Math. Systems Theory 11, 275-282 (1978)

Mathematical
Systems Theory

Measure-Preserving Homeomorphisms of the Torus Represent All Finite Entropy Ergodic Transformations

D. A. Lind[1]* and J.-P. Thouvenot[2]*

[1]Department of Mathematics, University of Washington, Seattle, Washington 98195
[2]Laboratoire de Calcul de Probabilites, 9 Quai Saint Bernard-Tour 56, 75230 Paris, France

Abstract. We show that every finite entropy ergodic transformation can be represented as a Lebesgue measure-preserving homeomorphism of the two-dimensional torus. Whether this is possible with diffeomorphisms is still unknown. Our proof suggests a notion of "universality" for homeomorphisms, about which we formulate several natural but unanswered questions.

1. Introduction and Statement of the Theorem

Ergodic theory originated with Liouville's result that the smooth (i.e. C^∞) Hamiltonian flow on the manifold of constant energy in the phase space of a mechanical system preserves a smooth measure. The theory was quickly generalized to the study of measure-preserving transformations of an arbitrary measure space. Has this singling out of the measure-preserving property expanded the class of objects studied? Kushnirenko [13] showed that smooth systems have finite entropy, so that infinite entropy transformations have been added. But no one knows whether every finite entropy transformation is isomorphic to a differomorphism of a manifold that preserves a smooth measure. A negative answer for diffeomorphisms of a fixed manifold could be physically important, for then the topology of the phase space of a mechanical system alone might rule out certain ergodic behavior.

Anosov and Katok have found a few positive results. In [2] they construct ergodic diffeomorphisms of the two-dimensional disc preserving Lebesgue measure which are isomorphic to certain irrational rotations of the circle. No one knows whether all irrational rotations can be so represented. Their technique works on any manifold with a nontrivial circle action, and also produces zero entropy diffeomorphisms that are weakly mixing, or have any finite or

*Both authors were supported in part by National Science Foundation Grant MCS 76-09159.

0025-5661/78/0011-0$01.60

infinite number of rationally independent eigenfrequencies (thereby demolishing the conjecture of Arnold and Avez [3, Appendix 16] that the number of independent eigenfrequencies is bounded by the dimension of the manifold). Recently, Katok [oral communication] by different methods has concocted diffeomorphisms of all two-dimensional manifolds except the disc and sphere that are isomorphic to Bernoulli shifts. These new methods may work for the disc and sphere as well.

We prove here that every finite entropy ergodic transformation is modelled by a continuous (but not necessarily differentiable) map of the two-dimensional torus $\mathbf{T}^2$.

Theorem 1. *Every finite entropy ergodic transformation is isomorphic to a Lebesgue measure-preserving homeomorphism of* $\mathbf{T}^2$.

Proof. The proof is in three steps. We begin with an arbitrary finite entropy ergodic transformation T of a Lebesgue space (X,μ) that is assumed from now on to be nonatomic. Suppose that M is a transition matrix of zeros and ones, and let S be the shift on the corresponding Markov shift space Σ_M. Further suppose that the topological entropy $h(S)$ is strictly greater than the measure entropy $h_\mu(T)$. The first step is to show in **2** that by a variant of a generator theorem of Krieger we can find an isomorphism Φ of (T,X,μ) with (S,Σ_M,μ_M), where μ_M is shift invariant and has (closed) support all of Σ_M.

The second step begins in **3** with an ergodic automorphism V of $\mathbf{T}^2$. This V automatically preserves Lebesgue measure λ on $\mathbf{T}^2$, and since $h_\lambda(V)>0$, we can assume (by taking powers of V if necessary) that $h_\lambda(V)>h_\mu(T)$. The Markov partition for V discovered by Berg [4] shows that there is a matrix M of zeros and ones, a shift invariant measure λ_M on Σ_M, and an isomorphism Ψ of (S,Σ_M,λ_M) with $(V,\mathbf{T}^2,\lambda)$ that is injective except on the inverse image N of a certain λ-null set. Also, $h(S)=h_\lambda(V)>h_\mu(T)$, so the first step yields a measure μ_M on Σ_M. An argument shows that $\mu_M(N)=0$. Hence if $\mu_2=\Psi(\mu_M)$, then $\Psi\Phi$ is an isomorphism of (T,X,μ) with $(V,\mathbf{T}^2,\mu_2)$, and μ_2 is nonatomic and has full support on $\mathbf{T}^2$.

We complete the proof of Theorem 1 in **4** by observing that these properties of μ_2 are exactly what is required to apply a theorem of Oxtoby and Ulam [16] to obtain a homeomorphism H of $\mathbf{T}^2$ with $H(\mu_2)=\lambda$. Then $U=HVH^{-1}$ is a homeomorphism of $\mathbf{T}^2$ that preserves λ, because V preserves μ_2, and H is an isomorphism of $(V,\mathbf{T}^2,\mu_2)$ with $(U,\mathbf{T}^2,\lambda)$. The original (T,X,μ) is thus isomorphic to $(U,\mathbf{T}^2,\lambda)$ via the composition $H\Psi\Phi$ of three isomorphisms. □

Theorem 1 undoubtedly extends to aperiodic transformations as well. What is needed is an analogue of Krieger's generator theorem for aperiodic transformations, where the natural necessary condition, that the entropy of the transformation on ergodic components be essentially bounded, should also be sufficient. For further details see [5, Chapter 30].

It is unlikely that these arguments will yield diffeomorphisms of $\mathbf{T}^2$, for the measure μ_2 on $\mathbf{T}^2$ can be singular with respect to λ, resulting in a nondifferentiable conjugacy H. However, this alone does not necessarily mean that $U=HVH^{-1}$ is not smooth. No one knows whether diffeomorphisms of $\mathbf{T}^2$ that preserve λ represent all finite entropy transformations.

Infinite entropy homeomorphisms of $\mathbf{T}^2$ probably exist. We are unaware of any examples, much less whether they represent all infinite entropy transformations.

Call a homeomorphism U of a compact metric space Y *universal* if every ergodic transformation of entropy less the topological entropy $h(U)$ is isomorphic to U acting on Y equipped with an invariant measure with full support. In **5** we observe that our proof shows that ergodic atuomorphisms of $\mathbf{T}^2$ and subshifts of finite type are universal. Universality is the opposite of unique ergodicity. Jewett's result on the prevalence of unique ergodicity shows that there are many nonuniversal homeomorphisms. We end with some speculations about universality, probably the most interesting of which is whether every ergodic transformation has a universal model.

2. Markov Embeddings.

Here we represent a measure-preserving transformation by a Markov shift with larger topological entropy acting on an invariant measure whose support is the entire shift space. The proof involves a modification of Denker's proof (as given in ([5], Chapter 28) of a generator theorem of Krieger [12].

Let $M=(M_{ij})$ be an $r\times r$ matrix of zeros and ones. The Markov shift space for M is the compact, totally disconnected space $\Sigma_M=\{(x_i)\in\{1,\ldots,r\}^{\mathbf{Z}}: M_{x_ix_{i+1}}=1\}$. The shift transformation S acts on Σ_M by $S(x_i)=(y_i)$, where $y_i=x_{i+1}$. Parry [17] discovered that if M is aperiodic, so that S is topologically mixing on Σ_M, then there is a unique invariant measure λ_M on Σ_M of maximal entropy. The measure entropy $h_{\lambda_M}(S)$ coincides with the topological entropy $h(S)$, and we denote both by $h(M)$.

Suppose that T is an ergodic transformation of the Lebesgue space (X,μ). Say that a partition $\gamma=\{E_1,\ldots,E_r\}$ of X is M-consistent if, for all $n\geqslant 0$, $\mu(\cap_{i=0}^{n}T^{-i}E_{x_i})>0$ implies that $M_{x_ix_{i+1}}=1$ for $0\leqslant i\leqslant n-1$. Say that γ is M-compatible if, for all $n\geqslant 0$, $\mu(\cap_{i=0}^{n}T^{-i}E_{x_i})>0$ holds if and only if $M_{x_ix_{i+1}}=1$ for $0\leqslant i\leqslant n-1$.

Let γ be an M-consistent generator for T. Then the transformation $\Phi: X\rightarrow\{1,\ldots,r\}^{\mathbf{Z}}=\Sigma_r$ given by $\Phi(x)=(x_i)$, where $T^ix\in E_{x_i}$, maps μ to the measure $\Phi(\mu)=\mu_M$ whose support (i.e. the smallest closed set of full measure) is contained in Σ_M. If γ is also M-compatible, then the support of μ_M is exactly Σ_M, i.e. μ_M is positive on nonempty open subsets of Σ_M. In either case, Φ is an isomorphism of (T,X,μ) with (S,Σ_M,μ_M). Since λ_M is the unique measure of maximal entropy for S, the condition $h_\mu(T)\leqslant h(M)$ is necessary for the existence of an M-compatible, or even an M-consistent, generator. Strict inequality turns out to be sufficient.

Theorem 2. *If T is an ergodic measure-preserving transformation of (X,μ), and M is an aperiodic matrix with $h_\mu(T)<h(M)$, then there is an M-compatible generator for T.*

Proof. Our proof first sketches the relevant features of Denker's proof of the existence of an M-consistent generator, and then describes a modification to obtain M-compatibility. We refer to [5] for unfamiliar terminology. The reader who understands Denker's proof should have little difficulty supplying the details omitted here.

By putting more conditions on the blocks of symbols occurring in the elements of Σ_M, a smaller closed shift invariant subset $\bar{\Sigma}$ can be produced on which S is a mixing subshift of finite type and such that $h_\mu(T) < h(S|\bar{\Sigma}) < h(S)$ (see [5, Lemma 26.17]). Suppose that A is a block of symbols of length $|A|$ that occurs in Σ_M but not in $\bar{\Sigma}$. Choose a transition length L so large that if C is any block occurring in Σ_M, then there are blocks B, B' of length L such that the concatenated blocks ABC and $CB'A$ occur in Σ_M. This L can also be chosen so that if C occurs in $\bar{\Sigma}$, then A occurs in only one place in ABC and in $CB'A$.

The construction of an M-consistent generator γ begins as follows. Let $c = 2(|A| + 2L)$, $k > 0$, and $n > 2(c+k)$. Choose a Rohlin set F such that $\{T^iF\}_0^{n-1}$ is disjoint and fills most of X. We omit subscripts on k, n, and F since only the first stage of the construction concerns us. Divide X into eight sets, labeled I through VIII, by taking in order the unions of successive blocks of iterates of T on F of length L, $|A|$, L, k, L, $|A|$, L, $n-c-k$, respectively. Let IX denote the remainder of X. The blocks of length L are kept for transition purposes. The partition γ is determined on II and VI by using A to label each. On VIII, enough $\bar{\Sigma}$-blocks are used in defining γ to code a first approximation to a generator for T. The partition is inductively defined on an increasing portion of the set IV using a technique that codes better approximations to a generator. Finally, γ is defined on set IX by using arbitrary $\bar{\Sigma}$-blocks.

The essential feature for us is that A occurs in sets II and VI with spacing $2L + |A| + k$, and since $n > 2(c+k)$ no other two occurrences of A in a γ-name have this spacing. Thus A codes the set F. Hence $F \in \bigvee_{-\infty}^{\infty} T^i\gamma = \gamma_T$, and this implies that IX is in γ_T, and hence that γ generates. Thus any extension of γ from the Rohlin stack I–VIII to the remaining set IX using names from Σ_M and such that $F \in \gamma_T$ will still be an M-consistent generator.

Our goal is therefore to define γ on IX such that every block in Σ_M occurs with positive measure, while preserving the property that $F \in \gamma_T$.

Let $r_F(x) = \min\{j > 0 : T^jx \in F\}$ denote the return time function of F. It is easy to arrange F so that $\|r_F\|_\infty = \infty$. This means that arbitrarily long blocks in IX are as yet unlabelled.

The obvious way to try to produce the desired γ is simply to list all possible blocks in Σ_M, and use them sequentially to label subsets of IX of positive measure. This certainly makes γ M-compatible. However, the block A may now occur not only to code F, but also in the blocks in IX used for M-compatibility. This invalidates the coding argument which gives $F \in \gamma_T$, and so γ may not generate. We avoid this by coding the blocks of Σ_M into IX in the following special way.

Enumerate all finite blocks in Σ_M by a list $\{B_m : m \geqslant n\}$. Since L is a transition length for Σ_M, for each m there are blocks $D_{j,m}$, D'_m, and D''_m of length L such that

$$C_m = AD_{1,m}AD_{2,m}A \ldots AD_{m,m}D'_mB_mD''_mD_{m+1,m}A \ldots AD_{2m,m}A$$

occurs in Σ_M, and also such that A occurs in C_m either in B_m or in one of the $2m$ positions outside B_m already indicated, but nowhere else. Because of the spacers

D'_m and D''_m, the blocks C_m are all distinct, for regardless of how A occurs in B_m, C_m has exactly m copies of A preceding and following B_m with spacing L. Choose sets $G_m \subset F$, pairwise disjoint and of positive measure, such that

$$(*) \qquad r_F > 4|C_m| + n \quad \text{on} \quad G_m.$$

Let $\bar{G}_m = \cup_{i=0}^{|C_m|-1} T^{n+|C_m|+L+i} G_m$. Label the levels

$$\{T^i G_m : n + |C_m| + L \leqslant i \leqslant n + 2|C_m| + L - 1\}$$

of $\bar{G}_m$ with C_m, $m \geqslant n$. Fill in the rest of IX with $\bar{\Sigma}$-blocks in order to complete the definition of γ.

Since every block in Σ_M occurs in the stack with positive measure, γ is M-compatible.

The occurrence of A in the γ-name of a point x is due to one of three reasons: (1) A is coding F (i.e. this part of the orbit of x is in II or VI) or is in IV (used in later stages), (2) A is coding B_m (i.e. A is one of the $2m$ blocks in C_m outside of B_m), or (3) A occurs in a B_m. Because of (*) and the spacers D'_m and D''_m, the first half of C_m cannot be confused with the terminal part of another C_j, and the last half of C_m cannot be confused with the initial part of another C_j. Hence the occurrences of A in the γ-name of a point determine the occurrences of the blocks C_m. Also, since $m > n$, the occurrences of A due to (1) cannot be confused with those due to (2). Thus if C_m occurs at position i in the γ-name of x, it follows that either $T^i x \in T^{n+|C_m|+L} G_m \subset \bar{G}_m$, or that C_m is a subblock of some B_j, and $T^i x \in \bar{G}_j$; in either case $T^i x$ is in IX. This accounts for the occurrences of A due to reasons (2) and (3) above. The remaining A blocks are due to (1), and they determine F as before. Hence $F \in \gamma_T$, and this completes our argument. □

3. Markov Partitions for Toral Automorphisms.

Let $\mathbf{T}^2$ denote the two-dimensional torus $\mathbf{R}^2/\mathbf{Z}^2$, written additively. An algebraic automorphism V of $\mathbf{T}^2$ automatically preserves Lebesgue measure λ. We will henceforth assume that V is ergodic with respect to λ. Such a V can be represented as a 2×2 matrix which acts on $\mathbf{R}^2$ preserving $\mathbf{Z}^2$. Ergodicity of V means that, considered as a linear transformation of $\mathbf{R}^2$, V has a real eigenvalue λ_1 of modulus greater than 1, and another real eigenvalue λ_2 of modulus less than 1. Let L_1 and L_2 be the projection onto $\mathbf{T}^2$ of the eigenspaces of V in $\mathbf{R}^2$ corresponding to λ_1 and λ_2. Thus L_1 and L_2 each wrap densely around $\mathbf{T}^2$.

Ken Berg [4] discovered that V has a generator $\alpha = \{D_1, \ldots, D_r\}$ whose atoms are parallelograms, with respect to which V is a Markov shift. The associated transition matrix is $M = (M_{ij})$, where $M_{ij} = 1$ if and only if $\lambda(D_i \cap V^{-1} D_j) > 0$. There is a natural surjection $\Psi : \Sigma_M \to \mathbf{T}^2$ defined as follows. If $(x_i) \in \Sigma_M$, then $\cap_{i=-n}^{n} V^{-i} \bar{D}_{x_i}$ is a nonempty compact set of diameter bounded by a multiple of λ_1^{-n}. Thus $\cap_{-\infty}^{\infty} V^{-i} \bar{D}_{x_i} = \{y\}$, and we define $\Psi(x_i) = y$. Clearly $\Psi S = V \Psi$. It is also true that $\Psi(\lambda_M) = \lambda$. The boundaries of the D_i are contained in $L_1 \cup L_2$. Thus $\partial\alpha = \cup_{i=1}^{r} \partial D_i \subset L_1 \cup L_2$, the latter set being invariant under V.

If $\Psi^{-1}(y)$ has more than one element, then some $V^i y$ must lie in the overlapping boundaries of two atoms of α. Thus Ψ is injective off $\Psi^{-1}(\cup_{-\infty}^{\infty} V^i(\partial\alpha)) \subset \Psi^{-1}(L_1 \cup L_2)$.

Such Markov partitions were independently discovered by Adler and Weiss [1] and used to show that entropy classifies ergodic automorphisms of $\mathbf{T}^2$ up to measure theoretic isomorphism.

Suppose now that V is an ergodic automorphism of $\mathbf{T}^2$ with $h_\lambda(V) > h_\mu(T)$. Let α be the Markov partition for V described above, and M be the corresponding transition matrix. Then $h(M) = h_\lambda(V) > h_\mu(T)$. By **2** there is an M-compatible generator γ for T, which yields an isomorphism Φ of (T, X, μ) with (S, Σ_M, μ_M).

Now μ_M is mapped under Ψ to $\Psi(\mu_M) = \mu_2$ on $\mathbf{T}^2$, and μ_2 is invariant under V. Since Ψ is surjective, the support of μ_2 is all of $\mathbf{T}^2$. For Ψ to be an isomorphism on (S, Σ_M, μ_M), it is sufficient that Ψ be injective a.e. with respect to μ_M, i.e. that $\mu_M(\Psi^{-1}(L_1 \cup L_2)) = \mu_2(L_1 \cup L_2) = 0$. To show this, we first use the observation of Adler and Weiss [1, §10] that the restriction of V to $(L_1 \cup L_2) \setminus \{0\}$ is dissipative. Hence there is no finite positive measure there invariant under V, and so $\mu_2((L_1 \cup L_2) \setminus \{0\}) = 0$. Suppose that $\mu_2(\{0\}) > 0$. Then since $\mathbf{T}^2 \setminus \{0\}$ is open, $\Psi^{-1}(\{0\})$ and $\Psi^{-1}(\mathbf{T}^2 \setminus \{0\})$ would be disjoint S-invariant sets of positive μ_M measure. This contradicts the ergodicity of S on (Σ_M, μ_M) which results from the assumed ergodicity of T on (X, μ). The same kind of argument shows that μ_2 is nonatomic.

Thus the composition $\Psi\Phi$ is an isomorphism of (T, X, μ) with $(V, \mathbf{T}^2, \mu_2)$, where μ_2 is a nonatomic V-invariant measure which is positive on open sets.

4. The Oxtoby-Ulam Theorem.

We are now ready for the last piece of the proof. While constructing measure-preserving homeomorphisms of manifolds, Oxtoby and Ulam proved a result [16, Corollary 1] of which we need the following special case.

Theorem 3. *Suppose that μ is a probability measure on $\mathbf{T}^2$ which is nonatomic, and which assigns positive measure to nonempty open sets. Then there is a homeomorphism H of $\mathbf{T}^2$ such that $H(\mu) = \lambda$.*

We note that if μ is assumed to be given by an everywhere positive C^∞ density function $d\mu/d\lambda$, then Moser [15] proved that H could be taken to be a C^∞ diffeomorphism. However, in our case μ need not even be absolutely continuous with respect to λ.

As we indicated in **1**, composing the results of the last three sections proves Theorem 1.

5. Universal Homeomorphisms.

Let S denote the shift on the Markov shift space Σ_M. Krieger's generator theorem implies that if $h_\mu(T) < h(S)$, then there is an S-invariant measure ν such that (T, X, μ) is isomorphic to (S, Σ_M, ν). Indeed, our Theorem 2 shows that ν can be chosen to have full support. Also, the arguments in **3** prove that ergodic automorphisms of $\mathbf{T}^2$ share this property. We isolate this phenomenon in the following definition.

Definition. *A homeomorphism U of a compact metric space Y is called* universal *if for every ergodic transformation T of a Lebesgue space (X,μ) with $h_\mu(T)<h(U)$, there is a U-invariant Borel measure ν on Y with full support such that (T,X,μ) is isomorphic to (U,Y,ν).*

Universality is the opposite of unique ergodicity. Hahn and Katznelson [7] constructed examples of uniquely ergodic homeomorphisms with positive entropy, and these clearly are not universal. Indeed, Jewett's result [8] that every weakly mixing transformation has a uniquely ergodic model (extended to ergodic transformations by Krieger [11]) yields many examples of nonuniversal homeomorphisms.

There are several natural questions about universality which cannot be decided on the basis of known examples.

Samples:

(1) Does every ergodic transformation have a universal model (complementing Jewett's result)? Or is it the case that universal homeomorphisms automatically have a unique invariant measure of maximal entropy with respect to which they are Bernoulli?

(2) Is universality preserved under taking powers or factors? Benjamin Weiss has pointed out with the following example that in general direct products of universal systems are not universal. Let U be a universal homeomorphism of positive entropy and R be an irrational rotation of the circle (which is trivially universal). Then the projection of every $U\times R$-invariant measure ν onto the second coordinate must be R-invariant, and hence be Lebesgue measure. Thus $(U\times R,\nu)$ is not weakly mixing, so that $U\times R$ is not universal. However, it is not known whether the product of positive entropy universal homeomorphisms is universal.

(3) Are ergodic automorphisms of compact groups universal? In particular, are ergodic nonhyperbolic toral automorphisms universal?

(4) Is a compact group extension of a universal homeomorphism universal?

References

1. R. Adler and B. Weiss, Similarity of automorphisms of the torus, *Memoirs of the Amer. Math. Soc.* **98** (1970).
2. D. V. Anosov and A. B. Katok, New examples in smooth ergodic theory. Ergodic diffeomorphisms, *Trans. Moscow Math. Soc., AMS Transl.* **23** (1972), 1–35.
3. V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, Benjamin, New York, 1968.
4. K. Berg, *On the conjugacy problem for K-systems*, Ph.D. Thesis, University of Minnesota, Minneapolis, Minn., 1967.
5. M. Denker, C. Grillenberger, and K. Sigmund, Ergodic theory on compact spaces, *Springer Lecture Notes in Mathematics* **527** (1976).

6. L. Goodwyn, Topological entropy bounds measure theoretic entropy, *Proc. Amer. Math. Soc.* **23** (1969), 679–688.
7. F. Hahn and Y. Katznelson, On the entropy of uniquely ergodic transformations, *Trans. Amer. Math. Soc.* **126** (1967), 335–360.
8. R. I. Jewett, The prevalence of uniquely ergodic systems, *J. Math. Mech.* **19** (1970), 717–729.
9. A. N. Kolmogorov, On dynamical systems with an integral invariant on the torus, *Dokl. Akad. Nauk SSSR (N.S.)* **93** (1953), 163–766 (Russian).
10. A. N. Kolmogorov, General theory of dynamical systems and classical mechanics, *Proc. 1954 Intern. Congr. Math.*, North-Holland, Amsterdam, 1957, 315–333 (Russian).
11. W. Krieger, On unique ergodicity, *Proc. of the Sixth Berkeley Symp. on Math. Stat. and Prob.*, University of Calif. Press I (1972), 327–346.
12. W. Krieger, On generators in ergodic theory, *Proc. Intern. Congr. Math.* (1975) **2**, 303–308.
13. A. G. Kushnirenko, An estimate from above for the entropy of a classical system, *Dokl. Akad. Nauk SSSR* **161** (1975), 37–38; trans. *Sov. Math. Dokl.* **6** (1965), 360–362.
14. D. A. Lind, Spectral invariants and smooth ergodic theory, *Springer Lecture Notes in Physics* **38**, 296–308.
15. J. Moser, On the volume elements on a manifold, *Trans. Amer. Math. Soc.* **120** (1965), 286–294.
16. J. C. Oxtoby and S. M. Ulam, Measure-preserving homeomorphisms and metrical transitivity, *Ann. of Math.* (2) **42** (1941), 874–920.
17. W. Parry, Intrinsic Markov chains, *Trans. Amer. Math. Soc.* **112** (1964), 55–66.

Received Sept. 14, 1977.