_1] > [mc_2]$, then there is at least one f-invariant set with braid type $[mc_2]$. This defines a partial order that is the natural generalization of the Scharkovski order [**Sch**] on periods for maps of the interval. From the definition, it seems that one must check all conjugacy classes of homeomorphisms exhibiting $[mc_1]$ in order to determine if $[mc_1]$ forces $[mc_2]$. However, Asimov and Franks [**AF**] (with an extension to the C^0 case by Hall [**Hal**]) showed that if ϕ is a pseudo-Anosov representative of $[mc_1]$, then $[mc_1]$ forces $[mc_2]$ if and only if ϕ, thought of as a homeomorphism of D^2, exhibits $[mc_2]$.

The 'forcing' partial ordering on essential pseudo-Anosov conjugacy classes therefore reduces to determining all the braid types that are exhibited by a given essential pseudo-Anosov homeomorphism $\phi : D_n \to D_n$. For a more detailed discussion of this partial order see [**Boy**, §9].

As far as computation goes, there are two major obstacles. The first is that ϕ has infinitely many periodic orbits. In fact, the number of periodic points of period k grows exponentially in k. The second is that the information that can be *directly* read off a train track map is symbolic dynamic in nature, and does not include the braid type of the periodic orbits of ϕ. We can list all the periodic orbits of ϕ but we do not know what their braid types are. Because of these difficulties, not very much is known about the general nature of this partial order. (See [**Boy**] for a discussion and bibliography of known results.)

One way to simplify the problem, and thereby begin to attack it, is to restrict attention to invariant sets of a fixed cardinality n. The partial order on braid types then restricts to a partial order on certain conjugacy classes in $\mathrm{MCG}(D_n)$. A great deal is known about the mapping class group and it is natural to ask if this dynamically defined partial order on essential pseudo-Anosov conjugacy classes has a purely topological, and hopefully more computable, description. More specifically, we would like to know if the induced partial order has a nice description in terms of the known solutions to the conjugacy problem for $\mathrm{MCG}(D_n)$.

In this paper, we answer these questions in the case $n = 3$ by giving an explicit description of the partial order. The $n = 3$ case is special; if there is an explicit description of the partial order for larger n, it will not be nearly as simple as the one presented here.

A homeomorphism ϕ is *pseudo-Anosov relative to a finite invariant set* P if it satisfies all of the properties of a pseudo-Anosov homeomorphism except that it may have one prong singularities at some points of P. A mapping class relative to P is pseudo-Anosov if it contains a homeomorphism that is pseudo-Anosov relative to P. If every point in P is a one prong singularity for the foliations of ϕ, then ϕ is *essential* relative to P and the mapping class of ϕ is an essential pseudo-Anosov class. When P has cardinality three, each relative pseudo-Anosov conjugacy class is essential [**FLP**, Expose 5]. Thus for the purposes of this paper, pseudo-Anosov and essential pseudo-Anosov are the same thing.

The first step then is to enumerate the pseudo-Anosov conjugacy classes. Let $mc_L \in \mathrm{MCG}(D_3)$ be the element determined by the homeomorphism that interchanges the punctures a and b by a counter-clockwise motion as shown below. Define $mc_R \in$

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

MCG(D_3) similarly using the homeomorphism that interchanges the punctures b and c by a clockwise motion as shown in Figure 1. For any word $w = e_k \cdots e_1$ in L and R, define $mc_w = mc_{e_k} \cdots mc_{e_1} \in \text{MCG}(D_3)$. We say that w is a *pseudo-Anosov word* if it contains at least one L and one R.

LEMMA 0.1. (Matsuoka) *A conjugacy class in* MCG*(D_3) is pseudo-Anosov if and only if it is represented by mc_w for some pseudo-Anosov word w. If mc_{w_1} is conjugate to mc_{w_2}, then w_1 and w_2 differ by a cyclic permutation.*

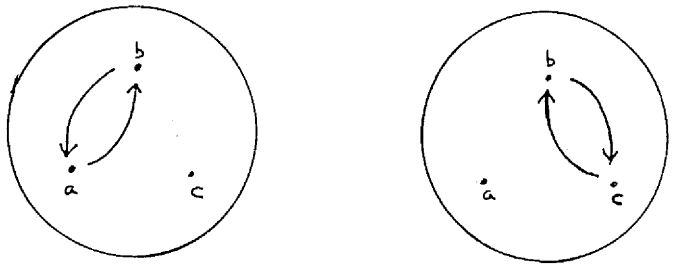


FIGURE 1.

The forcing partial order on pseudo-Anosov conjugacy classes in MCG(D_3) is easily described in terms of pseudo-Anosov words in L and R. The following theorem extends some partial results of Matsuoka [**Mat**].

THEOREM 0.2. $[mc_w] > [mc_v]$ *if and only if v is obtained from w by removing letters and by cyclic permutation.*

For example, if $w = LRLLR$, then $[mc_w] > [mc_v]$ for the following cyclic values of v : $LRLL$, $LRLR$, $RLLR$, LRL, LRR and LR.

Theorem 0.2 can be slightly strengthened by working in the relative mapping class group MCG(D_3; ∂D_3); this differs from MCG(D_3) in that all homeomorphisms and isotopies are assumed to pointwise fix ∂D_3. The center of MCG(D_3; ∂D_3) is an infinite cyclic group generated by an element Δ that is represented by a left Dehn twist about ∂D^2. The inclusion map induces a surjective homomorphism MCG(D_n, ∂D_n) $\to$ MCG(D_n) with infinite cyclic kernel generated by Δ.

For each pseudo-Anosov word w, let $\phi_w : D_3 \to D_3$ be a pseudo-Anosov homeomorphism representing mc_w. Define b_w to be the element of MCG(D_3; ∂D_3) determined by ϕ_w. The forcing partial order in MCG(D_3; ∂D_3) is defined in exactly the same way as in MCG(D_3). In particular, $[b_w] > [b]$ if and only if ϕ_w exhibits $[b]$. Moreover, the Asimov–Franks theorem implies that $[b_w] > [b_v\Delta^s]$ if and only if $[b_w\Delta^m] > [b_v\Delta^{s+m}]$. It therefore suffices to understand which conjugacy classes are forced by $[b_w]$.

If v and w are pseudo-Anosov words, then Theorem 0.2 gives necessary and sufficient conditions for the existence of an integer s such that $[b_w] > [b_v\Delta^s]$. The extra information that we gain by considering relative mapping classes is an explicit description of the possible values of s.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

Given a pair (w, s), where w is a word in the letters L and R and s is an integer, consider the following moves.

($\mathcal{R}1$) Cyclically permute w.

($\mathcal{R}2$) Remove a letter from w.

($\mathcal{R}3$) Remove LL from w and replace s by $s-1$.

($\mathcal{R}4$) Remove RR from w and replace s by $s+1$.

THEOREM 0.3. *If w and v are pseudo-Anosov words, then $[b_w] > [b_v\Delta^s]$ if and only if (v, s) is obtained from $(w, 0)$ by successive applications of ($\mathcal{R}1$)–($\mathcal{R}4$).*

For example, if $w = LRLLR$, then $[b_w] > [b_v\Delta^s]$ for the following cyclic values of (v, s): $(LRLL, 0)$, $(LRLR, 0)$, $(RLLR, 0)$, $(LRR, -1)$, $(LRL, 0)$, $(LRR, 0)$, $(LR, 0)$ and $(LR, -1)$.

Theorem 0.3 implies Theorem 0.2 so we will work in $\mathrm{MCG}(D_3; \partial D_3)$ and prove Theorem 0.3.

§1 contains some definitions, preliminary results and the proof of Lemma 0.1. The proofs of our main results are contained in §§3 and 5. In §2 we give a method for computing exhibited braid types and in §4 we use symbolic dynamics to produce invariant sets with prescribed dynamics.

1. *Preliminaries*

The region F shown in Figure 2 is divided into three 'disks', B_α, B_β and B_γ, and two 'rectangles', S_A and S_B, that are decomposed into arcs called *fibers*. The center of the disks are a, b and c respectively.

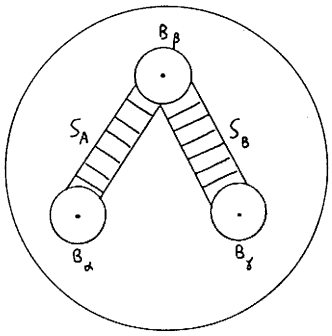


FIGURE 2.

Choose homeomorphisms f_L and f_R of D^2 with images $f_L(F)$ and $f_R(F)$ as shown in Figure 3. We assume that f_L and f_R:

(1) map a fiber into either a disk or a fiber;

(2) expand length and contract width in the rectangles;

(3) uniformly contract $B_\alpha \cup B_\beta \cup B_\gamma$ to $\{a\} \cup \{b\} \cup \{c\}$;

(4) have no periodic points in the complement of $F \cup \partial D_3$.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

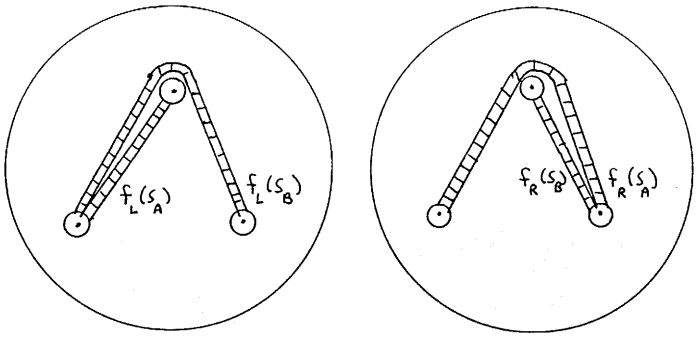


FIGURE 3.

Given any word $w = e_k \cdots e_1$ in the letters L and R, define $f_w : D_3 \to D_3$ to be the composition $f_w = f_{e_k} \cdots f_{e_1}$.

The set $\{a, b, c\}$ is f_w-invariant. Let D_3 be the disk D^2 punctured at the three points a, b and c. The mapping class group, $\mathrm{MCG}(D_3)$, of D_3 is the group of isotopy classes of orientation-preserving homeomorphisms of D_3. The relative mapping class group $\mathrm{MCG}(D_3; \partial D_3)$ is defined similarly with the added assumption that all homeomorphisms and isotopies pointwise fix ∂D^2. The restriction of f_w to D_3 (also called f_w) defines an element $mc_w \in \mathrm{MCG}(D_3)$ and an element $b_w \in \mathrm{MCG}(D_3; \partial D_3)$.

If $\{x, y, z\} \in \mathrm{int}(D^2)$ is an f_w-invariant set, then the *induced braid or orbit type* $bt(f_w, \{x, y, z\})$ is defined to be the conjugacy class in $\mathrm{MCG}(D_3; \partial D_3)$ represented by $hfh^{-1} : D_3 \to D_3$, where $h : (D \setminus \{x, y, z\}) \to D_3$ is an orientation-preserving homeomorphism.

LEMMA 1.1. *If w contains at least one L and one R, then mc_w is pseudo-Anosov. The maximal invariant set*

$$C_f(w) = \{x \in S_A \cup S_B : f_w^i(x) \in S_A \cup S_B \text{ for } \infty < i < \infty\}$$

is a Cantor set and the periodic orbits of f_w that are contained in $C_f(w)$ exhibit exactly the same braid types as any pseudo-Anosov homeomorphism representing mc_w.

Proof of Lemma 1.1. Let $G \subset F$ be the graph shown in Figure 4. G has three peripheral

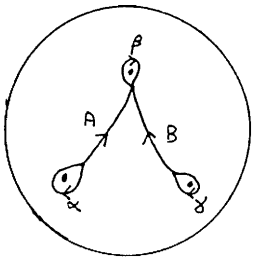


FIGURE 4.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

loops α, β and γ and two non-peripheral edges A and B. The edges A and B cross all the way through S_A and S_B and intersect the fibers transversely. The orientation on the peripheral loops is clockwise and the orientation on A and B is chosen so that β contains the terminal vertex of both edges. We use $\bar{E}$ to refer to an edge E of G with its orientation reversed.

There is a deformation retraction $r : D_3 \to G$ that sends a fiber in S_A or S_B to its unique intersection with G. Define $g_w = r \circ f_w : G \to G$. The restriction of g_w to the edges A and B need not be an immersion but we may assume that g_w maps vertices to vertices. Define $\hat{g}_w : G \to G$ to be the map that is homotopic to $g_w : G \to G$ rel vertices and that is an immersion when restricted to each edge of G.

For example, $\hat{g}_L$ is given by $A \mapsto \bar{A}, B \mapsto B\bar{\beta}\bar{A}, \alpha \mapsto \beta, \beta \mapsto \alpha, \gamma \mapsto \gamma$; and $\hat{g}_R$ is given by $A \mapsto A\beta\bar{B}, B \mapsto \bar{B}, \alpha \mapsto \alpha, \beta \mapsto \gamma, \gamma \mapsto \beta$. Note that $\hat{g}_w = \hat{g}_{e_k} \cdots \hat{g}_{e_1}$.

An easy induction argument shows that the edges in $\hat{g}_w(A)$ and $\hat{g}_w(B)$ are alternately peripheral and non-peripheral, with the first and last being non-peripheral. This implies, in the language of [**BH**], that $\hat{g}_w : G \to G$ is a train track map. The *transition matrix* for $\hat{g}_w$ is the two dimensional matrix whose ijth-entry is the number of times that $\hat{g}_w$ maps the jth non-peripheral edge across the ith non-peripheral edge in either direction. (We will follow the convention that A is the first non-peripheral edge and B is the second non-peripheral edge.) For example, the transition matrices for $\hat{g}_L$ and $\hat{g}_R$ are $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ respectively. The transition matrix for $\hat{g}_w = \hat{g}_{e_k} \cdots \hat{g}_{e_1}$ is the product of the transition matrices for the $\hat{g}_{e_i}$'s. Since w contains at least one L and one R, each entry of the transition matrix for $\hat{g}_w$ is positive. The first part of the lemma now follows from the construction in [**BH**] of a pseudo-Anosov homeomorphism ϕ_w isotopic to f_w. (A positive transition matrix for a train track map is not, in general, sufficient to guarantee that an isotopy class contains a pseudo-Anosov map; see [**BH**, Proposition 3.3.2]. In the present case, the extra 'link condition' is satisfied because $\hat{g}_w(A)$ and $\hat{g}_w(B)$ cross both A and B.)

It is an easy exercise in symbolic dynamics to check that there is a period preserving, one-to-one correspondence between the periodic orbits of f_w that are contained in $C_f(w)$ and the periodic orbits of ϕ_w. The only periodic orbits of f_w that lie in the complement of $S_A \cup S_B$ are the punctures $\{a, b, c\}$ and points in ∂D^2. The second part of the lemma now follows directly from [**AF**] and [**Hal**]. □

Following Lemma 1.1, we say that w is *a pseudo-Anosov word* if it contains at least one L and one R.

The inclusion map induces a surjective homomorphism $\mathrm{MCG}(D_3, \partial D_3) \to \mathrm{MCG}(D_3)$ with infinite cyclic kernel generated by an element Δ that is represented by a left Dehn twist about ∂D^2. Note that Δ is in the center of $\mathrm{MCG}(D_3; \partial D_3)$. The following lemma gives presentations of the mapping class groups that will be used in evaluating induced braid types.

LEMMA 1.2.

(a) $\mathrm{MCG}(D_3) \cong Z_3 * Z_2$; *$mc_L = st$ and $mc_R = s^2t$ where s generates Z_3 and t generates Z_2.*

(b) $\mathrm{MCG}(D_3; \partial D_3) \cong \{S, T | S^3 = T^2\}$; *$b_L = ST^{-1}$; $b_R = S^2T^{-1}$ and $\Delta = T^2$.*

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

Proof of Lemma 1.2. MCG(D_3; ∂D_3) is naturally identified [**Bir**] with the braid group B_3, which has the standard presentation $\{\sigma_1, \sigma_2 : \sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2\}$, where b_L is identified with σ_1^{-1}, b_R is identified with σ_2 and $\Delta = (\sigma_1\sigma_2\sigma_1)^2 = (\sigma_2\sigma_1\sigma_2)^2$.

Let S and T be the elements of MCG(D_3; ∂D_3) that correspond to $\sigma_2\sigma_1$ and $\sigma_1\sigma_2\sigma_1$ respectively. Then $b_L = ST^{-1}$, $b_R = S^2T^{-1}$, $\Delta = T^2$ and $\{TS^{-1}, S^2T^{-1} : T = S^3T^{-1}\}$ is a presentation for MCG(D_3; ∂D_3). Part (b) now follows from the fact that S and T can be expressed in terms of TS^{-1} and S^2T^{-1}.

Part (a) follows from the fact that MCG(D_3) is the quotient of MCG(D_3; ∂D_3) by the cyclic subgroup generated by $\Delta = T^2$. □

We can now describe the pseudo-Anosov conjugacy classes.

LEMMA 1.3. (Matsuoka)

(a) *Each pseudo-Anosov conjugacy class in* MCG(D_3) *is represented by* mc_w *for some pseudo-Anosov word* w. *If* mc_{w_1} *is conjugate to* mc_{w_2}, *then* w_1 *and* w_2 *differ by a cyclic permutation.*

(b) *If* $b_{w_1}\Delta^{s_1}$ *is conjugate to* $b_{w_2}\Delta^{s_2}$, *then* $s_1 = s_2$ *and* w_1 *differs from* w_2 *by a cyclic permutation.*

Proof of Lemma 1.3. Suppose that m is a pseudo-Anosov element of MCG(D_3). Apply Lemma 1.2 to decompose m into an alternating concatenation of s^i, ($i = 1, 2$), and t. Since m is pseudo-Anosov but s and t are not, there must be at least one s^i and one t in this decomposition. After a conjugacy, we may assume that the leading term is s or s^2 and that the last term is t. Replacing st with mc_L and s^2t with mc_R proves that $m = mc_w$ for some, necessarily pseudo-Anosov, word w.

If mc_{w_1} is conjugate to mc_{w_2}, then Lemma 1.2 implies that the words in s and t determined from w_1 and w_2 by replacing L with st and R with s^2t are conjugate. Theorem 2.8 of [**LS**, p. 187] then implies that w_1 and w_2 differ by a cyclic permutation. This proves (a).

To prove (b) note that the condition on w_1 and w_2 follows from part (a) so that it suffices to show that b is never conjugate to $b\Delta^s$ for $s \neq 0$. This follows from the fact that $\Delta^s = T^{2s}$ does not abelianize to the trivial element. □

COROLLARY 1.4. *Suppose that* w *and* v *are pseudo-Anosov words. Then* $[b_w] > [b_v\Delta^s]$ *if and only if* $[b_v\Delta^s] = bt(f_w, \{x, y, z\})$ *for some* f_w*-invariant set* $\{x, y, z\} \in S_A \cup S_B$.

Proof of Corollary 1.4. This follows immediately from Lemma 1.1 and the Asimov–Franks theorem ([**AF**] and [**Hal**]). □

Remark. Lemma 1.3 was originally proved by Matsuoka [**Mat**]. There is also a proof that uses train track maps and Stallings' folding operation [**Sta**]. (The connection is that $\hat{g}_L$ and $\hat{g}_R$ are the two possible full folds of G that begin by identifying non-peripheral segments.) This folding approach generalizes to provide a solution to the conjugacy problem for all n.

2. *Factoring an induced braid*

In this section, we compute induced braid types for f_L and f_R and show how these can be used to compute induced braid types for any f_w. We use the graph G and the maps

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

g_w defined in the proof of Lemma 1.1.

Throughout this and the following sections, we make use of a linear ordering on points in $A \cup B$ that is induced by identifying the oriented edge path $A\bar{B}$ in the graph G with the unit interval. Thus, for example, $\hat{x} < \hat{y}$ whenever $\hat{x} \in A$ and $\hat{y} \in B$.

If $x \in S_A \cup S_B \subset F$, denote its projection $r(x) \in A \cup B \subset G$ by $\hat{x}$.

Suppose that $\{x, y, z\} \subset \text{int}(S_A \cup S_B)$ and that $\hat{x}$, $\hat{y}$, and $\hat{z}$ are distinct points. Choose an isotopy $\phi_t : D^2 \to D^2$ satisfying:

- $\phi_0 =$ identity;
- $\phi_1(\{x, y, z\}) = \{a, b, c\}$;
- $\phi_t|(D_3 \setminus F) =$ identity for all $0 \le t \le 1$;
- $\widehat{\phi_t(x)}, \widehat{\phi_t(y)}$ and $\widehat{\phi_t(z)}$ are distinct for all $0 \le t \le 1$.

Less formally, define ϕ_t by sliding the element of $\{x, y, z\}$ nearest a to a, the element of $\{x, y, z\}$ nearest c to c and the remaining element to b; do this in a way that preserves the ordering on $\{\hat{x}, \hat{y}, \hat{z}\}$ (see Figure 5).

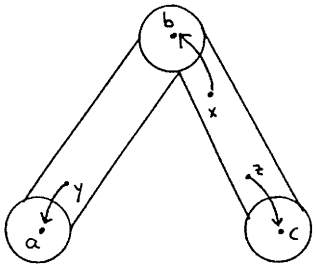


FIGURE 5.

Define $h_{\{x,y,z\}} = \phi_1 : (D^2 \setminus \{x, y, z\}) \to D_3$.

If $\{x, y, z\}$, $f_w(\{x, y, z\}) \subset \text{int}(S_A \cup S_B)$, then $h_{f_w(\{x,y,z\})} f_w h^{-1}_{\{x,y,z\}} : D_3 \to D_3$ defines an element of $\text{MCG}(D_3; \partial D_3)$ that we denote $b(w; \{x, y, z\})$.

The following lemma reduces the computation of induced braid types for any f_w to consideration of induced braid types for f_L and f_R. Recall that $w = e_k \cdots e_1$, where each e_i is L or R, and let $w_i = e_i \cdots e_1$.

LEMMA 2.1. *If $\{x, y, z\} \subset S_A \cup S_B$ is f_w-invariant, then $bt(f_w, \{x, y, z\})$ is represented by $b(w; \{x, y, z\})$. Moreover, if we denote $f_{w_i}(\{x, y, z\})$ by $\{x_i, y_i, z_i\}$, then*

$$b(w; \{x, y, z\}) = b(e_k; \{x_{k-1}, y_{k-1}, z_{k-1}\}) \cdot b(e_{k-1}; \{x_{k-2}, y_{k-2}, z_{k-2}\}) \cdots b(e_1; \{x, y, z\}).$$

Proof of Lemma 2.1. Since $\{x, y, z\}$ is f_w-invariant, $h_{\{x,y,z\}} = h_{f_w(\{x,y,z\})}$ and $b(w; \{x, y, z\}) = h_{\{x,y,z\}} f_w h^{-1}_{\{x,y,z\}}$ represents $bt(f_w, \{x, y, z\})$. For the second part, note that

$$\begin{aligned} & h_{\{x_k,y_k,z_k\}} f_w h^{-1}_{\{x,y,z\}} \\ &= (h_{\{x_k,y_k,z_k\}} f_{e_k} h^{-1}_{\{x_{k-1},y_{k-1},z_{k-1}\}}) \cdot (h_{\{x_{k-1},y_{k-1},z_{k-1}\}} f_{e_{k-2}} h^{-1}_{\{x_{k-2},y_{k-2},z_{k-2}\}}) \\ &\quad \cdots (h_{\{x_1,y_1,z_1\}} f_{e_1} h^{-1}_{\{x,y,z\}}) \\ &= b(e_k; \{x_{k-1}, y_{k-1}, z_{k-1}\}) \cdot b(e_{k-1}; \{x_{k-2}, y_{k-2}, z_{k-2}\}) \cdots b(e_1; \{x, y, z\}). \end{aligned}$$

□

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

We now turn our attention to computing $b(e_i, \{x_j, y_j, z_j\})$.

Each fiber in S_A contains the f_L-image of a fiber from S_A and a fiber from S_B. If $x \in S_A$ satisfies $f_L(x) \in S_A$, then define $j_L(\hat{x})$ to be the unique point in B satisfying $g_L(j_L(\hat{x})) = g_L(\hat{x})$. It is convenient to draw the edge A shorter than the edge B so that $\hat{x}$ and $j_L(\hat{x})$ lie on the same horizontal line. (See Figure 6.)

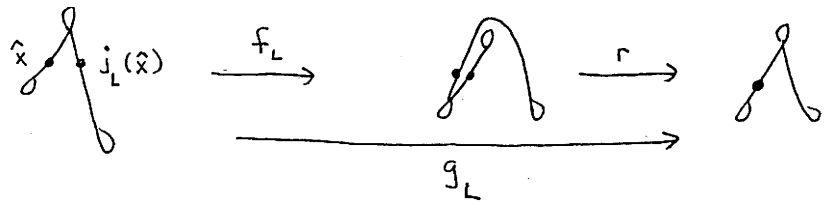


FIGURE 6.

LEMMA 2.2. *Suppose that* $\{x, y, z\}$, $f_L\{x, y, z\} \subset \operatorname{int} S_A \cup S_B$ *and that* $\hat{x} < \hat{y} < \hat{z}$.

(1) *If* $\hat{x}, \hat{y} \in A$ *and* $\hat{z} < j_L(\hat{y})$, *then* $g_L(\hat{z}) < g_L(\hat{y}) < g_L(\hat{x})$ *in* A *and* $b(L; \{x, y, z\}) = T^{-1}$.

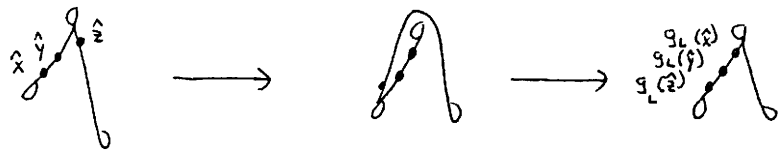


FIGURE 7.

(2) *If* $\hat{x} \in A, \hat{z} \in B$ *with* $\hat{z} < j_L(\hat{x})$ *and either* $\hat{y} \in A$ *with* $j_L(\hat{y}) < \hat{z}$ *or* $\hat{y} \in B$, *then* $g_L(\hat{y}) < g_L(\hat{z}) < g_L(\hat{x})$ *in* A *and* $b(L; \{x, y, z\}) = S^{-1}$.

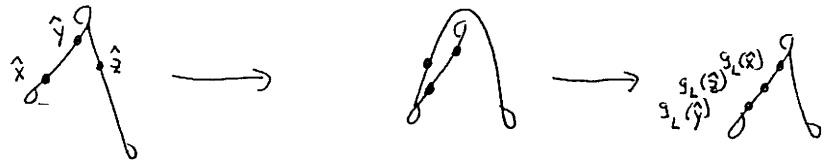


FIGURE 8.

(3) *If* $\hat{x} \in A$, $\hat{z} \in B$ *with* $\hat{z} > j_L(\hat{x})$ *and* $\hat{y} < j_L(\hat{x})$, *then* $g_L(\hat{y}) < g_L(\hat{x}) < g_L(\hat{z})$ *and* $b(L; \{x, y, z\}) = b_L$.

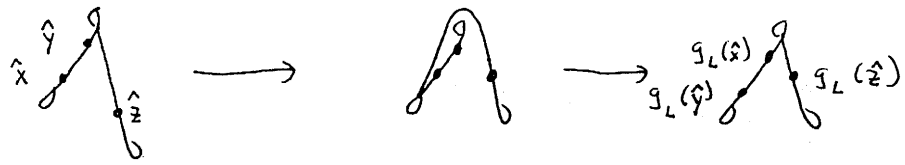


FIGURE 9.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

(4) *If $\hat{y}, \hat{z} \in B$ and either $\hat{x} \in B$ or $\hat{x} \in A$ with $\hat{y} > j_L(\hat{x})$, then $g_L(\hat{x}) < g_L(\hat{y}) < g_L(\hat{z})$ and $b(L; \{x, y, z\}) = b_\emptyset$ is the identity element.*

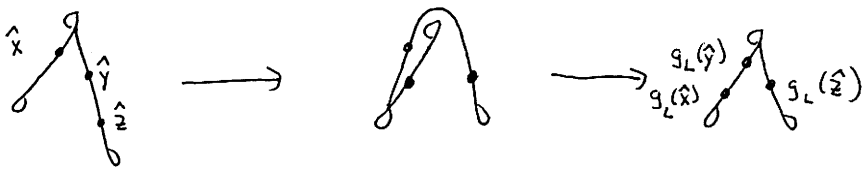


FIGURE 10.

Proof of Lemma 2.2. Suppose that $\hat{x}, \hat{y} \in A$ and $\hat{z} < j_L(\hat{y})$. Let $\hat{z}' = \hat{z}$ if $\hat{z} \in A$ and $\hat{z}' = j_L^{-1}(\hat{z})$ if $\hat{z} \in B$. Then $\hat{x} < \hat{y} < \hat{z}'$ in A so $g_L(\hat{z}) < g_L(\hat{y}) < g_L(\hat{x})$ in A.

Let σ_a, σ_b and σ_c be arcs connecting ∂D_3 to a, b and c respectively as shown in Figure 11. Let τ_x, τ_y and τ_z be arcs connecting ∂D_3 to x, y and z respectively as shown. Then $h^{-1}_{\{x,y,z\}}(\sigma_a)$ (respectively $h^{-1}_{\{x,y,z\}}(\sigma_b), h^{-1}_{\{xyz\}}(\sigma_c)$) is isotopic rel $(\{x, y, z\} \cup \partial D_3)$ to τ_x (respectively τ_y, τ_z). The arcs $h_{f_L(\{x,y,z\})} f_L(\tau_x), h_{f_L(\{x,y,z\})} f_L(\tau_y)$ and $h_{f_L(\{x,y,z\})} f_L(\tau_z)$ are, respectively, isotopic rel $(\{a, b, c\} \cup \partial D_3)$ to the arcs μ_a, μ_b and μ_c as shown. Thus the actions of $h_{f_L(\{x,y,z\})} f_L h^{-1}_{\{x,y,z\}}$ and of T^{-1} on σ_a, σ_b and σ_c agree up to isotopy. This implies that $h_{f_L(\{x,y,z\})} f_L h^{-1}_{\{x,y,z\}}$ and T^{-1} determine the same element of $\mathrm{MCG}(D_3; \partial D_3)$ and completes the proof of part (1).

The remaining three cases are very similar to case (1) and we leave the details to the reader. □

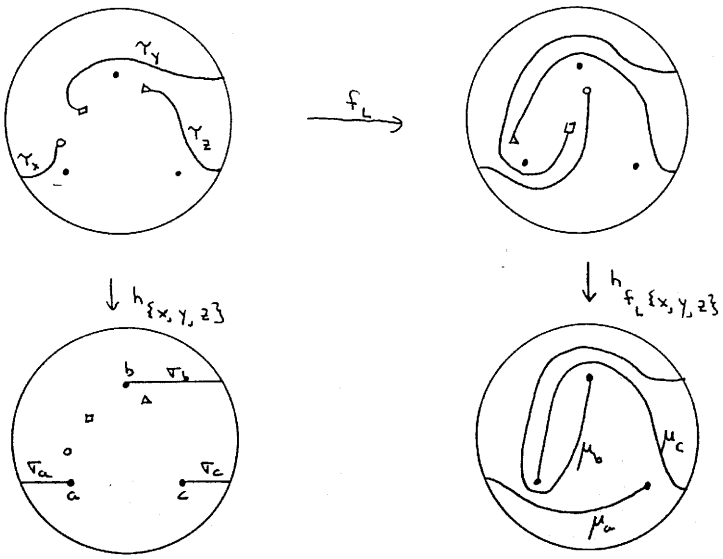


FIGURE 11.

We now give the analogous result for f_R. If $x \in S_B$ satisfies $f_L(x) \in S_B$, then define $j_R(\hat{x})$ to be the unique point in A satisfying $g_R(j_R(\hat{x})) = g_R(\hat{x})$.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

LEMMA 2.3. *Suppose that* $\{x, y, z\}$, $f_R\{x, y, z\} \subset \operatorname{int} S_A \cup S_B$ *and that* $\hat{x} < \hat{y} < \hat{z}$.

(1) *If* $\hat{x}, \hat{y} \in A$ *and either* $\hat{z} \in A$ *or* $\hat{z} \in B$ *with* $j_R(\hat{z}) > \hat{y}$, *then* $g_R(\hat{x}) < g_R(\hat{y}) < g_R(\hat{z})$ *and* $b(R; \{x, y, z\}) = b_\emptyset$ *is the identity element.*

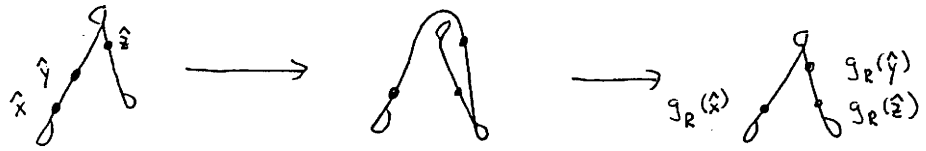


FIGURE 12.

(2) *If* $\hat{x} \in A, \hat{z} \in B$ *with* $\hat{x} < j_R(\hat{z})$ *and either* $\hat{y} \in B$ *or* $\hat{y} \in A$ *with* $j_R(\hat{z}) < \hat{y}$, *then* $g_R(\hat{x}) < g_R(\hat{z}) < g_R(\hat{y})$ *and* $b(R; \{x, y, z\}) = b_R$.

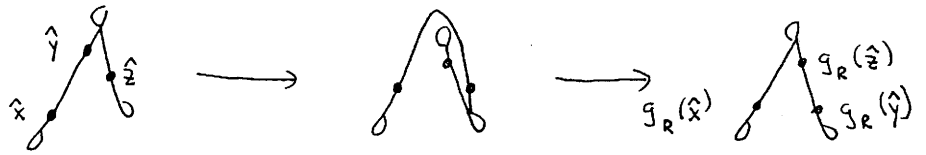


FIGURE 13.

(3) *If* $\hat{x} \in A$, $\hat{z} \in B$ *with* $j_R(\hat{z}) < \hat{x}$ *and either* $\hat{y} \in A$ *or* $\hat{y} \in B$ *with* $j_R(\hat{y}) > \hat{x}$, *then* $g_R(\hat{z}) < g_R(\hat{x}) < g_R(\hat{y})$ *in* B *and* $b(R; \{x, y, z\}) = S$.



FIGURE 14.

(4) *If* $\hat{y}, \hat{z} \in B$ *and* $\hat{x} > j_R(\hat{y})$, *then* $g_R(\hat{z}) < g_R(\hat{y}) < g_R(\hat{x})$ *in* B *and* $b(R; \{x, y, z\}) = T$.

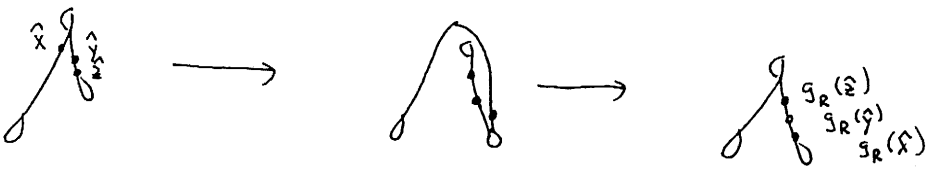


FIGURE 15.

Proof of Lemma 2.3. The proof is completely analogous to that of Lemma 2.2 and is left to the reader. □

Remark 2.4. It follows from Lemmas 2.2 and 2.3 that $b(e_i; \{x, y, z\})$ is determined by e_i, the relative order of $\{\hat{x}, \hat{y}, \hat{z}\}$ and the relative order of $\{g_{e_i}(\hat{x}), g_{e_i}(\hat{y}), g_{e_i}(\hat{z})\}$.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

3. *Only if part of Theorem 0.3*

In this section we prove half of our main theorem.

PROPOSITION 3.1. *If w and v are pseudo-Anosov words such that $[b_w] > [b_v\Delta^s]$, then (v, s) is obtained from $(w, 0)$ by successively applying $(\mathcal{R}1)$–$(\mathcal{R}4)$.*

As a first step toward proving Proposition 3.1, we look more closely at Lemmas 2.1–2.3.

LEMMA 3.2. *If $w = e_k \cdots e_1$ is a pseudo-Anosov word and $\{x, y, z\} \subset S_A \cup S_B$ is an f_w-invariant set, then $b(w; \{x, y, z\}) = d_k \cdots d_1$ where:*

(1) *if $e_i = L$ then $d_i = S^{-1}, T^{-1}, L$ or ϕ;*
(2) *if $e_i = R$ then $d_i = S, T, R$ or ϕ;*
(3) *if $e_i = L$, if $d_i = S^{-1}$ or T^{-1} and if l is the smallest parameter larger than i for which $e_l = R$, then $d_j = T^{-1}$ for $i < j < l$ and $d_l = \emptyset$;*
(4) *if $e_i = R$, if $d_i = S$ or T and if l is the smallest parameter larger than i for which $e_l = L$, then $d_j = T$ for $i < j < l$ and $d_l = \emptyset$.*

Proof of Lemma 3.2. Lemma 2.2 implies that if $b(L; \{x, y, z\}) = S^{-1}$ or T^{-1} then $g_L\{\hat{x}, \hat{y}, \hat{z}\} \subset A$. Lemmas 2.2 and 2.3 then imply that $b(L; f_L(\{x, y, z\})) = T^{-1}$ and that $b(R; f_L(\{x, y, z\})) = b_\emptyset$ is the identity element.

A similar argument shows that if $b(R; \{x, y, z\}) = S$ or T, then $b(R; f_R(\{x, y, z\})) = T$ and $b(L; f_R(\{x, y, z\})) = b_\emptyset$ is the identity element.

Lemma 3.2 is now an immediate corollary of Lemmas 2.1–2.3. □

Proof of Proposition 3.1. By Corollary 1.4, Lemma 2.1 and Lemma 3.2, we may assume that $b_v\Delta^s = d_k \cdots d_1$, where the d_i satisfy (1)–(4) of Lemma 3.2.

We first show that, after a cyclic permutation if necessary, w can be divided into blocks $w = \beta_1 \cdots \beta_r$, where the replacement words c_i corresponding to β_i are of the form $L\Delta^t$, $R\Delta^t$ or Δ^t. (If $\beta_i = e_j \cdots e_l$, then $c_i = d_j \cdots d_l$.) Moreover, if β_i contains more than one letter, then β_i contains at least one L and one R.

Begin with the following cyclic block decomposition of w. If $e_i = L$ and $d_i = S^{-1}$ or T^{-1}, if e_{i-1} and d_{i-1} do not satisfy these properties, and if l is the smallest parameter larger than i for which $e_l = R$, then $e_i \cdots e_l$ is a single block. Similarly, if $e_i = R$ and $d_i = S$ or T, if e_{i-1} and d_{i-1} do not satisfy these properties, and if l is the smallest parameter larger than i for which $e_l = L$, then $e_i \cdots e_l$ is a single block. The blocks so constructed are disjoint. Treat each e_i not in one of these blocks as a block. This gives a cyclic block decomposition $w = \beta_1 \ldots \beta_r$, where the replacement words c_i are of the form $S^{-1}T^{-k}$, T^{-k}, ST^k, T^k, L, R or $\emptyset$. In particular, the replacement words c_i are of the form

$$S^jT^{-k}\Delta^t \quad \text{for } 0 \le j \le 2, 0 \le k \le 1, \text{ and some } t. \qquad (*)$$

If $c_i = S^j\Delta^t$, then c_ic_{i+1} is of the form $(*)$. If $c_i = T^k\Delta^t$, then $c_{i-1}c_i$ is of the form $(*)$. We may therefore amalgamate blocks until the replacement words are of the desired form $ST^{-1}\Delta^t$, $S^2T^{-1}\Delta^t$ or Δ^t.

We will now show that each c_i can be obtained from β_i by applying the moves $(\mathcal{R}2)$–$(\mathcal{R}4)$. Define a homomorphism $\Psi : \mathrm{MCG}(D_3; \partial D_3) \to Z/6$ by $\Psi(S) = \frac{1}{3}$ and

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

$\Psi(T) = \frac{1}{2}$. Since $\Psi(S^3) = \Psi(T^2)$, Ψ is well defined. Note that $\Psi(S^{-1}), \Psi(T^{-1})$ and $\Psi(L)$ are negative, while $\Psi(S), \Psi(T)$ and $\Psi(R)$ are positive. Using the rules (1)–(4) of Lemma 3.2, L is replaced by an element with Ψ value between $-\frac{1}{2}$ and 0 and R is replaced by an element with Ψ value between 0 and $\frac{1}{2}$.

The proof concludes with a case by case check. Suppose at first that $c_i = \Delta^s$. If $s = 0$, then we can obtain c_i from β_i by erasing all the letters. If $s > 0$ (respectively $s < 0$), then, since $\Psi(c_i) = s$, β_i must contain at least $2s$ R's (respectively L's). In this case, erase all but $2s$ R's (respectively L's) and then replace them pairwise by Δ (respectively Δ^{-1}).

Suppose next that $c_i = L\Delta^s$. If $s = 0$, then $\Psi(c_i) < 0$ and so β_i must contain at least one L. We obtain c_i from β_i by erasing all but one L. If $s < 0$, then $\Psi(c_i) < s$ implies that β_i contains at least $(2s + 1)$ L's; we can erase all other letters and then replace s LL's with Δ^{-1}. If $s > 0$, then $\Psi(c_i) = s - \frac{1}{6}$ and so β_i must contain at least $2s$ R's. Since β_i cannot be a single letter, b_i also contains at least one L. Erase all other letters and replace s RR's with Δ.

Finally, suppose that $c_i = R\Delta^s$. If $s = 0$, then $\Psi(c_i) > 0$ and so β_i must contain at least one R; erase all other letters. If $s > 0$, then $\Psi(c_i) > s$ implies that β_i contains at least $(2s + 1)$ R's; erase all other letters and then replace s RR's with Δ. If $s < 0$, then $\Psi(c_i) = s + \frac{1}{6}$ and so β_i must contain at least $2s$ L's. Since β_i cannot be a single letter, β_i also contains at least one R. Erase all other letters and replace s LL's with Δ^{-1}. □

4. *Symbolic dynamics*

In this section we recall the connection between f_w-invariant sets in $S_A \cup S_B$ and closed paths in the transition graph Γ associated to $g_w : G \to G$. Our main result is Corollary 4.2 which we use in §5 to produce f_w-invariant sets $\{x, y, z\}$ with prescribed $b(w, \{x, y, z\})$.

We assume that $w = e_k \cdots e_1$ is a pseudo-Anosov word and let $w_i = e_i \cdots e_1$.

Beginning with $G_k = G$, inductively define subdivisions G_{i-1} of G_i by declaring the vertex set for G_{i-1} to be the full g_{e_i} pre-image of the vertex set for G_i. An edge of G_{i-1} is inductively declared to be peripheral (respectively non-peripheral) if its g_{e_i}-image in G_i is peripheral (respectively non-peripheral). Non-peripheral edges in G_{i-1} are inductively oriented and labelled A or B according to their g_{e_i}-image in G_i. (To simplify pictures, the label and orientation are combined, so that an edge labelled A with reverse orientation is shown as $\bar{A}$.) Peripheral edges are unlabelled and alternate with non-peripheral edges. (See Figure 16.)

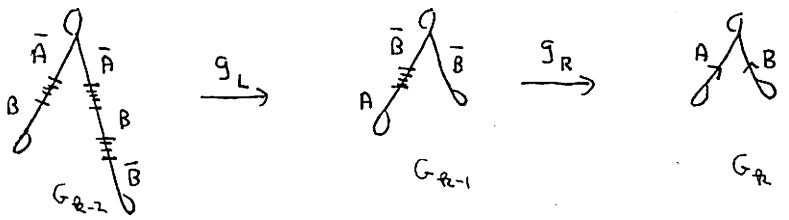


FIGURE 16.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

Define the transition graph Γ for $g_w : G \to G$ to be the graph with two vertices, labelled A and B, and an edge E for each non-peripheral edge $\epsilon \subset G_0$. If ϵ, when considered as a subinterval of G, is contained in A (respectively B), then the initial vertex of E is A (respectively B). If ϵ is labelled A (respectively B), then the terminal vertex of E is A (respectively B). Figure 17 shows Γ when $w = RL$.

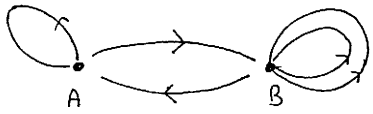


FIGURE 17.

LEMMA 4.1. *Suppose that ϵ_x, ϵ_y and ϵ_z are distinct edges of G_0 and that π is a permutation of $\{x, y, z\}$ such that the label on ϵ_α agrees with the edge of G that contains $\epsilon_{\pi(\alpha)}$ for $\alpha = x, y, z$. (In other words, the terminal vertex of E_α equals the initial vertex of $E_{\pi(\alpha)}$). Then:*

(1) *there is an f_w-invariant set $\{x, y, z\} \in S_A \cup S_B$ such that $\hat{\alpha} \in \epsilon_\alpha$ and such that $f_w(\alpha) = \pi(\alpha)$;*

(2) *if x_0, y_0, z_0 is another triple of points such that $\hat{\alpha}_0 \in \epsilon_\alpha$ and such that the relative ordering on $g_w(\hat{x}_0), g_w(\hat{y}_0)$ and $g_w(\hat{z}_0)$ agrees with the relative ordering on $g_w(\hat{x}), g_w(\hat{y})$ and $g_w(\hat{z})$, then the relative ordering on $g_{w_i}(\hat{x}_0), g_{w_i}(\hat{y}_0)$ and $g_{w_i}(\hat{z}_0)$ agrees with the relative ordering on $g_{w_i}(\hat{x}), g_{w_i}(\hat{y})$ and $g_{w_i}(\hat{z})$ for all $0 \le i \le k$.*

Proof of Lemma 4.1. Part (1) is the usual fact that closed loops in the transition graph for a Markov partition correspond to periodic orbits of the map.

For part (2), note that if μ_x and μ_y are distinct edges in some G_i, then the relative ordering on $(\hat{p}_x, \hat{p}_y)$ is independent of the choice of $\hat{p}_x \in \mu_x$ and $\hat{p}_y \in \mu_y$. Note also that if $(\hat{p}_x, \hat{p}_y)$ and $(\hat{q}_x, \hat{q}_y)$ are pairs of points in an edge μ of some G_i and that if the relative ordering on $(\hat{p}_x, \hat{p}_y)$ differs from the relative ordering on $(\hat{q}_x, \hat{q}_y)$, then the relative ordering on $(g_{e_{i+1}}(\hat{p}_x), g_{e_{i+1}}(\hat{p}_y))$ differs from the relative ordering on $(g_{e_{i+1}}(\hat{q}_x), g_{e_{i+1}}(\hat{q}_y))$.

Let $l \ge 0$ be the largest parameter value for which $g_{w_l}(\epsilon_x) \ne g_{w_l}(\epsilon_y)$. By our first observation, the relative orderings on $(g_{w_i}(\hat{x}), g_{w_i}(\hat{y}))$ and on $(g_{w_i}(\hat{x}_0), g_{w_i}(\hat{y}_0))$ agree for $0 \le i \le l$. Suppose that $l < k$. Since the relative orderings on $(g_{w_k}(\hat{x}), g_{w_k}(\hat{y}))$ and on $(g_{w_k}(\hat{x}_0), g_{w_k}(\hat{y}_0))$ agree, our second observation, applied $k-l-1$ times, implies that the relative orderings on $(g_{w_i}(\hat{x}), g_{w_i}(\hat{y}))$ and on $(g_{w_i}(\hat{x}_0), g_{w_i}(\hat{y}_0))$ agree for $l+1 \le i \le k$.

Applying this argument to the other pairs ϵ_x, ϵ_z and ϵ_y, ϵ_z completes the proof. □

COROLLARY 4.2. *Suppose that ϵ_x, ϵ_y and ϵ_z are distinct edges of G_0, that $\hat{x}_0 \in \epsilon_x$, $\hat{y}_0 \in \epsilon_y$ and $\hat{z}_0 \in \epsilon_z$, and that $\{\hat{x}_0, \hat{y}_0, \hat{z}_0\} \cap A$ has the same cardinality as $\{g_w(\hat{x}_0), g_w(\hat{y}_0), g_w(\hat{z}_0)\} \cap A$. Then there is an f_w-invariant set $\{x, y, z\} \in S_A \cup S_B$ such that $b(w; \{x, y, z\}) = b(e_1, \{x, y, z\}) \cdot b(e_2, f_{w_1}\{x_0, y_0, z_0\}) \cdot b(e_3, f_{w_2}\{x_0, y_0, z_0\}) \cdots b(e_k, f_{w_{k-1}}\{x_0, y_0, z_0\})$.*

Proof of Corollary 4.2. We assume without loss that $\epsilon_x < \epsilon_y < \epsilon_z$. Let π be the permutation of $\{\hat{x}, \hat{y}, \hat{z}\}$ determined by the relative ordering of $\{g_w(\hat{x}), g_w(\hat{y}), g_w(\hat{z})\}$. In

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

other words, if $g_w(\hat{x})$ is first (respectively second, third) in the relative ordering, then $\pi(x) = x$ (respectively y, z); $\pi(y)$ and $\pi(z)$ are defined by the same rule. For example, if $g_w(\hat{z}) < g_w(\hat{x}) < g_w(\hat{y})$, then $\pi(x) = y, \pi(y) = z$ and $\pi(z) = x$.

To check the hypothesis of Lemma 4.1, identify $\hat{x}$, $\hat{y}$ and $\hat{z}$ with 1, 2 and 3 respectively. Then $e_{\pi(\alpha)} \subset A$ if and only if $\pi(\alpha) \leq \text{Card}(\{\hat{x}_0, \hat{y}_0, \hat{z}_0\} \cap A)$ and the label on e_α is A if and only if $\pi(\alpha) \leq \text{Card}(\{g_w(\hat{x}_0), g_w(\hat{y}_0), g_w(\hat{z}_0)\} \cap A)$. Since these cardinalities are equal, the hypothesis of Lemma 4.1 are satisfied.

Let $\{x, y, z\}$ be the f_w-invariant set produced by Lemma 4.1. The fact that $b(w; \{x, y, z\})$ is as described follows from Lemma 2.1, Remark 2.4 and part (2) of Lemma 4.1. □

5. *If part of Theorem 0.3*

In this section we prove the second half of our main theorem.

PROPOSITION 5.1. *If v and w are pseudo-Anosov words and if (v, s) is obtained from $(w, 0)$ by successively applying $(\mathcal{R}1)$–$(\mathcal{R}4)$, then $[b_w] \geq [b_v \Delta^s]$.*

The main tool in proving Proposition 5.1 is Corollary 4.2. We will also need the following pair of lemmas.

LEMMA 5.2. *Suppose that $\hat{x}_2 < \hat{y}_2 < \hat{z}_2$ with $\hat{x}_2, \hat{y}_2 \in A$ and $\hat{z}_2 \in B$. Let $\hat{y}_1$ be the g_L pre-image of $\hat{y}_2$ in A, let $\hat{z}_1 \in B$ be the unique g_L pre-image of $\hat{z}_2$ and let $\hat{x}_1$ be either of the two g_L pre-images of $\hat{x}_2$. Then $b(L; \{x_1, y_1, z_1\}) = b_L$.*

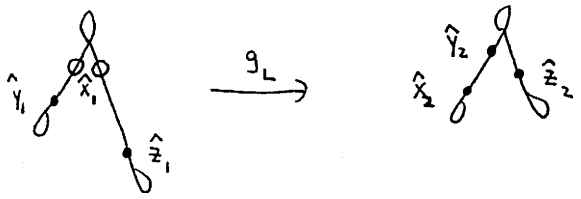


FIGURE 18.

Proof of Lemma 5.2. By hypothesis, $\hat{y}_1 \in A$, $\hat{z}_1 \in B$ and $\hat{y}_1 < \hat{x}_1 < j_L(\hat{y}_1) < \hat{z}_1$. Lemma 5.2 therefore follows from Lemma 2.2(3). □

There is an analogous result for f_R.

LEMMA 5.3. *Suppose that $\hat{x}_2 < \hat{y}_2 < \hat{z}_2$ with $\hat{x}_2 \in A$ and $\hat{y}_2, \hat{z}_2 \in B$. Let $\hat{x}_1 \in A$ be the unique g_R pre-image of $\hat{x}_2$, let $\hat{y}_1$ be the g_R pre-image of $\hat{y}_2$ in B and let $\hat{z}_1$ be either of the two g_R pre-images of $\hat{z}_2$. Then $b(R; \{x_1, y_1, z_1\}) = b_R$. (See Figure 19.)*

Proof of Lemma 5.3. By hypothesis, $\hat{x}_1 \in A$, $\hat{y}_1 \in B$ and $\hat{x}_1 < \hat{j}_R(y_1) < \hat{z}_1 < \hat{y}_1$. Lemma 5.3 therefore follows from Lemma 2.3(2). □

Proof of Proposition 5.1. By induction, it suffices, to assume that (v, s) is obtained from (w, m) by applying one move $\mathcal{M}$ from $(\mathcal{R}1)$–$(\mathcal{R}4)$.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

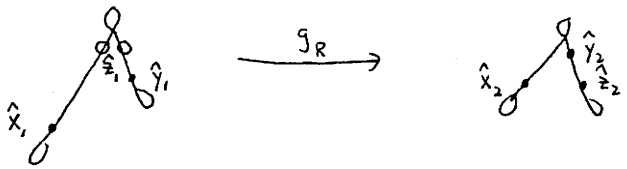


FIGURE 19.

If $\mathcal{M} = (\mathcal{R}1)$, then $s = 0$ and $[b_v] = [b_w]$.

Suppose that $\mathcal{M} = (\mathcal{R}2)$ and that L is the letter that is removed. If the L that we wish to remove is contained in a consecutive string of L's, there is no loss in removing the first L in that string. Thus, after cyclically permuting if necessary, we may assume that $e_k = L, e_{k-1} = R$ and that $v = e_{k-1} \cdots e_1$.

Choose points $x_{k-1}, y_{k-1}, z_{k-1}$ so that $\hat{x}_{k-1}, \hat{y}_{k-1}, \hat{z}_{k-1} \in G_{k-1}$ are as shown in Figure 20. The key features are that $\hat{x}_{k-1} \in A$; $\hat{y}_{k-1} \in B$ with $j_L(\hat{x}_{k-1}) < \hat{y}_{k-1}$ and with $g_L(\hat{y}_{k-1}) \in A$; and $\hat{z}_{k-1} \in B$ with $g_L(\hat{z}_{k-1}) \in B$. Lemma 2.2(4) implies that $b(L; \{x_{k-1}, y_{k-1}, z_{k-1}\}) = b_\emptyset$. Note that $\{g_L(\hat{x}_{k-1}), g_L(\hat{y}_{k-1}), g_L(\hat{z}_{k-1})\} \cap A$ contains two points.

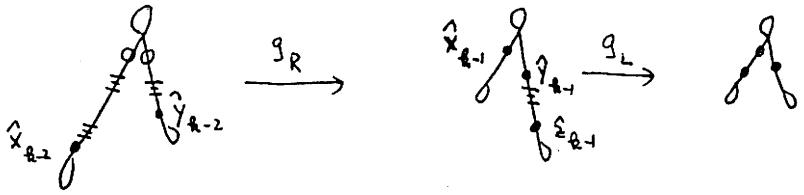


FIGURE 20.

We now inductively apply Lemmas 5.2 and 5.3 to choose f_{e_i} pre-images x_{i-1}, y_{i-1} and z_{i-1} of x_i, y_i and z_i respectively such that $b(e_i; \{x_{i-1}, y_{i-1}, z_{i-1}\}) = b_{e_i}$ for $1 \leq i \leq k-1$. Since v is a pseudo-Anosov word, $k > 2$. If $e_{k-2} = L$, choose z_{k-2} so that $\hat{z}_{k-2} \in A$; if $e_{k-2} = R$, choose z_{k-2} so that $\hat{z}_{k-2} \in B$. If $k = 3$, or if $k > 3$ and $e_{k-3} = L$, choose pre-images so that $\{\hat{x}_{k-3}, \hat{y}_{k-3}, \hat{z}_{k-3}\} \cap A$ contains two points; if $k > 3$ and $e_{k-3} = R$, choose pre-images so that $\{\hat{x}_{k-3}, \hat{y}_{k-3}, \hat{z}_{k-3}\} \cap B$ contains two points. Continue this process to inductively define $\{x_i, y_i, z_i\} \subset G_i$ as above with the additional feature that $\{\hat{x}_0, \hat{y}_0, \hat{z}_0\} \cap A$ contains two points. Corollary 4.2 implies the existence of a f_w-invariant set $\{x, y, z\} \subset S_A \cup S_B$ such that $b(w, \{x, y, z\}) = b_{e_{k-1}} \cdot b_{e_{k-2}} \cdots b_{e_1} = b_v$. Corollary 1.4 and Lemma 2.1 imply that $[b_w] > [b_v]$.

The case that $\mathcal{M} = (\mathcal{R}2)$ and that R is the letter being removed is completely symmetric and the details are left to the reader.

Suppose now that $\mathcal{M} = (\mathcal{R}3)$. After cyclically permuting if necessary, we may assume that $e_k = R, e_{k-1} = L, e_{k-2} = L$ and that $v = R \cdot e_{k-3} \cdots e_1$.

Choose points x_k, y_k, z_k so that $\hat{x}_k < \hat{y}_k < \hat{z}_k \in G_k$ with $\hat{x}_k, \hat{y}_k \in A$ and $\hat{z}_k \in B$. As in the previous case, x_{i-1}, y_{i-1} and z_{i-1} are f_{e_i} pre-images of

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

x_i, y_i and z_i respectively. Choose x_{k-1}, y_{k-1} and z_{k-1} so that $\hat{x}_{k-1}, \hat{y}_{k-1}, \hat{z}_{k-1} \in A$; by Lemma 2.3(1), $\hat{x}_{k-1} < \hat{y}_{k-1} < \hat{z}_{k-1}$ and $b(R; \{\hat{x}_{k-1}, \hat{y}_{k-1}, \hat{z}_{k-1}\}) = b_\emptyset$. Choose x_{k-2}, y_{k-2} and z_{k-2} so that $\hat{x}_{k-2}, \hat{y}_{k-2}, \hat{z}_{k-2} \in A$; by Lemma 2.2(1), $\hat{z}_{k-2} < \hat{y}_{k-2} < \hat{x}_{k-2}$ and $b(L, \{\hat{x}_{k-2}, \hat{y}_{k-2}, \hat{z}_{k-2}\}) = T^{-1}$. Choose x_{k-3}, y_{k-3} and z_{k-3} so that $\hat{x}_{k-3} \in A$ and $\hat{y}_{k-3} \in B$; then $\hat{x}_{k-3} < \hat{z}_{k-3} < \hat{y}_{k-3} < j_L(\hat{x}_{k-3})$ and, by Lemma 2.2(2), $b(L, \{\hat{x}_{k-3}, \hat{y}_{k-3}, \hat{z}_{k-3}\}) = S^{-1}$. Combining these three stages, we have $b(RLL, \{x_{k-3}, y_{k-3}, z_{k-3}\}) = S^{-1}T^{-1} = S^2T^{-1}\Delta^{-1} = R\Delta^{-1}$. (See Figure 21.)

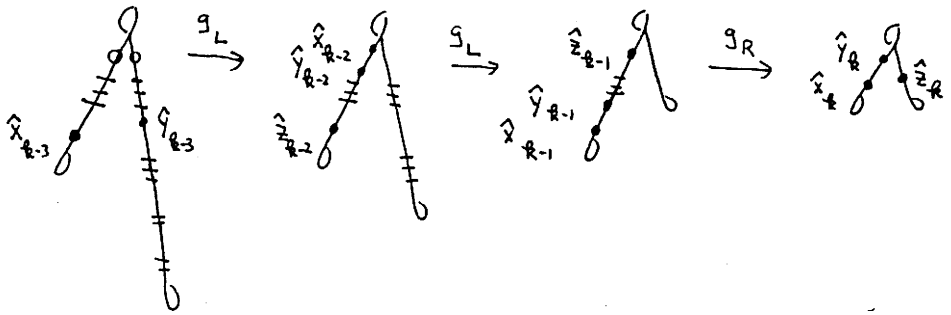


FIGURE 21.

We have not yet specified the exact location of z_{k-3}. Since v is a pseudo-Anosov word, $k > 3$. If $e_{k-3} = L$, choose $\hat{z}_{k-3} \in A$; if $e_{k-3} = R$, choose $\hat{z}_{k-3} \in B$. The proof now concludes exactly as in the previous case. Use Lemmas 5.2 and 5.3 to define x_i, y_i and z_i so that $b(e_i; \{x_i, y_i, z_i\}) = b_{e_i}$ for $1 \leq i \leq k-3$ and so that $\{\hat{x}_0, \hat{y}_0, \hat{z}_0\} \cap A$ contains two points. Then apply Corollary 4.2 to produce an f_w-invariant set with $b(w, \{x, y, z\}) = Re_{k-3}e_{k-4}\cdots e_1\Delta^{-1} = b_v\Delta^{-1}$. Corollary 1.4 and Lemma 2.1 imply that $[b_w] > [b_v\Delta^{-1}]$.

The case that $\mathcal{M} = (\mathcal{R}4)$ is completely symmetric to the case that $\mathcal{M} = (\mathcal{R}3)$; the details are left to the reader. □

REFERENCES

[AF] D. Asimov and J. Franks. Unremovable closed orbits. *Geometric Dynamics (Lecture Notes in Math. 1007)*. Ed. J. Palis. Springer, 1983, pp. 22–29

[BH] M. Bestvina and M. Handel. Train tracks for surface homeomorphisms. *Topology* **32** (1995), 109–140.

[Bir] J. Birman. *Braids, Links and Mapping Class Groups*. Princeton University Press, 1975.

[Boy] P. Boyland. Topological methods in surface dynamics. *Topology and its Applications* **58** (1994), 223–298.

[FLP] A. Fathi, F. Laudenbach and V. Poenaru. Travaux de Thurston sur les surfaces. *Asterisque* **66** (1979).

[Hal] T. Hall. Unremovable periodic orbits of homeomorphisms. *Mat. Proc. Camb. Phil. Soc.* **19** (1991), 523–531.

[LS] R. Lyndon and P. Schupp. *Combinatorial Group Theory*. Springer, 1977.

[Mat] T. Matsuoka. Braids of periodic points and a 2-dimensional analogue of Sharkovskii's ordering. *Dynamical Systems and Nonlinear Oscillations*. Ed. G. Ikegami. World Scientific Press, 1986, pp. 58–72.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press

[Sch] A. N. Scharkovski. Coexistence of cycles of a continuous map of a line into itself. *Ukrainian Math. J.* **16** (1964), 61–71.
[Sta] J. Stallings. Topology of finite graphs. *Inv. Math.* **71** (1983), 551–565.

https://doi.org/10.1017/S0143385797084940 Published online by Cambridge University Press