TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 306, Number 1, March 1988

THE AUTOMORPHISM GROUP OF A SHIFT OF FINITE TYPE

MIKE BOYLE, DOUGLAS LIND AND DANIEL RUDOLPH

ABSTRACT. Let (X_T, σ_T) be a shift of finite type, and $G = \mathrm{aut}(\sigma_T)$ denote the group of homeomorphisms of X_T commuting with σ_T. We investigate the algebraic properties of the countable group G and the dynamics of its action on X_T and associated spaces. Using "marker" constructions, we show G contains many groups, such as the free group on two generators. However, G is residually finite, so does not contain divisible groups or the infinite symmetric group. The doubly exponential growth rate of the number of automorphisms depending on n coordinates leads to a new and nontrivial topological invariant of σ_T whose exact value is not known. We prove that, modulo a few points of low period, G acts transitively on the set of points with least σ_T-period n. Using p-adic analysis, we generalize to most finite type shifts a result of Boyle and Krieger that the gyration function of a full shift has infinite order. The action of G on the dimension group of σ_T is investigated. We show there are no proper infinite compact G-invariant sets. We give a complete characterization of the G-orbit closure of a continuous probability measure, and deduce that the only continuous G-invariant measure is that of maximal entropy. Examples, questions, and problems complement our analysis, and we conclude with a brief survey of some remaining open problems.

TABLE OF CONTENTS

1. Introduction. Let T be a square nonnegative integral matrix. Following Williams [**Wi**], we associate to T a homeomorphism σ_T of a totally disconnected compact space X_T as follows. If T is $r \times r$, form the directed graph with r *states* or nodes, and with T_{ij} *symbols* or edges from state i to state j. Let $\mathcal{L}$ be the set of

Received by the editors February 23, 1987.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 20B27, 54H20, 58F11, 58F20; Secondary 20E26, 34C35, 68Q75.

Key words and phrases. Shift of finite type, symbolic dynamics, automorphism group, dimension group, invariant measure, residual finiteness, gyration function, periodic points.

The authors were supported, respectively, by NSF Grants DMS-8601619, DMS-8320356, and DMS-8504701.

symbols of this graph. Then $X_T \subset \mathcal{L}^{\mathbb{Z}}$ consists of those $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ with the terminal state of x_{i-1} matching the initial state of x_i for all $i \in \mathbb{Z}$. Points in X_T may be thought of as infinite trips on the graph. Clearly X_T is compact in the topology induced from the product topology on $\mathcal{L}^{\mathbb{Z}}$, and the shift $\sigma_T: X_T \to X_T$ defined by $(\sigma_T x)_i = x_{i+1}$ is a homeomorphism. This dynamical system (X_T, σ_T) is called a *shift of finite type* or *topological Markov shift*. Such systems are intrinsically characterized as expansive homeomorphisms of totally disconnected compact spaces with canonical coordinates [**Bo**]. They play a prominent role not only in topological dynamics [**DGS**] and coding theory [**ACH**], but are also crucial to the analysis of hyperbolic diffeomorphisms [**Sm**]. We shall assume throughout that σ_T is mixing, or, equivalently, that some power of T is strictly positive. To avoid trivial exceptions, we also require $T \neq [1]$.

Let $G = \mathrm{aut}(\sigma_T)$ denote the group of homeomorphisms of X_T commuting with σ_T. If $\varphi \in G$, then the fundamental observation of Curtis, Lyndon, and Hedlund [**H**, Theorem 3.4] shows there is an n and a finite block map $f: \mathcal{L}^{2n+1} \to \mathcal{L}$ so that $(\varphi x)_i = f(x_{i-n}, \ldots, x_{i+n})$. It follows that G is countable, and is discrete in the compact-open mapping topology. Despite the finite character of such mappings, very little is known about the algebraic structure of G. Hedlund [**H**] showed that for the full k-shift, $\mathrm{aut}(\sigma_{[k]})$ contains two involutions whose product has infinite order, and also a copy of every finite group. Ryan [**Ry2**] showed that the center of $\mathrm{aut}(\sigma_T)$ contains only the group Σ of powers of σ_T. However, it is still an open problem whether the automorphism group of the 2-shift is generated by the shift and involutions in the group. Another example of our ignorance is the inability to settle the question whether the automorphism groups of the 2-shift and 3-shift are isomorphic.

Recently two new approaches to the structure of $G = \mathrm{aut}(\sigma_T)$ have been made. Boyle and Krieger [**BK**] used the action of G on the invariant set of periodic points for σ_T to construct a nontrivial homomorphism from G to $\prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$ called the gyration function, and used this function to study G. Wagoner [**Wa2**], in analogy with K-theory, constructed a nontrivial representation of G by using its action on the space of Markov partitions of X_T. Furthermore, he has shown [**Wa1**] that G can be modelled by homeomorphisms commuting with a special diffeomorphism of the sphere S^q for $q \geq 5$.

Our purpose is to study the algebraic properties of G and the dynamics of its action on X_T and some associated spaces. The general method used here to construct elements of G goes back at least to Hedlund, and is usually called the "marker method." Roughly speaking, this method divides the symbols of a doubly infinite sequence into program and data, and the automorphism makes the program act on the data. The requirement of keeping program and data separated leads to certain complications. Special cases of this idea are used in §2 to show that G contains the free group on countably many generators, as well as the direct sum of countably many copies of $\mathbb{Z}$ and of any countable collection of finite groups. However, G does not contain a group with unsolvable word problem. Also, every subgroup of G is residually finite, implying G cannot contain a nontrivial divisible group, nor the infinite symmetric group. The finite type character of the shifts is related to the failure of divisibility in their automorphism groups, for in Example 3.9 we construct

a subshift whose automorphism group contains $\mathbb{Q}$ with 1 corresponding to the subshift. Automorphisms constructed using markers have finite order, although their composition may not. Conversely, we show that any finite-order element in G is obtained from a marker construction by using appropriate coordinates (Proposition 2.6).

Divisibility of elements of G is discussed in §3. The main question, which remains open, is whether an infinite-order element can have nth roots for infinitely many n. An argument using Ryan's theorem on the center of G shows that $\mathrm{aut}(\sigma_{[4]})$ is not algebraically isomorphic to $\mathrm{aut}(\sigma_{[2]})$. However, we are not able to decide whether $\mathrm{aut}(\sigma_{[3]})$ and $\mathrm{aut}(\sigma_{[2]})$ are isomorphic. In §4 we present an example of two shifts of finite type with equal zeta-functions that have nonisomorphic automorphism groups.

The growth of the set $G_n(\sigma_T)$ of automorphisms depending on the central n coordinates is doubly exponential in n. In §5 we define the *symmetry* of σ_T to be

$$s(\sigma_T) = \limsup_{n\to\infty} \frac{1}{n} \log\log |G_n(\sigma_T)|,$$

show that symmetry is a topological invariant, and prove that $\frac{1}{2}h(\sigma_T) \le s(\sigma_T) \le h(\sigma_T)$, where $h(\sigma_T)$ is the topological entropy of σ_T. The precise value of $s(\sigma_T)$ is not known to us for any $T \neq [1]$.

Krieger has associated to σ_T an automorphism $\widehat{T}$ of a countable ordered abelian group $(\mathcal{G}_T, \mathcal{G}_T^+)$ called the dimension group. In §6 we outline this construction, and show that each $\varphi \in G$ induces an automorphism of $(\mathcal{G}_T, \mathcal{G}_T^+, \widehat{T})$. The main question is whether this dimension group representation $\delta\colon \mathrm{aut}(\sigma_T) \to \mathrm{aut}(\widehat{T})$ is surjective. In Theorem 6.8 we prove that if the eigenvalues of $\widehat{T}$ are simple and no ratio of them is a root of unity, then for all sufficiently large n the map $\delta\colon \mathrm{aut}(\sigma_T^n) \to \mathrm{aut}(\widehat{T}^n)$ is surjective. The argument uses the fact that if the eigenvalues of T are simple, then $\mathrm{aut}(\widehat{T}^n)$ is finitely generated. The proof of this has the Dirichlet unit theorem as its priipal ingredient. Two examples complement the discussion, one of which shows that $\mathrm{aut}(\widehat{T})$ is not always finitely generated.

The period of a point is not altered under an automorphism, so G acts on the set Q_n of points with least σ_T-period n. Is this action transitive? In Theorem 7.2 we prove a strong form of transitivity on Q_n for all sufficiently large n. However, we give an example with two fixed points which we show cannot be interchanged by any composition of finite-order automorphisms. Our analysis of the action of G on periodic points implies certain algebraic properties of G, including that none of G, $G/[G,G]$, and G/Σ are finitely generated. We conclude §7 with an example showing that the profinite topology on G does not always coincide with that induced from the action of G on periodic points.

Boyle and Krieger [**BK**] introduced the *gyration function* of $\varphi \in G$ to be the number $g(\varphi, \sigma_T)(n) \in \mathbb{Z}/n\mathbb{Z}$ indicating the total amount of twist given by φ to the orbits of length n. The map $g\colon G \to \prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$ defined by

$$g(\varphi) = \big(g(\varphi,\sigma_T)(1), g(\varphi,\sigma_T)(2), \dots\big)$$

is a homomorphism called the *gyration representation.* They prove that $g(\sigma_{[k]}, \sigma_{[k]})$ has infinite order for $k \ge 2$. Using computer experimentation, we stumbled on the fact that $g(\sigma_T, \sigma_T)(p^n)$ converges p-adically, usually to a nonzero limit that turns

out to be transcendental. This is the basis of our proof in §8 that if the product of the nonzero eigenvalues of T is not ± 1, then $g(\sigma_T, \sigma_T)$ has infinite order. It is also a key ingredient in our proof that σ_T is not a limit, in the periodic point topology on G, of products of finite-order elements (Proposition 8.3).

The search for nontrivial representations of G leads naturally to studying G-invariant sets and measures on X_T. In Theorem 9.2 we prove that if a point is not σ_T-periodic, then its G-orbit is dense. A modification of this argument shows that if $Y \subset Z$ are σ_T-invariant compact subsets of σ_T, then Z is in the G-orbit closure of Y under the Hausdorff metric on compact subsets. In §10 we obtain a complete characterization of the G-orbit closure of a probability measure on X_T. Roughly speaking, a measure ν is in the G-orbit closure of μ when it has enough cumulative entropy to accommodate an approximate image of μ. The precise formulation is given in Theorem 10.1. One simple consequence is that the measure of maximal entropy on X_T is the only continuous G-invariant measure.

This work has benefitted substantially from numerous conversations with many people. We would like in particular to thank Ethan Coven, John Franks, Ralph Greenberg, Hang Kim, Bruce Kitchens, Neal Koblitz, Wolfgang Krieger, Gopal Prasad, Frank Rhodes, Jonathan Rosenberg, Fred Roush, John Smillie, and Jack Wagoner. We would also like to thank the Mathematical Science Research Institute, the IBM Thomas J. Watson Research Center, and the National Science Foundation for their support.

2. Markers and subgroups. We shall describe a method for building automorphisms of $G = \mathrm{aut}(\sigma_T)$, and use this method to construct subgroups of $\mathrm{aut}(\sigma_T)$ isomorphic to such groups as the direct sum of any countable collection of finite groups, the free group on infinitely many generators, and the direct sum of countably many copies of $\mathbb{Z}$. However, we show G does not contain a group with unsolvable word problem, and in the next section that it does not contain a nontrivial divisible group. We next discuss some elaborations of this marker method used throughout this paper, and show that automorphisms of finite order coincide with those obtained from a marker construction by using an appropriate symbolic presentation of X_T.

Recall from §1 that the set of symbols $\mathcal{L}$ for X_T is the collection of edges for the graph of T, and that points of X_T are just allowed bi-infinite sequences of symbols. For $x \in X_T$ let $x[m,n] = x_m \cdots x_n \in \mathcal{L}^{n-m+1}$ be the block of coordinates of x from m to n. Let $\mathcal{B}_n(X_T)$ be the set of allowed blocks of symbols of length n for T. Put $\mathcal{B}(X_T) = \bigcup_{n=1}^{\infty} \mathcal{B}_n(X_T)$. Suppose $M \in \mathcal{B}_m(X_T)$ and that $\mathcal{D} \subset \mathcal{B}_k(X_T)$ is a collection of blocks with $M\mathcal{D}M = \{MDM\colon D \in \mathcal{D}\} \subset \mathcal{B}_{2m+k}(X_T)$ such that for every $D \in \mathcal{D}$ the block M can overlap the concatenation MDM in only the initial and final segments of length m (this disallows even partial overlaps at the ends). Let π be an arbitrary permutation of $\mathcal{D}$. Define the action of a block map φ_π on $x \in X_T$ as follows. For each i, if $x[i, i+2m+k-1] = MDM$, define $(\varphi_\pi x)[i, i+2m+k-1] = M\pi(D)M$. Require φ_π to have no other action. Because the blocks from $M\mathcal{D}M$ cannot overlap except for the marker M, this is a well-defined σ-invariant map of finite order, so $\varphi_\pi \in \mathrm{aut}(\sigma_T)$. The correspondence $\pi \leftrightarrow \varphi_\pi$ hence embeds the symmetric group $\mathrm{sym}\,\mathcal{D}$ of $\mathcal{D}$ into $\mathrm{aut}(\sigma_T)$. We shall show shortly that $|\mathcal{D}|$ can be made arbitrarily large by an appropriate choice of M. This implies by Cayley's theorem that every finite group embeds into $\mathrm{aut}(\sigma_T)$, and

generalizes Hedlund's argument [**H**, Theorem 6.13] from the automorphism group of a full shift to that of a shift of finite type.

Blocks, or collections of blocks, with the kind of nonoverlapping property used above play a role for constructing continuous maps similar to that of Rohlin bases used to define measurable isomorphisms in ergodic theory [**Sh**, Chapter 10]. The idea has surfaced in several guises, such as prefix synchronization codes in information theory [**G**].

DEFINITION 2.1. Two blocks *overlap* if an initial segment of one coincides with a terminal segment of the other. A collection of blocks in $\mathcal{B}(X_T)$ has *only trivial overlaps* if distinct blocks do not overlap and each block overlaps itself only in the entire block.

The following gives an ample supply of blocks with only trivial overlaps.

LEMMA 2.2. *There is a collection* $\mathcal{M} = \bigcup_{n=1}^{\infty} \mathcal{M}_n \subset \mathcal{B}(X_T)$ *such that* $\mathcal{M}_n$ *contains* n *blocks of equal length,* $\mathcal{M}$ *has only trivial overlaps, and*

$$\mathcal{M}\mathcal{M} = \{MM' : M, M' \in \mathcal{M}\} \subset \mathcal{B}(X_T).$$

PROOF. Since σ_T is mixing, and $T \neq [1]$ by our convention, there must be a loop $i_0 i_1 \cdots i_k i_0 \in \mathcal{B}(X_T)$ of distinct symbols with $k \geq 1$. Furthermore, one of these symbols, which we can assume is i_0, is followed by a symbol $j_1 \neq i_1$.

First suppose $j_1 \neq i_0$. Choose a path of minimal length from j_1 to the loop, say $j_1 j_2 \cdots j_r i_s$. The case $r = 0$ is possible and corresponds to $j_1 = i_s$ for some $s \neq 0, 1$. Define $A = i_0 \cdots i_k$, $B = i_0 j_1 j_2 \cdots j_r i_s \cdots i_k \in \mathcal{B}(X_T)$. For $1 \leq q \leq n$ define $M_{nq} = A^2 B^q A B^{n-q+1}$. Noting the positions of i_0 in these blocks, it follows from the above minimality of paths that $\mathcal{M} = \{M_{nq} : 1 \leq q \leq n, n \geq 1\}$ has only trivial overlaps, and that $\mathcal{M}\mathcal{M} \subset \mathcal{B}(X_T)$ by construction. Since the lengths $|M_{nq}|$ are equal for $1 \leq q \leq n$, the collections $\mathcal{M}_n = \{M_{nq} : 1 \leq q \leq n\}$ satisfy the conclusions.

The remaining possibility is for $j_1 = i_0$. In this case let $A = i_1 \cdots i_q i_0$, and put $M_{nq} = A^2 i_0^q A i_0^{n-q+1}$. Again noting the positions of i_0 in the M_{nq} shows that $\mathcal{M}_n = \{M_{nq} : 1 \leq q \leq n\}$ for $n \geq 1$ satisfy the conclusions. □

We shall say that $G = \mathrm{aut}(\sigma_T)$ *contains* a group H if there is an isomorphism of H to a subgroup of G. Using the markers constructed in Lemma 2.2, we will show that G contains several kinds of infinite groups. For clarity, the constructions are first carried out on convenient full shifts, then extended to general σ_T by a substitution map.

THEOREM 2.3. *The group* $\mathrm{aut}(\sigma_T)$ *contains the direct sum of every countable collection of finite groups.*

PROOF. We first obtain the embedding when σ_T is the full 3-shift on $\{0, 1, 2\}$, then extend to general σ_T by a substitution map using markers from Lemma 2.2.

First suppose $X_T = \{0, 1, 2\}^{\mathbf{Z}}$, and let $\mathcal{D}_n = \{0, 1\}^n$, $M = 2$. For $\pi \in \mathrm{sym}(\mathcal{D}_n)$ define $\varphi_\pi \in G$ using blocks $M\mathcal{D}_n M = 2\mathcal{D}_n 2$ as above. This yields an embedding of $\mathrm{sym}\, \mathcal{D}_n$ to a subgroup of H_n of G. Since $(\varphi_\pi x)_i = 2$ iff $x_i = 2$ and blocks from $2\mathcal{D}_n 2$ can overlap those from $2\mathcal{D}_m 2$ only in the end symbols when $n \neq m$, elements of H_n commute with those of H_m for $n \neq m$. Thus G contains $\bigoplus_{n=1}^{\infty} \mathrm{sym}\, \mathcal{D}_n$, which clearly contains the direct sum of every countable collection of finite groups.

For general σ_T, use Lemma 2.2 to find three markers M_0, M_1, and M_2 of equal length with only trivial overlaps, such that $M_i M_j \in \mathcal{B}(X_T)$ for all i, j. If $\pi \in \operatorname{sym} \mathcal{D}_n = \operatorname{sym}\{0,1\}^n$, define φ_π to replace a block of the form $M_2 M_{i_1} \cdots M_{i_n} M_2$, where $i_1 \cdots i_n \in \mathcal{D}_n$, with $M_2 M_{j_1} \cdots M_{j_n} M_2$, where $\pi(i_1 \ldots i_n) = j_1 \cdots j_n$, and have no other action. The nonoverlapping nature of the M_i shows that φ_π is well-defined, that $\pi \mapsto \varphi_\pi$ embeds $\operatorname{sym}(\mathcal{D}_n)$ into $\operatorname{aut}(\sigma_T)$, and that the embedded subgroups commute. The proof now concludes as in the first case. □

THEOREM 2.4. *The group* $\operatorname{aut}(\sigma_T)$ *contains the free product of any finite number of* 2*-element groups. Thus it contains the free group on two generators, hence the free group on a countable number of generators.*

PROOF. We embed the free product of three copies of $\mathbb{Z}/2\mathbb{Z}$, the generalization to more copies being routine. We first work on a special full shift, then carry this over to a general σ_T.

Let the alphabet be $\mathcal{L} = \{0,1,2,3,*\}$, and $\sigma_\mathcal{L}$ be the full shift on $\mathcal{L}$. Define involutions φ_j for $j = 1,2,3$ as follows. All will be 2–block maps, and each will exchange three pairs of 2–blocks. Specifically, φ_j exchanges $s0$ with sj for $s \in \mathcal{L} \setminus \{0, j\}$. Thus each φ_j uses three markers for its definition, and has the important property that markers defining its action are not affected by it. It follows that each $\varphi_j \in \operatorname{aut}(\sigma_\mathcal{L})$. Let P be the free product of the 2-element groups $\{e, j\}$ for $j = 1,2,3$. Define a homomorphism from P to $\operatorname{aut}(\sigma_\mathcal{L})$ by mapping a reduced word $w = j_n \cdots j_1 \in P$ to $\psi = \varphi_{j_n} \cdots \varphi_{j_1}$. Since each $\varphi_j^2 = I$, the identity, this is well-defined. Consider the point $x = \cdots 0000{*}0000 \cdots$, with $x_0 = *$. Then $(\psi x)_n = j_n$, $(\varphi_{j_n}^{-1} \psi x)_{n-1} = j_{n-2}$ and so on. This means that inductively ψ determines the spelling of w, so this mapping embeds P into $\operatorname{aut}(\sigma_\mathcal{L})$. It is elementary group theory that P contains the free group F_2 of two generators [**MKS**, §1.4], and it is known [**Ro**, Theorem 11.27] that the commutator subgroup of F_2 is the free group on a countable number of generators.

This idea generalizes to arbitrary σ_T by using markers instead of symbols. If $\mathcal{L}_T$ is the alphabet for σ_T, for each $a \in \{0,1,2,3,*\}$ use Lemma 2.2 to construct a marker M_a over $\mathcal{L}_T$, all of equal length with only trivial overlaps, and beginning with and followed by i_0. Define involutions φ_j, $1 \le j \le 3$, to exchange $M_s M_0 i_0$ with $M_s M_j i_0$ for $s \in \mathcal{L} \setminus \{0, j\}$, and have no other effect. Since these markers have only trivial overlaps, the φ_j are well-defined. The argument that they generate the free product of three copies of $\mathbb{Z}/2\mathbb{Z}$ is exactly as before. □

REMARK 2.5. This theorem shows that G is not amenable.

THEOREM 2.6. *The group* $\operatorname{aut}(\sigma_T)$ *contains the countable direct sum of copies of* $\mathbb{Z}$.

PROOF. We first perform the embedding when σ_T is the full shift on the alphabet $\mathcal{L} = \{0,1,a,b,c\}$, then generalize. Let $M_n = ab^n c$. Then $\mathcal{M} = \{M_n : n \ge 1\}$ has only trivial overlaps. For each $n \ge 1$ we will define two involutions α_n, β_n, and then put $\varphi_n = \alpha_n \beta_n$. The idea behind this construction is contained in [**L2**], where it is used to construct automorphisms with interesting entropies. Define α_n to switch $M_n ij M_n$ with $M_n ji M_n$, where $i, j \in \{0,1\}$, and to have no other action. Define β_n to map $r i M_n j s$ to $j M_n i$, where $i, j, r, s \in \{0,1\}$, and M_n does not move. Declare β_n to have no other action. Clearly α_n and β_n are well-defined involutions that do not

move any M_k, and only affect those symbols 0 and 1 adjacent to the appearances of M_n and that are not adjacent to markers with subscript distinct from n. If $\varphi_n = \alpha_n\beta_n$, it follows that the φ_n commute, but have no other relations. The action of φ_n on the point $(M_n00)^\infty(M_n01)(M_n00)^\infty$ is to shift the block 01 to the left by $|M_n|+2$, proving that φ_n has infinite order. In fact, each φ_n has topological entropy log 4 [**L2**]. Thus the subgroup of G generated by the φ_n is isomorphic to the countable direct sum of copies of $\mathbb{Z}$.

The generalization to arbitrary σ_T using markers M_0, M_1, M_a, M_b, and M_c should now be routine. □

This method allows the embedding of many kinds of countable groups into G. But is there a reasonable answer to the following?

PROBLEM 2.7. *Characterize the subgroups of* $\mathrm{aut}(\sigma_T)$.

At least two properties of countable groups prevent them from being embeddable into $\mathrm{aut}(\sigma_T)$. One is the lack of residual finiteness, which we consider in the next section. For the other, recall that a finitely presented group is said to have solvable word problem if there is an algorithm to decide whether a word in the generators represents the identity. There are countable groups without this property [**Ro**, Chapter 12]. We are endebted to Bruce Kitchens for the following observation.

PROPOSITION 2.8. *The group* $\mathrm{aut}(\sigma_T)$ *contains no finitely generated group with unsolvable word problem.*

PROOF. Suppose a subgroup K of $G = \mathrm{aut}(\sigma)$ has n generators. The inverse of an automorphism is explicitly computable, if only by trying all block maps using coordinates from $-k$ to k and increasing k until the inverse is found. Say that $\varphi \in G$ has *range at most* m if $(\varphi x)_i$ depends on only $x_{i-m}, \ldots, x_{i+m}$. There is an m so that all the generators and their inverses have range at most m. Then a word ψ of length r in the generators and their inverses has range at most rm. As a block map, ψ is explicitly determined by the block maps inducing the generators and their inverses. To check whether $\psi = I$, it is only necessary to see if $\psi(x_{-rm}, \ldots, x_{rm}) = x_0$ for all allowed blocks of length $2rm+1$, a finite procedure. Thus K has solvable word problem. □

That a finitely presented subgroup of $\mathrm{aut}(\sigma_T)$ has solvable word problem follows from the residual finiteness of $\mathrm{aut}(\sigma_T)$, considered in the next section. Proposition 2.8 is stronger. There exist finitely generated residually finite groups with unsolvable word problem. For these facts see Theorem 4.6 in Chapter 4 of [**LS**] and the remarks that follow.

In each of the constructions above we have used a version of the marker method. By this we mean that the automorphism permutes certain blocks when they occur in the context of certain finite marking patterns, so that the marking patterns are not altered by the permutation. A useful point of view is that these marking patterns act as "program" on the "data" of blocks to be permuted. Invariance of the marking patterns is a reflection of the necessary separation of program from data. For the automorphisms φ_π constructed at the beginning of this section, the marking pattern is a pair of M's separated by k symbols, while the data is the collection $\mathcal{D}$ of blocks permuted by π. In more elaborate constructions the marking patterns can be quite complicated (see the proofs of Theorem 9.2 and Lemma 10.7),

but the automorphisms produced will have finite order. At the other extreme, an empty set of marking patterns corresponds to an automorphism that permutes symbols. The following result, apparently first observed by John Franks, shows that the finite-order elements of $\mathrm{aut}(\sigma_T)$ are precisely those that are obtained from a marker construction on a conjugate shift.

PROPOSITION 2.9. *Suppose $\varphi \in \mathrm{aut}(\sigma_T)$ has finite order. There is a shift of finite type σ_U and a conjugacy $\psi\colon X_T \to X_U$ so that $\psi\varphi\psi^{-1}$ is a 1-block permutation of symbols in X_U.*

PROOF. Let $\mathcal{P}_0$ be the partition of X_T into sets $\{x \in X_T\colon x_0 = a\}$ for $a \in \mathcal{L}_T$. Suppose $\varphi^k = I$, and put $\mathcal{P} = \bigvee_{j=0}^{k-1} \varphi^{-j}\mathcal{P}_0$. Note that φ permutes the atoms of $\mathcal{P}$. Since $\mathcal{P}$ has atoms that are compact and open, there is an $n \geq 1$ so that $\bigvee_{j=-n}^{n} \sigma^{-j}\mathcal{P}_0$ refines $\mathcal{P}$. Let $\mathcal{P}_1 = \bigvee_{j=-n}^{n} \sigma^{-j}\mathcal{P}$. Then $\mathcal{P}_1$ is a compact open partition of X_T refining $\mathcal{P}_0$, and a standard argument [**B1**] from symbolic dynamics shows that if n is sufficiently large, $\mathcal{P}_1$ is a 1-step Markov partition for σ_T with transition matrix, say, U indexed by the atoms of $\mathcal{P}_1$. This gives a conjugacy $\psi\colon X_T \to X_U$. Since φ commutes with σ_T, it will permute the atoms of $\mathcal{P}_1$. Hence $\psi\varphi\psi^{-1}$ acts by permuting the symbols of X_U. □

3. Residual finiteness and divisibility. In this section we will prove that $G = \mathrm{aut}(\sigma_T)$ is residually finite. Since this property is inherited by subgroups, it will follow that G does not contain nontrivial divisible groups, nor the infinite symmetric group. We then discuss some divisibility properties of G. We conclude with a construction of a subshift whose automorphism group contains a copy of the rationals.

Recall [**MKS**, p. 116] that an abstract group H with identity I is called *residually finite* if the intersection of all its normal subgroups of finite index is $\{I\}$. This is equivalent to H having enough homomorphisms to finite groups to separate points, and also to being able to embed H into a product of finite groups. The profinite topology on H is the coarsest making all homomorphisms from H to finite groups continuous. Then H is residually finite exactly when the profinite topology on H is Hausdorff [**MKS**, Problem 2.4.24(*a*)]. Clearly a subgroup of a residually finite group is itself residually finite.

THEOREM 3.1. *The group $\mathrm{aut}(\sigma_T)$ is residually finite.*

PROOF. Let $Q_n = Q_n(\sigma_T)$ denote the set of points in X_T with least σ_T-period n. Since σ_T is mixing, each Q_n is finite. An automorphism $\varphi \in G = \mathrm{aut}(\sigma_T)$ is a topological conjugacy of σ_T with itself, hence preserves Q_n. Thus for each $n \geq 1$, an automorphism φ induces a permutation $\varphi|_{Q_n}$ in the symmetric group $\mathrm{sym}\, Q_n$. Let K_n denote the kernel of the map $\varphi \mapsto \varphi|_{Q_n}$. The K_n are normal in G, and since $\bigcup_{n=1}^{\infty} Q_n$ is dense in X_T, it follows that $\bigcap_{n=1}^{\infty} K_n = \{I\}$. Hence G is residually finite. □

A group D is *divisible* if every element has roots of all orders [**MKS**, §6.2]. A consequence of residual finiteness is that complete divisibility cannot occur in G.

COROLLARY 3.2. *The group $\mathrm{aut}(\sigma_T)$ contains no nontrivial divisible groups.*

PROOF. Suppose D is a nontrivial divisible subgroup. Then D is residually finite by Theorem 3.1. Let $\varphi \neq I$ be in D, and N be a normal subgroup of finite

index in D with $\varphi \notin N$. Put $n = |D/N|$. Suppose there were a ψ in D with $\psi^n = \varphi$. Then $N = (\psi N)^n = \psi^n N = \varphi N$, contradicting $\varphi \notin N$. □

To point out the role that finite type plays in Corollary 3.2, we construct in Example 3.9 a subshift not of finite type whose automorphism group contains $\mathbb{Q}$. This construction can be amplified so the resulting automorphism group is exactly $\mathbb{Q}$.

Denote by S_∞ the group of permutations of the natural numbers fixing all but finitely many elements. J. Wagoner has raised the question of whether G contains S_∞. The following negative answer has also been found, independently, by Kim and Roush.

COROLLARY 3.3. *The group* $\mathrm{aut}(\sigma_T)$ *does not contain* S_∞.

PROOF. If A_∞ is the infinite subgroup of S_∞ consisting of the even permutations, then A_∞ is the union of the finite simple alternating groups, so is also simple. If S_∞ were contained in $\mathrm{aut}(\sigma_T)$, then by Theorem 3.1 the subgroup A_∞ would also be residually finite. But A_∞ is infinite and simple, so is not residually finite. □

Corollary 3.2 shows for example that G does not contain $\mathbb{Q}$ or $\mathbb{Z}(p^\infty) = \mathbb{Z}[1/p]/\mathbb{Z}$ for primes p. However, the proof does not rule out partial divisibility.

PROBLEM 3.4. *Is* $\mathbb{Z}[1/p]$ *contained in* $\mathrm{aut}(\sigma_T)$ *for any prime* p?

This amounts to asking whether there is an automorphism of infinite order with an infinite chain of pth roots. Indeed, we are unable to decide the following.

PROBLEM 3.5. *Is there an automorphism in* $\mathrm{aut}(\sigma_T)$ *of infinite order having an* n*th root for infinitely many* n?

Note that if φ is such an automorphism, then it cannot be topologically conjugate to a mixing shift of finite type. For if $\varphi \cong \sigma_U$ and $\psi^n = \varphi$, then ψ is also a mixing shift of finite type ([**BK**, Lemma 2.5] or [**L1**, Theorem 8]), hence ψ is conjugate to some σ_V. Then the spectral radius λ_U of U would be a Perron number [**L1**, §1] with nth root λ_V, which is also a Perron number. But λ_U has only finitely many nontrivial factorizations into Perron numbers [**L1**, Theorem 4], so it has only finitely many Perron roots.

Although we cannot characterize the subgroups of $\mathbb{Q}$ contained in G, there is a complete answer for subgroups of $\mathbb{Q}/\mathbb{Z}$.

PROPOSITION 3.6. *A subgroup of* $\mathbb{Q}/\mathbb{Z}$ *is contained in* $\mathrm{aut}(\sigma_T)$ *iff its* p*-torsion subgroup is finite for every prime* p.

PROOF. Recall that $\mathbb{Q}/\mathbb{Z} \cong \bigoplus_p \mathbb{Z}(p^\infty)$ is the primary decomposition of $\mathbb{Q}/\mathbb{Z}$ [**Ka**, §3]. If $H \subset \mathbb{Q}/\mathbb{Z}$, then H has primary decomposition $H \cong \bigoplus_p H_p$ with $H_p \subset \mathbb{Z}(p^\infty)$. If some H_p is infinite, then it is $\mathbb{Z}(p^\infty)$, and then this divisible group would be contained in G, contradicting Proposition 3.1. If all the H_p are finite, then H is contained in G by Theorem 2.3. □

We now turn to examining chains of roots of the identity I. If $\{n_j : j \geq 1\}$ is a sequence of integers $n_j > 1$ for all j, call a prime p *good* for the sequence if it divides at least one, but only finitely many, of the n_j. Call p *bad* if it divides infinitely many of the n_j. Some primes may be neither good nor bad. We first discuss the case when the roots generate a finite subgroup of G.

PROPOSITION 3.7. *Let $\{n_j : j \geq 1\}$ be a sequence with $n_j > 1$. Then there are $\varphi_j \in \mathrm{aut}(\sigma_T)$ with $\varphi_0 = I$, $\varphi_j^{n_j} = \varphi_{j-1}$ for $j \geq 1$, and which generate a finite nontrivial subgroup of $\mathrm{aut}(\sigma_T)$ iff $\{n_j\}$ has a good prime.*

PROOF. First suppose there are φ_j having the properties mentioned. By skipping to the first $\varphi_j \neq I$ and adjusting indices, we may assume $\varphi_1 \neq I$. Let p be a prime dividing the order $o(\varphi_1)$ of φ_1, and Φ denote the group generated by the φ_j. If p were bad, then φ_1 would have a p^kth root in Φ for every $k \geq 1$. This contradicts finiteness of Φ. Since $p \mid o(\varphi_1)$ and $o(\varphi_1) \mid n_1$, it follows that p is good.

Conversely, suppose p is good. Choose the largest j_0 so that $p \mid n_{j_0}$. Let ψ be a pth root of I, say constructed using Theorem 2.3. Define $\varphi_j = I$ for $0 \leq j < j_0 - 1$, $\varphi_{j_0} = \psi$, and put $\varphi_j = \psi^k$ for $j > j_0$, where $k \equiv (n_{j_0+1} \cdots n_j)^{-1} \bmod p$. An easy calculation shows these φ_j work. □

PROPOSITION 3.8. *Suppose $\{n_j : j \geq 1\}$ is a sequence with $n_j > 1$. Then there are $\varphi_j \in \mathrm{aut}(\sigma_T)$ with $\varphi_0 = I$, $\varphi_j^{n_j} = \varphi_{j-1}$ for $j \geq 1$, and which generate an infinite subgroup of $\mathrm{aut}(\sigma_T)$ iff $\{n_j\}$ has infinitely many good primes.*

PROOF. First suppose there are φ_j as described, and let Φ denote the subgroup of $G = \mathrm{aut}(\sigma_T)$ they generate. Since Φ is a union of cyclic groups, it is abelian. Suppose p divides $o(\varphi_j)$. If p were bad, then every element in Φ would have a pth root. This would force G to contain $\mathbb{Z}(p^\infty)$, contradicting Proposition 3.1. Thus every bad prime is relatively prime to every $o(\varphi_j)$. If $\{n_j\}$ had only finitely many good primes, then there is a j_0 so that for $j \geq j_0$ each n_j is a product of bad primes. Thus $(o(n_j), n_j) = 1$ for $j \geq j_0$, so $o(\varphi_{j+1}) = o(\varphi_j^{n_j}) = o(\varphi_j)$ for $j \geq j_0$, implying Φ is finite. This contradiction proves $\{n_j\}$ has infinitely many good primes.

Conversely, suppose $\{n_j\}$ has infinitely many good primes, say $p_1 < p_2 < \cdots$. Since each p_j divides only finitely many n_k, by passing to a subsequence we can assume there are $n_{i_1} < n_{i_2} < \cdots$ such that $p_j \mid n_{i_j}$ for $j \geq 1$. Put $i_0 = 1$, and let $m_j = n_{i_{j-1}+1} \ldots n_{i_j}$. Clearly it suffices to find a chain of roots for the m_j that generate an infinite subgroup.

Since p_j is good, for each j there is a d_j so that $m_1 \cdots m_k \not\equiv 0 \bmod p_j^{d_j}$ for all $k \geq 1$. Let $H_j = \mathbb{Z}/p_j^{d_j}\mathbb{Z}$, and put $H = \bigoplus_{j=1}^\infty H_j$. Let $a_{j,0} = p_j^{d_j - 1} \in H_j$. Inductively we can find $a_{j,k} \in H_j$ so that $m_{j+k} a_{j,k} = a_{j,k-1}$. Let $b_j = a_{1,j-1} + a_{2,j-2} + \cdots + a_{j,0} \in H$. Then the b_j generate an infinite subgroup of H, and recalling that $p_j \mid m_j$ we find that $m_j b_j = b_{j-1}$. By Theorem 2.6, H embeds into G. If φ_j is the image of b_j under this embedding, then the φ_j satisfy the requirements. □

In Corollary 3.2 we proved that the automorphism group of a shift of finite type cannot contain a divisible group such as $\mathbb{Q}$. If we drop the "finite type" hypothesis, divisible subgroups are possible. The following construction yields a minimal subshift (X, σ) and an embedding of $\mathbb{Q}$ into $\mathrm{aut}(X, \sigma)$.

The details of Example 3.9 are intricate, and the reader may wonder whether a more "natural" action of $\mathbb{Q}$ would suffice. Unfortunately, because $\mathbb{Q}$ is not locally compact, most natural actions of $\mathbb{Q}$ fail to be expansive, so cannot yield subshifts. For example, $\mathbb{Q}$ acts on $\{0,1\}^{\mathbb{Q}}$, but individual elements of this action are not expansive. Expansive maps with roots of arbitrary order are harder to come by.

EXAMPLE 3.9. *A minimal subshift (X, σ) and an embedding of $\mathbb{Q}$ into $\mathrm{aut}(X, \sigma)$ so that $1 \in \mathbb{Q}$ corresponds to $\sigma \in \mathrm{aut}(X, \sigma)$.*

We first sketch how to construct a subshift (X,σ) with an nth root. Suppose there are n symbols $a_0, a_1, \ldots, a_{n-1}$, and one "spacer" symbol s. Further suppose that allowed blocks in X have a_{n-1} always preceded by an s, and a_0 always followed by an s. Define a block map φ by $\varphi(a_j) = a_{j+1}$ for $0 \le j \le n-2$, and $\varphi(sa_{n-1}) = a_0 s$. In general, $\varphi(X)$ need not be X. However, if X is designed so $\varphi(X) = X$, then φ gives an automorphism of (X,σ). As the iterates of φ act on $x \in X$, different parts of x are moved left one position at different iterates, much like the familiar slinky toy. The cumulative effect of φ^n is to move every symbol to the left once, so φ is an nth root of σ.

Before giving the detailed construction, let us describe the role of the objects and maps obtained. The construction will proceed by stages, starting with an initial alphabet $\mathcal{L}$. At stage $q \ge 3$ we will have $q!$ words $w_0^{(q)}, \ldots, w_{q!-1}^{(q)}$ from $\mathcal{L}$ forming the set $\mathcal{W}^{(q)}$. Each $w_j^{(q)}$ will be a concatenation of words from $\mathcal{W}^{(q-1)}$ separated by 0, 1, or 2 spacer symbols s. The subshift X will consist of all $x \in \mathcal{L}^{\mathbf{Z}}$ so that every subblock of x is also a subblock of some constructed word. Every word from $\mathcal{W}^{(q-1)}$ will occur in every word from $\mathcal{W}^{(q)}$, and s^3 will never occur in any word. From this it will follow that every allowed block in X occurs syndetically, so (X,σ) will be minimal. It will also follow that if $x \in X$ and $q \ge 3$, then $x[-\infty,\infty]$ can be uniquely decomposed into a concatenation of words from $\mathcal{W}^{(q)}$ separated by 0, 1, or 2 spacers s. For each $q \ge 3$ there will be maps $\varphi_k^{(q)}$ $(3 \le k \le q)$ defined on words in $\mathcal{W}^{(q)}$. They will have the properties that $(\varphi_k^{(q)})^k = \varphi_{k-1}^{(q)}$ for $4 \le k \le q$, that $(\varphi_3^{(3)})^6 = \sigma$, and the consistency condition that $\varphi_k^{(q)}$ applied to a word in $\mathcal{W}^{(q+r)}$ gives the same result as $\varphi_k^{(q+r)}$ would. Thus on X the $\varphi_k^{(q)}$ $(q \ge 3)$ consistently define a block map φ_k, and these obey $\varphi_k^k = \varphi_{k-1}$ $(k \ge 4)$, $\varphi_3^6 = \sigma$. Hence mapping $1/k!$ to φ_k for $k \ge 3$ embeds $\mathbf{Q}$ into $\mathrm{aut}(X,\sigma)$, with 1 corresponding to σ.

To begin the construction, let the alphabet be $\mathcal{L} = \{a_0, a_1, \ldots, a_5, s\}$. For the initial stage $q = 3$, put $w_j^{(3)} = a_j$ $(0 \le j \le 5)$, and $\mathcal{W}^{(3)} = \{w_j^{(3)} : 0 \le j \le 5\}$. Define $\varphi_3^{(3)}(w_j^{(3)}) = w_{j+1}^{(3)}$ $(0 \le j \le 4)$, and $\varphi_3^{(3)}(sw_5^{(3)}) = w_0^{(3)} s$. This is the method outlined in the first paragraph to obtain a 6th root of σ, so $(\varphi_3^{(3)})^6 = \sigma$, and we need only make sure that $\varphi_3^{(3)}(X) = X$.

Next we give the first inductive step, to $q = 4$. Begin by defining for $0 \le m \le 3$ the 4 words

$$w_m^{(4)} = w_0^{(3)}(sw_0^{(3)})^3 \left[(sw_m^{(3)})(sw_{m+1 \bmod 6}^{(3)}) \cdots (sw_{m+5 \bmod 6}^{(3)})\right] (sw_0^{(3)})^4.$$

We then obtain the 4! words in $\mathcal{W}^{(4)}$ by defining

$$w_{4r+m}^{(4)} = (\varphi_3^{(3)})^r(w_m^{(4)}) \qquad (0 \le r \le 5,\ 0 \le m \le 3).$$

Note that every word in $\mathcal{W}^{(3)}$ occurs in every word in $\mathcal{W}^{(4)}$. Next put $\varphi_4^{(4)}(w_j^{(4)}) = w_{j+1}^{(4)}$ $(0 \le j \le 4! - 2)$, and $\varphi_4^{(4)}(sw_{4!-1}^{(4)}) = w_0^{(4)} s$. Then let $\varphi_3^{(4)} = (\varphi_4^{(4)})^4$. Words from $\mathcal{W}^{(4)}$ are cyclically moved by $\varphi_4^{(4)}$ but break into 4 groups of 6 each on which $\varphi_3^{(4)}$ acts exactly as $\varphi_3^{(3)}$. Thus $\varphi_3^{(4)}$ is consistent with $\varphi_3^{(3)}$, and $(\varphi_4^{(4)})^4 = \varphi_3^{(4)}$ by construction. Since every word from $\mathcal{W}^{(3)}$ occurs in every word from $\mathcal{W}^{(4)}$, and s^3 never occurs here, the minimality conditions are satisfied at stage $q = 4$.

Suppose at stage $q-1$ we have defined $(q-1)!$ words $w_j^{(q-1)}$ for $0 \le j \le (q-1)!-1$ and maps $\varphi_k^{(q-1)}$ $(3 \le k \le q-1)$ so that $(\varphi_k^{(q-1)})^k = \varphi_{k-1}^{(q-1)}$ and the $\varphi_k^{(3)}, \ldots, \varphi_k^{(q-1)}$ are consistent. Construct stage q as follows. First define, for $0 \le j \le q-1$,

$$w_m^{(q)} = w_0^{(q-1)}(sw_0^{(q-1)})^{q-1}\left[\prod_{i=0}^{(q-1)!-1} sw_{m+i \bmod (q-1)!}^{(q-1)}\right](sw_0^{(q-1)})^q.$$

We then obtain the $q!$ words in $\mathcal{W}^{(q)}$ by putting

$$w_{qr+m}^{(q)} = (\varphi_{q-1}^{(q-1)})^r(w_m^{(q)}) \qquad (0 \le r < (q-1)!,\ 0 \le m \le q-1).$$

Next define $\varphi_q^{(q)}$ by $\varphi_q^{(q)}(w_j^{(q)}) = w_{j+1}^{(q)}$ for $0 \le j \le q!-1$, and $\varphi_q^{(q)}(sw_{q!-1}^{(q)}) = w_0^{(q)}s$. Finally, set

$$\varphi_k^{(q)} = (\varphi_q^{(q)})^{q!/k!} \qquad (3 \le k \le q).$$

Then, as in the $q=4$ case, for fixed k each $\varphi_k^{(p)}$ $(k \le p \le q)$ is consistent with $\varphi_k^{(q)}$, and $(\varphi_k^{(q)})^k = \varphi_{k-1}^{(q)}$. Every word in $\mathcal{W}^{(q-1)}$ occurs in every word of $\mathcal{W}^{(q)}$, and s^3 never occurs. Furthermore, because of the repetitions of $sw_0^{(q-1)}$ at the ends of the $w_m^{(q)}$, every $x \in X$ has $x[-\infty,\infty]$ decomposed uniquely as a concatenation of words from $\mathcal{W}^{(q)}$ separated by no more than 2 s's. This completes the construction of stage q.

The $\mathcal{W}^{(q)}$ and $\varphi_k^{(q)}$ constructed obey the conditions described in the second paragraph, and we thus obtain the required minimal subshift (X,σ) and embedding of $\mathbb{Q}$ into $\mathrm{aut}(X,\sigma)$. □

REMARK 3.10. In our construction of $w_j^{(q)}$, the words from $\mathcal{W}^{(q-1)}$ are cyclically listed once. However, any arrangement of words from $\mathcal{W}^{(q-1)}$, with arbitrary repetitions, would also work, provided each word is used at least once. By using a long and highly recurrent listing of words from $\mathcal{W}^{(q-1)}$, it is possible to construct a uniquely ergodic subshift (X,σ) so that $\mathrm{aut}(X,\sigma) \cong \mathbb{Q}$, and such that every Borel measurable mapping $\varphi: X \to X$ commuting with σ is continuous. This should be contrasted with the uniquely ergodic Morse minimal subshift, whose automorphism group is just $\mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})$ with generators the shift and coordinate complementation (see [**CK**, **C**]).

4. Nonisomorphic automorphism groups. Conjugate subshifts of finite type have isomorphic automorphism groups. Also, clearly $\mathrm{aut}(\sigma_T) = \mathrm{aut}(\sigma_T^{-1})$. Since there are shifts of finite type not conjugate to their inverse [**PT**, Proposition 3.30], there are nonconjugate shifts of finite type with isomorphic automorphism groups. However, the following question remains open.

QUESTION 4.1. *If σ_U is not conjugate to σ_T or σ_T^{-1}, can* $\mathrm{aut}(\sigma_U)$ *and* $\mathrm{aut}(\sigma_T)$ *be isomorphic as abstract groups?*

In this regard, we know of only one useful invariant, namely the theorem of Ryan [**Ry1**, **Ry2**] that the center of $\mathrm{aut}(\sigma_T)$ is precisely the set of powers of σ_T. Thus, for example, as an abstract group $\mathrm{aut}(\sigma_{[4]}) = \mathrm{aut}(\sigma_{[2]}^2)$ has center generated by an element with a square root in the group, while $\mathrm{aut}(\sigma_{[2]})$ does not since the

2-shift has no square root. The following example, a more refined application of this idea, shows that the isomorphism class of $\mathrm{aut}(\sigma_T)$ is not determined by the zeta-function. The motivation for the specific matrices used lies in a consequence of [**B2**] that there is only one shift equivalence class over $\mathbb{Z}$ for matrices whose nonzero spectrum is $\{1,2\}$, while there are exactly two such classes for $\{1,8\}$, only one of which can correspond to the cube of the first class.

EXAMPLE 4.2. *Two mixing shifts of finite type with equal ζ-function and having nonisomorphic automorphism groups.*

Let

$$V = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \qquad T = V^3, \qquad U = \begin{bmatrix} 7 & 6 \\ 1 & 2 \end{bmatrix}.$$

Then $\zeta_T(t) = \zeta_U(t) = \left[(1-8t)(1-t)\right]^{-1}$. Clearly the generator σ_T for the center of $\mathrm{aut}(\sigma_T)$ has a cube root σ_V. Suppose σ_U has a cube root φ in $\mathrm{aut}(\sigma_U)$. Then φ is again a shift of finite type ([**BK**, Lemma 2.5] or [**L1**, Theorem 8]), say $\varphi = \sigma_W$. Since $\sigma_U = (\sigma_W)^3 \cong \sigma_{W^3}$, the nonzero spectrum of W is $\{1,2\}$ counting multiplicity. Hence by [**B2**], there is an integer j so that W is shift equivalent over $\mathbb{Z}$ to $\left[\begin{smallmatrix} 2 & 0 \\ j & 1 \end{smallmatrix}\right] = W_1$. Since

$$W_1 = \begin{bmatrix} 1 & 0 \\ j & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -j & 1 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 0 \\ -j & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ j & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$U = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 8 & 8 \\ 1 & 2 \end{bmatrix}, \qquad \begin{bmatrix} 8 & 8 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & 1 \end{bmatrix},$$

we would have

$$W_1^3 = N = \begin{bmatrix} 8 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 8 & 0 \\ 1 & 1 \end{bmatrix}$$

shift equivalent over $\mathbb{Z}$. Thus there would be matrices R and S over $\mathbb{Z}$ and a positive integer l with $RS = M^l$, $SR = N^l$, $NR = RM$, and $SN = MS$. Now $NR = RM$ forces $R_{12} = 0$ and $7R_{21} = -R_{22}$. Also, $SN = MS$ forces $S_{12} = 0$. Then $RS = M^l$ implies $R_{22}S_{22} = 1$, so $R_{22} = \pm 1$. This contradicts $7R_{21} = -R_{22}$. □

5. Symmetry. Let σ_T be a mixing shift of finite type. As in the proof of Proposition 2.8, say that $\varphi \in G = \mathrm{aut}(\sigma_T)$ has *range at most* n if $(\varphi x)_i$ depends only on $x_{i-n}, \ldots, x_{i+n}$. A natural measure of the symmetry of σ_T is the rate of growth of the subset $G_n(\sigma_T)$ of those automorphisms in G with range at most $[n/2]$. This growth turns out to be doubly exponential, so we define the *symmetry* of σ_T as

$$s(\sigma_T) = \limsup_{n\to\infty} \frac{1}{n} \log\log |G_n(\sigma_T)|.$$

The definition of symmetry depends on a particular presentation of σ_T, but the following proves that symmetry is a conjugacy invariant. Suppose σ_T and σ_U are conjugate via a mapping $\psi: X_T \to X_U$. Choose m so that both ψ and ψ^{-1} have range at most m. If $\varphi \in G_n(\sigma_T)$, then $\psi\varphi\psi^{-1} \in G_{n+4m}(\sigma_U)$, so that $|G_{n+4m}(\sigma_U)| \geq |G_n(\sigma_T)|$. Thus $s(\sigma_U) \geq s(\sigma_T)$, and by symmetry they coincide.

THEOREM 5.1. *If σ is a mixing shift of finite type, then*

$$\tfrac{1}{2}h(\sigma) \le s(\sigma) \le h(\sigma).$$

PROOF. We first establish the upper bound. Represent σ as σ_T with r states, and put $X = X_T$. Now $|G_n(\sigma_T)|$ is trivially bounded above by the total number of block maps $\mathcal{B}_{n+1}(X) \to \{0, \dots, r-1\}$, or $r^{|\mathcal{B}_{n+1}(X)|}$. If $h(\sigma_T) = \log\lambda$, where λ is the dominant eigenvalue for T, there is a $\kappa > 0$ so that $|\mathcal{B}_{n+1}(X)| < \kappa\lambda^n$. Thus $|G_n(\sigma)| < r^{\kappa\lambda^n}$, so that $s(\sigma) \le \log\lambda = h(\sigma)$.

The lower bound uses the automorphisms constructed at the beginning of §2. For this we first need to show that the number of marker-free blocks of length m between two markers grows faster than $(\lambda^{1-\varepsilon})^m$, where ε can be made small by choosing long enough markers.

Fix a small $\varepsilon > 0$ and a marker $M \in \mathcal{B}_n(X)$, where $n \ge n_0$ with n_0 to be determined. Let l be a transition length for T, so that $T^l > 0$. There is an $a > 0$ so that $|\mathcal{B}_k(X)| > a\lambda^k$ for $k \ge 1$. If $k \le n$, then M has $n-k+1$ subblocks of length k. Hence if k is chosen so that $a\lambda^k > n-k+1$, there will be a block $C \in \mathcal{B}_k(X)$ that does not appear in M. If n_0 is large enough, a choice of $k < \frac{\varepsilon}{2}n - l$ is possible.

Consider blocks $MDM \in \mathcal{B}_N(X)$ of the form

$$MECB_1CB_2 \cdots CB_KCFM,$$

where $B_j \in \mathcal{B}_{n-2k}(X)$, $N = (n-k)K + 2n + 2l + k$, and E, F are fixed transition blocks in $\mathcal{B}_l(X)$. Since l is a transition length, every block of $\mathcal{B}_{n-2k-2l}(X)$ can be the central part of each B_j. Hence the collection $\mathcal{D}$ of blocks D with the required form has cardinality

$$|\mathcal{D}| \ge \left(a\lambda^{n-2k-2l}\right)^K \ge \left(a\lambda^{(1-\varepsilon)n}\right)^K.$$

Since $2|C| + |B_j| = n$, a subblock of $D \in \mathcal{D}$ with length n must contain C. Now M does not contain C, and has only trivial self-overlap. Thus M can only occur in MDM as the initial or terminal segment. As at the beginning of §2, distinct permutations $\pi \in \operatorname{sym}\mathcal{D}$ determine distinct automorphisms $\varphi_\pi \in \operatorname{aut}(\sigma_T)$ whose range is clearly at most N. Thus

$$|G_{2N+1}(\sigma_T)| \ge \left[\left(a\lambda^{(1-\varepsilon)n}\right)^K\right]!\,.$$

Now Stirling's formula implies that $\log m! > m(\log m - 1)$, which applied to the above yields

$$\begin{aligned} s(\sigma_T) &\ge \limsup_{N\to\infty} \frac{1}{2N+1} \log\log|G_{2N+1}| \\ &\ge \limsup_{K\to\infty} \frac{1}{2(n-k)K + 4(n+l) + 2k + 1} \log\left(a\lambda^{(1-\varepsilon)n}\right)^K \\ &\ge \frac{1-\varepsilon}{2}\log\lambda = \frac{1-\varepsilon}{2}h(\sigma_T). \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the lower estimate is proved. □

We do not know the exact value of $s(\sigma_T)$ for any $T \ne [1]$, nor whether the definition's lim sup is actually a limit.

PROBLEM 5.2. *Compute* $s(\sigma_T)$.

For full shifts $\sigma_{[r]}$ this problem can be viewed as a quantitative inquiry into the relative sparseness of $\mathrm{aut}(\sigma_{[r]})$ in the semigroup of all block maps from $X_{[r]}$ to itself. A topological measure of this sparseness has been given by Sears [**Se**]. We remark that for $\sigma_{[r]}$, if we replace "automorphism" by "surjective map" in the definition of symmetry, the value is easily shown to be $\log r = h(\sigma_{[r]})$. For any block map involving k symbols can be used to define a right permutive block map using $k+1$ symbols by adding modulo r the value of the block map on the initial k symbols to the last coordinate. Such right permutive maps are surjective [**H**, Theorem 6.6].

NOTE ADDED IN PROOF. H. Kim and F. Roush have shown that $s(\sigma_T) = h(\sigma_T)$ for all irreducible T. Their solution to Problem 5.2 uses a modification of the construction used to prove Theorem 5.1.

Let μ_T be the measure of maximal entropy for σ_T. Coven and Paul [**CP**] proved that μ_T is φ-invariant for every $\varphi \in G$. Since G is not amenable (Remark 2.5), the sets G_n may provide a replacement for averaging sets for the action of G on (X_T, μ_T).

QUESTION 5.3. *If* $f \in C(X_T)$, *does* $|G_n|^{-1} \sum_{\varphi \in G_n} f(\varphi x)$ *converge to*

$$\int_{X_T} f \, d\mu_T$$

in any reasonable sense for most or all $x \in X_T$?

6. Induced action on the dimension group. For this section only, we drop our standing assumption that T be aperiodic, and assume only that T is nonnegative integral and not nilpotent. Building on the fundamental work of Williams [**Wi**] and Elliot [**E**], Krieger [**Kr2**] associated to each shift of finite type σ_T an order-preserving automorphism $\widehat{T}$ of an ordered group $(\mathcal{G}_T, \mathcal{G}_T^+)$ called the dimension group of σ_T. The triple $(\mathcal{G}_T, \mathcal{G}_T^+, \widehat{T})$ is a topological invariant since it can be defined using only topological notions. Two shifts of finite type are shift equivalent exactly when their dimension triples are isomorphic [**Kr2**].

We first review an algebraic description of the dimension triple, and indicate its relationship to underlying topological notions. Next we show how an automorphism φ of σ_T induces an automorphism $\delta(\varphi)$ of the dimension triple. The basic problem is to determine whether δ is surjective. We do not settle this. We show in Theorem 6.8 that if the nonzero eigenvalues of T are simple, and if the ratio of distinct eigenvalues is not a root of unity, then for all sufficiently large n the map $\delta\colon \mathrm{aut}(\sigma_T^n) \to \mathrm{aut}(\widehat{T}^n)$ is surjective. We also show that these hypotheses on the eigenvalues of T imply $\mathrm{aut}(\widehat{T}) = \mathrm{aut}(\widehat{T}^n)$ for all n. Example 6.7 shows that in general $\mathrm{aut}(\widehat{T}^2)$ can be larger than $\mathrm{aut}(\widehat{T})$, though still finitely generated. Also, using an idea suggested to us by Gopal Prasad, in Example 6.9 we exhibit a T for which $\mathrm{aut}(\widehat{T})$ is not finitely generated.

Suppose T is an $r \times r$ nonnegative integral matrix. For the moment we drop our standing assumption that T be aperiodic. It will be convenient to have matrices act on the right, and for vectors to be row vectors. Say that $v \in \mathbb{Q}^r$ is *eventually integral* (under T) if $vT^n \in \mathbb{Z}^r$ for large enough n. Call two eventually integral vectors v and w *equivalent* if $vT^n = wT^n$ for large enough n. The set $\mathcal{G}_T$ of equivalence

classes $[v]$ of eventually integral vectors v inherits an additive group structure from $\mathbb{Q}^r$. The positive cone $\mathcal{G}_T^+$ is the set of $[v]$ for the $vT^n \geq 0$ eventually.

Let $R = T^r\mathbb{Q}^r$ be the eventual range of T. Each class in $\mathcal{G}_T$ has a unique representative in R, so it will sometimes be convenient to regard $\mathcal{G}_T$ as embedded in R. Define $\widehat{T}$ on $\mathcal{G}_T$ by $\widehat{T}([v]) = [vT]$. By considering $\mathcal{G}_T$ as a subgroup of R, the extension of $\widehat{T}$ to R is invertible, so $\widehat{T}$ defines an automorphism of $\mathcal{G}_T$. It is clear that $\widehat{T}$ is order-preserving. By the *dimension triple* of σ_T we shall mean $(\mathcal{G}_T, \mathcal{G}_T^+, \widehat{T})$.

We now sketch Krieger's topological construction of the dimension triple. For a detailed introduction to these ideas, see [**BMT**, Chapters 2, 5, 11].

By an *n-ray* we shall mean a set of the form

$$x(-\infty, n]^* = \{y \in X_T : y_j = x_j \text{ for } -\infty < j \leq n\},$$

where $x \in X_T$. By an *n-beam* we mean a finite disjoint union of n-rays. Note that although the notion of n-ray is tied to the presentation T, that of n-beam is not, since an n-beam can be topologically described as a finite union of compact open subsets of unstable sets in X_T in the natural inductive limit topology. If $C = \bigcup_j x^j(-\infty, n]^*$ is an n-beam, define a vector $v_{C,n} \in \mathbb{Z}^r$ whose ith entry is the number of n-rays x^j such that x_n^j has terminal state i. If C is an n-beam, then it is also an m-beam for $m \geq n$, and $v_{C,m} = v_{C,n}T^{m-n}$. Define beams C and D to be equivalent if for large enough k we have $v_{C,k} = v_{D,k}$. The set of equivalence classes generate the positive cone $\mathcal{G}^+$ of an ordered group $\mathcal{G}$ using the definition $[C] + [D] = [C \cup D]$ if $C \cap D = \emptyset$. Now σ_T acts on beams, preserves equivalence, so induces an order-preserving automorphism $\widehat{\sigma}_T$ of $\mathcal{G}$. Since equivalence of beams corresponds to equivalence of eventually integral vectors, it is routine to verify that the map sending the class $[C]$ of an n-beam to $\widehat{T}^{-n}([v_{C,n}])$ is an isomorphism of $(\mathcal{G}, \mathcal{G}^+, \widehat{\sigma}_T)$ to the dimension triple $(\mathcal{G}_T, \mathcal{G}_T^+, \widehat{T})$ defined above.

Therefore, to prove that the dimension triple is a topological invariant, it suffices to prove that if $\psi : \sigma_T \to \sigma_U$ is a topological conjugacy, then ψ maps beams to beams, respects equivalence of beams, and intertwines $\widehat{\sigma}_T$ with $\widehat{\sigma}_U$. The routine verifications are omitted. However, we add one note of caution. If $\psi : X_T \to X_U$ is merely surjective, it is not necessarily true that ψ maps beams to beams.

We shall denote the group of order-preserving automorphisms of the dimension triple by $\mathrm{aut}(\widehat{T})$. Suppose $\varphi \in \mathrm{aut}(\sigma_T)$. Since φ is a self-conjugacy of σ_T, the argument above about topological invariance proves that φ induces an automorphism $\delta(\varphi) \in \mathrm{aut}(\widehat{T})$. A routine calculation shows $\delta(\varphi\psi) = \delta(\varphi)\delta(\psi)$, so $\delta : \mathrm{aut}(\sigma_T) \to \mathrm{aut}(\widehat{T})$ is a homomorphism.

QUESTION 6.1. *Is the dimension representation* $\delta : \mathrm{aut}(\sigma_T) \to \mathrm{aut}(\widehat{T})$ *always surjective?*

If U is an integral matrix commuting with T, we define $\widehat{U}$ on $\mathcal{G}_T$ by $\widehat{U}([v]) = [vU]$.

LEMMA 6.2. *Suppose U and V are nonnegative integral matrices so that $T = UV = VU$ is an elementary strong shift equivalence of T to itself. Then there is a $\varphi \in \mathrm{aut}(\sigma_T)$ with $\delta(\varphi) = \widehat{U}$.*

PROOF. This is mainly a matter of checking that the induced automorphism defined by Williams [**Wi**] acts correctly on beams.

Suppose T is $r \times r$. Let $W = \left[\begin{smallmatrix} 0 & U \\ V & 0 \end{smallmatrix}\right]$, so $W^2 = \left[\begin{smallmatrix} T & 0 \\ 0 & T \end{smallmatrix}\right]$, indexed by states $\{0, 1, \ldots, 2r-1\}$. Then $\mathcal{B}_2(\sigma_W)$ is the disjoint union of the set $\mathcal{B}_{U,V}$ of paths beginning and ending in $\{0, \ldots, r-1\}$ and the set $\mathcal{B}_{V,U}$ beginning and ending in $\{r, \ldots, 2r-1\}$. Thus X_{W^2} is the disjoint union of $X_{U,V}$ defined from $\mathcal{B}_{U,V}$ and $X_{V,U}$ defined from $\mathcal{B}_{V,U}$. Note that σ_W switches these sets. From the form of W^2, it follows that there are bijections $\theta_{U,V}\colon \mathcal{B}_1(\sigma_T) \to \mathcal{B}_{U,V}$ and $\theta_{V,U}\colon \mathcal{B}_1(\sigma_T) \to \mathcal{B}_{V,U}$ that respect initial and terminal states mod r. These induce 1-block conjugacies, denoted by the same symbol, $\theta_{U,V}\colon (X_T, \sigma_T) \to (X_{U,V}, \sigma_W^2)$ and $\theta_{V,U}\colon (X_T, \sigma_T) \to (X_{V,U}, \sigma_W^2)$. Let $\varphi = \theta_{V,U}^{-1} \sigma_W \theta_{U,V} \in \mathrm{aut}(\sigma_T)$.

We compute the action of φ on a left-infinite ray $x(-\infty, n]^*$. Suppose x_n has terminal state i. Then $\theta_{U,V}$ maps this to a ray in $X_{U,V}$ ending with state i. Next σ_W maps this to a union of rays, with U_{ij} of them ending in state j. Finally, $\theta_{V,U}^{-1}$ maps each ray ending in state j to one in X_T ending in state j. Passing to the action of φ on $\mathcal{G}_T$, we see the standard unit vector e_i is mapped by $\delta(\varphi)$ to $\sum_{j=1}^{r} U_{ij} e_j$. Thus $\delta(\varphi) = \widehat{U}$. □

Note that $\mathrm{aut}(\sigma_T)$ is naturally a subgroup of $\mathrm{aut}(\sigma_T^n)$. Furthermore, the restriction of $\delta\colon \mathrm{aut}(\sigma_T^n) \to \mathrm{aut}(\widehat{T}^n)$ to $\mathrm{aut}(\sigma_T)$ coincides with the definition of δ of $\mathrm{aut}(\sigma_T)$.

PROPOSITION 6.3. *Suppose $\Phi \in \mathrm{aut}(\widehat{T})$. Then for all sufficiently large n, there is a $\varphi \in \mathrm{aut}(\sigma_T^n)$ with $\delta(\varphi) = \Phi$.*

PROOF. Let T be $r \times r$ acting on $\mathbb{Q}^r$. Then $R = T^r \mathbb{Q}^r$ is the eventual range of T, and $K = \ker T^r$ is its eventual kernel. Hence $\mathbb{Q}^r = R \oplus K$. As before, we may consider $\mathcal{G}_T$ as embedded in R. Since $\mathcal{G}_T$ is torsion-free, the automorphism Φ extends to a nonsingular $\mathbb{Q}$-linear map W of $\mathcal{G}_T \otimes \mathbb{Q} = R$. If 0_K denotes the zero map on K, and U_0 is the matrix for $0_K \oplus W$ with respect to the standard basis, then $U_0 = 0_K \oplus W$ has rational entries, and $\widehat{U}_0 = \Phi$. Let $V_0 = 0_K \oplus W^{-1}$. Both U_0 and V_0 have rows in $\mathcal{G}_T^+$, so for k and $n-k$ large enough, $U = U_0 T^k$ and $V = V_0 T^{n-k}$ have $\mathbb{Z}^+$ entries. Thus $UV = (0_K \oplus I_R) T^n = T^n = VU$ is an elementary strong shift equivalence of T^n to itself. By Lemma 6.1, there is a $\psi \in \mathrm{aut}(\sigma_T^n)$ with $\delta(\psi) = \widehat{U}$. Put $\varphi = \psi \sigma_T^{-k}$. Then $\delta(\varphi) = \delta(\psi)\delta(\sigma_T^{-k}) = \widehat{U}_0 \widehat{T}^k \widehat{T}^{-k} = \widehat{U}_0 = \Phi$. □

To use Proposition 6.3, we will establish some results about finite generation of $\mathrm{aut}(\widehat{T})$. In what follows, $\mathrm{aut}\,(\mathcal{G}_T, \widehat{T})$ refers to the group of those automorphisms of $\mathcal{G}_T$ commuting with $\widehat{T}$, not necessarily preserving the positive cone. If U has nonzero eigenvalues $\lambda_1, \ldots, \lambda_k$, let $\chi_U^{\times}(t)$ denote $\prod_{j=1}^{k}(t - \lambda_j)$.

LEMMA 6.4. *Suppose $\chi_U^{\times}$ is irreducible. Then $\mathrm{aut}(\mathcal{G}_U, \widehat{U})$ is finitely generated and abelian.*

PROOF. By passing to the eventual range of U, we may suppose U is nonsingular. Let U be $r \times r$, and let λ be an eigenvalue of U. By a theorem of Taussky [**T**], there is an ideal $J \subset \mathbb{Z}[\lambda]$ so that $(\mathbb{Z}^r, U) \cong (J, M_\lambda)$, where M_λ denotes multiplication by λ. Hence $(\mathcal{G}_U, \widehat{U}) \cong (\mathbb{Z}[1/\lambda]J, M_\lambda)$. For more on this correspondence, see [**BMT**, Chapter 5]. Under this isomorphism, an automorphism Φ of $\widehat{U}$ corresponds to an automorphism of $\mathbb{Z}[1/\lambda]J$ commuting with M_λ, i.e. a $\mathbb{Z}[1/\lambda]$-module isomorphism. Since the quotient field of $\mathbb{Z}[1/\lambda]$ is $\mathbb{Q}(\lambda)$, there is an $\alpha \in \mathbb{Q}(\lambda)$ so that Φ corresponds to the restriction of M_α on $\mathbb{Z}[1/\lambda]J$. Let S be the set of prime divisors of the ideal

generated by λ, and $\mathcal{O}_S$ denote the ring of S-integral elements in $\mathbb{Q}(\lambda)$. Since both M_α and $M_{\alpha^{-1}}$ are automorphisms of $\mathbb{Z}[1/\lambda]J$, it follows that α is in the unit group $\mathcal{O}_S^\times$ of $\mathcal{O}_S$. By the Dirichlet unit theorem [**We**, Theorem 5–3–10], the group $\mathcal{O}_S^\times$ of units is a finitely generated abelian group. Thus $\mathrm{aut}(\widehat{U})$ corresponds to a subgroup of $\mathcal{O}_S^\times$, so is also finitely generated and abelian. □

REMARK. The Dirichlet unit theorem shows that $\mathrm{aut}(\mathcal{G}_U, \widehat{U})$ is the product of a finite cyclic group and a free abelian group on $e + r + s - 1$ generators, where $\mathbb{Q}(\lambda)$ has r real and $2s$ complex embeddings, and the factorization of the principal ideal generated by λ uses e distinct primes.

PROPOSITION 6.5. *If all the nonzero eigenvalues of T are simple, then* $\mathrm{aut}(\widehat{T})$ *is finitely generated and abelian.*

PROOF. Factor $\chi_T(t) = t^m p_1(t) \cdots p_k(t)$, where the $p_j(t)$ are distinct irreducibles. If T is $r \times r$, then $\mathbb{Q}^r$ is the direct sum of the eventual kernel of T and the rational subspaces R_j corresponding to the $p_j(t)$. Let $\mathcal{G}_j = \mathcal{G}_T \cap R_j$, and $\widehat{T}_j = \widehat{T}|_{\mathcal{G}_j}$. Suppose $\Phi \in \mathrm{aut}(\widehat{T})$. As before, Φ extends to a $\mathbb{Q}$-linear map of $\mathbb{Q}^r$ which is invertible on $\bigoplus_{j=1}^k R_j$. Since the $p_j(t)$ are distinct and irreducible, each R_j is Φ-invariant. It follows $\Phi|_{\mathcal{G}_j}$ is an automorphism of $(\mathcal{G}_j, \widehat{T}_j)$. Hence the map $\Phi \mapsto \bigoplus_{j=1}^k \Phi|_{\mathcal{G}_j}$ takes $\mathrm{aut}(\mathcal{G}_T, \widehat{T})$ to $\bigoplus_{j=1}^k \mathrm{aut}(\mathcal{G}_j, \widehat{T}_j)$. Since $\bigoplus_{j=1}^k \mathcal{G}_j$ has finite index in $\mathcal{G}_T$, this mapping is injective. By using an integral basis for $\mathbb{Z}^r \cap R_j$, the map $\widehat{T}_j$ is seen to be a dimension group automorphism, so $\mathrm{aut}(\mathcal{G}_j, \widehat{T}_j)$ is finitely generated abelian by Lemma 6.4. This proves $\mathrm{aut}(\mathcal{G}_T, \widehat{T})$ is finitely generated abelian. The group $\mathrm{aut}(\widehat{T})$ of order-preserving automorphisms is therefore also finitely generated abelian. □

LEMMA 6.6. *Suppose the nonzero eigenvalues of T are simple. If T does not have distinct eigenvalues whose ratio is an nth root of unity, then* $\mathrm{aut}(\widehat{T}^n) = \mathrm{aut}(\widehat{T})$.

PROOF. As noted above, $\mathrm{aut}(\widehat{T}) \subset \mathrm{aut}(\widehat{T}^n)$ is trivial.

Using the notations from the proof of Proposition 6.5, we see the hypotheses on the eigenvalues of T mean that the spectra of $T^n|_{R_j}$ are disjoint for $1 \le j \le k$. Thus any $\Phi \in \mathrm{aut}(\widehat{T}^n)$ extends to a $\mathbb{Q}$-linear map under which the R_j are invariant. Since the eigenvalues of T^n on R_j are distinct, and Φ commutes with T^n on R_j, it follows by linear algebra that Φ commutes with T on R_j, completing the proof. □

EXAMPLE 6.7. *An aperiodic matrix T for which* $\mathrm{aut}(\widehat{T}^2)$ *is larger than* $\mathrm{aut}(\widehat{T})$, *but both are finitely generated.*

Let

$$T = \begin{bmatrix} 5 & 2 & 2 \\ 4 & 1 & 4 \\ 0 & 6 & 3 \end{bmatrix}.$$

Then over $\mathbb{Z}[1/3]$ we find

$$V^{-1}TV = U = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{bmatrix}, \qquad \text{where } V = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix}.$$

Since the eigenvalues of U are distinct, we obtain $\mathcal{G}_U \cong \mathbb{Z}[1/3]^3$, and $\mathrm{aut}(\mathcal{G}_U, \widehat{U}) \cong GL(1, \mathbb{Z}[1/3])^3 \cong [\mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})]^3$. Since order-preserving automorphisms of $\mathcal{G}_T$ just

need to preserve the positive dominant eigendirection, $\mathrm{aut}(\mathcal{G}_T, \mathcal{G}_T^+, \widehat{T})$ is a subgroup of index 2 in $\mathrm{aut}(\mathcal{G}_T, \widehat{T})$. The conjugacy of T and U takes place over $\mathbb{Z}[1/3]$, so $\mathrm{aut}(\mathcal{G}_U, \widehat{U}) \cong \mathrm{aut}(\mathcal{G}_T, \widehat{T})$. Hence $\mathrm{aut}(\widehat{T}) \cong \mathbb{Z}^3 \oplus (\mathbb{Z}/2\mathbb{Z})^2$.

On the other hand, U^2 has a repeated eigenvalue of 3^2, and any automorphism of the corresponding subgroup of $\mathcal{G}_U$ extends to one of $\mathcal{G}_U$. Thus

$$\mathrm{aut}(\mathcal{G}_U, \widehat{U}^2) \cong GL(1, \mathbb{Z}[1/3]) \oplus GL(2, \mathbb{Z}[1/3]).$$

As before, we conclude $\mathrm{aut}(\widehat{T}^2) \cong \mathbb{Z} \oplus GL(2, \mathbb{Z}[1/3])$, which is larger than $\mathrm{aut}(\widehat{T})$. □

We remark that by using elementary matrix operations, one can show that $GL(2, \mathbb{Z}[1/3])$, and hence $\mathrm{aut}(\widehat{T}^2)$, is a finitely generated nonabelian group.

Assembling the pieces, we now state the main result of this section.

THEOREM 6.8. *Suppose the nonzero eigenvalues of T are simple. If no ratio of distinct eigenvalues is a root of unity, then for all sufficiently large n we have that $\delta\colon \mathrm{aut}(\sigma_T^n) \to \mathrm{aut}(\widehat{T}^n) = \mathrm{aut}(\widehat{T})$ is surjective. If some ratios of eigenvalues are roots of unity, the conclusion still holds for infinitely many n.*

PROOF. Let us first suppose no ratio of distinct eigenvalues is a root of unity. By Lemma 6.6, we have $\mathrm{aut}(\widehat{T}^n) = \mathrm{aut}(\widehat{T})$. By Proposition 6.5, $\mathrm{aut}(\widehat{T})$ is finitely generated. Using Proposition 6.3, for all sufficiently large n, each generator is in $\delta(\mathrm{aut}(\sigma_T^n))$, implying the conclusion.

If some ratios of distinct eigenvalues are roots of unity, the preceding argument still holds for all n relatively prime to the orders of these roots of unity. □

We conclude with an example to show that finite generation of $\mathrm{aut}(\widehat{T})$ does not generally hold.

EXAMPLE 6.9. *An aperiodic matrix T with $\mathrm{aut}(\widehat{T})$ not finitely generated.*

We are grateful to Gopal Prasad for the idea behind the following. Let

$$T = \begin{bmatrix} 5 & 2 & 2 \\ 4 & 4 & 1 \\ 0 & 3 & 6 \end{bmatrix}.$$

Then T is conjugate over $\mathbb{Z}[1/3]$ to

$$U = V^{-1}TV = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 3 & 3 \end{bmatrix}, \qquad \text{where } V = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix}.$$

As in the argument for Example 6.7, $\mathrm{aut}(\mathcal{G}_U) \cong GL(1, \mathbb{Z}[1/3]) \times H$, where H is the group of automorphisms of $\mathbb{Z}[1/3]^2$ commuting with $W = \left[\begin{smallmatrix} 3 & 0 \\ 3 & 3 \end{smallmatrix}\right]$. By analogy with the KAN decomposition of Lie groups, it is easy to see that $H \cong K \times A \times N$, where $K = \{\pm 1\}$, $A = \{3^n I\colon n \in \mathbb{Z}\}$, and

$$N = \left\{ \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} : \alpha \in \mathbb{Z}[1/3] \right\}.$$

Now N is isomorphic to the additive group $\mathbb{Z}[1/3]$, which is not finitely generated. Thus H is not finitely generated, so $\mathrm{aut}(\widehat{T}) \cong \mathbb{Z} \times H$ is also not finitely generated. □

The following question appears to us basic to understanding the structure of $\mathrm{aut}(\sigma_T)$.

QUESTION 6.10. *If T is aperiodic, is the kernel of the dimension representation of* $\mathrm{aut}(\sigma_T)$ *generated by elements of finite order?*

This question is especially pertinent to understanding the action of $\mathrm{aut}(\sigma_T)$ on the periodic points of σ_T (see §7 and [**BK**]). It is a natural generalization of an older one, which to the best of our knowledge was first explicitly conjectured by F. Rhodes in correspondence with G. Hedlund.

CONJECTURE 6.11 (F. RHODES). *The automorphism group of the 2-shift is generated by the 2-shift and involutions.*

7. Induced action on periodic points. Each automorphism of σ_T maps a periodic point to another with the same period. This action of G on periodic points was studied by Boyle and Krieger [**BK**]. We begin by discussing a fundamental question about this action, and give a partial answer in Theorem 7.2. We then discuss several representations of G provided via periodic points, and conclude with remarks about some related topologies on G.

Let $P_n = P_n(\sigma_T)$ denote the fixed points of σ_T^n, and $Q_n = Q_n(\sigma_T)$ be those points with least σ_T-period n. Since Q_n is σ_T-invariant, we can define $\mathrm{aut}(Q_n, \sigma_T)$ to be the group of bijections of Q_n commuting with σ_T. Let $F_n = \bigcup_{j=1}^{n} Q_j$.

QUESTION 7.1. *When is an automorphism in* $\mathrm{aut}(F_n, \sigma_T)$ *the restriction of one in* $\mathrm{aut}(X_T, \sigma_T)$?

This question is the natural generalization of a long-standing problem of R. F. Williams, namely when can fixed points be switched by an automorphism. The latter problem, posed originally as a potential refinement of shift equivalence, has withstood serious attacks for a number of years. Example 7.3 below provides a concrete case of a shift of finite type having two fixed points which no composition of finite-order elements can switch. It cannot be ruled out at present that the answer to Question 7.1 is "always".

THEOREM 7.2. *Let σ_T be a mixing shift of finite type. There is an $n_0(T)$ so that if $n \geq n_0$ and $x, y \in Q_n(\sigma_T)$ have disjoint orbits, there is a composition φ of involutions in* $\mathrm{aut}(\sigma_T)$ *so that $\varphi x = y$, $\varphi y = x$, and φ fixes all points whose orbit has length $\leq n$ and does not contain x or y.*

PROOF. The proof elaborates ideas from [**BK**], where this is proved for full shifts with $n_0 = 1$. Let us first describe some convenient notation. Shorten σ_T to σ, X_T to X, $\mathcal{B}_n(X_T)$ to $\mathcal{B}_n$, and $\mathcal{B}(X_T)$ to $\mathcal{B}$. If $A = a_0 \cdots a_{n-1} \in \mathcal{B}_n$, let $A[i,j]$ denote $a_i \cdots a_j$. If $A^2 \in \mathcal{B}_{2n}$, let A^∞ denote the point $x \in X$ with $x_i = A[i \bmod n]$. If $0 \leq i < n$, let $\sigma^i(A)$ be the cyclic permutation $A[i, n-1]A[0, i-1]$ of A. Then $\sigma^i(A^\infty) = (\sigma^i A)^\infty$.

First suppose A and B are distinct blocks in $\mathcal{B}_n$ with $A^2B^2A^2 \in \mathcal{B}$. Although $(AB)^\infty$ need not have least period $2n$, we claim there are cyclic permutations $\sigma^i(A)$, $\sigma^i(B)$ so that $\left(\sigma^i(A)\sigma^i(B)\right)^\infty \in Q_{2n}$. To prove this, suppose $(AB)^\infty$ has least period $2n/m < 2n$, so $AB = C^m$ with $m \geq 3$ and odd since $A \neq B$. Thus $|C|$ is even, so write $C = DE$ with $|D| = |E|$ and $D \neq E$. Thus $A = D(ED)^q$ and $B = E(DE)^q$ with $q = \frac{1}{2}(m-1)$. Let $A_1 = \sigma^{|D|}(A)$ and $B_1 = \sigma^{|D|}(B)$, and put $F = A_1B_1 = (ED)^{q-1}EDDDE(DE)^{q-1}E$. Suppose F^∞ has period $k < 2n$. Since

$D \neq E$, we conclude $k < n$, hence

$$k \leq \frac{2n}{3} = \frac{4q+2}{3}|D| \leq 2q|D|.$$

Now choose $j \geq 1$ such that $|D| \leq jk \leq 2q|D|$, and translate the central DDD block in F by jk to the left. It follows that this block contains either DE or ED as a subblock, implying $D = E$, a contradiction. Hence $F^\infty \in Q_{2n}$, verifying our claim. Note that the cyclic permutations A_1 and B_1 retain the property that $A_1^2B_1^2A_1^2 \in \mathcal{B}$, so A_1^∞, B_1^∞, and $(A_1B_1)^\infty$ are allowed points in X.

If $x, y \in Q_n$, say that $\varphi \in G$ *switches* x and y if $\varphi x = y$, $\varphi y = x$, and φ fixes all points whose σ-orbit has length $\leq n$ and does not contain x or y. We shall write in this case $\varphi: x \leftrightarrow y$. Suppose now that $A, B \in \mathcal{B}_n$ with $A^2, B^2 \in \mathcal{B}$ and $A^\infty, B^\infty \in Q_n$ defining disjoint orbits with $A[i] = B[i]$ for some $0 \leq i < n$. We will construct an involution $\varphi_{AB}: A^\infty \leftrightarrow B^\infty$. First replace A by $\sigma^i A$ and B by $\sigma^i B$, so now $A^2B^2A^2 \in \mathcal{B}$. By the above, we can further replace A by $\sigma^j A$ and B by $\sigma^j B$ so that still $A^2B^2A^2 \in \mathcal{B}$ and also $(AB)^\infty \in Q_{2n}$. To define φ_{AB}, let a *frame* be a word in $\{A, B\}^5$. If $z[i-2n, i+3n-1]$ is a frame, define

$$\varphi_{AB}(z)[i, i+n-1] = \begin{cases} B & \text{if } x[i, i+n-1] = A, \\ A & \text{if } x[i, i+n-1] = B, \end{cases}$$

and require φ_{AB} to have no other effect. To see that φ_{AB} is well-defined, suppose the contrary. Then there are two frames C and D with $C[-2n, 3n-1-k] = D[-2n+k, 3n-1]$ for some k with $0 < k < n$. Since $A^\infty \in Q_n$, it follows that A occurs in AA only in the initial and terminal halves, and also B does not occur in AA. Thus C and D have the form $ABABA$ or $BABAB$. The overlapping of C and D force $(AB)^\infty$ to have period $< 2n$, contradicting $(AB)^\infty \in Q_{2n}$. Thus φ_{AB} is well-defined. Clearly it preserves frames. It follows that φ_{AB} is an involution in G, and obviously it switches A^∞ with B^∞.

To complete the proof, let $l > 0$ be a transition length for σ_T, so that $(T^l)_{ij} > 0$ for all $i, j \in \mathcal{L}$. There is an $n_0 = n_0(T)$ so that for all $a, b \in \mathcal{L}$ and $n \geq n_0$, there are at least $3n$ blocks D in $\mathcal{B}_n$ satisfying $D[0] = a$, $D[l] = b$, and $D^\infty \in Q_n$. This follows by counting, since the number of n-blocks satisfying the first two conditions is at least $\kappa\lambda_T^n$ for suitable $\kappa > 0$, while the number of n-blocks D with D^∞ having period $< n$ is no more than $\sum_{d|n, d<n} |\mathcal{B}_d| = O(\lambda_T^{n/2})$. Now fix $n \geq n_0(T)$, and suppose $x, y \in Q_n$ have distinct orbits. Put $A = x[0, n-1]$ and $B = y[0, n-1]$. By the choice of n_0, there is $C \in \mathcal{B}_n$ with $C[0] = A[0]$, $C[l] = B[l]$, $C^\infty \in Q_n$, and the orbit of C^∞ missing x and y. By the above, $\varphi_{AC}: A^\infty \leftrightarrow C^\infty$ and $\varphi_{BC}: B^\infty \leftrightarrow C^\infty$. Then $\varphi = \varphi_{AC}\varphi_{BC}\varphi_{AC}: A^\infty \leftrightarrow B^\infty$, and no other orbits of length $\leq n$ are affected. □

Theorem 7.2, with a smaller estimate for $n_0(T)$, has been proved independently by M. Nasu [N].

The following example shows it is possible to have periodic orbits that cannot be switched by products of finite-order automorphisms.

EXAMPLE 7.3. *A mixing σ_T with exactly two fixed points that are also fixed by every finite-order automorphism of σ_T.*

Let

$$T = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 3 & 0 & 1 \end{bmatrix},$$

and $Q_1(\sigma_T) = \{x, y\}$, where $x = 0^\infty$ and $y = 2^\infty$. Suppose $\psi \in G$ with $\psi^k = I$ and $\psi\colon x \leftrightarrow y$. Now $(\psi\sigma_T)^k = \sigma_T^k$, so $\psi\sigma_T$ is a root of a mixing shift of finite type, hence itself is a mixing shift of finite type ([**L1**, Theorem 8] or [**BK**, Lemma 2.5]). Since the characteristic polynomial of T is irreducible over $\mathbb{Q}$, it follows that σ_T and $\psi\sigma_T$ have equal ζ-functions. But $x \in Q_2(\psi\sigma_T)$, while $Q_2(\sigma_T)$ is empty. This contradiction shows such a ψ cannot exist. Furthermore, σ_T has no roots. For suppose $\psi^k = \sigma_T$. Then as before ψ is a shift of finite type, say $\psi = \sigma_U$. But then the product of the nonzero eigenvalues of U is $3^{1/k}$, contradicting integrality of this product. In this example, no combination of marker constructions and roots of the shift can switch x and y, and it remains open whether this is possible. □

Recall that $P_n = P_n(\sigma_T)$ denotes the fixed points of σ_T^n, and that $Q_n = Q_n(\sigma_T)$ is the set of points with least σ_T-period n. Clearly each $\varphi \in G$ restricts to a permutation $\pi_n(\varphi) = \varphi|_{Q_n} \in \mathrm{aut}(Q_n, \sigma_T)$. If $Q_n = \varnothing$, set $\mathrm{aut}(Q_n, \sigma_T)$ to the trivial group, so then π_n is also trivial. Call

$$\pi(\varphi) = \big(\pi_1(\varphi), \pi_2(\varphi), \dots\big) \in \prod_{n=1}^{\infty} \mathrm{aut}(Q_n, \sigma_T)$$

the *periodic point* representation of $\varphi \in G$. The proof of Theorem 3.1 shows that it is faithful.

PROPOSITION 7.4. *The periodic point representation π of G is faithful.* □

The next two representations are derived from π. Let $\Sigma = \{\sigma_T^n\colon n \in \mathbb{Z}\}$, and $\mathcal{Q}_n = Q_n/\Sigma$ be the set of orbits of length n. If $\varphi \in G$, then $\pi_n(\varphi)$ commutes with $\pi_n(\sigma)$, so induces a permutation $\rho_n(\varphi) \in \mathrm{aut}(\mathcal{Q}_n, \sigma_T)$. Call

$$\rho(\varphi) = \big(\rho_1(\varphi), \rho_2(\varphi), \dots\big) \in \prod_{n=1}^{\infty} \mathrm{aut}(\mathcal{Q}_n, \sigma_T)$$

the *periodic orbit* representation of $\varphi \in G$, where again we use the convention that the symmetric group of the empty set is the trivial group. Clearly $\rho(\sigma^k) = I$, but Σ exhausts the kernel of ρ.

PROPOSITION 7.5. *The periodic orbit representation ρ of G is faithful on G/Σ.*

PROOF. Suppose $\rho_n(\varphi) = I$ for $n \geq 1$. By Theorem 2.5 of [**BK**], it follows that $\varphi \in \Sigma$. □

We next prove that the range of ρ is large, in that its closure in the compact group $\prod_{n=1}^{\infty} \mathrm{aut}(\mathcal{Q}_n, \sigma_T)$ is a subgroup of finite index.

THEOREM 7.6. *The closure of $\rho(G)$ in $\prod_{n=1}^{\infty} \mathrm{aut}(\mathcal{Q}_n, \sigma_T)$ is a subgroup of finite index.*

PROOF. Let $n_0 = n_0(T)$ from Theorem 7.2. We assume without loss that $|\mathcal{Q}_n| \geq 1$ for $n \geq n_0$. Suppose $n \geq n_0$, and let $\theta_j \in \mathrm{aut}(\mathcal{Q}_j, \sigma_T)$ for $1 \leq j \leq n$ with $\theta_j = I$ if $1 \leq j \leq n_0$. We show there is a $\varphi \in G$ with $\rho_j(\varphi) = \theta_j$, $(1 \leq j \leq n)$. It will follow that $\overline{\rho(G)}$ is a finite union of cosets of $\prod_{n=n_0}^{\infty} \mathrm{aut}(\mathcal{Q}_n, \sigma_T)$. Set $\varphi_1 = I$, and suppose inductively that $\varphi_j \in G$ has been found with $\rho_i(\varphi_j) = \theta_j$ for $1 \leq i \leq j$. Since $\rho_{j+1}(\varphi_j)^{-1}\theta_{j+1} \in \mathrm{aut}(\mathcal{Q}_{j+1}, \sigma_T)$ is a product of transpositions, by Theorem 7.2 there is a product ψ_{j+1} of involutions in G with $\rho_i(\psi_{j+1}) = I$ for $1 \leq i \leq j$ and

$\rho_{j+1}(\psi_{j+1}) = \rho_{j+1}(\varphi_j)^{-1}\theta_{j+1}$. Set $\varphi_{j+1} = \varphi_j\psi_{j+1}$ and continue. Then $\varphi = \varphi_n$ has the required properties. □

One implication of this density result is another proof of Ryan's theorem [**Ry2**] on the center of G. This proof was obtained jointly with W. Krieger.

THEOREM 7.7 (RYAN). *The center of G is Σ.*

PROOF. Choose $n_1 \geq n_0(T)$ so that if $n \geq n_1$, then $|\mathcal{Q}_n| \geq 3$. If φ commutes with every element in G, then by Theorem 4.5 its restriction $\rho_n(\varphi)$ to orbits commutes with all of $\mathrm{aut}(\mathcal{Q}_n, \sigma_T)$ for $n \geq n_1$. Hence $\rho_n(\varphi) = I$ for $n \geq n_1$, so by Theorem 2.5 in [**BK**] it follows that $\varphi \in \Sigma$. □

THEOREM 7.8. *The groups G, G/Σ, and $G/[G,G]$ are not finitely generated.*

PROOF. We show each factors onto an arbitrarily large product of two-element groups, which shows each is not finitely generated. By Theorem 4.5 there is an $n_1 \geq n_0(T)$ so that for $n \geq n_1$ we have $|\mathcal{Q}_n| \geq 2$ and that $(\rho_{n_1}, \ldots, \rho_n)$ maps G onto $\prod_{j=n_1}^{n} \mathrm{aut}(\mathcal{Q}_j, \sigma_T)$. Denoting the sign of a permutation ρ by $\operatorname{sgn}\rho \in \{\pm 1\}$, it follows that $(\operatorname{sgn}\rho_{n_1}, \ldots, \operatorname{sgn}\rho_n)$ maps G onto $\{\pm 1\}^{n-n_1}$. This mapping factors through $G \to G/\Sigma$ and $G \to G/[G,G]$, completing the proof. □

We now recall the gyration function g introduced in [**BK**]. To define g, let $\varphi \in G$, and from each orbit $\gamma \in \mathcal{Q}_n$ pick an element $x_\gamma \in \gamma$. Since $\varphi(x_\gamma) \in \varphi(\gamma)$, there is an integer $n(\gamma, \varphi)$ defined modulo n so that $\varphi(x_\gamma) = \sigma^{n(\gamma,\varphi)}x_{\varphi\gamma}$. Put $g(\varphi)(n) = g(\varphi, \sigma)(n) = \sum_{\gamma \in \mathcal{Q}_n} n(\gamma, \varphi)$, where by convention an empty sum is 0. The value of $g(\varphi)(n)$ is shown in [**BK**] to be independent of the choice of the x_γ, and also $g(\varphi\psi)(n) \equiv g(\varphi)(n) + g(\psi)(n) \bmod n$. Call

$$g(\varphi) = \big(g(\varphi)(1), g(\varphi)(2), \ldots\big) \in \prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$$

the *gyration function* of $\varphi \in G$, and call the homomorphism $\varphi \mapsto g(\varphi)$ the *gyration representation.* This function is studied in detail by Boyle and Krieger [**BK**]

The periodic orbit and gyration representations are related. If $\theta \in \mathrm{aut}(Q_n, \sigma_T)$ commutes with $\pi_n(\sigma_T) = \sigma|_{Q_n}$, its gyration number $g(\theta) \in \mathbb{Z}/n\mathbb{Z}$ is defined as above. Also, let $\rho(\theta)$ be the permutation induced by θ on the orbit space $\mathcal{Q}_n = Q_n/\Sigma$. Call a sequence $(\theta_1, \ldots, \theta_n)$ with $\theta_j \in \mathrm{aut}(Q_j, \sigma_T)$ *consistent* if whenever $2^m q \leq N$ with q odd, then

$$g(\theta_{2^m q}) \equiv \begin{cases} 0 & \bmod 2^m q \quad \text{if } \prod_{j=1}^{m-1} \operatorname{sgn}\rho(\theta_{2^j q}) = 1, \\ 2^{m-1}q & \bmod 2^m q \quad \text{if } \prod_{j=1}^{m-1} \operatorname{sgn}\rho(\theta_{2^j q}) = -1. \end{cases}$$

In particular, if $m = 0$ the empty product being 1 means $g(\theta_q) \equiv 0$. If σ_T is a full shift, it is shown in [**BK**] that a necessary and sufficient condition for a sequence $(\theta_1, \ldots, \theta_n)$ with $\theta_j \in \mathrm{aut}(Q_j(\sigma_T), \sigma_T)$ to be the restriction $\theta_j = \pi_j(\varphi)$ of a product φ of involutions is that it be consistent. Necessity was also shown for a broad class of σ_T. One consequence is that for a product φ of involutions, $\operatorname{sgn}\rho(\varphi)$ and $g(\varphi)$ determine each other. This is significant since many marker constructions are involutions. We show a modified form of the sufficiency condition is true for general shifts of finite type.

THEOREM 7.9. *Let σ_T be a mixing shift of finite type, and fix $N > n_0(T)$ with $n_0(T)$ as in Theorem 7.2. If $\theta_j \in \mathrm{aut}(Q_j(\sigma_T), \sigma_T)$ for $1 \le j \le N$ with $\theta_j = I$ if $1 \le j \le n_0$, and if $(\theta_1, \ldots, \theta_N)$ is consistent, then there is a product φ of involutions with $\pi_j(\varphi) = \theta_j$ for $1 \le j \le N$.*

PROOF. We first observe that the involution φ_{AB} in the proof of Theorem 7.2 can be constructed so that $\big(\pi_1(\varphi_{AB}), \ldots, \pi_N(\varphi_{AB})\big)$ is consistent. Let $k = |A|$, and change the definition of "frame" in the construction from $\{A, B\}^5$ to $\{A, B\}^{2N+1}$. The only points of period $\le N$ affected by φ_{AB} have period a multiple of k, and are concatenations of A's and B's. Let σ_2 be the 2-shift on $\{0, 1\}$, and $\tau \in \mathrm{aut}(\sigma_2)$ be the 1-block map exchanging 0 and 1. Mapping $A \to 0$ and $B \to 1$ shows that $\rho_{nk}(\varphi_{AB}) \in \mathrm{aut}(\mathcal{Q}_{nk}(\sigma_T), \sigma_T)$ and $\rho_n(\tau) \in \mathrm{aut}(\mathcal{Q}_n(\sigma_2), \sigma_2)$ have the same sign. Furthermore, $g(\varphi_{AB}, \sigma_T)(2nk) \equiv kg(\tau, \sigma_2)(2n) \bmod 2nk$. Since $\rho_m(\varphi_{AB}) = I$ if $m \not\equiv 0 \bmod k$, and $\big(\pi_1(\tau), \ldots, \pi_N(\tau)\big)$ is consistent from Lemma 3.3 of [**BK**], it follows that $\big(\pi_1(\varphi_{AB}), \ldots, \pi_N(\varphi_{AB})\big)$ is also consistent.

To construct φ, assume inductively that φ_m has been found with $\pi_j(\varphi_m) = \theta_j$ for $1 \le j \le m$ and so that $\big(\pi_1(\varphi_m), \ldots, \pi_N(\varphi_m)\big)$ is consistent. As in the proof of Lemma 3.7 of [**BK**], there is a product ψ_m of involutions of the above type fixing points with period $\le m$ such that $\pi_{m+1}(\psi_{m+1}) = \pi_{m+1}(\varphi_m^{-1}\theta_{m+1})$ if the gyration number of $\varphi_m^{-1}\theta_{m+1}$ vanishes. However, this follows from consistency of the $\theta_j = \pi_j(\varphi_m)$. □

Although G is discrete is the compact-open mapping topology, there are at least two other natural topologies on G making it a nondiscrete topological group. The first is the *periodic point* topology, defined as the weakest topology making the restriction homomorphisms $G \to \mathrm{aut}(Q_n, \sigma_T)$ continuous for all n. In this topology an automorphism is close to the identity it if fixes all points with period less than some large bound. The second is the *profinite* topology on G, which has as basic open sets the cosets of finite-index normal subgroups of G. Clearly the profinite topology refines the periodic point topology, and the proof of Proposition 7.4 shows both are Hausdorff. Do they coincide? The following example shows the answer in general is "no".

EXAMPLE 7.10. *A mixing shift of finite type such that the profinite topology on its automorphism group strictly refines the periodic point topology.*

Let

$$T = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

indexed with states a, b, and c. We will use the dimension group representation of G to define a homomorphism θ from G to $\{\pm 1\}$, and show that for every n there is a $\varphi_n \in G$ fixing all points of period $\le n$ for which $\theta(\varphi_n) = -1$. This will show that the profinite neighborhood $\ker\theta$ of I in G contains no periodic point neighborhood of I.

Now T has characteristic polynomial $\chi_T(t) = (t+1)(t^2 - 2t - 1)$, so -1 is a simple eigenvalue of T, with corresponding eigenvector $v = [1 \quad -1 \quad 0]$. Recall from §6 the dimension group representation $\delta: G \to \mathrm{aut}(\mathcal{G}_T, \mathcal{G}_T^+, \widehat{T})$. Now $\mathbb{Z}v \subset \mathcal{G}_T$, and any element $\varphi \in \mathrm{aut}(\mathcal{G}_T, \mathcal{G}_T^+, \widehat{T})$ must map v to either v or $-v$. Let $\theta(\varphi)$ be the choice

of sign used, so $\theta: G \to \{\pm 1\}$ with $\delta(\varphi)(v) = \theta(\varphi)v$. Since δ is a homomorphism, so is θ.

We now show that θ has the required property. Fix $n \geq 1$. The blocks $M_1 = c^n a$ and $M_2 = c^n b$ have only trivial overlaps. Define $\varphi \in \mathrm{aut}(\sigma_T)$ to switch these blocks, and have no other effect. Clearly φ_n fixes all points with period $\leq n$. We compute the action of $\delta(\varphi_n)$ on v. For $j = a, b, c$ choose the points x^j so that

$$x_i^j = \begin{cases} c & \text{if } i \neq 0, \\ j & \text{if } i = 0. \end{cases}$$

Consider the 0-rays $C_j = x^j(-\infty, 0]^*$ for $j = a, b, c$. Under the correspondence of equivalence classes of beams to classes of eventually nonnegative vectors in $\mathcal{G}_T$ described in §6, each class $[C_j]$ corresponds to the class of the standard unit basis vector e_j. Now φ_n exchanges the 0-rays C_a with C_b, and fixes C_c. Thus

$$\delta(\varphi_n) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and $\delta(\varphi_n)v = -v$, so $\theta(\varphi_n) = -1$. □

8. p-adic aspects of the gyration representation. Let us begin by recalling some notation from the previous section. We let $Q_n = Q_n(\sigma_T)$ denote the points in X_T with least σ_T-period n, and $\mathcal{Q}_n = \mathcal{Q}_n(\sigma_T) = Q_n/\Sigma$ be the orbit space of Q_n under σ_T. For each orbit $\gamma \in \mathcal{Q}_n$ pick $x_\gamma \in \gamma$. If $\varphi \in G$, then $\varphi x_\gamma \in \varphi(\gamma)$, so there is an integer $n(\gamma, \varphi)$ defined modulo n so that $\varphi x_\gamma = \sigma^{n(\gamma,\varphi)} x_{\varphi\gamma}$. The nth *gyration number* of φ is

$$g(\varphi, \sigma_T)(n) \equiv \sum_{\gamma \in \mathcal{Q}_n(\sigma_T)} n(\gamma, \varphi) \in \mathbb{Z}/n\mathbb{Z}.$$

If $Q_n = \emptyset$, this number is defined by convention to be 0. The *gyration representation* of φ is the homomorphism sending $\varphi \in G$ to its gyration function

$$g(\varphi, \sigma_T) = \big(g(\varphi, \sigma_T)(1), g(\varphi, \sigma_T)(2), \dots\big) \in \prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}.$$

If $\sigma_{[r]}$ is the full r-shift, Boyle and Krieger [**BK**, Corollary 2.3] showed that $g(\sigma_{[r]}, \sigma_{[r]})$ has infinite order in $\prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$. It follows that $\sigma_{[r]}$ is not a product of finite-order elements in G. Indeed, $\sigma_{[r]}$ is not even the limit in the periodic point topology of such products [**BK**, Theorem 2.8]. Also, $g(\sigma_{[r]} \times I, \sigma_{[r]} \times \sigma_{[s]})$ has infinite order [**BK**, Proposition 2.4]. Using p-adic analysis, we shall extend these results to matrices T for which the product $\det^{\times} T$ of the nonzero eigenvalues is not ± 1. The crucial fact is that, roughly speaking, for most primes p the numbers $g(\sigma_T, \sigma_T)(p^n)$ converge p-adically to a transcendental limit in the p-adic completion $\mathbb{Q}_p$ of $\mathbb{Q}$.

To see the computational significance of this convergence, consider Table 1 of the gyration function of the 2-shift acting as a self-automorphism, evaluated at powers of 3 and of 5. We have expanded the values to the appropriate base.

The evident stabilization of the digit coefficients for, say, 5^n can be expressed by saying that $g(\sigma_{[2]}, \sigma_{[2]})(5^n)$ converges 5-adically to some $\alpha \in \mathbb{Q}_5$. To prove this

n	$g(\sigma_{[2]},\sigma_{[2]})(3^n)$	$g(\sigma_{[2]},\sigma_{[2]})(5^n)$
1	$2 = 2$	$1 = 1$
2	$2 = 2 + 0\cdot 3$	$1 = 1 + 0\cdot 5$
3	$11 = 2 + 0\cdot 3 + 1\cdot 3^2$	$26 = 1 + 0\cdot 5 + 1\cdot 5^2$
4	$38 = 2 + 0\cdot 3 + 1\cdot 3^2 + 1\cdot 3^3$	$276 = 1 + 0\cdot 5 + 1\cdot 5^2 + 2\cdot 5^3$
5	$38 = 2 + 0\cdot 3 + 1\cdot 3^2 + 1\cdot 3^3$	$2776 = 1 + 0\cdot 5 + 1\cdot 5^2 + 2\cdot 5^3 + 4\cdot 5^4$

TABLE 1. Values of the gyration function for the 2-shift.

convergence, and identify the limit, note that since each orbit contributes 1 to the gyration function, we have

$$g(\sigma_{[2]},\sigma_{[2]})(5^n) \equiv |\mathcal{Q}_{5^n}(\sigma_{[2]})| \equiv \frac{|P_{5^n}(\sigma_{[2]})| - |P_{5^{n-1}}(\sigma_{[2]})|}{5^n}$$
$$\equiv \frac{2^{5^n} - 2^{5^{n-1}}}{5^n} \equiv \left(1 - \frac{1}{5}\right) 2^{5^{n-1}} \left(\frac{2^{5^n - 5^{n-1}} - 1}{5^n - 5^{n-1}}\right).$$

As $n \to \infty$, we have $2^{5^{n-1}} \to \omega(2) \in \mathbb{Q}_5$, a 4th root of unity, the so-called Teichmüller representative of 2. Also, since $5^n - 5^{n-1} \to 0$ in $\mathbb{Q}_5$, the third factor converges to the 5-adic derivative of 2^x at $x = 0$, namely the 5-adic logarithm $\log_5 2$ (see [**Ko**, Chapter 5, §1] for details). Thus $g(\sigma_{[2]},\sigma_{[2]})(5^n)$ converges 5-adically, with limit $\alpha = (1 - \frac{1}{5})\omega(2)\log_5 2$. If we denote $a_0 + a_1 5 + a_2 5^2 + \cdots$ by $.a_0a_1a_2\ldots$, then $\omega(2) = .212134\ldots$, $\frac{4}{5}\log_5 2 = .330333\ldots$, and the product is $\alpha = .101240\ldots$, in agreement with the table entry. Furthermore, $\log_5 2$ is transcendental [**Br**]. Since $\frac{4}{5}\omega(2)$ is algebraic, the limit α is also transcendental. Since $\alpha \bmod 5^n$ is a unit in $\mathbb{Z}/5^n\mathbb{Z}$, it follows that $g(\sigma_{[2]},\sigma_{[2]})(5^n)$ generates $\mathbb{Z}/5^n\mathbb{Z}$ for $n \geq 1$. Hence $\sigma_{[2]}$ could not be a product of finite-order elements in G, for otherwise the orders of $g(\sigma_{[2]},\sigma_{[2]})(n)$ would be uniformly bounded in n. This analysis fails at the prime 2, since $g(\sigma_{[2]},\sigma_{[2]})(2^n) \to 0$ in $\mathbb{Q}_2$. However, 2 is the only exceptional prime.

It is perhaps interesting to note that the p-adic convergence of the gyration function was first discovered by computer experimentation.

The ideas above extend to more general σ_T by using the p-adic eigenvalues of T.

THEOREM 8.1. *Suppose the product of the nonzero eigenvalues of T is not ± 1. Then the gyration representation $g(\sigma_T,\sigma_T)$ has infinite order in $\prod_{n=1}^{\infty}\mathbb{Z}/n\mathbb{Z}$.*

PROOF. Let the characteristic polynomial of T be

$$\chi_T(t) = t^m(t^d + a_1 t^{d-1} + \cdots + a_d),$$

where $a_d \neq \pm 1$ by assumption. Choose an odd prime p relatively prime to a_d. A good account of the basic p-adic analysis used here is contained in [**Ko**]. Let K be the splitting field of χ_T over the p-adic completion $\mathbb{Q}_p$ of $\mathbb{Q}$. The p-adic valuation $|\cdot|_p$ extends uniquely to K. Thus $\chi_T(t)$ factors over K as $t^m\prod_{j=1}^{d}(t - \lambda_j)$. Since a_d is a unit in the ring of integers in K, each $|\lambda_j|_p = 1$. It follows there is an integer $r > 0$ so that $|1 - \lambda_j^r|_p < 1$ for $1 \leq j \leq d$. We will prove that $g(\sigma_{T^r},\sigma_{T^r})(p^n)$

converges p-adically to a nonzero limit, show this forces $g(\sigma_{T^r}, \sigma_{T^r})$ to have infinite order, and conclude from this that $g(\sigma_T, \sigma_T) = g(\sigma_T)$ also has infinite order.

For arbitrary U, we have $|Q_n(\sigma_U)| = \sum_{d|n} \mu(n/d) \operatorname{tr} U^d$, where μ is the Möbius function, and $g(\sigma_U, \sigma_U)(n) = |Q_n(\sigma_U)|/n$. Hence

$$g(\sigma_{T^r}, \sigma_{T^r})(p^n) = \frac{1}{p^n}\left\{\operatorname{tr}(T^r)^{p^n} - \operatorname{tr}(T^r)^{p^{n-1}}\right\}$$
$$= \left(1 - \frac{1}{p}\right) \sum_{j=1}^{d} \lambda_j^{rp^{n-1}} \left(\frac{\lambda_j^{r(p^n - p^{n-1})} - 1}{p^n - p^{n-1}}\right).$$

Since $|1 - \lambda_j^r|_p < 1$, it follows that $\lambda_j^{rp^n} \to 1$ in K. The second factor in the sum converges to the p-adic derivative of λ_j^{rx} with respect to x at $x = 0$, which is $r \log_p \lambda_j$. Here $\log_p y$ is the p-adic logarithm defined for $y \in K$ with $|y - 1|_p < 1$ by the convergent series $\log_p y = \sum_{n=1}^{\infty} (-1)^{n+1}(y-1)^n/n$. Thus as $n \to \infty$,

$$g(\sigma_{T^r}, \sigma_{T^r})(p^n) \to \left(1 - \frac{1}{p}\right) \log_p a_d^r \neq 0,$$

the nonvanishing following since $a_d \neq \pm 1$. This means that as n increases, the p-adic expansion of $g(\sigma_{T^r}, \sigma_{T^r})(p^n)$ has low order coefficients that stabilize to nonzero values. In particular, if

$$\left|\left(1 - \frac{1}{p}\right) \log_p a_d^r\right|_p > p^{-k},$$

then $p^{n-k} g(\sigma_{T^r}, \sigma_{T^r})(p^n) \not\equiv 0 \bmod p^n$ for n sufficiently large. This clearly implies that $g(\sigma_{T^r}, \sigma_{T^r})$ has infinite order. We conclude by showing this forces $g(\sigma_T, \sigma_T)$ to also have infinite order. First note that $g(\sigma_{T^r}, \sigma_{T^r}) = g(\sigma_T^r, \sigma_T^r) = r g(\sigma_T, \sigma_T^r)$. The calculation of Proposition 1.6 of [**BK**] shows that if $a_j = a_j(r, n) \in \mathbb{Z}$ is defined so $r a_j \equiv j \bmod jn$ when $j = (r, jn)$ and 0 otherwise, then

$$g(\varphi, \sigma_T^r)(n) \equiv \sum_{j|r} a_j g(\varphi, \sigma_T)(jn) \mod n.$$

Apply this to $\varphi = \sigma_T$ in the above to express $g(\sigma_{T^r}, \sigma_{T^r})(n)$ as an integral combination of the $g(\sigma_T, \sigma_T)(jn)$ for $j \mid r$. Thus if $g(\sigma_T, \sigma_T)$ had finite order, so would $g(\sigma_{T^r}, \sigma_{T^r})$. This contradiction establishes the theorem. □

REMARKS. 1. If λ is a unit in K, the splitting field of χ_T over $\mathbb{Q}_p$, then λ^{p^n} converges to a $(p^e - 1)$st root of unity $\omega(\lambda)$ in K, where e is the ramification index of K over $\mathbb{Q}_p$. The above shows that we always have in $\mathbb{Q}_p$ that

$$\lim_{n\to\infty} g(\sigma_T, \sigma_T)(p^n) = \left(1 - \frac{1}{p}\right) \log_p \left(\prod_{j=1}^{d} \lambda_j^{\omega(\lambda_j)}\right). \tag{8–1}$$

However, we cannot conclude from this that the limit is nonzero. By taking rth powers, we force $\omega(\lambda_j^r) = 1$, so the product in (8–1) is $a_d^r \neq 1$, and the limit does not vanish. The general question of whether (8–1) is nonzero hinges on how the λ_j and the corresponding root of unity $\omega(\lambda_j)$ interact. For example, if the λ_j are multiplicatively independent, then the p-adic versions of Baker's transcendence results [**Br**] show that (8–1) is not zero.

2. If the product of the nonzero eigenvalues of T is ± 1, all that is needed to prove the theorem is to find one prime p for which (8–1) does not vanish. Since this is the "typical" case, finding such a p can usually be done by hand. For example, if $T = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ we can take $p = 3$ and simply check that the product in (8–1) is not ± 1. Indeed, all but finitely many primes will work in this case.

3. If $\det^{\times} T \neq \pm 1$, the roots of unity $\omega(\lambda_j)$ in (8–1) are merely a nuisance, to be eliminated by passing to rth powers. However, when $\det^{\times} T = \pm 1$ it is essential that at least some of the $\omega(\lambda_j)$ not be 1 to obtain a nonzero limit in (8–1).

The proof of Theorem 8.1 can be modified to obtain the following.

THEOREM 8.2. *Suppose the product of the nonzero eigenvalues of T is not ± 1. Then $g(\sigma_T \times I, \sigma_T \times \sigma_U)$ has infinite order.*

PROOF. Choose a prime p relatively prime to $\det^{\times} T$ and to $\det^{\times} U$. We first compute the gyration function $g(\sigma_T \times I, \sigma_T \times \sigma_U)(p^n)$. Now

$$Q_{p^n}(\sigma_T \times \sigma_U) = \bigcup_{k=0}^{n} \{Q_{p^n}(\sigma_T) \times Q_{p^k}(\sigma_U) \cup Q_{p^k}(\sigma_T) \times Q_{p^n}(\sigma_U)\}.$$

If $(x, y) \in Q_{p^k}(\sigma_T) \times Q_{p^n}(\sigma_U)$, then the smallest $q \geq 0$ for which $(\sigma_T \times I)^q(x, y)$ is in the $\sigma_T \times \sigma_U$-orbit of (x, y) is $q = p^k$, and this element is (x, y) itself. It follows that such orbits contribute 0 mod p^n to the gyration function. Let $0 \leq k \leq n-1$. For each orbit $\gamma \in \mathcal{Q}_{p^n}(\sigma_T)$ pick $x_\gamma \in \gamma$. Then

$$\bigcup_{\gamma \in \mathcal{Q}_{p^n}(\sigma_T)} \{x_\gamma\} \times Q_{p^k}(\sigma_U)$$

is a complete set of orbit representatives for $Q_{p^n}(\sigma_T) \times Q_{p^k}(\sigma_U)$. Since

$$(\sigma_T \times I)(x, y) = (\sigma_T \times \sigma_U)(x, \sigma_U^{-1} y),$$

each representative contributes 1 to the gyration function. Hence

$$\begin{aligned} g(\sigma_T \times I, \sigma_T \times \sigma_U)(p^n) &\equiv |\mathcal{Q}_{p^n}(\sigma_T)| \sum_{k=0}^{n-1} |Q_{p^k}(\sigma_T)| \\ &\equiv \frac{1}{p^n} \left(\operatorname{tr} T^{p^n} - \operatorname{tr} T^{p^{n-1}}\right) \left(\operatorname{tr} U^{p^{n-1}}\right). \end{aligned}$$

Suppose K is a splitting field for $\chi_T(t)$ and $\chi_U(t)$ over $\mathbb{Q}_p$. Let T have nonzero p-adic eigenvalues $\lambda_1, \ldots, \lambda_d$, and let those of U be $\mu_1, \ldots, \mu_e$. Then by the above

$$g(\sigma_T \times I, \sigma_T \times \sigma_U)(p^n) \equiv \sum_{i,j} \left(\frac{\lambda_i^{p^n} - \lambda_i^{p^{n-1}}}{p^n} \right) \mu_j^{p^{n-1}}.$$

The p-adic convergence of this expression follows as before. To obtain a nonzero limit, pass to rth powers to obtain $\lambda_i^{rp^n} \to 1$, $\mu_j^{rp^n} \to 1$, and argue as in the proof of Theorem 8.1. □

Theorem 2.9 of [**BK**] implies that if $\det^{\times} T \neq \pm 1$, then σ_T is not the product of finite-order elements in G. This is also a consequence of Theorem 8.1. The following, which generalizes Theorem 2.8 of [**BK**], shows σ_T is not even a limit of such products.

PROPOSITION 8.3. *Suppose the product of the nonzero eigenvalues of σ_T is not ± 1. Then σ_T is not the limit, in the periodic point topology on G, of products of finite-order elements.*

PROOF. The periodic point topology was defined in §7 to be the coarsest making the restriction maps $G \to \mathrm{aut}(Q_n, \sigma_T)$ continuous. From the proof of Theorem 8.1 we can conclude there is a prime p and an integer m so that for all $n \geq m$ we have

$$g(\sigma_T, \sigma_T)(p^n) \not\equiv 0 \mod p^m.$$

The result now follows exactly as in the proof of Theorem 2.8 of [**BK**] by using Proposition 2.7 there. □

REMARK 4. All of the above results will hold even if $\det^{\times} T = \pm 1$, provided the prime p used in the proofs can be chosen so that (8–1) is nonzero. As indicated in Remark 2, in specific case this can usually be done *ad hoc*.

PROBLEM 8.4. *Remove the restriction* $\det^{\times} T \neq \pm 1$ *from the above.*

The proof of Theorem 8.1 shows that $g(\sigma_T, \sigma_T)(p^n)$ converges p-adically for all primes p. It is also easy to see that if φ has finite order, then $g(\varphi, \sigma_T)(p^n) \to 0$ in $\mathbb{Q}_p$. Thus for automorphisms φ constructed by known methods, $g(\varphi, \sigma_T)(p^n)$ always converges p-adically.

QUESTION 8.5. *Does the gyration function $g(\varphi, \sigma_T)(p^n)$ converge p-adically for every $\varphi \in \mathrm{aut}(\sigma_T)$?*

9. Compact invariant sets. In §7 we investigated $G = \mathrm{aut}(\sigma_T)$ by studying its action on the compact G-invariant sets $Q_n(\sigma_T)$ of points with least σ_T-period n. Are there other compact G-invariant subsets of X_T? In this section we show that, modulo possibly a few orbits of low period, every compact G-invariant subset is a finite union of various $Q_n(\sigma_T)$ or all of X_T. We also discuss the implications for the action of G on the compact space of subshifts equipped with the Hausdorff metric.

Several constructions and proofs to follow are made clearer by the process of "passing to a higher order block presentation", described as follows. Recall from §1 that if T is $r \times r$, then the graph of T has r *states* or nodes, and $\sum_{i,j=1}^{r} T_{ij}$ *symbols* or edges. Order the set $\mathcal{L}$ of symbols arbitrarily. Let $E(T)$ be the transition matrix for symbols, defined by $E(T)_{xy} = 1$ if the terminal state of x matches the initial state of y, and 0 otherwise. Iterating this procedure yields the transition matrix $E^n(T) = E(E^{n-1}(T))$ for allowed n-blocks of symbols. Specifically, we define the *n-block presentation* of T to be $T^{[n]} = E^{n-1}(T)$, where $E^0(T) = T$. The symbols for $T^{[n]}$ are then the allowed n-blocks of σ_T, i.e. $\mathcal{L}_{T^{[n]}} = \mathcal{B}_n(\sigma_T)$. If $B, C \in \mathcal{L}_{T^{[n]}}$, then $T^{[n]}_{BC} = 1$ if $B[1, n-1] = C[0, n-2]$, and 0 otherwise. Thus the symbols in a point $x \in X_T$ have simply been recoded in $X_{T^{[n]}}$ by using n-blocks of symbols.

There is a natural conjugacy $\beta_n \colon X_T \to X_{T^{[n]}}$ defined by

$$(\beta_n x)_i = x[i-k, i-k+n-1], \qquad \text{where } k = \left[\frac{n}{2}\right].$$

The metric on $X_{T^{[n]}}$ is understood to be that induced from X_T under β_n. With this convention, the maximum of the diameters of the sets $B^* = \{y \in X_{T^{[n]}} \colon y_0 = B\}$ over $B \in \mathcal{L}_{T^{[n]}}$ tends to 0 as $n \to \infty$.

If K is a subshift of X_T, let $\mathcal{B}(K)$ be the collection of all blocks that occur in some point in K. For $x \in X_T$, put $\mathcal{B}(x) = \mathcal{B}(\overline{\Sigma x})$. If $B \in \mathcal{B}_n(X_T)$, put $B^* = \{x \in X_T : x[0, n-1] = B\}$. Here the reader should distinguish B considered as an n-block from X_T from B considered as a symbol from $X_{T^{[n]}}$. The homeomorphism β_n introduces a factor of $\sigma_T^{[n/2]}$ between the two meanings of B^*. In particular,

$$\max\{\operatorname{diam}(B^*) : B \in \mathcal{L}_{T^{[n]}}\} \to 0 \quad \text{as } n \to \infty.$$

Thus if $Y_1, \ldots, Y_k$ are disjoint subshifts of X_T, for all large enough n the sets $\mathcal{L}_{T^{[n]}}(Y_j)$ of symbols from $\mathcal{L}_{T^{[n]}}$ occurring in Y_j are disjoint. Furthermore, if $Z \subset X_T$ is a shift of finite type, for all large enough n we can represent Z as 1-step on the symbols $\mathcal{L}_{T^{[n]}}(Z)$.

For completeness, we prove a standard technical result. A stronger version is contained in [**DGS**, 26.17].

LEMMA 9.1. *Suppose $B \in \mathcal{B}(X_T)$ and that $B \notin \mathcal{B}(x)$. Then there is an infinite mixing shift of finite type $Z \subset X_T$ with $B \in \mathcal{B}(Z)$ and Z disjoint from $\overline{\Sigma x}$.*

PROOF. Since our standing assumption is that σ_T is mixing, there are $C_j \in \mathcal{B}(X_T)$, $j = 0, 1$, with relatively prime lengths, with $BC_jB \in \mathcal{B}(X_T)$, and such that B occurs only as the initial and terminal block in each BC_jB. Form the shift of finite type $Z \subset X_T$ of points of the form $\cdots BC_{i_{-1}}BC_{i_0}BC_{i_1}B \cdots$ with $i_k = 0$ or 1 for $k \in \mathbb{Z}$. Since the C_j have relatively prime lengths, $\sigma_T|_Z$ is mixing with positive entropy. Let $n \geq 2(|C_0| + |C_1| + |B|)$. Every block in $\mathcal{B}_n(Z)$ contains B as a subblock, while B does not occur in x. Thus, by passing to a higher order block presentation, for large enough n, we can realize Z as a 1-step shift of finite type on an alphabet disjoint from the symbols occurring in x. Thus $Z \cap \overline{\Sigma x} = \varnothing$. □

THEOREM 9.2. *The automorphism group orbit Gx of $x \in X_T$ is dense if and only if x is not σ_T-periodic.*

PROOF. Since the period of a point is preserved under an automorphism, a σ_T-periodic point x clearly has Gx finite.

Now assume x is not σ_T-periodic. To show $\overline{Gx} = X_T$, it is enough to prove that for every $B \in \mathcal{B}(X_T)$ there is a $\varphi \in G$ with $B \in \mathcal{B}(\varphi x)$. If $B \in \mathcal{B}(x)$, take $\varphi = I$. Thus we assume $B \notin \mathcal{B}(x)$, and find φ. By Lemma 9.1, there is a mixing shift of finite type Z disjoint from $\overline{\Sigma x}$ with $B \in \mathcal{B}(Z)$.

We first treat the case that a block in $\mathcal{B}(x)$ occurs exactly once in x. By passing to the n-block presentation of X_T, we can assume Z is a 1-step shift of finite type on an alphabet $\mathcal{L}(Z) \subset \mathcal{L}_{T^{[n]}} = \mathcal{B}_n(\sigma_T)$ disjoint from the symbols occurring in x, that B occurs as a subword of $b \in \mathcal{L}(Z)$, and that a symbol $a \in \mathcal{L}_{T^{[n]}}$ occurs in x at a unique index, which by applying a power of $\sigma_{T^{[n]}}$ we can assume is 0. For the rest of this case, all symbols and blocks are from the n-block presentation. Let l be a transition length for $X_{T^{[n]}}$ and for Z. Since both are mixing, there is a block $E = aAbCbA'x[4l, 6l]$ such that $|A| < l$, $|A'| < l$, a does not occur in A and occurs in A' at most once, $C \in \mathcal{B}(Z)$, and $|E| = 6l + 1$. We claim that E and $F = x[0, 6l]$ overlap only trivially in the sense of §2. Since a occurs only as the initial symbol in F while E contains symbols from Z never used in F, it follows E does not overlap F, and that F only overlaps itself trivially. The only nontrivial overlap of E with itself could occur with the occurrence of a in A', but this is ruled

out since this would force $x[4l, 6l]$ and C to have symbols in common. Thus, as in §2, the map $\varphi \in G$ defined by interchanging E and F and having no other effect is a well-defined involution. Since the symbol b from the n-block presentation occurs in φx, it follows that the original block B occurs in φx, completing this case.

We next turn to the recurrent case. Here blocks in x can recur with distressing frequency, making the nonoverlapping condition for markers more difficult to achieve. The reader should keep in mind the Thue-Morse sequence, where every allowed block of length n recurs within $8n$, to better understand the complications below.

Let x, B, and Z retain their meanings, and assume we have already passed to a higher order block presentation with B a subword of $b \in \mathcal{L}(Z)$ and $\mathcal{L}(Z) \cap \mathcal{L}(x) = \varnothing$. Choose symbols $a, c \in \mathcal{L}(x) = \mathcal{B}_1(x)$ so that there are words aAb, $bA'c \in \mathcal{B}(X_T)$, where A and A' use no symbols from $\mathcal{L}(x)$, and so that some word aFc occurs in x. We claim we can assume there are arbitrarily long words of the form aFc in $\mathcal{B}(x)$. For if not, the longest such word could occur only once, and the previous case applies. Let l be a transition length for X_T and for Z. Adjusting x by a power of σ_T if necessary, there is a $k \geq 4l$ with $x_0 = a$ and $x_k = c$, and a word $C \in \mathcal{B}(Z)$ so that $D = aAbCbA'c \in \mathcal{B}(X)$ with $|D| = k + 1$. Note that the only symbols in D that are also in $\mathcal{L}(x)$ are the initial and terminal symbols. Let $E = x[0, k]$. Since x is not σ_T-periodic, there is a $p > k$ so that $x[-p, p-j] \neq x[-p+j, p]$ for $0 < j < k$. Define $\xi\colon \mathcal{B}_{2p+1}(X) \to \mathcal{B}_{2p+1}(X)$ by replacing each occurrence of D by E; if a block begins with a terminal segment of D, replace it with the corresponding terminal segment of E, and similarly at the other end. Since D cannot overlap itself except possibly in one symbol, ξ is well-defined. Note that $\xi(x[-p, p]) = x[-p, p]$.

Let $\mathcal{M} = \xi^{-1}(x[-p, p]) \cap \mathcal{B}_p(X)\{D, E\}\mathcal{B}_{p-k}(X)$. Then $\mathcal{M}$ has the nonoverlapping property that if $M, M' \in \mathcal{M}$, then $M[-p, p-j] \neq M'[-p+j, p]$ for $0 < j < k$, since otherwise applying ξ would contradict the choice of p. Define $\varphi \in G$ as follows. If $y[-p, p] \in \mathcal{M}$, then put

$$(\varphi y)[0, k] = \begin{cases} E & \text{if } y[0, k] = D, \\ D & \text{if } y[0, k] = E. \end{cases}$$

Declare φ to have no other action. Because of the nonoverlapping property of the words in $\mathcal{M}$ mentioned above, a symbol in y can be affected at most once by φ, so φ is well-defined. If $y[-p, p] \in \mathcal{M}$, then $(\varphi y)[-p, p] \in \mathcal{M}$, proving that φ is an involution. Finally, since the symbol b from the higher order presentation occurs in D, which is a subblock of $(\varphi x)[-p, p]$, we see that the original block $B \in \mathcal{B}(x)$, completing the proof. □

THEOREM 9.3. *Let σ_T be a mixing shift of finite type, and define $n_0(T)$ as in Theorem 7.2. If K is a compact G-invariant set, then either $K = X_T$, or K is a finite union of $Q_{n_j}(\sigma_T)$ with $n_j \geq n_0(T)$ together with possibly a finite number of orbits with periods $< n_0(T)$.*

PROOF. If K contains a nonperiodic point, then $K = X_T$ by Theorem 9.2. If $x \in K \cap Q_n(\sigma_T)$ for some $n \geq n_0(T)$, then by Proposition 7.3 we have $Q_n(\sigma_T) \subset K$. If $n_j \to \infty$, then $\bigcup_{j=1}^{\infty} Q_{n_j}(\sigma_T)$ is dense, so either K contains only a finite number of the sets $Q_n(\sigma_T)$ for $n \geq n_0(T)$, or $K = X_T$. □

REMARKS 1. Until the question of switching points with small periods, such as in Example 7.3, is settled, the possibility remains that for some $n < n_0(T)$ the set $Q_n(\sigma_T)$ could have proper G-invariant subsets.

2. An irreducible shift of finite type is a finite tower over a mixing base shift of finite type, so Theorem 9.3 easily extends to this case.

Let $\mathcal{S}$ denote the compact space of subshifts of (X_T, σ_T) equipped with the Hausdorff metric. In this context, two subshifts $Y, Z \in \mathcal{S}$ are close if for a large n the sets $\mathcal{B}_n(Y)$ and $\mathcal{B}_n(Z)$ coincide. This is the topological analogue of the weak topology on σ_T-invariant measures used in the next section. From this description it is evident that G acts continuously on $\mathcal{S}$. Theorem 9.3 identifies all but a finite number of the fixed points of this action. The result also shows that if $Y \in \mathcal{S}$ is infinite, then X_T is in the G-orbit closure of Y in $\mathcal{S}$. We now generalize this argument.

PROPOSITION 9.4. *If $Y \subset Z \subset X_T$ are infinite subshifts with Z mixing, then Z is in the G-orbit closure of Y in the Hausdorff metric.*

PROOF. Fix an $n \geq 1$, and enumerate the blocks of $\mathcal{B}_n(Z)$ as $B_1, \ldots, B_k$. Since $\sigma_T|_Z$ is mixing, there is $B \in \mathcal{B}(Z)$ which contains each B_j. Pick $y \in Y$ with infinite orbit. Let $Z' \subset X_T$ be the shift of finite type so that $\mathcal{B}_n(Z') = \mathcal{B}_n(Z)$. Then Z' is mixing since Z is. The construction of φ with $B \in \mathcal{B}(\varphi y)$ in the proof of Theorem 9.2 can be carried out with Z' as the ambient space, so $\varphi Y \subset Z'$. Since each B_j occurs in B, we can conclude that $\mathcal{B}_n(\varphi Y) = \mathcal{B}_n(Z') = \mathcal{B}_n(Z)$. Thus every neighborhood of Z in $\mathcal{S}$ contains an image φY for some $\varphi \in G$, completing the proof. □

It is more difficult to determine when a subshift Z in general position is in the G-orbit closure of Y. Two necessary conditions follow.

PROPOSITION 9.5. *If Y and Z are subshifts of X_T and Z is in the G-orbit closure of Y, then*

(1) $h(\sigma_T|_Z) \geq h(\sigma_T|_Y)$,
(2) $|Q_n(\sigma_T|_Z)| \geq |Q_n(\sigma_T|_Y)|$ *for* $n \geq 1$.

PROOF. Suppose $\varphi_k \in G$ with $\varphi_k Y \to Z$. Since topological entropy is upper semicontinuous on $\mathcal{S}$, and $h(\sigma_T|_{\varphi_k Y}) = h(\sigma_T|_Y)$, property (1) holds. Also, for fixed n we must have for large enough k that $\varphi_k[Q_n(\sigma_T|_Y)] \subset Q_n(\sigma_T|_Z)$, from which (2) follows. □

PROBLEM 9.6. *Determine necessary and sufficient conditions on $Y, Z \in \mathcal{S}$ for Z to be in the G-orbit closure of Y.*

Of course, Example 7.3 shows that this problem is not solved even for subshifts with a finite number of points.

10. Orbits of measures. Let $P(X_T)$ denote the compact convex set of σ_T-invariant probability measures on X_T, and $M(X_T)$ be the σ_T-invariant nonnegative measures, each equipped with the weak topology. The group $G = \mathrm{aut}(\sigma_T)$ acts naturally on both spaces, and the unique measure μ_T of maximal entropy is invariant under every $\varphi \in G$ [**CP**]. Are there other continuous G-invariant measures? We show there are none, and obtain a complete characterization of the G-orbit closure in $P(X_T)$ of a continuous probability measure.

To state our result, let $h(\mu)$ denote the entropy of σ_T with respect to $\mu \in M(X_T)$. If $\mathcal{P}_0$ is the partition of X_T by the 0th coordinate, then

$$h(\mu) = \lim_{n\to\infty} \frac{1}{n} \sum_{A \in \bigvee_{j=0}^{n-1} \sigma_T^{-j} \mathcal{P}_0} -\mu(A) \log \mu(A).$$

Observe that this formula applies even to nonnegative measures with total mass different from 1, and that if $r, s \geq 0$ and $\mu, \nu \in M(X_T)$, then [**W**, Theorem 8.1] $h(r\mu + s\nu) = rh(\mu) + sh(\nu)$. We shall use a version of the ergodic decomposition suited to our needs. For $\mu \in P(X_T)$ there is a measurable function $t \mapsto \mu^t$ from $[0,1]$ to $P(X_T)$ so that each μ^t is ergodic under σ_T, if $E \subset X_T$ is Borel then $\mu^t(E)$ is measurable in t with $\mu(E) = \int_0^1 \mu^t(E)\, dt$, and $h(\mu^t)$ is decreasing in t. Note that the μ^t need not be distinct. Indeed, if μ is already ergodic, then $\mu^t = \mu$ for a.e. t. All but the entropy statement follows from the standard ergodic decomposition [**DGS**, Chapter 13]; $h(\mu^t)$ can be arranged to be decreasing by a measurable rearrangement of $[0,1]$. A limiting form of the linearity of h mentioned above shows that if $E \subset [0,1]$, then $h(\int_E \mu^t\, dt) = \int_E h(\mu^t)\, dt$.

Define the cumulative entropy $H_\mu(t) = \int_0^t h(\mu^s)\, ds$ for $0 \leq t \leq 1$. This function can be defined without reference to the ergodic decomposition by

$$H_\mu(t) = \sup\{h(\nu) \colon \nu \in M(X_T),\ \nu \leq \mu,\ \nu(X_T) = t\}.$$

For by ergodicity of the μ^t, if $\nu \leq \mu$ with $\nu(X_T) = t$, then $\nu = \int_0^1 \rho(s)\mu^s\, ds$ with $\int_0^1 \rho(s)\, ds = t$. Thus $h(\nu) = \int_0^1 \rho(s) h(\mu^s)\, ds$ which, by monotonicity of $h(\mu^t)$, has maximum value $H_\mu(t)$ when ρ is the indicator function of $[0,t]$. Our characterization of orbit measures is the following.

THEOREM 10.1. *Let σ_T be a mixing shift of finite type on X_T, let $G = \mathrm{aut}(\sigma_T)$, and suppose $\mu \in P(X_T)$ is continuous. Then ν is in the G-orbit closure of μ iff $H_\mu \leq H_\nu$ on $[0,1]$.*

COROLLARY 10.2. *The measure μ_T of maximal entropy for σ_T is the only G-invariant continuous probability measure on X_T.*

PROOF OF THE COROLLARY. Suppose $\mu \in P(X_T)$ is continuous and G-invariant. Then $h(\mu^t) \leq h(\mu_T^t) = h(\sigma_T)$ for all t, so $H_\mu \leq H_{\mu_T}$. Thus $\mu_T \in \overline{G\mu} = \{\mu\}$. □

Note that if μ and ν are ergodic, the condition of Theorem 10.1 reduces to $h(\mu) \leq h(\nu)$.

We will first prove the necessity of the condition $H_\mu \leq H_\nu$ by using upper semicontinuity of entropy. The main work is to prove its sufficiency. We first find in Lemma 10.3 a partition $0 = t_0 < t_1 < \cdots < t_K = 1$ so that if $J_k = [t_{k-1}, t_k)$, $\alpha_k = |J_k|$, and $\mu_k = \alpha_k^{-1} \int_{J_k} \mu^t\, dt$, then μ is a convex combination $\sum_{k=1}^K \alpha_k \mu_k$, where the μ_k are mutually singular and $h(\mu_k^t)$ is almost constant. We then use the Convex Approximation Lemma 10.4 to perturb ν to $\tilde{\nu} = \sum_{k=1}^K \alpha_k \tilde{\nu}_k$, so that the $\tilde{\nu}_k$ are mixing Markov measures with disjoint supports and $h(\tilde{\nu}_k) > h(\mu_k)$. Using a Rohlin stack argument in Lemma 10.5 based on a marker, and assuming that almost every μ^t has full support, we construct a marker automorphism in Lemma 10.6 mapping μ close to ν that does not affect a preassigned subshift. The case when

μ does not have full support, involving difficulties similar to those in the second half of the proof of Theorem 9.2, is treated in Lemma 10.7. Using a decomposition from Lemma 10.8, the pieces of the proof for sufficiency are, at last, assembled.

PROOF OF THEOREM 10.1 (NECESSITY). Suppose $\mu \in P(X_T)$ is continuous with ergodic decomposition $\mu = \int_0^1 \mu^s \, ds$, and $\varphi_n \in G$ has $\varphi_n \mu \to \nu$ weakly. Fix $t \in [0,1]$ and put $\mu' = \int_0^t \mu^s \, ds$. By compactness, a subsequence of $\varphi_n \mu'$ converges to some ν', where $\nu'(X_T) = t$ and $\nu' \leq \nu$. By the intrinsic definition of H_ν given above, we get $h(\nu') \leq H_\nu(t)$. Using the upper semicontinuity of entropy [**W**, Theorem 8.2] and that $h(\mu') = h(\varphi_n \mu')$ for all n, we have

$$H_\mu(t) = h(\mu') = \limsup_{n \to \infty} h(\varphi_n \mu') \leq h(\nu') \leq H_\nu(t). \quad \square$$

We begin building the machinery to prove sufficiency. For $\mu \in M(X_T)$ let $\operatorname{supp}(\mu)$ denote the complement of the largest open μ-null set.

LEMMA 10.3. *Let $\mu, \nu \in P(X_T)$, let $Z \subset X_T$ be a proper subshift, and suppose $h(\nu) > h(\mu)$. For every neighborhood V of ν in $P(X_T)$ and $\varepsilon > 0$, there is a mixing Markov measure $\tilde{\nu} \in V$ with $|h(\tilde{\nu}) - h(\nu)| < \varepsilon$, $\operatorname{supp}(\tilde{\nu}) \cap Z = \emptyset$, and $\tilde{\nu} \perp \mu$.*

PROOF. Let V be a neighborhood of ν in $P(X_T)$ and $\varepsilon > 0$. Replacing T by a higher order block presentation if necessary, there is a δ with $0 < \delta < \varepsilon$ so that if $|\nu'(a^*) - \nu(a^*)| < \delta$ for all $a \in \mathcal{L}_T$, then $\nu' \in V$. Since Z is proper, there is a $D \in \mathcal{B}_d(X_T)$ so that $D^* \cap Z = \emptyset$, where we continue to use the notation $D^* = \{x \in X_T : x[0, d-1] = D\}$. Let l be a transition length for T, and let $\gamma > 0$ whose value will be determined later. By the ergodic and Shannon-McMillan-Breiman theorems, there is an m large enough so $m/(2l+d+m+1) > 1 - \delta/3$ and collections $\mathcal{C}_j \subset \mathcal{B}_j(X_T)$ $(j = m, m+1)$ so that

$$\exp[m(h(\nu) - \gamma)] < |\mathcal{C}_j| < \exp[mh(\nu)] \tag{10–1}$$

and

$$|f_a(C) - \nu(a^*)| < \frac{\delta}{3} \quad \text{for all } C \in \mathcal{C}_m \cup \mathcal{C}_{m+1}, \tag{10–2}$$

where $f_a(C)$ denote the frequency of the symbol a in C. For each $C \in \mathcal{C} = \mathcal{C}_m \cup \mathcal{C}_{m+1}$ choose $A_0, A_1 \in \mathcal{B}_l(X_T)$ with $C' = A_0 C A_1$ so that $DC'D \in \mathcal{B}(X_T)$. Form the collection $\mathcal{C}'$ of such blocks. Let Y be the shift of finite type with the subblocks of concatenations of elements of $D\mathcal{C}'$ as allowed blocks. Clearly Y is topologically mixing since m and $m+1$ are relatively prime. Let $\overline{\nu}$ be the measure of maximal entropy on Y. We show first that $|h(\overline{\nu}) - h(\nu)| < \varepsilon$ and $\operatorname{supp}(\overline{\nu}) \cap Z = \emptyset$, then perturb $\overline{\nu}$ to get singularity with respect to μ.

First note that $Y \subset \bigcup_{j=0}^{2l+d+m} \sigma_T^j(D^*)$ and $Z \cap \sigma_T^j(D^*) = \emptyset$ for all j, so $Y \cap Z = \emptyset$.

Let ν' be any σ_T-invariant measure with $\nu'(Y) = 1$. By the ergodic theorem, there is a $y \in Y$ so that $y[0, d-1] = D$ and $|f_a(y[0, n-1]) - \nu'(a^*)| < \delta/3$ for all $a \in \mathcal{L}_T$ and sufficiently large n. Pick n so that $y[0, n-1]$ has the form $DC_1'DC_2' \cdots DC_k'$, $C_j' \in \mathcal{C}'$. Then $k > n/(2l+d+m)$, and if C_j' has central block

C_j then

$$\left|nf_a(y[0,n-1]) - n\nu(a^*)\right| \le (2d+l)k + \sum_{j=1}^{k} |C_j|\left(f_a(C_j) - \nu(a^*)\right)$$
$$\le \left(\frac{2l+d}{2l+d+m}\right)n + \frac{\delta}{3}n < \frac{2}{3}\delta n.$$

In particular, $|\overline{\nu}(a^*) - \nu(a^*)| < \delta$ for $a \in \mathcal{L}_T$, proving $\overline{\nu} \in V$.

Next we estimate $h(\overline{\nu}) = h(\sigma_T|_Y)$ by estimating $|\mathcal{B}_n(Y)|$. Every $E \in \mathcal{B}_n(Y)$ has the form $E = ADC_1'DC_2' \cdots DC_k'B$, where $|A|, |B| \le 2l+d+m$, $C_j' \in \mathcal{C}'$, and

$$\frac{n}{2l+d+m+1} - 2 \le k \le \frac{n}{2l+d+m}.$$

Making independent choices of the C_j', it follows from (10–1) that

$$|\mathcal{B}_n(Y)| \ge \left(\exp[m(h(\nu)-\gamma)]\right)^k$$
$$\ge \kappa_1 \exp\left[\left(\frac{m}{2l+d+m+1}\right)n(h(\nu)-\gamma)\right]$$
$$\ge \kappa_1 \exp\left[\left(1-\frac{\delta}{3}\right)n(h(\nu)-\gamma)\right],$$

where $\kappa_1 > 0$. Hence

$$h(\sigma_T|_Y) = \lim_{n\to\infty}\frac{1}{n}\log|\mathcal{B}_n(Y)| \ge \left(1-\frac{\delta}{3}\right)(h(\nu)-\gamma) > h(\nu) - \varepsilon$$

for γ small enough. Similar estimates show

$$|\mathcal{B}_n(Y)| \le |\mathcal{B}_{2l+d+m}(X_T)|^2\left(\exp[mh(\nu)]\right)^k$$
$$\le \kappa_2 \exp\left[\left(\frac{m}{2l+d+m}\right)nh(\nu)\right]$$

for suitable $\kappa_2 > 0$, showing $h(\sigma_T|_Y) \le h(\nu)$. Thus $h(\nu) - \varepsilon < h(\overline{\nu}) \le h(\nu)$.

We have hence found a mixing Markov measure $\overline{\nu} \in V$ so that $|h(\overline{\nu}) - h(\nu)| < \varepsilon$ and $\operatorname{supp}(\overline{\nu}) \cap Z = \varnothing$. By perturbing the transition probabilities of $\overline{\nu}$, we can obtain a continuous family $\{\overline{\nu}_\alpha\}$ of mixing Markov measures each supported on Y and with $\overline{\nu}_0 = \overline{\nu}$. For small enough α the $\overline{\nu}_\alpha$ remain in V and have $|h(\overline{\nu}_\alpha) - h(\nu)| < \varepsilon$. Now the $\overline{\nu}_\alpha$ are ergodic and distinct, so they are mutually singular. It follows that all but countably many of the $\overline{\nu}_\alpha$ are singular with respect to μ. Set $\tilde{\nu}$ to be any of these. □

If μ has ergodic decomposition $\mu = \int_0^1 \mu^t\,dt$, define $h_\infty(\mu) = \operatorname{ess\,sup}_{0\le t\le 1} h(\mu^t)$.

CONVEX APPROXIMATION LEMMA 10.4. *Let $\mu, \nu \in P(X_T)$ with $H_\mu \le H_\nu$ and $h_\infty(\mu) < h(\sigma_T)$, and let Z be a proper subshift. For every neighborhood V of ν in $P(X_T)$ there is a $\tilde{\nu} = \sum_{k=1}^K \alpha_k\tilde{\nu}_k \in V$ and a corresponding convex decomposition $\mu = \sum_{k=1}^K \alpha_k\mu_k$, where $\alpha_k > 0$ with $\sum_{k=1}^K \alpha_k = 1$, such that the μ_k and $\tilde{\nu}_k$ are all mutually disjoint probability measures, the $\tilde{\nu}_k$ are mixing Markov measures with disjoint supports that are also disjoint from Z, and $h_\infty(\mu_k) < h(\tilde{\nu}_k)$.*

PROOF. Let $\mu = \int_0^1 \mu^t\,dt$ and $\nu = \int_0^1 \nu^t\,dt$ be the monotone ergodic decompositions as described above. Put $\beta = h(\sigma_T) - h_\infty(\mu) > 0$. Choose $\varepsilon > 0$ so that

if $\nu' \in P(X_T)$ and $\|\nu - \nu'\| < 2\varepsilon$, then $\nu' \in V$. By partitioning $[0, h(\sigma_T)]$ into intervals of length $< \varepsilon\beta/2$, we can find $0 = t_0 < t_1 < \cdots < t_{K-1} < t_K = 1$ so that if $J_k = [t_{k-1}, t_k)$ and $\mu_k = |J_k|^{-1} \int_{J_k} \mu^t \, dt$, then the μ_k are mutually singular and

$$\sup_{t \in J_k} \{h(\mu^t)\} - \inf_{t \in J_k} \{h(\mu^t)\} < \varepsilon\beta/2. \tag{10–3}$$

Put $\alpha_k = |J_k|$. We now inductively define $\nu_k \in P(X_T)$ so that $\nu = \sum_{k=1}^K \alpha_k \nu_k$ and $h(\nu_k) \geq h(\mu_k)$. Let

$$s_1 = \sup \left\{ s \leq 1 - t_1 \colon \int_s^{s+t_1} h(\nu^t) \, dt \geq \int_0^{t_1} h(\mu^t) \, dt \right\}.$$

Since $H_\nu(t_1) \geq H_\mu(t_1)$, the set defining s_1 is nonvoid. Put $\nu_1 = t_1^{-1} \int_{s_1}^{s_1+t_1} \nu^t \, dt$. By definition $h(\nu_1) \geq h(\mu_1)$, and we claim $H_{\nu-\nu_1} \geq H_{\mu-\mu_1}$ on $[0, 1 - t_1]$, which allows inductive construction of all the ν_k. To prove the claim, first suppose $s_1 = 1 - t_1$. Then using monotonicity

$$\begin{aligned} \inf_{0 \leq t \leq 1-t_1} \{h(\nu^t)\} &\geq \frac{1}{t_1} \int_{1-t_1}^1 h(\nu^t) \, dt \\ &\geq \frac{1}{t_1} \int_0^{t_1} h(\mu^t) \, dt \geq \sup_{t_1 \leq t \leq 1} \{h(\mu^t)\}, \end{aligned}$$

so $H_{\nu-\nu_1} \geq H_{\mu-\mu_1}$ follows trivially. Now assume $s_1 < 1 - t_1$, so by continuity $\int_{s_1}^{s_1+t_1} h(\nu^t) \, dt = \int_0^{t_1} h(\mu^t) \, dt$. If $u < s_1$, then

$$\begin{aligned} H_{\nu-\nu_1}(u) &= \int_0^u h(\nu^t) \, dt \geq \int_0^u h(\mu^t) \, dt \\ &\geq \int_{t_1}^{t_1+u} h(\mu^t) \, dt = H_{\mu-\mu_1}(u), \end{aligned}$$

while if $s_1 < u < 1 - t$, then

$$\begin{aligned} H_{\nu-\nu_1}(u) &= \int_0^{t_1+u} h(\nu^t) \, dt - \int_{s_1}^{s_1+t_1} h(\nu^t) \, dt \\ &\geq \int_0^{t_1+u} h(\mu^t) \, dt - \int_0^{t_1} h(\mu^t) \, dt = H_{\mu-\mu_1}(u), \end{aligned}$$

verifying our claim.

For each k let $\nu_k = \int_0^1 \nu_k^t \, dt$ be the ergodic decomposition, put $\nu_k' = \int_0^{1-\varepsilon} \nu_k^t \, dt + \varepsilon\mu_T$, and sum $\nu' = \sum_{k=1}^K \nu_k'$. We show $h(\nu_k') > h_\infty(\mu_k)$. For by (10–3)

$$\begin{aligned} h(\nu_k') &= h(\nu_k) - \int_{1-\varepsilon}^1 h(\nu_k^t) \, dt + \varepsilon h(\sigma_T) \\ &\geq h(\mu_k) - \varepsilon h_\infty(\mu_k) + \varepsilon h(\sigma_T) \\ &\geq h_\infty(\mu_k) + \varepsilon\beta/2 > h_\infty(\mu_k). \end{aligned}$$

Since $\|\nu - \nu'\| \leq 2\varepsilon$, it follows by our choice of ε that $\nu' \in V$. Finally, by Lemma 10.3 we can inductively modify each ν_k' to $\tilde{\nu}_k$ so that $\tilde{\nu} = \sum_{k=1}^K \alpha_k \tilde{\nu}_k$ satisfies the conclusions. □

If $F \subset X_T$ and $x \in F$, define $r_F(x) = \min\{n > 0 \colon \sigma_T^n x \in F\}$, and put $F_n = \{x \in F : r_F(x) = n\}$. For a collection $\mathcal{C} \subset \mathcal{B}(X_T)$ let $\mathcal{C}^* = \bigcup_{C \in \mathcal{C}} C^*$.

LEMMA 10.5. *Suppose $\mu_k \in P(X_T)$ for $1 \le k \le K$ are mutually singular, and $\varepsilon > 0$. There is an $M = M(\varepsilon, \mu_1, \dots, \mu_K)$ so that if F is compact open and $N \ge n \ge M$, then there are collections $\mathcal{C}_n(\mu_k) \subset \mathcal{B}_n(X_T)$, disjoint in n and k, such that $|\mathcal{C}_n(\mu_k)| < \exp[n(h_\infty(\mu_k) + \varepsilon)]$ for $1 \le k \le K$ and*

$$\sum_{n=M}^{N} n\mu_k(F_n \cap \mathcal{C}_n(\mu_k)^*) \ge \sum_{n=M}^{N} n\mu_k(F_n) - \varepsilon.$$

PROOF. The rough idea is to use singularity of the μ_k to first find disjoint collections $\mathcal{D}_k$ of atoms having the right exponential size and containing most of the μ_k-mass. Using the idea that a partial orbit can intersect at most one of these atoms with frequency $> \frac{1}{2}$, we obtain from the $\mathcal{D}_k$ the required disjoint collections $\mathcal{C}_n(\mu_k)$.

Since the μ_k are mutually singular, there is an s and disjoint collections $\mathcal{D}_k \subset \mathcal{B}_{2s+1}(X_T)$ such that $\mu_k(\mathcal{D}_k^*) > 1 - \varepsilon/12$ for $1 \le k \le K$. Applying the version of the Shannon-McMillan-Breiman theorem from [**DGS**, Theorem 13.4] to σ_T and to σ_T^{-1}, for all large enough M and $1 \le k \le K$ we have

$$\mu_k(E_k^+) = \mu_k\left\{x\colon \sup_{m \ge M/6} -\frac{1}{m}\log \mu_k(x[0, m-1]^*) < h_\infty(\mu_k) + \frac{\varepsilon}{2}\right\} > 1 - \frac{\varepsilon}{12},$$

$$\mu_k(E_k^-) = \mu_k\left\{x\colon \sup_{m \ge M/6} -\frac{1}{m}\log \mu_k(x[-m+1, 0]^*) < h_\infty(\mu_k) + \frac{\varepsilon}{2}\right\} > 1 - \frac{\varepsilon}{12}.$$

Let $E_k = E_k^+ \cap E_k^-$, and fix M so that $\mu_k(E_k) > 1 - \varepsilon/6$, and also satisfying $M > 6s$ and

$$\frac{n}{3}\exp\left[h_\infty(\mu_k) + (n+1)\frac{\varepsilon}{2}\right] < \exp[n\varepsilon] \tag{10–4}$$

for $n \ge M/6$ and $1 \le k \le K$.

Suppose now F is compact open, and $M < n < N$. If $x \in F_n$ with $\sigma_T^j x \in E_k$ for some $n/3 \le j \le 2n/3$, then

$$\mu_k(x[j, n-1]^*) > \exp\left[-(n-j)\left(h_\infty(\mu_k) + \frac{\varepsilon}{2}\right)\right].$$

The number of blocks $x[j, n-1]$ arising from such an x is thus bounded above by $\exp[(n-j)(h_\infty(\mu_k) + \varepsilon/2)]$. Similarly the number of blocks $x[0, j]$ from such an x is $\le \exp[(j+1)(h_\infty(\mu_k) + \varepsilon/2)]$. It follows using (10–4) that the number of blocks $x[0, n-1]$, where $x \in F_n$ and $\sigma_T^j x \in E_k$ for some $n/2 \le j \le 2n/3$ is bounded above by

$$\frac{n}{3}\exp\left[(n+1)\left(h_\infty(\mu_k) + \frac{\varepsilon}{2}\right)\right] \le \exp\left[n(h_\infty(\mu_k) + \varepsilon)\right].$$

Now let $\mathcal{C}_n(\mu_k)$ denote the collection of blocks $x[0, n-1]$, where $x \in F_n$, $\sigma_T^j x \in E_k$ for some $n/3 \le j \le 2n/3$, and

$$\left|\left\{j : \frac{n}{3} \le j \le \frac{2n}{3},\ x[j-s, j+s] \in \mathcal{D}_k\right\}\right| > \frac{n}{6}.$$

Since the $\mathcal{D}_k$ are disjoint, clearly the $\mathcal{C}_n(\mu_k)$ are disjoint in k and trivially in n since the lengths differ. The discussion above proves $|\mathcal{C}_n(\mu_k)| < \exp[n(h_\infty(\mu_k) + \varepsilon)]$.

Finally, since

$$n\mu_k\{x \in F_n : x[0,n-1] \notin \mathcal{C}_n(\mu_k)\} \leq 3\mu_k\left(E_k^c \cap \bigcup_{j=n/3}^{2n/3} \sigma_T^j F_n\right) + 6\mu_k\left(\mathcal{D}_k^{*c} \cap \bigcup_{j=n/3}^{2n/3} \sigma_T^j F_n\right),$$

we have

$$\sum_{n=M}^{N} n\mu_k(F_n \cap \mathcal{C}_n(\mu_k)^*) \geq \sum_{n=M}^{N} n\mu_k(F_n) - 3\mu_k(E_k^c) - 6\mu_k(\mathcal{D}_k^{*c}) \geq \sum_{n=M}^{N} n\mu_k(F_n) - \varepsilon. \qquad \square$$

LEMMA 10.6. *Suppose $\mu, \nu \in P(X_T)$ with $H_\mu \leq H_\nu$, that μ is continuous, and that $\operatorname{supp}(\mu^t) = X_T$ for a.e. t. Let Z be a proper subshift of X_T and V be a neighborhood of ν in $P(X_T)$. Then there is an involution $\varphi \in \operatorname{aut}(\sigma_T)$ such that $\varphi\mu \in V$ and φ is the identity on Z.*

PROOF. We first reduce to the case $h_\infty(\mu) < h(\sigma_T)$. Let $t_0 = \sup\{t : h(\mu^t) = h(\sigma_T)\}$. Since σ_T is intrinsically ergodic, for $0 \leq t \leq t_0$ we have $\mu^t = \nu^t = \mu_T$. Since μ_T is G-invariant, any $\varphi \in G$, in particular the φ we construct below, maps $\int_0^{t_0} \mu^t\, dt = t_0\mu_T$ to $\int_0^{t_0} \nu^t\, dt$. Removing this part from each, we are reduced to the case $h(\mu^t) < h(\sigma_T)$ for all t. Since it is enough to map $(1-\varepsilon)^{-1}\int_\varepsilon^1 \mu^t\, dt$ to V for a sufficiently small $\varepsilon > 0$, we can and do assume $h_\infty(\mu) < h(\sigma_T)$.

Using the Convex Approximation Lemma, there is $\tilde{\nu} = \sum_{k=1}^K \alpha_k \tilde{\nu}_k \in V$ and $\mu = \sum_{k=1}^K \alpha_k \mu_k$ with $\alpha_k > 0$, $\sum_{k=1}^K \alpha_k = 1$, so that the μ_k and $\tilde{\nu}_k$ are all mutually disjoint, the $\tilde{\nu}_k$ are mixing Markov measures with supports Y_k that are mutually disjoint and disjoint from Z, and $h_\infty(\mu_k) < h(\tilde{\nu}_k)$.

By passing to a higher block presentation, we can assume that the $\sigma_T|_{Y_k}$ and σ_T are 1-step, $\mathcal{L}_{Y_k}$ and $\mathcal{L}_Z$ are all disjoint, and that there is a $\delta_0 > 0$ so that if $|\nu'(a^*) - \tilde{\nu}(a^*)| < \delta_0$, then $\nu' \in V$. Let $\delta > 0$, which will eventually be made small enough for everything which follows to work. The first requirement on δ is that

$$(1-\delta)[h(\tilde{\nu}_k) - \delta] > h(\mu_k) + \delta \tag{10–5}$$

for all k.

By the ergodic and Shannon-McMillan-Breiman theorems, there exists an M_0 so that for $n > M_0$ there is a collection $\mathcal{C}_n(\tilde{\nu}_k) \subset \mathcal{B}_n(Y_k)$ so that

$$|\mathcal{C}_n(\tilde{\nu}_k)| > \exp[n(h(\tilde{\nu}_k) - \delta)],$$
$$\tilde{\nu}_k(\mathcal{C}_n(\tilde{\nu}_k)^*) > 1 - \delta,$$

and

$$|f_a(C) - \tilde{\nu}_k(a^*)| < \delta \quad \text{for all } a \in \mathcal{L}_T. \tag{10–6}$$

There is a block $A_0 \in \mathcal{B}(X_T)$ so that $A_0^* \cap \left(Z \cup \bigcup_{k=1}^K Y_k\right) = \varnothing$. Since the number of blocks in $\mathcal{B}_n(X_T)$ beginning with A_0 grows exponentially in n, and each μ_k has full support, there is a block A beginning with A_0 so that $0 < \mu_k(A^*) < \delta^2/4|A|$

for $1 \le k \le K$. Let $|A| = d$ and put $M = [d/4\delta]$. Let l be a transition length for each $\sigma_T|_{Y_k}$ and for σ_T. We may assume d is large enough so that $M \ge M_0$, that $M \ge M(\delta, \mu_1, \dots, \mu_K)$ from Lemma 10.5, and that $n-2d-4l > (1-\delta)n$ for $n \ge M$. Since $\operatorname{supp}(\mu_k) = X_T$, there is an $N > M$ so that $\mu_k\left(\bigcup_{j=0}^{N-1} \sigma_T^j A^*\right) > 1-\delta$ for $1 \le k \le K$. Let $F = A^*$. If $F_n = \{x \in F : r_F(x) = n\}$, then

$$\sum_{n=M}^{N} n\mu_k(F_n) > 1 - M\mu_k(A^*) - \delta > 1 - 2\delta.$$

By Lemma 10.5, there are mutually disjoint collections $\mathcal{C}_n(\mu_k) \subset \mathcal{B}_n(X_T)$ so that $|\mathcal{C}_n(\mu_k)| \le \exp[n(h_\infty(\mu_k) + \delta)]$ and

$$\sum_{n=M}^{N} n\mu_k\big(F_n \cap \mathcal{C}_n(\mu_k)^*\big) > 1 - 3\delta. \tag{10–7}$$

We now construct φ. For each k choose a block A_k so $AA_k \in \mathcal{B}(X_T)$ has minimal length such that it ends with a block from $\mathcal{B}_d(Y_k)$. Note that $|A_k| < d$ is possible. Clearly $|A_k| < 2d + l$ since l is a transition length for σ_T. Similarly choose B_k so $B_kA \in \mathcal{B}(X_T)$ and has minimal length for such a block starting with a block from $\mathcal{B}_d(Y_k)$. For each $C \in \mathcal{C}_{n-3d-4l}(\tilde{\nu}_k)$ there is a C' containing C as a subblock so that $AA_kC'B_kA \in \mathcal{B}_{n+d}(X_T)$, and so that A occurs only as the initial and terminal d-block. Form $\mathcal{D}_{nk} = \{AA_kC'B_kA : C \in \mathcal{C}_{n-3d-4l}(\tilde{\nu}_k)\}$, so

$$\begin{aligned} |\mathcal{D}_{nk}| &= |\mathcal{C}_{n-3d-4l}(\tilde{\nu}_k)| > \exp[(n-3d-4l)(h(\tilde{\nu}_k) - \delta)] \\ &> \exp[n(1-\delta)(h(\tilde{\nu}_k) - \delta)] \\ &> \exp[n(h_\infty(\mu_k) + \delta)] > |\mathcal{C}_n(\mu_k)|. \end{aligned}$$

Thus it is possible to define a permutation θ_{nk} of $A\mathcal{B}_{n-d}(X_T)A$ so that $\theta_{nk}^2 = I$, $\theta_{nk}(\mathcal{C}_n(\mu_k)A) \subset \mathcal{D}_{nk}$, and θ_{nk} is the identity off $\mathcal{C}_n(\mu_k)A \cup \theta_{nk}(\mathcal{C}_n(\mu_k)A)$ Note that each block in $\mathcal{C}_n(\mu_k)$ begins with A, so these θ_{nk} are exactly the type introduced in §2, where here A is the marker block.

If $x \in X_T$ with $x[i, i+n+d] \in \mathcal{C}_n(\mu_k)A \cup \theta_{nk}(\mathcal{C}_n(\mu_k)A)$, where $M \le n \le N$ and $1 \le k \le K$, define $(\varphi x)([i, i+n+d]) = \theta_{nk}(x[i, i+n+d])$. Since the $\mathcal{C}_n(\mu_k)$ are disjoint in n and k, and $\theta_{nk}^2 = I$, it follows that φ is a well-defined involution.

We complete this proof by showing that $|\varphi\mu_k(a^*) - \tilde{\nu}_k(a^*)| < \delta_0$ for each k and $a \in \mathcal{L}_T$. This will imply that $|\varphi\mu(a^*) - \tilde{\nu}(a^*)| < \delta_0$ for $a \in \mathcal{L}_T$, so $\varphi\mu \in V$, the required conclusion. If $C \in \mathcal{C}_n(\mu_k)$ then

$$\begin{aligned} \varphi\mu_k\left(a^* \cap \bigcup_{j=0}^{n-1} \sigma_T^i(F_n \cap \theta_{nk}(C^*))\right) &= nf_a\big(\theta_{nk}(C)\big)\varphi\mu_k\big(F_n \cap \theta_{nk}(C)^*\big) \\ &= nf_a\big(\theta_{nk}(C)\big)\mu_k(F_n \cap C^*). \end{aligned}$$

Letting

$$E = \bigcup_{n=M}^{N} \bigcup_{j=0}^{n-1} \bigcup_{C \in \mathcal{C}_n(\mu_k)} \sigma_T^j\big(F_n \cap \theta_{nk}(C)^*\big),$$

then (10–7) above shows that $\varphi\mu_k(E) > 1 - 3\delta$ while by (10–6) $|\varphi\mu_k(a^* \cap E) - \tilde{\nu}_k(a^*)| < 2\delta$. For δ sufficiently small this forces $|\varphi\mu_k(a^*) - \tilde{\nu}_k(a^*)| < \delta_0$, concluding the proof. □

The following result deals with the case that μ does not have full support. Complications arise because μ may be highly recurrent, so that constructing markers with the necessary disjointness is more difficult. To appreciate the problems, an excellent example to keep in mind while reading the proof is for μ to be the unique invariant measure supported on the Morse minimal set and ν to be the measure of maximal entropy on $X_{[2]}$. In this sense, the difficulties parallel those in the second half of the proof of Theorem 9.2.

LEMMA 10.7. *Suppose $\mu, \nu \in P(X_T)$ with $H_\mu \le H_\nu$, that μ is continuous, and* $\operatorname{supp}(\mu) \neq X_T$. *For every neighborhood V of ν in $P(X_T)$ there is an involution $\varphi \in \operatorname{aut}(\sigma_T)$ with $\varphi\mu \in V$.*

PROOF. Let $Z = \operatorname{supp}(\mu) \neq X_T$. Since $\operatorname{supp}(\mu^t) \subset Z$ for a.e. t, it follows that $h_\infty(\mu) \le h(\sigma_T|_Z) < h(\sigma_T)$. By the Convex Approximation Lemma, there is a $\tilde{\nu} = \sum_{k=1}^K \alpha_k \tilde{\nu}_k \in V$ and $\mu = \sum_{k=1}^K \alpha_k \mu_k$, where $\alpha_k > 0$ with $\sum_{k=1}^K \alpha_k = 1$, the μ_k and $\tilde{\nu}_k$ are all mutually singular, each $\tilde{\nu}_k$ is mixing Markov supported on a mixing shift of finite type Y_k with the Y_k and Z mutually disjoint, and $h_\infty(\mu_k) < h(\tilde{\nu}_k)$ for all k.

By passing to a higher order block presentation, we may arrange $\mathcal{L}_{Y_k}$ and $\mathcal{L}_Z$ to be disjoint, $\sigma_T|Y_k$ and σ_T to be 1-step, and for there to exist a $\delta_0 > 0$ so that if $|\nu'(a^*) - \tilde{\nu}(a^*)| < \delta_0$, then $\nu' \in V$. Introduce $\delta > 0$, which will eventually be made small enough for the following to work.

Before starting the main argument, we first recode σ_T so that for each $a \in \mathcal{L}_Z$ and all k there are "escape" blocks A_{ak} and A_{ka} so that A_{ak} starts with a, has no other symbol in $\mathcal{L}_Z$, and ends with a symbol from $\mathcal{L}_{Y_k}$, while A_{ka} has these properties in reverse order. To arrange this recoding, first choose A'_{ak} starting with a, no other symbol is a, and ending with a symbol from $\mathcal{L}_{Y_k}$. Choose A'_{ka} to have these properties in reverse order. Introduce entirely new distinct symbols b_{ak}, $b_{ka} \notin \mathcal{L}_T$ for all a and k. Form a new shift of finite type from X_T by replacing all symbols in A'_{ak} except the first and last by the same number of b_{ak}'s, and similarly with the A'_{ka} and b_{ka}'s. This recoding of X_T has the property sought, but is not 1-step. Pass to a higher order block presentation so the resulting shift is 1-step, and the existence of the escape blocks as above is preserved.

By the ergodic and Shannon-McMillan-Breiman theorems, there is an M_0 so that for $n \ge M_0$ and all k there is a collection $\mathcal{C}_n(\tilde{\nu}_k) \subset \mathcal{B}_n(Y_k)$ so that

$$\tilde{\nu}_k\big(\mathcal{C}_n(\tilde{\nu}_k)^*\big) > 1 - \delta, \qquad |\mathcal{C}_n(\tilde{\nu}_k)| > \exp[n(h(\tilde{\nu}_k) - \delta)], \qquad \text{and}$$

$$|f_a(C) - \tilde{\nu}_k(a^*)| < \delta \quad \text{for all } C \in \mathcal{C}_n(\tilde{\nu}_k) \text{ and } a \in \mathcal{L}_T \tag{10-8}$$

Let $M_1 = M(\delta, \mu_1, \dots, \mu_K)$ from Lemma 10.5, and let l be a transition length for σ_T and each $\sigma_T|_{Y_k}$. Now fix $M > \max\{4l/\delta, M_0, M_1\}$. For $n \ge M$, each $c, d \in \mathcal{L}_Z$, and $C \in \mathcal{C}_{n-4l}(\tilde{\nu}_K)$ there is a block $A_{ck}C'A_{kd} \in \mathcal{B}_n(X_T)$ with C a subblock of C'. Form the collection $\mathcal{D}_{nk}(c,d)$ of such blocks. Note that each $D \in \mathcal{D}_{nk}(c,d)$ starts with c, ends with d, has no other symbols in $\mathcal{L}_Z$, and $|f_a(D) - \tilde{\nu}_k(a^*)| < 2\delta$ by (10-8). Also,

$$|\mathcal{D}_{nk}(c,d)| = |\mathcal{C}_{n-4l}(\tilde{\nu}_k)| \ge \exp[(n-4l)(h(\tilde{\nu}_k) - \delta)]. \tag{10-9}$$

We now construct φ. Recall that $C \in \mathcal{B}_n(X_T)$ is called j-periodic if $C[0, n-j] = C[j, n]$. Since μ is continuous, there is a $p > M$ so that for all k,

$$\mu_k\left(\bigcup\{C^*: C \in \mathcal{B}_{2p+1}(X_T) \text{ is } j\text{-periodic for some } j < M\}\right) < \delta.$$

By [**Kr1**] or [**B1**, Lemma 2.2], there is a compact open set $F \subset Z$ such that $\{\sigma_T^j F: 0 \le j \le M\}$ is disjoint, and

$$\{x \in Z: x[-p, p] \text{ is not } j\text{-periodic for all } j < M\} \subset \bigcup_{j=0}^{2M-1} \sigma_T^j F.$$

Recall that r_F is the return time function for F, and put $F_n = \{x \in F: r_F(x) = n\}$. By the above we have

$$\sum_{n=M}^{2M-1} n\mu_k(F_n) > 1 - \delta. \tag{10–10}$$

Since F is compact open, there is a $q > p$ so that F is a union of sets $\sigma_T^{-q} B^*$ for $B \in \mathcal{B}_{2q+1}(Z)$. Let

$$\mathcal{F} = \{A \in \mathcal{B}_{2q+4M+1}(Z): \sigma_T^{-(q+2M)} A^* \subset F\}.$$

Since $M > M(\delta, \mu_1, \ldots, \mu_k)$ from Lemma 10.5, choose disjoint collections $\mathcal{C}_n(\mu_k) \subset \mathcal{B}_n(Z)$ for $M \le n < 2M$ and all $1 \le k \le K$ so that $|\mathcal{C}_n(\mu_k)| \le \exp[n(h_\infty(\mu_k) + \delta)]$ and

$$\sum_{n=M}^{2M-1} n\mu_k\left(F_n \cap \mathcal{C}_n(\mu_k)^*\right) > 1 - 2\delta.$$

Thus for δ small enough $|\mathcal{C}_n(\mu_k)| < |\mathcal{D}_{nk}(c, d)|$ for all $c, d \in \mathcal{L}_Z$. It follows there is a injection

$$\theta_n: \mathcal{C}_n(\mu_k) \to \bigcup_{c,d \in \mathcal{L}_Z} \mathcal{D}_{nk}(c, d)$$

that fixes the first and last symbols.

Now define $\xi: \mathcal{B}(X_T) \to \mathcal{B}(X_T)$ to replace any occurrence of $\theta_n(C)$ by C, where $M \le n \le 2M - 1$ and $C \in \mathcal{C}_n(\mu_k)$. This map is well-defined because blocks from $\mathcal{D}_{nk}(c, d)$ can only overlap in the end symbols, and these symbols are fixed by θ_n. Let

$$\mathcal{E} = \{E[2M, 2q + 6M + 1]: E \in \xi^{-1}(\mathcal{B}_{2M}(Z)\mathcal{F}\mathcal{B}_{2M}(Z))\}$$

and put $E = \sigma_T^{-(q+2M)}\mathcal{E}^*$. We claim that $\{\sigma_T^j E: 0 \le j < M\}$ is disjoint. For if $x = \sigma_T^j y$ for $x, y \in E$ and some $0 \le j < M$, applying ξ to $x[-q - 2M, q + 2M]$ and $y[-q - 2M - j, q + 2M - j]$ shows this would contradict disjointness of $\{\sigma_T^j F: 0 \le j < M\}$. The mapping φ is now defined as follows. If $\sigma_T^i x \in E$ and $\sigma_T^{i+n-1} \in E$ for $M \le n < 2M$, then

$$(\varphi x)[i, i + n - 1] = \begin{cases} \theta_n(x[i, i + n - 1]), & \text{if } x[i, i + n - 1] \in \mathcal{C}_n(\mu_k), \\ \theta_n^{-1}(x[i, i + n - 1]), & \text{if } x[i, i + n - 1] \in \theta_n(\mathcal{C}_n(\mu_k)). \end{cases}$$

Because $\{\sigma_T^j: 0 \le j < M\}$ is disjoint, φ is well-defined. The definition of $\mathcal{E}$ is made so that $x[-q - 2M, q + 2M] \in \mathcal{E}$ iff $(\varphi x)[-q - 2M, q + 2M] \in \mathcal{E}$. Hence $\varphi^2 = I$. Finally, the estimate that $|\varphi\mu(a^*) - \tilde{\nu}(a^*)| < \delta_0$ for all $a \in \mathcal{L}_T$ follows from (10–8) and (10–10) exactly as in the previous lemma. □

LEMMA 10.8. *Let $\mu, \nu \in P(X_T)$ with $H_\mu \le H_\nu$, let $E \subset [0,1]$ have positive measure, and set $\mu_0 = \int_E \mu^t\, dt$, $\mu_1 = \mu - \mu_0$. Then there is a decomposition $\nu = \nu_0 + \nu_1$ with $H_{\nu_j} \ge H_{\mu_j}$ for $j = 0, 1$.*

PROOF. Let a be the measure of E, and put

$$u_0 = \sup\{u \le 1 - a\colon H_\nu(t+u) - H_\nu(u) \ge H_{\mu_0}(t) \text{ for } 0 \le t \le a\}.$$

The defining set clearly contains 0, so is nonempty. Put $\nu_0 = \int_{u_0}^{u_0+a} \nu^t\, dt$ and $\nu_1 = \nu - \nu_0$. By definition, for $0 \le t \le a$,

$$H_{\nu_0}(t) = H_\nu(t+u) - H_\nu(u) \ge H_{\mu_0}(t).$$

We now show $H_{\nu_1} \ge H_{\mu_1}$. If $u_0 = 1 - a$, then since $h(\mu^t)$ is decreasing in t, for $0 \le t \le 1 - a$ we have

$$H_{\nu_1}(t) = H_\nu(t) \ge H_\mu(t) \ge H_{\mu_1}(t). \tag{10–11}$$

Now suppose $u_0 < 1 - a$. Then by continuity

$$\int_{u_0}^{u_0+a} h(\nu^t)\, dt = \int_E h(\mu^t)\, dt. \tag{10–12}$$

If $t < u_0$, then (10–11) shows $H_{\nu_1}(t) \ge H_{\mu_1}(t)$. If $u_0 < t \le 1 - a$, then using (10–12) we obtain

$$\begin{aligned} H_{\nu_1}(t) &= H_\nu(t+a) - \int_{u_0}^{u_0+a} h(\nu^s)\, ds \\ &\ge H_\mu(t+a) - \int_E h(\mu^s)\, ds \ge H_{\mu_1}(t). \end{aligned}$$

□

PROOF OF THEOREM 10.1 (SUFFICIENCY). Suppose μ is continuous with $H_\mu \le H_\nu$, and that V is a neighborhood of ν in $P(X_T)$. There are $\varepsilon > 0$ and a neighborhood V_0 of ν in $M(X_T)$ so that if $\nu' \in V_0$ and $\nu'' \in P(X_T)$, then $[(1-\varepsilon)\nu' + \varepsilon\nu'']/[(1-\varepsilon)\nu'(X_T) + \varepsilon] \in V$.

Since every proper subshift is contained in a proper shift of finite type, and there are only countably many of the latter, there is a proper subshift $Z \subset X_T$ such that

$$m(E_1) = m\big(\{t \in [0,1]\colon \operatorname{supp}(\mu^t) \not\subset Z, \operatorname{supp}(\mu^t) \ne X_T\}\big) < \varepsilon,$$

where m is Lebesgue measure. Let $E_0 = \{t \in [0,1]\colon \operatorname{supp}(\mu^t) \subset Z\}$ and $E_2 = \{t \in [0,1] : \operatorname{supp}(\mu^t) = X_T\}$. These sets are clearly measurable. If $m(E_0) = 0$, then by Lemma 10.6 all but ε of μ can be mapped to a measure in V_0, and we are done.

Thus suppose $m(E_0) = a > 0$. Let $\mu_j = \int_{E_j} \mu^t\, dt$. Two applications of Lemma 10.8 show that there is a decomposition $\nu = \nu_0 + \nu_1 + \nu_2$ with $H_{\nu_j} \ge H_{\mu_j}$ for $0 \le j \le 2$. By Lemma 10.7, there is a $\varphi_1 \in G$ so that $\varphi_1\mu_0$ is close to ν_0. By Lemma 10.6, there is a $\varphi_2 \in G$ that is the identity on $\varphi_1(Z)$ and that maps μ_2 close to ν_2. It follows that the φ_j can be chosen so that $\varphi_2\varphi_1$ maps $\mu_0 + \mu_2$ into V_0, and since $\mu_2(X_T) < \varepsilon$, we obtain finally $(\varphi_2\varphi_1)\mu \in V$. □

11. Problems and questions. During the course of this paper we have indicated several open problems and questions. Two of these seem to us the most important. Recall that F_n is the set of points in X_T with period $\le n$.

QUESTION 6.10. *Is the kernel of the dimension representation generated by elements of finite order?*

QUESTION 7.1. *When is an automorphism in* $\mathrm{aut}(F_n, \sigma_T)$ *the restriction of one in* $\mathrm{aut}(X_T, \sigma_T)$?

Several others seem to be particularly interesting. Problem 3.3 asks whether there is an automorphism of infinite order with nth roots for infinitely many n. This is basic to understanding the kind of divisibility present in G. Our lack of knowledge about the algebraic structure of G is pointed out in Question 4.1, which asks for a nontrivial case of two shifts of finite type having isomorphic automorphism groups. In particular, are the automorphism groups of the 2-shift and the 3-shift isomorphic? Problem 6.1 asks whether the dimension group representation δ of G is always surjective. The p-adic behavior of the gyration function is quite interesting. Specifically (Question 8.5), does $g(\varphi, \sigma_T)(p^n)$ always converge p-adically for every $\varphi \in G$? Problem 9.6 generalizes Question 7.1 above to characterizing the orbits of subshifts of X_T.

REFERENCES

[**ACH**] R. Adler, D. Coppersmith, and M. Hassner, *Algorithms for sliding block codes*, IEEE Trans. on Information Theory **29** (1983), 5–22.

[**Bo**] R. Bowen, *Topological entropy and Axiom* A, Proc. Sympos. Pure Math., vol. 14, Amer. Math. Soc., Providence, R.I., 1970, pp. 23–41.

[**B1**] M. Boyle, *Lower entropy factors of sofic systems*, Ergodic Theory Dynamical Systems **4** (1984), 541–557.

[**B2**] ——, *Shift equivalence and the Jordan form away from zero*, Ergodic Theory Dynamical Systems **4** (1984), 367–379.

[**BK**] M. Boyle and W. Krieger, *Periodic points and automorphisms of the shift*, Trans. Amer. Math. Soc. **302** (1987), 125–149.

[**BMT**] M. Boyle, B. Marcus, and P. Trow, *Resolving maps and the dimension group for shifts of finite type*, Mem. of Amer. Math. Soc. (to appear).

[**Br**] A. Brumer, *On the units of algebraic number fields*, Mathematika **14** (1967), 121–124.

[**C**] E. Coven, *Endomorphisms of substitutions minimal sets*, Z. Wahrsch. Verw. Gebiete. **20** (1971), 129–133.

[**CK**] E. Coven and M. Keane, *The structure of substitution minimal sets*, Trans. Amer. Math. Soc. **162** (1971), 89–102.

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

[**DGS**] M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic theory on compact spaces*, Lecture Notes in Math., vol. 527, Springer, New York, 1976.

[**E**] G. A. Elliott, *On the classification of inductive limits of sequences of semisimple finite-dimensional algebras*, J. Algebra **38** (1976), 29–44.

[**G**] E. Gilbert, *Synchronization of binary messages*, IRE Trans. Inform. Theory **6** (1960), 470–477.

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

[**Ka**] I. Kaplansky, *Infinite Abelian groups*, University of Michigan Press, Ann Arbor, 1954.

[**Ko**] N. Koblitz, *p-adic numbers, p-adic analysis, and zeta-functions*, Springer, New York, 1977.

[**Kr1**] W. Krieger, *On the subsystems of topological Markov chains*, Ergodic Theory Dynamical Systems **2** (1982), 195–202.

[**Kr2**] ——, *On dimension functions and topological Markov chains*, Invent. Math. **56** (1980), 239–250.

[**L1**] D. Lind, *The entropies of topological Markov shifts and a related class of algebraic integers*, Ergodic Theory Dynamical Systems **4** (1984), 283–300.

[L2] ——, *Entropies of automorphisms of a topological Markov shift*, Proc. Amer. Math. Soc. (to appear).

[LS] R. C. Lyndon and P. E. Schupp, *Combinatorial group theory*, Ergebnisse der Math. 89, Springer, Heidelberg, 1977.

[MKS] W. Magnus, A. Karrass, and D. Solitar, *Combinatorial group theory*, Dover, New York, 1976.

[N] M. Nasu, *Topological conjugacy for sofic systems and extensions of automorphisms of finite subsystems of topological Markov shifts*, Proc. Univ. Maryland Special Year in Dynamical Systems 1986–87, Springer (to appear).

[PT] W. Parry and S. Tuncel, *Classification problems in ergodic theory*, London Math. Soc. Lecture Notes 67, Cambridge University Press, Cambridge, 1982.

[Ro] J. Rotman, *The theory of groups*, Allyn and Bacon, Boston, Mass., 1973.

[Ry1] J. Ryan, *The shift and commutativity*, Math. Systems Theory **6** (1972), 82–85.

[Ry2] ——, *The shift and commutativity* II, Math. Systems Theory **8** (1974), 249-250.

[Se] M. Sears, *The automorphisms of the shift dynamical system are relatively sparse*, Math. Systems Theory **5** (1971), 228-231.

[Sh] P. Shields, *The theory of Bernoulli shifts*, Univ. of Chicago Press, Chicago, 1973.

[Sm] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817.

[T] O. Taussky, *On a theorem of Latimer and MacDuffee*, Canad. J. Math. **1** (1949), 300-302.

[Wa1] J. Wagoner, *Realizing symmetries of a subshift of finite type by homeomorphisms of spheres*, Bull. Amer. Math. Soc. **14** (1986), 301–03.

[Wa2] J. Wagoner, *Markov partitions and K_2*, preprint, Univ. of California, Berkeley, 1985.

[W] P. Walters, *An introduction to ergodic theory*, Springer, New York, 1982.

[We] E. Weiss, *Algebraic number theory*, McGraw-Hill, New York, 1963.

[Wi] R. Williams, *Classification of subshifts of finite type*, Ann. of Math. **98** (1973), 120–153; Errata **99** (1974) 380–381.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, COLLEGE PARK, MARYLAND 20742 (Current address of Mike Boyle and Daniel Rudolph)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195 (Current address of Douglas Lind)