ISRAEL JOURNAL OF MATHEMATICS **160** (2007), 119–141
DOI: 10.1007/s11856-007-0057-z

UNIVERSAL MINIMAL TOPOLOGICAL DYNAMICAL SYSTEMS

BY

OFEK SHILON

Simbionix Ltd., Hamelacha 6, Northern Industrial Zone
Lod, Israel 71520
e-mail: ofek@simbionix.com

AND

BENJAMIN WEISS

Institute of Mathematics, Hebrew University of Jerusalem
Givat Ram, Jerusalem 91904, Israel
e-mail: weiss@math.huji.ac.il

ABSTRACT

Rokhlin (1963) showed that any aperiodic dynamical system with finite entropy admits a countable generating partition. Krieger (1970) showed that aperiodic ergodic systems with entropy $< \log a$, admit a generating partition with no more than a sets. In Symbolic Dynamics terminology, these results can be phrased— $\mathbb{N}^{\mathbb{Z}}$ is a universal system in the category of aperiodic systems, and $[a]^{\mathbb{Z}}$ is a universal system for aperiodic ergodic systems with entropy $< \log a$. Weiss ([We89], 1989) presented a **Minimal** system, on a **Compact** space (a subshift of $\overline{\mathbb{N}}^{\mathbb{Z}}$) which is universal for aperiodic systems. In this work we present a joint generalization of both results: given ε, there exists a minimal subshift of $[a]^{\mathbb{Z}}$, universal for aperiodic ergodic systems with entropy $< \log a - \varepsilon$.

1. Introduction

A Topological Dynamical System is a pair (X, T) where X is a topological space (not necessarily compact), and $T: X \to X$ is a homeomorphism. A Measure-Theoretic Dynamical System (MT-Dynamical system, or Dynamical System) is a quadruplet $(X, \mathcal{B}, \mu, T)$, where $\mathcal{B}$ is a σ-algebra structure on X, μ is a $\mathcal{B}$-Probability Measure and for our purposes, $T: X \to X$ is an invertible, bi-measurable, measure-preserving transformation. We further assume throughout

Received August 2, 2005

this work that the measure space $(X, \mathcal{B}, \mu)$ is standard. An MT-system is called ergodic, when it possesses no true subsystems in the measure theoretic sense: $Y \subset X, T^{-1}Y \subset Y \Longrightarrow \mu(Y) \in \{0,1\}$. A topological system is called minimal, when it possesses no true subsystems in the topological sense: $Y \subset X$, Y is closed and $T^{-1}Y \subset Y \Longrightarrow Y \in \{\varnothing, X\}$.

The first context of this work is the exploration of the relations between topological and measure-theoretic dynamics. The thread of analysis explored here was initiated by J. Auslander, who posed the problem whether a minimal system can carry two disjoint MT-systems. Weiss and Katznelson constructed a positive answer, and Weiss continued in [We89] to present an ultimate answer; a minimal system which carries all ergodic (aperiodic) MT-systems. This first *universal* minimal system was realized as a minimal subshift of $\overline{\mathbb{N}}^{\mathbb{Z}}$. In this work we present new variants on this result.

A second context for this work is the ongoing search for interesting universal systems. Rokhlin showed in 1963 that $\mathbb{N}^{\mathbb{Z}}$ is universal for all aperiodic systems with finite entropy. As mentioned, Weiss constructed a *minimal* subshift of $\overline{\mathbb{N}}^{\mathbb{Z}}$ ($\overline{\mathbb{N}}$ is even compact), universal for all aperiodic ergodic systems; without requiring the entropy to be finite. Krieger showed (1970) that the shift $\{1, \ldots, a\}^{\mathbb{Z}}$ (abbreviated $[a]^{\mathbb{Z}}$) is universal in the category of aperiodic ergodic systems of entropy strictly less than $\log a$. A natural question arises: can *minimal* universality be achieved in this smaller category? (i.e., is there a minimal subshift of $[a]^{\mathbb{Z}}$ universal for all ergodic aperiodic systems in this entropy range?) As posed, the assertion is readily seen to be false: any subtraction of blocks from $[a]^{\mathbb{Z}}$ to form a minimal system, would emit a positive loss of entropy, say δh, and by the variational principle this would result in a system unable to carry an MT system of entropy $\log a - 1/2\delta h$. So, the closest minimal analogue of Krieger's theorem to which we can aspire is — given ε, build a minimal subshift of $[a]^{\mathbb{Z}}$ universal for all aperiodic ergodic systems with entropy $< \log a - \varepsilon$. Such a construction is the main result presented here. Along the way we modify the approach to the construction in [We89], that left a certain technical issue unaddressed.

In Section 2 we review some well-known facts from ergodic theory and topological dynamics, to establish the framework for the rest of the paper. The details can be found in one form or another in the texts cited in the references. In Section 3 we present a detailed proof of Krieger's theorem, following Ornstein's guidelines, thereby demonstrating our basic tools and visual paradigm. The simplest case, zero entropy, is presented first and the general case is presented in a less detailed manner, emphasizing only the additional characteristics. In

Section 4 we present our novel results in the same fashion: from the simple to the complex, surveying mainly new ideas along the way.

This work is the outcome of the Thesis work of the first author, under the supervision of the second author.

2. Preliminaries

2.1. Symbolic dynamics. We will denote by $\prec$ occurrences of a block both in a sequence and in a subshift. Given an alphabet S, we say that a block $A \in S^n$ **occurs syndetically** in a sequence $x \in S^{\mathbb{Z}}$ (and write $A \underset{d}{\prec} x$) if there exists k such that A occurs in every k-segment of x. Given a subshift $M \subset S^{\mathbb{Z}}$, we say that A occurs syndetically in M (and write $A \underset{d}{\prec} M$), if A occurs syndetically in every $x \in M$. We denote by $BL_n(M)$ the set of all n-blocks occuring in M. We recall some basic facts about closed subshifts (cf. [DGS]):

Theorem 2.1: *Let S be a finite (discrete) alphabet. Then $M \subset S^{\mathbb{Z}}$ is minimal iff any block that occurs in M, occurs syndetically in M.*

We will make use also of the non-discrete alphabet $S = \overline{\mathbb{N}} = \{1, 2, 3, \dots, \infty\}$, where the topology on $\overline{\mathbb{N}}$ is the usual one with ∞ the limit point of $n \in \mathbb{N}$. We will call such a set **semi-discrete**. An easy generalization of the above result reads:

Theorem 2.2: *Assume $M \subset \overline{\mathbb{N}}^{\mathbb{Z}}$ satisfies $|BL_1(M)| = \infty$. Then M is minimal iff any ∞-free block that occurs in M, occurs syndetically in M.*

The usage of non-finite, non-discrete alphabets takes some careful revision of basic definitions and theorems. Let S be any alphabet. A subshift $M \subset S^{\mathbb{Z}}$ will be called a **Finite-Type (FT) Subshift** if it has a defining block system with bounded length (but not necessarily finite). This means that for some N, $x \in M$ iff every block of length N in x belongs to $BL_N(M)$. Such minimal N will be called the **Memory** of the subshift. If $A, B \prec M$, an **M-transition block** from A to B is a block $U \prec M$ such that $A \cdot U \cdot B \prec M$. ($\cdot$ denoting concatenation).

Theorem 2.3: *Let $M \subset S^{\mathbb{Z}}$ be a subshift, with S discrete. Then M is topologically mixing iff for any $A, B \prec M$ there exists $N_{A,B}$ such that for any $n > N_{A,B}$ there exists an M-transition block $U \in BL_n(M)$ from A to B.*

We will abbreviate MFT-subshift for a mixing-finite-type-subshift. We add an ad-hoc notation adapted to our needs: A subshift $M \subset S^{\mathbb{Z}}$ will be called

uniformly mixing (UM), if there exists some N such that for any $n > N$ and any $A, B \prec M$, there exists an M-transition block $U \in BL_n(M)$ from A to B.

Remark:

(1) Clearly for a finite alphabet S, an MFT-subshift M is UM.

(2) The condition "there exist N such that any $n > N$ is a transition length between any 2 blocks" can be relaxed to the existence of a single N such that both N and $N+1$ are uniform transition lengths between any two blocks.

An immediate generalization of Theorem 2.3 to the semi-discrete case reads:

THEOREM 2.4: *Let $M \subset \overline{\mathbb{N}}^{\mathbb{Z}}$ be an FT subshift, with $|BL_1(M)| = \infty$. Then M is topologically mixing iff for every two ∞-free blocks $A, B \prec M$, there exists $N_{A,B}$ such that for any $n > N_{A,B}$ there is a transition block $U \in BL_n(M)$ from A to B.*

Example: Let M be an MFT-subshift defined on the finite alphabet $\{1, 2, \ldots, N, \infty\}$. Let $\tilde{M}$ be the subshift obtained from M by letting the symbol ∞, whenever it occurs in M, assume the values $\{N+1, N+2, \ldots, \infty\}$. Than $\tilde{M}$ is a uniformly mixing subshift of $\overline{\mathbb{N}}^{\mathbb{Z}}$.

2.1.1. A Construction of a minimal subshift. We describe now in detail a standard construction of minimal subshifts that we will use later on. Let M_1 be some UM subshift of $\overline{\mathbb{N}}^{\mathbb{Z}}$ (may be $\overline{\mathbb{N}}^{\mathbb{Z}}$ itself), and let $\{N_i\}$ be any monotonically increasing sequence in $\mathbb{N}$.

We will construct inductively a descending sequence of FT-UM-subshifts $M_1 \supset M_2 \supset \cdots$, such that in M_i all blocks of length i that make use only of the symbols $\{1, 2, \ldots, N_i\}$ occur syndetically. $M = \bigcap_{i=1}^{\infty} M_i$ (which is non-empty by compactness), is readily seen to be minimal.

Suppose M_{i-1} is given, and is an FT-UM-subshift. Consider the collection of blocks

$$\tilde{B} = \{A_1, A_2, \ldots, A_{n_i}\} = BL_i(M_{i-1}) \cap \{1, 2, \ldots, N_i\}^i,$$

i.e., all blocks of length i which contain only the symbols $1, 2, \ldots, N_i$. Since M_{i-1} is mixing we can find a transition block U_1 from A_1 to A_2, a transition block U_2 from $A_1 \cdot U_1 \cdot A_2$ to A_3 and so forth; we can eventually build a single block $w_{i-1} \prec M_{i-1}$ in which all the blocks in $\tilde{B}$ occur. Assume w.l.o.g that $l(w_{i-1})$ is larger than M_{i-1}'s memory.

Next, choose some length L_i larger than $l(w_{i-1})$, and set the following defining block system for M_i: all blocks in $BL_{L_i}(M_{i-1})$ in which w_{i-1} occurs. M_i is obviously FT, and is easily seen to be UM by noting that $x \in M_i$ is constructed by freely assigning transition blocks from w_{i-1} to itself, from a given list of allowed blocks. Thus, the construction is complete.

The analogous construction for the finite alphabet case $[a]^{\mathbb{Z}}$ is much easier: it can be treated as the case $M_1 = [a]^{\mathbb{Z}}$, and we can drop the gradual exhaustion of the symbols (namely, N_i) altogether.

Remark: Ambiguity may arise in the case that w_{i-1} overlaps itself (e.g., what is the distance between the first two occurrences of 121 in the string 1212121?). To avoid inconvenience, we may concatenate to the end of w_{i-1} a string that would separate such occurrences.

2.2. Partitions, generators and Rokhlin towers. Given a measure space $(X, \mathcal{B}, \mu)$, we denote by $\overset{0}{=}$ equivalence modulo μ-null sets, of sets, partitions and σ-algebras. By d we denote the semimetric induced by symmetric difference, on sets or partitions, i.e., $d(A, B) = 1/2\mu(A\Delta B)$, and for labelled partitions: $d(P, Q) = 1/2 \sum \mu(P_i \Delta Q_i)$. We denote by $P \vee Q$ the join of the two partitions P and Q, and by $\bigvee_{i=1}^{\infty} P_i$ the infinite join of the partitions P_i, i.e., the σ-algebra generated by all finite joins of the P_i's. Call a partition P a **generator** if $\bigvee_{i=1}^{\infty} T^i P \overset{0}{=} \mathcal{B}$. It is well-known that given a standard space $(X, \mathcal{B}, \mu)$, there exists a sequence of measurable sets $\{A_i\}$ dense in $\mathcal{B}$ in the metric d. (Equivalently, there exists a generating sequence of finite partitions.)

Next, we facilitate the transition between symblic and general dynamics.

Definition 2.5: Let $(X, \mathcal{B}, \mu, T)$ be a dynamical system, P a partition of X with index set I, M a subshift of $I^{\mathbb{Z}}$.

(1) Denote by Φ_P the map that matches each $x \in X$ with its bi-infinite P name, i.e., $\Phi_P(x) \in I^{\mathbb{Z}}$ is defined via $(\Phi_P(x))_n = j \iff T^n x \in P_j$.
(2) P will be called M-compatible, if almost every P-name belongs to M. Specifically, when for μ-almost every $x \in X$ holds: $\Phi_P(x) \in M$.

Theorem 2.6: *Let $(X, \mathcal{B}, \mu, T)$ an MT dynamical system with $(X, \mathcal{B}, \mu)$ a standard space, P a partition of X with index set S. The following are equivalent:*

(1) *P is a generator;*
(2) *there exists a subset $Y \subseteq X$ with $\mu(Y) = 1$ such that Φ_P is injective in Y, i.e., the P-names of almost all points in X are distinct;*
(3) *Φ_P is an MT-conjugacy between $(X, \mathcal{B}, \mu, T)$ and $(S^{\mathbb{Z}}, \mathcal{F}, \Phi_P\mu, \sigma)$.*

We formulate explicitly a natural albeit technical lemma that will be used shortly.

LEMMA 2.7: *Let M_n be a descending sequence of subshifts of $\overline{\mathbb{N}}^{\mathbb{Z}}$, $(X, \mathcal{B}, \mu, T)$ be a dynamical system, $P^{(n)}$ a sequence of partitions that converges to P, such that $P^{(n)}$ is M_n compatible. Then P is M compatible.*

Also recall the Rokhlin tower construction, and a helpful modification of it

THEOREM 2.8: *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic dynamical system, $0 < \varepsilon$ and $n \in \mathbb{N}$. Then:*

(1) *There exists an (n, ε)-Rokhlin base, i.e., a set B such that B, TB, $T^2B, \ldots, T^{n-1}B$ are pairwise disjoint, and their union is of measure $> 1 - \varepsilon$.*
(2) *There exists a base $\tilde{B}$ such that $\tilde{B}, T\tilde{B}, \ldots, T^{n-1}\tilde{B}$ are pairwise disjoint, furthermore: $\mu(\bigcup_{i=0}^{n} T^i \tilde{B}) = 1$, i.e., the ε remainder can be covered in a single additional iteration of T. Such a base will be denoted a* **Special Rokhlin Base of Heights** *$(n, n+1)$.*

For a more detailed survey of the second variant above, cf. [We00], §3.

Given a special Rokhlin $(n, n+1)$ base B, **purifying B with respect to a partition** P means partitioning B into sets $B_m, 1 \leq m \leq M$, such that for $0 \leq i \leq n$, $T^i B_m$ is contained entirely in a single atom of P. Each collection $\{B_m, TB_m, \ldots, T^n B_m\}$ is called a **pure column** with respect to P.

2.2.1. A Construction of a Generator. We describe now a standard scheme of constructing a generating partition as a limit of other partitions. Let $\{A_j\}$ be a sequence of sets dense in $\mathcal{B}$. Suppose a sequence of partitions $\{P_i\}$ with the same index set I satisfies for every large enough i:

$$A_j \overset{\varepsilon_j^{(i)}}{\subset} \bigvee_{k=0}^{n_j - 1} T^k P_i$$

for some fixed sequence of integer 'window lengths' $\{n_j\}$ and some fixed family (indexed by i) of sequence of positive estimates $\{\varepsilon_j^{(i)}\}_j$ which tends to 0 as j tends to ∞. Furthermore, if it is shown that $P_i \to P$, then P is a generator.

We will actually employ a more specific form of the estimates $\varepsilon_j^{(i)}$, namely $\varepsilon_j^{(i)} = 2^{-j} + 2^{-j-1} + \cdots + 2^{-i}$.

Obviously, the d-dense sequence of sets $\{A_j\}$ can be replaced by a generating sequence of partitions $\{Q_j\}$.

2.3. Entropy, equipartition and Robust codes. For our purposes it is worth recalling:

(1) Given an ergodic system $(X, \mathcal{B}, \mu, T)$ and a partition P, The (measure theoretic) entropy of the partition $H = H_P^T$ satisfies the asymptotic equipartition property (AEP), namely, for every ε there exists N such that for every $n > N$

$$\mu\{x|2^{-n(H+\varepsilon)} < \mu([x_1^n]_P) < 2^{-n(H-\varepsilon)}\} > 1 - \varepsilon$$

(where $[x_1^n]_P$ denotes the set of all points with n-P-name identical to that of $x \in X$). Standard bounds on cardinality readily follow.

(2) **The Relative AEP**, is a straightforward corollary to the regular AEP: given an ergodic system $(X, \mathcal{B}, \mu, T)$ and $Q \subset P$ two partitions of X with entropies H_Q^T and H_P^T respectively, set $h_{P|Q} = H_P^T - H_Q^T$ (the system's **relative entropy**). Then, for every ε there exists N such that for $n > N$ there exists a set $G_{n,P}$ of measure $> 1-\varepsilon$, which is a union of words $[x_1^n]_Q$ which satisfy:

$$2^{n(h_{P|Q}-\varepsilon)} < |\{[y_1^n]_P | [y_1^n]_P \cap [x_1^n]_Q \neq \varnothing\}| < 2^{n(h_{P|Q}+\varepsilon)}.$$

Literally: ε-almost every Q-word, is ε-almost covered by about $2^{nh_{P|Q}}$ P-words.

(3) The Topological Entropy of a top. dyn. system (X, T) is the supremum of all measure theoretic entropies of ergodic systems carried by (X, T): $h_{\mathrm{TOP}}(T) = \sup\{h_\mu(T) | \mu \in \mathcal{M}_T(X), \mu \text{ ergodic}\}$ (the variational principle).

(4) Specifically for a subshift $M \subset S^{\mathbb{Z}}$, $h_{\mathrm{TOP}}(\sigma) = \lim_{n\to\infty} \frac{1}{n} \log |BL_n(M)|$, (i.e., $|BL_n(M)|$ is logaritmically asymptotic to $2^{nh(\sigma)}$).

We will also utilize a special case of a fairly standard result from information theory, concerning error-correcting codes in a topological context.

Definition 2.9: a **codebook** of length N in an alphabet M is a subset of N-words in the alphabet $C \subset BL_N(M)$. Each codebook element is called **codeword**.

Theorem 2.10: *Let $M \subset \{1, \ldots, a\}^{\mathbb{Z}}$ be a fixed subshift (with $a \geq 2$), with positive topological entropy $h = h_{TOP}(\sigma) > 0$, and let K be a fixed vocabulary size. Then, there exists some distortion threshold $0 < \delta_0(h) < 1$ and some $N = N(h, K)$ such that for any $n > N$, there is a δ_0-robust n-codebook in M, of size K. (I.e., we can find a codebook $C \subset BL_N(M)$ with $|C| = K$ and for any $a, b \in C$, the average hamming distance $\Delta(a, b)$ is more than $2\delta_0$.)*

Proof: The number of ways to distort an n-word at δn places is bounded by $\binom{n}{\delta n} a^{\delta n}$, and by the Stirling formula this bound can be approximated as $\sim 2^{U(n)+V(\delta)+nW(\delta)}$, where $U(n) = -\frac{1}{2}\log 2\pi n$, $V(\delta) = -\frac{1}{2}[\log\delta + \log(1-\delta)]$ and $W(\delta) = -\delta\log\delta - (1-\delta)\log(1-\delta) + \delta\log a$. To form a δ-robust code we fix an arbitrary first word in $BL_n(M)$, erase all of its 2δ-neighbors, pick some second word and erase all of its 2δ-neighbors, etc. We require n to be large enough so that $|BL_n(M)| > 2^{n(h/2)}$. Then, a condition that would suffice to keep constructing K words of this code without running out of words, is $U(n) + V(2\delta) + n\tilde{W}(2\delta) + \log K < n(h/2)$, and an elementary analysis of these functions reveals that we can indeed set δ_0, N with the desired properties. ∎

We will make use of a specific setup of the theorem's conditions:

Corollary 2.11: *Fix some alphabet size a, and fix an allowed error $\delta = 0.001$. If $h(M) > (1/2)\log a$. Then for any vocabulary size K we can find a codebook in M, with K words of length $N = 36 + 4\log K$ (for example), that can overcome distortions of magnitude δ.*

3. Krieger's Theorem

3.1. The zero entropy case.

Theorem 3.1: *The topological dynamical system $\{0,1\}^{\mathbb{Z}}$ is universal for all standard ergodic aperiodic dynamical systems with entropy 0.*

Given some standard ergodic dynamical system $(X, \mathcal{B}, \mu, T)$ with entropy 0, by Theorem 2.6 our mission reduces to finding a two-set generator for $\mathcal{B}$. We will construct such a generator using the scheme surveyed in 2.2.1: we fix $\{A_i\}$ a sequence of sets dense in $\mathcal{B}$, and strive to form a convergent sequence of partitions P_i that estimate A_i well in the aforementioned sense, for $i \geq j$:

$$A_j \overset{\varepsilon_j^{(i)}}{\subset} \bigvee_{k=0}^{n_j - 1} T^k P_i$$

where n_j is some ascending sequence in $\mathbb{N}$, and $\varepsilon_j^{(i)} = 2^{-j} + 2^{-j-1} + \cdots + 2^{-i}$.

Proof: Let $P_1 = \{A_1, A_1^C\}$, $n_1 = 1$. Suppose that $P_1, P_2, \ldots, P_{i-1}$ and $n_1, n_2, \ldots, n_{i-1}$ were built with the desired properties and let us form P_i and n_i. By the AEP and the zero entropy assumption, we can fix n_i so that there exists a set G_i of measure $> (1 - 2^{-i-4})$ that is a union of less than $2^{(2^{-i-4})n_i}$ atoms in $\bigvee_{k=0}^{n_i-1} T^k(\{A_i, A_i^C\})$. We may assume that $n_i > n_{i-1}$. Fix some special (n_i, n_i+1) Rokhlin base B, and assume w.l.o.g. that

$$\mu(B \cap G_i) > (1 - 2^{-i-4})\mu(B)$$

(for, otherwise, we can switch to some $T^m B$ with this property). Purify this tower with respect to A_i. We need to make a small modification of P_{i-1} into P_i, so that A_i would be well estimated by $\bigvee_{k=0}^{n_i-1} T^k P_i$. This modification would be carried out in stages, by relabelling the bottom few cells in every 'good' pure column in the tower above B (i.e., every pure column above $B \cap G_i$). We'll label a cell 0 or 1, according to its intended label in the new partition P_i.

Modification (1): Synchronizing

By synchronizing we mean — identifying the base B uniquely within a long P_i name. The easiest way to achieve this, is to label the lowest $2^{-i-4}n_i$ cells by 0, and of the remaining cells label 1 once every $2^{-i-5}n_i$ cells. The overall modification in this stage is less than

$$2^{-i-4} + 1/(2^{-i-5}n_i)$$

and by demanding in advance that $n_i > 2^{2i+9}$, we can make this distortion less than 2^{-i-3}.

Modification (2): Coding A_i

By an 'A_i configuration' of a column above B, we mean a string of 0's and 1's describing whether the current level is in A_i or not (i.e., an $\{A_i, A_i^C\}-n_i$-name). For a column of height n_i, there are 2^{n_i} theoretically possible A_i-configurations, but n_i was chosen such that almost all of B is covered by at most $2^{(2^{-i-4})n_i}$ such configurations, and so, to properly describe the large majority of used configurations we need a string of length at most $2^{-i-4}n_i$. To that purpose, reserve the next $2^{-i-4}n_i$ cells (skipping the single cells already forced to be 1), and relabel them in every column to uniquely code the A_i configuration of that column. The magnitude of the modification in this stage is 2^{-i-4}.

By now we have achieved some of our goals: we can positively identify B by recognizing a long string of 0's, and inside $B \cap G_i$ we can distinguish between the 'good' pure columns (which makes the most of the columns) by observing the coded configuration. Thus every pure cell $C \subset B \cap G_i$ belongs to $\bigvee_{k=0}^{n_i-1} T^k P_i$, and obviously every $T^m C$ belongs to $\bigvee_{k=0}^{n_i-1} T^k P_i$. Hence:

$$A_i \overset{2^{-i}}{\subset} \bigvee_{k=0}^{n_i-1} T^k P_i$$

and by bounding the magnitude of modifications we easily achieved that $d(P_{i-1}, P_i) < 2^{-i}$, which is enough to make the sequence P_i converge.

The rest of our goals will not be achieved by additional modifications of the construction, but rather by calibrating its parameters.

Modification (3): Controlling the loss of former estimates

The graphical interpretation of the estimate $A_j \overset{\varepsilon_j^{(i)}}{\subset} \bigvee_{k=0}^{n_j-1} T^k P_i$ is that by looking at a 'sliding window' of width n_j starting at a point x, we can make a prediction whether x belongs to A_j or not, and be wrong only with probability $\varepsilon_j^{(i)}$. If we assume $n_j < n_i$, we can bound the region affected by the modification $P_{i-1} \to P_i$ to lie within a radius n_j around modified cells , and adjust the parameters (meaning, the n_k's) so that this region would be less than 2^{-i}. By assuming that the sequence $\{n_k\}$ is ascending, it is enough to ensure that the region within radius n_{i-1} of modified cells is small. In other words, it is enough to guarantee that

$$d\Big(\bigvee_{k=0}^{n_{i-1}-1} T^k P_{i-1}, \bigvee_{k=0}^{n_{i-1}-1} T^k P_i\Big) < 2^{-i}.$$

The portion of modified n_{i-1}-strings within an n_i column is no more than

$$2^{-i-3} + 2 \cdot \frac{n_{i-1}}{n_i} + 2^{i+5} \cdot \frac{2n_{i-1}}{n_i}$$

(taking into account the sync block + A_i coding block, as well as the scattered '1' pollution). Bounding this portion under 2^{-i} amounts to demanding a growth condition of the form:

$$(2^{2i+6} + 2^{i+1})n_{i-1} < n_i.$$

This growth condition can be demanded in advance while obtaining n_i, say in the simpler form $2^{3i} n_{i-1} < n_i$.

And so, for every $j \le i$, the estimate $A_j \overset{\varepsilon_j^{(i)}}{\subset} \bigvee_{k=0}^{n_j-1} T^k P_i$ holds, and our construction is complete. ∎

3.2. The Positive Entropy Case.

Motivation. Let $M \subset \{1, \dots, a\}^{\mathbb{Z}}$ be a subshift. For a large n, $|BL_n(M)| \sim 2^{nh(M)}$ (in a logarithmic asymptotic sense). A 'coding' of these n-words is a scheme of renaming them, preferrabaly using as few letters as possible, while maintaining the distinction between them. An optimal coding utilizing a possible letters per cell, would result in words of length x, where $a^x = 2^{nh(M)}$, i.e. $x = (nh(M))/\log a$. Therefore, we can think of $h(M)/\log a$ as the asymptotic compression ratio achievable for a word in M. In the formation of partitions that follow, similarly to the zero entropy case, we would have an a-atoms partition P with entropy close to $\log a$, and we would want to modify it on a small set into P' to estimate well another partition Q. Now, within a fixed typical P-word of length n, by the *relative* AEP, the number of distinct $P \vee Q$-words of length n is $\sim 2^{nh_{P\vee Q|P}}$, which offers a much better compression ratio: $h_{P\vee Q|P}/\log a$ (we will denote $h_0/\log a$ for short). Our coding scheme will consist of fixing a certain $0 < \alpha < 1$, choosing an α-prefix of the long $P \vee Q$ word of length n, compressing slightly the P-configuration of the α prefix itself, and using the remainder of the prefix to compress the $P \vee Q$ configuration of the entire n word.

So, the α-segment used for the coding needs to be large enough so that the remainder after the P-compression would suffice to encompass the $P \vee Q$-compression of the entire n column. In other words:

$$\alpha n \cdot \left(1 - \frac{H_T^P}{\log a}\right) > n \cdot \frac{h_0}{\log a}$$

$$\alpha > \frac{h_0}{\log a - H_T^P}.$$

In what follows, we will indeed utilize an α segment of the form $C \cdot h_0$, where $C = 10/(\log a - h(T))$ is chosen to be a system's constant. The α-modified segment is still small enough to guarantee convergence and even control loss of former estimates.

Another feature that must play a role, is bounding the entropy of the built partitions P_i from below: by the above discussion, we change P_{i-1} by a magnitude of $C \cdot h_{P_{i-1}\vee Q_i|P_{i-1}} = C \cdot h_i$ to capture Q_i well. Thus, h_i can be seen as a measure of how much P_{i-1} must change in order to estimate Q_i well. To make this size small in the next stage we will, in some sense, use Q_i to increase the entropy of the newly formed partition P_i.

The Theorem.

THEOREM 3.2: *The topological dynamical system $\{1, 2, \ldots a\}^{\mathbb{Z}}$ is universal for all standard ergodic aperiodic dynamical systems with entropy strictly less than $\log a$.*

Since entropy will be a key player in the construction, it is more natural to work with a generating sequence of partitions $Q_1 \subset Q_2 \subset Q_3 \subset \cdots$ (i.e. $\bigvee Q_i \overset{0}{=} \mathcal{B}$), than with a dense sequence of sets $\{A_i\}$. Assume w.l.o.g. that $\{H_T^{Q_i}\}$ is ascending, and of course converges to $h(T)$.

We will separate the core of the construction into a

Lemma 3.3: *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic, aperiodic, standard dynamical system with $h(T) < \log a$. Let P be a partition with a atoms, Q any partition, $0 < \delta$, $N \in \mathbb{N}$. Denote $h_0 = H_T^{P \vee Q} - H_T^P$ and fix $C = 10/(\log a - h(T))$. Then there exists n and a partition P' with a atoms, such that*

$$Q \overset{\delta}{\subset} \bigvee_{k=0}^{n} T^k P' \quad \text{and} \quad d\Big(\bigvee_{k=0}^{N-1} T^k P, \bigvee_{k=0}^{N-1} T^k P'\Big) < Ch_0 + 2\delta.$$

Proof: By the relative AEP we can find n so large that $\delta/4$-almost every atom in $\bigvee_{k=0}^{n-1} T^k P$ is $\delta/4$-almost covered by no more than $2^{n(h_0+\delta/4)}$ atoms in $\bigvee_{k=0}^{n-1} T^k(P \vee Q)$. By the regular AEP, we can also make n so large that a $(1-\delta/4)$ portion of the space is covered by no more than $2^{nCh_0(H_T^P+\delta/4)}$ P-words of length nCh_0. As before, fix some special $(n, n+1)$ Rokhlin base B, with the additional property that only a fraction up to $\delta/2$ of B itself is uncovered by the 'good' $P \vee Q$-words of length n and good P words of length nCh_0. Label each cell by a symbol in $\{1, 2, \ldots, a\}$, according to it's location in P, and let us focus on a specific good P-column above B.

Modification (1) — Synchronizing.

This modification is identical to its parallel in the zero-entropy case: label the lowest $\delta/8$ cells by zeros, and label 1 once every $(\delta/16)n$ cells. The overall modification this far is of magnitude

$$\frac{\delta}{8} + \frac{16}{\delta n}$$

that can be made smaller than, say, $\delta/4$, by demanding that n be large enough to begin with.

Modification (2) — Coding the P prefix and the entire $P \vee Q$ name.

For this modification we reserve the first $n(Ch_0+\delta/2)$ cells untouched by modification (1). We need to modify the first nCh_0 of these cells to code uniquely the P-name of the entire $n(Ch_0 + \delta/2)$ — while using the compression ratio achievable by the choice of n — and using the saved $\delta/2$ suffix space to code uniquely the $P \vee Q$-name of the entire n-column.

The amount of different names we need, is the amount of possible P-names in the $n(Ch_0 + \delta/2)$ prefix, times the number of possible $P \vee Q$-names of length n (over a given P prefix). By the choice of n and B, this is less than

$$2^{nCh_0(H_T^P+\delta/4)} 2^{n(h_0+\delta/4)}$$

and since we have at our disposal $2^{n(Ch_0+\delta/2)\log a}$ possible codes, we can claim that our stock suffices:

$$2^{nCh_0(H_T^P+\delta/4)}2^{n(h_0+\delta/4)} < 2^{n(Ch_0+\delta/2)\log a}$$

$$Ch_0\Big(H_T^P+\frac{\delta}{4}\Big)+h_0-\frac{\delta}{4} < Ch_0\log a.$$

We can assume w.l.o.g. that δ satisfies $(\log a - H_T^P) - \delta/4 > (1/2)(\log a - H_T^P)$, and hence

$$\frac{h_0-\delta/4}{\log a - H_T^P - \delta/4} < \frac{2h_0}{\log a - H_T^P} < \frac{10h_0}{\log a - H_T^P} = Ch_0;$$

which implies the desired inequality, and thus meets our need.

Notice again that the overall modification in stage (2) is $Ch_0+\delta/2$.

Remark: The estimates given above assume $h_0 \neq 0$. The construction in the case that $h_0 = 0$ will not be elaborated here, since it is practically identical to the construction in Section 3.1.

Notice that some of our assertions already hold: while observing the atoms of $\bigvee_{k=0}^{n-1} T^kP'$, we can single out those beginning with a long string of zeros and identify their base as B. We can then deduce the P-configuration of the next $n(Ch_0+\delta/2)$ cells by deciphering the coded nCh_0 prefix, and afterwards we can decipher the entire n-$P\vee Q$ configuration within that column, and thus reconstruct Q entirely, on the 'good' portion of B, i.e., on a set of measure at least $1-\delta$. So, we already have $Q\overset{\delta}{\subset}\bigvee_{k=0}^{n} T^kP'$, as well as $d(P,P') < Ch_0+\delta$, which is still weaker than the desired bound.

Modification (3) — Controlling $d(\bigvee_{k=0}^{N-1} T^kP, \bigvee_{k=0}^{N-1} T^kP')$.

This modification is entirely identical to its parallel in the zero entropy case; it amounts to demanding a condition of the form $N \ll n$, and bounding the measure of the region of radius N around modified cells. Since the modified prefix on all good columns is of magnitude $Ch_0+\delta$, we can easily make its N-neighbourhood smaller than $Ch_0+2\delta$ (taking into account the scattered 'pollution' as well). The good columns — where the entire modification was carried out — measure at least $1-\delta$, so we must add a factor of δ to the error in estimating Q. ∎

We return now to the proof of the Theorem 3.2, by use of the lemma.

Proof: Given an ergodic, aperiodic, standard dynamical system $(X,\mathcal{B},\mu,T)$ with $h(T) < \log a$, let $\{Q_i\}$ be a generating sequence of partitions. We will form partitions P_i which estimate the Q_i well. Assume w.l.o.g. that $H_T^{Q_i} > h(T) - 3^{-i}$ (otherwise switch to a subsequence of $\tilde{Q}_i = \bigvee_{k=0}^{i} Q_k$). Suppose that $P_1, \dots, P_{i-1}$ and $n_1, \dots, n_{i-1}$ are given, and satisfy for every $j < i$

$$Q_j \overset{\varepsilon_j^{(i-1)}}{\subset} \bigvee_{k=0}^{n_j-1} T^k P_{i-1}$$

$$d\Big(\bigvee_{k=0}^{n_j-1} T^k P_{j-1}, \bigvee_{k=0}^{n_j-1} T^k P_j\Big) < C \cdot 3^{-i} + 2^{-i}$$

where, as before, $\varepsilon_j^{(i-1)} = 2^{-j} + \cdots + 2^{-(i-1)}$.

We must now use Q_i to increase the entropy of the estimating partition P_i: by continuity of the entropy, we can pick $0 < \delta_i$ such that if for some P holds $Q_i \overset{\delta_i}{\subset} \bigvee_{k=0}^{n_{i-1}} T^k P$, then $H_T^P > H_T^{Q_i} - 3^{-i}$. Without loss of generality, $\delta_i < 2^{-i}$. We necessarily have for such a P:

$$\begin{aligned} h_{P \vee Q_i | P} &= H_T^{P \vee Q_i} - H_T^P < h(T) - H_T^{Q_i} + 3^{-i} \\ &< 2 \cdot 3^{-i}. \end{aligned}$$

Now, apply the lemma with $\delta = \delta_i$ and $N = n_{i-1}$ to obtain $P = P_i$ and $n = n_i$ such that:

$$Q_i \overset{\delta_i}{\subset} \bigvee_{k=0}^{n_i-1} T^k P_i$$

$$d\Big(\bigvee_{k=0}^{n_{i-1}-1} T^k P_{i-1}, \bigvee_{k=0}^{n_{i-1}-1} T^k P_i\Big) < C \cdot 2 \cdot 3^{-i} + 2^{-i}.$$

And we are almost done, the second inequality enables control over the loss of former estimates. For a given ε and large enough j

$$d\Big(\bigvee_{k=0}^{n_j-1} T^k P_j, \bigvee_{k=0}^{n_j-1} T^k P_i\Big) < \sum_{k=j}^{i} (C \cdot 2 \cdot 3^{-k} + 2^{-k}) < \varepsilon$$

hence,

$$Q_j \overset{\delta_j+\varepsilon}{\subset} \bigvee_{k=0}^{n_i-1} T^k P_i$$

The same bound applies for the limit partition P (which is readily seen to exist), and so P must be a generator and our construction is complete. ∎

4. Minimal Universal Systems

We strive next to form a *minimal* subshift M of $\{1,\ldots,a\}^{\mathbb{Z}}$, which is universal for some wide category of dynamical systems. We have described in Section 2.1.1 a detailed construction of a minimal subshift: it can be constructed as an intersection of subshifts M_i, such that in M_i all words up to length i in M_{i-1} appear at a bounded distance L_i (specifically: a single word w_{i-1} which contains them all appears at least every $(1/2)L_i$ places). We next described in 2.2.1 a scheme of constructing a sequence of partitions P_i which estimate the generating family of sets $\{A_i\}$ in an efficient manner, and thus converge to a generator. In the three constructions which follow, the basic approach is similar. However, we must now concern ourself with an additional complication: each P_i must be M_i compatible, to make the limit partition P M-compatible (by Lemma 2.7). The nature of the subshifts M_i (all being UM) makes this task manageable: instead of replacing a word with another at whim, we must now use only legal words and pad the replaced region with preceding and succeeding transition blocks (whose length is bounded). A delicate issue will arise in the modification of P_{i-1} to become M_i compatible; this modification involves injecting w_{i-1} at distances $(1/2)L_i$, which are independent of the heights $\{n_j\}$ of the 'windows' used to code A_j in former partitions. At first glance it is hard to see how can we bound the distortion that such a modification causes in previous estimates, but a second glance reveals the solution: the usage of error-correcting codes.

In the last subsection we present a new construction of a minimal universal space for all aperiodic standard ergodic systems (of any entropy), realized as a subshift of $\overline{\mathbb{N}}^{\mathbb{Z}}$. This construction involves yet another dimension of complication, as each partition P_i will pose a condition only upon a finite subset of symbols, $\{1,2,\ldots,N_i\}$.

4.1. The zero entropy case.

Theorem 4.1: *There exists a* **minimal** *subshift M of $\{0,1\}^{\mathbb{Z}}$ which is universal for all standard aperiodic ergodic dynamical systems with entropy* 0.

Proof: To be specific, we go over the details of the construction of the M_i in the present context.

Construct a descending sequence of subshifts $M_i \subset \{0,1\}^{\mathbb{Z}}$ in the following scheme:

- M_1 is the full shift $\{0,1\}^{\mathbb{Z}}$
- Fix a sequence of words $w_i \prec BL(M_i)$ and lengths L_i such that

$$B \in BL_i(M_{i-1}) \Rightarrow B \prec w_{i-1} \quad C \in BL_{L_i}(M_i) \iff w_{i-1} \prec C \prec M_{i-1};$$

i.e., w_{i-1} contains all i-words which actually occur in M_{i-1}, and M_i is constructed from M_{i-1} by the additional restriction that w_{i-1} appears every L_i places.

- We recall that any M_i constructed in such a manner is MFT, so in order to modify an M_{i-1} sequence to become M_i compatible, we need only make modifications of length $K_{i-1} = l(\omega_{i-1}) + 2(M_{i-1}$'s memory) once every $\frac{1}{2}L_i$ symbols. $M = \bigcap_i M_i$ is necessarily minimal.
- Lastly we will assume that for each i:

$$h_{\mathrm{TOP}}(M_i) > 1/2.$$

This entropy ensures that each M_i is roomy enough to contain many properly spaced codewords, to generate the robust coding used later. This requirement can be easily guaranteed by a condition of the form: $K_{i-1} \ll L_i$.

Now, let $(X, \mathcal{B}, \mu)$ be a standard space with some ergodic Borel automorphism T, such that $h(T) = 0$. Fix a sequence of sets $\{A_j\}$ dense in $\mathcal{B}$. We now need to construct a sequence of partitions P_i such that:

(1) P_i would be M_i compatible,

(2) $\{P_i\}$ would converge to some partition P (which would be automatically M-compatible) — this would require some growth demands on the L_i, to be made specific soon.

(3) For every $i \geq j$, $A_j \overset{\varepsilon_j^{(i)}}{\subset} \bigvee_{k=0}^{n_j} T^k P_i$ where $\varepsilon_j^{(i)} = 2^{-j} + 2^{-(j+1)} + \cdots + 2^{-i}$, to make P a generator.

Suppose then that we are given P_{i-1} and $n_1, \ldots, n_{i-1}$ such that for any $j \leq i-1$ holds $A_j \overset{\varepsilon_j^{(i-1)}}{\subset} \bigvee_{k=0}^{n_j} T^k P_{i-1}$. We need to construct a P_i that has the same property and in addition is M_i-compatible and not too far from P_{i-1}. Take Q to be the partition generated by P_{i-1} and A_i. Since the entropy of the system X is zero, by the AEP we can find a large enough n_i such that there is a set G_i of measure $(1 - 2^{-i-4})$ that is covered by at most $2^{2^{-i-4} n_i}$ words in $\bigvee_{k=0}^{n_i-1} T^k Q$. Fix some special $(n_i, n_i + 1)$ Rokhlin Base B, with the additional property that only a 2^{-i-4}-fragment of B is outside G_i. Construct the $n_i + 1$ Rokhlin tower above B and purify it with respect to Q. Label each cell in every pure column by 0 or 1 according to it's location in P_{i-1}. The modification of P_{i-1} into P_i will be carried out in stages, by relabeling some of the cells in the good ($\subset G_i$) columns. As the modification progresses, we will encounter some additional conditions that the lengths $\{L_j\}$ and $\{n_j\}$ must satisfy.

Modification (1): M_i compatibility.

We modify the partition to be M_i compatible, by Modifying the lowest K_{i-1} levels to begin with w_{i-1}, modifying the uppermost K_{i-1} levels to end with an M_{i-1}-transition block to w_{i-1}, and injecting the word w_{i-1} with appropriate transition blocks prior and afterwards (with overall length K_{i-1}) once every $(1/2)L_i$ places. This modification is of magnitude $K_{i-1}/(\frac{1}{2}L_i) + (2K_{i-1})/n_i$, and w.l.o.g. this is less then $(3K_{i-1})/L_i$ (i.e., $n_i > L_i$). Soon we will demand a bound on this magnitude, of the form $(3K_{i-1})/L_i < 0.001$ (and in fact, much stronger). We need to maintain this newly achieved M_i compatibility throughout the modifications at stage i, described next.

Modification (2): Synchronizing

To that purpose we reserve $2^{-i-4}n_i$ places. We fix some legitimate M_i-transition block v' between w_{i-1} to itself, and relabel the reserved levels by repeated occurrences of $v = w_{i-1} \cdot v'$.

To make this identification unique, we again 'pollute' accordingly the remainder of the tower. An easy way to achieve this is to pick another M_i-transition block u' between w_{i-1} to itself, of length identical to v' (again, denote $u = w_{i-1} \cdot u'$) and inject successive occurrences of it once every $2^{-i-5}n_i$ places. The injected sequence of u's must withstand future distortions, and so must have a nominal substantial thickness; say overall length of $2^{-2i-9}n_i$.

We must make sure, of course, that u does not occur in $v\cdot v$, but foresight tells that this property ($u \not\prec v\cdot v$) needs to withstand a multitude of future distortions as well, so let us pick u and v such that the minimal average Hamming distance from u to $BL_{l(v)}(v \cdot v)$ is larger than 0.001 (this is possible by Corollary 2.11).

The overall modification in this stage of the modification is of magnitude $2^{-i-4} + 2^{-2i-9}/2^{-i-5} = 2^{-i-3}$.

Modification (3): Estimating A_i.

Now we come to the actual coding of A_i. Although there are 2^{n_i} possible A_i configurations of each pure column, the choices of n_i and the AEP ensure that there are essentially (up to a set of measure $(1-2^{-i-4})$) only $k = 2^{2^{-i-4}n_i}$ such configurations, so we need only this amount of labels. Now, M_i-compatibility prohibits us from using any set of k different names we want, but if we reserve to this purpose additional $2^{-i-2}n_i$ places, Corollary 2.11 will come to our rescue again and assert that there exists a set $G \subset BL_{2^{-i-2}n_i}(M_i)$, of size $|G| = k$, and with the additional useful property that any two of them are more than 0.002 apart (in average Hamming distance). We can now modify the next $2^{-i-2}n_i$ places (along with appropriate preceding and succeeding transition blocks, whose length will be omitted from the calculation for convenience), to contain unique identifiers of all the good columns — and as such, unique identifiers of all the A_i configurations of all good columns. The overall modification in this stage is of measure 2^{-i-2}.

The modification of a typical column thus far can be sketched (unproportionally) as in Figure 4.1.

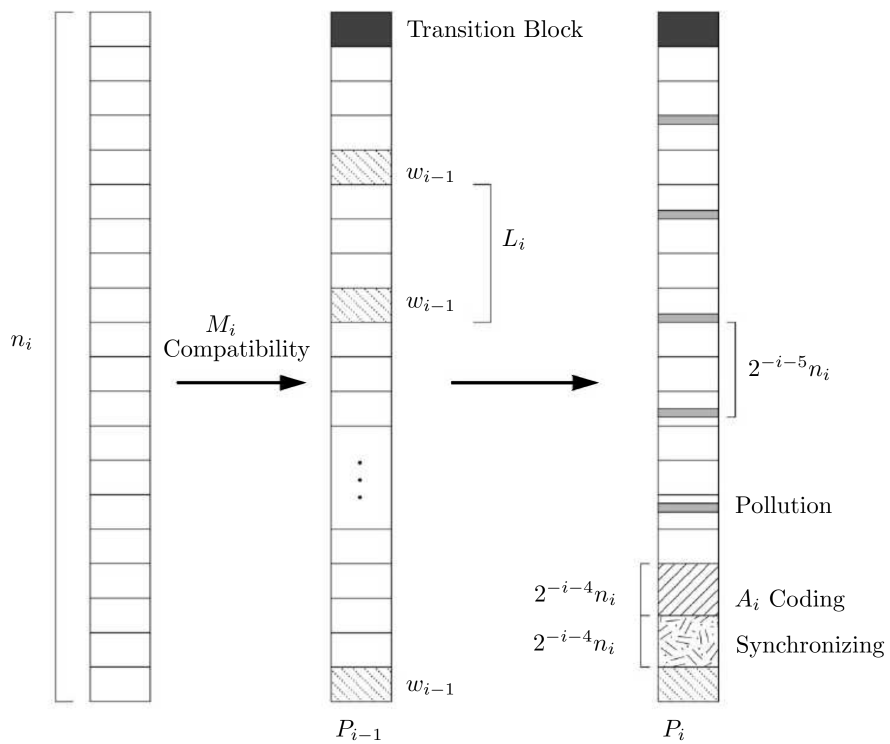

Figure 4.1. Column modification stages

Modification (4): Controlling the loss of former estimates

In this new setup, we need to adjust the *two* parameter families in the construction (L_j and n_j) so that at any stage of the construction the estimates $A_j \overset{\varepsilon_j^{(i)}}{\subset} \bigvee_{k=0}^{n_j-1} T^k P_i$ would still hold. We will extract some conditions on $\{L_i\}$ and $\{K_i\}$ in terms of i (that is, a demand on the universal system M to begin with), and some conditions on $\{n_i\}$ in terms of i, $\{L_i\}$ and $\{K_i\}$ (that is, demands on the construction of P_i for the specific dynamical system $(X, \mathcal{B}, \mu, T)$).

Let us first account for the measure of n_j blocks damaged by modifications (2) and (3) at stage i (i.e., the synchronization and the coding of A_i). These distortions are grouped into a continuous block of overall length $2^{-i-3}n_i$, and $\sim 2^{i+5}$ contamination blocks of length $2^{-2i-9}n_i$, all $2^{-i-5}n_i$ apart. The distortions affect only n_j-blocks within an n_j radius from this area, which measure no more than

$$\frac{2^{-i-3}n_i + 2n_j}{n_i} + 2^{i+5}\frac{2^{-2i-9}n_i + 2n_j}{n_i} = 2^{-i-3} + 2^{-i-4} + (2^{i+6}+2)\frac{n_j}{n_i}$$

By placing a growth demand on the sequence $\{n_i\}$ this magnitude can easily be made less than 2^{-i-1}.

In order to control the estimate of A_j lost in modification (1) in the construction of P_i, the key idea is to separate the handling of j's for which $n_j \ll L_i$ from the j's for which $L_i \ll n_j$. To carry out such a separation, we will require some growth condition of the type $L_{i-1} \lll L_i$, and make an additional requirement from the n_j 's, not to 'land too near' any L_i.

Suppose then, that $n_j \ll L_i$, by looking n_j places forward at the P_i-name of a point $x \in X$, we can make a prediction whether $x \in A_j$, and be wrong only at probability $\varepsilon_j^{(i-1)}$. Since $n_j/L_i \ll 1$, and we modify only a short (K_{i-1}) segment once every $(1/2)L_i$ places, we can safely bound the probability of predictions gone bad: only an order of magnitude of $\frac{K_{i-1}+2n_j}{(1/2)L_i}$ of the predictions is even effected by the M_i-compatibility modification (Modification 1). This bound can be made small enough by placing the restriction

$$\frac{K_{i-1}}{(1/2)L_i} < 2^{-i-1},$$

a priori, and replacing the vague condition $n_j \ll L_i$ by the concrete:

$$\frac{2n_j}{(1/2)L_i} < 2^{-i-1}.$$

So, the overall addition to the probability of error in estimating A_j from P_i in a sliding window of width n_j, is less then 2^{-i}, as needed.

Next, suppose that $L_i \ll n_j$. More specifically, suppose that

$$L_i < L_{i+1} < \cdots < L_{i+k} \ll 2^{-2j-9}n_j < n_j \ll L_{i+k+1}.$$

The initial coding of A_j was done via relabeling blocks of length at least $2^{-2j-9}n_j$, and in each such block we polluted a $\frac{K_{i-1}}{(1/2)L_i}$ fraction in the i-th compatibility modification, another $\frac{K_i}{(1/2)L_{i+1}}$ fraction in the $(i+1)$-th compatibility modification, etc. We can choose the lengths $\{L_i\}$ to begin with, so that

$$\sum_m \frac{K_m}{(1/2)L_{m+1}} < 0.001$$

and specifically,

$$\sum_{m=i}^{i+k} \frac{K_m}{(1/2)L_{m+1}} < 0.001.$$

Since the words $v = v_i$ and $u = u_i$ chosen at stage i for the synchronization are more than 0.001 apart, they can still be distinguished after the $(i, \ldots, i+k)$-th compatibility modifications. In this case, the compatibility modifications do not take any toll either in the ability to synchronize the base nor in the ability to estimate A_j. In any case, the construction is complete. One can also easily make precise now the vague demands posed earlier: $L_{i-1} \lll L_i$, and the n_j's do not 'land too near' any L_i; e.g., it would suffice to require that for every n_j there would exist L_m such that

$$L_m < \frac{1}{100} 2^{-2j-9} n_j < n_j < \frac{1}{100} L_{m+1}.$$

To sum up, we set out to form inductively a convergent sequence of binaric partitions $\{P_i\}$, that would satisfy for every $j < i$:

$$A_j \overset{2^{-j}+2^{-j-1}+\cdots+2^{-i}}{\subset} \bigvee_{k=0}^{n_j-1} T^k P_i.$$

The induction can begin with $P_1 = \{A_1, A_1^C\}$ since M_1 is the full shift. The modification of P_{i-1} into P_i was held while enforcing (a) M_i compatibility, and (b) an overall loss of accuracy in estimating A_j (for any former j) of magnitude bounded by 2^{-i}. And thus, the construction is complete. ∎

4.2. The positive entropy case.

Theorem 1: *Fix some $0 < \varepsilon$. There exists a minimal subshift $M \subset \{1, \ldots, a\}^{\mathbb{Z}}$ which is universal for aperiodic standard ergodic systems with entropy $<$ $(\log a - \varepsilon)$.*

Proof: The basis for the proof is the construction in Section 3.2, along with techniques for M_i compatibility taken from Section 4.1. Basically, all injected blocks turn a bit thicker. We will survey the construction more briefly, emphasizing only its novel elements. M would be constructed again as $\bigcap_i M_i$, where

$M_1 = \{1, \ldots, a\}^{\mathbb{Z}}$ and each M_i is generated from M_{i-1} by the additional demand that every $w \in BL_i(M_{i-1})$ occurs in every $w' \in BL_{L_i}(M_i)$. As before, any word in M_{i-1} can be modified to belong to M_i by changing K_{i-1} places once every $(1/2)L_i$ (injecting the concatenated word w_{i-1} along with padding transition blocks, once every $(1/2)L_i$ places). We demand further that for every i, $h_{\mathrm{TOP}}(M_i) > (\log a - \varepsilon/2)$.

Let $(X, \mathcal{B}, \mu, T)$ be an aperiodic, standard, ergodic dynamical system. Fix a generating sequence of partitions $Q_1 \subset Q_2 \subset Q_3 \subset \cdots$, (i.e., $\bigvee Q_i \stackrel{0}{=} \mathcal{B}$), and assume that $\{H_T^{Q_i}\}$ is ascending to $h(T)$. We can assume for convenience that Q_1 contains precisely a sets, and since M_1-compatibility is not an issue, we can simply set P_1 to be Q_1.

Suppose $P_1, \ldots, P_{i-1}$ are formed. Set $h_i = H_T^{P_{i-1} \vee Q_i} - H_T^{P_{i-1}}$, $C = 10/\varepsilon$. Reserve the lowest cells to identify the base uniquely, again, by labeling them with successive occurrences of a periodic short word v, and polluting the remainder of the tower with another, 2δ-distant word u (δ will be set immediately).

For large enough n_i, there are approximately $\sim 2^{n_i(Ch_i + 2^{-i}) \cdot H_T^{P_{i-1}}}$ typical P_{i-1}-configurations over the lowest unmodified $n_i(Ch_i + 2^{-i})$ cells, and upon every such prefix, there are $\sim 2^{n_i h_i}$ typical Q_i-configurations of length n_i. As before, we need to compress slightly the prefix's P_{i-1} configuration, and squeeze the compressed Q_i configuration (of the entire n_i word) into the saved space. An additional degree of complexity arises, since we must perform this compression using a robust M_i-code, i.e., enough properly spaced M_i-words. To convince ourself that this is possible, let us check again on the key estimate in Theorem 2.10, or rather an easy improvement of it

$$U(n) + V(2\delta) + nW(2\delta) + \log K < nh(M_i),$$

where K is the required vocabulary size. In the context of coding the prefix's P_{i-1} configuration, $\log K \sim nCh_i \cdot H_T^{P_{i-1}}$, and so if $Ch_i \cdot H_T^{P_{i-1}} < h(M_i)$ then δ can be made small enough for the entire estimate to hold. Since by an initial condition on each of the subshifts: $H_T^{P_{i-1}} < h(M)$, this bound can be achieved if

$$Ch_i = \frac{10(H_T^{P_{i-1} \vee Q_i} - H_T^{P_{i-1}})}{\log a - h(T)} < 1$$

which boils down to

$$H_T^{P_{i-1}} > h(T) - \frac{1}{10}(\log a - h(T)).$$

This can be guaranteed by a proper selection of, say, P_2 — it only needs to estimate well enough an appropriate Q_{i_k}.

To sum up: we can indeed code the P_{i-1} configuration of the prefix while maintaining a slight compression and M_i compatibility. In the process of doing that, we may need to set a δ smaller than 0.001, but δ can be bounded from below and thus be chosen in advance (the lengths L_i, and thus the subshifts M_i themselves, need to be chosen a priori accordingly). We can squeeze an M_i-compatible coding of the Q_i configuration of the entire column into the remainder of the prefix. The rest of the modification procedure goes practically unaltered. ∎

4.3. The finite/infinite entropy case. In [We89], Weiss presented a minimal system which is universal in the category of ergodic systems with any entropy, which was realized as a subshift of $\overline{\mathbb{N}}^{\mathbb{Z}}$. The approach we will present here is somewhat different from the original one, the difference arising in the stage of maintaining former estimates: the original proof relied on *recoding* information lost in the i-th modification, within the remaining ∞ symbols. In this work, instead of recoding lost information, we will just use robust codes to begin with.

The minimal subshift we use is again an intersection $\bigcap M_i$, with M_i of the type described in Example 2.1: for some fixed sequence N_i, M_i is actually derived from a subshift $\tilde{M}_i$ defined on the finite symbol set $\{1, 2, \ldots, N_i, \infty\}$, and then the symbol ∞ is let loose to assume any value in $\{N_i + 1, N_i + 2, \ldots, \infty\}$. M_i is obtained from M_{i-1} by the additional restriction that the single word w_{i-1}, which contains all words of length $i-1$ with symbols $\{1, 2, \ldots, N_{i-1}\}$ that actually occur in M_{i-1}, appears in every block of length L_i. We recall that in order to modify an M_{i-1} sequence to become M_i compatible, we need only make modifications of length K_{i-1} once every $(1/2)L_i$. We also demand that for every i, $h(\tilde{M}_i) > \log N_i - \varepsilon$ for a small prefixed ε.

All the M_i thus constructed have infinite entropy; so there is no a-priori bound on the size of the vocabulary that requires coding at stage i. However, the structure of M_i comes to our aid — P_i will contain only $N_i + 1$ nonempty atoms (this is the minimum required to guarantee M_i compatibility). So, we bound the amount of relevant n_i-words at the i stage, and are able to code an A_i configuration of a long n_i column using just the symbols $\{N_{i-1}, \ldots, N_i - 1\}$ within a short (say $2^{-i-3}n_i$) prefix, and a δ-robust code.

Given $(X, \mathcal{B}, \mu)$ a standard aperiodic space with an ergodic Borel automorphism T, suppose that $P_1, P_2, \ldots, P_{i-1}$ and $n_1, \ldots, n_{i-1}$ have been constructed such that every P_j is M_j-compatible, and for every $j < i$

$$A_j \overset{2^{-j}+2^{-j-1}+\cdots+2^{-i+1}}{\subset} \bigvee_{k=1}^{n_j} T^k P_{i-1}.$$

We need to construct P_i with identical properties.

To that purpose, take n_i to be large with respect to n_{i-1} (exactly how large will become clear in the process), and a special $(n_i, n_i + 1)$ Rokhlin base B.

We reserve $2^{-i-4}n_i$ places to code the location of the base, in the manner demonstrated in 4.1; i.e. by usage of successive occurrences of a periodic word v and pollution of the column with a δ-distant word u.

M_i compatibility would be achieved again via changes of K_{i-1} cells once every $(1/2)L_i$, and in addition fixing the first and last K_{i-1} cells in every column to begin with w_{i-1} and end with a transition block to it. We can set the ratio $\frac{K_{i-1}}{(1/2)L_i}$ in advance to be less than δ, and in fact to make each coding resistant to all future compatibility modifications we would require $\sum_i \frac{K_{i-1}}{(1/2)L_i} < \delta$.

To code the A_i configuration using 2δ-apart M_i words, we reserve, say, another $2^{-i-4}n_i$ places (including two transition blocks, of fixed length). We actually utilize only $\tilde{M}_i$ words with N_i symbols. An easy adaptation of Theorem 2.10 to the new setup, reveals that a demand of the form $2 \cdot 2^{i+4} < N_i$ would suffice to that purpose.

The rest of the proof — specifically, the control of former estimates — goes on similarly to Section 4.1, i.e.: separating the cases $n_j \ll L_i$ from $L_i \ll n_j$, bounding the damage caused in the first case, and in the second case — using the fact that the overall compatibility modifications are less than δ in magnitude.

References

[We89] B. Weiss, *Countable Generators in Dynamics - Universal Minimal Models*, Contemporary Mathematics **94** (1989), 321–326.

[We00] B. Weiss, *Single Orbit Dynamics*, American Mathematical Society, Providence, RI, 2000.

[DGS] M. Denker, C. Grillenberger and C. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes in Mathematics 527, Springer-Verlag, Berlin–New York, 1976.

[Sh] P. Shields, *The ergodic theory of discrete sample paths*, American Mathematical Society, Providence, RI, 1996.

[Bi] P. Billingsley, *Ergodic Theory and Information*, John Wiley & Sons, Inc., New York–London–Sydney, 1965.

[Ha] P. Halmos, *Lectures on Ergodic Theory*, Chelsea Publishing, New York, 1956.