On notions of determinism in topological dynamics

MICHAEL HOCHMAN

Mathematics Department, Princeton University, Fine Hall, Washington Road,
Princeton, NJ 08544, USA
(e-mail: hochman@math.princeton.edu)

Abstract. We examine the relations between topological entropy, invertibility, and prediction in topological dynamics. We show that topological determinism in the sense of Kamiński, Siemaszko, and Szymański imposes no restriction on invariant measures except zero entropy. Also, we develop a new method for relating topological determinism and zero entropy, and apply it to obtain a multidimensional analog of this theory. We examine prediction in symbolic dynamics and show that while the condition that each past admits a unique future only occurs in finite systems, the condition that each past has a bounded number of futures imposes no restriction on invariant measures except zero entropy. Finally, we give a negative answer to a question of Eli Glasner by constructing a zero-entropy system with a globally supported ergodic measure in which every point has multiple preimages.

1. Introduction

There are several ways to define 'determinism' of a dynamical system, all of which express the idea that the past determines the future (and *vice versa*). In ergodic theory, a measure-preserving map T of a probability space $(X, \mathcal{B}, \mu)$ is deterministic if, for every measurable $f : X \to \mathbb{R}$ (or, equivalently, every finite-valued f), the sequence $f(Tx), f(T^2x), \ldots$ determines $f(x)$ with probability one, that is, $f \in \sigma(Tf, T^2f, \ldots)$, where $\sigma(\mathcal{F})$ is the σ-algebra generated by $\mathcal{F}$. Another equivalent condition is that every factor $(Y, \mathcal{C}, \nu, S)$ of $(X, \mathcal{B}, \mu, T)$ is essentially invertible, i.e. there is an invariant set $Y_0 \subseteq Y$ of full measure such that $S|_{Y_0}$ is invertible. Yet another equivalent condition which is widely used is that entropy vanishes: $h(T, \mu) = 0$.

In this work, we examine the relations between prediction, invertibility, and entropy in the category of topological dynamics, where by a topological dynamical system (X, T) we mean a continuous onto map $T : X \to X$ of a compact metric space. One can find analogs of these three conditions, but the relations between them are more complex. We present here several results that underscore the independence of these notions, complementing some of the recent works on the subject, e.g. [**3, 5, 9**].

1.1. *Topological predictability.* Kamiński *et al* introduced in [**6**] an interesting and natural notion of predictability for topological systems. A system (X, T) is *topologically predictable*†, or TP, if, for every continuous function $f \in C(X)$, we have $f \in \langle 1, Tf, T^2 f, \ldots \rangle$, where $\langle \mathcal{F} \rangle \subseteq C(X)$ denotes the closed algebra generated by a family $\mathcal{F} \subseteq C(X)$. Kamiński *et al* showed that (X, T) is topologically predictable if and only if every factor of (X, T) is invertible, where a factor is a system (Y, S) and a continuous onto map $\pi : X \to Y$ such that $\pi T = S\pi$.

One would like to understand what other dynamical implications topological predictability has. In [**7**], it was shown that TP systems have zero topological entropy (correcting a gap, as the authors note, in their earlier proof from [**6**]), but the converse to this is false. Indeed, every TP system on a totally disconnected space is equicontinuous, whereas every zero-entropy measure can be realized as an invariant measure on a totally disconnected space (and, hence, for measures with irrational or continuous spectrum, not TP).

Nonetheless, although 'not TP' seems to say little about the invariant measures, TP is a rather strong condition, and one might suppose it to impose restrictions on the measurable dynamics. In previous work on the subject, the main tool used to establish that a system is TP was the fact that, if every point in the product $(X \times X, T \times T)$ is forward recurrent, then (X, T) is TP. Consequently, distal systems and the pointwise rigid systems are TP; but no others were known.

Our first result, which may be of independent interest, is that TP imposes no restrictions on invariant measures except zero entropy.

THEOREM 1.1. *For every zero-entropy, ergodic measure-preserving system* $(X, \mathcal{B}, \mu, T)$ *there are a topological system* (Y, S) *and an invariant measure* ν *on* Y *such that* $(Y, \nu, S) \cong (X, \mathcal{B}, \mu, T)$ *and, for every* y', y'' *in* Y, *the point* (y', y'') *is forward recurrent for* $S \times S$. *In particular,* (Y, S) *is TP.*

This construction is related to the construction in Weiss [**12**]. For any zero-entropy measure-preserving system, that construction produces, as a by-product, a topological model in which every pair is two-sided recurrent in the product system. However, that is a far weaker statement than forward recurrence. In fact, the realization in [**12**] is on a subshift, which is totally disconnected, and one cannot hope that such a system will be TP (for then the action would be equicontinuous, and the invariant measures would have pure point spectrum).

As a consequence of Theorem 1.1, one gets a new functional characterization of the vanishing of entropy in measure-preserving systems.

COROLLARY 1.2. *A measure-preserving system* $(X, \mathcal{B}, \mu, T)$ *has entropy* 0 *if and only if there exists a separable subalgebra* $\mathcal{A} \subseteq L^\infty(\mu)$ *which separates points and such that* $f \in \langle 1, Tf, T^2 f, \ldots \rangle$ *for every* $f \in \mathcal{A}$.

Next, we discuss the notion of TP for $\mathbb{Z}^d$ actions. Such an action $\{T^u\}_{u \in \mathbb{Z}^d}$ of $\mathbb{Z}^d$ by homeomorphisms on X is topologically predictable (TP) if $f \in \langle 1, T^u f : u < 0 \rangle$ for every $f \in C(X)$; here $<$ is the lexicographical ordering on $\mathbb{Z}^d$. One can also work with other

---

† Kamiński *et al* use the term topological determinism, but this seems to us confusing in the present context.

orderings, e.g. lexicographic orderings with respect to other coordinate systems. One may ask whether this notion is independent of the generators (the lexicographic ordering certainly is not). It is not; even in dimension one, the property TP depends on the generator, i.e. TP for T does not imply it for T^{-1}. Thus, TP is a property of a group action and a given set of generators.

The proof in [**6, 7**] that TP implies zero entropy for a single transformation used the non-trivial theory of extreme partitions and entropy pairs. In §3.2, we give a new and direct argument for this implication, which is somewhat more transparent. Furthermore, our proof can be used to generalize the result to actions of $\mathbb{Z}^d$.

THEOREM 1.3. *For a $\mathbb{Z}^d$ action, TP implies zero topological entropy.*

There is a rather complete theory of entropy, developed by Ornstein and Weiss, for actions of amenable groups on probability spaces. One feature which is absent from the general theory (and which we utilized for $\mathbb{Z}$ and $\mathbb{Z}^d$ actions) is a good notion of the 'past' of an action, and the ability to represent the entropy of a partition as a conditional entropy of the partition with respect to the 'past'. However, by analogy to the abelian case, the following question is natural.

Problem 1.4. Suppose that an infinite discrete amenable group G acts by homeomorphisms on X. Let $S \subseteq G$ be a subsemigroup not containing the unit of G, and such that $S \cup S^{-1}$ generates G. Suppose that for every $f \in C(X)$, we have $f \in \langle 1, sf : s \in S \rangle$. Does this imply that $h(X, G) = 0$?

1.2. *Prediction for symbolic systems.* Let Σ be a finite set of symbols and consider the space $\Sigma^{\mathbb{Z}}$ of bi-infinite sequences over Σ. Denote by $\sigma : \Sigma^{\mathbb{Z}} \to \Sigma^{\mathbb{Z}}$ the shift map. A symbolic system is a closed, non-empty, σ-invariant subset of $\Sigma^{\mathbb{Z}}$.

Let $X \subseteq \Sigma^{\mathbb{Z}}$ be a subshift and let $x^- \in \Sigma^{-\mathbb{N}}$, where $\mathbb{N} = \{1, 2, 3, \ldots\}$; for $x \in \Sigma^{\mathbb{Z}}$, we also write $x^- = x|_{-\mathbb{N}}$. A finite or infinite sequence $x^+ \in \bigcup_{0 \leq n \leq \infty} \Sigma^n$ is an admissible extension of x^- (with respect to X) if the concatenation $x^- x^+$ is in X. If $h(X) = 0$, then $h(\mu) = 0$ for every invariant measure μ on X, and so there is a set of points $X_0 \subseteq X$, having full measure with respect to every invariant measure, such that x^- has a unique extension for every $x \in X_0$; that is, if $y \in X_0$ is another point, then $y^- = x^-$ implies that $x = y$. A natural question is whether this can occur for every $x, y \in X$. The answer is no: in fact, it is well known that the only subshifts for which every admissible past x^- admits a unique continuation are finite unions of periodic orbits (we give a proof in Lemma 4.1).

However, there do exist subshifts where each $x^- \in \Sigma^{-\mathbb{N}}$ has only finitely many extensions; the best known are probably the Sturmian subshifts. Such subshifts must have zero entropy. It turns out that such systems are not uncommon, and that entropy is again the only restriction to the dynamics of their invariant measures.

THEOREM 1.5. *Every ergodic measure-preserving system with entropy zero is isomorphic to a shift-invariant Borel measure on a uniquely ergodic subshift $X \subseteq \{0, 1\}^{\mathbb{Z}}$ with the property that every $x^- \in \{0, 1\}^{-\mathbb{N}}$ has at most two infinite extensions.*

This may be viewed as a sharpening of the Jewett–Krieger generator theorem, which states that every measure-preserving system with finite entropy h can be realized as the

unique invariant measure on a uniquely ergodic subshift on k symbols, provided that $\log k > h$. In zero entropy, one cannot use fewer than two symbols. This theorem says that one can do the next best thing.

1.3. *Non-invertibility and entropy.*   Consider a symbolic system $X \subseteq \Sigma^{\mathbb{N}}$ (note that we now have a one-sided shift) and an invariant probability measure μ on X. Recall that, since the partition of X according to the first symbol generates the σ-algebra, the entropy $h(\mu)$ is the average of the entropy of the conditional measures, given x, induced on the preimage set $\sigma^{-1}(x)$. Thus, if $h(\mu) > 0$, then with positive probability $\sigma^{-1}(x)$ is not concentrated on a single point, and consequently there is a large set of points in X with multiple preimages. It is therefore natural to ask what 'degree' of non-invertibility is necessary to guarantee positive entropy.

One plausible condition is that each point have multiple preimages; we call such a system *everywhere non-invertible*. Indeed, for subshifts this is enough to imply positive entropy, because, for symbolic systems, everywhere non-invertibility implies a stronger condition: the preimage of every point has diameter $> \delta$ for some $\delta > 0$. Whenever this condition is satisfied, we say that the system has *no small preimages.* An easy argument shows that a map with no small preimages has entropy at least $\log 2$ (see Proposition 5.1 below).

Everywhere non-invertibility does not guarantee positive entropy in general, though in some special cases it does, e.g. maps of the interval [**1**]. One would hope to find additional hypotheses, which, together with everywhere non-invertibility, imply positive entropy. One candidate is the presence of a globally supported ergodic measure. In an everywhere non-invertible system there is always an open set of points whose preimages have diameter which is bounded below by some positive constant and, when there is a globally supported ergodic measure, almost every orbit spends a positive fraction of its time in this set. One would hope to use this fact to construct many well-separated orbits. Eli Glasner has raised the question of whether this hypothesis indeed implies positive entropy. We show that it does not.

Example 1.6. There exists a zero-entropy, everywhere non-invertible system with a globally supported ergodic measure.

For an integer $k > 0$, we say that a system (X, T) is at least *k-to-one* if the preimage set of every point is of size at least k. Bobok has shown that if a map of the circle (or the interval) is k-to-one, then $h(T) \geq \log k$, and has asked if this holds in general, at least under the assumption that there are no small preimages. We can give a negative answer to this question.

Example 1.7. There exists an infinite-to-one system (X, T) with no small preimages, and which supports a global ergodic invariant measure, but $h(X, T) = \log 2$.

There seems to be no obstruction in our examples to making the measures weakly mixing, and possibly strongly mixing, but we do not pursue this here.

The question remains whether such examples exist for a continuous map on a manifold. For smooth maps they do not; see [**2**].

2. Notation

We will use freely standard facts about topological dynamics and entropy which can be found e.g. in [**11**]. This section contains some further notation for dealing with sequence spaces.

Let Σ be a set and write Σ^* for the set of all finite words over Σ. The ith letter of a word $a \in \Sigma^*$ is denoted by $a(i)$. If $a = a(1)a(2) \cdots a(k)$, then k is the length of a and is denoted by $\ell(a)$. We denote concatenation of the words $a, b \in \Sigma^*$ by ab.

Similarly, we define the spaces of one-sided sequences $\Sigma^{\mathbb{N}}, \Sigma^{-\mathbb{N}}$ (we use the convention $\mathbb{N} = \{1, 2, 3, \ldots\}$) and of two-sided sequences $\Sigma^{\mathbb{Z}}$. If a topology is given on Σ, these sequence spaces carry the product topology; for finite Σ, we take the discrete topology for Σ. We denote by σ the shift map on both these spaces, which is defined by the formula $(\sigma(x))(i) = x(i + 1)$; this map restricted to $\Sigma^{\mathbb{N}}$ and $\Sigma^{\mathbb{Z}}$ is continuous and onto, and is a homeomorphism in the two-sided case. In the one-sided case, the preimage set of every point is identified with Σ. We also define the shift on Σ^* in the obvious way, by

$$\sigma(x(1)x(2) \cdots x(k)) = x(2)x(3) \cdots x(k)$$

(note that $\sigma^n(ab) = (\sigma^n a)b$ if $n \leq \ell(a)$ but is equal to $\sigma^{n-\ell(a)}(b)$ if $\ell(a) < n \leq \ell(a) + \ell(b)$; otherwise it is the empty word). When concatenating infinite sequences, we adopt the convention that, if $x \in \Sigma^{-\mathbb{N}}$ and $y \in \Sigma^{\mathbb{N}}$, then $xy \in \Sigma^{\mathbb{Z}}$ is the sequence z with $z(i) = x(i)$ for $i < 0$ and $z(i) = y(i + 1)$ for $i \geq 0$ (note that $0 \notin \mathbb{N}$, which is the reason for this shift of y).

For a word x (finite or infinite), if $x = ab$, then a is called a *front segment* of x (if $\ell(a) = k$, then a is a front k-segment of x), and b a *back segment* of x. For $a, b \in \Sigma^*$, we say that a is a subword of b at index i if $i \leq \ell(b) - \ell(a) + 1$ and $a(j) = b(i + j)$ for $j = 1, \ldots, \ell(a)$. The index i is called the *alignment* of a in b. If such an i exists, we say that a appears in b, or that it is a subword of b.

We denote by $[i; j]$ the segment of consecutive integers $[i, j] \cap \mathbb{Z}$, and denote by $x|_{[i; j]} = x(i)x(i + 1) \cdots x(j)$ the subword of x determined by $[i; j]$, provided that x is long enough for this to make sense.

All measures are assumed to be Borel probability measures.

3. Topological predictability

3.1. Realization of measures on TP systems.
A topologically predictable system has zero topological entropy and, therefore, by the variational principle, every invariant measure on it has entropy zero. In this section, we prove Theorem 1.1, showing that this is the only restriction on invariant measures. The construction is rather technical. We remark that this section is not used in the following sections.

A point x in a dynamical system (X, T) is forward recurrent if $T^{n(k)}x \to x$ for some sequence of times $n(k) \to \infty$. Note that if every point in a system is forward recurrent, then every closed subset $A \subseteq X$ which is forward invariant, i.e. $TA \subseteq A$, is invariant, i.e. $T^{-1}A = TA = A$.

In order to construct a TP system supporting a given measure, we shall construct an isomorphic measure on a topological system (X, T) for which every point in $X \times X$ is forward recurrent. Indeed, by the remark above, this implies that every forward-invariant,

closed equivalence $R \subseteq X \times X$ is also invariant under T^{-1}, and this is equivalent to the property that every factor is invertible, so (X, T) is topologically predictable [6]. Our construction cannot be symbolic, since infinite symbolic systems always contain forward-asymptotic pairs. We shall instead construct a connected subshift of $[0, 1]^{\mathbb{N}}$.

Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system with zero entropy. We wish to construct a space Y and a homeomorphism $S : Y \to Y$ for which every pair is forward recurrent and which support a measure isomorphic to $(X, \mathcal{B}, \mu, T)$.

For the construction, we may assume, by e.g. [12], that T is a minimal, topologically weakly mixing, strictly ergodic homeomorphism of a totally disconnected metric space X, and that there exists a clopen generator for T.

Given a measurable function $f : X \to [0, 1]$, let $f^{(m)} : X \to [0, 1]^m$ denote the function

$$x \mapsto (f(x), f(Tx), \ldots, f(T^{m-1}x)),$$

and similarly let $f^{(\infty)} : X \to [0, 1]^{\mathbb{N}}$ denote the map

$$x \mapsto (f(x), f(Tx), f(T^2x), \ldots).$$

We use the notation $\|a\|_{\infty} = \sup |a(i)|$ for $a \in \mathbb{R}^m$ or $a \in \mathbb{R}^{\mathbb{N}}$.

For integers m, r, we say that f is (m, r)-good if there is a subset $X_{f,m,r} \subseteq X$ of full measure such that, for every $x', x'' \in X_{f,m,r}$, there is an integer $0 < k < r$ (which may depend on x', x'') satisfying

$$\| f^{(m)}(x') - f^{(m)}(T^k x') \|_{\infty} < \frac{1}{m},$$

$$\| f^{(m)}(x'') - f^{(m)}(T^k x'') \|_{\infty} < \frac{1}{m}.$$

Suppose that f is $(m, r(m))$-good for some sequence $r(m)$. Setting $X_0 = \bigcap_{m=1}^{\infty} X_{f,m,r(m)}$, the relation above holds for every $x', x'' \in X_0$ and all $m \in \mathbb{N}$. If we set $\nu = f^{(\infty)}\mu$ and $Y = \operatorname{supp} \nu \subseteq [0, 1]^{\mathbb{N}}$, it follows that each pair of points in Y is forward recurrent for the shift σ. Also, ν is shift invariant on (Y, σ), $f^{(\infty)}$ is a factor map from X to Y, and, if the partition induced by f on X generates for T, then this is an isomorphism. Thus, the theorem will follow once we construct a function f as above.

We construct f by approximation. More specifically, we define a sequence of functions $f_n : X \to [0, 1]$ and integers $r(n)$ such that f_n is $(m, r(m))$-good for each $m \leq n$. The sequence f_n will be constructed so that it converges almost everywhere to a function f, which is clearly $(m, r(m))$-good for $m \in \mathbb{N}$. Also, each f_n will generate for T and we will guarantee that f generates by controlling the speed of convergence of f_n to f. The f_n's will be continuous and each will take on only finitely many values, so we may identify them with finite partitions P_n of X into clopen sets, where $f_n(x) = i$ if and only if x is in the partition element of P_n indexed by i (we allow i to take non-integer values).

The construction proceeds by induction. Our induction hypothesis will be that we are given a function f_n arising from a finite clopen generating partition P_n, and integers $r(1), \ldots, r(n)$, such that f_n is $(m, r(m))$-good for $m = 1, \ldots, n$. For any ε, we will show how to define f_{n+1} and $r(n + 1)$ satisfying the same condition with $n + 1$ in place of n, and such that

$$\mu(x \in X : f_n(x) \neq f_{n+1}(x)) < \varepsilon.$$

By choosing $\varepsilon = \varepsilon(n)$ to decrease rapidly enough, this last condition guarantees that $f_n \to f$ almost surely, and that f generates for T (see e.g. [**10**]).

Suppose then that we are given f_n, $r(1), \ldots, r(n)$, and $\varepsilon > 0$ as above. First, note that the properties of these objects are completely determined by the itineraries of length $r(n) + n$ associated under f_n to points in X, i.e. by the image of $f_n^{(r(n)+n)}$. The following lemma, whose proof we omit, says that the desired properties of the blocks continue to hold if we modify itineraries in a sufficiently slow way.

LEMMA 3.1. *For f_n, P_n, $r(1), \ldots, r(n)$ as above, there is a number $0 < \rho < 1/(n+1)$ with the following property. Suppose that $y', y'' \in [0, 1]^{r(n)+n}$ are blocks appearing in $f_n^{(\infty)}(X)$ and $\alpha', \alpha'' \in [0, 1]^{r(n)+n}$ have the property that $|\alpha'(i) - \alpha'(i+1)| < \rho$ and $|\alpha''(i) - \alpha''(i+1)| < \rho$ for all $1 \le i \le r(n) + n - 1$. Define $z', z'' \in [0, 1]^{r(n)+n}$ by $z'(i) = \alpha'(i) \cdot y'(i)$ and $z''(i) = \alpha''(i) \cdot y''(i)$. Then there exists $0 < k \le r(m)$ with $|z'(i) - z'(i+k)| < 1/m$ and $|z''(i) - z''(i+k)| < 1/m$ for $i = 1, 2, \ldots, n$.*

Let $Y \subseteq [0, 1]^{\mathbb{N}}$ be the symbolic subshift defined by the property that every block of length $r(n) + n$ in Y appears in $f_n^{(\infty)}(X)$. Note that Y is a shift of finite type and is irreducible because X is topologically mixing. In particular, there is an integer D such that given two blocks a, c appearing in Y, there is a block b_k for every $k \ge D$ such that ab_kc appears in Y. We can also fix a block a^* appearing in Y which contains a copy of every n-block in Y. Increasing D or lengthening a^* if necessary, we may assume that $D > 1/\varepsilon$ and that a^* is of length D.

We need the following, which is a specialized version of Lemma 2 from [**12**].

LEMMA 3.2. *There exist $\delta > 0$ and $T_0 \in \mathbb{N}$ such that, for all $T \ge T_0$, there is a family I of subsets of $\{0, \ldots, T - 1\}$ satisfying:*
(1) $|I| \ge 2^{\delta T}$;
(2)   *for $A \in I$ and distinct $u, v \in A$, we have $|u - v| \ge 10D/\varepsilon$;*
(3)   *for each $A, B \in I$ and $k \le 9T/10$, we have $A \cap (B + k) \ne \emptyset$.*

We use the lemma in conjunction with the following simple fact.

LEMMA 3.3. *Fix T and let $A, B \subseteq \{0, 1, \ldots, T\}$ satisfy the three conditions of the previous lemma. Fix $0 \le k \le (9T/10) - n$, and let $z', z'' \in [0, 1]^{\mathbb{N}}$ be such that a^* appears in z' at each index $i \in A$ and in z'' at each index $j \in B + k$. Then, for every pair a, b of n-blocks from Y, there is an index u such that a appears in z' at u, and b appears in z'' at u.*

Let ρ, δ, T_0 be as in the preceding lemmas. Since (X, T, μ) has zero topological entropy, it follows that we can choose an integer $H \ge (10/\rho\varepsilon)T_0$ large enough so that $2^{\delta(\varepsilon\rho/10)H}$ is greater than the number of (P_n, H)-names in X. We fix such an integer H and construct an Alpern tower [**4**] over some clopen set $B \subseteq X$, with columns of heights H and $H + 1$. This means that every point in B returns to B for the first time after either H or $H + 1$ applications of T. The ith level of the tower is the set of points $T^i B \setminus B$, and the disjoint union of these levels for $0 \le i \le H + 1$ is all of X. The last property can be obtained because (X, T) is minimal. This is a standard modification of the construction

of Alpern towers: one begins the construction with a clopen set and notes that, due of minimality, all points eventually return to it.

Purify the columns according to P_n, and let $B_1 \cdots B_N$ be the bases of the purified columns. Thus, $\{B_1, \ldots, B_N\}$ is a clopen partition of B which refines the partition according to return time and, if $h(i)$ denotes the height of the column over B_i, then all $x \in B_i$ have the same P_n-itinerary up to time $h(i)$, and these itineraries are distinct for different i. Note that the P_n-name of each column appears in Y.

Divide each column into $10/\varepsilon\rho$ blocks of length $(\varepsilon\rho/10)H$ (which we assume for convenience is an integer), and possibly an additional level in those columns which are of height $H + 1$. We proceed to modify P_n as follows.

- In each column, rename the bottom $1 + (1/\rho)$ blocks so that they are identical, and similarly for the top $1 + (1/\rho)$ blocks; and do so in such a way that the name of the entire column is admissible for Y. This can be done because $(\rho\varepsilon/10)H$, the length of each block, is much larger than D. Notice that by choice of ρ, the first and last $n + 1$ blocks in each column are identical.

- To each block, except the top and bottom n blocks of each column, assign a distinct set $A \subseteq \{0, \ldots, (\varepsilon\rho/10)H - 1\}$ such that $|u - v| \geq 10D/\varepsilon$ for distinct $u, v \in A$ and, if A, B are assigned to distinct blocks and $1/10 \cdot (\varepsilon\rho/10)H \leq k \leq 9/10 \cdot (\varepsilon\rho/10)H$, then $A \cap (B + k) \neq \emptyset$. We can do this by the choice of H and the lemma. To the bottom n blocks in each column assign the same set A which is assigned to the $(n + 1)$th block of that column, and similarly to the top n blocks assign the same set which is assigned to the $(n + 1)$th block from the top. We have thus assigned a set to each block.

- For a block b appearing in one of the columns and the set A associated to it, we modify b as follows. For convenience, in this paragraph we renumber the coordinates of b from 0 to $(10/\varepsilon\rho) - 1$, no matter where in the column b actually appears. For each $i \in A$, we replace the block of length D in b starting at i with the block a^*. Next, modify the symbols from $i - D$ to $i - 1$ and from $i + D$ to $i + 2D - 1$ in such a way that the entire block from $i - 2D$ to $i + 3D$ appears in Y; we can do this by the definition of D. All in all, we have changed b from index $i - D$ to index $i + 2D - 1$. Because of the distance between successive elements of A, these changes for different $i \in A$ occur at different places in b and the changes do not interfere with each other.

  Note that the bottom $n + 1$ blocks of each column are still identical, as are the $n + 1$ top blocks.

  Denote by $\tilde{P}_{n+1}$ the partition obtained so far, and by $\tilde{f}_{n+1}$ the corresponding function.

- If $b_1, b_2, \ldots, b_{1/\rho}$ are the bottom $1/\rho$ blocks of some column, replace b_k with $(k - 1)\rho \cdot b_k$, where $\alpha \cdot b_i$ is the block obtained by multiplying each coordinate of b_i by α. Similarly, if $c_1, c_2, \ldots, c_{1/\rho}$ are the top n blocks of a column, replace c_k with $(1/\rho - k)\rho c_k$.

- For columns of height $H + 1$, replace the top symbol with 0.

- Perturb the first symbol of each column by less than ε in such a way that the name of each column is unique.

Let f_{n+1} be the function defined by the revised partition; we claim that it has the desired properties for some integer $r(n + 1)$.

We first estimate the measure of points on which f_n and f_{n+1} differ. It suffices to show that in each column the fraction of levels modified is less than ε. The change to the top and bottom $1/\rho$ blocks amounts to $2/\rho$ blocks out of $10/\varepsilon\rho$, which is $\varepsilon/5$ of the levels. Consider now the intermediate levels. Since in the sets A associated to the blocks the distance between elements is at least $10D/\varepsilon$, and each element causes a change of $3D$ symbols to its block, here too we have caused a change to at most a $(3\varepsilon/10)$ fraction of the levels. The change to the top symbol of columns of height $H + 1$ amounts to less than $1/H$ of the space. Thus, we have indeed modified f_n on a set of measure less than ε.

We now show that we can choose $r(n + 1)$ so that f_{n+1} is $(m, r(m))$-good for each $m \le n + 1$. Note that every block in $f_{n+1}^{(\infty)}(X)$ of length $r(n) + n$ is of the form described in Lemma 3.1, so for $m \le n$ the conclusion follows immediately from that lemma.

We must show that f_{n+1} is $(n + 1, r)$-good for some r. Let $x', x'' \in X$. We must show that there is a k of bounded size such that

$$\|f_{n+1}^{(n+1)}(x') - f_{n+1}^{(n+1)}(\sigma^k x')\|_\infty < \frac{1}{n + 1}$$

and

$$\|f_{n+1}^{(n+1)}(x'') - f_{n+1}^{(n+1)}(\sigma^k x'')\|_\infty < \frac{1}{n + 1}.$$

Denote $y' = \tilde{f}_{n+1}(x')$ and $y'' = \tilde{f}_{n+1}(x'')$, and also $z' = f_{n+1}^{(n+1)}(x')$ and $z'' = f_{n+1}^{(n+1)}(x'')$. We distinguish several cases.

Case 1. Both x', x'' are in the top block or level $H + 1$ of their respective columns. Then the first $10/\rho\varepsilon$ symbols of $f_{n+1}^{(\infty)}(x')$, $f_{n+1}^{(\infty)}(x'')$ are 0, and the conclusion holds for $k = 1$.

Case 2. Exactly one of the points, say x', is in the top block or level $H + 1$ of its column, so the first $10/\rho\varepsilon$ symbols of $f_{n+1}^{(\infty)}(x')$ are 0. Note that in $y'' = \tilde{f}_{n+1}(x'')$ the block a^* appears somewhere between index 1 and $10/\rho\varepsilon$; hence, there is a $0 < k \le 10/\rho\varepsilon$ with

$$\|\tilde{f}_{n+1}^{(n+1)}(x'') - \tilde{f}_{n+1}^{(n+1)}(\sigma^k x'')\|_\infty = 0.$$

If we replace $\tilde{f}_{n+1}^{(n+1)}$ with $f_{n+1}^{(n+1)}$, the left-hand side changes by at most ρ and we get

$$\|f_{n+1}^{(n+1)}(x'') - f_{n+1}^{(n+1)}(\sigma^k x'')\|_\infty < \rho.$$

On the other hand,

$$\|f_{n+1}^{(n+1)}(x') - f_{n+1}^{(n+1)}(\sigma^k x')\|_\infty = 0,$$

because the first $10/r\varepsilon$ symbols of the itinerary of x' are 0, as desired.

Case 3. x', x'' are in different columns or the same column but at least $1/9 \cdot 10/\varepsilon\rho$ levels apart, and neither is in the top block or top level. By looking at the blocks to which x', x'' belong and at the next block, by Lemma 3.3 we see that for every pair of $n + 1$-blocks, and in particular the one appearing at the start of the itineraries of x', x'', there is a k in the range we want such that these blocks appear again in the $\tilde{f}_{n+1}$ itinerary of both x' and x'' at index k. As in Case 2, this gives the conclusion for the f_{n+1} itinerary, because the change from $\tilde{f}_{n+1}$ to f_{n+1} is 'too slow' to affect the inequality very much.

Case 4. x', x'' belong to the same column and are within $1/9 \cdot 10/\varepsilon\rho$ levels of each other. If they are in one of the bottom $1/\rho$ levels, then we are done by the periodicity of these blocks (again, there is some slow 'drift', which does not affect us). Otherwise, the initial $n+1$-block of both itineraries belongs to Y. We claim that there is an M such that either for some $0 < i < M$ the points $T^i x'$, $T^i x''$ belong to different columns but not to the top or bottom $1/\rho$ blocks of those columns, or else there exists a $k < M$, as desired. This suffices because in the former case we can argue as in Case 3, and deduce that as k ranges over $1, \dots, M + 10/\rho\varepsilon$, every pair of $n+1$-blocks from Y appears at index k in the f_{n+1}-itineraries of x', x''. This gives the conclusion we want.

It remains to show that there is such an M. This follows from the fact that $f_{n+1}^{(\infty)}(X)$ is a minimal symbolic system. Indeed, suppose the contrary. Then for every M there exist points $x'_M, x''_M \in X$ such that whenever $1 \leq i \leq M$ and $T^i x'_M$, $T^i x''_M$ are in different columns it is because they are within $10/\varepsilon\rho$ of the top or bottom of a column, and also the initial $n+1$ blocks of the itineraries of x', x'' do not appear again together before time M. We may assume that $x'_M \to x'$ and $x''_M \to x''$. Now x', x'' have these properties as well, for all M. Assuming as we may that x' is above x'' in the column they belong to, it follows that the itinerary of x' is a shift of the itinerary of x'', so the pair $(f_{n+1}^{(n+1)}(x'), f_{n+1}^{(n+1)}(x'')) \in f_{n+1}^{(n+1)}(X)$ is of the form $(y, T^r y)$ for some $r \leq 1/9 \cdot 10/\rho\varepsilon$. But since $f_{n+1}^{(\infty)}(X)$ is minimal this point must be recurrent, a contradiction. This completes the proof of Theorem 1.1. $\qquad\square$

Notice that the construction has introduced a fixed point $000\dots$ in the resulting subshift. We do not know if this can be avoided; more specifically, we do not know if the subshift can be made to be minimal.

3.2. *Partitions derived from continuous functions and predictable $\mathbb{Z}^d$ actions.* In this section, we prove a purely measure-theoretic and topological lemma which involves no dynamics. Let X be a normal topological space and μ a regular probability measure on the Borel σ-algebra of X. The entropy and conditional entropy of finite and countable partitions are defined as usual [11]. For finite or countable measurable partitions $\mathcal{P} = (P_1, P_2, \dots)$ and $\mathcal{Q} = (Q_1, Q_2, \dots)$ of X with finite entropy, the Rohlin metric is defined by
$$d(\mathcal{P}, \mathcal{Q}) = H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{P}).$$
This metric has the property that if $\mathcal{P} = (P_1, P_2, \dots)$ and we define
$$\mathcal{P}^{(n)} = \left(P_1, \dots, P_n, \bigcup_{k=n+1}^{\infty} P_k \right),$$
then $\mathcal{P}^{(n)} \to \mathcal{P}$ in d.

We say that a partition $\mathcal{P}$ is *continuous* if there is a continuous function $f \in C(X)$ which is constant almost surely on each atom of P_i. Equivalently, $\mathcal{P}$ agrees with the partition of X into level sets of some $f \in C(X)$, up to measure zero.

PROPOSITION 3.4. *The continuous partitions are dense with respect to the Rohlin metric in the space of finite-entropy countable partitions.*

Proof. The proof is a variation on Urisohn's lemma, which states that given two closed disjoint sets C_0, C_1 in a normal space, there is a continuous function $0 \leq f \leq 1$ such that $f^{-1}(0) = C_0$ and $f^{-1}(1) = C_1$.

Let $\mathbb{D} \subseteq \mathbb{Q} \cap [0, 1]$ denote the dyadic rationals. Let $\mathcal{P} = (P_0, P_1)$ be a partition into two sets and let $\varepsilon > 0$. We construct a continuous function $f : X \to [0, 1]$ with $\mu(\bigcup_{r \in \mathbb{D}} f^{-1}(r)) = 1$ such that the countable partition $\mathcal{Q} = \{f^{-1}(r) : r \in \mathbb{D}\}$ satisfies $d(\mathcal{P}, \mathcal{Q}) < \varepsilon$. The proof in the case where $\mathcal{P}$ has more than two atoms is similar; this is sufficient, because the finite partitions are dense in the Rohlin metric.

We construct a family of open sets $\{U_r\}_{r \in \mathbb{D}}$ with $\overline{U_r} \subseteq U_s$ for $r \leq s$ and with $\mu(\partial U_r) = 0$. We will also define closed disjoint sets $(C_r)_{r \in \mathbb{D}}$ such that $C_s \subseteq U_t \setminus U_r$ for all $r < s < t$ and $\mu(\cup C_r) = 1$. We will then define f by

$$f(x) = \inf(\{1\} \cup \{r : x \in U_r\}).$$

This defines a continuous function with $f|_{C_r} = r$, and so $\{f^{-1}(x) : x \in [0, 1]\}$ equals $\{C_r\}$ up to measure 0.

Fix a sequence (ε_k) to be determined later. For $i = 0, 1$, let C_i be disjoint closed sets with null boundary and $\mu(C_i \Delta P_i) < \varepsilon$. Set $U_0 = \emptyset$ and $U_1 = [0, 1] \setminus C_1$.

Let $\mathbb{D}_k \subseteq \mathbb{D}$ be the set of reduced dyadic rationals with denominator 2^k. We proceed by induction on k, defining at each step the sets U_r, C_r for $r \in \mathbb{D}_k$ under the assumption that they have been defined already for $r \in \bigcup_{j<k} \mathbb{D}_j$. Write $\mathbb{E}_k = \bigcup_{j<k} \mathbb{D}_j = \{r_1, \ldots, r_n\}$ with $r_1 < \cdots < r_n$ and let $r \in \mathbb{D}_k$. Then there are $r', r'' \in \mathbb{E}_k$ with $r' < r < r''$ and $(r', r'') \cap \mathbb{E}_k = \emptyset$. Let $V = U_{r''} \setminus \overline{U}_{r'}$ and choose $C_r \subseteq V$ with

$$\mu(C_r) > (1 - \varepsilon_k)\mu(V) = (1 - \varepsilon_k)\mu(U_{r''} \setminus U_{r'}).$$

Choose U_r such that it contains $C_r \cup U_{r'}$; it satisfies $\mu(\partial U_r) = 0$ and $\overline{U}_r \subseteq U_{r''}$.

Write $\mathcal{Q} = \{C_r\}_{r \in \mathbb{D}}$. Set $\widetilde{C}_k = \bigcup_{i \geq k} \bigcup_{r \in \mathbb{D}_i} C_r$ and let $\mathcal{Q}_k = \{C_r\}_{r \in \mathbb{E}_k} \cup \{\widetilde{C}_k\}$ be the partition obtained by merging all the atoms C_r in $\mathcal{Q}$ with $r \in \bigcup_{j \geq k} \mathbb{D}_j$. Let $C_k^* = \bigcup_{r \in D_k} C_r$. The sequence (ε_k) controls the convergence of the sequence $(\mu(C_k^*))$ to 1, and the latter can be made to converge arbitrarily quickly. In particular, we can guarantee that Q has finite entropy. Now $\mathcal{Q}_k \to \mathcal{Q}$ in the Rohlin metric, so

$$d(\mathcal{P}, \mathcal{Q}) = \lim_{k \to \infty} d(\mathcal{P}, \mathcal{Q}_k)$$

$$\leq \lim_{k \to \infty} \left(d(\mathcal{P}, \mathcal{Q}_1) + \sum_{i=1}^{k-1} d(\mathcal{Q}_i, \mathcal{Q}_{i+1}) \right)$$

$$= d(\mathcal{P}, \mathcal{Q}_1) + \sum_{i=1}^{\infty} d(\mathcal{Q}_i, \mathcal{Q}_{i+1})$$

and the last line can be made arbitrarily small by prudent choice of (ε_k), since $\mathcal{Q}_{i+1}$ refines $\mathcal{Q}_i$ by splitting C_k^* into at most 2^k atoms whose relative mass is determined by ε_k. $\square$

We can now prove Theorem 1.3. Note that even for $d = 1$ this proof is more direct than that given in [**6**].

Proof of Theorem 1.3. Let $\mathbb{Z}^d$ act on X and suppose that for every $f \in C(X)$ one has

$$f \in \langle 1, T^u f : u < 0 \rangle,$$

where $<$ is the lexicographical order on $\mathbb{Z}^d$. This implies that f is measurable with respect to the σ-algebra generated by $\{T^u f : u < 0\}$ and, in particular, this shows that for any T-invariant measure μ there is a dense (in the Rohlin metric) set of partitions $\mathcal{Q}$ for which $h(\mathcal{Q}) = 0$, namely those which come from continuous functions (Proposition 3.4). Since $h(\mu, \mathcal{P})$ is continuous in $\mathcal{P}$ under the Rohlin metric, we conclude that $h(\mu, \mathcal{P}) = 0$ for every two-set partition and hence $h(\mu) = 0$. By the variational principle, $h_{\text{top}}(T) = 0$. $\quad\square$

4. Prediction in symbolic systems

4.1. *Generalities about subshifts and prediction.* Let Σ be a finite alphabet and $\sigma : \Sigma^{\mathbb{Z}} \to \Sigma^{\mathbb{Z}}$ the shift transformation. For $x \in \Sigma^{\mathbb{Z}}$, set $x^- = (\ldots, x_{-2}, x_{-1})$, and for a subshift $X \subseteq \Sigma^{\mathbb{Z}}$ let $X^- = \{x^- : x \in X\}$. A finite or right-infinite word a is an extension of $x^- \in X^-$ if $x^- a$ appears in X. Let $L(X)$ be the set of finite words appearing in X and $L_m(X) = L(X) \cap \Sigma^m$.

The following fact is well known.

LEMMA 4.1. *A subshift X is the union of periodic orbits if and only if every $x^- \in X^-$ extends uniquely to $x \in X$.*

Proof. If X is a finite union of periodic orbits, the conclusion is clear.

For the converse, we rely on the simple fact that, if there is some n such that $x_{-n}, \ldots, x_{-1}$ determines x_0 for all $x \in X$, then X is the finite union of periodic orbits. Thus, if $X \subseteq \Sigma^{\mathbb{Z}}$ is not the union of periodic orbits, then for every n there are a word $a_n \in L_n(X)$ and distinct symbols $u_n, v_n \in \Sigma$ such that $a_n u_n, a_n v_n \in L_{n+1}(X)$. Therefore, there are words $b_n, c_n \in \Sigma^{\mathbb{N}^+}$ beginning with u_n, v_n respectively such that $a_n b_n, a_n c_n$ appear in X. By compactness, we can choose a subsequence $n(k)$ such that $u = u_{n(k)}$ and $v = v_{n(k)}$ are constant, $a_{n(k)} \to x^- \in X^-$, $b_{n(k)} \to b \in \Sigma^{\mathbb{N}^+}$, and $c_{n(k)} \to c \in \Sigma^{\mathbb{N}^+}$. But then a, b begin with the distinct symbols u, v and $x^- a, x^- b \in X$, so x^- has at least two extensions in X. $\quad\square$

Thus, every infinite subshift, including zero-entropy ones, has at least one past with multiple extensions. On the other hand, the following observation was pointed out to us by Weiss. Note that it is a special case of the general fact that minimal systems are invertible on a dense G_δ.

LEMMA 4.2. *If X is a minimal subshift, then, for every $a \in L(X)$ and $k \in \mathbb{N}$, there is a word $b \in L(X)$ such that $ba \in L(X)$, and every occurrence of ba in X is followed by a unique word $c \in \Sigma^k$.*

Proof. It suffices to show this for $k = 1$, as the general case then follows by induction. Let $a \in L(X)$ and $u \in \Sigma$ be such that $au \in L(X)$. Consider all b's such that $bu \in L(X)$ and au appears in bu exactly twice, as a front segment and a back segment. By minimality, the lengths of such b's is bounded above and we can choose a maximal such b. If $x^+ \in X^+$ and $bx^+ \in X^+$, then by minimality au appears in x^+; thus, by maximality of b, we must have $x^+(1) = u$, for otherwise there is a front segment c of x^+ such that au appears in bc only as a front and a back segment, which is impossible by maximality of b. Thus, b is always followed by u in X. $\quad\square$

COROLLARY 4.3. *If X is a minimal subshift and $u \in L(X)$, then there is a word $v \in L(X)$ such that every occurrence of v is followed by u.*

Proof. Let u be given and let k be large enough that every $c \in L_k(X)$ contains u. In the previous lemma, let a be the empty word and let b, c be the words obtained. Then b is always followed by c and $c = c'uc''$ for some c', c''. The word $v = bc'$ has the desired property. $\qquad\square$

4.2. *Realization theorem.* We now begin the proof of Theorem 1.5. We start with a measure-preserving system $(X, \mathcal{B}, \mu, T)$ of entropy zero, and wish to construct a strictly ergodic subshift, supporting an isomorphic measure, in which each past has at most two futures. We may assume that μ is aperiodic (i.e. the set of periodic points has measure zero), otherwise the statement is trivial. By e.g. [12], we may assume that μ is an invariant measure on a uniquely ergodic, topologically weakly mixing, minimal subshift $X \subseteq \{0, 1\}^{\mathbb{Z}}$.

We construct a sequence of two-set generating clopen partitions $\mathcal{P}_n$ for $n = 0, 1, 2, \ldots$ such that $\mathcal{P}_n \to \mathcal{P}_*$, where $\mathcal{P}_*$ generates for μ. Denote by X_n the symbolic system arising from X and $\mathcal{P}_n$. Note that since $\mathcal{P}_n$ is clopen, X_n is minimal and uniquely ergodic. The two-sided $\mathcal{P}_n$-name of a point $x \in X$ is a point in X_n.

We will define a sequence of integers $m(n) \geq n$ such that $L_{m(n)}(X_n) = L_{m(n)}(X_{n+1})$, and another sequence $k(n) \geq n$ with the property that for every $u \in \Sigma^{k(n)}$,

$$\#\{w \in \Sigma^n : uw \in L(X_n)\} \leq 2;$$

these numbers will satisfy $m(n) \geq k(n) + n$, so that the system X_* arising from $\mathcal{P}_*$ will have the property that for every $u \in \Sigma^{k(n)}$,

$$\#\{w \in \Sigma^n : uw \in L(X_*)\} \leq 2.$$

This implies the desired result. By choosing the $m(n)$ large enough at each stage, we can furthermore guarantee that X_* is minimal and uniquely ergodic, but for simplicity we do not go into the details of this.

The construction is by induction. Define $\mathcal{P}^{(0)}$ to be the clopen generating partition according to the zeroth symbol; set $m(0) = 0$ and $k(0) = 0$.

We describe now the inductive step of the construction. We are given a two-set generating partition $\mathcal{P}_n$ of X into clopen sets and an integer $m(n)$. Given $\varepsilon > 0$, we will construct a new partition $\mathcal{P}_{n+1}$ which is ε-close to $\mathcal{P}_n$. We will ensure that $L_{m(n)}(X_n) = L_{m(n)}(X_{n+1})$ and define an integer $k(n + 1)$ with the properties above. Finally, we will be free to choose $m(n + 1)$ arbitrarily, since it only affects the next step of the construction.

Let Y_n be the shift of finite type whose allowed blocks of length $m(n) + 1$ are those appearing in $L_{m(n)+1}(X_n)$. Since X_n is infinite and transitive, and $X_n \subseteq Y_n$, it follows from basic properties of shifts of finite type that Y_n has positive entropy. Using the fact that X_n is mixing and has zero entropy (whereas Y_n has positive entropy), we can find a word $a \in L_{m(n)+1}(X_n)$, a word $b_{\text{old}} \in L(X_n)$, and a word $b_{\text{new}} \in L(Y_n) \backslash L(X_n)$ such that $b_{\text{old}}, b_{\text{new}}$ have the same length, and both begin and end with the word a. Furthermore, using standard marker arguments (see e.g. [8]), we may assume that if $x \in X_n$ and we

replace some sequence of occurrences of b_{old} in x with b_{new}, and if these occurrences were at least $2\ell(b_{old})$ apart, then we can identify the location of the changes from the modified sequence.

The partition of $\mathcal{P}_{n+1}$ will be constructed by replacing some of the occurrences of b_{old} in X_n with b_{new}. This is done as follows. First, using Corollary 4.3, choose $c \in L(X_n)$ such that every time c appears in X_n it is followed by b_{old}. We can extend c backwards arbitrarily while preserving this property, so we may assume that c is arbitrarily long. Since X_n is minimal, there is an R such that the gap between occurrences of c in X_n is at most R.

Next, choose a large N (how large will depend on R, $\ell(b_{old})$ and on the growth of words in the system X_n, and will be explained below) and choose a clopen bounded Alpern tower in X_n all of whose columns are of height $N - 1$ or N, and such that the base is contained in the cylinder set defined by cb_{old}. Purify each column of the tower according to the clopen partition $\bigvee_{i=0}^{4N} T^{-i}\mathcal{P}_n$. Consider one such column, which corresponds to the $\mathcal{P}_n$-name w. We proceed to modify the $\mathcal{P}_n$-name of the column; doing this for each column defines a new partition $\mathcal{P}_{n+1}$.

Fix $x \in X$ and its corresponding column. Let $i(1) = 0$ denote the height in the column of the first occurrence of cb_{old} in w, let $i(2)$ be the index of the next occurrence which does not intersect the first occurrence, and so on until $i(r)$, the index of the last occurrence of cb_{old} which is contained completely in the current column. Replace the occurrences of cb_{old} at indices $i(1)$, $i(2)$ with cb_{new}.

Using the syndeticity of occurrences of c, for some $\alpha > 0$ we have $r \geq \alpha N$, where α depends on R but not N. We next encode the $\mathcal{P}_n$-name of x from time 0 to $4N$. We do so by replacing the word cb_{old} at some of the levels $i(4), i(6), \ldots, i(r-2)$ with cb_{new}. We use only locations $i(j)$ where j is even; thus, no new consecutive occurrences of cb_{new} are introduced, and the consecutive occurrences of cb_{new} at the bottom of the column are unique and serve to identify it. We can encode the atom of $\bigvee_{i=0}^{4N} T^{-i}\mathcal{P}_n$ to which x belongs in the approximately $\frac{1}{2}\alpha N$ bits available because $h(X_n) = 0$, so the number of $\bigvee_{i=0}^{4N} T^{-i}\mathcal{P}_n$-names is $< 2^{\alpha N/4}$ assuming that N is large enough.

We have defined a partition $\mathcal{P}_{n+1}$. Note that we have modified w along a set of density at most $\ell(b_{old})/\ell(cb_{old})$, which can be made arbitrarily small by making c long; thus, $\mathcal{P}_{n+1}$ can be made ε-close to $\mathcal{P}_n$.

Since b_{new} does not appear in $L(X_n)$, we can recover the $\mathcal{P}_n$-name of a point $x \in X$ simply by replacing every occurrence of cb_{new} with cb_{old}. Thus, since $\mathcal{P}_n$ generates, so does $\mathcal{P}_{n+1}$.

Because b_{old}, b_{new} agree on their first and last $m(n)$ symbols, and because $b_{new} \in Y_n$ and all $m(n)$-blocks in Y_n are in $L_n(X_n)$, we also have $L_{m(n)}(X_n) \subseteq L_{m(n)}(X_{n+1})$.

Consider a point $x \in X$. We will show that by looking $2N$ symbols into the past of the $\mathcal{P}_{n+1}$-name of x, we can determine that the $\mathcal{P}_{n+1}$-name of x from time 1 to $\ell(b_{new})$ takes on one of at most two possible values. Thus, setting $k(n) = 2N$ and noting that $\ell(b_{new}) \geq m(n) \geq n$, we will have completed the inductive step.

Look into the $\mathcal{P}_{n+1}$-past of x until we find a sequence of two consecutive occurrences of cb_{new}; this must happen after at most N symbols at some index i. Looking back at most N symbols more, we find the next group of two or five consecutive cb_{new}'s at some index j. Between j and i we have coded the $\mathcal{P}_n$-name of x from time j to time $j + 3N$ (and even

a little bit more). In any case, assuming as we may that $N > \ell(b_{\text{new}})$, and since $j \geq -2N$, we can certainly recover the $\mathcal{P}_n$-name of x from time j to time $\ell(b_{\text{new}})$.

We now claim that there are at most two choices for the $\mathcal{P}_{n+1}$-name of x from time 1 to $m(n + 1)$. Note that the $\mathcal{P}_n$-name of x and the $\mathcal{P}_{n+1}$-name of x differ only at points which lie in the $\ell(b_{\text{new}})$ symbols following certain occurrences of c. But if some such occurrence of c intersects the $\mathcal{P}_n$-name of x from times $-\ell(b_{\text{new}}) + 1$ to $\ell(b_{\text{new}})$, then from space considerations there is a unique such c; and in this case the next $\ell(b_{\text{new}})$ symbols of x are either b_{new} or b_{old}. Thus, there are at most two possible choices for the atom of $\bigvee_{s=1}^{m(n)+1} T^s \mathcal{P}_{n+1}$ to which x belongs.

This completes the discussion of the induction step. By choosing ε small enough at each stage, we can arrange that $\mathcal{P}_n \to \mathcal{P}_*$ with $\mathcal{P}_*$ a generating partition for μ, and X_* will be 2-branching. By a proper choice of $m(n)$ and using the unique ergodicity and minimality of X (and hence of all the X_n), we can also ensure that X_* is minimal and uniquely ergodic.

5. An extremely non-invertible zero-entropy system

5.1. Generalities.
In this section, we address the relation between entropy and the structure of preimage sets of points in non-invertible topological systems. The motivation for this is the following simple fact, whose proof is a good illustration of why one expects there to be a connection between entropy and large preimage sets.

PROPOSITION 5.1. *A system with no small preimages has entropy at least* $\log 2$.

Proof. Let (X, T) be a system and $\delta > 0$ be such that for every $x \in X$ there are $x', x'' \in T^{-1}(x)$ with $d(x', x'') > \delta$. We can define functions $\tau_0, \tau_1 : X \to X$ such that $\tau_0(x), \tau_1(x) \in T^{-1}(x)$ and $d(\tau_0(x), \tau_1(x)) > \delta$; note that τ_0, τ_1 need not be continuous. For $n \in \mathbb{N}$ and a sequence $a = a_n a_{n-1} \cdots a_1 \in \{0, 1\}^n$, let

$$\tau_a(x) = \tau_{a_n}(\tau_{a_{n-1}}(\cdots \tau_{a_1}(x) \cdots)).$$

Note that $T(\tau_a(x)) = \tau_b(x)$, where $b \in \{0, 1\}^{n-1}$ is obtained by deleting the first symbol of a.

For a fixed $x \in X$, consider the set

$$A_n(x) = \{\tau_a(x) : a \in \{0, 1\}^n\}.$$

If $a, b \in \{0, 1\}^n$ and $a \neq b$, then there is a maximal index $i < n$ such that $a_j = b_j$ for $1 \leq j \leq i$ but $a_{i+1} \neq b_{i+1}$. Let $y = \tau_{a_i a_{i-1} \cdots a_1}(x) = \tau_{b_i b_{i-1} \cdots b_1}(x)$; then

$$T^{n-i-1}(\tau_a(x)) = \tau_{a_{i+1}}(y),$$
$$T^{n-i-1}(\tau_b(x)) = \tau_{b_{i+1}}(y),$$

so $d(T^{n-i+1}\tau_a(x), T^{n-i+1}\tau_b(x)) > \delta$. It follows that all the points in $A_n(x)$ are distinct and the set $A_n(x)$ is (n, δ)-separated; since this is true for all n, this implies that $h(X, T) > \log 2$. $\square$

One easy consequence of this is that for finite alphabets Σ every extremely non-invertible subshift of $\Sigma^{\mathbb{Z}}$ has entropy at least $\log 2$, because once a metric is fixed there is a δ such that every two distinct preimages of a point are δ apart.

As was mentioned in the introduction, Bobok has shown that for maps of the interval if a map is k-to-one then it has entropy $>\log k$ [**1**].

It is not hard to construct examples of zero-entropy systems where every point has multiple preimages, but it is not so easy to construct such a system with a globally supported ergodic measure, and Eli Glasner has asked whether this is possible. The construction below gives an affirmative answer to this question.

5.2. *The construction.*

Let σ be the shift on the one-sided Bebutov system $[0, 1]^{\mathbb{N}}$. We will construct a subshift of the Bebutov system by specifying a point $x_* \in [0, 1]^{\mathbb{N}}$ and taking its orbit closure $X = \overline{\{\sigma^n x_*\}_{n \in \mathbb{N}}}$. Things will be engineered so that X has zero topological entropy, and x_* is generic for an ergodic measure μ on X having support X.

For words $x, y \in [0, 1]^{\mathbb{N}}$, we set

$$d(x, y) = \sum_{i=1}^{\infty} |x(i) - y(i)| \cdot 2^{-i};$$

this defines a metric on $[0, 1]^{\mathbb{N}}$ which is compatible with the compact product topology. We also write

$$\|x\| = d(x, \bar{0}),$$

where $\bar{0} = (0, 0, \ldots)$. For a finite word x, we define

$$\|x\| = \sum_{i=1}^{\ell(x)} |x(i)| \cdot 2^{-i} = \inf\{\|y\| : y \in [0, 1]^{\mathbb{N}} \text{ and } x \text{ is a front segment of } y\}.$$

Note that $\|ab\| \geq \|a\|$ and that if x_n are finite words and $x_n \to x \in [0, 1]^{\mathbb{N}}$ in the obvious sense, then $\|x_n\| \to \|x\|$.

Suppose that $x \in [0, 1]^*$ is a finite word. We define $\theta_0(x), \theta_1(x) \in [0, 1]$ by

$$\theta_0(x) = \tfrac{1}{8}\|x\|, \qquad \theta_1(x) = \tfrac{1}{4}\|x\|$$

and we define $\tau_0, \tau_1 : [0, 1]^* \to [0, 1]^*$ by

$$\tau_0(x) = \theta_0(x)x, \qquad \tau_1(x) = \theta_1(x)x,$$

i.e. the symbols $\theta_i(x)$ are appended to the beginning of x.

For a sequence $b = b_M b_{M-1} \cdots b_1 \in \{0, 1\}^M$, define τ_b inductively by

$$\tau_{b_M \cdots b_1}(x) = \tau_{b_M}(\tau_{b_{M-1} \cdots b_1}(x))$$

and set $T_\emptyset(x) = x$. Note that if $b = b_M \cdots b_1$, then

$$\sigma^i(\tau_b(x)) = \tau_{b_{M-i} \cdots b_1}(x)$$

and in particular $\sigma^M(\tau_b(x)) = x$. One verifies that $\|\tau_b(x)\| \to 0$ exponentially as the length of b tends to ∞, uniformly in b and x.

We define τ_b on $[0, 1]^{\mathbb{N}}$ by the same formula. In the subshift we are about to construct, the preimage set of a point x will contain at least $\tau_0(x), \tau_1(x)$. Since $\tau_b(x) \to \bar{0}$ as $\ell(b) \to \infty$, the preimage tree of each point will be 'narrow', and not contribute to the entropy. Note however that there will also be preimages which do not come from applications of τ_b.

We construct x_* recursively. At the nth stage, we will be given a finite word x_n of length L_n and construct a word x_{n+1} of length L_{n+1} such that $x_{n+1} = x_n x_n'$ for some word x_n'. We then take x_* to be the limit of this increasing sequence of finite words.

We begin with an arbitrary finite word x_0 of length $L_0 > 0$. Our only assumption about x_0 is that it is strictly positive.

The passage from stage n to stage $n+1$ is as follows. Given x_n of length L_n, for $0 \leq k < L_n$ let w_k be the back segment of x_n starting at index k, that is,

$$w_k = x_n(k)x_n(k+1)\cdots x_n(L_n),$$

so $\ell(w_k) = L_n - k + 1$. For $b \in \{0, 1\}^{3^{L_n}}$, set

$$w_{b,k} = \tau_b(w_k).$$

Define y_n to be some concatenation of the words $w_{b,k}$ as b varies over $\{0, 1\}^{3^{L_n}}$ and $0 \leq k < L_n$ (the order is not important).

Now choose a large integer M_n which we will specify later. For now, we note that M_n may be chosen to depend not only on all the previous stages but also on y_n. Define

$$x_{n+1} = \underbrace{(x_n x_n \cdots x_n)}_{M_n \text{ times}} y_n.$$

Set $x_* = \lim x_n$ and let X be the orbit closure of x_*. In the next few subsections, we will show that (X, σ) has the advertised properties.

5.3. (X, σ) *is extremely non-invertible.* The point x_* has been constructed in such a way that if some finite word a appears in x_*, then it appears in at least two different configurations, preceded by symbols $r, r' \in [0, 1]$ such that $|r - r'| \geq \frac{1}{16}\|a\|$. This is because if a is a subword of x_n, then a is a front segment of some back segment b of x_n, and so $\tau_0(b)$ and $\tau_1(b)$ appear in x_{n+1}, and by definition the first symbols of $\tau_0(b)$ and $\tau_1(b)$ differ by $\frac{1}{16}\|b\|$, and $\|b\| \geq \|a\|$.

Thus, if y is a limit point of x_* and $y \neq 0$, then y is a limit point of finite subwords a_n of x_*, and since $\|y\| > c > 0$ for some c, we have $\|a_n\| > c$ for all large enough n. Therefore, we can find symbols $r'_n, r''_n \in [0, 1]$ such that $|r'_n - r''_n| > \frac{1}{16}c$ and $r'_n a_n, r''_n a_n$ appear in x_*. Passing to a subsequence, we get $r'_n a_n \to r' y$ and $r''_n a_n \to r'' y$ for some $r', r'' \in [0, 1]$ with $|r' - r''| \geq \frac{1}{16}c$, and so $r' y, r'' y$ are distinct preimages of y in X.

It remains to check that $\bar{0}$ has two preimages (it is clear from the construction that $\bar{0} \in X$, since x_* has arbitrarily long sequences of small numbers, consisting of front segments of the $w_{b,k}$). Since $\bar{0}$ is a fixed point of σ, one preimage is $\bar{0}$ itself. To see that there are other preimages, note that the words x_n all end in the same positive letter ε, the last letter of x_0, and this is also the last letter of all the words $w_{b,k}$ we constructed at each stage. On the other hand, as $\ell(b) \to \infty$, the front segments of $w_{b,k}$ approach $\bar{0}$, so there are arbitrarily long sequences of arbitrarily small numbers in x_*, each sequence preceded by an occurrence of ε. Thus, $\varepsilon 000\ldots$ is also a preimage of 0 in X.

5.4. (X, σ) *has zero topological entropy.* We verify this by estimating the number of ε-separated orbits. For words a, a' (either finite or infinite), we write

$$\|a - a'\|_\infty = \sup_i |a(i) - a'(i)|.$$

Note that for $x, x' \in X$,

$$\|x|_{[1;n]} - x'|_{[1;n]}\|_\infty > \varepsilon \implies \max\{d(\sigma^i x, \sigma^i x') : i = 1, \ldots, n\} > \varepsilon.$$

Fix $\varepsilon > 0$, and let A_n be the set of all subwords of x_* of length n. Set

$$C_\varepsilon(n) = \max\{|A| : A \subseteq A_n, \forall a, a' \in A \ \|a - a'\|_\infty > \varepsilon\}.$$

The topological entropy of (X, S) is

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log C_\varepsilon(n).$$

For a finite or infinite word a with symbols in $[0, 1]$, let $[a]_\varepsilon$ denote the word b of the same length such that

$$b(i) = [a(i)/\varepsilon] \cdot \varepsilon$$

(here $[r]$ denoted the integer part of r). Thus, the coordinates of $[a]_\varepsilon$ belong to the finite set $\{0, \varepsilon, 2\varepsilon, \ldots, [1/\varepsilon]\varepsilon\}$. Note that if $\|a - a'\|_\infty \geq \varepsilon$, then $\|[a]_{\varepsilon/2} - [a']_{\varepsilon/2}\|_\infty \geq \varepsilon/2$. It is therefore sufficient to prove the following.

CLAIM 5.2. *For every $\varepsilon > 0$, the number of length-n subwords of $[x_*]_{\varepsilon/2}$ which are at least $\varepsilon/2$ apart in $\|\cdot\|_\infty$ grows subexponentially with n.*

We will use the following property of x_*.

LEMMA 5.3. *For every n, we can write $x_* = a_1 a_2 a_3 \cdots$, where each a_i is of length at least 3^{L_n} and, for each i, either:*
(1) $a_i = x_n$; or
(2)   *for each $1 \leq j \leq \ell(a_i) - L_n$, we have $a_i(j) \leq \frac{7}{8} a_i(j+1)$.*
In particular, for any $\varepsilon > 0$, for n large enough either each a_i is equal to x_n or else all the coordinates of a_i, except the last $2L_n$ coordinates, are of magnitude $<\varepsilon$.

The proof of the lemma is an elementary induction from the definitions, and is omitted.

Proof of Claim 5.2. Fix $\varepsilon > 0$ and let $z_* = [x_*]_{\varepsilon/2}$ and $z_m = [x_m]_{\varepsilon/2}$. From the lemma, we see that for the given ε for large enough m, we can write

$$z_* = v_1 v_2 v_3 \cdots$$

and for each i either the word v_i is equal to z_m or else $\ell(v_i) \geq 3^{L_m}$ and at least a $(1 - 2^{-L_m})$-fraction of the coordinates of v_i are 0. In view of this, the fact that the number of subwords of z_* of length n grows subexponentially is now a standard counting argument, and the claim follows. This shows that $h_{\text{top}}(X, \sigma) = 0$. $\square$

5.5.   *x_* is generic for a globally supported measure μ on X.*   A point y in a dynamical system (Y, S) is a generic point for a measure μ if for every continuous function $f \in C(Y)$ $\lim_{N \to \infty}(1/N) \sum_{n=1}^N f(S^i y)$ exists. When this is true, then $(1/N) \sum_{n=1}^N \delta_{S^i y}$ converges in the weak-$*$ topology to an invariant measure μ on Y (here δ_x is the point mass at x). One condition that guarantees that y is generic is that for every open set $U \subseteq Y$ the averages $\lim_{N \to \infty}(1/N) \sum_{n=1}^N 1_U(S^i y)$ exist; in fact, it is sufficient to verify this for U coming from a basis for the topology of X.

For $U \subseteq [0, 1]^k$, let

$$[U] = U \times [0, 1]^{\mathbb{N}\setminus\{1,\ldots,k\}} \subseteq [0, 1]^{\mathbb{N}}$$

be the cylinder determined by U. Sets of this form for open U constitute a basis for the topology of $[0, 1]^{\mathbb{N}}$. We will show that for every such U, the sequence

$$p(m) = \frac{1}{m} \sum_{i=1}^{m} 1_{[U]}(\sigma^i x_*) \tag{5.1}$$

converges. This implies that the weak* limit measure

$$\mu = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\sigma^i x_*}$$

exists, and is a shift-invariant measure on X. In fact, we will show that $\mu(U) > 0$ if and only if $p(n) > 0$ for some n. From this, it will follow that μ has global support in X.

For a finite word a, we will say that $a \in [U]$ if $ab \in [U]$ for every infinite $b \in [0, 1]^{\mathbb{N}}$. Thus, if $a \in [U]$, then $ab \in [U]$ for every finite b. The property $a \in [U]$ depends only on the first k coordinates of a (recall that $U \subseteq [0, 1]^k$). Note that if $\ell(a) < k$ it is possible that $a \notin [U]$ but that $ab \in [U]$ for some (finite of infinite) b.

CLAIM 5.4. *Let $U \subseteq [0, 1]^k$ and $p(n)$ be as above. The limit $\lim_{s \to \infty} p(L_s)$ exists; furthermore, if $p(n) > 0$ for some n then the limit is positive.*

Proof. If $\sigma^n x_* \notin [U]$ for every n, then clearly $\lim p(n) = 0$. Therefore, we must check only the case when $\sigma^n x_* \in [U]$ for some n. Note that in this case, $p(m) > 0$ for all $m \geq n$. We prove first that $p(L_r)$ converges at $r \to \infty$, and then the general claim.

For a word a, let $I(a)$ be the number of indices $0 \leq n < \ell(a)$ such that $\sigma^n a \in [U]$. If we let a_m be the front m-segment of x_*, we have

$$\frac{I(a_m)}{m} \leq p(m) \leq \frac{I(a_m) + k}{m}$$

(the right inequality is because of edge effects; it is possible for $\sigma^n a \notin [U]$ but $\sigma^n x_* \in [U]$ if $\ell(a) - k < n < \ell(a)$). In particular, for any r we have

$$\frac{I(x_r)}{L_r} \leq p(L_r) \leq \frac{I(x_r) + k}{L_r}. \tag{5.2}$$

If $p(L_r) > 0$, then also $p(L_{r+1}) > 0$, and x_{r+1} contains at least M_r copies of x_r. Thus, if we assume that $M_s \geq 2^s$ for every s, we may fix r such that $I(x_s) \geq 2^s$ for every $s \geq r$.

For an s as above, write

$$x_{s+1} = x_s x_s \cdots x_s y_s$$

as in the construction of x_{s+1}, with the x_s's repeating M_s times. We can write $I(x_{s+1}) = I_1 + I_2$, where

$$I_1 = \#\{0 \leq n < M_s L_s : \sigma^n x_{s+1} \in [U]\},$$
$$I_2 = \#\{M_s L_s \leq n < L_{s+1} : \sigma^n x_{s+1} \in [U]\}.$$

We have

$$M_s \cdot I(x_s) \leq I_1 \leq M_s \cdot (I(x_s) + k),$$

since we may gain at most $M_s k$ occurrences at the edges of the x_s's but we cannot lose occurrences. Also, we have the trivial bound $I_2 \leq \ell(y_s)$. Therefore,

$$M_s I(x_s) \leq I(x_{s+1}) \leq M_s(I(x_s) + k) + \ell(y_s)$$

and, substituting this and $L_{s+1} = M_s L_s + \ell(y_s)$ into inequality (5.2), we get

$$\frac{M_s \cdot I(x_s)}{M_s L_s + \ell(y_s)} \leq p(L_{s+1}) \leq \frac{M_s \cdot I(x_s) + \ell(y_s) + (M_s + 1)k}{M_s L_s + \ell(y_s)}.$$

Dividing the middle term by $p(L_s)$ and using (5.2) again, we get

$$\frac{1}{1 + k/I(x_s)} \cdot \frac{1}{1 + \ell(y)/M_s L_s} \leq \frac{p(L_{s+1})}{p(L_s)} \leq \frac{1 + k/I(x_s) + (\ell(y) + k)/M_s I(x_s)}{1 + \ell(y)/M_s L_s}.$$

We saw above that $k/I(x_s)$ is exponentially small in s. Thus, if $\{M_n\}$ grows quickly enough, both the expression on the left, which we denote α_s, and the expression on the right, which we denote β_s, converge to 1 rapidly enough for their product to converge to a finite positive number. Now the relation $\alpha_s \leq p(L_{s+1})/p(L_s) \leq \beta_s$ and the fact that $0 < \prod_r^\infty \alpha_s, \prod_r^\infty \beta_s < \infty$ imply that $p(L_s)$ converges to a positive limit as $s \to \infty$. □

CLAIM 5.5. *For U and $p(n)$ as above, $\lim_{n\to\infty} p(n)$ exists and is positive if $p(n) > 0$ for some n.*

Proof. Let $p = \lim p(L_s)$, the limit of $p(n)$ along the subsequence L_s. To show that $p(n) \to p$, we show that if $L_s \leq n < L_{s+1}$, then $p(n)/p(L_{s-1})$ is close to 1, in a manner depending on s and tending to 1 with s. To see this, recall that

$$x_{s+1} = (x_s x_s \cdots x_s)y_s$$
$$= ((x_{s-1} \cdots x_{s-1} y_{s-1}) \cdots (x_{s-1} \cdots x_{s-1} y_{s-1}))y_s.$$

Write a_n for the front n-segment of x_{s+1}. Then there is a unique way to write a_n as

$$a_n = (x_s \cdots x_s)(x_{s-1} \cdots x_{s-1})w$$

with w a front segment of either x_{s-1}, y_{s-1} or y_s.

For $n \geq L_s$, the number of x_s's appearing is at least 1. Now consider two alternatives: if w is a front segment of either x_{s-1} or y_{s-1}, then $\ell(w)$ is negligible compared to $\ell(a_n)$ because $\ell(a_n) \geq \ell(x_s) \geq M_{s-1}\ell(x_{s-1})$ and M_{s-1} has been chosen large. On the other hand, if $w = y_s$, then all M_s repetitions of x_s appear in a_n, and again we have that $\ell(w)$ is negligible compared to $\ell(a_n)$.

An estimate like the one carried out for $p(L_s)$ shows that we can ignore edge effects and write $p(n)$ as some weighted average of $p(L_s)$ and $p(L_{s-1})$. But we know already that $p(L_s)/p(L_{s-1}) \to 1$, so $p(n) \approx p(L_{s-1}) \to p$. □

5.6. *The only ergodic measures on X are μ and the point mass $\delta_{\bar{0}}$.* *A-priori the measure μ for which x_* is generic need not be ergodic. Rather than prove directly that μ is ergodic, we will show that if ν is any ergodic measure on (X, σ), then ν is a convex combination of μ and $\delta_{\bar{0}}$. This implies that μ is an extreme point of the convex set of invariant measures on X, so it is ergodic and is the only ergodic measure on X other than $\delta_{\bar{0}}$.*

THEOREM 5.6. *The only ergodic measures for (X, σ) are μ and δ_0.*

Proof. Using Lemma 5.3, we can select a sequence $r(n) \to \infty$ and write

$$x_* = b_{1,n} b_{2,n} b_{3,n} \cdots$$

such that each $b_{i,n}$ is either equal to $x_{r(n)}$ or has the property that $\ell(b_{i,n}) \geq 3^{L_{r(n)}}$ and all but the final $2L_{r(n)}$ coordinates are $<1/n$.

If ν is an ergodic measure for (X, σ), then for some sequence with $m(n) - k(n) \to \infty$ we have

$$\nu = \lim_{n \to \infty} \frac{1}{m(n) - k(n) + 1} \sum_{i=k(n)}^{m(n)} \delta_{\sigma^i x_*}$$

(this follows from the fact that by the ergodic theorem ν has generic points, and these can be approximated arbitrarily well by shifts $\sigma^i(x_*)$ of x_*). By passing to subsequences, we can assume that $m(n) - k(n) > 2^{L_{r(n)}}$; denote $w_n = x_*|_{[k(n),m(n)]}$ so that $\ell(w_n) > 2^{L_{r(n)}}$. Write λ_n for the total number of indices $i = 1, \ldots, \ell(w_n)$ such that i is in a word $b_{j,n}$ with $b_{j,n} = x_{r(n)}$. We may further assume, by passing to a subsequence, that $\lambda_n \to \lambda \in [0, 1]$.

Now we can write $w_n = b' b_{i(n),n} \cdots b_{j(n),n} b''$ for some $i(n) < j(n)$ and b', b'' as short as possible. Notice that if $b_{i(n)-1,n}$ or $b_{j(n)+1,n}$ are $x_{r(n)}$, then their lengths, respectively, are negligible (logarithmic) compared to $\ell(w_n)$, and so also are the lengths of b', b'', respectively. On the other hand, if $b_{i(n)-1,n}$ is not $x_{r(n)}$ and if the length of b' is more than $(1/n)\ell(w_n)$, then that word is made up almost entirely of coordinates of magnitude less than $1/n$. Similar reasoning holds for b''. It is now simple to verify the following.

- If $\lambda = 0$, then for large n most of w_n is made up of coordinates of magnitude $<1/n$, so in this case we have $\nu = \delta_0$.
- If $\lambda = 1$, then for large n the distribution of words of length $\sqrt{L_{r(n)}}$ in w_n is very close to their distribution in $x_{r(n)}$, and since $r(n) \to \infty$ we have $\nu = \mu$ in this case.
- Finally, for $0 < \lambda < 1$ the same reasoning as above shows that

$$\nu = \lambda\mu + (1 - \lambda)\delta_0$$

(note that because the lengths of the $b_{i,n}$ tend to infinity with n, the statistics of subwords of w_n of length $\sqrt{L_{r(n)}}$ are only very slightly affected by the places where two $b_{i,n}$'s meet. Since we assumed that ν is ergodic, this is impossible).

Thus, $\nu = \delta_0$ or $\nu = \mu$. Since $\mu \neq \delta_0$, this implies that μ is ergodic. This completes the proof. $\qquad\square$

5.7. *Further comments.* This example is optimal in the following sense. Any minimal system (X, T) has the property that on some dense G_δ subset of X the preimage of any point is a single point. Thus, there are no minimal extremely non-invertible systems. Thus, if we want an extremely non-invertible system supporting a global ergodic measure, we cannot hope for a uniquely ergodic example. The example we have given is the next best thing: it has only two invariant measures and a unique minimal subsystem, the fixed point $\bar{0}$.

The construction can be modified in several ways. For instance, one can guarantee that the preimage set of every point is large: by augmenting the two functions θ_0, θ_1 at each

stage of the construction with other functions, it is not hard to make the preimage set of every point of cardinality $2^{\aleph_0}$. By modifying θ_0, θ_1 in a more complex way, one can replace the minimal subsystem $\{\overline{0}\}$ with other systems.

Using the last modification, one can establish Example 1.7 by taking the product of the resulting system with the one-sided two-shift $\{0, 1\}^{\mathbb{N}}$, and the product (Bernoulli) measure. This yields a system with infinitely many preimages for every point, no small preimages, and a globally supported ergodic measure of entropy $\log 2$. In this example, there are many other invariant measures; by a more careful choice of the system we multiply with, for example, a minimal, uniquely ergodic subshift with a weakly mixing invariant measure of entropy $\log 2$, this can be avoided.

Finally, in the construction we defined words $w_{b,k} = \tau_b(w_k)$, where b varies over all 0, 1-valued sequences of a fixed length. By varying this length in a 'random' way, the measure μ can be made to be weakly mixing, and perhaps even strongly mixing.

REFERENCES

[1] J. Bobok. The topological entropy versus level sets for interval maps. *Studia Math.* **152**(3) (2002), 249–261.

[2] J. Bobok and Z. Nitecki. Topological entropy of m-fold maps. *Ergod. Th. & Dynam. Sys.* **25**(2) (2005), 375–401.

[3] W.-C. Cheng and S. E. Newhouse. Pre-image entropy. *Ergod. Th. & Dynam. Sys.* **25**(4) (2005), 1091–1113.

[4] S. J. Eigen and V. S. Prasad. Multiple Rokhlin tower theorem: a simple proof. *New York J. Math.* **3A** (1997–1998), 11–14 (Proceedings of the New York Journal of Mathematics Conference, 9–13 June 1997) (electronic).

[5] D. Fiebig, U.-R. Fiebig and Z. H. Nitecki. Entropy and preimage sets. *Ergod. Th. & Dynam. Sys.* **23**(6) (2003), 1785–1806.

[6] B. Kamiński, A. Siemaszko and J. Szymański. The determinism and the Kolmogorov property in topological dynamics. *Bull. Pol. Acad. Sci. Math.* **51**(4) (2003), 401–417.

[7] B. Kamiński, A. Siemaszko and J. Szymański. Extreme relations for topological flows. *Bull. Pol. Acad. Sci. Math.* **53**(1) (2005), 17–24.

[8] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.

[9] Z. Nitecki and F. Przytycki. Preimage entropy for mappings. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **9**(9) (1999), 1815–1843; Discrete dynamical systems.

[10] P. Shields. *The Theory of Bernoulli Shifts (Chicago Lectures in Mathematics)*. The University of Chicago Press, Chicago–London, 1973.

[11] P. Walters. *An Introduction to Ergodic Theory (Graduate Texts in Mathematics, 79)*. Springer, New York, 1982.

[12] B. Weiss. Multiple recurrence and doubly minimal systems. *Topological Dynamics and Applications (Minneapolis, MN, 1995) (Contemporary Mathematics, 215)*. American Mathematical Society, Providence, RI, 1998, pp. 189–196.