INFINITE-TO-ONE CODES AND MARKOV MEASURES

BY

MIKE BOYLE AND SELIM TUNCEL[1]

ABSTRACT. We study the structure of infinite-to-one continuous codes between subshifts of finite type and the behaviour of Markov measures under such codes. We show that if an infinite-to-one code lifts one Markov measure to a Markov measure, then it lifts each Markov measure to uncountably many Markov measures and the fibre over each Markov measure is isomorphic to any other fibre. Calling such a code Markovian, we characterize Markovian codes through pressure. We show that a simple condition on periodic points, necessary for the existence of a code between two subshifts of finite type, is sufficient to construct a Markovian code. Several classes of Markovian codes are studied in the process of proving, illustrating and providing contrast to the main results. A number of examples and counterexamples are given; in particular, we give a continuous code between two Bernoulli shifts such that the defining vector of the image is not a clustering of the defining vector of the domain.

0. Introduction. This paper is concerned with the structure of infinite-to-one continuous codes between subshifts of finite type and with the behaviour of Markov measures under such codes. The emergence of notable interest in infinite-to-one codes is recent. This interest was partly fueled by the development of Krieger's marker method [11, 12] and is reflected by the papers [3 and 15]. Previous research had concentrated on bounded-to-one codes and yielded considerable information about these. We list [1, 4, 7–10, 16–20] for examples of material dealing with bounded-to-one codes. The properties of bounded-to-one codes constituted one of our main motivations; it may be useful to compare our results with these.

After a section on definitions and notation, we will give three examples. The first example concerns infinite-to-one analogues of the resolving codes used in [1, 16 and 17]. The main point about these is that they lift each Markov measure to uncountably many Markov measures. In [15], Marcus, Petersen and Williams construct infinite-to-one codes through which no Markov measure lifts to a Markov measure. Recall that, through a bounded-to-one code, each Markov measure lifts to a unique Markov measure [20]. This lifting process does not alter the memory of the measure when the code is 1-block. However, in the infinite-to-one case there may be strict increases in memory not warranted by the block length of the code: our third (peculiar memory) example is a 1-block code such that no 1-step Markov measure lifts to a 1-step Markov measure, while each 1-step Markov measure has uncountably many 2-step preimages. In addition, this code sends every 1-step Markov

measure on the domain to a 2-step Markov measure. Our second example concerns Bernoulli shifts. Between two Bernoulli shifts, we construct an infinite-to-one block code which is a counterexample to one of the conjectures in [7].

In §3 we use pressure to study the fibres of Markov measures. We describe the fibres in terms of pressure and information cocycles, and show that all fibres are isomorphic. In particular, we show that if one Markov measure lifts to a Markov measure then so must all other Markov measures; let us agree to call such codes Markovian. There is a characterization of Markovian codes in terms of pressure. In §7 we consider certain codes which, for reasons that will be made clear later, we dub "magic diamond codes". Magic diamond codes send every 1-step Markov measure down to a Markov measure. We show that every infinite-to-one code may be intercepted by a code of this kind, and it follows that for a Markovian code the fibres must contain uncountably many Markov measures.

Parry has introduced and exploited a cocycle-coboundary equation to use information, rather than just entropy, as an invariant of various entropy-preserving codes (see [19]). In §4 we focus on infinite-to-one (hence, entropy-reducing [16]) block codes between Markov chains and we show that an inequality version of Parry's cocycle-coboundary equation is valid. Combined with the pressure conditions of §3, this inequality leads to necessary conditions on the β-functions [17, 20] of the Markov chains.

It was proved in [3] that, from a purely topological point of view, the existence of infinite-to-one codes between subshifts of finite type is governed by a simple condition on periodic points. We demonstrate in §5 that this simple condition persists after restricting to Markovian codes. To do this, we introduce infinite-to-one analogues, which we call continuing codes, of the closing codes of Kitchens [8] and Marcus [14]. Continuing codes are conjugates of the resolving codes of §2 and, hence, are Markovian. We show that the techniques of [3] may be refined to construct a continuing code whenever the necessary periodic point condition is satisfied.

A Perron number is a positive algebraic integer strictly greater than the modulus of any of its conjugates. Lind recently proved [13] that the topological entropies taken on by aperiodic subshifts of finite type are precisely the logarithms of Perron numbers. In §6 we consider "uniform" codes, which send the measure of maximal entropy to the measure of maximal entropy. We prove that if there exists such a code from an aperiodic subshift of finite type with topological entropy $\log \alpha$ onto one of topological entropy $\log \beta$ then α/β must be a Perron number. It follows that only finitely many values can occur as the topological entropies of uniform factors of subshifts of finite type with a given topological entropy $\log \alpha$.

We would like to thank Ramesh Gangolli and Doug Lind for their hospitality, interest and encouragement.

1. Definitions and notation. We shall assume that the reader is familiar with basic information theory and entropy theory, and use the usual related notation for partitions (see, for instance, Walters' book [21]).

Consider a finite set of k elements, $\{1, 2, \ldots, k\}$ say, and give it the discrete topology. The (*full*) *k-shift* is obtained by furnishing the set $\prod_{-\infty}^{\infty}\{1, 2, \ldots, k\}$ with

the product topology and the left shift transformation. A *subshift* S is obtained by restricting a k-shift, for some $k \in \mathbf{N}$, to a closed shift-invariant subset. We shall use the same symbol to denote the underlying space of a subshift and the restriction of the shift transformation to this space. The *alphabet* $\mathscr{A}(S)$ of the subshift S is the subset of $\{1, 2, \ldots, k\}$ consisting of those elements that appear in some point of S. An element of $\mathscr{A}(S)$ is a *symbol* of S. If $i_0, i_1, \ldots, i_{l-1} \in \mathscr{A}(S)$ are such that the string $i_0 i_1 \cdots i_{l-1}$ appears in some point of S, then $i_0 i_1 \cdots i_{l-1}$ is called a *word* (or *block*) of length l; when there is a need to emphasize S, we use the terms S-word and S-block. The closed-open set

$$[i_0 i_1 \cdots i_{l-1}] = \{x = (x_n) \in S : x_0 = i_0, x_1 = i_1, \ldots, x_{l-1} = i_{l-1}\}$$

is a *cylinder*; the *length* of this cylinder is l. A base for the topology of S is given by these sets and their translates. The *state partition* of S consists of the cylinders $[i]$, $i \in \mathscr{A}(S)$.

Let S be a $k \times k$ irreducible matrix of zeros and ones. The *subshift of finite type* (or *topological Markov chain*) defined by S is the restriction of the full k-shift to those sequences $x = (x_n)$ with $S(x_n, x_{n+1}) = 1$ for all $n \in \mathbf{Z}$. We shall use the same symbol to denote a subshift of finite type, its space, and its defining matrix.

If a subshift obtained from a full shift by disallowing finitely many words is topologically transitive (in the sense that it contains some point with a dense forward orbit), then it may be identified with a subshift of finite type. Except for a brief spell in §5, all the subshifts we consider will be of finite type.

A $k \times k$ zero-one matrix S yields a directed graph by taking k vertices and drawing a path from i to j if and only if $S(i, j) = 1$. Conversely, a directed graph with at most one path between any two vertices gives a zero-one transition matrix. In a number of examples, we shall find it convenient to specify subshifts of finite type through directed graphs.

If S is a subshift of finite type, its *n-block system* is the subshift of finite type S_n defined by the 0-1 transition matrix for S-words of length n: the rows and columns of the matrix S_n are indexed by S-words of length n and $S_n(x_1 x_2 \cdots x_n, x_1' x_2' \cdots x_n') = 1$ if and only if $x_1' = x_2, x_2' = x_3, \ldots, x_{n-1}' = x_n$. By definition, $S_1 = S$. The subshifts of finite type S_n, $n \geq 1$, are topologically conjugate. Recall that two matrices A, B are said to be *compatible* if they have the same size and $A(i, j) = 0 \Leftrightarrow B(i, j) = 0$. A stochastic matrix M compatible with one of the matrices S_n, $n \geq 1$, defines a shift-invariant (multiple) Markov measure m on the subshift of finite type S. Our compatibility assumption implies that the support of m is the whole of S; we shall always take Markov measures to be defined on their supporting subshifts of finite type. The pair (S, m) is a *Markov chain*. The Markov measure m is said to be *n-step* (or, to have *memory n*) if m may be defined by a stochastic matrix compatible with S_n, but not by one compatible with S_{n-1}. The *information cocycle* of m is

$$I_m = I\left(\alpha \Big| \bigvee_{i=1}^{\infty} S^{-i}\alpha\right) = I\left(\alpha \Big| \bigvee_{i=1}^{n} S^{-i}\alpha\right)$$

where α is the state partition of S and m has memory n.

As usual, $C(S)$ denotes the Banach space of continuous real-valued functions of a subshift S. A function of the form $g \circ S - g$, with $g \in C(S)$, is a *coboundary*. Two functions $f_1, f_2 \in C(S)$ are *cohomologous* if they differ by a coboundary; we shall write $f_1 \sim f_2$ in this case. The *pressure* of $f \in C(S)$ is defined to be

$$\mathscr{P}(f) = \sup\left\{ h(\mu) + \int f \, d\mu : \mu \text{ is } S\text{-invariant Borel probability}\right\},$$

where $h(\mu)$ denotes the entropy of μ. Evidently, $\mathscr{P}(f + c) = \mathscr{P}(f) + c$ for any constant $c \in \mathbf{R}$, and $\mathscr{P}(f_1) = \mathscr{P}(f_2)$ when $f_1 \sim f_2$. The map $\mathscr{P}: C(S) \to \mathbf{R}$ is positive, continuous and convex. An invariant Borel probability μ is said to be an *equilibrium state* of $f \in C(S)$ if $h(\mu) + \int f \, d\mu = \mathscr{P}(f)$. We refer the reader to [21] for a general treatment of pressure and equilibrium states. Since the shift is expansive, every $f \in C(S)$ has an equilibrium state, but not always a unique one (see [5, 6, 21]). When S is of finite type, the following theorem gives a dense set of functions with a unique equilibrium state.

Lanford-Ruelle Variational Principle [19]. Let S be a subshift of finite type and let $f: S \to \mathbf{R}$ depend on finitely many coordinates, so that there exist k, n with $f(S^k x) = f(x_0, x_1, \ldots, x_n)$ for all $x = (x_i) \in S$. Then there exists a unique Markov measure m on S such that $f \sim -I_m + \mathscr{P}(f)$. m is the unique equilibrium state of f, and the memory of m is at most n when, as above, f depends on $n + 1$ coordinates.

A further consequence of expansivenes is the fact that

$$h(\mu) = \inf\left\{ \mathscr{P}(f) - \int f \, d\mu : f \in C(S)\right\}$$

for any invariant probability μ. The *topological entropy* of S is $h(S) = \mathscr{P}(0)$. We assume the reader is familiar with the properties of topological entropy. When S is of finite type we denote by $\max_S$ the unique measure of maximal entropy, the 1-step Markov measure with $h(\max_S) = h(S)$ and $h(\max_S) > h(\mu)$ for $\mu \neq \max_S$.

We shall use the term (*continuous*) *code* to mean a continuous shift-commuting surjection ϕ between subshifts S, T; we shall write $\phi: S \downarrow T$ and say that T is a *factor* of S. Up to composition with a power of the shift, every continuous code $\phi: S \downarrow T$ may be expressed as a k-block code for some $k \in \mathbf{N}$ in the following way: there exists a map, which we also denote by ϕ, from S-blocks of length k onto the symbols of T such that

$$\phi(x)_n = \phi(x_n x_{n+1} \cdots x_{n+k-1}) \quad \text{for all } x = (x_i) \in S \text{ and } n \in \mathbf{Z}.$$

Two homeomorphisms S and T are *topologically conjugate* if there exists a homeomorphism ψ such that $\psi S = T \psi$. Two continuous codes $\phi_1: S_1 \downarrow T_1$ and $\phi_2: S_2 \downarrow T_2$ are *topologically equivalent* if there exist topological conjugacies $\psi_1: S_1 \downarrow S_2$ and $\psi_2: T_1 \downarrow T_2$ such that $\phi_2 \psi_1 = \psi_2 \phi_1$.

A continuous code $\phi: S \downarrow T$ between subshifts of finite type will be called a *bounded-to-one code* if there exists a number K such that $|\phi^{-1}(y)| \leqslant K$ for all $y \in T$. If ϕ is not bounded-to-one then there exists $y \in T$ such that $\phi^{-1}(y)$ contains uncountably many points (see [1]), and we call ϕ an *infinite-to-one code*. A code

ϕ: $S \downarrow T$ is bounded-to-one if and only if $h(S) = h(T)$ (see [**4, 16**]). Another characterization may be stated as follows. A 1-block code ϕ: $S \downarrow T$ is said to *collapse a diamond* if there exist distinct S-words $x_0 x_1 \cdots x_n$ and $x_0' x_1' \cdots x_n'$ such that $x_0 = x_0'$, $x_n = x_n'$ and $\phi(x_0 x_1 \cdots x_n) = \phi(x_0' x_1' \cdots x_n')$; and ϕ is infinite-to-one if and only if it collapses a diamond (see [**1, 10**]). This characterization may be used to formulate similar conditions for higher-block codes.

A *block code* between Markov chains (S, m), (T, p) is a continuous code ϕ: $S \downarrow T$ such that $m \circ \phi^{-1} = p$; we write ϕ:$(S, m) \downarrow (T, p)$ in this case. A block code ϕ:$(S, m) \downarrow (T, p)$ is bounded-to-one if and only if $h(m) = h(p)$ (see [**16**]).

Details of most of the above, and other related material, may be found in [**19**].

2. Examples.

Resolving codes. Let S, T be subshifts of finite type. A 1-block code ϕ: $S \downarrow T$ is called *right resolving* if, given $(j_1, j_2) \in \mathscr{A}(T)^2$ with $T(j_1, j_2) = 1$ and $i_1 \in \mathscr{A}(S)$ with $\phi(i_1) = j_1$, there exists $i_2 \in \mathscr{A}(S)$ with $\phi(i_2) = j_2$ and $S(i_1, i_2) = 1$. Note that we do not insist that i_2 be unique. In fact, a right resolving code ϕ: $S \downarrow T$ is bounded-to-one if and only if, for any i_1, j_1, j_2 as above, there exists a unique $i_2 \in \mathscr{A}(S)$ with the properties above. Clearly, if this uniqueness condition holds then no cylinder (and hence no point) of T can have more than $|\mathscr{A}(S)|$ preimages. Conversely, suppose that the right resolving code ϕ sends the distinct S-words $i_1 i_2$ and $i_1 i_2'$ to the same T-word. Define a subshift $\bar{S}$ of S by disallowing the word $i_1 i_2'$. Then ϕ takes $\bar{S}$ onto T, we have $h(S) > h(\bar{S}) \geqslant h(T)$, and ϕ: $S \downarrow T$ is infinite-to-one. Hence, our definition generalizes the bounded-to-one right resolving codes used in [**1, 16** and **17**]. The following statements should be compared with the propositions and lemmas numbered (12), (13) and (16)–(19) in reference [**17**], and with the remarks immediately after Proposition (13) of [**17**].

A rectangular 0-1 matrix is called an *amalgamation* matrix if its columns are nontrivial and each row contains precisely one nonzero entry. Two matrices A, B have the same *shape* if they have the same size and if $A(i, j) = 0 \Leftrightarrow B(i, j) = 0$. The following proposition should be clear.

(2.1) PROPOSITION. *If ϕ: $S \downarrow T$ is a (1-block) right resolving code and A_ϕ is the $\mathscr{A}(S) \times \mathscr{A}(T)$ amalgamation matrix defined by putting*

$$A_\phi(i, j) = 1 \Leftrightarrow \phi(i) = j,$$

then the matrices SA_ϕ and $A_\phi T$ have the same shape (and therefore we have $SA_\phi \geqslant A_\phi T$). Conversely, if A is an amalgamation matrix such that SA and AT have the same shape, then a 1-block right resolving code ϕ: $S \downarrow T$ (with $A_\phi = A$) may be defined by setting $\phi(i) = j$ if and only if $A(i, j) = 1$. The right resolving code ϕ is bounded-to-one if and only if $SA_\phi = A_\phi T$.

(2.2) PROPOSITION. *Suppose that ϕ: $S \downarrow T$ is a (1-block) right resolving code and let $A = A_\phi$ denote the amalgamation matrix associated to ϕ in (2.1). Let M, P be stochastic matrices compatible with S, T, respectively, and let m, p denote the 1-step Markov measures defined by M, P. If $MA = AP$, then $m \circ \phi^{-1} = p$.*

PROOF. Suppose $MA = AP$. Let $\bar{m}$, $\bar{p}$ be the invariant probability vectors of M, P, and notice that $\bar{m}A = \bar{p}$. Let $[j_0 \cdots j_n]$ be a cylinder of T. Then

$$m\left(\phi^{-1}[j_0 \cdots j_n]\right) = \sum_{i_0 \cdots i_n \in \phi^{-1}(j_0 \cdots j_n)} \bar{m}(i_0) M(i_0, i_1) \cdots M(i_{n-1}, i_n)$$

$$= \sum_{i_0 \cdots i_{n-1} \in \phi^{-1}(j_0 \cdots j_{n-1})} \left\{ \bar{m}(i_0) M(i_0, i_1) \cdots \right.$$

$$\left. M(i_{n-2}, i_{n-1}) \left(\sum_{i \in \mathscr{A}(S)} M(i_{n-1}, i) A(i, j_n) \right) \right\}$$

$$= \sum_{i_0 \cdots i_{n-1} \in \phi^{-1}(j_0 \cdots j_{n-1})} \bar{m}(i_0) M(i_0, i_1) \cdots M(i_{n-2}, i_{n-1}) P(j_{n-1}, j_n),$$

since

$$\sum_{i \in \mathscr{A}(S)} M(i_{n-1}, i) A(i, j_n) = (MA)(i_{n-1}, j_n) = (AP)(i_{n-1}, j_n) = P(\phi i_{n-1}, j_n).$$

Repeating the above argument $n - 1$ times, we have

$$m\left(\phi^{-1}[j_0 \cdots j_n]\right) = \left(\sum_{i_0 \in \phi^{-1}(j_0)} \bar{m}(i_0) \right) P(j_0, j_1) \cdots P(j_{n-1}, j_n)$$

$$= (\bar{m}A)(j_0) P(j_0, j_1) \cdots P(j_{n-1}, j_n)$$

$$= p([j_0 \cdots j_n]) \quad (\text{since } \bar{m}A = \bar{p}). \quad \square$$

When ϕ is infinite-to-one, (2.2) implies that uncountably many preimages of p may be obtained by splitting its defining matrix P: if M is a matrix such that for any $i_1 \in \mathscr{A}(S)$ and $j_2 \in \mathscr{A}(T)$ with $T(\phi i_1, j_2) = 1$ we have

$$\sum_{i_2 \in \phi^{-1}(j_2)} M(i_1, i_2) = P(\phi i_1, j_2),$$

then M gives a preimage of p. Unfortunately, the measures m obtained in this way need not constitute all the (1-step) Markov preimages of p; this fact will follow from our next example. The following corollary can be deduced from (2.2) by noting that a right resolving code $\phi: S \downarrow T$ induces a right resolving code from S_k onto T_k for each $k \in \mathbf{N}$.

(2.3) COROLLARY. *If $\phi: S \downarrow T$ is a (1-block) right resolving code then every k-step Markov measure on T has uncountably many Markov preimages of memory at most k.*

We have concentrated on right resolving codes, left resolving codes may be defined in a similar fashion. A left resolving code is right resolving with respect to the inverses of the shifts. For left resolving codes, (2.3) is valid, and analogues of (2.1), (2.2) may be formulated by using division and stochastic division matrices as in [17].

Bernoulli shifts. Let $B(p_1, \ldots, p_k)$ denote the Bernoulli shift defined by the probability vector $(p_1, \ldots, p_k)$. It has been shown [7, **20**] that if there exists a bounded-to-one block code from $B(p_1, \ldots, p_k)$ onto $B(q_1, \ldots, q_l)$ then $k = l$ and

$(p_1,\ldots,p_k)$ is a permutation of $(q_1,\ldots,q_l)$. It was conjectured in [7] that the natural generalization of this result to infinite-to-one block codes was valid: the existence of a block code from $B(p_1,\ldots,p_k)$ onto $B(q_1,\ldots,q_l)$ was conjectured to be equivalent to the existence of a partition of $\{1,\ldots,k\}$ into l sets $C_1,\ldots,C_l$ such that $\sum_{i\in C_j} p_i = q_j$ for $j = 1,\ldots,l$. In [7] the conjecture was shown to be valid in the special case where $(p_1,\ldots,p_k) = (1/k,\ldots,1/k)$, and Gerhard Keller (oral communication) has shown that it holds under algebraic independence assumptions on the ratios $p_2/p_1,\ldots,p_k/p_1$. We give here a counterexample to the general conjecture by constructing a block code from $B(\frac{1}{6},\frac{1}{6},\frac{2}{9},\frac{2}{9},\frac{2}{9})$ onto $B(\frac{1}{3},\frac{1}{3},\frac{1}{3})$. Consider the following 2-block code ϕ from the full 5-shift on the symbols $\{1,2,3,4,5\}$ to the full 3-shift on the symbols $\{a,b,c\}$:

11	21	33	45	54
12	22			
13	23			

$$\downarrow \phi$$
$$a$$

14	25	31	41	51
		35	44	53

$$\downarrow \phi$$
$$b$$

15	24	32	42	52
		34	43	55

$$\downarrow \phi$$
$$c$$

Here, we have listed the 2-blocks of the 5-shift according to their images under ϕ. Observe that, when regarded as a 1-block code from the 2-block system of the 5-shift, ϕ is right resolving. It follows that ϕ is surjective. To prove that ϕ is measure-preserving, we need a lemma. If w is a word of the 3-shift and it has length n, then $\phi^{-1}(w)$ consists of words v of length $n + 1$; for $i = 1, 2, 3, 4, 5$ let w_i denote the measure of the set $\{v \in \phi^{-1}(w): v$ ends with the symbol $i\}$.

(2.4) LEMMA. *For any word w of the 3-shift we have* $2(w_1 + w_2) = w_3 + w_4 + w_5$.

PROOF. Use induction on the length of w. If the length is 1, this is clear. Assume for words of length $n - 1$, $n \geqslant 2$. Let $w = j_1 \cdots j_{n-1} j_n$, $w' = j_1 \cdots j_{n-1}$. By assumption

$$(\ast) \qquad\qquad 2(w_1' + w_2') = w_3' + w_4' + w_5'.$$

There are three cases.

(i) $j_n = a$. Then $w_1 = w_2 = \frac{1}{6}(w_1' + w_2')$, $w_3 = \frac{2}{9}(w_1' + w_2' + w_3')$, $w_4 = \frac{2}{9}w_5'$, $w_5 = \frac{2}{9}w_4'$. Hence, using $(\ast)$,

$$w_3 + w_4 + w_5 = \frac{1}{3}(w_3' + w_4' + w_5') = \frac{2}{3}(w_1' + w_2') = 2(w_1 + w_2).$$

(ii) $j_n = b$. Then $w_1 = \frac{1}{6}(w_3' + w_4' + w_5')$, $w_2 = 0$, $w_3 = \frac{2}{9}w_5'$, $w_4 = \frac{2}{9}(w_1' + w_4')$, $w_5 = \frac{2}{9}(w_2' + w_3')$ and we have

$$w_3 + w_4 + w_5 = \frac{2}{9}(w_1' + w_2' + w_3' + w_4' + w_5')$$

$$= \frac{3}{9}(w_3' + w_4' + w_5') = 2(w_1 + w_2).$$

(iii) $j_n = c$ is similar to (ii). $\square$

Now we use (2.4) and induction to prove that ϕ is measure-preserving. This is clear for words of length 1, assume for words of length $n - 1$. Let $w = j_1 \cdots j_{n-1} j_n$, $w' = j_1 \cdots j_{n-1}$. By assumption

$$\left(\frac{1}{3}\right)^{n-1} = w_1' + w_2' + w_3' + w_4' + w_5'.$$

Again there are three cases.

(i) If $j_n = a$ then

$$w_1 + w_2 + w_3 + w_4 + w_5 = \frac{5}{9} w_1' + \frac{5}{9} w_2' + \frac{2}{9}(w_3' + w_4' + w_5')$$

$$= \frac{3}{9}(w_1' + w_2') + \frac{3}{9}(w_3' + w_4' + w_5') \quad \text{(by (2.4))}$$

$$= \frac{1}{3}(w_1' + w_2' + w_3' + w_4' + w_5') = \left(\frac{1}{3}\right)^n.$$

(ii) If $j_n = b$ then

$$w_1 + w_2 + w_3 + w_4 + w_5 = \frac{2}{9}(w_1' + w_2') + \frac{7}{18}(w_3' + w_4' + w_5')$$

$$= \frac{1}{3}(w_1' + w_2' + w_3' + w_4' + w_5') = \left(\frac{i}{3}\right)^n \quad \left(\text{using (2.4)}\right).$$

(iii) $j_n = c$ follows from (i), (ii) and the inductive hypothesis. The proof that ϕ is a block code from $B(\frac{1}{6}, \frac{1}{6}, \frac{2}{9}, \frac{2}{9}, \frac{2}{9})$ onto $B(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is now complete.

Write ϕ as a 1-block right resolving code from the 2-block system of the 5-shift. Consider a preimage of the Bernoulli measure $B(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ that is obtained from (2.2). The nonzero entries of any row of the defining matrix of such a measure must be obtained by splitting the vector $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and the Bernoulli measure $B(\frac{1}{6}, \frac{1}{6}, \frac{2}{9}, \frac{2}{9}, \frac{2}{9})$ does not have this property. This shows that (2.2) need not give all 1-step preimages of a 1-step Markov measure.

Peculiar memory example. Under a bounded-to-one code, the memory of the unique preimage of a Markov measure is regulated by the block-length of the code (see [17, 20]). For instance, under a 1-block bounded-to-one code, memory does not increase. We now give an infinite-to-one 1-block code $\phi: S \downarrow T$ which forces an increase in memory as it lifts 1-step Markov measures: each 1-step Markov measure on T has 2-step preimages on S, but no 1-step preimages. Furthermore, every 1-step Markov measure on S is sent to a 2-step Markov measure on T, so that we have a peculiar cross-over of memories. The subshift of finite type S is defined by the directed graph:
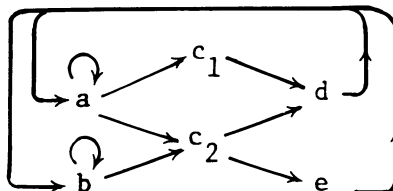
The alphabet of the subshift of finite type T is $\{a, b, c, d, e\}$; this subshift is the factor of S obtained by setting $\phi(c_1) = \phi(c_2) = c$ and letting the 1-block code ϕ map each of the other symbols to itself.

We first show that no 1-step Markov measure lifts to a 1-step Markov measure. Let m, p be Markov measures defined by stochastic matrices M, P compatible with S, T. Suppose, for a contradiction, that $m \circ \phi^{-1} = p$. Observe that $p[a] = m[a]$. Since $\phi^{-1}[ac] = [ac_1] \cup [ac_2]$ we have

$$P(a, c) = M(a, c_1) + M(a, c_2).$$

From the equalities $\phi^{-1}[bc] = [bc_2]$ and $\phi^{-1}[bce] = [bc_2 e]$ we deduce that $P(c, e) = M(c_2, e)$, and, using the equality $\phi^{-1}[ace] = [ac_2 e]$, we then find that $P(a, c) = M(a, c_2)$. Hence $P(a, c_1) = 0$, a contradiction.

Thus, no 1-step Markov measure lifts to a 1-step Markov measure. On the other hand, in the terminology of §5, ϕ is a right continuing code of retract one and (5.2) implies that every 1-step Markov measure on T lifts to uncountably many 2-step Markov measures on S. In addition, ϕ is an example of the magic diamond codes considered in §7, and, consequently, if m is a 1-step Markov measure on S then $m \circ \phi^{-1}$ is a Markov measure on T. (In fact, an application of the argument in the proof of (7.1) reveals that $m \circ \phi^{-1}$ has memory two.)

3. Pressure conditions for fibres of Markov measures. In this section we use pressure and information cocycles to describe the fibres of Markov measures, and to relate different fibres. We will use freely the notation and the properties of pressure given in §1. If $g: S \to \mathbf{R}$ is a function depending on only two coordinates of the subshift of finite type S then, by the Lanford-Ruelle variational principle,

$$g \sim -I_m + \mathscr{P}(g),$$

where the 1-step Markov measure m is the unique equilibrium state of g. To obtain $\mathscr{P}(g)$ and m from g, consider the matrix R compatible with S and defined by

$$R(i_0, i_1) = S(i_0, i_1) e^{g(i_0, i_1)},$$

let $\lambda > 0$ be its maximum eigenvalue provided by the Perron-Frobenius theorem, and let $\rho > 0$ be a corresponding right eigenvector, $R\rho = \lambda\rho$. Then $\mathscr{P}(g) = \log \lambda$ and m is the Markov measure defined by the stochastic matrix

$$M(i_0, i_1) = R(i_0, i_1)\rho(i_1)/\lambda\rho(i_0).$$

Fix two subshifts of finite type S, T with a 1-block code $\phi: S \downarrow T$. Suppose that m, p are 1-step Markov measures on S, T with $m \circ \phi^{-1} = p$. Let $\bar{p}$ be any 1-step Markov measure on T. Our first observation is

(3.1) LEMMA. $\mathscr{P}(I_p \circ \phi - I_m - I_{\bar{p}} \circ \phi) = 0$.

PROOF. Let $M, P, \bar{P}$ denote the defining matrices of $m, p, \bar{p}$. We have $I_m \sim -\log M(i_0, i_1)$ and similar statements hold for $p, \bar{p}$. Hence, the function

$$I_p \circ \phi - I_m - I_{\bar{p}} \circ \phi$$

depends on two coordinates of S, and its pressure is given by the logarithm of the maximum eigenvalue of the matrix R defined by

$$R(i_0, i_1) = M(i_0, i_1)\overline{P}(\phi i_0, \phi i_1)/P(\phi i_0, \phi i_1).$$

Choose constants k, K such that $0 < k \leqslant m[i]/p[j] \leqslant K$ for all $i \in \mathscr{A}(S), j \in \mathscr{A}(T)$, and observe that, since $m \circ \phi^{-1} = p$, we then have

$$(*) \qquad \frac{1}{K} \leqslant \sum_{i_0 \cdots i_n \in \phi^{-1}(j_0 \cdots j_n)} \frac{M(i_0, i_n) \cdots M(i_{n-1}, i_n)}{P(j_0, j_1) \cdots P(j_{n-1}, j_n)} \leqslant \frac{1}{k}$$

for all T-words $j_0 \cdots j_n$. It follows that

$$\mathscr{P}\left(I_p \circ \phi - I_m - I_{\overline{p}} \circ \phi\right) = \lim_{n \to \infty} \frac{1}{n} \log \sum_{i_0 \cdots i_n} R(i_0, i_1) \cdots R(i_{n-1}, i_n)$$

$$= \lim_{n \to \infty} \frac{1}{n} \log \sum_{j_0 \cdots j_n} \sum_{i_0 \cdots i_n \in \phi^{-1}(j_0 \cdots j_n)}$$

$$\times \frac{M(i_0, i_1) \cdots M(i_{n-1}, i_n)\overline{P}(j_0, j_1) \cdots \overline{P}(j_{n-1}, j_n)}{P(j_0, j_1) \cdots P(j_{n-1}, j_n)}$$

$$= \lim_{n \to \infty} \frac{1}{n} \log \sum_{j_0 \cdots j_n} \overline{P}(j_0, j_1) \cdots \overline{P}(j_{n-1}, j_n) \quad \text{(by } (*))$$

$$= \lim_{n \to \infty} \frac{1}{n} \log(\text{card } \mathscr{A}(T)) = 0. \quad \square$$

(3.2) LEMMA. *There exists a unique Markov measure $\overline{m}$ such that*

$$-I_{\overline{m}} \sim I_p \circ \phi - I_m - I_{\overline{p}} \circ \phi.$$

This measure $\overline{m}$ satisfies $\overline{m} \circ \phi^{-1} = \overline{p}$.

PROOF. The existence and uniqueness of $\overline{m}$ follows from (3.1) and the Lanford-Ruelle variational principle. Since the measures $\overline{m} \circ \phi^{-1}$, $\overline{p}$ are both ergodic, to establish $\overline{m} \circ \phi^{-1} = \overline{p}$ it suffices to show that $\overline{m} \circ \phi^{-1}$, $\overline{p}$ are equivalent. Let R be as in the proof of (3.1) and let $\rho > 0$ be a right invariant vector, $R\rho = \rho$. Then $\overline{m}$ is the 1-step Markov measure defined by the stochastic matrix $\overline{M}$ with

$$\overline{M}(i_0, i_1) = R(i_0, i_1)\rho(i_1)/\rho(i_0).$$

Letting c, C be constants such that

$$0 < c \leqslant \overline{m}[i_0]\rho(i')/\overline{p}[\phi i_0]\rho(i_0) \leqslant C$$

for all i_0, $i' \in \mathscr{A}(S)$, and using the definition of R and $(*)$ from the proof of (3.1), we find that

$$(c/K)\overline{p}[j_0 \cdots j_n] \leqslant \overline{m}(\phi^{-1}[j_0 \cdots j_n]) \leqslant (C/k)\overline{p}[j_0 \cdots j_n]$$

for all cylinders $[j_0 \cdots j_n]$ of T. $\quad \square$

(3.3) REMARK. We pause to outline an interpretation of (3.2) which seems to us central.

Let $\mathcal{M}_n(S)$ denote the Markov measures for S with memory less than or equal to n, and put $\mathcal{M}(S) = \bigcup_n M_n(S)$. Let $C_n(S)$ denote the vector subspace of $C(S)$ consisting of functions which depend only on coordinates 0 through n of S. Let $\tilde{C}_n(S)$ be the quotient space obtained by factoring out the constants and coboundaries in $C_n(S)$. The Lanford-Ruelle variational principle ensures that the map $\tilde{C}_n(S) \to \mathcal{M}_n(S)$ which sends an equivalence class to the equilibrium state of any representative is a bijection. (Its inverse is the map which sends $m \in \mathcal{M}_n(S)$ to the equivalence class of $-I_m$.) Thus we may think of a Markov measure as an equivalence class of continuous functions. In particular, we may impose on $\mathcal{M}_n(S)$ (and hence $\mathcal{M}(S)$) the vector space structure inherited from $C(S)$. The vector space operations then have a straightforward formulation in terms of matrix operations—see §5 of [**18**] and §4 of [**10**]. A continuous code $\phi: S \downarrow T$ induces a linear injection $\tilde{\phi}: \mathcal{M}(T) \to \mathcal{M}(S)$ via the natural injection $f \mapsto f \circ \phi$ of $C(T)$ into $C(S)$. When ϕ is bounded-to-one, each Markov measure p on T lifts to a unique Markov measure on S, and that measure is $\tilde{\phi}(p)$: Markov measures lift by the linear injection $\tilde{\phi}$. This viewpoint is due to W. Parry and the second-named author of this article (see [**18, 20**]); the flow presented in [**18**] corresponds to scalar multiplication in the space of Markov measures.

Now suppose $\phi: S \downarrow T$ is infinite-to-one. By passing to sufficiently high block systems of S, T, note that (3.1) and (3.2) are valid for Markov measures m, p, $\bar{p}$ of arbitrarily high memory and codes ϕ of any block length. Thus, if $m \in \mathcal{M}(S)$ and $p \in \mathcal{M}(T)$ are such that $m \circ \phi^{-1} = p$ then, for any $\bar{p} \in \mathcal{M}(T)$, we obtain a unique preimage $\bar{m} \in \mathcal{M}(S)$ by requiring

$$I_{\bar{p}} \circ \phi - I_{\bar{m}} \sim I_p \circ \phi - I_m.$$

By considering $\max_T$ and m with $m \circ \phi^{-1} = \max_T$, we see that, for each $p \in \mathcal{M}(T)$, the measure $m + \tilde{\phi}(p)$ is sent to p. In other words, each preimage of $\max_T$ determines an affine injection of $\mathcal{M}(T)$ into $\mathcal{M}(S)$. Call the image of such an injection a *leaf*. Then each leaf is an affine subspace of $\mathcal{M}(S)$, and a translate of any other leaf. Furthermore, the fibre of each $m \in \mathcal{M}(T)$ is a translate of any other fibre. In particular, if one Markov measure lifts to a Markov measure then so must all other Markov measures (and the map is Markovian).

(3.4) THEOREM. *Let (S, m) and (T, p) be Markov chains. If $\phi: (S, m) \downarrow (T, p)$ is a block code then*

$$\mathscr{P}\big(f \circ \phi + I_p \circ \phi - I_m\big) = \mathscr{P}(f) \quad \text{for all } f \in C(T).$$

PROOF. As we have already observed, (3.1) and (3.2) are valid for Markov measures m, p, $\bar{p}$, of arbitrarily high memory and codes ϕ of any block length. Let $f \in C(T)$ depend on finitely many coordinates. Write $f \sim -I_{\bar{p}} + \mathscr{P}(f)$ for some Markov measure $\bar{p}$ on T. Use (3.2) to find a Markov measure $\bar{m}$ on S with

$$-I_{\bar{m}} \sim I_p \circ \phi - I_m - I_{\bar{p}} \circ \phi.$$

Now

$$-I_{\bar{m}} \sim I_p \circ \phi - I_m + f \circ \phi - \mathscr{P}(f).$$

Applying pressure and using the fact that $\mathscr{P}(-I_{\overline{m}}) = 0$ we obtain

$$\mathscr{P}(f \circ \phi + I_p \circ \phi - I_m) = \mathscr{P}(f)$$

for f depending on finitely many coordinates, and the result follows for any $f \in C(T)$ by approximation and continuity of pressure. $\square$

By taking $f = 0$ in (3.4), notice that, in particular, $\mathscr{P}(I_p \circ \phi - I_m) = h(T)$.

(3.5) REMARKS. (i) An extension of the proof of (3.1) may be used to establish (3.4) directly, without recourse to (3.2). We believe the argument we have given is more transparent.

(ii) On a subshift of finite type S, an invariant probability measure is called a Gibbs measure (for the function $G \in C(S)$) if there exist constants $C_1, C_2 > 0$ such that

$$C_1 \leqslant \mu[x_0 \cdots x_n]/\exp(\Sigma_n G)(x) \leqslant C_2$$

for $\Sigma_n G = \Sigma_{i=0}^n G \circ S^i$, all $n \in \mathbf{N}$ and all $x \in S$. The unique equilibrium state for an exponentially decaying function is a Gibbs measure for that function (see [2] and its references). (3.4) is still valid when the Markov measures m, p are replaced by Gibbs measures μ, ν. Functions $G \in C(S)$ and $H \in C(T)$ for which μ, ν are Gibbs measures replace $-I_m$ and $-I_p$. The converse of (3.4) also holds for classes more general than Markov measures, and we present the more general result.

(3.6) THEOREM. *Let $\phi: S \downarrow T$ be a continuous code between subshifts of finite type. Suppose $g \in C(S)$ is such that $\mathscr{P}(f \circ \phi + g) = \mathscr{P}(f)$ for all $f \in C(T)$. If ν is the unique equilibrium state of a function $F \in C(T)$, then any equilibrium state μ of $g + F \circ \phi$ satisfies $\mu \circ \phi^{-1} = \nu$.*

PROOF. By assumption, (*) $\mathscr{P}(F) = \mathscr{P}(F \circ \phi + g) = h(\mu) + \int F \circ \phi \, d\mu + \int g \, d\mu$. We have

$$
\begin{aligned}
h(\mu \circ \phi^{-1}) &= \inf_{f \in C(T)} \left\{ \mathscr{P}(f) - \int f \, d(\mu \circ \phi^{-1}) \right\} \\
&= \inf_{f \in C(T)} \left\{ \mathscr{P}(f \circ \phi + g) - \int f \circ \phi \, d\mu \right\} \\
&\geqslant \inf_{f \in C(T)} \left\{ h(\mu) + \int f \circ \phi \, d\mu + \int g \, d\mu - \int f \circ \phi \, d\mu \right\} \\
&= h(\mu) + \int g \, d\mu = \mathscr{P}(F) - \int F \, d(\mu \circ \phi^{-1}) \quad \text{(by (*))}.
\end{aligned}
$$

Hence $h(\mu \circ \phi^{-1}) + \int F \, d(\mu \circ \phi^{-1}) \geqslant \mathscr{P}(F)$ and, since ν is the unique equilibrium state of F, we find that $\mu \circ \phi^{-1} = \nu$. $\square$

(3.7) REMARK. If ν is the unique equilibrium state of a function $F \in C(T)$, then ν must be ergodic. Conversely, Bob Phelps has shown us that any ergodic ν occurs as the unique equilibrium state of some function F. Thus, (3.6) applies to all ergodic measures ν.

(3.8) COROLLARY. *A code $\phi: S \downarrow T$ is Markovian if and only if there exists a locally constant function $g \in C(S)$ such that $\mathscr{P}(f \circ \phi + g) = \mathscr{P}(f)$ for all $f \in C(T)$.*

4. A cocycle-coboundary inequality. Let (S, m) and (T, p) be Markov chains. The preceding section contained conditions that must be satisfied by a block code $\phi: (S, m) \downarrow (T, p)$. The function $I_p \circ \phi - I_m$ played a central role in that section. We now obtain further necessary conditions by showing that $I_p \circ \phi - I_m$ is cohomologous to a nonpositive function.

Let $\phi: (S, m) \downarrow (T, p)$ be a block code, and let α, α' denote the state partitions of S, T. Put $\beta = \phi^{-1}\alpha'$. Replacing α, β by $\bigvee_{i=0}^{K} S^{-i}\alpha, \bigvee_{i=0}^{L} S^{-i}\beta$ if necessary, we may assume that $\beta \leqslant \alpha$ and that

$$I_m = I(\alpha|S^{-1}\alpha), \quad I_p \circ \phi = I(\beta|S^{-1}\beta).$$

We now have

$$I\left(\alpha\Big| \bigvee_{i=1}^{\infty} S^{-i}\beta\right) = I\left(\alpha \vee \beta\Big| \bigvee_{i=1}^{\infty} S^{-i}\beta\right) = I\left(\beta\Big| \bigvee_{i=1}^{\infty} S^{-i}\beta\right) + I\left(\alpha\Big| \bigvee_{i=0}^{\infty} S^{-i}\beta\right)$$

$$= I\left(\beta|S^{-1}\beta\right) + I\left(\alpha\Big| \bigvee_{i=0}^{\infty} S^{-i}\beta\right),$$

and also

$$I\left(\alpha\Big| \bigvee_{i=1}^{\infty} S^{-i}\beta\right) \leqslant I\left(\alpha \vee S^{-1}\alpha\Big| \bigvee_{i=1}^{\infty} S^{-i}\beta\right)$$

$$= I\left(S^{-1}\alpha\Big| \bigvee_{i=1}^{\infty} S^{-i}\beta\right) + I\left(\alpha\Big| \bigvee_{i=1}^{\infty} S^{-i}\beta \vee S^{-1}\alpha\right)$$

$$= I\left(\alpha\Big| \bigvee_{i=0}^{\infty} S^{-i}\beta\right) \circ S + I\left(\alpha|S^{-1}\alpha\right).$$

Therefore

$$I_m = I\left(\alpha|S^{-1}\alpha\right) \geqslant I\left(\beta|S^{-1}\beta\right) + I\left(\alpha\Big| \bigvee_{i=0}^{\infty} S^{-i}\beta\right) - I\left(\alpha\Big| \bigvee_{i=0}^{\infty} S^{-i}\beta\right) \circ S$$

$$= I_p \circ \phi + g \circ S - g,$$

where $g = -I(\alpha|\bigvee_{i=0}^{\infty} S^{-i}\beta)$. This inequality is improved in the following theorem, which will be proved by different means.

(4.1) THEOREM. *Let $\phi: (S, m) \downarrow (T, p)$ be a block code between Markov chains, and suppose that m, p have memories K, L and that ϕ is k-block. Then there exists a function $g \in C(S)$ depending on (at most) $\max\{K, L + k - 1\}$ coordinates and satisfying $I_m \geqslant I_p \circ \phi + g \circ S - g$.*

PROOF. By passing to higher block systems, we may assume that both m and p have memory 1 and that ϕ is 1-block; we must then find a function g which depends on only one coordinate of S. Put $J = I_m - I_p \circ \phi$. The following lemma was proved in [7] for the Bernoulli case. Its proof, which we omit, is a simple extension of the proof in [7].

(4.2) LEMMA. *If $x \in S$ satisfies $S^n x = x$, then $\sum_{i=0}^{n-1} J(S^i x) \geqslant 0$.*

Continuing with the proof of (4.1), note that J depends on two coordinates. Let $a \in \mathscr{A}(S)$ be a fixed symbol, and let $g(a)$ be arbitrary. For $b \in \mathscr{A}(S)$, $b \neq a$, put

$$g(b) = \inf\{ g(a) + J(a, c_1) + J(c_1, c_2) + \cdots + J(c_{n-1}, b):$$
$$ac_1 \cdots c_{n-1}b, n \geqslant 1, \text{ is } S\text{-word}\}.$$

Picking a path from b to a to complete a cycle from a to a, one sees from (4.2) that $g(b)$ exists as a finite number. We claim that $g(c) - g(b) \leqslant J(b, c)$ whenever $S(b, c) = 1$. If $b \neq a$, for any S-word $ac_1 \cdots c_{n-1}b$ we have

$$g(c) \leqslant g(a) + J(a, c_1) + J(c_1, c_2) + \cdots + J(c_{n-1}, b) + J(b, c)$$

and, taking the infimum over all paths $ac_1 \cdots c_{n-1}b$ from a to b, we obtain $g(c) \leqslant g(b) + J(b, c)$. If $b = a$, then c itself follows a and, by the definition of $g(c)$,

$$g(c) \leqslant g(a) + J(a, c). \quad \square$$

(4.3) REMARKS. (i) For a function $f \in C(S)$, let $[f]$ denote its cohomology class $[f] = \{ f' \in C(S): f' \sim f \}$. Write $[f_1] \geqslant [f_2]$ if $f_1 \geqslant f_2'$ for some $f_2' \in [f_2]$. This ordering is compatible with addition and with multiplication by nonnegative scalars, and $[f_1] \leqslant [f_2]$, $[f_1] \geqslant [f_2]$ together imply $[f_1] = [f_2]$. The conclusion of (4.1) then becomes $[I_m - I_p \circ \phi] \geqslant [0]$.

(ii) Parry has shown that a cocycle-coboundary equation is valid under various entropy-preserving codes (see [19]). In particular, Parry's equation is valid for bounded-to-one block codes. When the block code $\phi: (S, m) \downarrow (T, p)$ is bounded-to-one, (4.1) yields Parry's equation $I_m = I_p \circ \phi + g \circ S - g$ on noting that $\int I_m \, dm = h(m) = h(p) = \int I_p \circ \phi \, dm$. Conversely, if ϕ is infinite-to-one then $h(m) > h(p)$ and the inequality of (4.1) must be strict on a nonempty open set.

The β-function of a Markov chain (S, m) is the positive real analytic function $\beta_m(t)$ defined by $\log \beta_m(t) = \mathscr{P}(-tI_m)$. Various properties of the β-function may be found in [17 and 20]; we simply recall that it is invariant under bounded-to-one block codes and, for the infinite-to-one case, prove

(4.4) COROLLARY. *If* $\phi: (S, m) \downarrow (T, p)$ *is an infinite-to-one block code between Markov chains then*

$$\beta_m(t) - \beta_p(t) \begin{cases} > 0 & \text{for } t < 1, \\ = 0 & \text{when } t = 1, \\ < 0 & \text{for } t > 1. \end{cases}$$

PROOF. Clearly, $\beta_m(1) = \beta_p(1) = 1$. By (3.4),

$$\mathscr{P}(-tI_p) = \mathscr{P}(-tI_p \circ \phi + I_p \circ \phi - I_m) = \mathscr{P}((1 - t)I_p \circ \phi - I_m + tI_m - tI_m)$$
$$= \mathscr{P}((1 - t)(I_p \circ \phi - I_m) - tI_m).$$

Now, by (4.1), $I_p \circ \phi - I_m \leqslant g - g \circ S$ for some $g \in C(S)$. Moreover, since ϕ is infinite-to-one, this inequality is strict on a nonempty open set. Hence, if $t < 1$ then

$$\mathscr{P}(-tI_p) < \mathscr{P}((1 - t)g - (1 - t)g \circ S - tI_m) = \mathscr{P}(-tI_m).$$

Similarly, $\mathscr{P}(-tI_p) > \mathscr{P}(-tI_m)$ for $t > 1$, and the result follows. $\quad \square$

5. Continuing codes. Recall that two points x, $\bar{x}$ of a subshift of finite type are *left asymptotic* if there exists an integer N such that $x_n = \bar{x}_n$ for all $n \le N$.

Kitchens [8] characterized, in the following way, those codes topologically equivalent to bounded-to-one resolving maps. A code $\phi\colon S \downarrow T$ is called *right closing* if, given $x \in S$ and $y \in T$ such that $\phi(x)$ is left asymptotic to y, there exists exactly one $\bar{x} \in S$ such that $\bar{x}$ is left asymptotic to x and $\phi(\bar{x}) = y$. A closing code is necessarily bounded-to-one, and a code ϕ is right closing if and only if it is topologically equivalent to a bounded-to-one right resolving code. There is a natural extension of this characterization. Say that a code $\phi\colon S \downarrow T$ is *right continuing* if, given $x \in S$ and $y \in T$ such that $\phi(x)$ is left asymptotic to y, there exists at least one $\bar{x} \in S$ such that $\bar{x}$ is left asymptotic to x and $\phi(\bar{x}) = y$.

(5.1) PROPOSITION. *For a code* $\phi\colon S \downarrow T$ *between subshifts of finite type, the following are equivalent.*

(i) ϕ *is topologically equivalent to a right resolving code.*

(ii) ϕ *is right continuing.*

(iii) *There exists a nonnegative integer n such that, given* $x \in S$ *and* $y \in T$ *with* $(\phi x)_i = y_i$ *for all* $i \le 0$, *we can find* $\bar{x} \in S$ *such that* $\phi(\bar{x}) = y$ *and* $\bar{x}_i = x_i$ *for all* $i \le -n$.

PROOF. Since all three properties are invariant under topological equivalence, we may assume that ϕ is a 1-block code. It is clear that (i) implies (ii).

(ii) $\Rightarrow$ (iii). Suppose ϕ is right continuing and (iii) fails to hold. For each $N \in \mathbf{N}$ choose points $x^{(n)} \in S$ and $y^{(n)} \in T$ such that $(\phi x^{(n)})_i = (y^{(n)})_i$ for all $i \le 0$, and such that there exists no $\bar{x} \in S$ with $\phi(\bar{x}) = y^{(n)}$ and $\bar{x}_i = (x^{(n)})_i$ for all $i \le -n$. For each $n \in \mathbf{N}$, let $J_n \subset \mathscr{A}(S)$ be the set of symbols x_0 which can start a one-sided sequence $x_0 x_1 x_2 \cdots$ that maps to the one-sided sequence $y_0^{(n)} y_1^{(n)} y_2^{(n)} \cdots$. By restricting to a subsequence if necessary, we may assume that the sets J_n are all identical, each $J_n = J$ say, and also that the sequence $\{x^{(n)}\}$ converges to some point $x \in S$. Now for large enough integers k we have $(y^{(k)})_0 = (\phi x)_0$; fix k with this property and define a point $y \in T$ by setting

$$y_i = \begin{cases} (\phi x)_i & \text{for } i \le 0, \\ (y^{(k)})_i & \text{for } i \le 0. \end{cases}$$

Use the right continuing property of ϕ to find $\tilde{x} \in S$ and $l \in \mathbf{N}$ with $(\phi\tilde{x}) = y$ and $\tilde{x}_i = x_i$ for all $i \le -l$. Note that $\tilde{x}_0 \in J$. Using the fact that $x^{(n)} \to x$, pick some $n \ge l$ such that $(x^{(n)})_i = x_i$ for $-l \le i \le 0$. Define a point $\bar{x} \in S$ by setting

$$\bar{x}_i = \begin{cases} (x^{(n)})_i & \text{for } i \le -l, \\ \tilde{x}_i & \text{for } -l \le i \le 0, \end{cases}$$

and by taking $\bar{x}_0 \bar{x}_1 \bar{x}_2 \cdots$ to be any one-sided sequence which starts with $\bar{x}_0 = \tilde{x}_0 \in J$ and which ϕ maps to $y_0^{(n)} y_1^{(n)} y_2^{(n)} \cdots$. Now this point $\bar{x}$ has $\phi(\bar{x}) = y^{(n)}$ and $\bar{x}_i = (x^{(n)})_i$ for all $i \le -n$, and thus contradicts the choice of $x^{(n)}$ and $y^{(n)}$.

(iii) $\Rightarrow$ (i). Suppose (iii) holds, and let α denote the state partition of S. To deduce (i), define a partition β of S as in [8]: let β be the partition $\alpha \le \beta \le \bigvee_{i=0}^{n} S^{-i}\alpha$

obtained by clumping those cylinders $[x_0 \cdots x_n]$, $[\bar{x}_0 \cdots \bar{x}_n]$ such that $x_0 = \bar{x}_0$ and $\phi(x_0 \cdots x_n) = \phi(\bar{x}_0 \cdots \bar{x}_n)$. Consider the 0-1 transition matrix $\tilde{S}$ for β: the rows and columns of $\tilde{S}$ are indexed by the sets of β and, for sets $B_0, B_1 \in \beta$,

$$\tilde{S}(B_0, B_1) = 1 \Leftrightarrow B_0 \cap S^{-1}B_1 \neq \varnothing.$$

As Kitchens shows in [8], the subshift of finite type $\tilde{S}$ is topologically conjugate to S via the natural $(n + 1)$-block code $\psi: S \downarrow \tilde{S}$ with $(\psi x)_0 = B_0$ when $[x_0 \cdots x_n] \subset B_0$. Let $\tilde{T}$ be the $(n + 1)$-block system of T and let $\tilde{\phi}: \tilde{S} \downarrow \tilde{T}$ be the 1-block code induced by ϕ. Then ϕ and $\tilde{\phi}$ are topologically equivalent and the image of a symbol $B_0 \in \mathscr{A}(\tilde{S})$, where $B_0 \in \beta$, is determined by choosing a cylinder $[x_0 \cdots x_n] \subset B_0$ and putting $\tilde{\phi}(B_0) = \phi(x_0 \cdots x_n)$. We claim that $\tilde{\phi}$ is right resolving. Let $[x_0 \cdots x_n] \subset B_0 \in \beta$ and let $\tilde{\phi}(B_0) = \phi(x_0 \cdots x_n) = y_0 \cdots y_n \in \mathscr{A}(\tilde{T})$. Let $y_1 \cdots y_n y_{n+1} \in \mathscr{A}(\tilde{T})$ be a follower of $y_0 \cdots y_n$. By (iii), there exists an S-word $\bar{x}_0 \cdots \bar{x}_{n+1}$ such that $\bar{x}_0 = x_0$ and $\phi(\bar{x}_0 \cdots \bar{x}_{n+1}) = y_0 \cdots y_{n+1}$. Then we have $[\bar{x}_0 \cdots \bar{x}_n] \subset B_0$, and the set $B_1 \in \beta$ with $[\bar{x}_1 \cdots \bar{x}_{n+1}] \subset B_1$ is such that $\tilde{S}(B_0, B_1) = 1$. $\square$

(5.2) REMARK. Say that a right continuing code has *retract n* if the nonnegative integer n satisfies (iii) of Proposition (5.1). If ϕ is a k-block right continuing code with retract n, then every l-step Markov measure on T lifts to uncountably many Markov measures of memory at most $l + n + k - 1$. To see this, express ϕ as a composition $S \downarrow S_k \downarrow \tilde{S}_k \downarrow \tilde{T} \downarrow T$ where S_k is the k-block system of S and $\tilde{S}_k \downarrow \tilde{T}$ is the right resolving code furnished in the proof of (5.1). It also follows from (5.1) that a code is right closing if and only if it is right continuing and bounded-to-one.

We have concentrated on right continuing codes; left continuing codes may be defined in a similar way, and they have similar properties. Slight modifications of the proof of (5.1) show that a left continuing code may be recoded to a left resolving one, and that a left and right continuing code may be recoded to a left and right resolving one.

Let S, T be subshifts of finite type. Write $S \overset{\text{per}}{\to} T$ to mean the following: if S has a point of period n, then T has a point whose period divides n. The condition $S \overset{\text{per}}{\to} T$ is clearly necessary for the existence of a code from S onto T; it was proved in [3] that it is also sufficient when $h(S) > h(T)$. We now refine this result.

(5.3) THEOREM. *Suppose that S, T are subshifts of finite type such that $h(S) > h(T)$ and $S \overset{\text{per}}{\to} T$. Then we have the following.*

(i) *There exists a right continuing (and, hence, Markovian) code $\phi: S \downarrow T$.*

(ii) *If $\bar{S}$ is a proper subshift of S and $\bar{\phi}: \bar{S} \to T$ is a continuous shift-commuting map, then $\bar{\phi}$ extends to a right continuing code $\phi: S \downarrow T$.*

In (5.3), (i) follows immediately from (ii) by, for example, letting $\bar{S}$ consist of a periodic orbit and using the fact that $S \overset{\text{per}}{\to} T$ to construct a shift-commuting map from this orbit. We prove the stronger result (ii), which requires little extra work, to show that the property of being right continuing is global. That is, the right continuing property cannot be verified or ruled out by any property a code might exhibit on a proper subshift of the domain. The proof of (5.3) requires familiarity

with approximately the first half of [**3**]. More specifically, we shall assume familiarity with [**3**] through the proof of the main result.

PROOF OF (5.3). As remarked above, (i) follows from (ii). Once (ii) is known for aperiodic S, its extension to the periodic case is routine. So, we assume that S and T are aperiodic. We may find in S a subshift U such that $U \cap \bar{S} = \varnothing$, $h(U) > h(T)$ and U is (topologically conjugate to) an aperiodic subshift of finite type (see Lemma 26.17 of [**22**]). As in [**3**], we may then find a right closing code $\tilde{\phi} \colon \tilde{S} \downarrow T$, where $\tilde{S}$ is a subshift of U and $\tilde{S}$ is (topologically conjugate to) a subshift of finite type. Now $\tilde{S}$ and $\bar{S}$ are closed and disjoint, and, for large enough k, the set of k-blocks of $\tilde{S}$ is disjoint from the set of k-blocks of $\bar{S}$; otherwise we could use the compactness of S to construct sequences in $\tilde{S}$ and $\bar{S}$ with a common limit point. Thus, by passing to a high enough block system of S, we may assume that $\mathscr{A}(\tilde{S})$ and $\mathscr{A}(\bar{S})$ are disjoint subsets of $\mathscr{A}(S)$, that $\tilde{S}$ is a subshift of finite type obtained from S by disallowing certain symbols, and that $\bar{\phi}$ and $\tilde{\phi}$ are 1-block codes.

Let the right continuing code $\tilde{\phi} \colon \tilde{S} \downarrow \tilde{T}$ have retract n, given by (5.1)(iii). For $N \in \mathbf{N}$, consider the subshift of S obtained by disallowing $\tilde{S}$-words of length $N + n$; it is not hard to see that the topological entropy of this subshift increases to $h(S)$ as $N \to \infty$. Fix $N \in \mathbf{N}$ such that the matrices S^N and T^N are strictly positive and

$(*)$ the subshift of S obtained by disallowing $\tilde{S}$-words of length $N + n$ has topological entropy strictly greater than $h(T)$.

The reason for requiring $(*)$ will become clear later. Let $J \geqslant 2N$ be an integer, to be chosen later.

Let $x \in S$, and consider a (possibly infinite) segment of x. Call the segment a low stretch (of x) if it is composed of symbols from $\mathscr{A}(\tilde{S})$, its length is at least $N + n$, and it is not preceded or followed (in x) by a symbol from $\mathscr{A}(\tilde{S})$. Call the segment a high stretch (of x) if it contains no $\tilde{S}$-word of length $N + n$, and, when it has a beginning/ending, it begins/ends with a symbol outside $\mathscr{A}(\tilde{S})$ but is preceded/followed in x by a symbol from $\mathscr{A}(\tilde{S})$. This terminology reflects the fact that we are visualizing x as alternating between low stretches and high stretches:

Call a high stretch short or long according to whether its length is strictly less than J or not. Note that each point of $\bar{S}$ consists of a single high stretch, and that each point of $\tilde{S}$ consists of a single low stretch.

Roughly speaking, our construction of the right continuing extension ϕ will be as follows. We shall first use the machinery of [**3**] to define an extension $\psi \colon S \downarrow T$ of $\bar{\phi}$. The code ϕ will be obtained by alternating between ψ and $\tilde{\phi}$. We shall use ψ on long high stretches and $\tilde{\phi}$ on low stretches, except on the initial $N - 1$ symbols of a stretch and on a segment of length at most $J - N$ at the end of a long high stretch. The initial $N - 1$ symbols will be required for a transition (filler) code. The segment

at the end of a long high stretch will be required, and the integer J chosen, so that we can build a sufficiently rich transition code from long high stretches to low stretches to ensure that ϕ is right continuing. We shall use a filler code on short high stretches also.

Let $K = 3N$. Use Krieger's lemma (Lemma 2.2 in [3]) to define a closed-open marker set $F = F(N, K) \subset S$. Since F is closed-open, there exists $L \in \mathbf{N}$ such that whether a point x belongs to F depends only on its coordinates between $-L$ and L. Let $\tilde{F} = S^{L+2N}F$, so that

$$(**) \qquad \begin{array}{l} \text{whether a point } x \text{ belongs to } \tilde{F} \text{ is determined by its} \\ \text{coordinates between } -2L - 2N \text{ and } -2N. \end{array}$$

Using $\tilde{F}$ as a marker set, define an extension $\psi: S \downarrow T$ of $\bar{\phi}$ as in Lemma 2.4 of [3].

Denote the set of k-blocks of S by $\mathscr{B}_k(S)$, and put $\mathscr{B}(S) = \bigcup_k \mathscr{B}_k(S)$. Use similar notation for the subshifts of finite type $\tilde{S}$ and T. Let $B = b_0 \cdots b_{N+n-1} \in \mathscr{B}(S) \setminus \mathscr{B}(\tilde{S})$, and let j be the largest integer such that $0 \leqslant j \leqslant N + n - 1$ and $b_j \in \mathscr{A}(S) \setminus \mathscr{A}(\tilde{S})$. For an integer i and a symbol $b \in \mathscr{A}(\tilde{S})$, let $\mathscr{C}_i(B, b)$ denote the set of blocks $c_0 \cdots c_{i-1} \in \mathscr{B}_i(S)$ such that $c_0 \cdots c_j = b_0 \cdots b_j$, $c_{i-1} = b$, $c_{i-N} \cdots c_{i-1} \in \mathscr{B}(\tilde{S})$, $c_{i-N-1} \notin \mathscr{A}(\tilde{S})$, and such that $c_0 \cdots c_{i-1}$ contains no $\tilde{S}$-word of length $N + n$. (Observe that the sets $\mathscr{C}_i(B, b)$ contain those blocks that may be encountered around the end of a long high stretch.) For $\alpha, \beta \in \mathscr{A}(T)$, let $\mathscr{D}_i(\alpha, \beta)$ denote the set of blocks $\alpha_0 \cdots \alpha_{i-1} \in \mathscr{B}_i(T)$ with $\alpha_0 = \alpha$, $\alpha_{i-1} = \beta$. Use $(*)$ to find an integer I such that for any $i \geqslant I$ and any b, α, B as above we have

$$|\mathscr{C}_i(B, b)| \geqslant |\mathscr{D}_i(\alpha, \tilde{\phi}b)|.$$

Now define, for each $I < i < I + 2N$, a surjection $\Psi_i^{B,\alpha,b}$ of $\mathscr{C}_i(B, b)$ onto $\mathscr{D}_i(\alpha, \tilde{\phi}b)$. These surjections will be used for transition from long high stretches to low stretches. Take $J = I + 2N$. For each $N < i \leqslant J + N$, let Φ_i be a map from $\mathscr{A}(T)^2$ to $\mathscr{B}_i(T)$ such that $\Phi_i(\alpha, \beta) \in \mathscr{D}_i(\alpha, \beta)$ for all $\alpha, \beta \in \mathscr{A}(T)$. The maps Φ_i will be used as filler codes.

We now define the pieces of ϕ. Let $x \in S$.

(i) *Low stretches.* If x_0 is in a low stretch and $x_{-N+1} \cdots x_{-1}x_0 \in \mathscr{B}_N(\tilde{S})$ then put $(\phi x)_0 = \tilde{\phi}(x_0)$. This defines ϕ on a low stretch, except on the initial $N - 1$ symbols of a stretch.

(ii) *Long high stretches.* If $x_{-N+1} \cdots x_0 \cdots x_{I+N}$ is part of a high stretch, then put $(\phi x)_0 = (\psi x)_0$. This definition avoids the first $N - 1$ symbols and the last $I + N = J - N$ symbols of a long high stretch.

(iii) *Short high stretches.* Suppose that $x_1 \cdots x_j$, where $1 \leqslant j < J$, is a short high stretch. This means that x_0 ends a low stretch and x_{j+1} begins another low stretch, so that $(\phi x)_0$ and $(\phi x)_{j+N}$ have been defined by (i) (and by shift invariance). Define the coordinates of ϕx between 0 and $j + N$ by putting

$$(\phi x)_0(\phi x)_1 \cdots (\phi x)_{j+N} = \Phi_{j+N+1}\big(\tilde{\phi}(x_0), \tilde{\phi}(x_{j+N})\big).$$

(iv) *Transition from low stretches to long high stretches.* Suppose that $x_1 \cdots x_{N-1}$ are the initial symbols of a long high stretch. In this case, x_0 is the last symbol of a low stretch, $(\phi x)_0$ has been defined by (i), and $(\phi x)_N$ has been defined by (ii).

Define the coordinates of ϕx between 0 and N by putting

$$(\phi x)_0(\phi x)_1 \cdots (\phi x)_N = \Phi_{N+1}(\tilde{\phi}(x_0), (\psi x)_N).$$

(v) *Transition from long high stretches to low stretches.* Suppose that $x_{-I-N} \cdots x_{-1}$ is the end of a long high stretch, so that x_0 is the first symbol of a low stretch and $(\phi x)_{N-1}$ has already been defined. Choose an integer k between $I - N$ and $I + N$ as follows. If $S^i x \in \tilde{F}$ for some i such that $-I - N < i < I + N$, then define k by requiring $-k$ to equal the largest such i; otherwise, let $k = I$. For $-I - N \leqslant i \leqslant -k$, let $(\phi x)_i = (\psi x)_i$. Also let

$$(\phi x)_{-k}(\phi x)_{-k+1} \cdots (\phi x)_{N-1} = \Psi_{k+N}^{B,\alpha,b}(x_{-k} \cdots x_{N-1}),$$

where $B = x_{-k} \cdots x_{-k+N+n-1}, \alpha = (\psi x)_{-k}$ and $b = x_{N-1}$.

The construction of ϕ is now complete; we end the proof by verifying that ϕ is right continuing. Let $y \in T$, and let $x \in S$ be such that ϕx is left asymptotic to y. We must show that there exists $\bar{x} \in S$ such that $\bar{x}$ is left asymptotic to x and $\phi(\bar{x}) = y$. We may assume without loss of generality that $(\phi x)_i = y_i$ for all $i \leqslant 0$, and we need only deal with the following two cases.

Case 1. Suppose that the one-sided segment $\cdots x_{-1}x_0$ is part of a high stretch. Let $\tilde{x} \in \tilde{S}$ be such that $\phi(\tilde{x}) = y$. Choose an integer k between $I - N$ and $I + N$ as in (v): if $S^i x \in \tilde{F}$ for some i such that $-I - N < i < -I + N$, then let $-k$ equal the largest such i; otherwise let $k = I$. Put $B = x_{-k} \cdots x_{-k+N+n-1}, \alpha = (\psi x)_{-k}$ and $b = \tilde{x}_{N-1}$. Define a point $\bar{x} \in S$, left asymptotic to x, by letting

$$\bar{x}_i = \begin{cases} x_i & \text{if } i \leqslant -k, \\ \tilde{x}_i & \text{if } i \geqslant N - 1, \end{cases}$$

and by requiring $\bar{x}_{-k}\bar{x}_{-k+1} \cdots \bar{x}_{N-1}$ to be in $\mathscr{C}_{k+N}(B, b)$ and to satisfy

$$(\ast\ast\ast) \qquad \Psi_{k+N}^{B,\alpha,b}(\bar{x}_{-k} \cdots \bar{x}_{N-1}) = y_{-k} \cdots y_{N-1}.$$

The point $\bar{x}$ consists of two stretches, a high stretch up to $\bar{x}_0$ and a low stretch starting at $\bar{x}_0$. The choice of $\tilde{x}$ and $(\ast\ast\ast)$ guarantee that $(\phi\bar{x})_i = y_i$ for $i \geqslant -k$. By $(\ast\ast)$, for $i < -I + N$, we have $S^i x \in \tilde{F}$ if and only if $S^i \bar{x} \in \tilde{F}$. It follows from this, the definition of k, and the construction of ψ in [3] that $(\psi\bar{x})_i = (\psi x)_i$ for $i \leqslant -k$. That is, $(\phi\bar{x})_i = y_i$ for $i \leqslant -k$ as well.

Case 2. Suppose that, for some $j \leqslant -N - n$, the segment $x_{j+1} \cdots x_{j+N+n}$ is part of a low stretch. In other words, suppose that $x_{j+1} \cdots x_{j+N+n} \in \mathscr{B}_{N+n}(\tilde{S})$. Since n is the retract of $\tilde{\phi}$, we can find a point $\tilde{x} \in \tilde{S}$ such that $\tilde{x}_{j+N} = x_{j+N}$ and $\tilde{\phi}(\tilde{x}_i) = y_i$ for $i \geqslant j + N$. Now we obtain a point $\bar{x} \in S$, left asymptotic to x and with $\phi(\bar{x}) = y$, by taking

$$\bar{x}_i = \begin{cases} x_i & \text{if } i \leqslant j + N, \\ \tilde{x}_i & \text{if } i \geqslant j + N. \end{cases} \qquad \square$$

In [8], Kitchens constructed a bounded-to-one code which cannnot be expressed as a finite composition of closing codes, thereby refuting a conjecture of Adler and Marcus [1]. Unfortunately, despite (5.3), analogous complications arise in the infinite-to-one case. The following example gives an infinite-to-one Markovian code

which cannot be expressed as a finite composition of bounded-to-one codes and continuing codes.

(5.4) EXAMPLE. Consider the graph below. Define a subshift of finite type S by letting $\mathscr{A}(S)$ consist of the symbols appearing in the graph, allowing each symbol to follow itself, and also allowing the transitions given in the graph. (For purposes of visual clarity, the node labelled a, corresponding to a single symbol a, is repeated in the graph.)
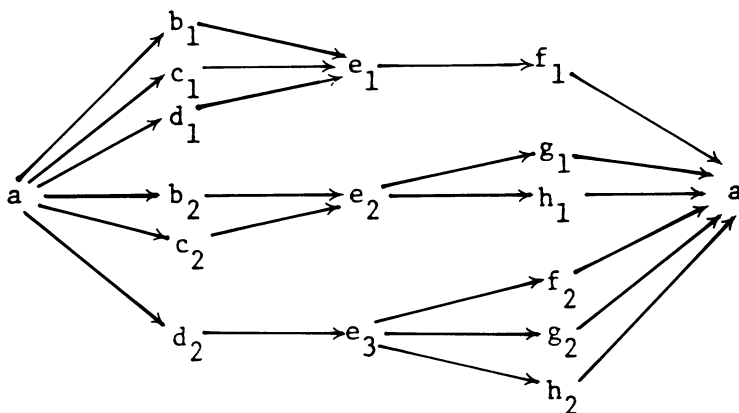
Define a 1-block code ϕ from S onto a subshift of finite type T with $\mathscr{A}(T) = \{a, b, c, d, e, f, g, h\}$ by mapping each symbol of S to the letter it uses. For instance, we have defined $\phi(e_1) = \phi(e_2) = \phi(e_3) = e$ and $\phi(a) = a$. Let $h(T) = \log \beta$. Define a function $f \in C(S)$ depending on only two coordinates of S by putting

$$f(x_0, x_1) = \begin{cases} \log(1/2\beta) & \text{if } x_0 x_1 = d_1 e_1 \text{ or } e_3 f_2, \\ \log(1/\beta) & \text{otherwise.} \end{cases}$$

It is easy to see that for every T-word $y_0 \cdots y_n$ with $y_0 = y_n = a$ we have

$$(*) \qquad \sum_{x_0 \cdots x_n \in \phi^{-1}(y_0 \cdots y_n)} \exp \sum_{i=0}^{n-1} f(x_1, x_{i+1}) = \frac{1}{\beta^n}.$$

It follows that $\mathscr{P}(f) = 0$. Hence, the equilibrium state m of f satisfies $-I_m \sim f$, and $(*)$ implies that $m \circ \phi^{-1} = \max_T$. Therefore, ϕ is Markovian.

Now suppose that $\phi = \phi_n \circ \phi_{n-1} \circ \cdots \circ \phi_1$ where each $\phi_i \colon S_{(i-1)} \downarrow S_{(i)}$ is a continuous code between subshifts of finite type, with $S_{(0)} = S$, $S_{(n)} = T$. We shall show that at least one of the codes $\phi_1, \ldots, \phi_n$ is not right continuing, left continuing, or bounded-to-one.

Consider the fixed points $\varepsilon_1 = (e_1)^\infty$, $\varepsilon_2 = (e_2)^\infty$ and $\varepsilon_3 = (e_3)^\infty$ in S. First assume that $\phi_1(\varepsilon_1) = \phi_1(\varepsilon_2)$. Choose $k \in \mathbf{N}$ so that both ϕ_1 and $\phi_n \circ \cdots \circ \phi_2$ can be expressed as k-block codes. That is, the zero coordinate of each $\phi_1(x)$ depends only on the k-block $x_0 \cdots x_{k-1}$, and a similar statement is valid for $\phi_n \circ \cdots \circ \phi_2$. Let $x' \in S$ be the point $\cdots d_1 d_1 e_1 e_1 \cdots$, where the arrow indicates the zero coordinate. Similarly, let $x'' \in S$ be the point $\cdots e_2 e_2 (e_2)^{2k-2} g_1 g_1 \cdots$, where $(e_2)^{2k-2}$

denotes the word $e_2 \cdots e_2$ of length $2k - 2$. Put $y' = \phi_1(x')$, $y'' = \phi_1(x'')$. Setting $\bar{e} = \phi_1((e_1)^k) = \phi_1((e_2)^k)$, observe that $y_i' = \bar{e}$ for $i \geqslant 1$ and $y_i'' = \bar{e}$ for $i \leqslant k - 1$. Define a point $y \in S_{(1)}$ by putting

$$y_i = \begin{cases} y_i' & \text{for } i \leqslant 0, \\ y_i'' & \text{for } i \geqslant 1. \end{cases}$$

Then $z = (\phi_n \circ \cdots \circ \phi_2)(y)$ is the point $\cdots d\overset{\downarrow}{d}(e)^{2k-2}gg \cdots$. Since $\phi^{-1}(z)$ consists of the point $x = \cdots d_2 \overset{\downarrow}{d_2}(e_3)^{2k-2}g_2 g_2 \cdots$, we must have $\phi_1(x) = y$. Hence, $\phi_1((e_3)^k) = y_1 = y_1'' = \bar{e}$ and we have $\phi_1(\varepsilon_3) = \phi_1(\varepsilon_1) = \phi_1(\varepsilon_2)$. In addition, similar arguments show that ϕ_1 must identify many other S-blocks. In particular, for $n \geqslant k$ we must have

$$\phi_1\big(a^n(d_1)^n\big) = \phi_1\big(a^n(d_2)^n\big), \quad \phi_1\big((d_1)^n(e_1)^n\big) = \phi_1\big((d_2)^n(e_3)^n\big),$$

$$\phi_1\big((e_1)^n(f_1)^n\big) = \phi_1\big((e_3)^n(f_2)^n\big) \quad \text{and} \quad \phi_1\big((f_1)^n a^n\big) = \phi_1\big((f_2)^n a_n\big).$$

Now let $1 \leqslant i \leqslant n$ be such that $\Psi = \phi_{i-1} \circ \cdots \circ \phi_1$ maps $\varepsilon_1, \varepsilon_2, \varepsilon_3$ to distinct points, while the points $\phi_i \Psi(\varepsilon_1)$, $\phi_i \Psi(\varepsilon_2)$, $\phi_i \Psi(\varepsilon_3)$ are not pairwise distinct. Arguments similar to those used above show that then we have $\phi_i \Psi(\varepsilon_1) = \phi_i \Psi(\varepsilon_2) = \phi_i \Psi(\varepsilon_3)$. (Above, we considered the case $i = 1$.) In addition, ϕ_i must identify many other blocks. In particular, for large $n \in \mathbf{N}$, consider the S-blocks

$$a^n(d_1)^n(e_1)^n(f_1)^n a^n \quad \text{and} \quad a^n(d_2)^n(e_3)^n(f_2)^n a^n.$$

Their images under Ψ are distinct words w, w', because $\Psi(\varepsilon_1) \neq \Psi(\varepsilon_3)$. However, just as in the case $i = 1$, we must have $\phi_i(w) = \phi_i(w)$. Thus, ϕ_i collapses a diamond, and it is infinite-to-one. Now let x denote the point $\cdots e_1 \overset{\downarrow}{e_1} f_1 f_1 \cdots$ in S, and put $y = \phi_i(\Psi x)$, $z = \phi(x)$. The distinct fixed points $\Psi(\varepsilon_1)$, $\Psi(\varepsilon_2)$, $\Psi(\varepsilon_3)$ are not left asymptotic, and their common image under ϕ_i is left asymptotic to y. However, since $\phi^{-1}(z)$ consists of x and the point $\cdots e_3 \overset{\downarrow}{e_3} f_2 f_2 \cdots$, the subshift of finite type $S_{(i-1)}$ contains at most two points that ϕ_i maps to y. Hence, ϕ_i is not right continuing. A similar argument shows that ϕ_i is not left continuing. (This may also be seen from the symmetry of the example.)

6. Uniform factors. Throughout this section, suppose that $\phi \colon S \downarrow T$ is a continuous code between subshifts of finite type S, T with positive topological entropies $h(S) = \log \alpha$, $h(T) = \log \beta$, and let $\gamma = \alpha/\beta$. If ϕ takes the measure $\max_S$ to the measure $\max_T$, then we say that ϕ is uniform and that T is a uniform factor of S. This terminology is justified by Proposition (6.1) below. Natural examples of uniform codes are bounded-to-one codes, projections of a direct product of subshifts of finite type onto a component, and compositions of these. We shall later present a uniform code that cannot be expressed as a finite composition of bounded-to-one codes and projections.

We consider positive functions on cylinders of T (including functions which depend only on the length $l(C)$ of a cylinder C), and define an equivalence relation $\sim$ on these functions as follows: put $f \sim g$ if there exist positive constants k, K such that

$$kf(C) \leqslant g(C) \leqslant Kf(C) \quad \text{for all cylinders } C.$$

For example, $\max_T(C) \sim \beta^{-l(C)}$. Notice that if $f \sim f'$ and $g \sim g'$ then $fg \sim f'g$ and $cf \sim f$ for any positive constant c. We use similar notation for positive functions on cylinders of S. We are also using the symbol $\sim$ to indicate cohomology, but this should cause no confusion.

(6.1) PROPOSITION. *Suppose that* $\phi: S \downarrow T$ *is a k-block code. For a cylinder* $C \subset T$, *let* $|\phi^{-1}C|$ *denote the number of cylinders* $[i_1 \cdots i_{l(C)+k-1}] \subset S$ *that make up* $\phi^{-1}C$. *The following are equivalent.*

(i) ϕ *is uniform.*

(ii) $|\phi^{-1}C| \sim \gamma^{l(C)}$ *as positive functions on the cylinders of* T.

(iii) *For every subshift* U *of* T *we have* $h(\phi^{-1}U) = h(U) + \log \gamma$.

PROOF. Suppose ϕ is uniform. For a cylinder $C \subset T$ we have
$$\max_S\left(\phi^{-1}C\right) \sim \alpha^{-(l(C)+k-1)}|\phi^{-1}C| \sim \alpha^{-l(C)}|\phi^{-1}C|$$
and
$$\beta^{-l(C)} \sim \max_T(C) = \max_S\left(\phi^{-1}C\right),$$
and we find that $\beta^{-l(C)} \sim \alpha^{-l(C)}|\phi^{-1}C|$. Hence, the function $|\phi^{-1}C| \sim \gamma^{l(C)}$, and (i) implies (ii). Conversely, suppose (ii) holds. We have
$$\max_S\left(\phi^{-1}C\right) \sim \alpha^{-l(C)}|\phi^{-1}C| \sim \alpha^{-l(C)}\gamma^{l(C)} = \beta^{-l(C)} \sim \max_T(C).$$

Thus, the measures $\max_S \circ \phi^{-1}$ and $\max_T$ are equivalent, and (i) follows by ergodicity. The fact that (ii) implies (iii) is clear, and we omit the argument that (iii) implies (i). $\square$

Various facts about Markovian codes have a kinder presentation when ϕ is uniform. In this case, (3.4) implies that
$$\mathscr{P}(f \circ \phi) = \mathscr{P}(f) + \log \gamma \quad \text{for all } f \in C(T).$$

If $f \in C(T)$ has a unique equilibrium state then, using (3.6), this equilibrium state may be lifted to any equilibrium state of $f \circ \phi$. The corresponding leaf over $\mathscr{M}(T)$ is given by a linear injection into $\mathscr{M}(S)$ (see (3.3)). On this leaf, $m \circ \phi^{-1} = p$ if and only if $I_m \sim I_p \circ \phi + \log \gamma$ (see (3.2)), and if $m \circ \phi^{-1} = p$ then $\beta_m(t) = \gamma^{1-t}\beta_p(t)$ (see the proof of (4.4)).

In addition, subshifts of finite type display striking rigidity under uniform codes. Our results in this direction rely on Kitchens' Jordan form theorem [9]. For matrices A, B let $A \subset B$ mean that a permutation of A may be obtained by deleting the rows and columns through some diagonal entries of B, and let $G(A)$ denote the matrix obtained from the Jordan form of A by deleting the rows and columns through zero entries of its diagonal. Then we have

(6.2) KITCHENS' THEOREM [9]. *If* $\phi: (S, m) \downarrow (T, p)$ *is a block code between Markov chains defined by the stochastic matrices* M, P, *then* $G(P) \subset G(M)$.

The following consequence of (6.1) is implicit in [9].

(6.3) JORDAN FORM CONDITION. *If* T *is a uniform factor of* S, *then* $G(\gamma T) \subset G(S)$.

In particular, a uniform factor T of S cannot have more nonzero eigenvalues than S; for example, the two-shift has no uniform factor of lower entropy. Less crudely, (6.3) gives strong and computable conditions on the eigenvalues of S and T. We prove related necessary conditions. We thank John Minardi for helpful discussions and the proof of the following lemma.

(6.4) LEMMA. *Suppose γ and β are algebraic numbers, $\chi(\lambda) \in \mathbf{Z}[\lambda]$ and $\chi(\gamma\beta_j) = 0$ for every conjugate β_j of β. Then $\chi(\gamma_i\beta_j) = 0$ for all conjugates γ_i of γ and all conjugates β_j of β.*

PROOF. Let K be a splitting field over $\mathbf{Q}$ containing γ and β. Given $\gamma_i\beta_j$, there is a field automorphism of K which fixes $\mathbf{Q}$ and takes γ_i to γ. This automorphism must take $\gamma_i\beta_j$ to $\gamma\beta_k$ for some conjugate β_k of β. So, $\gamma\beta_k$ is a conjugate of $\gamma_i\beta_j$. Since $\gamma\beta_k$ is a root of the polynomial χ, so is $\gamma_i\beta_j$. $\square$

For a matrix A, let χ_A denote its characteristic polynomial.

(6.5) LEMMA. *Suppose that the code $\phi\colon S \downarrow T$ is uniform and that δ is a root of χ_T. Then for any conjugate γ_i of γ and any conjugate δ_j of δ, the product $\gamma_i\delta_j$ is a root of χ_S.*

PROOF. If δ_j is a conjugate of δ then $\chi_T(\delta_j) = 0$ and, by (6.3), $\chi_S(\gamma\delta_j) = 0$. Apply (6.4). $\square$

Recall that a Perron number is an algebraic integer greater than or equal to one and greater than the modulus of any of its conjugates.

(6.6) THEOREM. *Suppose that the code $\phi\colon S \downarrow T$ is uniform and that S is aperiodic. Then γ is a Perron number.*

PROOF. The inverse image under ϕ of a periodic orbit is a subshift of finite type; by (6.1), this subshift has entropy $\log \gamma$. Therefore γ is an algebraic integer. Suppose γ is not Perron. Then there exists a conjugate γ_i of γ with $|\gamma_i| \geq \gamma$. By (6.5), the product $\gamma_i\beta$ is a root of χ_S. But $\gamma_i\beta \neq \gamma\beta = \alpha$ and $|\gamma_i\beta| \geq \gamma\beta = \alpha$, contradicting the aperiodicity of S. $\square$

For example, the six-shift cannot factor uniformly onto the four-shift, because $6/4$ is not an algebraic integer. Similarly, if $h(S) = \log(5 + \sqrt{5})$, $h(T) = \log(1 + \sqrt{5})$, and S is aperiodic, then S cannot factor uniformly onto T because $(5 + \sqrt{5})/(1 + \sqrt{5}) = \sqrt{5}$ and $\sqrt{5}$ is not Perron.

(6.7) REMARK. If $\phi\colon S \downarrow T$ is uniform and S has period k then α^k and β^k are Perron. Restricting ϕ to an irreducible component of S^k and applying (6.6), we find that γ^k is Perron.

(6.8) COROLLARY. *Suppose $\log \alpha$ is the entropy of some subshift of finite type. There exists a finite set B of numbers, depending only on α, such that if a subshift of finite type S with $h(S) = \log \alpha$ factors uniformly onto a subshift of finite type T, then $h(T) \in B$.*

PROOF. There are only finitely many ways of writing a Perron number as a product of Perron numbers (see [13]). Use (6.6) and (6.7). $\square$

We end the section with the example promised at the beginning, of a uniform code that is not a finite composition of bounded-to-one codes and projections. Note that

the composition of a bounded-to-one code ψ with a projection π,

$$U \times V \xrightarrow{\pi} V \xrightarrow{\Psi} W,$$

may be written as a composition of a projection with a bounded-to-one code as follows:

$$U \times V \xrightarrow{\mathrm{id} \times \psi} U \times W \to W.$$

Thus, a finite composition of bounded-to-one codes and projections may be expressed as a single bounded-to-one code followed by a projection, and below we need only check that our uniform map cannot be written in the latter form.

(6.9) EXAMPLE. Let

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad S = \begin{bmatrix} Q & R \\ Q & R \end{bmatrix}.$$

Then the 10×10 matrix S is irreducible and aperiodic; it defines a subshift of finite type S on the symbols $1, 2, \ldots, 10$. Define a 1-block code ϕ from S to the full two-shift T on the symbols a, b by mapping the symbols $1, 2, 3, 4, 5$ to a and the symbols $6, 7, 8, 9, 10$ to b.

The matrices Q, R have characteristic polynomials

$$\chi_Q(\lambda) = (\lambda^3 - \lambda - 1)(\lambda^2 - \lambda + 1), \quad \chi_R(\lambda) = (\lambda^3 - \lambda - 1)\lambda(\lambda - 1),$$

and hence the same spectral radius γ, the positive root of $\lambda^3 - \lambda - 1 = 0$. Furthermore, Q and R have a common right eigenvector for γ, the positive column vector

$$\rho = (\gamma + 1, \gamma^2, \gamma, 1, \gamma^2 + \gamma)^t.$$

It follows that $h(S) = \log(2\gamma)$, and that the ratio $e^{h(S)}/e^{h(T)}$ equals γ. To see that ϕ is surjective and uniform, let $C = [a_0 a_1 \cdots a_{n-1}]$, where each a_i equals a or b, be a cylinder of length n in T. It is easy to see that

$$|\phi^{-1}C| = \mathbf{1}^t \left(\prod_{i=1}^{n-1} S(a_i) \right) \mathbf{1},$$

where $S(a) = Q$, $S(b) = R$ and $\mathbf{1}$ stands for the 5×1 column vector of ones. Letting $c > 0$ be a constant such that $c\rho \leqslant \mathbf{1}$ and using the fact that $\mathbf{1} \leqslant \rho$, we have

$$|\phi^{-1}C| = \mathbf{1}^t \left(\prod_{i=1}^{n-1} S(a_i) \right) \mathbf{1} \leqslant \mathbf{1}^t \left(\prod_{i=1}^{n-1} S(a_i) \right) \rho$$

$$= \gamma^{n-1}(\mathbf{1}^t \cdot \rho) = (\mathbf{1}^t \cdot \rho/\gamma)\gamma^n$$

and

$$|\phi^{-1}C| \geqslant \mathbf{1}^t \left(\prod_{i=1}^{n-1} S(a_i) \right) (c\rho) = (\mathbf{1}^t \cdot \rho)(c/\gamma)\gamma^n.$$

Consequently, $|\phi^{-1}C| \sim \gamma^{l(C)}$, so that ϕ is surjective and uniform.

Now suppose that there exists a decomposition of ϕ into a bounded-to-one code followed by a projection. Then, since the subshifts of finite type R, Q are the inverse images under ϕ of the two fixed points of T, we obtain a subshift of finite type U as a common bounded-to-one factor of R and Q. Now (6.2) implies that (see [**9**]) the characteristic polynomial χ_U divides both χ_Q and χ_R. Thus, $\chi_U(\lambda) = \lambda^3 - \lambda - 1$, and U has zero trace. This is a contradiction since any factor of Q (or R) must have a fixed point. It follows that ϕ cannot be expressed as a finite composition of bounded-to-one codes and projections.

7. Magic diamonds. Let S be a subshift of finite type, and let $A = [a_0 \cdots a_k]$, $B = [b_0 \cdots b_k]$ be distinct cylinders of S such that $a_0 = b_0$, $a_k = b_k$, $A \cap S^i A = B \cap S^i B = \varnothing$ for $0 < i < k$, and $A \cap S^j B = \varnothing$ for $|j| < k$. We call such a pair (A, B) of cylinders a *magic diamond*. A *magic diamond code* ψ is then obtained by identifying A and B. Under this code, the image $\psi(x)$ of a point $x \in S$ is the sequence obtained from x by replacing any occurrence of the block $b_0 \cdots b_k$ with the block $a_0 \cdots a_k$. To determine $(\psi x)_0$ we need to look at the block $x_{-k+1} \cdots x_0 \cdots x_{k-1}$ to see whether x_0 is internal to an occurrence of $b_0 \cdots b_k$. The intersection conditions in the definition of a magic diamond ensure that ψ is well defined. Let $\overline{U}$ denote the image of S under ψ. The subshift $\overline{U}$ is topologically conjugate to a subshift of finite type, since $\overline{U}$ is the same as the subshift of S obtained by disallowing the block $b_0 \cdots b_k$. More precisely, the k-block system of $\overline{U}$ is a subshift of finite type U defined by a 0-1 matrix U, and we may view the magic diamond code as a code $S \downarrow U$, for which we will also use the symbol ψ.

(7.1) PROPOSITION. *Let* $\psi: S \downarrow U$ *be the magic diamond code defined above. If m is a 1-step Markov measure on S, then $m \circ \psi^{-1}$ is a Markov measure on U and the memory of $m \circ \psi^{-1}$ is at most k.*

PROOF. Let $y \in \overline{U}$ and $N \geqslant 1$. We shall prove the conditional probability

$$m \circ \psi^{-1}\big(y_0 \cdots y_N | y_{-2k+1-n} \cdots y_{-1}\big) = \frac{m \circ \psi^{-1}\big[y_{-2k+1-n} \cdots y_{-1} y_0 \cdots y_N\big]}{m \circ \psi^{-1}\big[y_{-2k+1-n} \cdots y_{-1}\big]}$$

is independent of $n \geqslant 0$, and the proposition will follow by recalling that U is the k-block system of $\overline{U}$. Observe that there can be at most one integer i with $-2k + 1 \leqslant i \leqslant -k - 1$ and $y_i \cdots y_{i+k} = a_0 \cdots a_k$. Define an integer I as follows. If $y_i \cdots y_{i+k} = a_0 \cdots a_k$ for some i such that $-2k + 1 \leqslant i \leqslant -k - 1$, then let $I = i$; otherwise let $I = -k$. By this definition

$(*)$ $\quad$ either y_I is the first or the last symbol of an occurrence of $a_0 \cdots a_k$ in y, or y_I is not part of an occurrence of $a_0 \cdots a_k$.

In any case, any $x \in S$ such that $(\psi x)_i = y_i$ for $-2k + 1 \leqslant i \leqslant -1$ must have $x_I = y_I$. Therefore, if $x, x' \in S$ are such that $(\psi x)_i = y_i$ for $-2k + 1 - n \leqslant i \leqslant -1$

and $(\psi x')_i = y_i$ for $-2k + 1 \leqslant i \leqslant N$ then we may define

$$\bar{x}_i = \begin{cases} x_i & \text{for } i \leqslant I, \\ x'_i & \text{for } i \geqslant I \end{cases}$$

to obtain a point $\bar{x} \in S$. It follows from $(*)$ that $\bar{x}$ has $(\psi\bar{x})_i = y_i$ for $-2k + 1 - n \leqslant i \leqslant N$. A straightforward computation using this fact reveals that we have

$$m \circ \psi^{-1}\left(y_0 \cdots y_N | y_{-2k+1-n} \cdots y_{-1}\right) = m\left(E_N | E_0\right)$$

for any $n \geqslant 0$, where

$$E_j = \left\{ x \in S : [x_I \cdots x_{N+k-1}] \cap \psi^{-1}[y_{-2k+1} \cdots y_j] \neq \varnothing \right\}, \quad j \geqslant 0. \quad \square$$

(7.2) REMARK ON TERMINOLOGY. A magic word for a 1-block map $\psi: S \downarrow U$ is a U-word $y_1 \cdots y_l$ for which there exists an integer I, $1 \leqslant I \leqslant l$, such that

$$\Psi(x_1 \cdots x_l) = \Psi(\bar{x}_1 \cdots \bar{x}_l) = y_1 \cdots y_l$$

implies that $x_I = \bar{x}_I$ (see [1, 14]). A magic diamond code is obtained by collapsing a judiciously chosen diamond and, as we have seen in the proof of (7.1), it has the property that every sufficiently long word in the image behaves like a magic word.

It is easy to see that a magic diamond code is both left and right continuing.

(7.3) PROPOSITION. *Let $\phi: S \downarrow T$ be an infinite-to-one 1-block code between subshifts of finite type, and let $\varepsilon > 0$. Then ϕ must collapse a magic diamond (A, B) such that the image U of S under the magic diamond code ψ obtained by identifying A and B satisfies $h(S) - h(U) < \varepsilon$.*

PROOF. Since ϕ is infinite-to-one, $h(S) > 0$. So, we may pick $a \in \mathscr{A}(S)$ and S-blocks $D = d_1 \cdots d_k$ and $E = e_1 \cdots e_{k+l}$ such that:
  (i) $k \geqslant 2$ and $l \geqslant 1$,
  (ii) $d_k = e_{k+l} = a$,
  (iii) the symbols $d_1, \ldots, d_{k-1}$ and $e_1, \ldots, e_{k+l-1}$ are different from a,
  (iv) $S(a, d_1) = S(a, e_1) = 1$.
Then, for any positive integers i and j, the blocks aD^i and aE^j are allowed and they can be overlapped only at their endpoints. Since the infinite-to-one code ϕ must collapse some diamond, we can find distinct S-blocks $C = c_1 \cdots c_n$ and $\bar{C} = \bar{c}_1 \cdots \bar{c}_n$ such that $\phi(C) = \phi(\bar{C})$, $S(a, c_1) = S(a, \bar{c}_1) = 1$ and $c_n = \bar{c}_n = a$. Now, for each $N \in \mathbf{N}$ the S-blocks $aD^{nN}ECDE^{nN}$ and $aD^{nN}E\bar{C}DE^{nN}$ give a magic diamond which is collapsed by ϕ. The required magic diamond code ψ is obtained by identifying the blocks $A = aD^{nN}ECDE^{nN}$ and $B = aD^{nN}E\bar{C}DE^{nN}$ for large enough N. $\quad \square$

Having shown that every infinite-to-one code may be intercepted by a magic diamond code, we recall the results of §3 to arrive at the following corollaries. These corollaries should be contrasted to the rigidity of subshifts of finite type under uniform codes.

(7.4) COROLLARY. *Let* (S, m) *be a Markov chain. Infinitely many numbers occur as the entropies of Markov chains* (T, p) *onto which* (S, m) *factors by block codes.* (*By* (6.2), *the defining matrices* P *of these factors* (T, p) *give rise to only finitely many Jordan forms away from zero,* $G(P)$.)

(7.5) COROLLARY. *Let* $\phi: S \downarrow T$ *be an infinite-to-one Markovian code. There exists a subshift of finite type* U *with* $h(S) > h(U) > h(T)$ *and Markovian codes* $\psi: S \downarrow U$ *and* $\theta: U \downarrow T$ *such that* $\phi = \theta \circ \psi$. *Every Markov measure on* T *lifts to uncountably many Markov measures on* S.

PROOF. We may assume that ϕ is 1-block and that there exists a 1-step Markov measure m such that $m \circ \phi^{-1} = \max_T$. (7.3) shows that we may write $\phi = \theta \circ \psi$ for a magic diamond code $\psi: S \downarrow U$ and an infinite-to-one code $\theta: U \downarrow T$. By (7.1), the measure $m \circ \psi^{-1}$ is Markov and, since $(m \circ \theta^{-1}) \circ \psi^{-1} = m \circ \phi^{-1} = \max_T$, the code θ is Markovian. Moreover, if $\overline{m}$ is any Markov measure on S with $\overline{m} \circ \psi^{-1} = m \circ \psi^{-1}$ then $\overline{m} \circ \phi^{-1} = \max_T$. But, as the magic diamond code ψ is right continuing, there exist uncountably many such Markov measures $\overline{m}$. $\square$

REFERENCES

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

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

3. M. Boyle, *Lower entropy factors of sofic systems*, Ergodic Theory Dynamical Systems (to appear).

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

5. F. Hofbauer, *Examples for the nonuniqueness of the equilibrium state*, Trans. Amer. Math. Soc. **228** (1977), 223–241.

6. R. B. Israel, *Convexity in the theory of lattice gasses*, Princeton Univ. Press, Princeton, N.J., 1979.

7. A. del Junco, M. Keane, B. Kitchens, B. Marcus and L. Swanson, *Continuous homomorphisms of Bernoulli schemes*, Ergodic Theory and Dynamical Systems. I, Progr. Math., Vol. 10, Birkhäuser, Boston, Basel and Stuttgart, 1981, pp. 91–111.

8. B. Kitchens, Ph.D. thesis, University of North Carolina, 1981.

9. _____, *An invariant for continuous factors of Markov shifts*, Proc. Amer. Math. Soc. **83** (1981), 825–828.

10. _____, *Linear algebra and subshifts of finite type*, Proceedings of the Conference on Modern Analysis and Probability (Yale University, 1982), in honor of S. Kakutani, Contemp. Math., vol. 26, Amer. Math. Soc., Providence, R.I., 1984.

11. W. Krieger, *On the subsystems of topological Markov chains*, Ergodic Theory Dynamical Systems **2** (1982), 195–202.

12. _____, *On certain notions of equivalence for topological Markov chains*, preprint 1982.

13. D. Lind, *Entropies and factorizations of topological Markov shifts*, Bull. Amer. Math. Soc. (N.S.) **9** (1983), 219–222.

14. B. Marcus, *Sofic systems and encoding data*, preprint 1982.

15. B. Marcus, K. Petersen and S. Williams, *Transmission rates and factors of Markov chains*, Proceedings of the Conference on Modern Analysis and Probability (Yale University, 1982), in honor of S. Kakutani, Contemp. Math., vol. 26, Amer. Math. Soc., Providence, R.I., 1984.

16. W. Parry, *A finitary classification of topological Markov chains and sofic systems*, Bull. London Math. Soc. **11** (1979), 170–176.

17. W. Parry and S. Tuncel, *On the classification of Markov chains by finite equivalence*, Ergodic Theory Dynamical Systems **1** (1981), 303–335.

18. _____, *On the stochastic and topological structure of Markov chains*, Bull. London Math. Soc. **14** (1982), 16–27.

19. _____, *Classification problems in ergodic theory*, London Math. Soc. Lecture Note Ser., Vol. 67, Cambridge Univ. Press, Cambridge, 1982.

20. S. Tuncel, *Conditional pressure and coding*, Israel J. Math. **39** (1981), 101–112.

21. P. Walters, *An introduction to ergodic theory*, Graduate Texts in Math., No. 79, Springer-Verlag, New York, 1982.

22. M. Denker, C. Grillenberger and K. Sigmund, *Ergodic theory on compact spaces*, Lecture Notes in Math., Vol. 527, Springer-Verlag, New York, 1976.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195.

Current address (M. Boyle): Mathematical Sciences Research Institute, 2223 Fulton Street, Room 603, Berkeley, California 94720

Current address (S. Tuncel): Mathematics Institute, University of Warwick, Coventry CV4 7AL, England