CRINKLY CURVES, MARKOV PARTITIONS AND DIMENSION.

by

Timothy John Bedford

Submitted for the degree of Ph.D at Warwick University.

Research conducted in the Mathematics Institute,
Warwick University.

September 1984

CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. C. Series for her years of patient advice and help. F.M. Dekking provided the stimulus for much of the work here, as will be obvious to any reader. F. Ledrappier showed me how to obtain the lower bound on dimension in Thrm 4.5. I would finally like to thank Rosalind for putting up with me during the preparation of this thesis, everybody at the Warwick Mathematics Institute for making my time there so enjoyable, and the SERC for providing financial support.

DECLARATION

No material contained in this thesis has been used before.

To Rosalind, my family,

and my friends.

SUMMARY

We consider the relationship between fractals and dynamical systems. In particular we look at how the construction of fractals in (D1) can be interpreted in a dynamical setting and additionally used as a simple method of describing the construction of invariant sets of dynamical systems. There is often a confusion between Hausdorff dimension and capacity -which is much easier to compute- and we show that simple examples of fractals, arising in dynamical systems, exist for which the two quantities differ.

In Chapter One we outline the mathematical background required in the rest of the thesis.

Chapter Two reviews the work of F.M. Dekking on generating 'recurrent sets', which are types of fractals. We show how to interpret this construction dynamically. This approach enables us to calculate Hausdorff dimension and describe Hausdorff measure for certain recurrent sets. We also prove a conjecture of Dekking about conditions under which the best general estimate of dimension actually equals dimension.

In Section One of Chapter Three recurrent sets are used to construct special Markov partitions for expanding endomorphisms of T^2 and hyperbolic automorphisms of T^3. These partitions have transition matrices closely related to the covering maps. It is also shown that Markov partitions can be constructed for the same map whose boundaries have different capacities. Section Two looks at the problem of coding between two Markov partitions for the same expanding endomorphism of T^2. It is shown that there is a relationship between mean coding time and the capacities of the boundaries. Section Three uses recurrent sets to construct fractal subsets of tori which have non-dense orbits under the above mappings.

Finally, Chapter Four calculates capacity and Hausdorff dimension for a class of fractals (which are also recurrent sets) whose scaling maps are not similitudes. Examples are given for which capacity and Hausdorff dimension give different answers.

GLOSSARY OF SYMBOLS. We give page numbers of definitions.

$\tilde{A}: T^n \to T^n$	1
$p: \mathbb{R}^n \to T^n$	1
E^s, E^u	1
W^s, W^u	1
$[x,y]$	2
$\underline{\Sigma}_n$, $\underline{\Sigma}(B)$	2
σ	2
$C_r(\underline{x})$	2
Σ, $\Sigma(B)$	2
G_n	2
G_B	3
h_μ	3
$P(f)$	4
$\pi: \underline{\Sigma} \to T^n$	5
$\bigvee \tilde{A}^{-i} \mathcal{R}$	6
HM_r	6
dim	6
m	7
$N(\varepsilon)$	8
cap	8
S	12
S*	12
$\theta: S^* \to S^*$	12
$G(S)$	12
$f: S^* \to \mathbb{R}^n$	12
$K[\cdot]$	12
$\mathcal{K}(\mathbb{R}^n)$	12
$\mathbb{K}_\theta(W)$	13
Q	14
E	14 /88
θ_{ab}	16
$\mathbb{Z}^{\lvert S \rvert}$	16
$\mathbb{Z}^{\lvert E \rvert}$	16
λ_E	16
λ_i	16
estimate (*)	16
$\lvert W \rvert_E$	17
A^ε	19
$\tilde{G}$	21
G	21
$\tilde{\Sigma}_\theta$	21
Σ_θ	21
$\mathbb{K}_\theta$	22
$\tilde{L}$	23
$\mathcal{R}_\theta$	24
$\tilde{R}(s,t)$	24
$\partial \mathcal{R}_\theta$	27
ls	30
$\partial^s \mathcal{R}$, $\partial^u \mathcal{R}$	44
a_ϕ	65
n_i	66
$Gr(n)$	89
$\underline{p}$	91
$H(\underline{p})$	91
$P(y,n,i)$	91
$\underline{P}_n(y)$	91

INTRODUCTION.

Each chapter has an introduction, so we shall not say too much about individual results here.

Since the publication of Mandelbrot's book (Ma1) there has been wide interest in fractals. The beauty of (Ma1) is that it provides a general language in which a wide range of complicated physical and mathematical phenomena can begin to be described. The appearence of (Ma1) stimulated several mathematical papers including (Hut) which analysed strictly self similar sets using constructions that we would call full shift spaces and Markov partitions. We feel that fractals can be used constructively as well as descriptively. For this a more general formalism for generating fractals than that provided by (Hut) is required. Such a formalism is provided by (D1) because it gives a large degree of control over the geometric properties of the fractal to be constructed.

We analyse the 'recurrent sets' of (D1) using subshifts of finite type (showing in Thrm. 2.16 how to link the different constructions of (Hut) and (D1)) and see that the scaling structure of recurrent sets is closely related to a dynamical structure. This approach enables us to use the ideas from (Fu) and (Bo6) which link ergodic and fractal properties.

Fractals arise naturally in dynamical systems as invariant subsets, for instance in hyperbolic toral automorphisms (Ub). Current constructions of such sets often give little indication of any geometric structure. We show here that it is possible to use Dekking's formalism to produce more examples. In particular we construct special Markov partitions for some hyperbolic automorphisms of T^3. Markov partitions are used to give a description of the dynamics of a map. One would like to have Markov partitions for which the transition matrix could be written down from knowledge of the map. Manning (Mn2) does exactly this for hyperbolic automorphisms of T^2 and Bowen (Bo7) shows that for an Axiom A diffeomorphism with a zero (topological) dimensional basic set, there is a relationship between the induced map on homology and the transition matrix of a Markov partition. Our result is an extension of Manning's to T^3. We also use the recurrent set structure to make statements about expected code time between different partitions for expanding endomorphisms of T^2 and about invariant subsets.

A key element to understanding fractals is calculation of Hausdorff dimension. The dynamical structure of recurrent sets was introduced to enable us to perform this calculation under certain kinds of scaling maps. However, when the scaling map is not of the correct form the dynamical techniques break down. Our final chapter is devoted to calculating Hausdorff dimension in some simple cases of this category.

CHAPTER ONE

MATHEMATICAL BACKGROUND.

Endomorphisms of the torus.

Let A be an nxn matrix with entries in $\mathbb{Z}$. A induces a linear map $A: \mathbb{R}^n \to \mathbb{R}^n$ in the usual way, in particular $A(\mathbb{Z}^n) \subset \mathbb{Z}^n$. We define the n-dimensional torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$. The covering map $p: \mathbb{R}^n \to T^n$ is defined by $x \mapsto x+\mathbb{Z}^n$, and there is an induced map $\tilde{A}: T^n \to T^n$ such that the following diagram commutes,

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^n \\ p \downarrow & & \downarrow p \\ T^n & \xrightarrow{\tilde{A}} & T^n \end{array}$$

Since A is linear, $\tilde{A}$ is an endomorphism, i.e.

$$\tilde{A}(x+y) = \tilde{A}(x) + \tilde{A}(y)$$

where addition is the group operation on T^n. We say that $\tilde{A}$ is a hyperbolic toral automorphism if $|\det A| = 1$ and no eigenvalues of A have unit modulus. $\tilde{A}$ is an expanding toral endomorphism if A has $|\det A| > 1$ and no eigenvalues with modulus less than or equal to one.

If A is hyperbolic, $\mathbb{R}^n$ splits into the direct sum of contracting (stable) and expanding (unstable) A-invariant subspaces, $\mathbb{R}^n = E^s \oplus E^u$. We define the stable and unstable manifolds of $p(x) \in T^n$ as

$$W^s(p(x)) = p(x + E^s) \;, \quad W^u(p(x)) = p(x + E^u)$$

respectively. Then $W^s(y) = \{z \in T^n : d(\tilde{A}^r y, \tilde{A}^r z) \to 0 \text{ as } r \to \infty\}$ and $W^u(y) = \{z \in T^n : d(\tilde{A}^{-r} y, \tilde{A}^{-r} z) \to 0 \text{ as } r \to \infty\}$. Write $W^s(y,\varepsilon) = \{z \in T^n : d(\tilde{A}^r y, \tilde{A}^r z) < \varepsilon \;\; \forall\, r \geq 0\}$ etc. If y,z are close and $\varepsilon > 0$ is small, $W^s(y,\varepsilon) \cap W^u(z,\varepsilon)$ is a single point which we denote $[y,z]$. We note that periodic points of $\tilde{A}$ are dense in T^n and that $\overline{W^s(y)} = T^n$ for any $y \in T^n$. Hyperbolic toral automorphisms are well known examples of Anosov diffeomorphisms and of Axiom A diffeomorphisms.

Symbolic dynamics

The full shift on n symbols (Σ_n, σ) is the shift map on $\Sigma_n = \prod_{-\infty}^{\infty}\{0,1,..,n-1\} = \{(x_i)_{i\in\mathbb{Z}}: x_i \in \{0,1,..,n-1\}\}$ defined by $(x_i) \overset{\sigma}{\mapsto} (y_i)$ where $x_{j+1} = y_j$, $\forall j \in \mathbb{Z}$. We metrize Σ_n by defining $d((x_i),(y_i)) = 2^{-m}$ where m is the largest integer such that $x_j = y_j$ for $|j| \leq m$. The topology induced by this metric is the same as the product topology coming from the discrete topology on $\{0,1,...,n-1\}$. By Tychonoff's Theorem Σ_n is compact. A basis for the topology is given by cylinders, an r-cylinder being

$$C_r(\underline{x}) = \{\underline{y} \in \Sigma_n : \underline{x} = (x_i),\ \underline{y} = (y_i),\ x_j = y_j \quad j = 0, \pm 1, .., \pm r\}.$$

Let B be an nxn matrix with entries in $\{0,1\}$. We define the subshift of finite type $\Sigma(B)$ by

$$\Sigma(B) = \{(x_i) \in \Sigma_n : b_{x_i\, x_{i+1}} = 1 \quad \forall i\}.$$

One also defines in a similar way one sided shift spaces Σ_n, $\Sigma(B)$ where $\Sigma_n = \prod_0^{\infty}\{0,..,n-1\}$.

We can consider Σ_n as the set of bi-infinite paths around the directed graph G_n with n vertices and a directed edge (v_i, v_j) for each ordered pair of vertices. If one labels the paths by $\mathbb{Z}$, $\sigma(\underline{x})$ is the same path as $\underline{x}$ but with the labelling moved on by one place. More generally, if B is an nxn matrix with entries in $\mathbb{Z}^+$ we can associate a directed graph G_B to B. G_B has n vertices labelled 0,..,n-1, and there are b_{ij} directed edges from vertex i to vertex j. We can now define $\underline{\Sigma}(B)$ as the set of bi-infinite paths around G_B labelled by $\mathbb{Z}$. In fact, any $(\underline{\Sigma}(B),\sigma)$ where B is a matrix over $\mathbb{Z}^+$ is topologically conjugate to a $(\underline{\Sigma}(A),\sigma)$ where A is a matrix over $\{0,1\}$, in other words there is a homeomorphism $\phi: \underline{\Sigma}(B) \to \underline{\Sigma}(A)$ so that $\phi\sigma_B = \sigma_A\phi$. The symbols of $\underline{\Sigma}(A)$ are the edges of G_B, and $a_{ij} = 1$ if and only if edge i ends at vertex v and edge j begins at vertex v. (Ad2) gives more details about this kind of construction.

The fractals we study have a scaling structure and will be modelled using subshifts of finite type. In particular, certain kinds of shift invariant measures will be used in Hausdorff dimension calculations. Write $\underline{\Sigma} = \underline{\Sigma}(B)$, and let $M(\underline{\Sigma},\sigma)$ be the set of shift invariant Borel probability measures on $\underline{\Sigma}$. Let $\zeta_n = \{C_n(\underline{x}) : \underline{x} \in \underline{\Sigma}\}$. With the convention that $0 \log 0 = 0$, we define the <u>entropy of ζ_n with respect to μ</u> as

$$H_\mu(\zeta_n) = -\sum_{A \in \zeta_n} \mu(A) \log \mu(A),$$

the <u>entropy of σ with respect to μ</u> is

$$h_\mu(\sigma) = h_\mu(\sigma,\zeta) = \lim_{n\to\infty} (1/n)\ H_\mu(\zeta_n) \quad .$$

Given a lipschitz function $f: \Sigma \to \mathbb{R}$, the <u>pressure</u> of f is

$$P(f) = \sup \left\{ h_\mu(\sigma) + \int f d\mu \; : \; \mu \in M(\Sigma,\sigma) \right\} .$$

The assumption that f is lipschitz implies that there is a unique ergodic member μ_f of $M(\Sigma,\sigma)$ maximizing $h_\mu(\sigma) + \int f d\mu$. We call μ_f the <u>equilibrium state</u> for f. The measure μ_f is a Gibbs measure i.e. there are constants $a, b > 0$ such that

$$a \leq \mu_f\left(C_{n+1}(x_0 \ldots x_n)\right)^{-1} \exp\left(-nP(f) + \sum_{i=0}^{n} f(\sigma^i \underline{x})\right) \leq b$$

where $\underline{x} = (x_0 \ldots x_n \ldots)$. In particular if $f \equiv 0$, μ_f is the <u>measure of maximal entropy</u>. Good references for the ergodic theory of shift spaces are (Bo3),(Wa1,Wa2).

<u>Markov partitions</u>.

The principle technique by which subshifts of finite type are used to study Axiom A diffeomorphisms is by the use of Markov partitions. We show how the construction works for a hyperbolic automorphism, $\tilde{A}: T^n \to T^n$.

For $R \subset T^n$ put $W^s(x,R) = W^s(x,\varepsilon) \cap R$, $W^u(x,R) = W^u(x,\varepsilon) \cap R$. A <u>Markov partition</u> $\mathcal{R} = \{R_0, \ldots, R_{m-1}\}$ is a finite collection of subsets covering T^n, with diameters small compared to ε and

i) $R_i = \overline{\text{int } R_i}$, and $R_i \neq \emptyset$

ii) $(\text{int } R_i) \cap (\text{int } R_j) \neq \emptyset \Rightarrow i=j$

iii) $x, y \in R_i \Rightarrow [x,y] \in R_i$

and satisfying the Markov conditions,

If $x \in \text{int } R_i$, and $\tilde{A}x \in \text{int } R_j$ then

iv) $\tilde{A}W^u(x,R_i) \supset W^u(Ax,R_j)$

v) $\tilde{A}W^s(x,R_i) \subset W^s(Ax,R_j)$.

(Ad1), (Si1, Si2), and (Bo1, Bo2) demonstrate the existence of Markov partitions in various settings including hyperbolic toral automorphisms.

The transition matrix for $\mathcal{R}$ is the mxm matrix B defined by

$$b_{ij} = 1 \quad \text{if} \quad \tilde{A}(\text{int } R_i) \cap \text{int } R_j \neq \emptyset$$
$$= 0, \quad \text{otherwise.}$$

The Markov properties imply that if $\underline{x} = (x_i)_{i \in \mathbb{Z}} \in \sum(B)$, $\bigcap_{i=0}^{k} \tilde{A}^i R_{x_{-i}}$ converges as $k \to \infty$ to a line segment that is a piece of unstable manifold in R_{x_0} (fig 1).

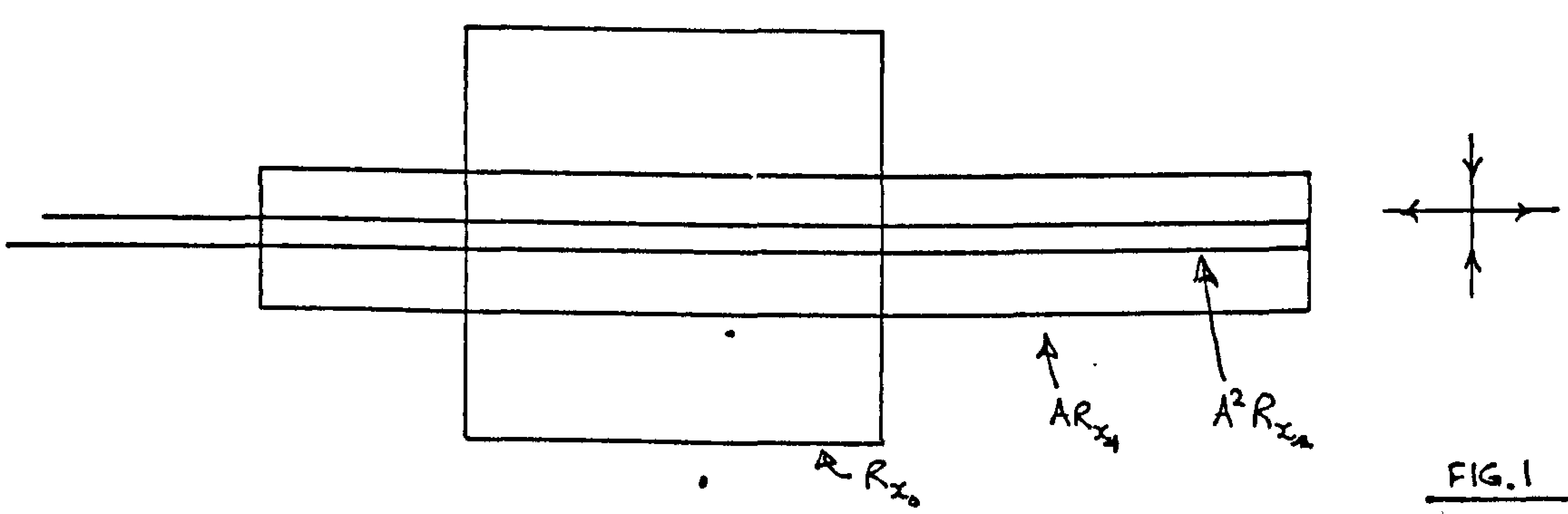


FIG. 1

Similarly $\bigcap_{i=0}^{+\infty} A^{-i}R_{x_i}$ is a piece of stable manifold in R_{x_0}. Thus, since these two submanifolds intersect in a single point of R_{x_0} we may define a map $\pi: \sum(B) \to T^n$ by

$$\pi((x_i)_{i \in \mathbb{Z}}) = \bigcap_{-\infty}^{\infty} \tilde{A}^{-i}R_{x_i}.$$

π is continuous, onto, boundedly finite to one, one to one on a residual set, and $\tilde{A}\pi = \pi\sigma$. Thus one can follow orbits of points on T^n by considering a corresponding symbol sequence in $\Sigma(B)$.

The map π is used to push measures on Σ down onto T^n, and to describe the dynamics of $\tilde{A}$ (for instance $\pi \underline{x} \in T^n$ is periodic if and only if $\underline{x} \in \Sigma(B)$ is). The constructions of Markov partitions referred to above give little indication of what form the matrix B can take. We shall return to this question in a later chapter. In the simpler case of $\tilde{A}$ being an expanding endomorphism a similar construction works (deleting (iii) and (v) from the definition of a Markov partition) to give a semiconjugacy from a one sided shift space. If $\mathcal{R} = \{R_1, \ldots, R_m\}$ is a Markov partition we shall write $\bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{R} = \{R_{j_0} \cap \tilde{A}^{-1}R_{j_1} \cap \cdots \cap \tilde{A}^{-i}R_{j_i} : (j_0 \ldots j_i \ldots) \in \Sigma(B)\}$.

Fractals

Suppose (X,d) is a metric space. Let $|U|$ denote the diameter of a set U. If $Y \subset X$, we define the <u>r-dimensional Hausdorff measure</u> of Y as

$$HM_r(Y) = \lim_{\varepsilon \to 0} \inf \left\{ \sum_{i=1}^{\infty} |U_i|^r : \bigcup_{i=1}^{\infty} U_i \supset Y, \ |U_i| < \varepsilon \ \forall i \right\}$$

Clearly $HM_r(Y) \in [0,\infty]$. The <u>Hausdorff dimension</u> of Y is

$$\dim(Y) = \inf\{r : HM_r(Y) = 0\}.$$

HM_r is an outer measure i.e. $HM_r(\bigcup_1^{\infty} A_n) \leq \sum_1^{\infty} HM_r(A_n)$

and $HM_r(\emptyset) = 0$. The set of HM_r-measurable sets forms a σ-algebra on which HM_r is σ-additive, and which includes the Borel sets. In particular when $r \in \mathbb{N}$, HM_r is equivalent to the usual outer Lebesgue measure. We shall always denote Lebesgue measure of the ambient space by m. Some useful facts about Hausdorff dimension are given below (their proofs are straightforward and can be found in (Ca) and (Ro))

i) $Y \subset W \Rightarrow \dim Y \leqslant \dim W$

ii) $\dim Y = \sup \{ r : HM_r(Y) = \infty \}$

iii) If $Y = \bigcup_1^\infty Y_n$, $\dim Y = \sup_n \dim Y_n$

iv) If $f : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $f(x) = t.x$ $(t \in \mathbb{R})$ then $t^r HM_r(W) = HM_r(f(W))$.

v) If $f : \mathbb{R}^n \to \mathbb{R}^n$ is Hölder continuous with exponent a, i.e. $|fx - fy| \leqslant c.|x - y|^a$, and $W \subset \mathbb{R}^n$, $a.\dim f(W) \leqslant \dim W$.

Two of the more useful results for obtaining lower bounds on dimension are

<u>Frostman's Lemma</u> Let K be a compact set in $\mathbb{R}^n$. Then $HM_r(K) > 0$ if and only if there is a probability measure μ with support on K such that for all balls B

$$\mu(B) \leqslant c.|B|^r$$

where c is a positive constant.

■

A proof of Frostman's lemma is given in (Ca, p7).

<u>Marstrand's Theorem</u> (Mr) Suppose E is a plane set and that p is a positive number such that for every point

x of a given set A, writing $E_x = \{(x,y) \in E : y \in \mathbb{R}\}$, we have $HM_t(E_x) > p$. Then $HM_{s+t}(E) \geq k.p.HM_s(A)$ where k is a positive absolute constant.

■

A corollary of this result is that $\dim(A \times B) \geq \dim A + \dim B$.

Dimension intuitively gives a measure of the 'rarity' of a set, but is often difficult to compute. Another measurement is that of capacity. Let $N(\varepsilon)$ be the minimum number of balls of diameter ε required to cover a compact set W in $\mathbb{R}^n$. Define the upper and lower capacities of W by

$$\overline{\text{cap}}\ W = \limsup_{\varepsilon \to 0} \log N(\varepsilon) / \log (1/\varepsilon) \quad \text{and}$$
$$\underline{\text{cap}}\ W = \liminf_{\varepsilon \to 0} \log N(\varepsilon) / \log (1/\varepsilon) \quad \text{respectively.}$$

When $\overline{\text{cap}}\ W = \underline{\text{cap}}\ W$, we call the common value the capacity of W. (Note that some authors use the terms limit capacity or logarithmic density in order to distinguish our capacity from potential theoretic capacity).

It is easy to see that if $W \subset \mathbb{R}^n$,

$$n \geq \overline{\text{cap}}\ W \geq \underline{\text{cap}}\ W \geq \dim W.$$

There are examples of sets W for which $\underline{\text{cap}}(W) > \dim(W)$, for instance $\mathbb{Q}$. This particular example works because $\text{cap}(W) = \text{cap}(\overline{W})$ for any W. We shall see, in chapter four, an example of an invariant set in a dynamical system for which cap $\neq$ dim.

Mandelbrot (Ma1) defines a <u>fractal</u> as a set whose topological dimension differs from its Hausdorff dimension. (Hur) give a definition of topological dimension and prove that

$$\dim W \geqslant \text{top. dim.} (W)$$

always holds. We shall be rather liberal in the use of the word 'fractal' owing to the difficulty of calculating dim in general. The term 'fractal dimension' will be taken by us to refer both to Hausdorff dimension and capacity.

Main definitions and results are numbered n.m where n is the chapter they occur in, and m numbers them consecutively within the chapter.

CHAPTER TWO

RECURRENT SETS

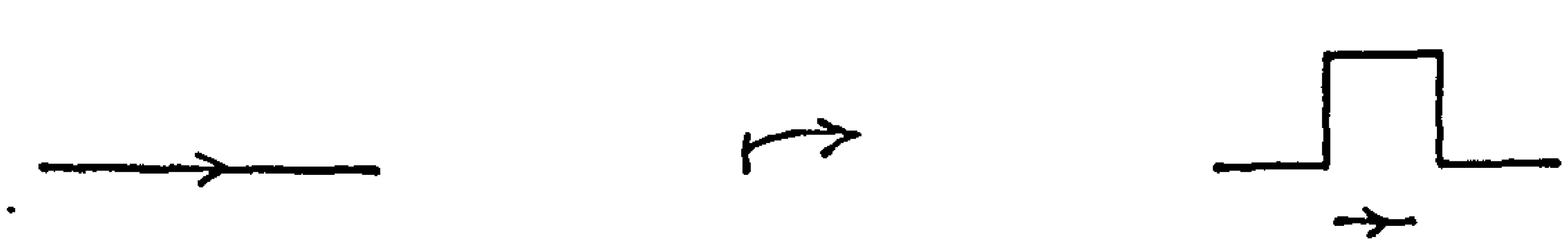
In this Chapter we consider Michel Dekking's construction of fractals, and see how to view them as dynamical systems. Subshifts of finite type are used to calculate dimension and describe Hausdorff measures.

§ 1: Constructing Recurrent Sets.

Dekking's construction of fractals (D1) is based upon the idea of "polygonal line substitution" (described by Mandelbrot in Ma 1, Ma 2). It is consequently a more flexible technique than that used by, for example, Hutchinson (Hut). It is more difficult to analyse from the point of view of calculating dimension and was developed prior to both (Ma 1) and (Hut).

A simple example is provided by the scheme in which straight (directed) line segments are transformed as follows (fig 2).

FIG. 2



We begin with a line segment and proceed inductively to transform the smaller sized line segments generated. In order to draw the pictures on a computer we use symbols to represent the various directed line segments occuring (fig 3).

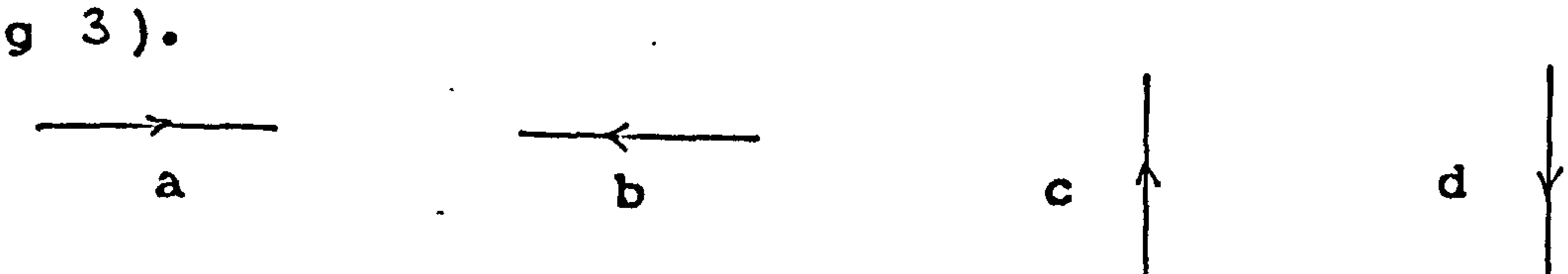
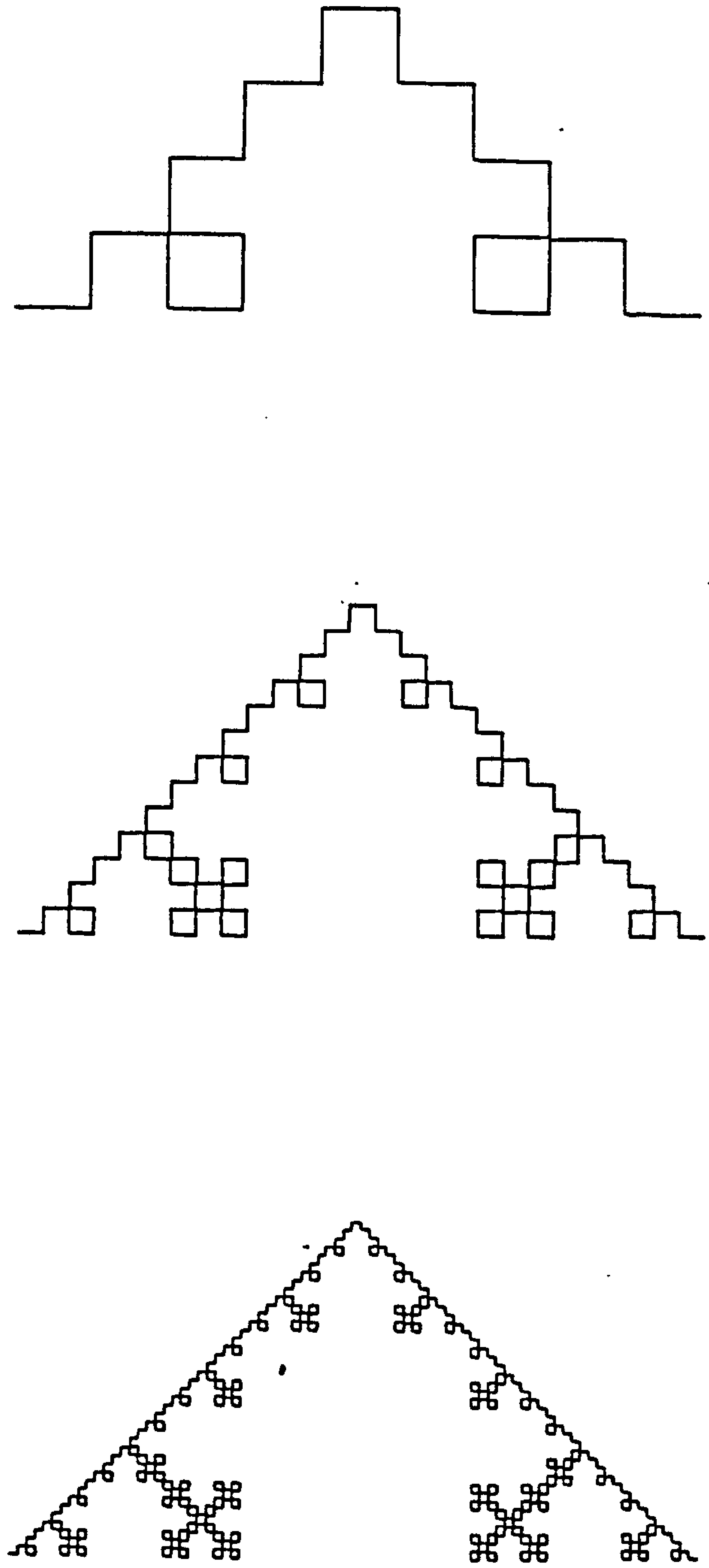


FIG. 3



$L^{-2}K[\theta^2 a]$

$L^{-3}K[\theta^3 a]$

$L^{-4}K[\theta^4 a]$

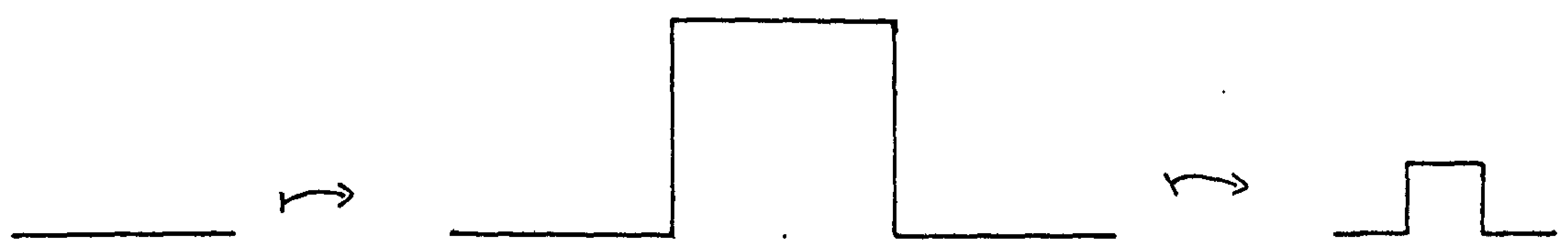
FIG. 5

Then the given transformation is given symbolically by

$$a \mapsto acada$$
$$b \mapsto bdbcb$$
$$c \mapsto cbcac$$
$$d \mapsto dadbd.$$

In order to obtain the picture of fig. 2 one has also to normalize.(fig 4).

FIG. 4



Further stages of the induction give pictures as shown in fig. 5 .

We now formalize the above approach. Let S be a finite alphabet of symbols. S* will be the free semigroup generated by S, and $\Theta: S^* \to S^*$ a semigroup endomorphism. G(S) is the free group generated by S. We denote by $f: S^* \to \mathbb{R}^n$ a homomorphism i.e. f satisfies

$$f(VW) = f(V) + f(W)$$

for all words $V, W \in S^*$. The map f is used to describe relative position. We also require a map to associate compact subsets of $\mathbb{R}^n$ with words in S*. Denote by $\mathcal{K}(\mathbb{R}^n)$ the space of compact subsets of $\mathbb{R}^n$. We require a map $K[\cdot]: S^* \to \mathcal{K}(\mathbb{R}^n)$ to have the property that

$$K[VW] = K[V] \cup (K[W] + f(V)) \qquad \forall\, V,W \in S^*.$$

The natural choice for $K[\cdot]$ when drawing fractals on a computer is

$$K[s] = [0, f(s)] \qquad \forall\, s \in S$$

where $[a,b] = \{at + (1-t)b : t \in [0,1]\}$. This makes $K[s_1 \ldots s_r]$ the polygonal line with vertices at 0, $f(s_1)$, $f(s_1)+f(s_2), \ldots, f(s_1)+\ldots+f(s_r)$. Other important choices for $K[\cdot]$ from a pure mathematical point of view are $K[s] = \mathbb{K}_\theta(s)$ (see below) and $K[s] = \{f(s)\}$. In fact, the choice of $K[\cdot]$ does not affect the recurrent set generated (2.16).

Suppose now that $L: \mathbb{R}^n \to \mathbb{R}^n$ is a linear map so that $f\theta(s) = Lf(s)$ for all $s \in S$. (This condition ensures that when we substitute a polygonal line for a line interval our two arcs have the same beginning and end points - in other words L is the 'normalizing' map referred to above). Then our approximations to a particular fractal, denoted $\mathbb{K}_\theta(W)$, are given by

$$L^{-m}K[\theta^m W] \quad .$$

In order for these approximations to converge in the Hausdorff metric, we require that L should have all its eigenvalues of modulus larger than one i.e. that L should be <u>expanding</u> (Dekking uses the word 'expansive' which we shall avoid due to its use in dynamical systems).

So far we have only seen how to produce fractal curves. In order to generate more complicated figures such as Cantor sets we need to use symbols that can give us 'gaps'.

Def. 2.1 A symbol $s \in S$ is virtual if $K[s] = \emptyset$.

It would now be possible for our approximating sets to eventually become empty if for some N, $\theta^N W$ contained only virtual symbols. We need to find some extra conditions that will ensure convergence to a non-empty set.

Def. 2.2 Let $Q \subset S$ (so $Q^* \subset S^*$). Then $\theta : S^* \to S^*$ is Q-stable if there exists $m > 0$ such that for all $s \in S$ either

i) $\theta^k s \in Q^* \quad \forall k \geq m$ or, ii) $\theta^k s \notin Q^* \quad \forall k \geq m$.

From now on Q will denote the set of virtual symbols. Any symbol that satisfies (ii) is called essential. The set of essential symbols is denoted E. The following Theorem asserts the existence of recurrent sets.

Thrm 2.3 (D1,3.3) Let $\theta, S^*, f, K[.]$, and L be as above with L expanding. Suppose that θ is Q-stable. Then there exists a non-empty compact set $K_\theta(W)$ such that

$$L^{-m}K[\theta^m W] \longrightarrow K_\theta(W) \quad \text{as} \quad m \to \infty$$

in the Hausdorff metric, for any word W which contains at least one essential symbol.

■

Remarks i) Since $f\theta = Lf$, we have that if $\theta(s)=s_1..s_r$,

$$L\mathbb{K}_\theta(s) = \bigcup_{i=1}^{r}\left(\mathbb{K}_\theta(s_i) + f(s_1..s_{i-1})\right).$$

ii) For any $m>0$, $f\theta^m = L^m f$, and so $\mathbb{K}_{\theta^m}(W) = \mathbb{K}_\theta(W)$.

It is possible to have a virtual symbol that is also essential. However, by (D3,2.1) we can generate any recurrent set in such a way that $E = S\backslash Q$ and $\theta Q^* \subset Q^*$. From now on we shall assume this is the case. Furthermore, we can now remove the conditions on Q-stability,

Prop. 2.4 (D1, 3.2i) $\theta Q^* \subset Q^*$ implies that θ is Q-stable.

■

Different Scaling Maps

The 'consistency' condition that allowed us, above, to perform the polygonal line substitution was $f\theta = Lf$. We now wish to define a recurrent set in which more than one scaling ('normalizing') map is used (c.f. Hut). For each $s \in S$, let $(Ls): \mathbb{R}^n \to \mathbb{R}^n$ be an expanding linear map. Suppose that $\theta(s) = s_1..s_r$. The consistency condition required to ensure that our polygonal line substitute fits where it should go is

$$\sum_{j=1}^{r} (Ls_j)^{-1}(fs_j) = f(s) .$$

Dekking (D3, p7) shows that one can define approximations to a recurrent set using this idea. He calculates dimension in the limited case that there are two similitudes L_1, L_2 and for all s, $(Ls) = L_1$ or L_2.

The current formalism is not very effective in this piecewise linear case. We shall see later that by supplementing this formalism we can deal with recurrent sets for some piecewise linear and non-linear scalings.

Dimension Estimates

Let $\theta : S^* \to S^*$ be an endomorphism, then we can abelianize θ to obtain a map $\theta_{ab} : \mathbb{Z}^{|S|} \to \mathbb{Z}^{|S|}$. Corresponding to the essential symbols E is a space $\mathbb{Z}^{|E|} \subset \mathbb{Z}^{|S|}$. Since $\theta Q^* \subset Q^*$, θ induces a map on E^* (or, more properly, on S^*/Q^*) and hence on $\mathbb{Z}^{|E|}$. The induced map on $\mathbb{Z}^{|E|}$ is given by a non-negative matrix with integer entries. We denote the eigenvalue with greatest modulus by λ_E. Let $L : \mathbb{R}^n \to \mathbb{R}^n$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ with $|\lambda_1| \geqslant \ldots \geqslant |\lambda_n| > 1$. Then

$$\dim K_\theta(W) \leqslant n-1 + \frac{\log \lambda_E - \sum_{i=2}^{n} \log |\lambda_i|}{\log |\lambda_1|} \qquad (*)$$

(We shall often refer to this inequality as the dimension estimate (*)). The estimate comes from a simple "box counting" process as follows.

Choose $M > 0$ so that for each $s \in E$, $\mathbb{K}_\theta(s)$ is contained in an n-cube of side length M (we assume that the edges of these n-cubes lie parallel to the eigenspaces of L). Notice that by the remark (i) after 2.3, if we choose $K[s] = \mathbb{K}_\theta(s)$ for all $s \in S$ then

$$L^{-m}K[\theta^m s] = \mathbb{K}_\theta(s) \quad \text{for all } m.$$

Thus there is a covering of $\mathbb{K}_\theta(s)$ by $|\theta^m s|_E$ boxes (where $|W|_E$ is the number of symbols of W in E). Each box is the image under L^{-m} of an n-cube of side M, and has sides of lengths

$$M|\lambda_1|^{-m}, \ldots\ldots, \quad M|\lambda_n|^{-m} .$$

Such a box can itself be covered by

$$\left|\frac{\lambda_1}{\lambda_2}\right|^m \left|\frac{\lambda_1}{\lambda_3}\right|^m \cdots \left|\frac{\lambda_1}{\lambda_n}\right|^m$$

n-cubes of side length $M|\lambda_1|^{-m}$. Since clearly there exists $N > 0$ such that $|\theta^m s|_E \leqslant N \lambda_E^m$, we can make an estimate of the capacity of $\mathbb{K}_\theta(s)$,

$$\operatorname{cap}(\mathbb{K}_\theta(s)) \leqslant \lim_{m\to\infty} \frac{\log(N.\lambda_E^m |\lambda_1|^{(n-1)m} |\lambda_2|^{-m} \cdots |\lambda_n|^{-m})}{-\log(M \lambda_1^{-m})}$$

$$= n-1 + \frac{\log \lambda_E - \sum_{i=2}^{n} \log|\lambda_i|}{\log|\lambda_1|}$$

The dimension estimate now follows because

$$\dim \mathbb{K}_\theta(s) \leqslant \operatorname{cap} \mathbb{K}_\theta(s) .$$

The estimate (*) is the best possible in general. Dekking (D1) gives several examples where

$$|\lambda_1| = |\lambda_2| = \ldots = |\lambda_n|$$

and equality holds in (*). He also gives an example of a plane filling curve for which $|\lambda_1| \neq |\lambda_2|$ and equality holds in (*). In Chapter 4 we give an example of a non-plane-filling curve where $|\lambda_1| \neq |\lambda_2|$ and equality holds in (*). As is pointed out in (D3) there are several possible obstructions to equality in (*).

Def. 2.5 θ is essentially mixing if there exists n such that for all $s, t \in E$, $s \in \theta^n t$.

If θ is not essentially mixing there may be some essential s with

$$\overline{\lim_{n\to\infty}} (1/n) \log |\theta^n s|_E < \lambda_E.$$

This would make (*) too big for $K_\theta(s)$.

Even if θ is essentially mixing, λ_E may give an incorrect estimate of the number of boxes required to cover $K_\theta(s)$. This is because the economy of our covering depends upon the different copies of $L^{-m} K_\theta(s)$ in $L^{-m} K[\theta^m W]$ not overlapping too much. If L is a similitude and $f(s) \in \mathbb{Z}^n$ for all $s \in S$ then the only thing that can go wrong is for us to draw two copies of $L^{-m} K_\theta(s)$ in the same place.

Def. 2.6 A symbol $s \in S$ duplicates if the word sUs occurs in some $\theta^n t$, and $f(sU) = 0$.

Thrm. 2.7 (D3, 6.1) Let $\theta: S^* \to S^*$ be essentially mixing and L an expanding similitude with eigenvalues of modulus λ. Suppose that $f(s) \in \mathbb{Z}^n$ for all symbols s. Then $\dim K_\theta = (\log \lambda_E)/(\log \lambda)$ if and only if no essential symbol duplicates.

■

This Theorem works in fact if f takes values in any lattice. In order to deal with the more general situation Dekking makes the following definition and conjecture.

Def. 2.8 A recurrent set $K_\theta(W)$ is resolvable if

$$\lim_{n\to\infty} \frac{m((K[\theta^n W])^\varepsilon)}{|\theta^n W|_E} > 0 \text{ for some } \varepsilon > 0.$$

where m is Lebesgue measure of the ambient space and $A^\varepsilon = \{x \in \mathbb{R}^n : |x - y| < \varepsilon, y \in A\}$ for $A \subset \mathbb{R}^n$.

Conjecture 2.9 (D3, 6.3) Let $\theta: S^* \to S^*$ be essentially mixing with L an expanding similitude. Then

$$\dim K_\theta = \frac{\log \lambda_E}{\log \lambda} = (*) \Leftrightarrow K_\theta \text{ is resolvable.}$$

Dekking proves this conjecture when $f(s) \in \mathbb{Z}^n$ for all s. We will prove it completely in the next section.

If the scaling map L is not a similitude the

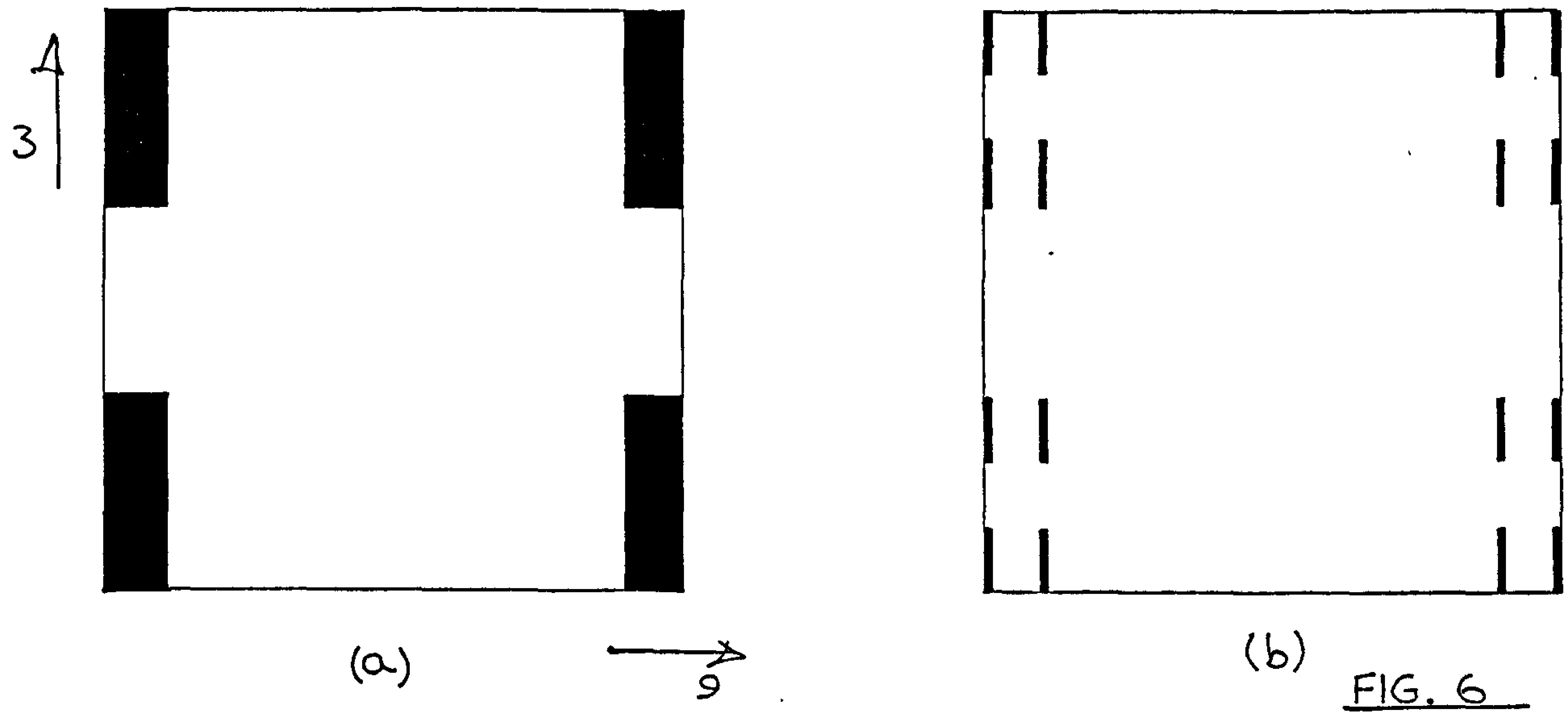


FIG. 6

dimension estimate (*) may not even be the capacity. As an example we give a set defined by polygonal substitution (this is a technique fully explained in Chapter 4. It works in a similar fashion to polygonal line substitution but with smaller sized polygons replacing the larger ones). The first two approximations to our set are shown in fig. 6 (a),(b),(the set is also a recurrent set,(D3, p15)). The covering argument used to obtain (*) would have made us cover each of the shaded regions of fig 6(a) with three squares of side length 1/9. At the next stage of approximation (fig 6(b)),however, we see that only two squares of side 1/9 were needed. Obviously the over-estimation becomes far worse as the size of our cover decreases. The capacity of the above example is $\frac{\log 12}{\log 9}$ (and equals the dimension) compared with the estimate (*) of $\frac{\log 18}{\log 9}$. We shall consider the problem of L not being a similitude more fully in Chapter 4.

§ 2: Symbolic Dynamics for Recurrent Sets.

Our basic idea is to set up a correspondence between a semigroup endomorphism θ and a subshift.

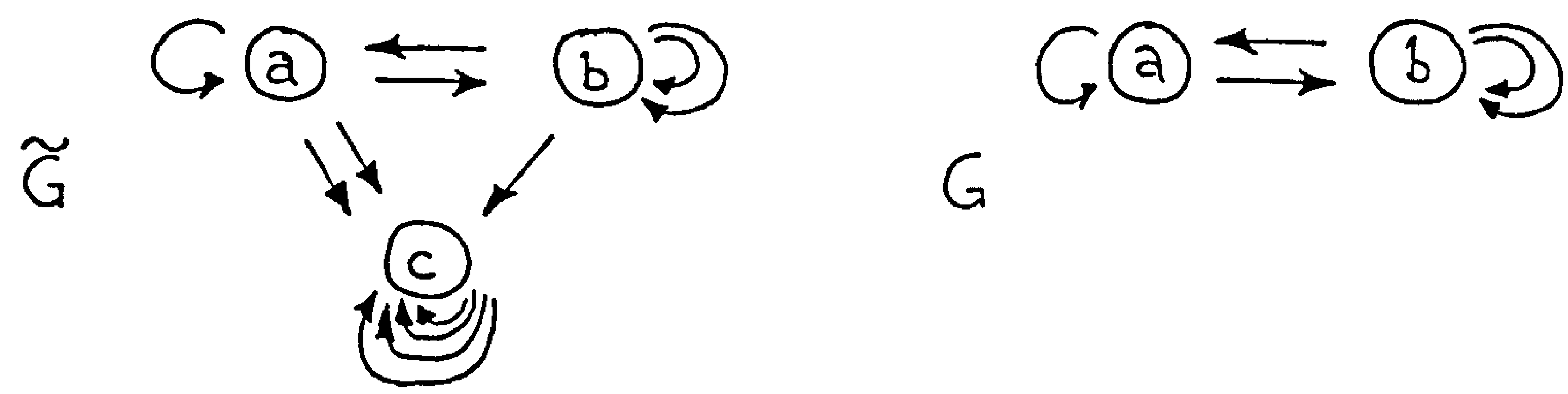
Let $S = \{0,1,..,n-1\}$ be our alphabet of symbols. We draw a directed graph $\tilde{G}$ corresponding to θ. The graph has n vertices labelled $0,..,n-1$. There are k paths from vertex s to vertex t if and only if t appears in the word $\theta(s)$ k times. Label the edges of the graph by triples (s,t,j) where $t \in \theta(s)$, $1 \leq j \leq k$. We define the (one-sided) shift space $\bar{\Sigma}_\theta$ as the set of infinite paths around $\tilde{G}$, i.e. $\bar{\Sigma}_\theta = \Sigma(\theta_{ab})$ where θ_{ab} is thought of as the $n \times n$ matrix that acts on $\mathbb{Z}^{|S|}$.

It is not true, in general, that $\bar{\Sigma}_\theta$ contains transitive points. In fact, since $\theta\Omega^* \subset \Omega^*$, the set of states corresponding to edges ending at 'virtual' vertices is absorbing.

<u>Def. 2.10</u> $\Sigma_\theta \subset \bar{\Sigma}_\theta$ is the subshift consisting of all paths that only visit vertices representing essential symbols.

Clearly Σ_θ is represented by a subgraph G of $\tilde{G}$ and so is of finite type. In fact, if A is the matrix such that $A: \mathbb{Z}^{|E|} \to \mathbb{Z}^{|E|}$ is the map on $\mathbb{Z}^{|E|} \subset \mathbb{Z}^{|S|}$ induced by θ_{ab}, then $\Sigma_\theta = \Sigma(A)$. Thus the topological entropy of $(\sigma_\theta, \Sigma_\theta)$ equals $\log \lambda_E$.

Example 2.11 Let $S = \{a,b,c,\}$, $E = \{a,b\}$ $Q = \{c\}$. $\theta : S^* \to S^*$ is defined by $a \mapsto cabc$, $b \mapsto bbca$, $c \mapsto cccc$. Then $Q^* = \langle c \rangle$ and $\theta Q^* \subset Q^*$. The matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.



Prop. 2.12 If θ is essentially mixing then $(\sigma_\theta, \Sigma_\theta)$ is aperiodic.

Proof: There is an $n > 0$ such that for all $s, t \in E$, $s \in \theta^n t$. Hence $(A^n)_{s,t} > 0$ for all $s,t \in E$,

∎

We now wish to project Σ_θ onto our recurrent set. From now on we take K_θ to be the disjoint union of $K_\theta(s)$, $s \in E$. In order to define

$$\pi : \Sigma_\theta \to K_\theta = \coprod_{s \in E} K_\theta(s) \subset \coprod_{s \in E} \mathbb{R}^m_s$$

we need only define the image of an n-cylinder of Σ_θ in K_θ (since each $\underline{x} \in \Sigma_\theta$ is a countable intersection of n-cylinders). An n-cylinder of Σ_θ corresponds to a path of length n around G and hence to a sequence

$$s_0 ; \quad (s_0,s_1,h_1),\ldots,(s_{n-2},s_{n-1},h_{n-1})$$

of labelled edges of G (together with a starting vertex) where for each $i, s_i \in \theta(s_{i-1})$ and s_i is the h_i th occurence of the symbol $s=s_i$ in $\theta(s_{i-1})$. We can now inductively define $s^*_i \in \theta^i E$ by $s^*_0 = s_0 \in E$ and s^*_i is the h_i th occurence of s_i in $\theta s^*_{i-1} \subset \theta^i s_0$. We define

$$\pi(\,C_n(s_0;\ldots,(s_{n-2},s_{n-1},h_{n-1})))$$
$$= \quad L^{-n}\Big(K_\theta(s^*_{n-1}) + f(W_1)\Big) \subset \mathbb{R}^m_{s_0}$$

where $\theta^n s_0 = W_1 s^*_{n-1} W_2$, some words W_1, W_2. Since for all $s \in S$, if $\theta s = s^1.\ldots.s^r$ then

$$K_\theta(s) = \bigcup_1^r L^{-1}\Big(K_\theta(s^i) + f(s^1\ldots.s^{i-1})\Big),$$

we see that

$$\pi(\,C_n(s_0;\ldots,(s_{n-2},s_{n-1},h_{n-1})))$$
$$\supset (\,C_{n+1}(s_0;\ldots,(s_{n-1},s_n,h_n)))$$

as required.

<u>Remark</u> i) π is clearly continuous and surjective.

ii) For each triple (s,t_k,h) (where $\theta(s) = t_1\ldots t_r$) used to label an edge of G define the <u>right shift</u> $\sigma_{(s,t_k,h)}: C_0(t_k) \to \Sigma_\theta$ by

$$(t_k;\ (t_k,s_1,h_1),\ldots) \mapsto (s;\ (s,t_k,h),(t_k,s_1,h_1),\ldots).$$

Then the following diagram commutes

$$\begin{array}{ccc} C_0(t_k) & \xrightarrow{\quad\sigma_{(s,t_k,h)}\quad} & C_0(s) \\ \pi\downarrow & & \downarrow\pi \\ K_\theta(t_k) & \xrightarrow{\quad\tilde{L}^{-1}\quad} & K_\theta(s) \end{array}$$

where $\tilde{L}^{-1}$ is defined by mapping $K_\theta(t_k)$ onto the copy of $K_\theta(t_k)$ in $K_\theta(s) = L^{-1}(\bigcup_{i=1}^{r}(K_\theta(t_i) + \sum_{j<i} f(t_j)))$. Unfortunately, in general, there is no mapping of K_θ to itself corresponding with the left shift on Σ_θ (Hutchinson (Hut) uses right shifts on $\prod_1^\infty \{1,..,n\}$ to describe self-similar sets for this reason). Such a map, $\tilde{L}$, would have to be multivalued at very many points. We are interested in the circumstances under which $\tilde{L}$ would be defined on a residual subset of K_θ.

<u>Def. 2.13</u> A recurrent set K_θ is <u>well matched</u> (to θ) and the collection $\mathcal{R}_\theta = \{K_\theta(s): s \in E\}$ is the <u>θ-Markov partition</u> if the following property holds. When $\theta(s) = t_1 \dots t_r$, $s \in E$, writing $\tilde{R}(s,t_i) = K_\theta(t_i) + f(t_1 \dots t_{i-1})$ for each $t_i \in E$ we have

$$\text{int}\left(\tilde{R}(s,t_i) \cap \tilde{R}(s,t_j)\right) = \emptyset \quad \text{if} \quad i \neq j$$

in the induced topology as a subset of $L\,K_\theta(s)$.

<u>Remarks</u> i) When defining Markov partitions in dynamical systems one usually requires further conditions which would correspond to a) $K_\theta(s) = \overline{\text{int } K_\theta(s)}$ in the induced topology and b) $L^{-1}\left(\bigcup_{i=1}^{r} \tilde{R}(s,t_i)\right) = K_\theta(s)$. Here these conditions hold because of the continuity of π and the structure of a recurrent set respectively. The difference in definition occurs because one is usually trying to set up the map π which we already have here.

ii) Even if a recurrent set K_θ is not well matched, it may be possible to construct the same set using a different endomorphism γ to which the set is well matched. When we say " K_θ is well matched" we shall always mean "well matched to θ".

We shall now assume that L is a similitude with eigenvalue λ, $|\lambda|>1$. The dimension estimate (*) can be written

$$\frac{\log \lambda_E}{\log \lambda} = \frac{h(\sigma_\theta)}{\log \lambda}$$

Thrm 2.14 The following are equivalent,

i) K_θ is resolvable,

ii) K_θ is well matched,

iii) π is bounded to one.

If these conditions hold, π is 1-1 and $\tilde{L}$ is defined on a residual set in K_θ.

Proof: ii) $\Rightarrow$ iii) This is just (Bo4, p23).

iii) $\Rightarrow$ ii) Suppose not. Then there are essential symbols s,a,b such that

$$\text{int}\left(\pi(C_1(s;(s,a,j))) \cap \pi(C_1(s;(s,b,k)))\right) \neq \emptyset$$

for some j,k. Any point in this intersection can be represented by a sequence in Σ_θ beginning with (s;(s,a,j)...) or (s;(s,b,k)...). Since θ is essentially mixing there is an $M>0$ such that

$$C_M(s;(s,b,k),\ldots,(\cdot,s,\cdot))$$

is mapped into the above intersection. Hence for any point x in

$$\operatorname{int}\left(\pi(C_{M+1}(s;(s,b,k),\ldots,(s,a,j)))\right.$$
$$\left.\cap \operatorname{int} \pi(C_{M+1}(s;(s,b,k),\ldots,(s,b,k)))\right),$$

we have now found three sequences of length M+1 that begin infinite sequences of Σ_θ mapping to x under π.

Applying this construction inductively, by considering cylinders upto length NM+1, we have found N different sequences in Σ_θ that map to a point x in K_θ. Thus π cannot be bounded to one.

i) $\Rightarrow$ ii) Suppose not. From the proof of iii) $\Rightarrow$ ii), and without loss of generality replacing θ^M by θ we may assume that there are essential symbols t, s_i with

$$\theta(t) = s_1 \ldots s_r \qquad \text{and} \qquad \bigcup_{j \neq i} \tilde{R}(t,s_j) \supset \tilde{R}(t,s_i).$$

We now 'virtualize' s_i (cf. D3 6.1). Define a new alphabet S' by $S' = S \cup \{\bar{s} : s \in S\}$. $\theta' : S'^* \to S'^*$ is defined by

$$\theta'(s) = \theta(s) \quad \text{for } s \in S, s \neq t,$$
$$\theta'(\bar{s}) = \overline{\theta(s)} \quad \text{and}$$
$$\theta'(t) = s_1 \ldots s_{i-1} \bar{s}_i s_{i+1} \ldots s_r .$$

Then $K_{\theta'} = K_\theta$ but $\lambda_{E'} < \lambda_E$ and so

$$\frac{m\left((K[\theta^j E])^\varepsilon\right)}{|\theta^j E|_E} \leqslant \text{const.}\left(\frac{\lambda_{E'}}{\lambda_E}\right)^j \longrightarrow 0$$

as $j \to \infty$. Thus K_θ is not resolvable.

ii) $\Rightarrow$ i) We may take $K[s] = K_\theta(s)$ for all essential s. No copy of $K_\theta(s)$ intersects a copy of $K_\theta(t)$ in its interior in $K_\theta(\theta^j v)$, any essential v. Thus given a small $\varepsilon > 0$ there is c, $1 > c > 0$ such that

$$m\left((K_\theta(\theta^j v))^\varepsilon\right) \geqslant c. \sum_{s \in \theta^j v} m\left((K_\theta(s))^\varepsilon\right)$$

Thus $$\frac{m\left((K_\theta(\theta^j v))^\varepsilon\right)}{|\theta^j v|_E} \geqslant \text{const.}\left(\frac{\lambda_E}{\lambda_E}\right)^j > 0$$

and K_θ is therefore resolvable.

We now assume that K_θ is resolvable. Define

$$\partial\mathcal{R}_\theta = \{x \in K_\theta : x \in L^{-1}(\tilde{R}(t,s_i) \cap \tilde{R}(t,s_j)),\ s_i, s_j \in \theta(t), t \in E\}$$

Then $\partial\mathcal{R}_\theta$ is nowhere dense, and the set

$$Y = K_\theta \setminus \bigcup_{n=0}^{\infty} \tilde{L}^{-n}\, \partial\mathcal{R}_\theta$$

on which π is 1-1 is thus residual. Since $\pi^{-1}Y$ is σ_θ-invariant we can define $\tilde{L}$ so that

$$\begin{array}{ccc} \pi^{-1}Y & \xrightarrow{\ \sigma\ } & \pi^{-1}Y \\ \pi \downarrow & & \downarrow \pi \\ Y & \xrightarrow{\ \tilde{L}\ } & Y \end{array}$$

commutes. (This also ensures that $\tilde{L}$, $\tilde{L}^{-1}$ are mutual inverses).

∎

<u>Remarks</u> (ii)$\Leftrightarrow$(iii) works without assuming that L is a similitude.

K_θ also satisfies Hutchinson's open set condition if there is a Markov partition.

Thrm. 2.15 If K_θ is not well matched there exists a countable sequence of recurrent sets K_{θ_n} such that

i) $K_{\theta_n} \subset K_{\theta_{n+1}}$ ii) $\bigcup_1^\infty K_{\theta_n} = K_\theta$

iii) K_{θ_n} has a θ_n-Markov partition

iv) $\dim K_\theta = \sup_n \dim K_{\theta_n}$

$$\leq \varliminf_{p\to\infty} \frac{(1/p) \log m((K[\theta^p E])^\varepsilon)}{\log \lambda} \quad \text{some small } \varepsilon > 0,$$

$$< \frac{h(\sigma_\theta)}{\log \lambda} = \frac{\log \lambda_E}{\log \lambda}$$

Proof: Since there is not a θ-Markov partition, we must have symbols s, s', s'' such that

$$\theta(s) = s_1 .. s' .. s_k s'' .. s_r \quad \text{and}$$

$$\text{int}\left(K_\theta(s') \cap (K_\theta(s'') + f(s'+..+s_k))\right) \neq \emptyset$$

We virtualize this occurence of s' as in the proof of 2.14. However, note that this time our new endomorphism may give us a (strict) subset of K_θ. Repeating the above for each symbol $s \in E$ (i.e. for each $s \in E$ check to see of an 'interior' intersection occurs as above - if it does, virtualize a symbol to remove the intersection).

After a finite number of steps we have arrived at an endomorphism θ_0 and a recurrent set K_{θ_0}. Clearly $K_{\theta_0} \subset K_\theta$ and K_{θ_0} has a θ_0-Markov partition (because we have removed all 'interior' intersections' from the essential symbols). We now gradually put back all the parts of K_θ that we removed. Inductively choose θ_n by comparing θ^2_{n-1} with θ^2 : If there is a symbol $t \in \theta^{2^n}(s)$ which has been virtualized in $\theta^2_{n-1}(s)$ but for which

$$\text{int}\left(\mathbb{K}_\theta(t) + f(W_1)\right) \cap \text{int}\left(\mathbb{K}_\theta(t') + f(W_1')\right) = \emptyset$$

in the induced topology for all $t' \in \theta^{2^n}(s)$, where $\theta^{2^n}(s) = W_1 t W_2 = W_1' t' W_2'$, unvirtualize t in defining θ_n. Then $\mathbb{K}_{\theta_n} \subset \mathbb{K}_{\theta_{n+1}}$ and as $n \to \infty$, $\mathbb{K}_{\theta_n} \nearrow \mathbb{K}_\theta$. Hence $\bigcup_n \mathbb{K}_{\theta_n} = \mathbb{K}_\theta$, $\dim \mathbb{K}_\theta = \sup_n \mathbb{K}_\theta$ and each $\mathbb{K}_{\theta_n}$ is well matched to θ_n.

The condition for resolvability says that

$$m\left((K[\theta^j S])^\varepsilon\right) \sim \text{const. } \lambda_E^j$$

Taking $K[s] = \mathbb{K}_\theta(s)$ we have $K[\theta^j S]^\varepsilon = L^j(\mathbb{K}_\theta)^\varepsilon$.

Applying this to each of the sets $\mathbb{K}_{\theta_n}$ gives

$$a_{j,n} = \frac{\log m\left((L^j \mathbb{K}_{\theta_n})^\varepsilon\right)}{\log \lambda^j} \longrightarrow \dim \mathbb{K}_{\theta_n}$$

for $j = l.2^n$ as $l \to \infty$ (L^{2^n} is the scaling map for $\mathbb{K}_{\theta_n}$).

Since $m\left((L^j \mathbb{K}_{\theta_n})^\varepsilon\right)$ is increasing in j we have

$$a_{j,n} \to \dim \mathbb{K}_{\theta_n} \quad \text{as } j \to \infty .$$

Now, $\mathbb{K}_{\theta_n} \nearrow \mathbb{K}_\theta$ so $L^j \mathbb{K}_{\theta_n} \nearrow L^j \mathbb{K}_\theta$. Thus as $n \to \infty$

$$(L^j \mathbb{K}_{\theta_n})^\varepsilon \nearrow (L^j \mathbb{K}_\theta)^\varepsilon \text{ and } m\left((L^j \mathbb{K}_{\theta_n})^\varepsilon\right) \nearrow m\left((L^j \mathbb{K}_\theta)^\varepsilon\right)$$

Hence $a_{j,n} \nearrow \dfrac{\log m\left((L^j \mathbb{K}_\theta)^\varepsilon\right)}{\log \lambda^j}$ as $n \to \infty$

and $\dim \mathbb{K}_\theta = \sup \dim \mathbb{K}_{\theta_n} \leqslant \underline{\lim}_{j \to \infty} \dfrac{\log m\left((L^j \mathbb{K}_\theta)^\varepsilon\right)}{\log \lambda^j}$

To see that $\overline{\lim_{j\to\infty}} \dfrac{\log m((L^j K_\theta)^\varepsilon)}{\log \lambda^j} < \dfrac{\log \lambda_E}{\log \lambda}$ = estimate(*)

apply the construction in the proof of i) $\Leftarrow$ ii) of the last theorem. This gives a new θ' such that $K_{\theta'} = K_\theta$ and $\lambda_{E'} < \lambda_E$. Hence

$$\dim K_\theta = \dim K_{\theta'} \leqslant \frac{\log \lambda_{E'}}{\log \lambda} < \frac{\log \lambda_E}{\log \lambda},$$

$$\dim K_\theta \leqslant \overline{\lim_{j\to\infty}} \frac{\log m((L^j K_\theta)^\varepsilon)}{\log \lambda^j} \leqslant \frac{\log \lambda_{E'}}{\log \lambda} < \frac{\log \lambda_E}{\log \lambda}$$

(cf D3 6.1) ∎

The above result proves half of the conjecture (2.9) about resolvability. We shall calculate the dimension of a recurrent set with a class of non-linear scaling maps replacing L. Dekking (D3) indicates how to go about the construction of such sets when L is replaced by a piecewise linear map. The general construction is not much more difficult.

<u>Thrm. 2.16</u> Let $\theta: S^* \to S^*$ be an endomorphism of a semigroup and for each $s \in S$, let $1s^{-1}: \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz with constant less than one and fixed point the origin. Suppose that $f: S^* \to \mathbb{R}^n$ satisfies

$$f(s) = \sum_{j=1}^{r} (1s_j)^{-1}(fs_j) \qquad \text{where } \theta(s)=s_1 \cdots s_r$$

for all $s \in S$. Then the transformation of $\prod_{s\in E} \mathcal{K}(\mathbb{R}^n_s)$ given by

$$A_s \longmapsto \bigcup_{\substack{j=1,\dots,r\\ s_j\in E}} \Big((1s_j)^{-1}(A_{s_j}) + \sum_{i<j} (1s_i)^{-1}(fs_i) \Big), \quad A_s \in \mathcal{K}(\mathbb{R}^n_s)$$

is a contraction map (using the Hausdorff metric on $\mathcal{K}(\mathbb{R}^n)$). The unique fixed set is a compact non-empty set (if $E \neq \emptyset$) denoted K_θ, and there is a continuous surjective map

$$\pi : \Sigma_\theta \to K_\theta \quad .$$

Proof: We shall ease our notation, by showing how we may assume that the subshift Σ_θ is defined by a matrix of zeros and ones. Suppose that $\theta(s) = W_1 t W_2 t W_3$ for some words W_1, W_2, W_3 where s and t are essential. Then define $S' = S \cup \{t'\}$, $E' = E \cup \{t'\}$ and $\theta' : S'^* \to S'^*$ by

$\theta'(u) = \theta(u), \quad u \in S, u \neq s$

$\theta'(s) = W_1 t W_2 t' W_3$ and

$\theta'(t') = \theta'(t)$.

Letting $lt' = lt$ and $f(t') = f(t)$, we have that $K_\theta(E)$ and $K_{\theta'}(E')$ differ only in that $K_{\theta'}(E')$ contains two copies of $K_\theta(t)$, namely $K_{\theta'}(t')$ and $K_{\theta'}(t)$. Proceeding in this fashion we can ensure that for each $s, t \in E$, t appears only once in $\theta(s)$. Now we can see that if A is the matrix so that $A : \mathbb{Z}^{|E|} \to \mathbb{Z}^{|E|}$ is the map induced by θ_{ab} then A is a zero-one matrix. Since $\Sigma_\theta = \Sigma(A)$, we can represent a point of Σ_θ by a sequence

$$(s_0, s_1, s_2, \ldots, s_i, \ldots) \qquad \text{where } s_i \in E \text{ and } s_i \in \theta(s_{i-1}).$$

During the rest of this proof, unless otherwise stated, $(s_0, \ldots, s_m \ldots)$ will be taken to be the point in Σ_θ where for $r \geq 0$, s_{m+r+1} is the first symbol of the word $\theta(s_{m+r})$. Let

$$D_m = \{(s_0, \ldots, s_m, \ldots) \in \Sigma_\theta\}, \qquad \bigcup_{m \geq 0} D_m = D.$$

D is clearly dense in Σ_θ, and we define $\pi : D \to \prod_{s\in E} \mathbb{R}^n_s$ inductively as follows,

$$D_0 : \pi(s_0 \ldots) = 0 \in \mathbb{R}^n_{s_0}$$

$$D_1 : \pi(s_0, t_i, \ldots) = \sum_{j<i} (lt_j)^{-1}(ft_j) \in \mathbb{R}^n_{s_0}$$

where $\theta(s_o) = t_1 \ldots t_r$.

$$D_m : \pi(s_0, \ldots, s_{m-1}, u_j, \ldots) = \pi(s_0, \ldots, s_{m-1}, \ldots) + (ls_1)^{-1} \ldots (ls_{m-1})^{-1}\left(\sum_{r<j}(lu_r)^{-1}(fu_r)\right) \in \mathbb{R}^n_{s_0},$$

where $\theta(s_{m-1}) = u_1 \ldots u_k$.

(Notice that if $s_{m-1}s'_{m-1}$ is a subword of $\theta(s_{m-2})$ then

$$\pi(s_0, \ldots, s'_{m-1} \ldots)$$

$$= \pi(s_0, \ldots, s_{m-1}, \ldots) + (ls_1)^{-1} \ldots (ls_{m-1})^{-1}\left(\sum_{r\leq k}(lu_r)^{-1}(fu_r)\right)$$

$$= \pi(s_0, \ldots, s_{m-1}, \ldots) + (ls_1)^{-1} \ldots (ls_{m-1})^{-1}(fs_{m-1})$$

This is easy to check using the definition of $\pi|D_{m-2}$.)

Claim: π is uniformly continuous on D. This is because there is an $M > 0$ so that

$$\left\| \sum_{j<i} (lu_j)^{-1}(fu_j) \right\| < M$$

for all i,s, where $\theta(s) = u_1 \ldots u_k$. Let $a < 1$ satisfy $a > \text{Lip}\,(ls)^{-1}$ for all $s \in S$. Then if $\underline{x}, \underline{y} \in \Sigma_\theta \cap D$ are in the same m-cylinder,

$$\| \pi\underline{x} - \pi\underline{y} \| \leq 2M \sum_{r=0}^{\infty} a^{m-1+r} \leq \text{const.}\, a^m$$

This proves the claim. Extend π continuously to a map

$$\pi : \Sigma_\theta \to \prod_{s\in E} \mathbb{R}^n_s$$

and define $K_\theta = \pi(\Sigma_\theta)$. In order to show that K_θ is fixed under the stated transformation we need the following lemma.

Lemma 2.17 Let $s \in E$ and σ_s be the right shift

$$\sigma_s : \bigcup_{t_i \in E,\ t_i \in \theta(s)} C_0(t_i) \rightarrow C_0(s).$$

Then if $\underline{x} = (t_i, s_1, s_2, \ldots, s_m, \ldots) \in D_{m+1}$, and

$\underline{y} = \sigma_s \underline{x} = (s, t_i, s_1, \ldots, s_m, \ldots) \in D_{m+2}$ where $\theta(t) = t_1 \ldots t_r$

we have $(1t_i)^{-1}(\pi\underline{x}) + \sum_{j<i} (1t_j)^{-1}(ft_j) = \pi\underline{y}$

Proof: By induction on m.

$\underline{m=0}$: $\pi\underline{y} = \sum_{j<i} (1t_j)^{-1}(ft_j) = (1t_i)^{-1}(\pi\underline{x}) + \sum_{j<i}(1t_j)^{-1}(ft_j)$

since $\pi\underline{x} = 0$.

$\underline{m=r}$: Let $\underline{z} = (t_i, s_1, \ldots, s_{r-1}, \ldots)$

$\underline{x} = (t_i, s_1, \ldots, s_r, \ldots)$

$\underline{z}' = (s, t_i, s_1, \ldots, s_{r-1}, \ldots)$

$\underline{y} = (s, t_i, s_1, \ldots, s_r, \ldots)$

Suppose that $\theta(s_{r-1}) = u_1 \ldots u_k$, with $u_k = s_r$. Then

$$\pi\underline{y} = \pi\underline{z}' + (1t_i)^{-1}(1s_1)^{-1} \ldots (1s_{r-1})^{-1}\left(\sum_{j<k}(1u_j)^{-1}(fu_j)\right)$$

$$= (1t_i)^{-1}(\pi\underline{z}) + \sum_{j<i}(1t_j)^{-1}(ft_j)$$

$$+ (1t_i)^{-1}(1s_1)^{-1} \ldots (1s_{r-1})^{-1}\left(\sum_{j<k}(1u_j)^{-1}(fu_j)\right)$$

(by our induction hypothesis)

$$= (1t_i)^{-1}(\pi\underline{x}) + \sum_{j<i}(1t_j)^{-1}(ft_j)$$

∎

The lemma tells us that under the transformation of subets of $\mathbb{R}^n$ given in the statement of 2.16,

$$\pi D_m \mapsto \pi D_{m+1}.$$

Since $D_m \subset D_{m+1}$ we have πD fixed. Hence K_θ is also fixed.

In order to prove the claim that any collection of non-empty compact sets $\{A_s : s \in E\}$ converges under iterations of the transformation to K_θ we consider the space $X = \prod_{s\in E} \mathcal{K}(\mathbb{R}^n_s)$. Giving $\mathcal{K}(\mathbb{R}^n)$ the Hausdorff metric d, we define a metric on X by

$$\underline{d}(\ (A_s)_{s\in E},\ (B_s)_{s\in E}\ ) = \max_{s\in E} d(A_s, B_s).$$

Then X is a complete metric space. If $\theta(s) = t_1 .. t_r$, then

$$\underline{d}\Big(\Big(\bigcup_{t_i\in E}\big((lt_i)^{-1}A_{t_i} + \sum_{j<i}(lt_j)^{-1}(ft_j)\big)\Big)_s\ ,$$

$$\Big(\bigcup_{t_i\in E}\big((lt_i)^{-1}B_{t_i} + \sum_{j<i}(lt_j)^{-1}(ft_j)\big)\Big)_s\Big)$$

$$\leq \max_{s\in E}\ \max\Big\{d(\ (lt_i)^{-1}A_{t_i},\ (lt_i)^{-1}B_{t_i}\ )\ :\ t_i\in E,\ t_i\in\theta(s)\Big\}$$

$$\leq \max_{s\in E}\ \max_{t_i\in E}\ a.d(A_{t_i}, B_{t_i}) \qquad \text{since Lip } (lt_i)^{-1} < a < 1$$

$$\leq a.\underline{d}((A_s), (B_s)) \ .$$

Thus our transformation is a contraction mapping with $K_\theta = (K_\theta(s))_{s\in E}$ as the unique fixed point.

■

<u>Remark</u>: If θ is essentially mixing then any collection of compact sets $\{(A_s) : s\in E\}$ where at least one $A_s \neq \emptyset$ will converge to K_θ.

We now define the property of being well matched for the non-linear scaling case.

Def. 2.18 A recurrent set K_θ with scaling maps ls, $s \in S$, is well matched and has the θ-Markov partition $\mathcal{R}_\theta = \{ K_\theta(s) : s \in E \}$ if whenever $\theta(s) = t_1 \dots t_r$, writing

$$\widetilde{R}(s,t_i) = (lt_i)^{-1} K_\theta(t_i) + \sum_{j<i} (lt_j)^{-1}(ft_j) \quad \text{if } t_i \in E$$

we have

$$\text{int}\left(\widetilde{R}(s,t_i) \cap \widetilde{R}(s,t_j) \right) = \emptyset \qquad i \neq j$$

in the induced topology as a subset of $K_\theta(s)$.

The following theorem is based on (Bo 6) and also uses ideas from (Mo),(Hut). The author has recently become aware of a similar result of Ruelle(Ru). In the following, for a matrix A with $|Ax| = r|x|, \forall x$, we shall let $|A| = r$.

Thrm 2.19 Let K_θ be a well matched recurrent set for which θ is essentially mixing. Suppose that for each $s \in E$ there is an open convex set $U(s) \subset \mathbb{R}^n_s$ with $K_\theta(s) \subset U(s)$. Suppose also that

i) There is a $b > 1$ independent of s and x such that $D(ls)_x$ is a similitude and $|D(ls)_x| > b > 1$ for $x \in ls^{-1}U(s)$.

ii) ls^{-1} is C^1 on $U(s)$ and $|D(ls^{-1})_x|$ is Lipschitz on $U(s)$.

iii) ls is C^1 and $|D(ls)_x|$ is Lipschitz on $ls^{-1}U(s)$.

iv) If $\theta(s) = t_1 \dots t_m$ then

$$lt_i^{-1}U(t_i) + \sum_{j<i} lt_j^{-1}(ft_j) \subset U(s), \quad \forall t_i \in E.$$

Then if we define $\varphi : \Sigma_\theta \to \mathbb{R}$ by

$$\varphi(\underline{x}) = \log \left| D(ls)^{-1}_{\pi\sigma x} \right| \quad \text{when } \underline{x} = (\cdot, s, \ldots)$$

there is a unique $a > 0$ so that $P(a\varphi) = 0$. We have $a = \dim K_\theta$, and $HM_a(K_\theta)$ is equivalent to $\pi_* \mu$ where μ is the equilibrium state for $a\varphi$.

<u>Proof</u>: Let $(s_0, \ldots, s_m)$ be a sequence of symbols with

$$\theta(s_{i-1}) = s^1_{i-1} \ldots s^{r_{i-1}}_{i-1} \quad \text{and} \quad s^{n_{i-1}}_{i-1} = s_i \; \forall i, \; 1 \le n_{i-1} \le r_{i-1}.$$

We define maps $l_1, \ldots, l_m$ for $s_1, \ldots, s_m$ by letting

$$l_i^{-1}(x) = ls_i^{-1}(x) + \sum_{j < n(i-1)} (ls^j_{i-1})^{-1}(fs^j_{i-1}) \quad \text{(fig 7)}.$$

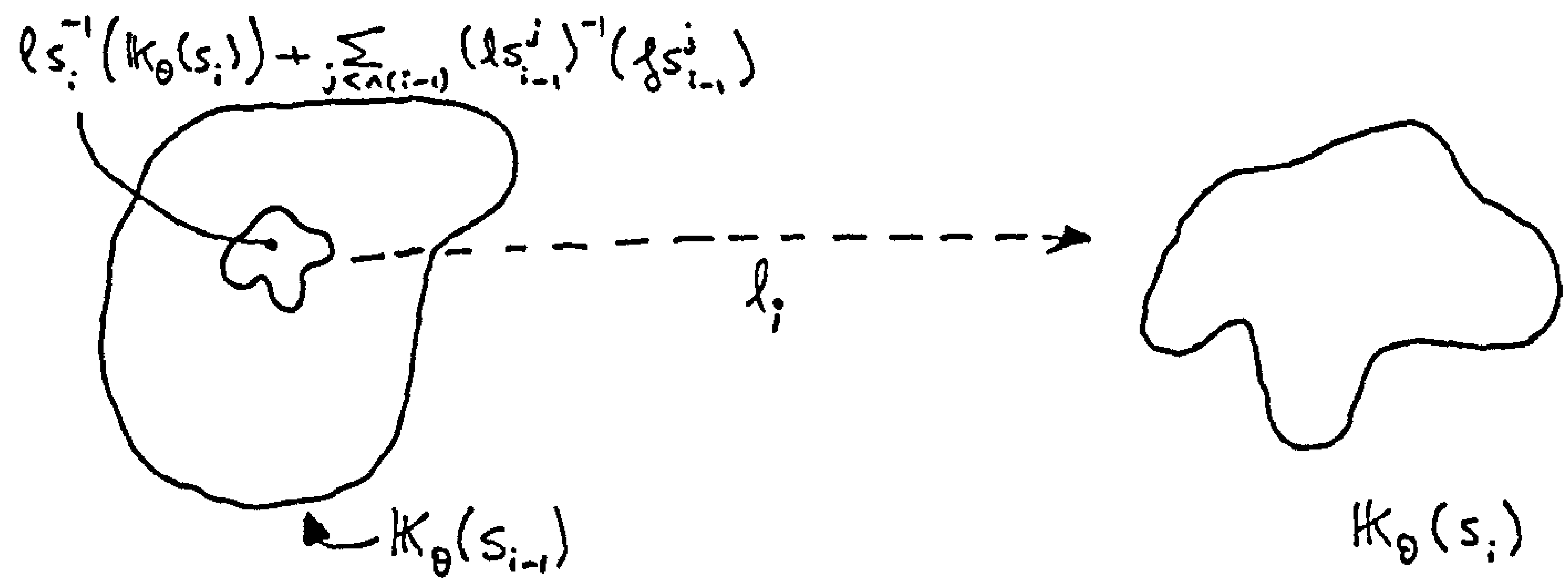


FIG. 7.

We thus have maps so that

$$\begin{array}{ccc} C_1(s_{i-1}, s_i, \ldots) & \xrightarrow{\sigma_\theta} & C_0(s_i, \ldots) \\ \downarrow \pi & & \downarrow \pi \\ \pi C_1(s_{i-1}, s_i, \ldots) & \xrightarrow{l_i} & K_\theta(s_i) \end{array}$$

commutes (l_i is only defined on a residual set).

Define $U(s_0, \ldots, s_m) = l_1^{-1} \ldots l_m^{-1}(U(s_m))$

and $\varphi_1 : U(s_0, \ldots, s_m) \to \mathbb{R}$ by

$$\varphi_1(x) = \log \left| D(l_1^{-1})_{l_1 x} \right|$$

$$= -\log \left| D(l_1)_x \right| \quad \text{since } Dl_1 \text{ is a similitude.}$$

In particular note that $\varphi_1(\pi\underline{x}) = \varphi(x)$, where $\pi\underline{x}=x$. As $|D(l_i)_x|$ is Lipschitz, so is φ. Hence $a\varphi$ has a unique equilibrium state that is a Gibbs measure. The variational formula for pressure (Wa 1) implies that, as a function of a, $P(a\varphi)$ is continuous and strictly decreases as a increases. Since $P(0) = P(0\varphi) > 0$ and for large enough a, $P(a\varphi) < 0$, there is a unique a such that $P(a\varphi) = 0$.

We prove the theorem in a series of steps.

i) There is a constant $c_1 > 0$ so that if $x, y \in U(s_0, \ldots, s_m)$,

$$|\varphi_1(x) - \varphi_1(y)| \leq c_1 b^{-m}.$$

Proof: Let $x' = l_m \ldots l_1(x)$, $y' = l_m \ldots l_1(y)$. We can find $z' \in U(s_m)$ by the Mean Value Theorem such that

$$|x - y| \leq \left| D(l_1^{-1} \ldots l_m^{-1})_{z'} \right| \cdot |x' - y'|$$

$$\leq \left(\prod_1^m \sup \left\{ \left| D(ls_i)_{z_i}^{-1} \right| : z_i \in U(s_i) \right\} \right) \cdot |x' - y'|$$

$$\leq b^{-m} |x' - y'| .$$

Since φ_1 is Lipschitz and $x', y' \in U(s_m)$, we can find $c_1 > 0$ such that $|\varphi_1 x - \varphi_1 y| \leq c_1 b^{-m}$.

ii) There is a $d > 0$ independent of $s_0, \ldots, s_m$ and m such that for any $(s_0, \ldots, s_m, \ldots) \in \Sigma_\theta$, $U(s_0, \ldots, s_m)$ is contained in a ball of radius

$$d \cdot \exp\left(\sum_{r=0}^{m-1} \varphi\left(\sigma^r(s_0, \ldots, s_m, \ldots)\right) \right)$$

Proof: Choosing x,y etc as in step (i),

$$|x - y| \leqslant \prod_{r=1}^{m} \left|D(l_r)^{-1}_{(l_{r+1}^{-1}\cdots l_m^{-1}z')}\right| . |x' - y'|$$

$$\leqslant \prod_{r=1}^{m} \left|D(l_r)_{(l_r^{-1}\cdot\cdot l_m^{-1}z')}\right|^{-1} |x' - y'|$$

$$\leqslant \prod_{r=1}^{m} \left|D(l_r)_{(l_{r-1}\cdot\cdot l_1 z)}\right|^{-1} |x' - y'| ,$$

where $l_m\cdot\cdot l_1 z = z'$

$$= \exp\left(\sum_{r=1}^{m} \varphi_1(l_{r-1}\cdot\cdot l_1 z)\right) . |x' - y'|$$

Now, if $w,z \in U(s_0,\dots,s_m)$ then

$$l_{r-1}\cdot\cdot l_1(w) \quad , \quad l_{r-1}\cdot\cdot l_1(z) \in U(s_r,\dots,s_m).$$

Hence,

$$\left| \sum_{r=1}^{m} \varphi_1(l_{r-1}\cdot\cdot l_1 w) - \sum_{r=1}^{m} \varphi_1(l_{r-1}\cdot\cdot l_1 z)\right|$$

$$\leqslant c_1 \sum_{r=1}^{m} b^{-(m-r)} < c_1 \sum_{j=0}^{\infty} b^{-j} < d_1 \text{ say, by step (i).}$$

Therefore

$$|x - y| < d.\exp\left(\sum_{r=0}^{m-1} \varphi(\sigma^r \underline{z})\right), \ \forall\, x,y \in U(s_0,\dots,s_m)$$

for any $\underline{z}$ with $\pi\underline{z} \in U(s_0,\dots;s_m)$.

iii) There is $d' > 0$ independent of $s_0,\dots,s_m$, m, such that for any $(s_0,\dots,s_m,\dots) \in \Sigma_\theta$, $U(s_0,\dots,s_m)$ contains a ball of radius

$$d'.\exp\left(\sum_{r=0}^{m-1} \varphi(\sigma^r(s_0,\dots,s_m,\dots))\right)$$

Proof: Choosing x,y etc as in (ii),

$$|x' - y'| \leqslant |D(l_m\cdot\cdot l_1)_z| . |x - y|$$

$$\leq \prod_{r=1}^{m} |D(l_r)(l_{r-1}..l_1 z)| . |x - y|$$

As in step (ii) there is $d_2 > 0$ such that

$$|x' - y'| \leq d_2 \exp\left(-\sum_{r=0}^{m-1} \varphi(\sigma^r(s_0,..s_m,..))\right).|x - y|$$

Thus we can find $d_3 > 0$ with

$$|x - y| \geq d_3 \exp\left(\sum_{r=0}^{m-1} \varphi(\sigma^r(s_0,...,s_m,..))\right).|x' - y'|$$

Let r' be such that every $U(s)$ contains a ball of radius r'.

Let B' be such a ball in $U(s_m)$, centre x', and let

$|x' - y'| = r'$. Then

$$|x - y| \geq r'd_3 \exp\left(\sum_{r=0}^{m-1} \varphi(\sigma^r(s_0,...,s_m,..))\right) .$$

Taking $d' < r'd_3$ implies that there is a ball of radius

$$d' \exp\left(\sum_{r=0}^{m-1} \varphi(\sigma^r(s_0,..s_m,..))\right)$$

with centre x in $U(s_0,..,s_m)$

iv) $\dim K_0 \leq a$

Proof: The equilibrium state, μ, for $a\varphi$ is unique and is a Gibbs measure and hence there is $c > 0$ such that

$$\mu(C_m(s_0,...,s_m)) \in [c^{-1},c] .\exp\left(\sum_{r=0}^{m-1} a\varphi(\sigma^r(s_0,...,s_m,.))\right)$$

For each m, $\mathcal{U}_m = \{U(s_0,...,s_m) : (s_0,...,s_m,...) \subset \Sigma_\theta\}$ covers K_θ and $|U(s_0,...,s_m)| \leq 2c_1 b^m$, so we can use $\mathcal{U}_m$ to give an estimate of Hausdorff measure,

$$\sum_{(s_0,...,s_m)} |U(s_0,...,s_m)|^a$$

$$\leq (2d)^a \sum_{(s_0,...,s_m)} \exp\left(\sum_{r=0}^{m-1} a\varphi(\sigma^r(s_0,...,s_m..))\right)$$

$$\leq c.(2d)^a \sum_{(s_0,...,s_m)} (\pi_*\mu)(\pi C_m(s_0,...,s_m))$$

$\leqslant \; c.(2d)^a \; < \; \infty$.

This argument shows that dim $K_\theta \leqslant a$, and (if we consider only sequences $(s_0,\ldots,s_m\ldots)$ beginning with $(s_0,\ldots,s_j)$) that

$$HM_a(\pi C_j(s_0,\ldots,s_j)) \leqslant \text{const}.\pi_*\mu(\pi C_j(s_0,\ldots,s_j)) \quad .$$

v) dim $K_\theta \geqslant a$

Proof: Given $t > 0$, we estimate the $\pi_*\mu$ measure of a ball of radius t. For each $\underline{x} \in \Sigma_\theta$ choose m such that

$$|U(s_0,\ldots,s_m)| \leqslant t \leqslant |U(s_0,\ldots,s_{m-1})| \quad , \quad \underline{x}=(s_0,\ldots,s_m,\ldots).$$

Let I be the set of (finite) admissible sequences satisfying both these inequalities. From steps (ii) and (iii) we have

$$d' \exp(\textstyle\sum_0^{m-1}\varphi\sigma^r\underline{x}) \leqslant t \quad ,$$

$$d \exp(\textstyle\sum_0^{m-2}\varphi\sigma^r\underline{x}) \geqslant t$$

Hence each $U(s_0,\ldots,s_m)$ with $(s_0,\ldots,s_m) \in I$ is contained in a ball of radius

$$d \exp(\textstyle\sum_0^{m-1}\varphi\sigma^r\underline{x}) \leqslant (d/d')\, t \quad = \quad c.t$$

and contains a ball of radius

$$\tfrac{1}{2}d' \exp(\textstyle\sum_0^{m-1}\varphi\sigma^r\underline{x}) \geqslant \tfrac{1}{2}(d'/d)\, e^{\varphi\sigma^{m-1}\underline{x}}\, d\exp(\textstyle\sum_0^{m-2}\sigma^r\underline{x})$$

$$\geqslant \tfrac{1}{2}(d'/d)\, e^{-\|\varphi\|}.t \quad = \quad c'.t$$

By lemma 5.3.1(a) of (Hut), at most

$$\left(\frac{1+2c}{c'}\right)^n \quad = \quad c''$$

of the $U(s_0,\ldots,s_m)$ with $(s_0,\ldots,s_m) \in I$ can meet a ball of radius t.

$$\pi_*\mu(\pi C_m(s_0,\ldots,s_m)) \leqslant (c/d')^a |U(s_0,\ldots,s_{m-1})|^a$$

so letting B be a ball of radius t we have

$$\pi_* \mu(B) \leqslant \sum_{(s_0,\ldots,s_m)\in I} \pi_* \mu(\pi C_m(s_0,\ldots,s_m))$$

$$\leqslant \sum_{(s_0,\ldots,s_m)\in I} (c/d'^a) \, |U(s_0,\ldots,s_m)|^a$$

$$\leqslant (c''c' \ \ /d'^a) \cdot t^a$$

Hence by Frostman's lemma, $\dim(K_\theta) > a$. We have also shown in steps (iv) and (v) that $HM_a|_{K_\theta}$ is equivalent to $\pi_* \mu$.

■

Corollary 2.20 If K_θ is as above, with $ls = L$ for all s, where L is a linear map with eigenvalue of modulus λ,

$$\dim K_\theta = h(\sigma_\theta)/\log \lambda = \text{estimate } (*)$$

and HM_a is equivalent to $\pi_* \mu$ where μ is the measure of maximal entropy.

Proof: Use the Variational principle (Wa.1).

■

This completes our proof of the conjecture of Dekking on resolvability. His proof of dimension when $f(s) \in \mathbb{Z}^n$ was by finding a measure to use with Frostman's lemma. This measure was constructed as a weak limit by averaging Dirac measures spread evenly across K_θ. Thus it is really the same as the measure of maximal entropy.

Corollary 2.21 Suppose K_θ is as in 2.19 and each ls is a similitude with $|ls| = \lambda_s$. Write

$$\Lambda_a = A . \begin{pmatrix} \lambda_{s_1}^{-a} & & 0 \\ & \ddots & \\ 0 & & \lambda_{s_r}^{-a} \end{pmatrix}$$

where $S = \{s_1 \cdots s_r\}$, $\Sigma_\theta = \Sigma(A)$. Then dim K_θ is the unique real, a, such that the maximal eigenvalue of Λ_a is one.

Proof: Extend Σ_θ to the two-sided shift space $\underline{\Sigma}_\theta$. Suppose $\underline{x} = (x_i) \in \underline{\Sigma}_\theta$. Then $\varphi(\underline{x}) = -\log \lambda_{x_1}$. Define $\phi(\underline{x}) = \varphi \circ \sigma^{-1}(\underline{x})$. Since the equilibrium state, μ, for is shift invariant,

$$\int \varphi \, d\mu = \int \phi \, d\mu$$

In particular $P(a\varphi) = 0 \Leftrightarrow P(a\phi) = 0$. Applying Lemma 4.7 of (Wa 1) gives the required answer. ∎

CHAPTER THREE

CONSTRUCTING MARKOV PARTITIONS AND INVARIANT SETS.

In this Chapter we use recurrent sets to construct special Markov partitions for certain maps of tori. We study the coding time between different partitions for the same map in simple cases, and see how the mean coding time depends upon the semigroup endomorphism used to generate the Markov boundary. We also generate fractal invariant subsets for these maps.

§1: Special Markov partitions.

Hyperbolic automorphisms of the torus have long been studied as examples of maps showing chaotic behaviour. A description of their dynamics was obtained when Markov partitions were constructed for them (Ad1, Si1, Si2). We know then that one can carry out the construction given in Chapter one to give a subshift of finite type $\Sigma(B)$ and a map π such that the following diagram commutes.

$$\begin{array}{ccc} \Sigma(B) & \xrightarrow{\sigma_B} & \Sigma(B) \\ \pi\downarrow & & \downarrow\pi \\ T^n & \xrightarrow{\tilde{A}} & T^n \end{array}$$

Unfortunately it is not known what matrices B can occur for a given $\tilde{A}$ (although there are restrictions, for instance $h(\tilde{A}) = h(\sigma_B)$). This question has been the motivation for much work on subshifts. Here we approach the problem from a different angle and show that for certain hyperbolic automorphisms of T^3 (and expanding endomorphisms of T^2) one can find Markov partitions so that B is obtained directly from A. The construction of such partitions involves the use of fractals. Bowen (B5) showed that the boundaries of a Markov partition $\mathcal{R}$ for a hyperbolic automorphism $\tilde{A}$ of T^3 cannot be smooth submanifolds with boundary. Furthermore, if A has eigenvalues $|\lambda_1| > |\lambda_2| > 1 > |\lambda_3|$ then

$$\overline{\text{cap}}(\partial^s\mathcal{R} \cap \partial^u\mathcal{R}) \geq 2 - \frac{\log|\lambda_2|}{\log|\lambda_1|}$$

where for $R \in \mathcal{R}$, $\partial^s R = \{x \in R : x \notin \text{int}W^u(x,R)\}$

$\partial^u R = \{x \in R : x \notin \text{int}W^s(x,R)\}$, and

$\partial^s\mathcal{R} = \bigcup_{R\in\mathcal{R}} \partial^s R$, $\partial^u\mathcal{R} = \bigcup_{R\in\mathcal{R}} \partial^u R$. This is because the Markov conditions imply that $\tilde{A}(\partial^u\mathcal{R}) \subset \partial^u\mathcal{R}$ and $\tilde{A}^{-1}(\partial^s\mathcal{R}) \subset \partial^s\mathcal{R}$. Hence

$$\overline{O(\partial^s\mathcal{R} \cap \partial^u\mathcal{R})} \subset \partial^s\mathcal{R} \cup \partial^u\mathcal{R} \subsetneq T^3.$$

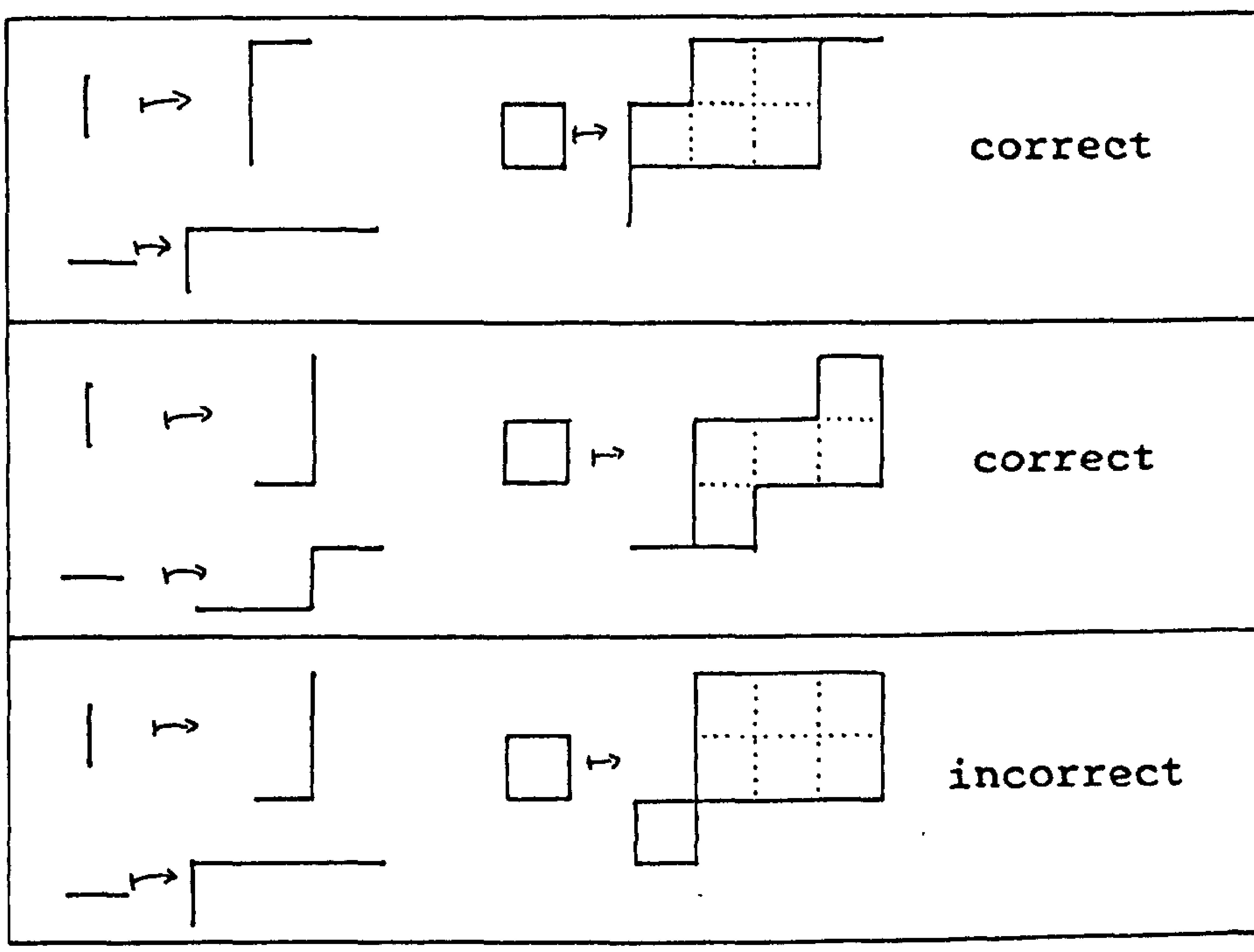
But a result of Urbański (Ub) says that if a curve C in T^3 has $\overline{\text{cap}}(C) < 2 - (\log|\lambda_2|)/(\log|\lambda_1|)$ then C has dense orbit. Urbański's result applies also to expanding endomorphisms of T^2 when the eigenvalues of the covering map are irrational.

More motivation for using fractals to generate Markov partitions is the idea of replicating 'fractiles' from (Ma1, p47) and (D2), both of which give tilings of the plane, $\mathcal{T}$, for which there is an expanding linear map $L_{\mathcal{T}}: \mathbb{R}^2 \to \mathbb{R}^2$ so that each tile of $L_{\mathcal{T}}(\mathcal{T})$ is exactly a union of tiles from $\mathcal{T}$. R. Mañé has told me that the following result is folklore.

<u>Thrm. 3.1</u> Let A be a 2x2 matrix of integers inducing an expanding endomorphism on T^2. Then there is a Markov partition for $\tilde{A}:T^2 \to T^2$ so that there is a semiconjugacy to $\tilde{A}$ from $\Sigma_{|\det A|}$, the full shift on $|\det A|$ symbols.

<u>Proof</u>: Let $e_1 = (1,0)$, $e_2 = (0,1)$ in $\mathbb{R}^2$. We make our construction in the covering plane using the recurrent set formalism. Let $S = \{s_1, s_2, s_1^{-1}, s_2^{-1}\}$ and set $f(s_i) = e_i$ $f(s_i^{-1}) = -e_i$, and $K[s] = [0, f(s)]$. We write $W = s_1 s_2 s_1^{-1} s_2^{-1}$ so $K[W]$ is the boundary of the unit square. The next step is to choose an appropriate endomorphism θ of S^* (note that $s_1 s_1^{-1}$ do not cancel in S^* because we are working in a semigroup). Given a line $l = [a,b]$ with endpoints $a,b \in \mathbb{Z}^2$ we define the <u>anticlockwise perturbation of l at a</u> as follows. Starting at $a = x_0$, choose inductively lattice points $x_i \in \mathbb{Z}^2$ such that $x_i = x_{i-1} \pm e_j$ $(j=1,2)$ and $[x_0, x_i]$ makes the smallest anticlockwise angle with l (in particular we might have $x_i \in l$, and we finish with $x_r = b \in l$). The sequence $a = x_0, x_1 \ldots x_r = b$ are the vertices of a polygonal line that we call the anticlockwise perturbation of l at a. Choose θ so that $K[\theta(s_i)]$ is the anticlockwise perturbation of $[0, A(e_i)]$ at 0, and define $\theta(s_i^{-1}) = t_r^{-1} \ldots t_1^{-1}$ if $\theta(s_i) = t_1 \ldots t_r$. Clearly the relation $Af = f\theta$ holds. Our choice of θ

means that the 'sides' of $K[\theta W]$ do not cross over and thus that the region bounded by $K[\theta W]$ has area $|\det A|$. In fact this property is all we require for the rest of the proof, and it is clear that there may be many different choices of θ with the required property (fig 8).



Different choices of θ for $\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$

FIG. 8

There is a natural orientation of line segments in $K[\theta^n W]$ given by symbol order in $\theta^n W$. This enables us to define the 'inside' of $K[\theta^n W]$ as all the points to the left of line segments having a single orientation defined. (We use this definition because $K[\theta^n W]$ may have multiple self intersections). Define V^n as the closure of the set of points inside $K[\theta^n W]$. We now define a transformation of certain subsets of the plane. Suppose R is the closure of the inside of $K[Y]$ (where $f(Y)=0$), then define $\underline{\theta}R$ to be the closure of the inside of $K[\theta Y]$. Extend the definition in the obvious way to finite unions of sets of the above form. In particular we have $\underline{\theta}V^n = V^{n+1}$. We will show that V^{n+1} is tiled by $|\det A|$ copies of V^n. Our choice of θ implies that V^1 can be tiled by $D = |\det A|$ copies of V^0,

$$V^1 = \bigcup_{i=1}^{D}(r_i + V^0), \quad r_i \in \mathbb{Z}^2, \quad \text{int}(r_i+V^0) \cap \text{int}(r_j+V^0)=\emptyset$$

if $i \neq j$, and that

$$\text{int}(V^1) \cap \text{int}(e_i+V^1) = \emptyset \quad i=1,2 .$$

We prove by induction that

$$V^n = \bigcup_1^D (A^{n-1}r_i + V^{n-1})$$

$$\text{int}(A^{n-1}r_i+V^{n-1}) \cap \text{int}(A^{n-1}r_j+V^{n-1}) \neq \emptyset \quad \Rightarrow \quad i=j$$

and $\quad \text{int}\, V^n \cap \text{int}(A^n e_i+V^n) = \emptyset \quad i=1,2.$ (fig. 9).

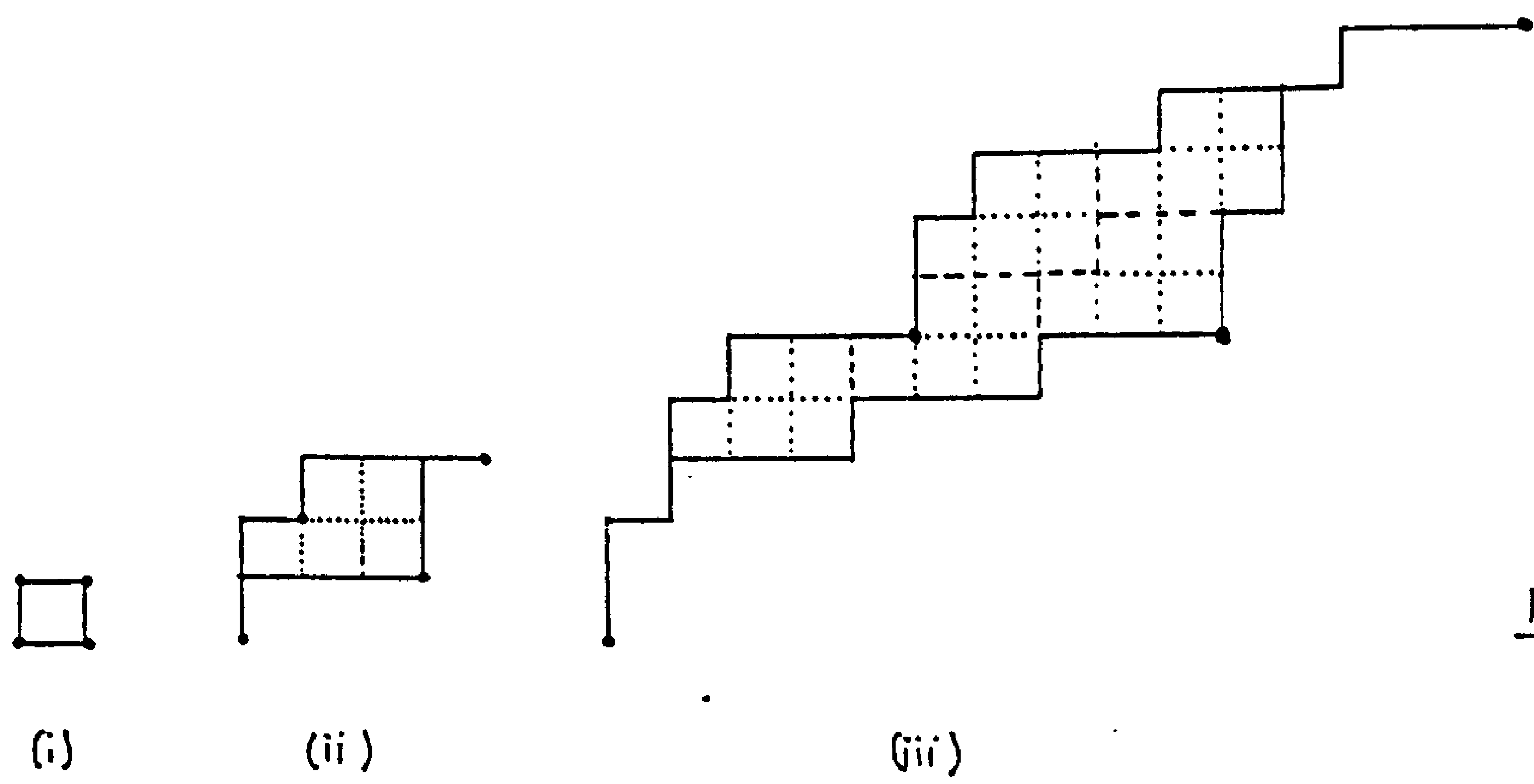


FIG. 9

K[θ^n W], n=0,1,2, for $\theta s_1 = s_2 s_1 s_1 s_1$, $\theta s_2 = s_2 s_2 s_1$. V^{n+1} is tiled by five copies of V^n.

Suppose the statement holds for $m \leq n$.

<u>Claim:</u> Let U be a finite union of tesselating V^{n-1}s, $U = \bigcup(p_i+V^{n-1})$, $p_i \in A^{n-1}\mathbb{Z}^2$. Then $\underline{\theta}U = \bigcup(q_i+V^n)$, $q_i=Ap_i$, is a union of tesselating V^ns.

Proof of claim: $\underline{\theta}(p_i+V^{n-1}) = q_i+V^n$, and the V^n tesselate since $\text{int}(A^n e_i+V^n) \cap \text{int}\, V^n = \emptyset$ by hypothesis. This proves the claim.

Set $U=V^n$, then $\underline{\theta}U=V^{n+1}$ so the claim proves the first part of the induction. Setting $U=(A^n e_i+V^n)$ we see that if $\text{int } V^{n+1} \cap \text{int}(A^{n+1}e_i+V^{n+1}) \neq \emptyset$ then the intersection contains $\text{int}(q_j+V^n)$ for some q_j. But this came from an r_j+V^{n-1}, which implies that

$$\text{int } V^n \cap \text{int}(A^n e_i+V^n) \supset \text{int}(r_j+V^{n-1}) \neq \emptyset,$$

a contradiction. This proves the induction.

Let $\underline{V}^n = A^{-n-1}V^n$. Then put $\mathcal{P}_n = \{q+\underline{V}^n : q \in A^{-1}\mathbb{Z}^2\}$. As $n\to\infty$, $A^{-n-1}K[\theta^n W] \to A^{-1}\,\mathbb{K}_\theta(W)$ and so the tilings $\mathcal{P}_n$ converge to a tiling $\mathcal{P}$. By construction $A(q+\underline{V}^n)$ is a union $\bigcup_{i=1}^{D}(r_i+\underline{V}^{n-1})$, $q,r_i \in A^{-1}\mathbb{Z}^2$, of sets with disjoint interiors for all n. Hence in the tiling $\mathcal{P}$,

$$A(q+\underline{V}) = \bigcup_{i=1}^{D}(r_i+\underline{V}) \qquad q,r_i \in A^{-1}\mathbb{Z}^2.$$

$\mathcal{P}$ is clearly invariant under integer translation and projects onto T^2 via the covering map, p, to give

$$\mathcal{R} = \{a+p(\underline{V}) : a \in \tilde{A}^{-1}(0)\}.$$

By construction, $R,R'\in\mathcal{R}$ and $\text{int}(R\cap R') \neq \emptyset$ implies $R=R'$. Furthermore each $\tilde{A}R$ is exactly the union $\bigcup_{R'\in\mathcal{R}} R'$. Hence $\mathcal{R}$ is a Markov partition. The standard construction now gives a semiconjugacy to $\tilde{A}$ from the full shift on $|\det A|$ symbols.

∎

<u>Remarks</u> i) Define $B = (b_{ij})_{2\times 2}$ by letting $b_{ij} = \#\{k : \theta(s_i) = t^1..t^l,\ t^k = s_j \text{ or } s_j^{-1}\}$. Then clearly $B \geq |A|$, i.e. $b_{ij} \geq |a_{ij}|$, $\forall i,j$. We shall prove in the final section of this Chapter that with the assumption that there is no essential symbol duplication,

$$\text{cap}(\mathbb{K}_\theta(W)) = 1 + \frac{\log\lambda_B - \log|\lambda_2|}{\log|\lambda_1|}$$

where λ_B is the maximal eigenvalue of B. In particular we note that by putting extra 'kinks' in our choice of θ we can construct Markov partitions for the same map with different boundary capacities. By Wielandt's Theorem (Ga p57) $\lambda_B \geqslant |\lambda_1|$. In the case of A being a non-negative matrix the choice of θ above defines a Markov partition with capacity equal to the minimum possible given by Urbański's Theorem.

ii) It is clear that there are often many choices for θ that will do. Thus we have a canonical class of Markov partitions rather than a single partition.

iii) Each of the approximating partition elements may be homeomorphic to discs, but the limiting boundary can still have multiple intersections. Thus the Markov partition elements are only homotopy equivalent to discs.

iv) If A has rational eigenvalues, the eigenspaces have rational slope. One can then obtain a Markov partition by using segments of the eigenspaces as the Markov boundary. Thus such a map could have Markov partitions with smooth or fractal boundaries.

In the next section we shall consider the question of which of the different Markov partitions one can construct as above is the best.

Expanding endomorphisms of T^2 are, geometrically, rather similar to hyperbolic automorphisms of T^3 with two-dimensional unstable manifolds. Thus the kind of construction used above to generate Markov partitions for expanding maps of T^2 can be used in the T^3 hyperbolic setting. We shall prove a generalization of the following unpublished result of Manning (Mn2).

<u>Thrm. 3.2</u> Let A be a hyperbolic matrix of positive integers and determinant +1. Then there is a semiconjugacy from $(\Sigma(A^t),\sigma)$ to the hyperbolic automorphism $(T^2,\tilde{A})$ (where t denotes transposition).

∎

<u>Thrm. 3.3</u> Let A be a hyperbolic 3x3 matrix of integers such that

i) det A = 1,

ii) A^{-1} is a non-negative matrix,

iii) top. dim.(E^s) = 1, top. dim.(E^u) = 2,

iv) the contracting eigenvalue of A is positive,

v) condition (‡) (defined below) is satisfied.

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

<u>Proof</u>: We proceed by stages.

a) First we establish the positions of the linear spaces E^s and E^u relative to the co-ordinate axes. Since A^{-1} is non-negative and $\tilde{A}$ has no non-trivial invariant subtori (by para. 1, p72), E^s lies in the interior of the positive cone, $\{(x_1,x_2,x_3): x_i \geqslant 0, \text{ all } i\}$, and E^u intersects the positive cone only at the origin.

b) We know that the Markov partition boundaries are not smooth (Bo5), so we shall show how to approximate the claimed partition arbitrarily well. In this first step we show how to define the first approximation. When $x \in E^s$ and $y \in E^u$ we shall move freely between writing $(x,y) \in E^s x E^u < \mathbb{R}^3$, $x \in E^s$, and $x \in \mathbb{R}^3$.

Define projection down stable manifolds $p_s: \mathbb{R}^3 \to E^u$ by $\{p_s(x)\} = (x+E^s) \cap E^u$. Let e_1,e_2,e_3 be the standard basis vectors. We define faces F_i $i=1,2,3$ by letting F_i be the square in $\mathbb{R}^3$ with vertices $0,e_j,e_k,e_j+e_k$ $(i \neq j,k)$. Projecting onto E^u gives us $t_i=p_s(e_i)$, $H_i=p_s(F_i)$ for $i=1,2,3$. Writing $F = \bigcup F_i$ and $H = \bigcup H_i$ notice that p_s maps F bijectively onto H. Hence we have a picture like fig 10.

Condition (I) is that $AH \supset H$.

Our first approximation to the Markov partition is given by three prisms P_i for $i=1,2,3$. Each P_i is the product of an interval of E^s with H_i and it lies in $\mathbb{R}^3$ so that its upper face is H_i. More precisely, let $a_i \in \mathbb{R}^3$ be such that $\{a_i\} = E^s \cap (-e_i+E^u)$ and define

$$P_i = [a_i, 0] \times H_i \subset E^s \times E^u = \mathbb{R}^3$$

(fig 11).

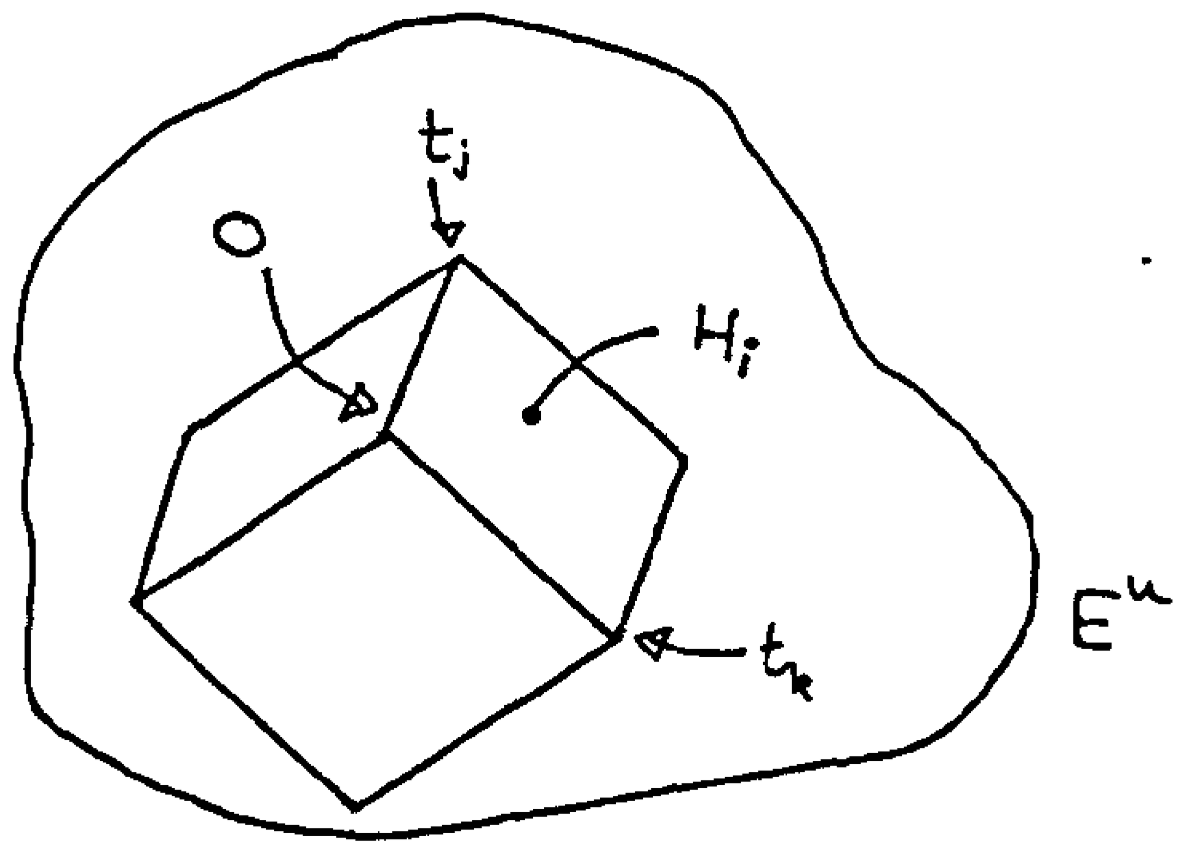


FIG. 10

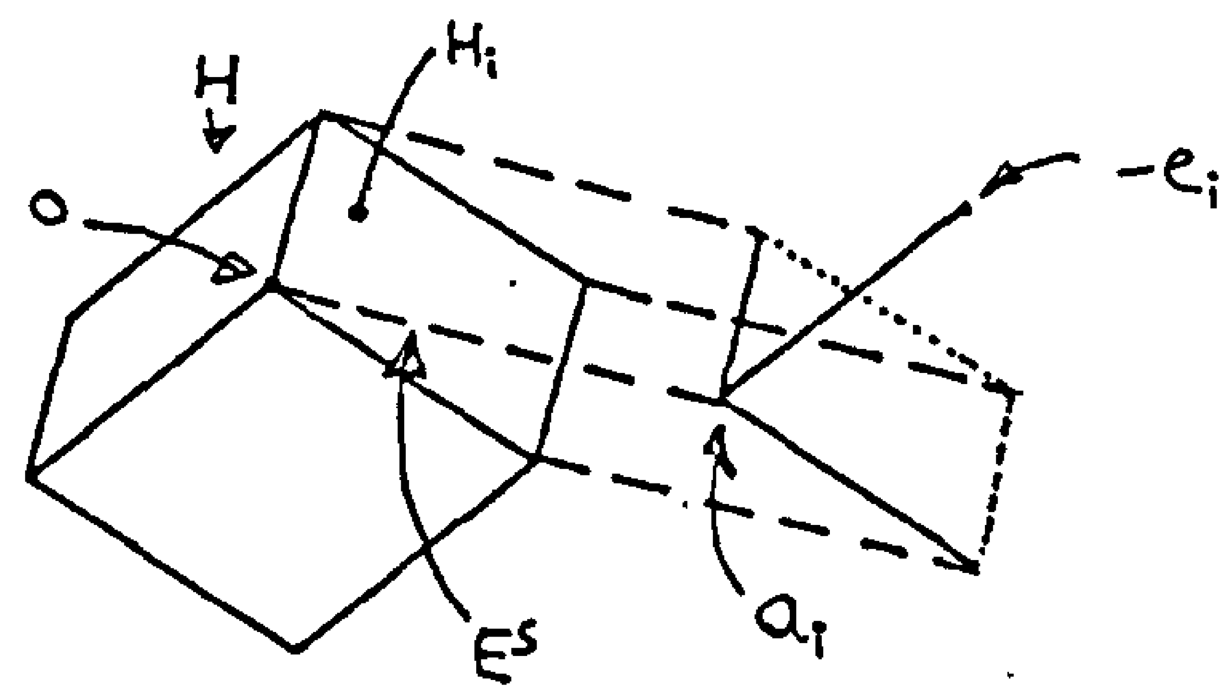


FIG. 11

Define $P = \bigcup_{i=1}^{3} P_i$. It is clear that $\text{int}(P_i)$ does not intersect $\text{int}(P_j)$ if $i \neq j$. We claim that

$$\mathcal{P} = \{q+P : q \in \mathbb{Z}^3\}$$

gives a tiling of $\mathbb{R}^3$ i.e. $\text{int}(q+P) \cap \text{int}(r+P) \neq \emptyset$ implies $r=q$, and each $x \in \mathbb{R}^3$ is a member of some $q+P$. The corresponding two dimensional picture is shown in fig 12.

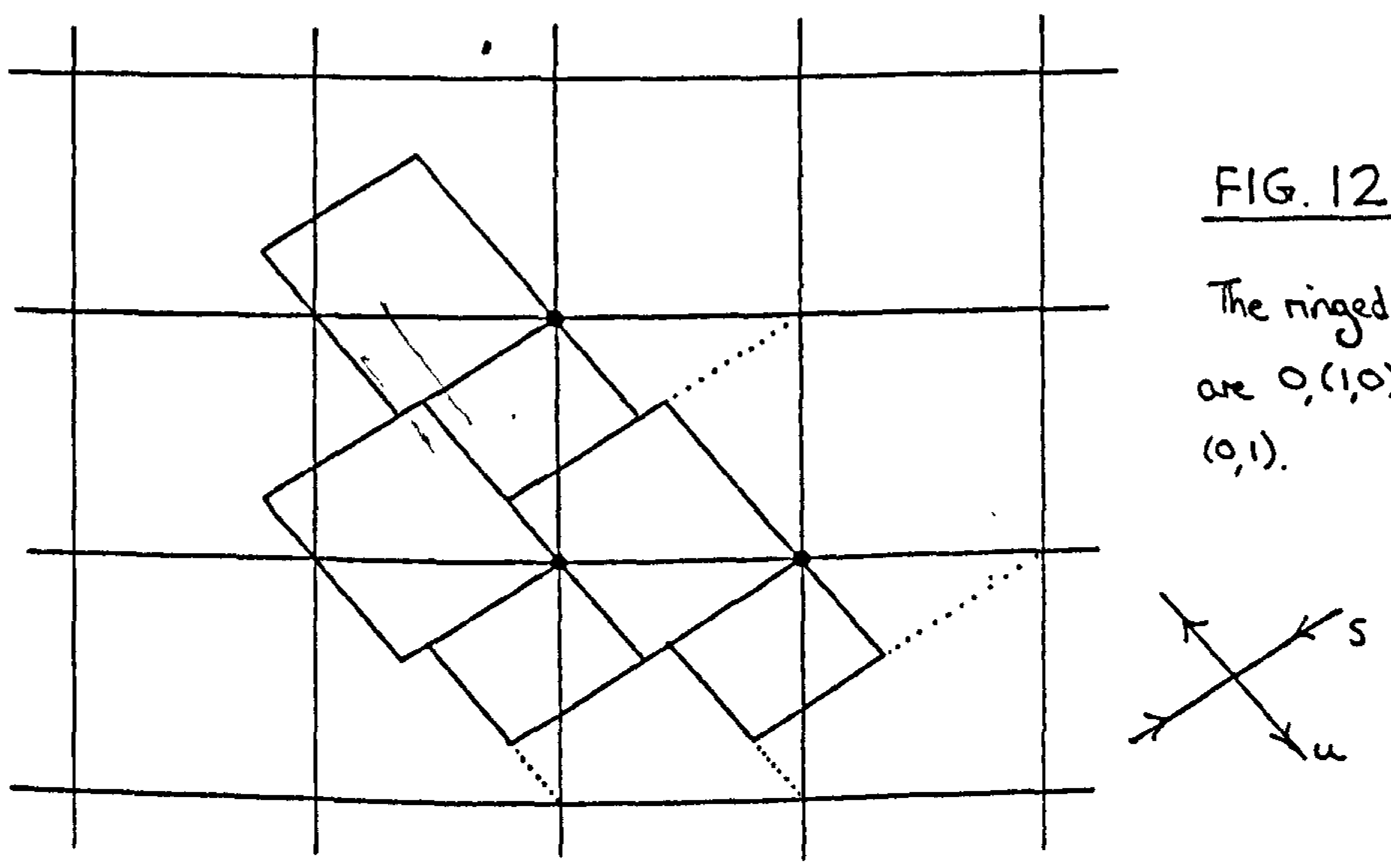


FIG. 12

The ringed vertices are 0, (1,0) and (0,1).

We call a face of q+P an <u>s-face</u> if it is a union of line segments parallel to E^s, or a <u>u-face</u> if it is parallel to E^u. All the faces of P are either s or u-faces. In order to show that $\mathcal{P}$ gives a tiling we have to show that P tesselates with its neighbours i.e. that each face of P contains or is contained in a face of a neighbour to P. We deal first with the u-faces of P.

The bottom face of P_i, a_i+H_i, (which is a u-face of P) satisfies $a_i+H_i \subset -e_i+H$ ($-e_i+H$ is the upper face of $-e_i+P$ a u-face of $-e_i+P$). For $\{t_i\} = (e_i+E^s) \cap E^u$ so $-e_i+t_i = a_i$ and $-e_i+(t_i+t_j) = a_i+t_j$. But a_i+t_j is a vertex of a_i+H_i and so three of the vertices of a_i+H_i are vertices of $-e_i+H$. Hence $a_i+H_i = (-e_i+t_i)+H_i \subset -e_i+H$, and we have the situation shown in figs 13 and 14.

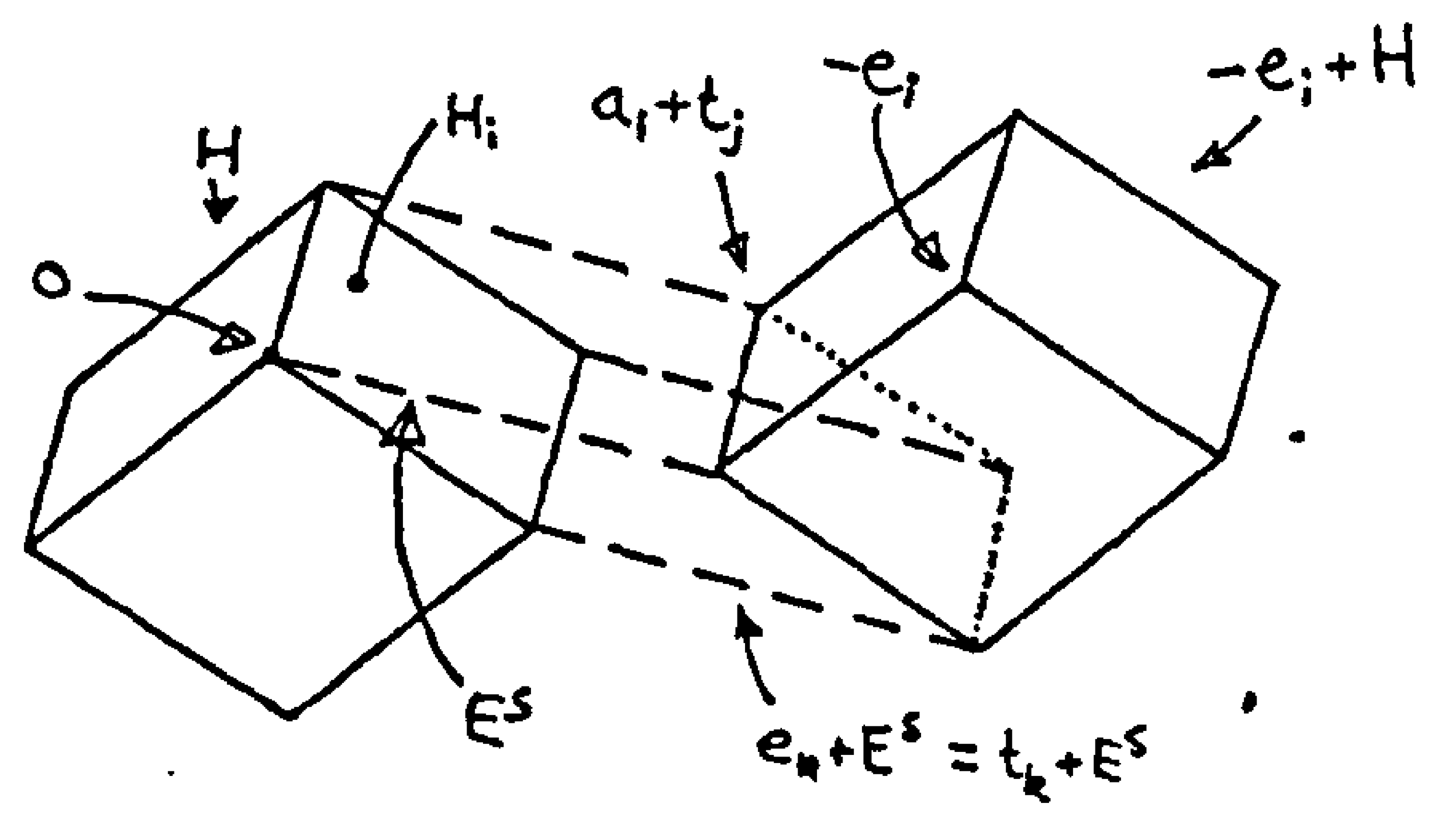


FIG. 13

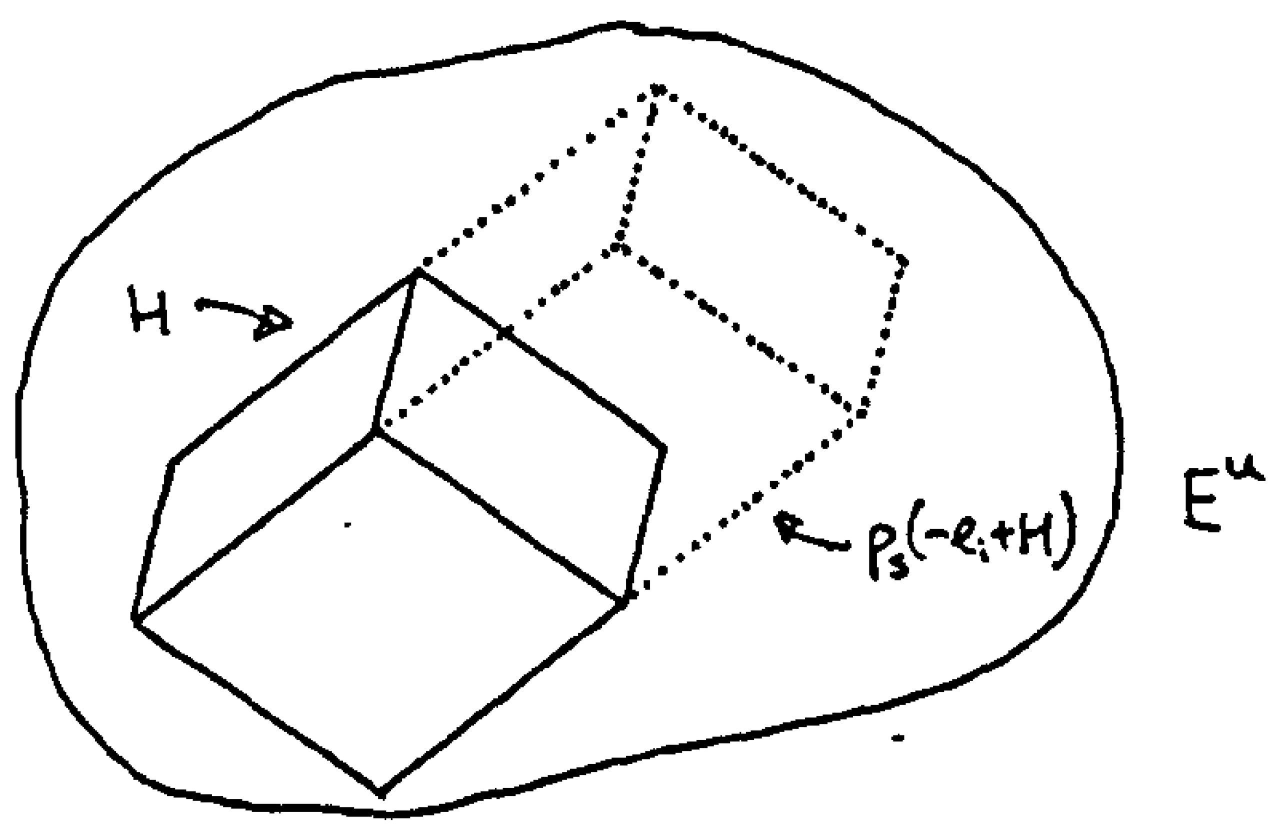


FIG. 14

Thus the bottom u-faces of P are subsets of (upper) u-faces of neighbours of P. The above argument also shows that the upper u-face of P, H, meets neighbours of P correctly, for

$$H = \bigcup_{i=1}^{3}(e_i + (a_i + H_i)) \qquad \text{(fig 15).}$$

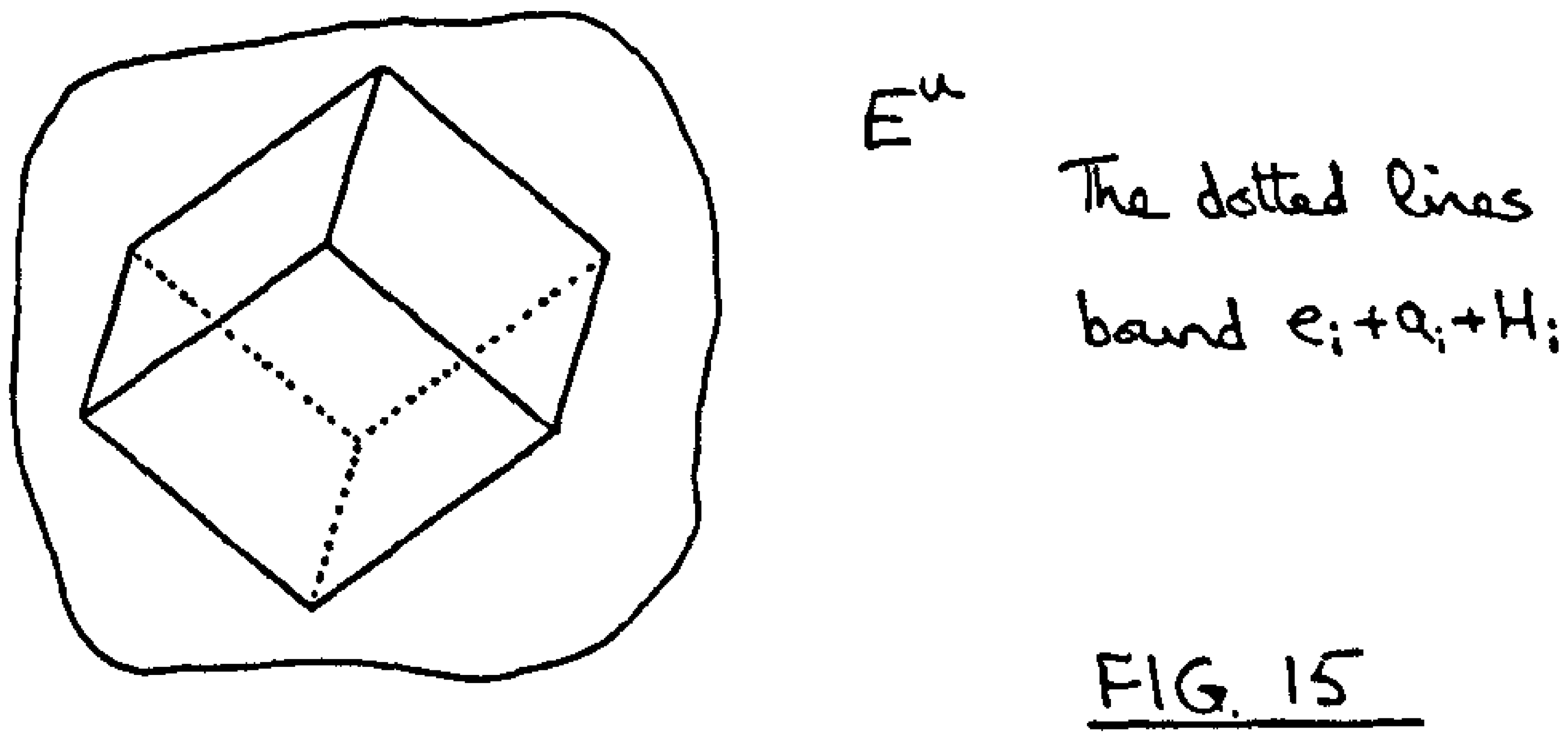


FIG. 15

We now have to deal with the s-faces of P. Since $p(E^u)$ is dense in T^3 it is enough to show that the intersection of E^u with $\{q+P_i: q \in \mathbb{Z}^3, i=1,2,3\}$ is made up of non-overlapping copies of H_i, i=1,2,3. However, using only the definition of P_i given above, the argument on pages 61-63 show this.

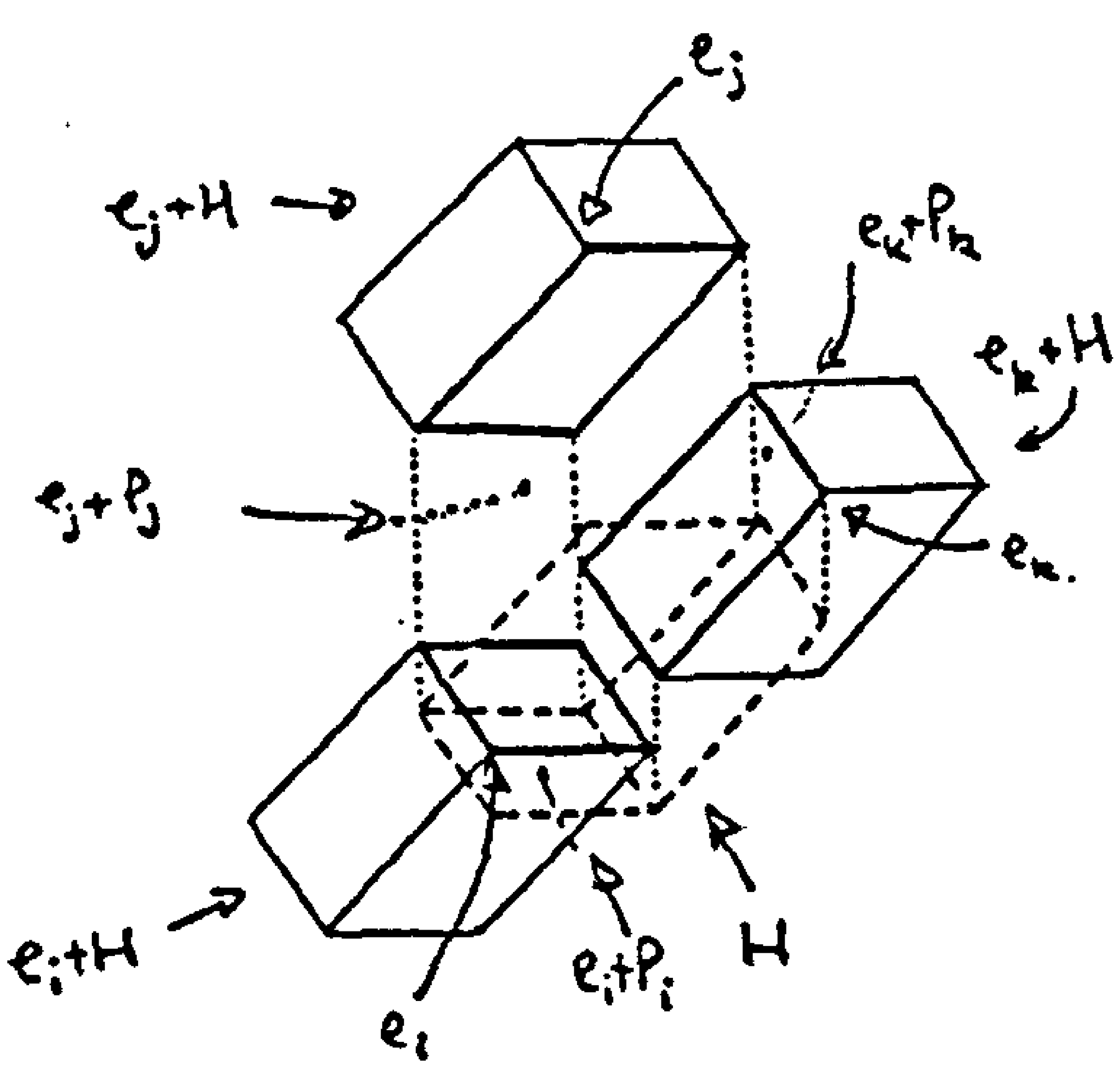


FIG 16a

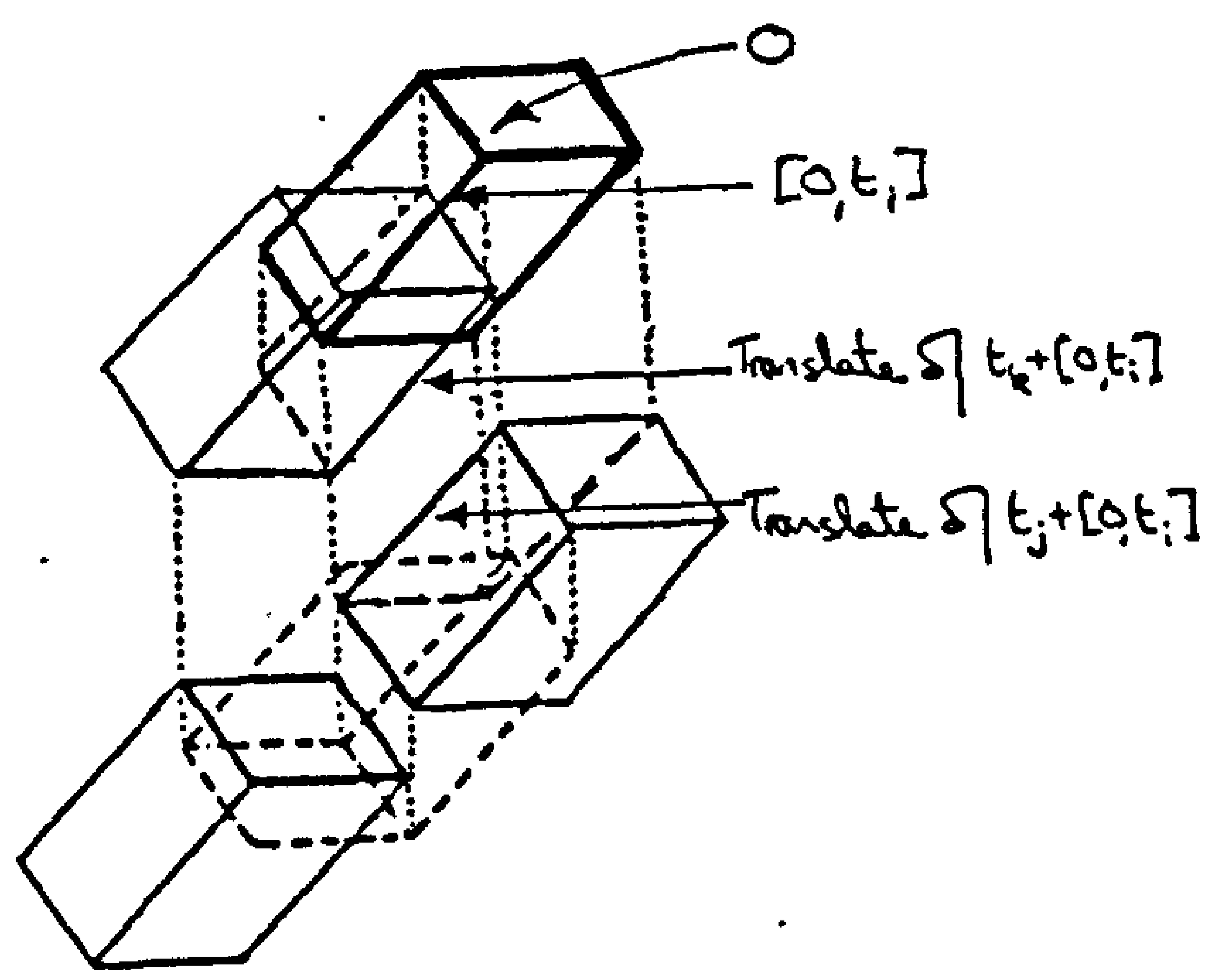


FIG 16b

This shows that $\mathcal{P}$ gives a tiling of $\mathbb{R}^3$. Figures 16a and 16b show how the various translates of P fit together in $\mathbb{R}^3$.

c) Closer approximations to the Markov partition. The boundaries of the P_i have to be altered to satisfy the Markov conditions. We introduce recurrent sets and choose an appropriate θ that will enable us to define closer approximations to the Markov partition.

Let V be a region of $x+E^u$. Write $Q_i = (a_i,0]\times\partial H_i$ and $\partial\mathcal{Q} = \bigcup_{q\in\mathbb{Z}^3}\bigcup_{i=1}^{3}(q+Q_i)$. Define the <u>patterning of V</u> as $\partial\mathcal{Q}\cap V$. The patterning of V is a tiling by copies of H_i. (In the final section of proof we shall see the geometrical significance of the patterning). We define the anticlockwise perturbation of a line in a similar fashion to the definition given in 3.1, replacing points of $\mathbb{Z}^2$ by vertices of the pattern in V. With vectors t_i we associate symbols s_i. Let $S = \{s_i, s_i^{-1} : i=1,2,3\}$, define $f(s_i) = t_i$, $f(s_i^{-1}) = -t_i$ and choose θ so that $K[\theta s_i]$ is the anticlockwise perturbation at 0 of $[0,At_i]$ (if $\theta s = r_1..r_1$ let $\theta(s^{-1}) = r_1^{-1}..r_1^{-1}$).

We now have to check two things. Firstly that $At_j+K[\theta s_i]$ lies in the pattern, and secondly that it is the anticlockwise perturbation at At_j of $At_j+[0,At_i]$ It is clear that these properties hold if the pattern around $At_j+[0,At_i]$ is locally the same as that around $[0,At_i]$. Recall that the pattern is defined by intersecting E^u with $\partial\mathcal{Q}$. Note that up to integer translation $[0,t_i]$, $t_j+[0,t_i]$ and $t_k+[0,t_i]$ all lie in the same s-face of either P_j or P_k (which one depends on whether $[a_j,0]\supset[a_k,0]$ or vice versa), fig 16b. We shall investigate the pattern near $At_j+[0,At_i]$ etc. by seeing how $\mathcal{P}$ intersects AP_k for each k.

Notice that we cannot have $q\in\mathbb{Z}^3$ such that

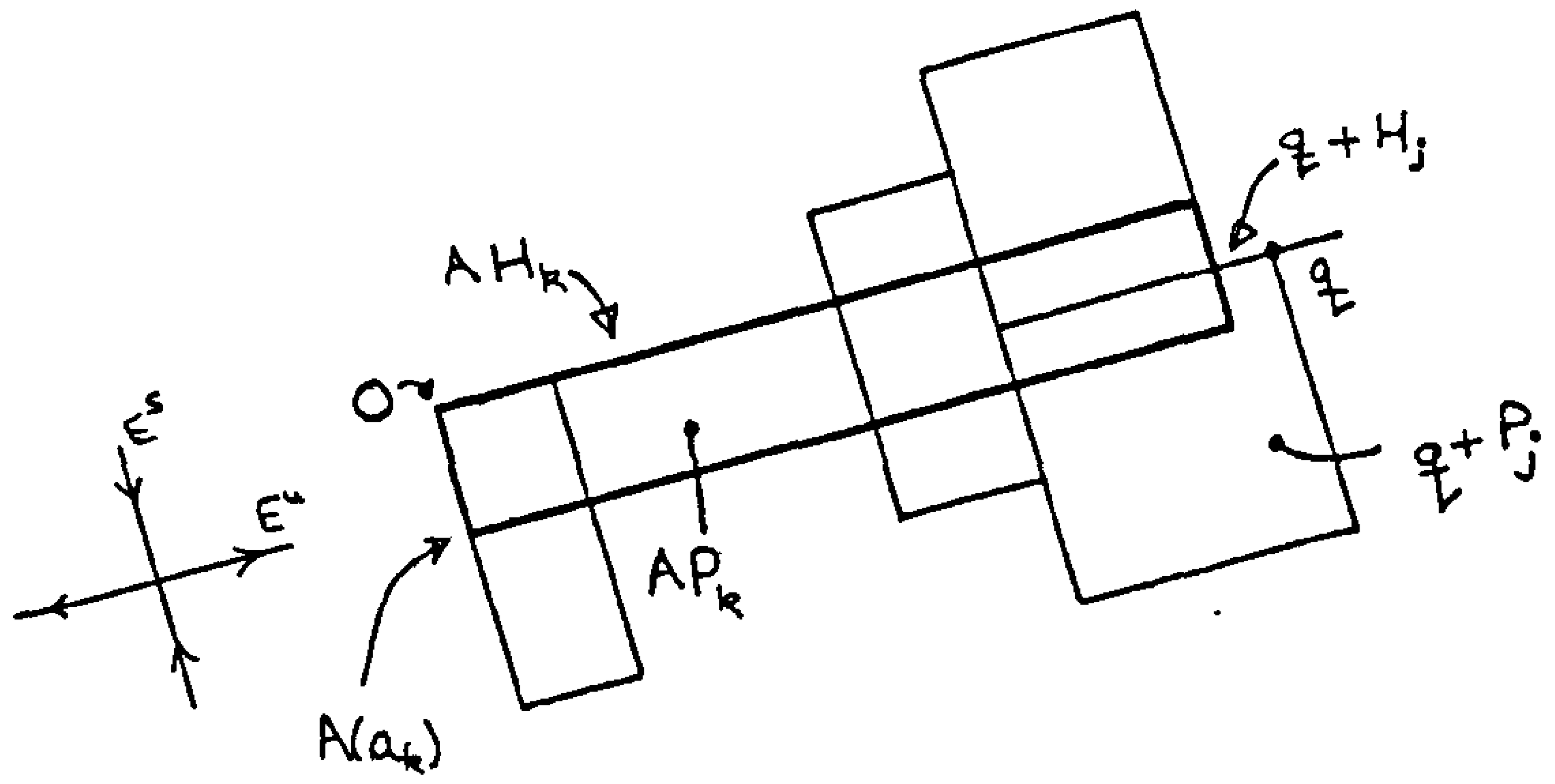


FIG. 17

$$(q+H_j) \cap \text{int}(AP_k) \neq \emptyset \qquad \text{(fig 17)}.$$

For since $A\mathbb{Z}^3 = \mathbb{Z}^3$, $A^{-1}q = r \in \mathbb{Z}^3$. By assumption $A^{-1}H \subset H$, so $A^{-1}(q+H_j) \subset r+H$. In particular there is a P_l such that

$$\text{int } A(r+P_l) \cap \text{int } AP_k \neq \emptyset.$$

But this implies $\text{int}(r+P_l) \cap \text{int } P_k \neq \emptyset$ which is a contradiction.

The above argument shows that if the lower u-faces of e_i+P_i for i=1,2,3 are included in the pattern, $At_j+K[\Theta s_i]$ etc lie in the pattern (we have to add to our original pattern because we have not excluded the possibility shown in fig 18).

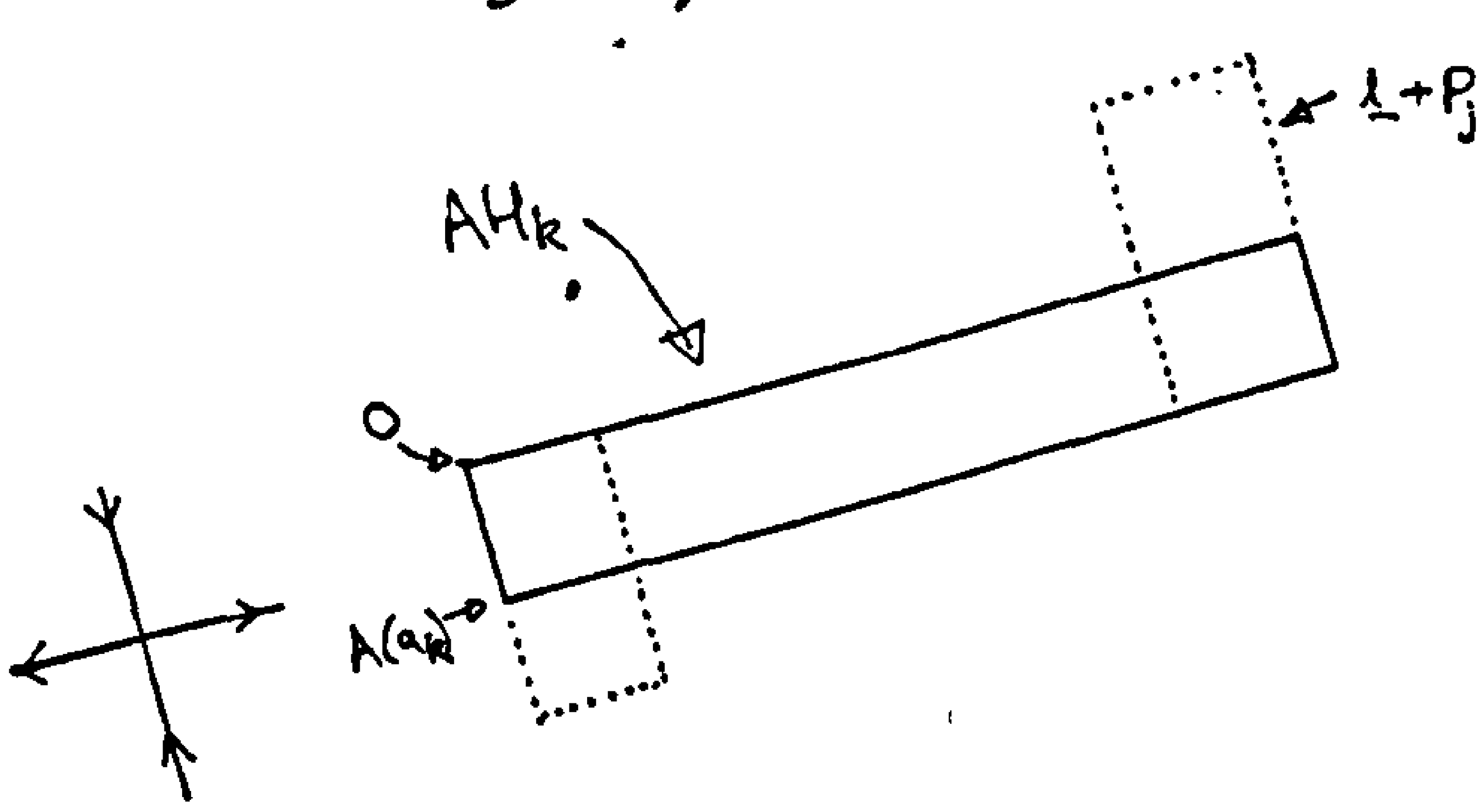


FIG. 18

However, our assumption that $H \subset AH$ implies that $At_j+[O,At_i]$ does not intersect int H, and hence that the anticlockwise perturbation of $At_j+[O,At_i]$ at At_j does not intersect int H. Thus $At_j+K[\Theta s_i]$ is in the pattern and is the anticlockwise perturbation of $At_j+[0,At_i]$ at At_j.

Write $W_i = s_j s_k s_j^{-1} s_k^{-1}$. Define V_i^1 as the set of points 'inside' $K[\theta W_i]$, as we did in the proof of 3.1 . Our choice of θ is such that the sides of V_i^1 do not overlap, in particular if we write

$$P_i^1 = [a_i,0] \times (A^{-1}V_i^1) \subset E^s \times E^u,$$

the different P_i^1 do not intersect in their interiors. Define $W^s(x,N) = (x+E^s) \cap N$ for $N \subset \mathbb{R}^3$. Since (as shown above) we cannot have $(q+H_j) \cap \mathrm{int}(AP_i) \neq \emptyset$ (fig 17),

$$W^s(Ax,AP_i) \subset W^s(Ax,p+P_j) \quad \text{if } x \in P_i,\ Ax \in P_j+p.$$

Hence

$$W^s(Ax,AP_i^1) \subset W^s(Ax,p+P_j) \quad \text{if } x \in P_i^1,\ Ax \in P_j+p.$$

Similarly define $W^u(x,N) = (x+E^u) \cap N$. We must have

$$W^u(Ax,AP_i^1) = a + V_i^1 \quad \text{some } a \in E^s.$$

Thus

$$W^u(Ax,AP_i^1) \supset W^u(Ax,p+P_j).$$

d) We now show how to define a sequence of partitions converging to the Markov partition which more and more nearly satisfy the Markov conditions.

Given $K[\theta^n W_i]$ define V_i^n and $\underline{\theta}$ analogously to the definitions in 3.1 . As before we have $\underline{\theta} V_i^n = V_i^{n+1}$. We must first prove that $K[\theta^n W_i]$ is in the pattern.

Because the pattern is a tiling, the slope of E^u and

E^s imply that if $r+H_i$ is in the tiling, adjoining $r+H_i$ along the edge $r+[0,t_k]$ is either $r-t_j+H_i$ or $r+H_j$ $(j\neq k)$. By choice of Θ,

$$\mathrm{int}(V_i^1)\cap\mathrm{int}(-t_j+V_i^1) = \emptyset$$
$$\mathrm{int}(V_i^1)\cap\mathrm{int}(V_j^1) = \emptyset .$$

Also by choice of Θ and since $(q+H_j)\cap\mathrm{int}(AP_i) = \emptyset$ for all $q\in\mathbb{Z}^3, j, i$,

$$[Aa_i,0]\times(K[s_j]+f(Y)) \subset ([a_k,0]\times K[s_j]) + q \quad \text{or}$$
$$([a_k,0]\times K[s_j]) + q + f(s_l)$$

where $\Theta W_i = Ys_jY'$ for some words $Y,Y'\in S^*$, k,j,l distinct, for some $q\in\mathbb{Z}^3$. We prove by induction that

$$[A^na_i,0]\times(K[s_j]+f(Z)) \subset ([a_k,0]\times K[s_j]) + q \quad \text{or}$$
$$([a_k,0]\times K[s_j]) + q + f(s_l)$$

when $\Theta^nW_i = Zs_jZ'$. Suppose it holds for $m\leq n$. Then applying A gives

$$[A^{n+1}a_i,0]\times(AK[s_j]+Af(Z)) \subset ([Aa_k,0]\times AK[s_j]) + Aq \quad \text{or}$$
$$([Aa_k,0]\times AK[s_j]) + Aq + Af(s_l)$$

which implies (using $Af=f\Theta$),

$$[A^{n+1}a_i,0]\times(K[\Theta s_j]+f(\Theta Z)) \subset ([Aa_k,0]\times K[\Theta s_j]) + r \quad \text{or}$$
$$([Aa_k,0]\times K[\Theta s_j]) + r + f(\Theta s_l)$$

Suppose $\Theta s_j = r_1..r_{m'}$, then since

$$K[r_1..r_{m'}] = \bigcup_{m=1}^{m'}(K[r_m]+f(r_1..r_{m-1})), \text{ we have}$$

$$[A^{n+1}a_i,0]\times(K[r_m]+f((\Theta Z)r_1..r_{m-1}))$$
$$\subset ([a_k,0]\times K[r]) + r' \quad \text{or}$$
$$([a_k,0]\times K[r]) + r' + f(s_{l'}).$$

This proves the induction. Hence for all n, $K[\Theta^nW_i]$ is a subset of the pattern.

The choice of θ is such that each V_i^1 is tiled in the pattern by b_{ij} copies (say) of each V_j^0 $(=H_j)$,

$$V_i^1 = \bigcup_{j=1}^{3} \bigcup_{k=1}^{b} (r(i,j,k) + V_j^0)$$

where $b = b_{ij}$. We prove by induction that

$$\text{int}(V_i^n) \cap \text{int}(-A^{n-1}t_j+V_i^n) = \emptyset \qquad (1)$$

$$\text{int}(V_i^n) \cap \text{int}(V_j^n) = \emptyset \qquad (2), \text{ and}$$

$$V_i^n = \bigcup_j \bigcup_k (A^{n-1}(r(i,j,k))+V_j^{n-1}) \qquad (3).$$

We have already checked (1),(2),(3) for n=1. Assume that they hold for $m \leq n$. Since V_i^1 is tiled using the pattern, the intersections between different tiles of V_i^1 are as in (1) and (2) for n=1. Thus (1) and (2) for n-1 imply that (3) gives a tiling of V_i^n.

Recall that $\underline{\theta}(A^{n-1}x+V_j^{n-1}) = A^n x + V_j^n$. The inductive step for (3) works simply by applying $\underline{\theta}$. The inductive step for (1) is as follows. By (1) and (3),

$$\text{int}(\bigcup_{j,k}(A^{n-1}r(i,j,k)+V_j^{n-1}))$$
$$\cap \text{ int}(\bigcup_{j,k}(A^{n-1}(r(i,j,k)-t_j)+V_j^{n-1})) = \emptyset.$$

Applying $\underline{\theta}$ and using (1),(2) gives

$$\text{int}(\bigcup_{j,k}(A^n r(i,j,k)+V_j^n)) \cap \text{int}(\bigcup_{j,k}(A^n(r(i,j,k)-t_j)+V_j^n)$$
$$= \emptyset.$$

By (3) we have

$$\text{int}(V_i^{n+1}) \cap \text{int}(-A^n t_j+V_i^{n+1}) = \emptyset.$$

A similar argument gives the inductive step for (2). Thus we have proved (1),(2),(3).

Define $P_i^n = [a_i,0] \times (A^{-n}V_i^n)$. We wish to show that as n tends to infinity the P_i^n more nearly satisfy the Markov conditions. We show

$$\bigcup_i([A^n a_i,0]\times K[\theta^n W_i]) \subset \bigcup_j \bigcup_{q\in \mathbb{Z}} (q + [A^{n-1}a_j,0]\times K[\theta^{n-1}W_j]).$$

This certainly holds for n=1. Suppose it holds for $m \leq n$. Applying A,

$$\bigcup_i [A^{n+1}a_i,0]\times AK[\theta^n W_i] \subset \bigcup_{j,q} (q + [A^n a_j,0]\times AK[\theta^{n-1}W_i])$$

which implies that

$$\bigcup_i [A^{n+1}a_i,0]\times K[\theta^{n+1}W_i] \subset \bigcup_{j,q} (q + [A^n a_j,0]\times AK[\theta^n W_i]) \qquad (4)$$

Thus if $P_i^n = [a_i,0]\times A^{-n}(V_i^n)$,

$$A(W^u(x,A^n P_i^{n+1})) \supset W^u(Ax,A^n P_j^n+q)$$

where $x \in \operatorname{int} A^n P_i^{n+1}$, $Ax \in \operatorname{int} A^n P_j^n+q$.

Therefore,

$$A(W^u(y,P_i^{n+1})) \supset W^u(Ay,P_j^n+r) \qquad y = A^{-n}x.$$

Writing $V^n = \bigcup_i V_i^n$, since $V^1 \supset H$ an easy induction using $\underline{\theta}$ implies that $V^{n+1} \supset V^n$. Hence

$$A(W^s(x,A^n P_i^{n+1})) \subset W^s(Ax,A^n P_j^n+q) \qquad \text{and so}$$

$$A(W^s(y,P_i^{n+1})) \subset W^s(Ay,P_j^n+r).$$

We now need to check that each

$$\mathcal{P}^n = \{q+P^n \,;\, P^n = \bigcup_i P_i^n \,,\, q\in \mathbb{Z}^3\}$$

is a tiling of $\mathbb{R}^3$. It is clear by induction that $A^{n-1}\mathcal{P}^{n-1}$ a tiling implies that $A^n\mathcal{P}^n$ is a tiling. For under the transformation $(A|_{E^s})\times(\underline{\theta})$ on $E^s\times E^u$, adjoining s-faces of tiles in $A^{n-1}\mathcal{P}^{n-1}$ are sent to adjoining s-faces of tiles in $A^n\mathcal{P}^n$. Also, adjoining u-faces are sent to adjoining u-faces because $H = \bigcup_i(e_i+(a_i+H_i))$ inductively implies that $V^n = \bigcup_i (A^n e_i+(A^n a_i+V_i^n))$. Hence for each n $A^n\mathcal{P}^n$

is a tiling, and so $\mathcal{P}^n$ is a tiling of $\mathbb{R}^3$.

As $n \to \infty$, $A^{-n} K[\theta^n W_i]$ converges in the Hausdorff metric. Hence $P_i^n \to \tilde{R}_i$,say. By the above arguments $\{\tilde{R}_i : i=1,2,3\}$ satisfies the Markov conditions and $\tilde{\mathcal{R}} = \{q+\tilde{R}_i : i=1,2,3, q\in \mathbb{Z}^3\}$ is a tiling of $\mathbb{R}^3$. By (3) and (4) $A\tilde{R}_i$ intersects some translate of $\tilde{R}_j$ b_{ij} times. Projecting via the covering map onto T^3 gives a tiling $\mathcal{R} = p(\tilde{\mathcal{R}})$ of T^3 which satisfies the Markov conditions. Clearly the transition matrix for $\mathcal{R}$ is (b_{ij}).

<u>Remark</u> The particular choice of θ that we made above was rather arbitrary - as when constructing a Markov partition for an expanding endomorphism of T^2 there are several possible choices for θ and hence a 'canonical' class of Markov partitions.

e) We now calculate $B = (b_{ij})$. This is done by considering what happens to the faces F_i, instead of the H_i. Intuitively $A(F_i)$ and $A(H_i)$ pass through approximately the same number of each $q+P_j$ $(q\in\mathbb{Z}^3)$ because H_i is the unstable component of F_i. They may go through the P_js in a different order.

Define a <u>stepped surface</u> $\tilde{U}\subseteq\mathbb{R}^3$ to be a surface such that $(x_1,x_2,x_3)\in\tilde{U}$ implies some $x_i\in\mathbb{Z}^3$. In other words $\tilde{U}$ is a union of $q+F_j$, $j=1,2,3$, $q\in\mathbb{Z}^3$. We say that

$\tilde{U}$ is <u>non-degenerate</u> if whenever $r+F_k, s+F_j \subset \tilde{U}$ and $\operatorname{int} p_s(r+F_k) \cap \operatorname{int} p_s(s+F_j) \neq \emptyset$, we have $r+F_k = s+F_j$. If $\tilde{U}$ is non-degenerate and has connected interior, then when $\tilde{U}$ is projected onto the (e_i, e_j) plane no two copies of F_k $(k \neq i,j)$ in $\tilde{U}$ project to the same square. This is because if $r+F_k, s+F_k \subset \tilde{U}$ we can find a sequence of faces in $\tilde{U}$ with adjoining edges going from $r+F_k$ to $s+F_k$. But non-degeneracy implies that if $m+F_l \subset \tilde{U}$ then $m+e_j+F_j \not\subset \tilde{U}$ $(l \neq j)$. Thus if an F_j $(j \neq k)$ appears in the sequence, $r+F_k$ and $s+F_k$ do not project to the same square. If the sequence is only of copies of F_k then clearly $r+F_k$ and $s+F_k$ do not project to the same square.

We construct a stepped surface $\tilde{U}$ corresponding to E^u in the following way. Remove from the half space of $\mathbb{R}^3$ containing the positive cone and with boundary E^u the interior of any cube $p+I^3$ ($I^3 = [0,1]^3$, $p \in \mathbb{Z}^3$) which has $\operatorname{int}(p+I^3) \cap E^u \neq \emptyset$. This gives a set X whose boundary is the stepped surface $\tilde{U}$. Because E^u intersects the positive cone only at the origin, if we write

$${}^{-}I_j = (-e_j, 0] \quad {}^{+}I_j = (0, +e_j],$$

then $E^u \cap (q+{}^{-}I_j) \neq \emptyset$ implies $E^u \cap (q+{}^{+}I_k) = \emptyset$ for any k. This implies that $\tilde{U}$ has the following special property:

$$E^u \cap (q+{}^{-}I_j) \neq \emptyset \iff q+F_j \subset \tilde{U}.$$

(fig. 19 shows the corresponding structure in two dimensions). This is because if $E^u \cap (q+{}^{-}I_j) \neq \emptyset$ then $q+I^3 \subset X$ and $q-e_j+I^3 \not\subset X$. The reverse implication works by seeing

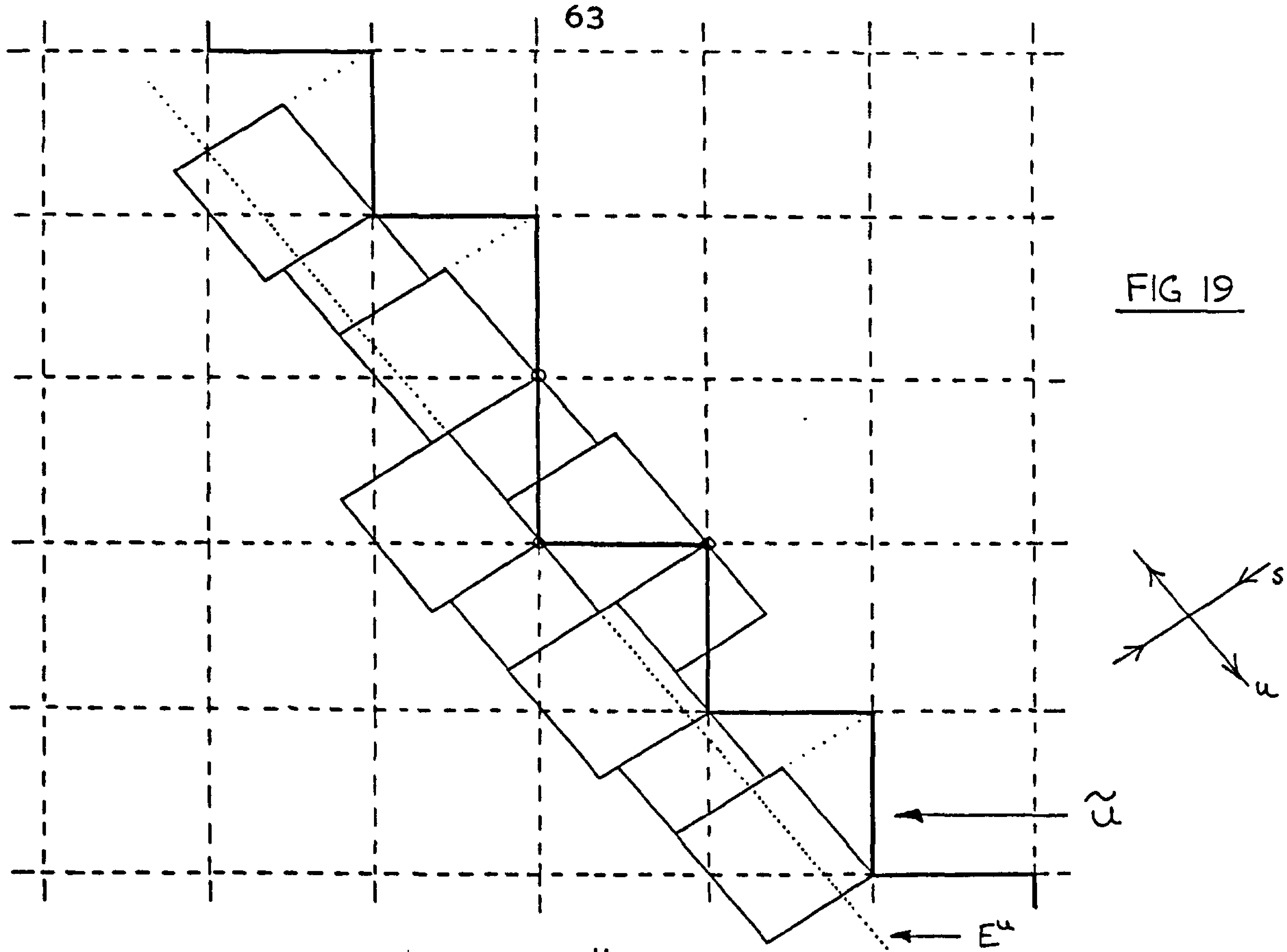


FIG 19

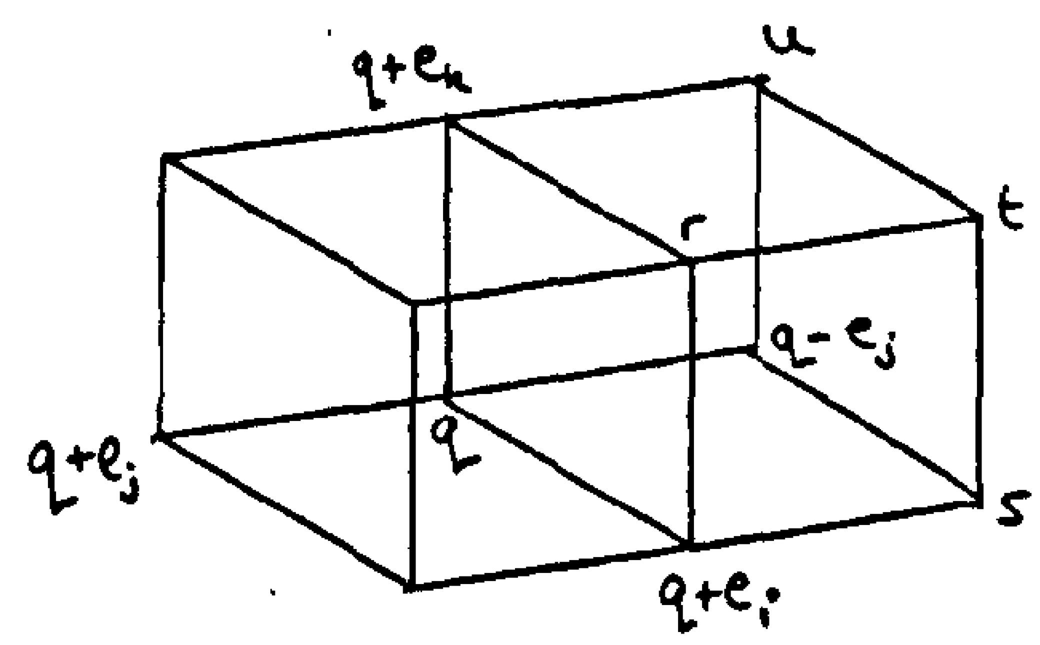


FIG. 20

how E^u can intersect the edges of $q-e_j+I^3$ (fig. 20). With the notation of fig. 20 , if $E^u \cap (r+{}^-I_j) \neq \emptyset$ then $E^u \cap (t+{}^-I_k)=\emptyset$ and $E^u \cap (r+{}^-I_k)=\emptyset$ so $E^u \cap (q+e_i+{}^-I_j) \neq \emptyset$ and similarly $E^u \cap (q+{}^-I_j) \neq \emptyset$. Similar arguments show that whichever edge of $q-e_j+I^3$ is intersected by E^u, E^u must also intersect $q+{}^-I_j$. This proves the special property.

Recall now that P_j was defined with upper face in H and lower face in $-e_j+H$. Hence $E^u \cap (q+{}^-I_j) \neq \emptyset$ if and only if $E^u \cap (q+((a_j,0]\times H_j)) \neq \emptyset$. Thus projecting the faces of $\tilde{U}$ onto E^u using the map p_s gives a tiling which equals the patterning of E^u. We can therefore think of $\tilde{U}$ as a non-degenerate lift of the patterning of E^u to a stepped surface. The lines $K[\theta s_j] \subset E^u$ lift to give polygonal lines $\tilde{L}_j \subset \tilde{U}$ joining 0 to $A(e_j)$. Now, V_i^1 lifts to the subset of $\tilde{U}$ bounded by $\tilde{L}_j$, $\tilde{L}_k$, $\tilde{L}_j+A(e_k)$, $\tilde{L}_k+A(e_j)$, which we call $\tilde{U}_i$.

In order to calculate the number of copies of F_l in $\tilde{U}_i$ we project $\tilde{U}_i$ onto the (e_m,e_n) plane ($l\neq m,n$ and $i\neq j,k$). Since $\tilde{U}$ is non-degenerate,

b_{il} = #(F_l in U_i) = #(squares in the projection of $\tilde{U}_i$ onto the (e_m,e_n) plane).

But the projection of $\tilde{U}_i$ onto the (e_m,e_n) plane is a figure $V_{i,l}$ bounded by L_j, L_k, $L_j+(a_{mk},a_{nk})$, $L_k+(a_{mj},a_{nj})$ (recall that $A(e_k) = \sum_{r=1}^{3} a_{rk}e_r$), where L_j is the projection of $\tilde{L}_j$ onto the (e_m,e_n) plane, etc. The different lines bounding $V_{i,l}$ do not overlap because of the non-degeneracy of $\tilde{U}$. Thus b_{il} = area of $V_{i,l}$

$$= \left| \det \begin{pmatrix} a_{mk} & a_{mj} \\ a_{nk} & a_{nj} \end{pmatrix} \right| = c_{li}$$

if $C = \text{adj } A = (c_{li})$ where $l\neq m,n$ and $i\neq j,k$. Hence

$$B = (\text{adj } A)^t = (A^{-1})^t \quad \text{since } \det A = 1$$

■

§ 2 Coding between Markov Partitions.

Theorem 3.1 showed how to construct Markov partitions with different boundary capacities for the same map of T^2. It is natural to ask if there is any sense in which one partition is a 'better' choice than another. One possible answer is as follows. Below we shall always work only with the measure of maximal entropy, denoting it μ. Let $D = |\det A|$, and suppose $\mathcal{X}$, $\mathcal{R}$ are two Markov partitions for $\tilde{A}$ constructed as in 3.1. There is an induced isomorphism $\phi: \Sigma_D \to \Sigma_D$ defined μ a.e. so that the following diagram commutes.

$$\begin{array}{ccc} (\Sigma_D, \sigma) & \xrightarrow{\ \phi\ } & (\Sigma_D, \sigma) \\ & {\scriptstyle \pi_{\mathcal{X}}} \searrow \quad \swarrow {\scriptstyle \pi_{\mathcal{R}}} & \\ & (T^2, A) & \end{array}$$

The map ϕ is a finitary isomorphism a.e., in other words if we define the anticipating function

$$a_\phi : \Sigma_D \to \mathbb{N} \cup \{\infty\}$$

by letting $a_\phi(\underline{x})$ be the smallest integer such that if $\underline{y} \in \Sigma_D$ and $y_i = x_i$ for $i \leq a_\phi(\underline{x})$ then $(\phi(\underline{y}))_0 = (\phi(\underline{x}))_0$, we have $a_\phi(\underline{x}) < \infty$ for a.a. $\underline{x}$. One says that ϕ has finite expected code length if $\int a_\phi \, d\mu < \infty$. An argument of Adler and Marcus (Ad2) shows that our map does indeed have finite expected code length. However, we shall see that the code length can be bounded in terms of the matrices $B(\mathcal{X})$, $B(\mathcal{R})$ (defined in Remark(i) after 3.1). Hence the expected code length depends on how crinkly the Markov partition boundaries are.

The Adler and Marcus (Ad2) argument formally works by using their Proposition 2.14 together with the argument on their page 78. It can be interpreted geometrically as follows. We denote an element of the partition $\tilde{A}^{-m}\mathcal{Y}$ by $\mathcal{T}(x_0,\ldots,x_{m-1})$ if there is an $(x_0,\ldots,x_{m-1},\ldots)\in\Sigma_D$ with $\pi_{\mathcal{Y}}(x_0,\ldots,x_{m-1},\ldots)\in \operatorname{int}\mathcal{T}(x_0,\ldots,x_{m-1})$. They show that there are blocks $x^1..x^r$ and y such that for each $x_0..x_m x^1..x^r$ there exists a block $y_0..y_m y$ with with

$$S(x_0..x_m x^1..x^r) \subset \operatorname{int}\ R(y_0..y_m y).$$

Thus if $\underline{x}\in\Sigma_D$ has $\pi_{\mathcal{X}}(\underline{x})\in S(x_0..x_m x^1..x^r)$ we know that $\phi(\underline{x}) = y_0..y_m y$, so in particular $a_\phi(\underline{x}) \leq (m+1+r)$.
Let $N_k = \{\underline{x}\in\Sigma_D : \underline{x} = (x_0..x_k..)$ with $x^1..x^r \not\subset x_0..x_k\}$, so $a_\phi(\underline{x}) > k$ implies that $\underline{x}\in N_k$. Hence

$$\mu\{\underline{x} : a_\phi(\underline{x}) = m\}$$
$$\leq \mu\{\underline{x} : a_\phi(\underline{x}) > m-1\} \leq \mu(N_{m-1}).$$

Finite expectation follows since $\mu(N_{m-1})$ converges exponentially to zero as $m\to\infty$. We shall make a more exact estimation below.

<u>Prop. 3.4</u> $\int a_\phi d\mu = \sum_{i=1}^{\infty} \frac{n_i}{|\det A|^i} < \infty$

where n_{i+1} = # (elements of $\bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{X}$ required to cover $\partial\mathcal{R}$).

<u>Proof:</u> Denote $C_i(\underline{x}) = \{\underline{y}\in\Sigma_D : (x_0..x_{i-1})=(y_0..y_{i-1})\}$.
We have to estimate $\mu\{\underline{x}\in\Sigma_D : a_\phi(\underline{x}) = i\}$.
Now, $i = a_\phi(\underline{x})$ means that

$$\pi_{\mathcal{S}}(C_i(\underline{x})) \subset \pi_{\mathcal{R}}(C_0(\underline{y})) \qquad \text{for some } \underline{y},$$

and $\quad \pi_{\mathcal{S}}(C_{i-1}(\underline{x})) \not\subset \pi_{\mathcal{R}}(C_0(\underline{z})) \qquad$ for any $\underline{z}$.

i.e. $\pi_{\mathcal{S}}(C_i(\underline{x}))$ does not intersect two distinct elements of $\mathcal{R}$ in sets of positive measure, but that $\pi_{\mathcal{S}}(C_{i-1}(\underline{x}))$ does. Since the measure we are considering is the measure of maximal entropy, Haar measure on T^2, this is the same as saying that $\pi_{\mathcal{S}}(C_i(\underline{x}))$ does not contain in its interior any part of the boundaries between elements of $\mathcal{R}$, but that $\pi_{\mathcal{S}}(C_{i-1}(\underline{x}))$ does.

$$\# \left\{ S \in \bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{S} \;:\; S = \pi_{\mathcal{S}}(C_i(\underline{x})) \subset \pi_{\mathcal{R}}(C_0(\underline{y})), \text{ for some } \underline{y} \right\}$$

$$= \# \left\{ S \in \bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{S} \right\} - \# \left\{ S \in \bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{S} \;:\; (\text{int } S) \cap \partial\mathcal{R} \neq \emptyset \right\}$$

$$= D^{i+1} - n_{i+1}$$

Certain elements S of $\bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{S}$ do not cover part of $\partial\mathcal{R}$ because they are subsets $S \subset S' \in \bigvee_{n=0}^{i-1} \tilde{A}^{-n}\mathcal{S}$ which did not cover part of $\partial\mathcal{R}$.

Now,

$$\# \left\{ S \in \bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{S} \;:\; S \subset S' \in \bigvee_{n=0}^{i-1} \tilde{A}^{-n}\mathcal{S} \text{ and } (\text{int } S') \cap \partial\mathcal{R} = \emptyset \right\}$$

$$= D\left(\# \left\{ S' \in \bigvee_{0}^{i-1} \tilde{A}^{-n}\mathcal{S} \;:\; (\text{int } S') \cap \partial\mathcal{R} = \emptyset \right\} \right)$$

$$= D\,(D^i - n_i)$$

Thus, $\# \left\{ S \;:\; (\text{int } S) \cap \partial\mathcal{R} = \emptyset,\ S \subset S',\ (\text{int } S') \cap \partial\mathcal{R} \neq \emptyset \right\}$

$$= D^{i+1} - n_{i+1} - D(D^i - n_i) \qquad = \quad Dn_i - n_{i+1}$$

and hence

$$\mu\{\underline{x} \in \Sigma_D : a_\phi(\underline{x}) = i\}$$
$$= \frac{Dn_i - n_{i+1}}{D^{i+1}} = \frac{n_i}{D^i} - \frac{n_{i+1}}{D^{i+1}}$$

Notice that we cannot have $a_\phi(\underline{x}) = 0$, for this would imply that $\mathcal{R} = \mathcal{S}$. Thus if $\mathcal{R} \neq \mathcal{S}$, $a_\phi(\underline{x}) > 0$ for all $\underline{x}$ and so $n_1 = D$.

$$\int a_\phi(\underline{x})\, d\mu = \sum_{i=1}^{\infty} i\mu\{\underline{x} \in \Sigma_D : a_\phi(\underline{x}) = i\}$$
$$= \lim_{N\to\infty} \sum_{i=1}^{N} i\left(\frac{n_i}{D^i} - \frac{n_{i+1}}{D^{i+1}}\right)$$
$$= \sum_1^\infty (n_i/D^i) \text{ if the series converges.}$$

■

We shall use the following lemma to obtain bounds on n_i.

<u>Lemma 3.5</u> There is a constant $r_{\mathcal{R}} > 0$ depending only on A and $B(\mathcal{R})$ 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 $\mathbb{R}^2$.

<u>Proof</u>: We only have to prove that if $|K|_i$ is the length of the λ_i component of a set K, there is a bound $r'_{\mathcal{R}}(i) > 0$ so that $|\mathbb{K}_\theta(W)|_i \leqslant r'_{\mathcal{R}}(i)$, where $W = s_1 s_2$ or $s_2 s_1$. We show by induction that,

$$|A^{-n}K[\theta^n W]|_i \leqslant 8bw \sum_{r=1}^{n} |\lambda_i|^{-r} \qquad (\ddagger)$$

where $b \geqslant b_{ij}$, $(b_{ij}) = B(\mathcal{R})$, and

$$w = \max_{j=1,2} \{|u_j| : u_j \text{ is the component of } e_j \text{ in the } \lambda_i\text{-eigenspace}\}.$$

Since the length of $\theta(s_1) \cup \theta(s_2)$ is less than 4b,

$$|K[\theta W]|_i \leqslant 2.4bw = 8bw.$$

Therefore $|A^{-1}K[\theta W]|_i \leqslant 8bw|\lambda_i|^{-1}$.

Suppose that $(\ddagger)$ holds for n. Then since

$$K[\theta^{n+1}W] = K[\theta(\theta^n W)] = \bigcup_{j=1}^{m} (Af(t^1..t^{j-1}) + K[\theta t^j])$$

when $\theta^n W = t^1..t^m$,

$$|A^{-1}K[\theta^{n+1}W]|_i \leqslant |K[\theta^n W]|_i + |A^{-1}K[\theta W]|_i$$
$$\leqslant |\lambda_i|^n (8bw \textstyle\sum_1^n |\lambda_i|^{-r}) + 8bw|\lambda_i|^{-1}$$

and so

$$|A^{-n-1}K[\theta^{n+1}W]|_i = 8bw \sum_1^{n+1} |\lambda_i|^{-r}$$

This proves the induction. Letting $n \to \infty$ shows

$$|K_\theta(W)|_i \leqslant 8bw \cdot \frac{1}{|\lambda_i| - 1}$$

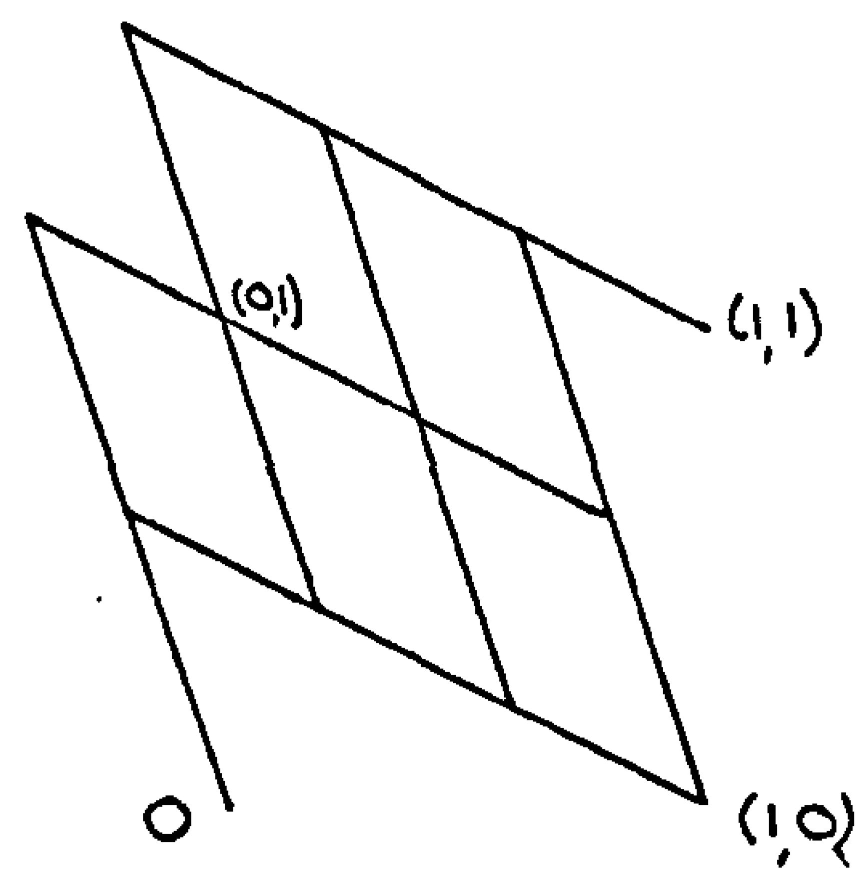
∎

<u>Prop. 3.6</u> There are constants $c_1, c_2 > 0$ depending only on A, $B(\mathcal{R})$, and $B(\mathcal{S})$ such that

$$\frac{c_1 \lambda_{B(\mathcal{R})}}{D - \lambda_{B(\mathcal{R})}} \leqslant \int a_\phi \, d\mu \leqslant \frac{c_2 \lambda_{B(\mathcal{R})}}{D - \lambda_{B(\mathcal{R})}}$$

<u>Proof:</u> We write $\hat{\mathcal{S}}$ for $p^{-1}\mathcal{S} \subset \mathbb{R}^2$, and let θ be the endomorphism generating $\mathcal{R}$. We slightly change the formalism used before, the object being to obtain $\partial\mathcal{R}$ immediately as

a recurrent set. Let $f(s_i) = A^{-1}(e_i)$ and introduce virtual symbols $\bar{s}_i$, $\bar{s}_i^{-1}$ corresponding to s_i, s_i^{-1} defining θ for these symbols by $\theta(\bar{s}_i) = \overline{\theta(s_i)}$ etc.. Then we can find a word V in our new S* so that $K_\theta(V)$ is a lift of $\partial\mathcal{R}$ (fig 23).



$K[V]$ for $\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$

FIG. 23

Recalling the definition of $\bar{\Sigma}_\theta$ and Σ_θ, and that $\log \lambda_{B(\mathcal{R})}$ is the entropy of Σ_θ, it is clear that there are constants q_1, q_2 so that

$$a_i = \#\{j : \theta^i V = t^1 .. t^r \text{ and } t^j \in \{s_1, s_1^{-1}, s_2, s_2^{-1}\}\} \in [q_1, q_2] \cdot \lambda^i_{B(\mathcal{R})}$$

Put $m_{i+1} = \#\{\hat{S} \in \hat{\mathcal{S}} : (\text{int } \hat{S}) \cap (A^i K_\theta(V)) \neq \emptyset\}$.
Recall $n_{i+1} = \#\{S \in \bigvee_{n=0}^{i} \tilde{A}^{-n}\mathcal{S} : (\text{int } S) \cap \partial\mathcal{R} \neq \emptyset\}$.
Then $m_{i+1} \geq n_{i+1}$ but also $4n_{i+1} \geq m_{i+1}$ (we may require more $\hat{S} \in \hat{\mathcal{S}}$ to cover $A^i K_\theta(V)$ than we needed $S \in \bigvee_0^i \tilde{A}^{-r}\mathcal{S}$ to cover $\partial\mathcal{R}$, but since applying $p_o A^{-i}$ to the former picture gives the latter we can bound the number of extra $\hat{S}$ required).

We easily obtain an upper bound on n_i. Let

$$k = \max_{i=1,2} \#\{\hat{S} \in \hat{\mathcal{S}} \text{ covering } K_\theta(s_i)\}$$

Then $n_i \leq m_i \leq ka_{i-1}$. Now

$k < \#\{\hat{S} \in \hat{\mathcal{S}}$ covering a lift of an $R \in \mathcal{R}\}$

$\leq \#\{\hat{S} \in \hat{\mathcal{S}}$ covering a square of side $r_{\mathcal{R}}\}$ by 3.5

But by 3.5 all such $\hat{S}$ would be contained in a square of side $r_{\mathcal{R}} + 2r_{\mathcal{S}}$, and since each $\hat{S}$ has area $1/D$,

$$k \leq \frac{(r_{\mathcal{R}} + 2r_{\mathcal{S}})^2}{D}$$

Letting $c_2 = (r_{\mathcal{R}} + 2r_{\mathcal{S}})^2 q_2/D$ gives the claimed upper bound.

A similar argument to the above tells us that a collection of t elements of $\hat{\mathcal{R}}$ intersects at least

$$t \Bigg/ \frac{(r_{\mathcal{S}} + 2r_{\mathcal{R}})^2}{D}$$

elements of $\hat{\mathcal{S}}$. Thus at least $a_i D/(r_{\mathcal{S}} + 2r_{\mathcal{R}})^2$ elements of $\hat{\mathcal{S}}$ cover $A^i(K_\theta V)$, i.e.

$$m_{i+1} \geq a_i \Bigg/ \frac{(r_{\mathcal{S}} + 2r_{\mathcal{R}})^2}{D}.$$

Setting $c_2 = q_1 D/4(r_{\mathcal{S}} + 2r_{\mathcal{R}})^2$ gives the claimed lower bound.

∎

Notice that finite expectation occurs if and only if $\lambda_{B(\mathcal{R})} < D$. Since $\log D = \log|\lambda_1| + \log|\lambda_2|$, Remark(i) of Theorem 3.1 implies $\lambda_{B(\mathcal{R})} < D \Leftrightarrow \mathrm{cap}(\partial\mathcal{R}) < 2$. The above proposition gives a formal way of saying that Markov partitions whose boundaries have high capacity are bad because it takes a long time to encode particular points.

§3 Invariant subsets for hyperbolic automorphisms of T^3 and expanding endomorphisms of T^2.

A series of papers (Hi1, Bo2, Bo5, Fr, Ha1, Ha2, Me1, Me2, Pr, Ir, Ub) have considered what kinds of invariant subsets can exist for hyperbolic automorphisms of T^n. The original motivations were the search for new types of Anosov diffeomorphisms and a question of Smale (in Hi1) who asked if there could exist a compact invariant set with topological dimension one. Figure 24 shows some of the results obtained. Some of the later results indicate, as various authors have commented, that invariant subsets (other than Mañé's invariant C^1 subtori) will have a complicated structure. A hyperbolic automorphism of T^3 in fact has no invariant C^0 subtori, for S^1 does not admit expansive homeomorphisms and there is no invariant 2-manifold by a result of Hirsch (Hi1).

The results of Mañé, Urbański and Irwin on the nature of paths with non-dense orbit also apply to expanding endomorphisms of T^2 when the eigenvalues of the covering map are not rational. In that case, as we remarked before, when the matrix A inducing the map is positive we can construct Markov partitions whose boundary has the minimal capacity, $2 - (\log|\lambda_2|)/(\log|\lambda_1|)$, allowed by Urbański's result. In the hyperbolic case on T^3 we cannot say if the same is true. Urbański constructs paths with non-dense orbit and capacity less than

$$2 - \frac{\log|\lambda_2|}{\log|\lambda_1|} + \varepsilon$$

FIG. 24

Invariant Submanifolds	
Hirsch (Hi1)	An invariant subset of $\tilde{A}$ cannot be homeomorphic to a sphere of dimension ≥ 1, a Klein bottle or a projective space. If the topological dimension of $E^s=1$, the only proper compact invariant submanifolds are periodic points.
Mañé (Me1)	If $f:T^n \to T^n$ is Anosov and $V \subset M$ is compact, invariant, $\partial V=\emptyset$, and is a C^1 manifold, then $f\|V$ is Anosov and each connected component is homeomorphic to a torus.
Invariant Sets	
Bowen (Bo2)	Minimal sets have zero topological dimension.
Hirsch (Hi1)	No compact invariant set has topological dimension equal to n-1.
Hancock (Ha1,Ha2)	$\{\sigma: I^m \to T^n \mid \overline{O(\sigma I^m)}=T^n\}$ is residual in the uniform topology on $C(I^m, T^n)$.
Przytycki (Pr)	For any k, $0 \leq k \leq n$, $k \neq n-1$, there is a compact invariant set N^k of topological dimension k.
Mañé (Me2)	The orbit closure of a non constant rectifiable arc contains a coset of a toral subgroup invariant under some power of $\tilde{A}$.
Irwin (Ir)	There exists a Holder continuous path in T^3 with a non-dense orbit.
Urbański (Ub)	There is a lower bound on $\overline{\text{cap}}$ of compact invariant sets.

Unless otherwise stated the above results refer to a hyperbolic automorphism of T^n, $\tilde{A}:T^n \to T^n$. Most of the results are not stated in their full generality and many results have been left out completely. The techniques used in (Pr), (Ir), and (Ub) largely stem from (Ha2).

and he asks if the ε can be removed. An easy covering argument (c.f. Be) shows that Irwin's Hölder continuous path with non-dense orbit has the required property.

Urbański's result shows the fractal nature of paths with non-dense orbits. However, capacity is not a good measure of the "infinitesimal wigglyness" of a curve. If a curve does not vary very much near one point it may have dense orbit even if its capacity is larger than $2 - (\log|\lambda_2|)/(\log|\lambda_1|)$.

<u>Def. 3.7</u> A path $C:[0,1] \to \mathbb{R}^2$ satisfies a <u>variation condition</u> f at $t_0 \in [0,1]$ (where $f: \mathbb{R}^+ \to \mathbb{R}^+$ is monotonic) if given a small $h \in \mathbb{R}$ we let t_0+s be the smallest parameter value with $s > 0$ for which $C_1(t_0+s) = C_1(t_0)+h$ and then

$$|C_2(t_0+s) - C_2(t_0)| \leqslant f(|h|)$$

where C_1, C_2 are the coordinate functions of C.

The following result (which was proven independently of Urbański's) applies equally well to the maps of T^2 (if the eigenvalues are irrational) and T^3 we have been studying, although we prove it in the more complicated T^3 setting.

We shall work in the covering plane using coordinates given by taking the coordinate axes along the eigenspaces and coordinates defined by distance.

Thrm 3.8 Let A be a 3x3 hyperbolic matrix of integers with $|\det A| = 1$ and eigenvalues with $|\lambda_1| > |\lambda_2| > 1 > |\lambda_3|$. Let $C:[0,1] \to \mathbb{R}^3$ be a path such that (C_1, C_2) satisfies a variation condition $f(h) = ch^a$ with $a > (\log|\lambda_2|)/(\log|\lambda_1|)$ at t_0. Then the orbit of $p(C)$ under $\tilde{A}$ is dense in T^3 where $p: \mathbb{R}^3 \to T^3$ is the covering map.

Proof: By taking a power of $\tilde{A}$, if necessary, we may assume that $\lambda_1 > \lambda_2 > 1$. (this does not affect $\log|\lambda_2| / \log|\lambda_1|$). We shall show that $p(C)$ has dense forward orbit. Hence we may, by considering only the unstable components of C, assume that $C_3 = 0$.

Suppose $z=C(t_0)$. For the moment we define $\tilde{C}=C-z$ in order to assume that $0=\tilde{C}(t_0)$. We first show that for all M and $\varepsilon/3 > 0$ there exists $m' > 0$ such that for all $m > m'$ $\forall r \in (0,M)\ \exists s > 0$ with $\lambda_1^m\tilde{C}_1(t_0+s)=r$ and $|\lambda_2^m\tilde{C}_2(t_0+s)| < \varepsilon/3$. Write $\tilde{C}=C^0$ and define a path C^1 as the image of C^0 under A,

$$C_1^1(s) = \lambda_1 C_1^0(s), \qquad C_2^1(s) = \lambda_2 C_2^0(s).$$

Since C^0 satisfies a variation condition, so does C^1. For take t_0,s,h as in the definition of a variation condition. Let $h'= \lambda_1 h$, then t_0+s is the smallest parameter value larger than t_0 for which

$$C_1^1(t_0) + h' = C_1^1(t_0+s).$$

Thus

$$|C_2^1(t_0+s) - C_2^1(t_0)| = \lambda_2|C_2^0(t_0+s) - C_2^0(t_0)|$$
$$\leqslant \lambda_2 ch^a = \lambda_2\lambda_1^{-a} c(\lambda_1 h)^a = \mu c(h')^a$$

where $\mu = \lambda_2\lambda_1^{-a}$. Note that $0<\mu<1$. If we inductively define paths C^n satisfies a variation condition f_n at t_0 where $f_n(h) = c\mu^n h^a$. Choose a small $h>0$ and a $k>0$ so that $\lambda_1^k h > M$. There exists $n>0$ such that

$$C_1^n(t_0+t) = h \quad \text{and} \quad |C_2^n(t_0+t)| < \lambda_2^{-k}(\varepsilon/3)$$

from the variation condition for C^n. Then if $m'=n+r$ and $m \geqslant m'$, $A^m\tilde{C}$ has the property that for all $r\in(0,M\rangle$ there exists $s>0$ with $\lambda_1^m C_1(t_0+s)=r$ and $|\lambda_2^m C_2(t_0+s)|< \varepsilon/3$.

Write $E_M = (0,M)\times\{0\}$. Since the eigenspace for λ_1 is dense in T^3, given $(\varepsilon/3)>0$ there exists $M>0$ such that for all $x\in T^3$ there exists $y'\in E_M$ with $d(x,y')< \varepsilon/3$. Hence for all p and $x\in T^3$ there exists $y\in p+E_M$ with $d(x,y)< 2\varepsilon/3$. For there is a $y''\in E_M$ with $d(y'-p,y'')< \varepsilon/3$ so $d(y',y''+p)< \varepsilon/3$ and putting $y = y''+p$, $d(x,y) < 2\varepsilon/3$.

We can now show that given $x\in T^3$ and $\varepsilon>0$ there is $m'>0$ such that for all $m>m'$ there exists $s>0$ with

$d(A^m C(t_0+s),x) < \varepsilon$. For take ε,M,m as above, then with the notation of the previous two paragraphs there exists $s > 0$ with

$$A^m(\ C_1(t_0),\ C_2(t_0+s)) \ = \ y$$

and $p=C(t_0)$. So

$$\begin{aligned} & d(x,\ A^m C(t_0+s)) \\ \leq\ & d(x,y) + d(y,A^m C(t_0+s)) \ = \ 2\varepsilon/3 + \varepsilon/3 \ = \varepsilon. \end{aligned}$$

Thus the forward orbit of C is dense.

This proves the theorem.

■

It is easy to give an example of a path C with

$$\overline{\text{cap}}(C) \ \geq \ 2 - \frac{\log|\lambda_2|}{\log|\lambda_1|}$$

but satisfying a variation condition $h \mapsto h^b$ where $b < (\log|\lambda_2|)/(\log|\lambda_1|)$. We work in E^u using coordinates as in the above proof. Let g be the graph of $h \mapsto h^b$. Let G_n be the closed region bounded by g, -g, $x=2^{-n}$, and $x=2^{-n-1}$. For each n choose a curve C_n with

$$\overline{\text{cap}}(C_n) \ \geq \ 2 - \frac{\log|\lambda_2|}{\log|\lambda_1|}$$

$C_n \subset G_n$ and $C_n \cap \{(x,y):\ x=2^{-r}\} \ = \ (2^{-r},0)$ for $r=n,n+1$. Let $C = \bigcup_n C_n$. Then clearly

$$\overline{\text{cap}}(C) \ \geq \ 2 - \frac{\log|\lambda_2|}{\log|\lambda_1|}$$

but C satisfies a variation condition $h \mapsto h^b$, $b < \dfrac{\log|\lambda_2|}{\log|\lambda_1|}$.

It is easy to use the recurrent set formalism to construct lots of invariant sets for expanding endomorphisms $\tilde{A}$ of T^2. Any finite alphabet $\bar{S}$ with $f:\bar{S} \to \mathbb{Z}^2$ and $\theta:\bar{S}^* \to \bar{S}^*$ such that $f\theta = Af$ will generate a set on T^2 invariant under $\tilde{A}$. Let $Y = p(\bigcup_{s\in\bar{S}} K_\theta(s))$. Suppose $\theta(s) = s_1 \ldots s_r$ then

$$A\, K_\theta(s) = \bigcup_{i=1}^{r} (K_\theta(s_i) + \textstyle\sum_{j<i} f(s_j))$$

Hence $\tilde{A}p(K_\theta(s)) = \bigcup_{i=1}^{r} p(K_\theta(s_i))$ and $\tilde{A}Y = Y$.

Clearly the fractal dimensions (i.e. Hausdorff dimension and capacity) of Y are the same as for K_θ. Whilst estimating capacity we shall use the notation of the previous chapter, in particular L = A, but $\tilde{L}$ is the map defined on page 23 and not the induced map on T^2. We need to know if K_θ is well matched, and we claim that K_θ is well matched if and only if no essential symbol duplicates (c.f. D3 5.3, 6.1). By the remark after 2.14 we only need to show that π_θ is bounded to one if and only if no essential symbol duplicates.

If we assume there is no essential duplication, since $f:\bar{S} \to \mathbb{Z}^2$,

$$\# \{ 1 \leq j \leq l(s,m) \;:\; \theta^m s = s_1 \ldots s_{l(s,m)},\ s_j \in E,\ f(s_1 \ldots s_{j-1}) = (p,q) \} \leq |E| \quad \text{for all } (p,q) \in \mathbb{Z}^2 \text{ and all } s \in E.$$

Suppose diam $K_\theta(s) \leq r$ for all s. Then if $y \in L^m K_\theta(s)$,

$$\# \{ 1 \leq j \leq l(s,m) : \theta^m s = s_1 \ldots s_{l(s,m)},\ s_j \in E,\ y \in K_\theta(s_j) + f(s_1 \ldots s_{j-1}) \} \leq 4r^2 |E|.$$

Hence $\# \{ \underline{x} \in \Sigma_\theta : \pi_\theta \underline{x} = y \} \leq 4r^2 |E|$, and π_θ is bounded to one. If an essential symbol duplicates, the argument of 2.14 (iii) $\Rightarrow$ (ii) shows that π_θ is not bounded to one.

If we have chosen θ so that the topological dimension of K_θ is one and K_θ is well matched, the dimension estimate (*) equals the capacity of K_θ. To see this notice that if V_n is any n-cylinder of K_θ, the projection of V_n onto the λ_2-eigenspace has non-empty interior. By the Markov property of K_θ we can find an $\varepsilon > 0$ and squares Q_s with the following properties,

i) $Q_s \cap K_\theta(s)$ projected onto the λ_2-eigenspace is an interval.

ii) $$(Q_{s_i} + f(s_1 \ldots s_{i-1})) \cap (Q_{s_j} + f(s_1 \ldots s_{j-1})) = \emptyset \quad \text{when } \theta^n(s) = s_1 \ldots s_r$$

iii) Q_s has sides parallel to the eigenspaces, with side length ε.

Any cover of K_θ by squares with sides parallel to the eigenspaces and side length l, $\varepsilon \lambda_1^{-n-1} < l \leq \varepsilon \lambda_1^{-n}$, must have at least $(\lambda_1 \lambda_2^{-1})^n$ squares covering

$$\tilde{L}^{-n}(Q_{s_i} + f(s_1 \ldots s_{i-1})) \cap K_\theta(s).$$

Since the number of n-cylinders is asymptotic to $e^{nh(\sigma_\theta)}$, we have at least $\text{const.}(\lambda_1 \lambda_2^{-1})^n \exp(nh(\sigma_\theta))$ squares of side 1 covering K_θ. This implies that

$$\text{cap } K_\theta \geqslant 1 + \frac{h(\sigma_\theta) - \log|\lambda_2|}{\log|\lambda_1|}$$

and hence that (*) equals cap K_θ in this case.

Application of the recurrent set formalism to the case of hyperbolic automorphisms of T^3 is more technical.

<u>Prop. 3.9</u> With the notation and conditions of 3.3, suppose that $\theta: S^* \to S^*$ is chosen so that

$$(\text{int } ([Aa_i, 0] \times K[W_i])) \cap \{q+H : q \in \mathbb{Z}^3\} = \emptyset \qquad (\$)$$

for each i, in the induced topology. Then $p(\bigcup_i K_\theta(s_i))$ has non-dense orbit under $\tilde{A}: T^3 \to T^3$.

<u>Proof</u>: Let $\bigcup_{j,m} (r_{jm} + [0, t_j])$ be the set of line intervals in the pattern. By (\$) we can use the argument of 3,13 (p58) to see that

$$K[\theta^n W_i] \subset \bigcup_{j,m} (r_{jm} + [0, t_j])$$

for all n. Hence $K_\theta(\theta^n W_i) \subset \bigcup_{j,m} (r_{j,m} + K_\theta(s_j)) \quad \forall n$, and so

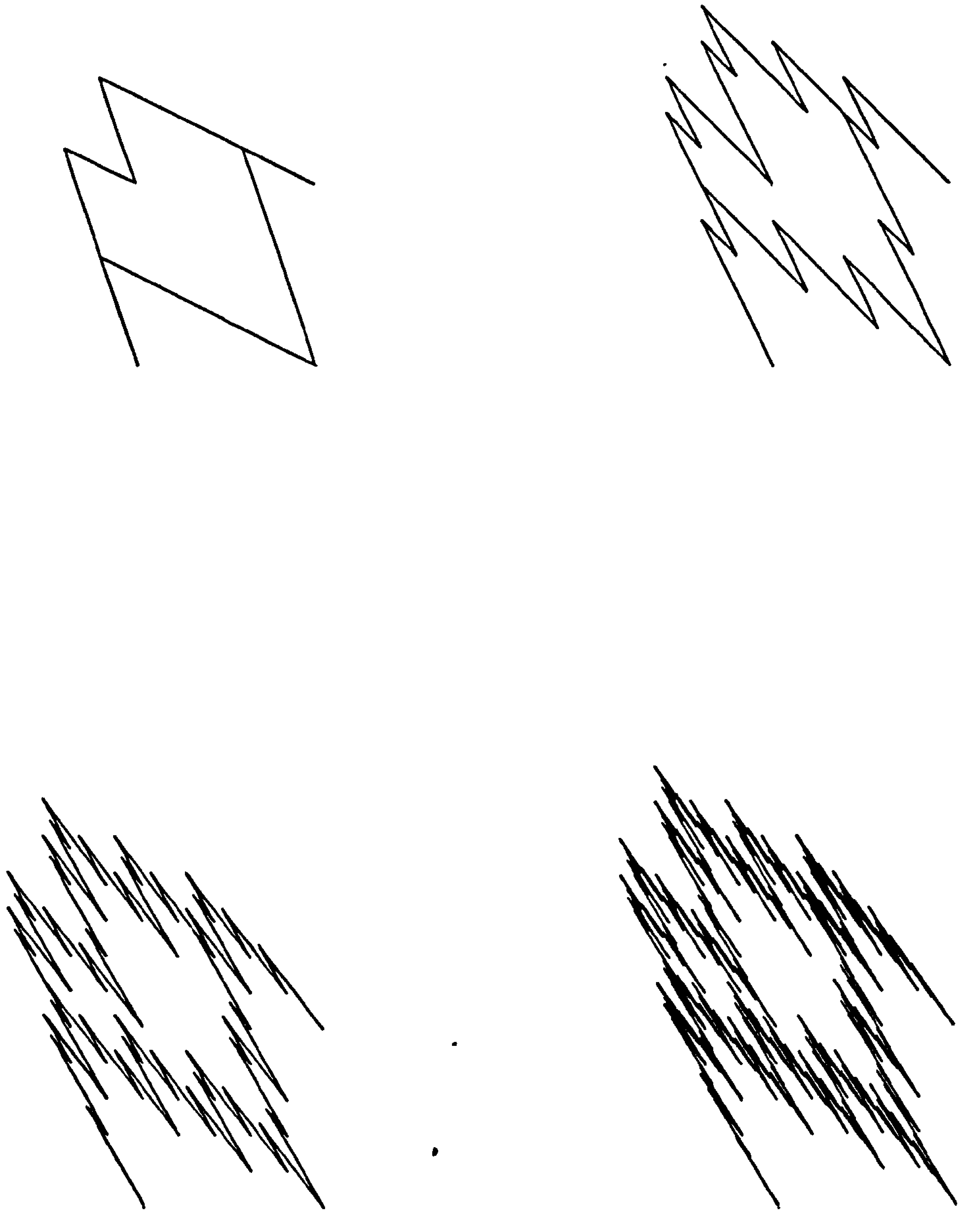
$$K_\theta(\theta^n W_i) \subset \bigcup_{q,k} (q + [a_k, 0] \times K_\theta(W_k)) \ , \ q \in \mathbb{Z}^3, \ k=1,2,3$$

Thus the forward orbit of $p(K_\theta(s_i))$ is not dense because $K_\theta(\theta^n W_i) = A^n K_\theta(W_i)$. Since $p(K_\theta(s_i)) \subset W^u(0)$, the backwards orbit is not dense either.

∎

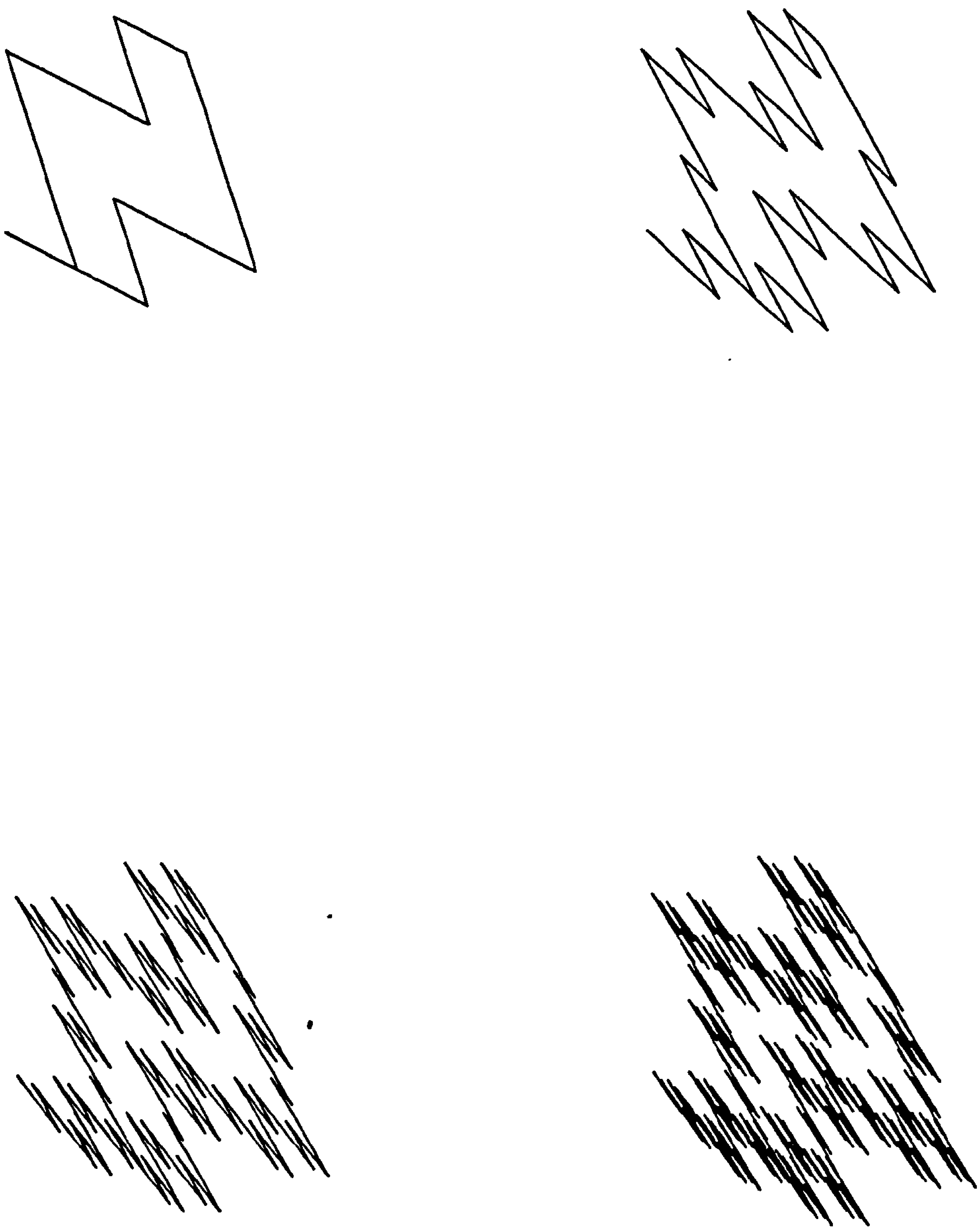
The above construction differs from constructions given in (Ha1,Ha2,Pr,Ir) because we have a single condition to check that ensures a non-dense orbit. Previous constructions were based on Hancock's idea of taking a curve C in T^3 and a region $U \subset T^3$, and inductively perturbing C if $\tilde{A}^n C$ entered U so that $\tilde{A}^n C \cap U = \emptyset$ for all n.

FIGURE 25



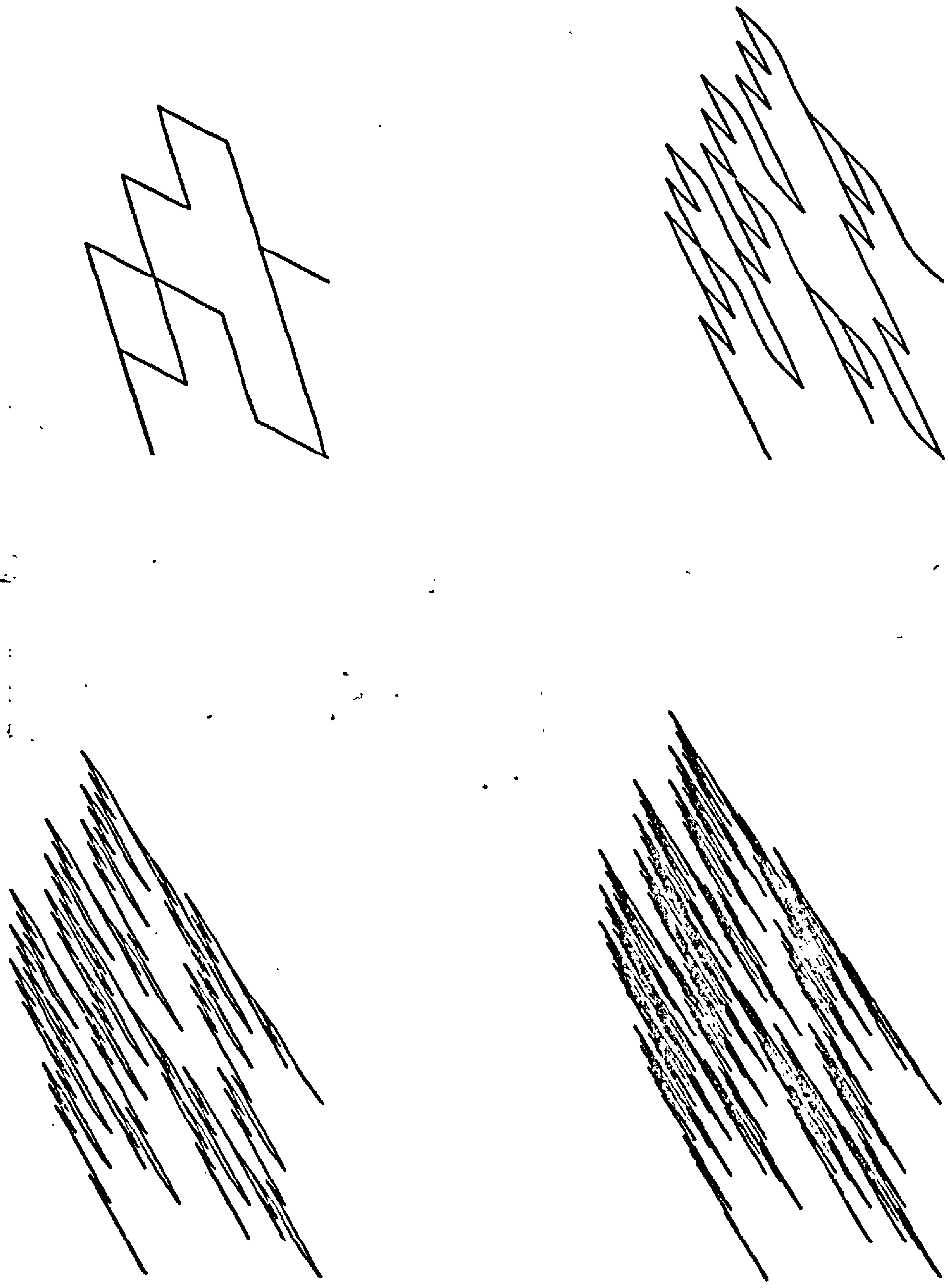
$A^{-n}K[\theta^n W]$ for n=1,.,4 where $W=s_1 s_2 s_1^{-1} s_2^{-1}$, and $\theta s_1=s_2 s_1 s_1 s_1$ $\theta s_2=s_2 s_2 s_1$ $A=\begin{pmatrix}3 & 1\\ 1 & 2\end{pmatrix}$. This choice of θ satisfies the conditions for 3.1. cap K_θ = 1.7484.

FIGURE 26



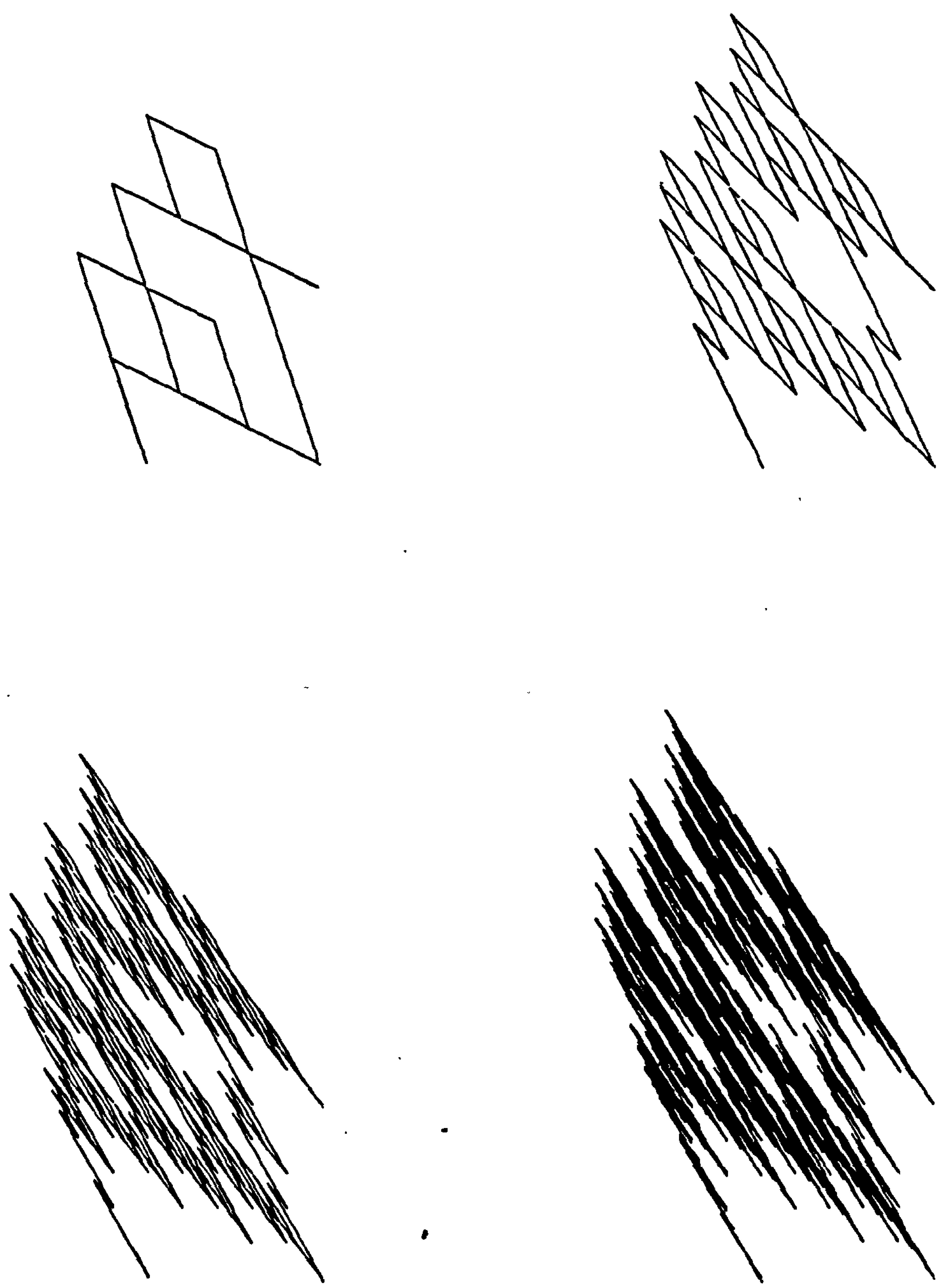
$A^{-n}K[\theta^n W]$ for n=1,.4 where $W=s_1 s_2 s_1^{-1} s_2^{-1}$, and $\theta s_1 = s_1 s_1 s_2 s_1$ $\theta s_2 = s_1 s_2 s_2$ $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$. This choice of θ satisfies the conditions for 3.1 . cap $K_\theta = 1.7484$

FIGURE 27



$A^{-n}K[\theta^n W]$ for n=1,.,4 where $W=s_1 s_2 s_1^{-1} s_2^{-1}$, and $\theta s_1 = s_2 s_1 s_2 s_1 s_2^{-1} s_1$ $\theta s_2 = s_2 s_2 s_1$ $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$. This choice of θ satisfies the conditions for 3.1 cap $K_\theta < 1.8832$, and this example has essential duplication.

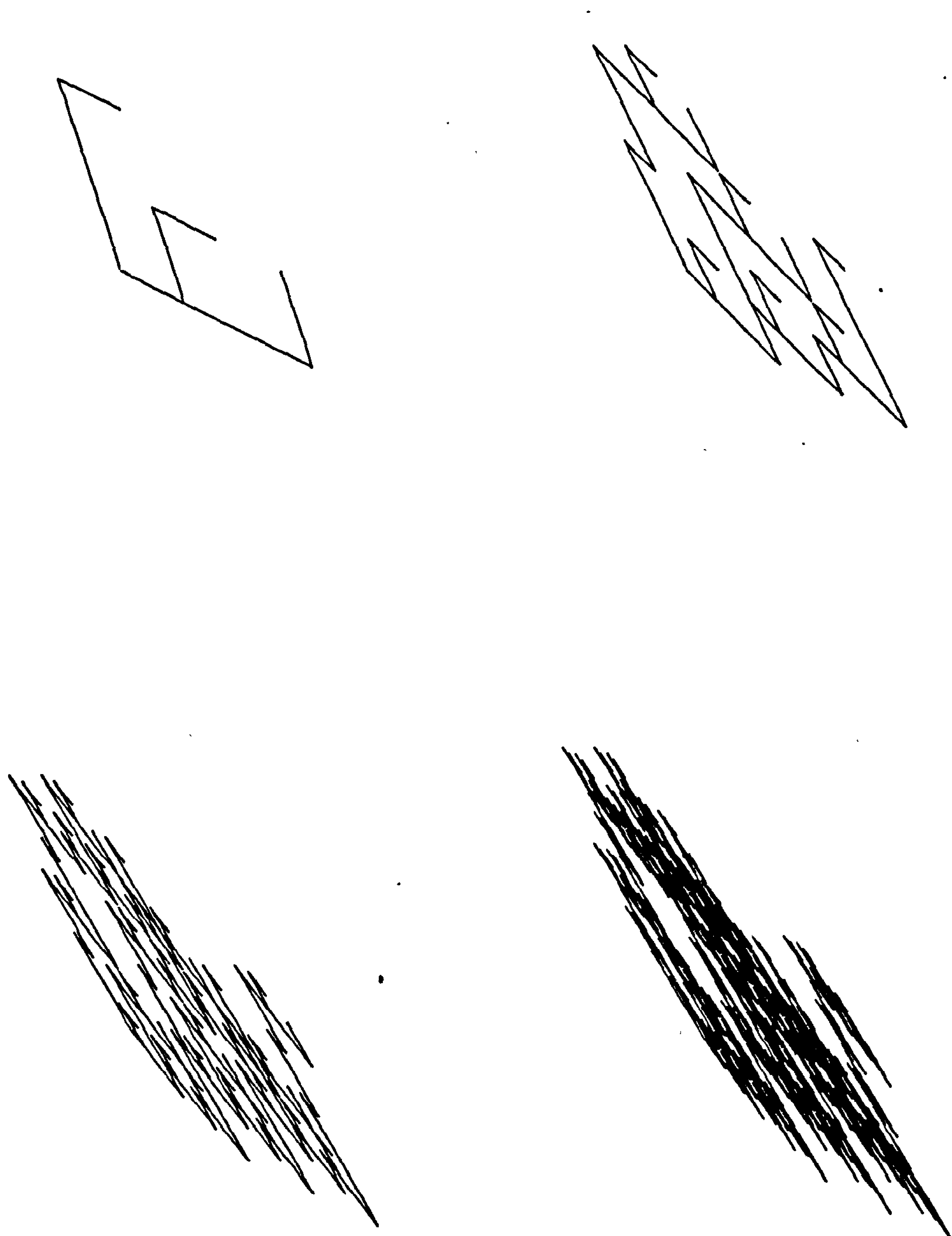
FIGURE 28



$A^{-n}K[\theta^n W]$ for n=1,.,4 where $W=s_1s_2s_1^{-1}s_2^{-1}$, and $\theta s_1=s_2s_1s_2s_1s_4^{-1}s_3^{-1}s_1s_1$ and $\theta s_2=s_2s_2s_1$ $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$. This choice of θ gives an invariant subset for $\hat{A}$.

In this example there is essential symbol duplication.

FIGURE 29



$A^{-n}K[\theta^n W]$ for n=1,.,4 where $W=s_1 s_1^{-1} s_2$., and $\theta s_1 = s_1 s_2 s_1 s_3^{-1} s_4^{-1} s_1 s_1 s_2$ $\theta s_2 = s_2 s_2 s_1$ $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$. Here we have s_1^{-1} and s_2^{-1} as virtual symbols. cap $K_\theta \leq 1.9572$.

CHAPTER FOUR

DIMENSION OF SELF AFFINE SETS

This Chapter goes back to the question of calculating dimension for fractals (e.g. recurrent sets) having a subshift of finite type structure when the scaling map is not a similitude. In this situation, as we saw in Chapter 2, even if a recurrent set K_θ is well matched the dimension estimate (*) may not equal cap(K_θ). Thus it is impossible to make fractal dimension calculations without geometric information in addition to knowledge of the subshift of finite type structure.

Many physicists (e.g. Br, Fa, Gr1, Gr2, Wi) have been interested in making fractal dimension calculations in numerical studies of fractal (strange) attractors. Takens (Ta) raised the question of whether in this context Hausdorff dimension equals capacity - an important question because of the relative ease of estimating capacity. Grassberger (Gr1) recently wrote, refering to (Fa),

> "It is generally accepted (Fa) that for almost all attractors D (capacity) is equal to the Hausdorff-Besicovitch dimension."

Here we shall give some simple examples demonstrating that this assumption is not correct in general. These examples are not themselves attractors, but could clearly be used to construct examples which are.

For most of this chapter we deal only with certain self affine sets in $\mathbb{R}^2$. We construct these sets as follows. Choose integers $r > s > 1$, and in the unit square $[0,1]^2$

draw a grid of lines $[0,1] \times \{i/s\}$ for $i=0,\ldots,s$ and $\{j/r\} \times [0,1]$ for $j=0,\ldots,r$. Shade some of the rectangles (fig 30) - the shaded set gives us our first approximation M_1 to the fractal. Let $L = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}$. Replacing each shaded rectangle with a copy of $L^{-1}(M_1)$ gives the second approximation M_2. Proceeding inductively we define M_n by replacing the shaded rectangles of M_{n-1} by copies of $L^{-(n-1)}(M_1)$. Then $M_n \to E$ a compact non-empty subset of $\mathbb{R}^2$ (covergence being in the Hausdorff metric). One can easily construct E as a recurrent set (by liberal use of virtual symbols). There is also a projection to E from a full shift, where an n-cylinder of E is the intersection of E with a shaded rectangle of M_n. Label the rows of the original grid $0,1,\ldots,s-1$. We write k_i for the number of shaded rectangles in row i of M_1 (fig 30).

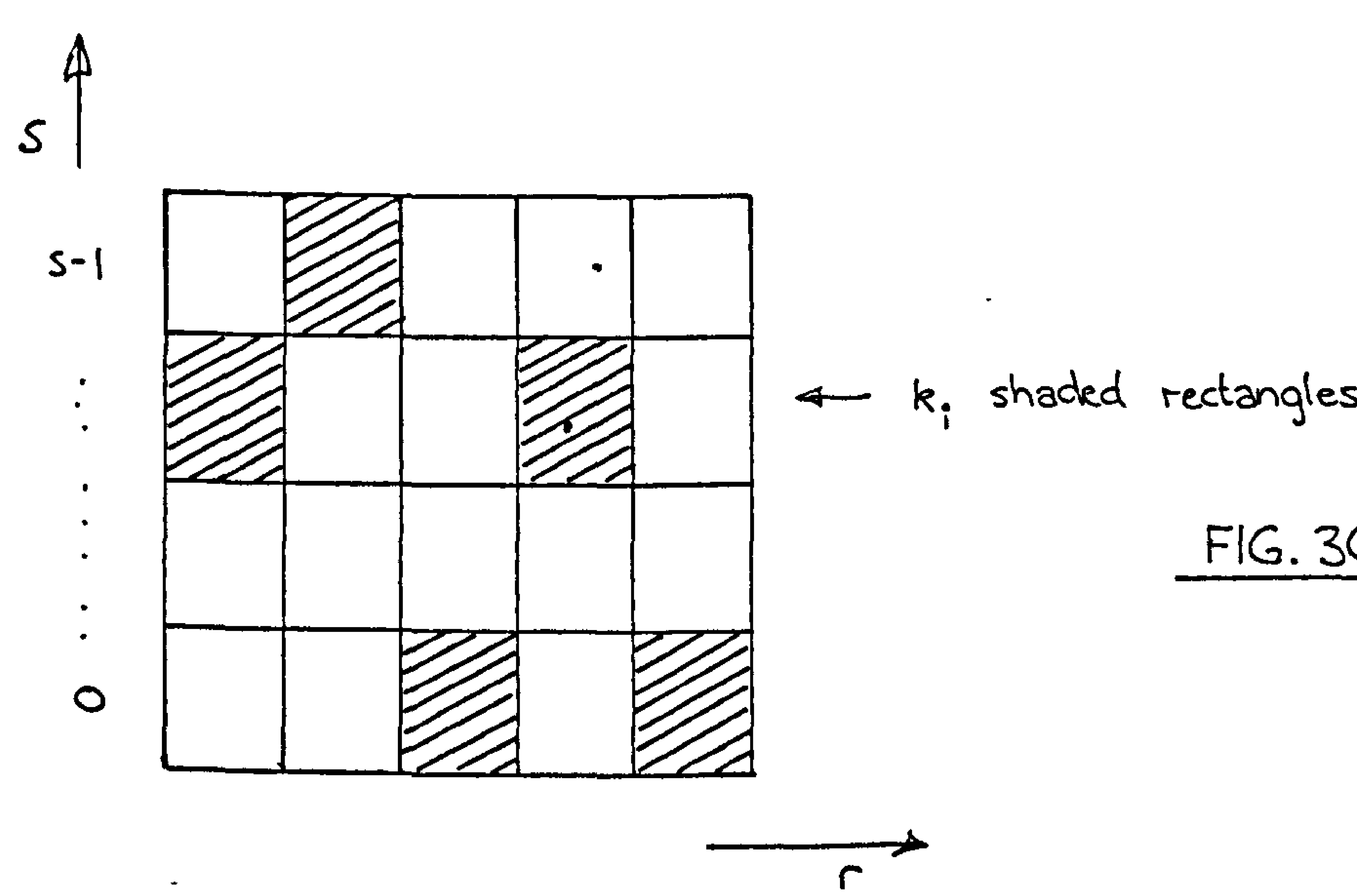


FIG. 30

The following proposition shows how insensitive capacity is to the geometry of E.

Prop. 4.1 Let $t = \#\{i \mid k_i \neq 0\}$. Then

$$\mathrm{cap}(E) = \left(\frac{1}{\log s} - \frac{1}{\log r}\right)\log t + \frac{\log(\sum_{i=0}^{s-1} k_i)}{\log r}$$

Proof: Let Gr(n) be the grid of squares in $[0,1]^2$ with side length r^{-n} whose corners have coordinates of the form $(i/r^n, j/r^n)$, $0 \leq i,j \leq r^n$. We only have to estimate capacity using coverings by squares from $\bigcup_n \mathrm{Gr}(n)$, for define $N'(r^{-n})$ to be the minimum number of squares in Gr(n) required to cover E. We claim that

$$\overline{\lim_{n\to\infty}} \frac{\log N'(r^{-n})}{n.\log r} = \overline{\lim_{\varepsilon\to 0}} \frac{\log N(\varepsilon)}{-\log \varepsilon} \quad \text{and}$$

$$\underline{\lim_{n\to\infty}} \frac{\log N'(r^{-n})}{n.\log r} = \underline{\lim_{\varepsilon\to 0}} \frac{\log N(\varepsilon)}{-\log \varepsilon}$$

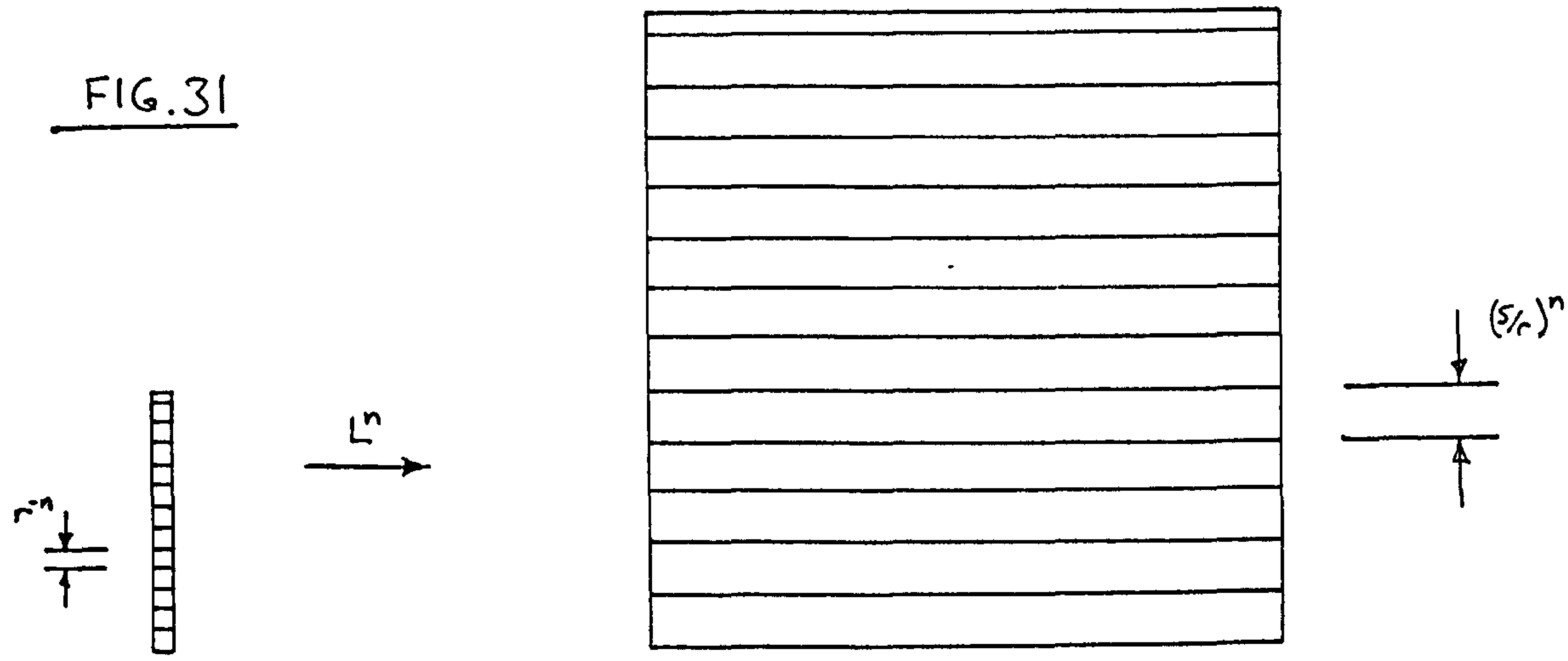
For given $\varepsilon > 0$ choose n such that $r^{-n} \leq \varepsilon < r^{-(n-1)}$. Then each ε-ball is contained in 9 squares of Gr(n-1) and therefore $9r^2$ squares of Gr(n). Each square of Gr(n) is contained in a ball of radius ε.

Thus $N(\varepsilon) \leq N'(r^{-n}) \leq 9r^2 N(\varepsilon)$. This, together with the fact that $\frac{\log \varepsilon}{-n.\log r} \longrightarrow 1$ as ε (or r^{-n}) $\to 0$, proves the claim.

Our set is self affine and so for any n-cylinder of E, C_n, $L^n(C_n)$ is just a copy of E. Thus a covering of C_n by squares of Gr(n) corresponds exactly to a covering of E by horizontal strips (fig 31).

Therefore we can count the number of squares from Gr(n) required to cover C_n by counting the number of horizontal strips of height $(s/r)^n = \lambda^n$ intersecting E.

FIG.31



We already know that exactly t^j horizontal strips of height s^{-j} intersect E. Choose n so that $\lambda^n \leqslant s^{-j} \leqslant \lambda^{n-1}$. Taking logs gives

$$n\left(\frac{\log r}{\log s} - 1\right) \geqslant j > (n-1)\left(\frac{\log r}{\log s} - 1\right) \qquad (1).$$

The horizontal strips intersect the y-axis in intervals of lengths λ^n and s^{-j}. Each s^{-j}-interval intersects at most $[s^{-j}. \lambda^{-n}] + 1$ λ^n-intervals, and each λ^{n-1}-interval intersects at most $[\lambda^{n-1}.s^j] + 1$ s^{-j}-intervals. Hence

$\#(\lambda^n\text{-strips required to cover E}) \leqslant t^j([s^{-j}\lambda^{-n}] + 1)$ and

$\#(\lambda^{n-1}\text{-strips required to cover E}) \geqslant t^j([\lambda^{n-1}s^j] + 1)$

Now, $\log(s^{-j}\lambda^{-n})$ is bounded above and below by constants (from (1)). Thus there are constants $c_1, c_2 > 0$ such that

$\#(\text{squares in } Gr(n) \text{ required to cover } C_n)$

$= \#(\lambda^n\text{-strips required to cover E})$

$\in [c_1, c_2] . t^j$

There are $(\sum_{i=0}^{s-1} k_i)^n$ distinct n-cylinders of E, so

$$\overline{\text{cap}}(E) \leqslant \overline{\lim_{n\to\infty}} \frac{\log (c_2 t^j (\sum k_i)^n)}{\log(r^{-n})}$$

$$= \underline{\lim_{n\to\infty}} \frac{\log (c_1 t^j (\sum k_i)^n)}{\log(r^{-n})} \leqslant \underline{\text{cap}}(E) .$$

Therefore

$$\text{cap}(E) = \frac{\left(\frac{\log r}{\log s} - 1\right)\log t + \log(\sum k_i)}{\log r}$$

(using (1)) as claimed.

■

We now go on to calculate the dimension of E, but first establish some notation:

$\underline{p} = (p_0, \ldots, p_{s-1})$ has $\sum_0^{s-1} p_i = 1$, $p_i \geqslant 0$ and $p_i = 0$ if $k_i = 0$.

We write $H(\underline{p}) = \sum_0^{s-1} p_i \log p_i$,

$B_\delta(p_i) = \{x : |x - p_i| < \delta\}$ if $p_i \neq 0$,

$B_\delta(0) = \{0\}$ and $B_\delta(\underline{p}) = \prod_0^{s-1} B_\delta(p_i)$.

We shall write $y = 0.y_1y_2\ldots$ base s, put

$P(y,n,i) = \# \{1 \leqslant j \leqslant n \mid y_j = i\}$ and $\underline{P}_n(y) = \left(\frac{P(y,n,i)}{n}\right)_{i=0}^{s-1}$

We use the sup metric when working in $\mathbb{R}^s$.

The following lemma is a slight generalization of a calculation due to Eggleston (Eg)

<u>Lemma 4.2</u> $\dim\{y : (\underline{P}_n(y))_{n=1}^\infty$ has a limit point in $B_\delta(\underline{p})\}$

$$= \sup \{-H(\underline{q})/\log s : \underline{q} \in B_\delta(\underline{p})\}$$

Proof: Call this set A_δ. Firstly note that if $y \in A_\delta$, then $\underline{P}_n(y) \in B_\delta(\underline{p})$ infinitely often (1)

for otherwise there is an N such that for all $n > N$,

$$|\underline{P}_n(y) - \underline{p}| > \delta$$

which implies that $|\underline{q} - \underline{p}| \geqslant \delta$ for any limit point $\underline{q}$ of $\underline{P}_n(y)$.

Let $X(n, \delta')$ be the set of intervals of length s^{-n} whose left hand end point x is of the form $x = k/s^n$ for some $k \in \mathbb{Z}$, and also satisfies $\underline{P}_n(x) \in B_{\delta'}(\underline{p})$.

Then if $\delta' > \delta$, for all N we have by (1),

$$\bigcup_{n=N}^{\infty} X(n, \delta') \supset A_\delta \qquad (2)$$

We shall show that the dimension estimate coming from (2) almost gives the required answer.

We can bound $\# X(n, \delta')$ combinatorially,

$$\# X(n, \delta') = \sum \frac{n!}{t_0! \ldots t_{s-1}!} \qquad (3)$$

where the summation is over s-tuples of integers $(t_0, \ldots, t_{s-1})$ such that $t_i/n \in B_{\delta'}(p_i)$ and $\sum_0^{s-1} t_i = n$.

Write $\alpha_{\underline{r}} = -H(\underline{r})/\log s$ and suppose that $\sup\{\alpha_{\underline{r}} \mid \underline{r} \in B_{\delta'}(p)\}$ equals $\alpha_{\underline{q}}$, $\underline{q} \in \overline{B}_{\delta'}(\underline{p})$.

By Stirlings formula, if $t_i/n \in B_{\delta'}(p_i)$

$$\frac{n!}{t_0! \ldots t_{s-1}!} \sim \frac{(2\pi)^{\frac{1}{2}(s-1)} . n^{n+\frac{1}{2}}}{(t_0)^{t_0+\frac{1}{2}} \ldots (t_{s-1})^{t_{s-1}+\frac{1}{2}}}$$

$$= (2\pi)^{\frac{1}{2}(s-1)} . n^{\frac{1}{2}} . (t_0 . . t_{s-1})^{-\frac{1}{2}} . ((t_0/n)^{t_0} . . . . (t_{s-1}/n)^{t_{s-1}})^{-1}$$

$$= (2\pi)^{\frac{1}{2}(s-1)} . n^{(1-s)\frac{1}{2}} . (r_0 . . r_{s-1})^{-\frac{1}{2}} . (r_0^{r_0} . . . . r_{s-1}^{r_{s-1}})^{-n}$$

where $r_i = t_i/n$

$$< (2\pi)^{\frac{1}{2}(s-1)} n^{(1-s)\frac{1}{2}} . (\textstyle\prod_0^{s-1} (p_i - \delta'))^{-\frac{1}{2}} . s^{\alpha_{\underline{r}} n} \quad \text{since } \underline{r} \in B_{\delta'}(\underline{p})$$

$$< (2\pi)^{\frac{1}{2}(s-1)} n^{(1-s)\frac{1}{2}} . (\textstyle\prod_0^{s-1} (p_i - \delta'))^{-\frac{1}{2}} . s^{\alpha_{\underline{q}} n}$$

by definition of $\underline{q}$

The summation (3) has less than $(2\delta' n)^{s-1}$ terms, and so

$$\# X(n,\delta') \leq K n^{(1-s)\frac{1}{2}} n^{s-1} . s^{\alpha_{\underline{q}} n} \quad \text{some constant } K > 0$$

$$= K n^{(s-1)\frac{1}{2}} s^{\alpha_{\underline{q}} n} \qquad (4)$$

An interval of $X(n,\delta')$ is of length s^{-n}. Thus the d-measure estimate for A_δ arising from (2) is

$$\sum_{n=N}^{\infty} \frac{\# X(n,\delta')}{s^{nd}} \leq K . \sum_{n=N}^{\infty} n^{(s-1)\frac{1}{2}} \left(\frac{s^{\alpha_{\underline{q}}}}{s^d}\right)^n$$

$$< \infty \qquad \text{if} \quad d > \alpha_{\underline{q}}$$

The above estimates hold for all large N, so we have shown that for all $\delta' > \delta$,

$$\dim A_\delta \leq \sup \left\{ -H(\underline{q})/\log s \;\middle|\; \underline{q} \in B_{\delta'}(\underline{p}) \right\}$$

Letting $\delta' \to \delta$ gives the required upper bound. Eggleston's result (Eg) is that

$$\dim \left\{ y : \underline{P}_n(y) \to \underline{q} \right\} = -H(\underline{q}) / \log s \quad .$$

This gives the lower bound.

∎

<u>Lemma 4.3</u> Given $\varepsilon > 0$ and $N > 0$ there is a cover $\mathcal{U}$ of A_δ using intervals u of the following form,

i) u has length r^{-n} for some $n(u) > N$

ii) $u \subset [ks^{-n}, (k+1)s^{-n}]$ some $k \in \mathbb{Z}$.

iii) $\sum_{u \in \mathcal{U}} |u|^a < 1$ where $a = \dim A_\delta + \varepsilon$

<u>Proof:</u> Since $\dim A_\delta < a$, we can find a cover of A_δ, $\mathcal{U}'$, so that

$$v \in \mathcal{U}' \Rightarrow |v| < r^{-(N+1)} \qquad (1)$$

$$\sum_{v \in \mathcal{U}'} |v|^a < 2^{-1} r^{-a} \qquad (2)$$

For $v \in \mathcal{U}'$ there is an $n \in \mathbb{N}$ such that

$$r^{-(n+1)} < |v| \leqslant r^{-n} \quad .$$

Thus we can enclose v in an interval v' of length $r^{-n} < r^{-N}$ by (1). The interval v' may not satisfy (ii). Suppose $k \in \mathbb{Z}$ and $ks^{-n} \in v'$. Define

$$u_1 = \left[\frac{k}{s^n} - \frac{1}{r^n} , \frac{k}{s^n} \right] \quad , \qquad u_2 = \left[\frac{k}{s^n} , \frac{k}{s^n} + \frac{1}{r^n} \right]$$

Clearly $|u_1|^a + |u_2|^a = 2|v'|^a < 2r^a |v|^a$ (3).

Transforming every interval v of $\mathcal{U}'$ into intervals $u_1(v)$, $u_2(v)$ as above gives a new cover of A_δ,

$$\mathcal{U} = \left\{ u_1(v), u_2(v) : v \in \mathcal{U}' \right\}$$

which by construction satisfies (i), (ii). Furthermore

$$\sum_{u\in\mathcal{U}} |u|^a < \sum_{v\in\mathcal{U}'} 2r^a|v|^a < 1 \quad \text{by (2) and (3)}$$

■

<u>Notation:</u> Given $\underline{p}$ and $\varepsilon > 0$, choose the largest $\delta > 0$ so that

$$\sum_{i=0}^{s-1} \left(B_\delta(p_i) \cdot \frac{\log k_i}{\log r}\right) \subset B_\varepsilon(p)$$

where $p = \sum_{i=0}^{s-1} p_i\left(\frac{\log k_i}{\log r}\right)$. Then write

$$E_{\underline{p},\varepsilon} = \left\{(x,y) \in E : \overline{\lim_{n\to\infty}} \left(\sum_{i=0}^{s-1} \frac{P(y,n,i)}{n} \frac{\log k_i}{\log r}\right) \in B_\varepsilon(p)\right.$$

$$\left. \text{and } \underline{P}_n(y) \text{ has a limit point in } B_\delta(\underline{p})\right\}$$

<u>Lemma 4.4</u>. $\dim E_{\underline{p},\varepsilon} \leqslant \sup\left\{(-H(\underline{q})/\log s) + p + \varepsilon \;\middle|\; \underline{q} \in B_\delta(\underline{p})\right\}$
$= d_1$

<u>Proof:</u> Choose $\varepsilon' > \varepsilon$. Suppose y is such that

$$\overline{\lim_{n\to\infty}} \left(\sum_{i=0}^{s-1} \frac{P(y,n,i)}{n} \frac{\log k_i}{\log r}\right) \in B_\varepsilon(p).$$

Then there exists an N such that

$$n > N \Rightarrow \sum_{i=0}^{s-1} \frac{P(y,n,i)}{n} \frac{\log k_i}{\log r} < p + \varepsilon' \quad (1).$$

Let $A_N = \{y : (x,y) \in E_{\underline{p},\varepsilon} \text{ and } \forall\, n > N \text{ (1) holds}\}$

$E_N = \{(x,y) \in E_{\underline{p},\varepsilon} : y \in A_N\}$ some x.

We have $\bigcup_{n=1}^{\infty} E_N = E_{\underline{p},\varepsilon}$, and show that for all N $\dim E_N$ is less than d_1, for then $\dim E_{\underline{p},\varepsilon} \leqslant d_1$.

Since $A_N \subset A_\delta$, the set from lemma 4.2, $\dim A_N \leqslant \dim A_\delta$.

Lemma 4.3 guarantees the existence of a cover $\mathcal{U}$ of A_N with the properties stated there. Given an interval $u \in \mathcal{U}$ we define a collection of squares $\mathcal{U}(u)$ in the following way. Consider $S = [0,1] \times u$. S intersects E, and in particular intersects some n-cylinders C_n. $S \cap C_n$ is covered by a square of side r^{-n}. Let

$$\mathcal{U}(u) = \{v : v \text{ is a square of side } r^{-n} \text{ covering } S \cap C_n, \text{ some n-cylinder } C_n\}$$

$$\# \mathcal{U}(u) = \# \{ \text{n-cylinders intersecting } S \} = k_0^{P(y,n,0)} k_1^{P(y,n,1)} \cdots k_{s-1}^{P(y,n,s-1)} \quad \text{if } y \in u \qquad (2).$$

Since u covers part of A_N, we may assume $y \in u \cap A_N$ in (2).

Take $a > \dim A_\zeta$, $b > p + \varepsilon'$. The collection $\mathcal{U}_* = \bigcup_{u \in \mathcal{U}} \mathcal{U}(u)$ covers E_N and we estimate (a+b)-dimensional Hausdorff measure using $\mathcal{U}_*$,

$$\sum_{v \in \mathcal{U}_*} |v|^{a+b} = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{U}(u)} |v|^{a+b} \qquad (3).$$

Now $\sum_{v \in \mathcal{U}(u)} |v|^{a+b} < r^{-n(a+b)} 2^{\frac{a+b}{2}} \prod_{i=0}^{s-1} k_i^{P(y,n,i)}$, $y \in u \cap A_N$

But

$$\frac{1}{n} \log \left(r^{-nb} \prod_{i=0}^{s-1} k_i^{P(y,n,i)} \right) < 0$$

$$\text{if and only if} \quad b > \sum_0^{s-1} \frac{P(y,n,i)}{n} \frac{\log k_i}{\log r} \qquad (4).$$

Since $y \in u \cap A_N$,

$$\sum_0^{s-1} \frac{P(y,n,i)}{n} \frac{\log k_i}{\log r} < p + \varepsilon' \qquad \text{from (1) and } n > N,$$

$$< b \quad \text{by choice of } b.$$

Thus (4) holds, so

$$r^{-nb} \prod_{i=0}^{s-1} k_i^{P(y,n,i)} < 1$$

which implies that $\sum_{v \in \mathcal{U}(u)} |v|^{a+b} < r^{-na} 2^{\frac{a+b}{2}} = |u|^a 2^{\frac{a+b}{2}}$

and

$$\sum_{v \in \mathcal{U}_*} |v|^{a+b} < 2^{\frac{a+b}{2}} \sum_{u \in \mathcal{U}} |u|^a \qquad \text{from} \qquad (3)$$

$$< 1.2^{\frac{a+b}{2}} \qquad \text{by choice of } \mathcal{U}.$$

This holds for all $\varepsilon' > \varepsilon$, which proves the lemma.

■

We are now able to calculate dim E. F. Ledrappier showed me how to obtain the lower bound using Marstrands theorem (Mr).

<u>Thrm 4.5</u> $$\dim E = \sup_{\underline{p}} \left(\frac{-\sum_0^{s-1} p_i \log p_i}{\log s} + \sum_0^{s-1} p_i \frac{\log k_i}{\log r} \right)$$

$$= d$$

<u>Proof:</u> We first obtain the upper bound $\dim E \leq d$. Consider the subspace of $\mathbb{R}^{s+1} = \{(\underline{p},p) \in \mathbb{R}^s \times \mathbb{R}\}$ defined by letting $A \subset \mathbb{R}^{s+1}$ be the set of $(\underline{p},p)$ such that

$0 = -p + \sum_0^{s-1}(p_i \log k_i) / \log r,$

Let $B = A \cap \left(\{(p_0,\dots,p_{s-1}) : \sum_0^{s-1} p_i = 1,\ p_i \geq 0\} \times [0,1] \right)$

For any point $(x,y) \in E$ there is a $(\underline{p},p) \in B$ such that

$$\overline{\lim_{n\to\infty}} \left(\sum_0^{s-1} \frac{P(y,n,i)}{n} \frac{\log k_i}{\log r} \right) = p$$

and $\underline{P}_n(y)$ has $\underline{p}$ as a limit point.

We estimate dim E by partitioning B into a finite number of subsets. Given $\varepsilon>0$ and $(\underline{p},p)\in B$, there exists $\delta>0$ so that $\sum_i B_\delta(p_i)(\log k_i)/\log r \subset B_\varepsilon(p)$. Thus given $\varepsilon>0$ we can cover B by $\bigcup_{(\underline{p},p)} B_\delta(\underline{p}) \times B_\varepsilon(p)$. Since B is compact we require only a finite number of such balls,

$$B \subset \bigcup_k \left(B_\delta(\underline{p}(k)) \times B_\varepsilon(p(k))\right) \quad k=1,..,m \quad \text{some } m.$$

By lemma 4.4,

$$\dim E \leqslant \sup_k \left(\sup\left\{(-H(\underline{q}) / \log s) + p(k) + \varepsilon : \underline{q} \in B_\delta(\underline{p}(k))\right\}\right)$$

We now let $\varepsilon \to 0$ taking finer and finer covers of B. This proves $\dim E \leqslant d$.

We now have to show $\dim E \geqslant d$. Given $y \in [0,1]$ define $E_y = E \cap \{(x,y) : x \in \mathbb{R}\}$. Suppose that $\underline{P}_n(y) \to \underline{p} = (p_0,..,p_{s-1})$. We first show that

$$\dim E_y \geqslant \sum_{i=0}^{s-1} p_i \frac{\log k_i}{\log r} = d_2 .$$

Each n-cylinder C_n of E intersects E_y in a set contained in an interval I_n of length r^{-n}. Define μ_n, a measure with support on $\bigcup I_n$, as Lebesgue measure on $\bigcup I_n$ multiplied by $r^{-n}.\#\{I_n\}$. The space of probability measures on $[0,1] \times \{y\}$ in the weak* topology is compact and so we can find a subsequence μ_{n_i} converging to a probability measure μ with support on E_y. Since for each I_n, I_n' we have $(I_n) \cap E = x + (I_n') \cap E$ for some x,

$$\mu(I_n) = \mu(I_n') = 1/\#\{I_n\} = 1/\prod_0^{s-1} k_i^{P(y,n,i)} .$$

Given $\delta > 0$, for all $n > N$

$$\sum_i \frac{P(y,n,i) \log k_i}{n} \geqslant \sum_i p_i \log k_i - \delta .$$

Take $\delta = \varepsilon . \log r$. We claim there is a constant $c > 0$ so that for $n > N$, $\mu(I_n) \leqslant c\,|I_n|^{d_2 - \varepsilon}$. Then Frostman's lemma shows that $\dim E_y \geqslant d_2$.

$$\mu(I_n) \leqslant c\,|I_n|^{d_2 - \varepsilon} = c\, r^{-n(d_2 - \varepsilon)}$$

if $\log \mu(I_n) \,/\, \log \left(r^{-n(d_2 - \varepsilon)} \right) \geqslant 1$. But

$$\frac{\log \mu(I_n)}{-n(d_2 - \varepsilon)\log r} = \frac{\sum_i P(y,n,i) \log k_i}{n\left(\sum_i p_i \left(\frac{\log k_i}{\log r} \right) - \varepsilon \right) \log r}$$

$$= \frac{\sum_i (1/n) \cdot P(y,n,i) \log k_i}{\sum_i p_i \left(\frac{\log k_i}{\log r} \right) - \delta} \geqslant 1$$

Thus $\dim E_y \geqslant d_2$. Eggleston 's result (Eg) says that $A_{\underline{p}} = \{ y : \underline{P}_n(y) \to \underline{p} \}$ has $\dim A_{\underline{p}} = -H(\underline{p}) / \log s$. Each $y \in A_{\underline{p}}$ has $\dim E_y \geqslant d_2$. By Marstrands theorem (Mr)

$$\dim \left(\bigcup_{y \in A_{\underline{p}}} E_y \right) \geqslant \frac{-H(\underline{p})}{\log s} + \sum_{i=0}^{s-1} p_i \frac{\log k_i}{\log r}$$

Since for all $\underline{p}$, $\left(\bigcup_{y \in A_{\underline{p}}} E_y \right) \subset E$, we have proved $\dim E \geqslant d$ ∎

We now give some examples,

i) <u>Products of Cantor sets</u>. In the example given in fig. 32 it is easy to see that E is the product of a Cantor set of

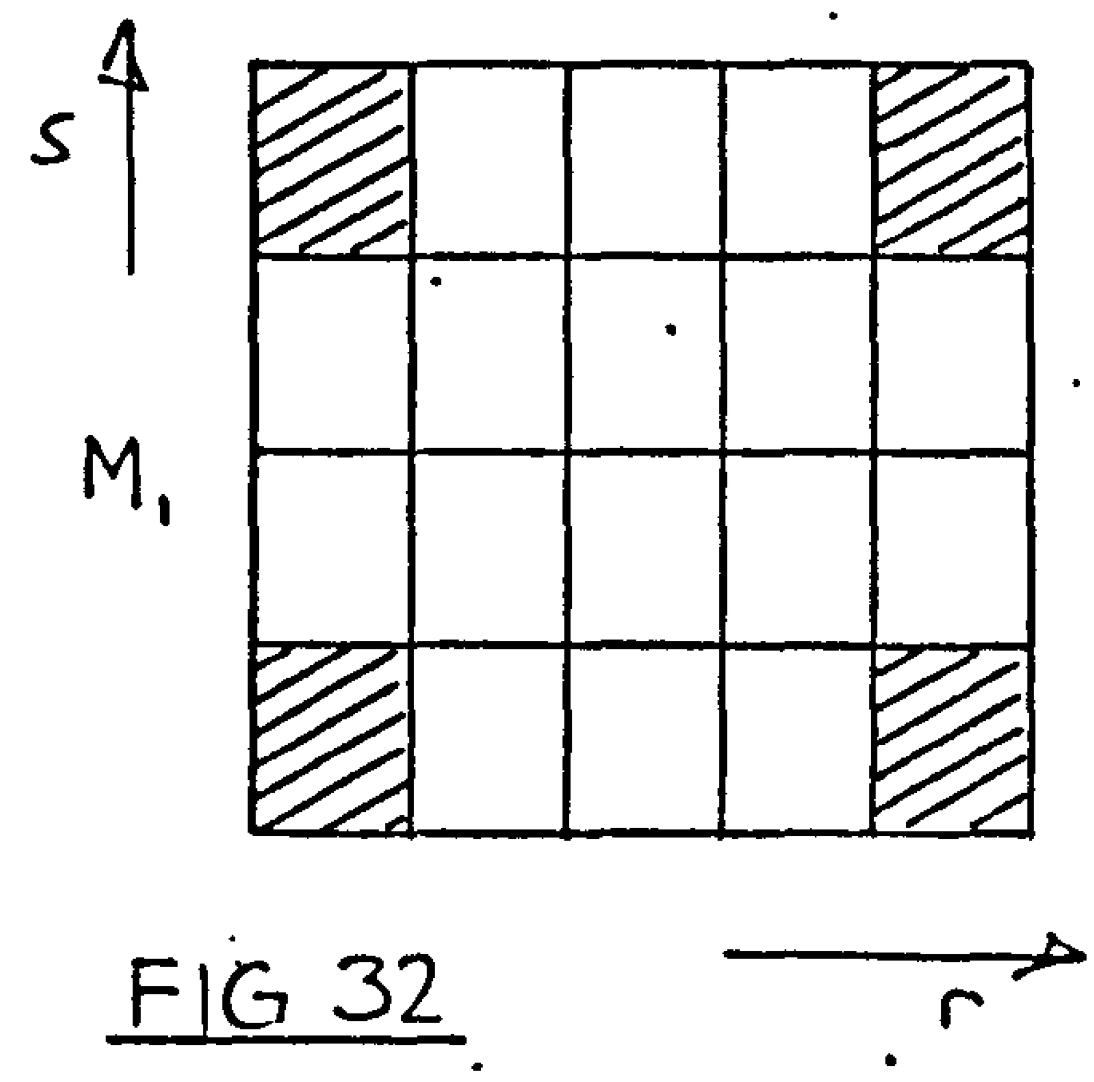


FIG 32

dimension $(\log 2)/(\log r)$ with another Cantor set of dimension $\log 2/\log s$. It is well known that the dimension of the product is $\dfrac{\log 2}{\log s} + \dfrac{\log 2}{\log r}$. Since $k_i = 2$ or 0 for each i, the sup in our formula occurs when $p_0 = p_{s-1} = \frac{1}{2}$. Thus our formula gives the dimension as $\dfrac{\log 2}{\log s} + \dfrac{\log 2}{\log r}$. Proposition 4.1 gives cap E = dim E.

2)

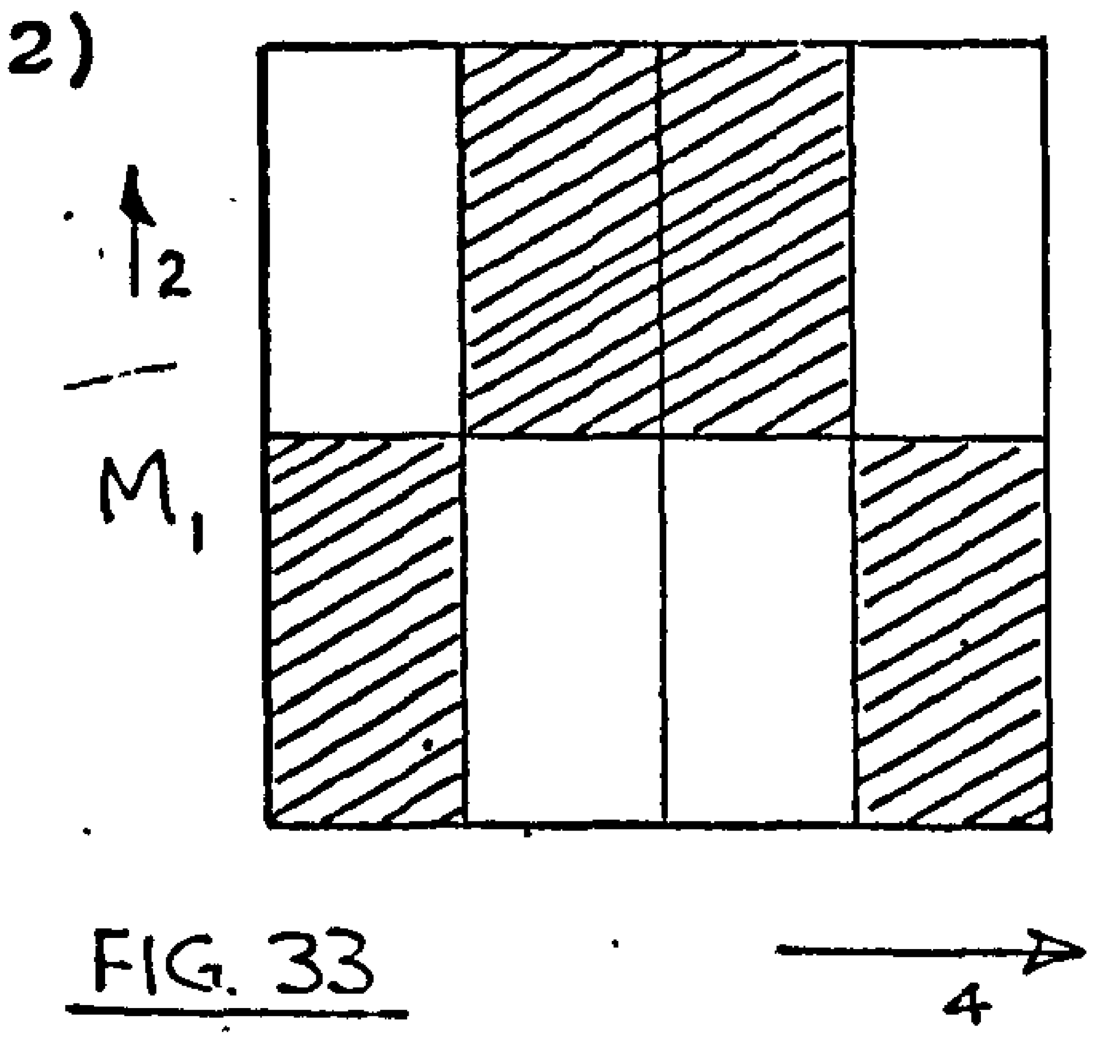


FIG. 33

We may have cap = dim when E is not a product of Cantor sets.

Here we have

cap E = dim E = 3/2

3)

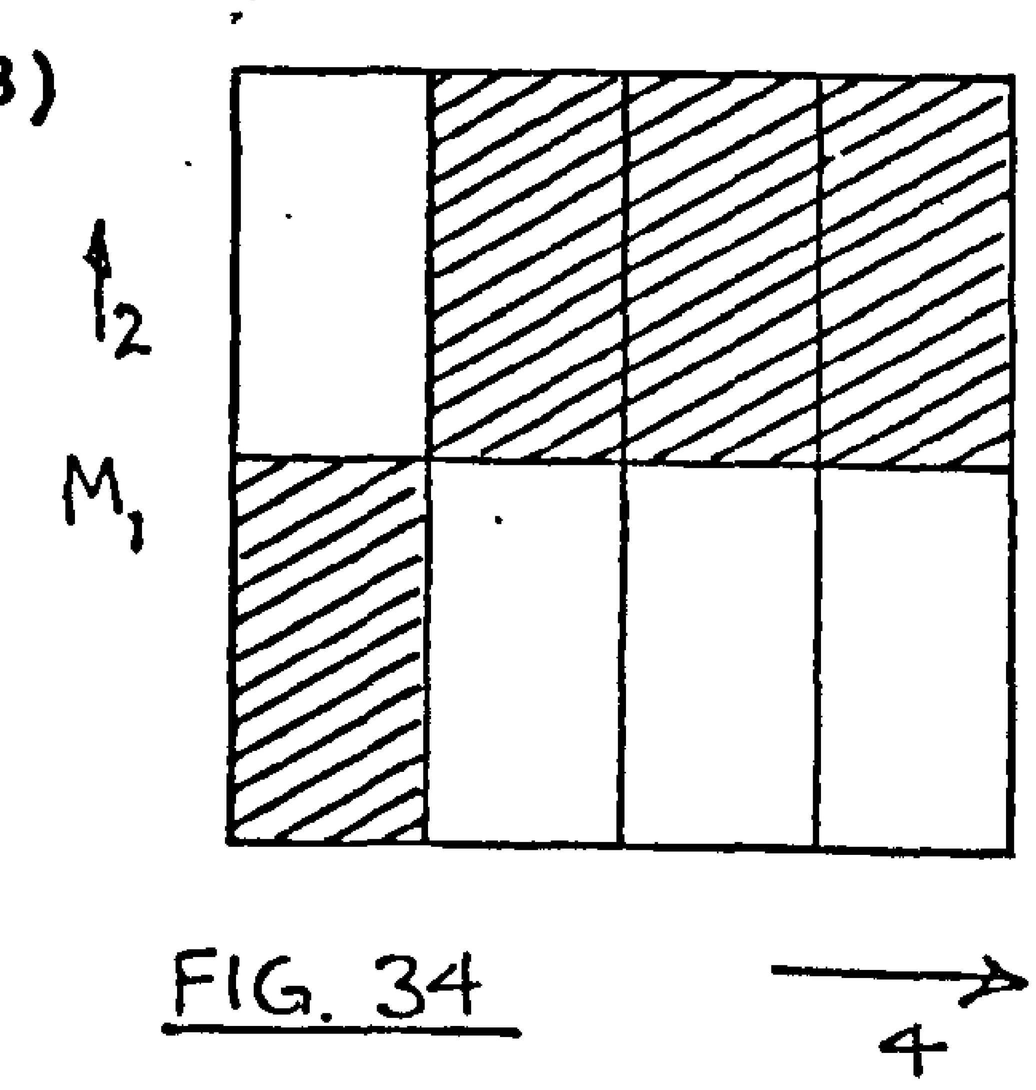


FIG. 34

Capacity may not equal dimension. Here cap E = 3/2, but dim E $\approx$ 1·45

It might be possible to calculate dimension for some recurrent sets not of the form assumed above by finding a lippeomorphism (a dimension preserving map) to a recurrent set with the right structure. In the case of Kiesswetter's curve, K,(fig 35; defined in (Ke)) no such obvious map exists because sets of the above form are not graphs of continuous functions. However we can still calculate dim K by a variant of 4.1. The same technique works for at least countably many Kiesswetter-type curves although we only give it here for K. Dekking (D1) showed how to construct K as a recurrent set and raised the question of calculating dim K as K was the only curve defined in (D1) for which he could not prove the dimension estimate to be correct.

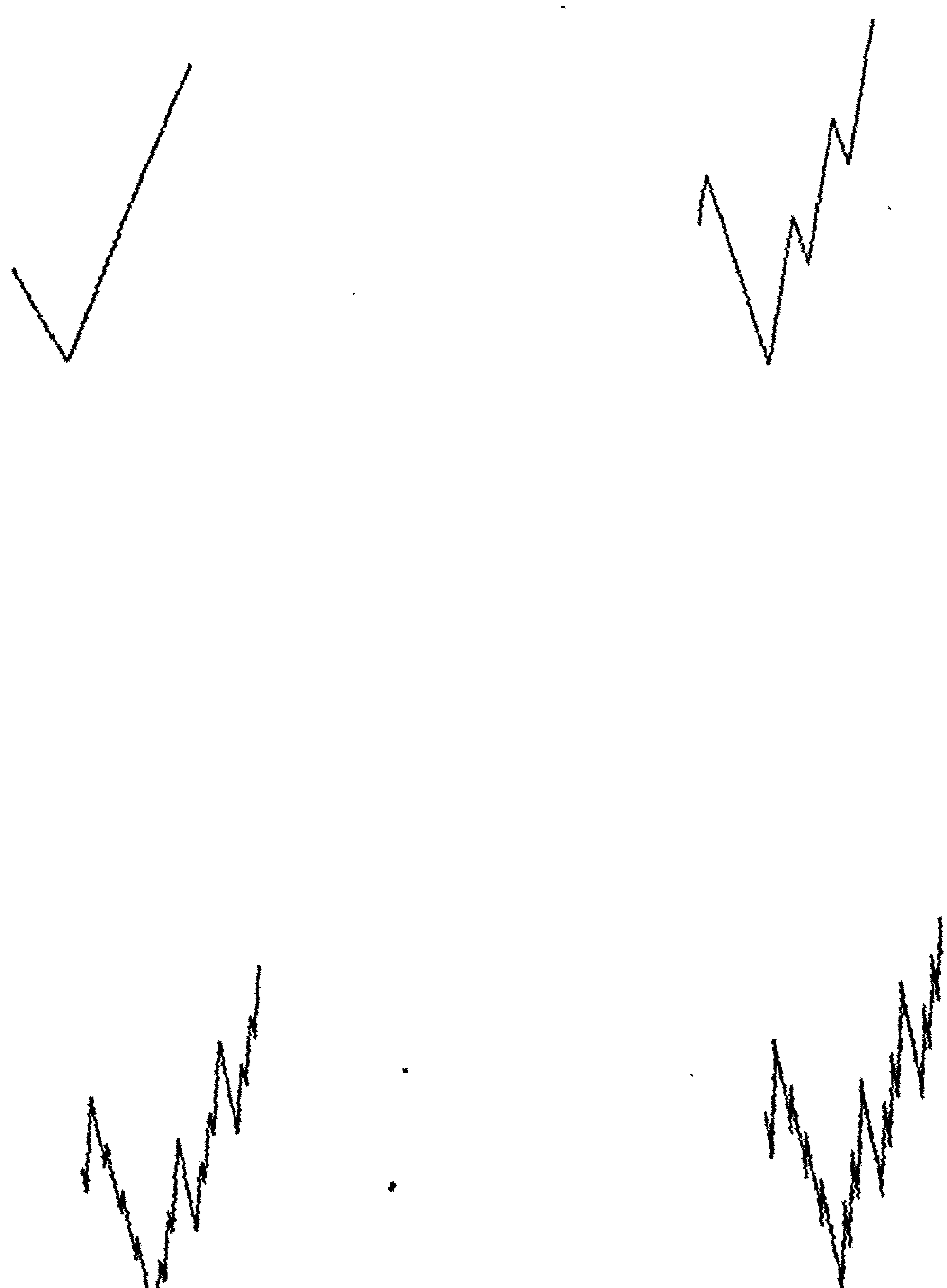
The Kiesswetter curve can be defined as follows (Ke). Given a point $x \in [0,1]$, write $x = 0 \cdot x_1 x_2 \ldots$ base 4. Define $X_i = 0$ if $x_i = 0$ or $X_i = x_i - 2$ otherwise. Let $N_i = \# \{ 1 \leq j \leq i : x_j = 0 \}$. Then write

$$k(x) = \sum_{1}^{\infty} (-1)^{N_i} X_i / 2^i \quad .$$

K is the graph of k(x). Alternatively let $S = \{a, b\}$, define $f(a) = (1,1)$, $f(b) = (1,-1)$ and $\theta(a) = baaa$, $\theta(b) = abbb$. If $L = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$, $K = K_\theta(a)$.

Our previous treatment of recurrent sets gives a subshift of finite type projecting onto K. However, because of the symmetry between $\theta(a)$ and $\theta(b)$,

FIG. 35



$$K_\theta(b) = A(K_\theta(a))$$

where $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and we can ignore the difference between a and b. There is thus a projection π to $K_\theta(a)$ from Σ_4,

$$\pi((x_1, x_2, \ldots)) = 0{\cdot}x_1 x_2 \ldots .$$

If we define $\tilde{L}: K \to K$ by identifying $K_\theta(b)$, $f(b) + K_\theta(a)$ $f(ba) + K_\theta(a)$, and $f(baa) + K_\theta(a) \subset L(K)$ in the obvious way, $\pi\sigma = \tilde{L}\pi$,

<u>Prop. 4.6</u> dim K = cap K = 3/2

<u>Proof:</u> For K the dimension estimate (*) is 3/2 so dim K $\leqslant$ cap K $\leqslant$ 3/2. We show dim K $\geqslant$ 3/2 by Frostman's lemma using $\pi_* \nu$ where ν is the Bernoulli measure on Σ_4 giving each n-cylinder measure 4^{-n} (the same measure is obtained by pushing Lebesgue measure onto K using the map $x \mapsto (x, k(x))$)..

From the formula for k(x) it is clear that $\sup_x k(x) = 1$ and that the sup is reached for x=1. Similarly $\inf_x k(x) = -1$ (when $x = 0{\cdot}11\dot{1}$). Let $D_{-1} = [0,1] \times [-1,1]$. Pulling D_{-1} back via $\tilde{L}^{-1}$ (extending the domain of $\tilde{L}$ in the obvious way) gives rectangles D_0 (fig 36). Continuing this process gives a sequence of covers $\mathcal{D}_n$ of K. Each $D_n \in \mathcal{D}_n$ has sides of length $4^{-(n+1)}$ and 2^{-n}. We divide each D_n into 2^{n+2} squares of side $4^{-(n+1)}$. Such a square we call an F_n. In order to apply Frostman's lemma we must calculate $\pi_* \nu(F_n)$.

COVERING KIESSWETTER'S CURVE

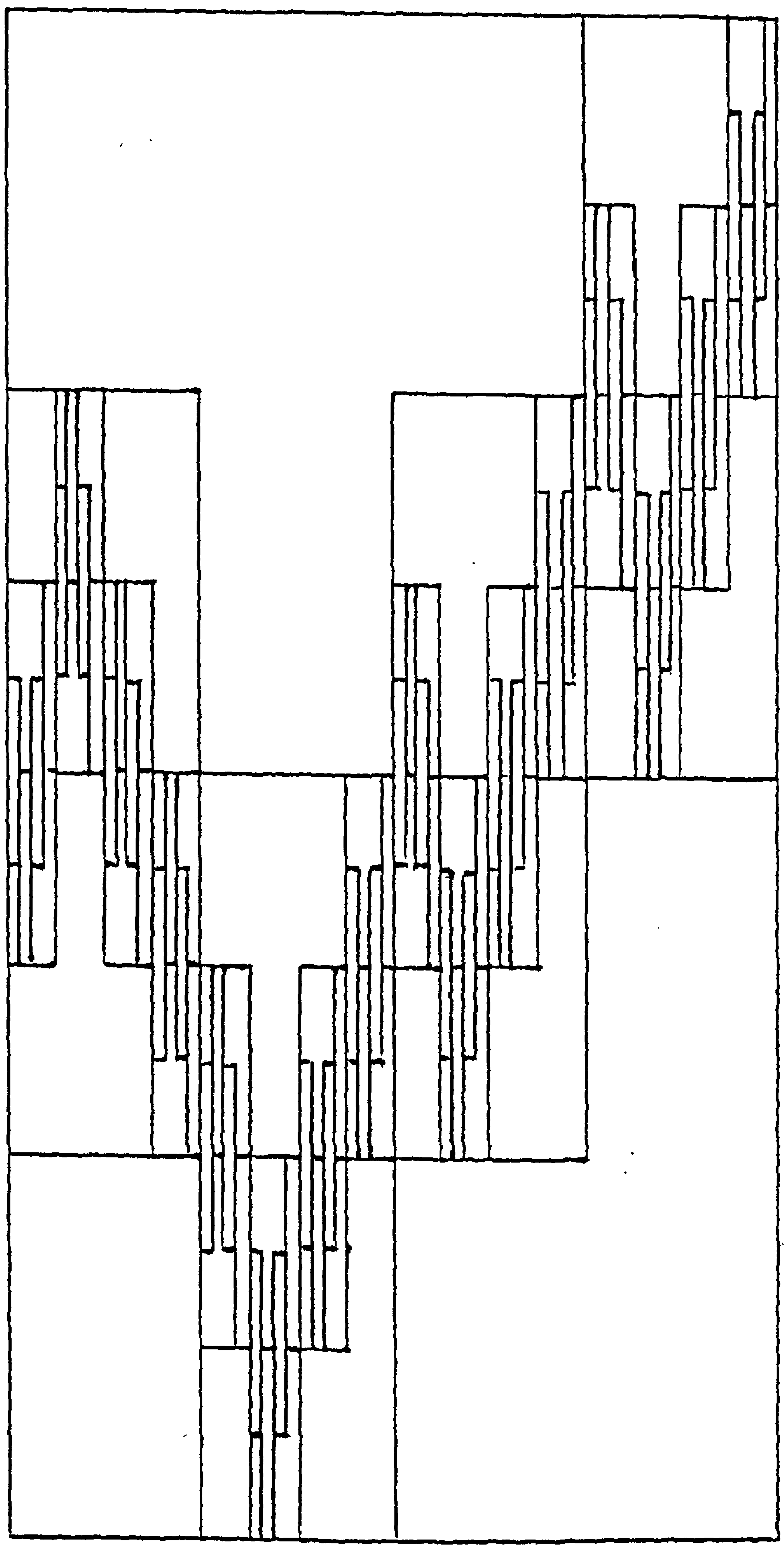


FIG. 36

Define $G_{n,i} = [0,1] \times [i/2^{n+1}, (i+1)/2^{n+1}]$ $i=-2^n,\ldots,2^n-1$. Then for any F_n, $\tilde{L}^{n+1}F_n = G_{n,i}$, some i, sub-cylinders of F_n map onto sub-cylinders of $G_{n,i}$ and $D_m \subset F_n$ map onto $D_{m'} \subset G_{n,i}$. In particular if $m = 2(n+1)$, $m' = n+1$ (fig 37).

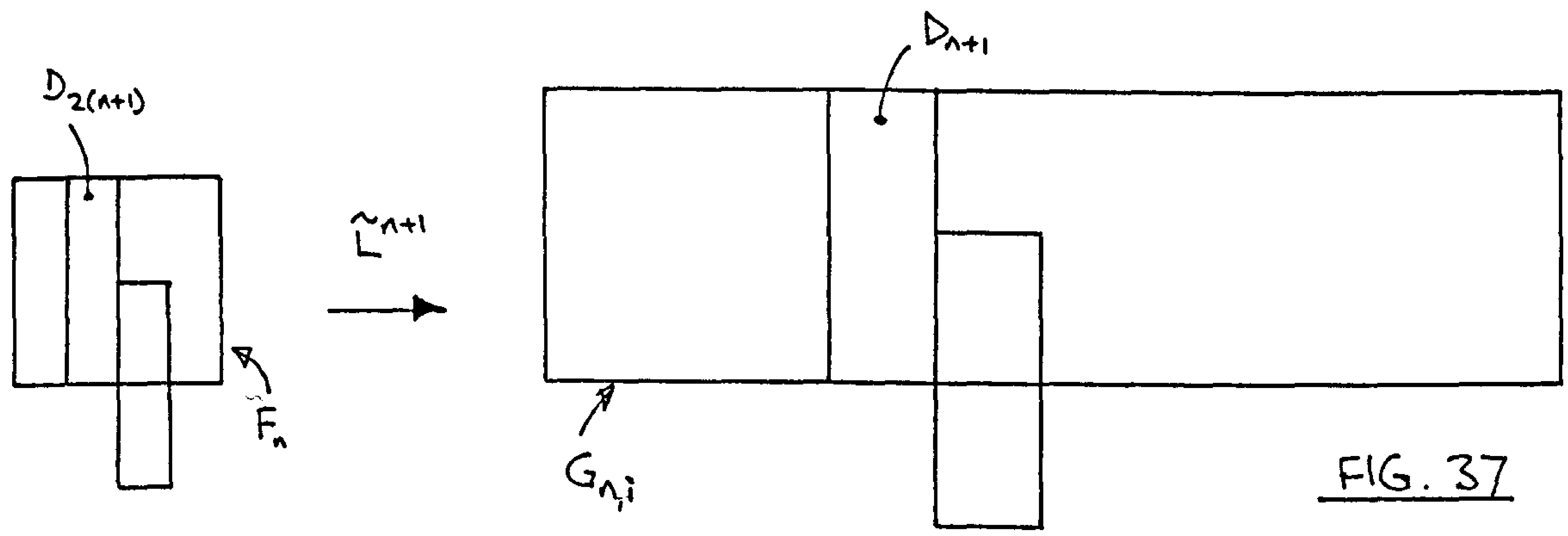


FIG. 37

We must now count the number of sub-cylinders of $G_{n,i}$. This number is greatest for $i=-1,0$ because each D_m has two D_{m+1} in the middle of D_m and one D_{m+1} at either end (fig 36). Let a_n denote the number of complete D_{n+1}'s in $G_{n,0}$, b_n denote the number of D_{n+1}'s whose lower half intersects $G_{n,0}$, and c_n denote the number of D_{n+1}'s whose upper half intersects $G_{n,0}$. Then

1) $a_{n+1} = a_n + c_n$ (since a complete D_{n+1} in $G_{n,0}$ contains one D_{n+2} lying in $G_{n+1,0}$, and each D_{n+1} with upper half lying in $G_{n,0}$ has one D_{n+2} in $G_{n+1,0}$).

2) $b_{n+1} = 2a_n$ (since a complete D_{n+1} in $G_{n,0}$ contains two D_{n+2}'s lying about the line $y=2^{-(n+2)}$.

3) $c_{n+1} = 2c_n$ (since each D_{n+1} with upper half in $G_{n,0}$ contains two D_{n+2}'s with upper half in G_{n+1}).

We have $a_0 = 3$, $b_0 = 2$, $c_0 = 4$ and it is easy to check that $c_n = 2^{n+1}$, $b_n = 6 + 8(2^{n-2}-1)$, $a_n = 3 + 4(2^{n-1}-1)$

From the way we defined ν and the symmetric distribution of D_m in a D_n $(m > n)$ it is easy to see that the $\pi_* \nu$ measure of the upper or lower half of a D_n is half that of the D_n. Hence,

$$\pi_* \nu(F_n) \leqslant \left(a_n + \frac{b_n + c_n}{2}\right) \pi_* \nu(D_m) \qquad m=2(n+1)$$

$$\leqslant \left(a_n + \frac{b_n + c_n}{2}\right) \cdot \frac{1}{4^{2n+3}}$$

$$= \text{const.} 2^{-3n}$$

Thus $\qquad \pi_* \nu(F_n) \leqslant \text{const.} |F_n|^{3/2}.$

Now if U is an open ball of diameter t choose n with $t \in [4^{n+1}, 4^n]$. Then $t \geqslant \text{const.} 4^{-n}$. Also since $t \leqslant 4^{-n}$ we need at most four F_{n-1}'s to cover U. We therefore have

$$|U|^{3/2} = t^{3/2} = \text{const.} 4 . 4^{-n3/2} = \text{const.} \pi_* \nu(F_{n-1})$$
$$\geqslant \text{const.} \pi_* \nu(U).$$

Frostman's lemma now gives $HM_{3/2}(K) > 0$ which proves the proposition.

■

REFERENCES.

(Ad1) R.L. Adler and B. Weiss, Similarity of automorphisms of the torus, AMS Memoirs 98 (1970).

(Ad2) R.L. Adler and B. Marcus, Topological equivalence of dynamical systems, AMS Memoirs 219 (1979).

(Be) A. Besicovitch and H.Ursell, Sets of fractional dimensions (V): On dimensional numbers of some continous curves, JLMS 12 (1937) 18-25.

(Bo1) R. Bowen, Markov partitions for Axiom A diffeomorphisms, AJM 92 (1970) 725-747.

(Bo2) R. Bowen, Markov partitions and minimal sets for Axiom A diffeomorphisms, AJM 92 (1970) 907-918.

(Bo3) R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Springer Lecture Notes in Maths 470 (1975).

(Bo4) R. Bowen, On Axiom A Diffeomorphisms, Conf. Board in Math. Sci. Regional Conf. Series, 35 (1977).

(Bo5) R. Bowen, Markov partitions are not smooth, Proc. AMS 71 (1978) 130-132.

(Bo6) R. Bowen, Hausdorff dimension of quasi-circles, IHES Publ. Math. 50 (1979) 11-25

(Bo7) R. Bowen, Entropy versus homology for certain diffeomorphisms, Topology 13 (1974) 61-67.

(Br) D.S. Broomhead and G. Rowlands, On the use of perturbation theory in the calculation of the fractal dimension of strange attractors, Physica 10D (1984) 340-352.

(Ca) L. Carleson, Selected problems on exceptional sets, Van Nostrand, 1967, Princeton.

(D1) F.M. Dekking, Recurrent sets, Adv. in Maths.44 (1982) 78-104.

(D2) F.M. Dekking, Replicating superfigures and endomorphisms of free groups, J. Combin. Thry. Ser. A 32 (1982) 315-320.

(D3) F.M. Dekking, Recurrent sets : A fractal formalism. Report 82-32 Technische Hogeschool Delft (1982).

(Eg) H. G. Eggleston, The Fractional Dimension of a set defined by decimal properties, Quart. J. Math. 20, (1949) 31-39

(Fa) J.D. Farmer, E. Ott, and J.A. Yorke, The dimension of chaotic attractors, Physica 7D (1983) 153-180.

(Fr) J. Franks, Invariant sets of hyperbolic toral automorphisms, AJM 99 (1977) 1089-1095.

(Fu) H. Furstenburg, Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation, Math. Systems Thry. 1 (1967) 1-49.

(Ga) F.R. Gantmacher, Matrix Theory Vol. 2, Chelsea NY, 1959.

(Gr1) P. Grassberger, On the fractal dimension of the Henon attractor, Phys. Lett. 97A (1983) 224-226.

(Gr2) P. Grassberger, Generalized dimensions of strange attractors, Phys. Lett. 97A (1983) 227-230.

(Ha1) S.G. Hancock, Construction of invariant sets for Anosov diffeomorphisms, JLMS(2) 18 (1978) 339-348.

(Ha2) S.G. Hancock, Ph.D Thesis, Warwick University 1979.

(Hi1) M.W. Hirsch, On invariant subsets of hyperbolic sets, Essays in Topology and related topics, Springer-Verlag Berlin, 1970.

(Hi2) M.W. Hirsch and C. Pugh, Stable manifolds and hyperbolic sets, Proc. Symp. Pure Maths, 14 (1970) 133-163.

(Hur) W. Hurewicz and H. Wallman, Dimension Theory, Princeton 1948.

(Hut) J.E. Hutchinson, Fractals and self-similarity, Indiana Univ. Maths Jour. 30 (1981) 713-747.

(Ke) K. Kiesswetter, Ein einfaches Beispiel fur eine Funktion, welche uberall stetig und nicht differenzierbar ist, Mathematisch Physikalische Semesterberichte 13

(Ir) M. Irwin, The orbit of a Holder continuous path under a hyperbolic toral automorphism, Erg. Thry and dyn. syst's 3 (1983) 345-351

(Ma1) B. Mandlebrot, Fractals, form, chance and dimension, W.H. Freeman, San Fransisco 1977.

(Ma2) B. Mandlebrot, The fractal geometry of nature, W.H. Freeman, San Fransisco 1982.

(Me1) R. Mañé, Invariant sets of Anosov diffeomorphisms, Inventiones Math. (46) 1978, 147-152.

(Me2) R. Mañé, Orbits of paths under hyperbolic toral automorphisms, Proc AMS 73 (1979) 121-125.

(Mn1) A.K. Manning, There are no new Anosov diffeomorphisms on tori, AJM 96 (1974) 422-429.

(Mn2) A.K. Manning, Lecture notes, M.Sc. course on dynamical systems, 1983.

(Mr) J.M. Marstrand, The dimension of Cartesian product sets, Proc. Camb. Phil. Soc., 50 (1954) 198-202.

(Mo) P.A.P. Moran, Additive functions of intervals and Hausdorff measures, Proc. Camb. Phil. Soc. 42 (1946) 15-23.

(Pr) F. Przytycki, Construction of invariant sets for Anosov diffeomorphisms and hyperbolic attractors, Studia Math. 68 (1980) 199-213.

(Ro) C.A. Rogers, Hausdorff Measures, CUP, 1970.

(Ru) D. Ruelle, Bowen's formula for the Hausdorff dimension of self similar sets, in Scaling and self similarity in Physics Ed. J. Frohlich, Progress in Physics Vol. 7 Birkhauser Verlag, Basel, 1983.

(Se) E. Senata, Non negative matrices, George Allen and Unwin, 1973.

(Si1) Ya Sinai, Markov partitions and C-diffeomorphisms, Func. Anal. and its Appl. 2 (1968) 64-89.

(Si2) Ya Sinai, Construction of Markov partitions, Func. Anal. and its Appl. 2 (1968) 245-253.

(Sm) S. Smale, Differentiable dynamical systems, Bull. AMS 73 (1967) 747-817.

(Ta) F. Takens, Detecting strange attractors in turbulence, Springer lecture notes in maths 898, Dynamical systems and turbulence, Warwick 1980 Proc..

(Ub) M. Urbański, On capacity of continuum with non-dense orbit under a hyperbolic toral automorphism, Preprint 289, Inst. of Maths., Polish Academy of Sciences, November 1983.

(Wa1) P. Walters, A variational principle for the pressure of continuous transformations, AJM 97 (1976) 937-971

(Wa2) P. Walters, An introduction to ergodic theory, Graduate Texts in Maths 79, Springer Verlag NY, 1981.

(Wi) T.A. Witten and L.M. Sander, Diffusion-limited aggregation, a kinetic critical phenomenon, Phys. Rev. Lett. 47 (1981) 1400-1403.