

Generating special Markov partitions for hyperbolic toral automorphisms using fractals

TIM BEDFORD

*Department of Pure Mathematics and Mathematical Statistics,
University of Cambridge, 16, Mill Lane, Cambridge, England*

(Received 7 March 1985 and revised 20 June 1985)

Abstract. We show that given some natural conditions on a 3×3 hyperbolic matrix of integers A ($\det A = 1$) there exists a Markov partition for the induced map $A(x + \mathbb{Z}^3) = A(x) + \mathbb{Z}^3$ on T^3 whose transition matrix is $(A^{-1})'$. For expanding endomorphisms of T^2 we construct a Markov partition so that there is a semiconjugacy from a full (one-sided) shift.

Introduction

A highly flexible method for generating fractals is provided in [6]. We use the method here to construct Markov partition boundaries for hyperbolic automorphisms of T^3 and expanding endomorphisms of T^2 . (Results of Urbanski [10] and Mañé [8] show that such boundaries must be fractal.) We start with a good initial tiling of the torus and are able to specify exactly and follow, via Dekking's fractal generation method, all the boundary perturbations needed to obtain a Markov partition. This makes it possible to calculate the transition matrix and to draw the partitions. For a definition of a Markov partition we refer the reader to Bowen [4], [5] (Bowen does not consider the case of expanding endomorphisms, but definitions similar to those he gives for the hyperbolic situation work if one deletes the conditions relating to stable manifolds). Bowen's definition requires that the sets in our partition be small. We shall relax this condition. The effect of this is that if $\tilde{A}: T^3 \rightarrow T^3$ is our hyperbolic automorphism and R_i, R_j are two elements of our Markov partition \mathcal{R} then $\tilde{A}(\text{int } R_i) \cap \text{int } R_j$ may have more than one connected component. The transition matrix for \mathcal{R} will be the $|\mathcal{R}| \times |\mathcal{R}|$ matrix $B = (b_{ij})$ where b_{ij} equals the number of connected components of $\tilde{A}(\text{int } R_j) \cap \text{int } R_i$. The usual type of construction of a semiconjugacy from the subshift of finite type can still be done (Adler and Marcus [1, p. 15] talk about transition matrices over \mathbb{Z}^+ instead of the usual $\{0, 1\}$).

Dekking's construction of fractals ([6], [7]) is a nice geometric method in which various directed line segments have symbols associated with them. A rule by which one directed line segment is transformed into a directed polygonal line is then written as an endomorphism θ of the free semigroup S^* generated by the set of symbols. $K[V]$ is the directed polygonal line corresponding to a word $V \in S^*$. A map $f: S^* \rightarrow \mathbb{R}^m$ gives the coordinates of the endpoint of $K[V]$. The polygonal line

$K[\theta s]$ is bigger than the line $K[s]$ and so we normalize by a linear map L so that the endpoints of $L^{-1}K[\theta s]$ are the same as those of $K[s]$. Dekking [6] proves that as $n \rightarrow \infty$, $\lim L^{-n}K[\theta^n s]$ exists (where convergence is in the Hausdorff metric). We denote this limit by $K_\theta(s)$. $K_\theta(s)$ is typically a fractal and has the useful property that it is connected. Furthermore, it has an invariance property that we shall often use,

$$LK_\theta(s) = K_\theta(\theta s).$$

The proofs given here are fairly detailed sketches - detailed proofs can be found in [2]. I would like to thank Caroline Series, who supervised this work, Michel Dekking, Colin Rourke, and the referee for their advice.

Constructing partitions

THEOREM 1. *Let A be a 2×2 matrix of integers inducing an expanding endomorphism on T^2 by $\tilde{A}(x + \mathbb{Z}^2) = A(x) + \mathbb{Z}^2$. Then there is a Markov partition for $\tilde{A}: T^2 \rightarrow T^2$ so that there is a semiconjugacy to A from the (one sided) full shift on $|\det A|$ symbols.*

Proof. Let $e_1 = (1, 0)$, $e_2 = (0, 1)$ in \mathbb{R}^2 . We make our construction in the covering plane. Let $S = \{s_1, s_2, s_1^{-1}, s_2^{-1}\}$, $f(s_i) = e_i$, $f(s_i^{-1}) = -e_i$ and $W = s_1 s_2 s_1^{-1} s_2^{-1}$. This makes $K[W]$ the boundary of the unit square. Choose θ so that the ‘sides’ of $K[\theta W]$ do not cross over. For instance one way would be to make $K[\theta s_i]$ the nearest polygonal line with vertices in \mathbb{Z}^2 in the anticlockwise direction from $AK[s_i]$ (we call this choice the *anticlockwise perturbation*). There can be many different choices for θ in general. There is a natural orientation of line segments in $K[\theta^n W]$ given by symbol order in $\theta^n W$. Thus we can define the *inside* of $K[\theta^n W]$ as all points to the left of line segments having a single orientation defined ($K[\theta^n W]$ might have multiple self intersections). Let V^n be the closure of the set of points inside $K[\theta^n W]$ (in particular V^0 is the unit square). $A^{-n}V^n$ is our n th level approximation to the Markov partition.

The idea now is to use θ to transform the tiling of the plane by copies of V^n into a tiling by copies of V^{n+1} . Notice that V^1 is the union of $|\det A|$ copies of V^0 that have mutually disjoint interiors and which tile the plane. An induction now shows that for each n , V^{n+1} tiles the plane and is the union of $|\det A|$ copies of V^n having mutually disjoint interiors. The idea here is that if two V^n 's intersect in their interior then the intersection must be a copy of V^{n+1} , contradicting the fact that the V^n tile. By construction each of the tilings of \mathbb{R}^2 by copies of $A^{-n-1}V^n$ and hence also the limiting tiling \mathcal{R}' is invariant under translation by elements of $A^{-1}\mathbb{Z}^2 \supset \mathbb{Z}^2$. In order to obtain the required Markov partition, project \mathcal{R}' onto the torus to give a partition \mathcal{P} . Since V^n is tiled by $|\det A|$ copies of V^{n-1} , $A(A^{-n-1}V^n)$ is tiled by $|\det A|$ copies of $A^{-n}V^{n-1}$ and this implies that \mathcal{P} satisfies the Markov property and its transition matrix has every entry 1. □

Remark. A quantity which provides a measurement of the amount of crinklyness of a curve C is *capacity*. Let $N(\varepsilon)$ be the minimum number of balls of size ε required

to cover C . Then

$$\overline{\text{cap}}(C) = \limsup_{\epsilon \rightarrow 0} (\log N(\epsilon)) / (-\log \epsilon)$$

and

$$\underline{\text{cap}}(C) = \liminf_{\epsilon \rightarrow 0} (\log N(\epsilon)) / (-\log \epsilon).$$

If $\overline{\text{cap}} = \underline{\text{cap}}$ we call this the capacity of C . A result of Urbanski [10] implies that the capacity of $\partial\mathcal{P}$ is not less than $2 - ((\log |\lambda_2|) / (\log |\lambda_1|))$, where $1 < |\lambda_2| < |\lambda_1|$ are the eigenvalues of A . It is not difficult to show [3] that if $B(\mathcal{P}) = (b_{ij})$ is defined by letting b_{ij} be the total number of times s_j and s_j^{-1} appear in $\theta(s_i)$, and λ is the maximal eigenvalue of B then

$$\overline{\text{cap}} \partial\mathcal{P} \leq 1 + (\log \lambda - \log |\lambda_2|) / \log |\lambda_1|.$$

In particular there are \mathcal{P} with this minimal capacity. In the case where A has complex eigenvalues there are Markov partitions whose boundaries have capacity arbitrarily close to the value 1. Figures 1 and 2 show two Markov partitions for the same map. The boundary of the partition shown in figure 2 has capacity greater than that for figure 1.

A similar construction to that used above can be used in the T^3 hyperbolic setting. We shall prove a generalisation of the following result of Manning [9].

THEOREM 2. *Let A be a hyperbolic 2×2 matrix of positive integers and determinant $+1$. Then there is a semiconjugacy from the subshift of finite type associated with the matrix A^t to the hyperbolic automorphism (T^2, \tilde{A}) (where t denotes transposition).*

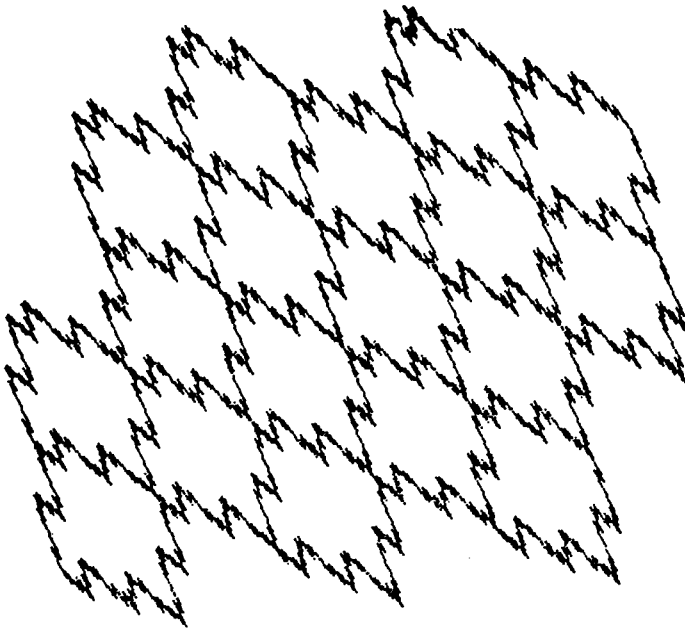


FIGURE 1. Markov partition boundary generated by the endomorphism $\theta_{s_1} = s_1 s_2 s_1 s_1 s_2 s_1 s_1$ and $\theta_{s_2} = s_2 s_2 s_1 s_2 s_2$.

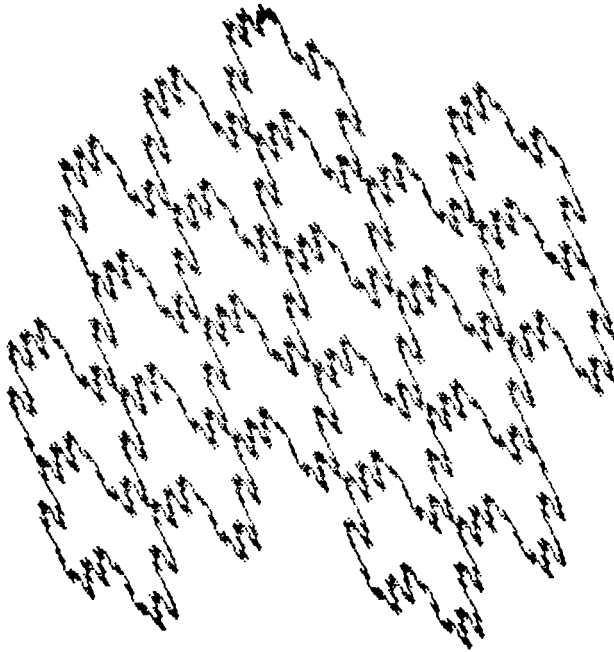


FIGURE 2. Markov partition boundary generated by the endomorphism $\theta_{s_1} = s_1 s_2 s_1 s_2 s_1 s_2^{-1} s_1 s_2 s_1$ and $\theta_{s_2} = s_2 s_2 s_1 s_2 s_2$.

N.B. In both cases the mapping of the torus is induced by the matrix $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. The figures show a fundamental region for the torus and 18 1-cylinders.

THEOREM 3. *Let A be a hyperbolic 3×3 matrix of integers such that:*

- (i) $\det A = 1$;
- (ii) A^{-1} is a non-negative matrix;
- (iii) A has a single real contracting eigenvalue λ_3 and $\lambda_3 > 0$;
- (iv) condition (*) (defined below) is satisfied.

Then the induced map $\tilde{A}: T^3 \rightarrow T^3$ has a Markov partition with transition matrix $(A^{-1})^t$.

Note. It will often help the reader to draw a two dimensional picture corresponding to the three dimensional situation described below.

Proof. Let E^u be the linear subspace of \mathbb{R}^3 spanned by eigenvectors for expanding eigenvalues, and E^s the subspace spanned by an eigenvector for the contracting eigenvalue. Define C as the positive cone $\{(x_1, x_2, x_3): x_i \geq 0\}$. By assumptions (ii) and (iii) of the theorem, E^u is a two dimensional space and intersects C only at the origin. Similarly E^s is one dimensional and intersects the interior of C .

(a) We now show how to approximate the claimed partition arbitrarily well. Here we define the first approximation. Let $p_s: \mathbb{R}^3 \rightarrow E^u$ be projection down stable manifolds and $\{e_i: i = 1, 2, 3\}$ be the standard basis for \mathbb{R}^3 . We write F_i for the face of the unit cube with edges e_j, e_k (where i, j, k are all different) and $H_i = p_s(F_i)$. Condition (*) is:

$$H = \bigcup_i H_i \text{ satisfies } AH \supset H. \tag{*}$$

Three prisms, P_i for $i = 1, 2, 3$, are defined by

$$P_i = [a_i, 0] \times H_i \subset E^s \times E^u = \mathbb{R}^3, \quad \text{where } a_i = E^s \cap (-e_i + E^u).$$

Writing $P = \bigcup_i P_i$ we claim that $\mathcal{P} = \{q + P : q \in \mathbb{Z}^3\}$ tiles \mathbb{R}^3 . This is proved by considering how the different faces of P meet translates of P . We call a face of $q + P$ an s -face if it is a union of line segments parallel to E^s , or a u -face if it is parallel to E^u .

Note first that the bottom (u -) face of P_i , $a_i + H_i$ satisfies $a_i + H_i \subset -e_i + H$ where $-e_i + H$ is a u -face of $-e_i + P$. Hence the u -faces of any $q + P$ are either subsets of, or contain, u -faces of translates of $q + P$.

In order to show that s -faces of $q + P$ meet s -faces of translates of $q + P$ it is enough to show (since $p(E^u)$ is dense in T^3) that the intersection of E^u with \mathcal{P} is made up of non-overlapping copies of H_i , $i = 1, 2, 3$, outside H . This is done by the following construction which is also used later.

Corresponding to E^u define a 'stepped' surface \tilde{U} in the following way. \tilde{U} is a union of faces $q + F_i$, in other words,

$$(x_1, x_2, x_3) \in \tilde{U} \text{ implies some } x_i \in \mathbb{Z}^3,$$

and is the unique lowest such surface sitting above E^u . Because E^u intersects the positive cone only at the origin, if we write ${}^-I_j = (-e_j, 0]$ then \tilde{U} has the following special property:

$$E^u \cap \text{int}(q + {}^-I_j) \neq \emptyset \text{ if and only if } q + F_j \subset \tilde{U}.$$

The special property implies that faces $q + F_i$ of \tilde{U} correspond under p_s to intersections of E^u with $q + P_i$. (Figure 3 shows the analogous picture in two dimensions.) Thus the intersection of E^u with \mathcal{P} is made up of non-overlapping copies of H_i

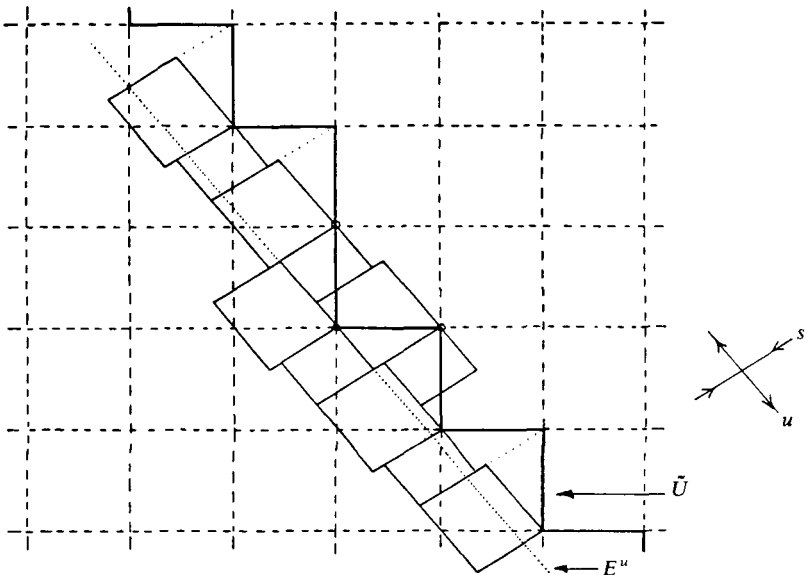


FIGURE 3

(we now call this the pattern of \mathcal{P} in E^u). We have now proved that \mathcal{P} gives a tiling of \mathbb{R}^3 . Figure 4 shows how various translates of the P_i 's fit together in \mathbb{R}^3 .

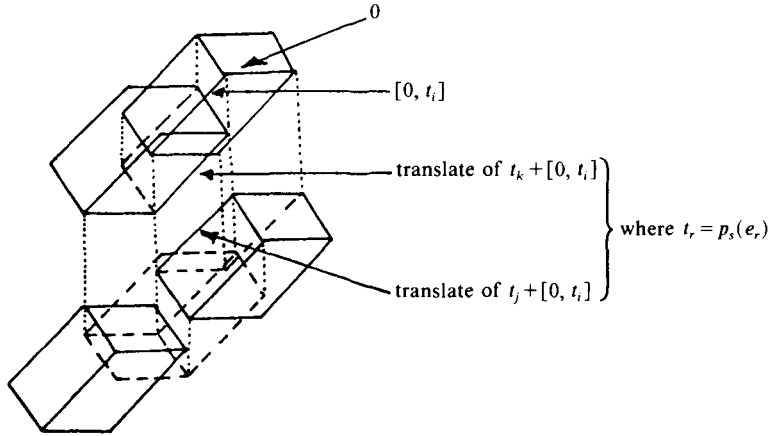


FIGURE 4

(b) The s -faces of the P_i now have to be altered so that they satisfy the Markov conditions. We choose a way to alter inductively the faces (as in theorem 1) by taking the anticlockwise perturbation of $A([0, p_s(e_i)])$ in the pattern for each i . However we have to check first that $A(p_s(e_j) + [0, p_s(e_i)])$ (an edge of $A(H_k)$) can be perturbed in the same way as $A([0, p_s(e_i)])$ whilst staying inside the pattern, and secondly that the same choice can be made for the corresponding lines in $(x + E^u) \cap A(P_i)$ for any $x \in \mathbb{R}^3$, i.e. for the whole s -face. Up to integer translation $[0, p_s(e_i)]$, $p_s(e_j) + [0, p_s(e_i)]$ and $p_s(e_k) + [0, p_s(e_i)]$ lie in the same s -face of either P_j or P_k . Thus we have only to prove that the choice of perturbation for the intersection of $x + E^u$ with an s -face of P_k can be made independently of $x \in P_k$. This is true if for all $q \in \mathbb{Z}^3$,

$$(q + H_j) \cap \text{int}(AP_k) = \emptyset.$$

That follows because otherwise, if $r = A^{-1}q$, since $A^{-1}H \subset H$ we have $A^{-1}(q + H_j) \subset r + H$ and in particular there is a P_m such that $\text{int} A(r + P_m) \cap \text{int} AP_k \neq \emptyset$ which contradicts the fact that \mathcal{P} tiles.

It is easy to check that the perturbed P_i give a new tiling \mathcal{P}^1 of \mathbb{R}^3 . Furthermore, writing $W^u(x, N) = (x + E^u) \cap N$ we have by construction that $W^u(Ax, AP_i^1) \supset W^u(Ax, p + P_j)$, some $p \in \mathbb{Z}^3$, $x \in \text{int} P_i^1 \in \mathcal{P}^1$ and $Ax \in \text{int} p + P_j$.

(c) We now introduce recurrent sets. Choose a symbol s_i corresponding to $p_s(e_i)$, an endomorphism θ corresponding to the anticlockwise perturbation and words $W_i = s_j s_k s_j^{-1} s_k^{-1}$. As in theorem 1 let V_i^n be the region inside $K[\theta^n W_i]$. Our closer approximations to the required Markov partition are $\mathcal{P}^n = \{q + P_i^n\}$ where $P_i^n = [a_i, 0] \times (A^{-n}V_i^n)$. An induction shows that

$$\begin{aligned} A(W^u(x, A^n P_i^{n+1})) &\supset W^u(Ax, q + A^n P_j^n) \\ \text{and} \quad A(W^s(x, A^n P_i^{n+1})) &\subset W^s(Ax, q + A^n P_j^n), \end{aligned} \tag{**}$$

where $x \in \text{int}(A^n P_i^{n+1})$ and $Ax \in \text{int}(q + A^n P_j^n)$. As $n \rightarrow \infty$ the tiles \mathcal{P}^n converge to a tiling \mathcal{R}' . Projecting \mathcal{R}' onto T^3 gives a partition which satisfies the Markov conditions by (**).

(d) To calculate the transition matrix $B = (b_{ij})$ for \mathcal{R} note that for all n ,

$$b_{ij} = \text{card} \{q + P_j^n: \text{int}(q + P_j^n) \cap \text{int} AP_i^{n+1} \neq \emptyset\},$$

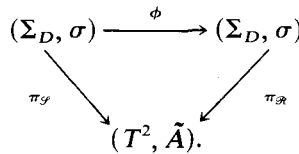
and that this equals the number of copies of H_j in the patterns tiling of V_i' . Lift the lines $K[\theta_{s_j}] \subset E^u$ into the stepped surface \tilde{U} to give polygonal lines L_j . Denoting by \tilde{U}_i the lift to \tilde{U} of V_i' notice that \tilde{U}_i is bounded by $L_j, L_k, L_j + A(e_k), L_k + A(e_j)$ and that b_{ir} equals the number of $q + F_r$ in \tilde{U}_i . We can easily calculate the number of such faces by projecting \tilde{U}_i onto the (e_m, e_n) -plane where $r \neq m, n$. Then we see that

$$b_{ir} = \det \begin{pmatrix} a_{mk} & a_{mj} \\ a_{nk} & a_{nj} \end{pmatrix},$$

which implies that $B = (\text{adj } A)^t = (A^{-1})^t$. □

Coding between Markov partitions

The proof of theorem 1 gave us lots of Markov partitions with different boundary capacities for the same map of T^2 . Let $D = |\det A|$, and suppose \mathcal{S}, \mathcal{R} are two Markov partitions for A constructed as in theorem 1. We denote the measure of maximal entropy for the full shift $\sigma: \Sigma_D \rightarrow \Sigma_D$ (which projects to Haar measure on the torus) by μ . There is an induced isomorphism $\phi: \Sigma_D \rightarrow \Sigma_D$ defined μ -a.e. so that the following diagram commutes



The anticipating function $a_\phi: \Sigma_D \rightarrow \mathbb{N} \cup \{\infty\}$ is defined by

$$a_\phi(x) = \min \{n: y \in \Sigma_D, y_i = x_i \text{ for } i \leq n \text{ implies } (\phi(y))_0 = (\phi(x))_0\}.$$

We say that ϕ has finite expected code length if $\int a_\phi d\mu < \infty$. An argument of Adler and Marcus [1] shows that our ϕ does indeed have finite expected code length. However a better estimate of the expected code length can be obtained in terms of the matrices $B(\mathcal{S}), B(\mathcal{R})$ (defined in the remark after theorem 1). Hence the expected code length depends on how crinkly the Markov partition boundaries are. Write \mathcal{S}_{i+1} for $\bigvee_{n=0}^i \tilde{A}^{-n} \mathcal{S}$.

PROPOSITION 3. $\int a_\phi d\mu = \sum_1^\infty n_i / D^i$, where n_{i+1} is the number of elements of \mathcal{S}_{i+1} covering $\partial \mathcal{R}$.

Proof. Write $C_i(x) = \{y \in \Sigma_D: (x_0, \dots, x_{i-1}) = (y_0, \dots, y_{i-1})\}$. We have to calculate $\mu\{x \in \Sigma_D: a_\phi(x) = i\}$. Now, $i = a_\phi(x)$ means that $\pi_{\mathcal{S}}(C_i(x))$ does not intersect two distinct elements of \mathcal{R} in a set of positive measure, but that $\pi_{\mathcal{S}}(C_{i-1}(x))$ does. This implies that $\pi_{\mathcal{S}}(C_i(x))$ does not contain in its interior any part of the boundary

between elements of \mathcal{R} , but that $\pi_{\mathcal{S}}(C_{i-1}(x))$ does. Thus

$$\begin{aligned} &\text{card} \{S \in \mathcal{S}_{i+1}: S = \pi_{\mathcal{S}}(C_{i+1}(x)) \subset \pi_{\mathcal{R}}(C_0(y)), \text{ some } y\} \\ &= \text{card} \{S \in \mathcal{S}_{i+1}\} - \text{card} \{S \in \mathcal{S}_{i+1}: (\text{int } S) \cap \partial \mathcal{R} \neq \emptyset\} \\ &= D^{i+1} - n_{i+1}. \end{aligned}$$

Hence

$$\begin{aligned} &\text{card} \{S \in \mathcal{S}_{i+1}: (\text{int } S) \cap \partial \mathcal{R} = \emptyset, S \subset S' \in \mathcal{S}_i, (\text{int } S') \cap \partial \mathcal{R} \neq \emptyset\} \\ &= \text{card} \{S \in \mathcal{S}_{i+1}: S \subset \pi_{\mathcal{R}}(C_0(y))\} - D \text{card} \{S' \in \mathcal{S}_i: S' \subset \pi_{\mathcal{R}}(C_0(y))\} \\ &= D^{i+1} - n_{i+1} - D(D^i - n_i) = Dn_i - n_{i+1}. \end{aligned}$$

We now have $\mu\{x \in \Sigma_D: a_\phi(x) = i\} = n_i/D^i - n_{i+1}/D^{i+1}$. Notice that we cannot have $a_\phi(x) = 0$ because this would imply that $\mathcal{R} = \mathcal{S}$. Hence if $\mathcal{R} \neq \mathcal{S}$ we have $a_\phi(x) > 0$ for all x and $n_1 = D$. We now have

$$\int a_\phi d\mu = \lim_{N \rightarrow \infty} \sum_{i=1}^N i\mu\{x \in \Sigma_D: a_\phi(x) = i\} = \sum_1^\infty n_i/D^i.$$

It is clearly possible to estimate the n_i by geometric means. The main problem in doing this is to obtain an estimate of how distorted the sets in \mathcal{R} and \mathcal{S} can be. A straightforward induction shows that (using the notation of theorem 1)

$$|A^{-n}K[\theta^n W]|_i < 8bw \sum_{i=1}^n |\lambda_i|^{-r},$$

where $|\cdot|_i$ is the length in the eigendirection corresponding to λ_i , $b = \max\{b_{ij}: (b_{ij}) = B(\mathcal{R})\}$ and w is a bound on the norms of components of the unit basis vectors in the λ_i eigenspaces. This proves

LEMMA 5. *There is a constant $r_{\mathcal{R}} > 0$ depending only on $B(\mathcal{R})$ such that an element \hat{R} of $\hat{\mathcal{R}}$ is contained in a square of side $r_{\mathcal{R}}$, where \hat{R} and $\hat{\mathcal{R}}$ are lifts of R and \mathcal{R} to \mathbb{R}^2 .*

Let θ be the endomorphism used to generate \mathcal{R} , and λ the maximal eigenvalue of $B(\mathcal{R})$. Then there exist $a, b > 0$ such that

$$a \cdot \lambda^n \leq |\theta^n W| \leq b \cdot \lambda^n.$$

Lifting to the covering plane and letting k be a bound on the number of $S \in \mathcal{S}$ covering $K_\theta(s_i)$ for $i = 1, 2$, we have $n_i \leq kb\lambda^i$. The lemma implies that $k \leq (r_{\mathcal{R}} + 2r_{\mathcal{S}})^2/D$. A similar bound can be made on the n_i from below with the extra assumption that θ has no essential symbol duplication (essential symbol duplication occurs if there exist n, s, t, U, U' and U'' , where t is an essential symbol, such that $\theta^n(s) = UtU'tU''$ and $K[tU'] = 0$), for this implies that distinct occurrences of t in $\theta^n W$ correspond to distinct copies of $K_\theta(t)$ in $K_\theta(\theta^n W)$. This shows

PROPOSITION 6. *There are constants $C_1, C_2 > 0$ computable from $A, B(\mathcal{R})$ and $B(\mathcal{S})$ such that $\int a_\phi d\mu \leq C_1$ and if there is no essential symbol duplication $C_2 \leq \int a_\phi d\mu$.*

REFERENCES

[1] R. L. Adler and B. Marcus. Topological equivalence of dynamical systems. *Mem. Amer. Math. Soc.* **219** (1979).

- [2] T. J. Bedford. Ph.D. Thesis, Warwick University (1984).
- [3] T. J. Bedford. Dimension and dynamics for fractal recurrent sets. Preprint, King's College Research Centre, Cambridge (1985).
- [4] R. Bowen. *Equilibrium States and the Ergodic Theory of Ansov Diffeomorphisms*. Lect. Notes in Maths. Vol 470. Springer-Verlag: Berlin, 1975.
- [5] R. Bowen. Markov partitions are not smooth. *Proc. Amer. Math. Soc.* **71** (1978) 130-132.
- [6] F. M. Dekking. Recurrent sets. *Advances in Math.* **44** (1982) 78-104.
- [7] F. M. Dekking. Replicating superfigures and endomorphisms of free groups. *J. Combin. Thry. Ser. A* **32** (1982) 315-320.
- [8] R. Mañé. Orbits of paths under hyperbolic toral automorphisms. *Proc. Amer. Math. Soc.* **73**, 121-125.
- [9] A. K. Manning. Lecture notes for M.Sc course on Dynamical Systems (1983).
- [10] M. Urbanski. On capacity of continuum with non-dense orbit under a hyperbolic toral automorphism. Preprint 289, Inst. of Maths., Polish Academy of Sciences, November (1983).