

**INFORMATION COMPRESSION AND RETENTION IN
DYNAMICAL PROCESSES
DRAFT—February 28, 2001**

KARL PETERSEN

ABSTRACT. We discuss some recent work on various constructions that accumulate or remove information within dynamical systems: tail fields, numeration systems and formal languages (especially of β -shifts), and factor mappings between symbolic or tiling dynamical systems.

1. INTRODUCTION

“Information” can be defined precisely and measured in terms of various kinds of entropy. In a measure space, information available can also be represented by a sub- σ -algebra \mathcal{F} of the full σ -algebra \mathcal{B} of measurable sets: the idea is that we do not know everything about a point x , but for each $F \in \mathcal{F}$ we do know whether or not $x \in F$. Information can be revealed, lost, transmitted, stored, presented in different forms—in a deterministic manner, or subject to random effects. In these notes we discuss some recent investigations of several kinds of information handling that occur in symbolic dynamics and abstract ergodic theory.

Many of our examples will be subshifts, closed shift-invariant subsets of the set of (one or two-sided) infinite sequences on a finite alphabet. To fix notation, with $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$ and $\mathbb{N} = \{1, 2, 3, \dots\}$ for each integer $d \geq 2$ let

$$(1.1) \quad \begin{aligned} \Sigma_d^+ &= \{0, 1, \dots, d-1\}^{\mathbb{Z}^+}, & \Sigma_d^1 &= \{0, 1, \dots, d-1\}^{\mathbb{N}}, \\ \text{and } \Sigma_d &= \{0, 1, \dots, d-1\}^{\mathbb{Z}}. \end{aligned}$$

Each is given the product topology, so that it is a compact metric space. The shift transformation defined by

$$(1.2) \quad (\sigma x)_k = x_{k+1} \quad \text{for all } k$$

is a homeomorphism on Σ_d and a continuous d -to-1 map on Σ_d^+ for each n . A *subshift* is a closed subset X such that $\sigma X \subset X$. A *subshift of finite type* is a subshift consisting of all sequences (in Σ_d^+ or Σ_d for some d) that do not contain any member of a certain finite list of forbidden words (finite strings on the alphabet $\{0, 1, \dots, d-1\}$). By passing to a higher block presentation, a standard type of recoding, we may assume (since we arrive in this way at a topologically conjugate system) that the forbidden words all have length 2. Thus the subshift consists of all sequences which do not include any disallowed transitions between symbols of the alphabet. The pattern of allowed transitions is described by a $d \times d$ -dimensional 0,1 matrix A : the entry A_{ij} is 1 if the word ij is allowed, 0 if it is forbidden. Subshifts and subshifts of finite type are used to encode complicated dynamical systems so as to make them available for combinatorial analysis. For general background on symbolic dynamics, see [64, 89].

We work with measure-preserving dynamical systems (X, \mathcal{B}, μ, T) , in which (X, \mathcal{B}, μ) is a measure space and $T : X \rightarrow X$ is a measure-preserving transformation, usually invertible. A topological dynamical system will usually be a pair (X, T) with X a compact metric space and $T : X \rightarrow X$ a homeomorphism, but we will also consider situations in which T is just a continuous map, or a group or more general set of continuous maps on X . Background on measure-theoretic, topological, and smooth dynamical systems can be found in [87, 53, 88].

I thank the colleagues with whom I have had the pleasure of investigating these topics, especially Terry Adams, Kim Johnson, Natalie Priebe, Klaus Schmidt, Kennan Shelton, Sujin Shin, and Jean-Paul Thouvenot. I am also grateful to the University François Rabelais (Tours, France) and the University of Chile (Santiago) for support and hospitality and for inviting the lecture series that induced me to write these notes.

2. TAIL FIELDS

2.1. Tail fields and equivalence relations. Let $\dots, x_{-1}, x_0, x_1, x_2, \dots$ be a stationary stochastic process that takes values in a finite set $D = \{0, 1, \dots, d-1\}$. The usual tail fields associated with the process are

$$(2.1) \quad \begin{aligned} \text{the future tail field } \mathcal{F} &= \bigcap_{n=1}^{\infty} \mathcal{B}(x_n, x_{n+1}, \dots), \\ \text{the past tail field } \mathcal{P} &= \bigcap_{n=1}^{\infty} \mathcal{B}(\dots, x_{-n-1}, x_{-n}), \\ \text{and the remote tail field } \mathcal{R} &= \bigcap_{n=1}^{\infty} \mathcal{B}\{x_j : |j| \geq n\}. \end{aligned}$$

These tail fields contain the information that remains when the transient present has passed from the scene. Each of these tail fields is also the σ -algebra of invariant (or saturated) sets for an associated Borel equivalence relation, for the action of a group of measurable transformations, and for the action of a single measurable transformation. (See [90] for a quick review of Borel equivalence relations and [29, 30, 100] for more background.) For example, \mathcal{F} is the family of invariant sets for the *homoclinic relation*

$$(2.2) \quad \begin{aligned} x \sim_{\Gamma^+} y &\text{ if and only if there is } n \text{ such that} \\ &x_j = y_j \text{ whenever } j \geq n; \end{aligned}$$

for the group Γ of *finite coordinate changes* (the same as adding in mod d coordinatewise an element of Σ_d that is identically 0 from some point on); and for the *odometer*, or *von Neumann-Kakutani adding machine* (the transformation that increases the first entry that is less than $d-1$ and sets all preceding entries to 0).

It is interesting and sometimes necessary to consider some finer tail fields which keep track of more information. For example, in statistical mechanics the symbols of the alphabet might represent individual particles which perhaps cannot be created or destroyed, so one might want to consider the sets that are invariant under the group Π of *permutations* of finitely many coordinates. The corresponding equivalence relation $x \sim_{\psi} y$ is generated by a cocycle, determined by a function $\psi : \Sigma_d \rightarrow \mathbb{Z}^d$ as follows. We let $\psi(x) = e^{x_0}$, where e^0, e^1, \dots, e^{d-1} are

the standard basis vectors in \mathbb{R}^d . Then

$$(2.3) \quad \psi_0^n(x) = \sum_{j=0}^n \psi(\sigma^j x)$$

is the d -dimensional vector whose i 'th entry tells the number of times that the symbol $i = 0, \dots, d-1$ appears in the string $x_0 x_1 \dots x_n$. We put

$$(2.4) \quad x \sim_\psi y \text{ if and only if } x \sim_\Gamma y \text{ and } \sum_{j=-\infty}^{\infty} [\psi(\sigma^j x) - \psi(\sigma^j y)] = 0.$$

(Note that the sum only includes finitely many nonzero terms.) Then the family \mathcal{F}_ψ of sets that are saturated with respect to the equivalence relation \sim_ψ coincides with the Π -invariant sets, the sets that are fixed by every permutation of finitely many coordinates. Definitions in the one-sided cases are analogous; for example,

$$(2.5) \quad x \sim_{\psi^+} y \text{ if and only if } x \sim_{\Gamma^+} y \text{ and } \sum_{j=0}^{\infty} [\psi(\sigma^j x) - \psi(\sigma^j y)] = 0,$$

with \mathcal{F}_ψ^+ the associated family of saturated sets. This family is often called the σ -algebra of *exchangeable* or *symmetric* sets. The definitions extend naturally to higher-dimensional actions and to cocycles generated by functions ψ taking values in other groups; more about this later.

2.2. The Pascal adic transformation. A very interesting single measurable transformation whose orbits are the same as Π -orbits is the *Pascal adic transformation* defined by Vershik and his collaborators [116, 66]. Let us describe this in case $d = 2$, when $T : \Sigma_2^+ \rightarrow \Sigma_2^+$ can be defined by its action on cylinders according to the formula

$$(2.6) \quad T(0^p 1^q 10 x_{p+q+2} x_{p+q+3} \dots) = 1^q 0^p 01 x_{p+q+2} x_{p+q+3} \dots$$

The action of T can also be defined in terms of a certain partial order on Σ_2^+ . Let us agree that $x, y \in \Sigma_2^+$ are *comparable* if they agree from some point on, i.e. $x \sim_{\Gamma^+} y$, and that $x < y$ if x and y are comparable and, if n is the last time when $x_n \neq y_n$, then $x_n < y_n$ (with $0 < 1$ as usual). Then we define Tx to be the smallest y that is $> x$, if there is one. There remain countably many points (the $1^k 0000 \dots$) for which T is not defined, and also countably many which have no preimages; T is a homeomorphism on the rest of Σ_2^+ .

The transformation T can also be thought of as acting on the space X of infinite paths $x_1 x_2 \dots$ in the *Pascal graph*. This is an infinite

directed graded graph with $n + 1$ vertices labeled (n, k) , $k = 0, 1, \dots, n$ at each level $n = 0, 1, 2, \dots$, and connections from (n, k) to $(n + 1, k)$ (labeled 1) and $(n + 1, k + 1)$ (labeled 0) for all $n = 0, 1, \dots$ and $k = 0, \dots, n - 1$. The labeling provides a one-to-one correspondence between Σ_2^+ and infinite paths starting at level 0. We denote the vertex at level n of a path x by $(n, k_n(x))$.

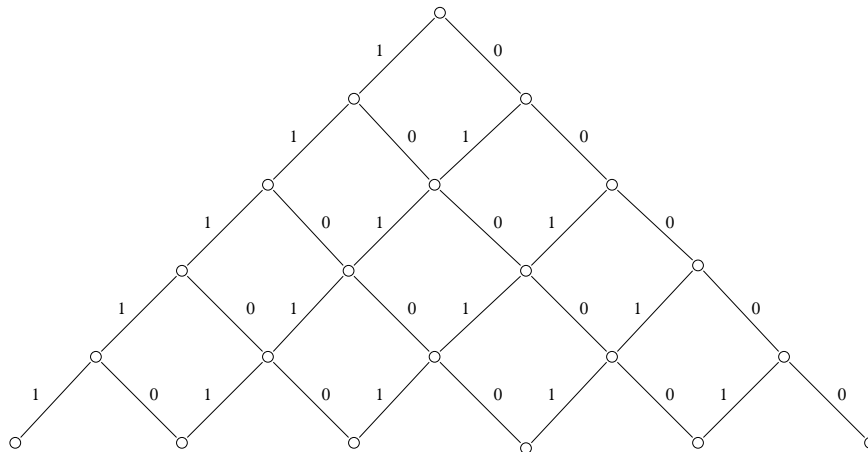


FIGURE 1. The Pascal graph.

The following figure illustrates the action of T . Notice that com-

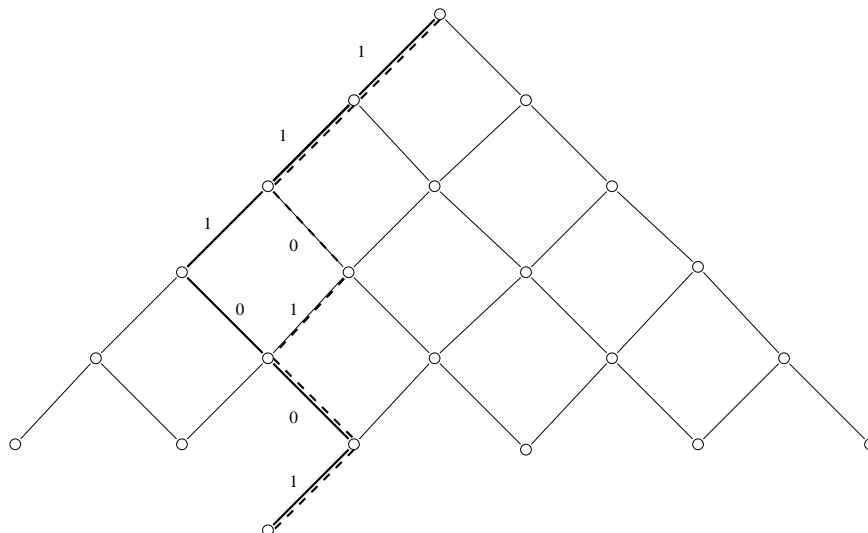


FIGURE 2. $T(111001\dots) = 110101\dots$

ing down to a vertex (n, k) there are $C(n, k)$ (the binomial coefficient

$n!/([k!(n-k)!])$ paths of length n , and that T maps the corresponding cylinder sets in order each to the next, beginning with $1^k 0^{n-k}$ and ending with $0^{n-k} 1^k$. Such sequences of cylinder sets, beginning with a minimal path down to a vertex at level n and ending with a maximal one into that vertex, each mapping to the next, correspond to columns in a cutting and stacking representation of the system (X, T) . Combinatorial, asymptotic, and divisibility properties of binomial coefficients determine the dynamical properties of the system.

2.3. Other adic transformations. The general class of adic transformations has been defined by A. Vershik and promoted as an alternative framework for the cutting and stacking constructions of ergodic theory that is maybe more amenable to combinatorial and geometric analysis. The viewpoint also connects with the theory of C^* algebras and group representations, leads quickly to many fascinating examples, and naturally facilitates the study of orbit equivalence in both the measure-theoretic and topological settings [43, 38, 39, 32, 18, 75, 26]. The basis for an adic system is a graded directed graph beginning with a single root vertex. Each level contains finitely many vertices, there are connections only from each level to the next one, and the set of edges into each vertex is ordered. Such a graph is sometimes called a “Bratteli diagram”; it can represent the ramification diagram of representations of a group, each connection showing the embedding of a representation in one of higher dimension. The space X is again the space of infinite paths that begin at the root. It is a compact metric space in the usual way. The transformation T is again defined as above, where possible.

The adic transformation T is in a natural way *transverse* to any shift transformation defined on a space of labelings of the paths in X by symbols from some alphabet, just as the horocycle flow on a surface acts in a sideways manner in relation to the geodesic flow. See [107, 58, 59, 60, 48] for previous discussions of transverse actions. We can see this relation already in the case of the simplest stationary adics. An adic system is called *stationary* if after the root vertex each level has the same number of vertices and the same system of edges from each level to the next. Stationary adic systems correspond in a natural way to substitution dynamical systems and odometers—see [116, 65, 108, 32, 46, 26]. In the case of the full 2-shift Σ_2^+ , the adic transformation T is the ordinary odometer or von Neumann-Kakutani adding machine. T is uniquely ergodic, with unique invariant measure $\mu_{1/2}$, the Bernoulli (i.i.d.) measure that assigns equal measure $1/2$ to

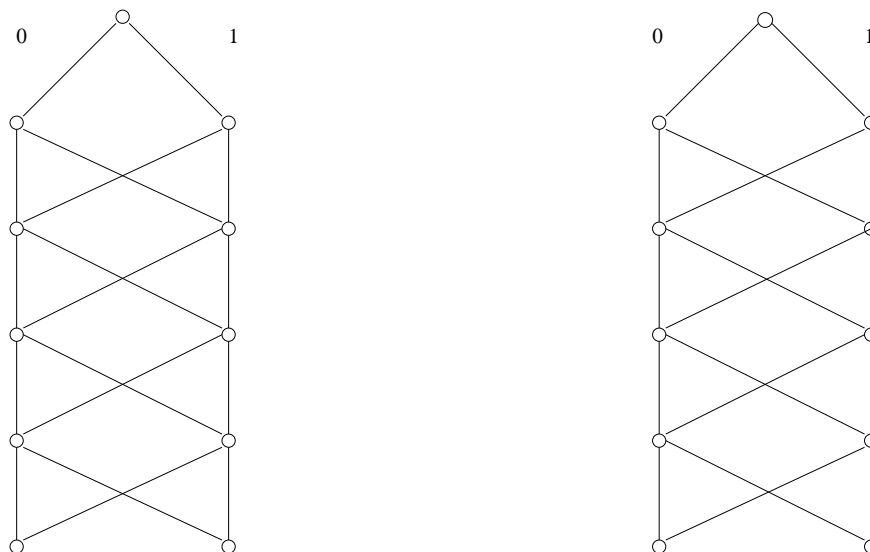


FIGURE 3. Bratelli diagrams for the odometer and Fibonacci substitution

each of the symbols 0 and 1, and thus equal measure $1/2^n$ to each n -block. The T -invariant sets are the future tail field \mathcal{F} ; thus a measure μ on Σ_2^+ that is quasi-invariant under the odometer T will be ergodic for T —i.e., every invariant set will have measure 0 or have complement with measure 0—if and only if it is K for the shift σ —i.e., the tail \mathcal{F} is trivial in the sense that it consists only of sets of measure 0 and their complements. T has entropy 0 with respect to its unique invariant measure, while the shift of course has completely positive entropy (with respect to $\mu_{1/2}$, all nontrivial factors have positive entropy).

A similar pattern holds for the stationary adic given by the graph on the right in Figure 3. The shift on the space of paths labeled by 0's and 1's is the golden mean SFT Σ_A consisting of all sequences which do not contain the block 11. The adic is isomorphic to the shift on the orbit closure of the fixed point of the primitive “Fibonacci substitution” $0 \rightarrow 01, 1 \rightarrow 0$, which is in turn isomorphic to translation modulo 1 by the golden mean. The adic is uniquely ergodic with unique invariant measure equivalent to the Shannon-Parry measure (measure of maximal entropy) on Σ_A ; it is a Markov measure with the same stochastic transition matrix (the “stochasticization” of A —see [81, p.23]) but a different initial probability vector, so as to give all cylinders down to a fixed vertex exactly the same measure, not just comparable measures (within constant multiples of one another) as does the maximal measure.

According to a version of the Jewett-Krieger Theorem due to Vershik, within the class of adic systems we can find representatives of all ergodic systems.

Theorem 2.1 (Vershik [112]). *Every ergodic measure-preserving transformation on a Lebesgue space is isomorphic to a uniquely ergodic adic transformation. Moreover, the isomorphism can be chosen so that a given countable dense invariant subalgebra of the measurable sets goes over into the algebra of cylinder sets.*

Question 2.2. Which ergodic nonsingular systems (X, \mathcal{B}, μ, T) arise as the adic transformations transverse to measure-preserving shifts? Are there such representations that are minimal in some way, for example such that the shift has minimal entropy? (If so, the minimal entropy may serve as a definition of entropy for the nonsingular system.)

As an example of the wonderful and mysterious dynamical systems that arise from the adic viewpoint we present the graph of Young diagrams.

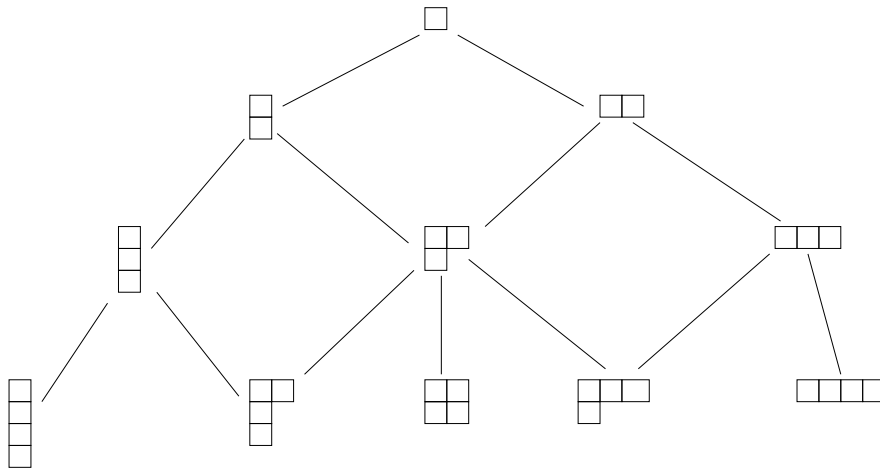


FIGURE 4. The beginning of the graph of Young diagrams.

The Pascal graph is a subgraph of this one. The dynamics of the adic and shift on this graph involve many fascinating topics, such as the *jeu de taquin*, the Robinson-Schensted-Knuth algorithm, distribution of longest monotonic subsequence of a random sequence, representations of symmetric groups, partitions of integers, and so on—see [113, 114, 115, 55].

2.4. The Pascal adic and the Hewitt-Savage 0,1 Law. We return now to our discussion of tail fields, the Pascal adic system (X, T) , and 0,1 laws.

Theorem 2.3 (de Finetti, Hajian-Ito-Kakutani, Vershik). *The ergodic invariant measures for the Pascal adic transformation are exactly the Bernoulli measures $\mu_\alpha = \mathcal{B}(\alpha, 1 - \alpha)$ for $0 \leq \alpha \leq 1$.*

We ascribe to de Finetti the statement that the only ergodic T -invariant measures are the μ_α , and that thus every exchangeable process (one with distribution invariant under all permutations of finitely many coordinates) is a mixture of independent processes (see [4] and the references given there and in [90]). The ergodicity of the μ_α for a transformation isomorphic to T was proved by Hajian, Ito, and Kakutani [42] and was used by Kakutani to prove the equidistribution of the division points in his interval-splitting procedure [52]. The connection with the Pascal adic transformation presented in this form was made by Vershik [111]. The ergodicity of the μ_α is equivalent to the Hewitt-Savage 0,1 Law, which can be regarded as a strengthening of the Kolmogorov 0,1 Law.

The Kolmogorov 0,1 Law states that if (x_k) is an independent process, then \mathcal{F} is trivial. So are \mathcal{P} and \mathcal{R} .

The Hewitt-Savage 0,1 Law states that if (x_k) is an independent process, then \mathcal{F}_ψ^+ is trivial. This was extended to finite-state Markov chains by Blackwell and Freedman [12] and to Gibbs states by Georgii [37]. Further results on these questions are found in [24, 7].

2.5. Some general 0,1 laws. Symbolic dynamics allows us to establish a theorem that unifies many 0,1 laws and extends them to subshifts of finite type.

Theorem 2.4. [90] *Let (Σ_A, σ) be a topologically mixing subshift of finite type and μ a shift-invariant Gibbs measure determined by a potential function with summable variation. Let G be a countable discrete group with finite conjugacy classes and $\psi : \Sigma_A \rightarrow G$ a continuous map. Then \mathcal{R}_ψ is trivial.*

A similar statement holds for the one-sided case, subject to a topological transitivity condition. Thus we find adic transformations on SFT's that are nonsingular and ergodic with respect to Gibbs measures.

Pushing the viewpoint in [90] farther, Schmidt extended the preceding theorem to the case of higher-dimensional actions (for abelian G)

[101]; he also *explained it* and extended other theorems about tail fields (see for example [7]) by showing that in many cases the usual remote field \mathcal{R} and the fine remote field \mathcal{R}_ψ coincide up to sets of measure 0.

For sub- σ -algebras \mathcal{F}_1 and \mathcal{F}_2 of the σ -algebra of measurable sets in a measure space (X, \mathcal{B}, μ) , write $\mathcal{F}_1 \stackrel{\circ}{=} \mathcal{F}_2$ if for each $F_1 \in \mathcal{F}_1$ there is $F_2 \in \mathcal{F}_2$ such that $\mu(F_1 \Delta F_2) = 0$ and vice versa.

Theorem 2.5. [102] *Let $\{x_n : n \in \mathbb{Z}\}$ be a two-sided stationary process taking values in a finite set D and $\psi : D \rightarrow G$ a continuous function from the discrete space D to a discrete countable group G with finite conjugacy classes. Define*

$$(2.7) \quad \mathcal{R} = \bigcap_{n=1}^{\infty} \mathcal{B}\{x_j : |j| \geq n\} \quad \text{and}$$

$$(2.8) \quad \mathcal{R}_\psi = \bigcap_{n=1}^{\infty} \mathcal{B}\{\psi(x_j) \dots \psi(x_0) \dots \psi(x_{-j}) : j \geq n\}.$$

Then $\mathcal{R} \stackrel{\circ}{=} \mathcal{R}_\psi$.

2.6. Philosophical interpretations. One may interpret these results as demonstrating the uselessness of history (of certain kinds, in certain circumstances) and the inefficacy of some types of scientific experiments. While little information may reside in the remote σ -algebra \mathcal{R} that lies beyond the horizons of all presents, we may try to preserve a limited amount of information by recording in a register (d -dimensional vector) the number of times that each symbol of the alphabet D has appeared. The more general products $\psi_{-j}^j(x) = \psi(x_j) \dots \psi(x_0) \dots \psi(x_{-j})$ can keep track of many different accumulated effects like twists and shapes. According to Theorem 2.5, the finer remote field \mathcal{R}_ψ will contain no more information than the original one \mathcal{R} . Moreover, for Gibbs measures there will probably be *no* residual information in either remote field.

The triviality of \mathcal{R}_ψ , which we know is equivalent to the ergodicity of an associated Borel equivalence relation and the action of a countable group of nonsingular transformations, has some striking interpretations in two examples discussed by K. Schmidt and S. Richardson [101, 102].

Suppose that the alphabet D labels a finite set of chemical building blocks, like the bases in DNA, so that a long string $x_{-n} \dots x_n$ on the symbols from D represents a macromolecule. If we are modeling strings that result from some chemical or evolutionary processes it may

be reasonable to assume that long strings look like the outputs of a stationary process whose distribution is a Gibbs measure. Membership of the string in a measurable set E corresponds to the string being in some particular physically observable state. Theorem 2.4 says that if we take a finite segment of such a string, break it up and spin it in a centrifuge, and count the number of each type of base present, we learn nothing—because of ergodicity of the equivalence relation, with probability 1, consistently with the information obtained the string could be in any state E of positive measure whatsoever. A continuous function ψ from Σ_d to a countable discrete group with conjugacy classes could indicate for example changes in direction as we move along the macromolecule. Then ergodicity would imply that even with added stereochemical information accumulated across a finite segment we could not determine anything about the actual state of the (effectively infinitely long) complete macromolecule.

A second example [101] relates to percolation in the plane. Suppose that at each site in the lattice \mathbb{Z}^2 , independently of everything else, either a black or white particle is deposited, each with probability $1/2$. The configuration space is $\Omega = \{0, 1\}^{\mathbb{Z}^2}$, and there is a measure-preserving action of \mathbb{Z}^2 on Ω generated by the horizontal and vertical shifts. It is known that with probability 1 the monochromatic clusters are bounded: each island of white is bounded and completely surrounded by black, each black island completely surrounded by white. It is possible now to form an abstract group G which keeps track of the colors, sizes, and shapes of connected components: let G be the free abelian group generated by the set of all pairs (i, C) where $i \in D = \{0, 1\}$ and C is a connected subset of the lattice graph of \mathbb{Z}^2 which contains the origin. Define $\psi : \Omega \rightarrow G$ by letting $\psi(\omega) = (i, C)$ with i giving the color that ω assigns to the origin (0 for black, 1 for white) and C the monochromatic component of ω that contains the origin. Then Schmidt's higher-dimensional extension of (2.4) says that given any positive-measure set $E \subset \Omega$ of configurations, for almost every configuration $\omega \in \Omega$ it is possible to snip out finitely many monochromatic components, rearrange them, and place them back down in the plane in such a way that (i) they again fit together exactly, (ii) again each island of white is completely surrounded by black and each island of black completely by white, and (iii) the resulting configuration $\omega' \in E$. In a physical interpretation, the model may describe a material consisting of particles of two kinds, a magnetic medium, or an array of particles with two kinds of spin. In an

equilibrium situation the statistics may be described by a Gibbs measure. The triviality of \mathcal{R}_ψ means that we cannot determine anything about the state of the system by counting the types of monochromatic clusters in a finite regions.

2.7. The idea behind the proofs. The symbolic-dynamic idea behind the proofs of these results is a development of arguments already used long ago to prove the Hewitt-Savage Theorem (see for example [31, p. 122]). Let E and F be measurable sets of configurations. Approximate E and F to a high degree of accuracy by unions of cylinder sets E' and F' . We change the cylinders comprising F' by changing finitely many coordinates near the origin to make them look like the cylinders comprising E' , compensating by changing some coordinates far enough out to be beyond the range of definition of the cylinder sets involved, so as to produce a finite *permutation* of the entries. By quasi-ergodicity we end up with an image of F under Π that hits E . To make this work precisely in the situations cited above and with relations determined by general $\psi : X \rightarrow G$ requires handling quite a few technical details.

2.8. Super- K . Return now to the special case when $\psi : \Omega \rightarrow \mathbb{Z}^d$ is the function that generates symbol counts. In spite of the preceding discussion, in general the triviality of \mathcal{F}_ψ^+ for a measure μ nonsingular with respect to the shift is a property stronger than the Kolmogorov property, and for the adic it is a stronger property than ergodicity.

A process may be thought of as a sequence of random variables on a probability space, or as the measure on (for example) $\mathbb{R}^{\mathbb{Z}}$ that describes the distribution of values of the process, or as a transformation and measurable partition on a measure space: every stationary process (x_k) taking values in a finite set $\{0, 1, \dots, d-1\}$ corresponds to a system (γ, T) , with $T : X \rightarrow X$ a measure-preserving transformation on a probability space (X, \mathcal{B}, μ) and γ a finite partition of X into measurable sets ($x_k = \gamma(T^k x) = i$ if and only if $T^k x \in G_i \in \gamma$). We will interchange these representations at will, displaying the partition involved when necessary. For example, $\mathcal{F}_{\psi_\gamma}^+$ denotes the fine future σ -algebra generated by keeping track of γ -symbol counts from time 0 to large times n .

If γ is a generating partition for T (the images of γ under all integer powers of T generate the full σ -algebra \mathcal{B}), then $\mathcal{F}(\gamma) = \mathcal{Z}$, the *Pinsker*

σ -algebra of the system (X, \mathcal{B}, μ, T) . For 0-entropy systems, $\mathcal{Z} \doteq \mathcal{B}$, while for Kolmogorov systems \mathcal{Z} is trivial.

Definition 2.6. We say that a process (γ, T) is *super-K* if $\mathcal{F}_{\psi_\gamma}^+$ is trivial.

Notice that the super-K property depends on the choice of partition γ , so that if two processes come from different generating partitions in the same measure-preserving system one might have this property and the other not. So we want to know which processes have super-K generators. But before leaving the example of the Pascal adic, we should also ask whether it might have any still stronger properties than ergodicity, with respect to any of its ergodic measures μ_α .

2.9. Point spectrum of the Pascal adic. It was conjectured by Vershik that the Pascal adic $T : X \rightarrow X$ is in fact weakly mixing with respect to each μ_α , and some efforts have been made to prove that it is even strongly mixing. Any such results would strengthen the 0,1 laws discussed above, and of course they would illuminate the dynamical properties of this fascinating system. We present now some partial results that we know of pertaining to this problem.

Proposition 2.7. *If $\lambda \in \mathbb{C}$ is an eigenvalue of the Pascal adic system (X, \mathcal{B}, μ, T) , then $\lambda^{C(n, k_n(x))} \rightarrow 1$ a.e. $d\mu_\alpha(x)$.*

Proposition 2.8. *The Pascal adic system has no eigenvalues that are roots of unity.*

Proof. In fact, if $\lambda \neq 1$, we cannot have $\lambda^{C(n, k_n(x))} \rightarrow 1$ along any path x . This follows from the well-known self-similar structures that result when all the binomial-coefficient entries in the Pascal triangle are reduced modulo a prime q , which are explained by the Kummer Carry Theorem and the resulting formula of Lucas.

Theorem 2.9 (Kummer's Carry Theorem [61]). *The exact power of a prime q that divides $C(n, k)$ is the number of carries used when adding k and $n - k$ base q .*

Theorem 2.10 (Lucas' Formula [67]:). *If $n = n_0 + n_1q + n_2q^2 + \dots$ and $k = k_0 + k_1q + k_2q^2 + \dots$ base q (so that $0 \leq n_j, k_j < q$ for all j), then*

$$(2.9) \quad C(n, k) \equiv_q C(n_0, k_0) C(n_1, k_1) \dots$$

Consequently, q divides $C(n, k)$ if and only if there is j with $k_j > n_j$. And so if all $n_j = q - 1$, which is the case when n is 1 less than a power of q , then $C(n, k)$ is relatively prime to q for all $k = 0, 1, \dots, n$.

Now if $\lambda^m = 1$, say $\lambda = e^{2\pi il/m}$, let q be a prime divisor of m . No matter what the path x is in the Pascal graph, whenever $n = q^r - 1$ for some r we have $C(n, k_n(x))$ not divisible by q , so that

$$(2.10) \quad \lambda^{C(n, k_n(x))} = e^{2\pi il C(n, k_n(x))/m} \neq 1$$

and hence is at least a distance $1/q$ from 1. \square

2.10. The distribution modulo 1 of binomial-coefficient multiples of an irrational. Pascal's triangle is built up by starting with a plane lattice full of 0 entries, introducing a single impurity 1, and then building rows downwards according to a law of development specified by the cellular automaton rule $x_0 + x_1$, equivalently the addition rule $C(n+1, k) = C(n, k-1) + C(n, k)$. This process can also be carried out modulo q , and in fact entire subtriangles can be regarded as units and added in pairs to produce lower ones. As the process evolves, we see larger and larger "voids" consisting of entries that are 0 modulo q , preceded by horizontal "blocking lines", at levels $n = q^r - 1$, of entries that are nonzero modulo q . Figures 5 and 6 illustrate the case $q = 3$.

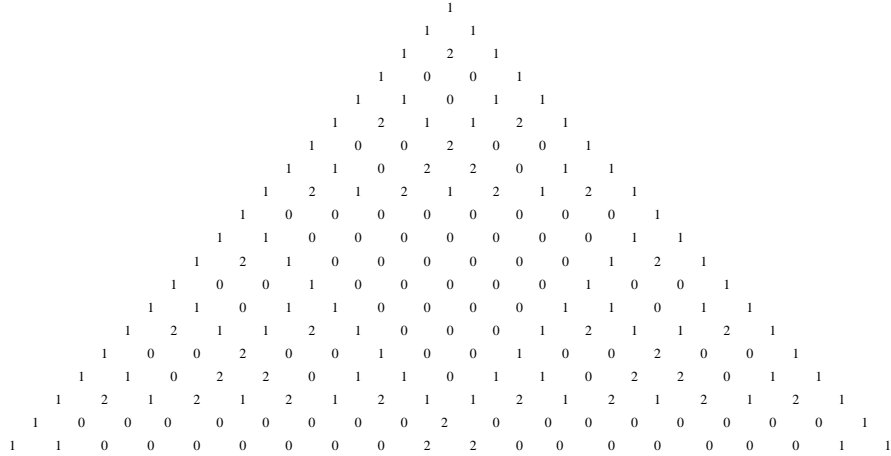


FIGURE 5. The Pascal triangle modulo 3.

What are the possibilities then for the distribution of $\lambda^{C(n, k_n(x))}$ on the unit circle if λ is not a root of unity, say for μ_α -a.e. x ? The following result shows that there are many λ which are candidates for eigenvalues, since for a.e. x the points $\lambda^{C(n, k_n(x))}$ spend too much time in every neighborhood of 1 (the upper frequency of visitation is 1). However, most of these also form a dense set in the circle, so these λ cannot be eigenvalues.

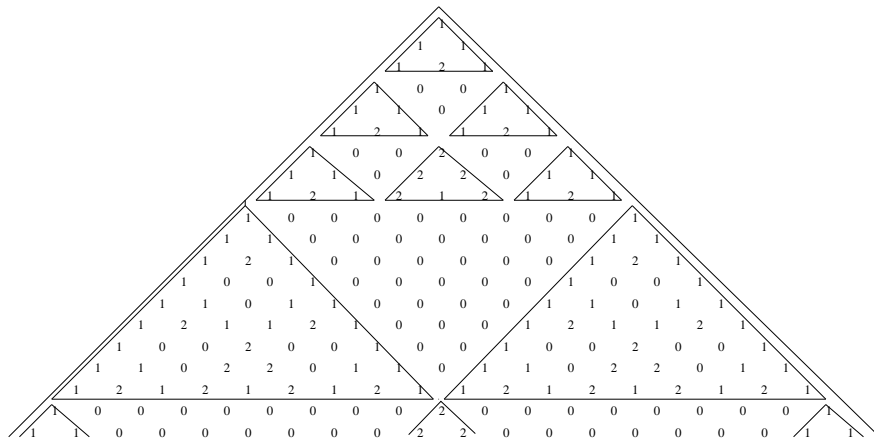


FIGURE 6. The self-similar, replicating structure of the Pascal triangle modulo 3.

Theorem 2.11. [3] *Given $\alpha \in (0, 1)$, there are a set $X_\alpha \subset X$ of paths in the Pascal graph of μ_α -measure 1 and a dense G_δ subset $\Lambda \subset S^1$ such that for all $x \in X_\alpha$ and $\lambda \in \Lambda$*

$$(2.11) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} |\lambda^{C(n, k_n(x))} - 1| = 0.$$

Proof. (1) The infinitely rescaled Pascal graph modulo q corresponds to a compact subset G_q of the plane which we call the Sierpinski q -gasket, the subset of an equilateral triangle which remains when smaller and smaller upside-down open equilateral triangles are removed. The removed triangles correspond to the voids (regions where $C(n, k)$ is divisible by q in the Pascal triangle), and displaying the gasket in finer and finer resolution corresponds to showing more and more of the Pascal triangle modulo q . The figures show the q -gasket in case $q = 2$, in two successive degrees of resolution, with a line L that cuts at a point α the bottom edge, taken to be the interval $(0, 1)$. By geometrical estimates we can see that most of the line L , in the sense of linear Lebesgue measure, lies in the removed triangles: $L \cap G_q$ has linear Lebesgue measure 0. When we interpret G_q as a representation of the full Pascal triangle, this implies that if x is an infinite path of “slope” α in the sense that $k_n(x)/n \rightarrow \alpha$, then

$$(2.12) \quad \frac{1}{n} |\{j : 0 \leq j < n, C(n, k_n(x)) \equiv_q 0\}| \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

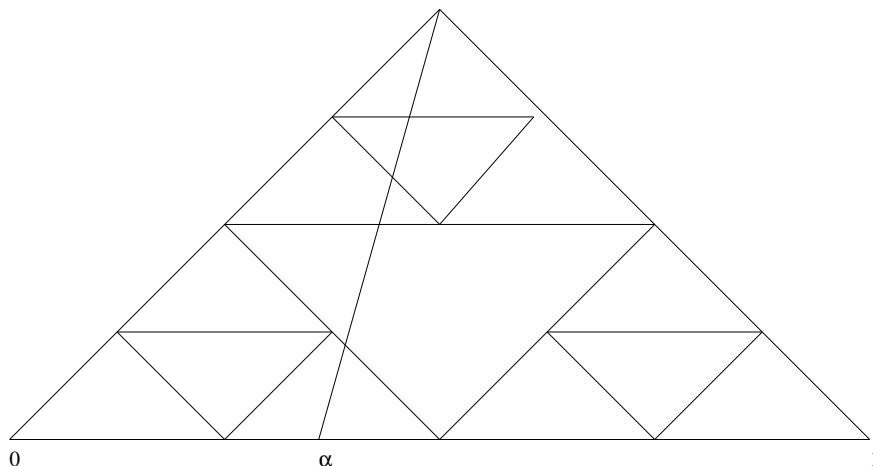
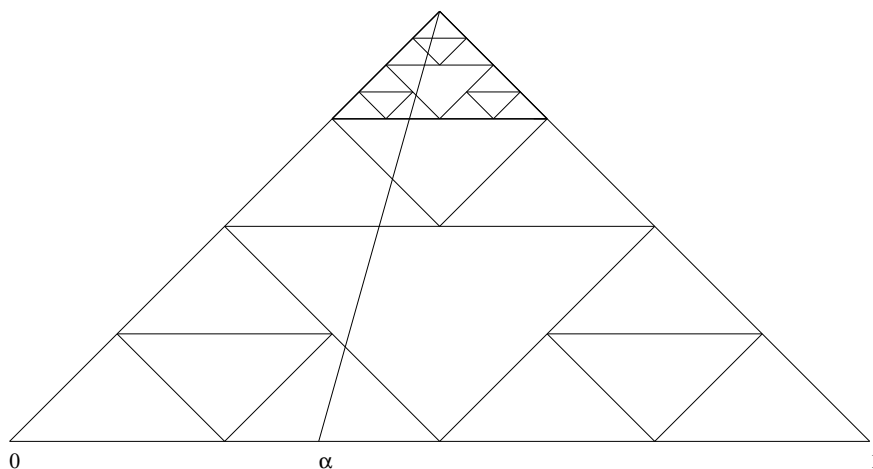


FIGURE 7. The Sierpinski 2-gasket.

FIGURE 8. The Sierpinski 2-gasket at better resolution.
The highlighted triangle corresponds to the one in the preceding figure.

(2) For a fixed prime q , for μ_α -a.e. x , for all $p = 0, 1, \dots, q - 1$,

$$(2.13) \quad \lim \frac{1}{n} \sum_{j=0}^{n-1} |e^{2\pi i C(j, k_j(x))p/q} - 1| = 0.$$

This is because by the Ergodic Theorem for a.e. x we have $k_n(x)/x \rightarrow \alpha$, so that if we look at a huge part of Pascal's triangle the path x will look like the line L in Figure 8: all of the wiggling that the path does occurs within the width of the line.

(3) Choose a sequence of primes q_m increasing to ∞ and positive ϵ_m with $\sum \epsilon_m < \infty$. Take R_m so large that if

$$(2.14) \quad X_m = \left\{ x \in X : \frac{1}{R_m} \sum_{j=0}^{R_m-1} |e^{2\pi i p C(j, k_j(x))/q_m} - 1| < \frac{1}{m} \right.$$

for all $p = 0, 1, \dots, q_m - 1$ }, then $\mu_\alpha(X_m) > 1 - \epsilon_m$.

(4) Now

$$(2.15) \quad \left\{ \frac{1}{R_m} \sum_{j=0}^{R_m-1} |e^{2\pi i C(j, k_j(x))\theta} - 1| : x \in X_m \right\}$$

is a *finite* set of continuous functions of $\theta \in [0, 1]$, since there are only finitely many possibilities for $\{C(j, k_j(x)) : j = 0, \dots, R_m - 1\}$, so if we let θ vary just a little from any p/q_m we will still have (2.14) holding with θ replacing p/q_m . So choose $\delta_m > 0$ such that if

$$(2.16) \quad \theta \in \Lambda_m = \bigcup_{p=0}^{q_m-1} \left(\frac{p}{q_m} - \delta_m, \frac{p}{q_m} + \delta_m \right),$$

then

$$(2.17) \quad \left| \frac{1}{R_m} \sum_{j=0}^{R_m-1} |e^{2\pi i C(j, k_j(x))\theta} - 1| < \frac{1}{m} \right.$$

(5) Define

$$(2.18) \quad \Lambda = \bigcap_{r=1}^{\infty} \bigcup_{m=r}^{\infty} \Lambda_m, \quad X_\alpha = \bigcup_{r=1}^{\infty} \bigcap_{m=r}^{\infty} X_m.$$

□

2.11. Questions about the distribution of binomial-coefficient multiples of irrationals modulo 1. The nature of the distribution of $\lambda^{C(n, k_n(x))}$ in the circle can also be studied in terms of a skew product transformation on the infinite torus. Let $\lambda = e^{2\pi i \alpha}$ with $\alpha \notin \mathbb{Q}$ and define $S : [0, 1]^\mathbb{N} \rightarrow [0, 1]^\mathbb{N}$ by

$$(2.19) \quad S(\theta_1, \theta_2, \dots) = (\theta_1 + \alpha, \theta_2 + \theta_1, \theta_3 + \theta_2, \dots) \pmod{1} \text{ in each coordinate.}$$

Then for example

$$(2.20) \quad S^n(0, 0, \dots) = (C(n, 1)\alpha, C(n, 2)\alpha, \dots).$$

It follows from Weyl's uniform distribution theorem that in each coordinate we see a sequence that is uniformly distributed in $[0, 1]$. Equivalently, if x is a path in the Pascal graph that is eventually diagonal, then for any irrational α the sequence $e^{2\pi i C(n, k_n(x))\alpha}$ is uniformly distributed in the circle. In fact, if we look at any finite set of coordinates we will see a sequence that is uniformly distributed in the torus of the appropriate dimension. Even more is true: $S : [0, 1]^{\mathbb{N}} \rightarrow [0, 1]^{\mathbb{N}}$ is *uniquely ergodic*, so that every orbit is uniformly distributed in the *full infinite-dimensional torus* [119, 33, 41, 91]. When we ask about the distribution of $e^{2\pi i C(n, k_n(x))\alpha}$ for other paths x in the Pascal graph, we are looking at the S -orbit of $(0, 0, \dots)$ but allowing our attention to shift one coordinate to the right from time to time, either at random or maybe according to some principle. A couple of the questions asked in [3] about the distributions of such observed sequences have been answered by D. Behrend, M. Boshernitzan, and G. Kolesnik [6]:

(1) $\{C(n, k_n(x))\alpha\}$ is uniformly distributed modulo 1 *for every irrational* α if and only if the path x in the Pascal graph is eventually diagonal.

(2) There is *no* α such that for every path x in the Pascal graph, except the two edge paths, $\{C(n, k_n(x))\alpha\}$ is uniformly distributed modulo 1.

Here are two more questions from [3]. The first might be easier to disprove than the existence of eigenvalues for T :

Conjecture 2.12. If $\lambda \in \mathbb{C}$ and there exists a path x in the Pascal graph such that $\lambda^{C(n, k_n(x))} \rightarrow 1$, then $\lambda = 1$.

The second was already asked by Erdős. It arises here from considering divisibility properties of binomial coefficients along the special central path in the Pascal triangle. (Thanks to M. Wierdl for discussions on this case.)

Question 2.13. Notice that $1 = 1$, $4 = 1 + 3$, and $256 = 1 + 3 + 9 + 243$. Are there any other solutions with $r \geq 0$ and distinct $s_1, \dots, s_m \geq 0$ of the equation

$$(2.21) \quad 2^r = 3^{s_1} + \dots + 3^{s_m}?$$

Finally, the Pascal adic (X, T) can be represented as a subshift of Σ_2^+ generated by a countable family ρ of substitutions, in an extension of the scheme described for example in [46] for associating an adic to a substitution (I believe that R. Burton and R. Kenyon have investigated this system from this point of view). For $n \geq k \geq 1$ define the mapping

ρ from symbols $\binom{n}{k}$ to strings of such symbols by

$$(2.22) \quad \binom{n+1}{k} \rightarrow \binom{n}{k-1} \binom{n}{k}$$

and

$$(2.23) \quad \binom{n}{0} \rightarrow \binom{n-1}{0}, \binom{n}{n} \rightarrow \binom{n-1}{n-1}, \binom{1}{1} \rightarrow 1, \quad \binom{1}{0} \rightarrow 0.$$

The substitution ρ expands strings by acting on the symbols that comprise them. For example,

$$(2.24) \quad \begin{aligned} \binom{4}{2} &\rightarrow \binom{3}{1} \binom{3}{2} \rightarrow \binom{2}{0} \binom{2}{1} \binom{2}{1} \binom{2}{2} \rightarrow \\ &\binom{1}{0} \binom{1}{0} \binom{1}{1} \binom{1}{0} \binom{1}{1} \binom{1}{1} \rightarrow 001011, \\ \binom{n}{1} &\rightarrow 0^{n-1}1 \quad \text{for each } n \geq 1, \\ \binom{n}{2} &\rightarrow \binom{n-1}{1} \binom{n-2}{1} \dots \binom{1}{1} \rightarrow 0^{n-2}1 0^{n-3}1 \dots 0^31 0^21 011, \\ \binom{n}{3} &\rightarrow \binom{n-1}{2} \binom{n-2}{2} \dots \binom{3}{2}, \quad \text{etc.} \end{aligned}$$

Define the family of *terminal blocks* to be the 0, 1-blocks that arise from any such chain of repeated substitutions beginning with some $\binom{n}{k}$, $0 \leq k \leq n$. The subshift X_ρ in question consists of all one-sided sequences on $\{0, 1\}$ all of whose subblocks can be found in these terminal blocks.

Proposition 2.14. *The Pascal adic system (X, T) and the countable-substitution subshift (X_ρ, σ) are topologically conjugate.*

Proof. Given $x \in X$, the n 'th vertex $x_n = (n, k_n(x))$, written as $\binom{n}{k_n(x)}$, expands under ρ to a terminal 0, 1-block $B(k, n)$. If $E(n, k_n(x))$ denotes the cylinder set in X determined by the minimal path down to x_n , then necessarily $x \in T^j E(n, k_n(x))$ for some $j = 0, 1, \dots, C(n, k_n(x)) - 1$. We define $W_n = \sigma^j B(k, n)$ and $\phi(x) = \lim_{n \rightarrow \infty} W_n \in X_\rho \subset \Sigma_2^+$ when this limit exists (i.e., when the lengths of the W_n grow without bound). The countably many remaining points are dealt with separately. \square

2.12. Making the remote field large. We return now to the question of the possible sizes of tail fields when the generator γ of a measure-preserving system (X, \mathcal{B}, μ, T) is varied. Above we saw examples of processes (the i.i.d. processes) for which the two-sided remote field \mathcal{R}

is trivial. According to a striking result of Ornstein and Weiss, every process can be recoded so that $\mathcal{R} \doteq \mathcal{B}$. Thus even completely nondeterministic Kolmogorov processes, for which $\mathcal{F} \doteq \mathcal{P} \doteq \{\emptyset, X\}$, including Bernoulli (i.i.d.) processes for which also $\mathcal{R} \doteq \{\emptyset, X\}$, can be recoded to be “bilaterally deterministic”: knowing the arbitrarily remote information in \mathcal{R} determines the present, and in fact the entire σ -algebra of the process.

Theorem 2.15. [76] *Given a m.p.t. $T : X \rightarrow X$ on a nonatomic Lebesgue probability space and a finite measurable partition α of X , there is a refinement β of α such that $\mathcal{R}(\beta) \supset \alpha$. Thus if α is a generator, (β, T) is isomorphic to (α, T) and bilaterally deterministic, in that $\mathcal{R}(\beta) \doteq \mathcal{B}$.*

Proof. The idea of the recoding is quite simple. Find a Rokhlin tower in the measure-preserving system (X, \mathcal{B}, μ, T) that consists of $k_1 n_1$ levels and covers all of X except for a set of measure ϵ_1 . The coding really only changes on three subblocks, each of length n_1 , of the tower, namely the top one and the two bottom ones. Suppose that $\alpha = \{A_1, \dots, A_d\}$. For each point x in the top n_1 levels of the tower, put x into the cell $B_{i,t}^1$ (“color” it with this one of finitely many colors) if $x \in A_i$. Similarly, for each point x in the bottom n_1 levels of the tower, put x into the cell $B_{i,b}^1$ if $x \in A_i$. For each point x in the tower, denote by $M_{n_1}(x)$ the set of “ n_1 -mates” of x , namely those y in the tower such that $y = T^{kn_1}x$ for some $k \in \mathbb{Z}$. Denote by $s_1(x)$ the sum modulo d of the α -indices of the n_1 -mates of x :

$$(2.25) \quad s_1(x) = \sum_{y \in M_{n_1}(x)} \alpha(y) \pmod{d}.$$

Then each point x in the central part of the tower, which we have not yet recoded, is put into the cell $B_{i,s_1(x)}^1$ if $x \in A_i \in \alpha$.

The process is iterated, with sequences n_i, k_i, ϵ_i chosen in such a way that the Borel-Cantelli Lemma guarantees that eventually we can remove the superscripts on the B 's and arrive at a finite measurable partition β of X which clearly refines α .

The partition β has $\mathcal{R}(\beta) \doteq \mathcal{B}$. This is because for almost all $x \in X$, given $N \in \mathbb{N}$ we can find a tower of height $k_j n_j$ with $n_j/2 > N$ and x not belonging to any of the top n_j levels or the bottom $2n_j$ levels of the tower. For such x not near the top or bottom of the tower, $\beta(x)$ is determined by $\beta(T^j x)$ for $|j| \geq n_j/2 > N$: knowing $\beta(T^j x)$ for $|j| \geq n_j/2 > N$ determines $s_j(x)$, since this is in the subscript

of the β -symbol of each of the n_j -mates of x , and it also determines $s_j(x) - \alpha(x)$, since we know the α -symbols of all the n_j -mates of x . Therefore $\mathcal{R}(\beta) \supset \alpha$. \square

Question 2.16. [86] Do there exist K -systems such that $\mathcal{R}(\beta) \doteq \mathcal{B}$ for every generator β ?

2.13. Making the one-sided fine future field small. Back now to the question of which K processes have super- K generators. Here the problem is that of destroying any information that might be preserved by recording symbol counts, rather than recoding so that all information persists forever. In joint work with J.-P. Thouvenot, we can accomplish this provided that there is a (positive-entropy) direct Bernoulli factor on which we can anchor the recoding algorithm.

Theorem 2.17. [86] *Let $T : X \rightarrow X$ be a m.p.t. on a nonatomic Lebesgue probability space and α a finite measurable partition of X . Suppose that the process (α, T) is isomorphic to the direct product of a positive-entropy Bernoulli system (B, σ, P) and another system (Y, S, ν) . Then there is a partition β of X which generates a process isomorphic to (α, T) and such that $\mathcal{F}_\psi^+(\beta) \doteq \mathcal{F}(\beta) \doteq \mathcal{F}(\alpha)$. Thus every K process with a direct Bernoulli factor has a super- K generator.*

The key property of independent processes which permits the recoding is the *asymptotic local flatness*, or *translation stability*, of the probabilities of probable symbol counts. This is expressed in the following lemma, which is proved by means of a small calculation with multinomial probabilities entirely like the Hajian-Ito-Kakutani proof of the ergodicity of the Pascal adic.

Lemma 2.18. [86] *Fix a Bernoulli system $\mathcal{B}(p_1, \dots, p_q)$ with shift-invariant probability measure P and let $L \in \mathbb{N}$. Given $\epsilon > 0$ there is $N \in \mathbb{N}$ such that if $n \geq N$ then*

$$(2.26) \quad P\{\omega : \text{for all } s \in \mathbb{Z}^q \text{ with } |s| \leq L, \\ \left| \frac{P\{\xi : \psi_0^n(\xi) = \psi_0^n(\omega)\}}{P\{\xi : \psi_0^{n+s-1}(\xi) = \psi_0^n(\omega) + s\}} - 1 \right| < \epsilon\} > 1 - \epsilon.$$

Again in words, the statement is that most i.i.d. sequences ω for large n have symbol-count vectors $\psi_0^n(\omega)$ such that any bounded translate of the vector has approximately the same probability of occurring as does $\psi_0^n(\omega)$.

Now we sketch the recoding method for producing super- K generators for K -processes which have a direct Bernoulli factor. Consider a typical point $x \in Y \times B$, which we may assume because of the isomorphism is coded by the partition α to a pair of strings (y, ω) on a finite generating partition γ for Y and the time-0 partition ρ for B . We recode the string y of γ -symbols of x to a string on a new alphabet β_0 that is chosen to be large enough that we have available enough permutations of each long-enough β_0 -block to be able to assign different ones to each γ -block of that length. Many sections of the ρ -string ω are not changed by the recoding. The coding uses a long marker block W in B of a special form.

(1) A string of symbols in y that appears against an appearance of W in ω is assigned, in a one-to-one way, a permutation of a single β_0 -block in such a way that across this “marked” set of coordinates each pair of symbols from the alphabet $\beta_0 \times \rho$ appears the *same* number of times.

(2) Across a “free” region of coordinates, between successive appearances of W in ω , we replace the string of $\gamma \times \rho$ symbols in (y, ω) by a permutation of a string on the new alphabet $\beta_0 \times \rho$ which has the property that every pair $(i, b), i \in \beta_0, b \in \rho$, appears the same number of times, except that one special pair appears exactly one extra time. The special pair is chosen to depend only on the string in ω that appears across the free interval under consideration. The result is that when counting pairs of symbols from $\beta_0 \times \rho$ across many free intervals, the count is asymptotically flat.

(3) If the counting of symbols in the new alphabet $\beta_0 \times \rho$ does not start or stop exactly with appearances of W in ω , we are dealing with a slight translation of the count (with high probability W appears in ω with bounded gap), and so the asymptotic local flatness of the probabilities of these symbol counts persists because of the Lemma.

To make the argument precise, considerable maneuvering is necessary with the right sorts of conditioning, conditional ϵ -independence of partitions given σ -algebras, and calculation of probabilities via disintegrations of measures. We do not attempt to delve into the details here, but end with another question from [86].

Question 2.19. If (X, \mathcal{B}, μ, T) has a generator α with $\mathcal{F}_{\psi_\alpha} \doteq \{\emptyset, X\}$, is there also a generator β with

$$(2.27) \quad \mathcal{P}_{\psi_\beta} = \bigcap_{n=1}^{\infty} \mathcal{B}\{\psi_\beta(x_0)\psi_\beta(x_{-1})\dots\psi_\beta(x_{-j}) : j \geq n\} \doteq \{\emptyset, X\}?$$

3. β -SHIFTS

3.1. The languages of β -shifts. A β -shift is a symbolic dynamical system that codes the transformation $T_\beta x = \beta x \pmod 1$ on the unit interval and therefore encapsulates the combinatorics of the numeration system consisting of (preferred) representations of numbers in the interval with respect to a base $\beta > 1$. In the following we will often concentrate on $\beta \in (1, 2)$, when the symbolic dynamics takes place in the 2-shift $\Sigma_2^1 = \{0, 1\}^{\mathbb{N}}$.

Let $\beta > 1$. Given $x \in \mathbb{R}$, the preferred (greedy) expansion $x_0 x_1 x_2 \dots$ of x base β is formed as follows:

$$(3.1) \quad \begin{aligned} x_0 &= \lfloor x \rfloor & r_0 &= x \pmod 1 \\ x_1 &= \lfloor \beta r_0 \rfloor & r_1 &= \beta x \pmod 1 \\ & & & \vdots \\ x_i &= \lfloor \beta r_{i-1} \rfloor & r_i &= \beta r_{i-1} \pmod 1 \quad \text{for } i > 1. \end{aligned}$$

Then for $x \in [0, 1]$

$$(3.2) \quad x = \frac{x_1}{\beta} + \frac{x_2}{\beta^2} + \dots \quad \text{and} \quad \beta x \pmod 1 = \frac{x_2}{\beta} + \frac{x_3}{\beta^2} + \dots$$

Let $d = \lfloor \beta \rfloor + 1$ unless $\beta \in \mathbb{N}$, in which case $d = \beta$, and let $D = \{0, 1, \dots, d-1\}$. Then $T_\beta x = \beta x \pmod 1$ on $[0, 1]$ corresponds to the shift σ on $D^{\mathbb{N}}$. More precisely, denote by Σ_β^+ the closure in $\Sigma_d^1 = D^{\mathbb{N}}$ (with d as above) of the set of all expansions $x_1 x_2 \dots$ as above of all $x \in [0, 1]$. Then the map $h : \Sigma_\beta^+ \rightarrow [0, 1]$ defined by

$$(3.3) \quad h(x_1 x_2 \dots) = \frac{x_1}{\beta} + \frac{x_2}{\beta^2} + \dots$$

is a continuous onto factor map from (Σ_β^+, σ) to $([0, 1], T_\beta)$ which is one-to-one except on a countable set.

That the class of β -shifts is actually representative of a large family of dynamical systems is evident from the following theorem of Parry:

Theorem 3.1. [80] *Every strongly transitive (for every nonempty open set U , there is $n \in \mathbb{N}$ with $\cup_{j=0}^n T^j U = X$) f -expansion map on $[0, 1]$ ($Tx = f^{-1}x \pmod 1$ for a strictly increasing continuous function mapping an interval $[0, a)$ onto $[0, 1)$) is topologically conjugate to a β -transformation.*

A special role is played by the expansion of 1 base β . We put

$$(3.4) \quad e_\beta(1) = a_1 a_2 a_3 \dots$$

with the a_i determined as above, unless this expansion terminates. In this case there is a first i such that $r_i = 0$, that is, $\beta r_{i-1} = n \in \mathbb{N}$, and then we put

$$(3.5) \quad a_i = n - 1, \quad r_i = 1,$$

and continue to apply the expansion algorithm with this modification. The result is that the preferred expansion, which would terminate with an infinite string of 0's, is replaced by the periodic expansion

$$(3.6) \quad e_\beta(1) = [a_1 a_2 \dots a_{i-1} (n - 1)]^\infty.$$

For example, the finite expansion 11 is replaced by 101010

The fundamental insights into the combinatorics of sequences in a β -shift, due to Bissinger, Rényi, and Parry, involve the lexicographic order \leq on infinite strings. (The order is extended to finite strings by agreeing that $u \leq v$ if $u0^\infty \leq v(d - 1)^\infty$.)

Theorem 3.2. [10, 97, 77]

- (1) $x \in D^\mathbb{N}$ is in Σ_β^+ if and only if $\sigma^k x \leq e_\beta(1)$ for all $k \geq 0$.
- (2) A sequence $a = a_1 a_2 \dots \in D^\mathbb{N}$ is the expansion of 1 base β , as above, for some $\beta > 1$ if and only if $\sigma^k a \leq a$ for all $k \geq 0$.

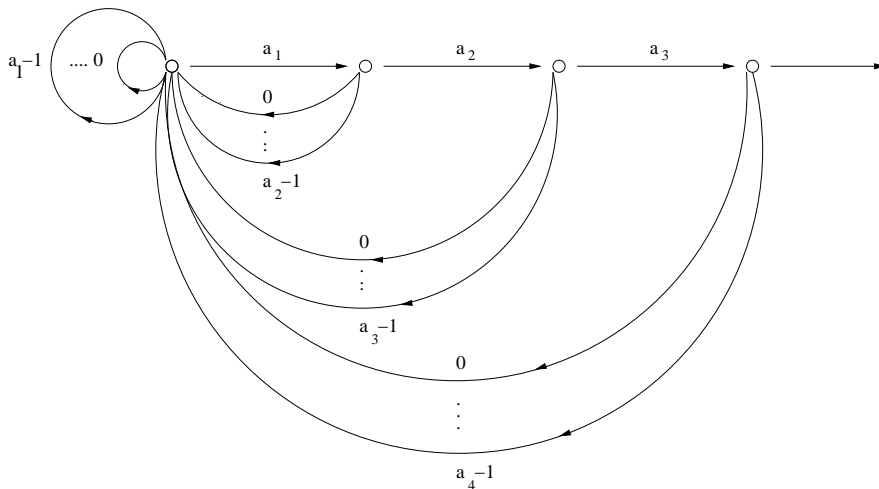
Question 3.3. What combinatorial characterizations can be found to characterize the languages of codings of the transformation $T_{\beta,\alpha}x = \beta x + \alpha \pmod 1$? (See [79, 78, 120, 20, 21].)

An important consequence is the identification of the *language* $\mathcal{L}(\beta)$ of the β -shift (the set of all finite subblocks of all the infinite sequences in Σ_β^+) as the set of labels of finite paths starting at the base vertex b in an infinite labeled directed graph $G(\beta)$: see Figure 9.

From these considerations emerges the following program for research: Study the connections among

- (1) Number-theoretic properties of β ;
- (2) Combinatorial properties of the expansion $e_\beta(1)$;
- (3) The dynamics of the β -shift (Σ_β^+, σ) ;
- (4) Formal-language properties of the associated language $\mathcal{L}(\beta)$;
- (5) related combinatorial, algebraic, and dynamical objects (enumeration systems, distributions of partial sums of expansions, measures, images of Σ_β^+ under mappings, etc.).

Background, motivation, and many results along this line can be found in [13, 9, 8, 99]. The following is known:

FIGURE 9. The graph that determines a β -shift.

- (1) $e_\beta(1)$ is periodic if and only if (Σ_β^+, σ) is a subshift of finite type, and the set of such β is dense in $(1, \infty)$.
- (2) If β is a Pisot-Vijayaraghavan number (an algebraic integer all of whose conjugates have modulus less than 1), then $e_\beta(1)$ is eventually periodic.
- (3) If $e_\beta(1)$ is eventually periodic, then (Σ_β^+, σ) is sofic; equivalently, $\mathcal{L}(\beta)$ is *regular*, or *rational*. This happens for countably many β .
- (4) If (Σ_β^+, σ) is sofic, then β is Perron (an algebraic integer which is larger in modulus than all of its conjugates).
- (5) If β is Perron, then there is $n \in \mathbb{N}$ such that the block 0^n is not found in $e_\beta(1)$.
- (6) The block 0^n is not found in $e_\beta(1)$ if and only if (Σ_β^+, σ) has *specification*: there is $s \in \mathbb{N}$ such that for each pair of blocks $u, v \in \mathcal{L}(\beta)$ there is a block $w \in \mathcal{L}(\beta)$ with length $|w| \leq s$ and $uwv \in \mathcal{L}(\beta)$. The set of all such β has Hausdorff dimension 1.
- (7) There is a word in $\mathcal{L}(\beta)$ that is not found in $e_\beta(1)$ if and only if (Σ_β^+, σ) is *synchronizing*: There is a block $u \in \mathcal{L}(\beta)$ such that if $wu \in \mathcal{L}(\beta)$ then $wuw' \in \mathcal{L}(\beta)$ if and only if $uw' \in \mathcal{L}(\beta)$. The class of all such β is meager and has Lebesgue measure 0.

Further progress on this program has been made by K. Johnson [50], and we now describe some of her work. A natural question, in light of the results listed above, might be what are the special properties of β and Σ_β^+ if the expansion $e_\beta(1)$ is *almost periodic* (every block that

appears in the sequence appears with bounded gap)? What property stronger than specification would this produce?

First, it is necessary to have examples of such β 's. Now in every subshift X (closed shift-invariant subset of Σ_d^1) by compactness there is a unique lexicographically maximal sequence $m(X)$. Since X is a subshift, the element $m(X)$ lexicographically dominates all of its shifts, and therefore it qualifies as the expansion $e_\beta(1)$ of 1 with respect to some base $\beta(X) > 1$, which is necessarily unique. Thus we obtain a map $X \rightarrow \beta(X)$ from subshifts to $(1, \infty)$. If we start with a *minimal* subshift (X, σ) , then $e_{\beta(X)}(1)$ will have the desired property of almost periodicity.

Example 3.4. [50] For the Prouhet-Thue-Morse minimal subshift (the orbit closure of $x = 01101001\dots$), the maximal element $m(X) = \sigma x = 1101001\dots$.

For minimal subshifts generated by constant-length substitutions and some variable-length ones, Johnson provides an algorithm involving substitutions whose fixed point is the maximal element $m(X)$. She also determines the conditions under which the maximal element is a shift of the fixed point of the original substitution.

What are the formal-language properties of the language $\mathcal{L}(\beta)$ for various β ? As mentioned above, $\mathcal{L}(\beta)$ is regular, (equivalently (Σ_β^+, σ) is sofic) if and only if the expansion $e_\beta(1)$ base β is eventually periodic. What about the other classes in the Chomsky hierarchy (see [45])? For which β is $\mathcal{L}(\beta)$ context-free? Context-sensitive? Recursive? Recursively enumerable? (The definitions are recalled below.)

Theorem 3.5. [50] *The language of a β -shift is context-free if and only if it is regular.*

Let us recall very briefly a description of the Chomsky hierarchy. Continue to let $D = \{0, 1, \dots, d-1\}$ be a finite alphabet of symbols, also called *terminals*. A *language* is any subset of the set D^* of all finite words (or blocks), including the empty word, on the alphabet D . D^+ denotes the set of all nonempty words on the symbols from D . A language \mathcal{L} is the language of a two-sided subshift if and only if it has the following two properties:

- (1) \mathcal{L} is *extractive*: if $w \in \mathcal{L}$ then also every subblock of w is in \mathcal{L} ;
- (2) \mathcal{L} is *insertive*: if $w \in \mathcal{L}$, then there are nonempty words $u, v \in D^+$ such that $uwv \in \mathcal{L}$.

(For one-sided subshifts there is a similar characterization.)

A language $\mathcal{L} \subset D^*$ is called *recursively enumerable* if it can be generated by a grammar as described now. A *grammar* is a 4-tuple $G = (V, D, \mathcal{P}, s)$ comprised of:

- a finite alphabet V of *variables* (thought of as temporary symbols);
- the finite alphabet D of *terminals*;
- a finite set \mathcal{P} of *productions* $w \rightarrow w'$, with $w, w' \in (V \cup D)^*$ and w including at least one variable among its symbols;
- a *start symbol* $s \in V$.

The *language* $\mathcal{L}(G)$ generated by a grammar G consists of all finite words in D^* which are the result of the composition of a finite string of productions to the start symbol s . (For the more precise description, see [45].)

A language \mathcal{L} is

- *regular* if it can be generated by a grammar all of whose productions are of the form $A \rightarrow tB$ with $A, B \in V$ and $t \in D$;
- *context-free* if it can be generated by a grammar all of whose productions are of the form $A \rightarrow w$, with $A \in V$ and $w \in (V \cup D)^*$;
- *context-sensitive* if it can be generated by a grammar all of whose productions are of the form $w \rightarrow w'$, with $w, w' \in (V \cup D)^*$, $|w| \leq |w'|$, and w containing at least one variable.

All of these languages are *recursively enumerable*, in that their words can be listed by a Turing machine. Equivalently, recursively enumerable languages can be *recognized* by Turing machines: starting from a designated initial state and given as input a word from D^* , after a finite number of steps the machine stops in a “good” state if and only if the word is in the language. A language is called *recursive* if it is recognized in this way by a Turing machine that eventually halts no matter what input it is given. A language is context-sensitive if and only if it is recognized by a Turing machine which for each input uses only a tape whose length is bounded by the length of the input. A language is regular if and only if it is recognized by a finite-state automaton, and context-free if and only if it is recognized by a finite-state automaton with a pushdown memory stack. (See [89] for equivalent characterizations of regular languages.)

Proof of Theorem 3.5. Suppose that $\mathcal{L}(\beta)$ is context-free. Our aim is to show that $e_\beta(1)$ is eventually periodic.

Every context-free language can be generated by a grammar which is in *Greibach normal form*: all productions are of the form

$$(3.7) \quad A \rightarrow t w, \quad \text{with } A \in V, t \in D, \text{ and } w \in V^*.$$

Moreover, every word is the result of a “leftmost derivation”, in each step of which a production is applied to the leftmost variable in the word at hand (starting with s), thereby adding at least one terminal at every step. For each $n = 1, 2, \dots$ let

$$(3.8) \quad \Delta_n = P_n^{(n)} \dots P_1^{(n)}$$

be a leftmost derivation that, when applied (on the left) to s produces the initial n -block $e_n = a_1 \dots a_n$ of $e_\beta(1)$.

By a diagonal process we choose $P_1 : s \rightarrow a_1 V_1^{(1)} \dots V_{r_1}^{(1)}$ to be a production which appears infinitely many times, say along a subsequence S_1 , among the $P_1^{(n)}$; then P_2 to be a production which appears infinitely many times along the $n \in S_1$ as a $P_2^{(n)}$, and so on. Note that $P_2 P_1 : s \rightarrow a_1 a_2 V_1^{(2)} \dots V_{r_2}^{(2)}$, etc..

The key idea in Johnson’s argument is that of the *last relevant variable*. In the process just described for generating $e_\beta(1)$ as the limit of $P_n \dots P_1 s$, some of the variables $V_i^{(j)}$ that appear might be unnecessary, since they never get acted on by the productions that we are applying (which always act on the leftmost variable in the word so far produced). For each $n = 1, 2, \dots$ let l_n be the smallest l such that

$$(3.9) \quad \begin{aligned} & \lim_{j \rightarrow \infty} P_{n+j} \dots P_{n+1} a_1 a_2 \dots a_n V_1^{(n)} V_2^{(n)} \dots V_l^{(n)} \\ &= \lim_{j \rightarrow \infty} P_{n+j} \dots P_{n+1} a_1 a_2 \dots a_n V_1^{(n)} V_2^{(n)} \dots V_{r_n}^{(n)} \\ &= \lim_{j \rightarrow \infty} P_{n+j} \dots P_{n+1} P_n \dots P_1 s \\ &= e_\beta(1). \end{aligned}$$

Then $V_{l_n}^{(n)}$ is the *last relevant variable* at stage n . We modify the sequence of words $P_n \dots P_1 s = a_1 \dots a_n V_1^{(n)} V_2^{(n)} \dots V_{r_n}^{(n)}$ arising as we apply our sequence of productions (P_n) by erasing from each the string $V_{l_n+1}^{(n)} \dots V_{r_n}^{(n)}$ of irrelevant variables. Then as we apply the P_n , from time to time we must arrive at the words

$$(3.10) \quad a_1 a_2 \dots a_n \dots a_{k_n} V_{l_n}^{(n)}.$$

Now we take advantage of the special property of lexicographic order. For each $A \in V$ denote by $\Sigma(A)$ the set of all sequences in $D^{\mathbb{N}}$ that can be generated by starting with A , applying productions in the grammar G , and taking limits. Denote by $M(A)$ the lexicographically maximal sequence in $\Sigma(A)$. Thus $\Sigma_{\beta}^{+} = \Sigma(s)$ and $e_{\beta}(1) = M(s)$. Then for each $n < m$

$$(3.11) \quad \begin{aligned} e_{\beta}(1) &= a_1 a_2 \dots a_{l_n-1} M(V_{l_n}^{(n)}) \\ &= a_1 a_2 \dots a_{l_n-1} a_{l_n} \dots a_{l_m-1} M(V_{l_m}^{(m)}). \end{aligned}$$

Since there are only finitely many symbols in the alphabet V of variables, there are n and m with $V_{l_n}^{(n)} = V_{l_m}^{(m)}$. Then (3.11) shows that $e_{\beta}(1)$ is eventually periodic.

Remark 1. A similar statement applies to the languages that arise from coding the orbits of unimodal maps of the interval by their visits to one side or the other of the critical point: answering a question of Wang, Yang, and Xie [118], Johnson shows that the language of a unimodal map is context-free if and only if it is regular [51].

In the literature one finds many *pumping lemmas* which describe the ability to repeat (pump) certain words repeatedly in some languages, under different circumstances. Pumping lemmas given the same name but different statements in different sources may not be equivalent. Pumping lemmas can be used to show that some languages are not regular or not context-free, but they usually do not suffice to characterize such properties. For the class of β -shift languages, Johnson [50] has the following results about pumping lemmas:

- (1) A characterization of which β -shift languages satisfy the Weak Regular Pumping Lemma: it is necessary and sufficient that the lexicographically minimal element in Σ_{β}^{+} occur in the orbit of $e_{\beta}(1)$. For example, the language corresponding to

$$(3.12) \quad e_{\beta}(1) = 1010^210^31\dots$$

does *not*.

- (2) Within the class of β -shift languages, regularity is characterized by satisfying the conclusion of the Strong Regular Pumping Lemma.
- (3) When $e_{\beta}(1)$ is the maximal point 1ω in the orbit-closure of the fixed point ω of the Fibonacci substitution $0 \rightarrow 01, 1 \rightarrow 0$, the associated language does *not* satisfy the Context-Free Pumping Lemma. We do not know whether the language associated to

the maximal point in the Prouhet-Thue-Morse subshift satisfies this pumping lemma.

- (4) The language associated to $1010^210^31\dots$ satisfies the Context-Free Pumping Lemma and Ogden's Lemma, but *not* the Generalized Ogden Lemma.
- (5) The languages of certain substitution subshifts (see [50, 4.4.16]) *do* satisfy the Context-Free Pumping Lemma, the Ogden Lemma, and the Generalized Ogden Lemma.

There are also further results in [50] about the positions of (the languages of) certain types of β -shifts in the Chomsky hierarchy:

- (1) There are non-context-free, even nonrecursive β -shifts.
- (2) Any β -shift coming from a constant-length aperiodic primitive substitution (for example, the one determined by the maximal element in the Prouhet-Thue-Morse subshift) is *context-sensitive*.
- (3) There are *nonrecursive* β -shifts, for example those for which β has a noncomputable expansion base 2.

3.2. Dynamical questions related to Erdős measures. Let $\beta > 1$ and $\alpha = 1/\beta$. We mostly assume that $\alpha \in (1/2, 1)$, so that coefficients of β -expansions are in $\{0, 1\}$.

Consider a *random walk* that starts at $0 \in \mathbb{R}$ and at each time $k \in \mathbb{N}$ takes a step of size 0 with probability $p \in (0, 1)$ or of size α^k with probability $1 - p$. The *Erdős measure* $\mu_{\alpha,p}$ is the limiting distribution of the walk, so that if ρ_p denotes the Bernoulli $1/2, 1/2$ measure on $\Sigma_2^1 = \{0, 1\}^{\mathbb{N}}$, for $E \subset \mathbb{R}$ we have

$$(3.13) \quad \mu_{\alpha,p}(E) = \rho_p\left\{\omega \in \Sigma_2^1 : \sum_{k=1}^{\infty} \omega_k \alpha^k \in E\right\}.$$

In terms of the map $h : \Sigma_2^1 \rightarrow [0, \alpha/(1 - \alpha)]$ defined by

$$(3.14) \quad h(\omega) = \sum_{k=1}^{\infty} \omega_k \alpha^k,$$

$$\mu_{\alpha,p}(E) = \rho_p(h^{-1}E).$$

As the limiting distribution of a random walk, the Erdős measure $\mu_{\alpha,p}$ is also an *infinite convolution of Bernoulli measures*. For each $k \geq 1$, let

$$(3.15) \quad \nu_k = p\delta_0 + (1 - p)\delta_{\alpha^k}$$

denote the convex combination of point masses at 0 and α^k . Then

$$(3.16) \quad \mu_{\alpha,p} = \nu_1 * \nu_2 * \dots$$

In 1935 Jessen and Wintner [49] proved that $\mu_{\alpha,p}$ is nonatomic and either singular with respect to Lebesgue measure m on \mathbb{R} or absolutely continuous with respect to m . In 1997 Mauldin and Simon [72] showed that $\mu_{\alpha,1/2}$ is absolutely continuous with respect to m if and only if it is equivalent to m . The question of which half of the Jessen-Wintner dichotomy applies to a given α has been the object of a lot of work over the years, much of it focusing on the symmetric case $p = 1/2$ but extendable also to other $p \in (0, 1)$.

- Jessen-Wintner, 1935 [49]: For $0 < \alpha < 1/2$, $\mu_{\alpha,p} \perp m$.
- Wintner, 1935 [121]: If $\alpha = (1/2)^{1/k}$ for some $k \in \mathbb{N}$, then $\mu_{\alpha,1/2} \ll m$.
- Erdős, 1939 [27]: If α is the reciprocal of a Pisot-Vijayaraghavan number, and $1/2 < \alpha < 1$, then $\mu_{\alpha,1/2} \perp m$.
- Erdős, 1940: [28]: There is $c \in (1/2, 1)$ such that for m -a.e. $\alpha \in (c, 1)$, $\mu_{\alpha,1/2} \ll m$.
- Garsia, 1963 [35]: Constructed a countable set of algebraic numbers α for which $\mu_{\alpha,1/2} \ll m$.
- Garsia, 1963 [34, 35]: For each $\alpha \in (0, 1)$ there is an *entropy* $H(\alpha)$ such that if $H(\alpha) < 1$, then $\mu_{\alpha,1/2} \perp m$.
- Alexander-Zagier, 1991 [5]: Gave a formula for $H(\alpha)$ when α is the reciprocal of the golden mean, showing that then $H(\alpha) < 1$ and hence $\mu_{\alpha,1/2} \perp m$.
- Solomyak, 1995 [109, 83]: For m -a.e. $\alpha \in (1/2, 1)$, $\mu_{\alpha,1/2} \ll m$.

A more detailed history, especially of recent work, can be found in [82].

Finally, the Erdős measure $\mu_{\alpha,1/2}$ is characterized by a *self-similarity property*, according to which it is the unique invariant measure of a certain iterated function system. Let $I = [0, \alpha/(1 - \alpha)]$ and define $S_0, S_1 : I \rightarrow I$ by

$$(3.17) \quad S_0x = \alpha x, \quad S_1x = \alpha x + \alpha.$$

Restricted to $[0, 1]$, S_0 and S_1 are the two branches of the inverse of $T_\beta x = \beta x \pmod{1}$ —see Figure 10. Then $\mu_{\alpha,p}$ is the unique Borel probability measure on I that satisfies

$$(3.18) \quad \mu(E) = p\mu(S_0^{-1}E) + (1 - p)\mu(S_1^{-1}E) \quad \text{for all Borel } E \subset I.$$

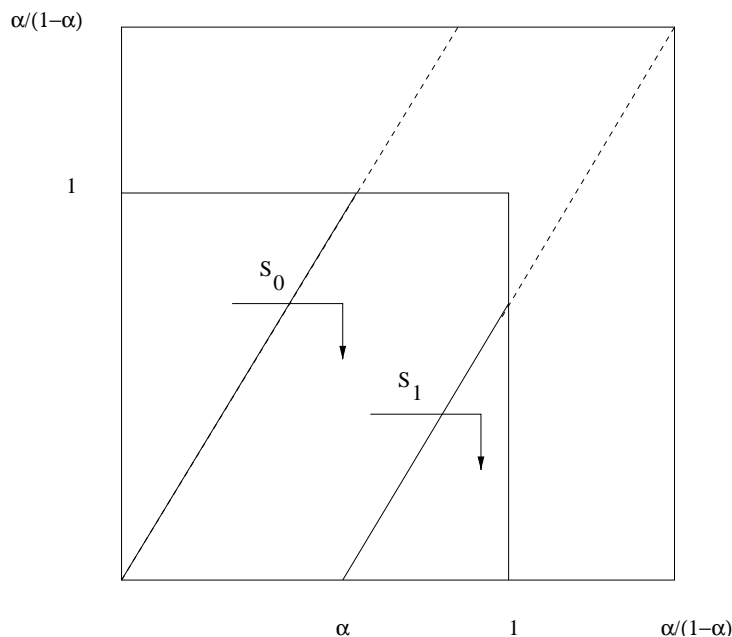


FIGURE 10. The iterated function system for an Erdős measure.

The support of the measure $\mu_{\alpha,p}$ is the unique nonempty compact *attractor* A for the iterated function system, which satisfies

$$(3.19) \quad A = S_0A \cup S_1A.$$

Let C denote the family of nonempty compact subsets of I , and define $S : C \rightarrow C$ by $SK = S_0K \cup S_1K$. Then S is a contraction on C with respect to the Hausdorff metric, and for any $K \in C$, $S^n K \rightarrow A$. If repeatedly and independently one of the two contractions S_0 and S_1 is selected with probabilities p and $1 - p$ respectively and the resulting sequence of maps is applied to any initial point, then we may think of $\mu_{\alpha,p}$ as describing the distribution of the limiting point in A . (If the initial point is 0 and $\omega \in \Sigma_2^1$ is the list of indices of the chosen maps, then the limiting point is given by (3.14).

These Erdős measures are to be contrasted with the *Rényi measures* m_β on $[0, 1]$. The Rényi measure m_β on $[0, 1]$ is the unique Borel probability measure on $[0, 1]$ which is invariant under $T_\beta x = \beta x \pmod 1$ and equivalent to Lebesgue measure; it is the image under the essentially one-to-one map $h : \Sigma_\beta^+ \rightarrow [0, 1]$, $h(\omega) = \sum \omega_k \alpha^k$, of the unique measure of maximal entropy (Shannon-Parry measure) on the β -shift (Σ_β^+, σ) .

There are two natural ways to convert the Erdős measure $\mu_{\alpha,p}$ on $I = [0, \alpha/(1-\alpha)]$ to a measure on $[0, 1]$ —and then, via h , to a measure on Σ_β^+ . For the case of the golden mean, Sidorov and Vershik [106] *scale* the interval I and the measure it carries back to $[0, 1]$ by multiplying by $(1-\alpha)/\alpha = \alpha$, obtaining a measure $\mu_{\alpha,1/2}^s$ on $[0, 1]$. In terms of the map S_0 above and the *normalization map*

$$(3.20) \quad \begin{aligned} n : \Sigma_2^1 &\rightarrow \Sigma_\beta^+, \\ n(\omega) &= \beta\text{-expansion of } S_0 h(\omega) = \Sigma_\beta^+ \cap h^{-1}h(0\omega), \end{aligned}$$

this measure is defined by

$$(3.21) \quad \mu_{\alpha,1/2}^s = \mu_{\alpha,1/2} \circ S_0^{-1} = \rho_{1/2} \circ n^{-1} \circ h^{-1}.$$

The second measure, used by K. Shelton [104], is the normalized restriction to $[0, 1]$ of Erdős measure:

$$(3.22) \quad \mu_{\alpha,p}^r = \frac{\mu_{\alpha,p}}{\mu_{\alpha,p}[0, 1]} = \frac{\rho_p \circ h^{-1}}{\mu_{\alpha,p}[0, 1]}.$$

It is not clear whether these two measures are always equivalent to one another. However, since they are equivalent on an interval, they are either both singular with respect to Lebesgue measure m or both equivalent to m , and Shelton has shown that in the case of the golden mean they *are* equivalent.

Denote by $\tilde{\mu}_{\alpha,p}^r = \mu_{\alpha,p}^r(h|_{\Sigma_\beta^+}^{-1})$ the lift of the restricted Erdős measure on $[0, 1]$ to the β -shift Σ_β^+ .

It was the idea of Sidorov and Vershik to study Erdős measures from a dynamical viewpoint. This permitted them in some cases to prove singularity with respect to Lebesgue measure on the basis of mutual singularity of inequivalent measures that are quasi-invariant and ergodic for the same transformation. It also presented dynamical systems and combinatorial phenomena ripe for study. In the thesis of Kennan Shelton [104] progress was made on several questions suggested by this dynamical viewpoint. In particular, for pseudo-golden mean numbers ($1 = \alpha + \alpha^2 + \cdots + \alpha^m$) Shelton independently showed that $\mu_{\alpha,p}^r$ is quasi-invariant for the β -transformation T_β . He also showed that it has a property stronger than ergodicity, namely partial mixing. We now discuss this work and some of the interesting problems that it raises.

Shelton's main results related to the system $([0, 1], T_\beta, \mu_{\alpha,p}^r)$ are as follows:

- The measure $\mu_{\alpha,p}$ is *not* T_β -invariant, but for α a pseudo-golden mean number, it is *quasi-invariant*: $\mu_{\alpha,p}(E) = 0$ if and only if $\mu_{\alpha,p}(T_\beta^{-1}E) = 0$ for each Borel $E \subset [0, 1]$.
- For α a pseudo-golden mean number, the system $([0, 1], T_\beta, \mu_{\alpha,p}^r)$ is *partially mixing*: there is $\delta > 0$ such that for all measurable $E, F \subset [0, 1]$,

$$(3.23) \quad \liminf \mu(E \cap T_\beta^{-k}F) \geq \delta \mu(E)\mu(F).$$

- Consequently,
 - (1) $([0, 1], T_\beta, \mu_{\alpha,p}^r)$ is *mildly mixing*: if $f \in L^\infty$, $k_j \rightarrow \infty$, and $fT_\beta^{k_j} \rightarrow f$ weak* in L^∞ , then f is (a.e.) constant;
 - (2) $([0, 1], T_\beta, \mu_{\alpha,p}^r)$ is ergodic (this was proved by Sidorov and Vershik [106] for the case when α is the reciprocal of the golden mean, with the direct consequence that the Erdős measure $\mu_{\alpha,1/2}$ is singular with respect to Lebesgue measure);
 - (3) $([0, 1], T_\beta, \mu_{\alpha,p}^r)$ is conservative (there are no wandering sets of positive measure);
 - (4) there is a unique T_β -invariant probability measure $\nu_{\alpha,p}^r \sim \mu_{\alpha,p}^r$;
 - (5) $\nu_{\alpha,p}^r$ is *lightly mixing* for T_β : if $\nu(A), \nu(B) > 0$, then

$$(3.24) \quad \liminf \nu(A \cap T^{-k}B) > 0$$

(equivalently, for all A, B with positive measure, $\nu(A \cap T^{-k}B) > 0$ for large enough k);

- (6) $([0, 1], T_\beta, \mu_{\alpha,p}^r)$ is *weakly mixing*: if $f \in L^\infty, \lambda \in \mathbb{C}$, and $fT = \lambda f$ a.e., then f is constant (a.e.).
- For $\alpha + \alpha^3 = 1$, $\mu_{\alpha,p}^r$ is *quasi-invariant* for T_β .

See [1] for a discussion of these properties in the general setting of non-probability-preserving dynamics.

Note that partial mixing need not be preserved when passing to an equivalent measure, so it would not follow from the Sidorov-Vershik proof [106] that their two-sided Erdős measure is Bernoulli in the golden-mean case.

These results lead to the following conjecture for a general dynamical system (X, \mathcal{B}, ν, T) .

Conjecture 3.6. If ν is invariant and lightly mixing for T , then there is $\mu \sim \nu$ which is quasi-invariant and partially mixing for T .

Note that T. Adams [2] has constructed a lightly mixing (T, ν) and a set of positive measure A such that for every $\delta > 0$ there is a measurable set B_δ such that

$$(3.25) \quad \liminf \nu(A \cap T^{-k} B_\delta) < \delta \nu(B_\delta).$$

Question 3.7. If μ is quasi-invariant for T and partially mixing on a generating semialgebra for the σ -algebra of measurable sets, is it partially mixing on the full σ -algebra?

For comparison, T. Adams [2] has constructed a T -invariant and weakly mixing measure ν which is lightly mixing on intervals but not on the full σ -algebra of measurable sets.

The analysis of these properties depends on understanding the *sum equivalence relation* \sim_α generated by the information-collapsing function h , defined, for $1/2 < \alpha < 1$ as we henceforth assume, on finite or infinite strings of 0's and 1's by $h(\omega) = \sum \omega_k \alpha^k$:

$$(3.26) \quad \omega \sim_\alpha \omega' \text{ if and only if } h(\omega) = h(\omega').$$

The relation can be pictured by a graded graph $\Gamma(\alpha)$ whose vertices mark on a horizontal scale the points on the real line arrived at by sums along the strings. The initial level L_0 consists of a single root vertex at 0, and for $n = 1, 2, \dots$ the set L_n of vertices at level n consists of all the points $\sum_{k=1}^n \omega_k \alpha^k$ in the interval I with each $\omega_k = 0$ or 1. There is an edge labeled 0 from $v \in L_n$ to $v' \in L_{n+1}$ if and only if, as elements of \mathbb{R} , $v = v'$; there is an edge labeled 1 from v to v' if and only if $v' - v = \alpha^{n+1}$. The graph can possess complicated connectivity when α is the root of a polynomial with coefficients $-1, 0$ and 1 . For example, for the case of the golden mean when $1 = \alpha + \alpha^2$, each appearance of 100 in a string ω can be replaced by 011 without affecting the sum $h(\omega)$; this is seen in the graph by the existence of many closed “diamonds”: see Figure 11. This graph is complicated enough, but for non-pseudo-golden mean numbers paths can cross, in terms of the order in \mathbb{R} , without passing through a common vertex: see Figure 12 for an example of the added intricacies.

For each vertex v in the graph for α , denote by $d_n(v)$ the number of paths down to that vertex. The combinatorics of these numbers is intricate, interesting, and important. The *Garsia entropy* of the graph

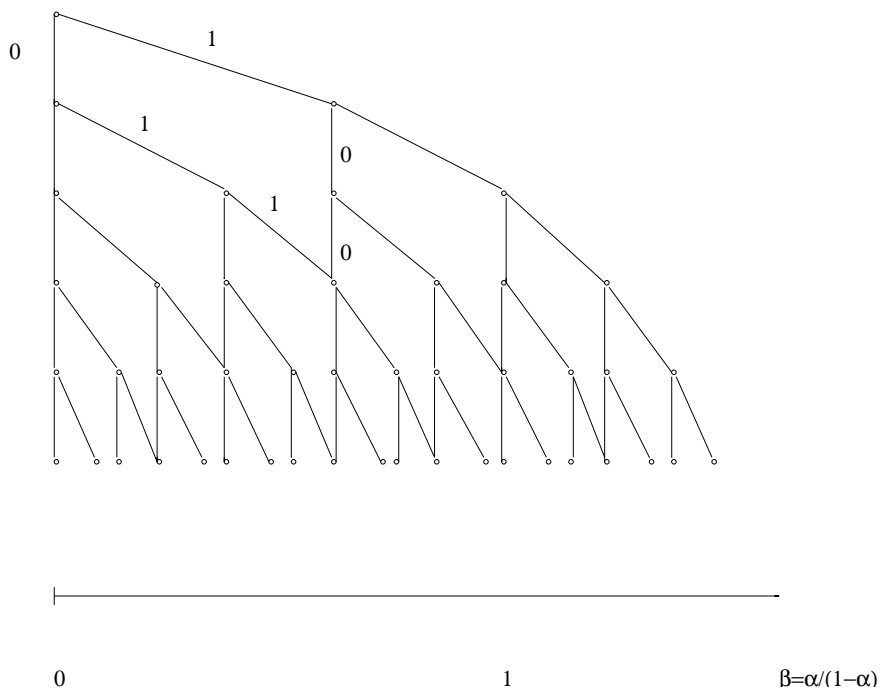


FIGURE 11. The graph of string equivalences for the golden mean.

is defined as follows, for the case $p = 1/2$. For each n let

$$(3.27) \quad H_n(\alpha) = - \sum_{v \in L_n} \frac{d_n(v)}{2^n} \log_2 \frac{d_n(v)}{2^n},$$

$$H(\alpha) = \lim_{n \rightarrow \infty} \frac{1}{n \log_2 \alpha} H_n(\alpha).$$

The computation and even estimation of these entropies for various α is not easy [35, 5, 106].

Question 3.8. Suppose for the rest of this section that 1 has a finite β -expansion,

$$(3.28) \quad 1 = a_1\alpha + a_2\alpha^2 + \cdots + a_s\alpha^s, \quad \text{with each } a_i = 0 \text{ or } 1,$$

so that α is a root of a polynomial with coefficients $-1, 0$ and 1 , and thus in the graph $\Gamma(\alpha)$ some vertices have more than one path entering from the root.

- (1) Are there pairs of *relatively transcendental* vertices $v, v' \in L_n$ with no common successor? That is, can it happen that there

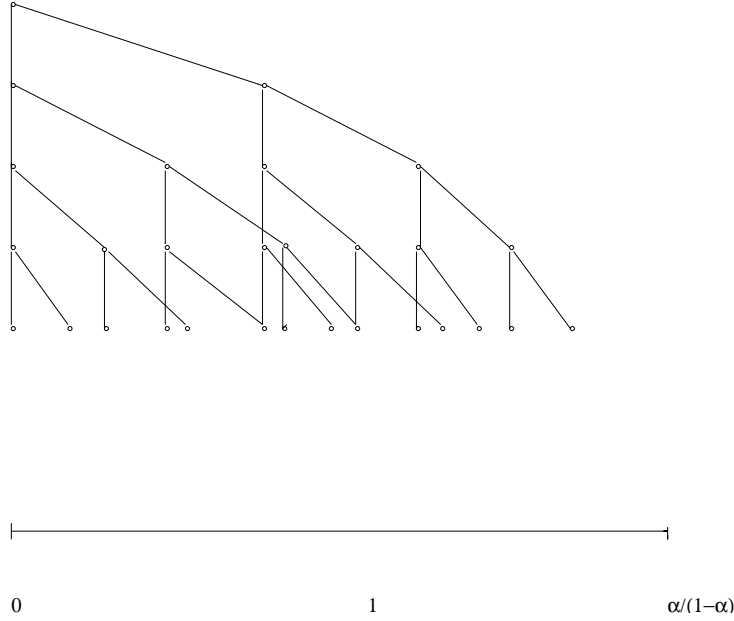


FIGURE 12. The graph of string equivalences for $\alpha + \alpha^3 = 1$.

do not exist $\omega_{n+1}, \dots, \omega_{n+k}, \omega'_{n+1}, \dots, \omega'_{n+k} \in \{0, 1\}$ with

$$(3.29) \quad v + \sum_{j=n+1}^{n+k} \omega_j \alpha^j = v' + \sum_{j=n+1}^{n+k} \omega'_j \alpha^j.$$

- (2) If $v, v' \in L_n$ are not relatively transcendental, what is the minimal length of a path from each to a common successor?
- (3) Estimate the minimal and maximal spacing between vertices at level n in $\Gamma(\alpha)$.
- (4) Describe the limiting distribution of the vertices of L_n in I .

The equation (3.28) satisfied by α leads to the equivalence of finite strings $10^s \sim_\alpha 0a_1a_2\dots a_s$: using these strings as coefficients of the same consecutive powers of α produces the same sum. Denote by $R_0(\alpha)$ the set of minimal-length pairs of strings (ω, ω') such that $\omega \neq \omega', l(\omega) = l(\omega')$, and $h(\omega) = h(\omega')$. For example, for $1 = \alpha + \alpha^2$, $R_0(\alpha) = \{(100, 011), (011, 100)\}$, while for $1 = \alpha + \alpha^3$,

$$(3.30) \quad R_0(\alpha) = \{(1000, 0101), (0101, 1000), (1010, 0111), (0111, 1010)\}.$$

Definition 3.9. Two finite 0, 1 strings ω, ω' are α -replacement equivalent, and we write $\omega \leftrightarrow_\alpha \omega'$, if there is a finite sequence of strings all

of the same length

$$(3.31) \quad \omega = \omega(1), \dots, \omega(k) = \omega'$$

such that for each $j = 1, \dots, k - 1$, $\omega(j + 1)$ is obtained from $\omega(j)$ by replacing a subblock by an $R_0(\alpha)$ -equivalent subblock.

Clearly $\omega \leftrightarrow_\alpha \omega'$ implies that $\omega \sim_\alpha \omega'$. What about the converse?

Proposition 3.10. [104] *If α is a pseudo-golden mean number, then $\omega \sim_\alpha \omega'$ implies that $\omega \leftrightarrow_\alpha \omega'$,*

Proof. The basic reason for this is the same as for the other good properties of the graphs $\Gamma(\alpha)$ for the α in this class: shadows of vertices have hard edges. To be more precise, define the *shadow* of a vertex $v \in L_n$ to be the set of all vertices $w \in \cup_{m>n} L_m$ such that $v < w < v + \alpha^{n+1} + \alpha^{n+2} + \dots$ (as real numbers). The two *edges* of the shadow of v are the sets of vertices $w = v + 0 + 0 + \dots + 0$ and $w = v + \alpha^{n+1} + \dots + \alpha^{n+k}$, $k \geq 1$. Some inequalities, combinatorics, and induction can be used to show that if w is a vertex in the shadow of a vertex v , then any path in $\Gamma(\alpha)$ from the root to w contains a vertex on an edge of the shadow of v . This implies that ω' can be extended to a string whose sum is in the shadow of ω , hence intersects an edge. Given this intersection, it is possible to construct the requisite chain of \leftrightarrow_α -equivalences. \square

It is surprising that even for the apparently simple case of $\alpha + \alpha^3 = 1$ these two equivalence relations are not the same, and in fact the matter is amazingly complicated. Quite possibly the relation \leftrightarrow_α is not even generated by a finite set of replacement pairs. A computer search produced the pair

$$(3.32) \quad 10010010010 \sim_\alpha 01100110011$$

and its dual pair

$$(3.33) \quad 01101101101 \sim_\alpha 10011001100.$$

Evidently these pairs do not contain any of the basic blocks 1000, 0101, 1010, 0111 and thus cannot possibly be \leftrightarrow_α equivalent. Shelton constructs the following SFT to create strings which are guaranteed not to include any of the 4-blocks 1000, 0101, 1010, 0111, as well as 000 or 111. (The block 000 cannot appear at the beginning of any string and still produce a sum equal to that of any block beginning with 1. Similarly, 111 cannot be the initial segment of a block whose sum equals that of one beginning with 0. And neither 000 nor 111 can

make a first appearance in any block after the first place unless one of the forbidden blocks 1000 or 0111 also appears.)

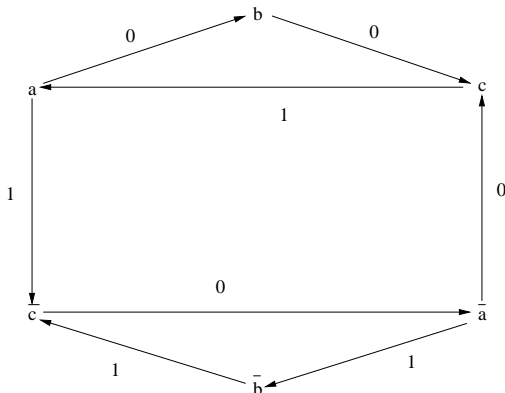


FIGURE 13. SFT for finding sum-equivalent strings which are not replacement equivalent.

Computer searches for pairs of paths corresponding to sum-equivalent finite strings produce many examples. For each $n \geq 1$ let $R_n(\alpha)$ denote the set of minimal length \sim_α -equivalent pairs neither of which have a subblock which for some $k < n$ is a member of a pair in $R_k(\alpha)$.

Conjecture 3.11. [104]

- (1) For each $n \geq 1$, $R_n(\alpha)$ consists of four dual pairs (u_n, v_n) , (\bar{u}_n, \bar{v}_n) , (v_n, u_n) , (\bar{v}_n, \bar{u}_n) of strings of length $5n + 6$. So in particular, \sim_α is not finitely generated.
- (2) As paths on the above SFT, the strings in $R_n(\alpha)$ end in cab , $ca\bar{c}$, $\bar{c}\bar{a}\bar{b}$, and $\bar{c}\bar{a}c$.
- (3) The $R_n(\alpha)$ generate the equivalence relation \sim_α on finite 0, 1 strings.

We indicate now the ideas behind the main results in [104].

Theorem 3.12. [104] *If α is a pseudo-golden mean number, then $\mu_{\alpha,p}^r$ is quasi-invariant for $T_\beta : [0, 1] \rightarrow [0, 1]$.*

Proof. We are assuming that $1/2 < \alpha < 1$. Let $J = [0, (1 - \alpha)/\alpha]$ if $\alpha + \alpha^2 > 1$, $J = [0, (2\alpha - 1)/(1 - \alpha)]$ if $\alpha + \alpha^2 \geq 1$.

Lemma 3.13. *$\mu_{\alpha,p}^r$ is quasi-invariant for T_β if and only if for each Borel $E \subset J$, $\mu_{\alpha,p}(E) = 0$ implies $\mu_{\alpha,p}(E + 1) = 0$.*

Proof. We illustrate the argument by assuming that $\alpha + \alpha^2 = 1$. If $E \subset [\alpha, 1]$, then $T_\beta^{-1}E = S_0E$, so by the self-similarity property of

$\mu_{\alpha,p}^r$,

$$\begin{aligned}
 \mu_{\alpha,p}^r(T_\beta^{-1}E) &= p\mu_{\alpha,p}^r(S_0^{-1}T_\beta^{-1}E) + (1-p)\mu_{\alpha,p}^r(S_1^{-1}T_\beta^{-1}E) \\
 (3.34) \qquad &= p\mu_{\alpha,p}^r(S_0^{-1}S_0E) + (1-p)\mu_{\alpha,p}^r(S_1^{-1}S_0E) \\
 &= p\mu_{\alpha,p}^r(E) + (1-p)\mu_{\alpha,p}^r(\emptyset) = p\mu_{\alpha,p}^r(E),
 \end{aligned}$$

which is 0 if and only if $\mu_{\alpha,p}^r(E) = 0$.

In case $E \subset [0, \alpha)$, $T_\beta^{-1}E = S_0E \cup S_1E$ and $S_1^{-1}S_0E = \emptyset$, so that

$$(3.35) \qquad \mu_{\alpha,p}^r(S_0E) = p\mu_{\alpha,p}^r(E) + (1-p)\mu_{\alpha,p}^r(S_1^{-1}S_0E),$$

which is 0 if and only if $\mu_{\alpha,p}^r(E) = 0$. On the other hand, $S_0^{-1}S_1y = \beta(\alpha y + \alpha) = y + 1$, so that

$$\begin{aligned}
 \mu_{\alpha,p}^r(S_1E) &= p\frac{\mu_{\alpha,p}(S_0^{-1}S_1E)}{\mu_{\alpha,p}[0, 1]} + (1-p)\mu_{\alpha,p}^r(E) \\
 (3.36) \qquad &= p\frac{\mu_{\alpha,p}(E + 1)}{\mu_{\alpha,p}[0, 1]} + (1-p)\mu_{\alpha,p}^r(E),
 \end{aligned}$$

from which the result follows. \square

Applying the Lemma, we take a measurable set $F \subset [1, \alpha/(1-\alpha)]$ of positive $\mu_{\alpha,p}^r$ -measure and indicate how to show that $\mu_{\alpha,p}^r(F-1) > 0$. Choose a cylinder set C in Σ_2^1 with $\rho_p(C \cap h^{-1}F) > 0$. Since $1 = \alpha + \alpha^2 + \cdots + \alpha^m$ and F lies to the right of 1, we may assume (by extending the path if necessary) that the vertex $v(C)$ in the graph $\Gamma(\alpha)$ that is at the end of the path labeled by C is in the shadow of the vertex $v(1^m)$ at the end of the path labeled by 1^m .

If in fact $C = [1^m]$, then we may form a subset of $F-1$ by replacing the initial block 1^m in each point ω of $C \cap h^{-1}F$ by the block 0^m . Since this is a nonsingular transformation with respect to the Bernoulli measure ρ_p , we arrive at a subset of $[0^m] \cap h^{-1}(F-1)$ which has positive measure, and thus $\mu_{\alpha,p}^r(F-1) > 0$.

If $v(C) \neq v(1^m)$, we use the hard edges of the shadow of $v(1^m)$ to perform an analogous cutting and pasting of initial blocks. See Figure 14. Simply stated, the path to $v(C)$ must hit one of the edges $1^m000\dots$ or $1^m111\dots$, say in a vertex $u = v(1^mz)$ with z either a finite string of 0's or a finite string of 1's. Replacing the initial block of length $l(1^mz)$ of the string defining C by 1^mz (if necessary) is a nonsingular transformation with respect to ρ_p (and in fact can be accomplished by finitely many replacements of one member of an $R_0(\alpha)$ -equivalent pair

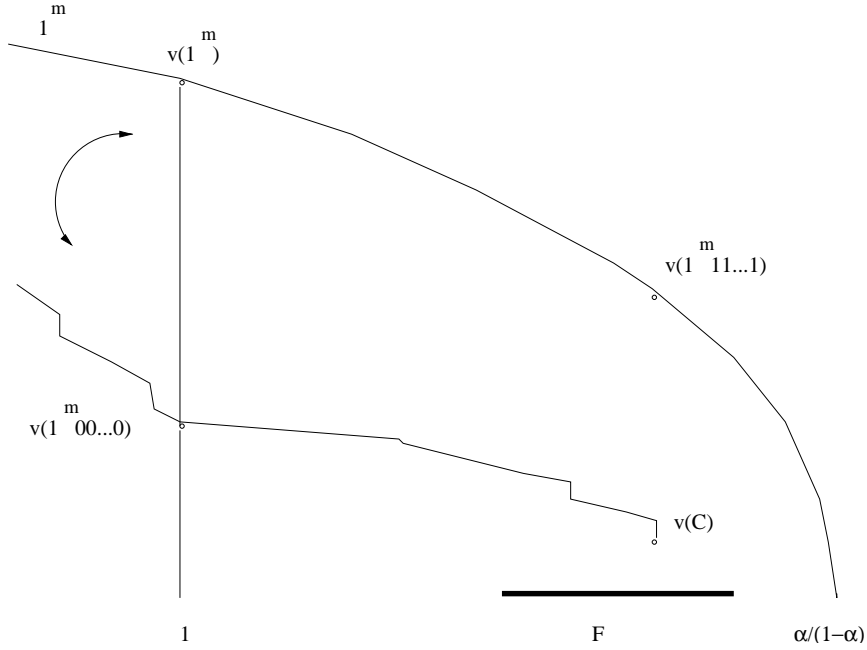


FIGURE 14. Cutting and pasting initial segments of paths in $h^{-1}F$.

by the other). Hence $\rho_p([1^m z] \cap h^{-1}F) > 0$, reducing to the previous case. \square

Question 3.14. Shelton has pointed out that this proof works for any Erdős-type measure μ_α^r defined by starting with any μ on Σ_2^1 quasi-invariant under the group Γ of finite coordinate changes (see Section 2) instead of a Bernoulli measure μ_p . What will be the dynamical properties of the β -transformation T_β with respect to such a measure μ_α^r if the original measure μ is a Markov or Gibbs measure (see [90])?

Theorem 3.15. [104] *If α is a pseudo-golden mean number, then $T_\beta : [0, 1] \rightarrow [0, 1]$ is partially mixing with respect to the quasi-invariant measure $\mu_{\alpha,p}^r$.*

Proof. For any cylinder set $C = [c_1 \dots c_r] \subset \Sigma_\beta^+$, denote by C^* the union of all cylinders in Σ_2^1 defined by sequences $\omega_1 \dots \omega_r$ of the same length r as the string defining C and having the same α -sum: $\sum_{j=1}^r \omega_j \alpha^j = \sum_{j=1}^r c_j \alpha^j$. Then there is a constant $K > 0$ (independent of C) such

that

$$(3.37) \quad \frac{1}{K} \tilde{\mu}_{\alpha,p}^r(C) \leq \rho_p((C0)^*) \leq K \tilde{\mu}_{\alpha,p}^r(C),$$

abbreviated $\tilde{\mu}_{\alpha,p}^r(C) \asymp \rho_p(C^*)$.

Since $\tilde{\mu}_{\alpha,p}^r(C) = \rho_p(h^{-1}hC)$ and $(C0)^* \subset h^{-1}hC$ for pseudo-golden mean numbers, one direction of this comparison is not too hard. It is much harder to bound the ρ_p -measure of $h^{-1}hC$ by the measure of $(C0)^*$, the cylinders corresponding to paths down to the vertex $v(C)$ in the graph $\Gamma(\alpha)$.

If $x = x_1 \dots x_n$ is a string in Σ_2^1 with $h(x) \in h(C)$, then the vertex $v(x)$ is in the shadow of $v(C)$. Again because the edges are hard the path x must hit an edge, and in fact the path cannot cross the level r of $v(C)$ too far from $v(C)$. These considerations lead to a decomposition of $h^{-1}hC$ into a union of cylinders whose ρ_p -measures can be estimated in terms of the ρ_p -measure of C^* .

Then partial mixing

$$(3.38) \quad \tilde{\mu}_{\alpha,p}^r(E \cap \sigma^{-k}F) \geq \delta \tilde{\mu}_{\alpha,p}^r(E) \tilde{\mu}_{\alpha,p}^r(F) \quad \text{for all large enough } k,$$

can be shown for E a cylinder set and F an arbitrary measurable set in Σ_β^+ using the mixing of the shift with respect to the Bernoulli measure ρ_p . The result for all measurable E, F follows by approximation. \square

Theorem 3.16. [104] *If $\alpha + \alpha^3 = 1$, then $\mu_{\alpha,p}^r$ is quasi-invariant for $T_\beta : [0, 1] \rightarrow [0, 1]$.*

Proof. The key element of this argument is that given a cylinder $C = [c_1 \dots c_r] \subset \Sigma_2^1$, for ρ_p -a.e. ω in the shadow of $v(C)$ there are $k \geq r$ and $c_{r+1} \dots c_k \in \{0, 1\}$ such that $c_1 \dots c_k \leftrightarrow_\alpha \omega_1 \dots \omega_k$ (each string can be converted to the other by a finite chain of replacements of $R_0(\alpha)$ -equivalent subblocks).

Thus even though the shadow of $v(C)$ does not have hard edges, by going to finite extensions of strings we can find \leftrightarrow_α -equivalent strings. Then given measurable $F \subset [1, \beta]$ with $\mu_{\alpha,p}(F) > 0$, we can again find a cylinder that intersects $F-1$ in a set of positive ρ_p -measure, implying that $\mu_{\alpha,p}(F-1) > 0$, and Lemma 3.13 yields the result. \square

4. FACTOR MAPS BETWEEN SFT'S

This section describes some joint work with Sujin Shin [85] that is currently in progress.

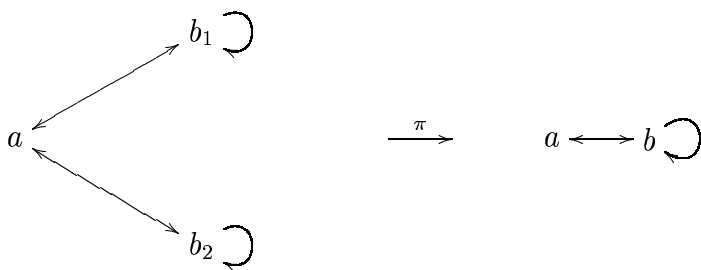
Let X and Y be topologically mixing subshifts of finite type (SFT's) and $\pi : X \rightarrow Y$ a factor map. By recoding if necessary, we may assume that X and Y are 1-step SFT's, so that each consists of all (2-sided) sequences on a finite alphabet consistent with the allowed transitions described by a directed graph with vertex set equal to the alphabet. We may also assume that π is a 1-block map, by passing if necessary to still higher block codings. The action of a factor map can represent the compression of information, either the intentional removal of redundancy in order to save space or unavoidable loss of some details of messages.

In the following $\mathcal{C}(X)$ denotes the set of continuous real-valued functions on X , $\mathcal{M}(X)$ the space of σ -invariant Borel probability measures on X , and $\mathcal{E}(X) \subset \mathcal{M}(X)$ the set of ergodic measures on X .

Question 4.1. Let $\nu \in \mathcal{E}(Y)$. How can one find the measures μ with $\pi\mu = \nu$ which have maximal entropy?

If there is only one such *relatively maximal measure* over ν , we will say that ν is π -*determinate*.

When Y consists of a single fixed point, the problem is that of finding the measures of maximal entropy on X . In that case it is known that there is a unique maximal measure \max_X , the Shannon-Parry measure; it is a Markov measure, and there is an explicit formula for it in terms of the maximum eigenvalue and corresponding eigenvector of the 0,1 transition matrix for X . But in general, a Markov measure ν on Y need not have any Markov preimages on X . Already Blackwell [11] showed that for the following factor map, a typical Markov measure on X will have a non-Markov image under π .



Example 1: Blackwell's Example.

The problem is that the measures of the cylinder sets formed by repeating a block in Y are given by sums of geometrically growing terms rather than single such terms as is the case for a Markov measure.

In [70] this phenomenon was discussed in connection with the problem of coding under restrictions of the kind imposed by forbidding the use of certain blocks. The factor map $\pi : X \rightarrow Y$ represents a channel with deterministic noise, that is, one which loses information in a predictable way. If input messages are governed by statistics described by a measure μ on X , then the information transmission rate is $h(\pi\mu)$ (h denotes the ergodic-theoretic entropy), and the difference $h(\mu) - h(\pi\mu)$ represents the amount of information that is lost (on average, over long messages) when signals are sent across the channel. In this interpretation we seek to identify the worst-case input statistics that lead to the most information loss when the output statistics $\nu = \pi\mu$ are fixed.

The problem also arises from work on a question in smooth dynamics formulated by Gatzouras and Peres [36], namely the search for measures of maximal Hausdorff dimension (hence in a sense the most important ones) for smooth maps $f : M \rightarrow M$ on manifolds which are expanding on compact invariant sets (for example on attractors). Gatzouras and Peres note that Manning and McCluskey [69] showed that there are Axiom A maps for which the Hausdorff dimension of the nonwandering set is *not* a limit of Hausdorff dimensions of ergodic invariant measures that it supports. For expanding maps, Gatzouras and Peres hope, the situation may be better, but existence and uniqueness of measures of maximal Hausdorff dimension is in general not settled. By using the Ledrappier-Young formula [63]

$$(4.1) \quad HD(\mu) = \frac{h_\mu(f)}{\lambda_\mu^2(f)} + \left[\frac{1}{\lambda_\mu^2(f)} - \frac{1}{\lambda_\mu^1(f)} \right] h_{\pi\mu}(f_*)$$

(where f_* is the action of f on the leaves of the strong unstable foliation, π is the projection onto this space of leaves, and $0 < \lambda_\mu^2 < \lambda_\mu^1$ are the Lyapunov exponents of f) and a coding by Markov partitions, they reduce these problems to questions like the following:

Question 4.2. For a factor map $\pi : X \rightarrow Y$ as above and $\alpha > 0$, is the expression

$$(4.2) \quad \phi_\alpha(\mu) = h(\mu) + \alpha h(\pi\mu)$$

maximized by a unique ergodic measure?

It was proved in [105] that in certain cases the answer is yes, there is a unique μ_α which maximizes $\phi_\alpha(\mu)$, and it is the unique equilibrium state of a certain potential function (see below for the precise statement). To handle the remaining cases one might try a two-step strategy:

(1) Find the measure(s) ν_α on Y that constitute the image under π of the optimal measure(s) μ_α , and then

(2) find the relatively maximal measure(s) over the ν_α .

Question (4.1) addresses part (2) of this plan.

Some of the important tools available for such an investigation are pressure and equilibrium states (see [98, 47, 54]), relative pressure and relative equilibrium states [62, 117], and compensation functions [117]. Recall that if $V : X \rightarrow \mathbb{R}$ is a continuous “potential function”, then the *pressure* of V is defined to be

$$(4.3) \quad P_X(V) = \lim_{\epsilon \searrow 0} \uparrow \limsup_{n \rightarrow \infty} \frac{1}{n} \log \sup \left\{ \sum_{x \in A} e^{S_n V(x)} : A \text{ is } n, \epsilon\text{-separated} \right\}$$

($S_n V(x) = \sum_{k=0}^{n-1} V(\sigma^k x)$). The Variational Principle states that

$$(4.4) \quad P_X(V) = \sup \left\{ h_\mu(\sigma) + \int_X V d\mu : \mu \in \mathcal{M}(X) \right\},$$

providing an interpretation of $P(V)$ as the minimum of free energy over all possible states μ for a fixed potential V . Any measure μ on X which achieves this supremum is called an *equilibrium state* for V .

For a factor map between compact topological dynamical systems Ledrappier and Walters [62] defined *relative pressure* and proved a relative variational principle, one consequence of which is that given $\pi : X \rightarrow Y$ as above and ν ergodic on Y , the ergodic measures μ that have maximal entropy among all measures in $\pi^{-1}\{\nu\}$ have fiber entropy given by

$$(4.5) \quad h_\mu(X|Y) = \int_Y \lim_{n \rightarrow \infty} \frac{1}{n} \log |\pi^{-1}[y_0 \dots y_{n-1}]| d\nu(y).$$

($|\pi^{-1}[y_0 \dots y_{n-1}]|$ is the number of n -blocks in X that map under π to the n -block $y_0 \dots y_{n-1}$.) By the Subadditive Ergodic Theorem, the limit inside the integral exists a.e. with respect to each ergodic measure ν on Y , and it is constant a.e.. The quantity

$$(4.6) \quad P(\sigma, \pi, 0)(y) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log |\pi^{-1}[y_0 \dots y_{n-1}]|$$

is the *relative pressure* of the function 0 over $y \in Y$. The maximum possible fiber entropy may be thought of as a “relative topological entropy over ν ”; we denote it by $h_{\text{top}}(X|\nu)$.

To understand when a Markov measure on Y has a Markov measure on X in its preimage under π , Boyle and Tuncel introduced the idea of a compensation function [19], and the concept was developed further by Walters [117]. Given a factor map $\pi : X \rightarrow Y$ between topological dynamical systems (for us always SFT’s), a *compensation function* is a continuous function $F : X \rightarrow \mathbb{R}$ such that

$$(4.7) \quad P_Y(V) = P_X(V \circ \pi + F) \quad \text{for all } V \in \mathcal{C}(Y).$$

The idea is that, because $\pi : \mathcal{M}(X) \rightarrow \mathcal{M}(Y)$ is many-to-one, we always have

$$(4.8) \quad P_Y(V) = \sup\{h_\nu(\sigma) + \int_Y V d\nu : \nu \in \mathcal{M}(Y)\}$$

$$(4.9) \quad \leq \sup\{h_\mu(\sigma) + \int_X V \circ \pi d\mu : \mu \in \mathcal{M}(X)\},$$

and a compensation function F can take into account, for all potential functions V on Y at once, this extra freedom, information, or free energy that is available in X as compared to Y , because of the ability to move around in fibers over points of Y . Among the important facts established in [19, 117] (see also [22]) we note the following:

- (1) For a *subshift* X , there is always a compensation function in $\mathcal{C}(X)$.
- (2) The following are equivalent:
 - (a) There are Markov measures $\mu \in \mathcal{M}(X)$ and $\nu \in \mathcal{M}(Y)$ such that $\pi\mu = \nu$.
 - (b) For every Markov $\nu \in \mathcal{M}(Y)$ there is a Markov $\mu \in \mathcal{M}(X)$ such that $\pi\mu = \nu$.
 - (c) There is a *locally constant* (continuous function taking only finitely many values) compensation function on X .
- (3) $\pi(\max_X) = \max_Y$ (π is *uniform*) if and only if there is a *constant* compensation function.
- (4) π is *finite-to-one* if and only if 0 is a compensation function.
- (5) A function F in the Walters class $\mathcal{F}(X)$ of fairly smooth functions on X (see [117]) is a compensation function if and only if there is a constant $c > 0$ such that for all $y \in Y$, for all $n \geq 1$,

and for every choice of a set of points $E_n(y)$ that includes exactly one point from each cylinder comprising $\pi^{-1}[y_0 \dots y_{n-1}]$,

$$(4.10) \quad \frac{1}{c} \leq \sum_{x \in E_n(y)} e^{S_n F(x)} \leq c.$$

(6) Consequently, if $G \in \mathcal{F}(Y)$, then $G \circ \pi$ is a compensation function if and only if there is a constant $c > 0$ such that

$$(4.11) \quad \frac{1}{c} \leq e^{S_n G(y)} |\pi^{-1}[y_0 \dots y_{n-1}]| \leq c \quad \text{for all } y \in Y, n \geq 1.$$

Having recalled this necessary background, we can state two relevant results from [105].

Theorem 4.3. [105] *If there is a saturated compensation function $G \circ \pi$ with $G \in \mathcal{C}(Y)$, then the measures which maximize*

$$(4.12) \quad \phi_\alpha(\mu) = h(\mu) + \alpha h(\pi\mu)$$

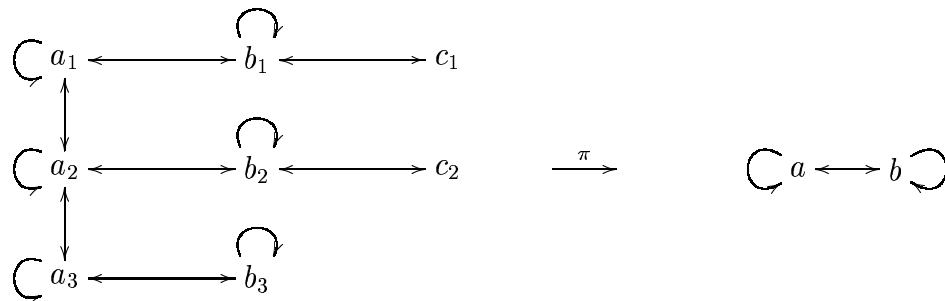
are the equilibrium states of $(\alpha/(\alpha + 1)) G \circ \pi$. Consequently, if there is such a G in $\mathcal{F}(Y)$, then there is a unique measure which maximizes $\phi_\alpha(\mu)$ (in which case we say that π is α -intrinsic).

Theorem 4.4. [105] *Suppose that $\nu \in \mathcal{E}(Y)$ and $\pi\mu = \nu$. Then μ is relatively maximal over ν if and only if there is $V \in \mathcal{C}(Y)$ such that μ is an equilibrium state of $V \circ \pi$.*

Notice that if there is a *locally constant* saturated compensation function $G \circ \pi$, then every Markov measure on Y is π -determinate with Markov relatively maximal lift, because in [117] it is shown that if there is a saturated compensation function $G \circ \pi$, then the relatively maximal measures over an equilibrium state of $V \in \mathcal{C}(Y)$ are the equilibrium states of $V \circ \pi + G \circ \pi$.

Further, \max_X is the unique equilibrium state of the potential function 0 on X , the unique maximizing measure for ϕ_0 ; and the relatively maximal measures over \max_Y are the equilibrium states of $G \circ \pi$, which can be thought of as the maximizing measures for ϕ_∞ .

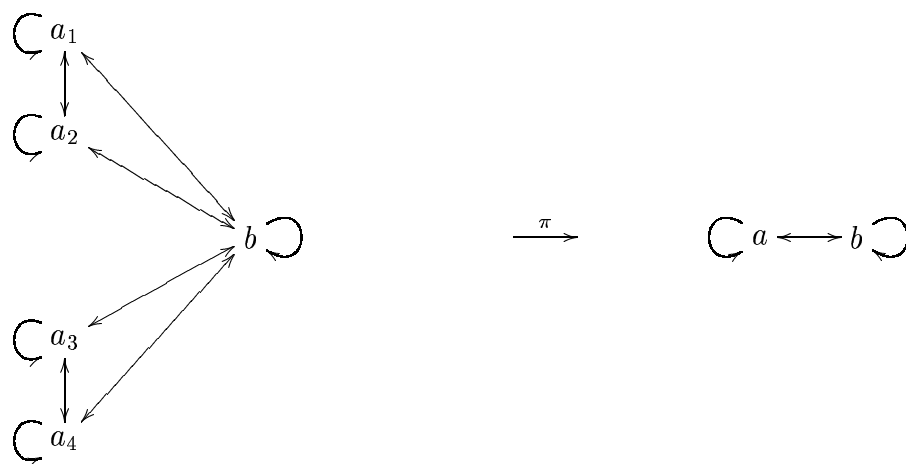
We begin with two examples of the failure of π -determinism. The first, from [105], is as follows:



Example 2: There is a non π -determinate (not fully supported) Markov measure.

The subshift X is given by the graph on the left and Y by the one on the right (the full 2-shift). The 1-block map π takes each a_i to a and each b_i and c_j to b . Let $\Gamma \subset Y$ be the golden mean SFT consist of all infinite sequences on the symbols a and b which do not contain the block aa , and let ν be a Markov measure on Γ , for example the Shannon-Parry measure \max_{Γ} . Then ν has two relatively maximal lifts, μ_1 and μ_2 , with each μ_i being the maximal measure supported on the subshift $X_i \subset X$ determined by the subgraph involving all the symbols with subscript i . (The subshift X_3 is included to show that entropy need not be constant in a fiber.)

Our second example involves measures supported on periodic orbits.



Example 3: There is a non π -determinate periodic-point measure.

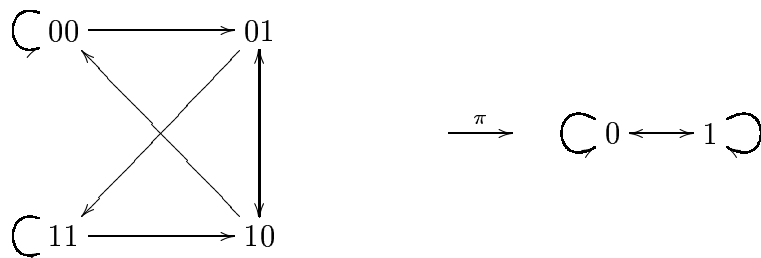
Let ν be the point mass on the fixed point $y = aaa \dots \in Y$. Then ν has exactly two relatively maximal ergodic lifts, namely the $1/2, 1/2$

Bernoulli measures on each of the two full 2-shifts in X , one on the symbols a_1 and a_2 , the other on the symbols a_3 and a_4 .

The situation is different if we consider the measure ν concentrated on the orbit of the periodic point $y = ababab \cdots \in Y$. Now the first-return map to $\pi^{-1}[a] = [a_1] \cup [a_2] \cup [a_3] \cup [a_4] \subset X$ is topologically the full 4-shift, since all transitions $a_i \rightarrow a_j$ are allowed (unlike in the preceding example, where $\pi^{-1}[a]$ is the union of two disjoint 2-shifts). Putting the maximal Bernoulli measure $\mathcal{B}(1/4, 1/4, 1/4, 1/4)$ on this first-return system defines the unique relatively maximal μ on X over this ν .

In each of the two preceding examples the failure of π -determinism can be blamed on lack of communication among fibers. An example suggested by Walters (see [117]) also shows that there can be *fully supported* ν on Y which are not π -determinate. Looking at Theorem 4.4, for such examples there are potential functions $V \in \mathcal{C}(Y)$ such that $V \circ \pi$ has two equilibrium states which project to the *same* ergodic measure on Y .

In this example, $X = Y = \Sigma_2$ =full 2-shift, and $\pi(x)_0 = x_0 + x_1 \pmod 2$ is a simple cellular automaton 2-block map. If we replace X by its 2-block recoding, so that π becomes a 1-block map, we obtain the following diagram:



Example 4: A factor map with fully supported non- π -determinate measures.

This is a finite-to-one map and hence is Markovian—for example, the Bernoulli $1/2, 1/2$ measure on Σ_2 is mapped to itself. The constant function 0 is a compensation function. Thus every Markov measure on Y is π -determinate: the equilibrium state μ_V of a locally constant V on Y lifts to the equilibrium state of $V \circ \pi$, which is the unique relatively maximal measure over μ_V (in fact it's the only measure in $\pi^{-1}\{\mu_V\}$).

For every ergodic ν on Y , all of $\pi^{-1}\{\nu\}$ consists of relatively maximal measures over ν , all of them having the same entropy as ν .

If $p \neq 1/2$, the two measures on the SFT X that correspond to the Bernoulli measures $\mathcal{B}(p, 1 - p)$ and $\mathcal{B}(1 - p, p)$ both map to the same measure ν_p on Y . Thus ν_p , which is fully supported on Y , is *not* π -determinate. (An entropy-decreasing example is easily produced by forming the Cartesian product of X with another SFT.)

Moreover, ν_p is the unique equilibrium state of some continuous function V_p on Y . Then the set of relatively maximal measures over ν_p , which is the entire set $\pi^{-1}\{\nu_p\}$, consists of the equilibrium states of $V_p \circ \pi + G \circ \pi = V_p \circ \pi$ [117], so this potential function $V_p \circ \pi$ has many equilibrium states.

Question 4.5. Construct such a function V_p explicitly.

These examples suggest that when looking for relatively maximal measures it may be advantageous to consider first-return maps to cylinder sets.

Theorem 4.6. *Suppose that $\pi : X \rightarrow Y$ has a singleton clump: there is a symbol a of Y whose inverse image is a singleton, which we also denote by a . Then every Markov measure on Y (fully supported, or at least assigning positive measure to $[a]$) is π -determinate.*

Proof. Denote the cylinder sets $[a]$ in X and in Y by X_a and Y_a , respectively. If ν is (1-step) Markov on Y , then the first-return map $\sigma_a : Y_a \rightarrow Y_a$ is countable-state Bernoulli with respect to the restricted and normalized measure $\nu_a = \nu/\nu[a]$: the states are all the loops or return blocks aC^i with $aC^i a = ac_1^i \dots c_r^i a$ appearing in Y and no $c_j^i = a$.

Under π^{-1} , the return blocks to $[a]$ expand into bands $aB^{i,j}$, with $aB^{i,j}a$ appearing in X and $\pi B^{i,j} = C^i$ for all i, j . Topologically, (X_a, σ_a) is a countable-state full shift on these symbols $aB^{i,j}$. We define μ_a to be the countable-state Bernoulli measure on (X_a, σ_a) which equidistributes the measure of each loop (state) of Y_a over its preimage band:

$$(4.13) \quad \mu_a[aB^{i,j}] = \frac{\nu_a[aC^i a]}{|\pi^{-1}[aC^i a]|} \quad \text{for all } i, j.$$

We show now that this choice of μ_a is relatively maximal over ν_a . Let λ_a be any probability measure on X_a which maps under π to ν_a . Then the countable-state Bernoulli measure on X_a which agrees with λ_a on all the 1-blocks $aB^{i,j}$ (its “Bernoullization”) has entropy no less than

that of λ_a and still projects to the Bernoulli measure ν_a , so we may as well assume that λ_a is countable-state Bernoulli. If $\lambda_a[aB^{i,j}] = q^{i,j}$ and $|\pi^{-1}(aC^i a)| = J_i$ for all i, j , then

$$(4.14) \quad h(X_a, \sigma_a, \lambda_a) = \sum_{i=1}^{\infty} \sum_{j=1}^{J_i} q^{i,j} \log q^{i,j}.$$

Note that for each i

$$(4.15) \quad \sum_{j=1}^{J_i} q^{i,j} = \nu_a[aC^i a]$$

is fixed at the same value for all λ_a . Thus for each i ,

$$(4.16) \quad \sum_{j=1}^{J_i} q^{i,j} \log q^{i,j}$$

is maximized by putting all the $q^{i,j}$ equal to one another.

Finally, this unique relatively maximal μ_a over ν_a determines the unique relatively maximal μ on X over ν on Y , since according to Abramov's formula

$$(4.17) \quad h(X, \sigma, \mu) = \mu[a] h(X_a, \sigma_a, \mu_a),$$

and $\mu[a] = \nu[a]$. □

As seen in Example 3, the relatively maximal measures over a periodic-point measure on Y can be found by analyzing an associated SFT on preimage blocks. Suppose that $C = c_1 \dots c_r$ is a block in Y which generates a periodic point $y = CCC \dots \in Y$ and ν is the ergodic invariant measure supported on the orbit of y . Not all of X is relevant now, just the subshift $X_y = \pi^{-1}\{y\}$. Paralleling the case of a singleton clump, denote by (X_C, σ_C) the first-return topological dynamical system to $\pi^{-1}[C]$. This system is in fact an SFT. Consider all blocks $b_i B b_j$ in X such that $\pi(b_i B b_j) = C c_1$. We form an SFT which has these blocks as symbols and allowed transitions $b_i B b_j \rightarrow b_{i'} B' b_{j'}$ if and only if $j = i'$, i.e., if and only if $b_i B b_j B' b_{j'} = b_i B b_{i'} B' b_{j'}$ is an allowed block in X (in $\pi^{-1}(CCc_1)$). This SFT, which keeps track of the allowed transitions in the fiber over y between the different preimages of C as we move from one repeat of C to the next, will be called the *block SFT above y* .

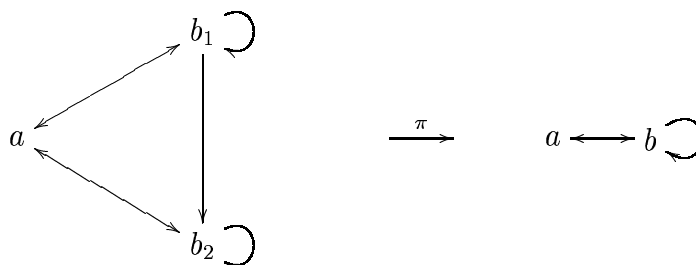
In (Y, σ, ν) , the first-return system to $[C]$ consists of a single fixed point, so any invariant measure μ_C on (X_C, σ_C) projects to the fixed point mass on $y \in [C]$. In the usual manner for induced transformations, μ_C uniquely determines an invariant measure μ on $\pi^{-1}\{y\} \subset X$,

and μ necessarily projects to the periodic-point measure ν . (In fact, $\mu = (\mu_C + \mu_C\sigma^{-1} + \dots + \mu_C\sigma^{-r+1})/r$ on X .)

Theorem 4.7. *Let ν be an ergodic measure on Y which is supported on the orbit of a periodic point $y = CCC\dots \in Y$. Then the relatively maximal measures over ν are determined by the maximal (Shannon-Parry) measures on the block SFT above y . Consequently, if the block SFT above y is irreducible, then the discrete measure on the orbit of y is π -determinate.*

Proof. Any relatively maximal measure over ν must determine a measure of maximal entropy on the block SFT (X_C, σ_C) . \square

It was shown in [105] that for the following factor map there is a saturated compensation function $G \circ \pi$ with $G \in \mathcal{C}(Y)$ but no such compensation function with $G \in \mathcal{F}(Y)$. There is a singleton clump, a .



Example 5: A factor map with a singleton clump.

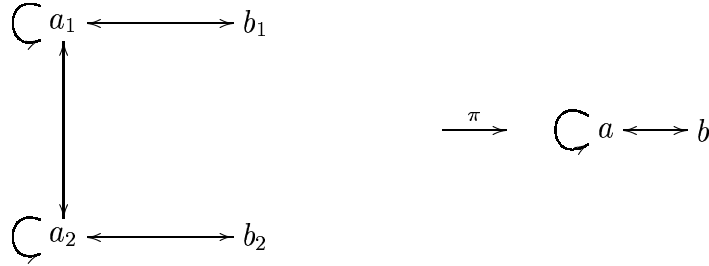
For each $k \geq 1$ the block $ab^k a$ in Y has $k + 1$ preimages, depending on when the subscript on b switches from 1 to 2. Let ν be Markov on Y . To each preimage $aB_1 aB_2 a \dots aB_r$ of $ab^{k_1} ab^{k_2} \dots ab^{k_r}$ the optimal measure μ_a assigns measure

$$(4.18) \quad \mu_a[aB_1 aB_2 a \dots aB_r] = \frac{1}{k_1 + 1} \dots \frac{1}{k_r + 1} \nu_a[ab^{k_1} ab^{k_2} \dots ab^{k_r}].$$

The unique relatively maximal measure over ν_a can be described in terms of fiber measures as follows. Given $y = ab^{k_1} ab^{k_2} \dots ab^{k_r} \dots \in Y_a$, $\mu_{a,y}$ chooses the preimages of each b^{k_i} with equal probabilities and independently of the choice of preimage of any other b^{k_j} . Then

$$(4.19) \quad \mu_a[aB_1 aB_2 a \dots aB_r] = \int_{Y_a} \mu_{a,y}[aB_1 aB_2 a \dots aB_r] d\nu_a(y).$$

In the following example there is no singleton clump, but the clumps are homogeneous with respect to π so there is a locally constant compensation function, and hence every Markov measure on Y is π -determinate and its unique relatively maximal lift is Markov.



Example 6: Every Markov measure is π -determinate.

In this case the return time to $[a]$ is bounded, so X_a is a finite-state SFT rather than the countable-state chain of the general case. There are six states, a_1a_1 , $a_1b_1a_1$, a_1a_2 , a_2a_2 , $a_2b_2a_2$, and a_2a_1 , according to the time 0 entries of $x \in X_a$ and $\sigma_a x$. Fix this order of the states for indexing purposes. It can be shown by direct calculation that for this example a stochastic matrix P determines a Markov measure on X_a that is relatively maximal over its image if and only if it is of the form

$$(4.20) \quad \begin{pmatrix} x & 1-2x & x & 0 & 0 & 0 \\ y & 1-2y & y & 0 & 0 & 0 \\ 0 & 0 & 0 & x & 1-2x & x \\ 0 & 0 & 0 & x & 1-2x & x \\ 0 & 0 & 0 & y & 1-2y & y \\ x & 1-2x & x & 0 & 0 & 0 \end{pmatrix}.$$

(In this case the image measure is also Markov.)

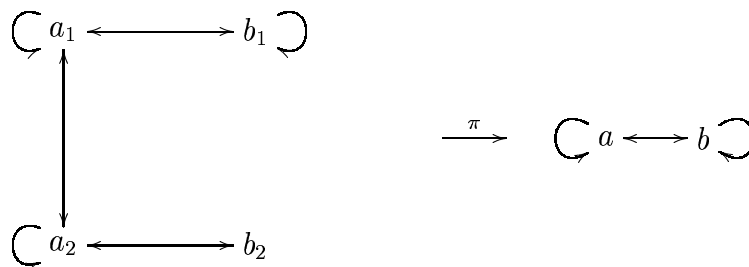
Here $0 < x, y < 1/2$ and the probability vector fixed by P is

$$(4.21) \quad p = \frac{1}{4y + 2(1-2x)}(y, 1-2x, y, y, 1-2x, y).$$

Further, given a (1-step) Markov measure ν on Y , put $K = \nu[aa]/\nu[aba]$. Then a stochastic matrix of the form (4.20) with fixed vector p satisfies $p_1+p_3+p_4+p_6 = \nu[aa]$ and $p_2+p_5 = \nu[aba]$ (so that the Markov measure μ that it determines projects to ν) if and only if $x = y = K/(2K + 2)$ (and then μ is relatively maximal over ν).

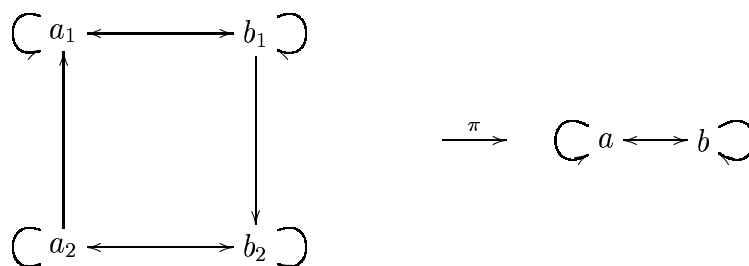
Let us make this example a little bit more complicated by adding a loop at b_1 , so that now the return time to $[a]$ is unbounded. It can

be verified that now there is still a continuous saturated compensation function, but there is no locally constant compensation function, so the code is not Markovian. However, if we look at higher block presentations of X and Y , we can find singleton clumps, for example $abba$. Therefore again every Markov measure on Y is π -determinate.



Example 7: Some higher block representations have singleton clumps.

Complicating this example a bit more, we can produce a situation in which there are no singleton clumps, not even for any higher block presentation.



Example 8: A factor map without singleton clumps.

For this example it can be shown that there is a continuous saturated compensation function $G \circ \pi$, but we do not know exactly which measures are π -determinate. Although the example appears simple, the question of how many fibers allow how much switching is complex. It seems that what is needed is a sort of relative Perron-Frobenius Theorem, giving the growth rates of the entries in restricted random matrix products. If $M_{y_0 y_1}$ is the 0, 1 transition matrix showing allowed transitions in X from the symbols in $\pi^{-1}\{y_0\}$ to those in $\pi^{-1}\{y_1\}$, then the entries in $M_{y_0 y_1} \dots M_{y_{n-2} y_{n-1}}$ count the numbers of paths above $y_0 \dots y_{n-1}$ that begin and end with given symbols in X . We need to

understand the asymptotics of these counts, even more delicate asymptotics than the difficult problems of Lyapunov exponents for random matrix products (cf. [15]).

Some of these considerations may have a connection with part of the theorem of Livsič about cocycle-coboundary equations (see [53, p. 609]):

Theorem 4.8 (Livsič). *Let (X, T) be a topological dynamical system and $G \in \mathcal{C}(X)$ a Hölder continuous function ($|G(x) - G(y)| \leq Kd(x, y)^\alpha$ for some constants K and $\alpha > 0$). Suppose that for every periodic point $y \in X$ we have*

$$(4.22) \quad \sum_{z \in \mathcal{O}(y)} G(z) = 0.$$

Then there is a Hölder continuous function g on X , with the same Hölder exponent and unique up to an additive constant, such that

$$(4.23) \quad G = g - gT.$$

For y a periodic point in Y , the set $X_y = \pi^{-1}\{y\}$ is a subshift whose usefulness we have already seen in analyzing the preimages of measures supported on periodic orbits. Recall the following theorem of Walters:

Theorem 4.9 (Walters [117]). *For $G \in \mathcal{C}(Y)$, $G \circ \pi$ is a compensation function if and only if for all $\nu \in \mathcal{E}(Y)$*

$$(4.24) \quad \int_Y \lim_{n \rightarrow \infty} \frac{1}{n} \log |\pi^{-1}[y_0 \dots y_{n-1}]| d\nu(y) = - \int_Y G d\nu.$$

In terms of the subshift X_y this says

$$(4.25) \quad |\mathcal{O}(y)|h_{\text{top}}(X_y) = - \sum_{z \in \mathcal{O}(y)} G(z), \text{ or}$$

$$\sum_{z \in \mathcal{O}(y)} [G(z) + h_{\text{top}}(X_z)] = 0.$$

In the finite-to-one case, $h_{\text{top}}(X_z) \equiv 0$, so (4.25) is satisfied if and only if $G \sim 0$, a consequence of the Livsič theorem. Thus in this case it suffices, for G Hölder, to check this condition on periodic orbits to see whether $G \sim 0$, so that $G \circ \pi$ is a compensation function. In other cases, failure to satisfy (4.25) would be enough to show that there does not exist a continuous saturated compensation function.

J.-P. Thouvenot (private communication) has suggested the possible relevance of the concept of “relatively Markov” in the search for

relatively maximal measures. When Y is a single point and μ is an invariant measure on X , the 1-step Markovization μ^1 of μ is the 1-step Markov measure on X that agrees with μ on 2-blocks; its entropy is at least as large as that of μ , and this leads to the conclusion that the maximal measure on X is Markov.

Definition 4.10. Let $\pi : (X, \mathcal{B}, \mu, T) \rightarrow (Y, \mathcal{C}, \nu, S)$ be a factor map between measure-preserving dynamical systems and α a finite generating partition for X . We say that μ is *relatively Markov for α over \mathcal{C}* if it satisfies one of the following two equivalent conditions:

- (1) $\alpha \perp_{T^{-1}\alpha \vee \mathcal{C}} \alpha_2^\infty$;
- (2) $H_\mu(\alpha | \alpha_1^\infty \vee \mathcal{C}) = H_\mu(\alpha | T^{-1}\alpha \vee \mathcal{C})$.

Question 4.11. For a 1-step factor map between SFT's with α equal to the time-0 partition, is every relatively maximal measure relatively Markov?

Question 4.12. Does every invariant measure μ on X have a *relative Markovization* over $\nu = \pi\mu$? If it exists, is it unique? Perhaps relative Markovizations can be formed as follows: find $V \in \mathcal{C}(Y)$ with ν the unique equilibrium state of V , and take for μ^1 an equilibrium state of $V \circ \pi$. When are these unique? Are they always relatively Markov?

Finally, we discuss several possible methods for constructing relatively maximal measures over a fixed ergodic measure ν on Y , based on Bowen's construction of measures of maximal entropy on SFT's [16, 17], [23, pp. 210–223]. Bowen constructs the Shannon-Parry measure on X by for each n defining μ_n to be the probability measure on X that is the sum of equal point masses on the periodic points of period n (fixed points of σ^n) and showing that the set of weak* limit points of the sequence (μ_n) consists of a single measure, the unique measure of maximal entropy on X . The idea is based on the fact that the maximal entropy measure on X assigns nearly equal measures, insofar as possible, to all cylinder sets of each length n . (An extension of this method to the construction of equilibrium states can be seen around p. 81 in [54].) We want to try something similar but make sure that we arrive at invariant measures μ on X that project to the given ν on Y .

Method 1: For each n and each cylinder set $[y_0 \dots y_{n-1}]$ in Y , equidistribute $\nu[y_0 \dots y_{n-1}]$ over the preimages $\pi^{-1}[y_0 \dots y_{n-1}]$, by selecting a point x_B in each cylinder set $B \subset \pi^{-1}[y_0 \dots y_{n-1}]$ and assigning to each x_B the point mass of measure $\nu[y_0 \dots y_{n-1}] / |\pi^{-1}[y_0 \dots y_{n-1}]|$. In terms

of fiber measures,

$$(4.26) \quad \mu_n = \int_Y \mu_y^{(n)} d\nu(y),$$

and the $\mu_y^{(n)}$ are chosen so as to equidistribute the available measure among the n -blocks within the fiber: for each cylinder set $[B] = [b_0 \dots b_n]$ in X with $\pi(B) = [y_0 \dots y_n]$, we choose $x_B \in B \cap \pi^{-1}[y_0 \dots y_n]$ and let

$$(4.27) \quad \mu_y^{(n)} = \frac{\sum_B \delta_{x_B}}{|\pi^{-1}[y_0 \dots y_{n-1}]|} .$$

To produce shift-invariant measures we seek weak* limits of time averages of these measures as well as limits as $n \rightarrow \infty$.

Method 2: For each n and each repeatable block $[y_0 \dots y_{n-1}]$ in Y (i.e., which forms a periodic point $y \in Y$), equidistribute $\nu[y_0 \dots y_{n-1}]$ over the periodic preimages of y . Measures produced in this way are automatically shift invariant, and hence so are their weak* limits.

With each of the methods, we would want to show that weak* limit points of time averages of the μ_n are shift-invariant, project under π to ν , and have maximal entropy among all measures in $\pi^{-1}\{\nu\}$. Moreover, we would like to show that all measures of maximal relative entropy are found in the closed convex hull of the set of ergodic measures associated with the weak* limit points of sequences of measures constructed in this way.

It seems that these methods work in the two cases analyzed above, namely when there is a singleton clump and when ν is a periodic-point measure with irreducible associated block SFT. Indeed, Method 1 works if there is a singleton clump a , since then for each block B in X_a we have

$$(4.28) \quad \lim_{n \rightarrow \infty} \mu_{a,n}[B] = \int_{Y_a} \frac{1}{|\pi^{-1}\pi B|} d\nu_a(y) = \frac{\nu_a[\pi B]}{|\pi^{-1}\pi B|}.$$

This implies that the method also works on X over Y .

Similarly, if ν is concentrated on the orbit of a periodic point $y = CCC \dots \in Y$, Method 2 works just as it does in Bowen's approach. As for Method 1, we can apply the Perron-Frobenius Theorem to the adjacency matrix A of the (first-return to $\pi^{-1}[C]$) block SFT above y . As before let the states of this SFT be the blocks $b_i B b_j$ such that $\pi(b_i B b_j) = C c_1$ (c_1 denoting the first entry of C). Suppose that the

adjacency matrix A has maximum eigenvalue λ and right and left eigenvectors r and l . It is known (see [103]) that

$$(4.29) \quad A^k = \lambda^k r l^{\text{tr}} + O(k^s |\lambda_2|^k) = \lambda^k (r_i l_j) + o(\lambda^n),$$

where λ_2 is the eigenvalue of next largest modulus after λ , with multiplicity m_2 , and $s = m_2 - 1$. Using Method 1, equidistributing the available measure in each fiber amounts to assigning equal measure to a chosen point in each preimage n -block of the initial block in the fiber. For example, for a particular symbol bBb' ,

$$(4.30) \quad \mu_C[bBb'] = \lim_{m \rightarrow \infty} \frac{|(bBb')(b_{i_1} B_{i_1} b'_{i_1}) \dots (b_{i_{m-1}} B_{i_{m-1}} b'_{i_{m-1}})|}{|\pi^{-1}(C^m)|}$$

$$(4.31) \quad = \lim_{m \rightarrow \infty} \frac{r_{bBb'} \lambda^m}{\sum_{s,t} l_s r_t \lambda^m}$$

$$(4.32) \quad = \frac{r_{bBb'}}{\sum_{s,t} l_s r_t} .$$

The fiber measure of a cylinder set determined by a longer initial block on the symbols bBb' is calculated in an analogous way. Looking at larger n and taking time averages so as to produce shift invariance, we obtain measures which assign approximately equal mass to all allowed cylinder sets of a given length. Therefore in the limit we obtain the maximal measure on X_C . If during this process each measure is pushed forward from X_C to X , the process works similarly on all of X .

5. FINDING ONE'S WAY WITHIN TILING DYNAMICAL SYSTEMS

In this section we discuss two types of information handling related to tiling dynamical systems: detecting imbedded hierarchical structure (following [92, 93, 94]) and the nature of factor maps (after [84, 95]).

5.1. Tiling dynamical systems and substitutions. To define a tiling dynamical system, one begins with a finite collection of subsets of \mathbb{R}^d , called *prototiles*. It is assumed that the prototiles are topologically fairly decent; for example, each should be compact and equal to the closure of its interior. A *tiling* of \mathbb{R}^d is a covering \mathcal{T} of \mathbb{R}^d by congruent copies of prototiles which intersect only along their boundaries. Frequently one restricts to coverings only by *translations* of prototiles. The copies of prototiles appearing in the covering are called *tiles*, and the *tile type* of a tile is the prototile of which it is a copy. In some situations the prototiles are endowed with labels, so that tiles may be distinguishable even though they have the same shape. The space of tilings is compact metrizable [96] when we consider two tilings to be close if in a large neighborhood of the origin the two unions of tile boundaries are close in Hausdorff metric. \mathbb{R}^d acts naturally on the space of tilings by translation. A *tiling dynamical system* consists of a pair (X, \mathbb{R}^d) , with X a closed set that is invariant under the action. Frequently X is the orbit closure $X(\mathcal{T})$ of a tiling \mathcal{T} .

There are many ways to construct interesting tilings and dynamical systems, two particularly important ones being imposing matching rules restricting tile adjacencies (analogous to SFT's in symbolic dynamics) and iterating a substitution/rescaling process (like substitution dynamical systems). The second method leads to tilings that are self-similar, and thus all tilings in their orbit closures embody a hierarchical structure that preserves some information about the substitution. Looking at pictures of self-similar tilings can be a dazzling experience. How can the hierarchical structure be recognized, and how can the original substitution (or one that does essentially the same job) be recovered by examining a self-similar tiling?

Symbolic dynamical systems may be thought of as one-dimensional tilings in which all tiles are identical in shape but may have different labels chosen from a finite alphabet. Already in this case it is an interesting problem to recognize hierarchical structure. F. Durand [25] proved that a sequence $\omega = \omega_1\omega_2\dots$ on a finite alphabet is *substitutive*, meaning that it is the image under a 1-block map of the fixed point

of a primitive substitution, if and only if its set of derived sequences is finite. A *derived sequence* is formed as follows. Fix n and look at the initial block $A_n = \omega_1 \dots \omega_n$ of ω and its occurrences in ω . If ω is almost periodic (meaning each block that appears appears with bounded gap), then ω factors uniquely into a concatenation of blocks $\xi = B_1 B_2 \dots$ with A_n the initial n -block of each B_i and not appearing elsewhere in ω , and with the lengths of the B_i bounded. Recode ξ by replacing B_1 by the symbol 0 wherever it appears in ξ , the next B_i which is not equal to B_1 by 1, and so on. The resulting sequence on a finite alphabet is the n 'th derived sequence of ω .

For tiling dynamical systems in higher dimensions we lack the concepts of order and first returns, plus shapes and adjacencies of tiles can get quite complicated. In the dissertation of N. Priebe [92] the concept of a Voronoi tiling determined by a patch in a given tiling was used as a substitute for a return sequence to generalize Durand's result. In order to state the generalization we have to establish some terminology.

Let \mathcal{T} be a tiling of \mathbb{R}^d , regarded as the collection of tiles it comprises. A \mathcal{T} -*patch* is a finite subset of \mathcal{T} . Given a set $A \subset \mathbb{R}^d$, we define the *outer patch determined by A* to be

$$(5.1) \quad \mathcal{T}(A) = \cup\{T \in \mathcal{T} : A \cap T \neq \emptyset\}$$

and the *inner patch determined by A* to be

$$(5.2) \quad \mathcal{T}_0(A) = \cup\{T \in \mathcal{T} : T \subset A\}.$$

An *elementary patch* is the outer patch of a point. The following generalizes the idea of a sliding block code in symbolic dynamics. Denote by $B_R(x)$ the ball of radius $R \geq 0$ about $x \in \mathbb{R}^d$. We say that a tiling \mathcal{S} is *locally derived* from a tiling \mathcal{T} , and write $\mathcal{T} \rightarrow_{LD} \mathcal{S}$, if there is $R > 0$ (the "window size") such that

$$(5.3) \quad \begin{aligned} x, y \in \mathbb{R}^d, \mathcal{T}(B_R(x)) = \mathcal{T}(B_R(y)) + (x - y) \\ \text{implies } \mathcal{S}(\{x\}) = \mathcal{S}(\{y\}) + (x - y). \end{aligned}$$

The idea is that the tile of \mathcal{S} containing each point $x \in \mathbb{R}^d$ can be determined, along with the precise position of x in that tile, by examining an R -neighborhood of x in the tiling \mathcal{T} . If $\mathcal{T} \rightarrow_{LD} \mathcal{S}$, then there is a continuous onto factor map of tiling dynamical systems, $(X(\mathcal{T}), \mathbb{R}^d) \rightarrow (X(\mathcal{S}), \mathbb{R}^d)$, which we call a *local code*: for each tiling \mathcal{T}' in the orbit closure $X(\mathcal{T})$ of \mathcal{T} and $x \in \mathbb{R}^d$, find $x' \in \mathcal{T}$ such that $\mathcal{T}'(B_R(x)) = \mathcal{T}(B_R(x')) + (x - x')$ and replace the patch $\mathcal{T}'(B_R(x))$ in \mathcal{T}' by $\mathcal{S}(\{x'\}) + (x - x')$. This is a well-defined process, independent of

x' . If each of \mathcal{T} and \mathcal{S} is locally derivable from the other, then we say that the two tilings are *mutually locally derivable*, abbreviated MLD.

A tiling \mathcal{T} is called *syndetic* (alternatively almost periodic, repetitive, locally isomorphic, . . .) if for every \mathcal{T} -patch P there is $R > 0$ such that every R -ball in \mathbb{R}^d contains a translate of P . The infimum of all such R for a given patch P is called the *repetitivity radius* of P . A tiling \mathcal{T} has *finite local complexity* (abbreviated FLC) if for each $R > 0$ every \mathcal{T} -patch of diameter R is a translate of one of a finite list of \mathcal{T} -patches.

An *expansion map* is a linear map $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ which preserves orientation and for which there is a constant $\lambda > 1$ such that

$$(5.4) \quad \|\phi(x)\| = \lambda \|x\| \text{ for all } x \in \mathbb{R}^d.$$

We define $\phi(\mathcal{T})$ to be the tiling consisting of the images under ϕ of the tiles of \mathcal{T} .

Definition 5.1. A tiling \mathcal{T} in \mathbb{R}^d is *self-similar* if it is syndetic, has FLC, and there is an expansion map ϕ on \mathbb{R}^d such that (i) for each $T \in \mathcal{T}$, $\phi(T)$ is a union of tiles in \mathcal{T} , and (ii) $T_1, T_2 \in \mathcal{T}$ are translates of one another if and only if the patches $\mathcal{T}_0(\phi(T_1))$ and $\mathcal{T}_0(\phi(T_2))$ are translates of one another (so that $\phi(T_1)$ and $\phi(T_2)$ are unions of tiles of \mathcal{T} in the same way).

We imagine that in the case of a self-similar tiling there is a rule for expanding each prototile by a factor λ and then decomposing it into a union of tiles. When this rule is applied simultaneously to each tile of \mathcal{T} we produce not a new tiling but just reproduce \mathcal{T} . (See Figure 16 for a 1-dimensional example.)

Definition 5.2. A tiling \mathcal{T} in \mathbb{R}^d is *pseudo-self-similar* if it is syndetic, has FLC, and there is an expansion map ϕ on \mathbb{R}^d such that $\phi(\mathcal{T}) \rightarrow_{LD} \mathcal{T}$.

The idea here is that the expansion of each tile of \mathcal{T} is only approximately a union of tiles of \mathcal{T} . Every self-similar tiling is pseudo-self-similar. For a nonperiodic pseudo-self-similar tiling with expansion map ϕ , the tilings \mathcal{T} and $\phi(\mathcal{T})$ are mutually locally derivable [94]. The orbit-closure of a pseudo-self-similar tiling is minimal, uniquely ergodic, and not strongly mixing [92, 110, 94].

We proceed now to define the derived Voronoi tilings $V_r, r > 0$ for a tiling \mathcal{T} which generalize the derived sequences $A_n, n = 1, 2, \dots$ of Durand. Let \mathcal{T} be a syndetic tiling and let $r > 0$. Copies of the patch $P_r = \mathcal{T}(B_r(0))$ appear throughout \mathbb{R}^d ; specify their locations precisely

by letting

$$(5.5) \quad \mathcal{L}_r = \{x \in \mathbb{R}^d : \mathcal{T}(B_r(x)) = P_r + x\}.$$

Then each \mathcal{L}_r is a *Delaunay set*: there are constants $\delta, K > 0$ such that each ball of radius K contains at least one point of \mathcal{L}_r , and each ball of radius δ contains at most one point of \mathcal{L}_r . For such a set its family of *Voronoi cells*, the cell C_x of $x \in \mathcal{L}_r$ consisting of all points of \mathbb{R}^d which are at least as close to x as to any other point of \mathcal{L}_r , forms a nice tiling V_r of \mathbb{R}^d (see Figure 15).

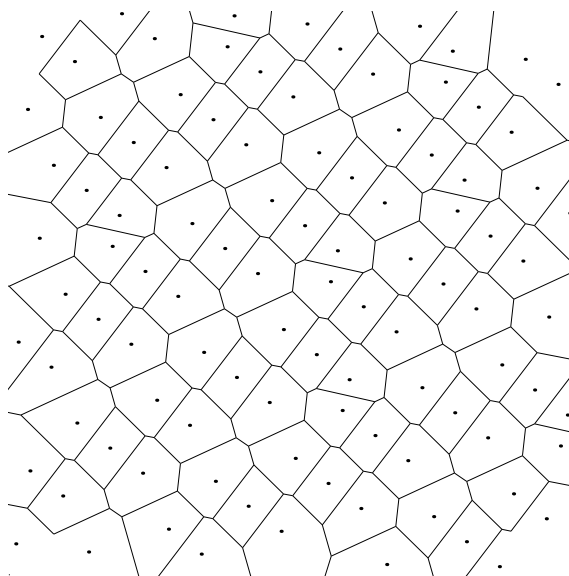


FIGURE 15. Part of a Delaunay set and its Voronoi tiling.

Because \mathcal{T} has finite local complexity, each of its derived Voronoi tilings V_r has only finitely many translation equivalence classes of tiles. However, some information about the original tiling can be lost if we only know the shapes of the Voronoi cells, so it is desirable to attach labels to these new tiles. Denote by $R(r)$ the repetitivity radius of P_r . Then we attach to the tile $C_x, x \in \mathcal{L}_r$, as a label the translation equivalence class of the patch $\mathcal{T}(B_{2R(r)}(x))$ which determines it.

Priebe [92, 93] shows that each of the tilings $V_r, r > 0$, is MLD with \mathcal{T} , and consequently the orbit-closure dynamical systems $(X(\mathcal{T}), \mathbb{R}^d)$ and $(X(V_r), \mathbb{R}^d)$ are topologically conjugate. Moreover, generalizing the two halves of the theorem of Durand separately, she proved the following:

Theorem 5.3. [92]

- (1) *Every nonperiodic syndetic self-similar tiling of \mathbb{R}^d has only a finite number of derived Voronoi tilings up to similarity in the following strong sense: there are an expansion mapping $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and a finite set $\{r_1, \dots, r_n\}$ of positive numbers such that for each $r > 0$ there are $i \in \{1, \dots, n\}$ and $j \in \{0, 1, 2, \dots\}$ such that $V_r = \psi^j V_{r_i}$.*
- (2) *If \mathcal{T} is a nonperiodic syndetic tiling of \mathbb{R}^d which has a finite number of derived Voronoi tilings in the above sense, with an expansion map ψ , then \mathcal{T} is pseudo-self-similar. Moreover, there is $k = 1, 2, \dots$ such that ψ^k serves as the expansion map in the definition of pseudo-self-similarity for \mathcal{T} .*

For planar tilings, Priebe and Solomyak [94] were able to “close the circle” on this question, first using derived Voronoi tilings and an iterative process to convert pseudo-self-similar tilings to self-similar ones, although possibly having messy boundaries.

Theorem 5.4. [94] *Any two-dimensional pseudo-self-similar tiling is mutually locally derivable with a bonafide self-similar tiling.*

Theorem 5.5. [94] *A two-dimensional, nonperiodic, syndetic tiling is pseudo-self-similar if and only if its set of similarity classes of derived Voronoi tilings is finite.*

Priebe also analyzed the hierarchical structures of planar tilings from a combinatorial viewpoint, regarding them (more properly their boundaries) as *graphs*. They also have *dual graphs* whose vertices are labelled tiles and whose edges join adjacent tiles. The *facets* correspond to tile interiors. The vertices, edges, and facets of the graph $G(\mathcal{T})$ associated to a tiling \mathcal{T} of \mathbb{R}^2 are labeled by elementary patches that they determine, registering how the graph elements reside in the tiling. Priebe calls two tilings *combinatorially isomorphic* if their labeled graphs are labeled-graph isomorphic. She defines graph substitutions for a wide class of graphs and calls a graph *substitutive* if it is the image under a labeled-graph homomorphism of a graph which is labeled-graph isomorphic to its image under a graph substitution. Then a tiling is defined to be *combinatorially substitutive* if its associated labeled graph is substitutive. The following theorem is another possible generalization of the Durand result; again it is not yet a characterization.

Theorem 5.6. [92]

- (1) *Every nonperiodic tiling of the plane that is locally derived from a self-similar tiling of the plane has a finite number of combinatorial isomorphism classes of derived Voronoi tilings.*
- (2) *Every nonperiodic syndetic tiling of the plane that has a finite number of isomorphism classes of derived Voronoi tilings is combinatorially substitutive.*

Question 5.7. What is a complete and simple characterization of the combinatorially substitutive tilings of \mathbb{R}^2 ?

5.2. Factor maps between tiling dynamical systems. Above we discussed local codes as factor maps between tiling dynamical systems. Surprisingly, not all factor maps between tiling dynamical systems are local codes: there are examples of factor maps which have the property that, while information within a finite window is sufficient to determine the tile type that contains the origin, it is not sufficient completely to specify the location of the origin within its tile. One class of examples [84] involves one-dimensional tiling systems and solvability of some cohomological equations, and Radin and Sadun [95] have given an example with substitution tilings that uses properties of the hierarchical structure.

A one-dimensional tiling dynamical system is the same as the flow built under a function on a subshift over a finite alphabet: the tiles are intervals of various lengths, and the translation flow slides along within a tile until it shifts to the next tile. Following [84], we establish some notation for working with these systems.

Let (X, T) and (Y, S) be subshifts on finite alphabets A and B , and take continuous height functions $g : X \rightarrow (0, \infty)$ and $h : Y \rightarrow (0, \infty)$ which take only finitely many values. By recoding if necessary, we may assume that g and h depend only on the central entry: $g(x) = g_0(x_0)$, $h(y) = h_0(y_0)$. Denote by $((X, T)_g, \mathbb{R})$ and $((Y, S)_h, \mathbb{R})$ the flows built under the ceiling functions g and h . Recall that the flow under a function $(X, T)_g$ is the quotient space of $(X \times \mathbb{R}, \mathbb{R})$ (with the action $(x, s)t = (x, s + t)$) under the equivalence relation \sim generated by identifying $(x, g(x))$ and $(Tx, 0)$. As in [56, 57], we denote by $[x, s]$ the equivalence class of a pair (x, s) when $x \in X, 0 \leq s < g(x)$. To each equivalence class $\xi \in (X, T)_g$ are associated a unique symbolic sequence $\pi_X \xi \in X$ and a unique $\pi_{\mathbb{R}} \xi \geq 0$ such that $0 \leq \pi_{\mathbb{R}} \xi < g(\pi_X \xi)$ and $\xi = [\pi_X \xi, \pi_{\mathbb{R}} \xi]$.

In some situations a factor map $\pi : (X, T) \rightarrow (Y, S)$ between the base dynamical systems (subshifts in this case) can give rise to a factor map

$\phi : ((X, T)_g, \mathbb{R}) \rightarrow ((Y, S)_h, \mathbb{R})$ between the flows built under functions (one-dimensional tiling dynamical systems). Suppose that there is a function $t : X \rightarrow \mathbb{R}$ such that

$$(5.6) \quad t(Tx) - t(x) = g(x) - h(\pi x) \quad \text{for all } x \in X.$$

Then putting

$$(5.7) \quad \phi[x, s] = [\pi x, 0](s + t(x)) \quad \text{for all } [x, s] \in ((X, T)_g, \mathbb{R})$$

defines a continuous map that commutes with the two translation actions. Factor maps that arise in this way are called *simple* [56, 57].

Example 5.8. A simple factor map between tiling dynamical systems which is not a local code:

For the base subshifts (X, T) and $(Y, S) = (X, T)$ we take a (generalized) Sturmian subshift coding translation modulo 1 by an irrational α . Specifically, define

$$(5.8) \quad \omega(n) = \chi_{[0,1/2)}(n\alpha \bmod 1) \quad \text{for all } n \in \mathbb{Z},$$

and let X be the orbit closure of ω under the shift transformation $T = \sigma$. For the factor map $\pi : (X, T) \rightarrow (Y, S)$ we take the identity.

The small trick here is to resist setting up the height functions h and g and then seeking the function t to satisfy the cohomological equation (5.6), but rather to define h and t first and then use (5.6) to define g .

To define the height function h , we use the (continuous) factor map $\rho : (X, \sigma) \rightarrow ([0, 1), R_\alpha)$, with R_α translation modulo 1 by α . Choose $\delta \in \mathcal{O}_{R_\alpha}(0)$ near 0, say $\delta = q\alpha \bmod 1 \in (0, 1/4)$. Choose distinct $\eta_1, \eta_2 > 5$, and let

$$(5.9) \quad h(y) = \begin{cases} \eta_1 & \text{if } \rho y \in (0, \delta) \\ \eta_2 & \text{if } \rho y \in (\delta, 1). \end{cases}$$

Recall that ρ is one-to-one on all of X except the orbits of the preimages of 0 and $1/2$, on which it is two-to-one. (For example, approaching the point $1/2$ from the right by a sequence $n_i\alpha \bmod 1$ yields a symbolic sequence $\xi = \lim(\sigma^{n_i}\omega)$ with $\xi(0) = 0$, while approaching from the left would produce a ξ' with $\xi'(0) = 1$.) In particular, $\rho^{-1}\{0\} = \{\omega, \omega'\}$, with $\omega(0) = 1, \omega'(0) = 0$; and $\rho^{-1}\{\delta\} = \{\zeta, \zeta'\}$, with $\zeta(0) = 1, \zeta'(0) = 0$. Defining

$$(5.10) \quad h(\omega) = \eta_1, h(\omega') = \eta_2, h(\zeta) = \eta_1, h(\zeta') = \eta_2$$

extends h to a continuous function on X .

Define

$$(5.11) \quad t(x) = \rho(x) \text{ if } 0 < \rho(x) < 1,$$

and extend t by continuity to the two points in $\rho^{-1}\{0\}$ (so that t maps to \mathbb{R} rather than to S^1). Now

$$(5.12) \quad t(Tx) - t(x) = \begin{cases} \alpha & \text{if } 0 < \rho(x) < 1 - \alpha \\ -1 + \alpha & \text{if } 1 - \alpha < \rho(x) < 1. \end{cases}$$

To satisfy (5.6), define

$$(5.13) \quad g(x) = t(Tx) - t(x) + h(\pi x).$$

Then g is continuous on X and takes finitely many positive values.

It can be checked that the factor map $\phi : ((X, T)_g, \mathbb{R}) \rightarrow ((Y, S)_h, \mathbb{R})$ defined by this setup is indeed onto and in fact one-to-one. However, this code is *not local*. By changing some entries far from the center in a symbolic sequence $x \in X$, we slightly change $t(x)$ and hence also $\phi[x, s] = [\pi x, 0](s + t(x))$. Thus while a central block of x can determine the central coordinate of πx and thus can determine which tile type in $\phi[x, s]$ is at the origin, it cannot determine where in this tile to place the origin—for this we need to know the full sequence x of tile types.

Example 5.9. Nonexistence of local or simple codes between tiling dynamical systems:

The preceding example can be modified in such a way that *no local code exists* between the tiling dynamical systems $((X, T)_g, \mathbb{R})$ and $((Y, S)_h, \mathbb{R})$. Choose $\gamma = 1 - \alpha$ and arrange that $1, \alpha, \eta_1, \eta_2$ are linearly independent over \mathbb{Z} . Then the ceiling function g takes values $\eta_1 + \alpha$ and $\eta_2 + \alpha - 1$. If there were a local code between the two systems, each time a long enough central block of tiles reappears in a tiling $[x, 0]$, at the corresponding translate of $\phi[x, 0]$ we should see exactly the same tile in exactly the same position. This forces a sum of tile lengths in the first system to equal a sum of tile lengths in the second, contradicting the linear independence.

It is easy to produce local codes between tiling dynamical systems which are not simple, and even local codes between tiling dynamical systems between which no simple code can exist. By just a slight adjustment of the foregoing example, one can also produce a factor map between one-dimensional tiling dynamical systems which is neither simple nor local, and moreover such that no simple nor local code can exist between the two systems. The idea is to select the base subshifts

(X, T) and (Y, S) so that no factor map can exist between them, for example by making them uniquely ergodic systems, one with purely discrete spectrum and the other weakly mixing. This precludes the existence of simple codes. The tile lengths can be again chosen linearly independent, so as to preclude the existence of local codes. But ceiling functions can still be selected and a continuous map defined so as to commute with the translation actions. The details are in [84].

Now we present the example of a non-local factor map due to Radin and Sadun [95]. The base of each flow under a function is the “Fibonacci” substitution subshift generated by the substitution

$$(5.14) \quad \zeta : 0 \rightarrow 01, 1 \rightarrow 0.$$

It may be preferable to work with the square of this substitution,

$$(5.15) \quad \zeta^2 : 0 \rightarrow 010, 1 \rightarrow 01,$$

which has a unique fixed point $\omega = 01001010 \cdots \in \Sigma_2^+$ and two fixed points in Σ_2 , generated by letting $\omega_{-1}\omega_0 = 00$ or 10 . The subshift X consists of all two-sided sequences $\xi \in \Sigma_2$ all of whose finite subblocks are found as subblocks of the fixed point ω ; it coincides with the orbit closure of either fixed point of ζ^2 . It is known that (X, σ) is uniquely ergodic and has purely discrete spectrum; it is in fact a Sturmian system that codes translation modulo 1 by the golden mean $\gamma = \gamma^2 - 1$.

The hierarchic structure of X (and of each sequence in it) is due to the generating substitution. Let us say that the alphabet elements 0 and 1 are *basic blocks of rank 0* and for each $k \in \mathbb{N}$ that the blocks $\zeta^k 0$ and $\zeta^k 1$ are *basic blocks of rank k* . Then for each k , each $x \in X$ factors uniquely into a concatenation of basic blocks of rank k [73, 74]. Put $H_k(x) = 0$ if the central (0th) place in x falls into a basic block $\zeta^k 0$ in this factorization, otherwise put $H_k(x) = 1$. Then each sequence $H \in \Sigma_2^+$ that is not eventually constant determines at least one $x \in X$ such that $H = H(x)$. (The eventually constant H 's correspond to shifts of the fixed points of ζ .)

Remark 5.10. In a hierarchical tiling system, the \mathbb{R}^d action is a sort of adic (because it is transverse) to the action of the substitution, which corresponds to the shift on the hierarchical codings of the tilings.

For each $n \geq 0$ denote by A_n the length of the basic block $\zeta^n 0$ and by B_n the length of $\zeta^n 1$. These numbers build up in the well-known

Fibonacci manner, $A_{n+1} = A_n + B_n, B_{n+1} = A_n$, so that

$$(5.16) \quad A_n = \frac{1}{\sqrt{5}} \left(\gamma^n - \frac{(-1)^n}{\gamma^n} \right), n \geq 1.$$

The usual Fibonacci tiling dynamical system consists of all tilings of \mathbb{R} obtained by choosing a sequence ξ in the Fibonacci substitution subshift X , replacing each 0 in ξ by an interval of length γ and each 1 in ξ by an interval of length 1, and selecting a placement of the origin in the resulting linear sequence of intervals. In terms of flow built under a function, this tiling system is obtained by using the height function g which takes the value γ on the cylinder $[0] = \{\xi \in \Sigma_2 : \xi_0 = 0\}$ and the value 1 on the cylinder $[1]$ to produce the flow $((X, T)_g, \mathbb{R})$. We think of the two kinds of tiles as having labels 0 and 1, respectively. The 0,1 basic blocks of tile labels of different levels in the hierarchical structure label patches of tilings, which we call *basic patches of ranks* 0, 1, 2, The fixed point ω of the substitution $\zeta : 0 \rightarrow 01, 1 \rightarrow 0$ labels the right half of a tiling \mathcal{T} of \mathbb{R} which has the beautiful self-similarity property illustrated in Figure 16. The expansion of \mathbb{R} consisting of multiplication

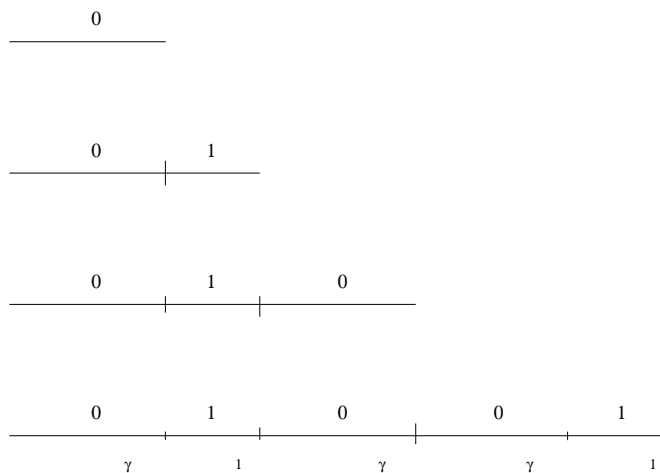


FIGURE 16. The self-similar Fibonacci tiling of \mathbb{R} .

by γ converts this tiling to one with larger tiles exactly composed of the previous smaller tiles. Alternatively stated, if we decompose each tile of \mathcal{T} labeled 0 into a tile labeled 0 and a tile labeled 1, with lengths having ratio γ , relabel each tile of \mathcal{T} labeled 1 by 0, and then expand the picture by a factor γ , we reproduce the original tiling \mathcal{T} .

Form the tiling dynamical system $((Y, S)_h, \mathbb{R})$ in the same way, except with different tile lengths: let $h \equiv \eta_0$ on $[0]$ and $h \equiv \eta_1$ on $[1]$. To

guarantee that the coding method works it is necessary to assume that

$$(5.17) \quad \gamma\eta_0 + \eta_1 = \gamma^2 + 1.$$

Looking at Equation (5.16), we see that the total length of the basic tile of rank n labeled by the block $\zeta^n 0$ is

$$(5.18) \quad \begin{aligned} & \eta_0 A_n + \eta_1 B_n = \eta_0 A_n + \eta_1 A_{n-1} = \\ & = \frac{1}{\sqrt{5}} \left[\gamma^n (\gamma\eta_0 + \eta_1) - \frac{(-1)^n}{\gamma^n} \left(\eta_1 - \frac{\eta_0}{\gamma} \right) \right]. \end{aligned}$$

Thus if two tiling systems of this kind are produced with tile lengths η_0, η_1 for one and η'_0, η'_1 for the other, and

$$(5.19) \quad \gamma\eta_0 + \eta_1 = \gamma\eta'_0 + \eta'_1,$$

then corresponding basic tiles in the two systems have asymptotic lengths, with the difference in length of corresponding rank n basic tiles being on the order of $1/\gamma^n$.

Now we define the factor map $\phi : ((X, T)_g, \mathbb{R}) \rightarrow ((Y, S)_h, \mathbb{R})$. Fix a tiling $\mathcal{T} = [x, s] \in ((X, T)_g, \mathbb{R})$. To define $\phi\mathcal{T}$, we let \mathcal{S} consist of a sequence of tiles in $((Y, S)_h, \mathbb{R})$ with the same 0, 1 labels as \mathcal{T} , and we are left only to position the origin in the central tile. That is, we will put $\phi[x, s] = [\sigma^k x, t]$, with k and t to be determined. This is accomplished by writing a sequence of tilings $\mathcal{S}_0, \mathcal{S}_1, \dots$, all of which have the same 0, 1 label sequence up to shifting, which converges to $\phi\mathcal{T}$ if the lengths of the tiles harmonize correctly.

For each $n = 0, 1, \dots$, \mathcal{S}_n is defined as follows. If the origin is contained in the interior of a rank n basic patch T_n of \mathcal{T} , we choose for \mathcal{S}_n any translation of \mathcal{S} which has a basic patch with the same (0, 1 block) label as T_n with its center located at the same position as the center of T_n . If the origin is the boundary between two rank n patches in an elementary patch $T_n^l T_n^r$ of \mathcal{T} , we choose for \mathcal{S}_n any translation of \mathcal{S} which has the center of the corresponding elementary patch of \mathcal{S} (with the same label as $T_n^l T_n^r$) at the same position as the center of $T_n^l T_n^r$ in \mathcal{T} .

When passing from \mathcal{S}_n to \mathcal{S}_{n+1} , the position of the origin in the basic rank n patch (or elementary patch) corresponding to T_n (or $T_n^l T_n^r$) is translated by no more than half the difference in the lengths of the corresponding basic rank $n+1$ patches in the two systems, which decreases exponentially because of (5.18) and (5.17). Moreover, the tile labels of \mathcal{S}_n and \mathcal{S}_{n+1} agree on a long central block, and therefore the sequence of tilings \mathcal{S}_n converges to a tiling $\phi\mathcal{T}$.

It is clear that ϕ is a one-to-one, onto map between the two tiling systems which commutes with the translation actions. Continuity follows from the observation that if two tilings \mathcal{T} and \mathcal{T}' have the same basic central blocks of tile labels of each rank $r \leq n$ for a large n , then their images have small central patches that are the same up to a translation of size bounded by the sum of a geometric series from the n 'th term on.

It is also not hard to see that if $\eta_1 \neq \eta_0/\gamma$, then ϕ cannot be a local code. For in such a case, the lengths of corresponding basic rank n tiles in the two systems are not equal, so at each stage n we apply to \mathcal{S}_n a nonzero translation, of size approximately $1/\gamma^n$, to obtain \mathcal{S}_{n+1} . Given any diameter R of central patches which is presumed to be a window size for a local code, we can find tilings \mathcal{T} and \mathcal{T}' which agree exactly on a central patch of size larger than R (including the precise location of the origin in the interior of the common patch) but for which $\phi\mathcal{T} \neq \phi\mathcal{T}'$ as follows. Specify \mathcal{T} by writing a 0, 1 sequence $H \in \Sigma_2^+$ of types of basic central patches which is not eventually constant. Position the tiling \mathcal{T} by putting the origin in the center of the central basic tile of rank 0. Specify \mathcal{T}' similarly by writing for $H(\mathcal{T}')$ a sequence H' which agrees with H on a long initial block, is not eventually constant, disagrees with H infinitely many times, but does not have two disagreements with H within 10 places of each other. (The last condition is included because it is possible for two polynomials in $\alpha = 1/\gamma$ with different sequences of coefficients ± 1 to be equal.) Since H and H' agree on a long initial block, \mathcal{T} and \mathcal{T}' agree on a large central patch. But because H and H' differ at an infinite, sparse set of places, lining up the centers of the corresponding basic tile patches is guaranteed to produce a limiting nonzero translation in the position of the origin in the central patches of radius R in $\phi\mathcal{T}$ and $\phi\mathcal{T}'$.

Question 5.11. What are the complete automorphism groups of these 1-dimensional Sturmian substitution tilings? Could cellular automata help in constructing automorphisms [14, 40, 44, 68, 71]? What is the interplay between the action of translation (a sort of adic), the action of the substitution (analogous to the shift), and the action of selected automorphisms?

REFERENCES

1. J. Aaronson, *An Introduction to Infinite Ergodic Theory*, Math. Surveys and Monographs, vol. 50, AMS, 1997.
2. T. Adams, *Lightly mixing on dense collections*, preprint, 1997.
3. T. M. Adams and K. E. Petersen, *Binomial-coefficient multiples of irrationals*, Mh. Math. **125** (1998), 269–278.
4. D. J. Aldous, *Exchangeability and related topics*, École d’Été de Probabilités de Saint-Flour XIII–1983, Lecture Notes in Math., vol. 1117, Springer-Verlag, 1985, pp. 2–198.
5. J. C. Alexander and D. Zagier, *The entropy of a certain infinitely convoluted Bernoulli measure*, J. London Math. Soc. **44** (1991), 121–134.
6. D. Behrend, M. Boshernitzan, and G. Kolesnik, *Irrational dilations of Pascal’s triangle*, preprint.
7. H. C. P. Berbee and W. Th. F. den Hollander, *Tail triviality for sums of stationary random variables*, Ann. Prob. **17** (1989), 1635–1645.
8. A. Bertrand-Mathis, *Le θ -shift sans peine*, preprint.
9. ———, *Développement en base θ ; répartition modulo un de la suite $(x\theta^n)_{n \geq 0}$; langages codés et θ -shift*, Bull. Soc. Math. France **114** (1986), 271–323.
10. B. H. Bissinger, *A generalization of continued fractions*, Bull. Amer. Math. Soc. **50** (1944), 868–876.
11. D. Blackwell, *The entropy of functions of finite state Markov chains*, Transactions First Prague Conference, 1957, pp. 13–20.
12. D. Blackwell and D. Freedman, *The tail σ -field of a Markov chain and a theorem of Orey*, Ann. Math. Stat. **35** (1964), 1291–1295.
13. F. Blanchard, *β -expansions and symbolic dynamics*, Theoretical Comp. Sci. **65** (1989), 131–141.
14. F. Blanchard and A. Maass, *Dynamical properties of expansive one-sided cellular automata*, Israel J. Math. **99** (1997), 149–174.
15. T. Bogenschütz and V. M. Gundlach, *Ruelle’s transfer operator for random subshifts of finite type*, Erg. Theory Dyn. Sys. **15** (1995), 413–447.
16. R. Bowen, *Periodic points and measures for Axiom A diffeomorphisms*, Trans. Amer. Math. Soc. **154** (1971), 377–397.
17. ———, *Some systems with unique equilibrium states*, Math. Systems Theory **8** (1974), 193–202.
18. M. Boyle and D. Handelman, *Entropy versus orbit equivalence for minimal homeomorphisms*, Pacific J. Math. **164** (1994), 1–13.
19. M. Boyle and S. Tuncel, *Infinite-to-one codes and Markov measures*, Trans. Amer. Math. Soc. **285** (1984), 657–683.
20. A. Broise, *Transformations dilatantes de l’intervalle et théorèmes limites*, Études Spectrales d’Opérateurs de Transfert et Applications (A. Broise, F. Dal’bo, and M. Peigné, eds.), Astérisque, no. 238, Soc. Math. de France, 1996, pp. 1–109.
21. Y. Bugeaud and J.-P. Conze, *Calcul de la dynamique de transformations linéaires contractantes mod 1 et arbre de Farey*, Acta Arith. **88** (1999), 201–218.
22. C. J. Burke and M. Rosenblatt, *A Markovian function of a Markov chain*, Ann. Math. Stat. **29** (1958), 1112–1122.

23. M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes in Math., no. 527, Springer-Verlag, Berlin, Heidelberg, New York, 1976.
24. P. Diaconis and D. Freedman, *Partial exchangeability and sufficiency*, Statistics: Applications and New Directions (J. K. Ghosh and J. Roy, eds.), Indian Statistical Institute, Calcutta, 1984, pp. 205–236.
25. F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math. **179** (1998), 169–185.
26. F. Durand, B. Host, and C. Skau, *Substitutional dynamical systems, Bratteli diagrams and dimension groups*, Erg. Th. Dyn. Sys. **19** (1999), 953–993.
27. P. Erdős, *On a family of symmetric Bernoulli convolutions*, Amer. J. Math. **61** (1939), 974–976.
28. ———, *On the smoothness properties of a family of Bernoulli convolutions*, Amer. J. Math. **62** (1940), 180–186.
29. J. Feldman and C. C. Moore, *Ergodic equivalence relations, cohomology, and von Neumann algebras. I*, Trans. Amer. Math. Soc. (1977), 289–324.
30. J. Feldman, C. Sutherland, and C. Moore, *Subrelations of ergodic equivalence relations*, Erg. Th. Dyn. Sys. **9** (1989), 289–324.
31. W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, J. Wiley and Sons, 1966.
32. A. H. Forrest, *K-groups associated with substitution minimal systems*, Israel J. Math. **98** (1997), 101–139.
33. H. Furstenberg, *Strict ergodicity and transformations of the torus*, Amer. J. Math. **83** (1961), 573–601.
34. A. Garsia, *Arithmetic properties of Bernoulli convolutions*, Trans. Amer. Math. Soc. **102** (1962), 409–432.
35. ———, *Entropy and singularity of infinite convolutions*, Pac. J. Math. **13** (1963), 1159–1169.
36. D. Gatzouras and Y. Peres, *The variational principle for Hausdorff dimension: A survey*, Ergodic Theory of \mathbb{Z}^d -Actions (M. Pollicott and K. Schmidt, eds.), LMS Lecture Notes, vol. 228, Cambridge Univ. Press, 1996, pp. 533–568.
37. H.-O. Georgii, *Canonical Gibbs states, their relation to Gibbs states, and applications to two-valued Markov chains*, Z. Wahr. v. Gb. **32** (1975), 277–300.
38. T. Giordano, I. Putnam, and C. Skau, *Topological orbit equivalence and C^* -crossed products*, J. reine angew. Math. **469** (1995), 51–111.
39. E. Glasner and B. Weiss, *Weak orbit equivalence of Cantor dynamical systems*, Int. J. Math. **6** (1995), 559–579.
40. E. Goles and S. Martinez, *Cellular Automata and Complex Systems*, Kluwer, 1999.
41. F. J. Hahn, *On affine transformations of compact abelian groups*, Amer. J. Math. **85** (1963), 428–446.
42. A. Hajian, Y. Ito, and S. Kakutani, *Invariant measures and orbits of dissipative transformations*, Adv. in Math. **9** (1972), 52–65.
43. R. H. Herman, I. F. Putnam, and C. F. Skau, *Ordered Bratteli diagrams, dimension groups and topological dynamics*, Int. J. Math. **3** (1992), 827–864.
44. G. Hernández and L. G. Torres, *Stochastic cellular automata*, Lect. Mat. **15** (1994), 167–191.

45. J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley, 1979.
46. B. Host, *Substitution subshifts and Bratteli diagrams*, Topics in Symbolic Dynamics and Applications (F. Blanchard, A. Maass, and A. Nogueira, eds.), LMS Lecture Notes, Cambridge Univ. Press, 1999.
47. Robert B. Israel, *Convexity in the Theory of Lattice Gases*, Princeton University Press, 1979.
48. S. Ito, *Construction of transversal flows for maximal Markov automorphisms*, Tokyo J. Math. **1** (1978), 305–324.
49. B. Jessen and A. Wintner, *Distribution functions and the Riemann zeta function*, Proc. Amer. Math. Soc. **30** (1935), 48–88.
50. K. Johnson, *Beta-shift Dynamical Systems and Their Associated Languages*, Ph.D. thesis, University of North Carolina at Chapel Hill, 1999.
51. ———, *Context-free implies regular for languages of unimodal maps*, preprint, 2000.
52. S. Kakutani, *A problem of equidistribution on the unit interval $[0, 1]$* , Proceedings of the Oberwolfach Conference on Measure Theory (1975), Lecture Notes in Math., vol. 541, Springer-Verlag, 1976, pp. 369–376.
53. A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.
54. G. Keller, *Equilibrium States in Ergodic Theory*, Cambridge Univ. Press, 1998.
55. S. V. Kerov and A. M. Vershik, *The characters of the infinite symmetric group and probability properties of the Robinson-Schensted-Knuth algorithm*, SIAM J. Alg. Disc. Meth. **7** (1986), 116–124.
56. H. B. Keynes, N. G. Markley, and M. Sears, *Automorphisms of suspension flows over the circle*, Dynamical Systems. Proceedings, University of Maryland, 1986–87 (J. C. Alexander, ed.), Lecture Notes in Math., vol. 1342, Springer-Verlag, 1988, pp. 342–353.
57. ———, *The structure of automorphisms of real suspension flows*, Ergod. Theory Dyn. Sys. **11** (1991), 349–364.
58. M. Kowada, *Spectral type of one-parameter group of unitary operators with transversal group*, Nagoya Math. J. **32** (1968), 141–153.
59. ———, *The orbit-preserving transformation groups associated with a measurable flow*, J. Math. Soc. Japan **24** (1972), 355–373.
60. I. Kubo, *Quasi-flows*, Nagoya Math. J. **35** (1969), 1–30.
61. E. E. Kummer, *Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen*, J. für Math. **44** (1852), 115–116.
62. F. Ledrappier and P. Walters, *A relativised variational principle for continuous transformations*, J. London Math. Soc. **16** (1977), 568–576.
63. F. Ledrappier and L.-S. Young, *The metric entropy of diffeomorphisms II*, Ann. Math. **122** (1985), 540–574.
64. D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, 1995.
65. A. N. Livshitz, *A sufficient condition for weak mixing of substitutions and stationary adic transformations*, Math. Notes **44** (1988), 920–925.
66. A.A. Lodkin and A.M. Vershik, *Approximation for actions of amenable groups and transversal automorphisms*, Operator Algebras and their Connections with Topology and Ergodic Theory (H. Araki, C.C. Moore, S. Stratila, and

- D. Voiculescu, eds.), Lecture Notes in Math., vol. 1132, Springer-Verlag, New York, 1985, pp. 331–346.
67. E. Lucas, *Théorie des fonctions numériques simplement périodiques*, Amer. J. Math. **1** (1878), 184–240.
 68. A. Maass, *On the sofic limit sets of cellular automata*, Erg. Theory Dyn. Sys. **15** (1995), 663–684.
 69. A. Manning and H. McCluskey, *Hausdorff dimension for horseshoes*, Erg. Th. Dyn. Sys. **3** (1983), 251–260.
 70. B. Marcus, K. Petersen, and S. Williams, *Transmission rates and factors of Markov chains*, Contemp. Math. **26** (1984), 279–293.
 71. J. L. Marroquín and A. Ramírez, *Stochastic cellular automata with Gibbsian invariant measures*, IEEE Trans. Inform. Theory **37** (1991), 541–551.
 72. R. D. Mauldin and K. Simon, *The equivalence of some Bernoulli convolutions to Lebesgue measure*, preprint, 1997.
 73. B. Mossé, *Puissances de mots et reconnaissabilité des points fixes d'une substitution*, Theor. Comp. Sci. **99** (1992), 327–334.
 74. ———, *Reconnaissabilité des substitutions et complexité des suites automatiques*, Bull. Soc. Math. France **124** (1996), 101–118.
 75. N. S. Ormes, *Strong orbit realization for minimal homeomorphisms*, J. Anal. Math. **71** (1997), 103–133.
 76. Donald S. Ornstein and Benjamin Weiss, *Every transformation is bilaterally deterministic*, Israel J. Math. **21** (1975), 154–158.
 77. W. Parry, *On the β -expansion of real numbers*, Acta Math. Sci. Acad. Hungar. **11** (1960), 401–416.
 78. ———, *On Rohlin's formula for entropy*, Acta Math. Sci. Acad. Hungar. **15** (1964), 107–113.
 79. ———, *Representations for real numbers*, Acta Math. Sci. Acad. Hungar. **15** (1964), 95–105.
 80. ———, *Symbolic dynamics and transformations of the unit interval*, Trans. Amer. Math. Soc. **122** (1966), 368–378.
 81. W. Parry and S. Tuncel, *Classification Problems in Ergodic Theory*, LMS Lecture Note Series, vol. 67, Cambridge University Press, Cambridge, 1982.
 82. Y. Peres, W. Schlag, and B. Solomyak, *Sixty years of Bernoulli convolutions*, Fractals and Stochastics II (C. Bandt, S. Graf, and M. Zaehle, eds.), Progress in Probability, vol. 46, Birkhäuser, 2000, pp. 39–65.
 83. Y. Peres and B. Solomyak, *Absolute continuity of Bernoulli convolutions, a simple proof*, Math. Res. Letters **3** (1996), 231–239.
 84. K. Petersen, *Factor maps between tiling dynamical systems*, Forum Math. **11** (1999), 503–512.
 85. K. Petersen and S. Shin, *Relatively maximal measures*, preprint, 2000.
 86. K. Petersen and J.-P. Thouvenot, *Tail fields generated by symbol counts in measure-preserving systems*, preprint, 2000.
 87. K. E. Petersen, *Ergodic Theory*, Cambridge University Press, 1983, 1989.
 88. ———, *Attractors and Attracting Measures*, lecture notes, 1997, <http://www.math.unc.edu/Faculty/petersen>.
 89. ———, *Symbolic Dynamics*, lecture notes, 1998, <http://www.math.unc.edu/Faculty/petersen>.

90. Karl Petersen and Klaus Schmidt, *Symmetric Gibbs measures*, Trans. Amer. Math. Soc. **349** (1997), 2775–2811.
91. A. G. Postnikov, *Ergodic problems in the theory of congruences and of Diophantine approximations*, Proc. Steklov Inst. Math. **82** (1966), 3–112.
92. N. Priebe, *Detecting Hierarchy in Tiling Dynamical Systems via Derived Voronoi Tessellations*, Ph.D. thesis, Univ. of N. Carolina at Chapel Hill, 1997.
93. N. Priebe, *Towards a characterization of self-similar tilings in terms of derived voronoi tessellations*, Geometriae Dedicata **79** (2000), 239–265.
94. N. Priebe and B. Solomyak, *Characterization of planar pseudo-self-similar tilings*, to appear in J. Disc. Comp. Geom., 2000.
95. C. Radin and L. Sadun, *Isomorphisms of hierarchical structures*, preprint.
96. C. Radin and M. Wolff, *Space tilings and local isomorphism*, Geom. Ded. **42** (1992), 355–360.
97. A. Rényi, *Representations for real numbers and their ergodic properties*, Acta Math. Sci. Acad. Hungar. **8** (1957), 401–414.
98. D. Ruelle, *Statistical mechanics on a compact set with F action satisfying expansiveness and specification*, Trans. Amer. Math. Soc. **185** (1973), 237–251.
99. J. Schmeling, *Symbolic dynamics for β -shifts and self-normal numbers*, Erg. Th. Dyn. Sys. **17** (1997), 675–694.
100. K. Schmidt, *Algebraic Ideas in Ergodic Theory*, CBMS Reg. Conf. Series. Math., vol. 76, AMS, Providence, R.I., 1990.
101. ———, *Invariant cocycles, random tilings, and the super- K and strong Markov properties*, Trans. Amer. Math. Soc. **349** (1997), 2812–2825.
102. Klaus Schmidt, *Tail-fields of products of random variables and ergodic equivalence relations*, Erg. Th. Dyn. Sys. **19** (1999), 1325–1341.
103. E. Seneta, *Non-negative Matrices and Markov Chains*, Springer-Verlag, New York, Heidelberg, Berlin, 1973, 1981.
104. K. Shelton, *Ergodic Properties of a Class of Erdős Measures*, Ph.D. thesis, University of North Carolina at Chapel Hill, 1998.
105. S. Shin, *Measures that maximize weighted entropy for factor maps between subshifts of finite type*, Ph.D. thesis, Univ. of N. Carolina, 1999.
106. N. Sidorov and A. Vershik, *Ergodic properties of the Erdős measure, the entropy of the goldenshift, and related problems*, Monatsh. f. Math. **126** (1998), 215–261.
107. Ja. G. Sinai, *Probabilistic ideas in ergodic theory*, Amer. Math. Soc. Transl. Ser. 2 **31** (1963), 62–84.
108. B. Solomyak, *On the spectral theory of adic transformations*, Adv. Soviet Math. **8** (1992), 217–230.
109. ———, *On the random series $\sum \pm \lambda^i$ (an Erdős problem)*, Ann. of Math. **142** (1995), 611–625.
110. B. Solomyak, *Dynamics of self-similar tilings*, Ergod. Th. Dyn. Sys. **17** (1997), 695–738.
111. A. M. Vershik, *Description of invariant measures for actions of some infinite groups*, Dokl. Akad. Nauk SSSR **28** (1974), 749–752.
112. A. M. Vershik, *A theorem on the Markov periodic approximation in ergodic theory*, J. Soviet Math. **28** (1985), 667–673.

113. A. M. Vershik and S. V. Kerov, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables*, Soviet Math. Dokl. **18** (1977), 527–531.
114. ———, *Asymptotic theory of characters of the symmetric group*, Func. An. Appl. **15** (1981), 246–255.
115. ———, *Asymptotic of the largest and typical dimensions of irreducible representations of a symmetric group*, Func. An. Appl. **19** (1985), 21–31.
116. A. M. Vershik and A. N. Livshitz, *Adic models of ergodic transformations, spectral theory, substitutions, and related topics*, Adv. Soviet Math. **9** (1992), 185–204.
117. P. Walters, *Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts*, Trans. Amer. Math. Soc. **296** (1986), 1–31.
118. Y. Wang, L. Yang, and H. Xie, *Complexity of unimodal maps with aperiodic kneading sequences*, Nonlinearity **12** (1999), 1151–1176.
119. H. Weyl, *Über die Gleichverteilung von Zahlen mod. Eins*, Math. Ann. **77** (1916), 313–352.
120. K. M. Wilkinson, *Ergodic properties of certain linear mod one transformations*, Adv. in Math. **14** (1974), 64–72.
121. A. Wintner, *On convergent Poisson convolutions*, Amer. J. Math. **57** (1935), 827–838.

DEPARTMENT OF MATHEMATICS, CB 3250, PHILLIPS HALL, UNIVERSITY OF
NORTH CAROLINA, CHAPEL HILL, NC 27599 USA

E-mail address: petersen@math.unc.edu