PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 102, Number 3, March 1988

ENTROPY AND EXPONENTIAL GROWTH OF π_1 IN DIMENSION TWO

JOHN M. FRANKS AND MICHAEL HANDEL

(Communicated by George R. Sell)

ABSTRACT. The authors show that if $f: M \to M$ is a $C^{1+\alpha}$ diffeomorphism of a compact surface and if the topological entropy of f is positive then there is a finite invariant set P such that the map induced by f on $\pi_1(M-P)$ has exponential growth.

Following work of Manning [**M**], several authors (Bowen, Gromov, Katok, Shub; see for example [**B**]) observed an important relationship between the topological entropy of a continuous map $f: M \to M$ and the action of f on the fundamental group $\pi_1(M)$. Specifically they showed that if for some $\alpha \in \pi_1(M)$ and some choice of generators of $\pi_1(M)$ it requires a word of length at least $C\lambda^n$ ($C > 0$, $\lambda > 1$) in these generators to represent the element $f_{\#}^n(\alpha)$ then the topological entropy of f is $\geq \log \lambda$. An $f_{\#}$ with this property is said to have *exponential growth* on $\pi_1(M)$. This has proved to be a powerful tool especially in the study of maps of surfaces.

The purpose of this article is to show that for smooth diffeomorphisms of surfaces the presence of positive topological entropy can always be detected by this result if an appropriate finite set of points is removed from the surface to increase the action on π_1.

THEOREM. *Let $f: M^2 \to M^2$ be a $C^{1+\alpha}$ diffeomorphism of a compact surface and suppose the topological entropy $h(f) > 0$. There exists a finite invariant set $P \subset M$ such that the map induced by f*

$$f_{\#}: \pi_1(M-P) \to \pi_1(M-P)$$

has exponential growth.

REMARK. Denote the exponential growth rate of $f_{\#}: \pi_1(M-P) \to \pi_1(M-P)$ by $\mathrm{EGR}(f, P)$. It is not always possible (Theorem 1 of [**F2**]) to choose P so that $\mathrm{EGR}(f, P) = h(f)$. It is unknown, even for the standard horseshoe in S^2, if P can be chosen so that $\mathrm{EGR}(f, P)$ is arbitrarily close to $h(f)$.

1. In this section, we review the necessary hyperbolic geometry (see [**F-L-P**] and [**H-T**] for more details).

Given a finite f-invariant set $P \subset M$, there is an induced homeomorphism(also called f) on the surface $N = M - P$. If P has at least three points (as it will for the cases that we consider) then N supports a finite area, complete metric of constant curvature -1. We assume that N has been equipped with such a metric.

Received by the editors April 26, 1985 and, in revised form, December 1, 1986.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 58F15, 58F20, 28D20.
Research supported in part by NSF grant MCS 8120790.

©1988 American Mathematical Society
0002-9939/88 $1.00 + $.25 per page

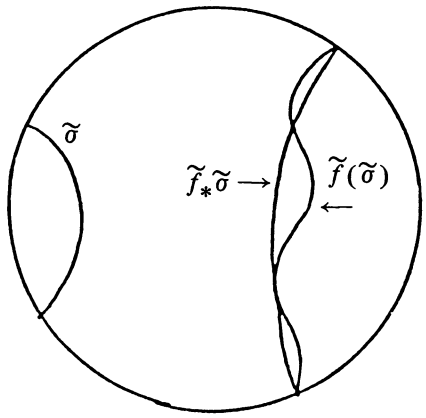

FIGURE 1

Each proper free homotopy class of closed curves or proper arcs in N contains a unique geodesic. Thus for any closed geodesic or proper geodesic arc σ, we may define $f_*(\sigma)$ to be the unique geodesic in the proper free homotopy class of $f(\sigma)$.

If σ and γ are simple proper arcs or simple closed curves in N, define the geometric intersection number $I(\sigma,\gamma)$ of σ and γ to be the minimum cardinality of $\sigma_0 \cap \gamma_0$ where σ_0 is properly homotopic to σ and γ_0 is properly homotopic to γ. When σ and γ are geodesics, then $I(\sigma,\gamma) = \mathrm{Card}(\sigma \cap \gamma)$, the cardinality of $\sigma \cap \gamma$. More generally (Expose 3 of [**F-L-P**]), if σ and γ do not have any endpoints in common, then $I(\sigma,\gamma) = \mathrm{Card}(\sigma \cap \gamma)$ if and only if no complementary region of $N - (\sigma \cup \gamma)$ is a bigon (i.e. a disk whose frontier consists of one arc in σ and one arc in γ). We will show that $f_\#$ has exponential growth by showing that there exists a simple proper geodesic arc γ, a simple closed geodesic σ and $k > 0$ such that $I(f_*^{k^n}(\sigma),\gamma)$ grows exponentially in n.

The universal cover $\tilde{N}$ of N can be identified with the hyperbolic plane H. We use the Poincaré disk model for H. Thus H is topologically equal to int D^2 and the geodesics of H are the segments of Euclidean circles and straight lines in E^2 which intersect ∂D^2 orthogonally. In particular, geodesics in H are in one-to-one correspondence with pairs of end points on S_∞.

We compactify H by adding $S_\infty = \partial D^2$ and choosing the neighborhoods of $P \in S_\infty$ in $H \cup S_\infty$ to be those of P in D^2. A fundamental result of Nielsen (Corollary 1.2 of [**H-T**]) states that any lift $\tilde{f}: H \to H$ of $f: N \to N$ extends uniquely to a homeomorphism of $H \cup S_\infty$. This provides a useful equivalent definition of $f_*(\sigma)$. Let $\tilde{\sigma} \subset H$ be any lift of σ and $\tilde{f}: H \cup S_\infty \to H \cup S_\infty$ any extended lift of f. Then $f_*(\sigma)$ is the projected image of the unique geodesic $\tilde{f}_*\tilde{\sigma}$ in H whose end points are the $\tilde{f}$ images of the end points of $\tilde{\sigma}$. See Figure 1. (The equivalence of these definitions is proved in §1 of [**H-T**].) We will use this latter definition in proving Proposition 3.1.

It will be convenient to choose a hyperbolic structure on N in which a given finite collection $\{A_i\}$ of embedded proper arcs are geodesics. This can be achieved so long as none of the components of $N - \bigcup A_i$ is a bigon or a triangle (cf. Expose 3, §3 of [**F-L-P**]).

2. The first step of our proof relies on an important result of A. Katok.

THEOREM 2.1 [**K**]. *If $f: M^2 \to M^2$ is a $C^{1+\alpha}$ diffeomorphism of a compact surface and $h(f) > 0$, then there exists a compact hyperbolic invariant set Λ which is perfect and has* $\mathrm{Per}(f/\Lambda)$ *dense in* Λ.

This set Λ has a great deal of structure which we wish to exploit. For a general reference on invariant sets with hyperbolic structure see [**S**] or Chapter 5 of [**G-H**]. Through each point $x \in \Lambda$ there are smooth (C^1) arcs $W^s_\varepsilon(x)$ and $W^u_\varepsilon(x)$ called local stable and unstable manifolds respectively and defined by

$$W^s_\varepsilon(x) = \{y | d(f^n(x), f^n(y)) \leq \varepsilon \text{ for } n \geq 0\}$$

and

$$W^u_\varepsilon(x) = \{y | d(f^n(x), f^n(y)) \leq \varepsilon \text{ for } n \leq 0\}.$$

There is a $\delta > 0$ such that whenever $d(x, y) < \varepsilon$ there is a unique point $[x, y] \in W^u_\varepsilon(x) \cap W^s_\varepsilon(y)$; moreover $[x, y] \in \Lambda$. This gives rise to so-called *local canonical coordinates* for Λ.

More precisely if $\hat{W}^s_\varepsilon(x) = W^s_\varepsilon(x) \cap \Lambda$ and $\hat{W}^u_\varepsilon(x) = W^u_\varepsilon(x) \cap \Lambda$ then the function $\Phi: \hat{W}^u_\varepsilon(x) \times \hat{W}^s_\varepsilon(x) \to \Lambda$ given by $\Phi(u, v) = [u, v]$ is a homeomorphism from $\hat{W}^u_\varepsilon(x) \times \hat{W}^s_\varepsilon(x)$ onto V_x, a neighborhood of x in Λ. See [**S**] for more details. It is this "local product structure" which we will make use of.

Since $f_\#$ has exponential growth on π_1 precisely when $(f^n)_\#$ does we can replace f by f^n and assume that we are dealing with a map that has a nonboundary fixed point p in Λ. In particular, there are sequences on both sides of p in $\hat{W}^u_\varepsilon(p) - \{p\}$ that limit on p and similarly for $\hat{W}^s_\varepsilon(p) - \{p\}$.

LEMMA 2.2. *If $\Lambda \subset M^2$ is a hyperbolic invariant set of a smooth map f with nonboundary fixed point p, then there is a smooth quadrilateral Q embedded in M^2 whose vertices x_1, x_2, x_3, x_4 are periodic points of f and which satisfies the following properties. Let A_1 and A_2 be opposite "vertical" sides of Q and B_1, B_2 the "horizontal" sides, then*

(Q1) $f(Q) \cap (B_1 \cup B_2) = \varnothing$.

(Q2) $Q \cap f(A_1 \cup A_2) = \varnothing$.

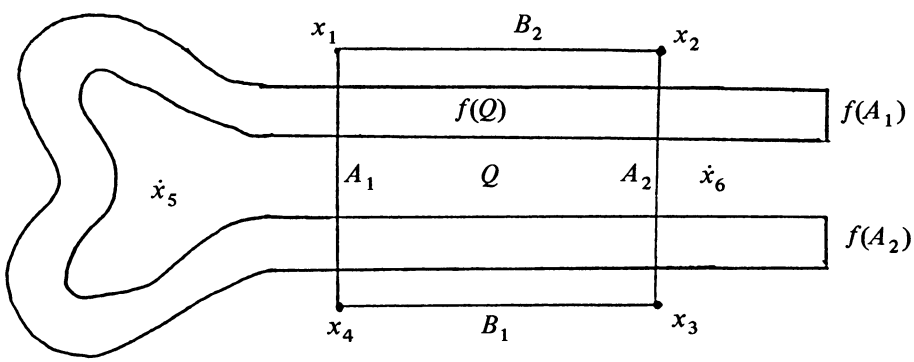

FIGURE 2

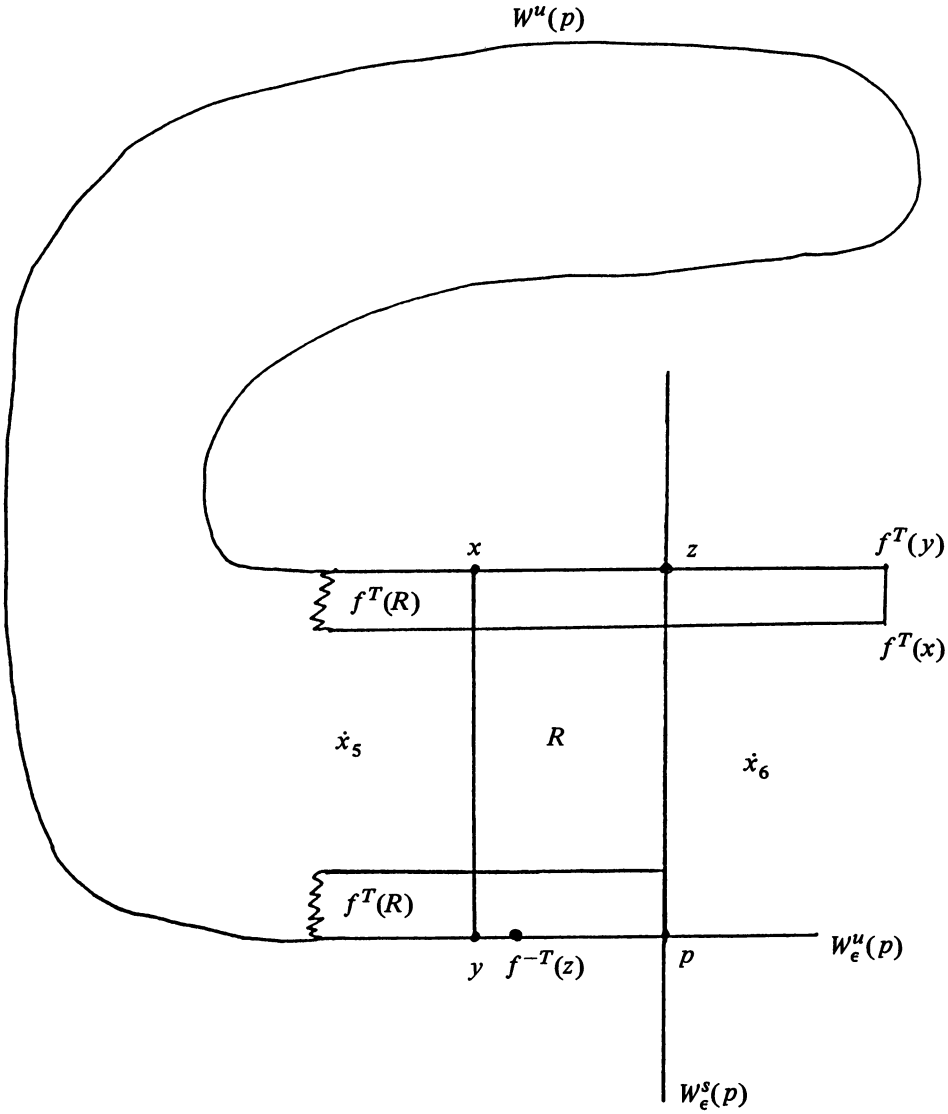

FIGURE 3

(Q3) *$Q \cap f(Q)$ has exactly two components, each of which has two sides in Q and two in $f(Q)$.*

(Q4) *There is a finite subset $P' \subset \mathrm{Per}(f)$ which contains the vertices $\{x_1, x_2, x_3, x_4\}$ of Q and such that no arc in $f(Q) - \mathrm{int}\, Q$ with end points in different components of $Q \cap f(Q)$ is homotopic rel end points in $M - P'$ to an arc in A_1 or A_2. (See Figure 2.)*

REMARK. The existence of Q satisfying (Q1)–(Q3) is well known. Property (Q4) guarantees that the horseshoe is homotopically nontrivial.

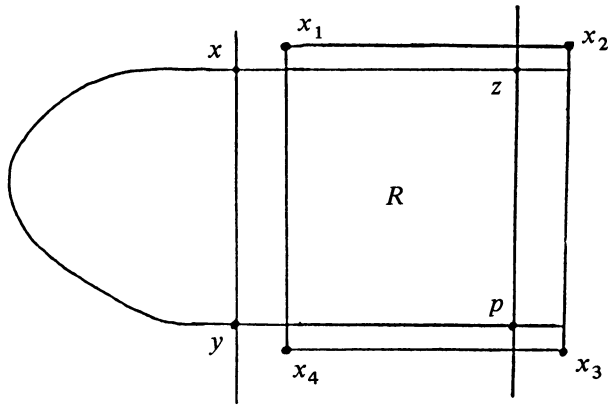

FIGURE 4

PROOF. We will make use of the global unstable manifold

$$W^u(p) = \bigcup_{n\geq 0} f^n(W^u_\varepsilon(p)).$$

Choose a point $z \in \hat{W}^s_\varepsilon(p) \cap W^u(p)$ with the property that there is no other point that is in W^s_ε between p and z and also in $W^u(p)$ between p and z (see Figure 3). We assume at first that $W^s(p)$ and $W^u(p)$ intersect with opposite intersection numbers at p and z. Choose some large T and $y \in \hat{W}^u_\varepsilon(p)$ near $f^{-T}(z)$ so that y is a limit of points on both sides of it in $\hat{W}^u_\varepsilon(p)$ and so that $f^{-T}(z)$ is between y and p on $W^u(p)$. Let $x = [z, y] = \hat{W}^u_\varepsilon(z) \cap \hat{W}^s_\varepsilon(y)$. Note that the points x, y, z and p form the four corners of a quadrilateral R whose sides are made up of segments of stable and unstable manifolds and that $f^T(R)$ is as shown in Figure 3. In particular $f^T(R) \cap R$ consists of two components.

We now choose nonboundary periodic points x_1, x_2, x_3 and x_4 in Λ that are quite close to the points x, z, p and y respectively and in the configuration shown in Figure 4. We assume without loss of generality that x_2 is higher than x_1 and that x_4 is farther to the left than x_1. The quadrilateral Q is formed by joining these points by smooth arcs that approximate the local stable and unstable manifolds. The hyperbolicity of Λ and the fact that $f(Q)$ approximates $f(R)$ makes it clear that Q has the properties (Q1)–(Q3).

To show that (Q4) holds we observe that if there is an arc in $f(Q) - \operatorname{int} Q$ which is homotopic rel end points to an arc in A_1 or A_2 then one (or both) of the components of $M^2 - (Q \cup f(Q))$ must be a disk. We can choose periodic points x_5 and x_6 in Λ which are, as in Figure 2, in each of the components of the complement of $Q \cup f(Q)$. For example, choose x_5 above, to the left and near to $[x_4, f^T x_2]$. If $P' = \{x_i\}_{i=1}^6$ property (Q4) will be satisfied.

The other case when $W^u(p)$ and $W^s(p)$ intersect at p and z with the same intersection number as in Figure 5 is handled similarly.

We first choose points x' and z' on $W^s_\varepsilon(y)$ and $W^u_\varepsilon(p)$ respectively as in Figure 5. They can be chosen so that that quadrilateral R' with vertices x', z', p and y will have an image under $f^{T'}$ as shown. In order to do this it may be necessary to increase T' (thereby making $f^{T'}$ much more contracting in the vertical direction)

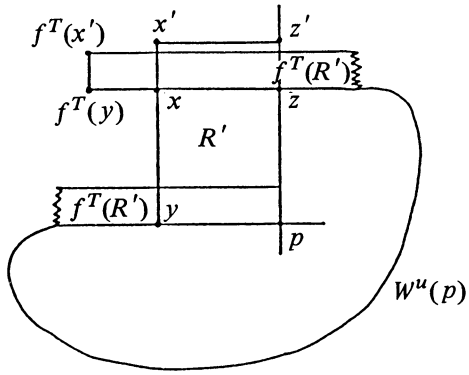

FIGURE 5

and subsequently rechoose y so that $f^{T'}(y)$ will be as shown. From this point the argument proceeds as in the earlier case. Q.E.D.

REMARK. Smoothness plays no role in the remainder of the proof; we use only the existence of the (generalized) horseshoe $f|Q$. Fried [**F**] showed that for the standard horseshoe on S^2, P can be a single periodic orbit of period 5. This is easy to verify using Thurston's classification theorem [**T**], because a homeomorphism of the five-times punctured sphere that cyclically permutes the punctures is never reducible and is isotopic to an isometry if and only if its fifth iterate is isotopic to the identity. This argument is known to Boyland, Kerckhoff and probably others. By combining [**Sm**] with [**T**], this sort of argument can be extended to the general case. The proof we give below is more constructive than the above alternative and introduces a new and useful technique for detecting exponential growth.

3. Using Lemma 2.2, we now complete the proof of the main theorem. Let P be the union of the orbits of the elements of P' and let $N = M - P$. Assuming without loss of generality that the interiors of A_1, A_2, B_1 and B_2 are disjoint from P, then Q becomes a punctured ideal quadrilateral.

We say that an *arc* in Q that connects A_1 to A_2 is *horizontal.* For each closed curve or proper arc γ, define $h(\gamma)$ to be the number of horizontal arcs contained in $\gamma \cap Q$. Proposition 3.1 is the main result of this section.

PROPOSITION 3.1. *For each simple proper geodesic arc β, $h(f_*\beta) \geq 2h(\beta)$.*

PROOF OF PROPOSITION 3.1. Properties (Q1)–(Q3) imply that every arc in $f(Q)$ that connects $f(A_1)$ to $f(A_2)$ contains at least two horizontal arcs. It therefore suffices to show that each horizontal arc in $\beta \cap Q$ determines an arc in $f_*(\beta) \cap f(Q)$ that connects $f(A_1)$ to $f(A_2)$.

It will be convenient to choose a hyperbolic metric in which A_i, B_i, $f(A_i)$ and $f(B_i)$ $(i = 1, 2)$ are geodesics. Properties (Q1)–(Q4) guarantee that any unpunctured component of $N - (\partial Q \cup f(\partial Q))$ is a quadrilateral. Thus the criterion stated in §1 is satisfied and the desired metric exists. This can be seen more directly as follows. Start with any hyperbolic metric μ. Choose homeomorphisms $h_1, h_2 \colon N \to N$ that are isotopic to the identity and such that: h_1 carries A_i, B_i and $f(A_i)$ to the unique geodesic in their proper homotopy classes; h_2 carries $h_1 f(B_i)$ to the unique

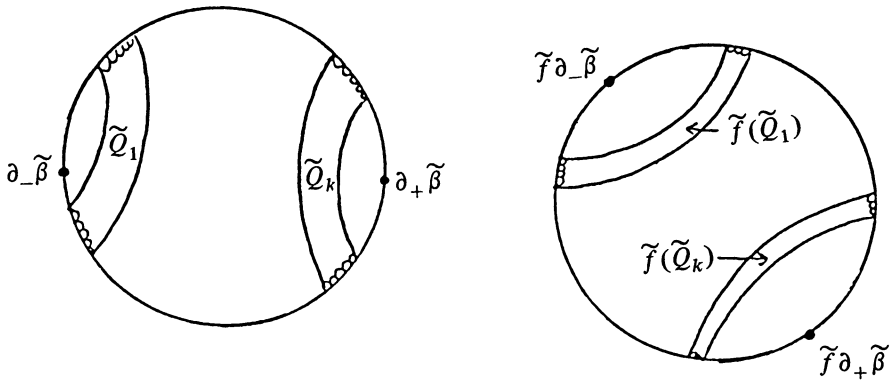

FIGURE 6

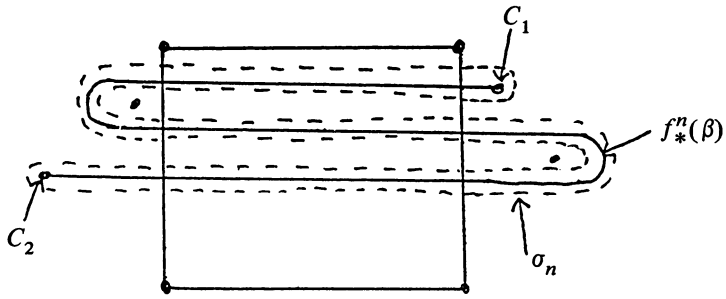

FIGURE 7

geodesic in its proper homotopy class; and h_2 fixes $h_1(A_i)$, $h_i(B_i)$ and $h_1 f(A_i)$ setwise. Then the pulled back metric $(h_2 \circ h_1)^*\mu$ has the desired properties.

Fix an extended lift $\tilde{f}: H \cup S_\infty \to H \cup S_\infty$ and a lift $\tilde{\beta}$ of β. Denote the endpoints of $\tilde{\beta}$ by $\tilde{\beta}_\pm$ and the full preimage of Q by $\bar{\bar{Q}}$. A component $\tilde{Q}_i$ of $\bar{\bar{Q}}$ is either disjoint from $\tilde{\beta}$ or intersects $\tilde{\beta}$ in an arc. We say that $\tilde{Q}_i$ is *horizontal with respect to* $\tilde{\beta}$ if $\tilde{\beta} \cap \tilde{Q}_i \neq \varnothing$ and if the endpoints of $\tilde{\beta} \cap \tilde{Q}_i$ project to a point on A_1 and a point on A_2. If $\tilde{Q}_i$ is horizontal with respect to $\tilde{\beta}$, then $\tilde{\beta} \cap \tilde{Q}_i$ projects onto a horizontal arc in $\beta \cap Q$. Conversely, each horizontal arc in $\beta \cap Q$ determines a component of $\bar{\bar{Q}}$ that is horizontal with respect to $\tilde{\beta}$.

Let $\tilde{Q}_1, \ldots, \tilde{Q}_k$ be the components of $\bar{\bar{Q}}$ that are horizontal with respect to $\tilde{\beta}$. Since $\tilde{f}$ carries components of $\bar{\bar{Q}}$ to components of the full preimage $\tilde{f}(\bar{\bar{Q}})$ of $f(Q)$ and preserves the cyclic ordering on S_∞, each $\tilde{f}(\tilde{Q}_i)$ is a component of $\tilde{f}(\bar{\bar{Q}})$ that separates $\tilde{f}(\tilde{\beta}_+)$ from $\tilde{f}(\tilde{\beta}_-)$. Thus $\tilde{f}_*(\tilde{\beta}) \cap \tilde{f}(\tilde{Q}_i)$ is an arc in $\tilde{f}(\tilde{Q}_i)$ with endpoints on the components of $\partial(\tilde{f}(\tilde{Q}_i))$ that are the $\tilde{f}$-images of the components of $\partial(\tilde{Q}_i)$ that contain the endpoints of $\tilde{\beta} \cap \tilde{Q}_i$. We conclude that each $\tilde{f}_*(\tilde{\beta}) \cap \tilde{f}(\tilde{Q}_i)$ projects onto an arc in $f_*(\beta) \cap f(Q)$ that connects $f(A_1)$ to $f(A_2)$.

PROOF OF THEOREM. Choose a proper geodesic arc β with endpoints on distinct punctures C_1 and C_2 such that $h(\beta) \geq 1$ and such that neither C_1 nor C_2 is an endpoint of A_1. Replacing f by some f^k $(k > 0)$ if necessary we may

assume that f fixes C_1 and C_2. Let σ_n be the simple closed geodesic determined by the frontier of a regular neighborhood of $f_*^n(\beta) \cup C_1 \cup C_2$. See Figure 7. Since $N - (f_*^n(\beta) \cup A_1)$ contains no bigons, the same is true for $N - (\sigma_n \cup A_1)$. Thus $I(f_*^n \sigma_0, A_1) = I(\sigma_n, A_1) = \mathrm{Card}(\sigma_n, A_1) = 2 \times \mathrm{Card}(f_*^n \beta, A_1) \geq 2^{n=1}$. □

REFERENCES

[**B**] R. Bowen, *Entropy and the fundamental group*, Structure of Attractors, Lecture Notes in Math., vol. 668, Springer-Verlag, Berlin and New York, 1978, pp. 21–29.

[**F-L-P**] A. Fathi, F. Laudenbach and V. Poeńaru, *Traveaux de Thurston sur les surfaces*, Asterisque **66–67** (1979), 181–207.

[**F**] D. Fried, *Entropy and twisted cohomology*, Topology **25** (1986), 455–470.

[**F2**] ——, *Growth rate of surface homeomorphisms and flow equivalence*, Ergodic Theory Dynamical Systems **5** (1985), 539–563.

[**H-T**] M. Handel and W. Thurston, *New proofs of some results of Nielsen*, preprint.

[**G-H**] J. Guckenheimer and P. Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Springer Series in Applied Math. Sci. 42, Springer-Verlag, New York, 1983.

[**K**] A. B. Katok, *Lyapunov exponents, entropy and periodic points for diffeomorphisms*, Publ. Math. Inst. Hautes Études Sci. **51** (1980), 137–173.

[**S**] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 797–817.

[**T**] W. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Preprint.

DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS 60201

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, LEHMAN COLLEGE, CITY UNIVERSITY OF NEW YORK, BRONX, NEW YORK 10468