TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 333, Number 1, September 1992

BOUNDARIES OF MARKOV PARTITIONS

JONATHAN ASHLEY, BRUCE KITCHENS AND MATTHEW STAFFORD

ABSTRACT. The core of a Markov partition is the nonwandering set of the map restricted to the boundary of the partition. We show that the core of a Markov partition is always a finitely presented system. Then we show that every one sided sofic system occurs as the core of a Markov partition for an n-fold covering map on the circle and every two sided sofic system occurs as the core of a Markov partition for a hyperbolic automorphism of the two dimensional torus.

0. INTRODUCTION

Markov partitions are often used to produce combinatorial models for dynamical systems. They have been shown to exist in many different settings. It is also known that even in simple seeming settings, the elements of the partition may be geometrically very complicated. One such result says that for a hyperbolic automorphism of T^3 the intersection of the stable boundaries and the unstable boundaries of the elements of a Markov partition cannot contain a C^1 arc [Bo3]. We are interested in the dynamics of the boundary. Given a Markov partition for a map, the core of the partition is the nonwandering set of the map restricted to the boundary. We see that the core is always a finitely presented system [F]. Then we concentrate on two simple cases. In the noninvertible case we take the n-fold covering maps of the circle. In the invertible case we take the hyperbolic automorphisms of T^2. In both cases we see that the core must always be a sofic system [W], and that any nonwandering sofic system occurs as the core of some partition.

Section 1 contains a discussion of Markov partitions, their associated subshift of finite type, and the covering map from the subshift of finite type to the dynamical system with the Markov partition.

Section 2 is concerned strictly with symbolic dynamics. There is a discussion of subshifts of finite type, sofic systems, factor maps between symbolic systems and the basic symbolic construction that allows us to realize sofic systems as cores of partitions.

In §3 we prove that every core is a finitely presented system. Then we restrict our attention to the covering maps of the circle and the automorphisms of T^2. We present basic examples, show that in both cases the cores must be nonwandering sofic systems and then show that every nonwandering sofic system occurs as the core of a Markov partition in this setting.

Received by the editors April 20, 1990.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 58F15.
Key words and phrases. Markov partitions, sofic systems.

0002-9947/92 $1.00 + $.25 per page

In §4 we discuss the connection between partitions of S^1 whose elements are intervals and the partitions we have produced.

We would like to thank R. F. Williams. Example 3.5 is his. It was the motivation for this work.

1. Markov partitions and their covers

Let X be a compact metric space and f a continuous onto map of X to itself. We say f is *positively expansive* if there is a $c > 0$ so that if $x \neq y$ are points in X there is an $n \in \mathbb{N}$ (we will use $\mathbb{N} = 0, 1, 2, \ldots$) so that $d(f^n(x), f^n(y)) > c$. When f is invertible, we say f is *expansive* if there is a $c > 0$ so that if $x \neq y$ are points in X there is an $n \in \mathbb{Z}$ so that $d(f^n(x), f^n(y)) > c$. In both cases we refer to c as an *expansive constant.* When f is a homeomorphism, for $\varepsilon > 0$ we define the *ε-stable set* to be

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

and the *ε-unstable set* to be

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

There are two metrization theorems that make dealing with these maps much easier.

Theorem. *Let X be a compact metric space and f a continuous onto map from X to itself*:

(i) (*Coven and Reddy* [CR]) *if f is positively expansive then there is a metric d, compatible with the original metric, an $\varepsilon > 0$ and $\lambda > 1$ so that when $d(x, y) < \varepsilon$, $d(f(x), f(y)) > \lambda d(x, y)$;*

(ii) (*Fried* [F]) *if f is an expansive homeomorphism then there is a metric d, compatible with the original metric, an $\varepsilon > 0$ and $\lambda > 1$ so that when $y \in W^u_\varepsilon(x)$, $d(f(x), f(y)) > \lambda d(x, y)$ when $y \in W^s_\varepsilon(x)$, $d(f^{-1}(x), f^{-1}(y)) > \lambda d(x, y)$.*

We will always assume that the metrics we use are adapted in this sense. In the case where f is a homeomorphism we say a closed set $R \subseteq X$ is a *rectangle* if there is an $\varepsilon > 0$ so that diameter(R) is less than ε and $[x, y] = W^s_\varepsilon(x) \cap W^u_\varepsilon(y)$ is a single point in R for all points x, y in R. Define $W^s(x, R) = W^s_\varepsilon(x) \cap R$ and $W^u(x, R) = W^u_\varepsilon(x) \cap R$. For each $x \in R$ there is a homeomorphism between R and $W^s(x, R) \times W^u(x, R)$. Say R is *proper* if $\mathrm{cl}(\mathrm{int}\, R) = R$. For a proper rectangle define the *stable boundary*, $\partial^s R$, to be $\{x \in R : W^s(x, R) \cap \mathrm{int}\, R = \varnothing\}$ and the *unstable boundary*, $\partial^u R$, to be $\{x \in R : W^u(x, R) \cap \mathrm{int}\, R = \varnothing\}$. It follows that $\partial R = \partial^s R \cup \partial^u R$. The familiar picture of a rectangle is a rectangle in the plane where the vertical sides are the stable boundary and the horizontal sides are the unstable boundary. Given $x \in R$, and ε in the definition of R we may define $\partial W^s(x, R)$ and $\mathrm{int}\, W^s(x, R)$ to be the boundary and the interior of $W^s(x, R)$ as subsets of $W^s_\varepsilon(x)$. We make analogous definitions for $\partial W^u(x, R)$ and $\mathrm{int}\, W^u(x, R)$. It follows that $\partial^s R \simeq W^s(x, R) \times \partial W^u(x, R)$ and $\partial^u R \simeq \partial^s W(x, R) \times W^u(x, R)$.

When f is a positively expansive map we say a collection of subset of X, $\mathscr{P} = \{R_1, \ldots, R_n\}$ is a *Markov partition* for (X, f) if:

(1) $\mathrm{cl}(\operatorname{int} R_i) = R_i$ for each R_i,
(2) $\operatorname{int} R_i \cap \operatorname{int} R_j = \varnothing$ for $i \neq j$;
(3) the R_i's cover X;
(4) if $f(\operatorname{int} R_i) \cap \operatorname{int} R_j \neq \varnothing$ then each point in $\operatorname{int} R_j$ has one preimage in R_i.

Condition (4) is the Markov condition.

When f is an expansive homeomorphism we say a collection of subsets of X, $\mathscr{P} = \{R_1, \ldots, R_n\}$ is a Markov partition for (X, f) if:
(1) each R_i is a proper rectangle;
(2) $\operatorname{int} R_i \cap \operatorname{int} R_j = \varnothing$ for $i \neq j$;
(3) the R_i's cover X;
(4) for $x \in \operatorname{int} R_i$, $f(x) \in R_j$,

$$f(W^s(x, R_i)) \subseteq W^s(f(x), R_j), \qquad f^{-1}(W^u(f(x), R_j)) \subseteq W^u(x, R_i).$$

Condition (4) is the Markov condition, and differs from the one for noninvertible maps. For a more detailed discussion of Markov partitions see [Bo1, F, S]. It is common to require the diameter of the elements of the partition to be less than the expansive constant for the map. This is unnecessarily restrictive and would rule out both Examples 1.1 and 1.2, which we want to allow. The seeming pathology that occurs in later examples is not due to this loosening of the definition. It would also occur under the more stringent definition.

Next we will define subshifts of finite type. Let $\{1, \ldots, n\}$ be an n-point space with the discrete topology. Form the product spaces $\{1, \ldots, n\}^{\mathbb{N}}$ and $\{1, \ldots, n\}^{\mathbb{Z}}$ with the product topologies. Each has a metric compatible with the topology. Let $d(x, y) = 1/n^l$ where $l = \max\{t: x_i = y_i \text{ for } i \leq t\}$ when $x, y \in \{1, \ldots, n\}^{\mathbb{N}}$ and $l = \max\{t: x_i = y_i \text{ for } |i| \leq t\}$ when $x, y \in \{1, \ldots, n\}^{\mathbb{Z}}$. Both of these spaces are Cantor sets. There is a shift transformation, σ, defined on each by $\sigma(x)_i = x_{i+1}$. In each space let

$$\mathscr{P} = \{[1]_0, \ldots, [n]_0\} \quad \text{where } [i]_0 = \{x: x_0 = i\}.$$

The partitions defined on $\{1, \ldots, n\}^{\mathbb{N}}$ and $\{1, \ldots, n\}^{\mathbb{Z}}$ are both Markov partitions for σ. If $(i_1, \ldots, i_t) \in \{1, \ldots, n\}^t$ then the set $[i_1, \ldots, i_t]_r = \{x: x_r = i_1, x_{r+1} = i_2, \ldots, x_{r+t-1} = i_t\}$ as a subset of $\{1, \ldots, n\}^{\mathbb{N}}$ or $\{1, \ldots, n\}^{\mathbb{Z}}$ is called a *cylinder set*. In $\{1, \ldots, n\}^{\mathbb{Z}}$, it is a rectangle when $-t + 1 \leq r \leq 0$. The dynamical system $(\{1, \ldots, n\}^{\mathbb{N}}, \sigma)$ is the *one sided full n-shift*. Here, the shift is a continuous, onto, n-to-one, positively expansive map. The dynamical system $(\{1, \ldots, n\}^{\mathbb{Z}}, \sigma)$ is the *two sided full n-shift*. Here, the shift is an expansive homeomorphism. Let A be an $n \times n$ 0-1 matrix. Define $X_A \subseteq \{1, \ldots, n\}^{\mathbb{N}}$ by $X_A = \{x: A_{x_i x_{i+1}} = 1 \text{ for all } i \in \mathbb{N}\}$, and give it the subspace topology. The dynamical system (X_A, σ) is the *one sided subshift of finite type* defined by A. Define $\Sigma_A \subseteq \{1, \ldots, n\}^{\mathbb{Z}}$ by $\Sigma_A = \{x: A_{x_i x_{i+1}} = 1 \text{ for all } i \in \mathbb{Z}\}$, and give it the subspace topology. The dynamical system (Σ_A, σ) is the *two sided subshift of finite type* defined by A.

Suppose (X, f) is as before and has Markov partition $\mathscr{P} = \{R_1, \ldots, R_n\}$. Define a *transition matrix* A by

$$A(i, j) = \begin{cases} 1 & \text{if } f(\operatorname{int}(R_i)) \cap \operatorname{int}(R_j) \neq \varnothing, \\ 0 & \text{if } f(\operatorname{int}(R_i)) \cap \operatorname{int}(R_j) = \varnothing. \end{cases}$$

When f is noninvertible we say $\mathscr{P}$ is *topologically generating* if for all $x \in X_A$, $\bigcap_{i \in \mathbb{N}} f^{-i}(R_{x_i})$ is a single point. When f is a homeomorphism we say $\mathscr{P}$ is *topologically generating* if for all $x \in \Sigma_A$, $\bigcap_{i \in \mathbb{Z}} f^{-i}(R_{x_i})$ is a single point. In

the noninvertible case, when $\mathscr{P}$ is topologically generating there is a continuous, onto, map defined in the obvious way, that makes the following diagram commute.

$$\begin{array}{ccc} X_A & \xrightarrow{\sigma} & X_A \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ X & \xrightarrow{f} & X \end{array}$$

Similarly when f is a homeomorphism we have

$$\begin{array}{ccc} \Sigma_A & \xrightarrow{\sigma} & \Sigma_A \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ X & \xrightarrow{f} & X. \end{array}$$

In these cases we say that (X_A, π) and (Σ_A, π) are the *covers* for $(X, f, \mathscr{P})$. Define the *stable boundary of* $\mathscr{P}$, $\partial^s\mathscr{P} = \bigcup_{j=1}^n \partial^s R_j$, the *unstable boundary of* $\mathscr{P}$ similarly, and the *boundary of* $\mathscr{P}$ to be $\bigcup_{i=1}^n \partial R_i$. It is easy to see that in the noninvertible case π is one-to-one on $X \setminus \bigcup_{i\in\mathbb{N}} f^{-i}(\partial\mathscr{P})$ and in the invertible case π is one-to-one on $X \setminus \bigcup_{i\in\mathbb{Z}} f^{-i}(\partial\mathscr{P})$. We will see that π is always boundedly finite-to-one. Our main intent is to investigate the properties of π and $\partial\mathscr{P}$. We say a Markov partition is *metrically generating* if for every $x \in X_A$ (or $x \in \Sigma_A$), $\bigcap_{i\in\mathbb{N}} f^{-i}(\operatorname{int} R_{X_i})$ (or $\bigcap_{i\in\mathbb{Z}} f^{-i}(\operatorname{int} R_{X_i})$) is at most a single point. We will see that when $\mathscr{P}$ is metrically generating, there is still a covering map, π, from X_A (or Σ_A) to X. First, some examples.

Example 1.1. Let f be the doubling map on the circle. If we think of S^1 as $[0, 1]$ with the endpoints identified then $f(x) = 2x \pmod 1$. Let $\mathscr{P} = \{R_1, R_2\}$ where $R_1 = [0, 1/2]$, $R_2 = [1/2, 1]$. This partition is metrically but not topologically generating. This is because $\{0, 1/2\}$ is contained in both $\bigcap_{i\in\mathbb{N}} f^{-i}(R_1)$ and $\bigcap_{i\in\mathbb{N}} f^{-i}(R_2)$. Nevertheless there is a continuous map $\pi\colon \{1, 2\}^{\mathbb{N}} \to S^1$ where the points $.1^\infty$ and $.2^\infty$ (all 1's and 2's) go to 0 in S^1 and the points $.12^\infty$ and $.21^\infty$ go to $1/2$ in S^1. Every other point goes where expected.

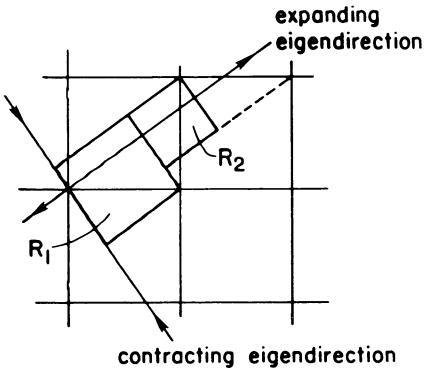

FIGURE 1

Example 1.2. Let f be the hyperbolic automorphism of the two dimensional torus defined by $\left[\begin{smallmatrix}1&1\\1&0\end{smallmatrix}\right]$ [AW]. A matrix $A \in GL(n, \mathbb{Z})$ defines an automorphism of T^n by letting A act on $\mathbb{R}^n$ and then taking the induced automorphism on $T^n \simeq \mathbb{R}^n/\mathbb{Z}^n$. Define a Markov partition on T^2 by Figure 1. Here $\mathscr{P} = \{R_1, R_2\}$ is topologically generating. The transition matrix is $\left[\begin{smallmatrix}1&1\\1&0\end{smallmatrix}\right]$ and the covering map $\pi: \Sigma_A \to T^2$ is defined as previously described.

Lemma 1.3. *Let $\mathscr{P} = \{R_1, \ldots, R_n\}$ be a metrically generating partition for (X, f) with transition matrix A. Let λ be an expansion coefficient for an adapted metric d. Then there is an $l \in \mathbb{N}$ so that:*

(i) *when f is noninvertible,* $\operatorname{diameter}(\bigcap_{i \le t} f^{-i}(\operatorname{int} R_{x_i})) < (1/\lambda)^{t-l}$ *for all $t > l$ and $x \in X_A$;*

(ii) *when f is invertible,* $\operatorname{diameter}(\bigcap_{i \le |t|} f^{-i}(\operatorname{int} R_{x_i})) < (1/\lambda)^{t-l}$ *for all $t > l$ and $x \in \Sigma_A$.*

Proof. For $x \in X_A$, $\bigcap_{i \in \mathbb{N}} f^{-i}(\operatorname{int} R_{x_i})$ is at most one point. Since X_A is compact, there exists an $l \in \mathbb{N}$ so that for all $x \in X_A$ $\operatorname{diameter}(\bigcap_{i \le l} f^{-i}(\operatorname{int} R_{x_i})) < \delta$ where the δ gives the neighborhood for local expansiveness in d. For any R_j, $\operatorname{diameter}(\operatorname{int} R_j \cap \bigcap_{i \le l} f^{-i}(\operatorname{int} R_{x_i})) < \delta/\lambda < 1/\lambda$. The assertion follows. The invertible case is similar. □

Observation 1.4. Suppose $\mathscr{P} = \{R_1, \ldots, R_n\}$ is a metrically generating Markov partition for (X, f) and A is its transition matrix. Then there is a natural covering map $\pi: (X_A, \sigma) \to (X, f)$ when f is noninvertible or from (Σ_A, σ) to (X, f) when f is invertible, with $\pi([j]_0) = R_j$, $j = 1, \ldots, n$.

Proof. For $x \in X_A$ consider $\bigcap_{i \in \mathbb{N}} f^{-i}(\operatorname{int} R_{x_i})$. It is at most a single point. If it is a single point define π by $\pi(x) = \bigcap_{i \in \mathbb{N}} f^{-i}(\operatorname{int} R_{x_i})$. As defined the range of π is $X \setminus \bigcup_{i \in \mathbb{N}} f^{-i}(\partial \mathscr{P})$. It is a dense G_δ containing all points with a dense orbit (or dense forward and backward orbit). Letting $\partial\pi = \{x \in X_A: \bigcap_{i \in \mathbb{N}} f^{-i}(R_{x_i}) \subseteq \partial \mathscr{P}\}$ we see that the domain of π is $X_A \setminus \bigcup_{i \in \mathbb{N}} \sigma^{-i}(\partial\pi)$. It is also a dense G_δ. As π is defined, $\pi \circ \sigma = \sigma \circ f$. Lemma 1.3 shows that π is Hölder continuous so it can be extended to all of X_A. It is clear that $\pi([j]_0) = R_j$ for $j = 1, \ldots, n$. A similar argument works for the invertible case. □

From now on whenever we speak of a Markov partition $\mathscr{P}$ for (X, f) and its cover (X_A, π) or (Σ_A, π) we mean that $\mathscr{P}$ is metrically generating, A is the transition matrix and π is the natural covering map. Notice that when f is not invertible π is one-to-one on all points with a dense orbit and when f is invertible π is one-to-one on all points with a dense forward and backward orbit.

Two points $x, y \in X_A$ (or Σ_A) are said to be *mutually separated* if $x_i \neq y_i$ for all $i \in \mathbb{N}$ $(i \in \mathbb{Z})$.

Theorem 1.5. *Let $\mathscr{P}$ be a Markov partition for (X, f) with transition matrix A:*

(i) *when f is not invertible (X, f) has cover (X_A, π) and*

$$\partial \mathscr{P} = \{x \in X: \pi^{-1}(x) \text{ contains a pair of mutually separated points}\};$$

(ii) *when* f *is invertible* (X, f) *has cover* (Σ_A, π) *and*

$$\bigcap_{i\in\mathbb{Z}} f^{-i}(\partial\mathscr{P}) = \{x \in X: \pi^{-1}(x) \text{ contains a pair of mutually separated points}\}.$$

Proof of (i). First we show that if $R_j \subseteq f(R_i)$ then $f: R_i \to f(R_i) \cap R_j$ is locally onto. This is immediate if the diameter R_i is small because then $f: (\operatorname{int} R_i) \to f(\operatorname{int} R_i) \cap \operatorname{int} R_j$ is a homeomorphism. If diameter of R_i is large $f(R_i)$ might "wrap around." By Lemma 1.3 for every $\varepsilon > 0$ there is an N so that for $x \in X_A$ and $m \geq N$ diameter $\bigcap_{i=0}^{m} f^{-i}(\operatorname{int} R_{x_i}) < \varepsilon$. This means $f: \bigcap_{i=0}^{m} f^{-i}(\operatorname{int} R_{x_i}) \to \bigcap_{i=1}^{m} f^{-i}(\operatorname{int} R_{x_i})$ is a homeomorphism. Push forward to see that $f: R_{x_{m-1}} \cap f^{-1}(R_{x_m}) \to R_{x_m}$ is a homeomorphism and we have the desired result.

Next we prove statement (i). Suppose $x \in R_i \cap R_{i'}$. Then $f(x)$ is also in $\partial\mathscr{P}$. This means there is an R_j so that $f(x) \in R_j$ and $f(R_i) \supseteq R_j$. Since f is onto and locally maps R_i onto R_j there must be a $R_{j'} \neq R_j$ so that $f(x) \in R_{j'}$ and $f(R_{i'}) \supset R_{j'}$. This means $\pi^{-1}(x)$ contains points y, z with $y_0 \neq z_0$ and $y_1 \neq z_1$. Then continue. Conversely, if $\pi^{-1}(x)$ contains two points whose 0th coordinates disagree, then x is in $\partial\mathscr{P}$. □

To prove Theorem 1.5(ii) we need some lemmas. We have $\mathscr{P}$ a Markov partition for (X, f) with f invertible and (Σ_A, π) its associated cover.

Lemma 1.6. *Let* $x \in \partial^s\mathscr{P}$, *and* $y, z \in \pi^{-1}(x)$. *Suppose* $\operatorname{int} W^u(x, R_{y_0}) \cap \operatorname{int} W^u(x, R_{z_0}) = \varnothing$. *Then* $y_i \neq z_i$ *for all* $i \geq 0$.

Proof. $f(\operatorname{int} W^u(x, R_{y_0})) \supseteq \operatorname{int} W^u(f(x), R_{y_1})$. This means $f(\operatorname{int} W^u(x, R_{y_0})) \cap \operatorname{int} W^u(f(x), R_{z_1}) = \varnothing$ and $f(\operatorname{int} R_{y_0}) \cap \operatorname{int} R_{z_1} = \varnothing$. □

Lemma 1.7. *Suppose* $y, z \in \pi^{-1}(x)$ *with* $y_{-1} = z_{-1}$ *and* $y_0 \neq z_0$, *then* $y_i \neq z_i$ *for all* $i \geq 0$.

Proof. If $f^{-1}(x) \in R_{y_{-1}}$, $f(\operatorname{int} R_{y_{-1}}) \cap \operatorname{int} R_{y_0} \neq \varnothing$, and $f(\operatorname{int} R_{y_{-1}}) \cap \operatorname{int} R_{z_0} \neq \varnothing$ then $x \in \partial^s R_{y_0} \cap \partial^s R_{z_0}$ and $\operatorname{int} W^u(x, R_{y_0}) \cap \operatorname{int} W^u(x, R_{z_0}) = \varnothing$. Then we can apply Lemma 1.6. □

Lemma 1.8. *If* $y, z \in \pi^{-1}(x)$ *and* $y_r = z_r$ *and* $y_s = z_s$ *for* $r < s$ *then* $y_i = z_i$ *for* $r \leq i \leq s$.

Proof. This follows immediately from Lemma 1.7. □

Remark 1.9. We say the map has no "diamonds." This was first observed by B. Marcus and gives a simple proof that π is boundedly finite-to-one. It means that cardinality of $\pi^{-1}(x)$ is no more than the cardinality of the symbol set of Σ_A squared. The rest of the proof of Theorem 1.5(ii) is just a combinatorial argument applying the fact that π has no diamonds.

Lemma 1.10. *Suppose* $y, z \in \pi^{-1}(x)$ *and* $y_0 = z_0$ *then* $[y, z] \in \pi^{-1}(x)$. *Here* $[y, z]$ *is the point with* $[y, z]_i = z_i$ *for* $i \leq 0$, $[y, z]_i = y_i$ *for* $i \geq 0$.

Proof. $\pi(\{w: w_i = y_i \text{ for } i \geq 0\}) \subseteq W^s(x, R_{y_0})$, $\pi(\{w: w_i = z_i \text{ for } i \leq 0\}) \subseteq W^u(x, R_{y_0})$. $W^s(x, R_{y_0}) \cap W^u(x, R_{y_0}) = x$. □

Proof of Theorem 1.5(ii). Assume that for each $y, z \in \pi^{-1}(x)$, $y_i = z_i$ for some i. For $y, z \in \pi^{-1}(x)$ assume by Lemma 1.10 that $y_i = z_i$ for $i \leq 0$ and $y_i \neq z_i$ for $i > 0$. There must be a $w \in \pi^{-1}(x)$ with $w_0 \neq y_0$ and $w_i \neq y_i$

for all $i \leq r$. If this were false then there would be an r so that $w_i = y_i$ for all $i \leq r$ and all $w \in \pi^{-1}(x)$ because by Remark 1.9 π is boundedly finite-to-one. Thus $\sigma^r(x) \notin \partial\mathscr{P}$. We may find $w \in \pi^{-1}(x)$ with $w_0 \neq y_0$, $w_i \neq y_i$ for all $i < r$ and $w_r = y_r$.

Suppose $r > 0$ (see Figure 2). There must be an $r' > 0$ that $w_{r'} = z_{r'}$. This forces a diamond. So $r < 0$ (see Figure 3).

Let $s < 0$ be so that $w_s = y_s$, $w_{s+1} \neq y_{s+1}$. There is a $u \in \pi^{-1}(x)$ with $u_r \neq y_r$. Since $u_j = w_j$, $u_k = y_k$, and $u_l = z_l$ for some $j, k, l \in \mathbb{Z}$ and the map has no diamonds we must have that $u_t = w_t = y_t = z_t$ for some t, $r < t \leq s$. (See Figure 4.)

Now the point v defined by $v_i = u_i$ for $i \leq t$, $v_i = w_i$ for $i > t$ is in $\pi^{-1}(x)$ by Lemma 1.10. We have a new point v that satisfies the same conditions as w but $t > r$. Continue and produce a point that is totally separated from y. The converse is clear. □

If X is a topological space and $g: Y \to Y$ is a continuous map, a point $x \in Y$ is *nonwandering* if for every neighborhood U of x there is a $k > 0$ so that $f^k(U) \cap U \neq \varnothing$. Suppose $\mathscr{P}$ is a Markov partition for (X, f). If f is not invertible then f maps $\partial\mathscr{P}$ into itself and we define the *core of* $\mathscr{P}$, $\mathscr{C}(\mathscr{P})$, to be the nonwandering points of f restricted to $\partial\mathscr{P}$. If f is invertible then $\partial\mathscr{P}$ is not invariant under f. The intersection, $\bigcap_{i\in\mathbb{Z}} f^{-i}(\partial\mathscr{P})$, is the largest invariant set contained in $\partial\mathscr{P}$. We define the *core of* $\mathscr{P}$, $\mathscr{C}(\mathscr{P})$, to be the nonwandering points of f restricted to $\bigcap_{i\in\mathbb{Z}} f^{-i}(\partial\mathscr{P})$.

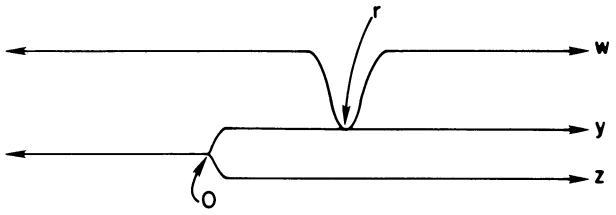

FIGURE 2

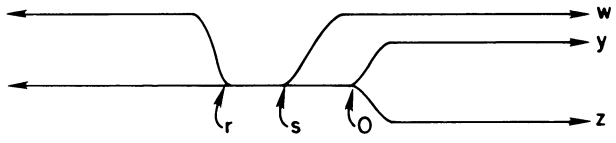

FIGURE 3

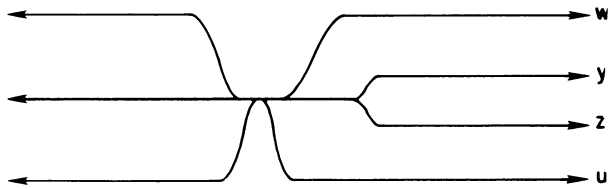

FIGURE 4

2. Symbolic dynamics

This section is about symbolic dynamics. We will make some definitions and prove some results that we will use in later sections. For a more detailed discussion of ideas about symbolic dynamics see [AM].

In §1 we defined the one sided n-shift $\{1, \dots, n\}^{\mathbb{N}}$ and the two sided n-shift $\{1, \dots, n\}^{\mathbb{Z}}$ together with their topologies, metrics, and shift transformations. Let Λ be a closed shift invariant subset of either fullshift. Put the subspace topology on it and call the dynamical system (Λ, σ) or Λ a *subshift*. Denote by $\mathscr{W}(\Lambda, n)$ the set of words of length n that occur in Λ, $\{x_0 \cdots x_{n-1} : x \in \Lambda\}$, and $\mathscr{W}(\Lambda) = \bigcup_n \mathscr{W}(\Lambda, n)$ the words that occur in Λ. For $w \in \mathscr{W}(\Lambda)$ define the *follower set* of w, $f_\Lambda(w)$ or $f(w)$, to be $\{u \in \mathscr{W}(\Lambda) : wu \in \mathscr{W}(\Lambda)\}$ and *predecessor set*, $p_\Lambda(w)$ or $p(w)$, to be $\{u \in \mathscr{W}(\Lambda) : uw \in \mathscr{W}(S)\}$. We define the k block presentation of Λ to be the subshift $\Lambda^{[k]} \subseteq (\mathscr{W}(\Lambda, k))^{\mathbb{N}}$ (or $(\mathscr{W}(\Lambda, k))^{\mathbb{Z}}$) consisting of all points $x = (x_i) = ((x_i^1 \cdots x_i^k)) \in (\mathscr{W}(\Lambda, k))^{\mathbb{N}}$ so that $x_i^2 \cdots x_i^k = x_{i+1}^1 \cdots x_{i+1}^{k-1}$ for all i and the point (x_i^1) is in Λ. This is a simple but useful renaming ((Λ, σ) and $(\Lambda^{[k]}, \sigma)$ are topologically conjugate) of Λ.

There are two types of subshifts that we will deal with. The first type is the subshifts of finite type and the second type is the sofic systems. We will generalize slightly the definition of subshifts of finite type given in §1. Start with a full n-shift, one or two sided, and let F be a finite list of forbidden blocks. Then the *subshift of finite type* defined by forbidding F is the set $\{x \in \{1, \dots, n\}^{\mathbb{N}}$: no subblock of x is in $F\}$. Suppose k is the length of the longest word in F then consider the k block presentation of Λ, define transition matrix A by

$$A(i_1 \cdots i_k, j_1 \cdots j_k) = \begin{cases} 1 & \text{if } i_{r+1} = j_r, \ r = 1, \dots, k-1, \\ 0 & \text{otherwise}, \end{cases}$$

for $i_1 \cdots i_k, \ j_1 \cdots j_k \in \mathscr{W}(\Lambda, k)$. Then $\Lambda^{[k]}$ is the subshift of finite type X_A defined by the transition matrix A. This agrees with the definition in §1. A subshift of finite type is a subshift so that for some k, the k block presentation is defined by a transition matrix. Equivalently, choose a subset $L \subseteq \{1, \dots, n\}^k$ of allowed blocks and define a subshift of finite type Λ to be $\{x \in \{1, \dots, n\}^{\mathbb{N}} : x_i \cdots x_{i+k-1} \in L \text{ all } i\}$.

The second type of subshift that interests us is the *sofic systems*. There are several equivalent definitions [W]. Say Λ is a sofic system if there are a finite number of different follower sets, $f_\Lambda(w)$, for $w \in \mathscr{W}(\Lambda)$. Or, equivalently, a finite number of predecessor sets $p_\Lambda(w)$ for $w \in \mathscr{W}(\Lambda)$. An equivalent definition is that a subshift is sofic if it is the continuous, shift commuting, image of a subshift of finite type. Each sofic system has associated to it two canonical covers (X_L, φ_L) and (X_R, φ_R) [Kr1]. This means $\varphi_L : X_L \to S$ is a continuous onto shift commuting map with some extra properties that will be discussed. A sofic system or a subshift of finite type is *irreducible* if there is a point with a dense forward orbit. A subshift of finite type is irreducible if and only if its transition matrix is irreducible. A matrix is *irreducible* if for each pair of vertices i, j there is an l so $(A^l)_{ij} > 0$. A sofic system or subshift of finite type is *nonwandering* if every point in the subshift is a nonwandering point.

FIGURE 5

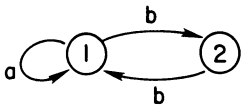

FIGURE 6

A subshift of finite type is nonwandering if and only if there is a permutation of the symbols of the subshift so that the transition matrix is a block diagonal matrix, with each of the square blocks on the diagonal irreducible.

Example 2.1. Let $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ the points of X_A (or Σ_A) are the infinite paths on the graph in Figure 5. We define a map from X_A (or Σ_A) into $\{a, b\}^{\mathbb{N}}$ (or $\{a, b\}^{\mathbb{Z}}$) by an *edge labelling* in Figure 6. This means $\varphi(x)_i$ is the label on the edge from x_i to x_{i+1}. The image under the map is the subshift consisting of all sequence where an even number of b's occur between two successive a's. This is a sofic system that is not a shift of finite type. It is usually referred to as the *even system.*

If (X, f) and (Y, g) are dynamical systems and $\varphi: X \to Y$ is a continuous onto map so that $\varphi \circ f = g \circ \varphi$ we say that φ is a factor map and that (Y, g) is a factor of (X, f). Any continuous shift commuting map between subshifts is a *block map* [H]. This means there is an n so that $\varphi(x)_i = \varphi(x_i \cdots x_{i+n})$ in the one sided case, and $\varphi(x)_i = \varphi(x_{i-n}, \ldots, x_{i+n})$ in the two sided case. The map, φ, depends on a fixed finite number of coordinates. We use φ both as a map on points and as a map on blocks. A well-known result [CP] says that a factor map, φ, between two irreducible sofic systems is either boundedly finite-to-one or some point has uncountably many preimages. For irreducible subshifts of finite type a factor map is finite-to-one if and only if the transition matrices have the same spectral radius [CP]. A factor map $\varphi: S \to T$ between sofic systems is *left-resolving* if it is a k block map and whenever $\varphi(i_0 i_1 \cdots i_k) = \varphi(i'_0 i_1 \cdots i_k)$, $i_0 = i'_0$. A factor map is *right-resolving* if it is a k block map and anytime $\varphi(i_0 \cdots i_{k-1} i_k) = \varphi(i_0 \cdots i_{k-1} i'_k)$, $i_k = i'_k$. Resolving maps are always finite-to-one and the left-resolving maps are particularly useful for one-sided shifts.

A finite-to-one factor map between two sided irreducible sofic systems has a *degree*, d. It is the smallest cardinality of the preimages of any point. Every point with a dense forward and backward orbit has exactly d preimages. Say such a map is d-to-1 a.e. A left-resolving map between one sided irreducible sofic systems also has a degree, d, and every point with a dense orbit has exactly d preimages. A d-to-1 a.e. factor map has a *magic word.* Let $\varphi: S \to T$ be a k block map between irreducible sofic systems. A word $w = w_1 \cdots w_n \in \mathscr{W}(T)$ is a *magic word* if there are d blocks $i_1^1 \cdots i_k^1, i_1^2 \cdots i_k^2, \ldots, i_1^d \cdots i_k^d \in \mathscr{W}(S, k)$ and a t so that every block $u = u_1 \cdots u_{n+k-1} \in \mathscr{W}(S, n+k-1)$ with $\varphi(u) = w$ has $u_t u_{t+1} \cdots u_{t+k-1} = i_1^r \cdots i_k^r$ for some $r = 1, \ldots, d$. If the map is left-resolving $t = 1$, if it is right-resolving $t = n - 1$.

Example 2.2. Examine Example 2.1. The map φ is a 2 block 1-to-1 a.e. factor map. It is both left- and right-resolving. The symbol a is a magic word for φ, only the block 11 maps to it. Such a symbol is a *magic symbol.*

An irreducible sofic system, as mentioned, has canonical left and right irreducible covers. $\varphi_L: X_L \to S$ is a one block, left-resolving one-to-one a.e. cover and $\varphi_R: X_R \to S$ is one block, right-resolving (one-to-one a.e. in the two sided case) cover. See [Kr1] for more information about these covers.

If $\varphi: X_A \to S$ is a factor map, then φ is a k block map for some k and by going to the k block presentation of X_A, $X_A^{[k]}$, we have a one block factor map $\varphi: X_A^{[k]} \to S$. This is often very useful. For a d-to-1 a.e. one block map between irreducible sofic systems, $\varphi: S \to T$, we say the *core of* φ, $\mathscr{C}(\varphi)$, is the set of points $\{y \in T: \varphi^{-1}(y)$ contains $d+1$ mutually separated points$\}$. A collection of points $x^1, \ldots, x^l$ is mutually separated if $x_i^r \neq x_i^s$ for any $1 \leq r \neq s \leq l$, $i \in \mathbb{N}$ (or $\mathbb{Z}$).

Observation 2.3 [F]. If $\varphi: X_A \to X_B$ (or $\Sigma_A \to \Sigma_B$) is a factor map and $\Lambda \subseteq \Sigma_B$ then Λ is a subshift of finite type if and only if $\varphi^{-1}(\Lambda)$ is a subshift of finite type.

Proof. Suppose Λ is a subshift of finite type defined by an allowed list of blocks $L = \{b_1, \ldots, b_n\}$. Then $\varphi^{-1}(\bigcup b_i)$, is open-closed and is a finite union of cylinder sets. $\varphi^{-1}(\Lambda)$ is defined by a finite list of allowed blocks.

Conversely, suppose $\varphi^{-1}(\Lambda)$ is a subshift of finite type and is defined by an allowable list of blocks $\{c_1, \ldots, c_m\}$. The complement of $\varphi(\bigcup c_j)$ is closed in X_B. For sufficiently large l, $\mathscr{W}(\Lambda, l) = \{b_1, \ldots, b_n\}$ and $\bigcup b_j \subseteq \bigcup c_i$. Λ is defined by an allowable list of blocks. □

Observation 2.4. If $\varphi: S \to T$ is a factor map between sofic systems and $\Lambda \subseteq T$, then Λ is sofic if and only if $\varphi^{-1}(\Lambda)$ is sofic.

Proof. If $\varphi^{-1}(\Lambda)$ is sofic then Λ is sofic by definition.

Suppose Λ is sofic. Assume φ is a one block map. For a word w that occurs in $\varphi^{-1}(\Lambda)$ $f_{\varphi^{-1}(\Lambda)}(w) = f_S(w) \cap \varphi^{-1}(f_\Lambda(\varphi(w)))$. Since there are finitely many distinct follower sets in S, and finitely many in Λ, there are finitely many in $\varphi^{-1}(\Lambda)$. □

Observation 2.5. Let $\varphi: X_A \to X_B$ (or $\Sigma_A \to \Sigma_B$) be a finite-to-one map. Then the core of φ, $\mathscr{C}(\varphi)$, is a sofic system.

Proof. Go to a higher block presentation of X_A so that φ is a one block map. Suppose φ has degree d. Define a new subshift of finite type, X_C, with symbol set $L_C = \{(i_1, \ldots, i_{d+1}): i_r \in L_A, i_r \neq i_s, r \neq s$, and $\varphi(i_r) = \varphi(i_s)$ all $r, s\}$. Define transitions $(i_1, \ldots, i_{d+1}) \to (j_1, \ldots, j_{d+1})$ if $i_r \to j_r$ for $r = 1, \ldots, d+1$. Define an into block map $\psi: X_C \to X_B$ by $\psi((i_1, \ldots, i_{d+1})) = \varphi(i_1)$. Then $\psi(X_C) = \mathscr{C}(\varphi)$ and $\mathscr{C}(\varphi)$ is sofic. □

Lemma 2.6. *Let X_B (or Σ_B) be an irreducible subshift of finite type and X_C (or Σ_C) a subshift of finite type contained in, but not equal to X_B (Σ_B). Then there is an irreducible subshift of finite type X_A (or Σ_A) and a one-to-one a.e. left-resolving (or right-resolving) map $\varphi: X_A \to X_B$ ($\Sigma_A \to \Sigma_B$) so that the core of φ is X_C (Σ_C).*

Proof. We first deal with the case where X_C is defined by an irreducible subgraph, G_C, of G_B with $G_B \backslash G_C$ still irreducible (any two vertices of $G_B \backslash G_C$ are connected by a path) and there are at least two edges beginning in G_C and ending outside of G_C. We will construct the graph G_A. Let $V_A = (V_B \backslash V_C) \cup (V_C \times \{1, 2\})$. Define the edges of G_A as follows. If $e \in E_B$ that begins at $i \notin V_C$ and ends at $j \notin V_C$ then put an edge in G_A beginning at i and ending at j. If $e \in E_B$ that begins at i and ends at $j \in V_C$ put two edges in G_A, $(e, 1)$ and $(e, 2)$. If $i \notin V_C$ then $(e, 1)$ and $(e, 2)$ begin at i and end at $(j, 1)$ and $(j, 2)$, respectively. If $i \in V_C$ then $(e, 1)$ begins at $(i, 1)$ and ends at $(j, 1)$ and $(e, 2)$ begins at $(i, 2)$ and ends at $(j, 2)$. Finally consider the edges that begin in V_C and end outside of it. Divide these into two nonempty disjoint sets, E_1 and E_2. For $e \in E_1$ beginning at $i \in V_C$ and ending at $j \notin V_C$ put an edge in G_A that begins at $(i, 1)$ and ends at j. Deal with edges in E_2 is the analogous way. This defines G_A. The fact that E_1 and E_2 are both nonempty means that G_A is irreducible. Define a graph homomorphism φ from G_A to G_B by sending each vertex or edge in G_A to the vertex or edge in G_B that caused it to be produced. Every vertex in $V_B \backslash V_C$ has one preimage and every vertex in V_C has two preimages under φ. Every edge ending in $V_B \backslash V_C$ has one preimage and every edge ending in V_C has two under φ. The map defined from X_A to X_B is left-resolving by construction. If $x \in X_B$ and $x_n \in V_B \backslash V_C$ then for $y, z \in \varphi^{-1}(x)$, $y_i = z_i$ for all $i \leq n$. This means φ is one-to-one a.e. and the core of φ is X_C. Example 2.7 illustrates this construction.

There are two problems left, one is that X_C may not be irreducible, the other is that X_C may not be defined by a subgraph of G_B meeting the above requirements.

When X_C is not irreducible we need to make sure that each terminal component G of G_C has at least two edges going from it to $V_B \backslash V_C$. A component G of G_C is terminal if no edge beginning in G ends at a vertex of G_C not in G. Then when we form E_1 and E_2 we make sure that each terminal component has an edge going from it to $V_B \backslash V_C$ in E_1 and in E_2. This is to insure that G_A is irreducible.

When X_C is not defined by a subgraph of G_B meeting the above requirements we go to a higher block presentation of X_B to get X_C defined by a subgraph, G_C, of G_B so that $G_B \backslash G_C$ is irreducible. If G_C has two edges beginning in G_C and ending in $V_B \backslash V_C$ we are done. If there is only one such edge we will do some symbol splits (by followers). Suppose the edge is e and it begins at $i \in V_C$ and ends at $j \in V_B \backslash V_C$ and that $f_B(j) > 1$. Partition $f_B(j)$ into two nonempty sets F_1 and F_2. Define a new graph $G_{B'}$ by leaving G_B alone except for j and its incoming and outgoing edges. Replace j by j_1 and j_2. Replace each incoming edge e by two edges, e_1 and e_2. If e is not a self-loop at j then e_1 and e_2 begin where e began, e_1 ends at j_1 and e_2 ends at j_2. If e is a self-loop then j is in say F_1. Both e_1 and e_2 will begin at j_1 and end at j_1 and j_2, respectively. For each outgoing edge, e, (not including a self-loop) ending at $k \in F_1$, replace it by a new edge beginning at j_1 and ending at k. Do the analogous thing for edges ending at elements of F_2. This is an elementary topological conjugacy. $X_{B'}$ is topologically conjugate to X_B, the graph G_C is the same in G_B and $G_{B'}$ but in $G_{B'}$ has two outgoing edges. This is an extremely important construction in symbolic dynamics (see [M] for

further discussion). If the only edge leaving G_B ends at a vertex with only one outgoing edge, we will need to do a sequence of these splittings. □

Example 2.7. Lemma 2.6 is illustrated by the graphs in Figure 7.

Lemma 2.8. *Let $S \subseteq \{1, \dots, n\}^{\mathbb{N}}$ (or $\{1, \dots, n\}^{\mathbb{Z}}$) be a sofic system, then there is a finite state automaton, $\mathscr{A}_{\bar{S}}$, with initial state I, terminal state T, so that the language accepted by $\mathscr{A}_{\bar{S}}$, $L_{\bar{S}}$, i.e. the labels of paths that begin at I and end at T is $\{w_0 \cdots w_k \in \{1, \dots, n\}^* : w_0 \cdots w_{k-1} \in \mathscr{W}(S)$ but $w_0 \cdots w_k \notin \mathscr{W}(S)\}$ (or we may choose $\mathscr{A}_{\bar{S}}$ so that $L_{\bar{S}} = \{w_0 \cdots w_k \in \{1, \dots, n\}^* : w_1 \cdots w_k \in \mathscr{W}(S)$ but $w_0 \cdots w_k \notin \mathscr{W}(S)\}$). This means $S = \{x : [x_i, \dots, x_{i+n}] \notin L_{\bar{S}} \forall i, n\}$.*

Proof. We will make this construction in a right-resolving manner because it is consistent with the usual automata notation and it is easier to follow. For one sided shifts we need left-resolving maps, that is $L_{\bar{S}} = \{w : w_1 \cdots w_k \in \mathscr{W}(S)$ but $w_0 \cdots w_k \notin \mathscr{W}(S)\}$. To translate from right- to left-resolving we use outgoing instead of incoming edges and read backwards along arrows in the finite state automaton.

Suppose G_B is a directed graph and S is obtained by an edge labelling of G_B. Define the vertices of $\mathscr{A}_{\bar{S}}$ to be subsets of V_B, letting $\phi = T$ and taking $I = V_B$. For $a \in \{1, \dots, n\}$ and each subset $U \neq \phi$ of V_B define an edge labelled a in $\mathscr{A}_{\bar{S}}$ from U to W, where W is the collection of vertices that can be reached by starting at some vertex in U and following an edge labelled a.

The vertex $T = \phi$ has incoming edges but no outgoing ones. This defines the finite state automata $\mathscr{A}_{\bar{S}}$ that recognizes all words of the form $w_0 \cdots w_k$ where $w_0 \cdots w_{k-1}$ occurs in S but $w_0 \cdots w_k$ does not. This collection of words is the regular language $L_{\bar{S}}$. Notice that if we take the subshift of finite type defined by $\mathscr{A}_{\bar{S}}$ minus T, the edge labelling defines a right-resolving map onto S. □

Example 2.9. Lemma 2.8 is illustrated by the graphs in Figure 8.

Theorem 2.10. *Let X_B (or Σ_B) be an irreducible subshift of finite type and S a sofic system contained in but not equal to X_B (or Σ_B). Then there is an irreducible subshift of finite type X_A (Σ_A) and a one-to-one a.e. left-resolving (or right) factor map $\varphi : X_A \to X_B$ ($\Sigma_A \to \Sigma_B$) so that the core of φ is S.*

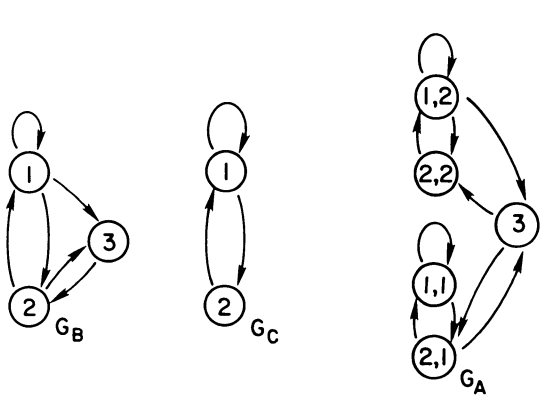

FIGURE 7

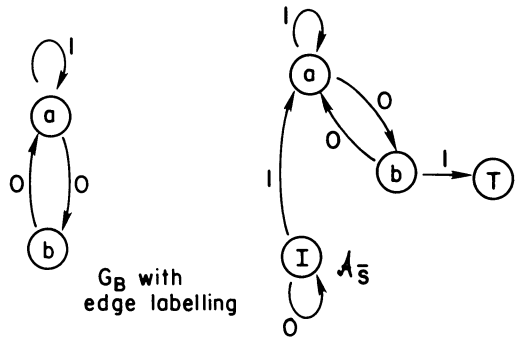

FIGURE 8

Proof. We will use Lemma 2.8 in its left-resolving form, $L_{\bar{S}} = \{w_0 \cdots w_k : w_1 \cdots w_k \in \mathscr{W}(S), w_0 \cdots w_k \notin \mathscr{W}(S)\}$ to construct a new subshift of finite type $X_{A'}$ containing another subshift of finite X_C and a one-to-one a.e. left-resolving factor mapping $\varphi' : X_{A'} \to X_B$ with $\varphi'(X_C) = S$ and the core of φ' contained in S. One important point is that, because of Observation 2.3, when S is not of finite type $(\varphi')^{-1}(S) \neq X_C$. But no periodic point in S will have a preimage outside of X_C. After this we will apply Lemma 2.7 to $X_{A'}$ and X_C to get $\psi : X_A \to X_{A'}$ with core X_C and $\varphi = \varphi' \circ \psi$ with core S.

Assume we have X_B defined by a graph G_B with $S \subseteq X_B$ and $\mathscr{A}_{\bar{S}}$ as in Lemma 2.7. Define a new graph, G, with labelled edges as follows. The vertices are pairs (a, W) where a is a symbol for X_B and W is a subset of the vertices of $\mathscr{A}_{\bar{S}}$. Say

$$(b, U) \xrightarrow{b} (a, W)$$

if a follows b in X_B and

$$U = \{i : i \xrightarrow{b} i' \text{ in } \mathscr{A}_{\bar{S}}, \text{ some } i' \in W \text{ or } i \xrightarrow{b} I \text{ in } \mathscr{A}_{\bar{S}}\}.$$

For word $w = w_0 \cdots w_k$, let $w_0 \cdots w_k(I)$ be the vertex in $\mathscr{A}_{\bar{S}}$ (if it exists) that you arrive at by starting at I and following w. (Note arrows are followed backwards in $\mathscr{A}_{\bar{S}}$.) If $w \in \mathscr{W}(S)$ then $w_0 \cdots w_k(I)$ is a vertex in $\mathscr{A}_{\bar{S}}$, but not T. If $w \in L_{\bar{S}}$, $w_0 \cdots w_k(I)$ is T. If w is in neither it is the empty set. Clearly the edge labelling of G defines a left-resolving factor map from the subshift of finite type defined by G onto X_B. If $w_0 \cdots w_k \in L_{\bar{S}}$ and (a, W) is a vertex in G with $w_0 \cdots w_k$ leading into it, the path must begin at the state (w_0, U_w) where $U_w = \{(w_0(I), w_0 w_1(I), \ldots, w_0 \cdots w_{n-1}(I), T\})$ (throw any repeated symbols out of the list). This means every $w_0 \cdots w_k \in L_{\bar{S}}$ is a magic word for the map defined by the edge labelling. So the core of φ' is contained in S. φ' is left-resolving and one-to-one a.e. There is a unique component of $X_{A'}$ of maximal entropy and every other component leads in backward time into it. Call this component $G_{A'}$. Let $X_{A'}$ be the subshift of finite type defined by $G_{A'}$, and φ' the map onto X_B defined by the edge labelling.

Consider the subgraph G_C of $G_{A'}$ obtained by throwing out any vertex (a, U) where $T \in U$. Let X_C be defined by G_C. Let $x \in S$ be a periodic point and (a, U) be any vertex in $G_{A'}$ that occurs in a cycle that maps onto x. The subset U cannot contain T. This means no periodic point in S has a preimage outside of X_C and $\varphi'(X_C) = S$.

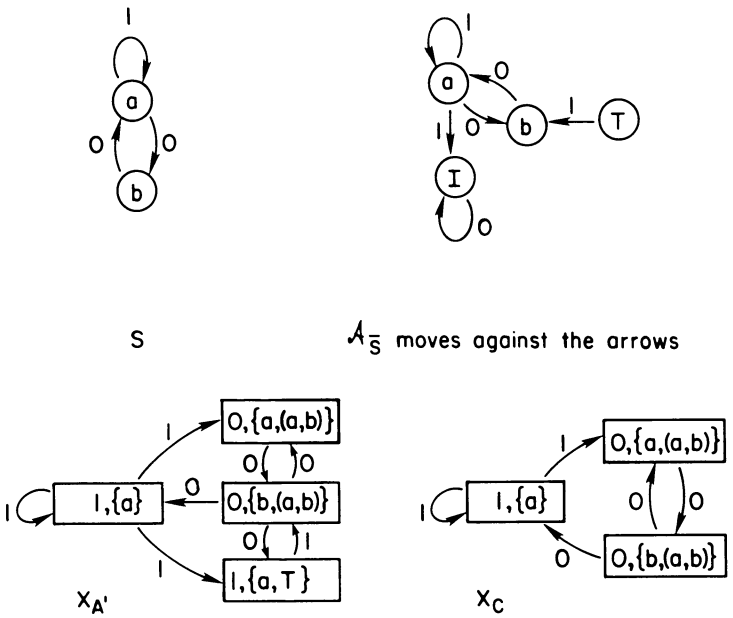

FIGURE 9

Finally, we apply Lemma 2.7 to $X_C \subseteq X_{A'}$ to get X_A and $\psi: X_A \to X_{A'}$ left-resolving one block map and with the property that every (a, U) where U contains T, that is every symbol not in X_C, has one preimage under ψ. We need to see that the core of the composition, φ, is S. By construction $\mathscr{C}(\psi) = X_C$ and $\varphi'(X_C) = S$. So $\mathscr{C}(\varphi) \supseteq S$. Suppose $w_0 \cdots w_k \in L_{\bar{S}}$. Then every path in $G_{A'}$ labelled by $w_0 \cdots w_k$ begins at (w_0, U_w) as described earlier. The subset U_w contains T so (w_0, U_w) has one inverse image under φ' and so $w_0 \cdots w_k$ is a magic word for φ. This means $\mathscr{C}(\varphi) = S$. □

Notice that constructing a factor map with S as a core is the same as finding a different Markov partition on X_B with S as the largest closed invariant set contained in the boundary.

Example 2.11. Theorem 2.10 is illustrated in Figure 9. It is from Example 2.9 but here $\mathscr{A}_{\bar{S}}$ is left-resolving. $S \subseteq \{0, 1\}^{\mathbb{N}} = \Sigma_B$.

Observation 2.12. Suppose S is an irreducible two sided sofic system (Σ_L, φ_L) and (Σ_R, φ_R) are its irreducible left and right covers. Then the core of the two maps is the same and if (Σ_A, φ) is any other cover the core of φ_R is contained in the core of φ.

Proof. Suppose $x \in S$ and x is not in the core of φ_R. This means there is an n so that $x_{-n} \cdots x_n \in \mathscr{W}(S)$ and $f(x_{-n} \cdots x_n) = f(wx_{-n} \cdots x_n)$ for all w such that $wx_{-n} \cdots x_n \in \mathscr{W}(S)$. This means $x_{-n} \cdots x_n$ is a magic word for φ_R. Consequently, $p(x_{-n} \cdots x_n) = p(x_{-n} \cdots x_n u)$ for all u so that $x_{-n} \cdots x_n u \in \mathscr{W}(S)$. The block $x_{-n} \cdots x_n$ is also a magic word for φ_L and x is not in the core of φ_L. By symmetry we see that the two cores are the same.

Suppose (Σ_A, φ) is a one block d-to-1 a.e. cover for S and $x \in S$ but not in the core of φ. As before there is an n so that $x_{-n} \cdots x_n$ is a magic word for φ. This means there is a t, $-n \leq t \leq n$, and $i^1, \ldots, i^d$ so that if $y_1 \cdots y_n \in \mathscr{W}(\Sigma_A)$ with $\varphi(y_1 \cdots y_n) = x_1 \cdots x_n$ then $y_t = i^r$ some $1 \leq r \leq d$. Observe that $\varphi(f_A(i^1)) = \cdots = \varphi(f_A(i^d))$, and $\varphi(p_A(i^1)) = \cdots = \varphi(p_A(i^d))$. If this were not true φ would fail to be d-to-1 somewhere. This means that $f_s(x_{-n} \cdots x_n) = f(ux_{-n} \cdots x_n)$ for all $u \in \mathscr{W}(S)$ with $ux_{-n} \cdots x_n \in \mathscr{W}(S)$. The block $x_{-n} \cdots x_n$ is a magic word for φ_R and x is not in the core of φ_R. □

In view of this observation the *core of* S is well defined. It is defined independently of the presentation of S and is an invariant of topological conjugacy.

3. Sofic systems and boundaries

An expansive factor of a subshift of finite type is called *finitely presented.* See [F] for a discussion of the dynamical properties of these systems. A totally disconnected (or 0-dimensional) finitely presented system is a sofic system.

Observation 3.1. Suppose (X, f) is a compact dynamical system and $\mathscr{P}$ is a Markov partition, then:

(i) if f is not invertible $(\partial\mathscr{P}, f)$ is a finitely presented system;

(ii) if f is invertible $(\bigcap_{n\in\mathbb{Z}} f^{-n}(\partial\mathscr{P}), f)$ is a finitely presented system.

Proof. Let A be the transition matrix for $\mathscr{P}$ and (X_A, π) or (Σ_A, π) be the cover. By Theorem 1.5 we have that

$$\partial\mathscr{P} = \{x\colon \pi^{-1}(x) \text{ contains a pair of mutually separated points}\}$$

in the noninvertible case and

$$\left(\bigcap_{n\in\mathbb{Z}} f^{-n}(\partial\mathscr{P})\right) = \{x\colon \pi^{-1}(x) \text{ contains a pair of mutually separated points}\}$$

in the invertible case. Now make the fibered product of X_A or Σ_A with itself over X, $X_{(A,A)} = \{(x, y) \in X_A \times X_A\colon \pi(x) = \pi(y)\}$. $(X_{(A,A)}, \sigma)$ is a subshift of finite type and the projection map $\pi_1\colon X_{(A,A)} \to X_A$ is a one-to-one a.e. factor map. Moreover, $\mathscr{C}(\pi_1) = \pi^{-1}(\partial\mathscr{P})$ and by Observation 2.5 is a sofic system. This means $(\partial\mathscr{P}, f)$ is an expansive image of a subshift of finite type. The same holds in the invertible case. □

We will look at two basic examples.

Construction 3.2. The *n-fold covering map* from S^1 to itself is $f_n(x) = nx \pmod 1$ where S^1 is $[0, 1]$ with the endpoints identified. Certain Markov partitions are very easy to construct for f_n. A finite set of points $0 \leq x_0 < x_1 < \cdots < x_k < 1$ partitions the circle into $k + 1$ intervals. If this finite set is invariant under f_n and the partition metrically generates then we have constructed a Markov partition for f_n. Any periodic point $z^0 \in S^1$ defines a Markov partition for f_n in the following way. Let $f_n(z^i) = z^{i+1}$ so the orbit is $z^0, \ldots, z^{p-1}$. This set is invariant and if the partition metrically generates we have a Markov partition. If this partition does not generate we add to the collection of points, the points $\{y^0, \ldots, y^{n-1}\} = f_n^{-1}(z^0)$. Assume $z^{p-1} = y_0$, then the collection of points $\{z^0, z^1, \ldots, z^{p-1}, y^1, \ldots, y^{n-1}\}$

is invariant and the partition it defines metrically generates. Here $\partial\mathscr{P} = \{z^0, \ldots, z^{p-1}, y^1, \ldots, y^{n-1}\}$ and $\mathscr{C}(\mathscr{P}) = f_n(\partial\mathscr{P}) = \{z^0, \ldots, z^{p-1}\}$. As an example of this take $z^0 = 0$ the fixed point for f_2. Then $y^1 = 1/2$ and the two points $\{0, 1/2\}$ define a Markov partition.

Construction 3.3. A *hyperbolic automorphism of* T^2, f_θ, is defined by a matrix $\theta \in GL(z, \mathbb{Z})$ with no eigenvalues of modulus one, see Example 1.2. A construction of Bowen [Bo2] allows us to construct some simple Markov partitions for these. The matrix θ has two real eigenvalues, λ with modulus greater than one and μ with modulus less than one. Let v_λ and v_μ be eigenvectors for λ and μ, respectively, see Figure 10.

The line, l_λ, passing through the origin containing v_λ when projected onto T^2 is $W^u(0)$ and the line, l_μ, passing through the origin containing v_μ when projected onto the torus is $W^s(0)$. Draw fairly long sections of l_λ and l_μ centered at the origin and project them onto T^2. Finally extend l_λ in each direction until each end intersects l_μ. Do the same for l_μ to end at intersections with l_λ. These two line segments partition the torus into a finite collection of connected rectangles, $\mathscr{P}$. Since the $\partial^u\mathscr{P}$ is a section of l_λ roughly centered at the origin, $f_\theta^{-1}(\partial^u\mathscr{P}) \subseteq \partial^u\mathscr{P}$. Similarly for $\partial^s\mathscr{P}$, $f_\theta(\partial^s\mathscr{P}) \subseteq \partial^s\mathscr{P}$. This guarantees the Markov condition. If the rectangles are small enough, i.e. if $\partial^u\mathscr{P}$ and $\partial^s\mathscr{P}$ long enough, $\mathscr{P}$ will generate metrically. This produces a Markov partition $\mathscr{P}$. Notice $\bigcap f_\theta^{-i}(\partial\mathscr{P}) = \mathscr{C}(\mathscr{P}) = \{0\}$, a single fixed point. We could equally well make this construction by taking $z^0 \in T^2$ periodic with orbit $z^0, \ldots, z^{p-1}$. Make the exact same construction using line segments l_λ^0, l_μ^0 through z^0 and parallel to v_λ, v_μ. Then take the partition of T^2 defined by the line segments $l_\lambda^0, f_\theta(l_\lambda^0), \ldots, f_\theta^{p-1}(l_\lambda^0)$ and $l_\mu^0, f_\theta(l_\mu^0), \ldots, f_\theta^{p-1}(l_\mu^0)$. This will again produce a Markov partition whose boundary is the union of these line segments and with $\bigcap f_\theta^{-i}(\partial\mathscr{P}) = \mathscr{C}(\mathscr{P}) = \{z^0, \ldots, z^{p-1}\}$.

These two constructions produce partitions with finite cores. We make the following observation and then devote the rest of the section to seeing exactly which sofic systems can occur in these settings.

Observation 3.4. (i) Suppose $\mathscr{P}$ is a Markov partition for f_n the n-fold covering map of S^1. Then $(\partial\mathscr{P}, f_n)$ and $(\mathscr{C}(\mathscr{P}), f_n)$ are sofic systems.

(ii) Suppose $\mathscr{P}$ is a Markov partition for a hyperbolic toral automorphism f of T^2. Then $(\bigcap f^{-i}(\mathscr{P}), f)$ and $(\mathscr{C}(\mathscr{P}), f)$ are sofic systems.

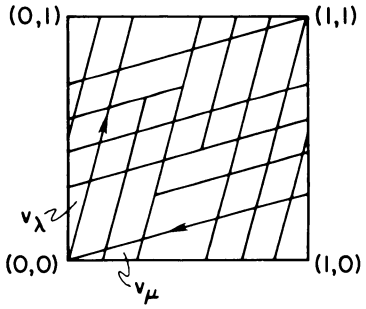

FIGURE 10

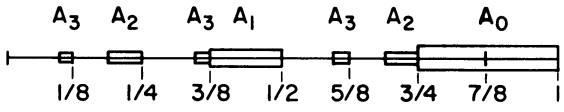

FIGURE 11

Example 3.5. This example is due to R. F. Williams and was the motivation for this work. We want to get the one sided *golden mean* subshift of finite type as the boundary of a Markov partition for f_2 on S^1. It is illustrated by Figure 11.

Let $A_0 = [3/4, 1]$ then $A_1 = f_2^{-1}(A_0)\backslash A_0$, and $A_n = f_2^{-1}(A_{n-1})\backslash(\bigcup_{i<n} A_i)$. Then let

$$P_0 = \mathrm{cl}\left(\left[0, \frac{1}{2}\right] \cap \bigcup_{i \text{ even}} A_i\right), \qquad P_2 = \mathrm{cl}\left(\left[\frac{1}{2}, 1\right] \cap \bigcup_{i \text{ even}} A_i\right),$$

$$P_1 = \mathrm{cl}\left(\left[0, \frac{1}{2}\right] \cap \bigcup_{i \text{ odd}} A_i\right), \qquad P_3 = \mathrm{cl}\left(\left[\frac{1}{2}, 1\right] \cap \bigcup_{i \text{ odd}} A_i\right).$$

The transition matrix, A, for $\mathscr{P}$ is

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Let $X_{[2]}$ be the full two shift and define $\varphi: X_A \to X_{[2]}$ by a one block map $\varphi(0) = \varphi(2) = 0$, $\varphi(1) = \varphi(3) = 1$. Observe that φ is left resolving one-to-one a.e. and the core of φ is the golden mean subshift of finite type. Taking $\pi_{[2]}: X_{[2]} \to S^1$ in the usual way and $\pi = \pi_{[2]} \circ \varphi: X_A \to S^1$ we see that $(\partial\mathscr{P}, f_2) = (\mathscr{C}(\mathscr{P}), f_2)$ is topologically conjugate to the golden mean. Intuitively, think of giving points in S^1 their diadic expansion. Then A_0 contains all points whose expansion begins with two ones. A_1 contains all points not in A_0 that have two ones beginning in the second entry and so forth. The boundary then consists of all points one of whose expansions does not contain a pair of adjacent ones. This is an example of a way to build a partition with a specified boundary. You get a subshift of finite type X_A and a factor map $\varphi: X_A \to X_{[2]}$ with the desired core. Make sure the core sits in $X_{[2]}$ so that $\pi_{[2]}: X_{[2]} \to S^1$ will inject the core of φ into S^1. Then the partition is the image of the cylinder sets under $\pi_{[2]} \circ \varphi$ and the boundary is as desired.

Example 3.6. Here we construct the even sofic system of Example 2.1 as the boundary of a Markov partition for f_2 on S^1. The construction is similar to Example 3.5. Give the points in the interval their diadic expansion. Let

$$A_0 = [.101, .11] \cup [.10001, .1001] \cup \cdots$$

a countable collection of intervals containing all points whose expansion begins $.10^n1\cdots$ for odd n (see Figure 12). Let $R_e = f_2(A_0) \cap [0, 1/2]$ and $R_o = f_2(A_0)^c \cap [0, 1/2]$. As before let $A_n = f_2^{-1}(A_{n-1}) \setminus \bigcup_{i<n} A_i$. Then take

$$P_0 = \mathrm{cl}\left(\bigcup_{i \text{ even}} A_i \cap R_e\right), \qquad P_1 = \mathrm{cl}\left(\bigcup_{i \text{ odd}} A_i \cap R_e\right),$$

$$P_2 = \mathrm{cl}\left(\bigcup_{i \text{ even}} A_i \cap R_o\right), \qquad P_3 = \mathrm{cl}\left(\bigcup_{i \text{ odd}} A_i \cap R_o\right),$$

$$P_4 = \mathrm{cl}\left(\bigcup_{i \text{ even}} A_i \cap \left[\frac{1}{2}, 1\right]\right), \qquad P_5 = \mathrm{cl}\left(\bigcup_{i \text{ odd}} A_i \cap \left[\frac{1}{2}, 1\right]\right).$$

Take A to be the transition matrix, define $\varphi\colon X_A \to X_{[2]}$ by $\varphi(0) = \varphi(1) = \varphi(2) = \varphi(3) = 0$, $\varphi(4) = \varphi(5) = 1$ and observe that the core of φ is the even system and that it injects into S^1 by $\pi_{[2]}$.

Observation 3.7. Suppose X, Y are compact metric spaces, f and g are positively expansive (or expansive) maps of X and Y, respectively, onto themselves. Suppose $\mathscr{P}$ is a metrically generating Markov partition for f on X and $\varphi\colon X \to Y$ is a factor map that is one-to-one on the transitive (doubly transitive if f is invertible) points of f. Then $\varphi(\mathscr{P})$ is a metrically generating Markov partition for g on Y.

Proof. We check the conditions needed for a Markov partition. Let $\mathscr{P} = \{P_1, \ldots, P_k\}$ and $\mathscr{Q} = \varphi(\mathscr{P})$ with $Q_i = \varphi(P_i)$. First, Q_i is closed because P_i is compact. If $U \subseteq X$ is a nonempty open set the $\varphi(U) \supseteq \varphi(U^c)^c$ which is nonempty and open so $\varphi(U)$ has nonempty interior. This means that $\mathrm{cl}(\mathrm{int}\, Q_i) = Q_i$. We have $\mathrm{int}\, Q_i \cap \mathrm{int}\, Q_j = \phi$ for $i \neq j$ otherwise φ would not be one-to-one on the points with dense orbits. The Q_i cover X.

Suppose f and g are not invertible. We need only to see that the Markov condition holds. If $g(\mathrm{int}\, Q_i) \cap \mathrm{int}\, Q_j$ and $y \in \mathrm{int}\, Q_i$ then there is a unique $x \in Q_i$ with $g(x) = y$. There must be at least one x since P_i satisfies this condition in X. Suppose there is a $y \in \mathrm{int}(Q_j)$ and $x^1, x^2 \in Q_i$ with $g(x^1) = g(x^2) = y$. Then there is a z close to y but with a dense orbit that has this property. That is clearly impossible.

Suppose f and g are invertible. Observe that since φ is a factor map the image of a rectangle must be a rectangle so the Q_i's are rectangles. We need to see that the Markov condition holds. This follows immediately from the fact that it holds in X and φ is a factor map. Since $\varphi^{-1}(\mathrm{int}\, Q_i) \subseteq P_i$, $\mathscr{Q}$ metrically generates. □

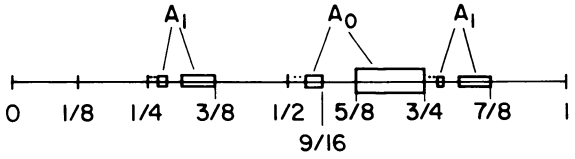

FIGURE 12

Lemma 3.8. *Suppose $X \subseteq S^1$ is a closed subset, invariant under f_n. Further, suppose (X, f_n) is topologically conjugate to a nonwandering sofic system (S, σ). Then there is a Markov partition $\mathscr{P}$ for f_n so that the core of the boundary is X.*

Proof. Pick a periodic point $z^0 \in X$. Use Construction 3.2 to produce a connected Markov partition $\mathscr{C}$ with standard cover (X_B, π_B). As noted there, $\partial\mathscr{C} = \{z^0, z^1, \ldots, z^{p-1}, y^1, \ldots, y^{n-1}\}$. Consider the subshift $\pi_B^{-1}(X) = T \subseteq X_B$. This is sofic, by reasoning similar to Observation 2.4, but is not conjugate to S. Each point in $\partial\mathscr{C}$ has two mutually separated preimages under π_B. Let $.(i_0 i_1 \cdots i_{p-1})^\infty$ and $.(i'_0 \cdots i'_{p-1})^\infty$ be the two points that map to z^0, $.(i_r \cdots i_{p-1} i_0 \cdots i_{r-1})^\infty$ and $.(i'_r \cdots i'_{p-1} i'_0 \cdots i_{r-1})^\infty$ the points that map to z^r, $0 \le r \le p-1$, and $.j_s(i_0 \cdots i_{p-1})^\infty$ and $.j'_s(i'_0 \cdots i'_{p-1})^\infty$ the points that map to y^s for $0 \le s \le n-1$. The other points with two preimages under π_B are points that eventually map to z^0 by f_n. Such a point has π_B preimages $.w j_s(i_0 \cdots i_{p-1})^\infty$ and $.w j'_s(i'_0 \cdots i_{p-1})^\infty$ or $.w(i_r \cdots i_{p-1} i_0 \cdots i_{r-1})^\infty$ and $.w(i_r \cdots i_{p-1} i_0 \cdots i_{r-1})^\infty$ for any $w \in \mathscr{W}(X_B)$ so that the preimages are allowable. All other points have single preimages under π_B.

Now apply Theorem 2.10 to $T \subseteq \Sigma_B$ to construct $\varphi: X_A \to X_B$ left resolving, one-to-one a.e. with $\mathscr{C}(\varphi) = T$. Let $\pi = \pi_B \circ \varphi$ and apply Observation 3.5 to π to get a Markov partition $\mathscr{P}$ for f_n. We want to see that the core of $\mathscr{P}$ is X. By Theorem 1.5 $\partial\mathscr{P} = \{x \in S^1 : \pi^{-1}(x)$ contains two mutually separated points$\}$. By construciton $X \subseteq \partial\mathscr{P}$ and since X is nonwandering $X \subseteq \mathscr{C}(\mathscr{P})$. To see the other containment we need to be more careful. The problem could arise from $x \in S^1$, $x \notin \partial\mathscr{C}$, but x having two preimages under π_B. See Figure 13.

First observe that if $w \notin \mathscr{W}(T)$ then x does not have mutually separated preimages under π. The problem could be that $w \in \mathscr{W}(T)$ but for some k, $w j_s(i_0 \cdots i_{p-1})^k, w j'_s(i'_0 \cdots i'_p)^k \notin \mathscr{W}(T)$. Then x could possibly have mutually separated preimages under π. Such a point $x \in \partial\mathscr{P}$ is wandering in $\partial\mathscr{P}$. There are only countably many such points. They accumulate only on points in X, and are eventually mapped onto z^0 by f_n. It means that $\partial\mathscr{P}$ may be larger than $X \cup \{y^1, \ldots, y^{n-1}\}$, but nevertheless $\mathscr{C}(\mathscr{P}) = X$. □

Lemma 3.9. *Suppose $X \subseteq T^2$ is a closed subset, invariant under a hyperbolic automorphism f_θ. Further suppose (X, f_θ) is topologically conjugate to a nonwandering sofic system (S, σ). Then there is a Markov partition $\mathscr{P}$ for f so that the core of the partition is X.*

Proof. The proof of this is very similar to the proof of Lemma 3.8. Pick a periodic point $z^0 \in X$ and use Construction 3.3 to get a connected Markov partition $\mathscr{C}$ with standard cover (Σ_B, π_B).

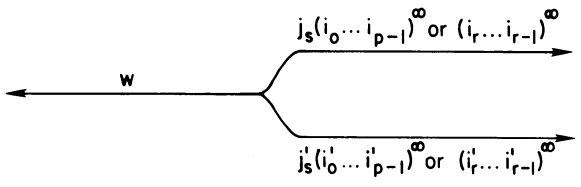

FIGURE 13

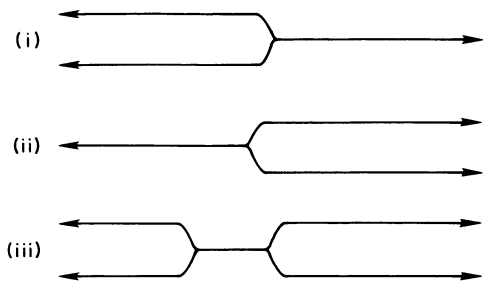

FIGURE 14

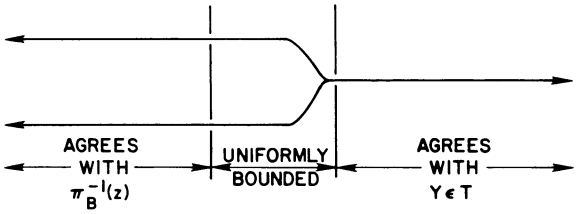

FIGURE 15

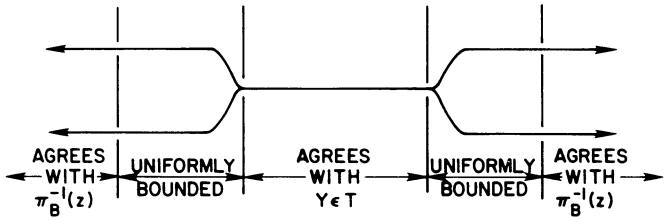

FIGURE 16

As in the previous proof, let $T = \pi_B^{-1}(X) \subseteq \Sigma_B$. Apply Theorem 2.10 to $T \subseteq \Sigma_B$ to construct $\varphi: \Sigma_A \to \Sigma_B$ left-resolving, one-to-one a.e. with $\mathscr{C}(\varphi) = T$. Let $\pi = \pi_B \circ \varphi$ and $\mathscr{P}$ be the Markov partition defined by Observation 3.5. We want to see the core is X. Again by Theorem 1.5 we see that X is in the core of $\mathscr{P}$. The problem is again the other containment. We must examine how points not in X may have mutually separated preimages. Such a point $x \in T^2$ must have more than one preimage under π_B. These preimages cannot be mutually separated. They could be one of three types which we represent schematically in Figure 14.

We will examine each type in turn. In the first type there are π_B preimages as represented but mutually separated π preimages. If any word $y_k \cdots y_{k+l} \notin \mathscr{W}(T)$ then the π preimages are not mutually separated. We have Figure 15.

Considered as points inside $\pi_B^{-1}(\bigcap f^{-i}(\partial\mathscr{P})) \subseteq \Sigma_B$ they are wandering and accumulate only on T. In case (ii) the reasoning is exactly the same. In case (iii) we see that the following holds. (See Figure 16.) The same reasoning holds and we see that the nonwandering set of $\pi_B^{-1}(\bigcap f^{-i}(\partial\mathscr{P}))$ is contained in T and X is the core of $\mathscr{P}$. □

Now we prove the main theorem of this section. Notice that the statement in the invertible case is stronger than in the noninvertible case. This is because

in the invertible (two sided) case we have Krieger's embedding theorem [Kr2]: A sofic system S is topologically conjugate to a proper subshift of a subshift of finite type Σ_A if and only if $h(S) < h(\Sigma_A)$ and the number of periodic points of each period in S does not exceed the number for the same period in Σ_A. In the noninvertible (one sided) case there is no corresponding theorem. We do not have a guess as to what the correct necessary and sufficient conditions are for one subshift of finite type to be conjugate to a proper subshift of another subshift of finite type. This seems to be a difficult question.

Theorem 3.10. (i) *A one sided subshift is conjugate to the core of a Markov partition for some f_n, the n-fold covering map of S^1, if and only if it is a nonwandering sofic system.*

(ii) *A two sided subshift is conjugate to the core of a Markov partition for f_θ, the hyperbolic automorphism of T^2 defined by θ, if and only if it is a nonwandering sofic system with $h(S) < h(f_\theta)$ and the number of periodic points of each period in S does not exceed the number in f_θ.*

Proof. (i) The core of a Markov partition must be a nonwandering sofic system. That is Observation 4.1(i). Fix n, then $0 < 1/n < 2/n < \cdots < n-1/n < 1$ defines a connected Markov partition for f_n by Construction 3.2. This partition generates metrically and the full n shift is the cover. The covering map identifies two fixed points, $.1^{\mathbb{N}}$ and $.n^{\mathbb{N}}$, preimages of the fixed point 0, and no other periodic points. If S is a transitive sofic system, then for all sufficiently large n, S sits inside the full n shift in such a way that it does not contain the fixed points $.1^{\mathbb{N}}$ and $.n^{\mathbb{N}}$. This means the covering map onto S^1 injects S into S^1. The map is a homeomorphism since S is compact. Now apply Lemma 3.8 to the image of S.

(ii) By Lemma 4.1(ii) the core of a Markov partition must be a nonwandering sofic system. Now fix $\theta \in GL(2, \mathbb{Z})$ and suppose S is a nonwandering sofic system with $h(S) < h(f_\theta)$ and the condition on the periodic points. Since $h(S) < h(f_\theta)$ there will be a k so that the number of points of period k for S is less than the number for f_θ. Pick such a k and let z^0 be a point of period k for f_θ in T^2. Apply Construction 3.3 to the orbit of z^0 to construct a Markov partition $\mathscr{P}$ for f_θ on T^2 with cover (Σ_A, σ).

We have that $h(\Sigma_A) = h(f_\theta) > h(S)$ and the number of periodic points of each period for Σ_A and f_θ is the same with the exception of the orbits the map onto the orbit of z^0. By the choice of z^0, S meets the hypotheses of Krieger's embedding theorem for Σ_A and it can be embedded so that $\pi^{-1}(z^0) \cap S = \varnothing$. This means π maps S into T^2 injectively and since S is compact π restricted to S is a homeomorphism. Finally we apply Lemma 3.9 to the image of S in T^2 to obtain the desired partition. □

4

Observation 4.1. A 0-1 matrix is a transition matrix for a Markov partition for f_n, the n-fold covering map on S^1 if and only if it is aperiodic and has column sum n.

Proof. It is easy to see that these conditions are necessary. It follows from a theorem in [AGW] that these conditions on a transition matrix, A, allow

the construction of a left resolving one-to-one a.e. factor map $\varphi\colon X_A \to X_{[n]}$, where $X_{[n]}$ is the full n-shift. Compose φ with the standard covering map $\pi_{[n]}\colon X_{[n]} \to S^1$ to get a one-to-one a.e. factor map $\pi\colon X_A \to S^1$. Now apply Observation 3.7 to get a partition with (X_A, π) as its cover. □

Remark 4.2. The transition matrix alone tells nothing about the core of the Markov partition. It is easy to construct a left-resolving one-to-one a.e. factor map $\varphi\colon X_{[2]} \to X_{[2]}$ with a nonfinite sofic core. This, in turn, produces a Markov partition for f_2 on S^1 with two elements and $\left[\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right]$ as a transition matrix with a nonfinite sofic core.

We have produced complicated Markov partitions for f_n on S^1. However, we have produced these by constructing complicated factor maps between subshifts of finite type and then pushing the Markov partition on the domain shift down onto S^1. In these cases the image shift came from a very simple, connected partition on S^1. This means that the complicated partition came from chopping up a simple one. We will show that some partitions do not arise in this way, they are inherently complicated.

Let $\mathscr{P}, \mathscr{Q}$ be Markov partitions for f_n on S^1 with covers (X_A, π_A) and (X_B, π_B), respectively. We say $\mathscr{P}$ *factors through* $\mathscr{Q}$ if there is a factor map $\varphi\colon X_A \to X_B$ so that $\pi_A = \pi_B \circ \varphi$. The examples we have constructed all factor through some connected partition. We will produce a partition that does not factor through any connected partition.

Lemma 4.3. *Let $\mathscr{P}, \mathscr{Q}$ be Markov partitions for f_n on S^1, then $\mathscr{P}$ factors through $\mathscr{Q}$ if and only if $\bigvee_{i=0}^{N} f^{-i}(\mathscr{P})$ refines $\mathscr{Q}$ for some N.*

Proof. A factor map $\varphi\colon X_A \to X_B$ is an N block map for some N. □

Let $\mathscr{P} = \{P_1, \ldots, P_k\}$ with cover (X_A, π_A) be a metrically generating Markov partition for f_n on S^1. For each P_j we will define *extreme points* (ζ_j, η_j), and the *interval of* P_j, with the property that $P_j \subseteq [\zeta_j, \eta_j]$ in a natural way. If the diameter of P_j is small it is obvious how to do this. Choose the ζ_j to the left hand (counter clockwise) endpoint and η_j to be the right hand (clockwise) endpoint of the smallest closed interval containing P_j. In general, we know by Lemma 1.3 that there is an $l \in \mathbb{N}$ so that diameter $(\bigcap_{i\le t} f^{-i}(P_{x_i})) < (1/n)^{t-l}$ for all $t > l$. For large t and $x \in X_A$ with $x_t = j$ we define the extreme points of $\bigcap_{i\le t} f^{-i}(P_{x_i})$ as above. There are n^t words, $[x_0, \ldots, x_t]$ in $\mathscr{W}(X_A, t+1)$ with $x_t = j$. If x and y are two such words there is a rotation of the circle that takes $\bigcap_{i\le t} f^{-i}(P_{x_i})$ to $\bigcap_{i\le t} f^{-i}(P_{y_i})$. Moreover the smallest closed intervals containing each of them are disjoint. If not, $\mathscr{P}$ would not be metrically generating. This means that diameter $\bigcap_{i\le t} f^{-i}(P_{x_i}) < (1/n)^t$. For $x, y \in X_A$ with $x_t = j$ and $f_n^s(y) = x$, f^s takes $\bigcap_{i\le t+s} f^{-i}(P_{y_i})$ to $\bigcap_{i\le t} f^{-i}(P_{x_i})$ by a homeomorphism. Define the extreme points of P_j and the interval of P_j to be the image of the extreme points and the image of the interval of $\bigcap_{i\le t} f^{-i}(P_{x_i})$, for any large t and x with $x_t = j$. Note that the interval of P_j is not all of S^1. We say the collection of extreme points and intervals are the extreme points and intervals for $\mathscr{P}$. The map f_n maps extreme points to extreme points and intervals completely over intervals. Each extreme point is either periodic or preperiodic.

Example 4.4. The partition in Example 3.5 has intervals $\{[0, 1/3], [0, 1/2], [1/2, 2/3], [1/2, 1]\}$.

Lemma 4.5. *Suppose $\mathscr{P}$ is a Markov partition for f_n on S^1 and every extreme point of $\mathscr{P}$ is contained in the interval of some P_j, where it is not an extreme point. Then $\mathscr{P}$ does not factor through any connected partition.*

Proof. First observe that if for every $x \in S^1$ and every $N \in \mathbb{N}$ there is $y \in X_A$ so that x is contained in the interval of $\bigcap_{i \le N} f_n^{-i}(P_{y_i})$ and is not an extreme point then $\mathscr{P}$ does not factor through any connected partition. This follows because if Q is the connected partition and x is a boundary point then for every N there is a $y \in X_A$ so that $\bigcap_{i \le N} f_n^{-i}(P_{y_i})$ intersects both intervals of Q that are separated by x. This means $\bigvee_{i \le N} f_n^{-i}(\mathscr{P})$ never refines Q.

Now let $y \in X_A$ and x an extreme point of $\bigcap_{i \le N} f_n^{-i}(P_{y_i})$. By the hypothesis there is a P_j whose interval contains $f_n^N(x)$ and does not have it as an extreme point. This means there is a $z \in X_A$ so that $z_n = P_j$, x is in the interval of $\bigcap_{i \le N} f_n^{-i}(P_{z_i})$ and is not an extreme point. Now we apply the observation. □

This says that if the extreme points of the partition are interleaved it cannot factor through a connected partition.

Theorem 4.6. *A Markov partition $\mathscr{P}$ for f_n on S^1 factors through a connected partition if and only if one of its extreme points is an extreme point for every element of $\mathscr{P}$ whose interval contains it.*

Proof. The necessity is proved in Lemma 4.5. To prove sufficiency let this extreme point be ξ. $f_n(\xi)$ is an extreme point with the same property, so we may assume ξ is periodic. Apply Construction 3.2 to the orbit of ξ to get a metrically generating connected partition, Q. We want to find an N so that $\bigvee_{i \le N} f^{-i}(\mathscr{P})$ refines Q and then apply Lemma 4.3. Let $X(t) = \{x \in X_A : \bigcap_{i \le t} f^{-i}(P_{x_i}) \not\subseteq Q_j \text{ any } j\}$. This is a closed subset of X_A with $X(t+1) \subseteq X(t)$. If $X(N)$ is empty for some N we are done. If no such N exists then by compactness there is a $z \in \bigcap X(t)$. This means $\bigcap_{i \le t} f^{-i}(P_{z_i}) \not\subseteq Q_j$ for any j and all t. So, $\pi_A(z) \in \partial Q$ and we can assume $\pi_A(z) = \xi$. But for each t, ξ is an extreme point for every element of $\bigvee_{i \le t} f^{-i}(\mathscr{P})$ whose interval contains it. Once the diameter of these sets is small, each set whose interval contains ξ must be contained in one of the two elements of Q separated by ξ. A contradiction. □

For a detailed discussion of connected partitions see [St].

Example 4.7. This is an example of a Markov partition for f_2 on S^1 that does not factor through any connected partition. We see this by finding the extreme points and then applying Theorem 4.6. The construction is similar to that in Examples 3.5 and 3.6. Choose a basic interval $A_0 = [5/8, 4/9]$ and let

$$A_n = f^{-1}(A_{n-1}) \setminus \bigcup_{i<n} A_i.$$

We want to see that if I is a preimage interval of A_0, $f^k(I) = A_0$ then either $I \subseteq A_0$ or $I \cap A_0 \subseteq \partial A_0$. This is a computation concerning the preimages of ∂A_0 and the lengths of preimage intervals.

Define

$$P_0 = \mathrm{cl}\left(\bigcup_{\text{even}} A_n \cap \left[\frac{2}{9}, \frac{5}{9}\right]\right), \qquad P_1 = \mathrm{cl}\left(\bigcup_{\text{odd}} A_n \cap \left[\frac{4}{9}, \frac{7}{9}\right]\right),$$

$$P_2 = \mathrm{cl}\left(\bigcup_{\text{even}} A_n \cap \left[\frac{5}{9}, \frac{8}{9}\right]\right), \qquad P_3 = \mathrm{cl}\left(\bigcup_{\text{odd}} A_n \cap \left[\frac{7}{9}, \frac{1}{9}\right]\right),$$

$$P_4 = \mathrm{cl}\left(\bigcup_{\text{even}} A_n \cap \left[\frac{8}{9}, \frac{2}{9}\right]\right), \qquad P_5 = \mathrm{cl}\left(\bigcup_{\text{odd}} A_n \cap \left[\frac{1}{9}, \frac{4}{9}\right]\right).$$

This is Markov partition because

$$f(P_0) = P_1 \cup P_2 \cup P_3, \qquad f(P_1) = P_4 \cup P_0, \qquad f(P_2) = P_5 \cup P_0,$$
$$f(P_3) = P_2 \cup P_4, \qquad f(P_4) = P_3 \cup P_5, \qquad f(P_5) = P_0.$$

We can compute the intervals of the P_j,

$$I(P_0) = [2/9, 8/15], \qquad I(P_1) = [4/9, 23/30], \qquad I(P_2) = [5/9, 53/60],$$
$$I(P_3) = [7/9, 1/15], \qquad I(P_4) = [8/9, 2/15], \qquad I(P_5) = [1/9, 4/15].$$

The extreme points are interleaved and by Theorem 4.6, $\mathscr{P}$ cannot factor through any connected partition.

References

[AGW] R. Adler, W. Goodwyn, and B. Weiss, *Equivalence of topological Markov shifts*, Israel J. Math. **27** (1977), 49–63.

[AW] R. Adler and B. Weiss, *Similarity of automorphisms of the torus*, Mem. Amer. Math. Soc., no. 98, 1965.

[AM] R. Adler and B. Marcus, *Topological entropy and the equivalence of dynamical systems*, Mem. Amer. Math. Soc., no. 219, 1979.

[Bo1] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Math., vol. 470, Springer-Verlag, 1975.

[Bo2] ____, *On Axiom A diffeomorphisms*, CBMS Regional Conf. Ser. in Math., no. 35, Amer. Math. Soc., Providence, R.I., 1978.

[Bo3] ____, *Markov partitions are not smooth*, Proc. Amer. Math. Soc. **71** (1978), 130–132.

[CP] E. Coven and M. Paul, *Endomorphisms of irreducible subshifts of finite type*, Math. Systems Theory **8** (1974), 167–175.

[CR] E. Coven and W. Reddy, *Positively expansive maps of compact manifolds.*

[F] D. Fried, *Finitely presented dynamical systems*, Ergodic Theory Dynamical Systems **7** (1987), 489–507.

[H] G. A. Hedlund, *Endomorphisms and automorphisms of the shift dynamical system*, Math. Systems Theory **3** (1969), 320–375.

[Kr1] W. Krieger, *On sofic systems.* I, Israel J. Math. **48** (1984), 305–330.

[Kr2] ____, *On the subsystems of topological Markov chains*, Ergodic Theory Dynamical Systems **2** (1982), 195–202.

[M] B. Marcus, *Factors and extensions of full shifts*, Monatsh. Math. **88** (1979), 239–247.

[S] M. Shub, *Global stability of dynamical systems*, Springer-Verlag, 1987.

[St] M. Stafford, *Markov partitions for expanding maps of the circle*, Trans. Amer. Math. Soc. **324** (1991), 385–403.

[W] B. Weiss, *Subshifts of finite type and sofic systems*, Monatsh. Math. **77** (1973), 462–478.

IBM Research Division, Almaden Research Center, San Jose, California 95120-6099

IBM Research Division, T. J. Watson Research Center, Yorktown Heights, New York 10598

Department of Mathematics, University of Texas at Austin, Austin, Texas 78713