Physica 10D (1984) 36–44
North-Holland, Amsterdam

APPLICATIONS OF ERGODIC THEORY AND SOFIC SYSTEMS TO CELLULAR AUTOMATA

D.A. LIND

Department of Mathematics, University of Washington, Seattle, Washington 98195, USA

We indicate a mathematical framework for analysing the evolution of cellular automata. Those automata obeying an additive rule are shown to be the same as endomorphisms of a compact abelian group, and therefore their statistical and dynamical behavior can be told exactly by using Fourier analysis and ergodic theory. Those obeying certain nonlinear rules are closely tied to a finitely-described object called a sofic system, but the underlying statistics appear to be more complicated and interesting. We conclude by formulating several conjectures about one such system.

1. Introduction

Cellular automata were originally introduced by von Neumann and Ulam [1] as potential idealized mathematical models of biological systems capable of self-reproduction. Since then they have been reintroduced and studied in many contexts, ranging from parallel processing in computers to the growth and evolution of crystals and organisms. A detailed description of cellular automata, together with many references to their applications, has been recently given by Wolfram [2].

This paper has two purposes. The first is to show how certain patterns and regularities observed in the evolution of cellular automata can be put into a mathematical framework capable of useful analysis. Using the basic tools of ergodic theory and Fourier analysis on compact groups, some of the empirical observations in [2] can be rigorously formulated and proved. Other quantitative statements there can be given an exact form.

The second purpose is to point out that some of the numerical observations lead to precise but apparently difficult mathematical problems. The computer simulations of Grassberger [3] on the behavior of one type of cellular automaton lead to likely looking conjectures in dynamical systems. However, even a crude model of the necessary theorem on random walks eluded Erdös and Ney [4], only to be settled later by Adelman [5].

Although treating cellular automata as dynamical systems is hardly new (see [6] for example), there are some novel features here. Additive rules are the same as endomorphisms of a compact abelian group, and the latter have been thoroughly investigated [7]. Virtually complete results are available for them. Also, the natural stage on which certain nonlinear cellular automata evolve is an object introduced by Weiss [8] in 1973 called a sofic system. Such systems can be described in terms of a finite-state automaton. Their theory leads to useful information on the statistics of nonlinear cellular automaton evolution.

2. Mathematical framework

We will deal with 1-dimensional cellular automata. Extensions to higher dimensions should be clear.

Let S be a finite set of states. A *configuration* on the lattice $\mathbb{Z}$ of integers is an assignment of an element of S to each site $i \in \mathbb{Z}$. Thus a configuration $x = \{x_i\}$ is a doubly-infinite sequence of elements $x_i \in S$. The set X of all possible configurations is denoted by $S^{\mathbb{Z}}$. For simplicity, we take $S = \{0, 1\}$ unless otherwise stated.

The topological glue holding the points of X together is determined by the metric

$$d(x, y) = \sum_{i=-\infty}^{\infty} 2^{-|i|}|x_i - y_i| .$$

Thus x and y are close when their components x_i and y_i agree for $|i| < N$ for a large N. The space X equipped with metric d is a Cantor set.

For $x \in X$, define $(\sigma x)_i = x_{i+1}$. Thus σ shifts the entries in a configuration one unit to the left. This shift is a continuous and one-to-one mapping of X onto itself, i.e. σ is a homeomorphism of X.

A cellular automaton rule gives the discrete time evolution of a configuration in terms of a local interaction of the site values. In other words, a configuration x evolves in one time step to a new configuration τx defined as follows. Fix a neighborhood size k and a function $\tau_0 : S^{2k+1} \to S$. Then define

$$(\tau x)_i = \tau_0(x_{i-k}, \dots, x_{i+k}) . \qquad (2.1)$$

One example of such a rule, using nearest neighbors only, is the additive rule

$$(\tau x)_i = x_{i-1} + x_{i+1} \pmod 2 . \qquad (2.2)$$

We shall study this rule in section 4.

It is clear that if x is close to y, then τx will be close to τy. This means that $\tau : X \to X$ is a continuous map. Also, spacial homogeneity of τ means that τ commutes with σ, i.e. $\sigma\tau = \tau\sigma$.

Conversely, using uniform continuity it is easy to show that if $\tau : X \to X$ is a continuous map commuting with σ, then τ must be given by a local rule of the form (2.1) [9]. Therefore cellular automata rules coincide with shift-commuting continuous maps of X.

Such continuous mappings have been studied before [9]. However, there is such a wealth of phenomena that simple or comprehensive results are rare.

3. Ergodic theory, entropy, and disorder

Before studying two cellular automata in detail, we make some general remarks.

The appropriate mathematical description of an initial distribution of configurations is a probability measure μ on X. We should also require μ to be spacially homogeneous, so $\mu(\sigma E) = \mu(E)$ for all measurable subsets E of X.

To describe such a measure, let $B = [b_0 b_1 \dots b_{k-1}]$, $b_j \in S$, $k \geq 1$, and put $C(B) = \{x \in X : x_0 = b_0, \dots, x_{k-1} = b_{k-1}\}$. Thus $C(B)$ specifies the set of configurations with prescribed values at a finite number of sites. Such a set is called a cylinder set. Clearly μ must satisfy the consistency relation

$$\mu(C[b_0, \dots, b_k]) = \sum_{j \in S} \mu(C[b_0, \dots, b_k, j]) . \qquad (3.1)$$

The Kolmogorov consistency theorem [10] shows that any assignment of measures to cylinders satisfying (3.1) will extend to a shift invariant measure on X.

For example, let $0 < p < 1$. If B has j 1's and $k - j$ 0's, define $\mu_p(C(B)) = p^j(1-p)^{k-j}$. This distribution μ_p specifies a random initial set of configurations with density p of 1's.

Since μ is assumed preserved under σ, we can apply the tools of ergodic theory. Say that μ is *ergodic* for σ if whenever $\sigma E = E$, then $\mu(E) = 0$ or 1. For example, each μ_p is ergodic for σ.

If μ is ergodic for σ, the ergodic theorem implies that for every continuous function $f : X \to \mathbb{R}$,

$$\frac{1}{2N+1} \sum_{i=-N}^{N} f(\sigma^i x) \to \int_X f(x)\, d\mu(x), \quad \text{as } N \to \infty ,$$

for μ-almost every x in X. As an example, to compute the frequency of 1's in x, let $f_1(x) = x_0$. Then the ergodic theorem implies that for μ-almost every x, the frequency of 1's in x exists and equals

$$\lim_{N \to \infty} \frac{1}{2N+1} \sum_{i=-N}^{N} f_1(\sigma^i x) = \int_X f_1(x)\, d\mu(x) .$$

Applying this to μ_p shows that μ_p-almost every point in X has frequency p of 1's. This restates the strong law of large numbers for Bernoulli trials.

Since τ is continuous, its image τX in X is compact. A σ-invariant measure μ is mapped to a measure $\tau\mu$ on X via the definition $(\tau\mu)(E) = \mu(\tau^{-1}E)$. Since $\sigma\tau = \tau\sigma$, $\tau\mu$ is σ-invariant as well, and ergodic for σ if μ was. The frequency of 1's in the first evolution τx of a typical point x is

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} f_1(\sigma^i \tau x)\,.$$

Since $f_1(\sigma^i\tau x) = (f_1\tau)(\sigma^i x)$, the ergodic theorem implies this limit equals

$$\int_X (f_1\tau)\,\mathrm{d}\mu = \int_X f_1\,\mathrm{d}(\tau\mu)\,, \tag{3.2}$$

for almost every x.

Similar results hold for further iterates τ^n. In fact, all statistical quantities, such as the frequency in $\tau^n x$ of the block [0110100], can be exactly computed once the image measure $\tau^n\mu$ is known.

Limiting behavior of the evolution as $n\to\infty$ can also be formulated mathematically. Say that a sequence of measures μ_n converges to μ (symbolically $\mu_n\to\mu$) if for every continuous real-valued function f on X, $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$. In probability theory this is called "weak convergence." The limiting behavior of a cellular automaton with given initial distribution μ of configurations is the study of the behavior of the measures $\tau^n\mu$ on $\tau^n X$ as $n\to\infty$. For example, we shall prove below that if τ is the additive rule (2.2) and $0 < p < 1$, then $N^{-1}\sum_{n=0}^{N-1} \tau^n\mu_p \to \mu_{1/2}$. This implies that regardless of the density of 1's in an initial random configuration, the limiting frequency of 1's is $\frac{1}{2}$. One consequence is a proof of the approximate result (3.7) of [2] concerning the density of triangles for additive rules (see (4.1) below).

Entropy measures disorder. In the context of Cantor sets, topological entropy of the shift σ on a σ-invariant compact subset Y of X can be defined as follows. Let $f\colon Y\to\{1,\ldots,k\}$ be continuous, and define $N_n(f,\sigma)$ to be the number of distinct sequences $\{f(y), f(\sigma y),\ldots, f(\sigma^{n-1}y)\}$ for $y\in Y$. Clearly $N_n(f,\sigma)\leqslant k^n$. Define $H(f,\sigma) = \limsup_{n\to\infty} (1/n)\log N_n(f,\sigma)$, where log is the natural logarithm. Finally, define the entropy of σ on Y to be $h(\sigma, Y) = \sup_f H(f,\sigma)$, where the supremum is taken over all continuous functions f taking a finite number of values. Thus $h(\sigma, Y)$ is the exponential rate of growth of the number of approximate orbits in Y of a given length.

Since $\tau X\subset X$, the images $\tau^n X$ are σ-invariant compact sets that decrease to a compact set $\tau^\infty X = \bigcap_{n=1}^{\infty}\tau^n X$. Thus $h(\sigma,\tau^n X)$ decreases, and the difference $h(\sigma, X) - h(\sigma,\tau^n X)$ measures the "increase in order" introduced by the cellular automaton rule. This is a mathematical description of the "self-organization" observed in [2].

One quantitative consequence of these observations is the following. Let $P_n(\tau)$ denote the number of blocks of length n with predecessors under τ. If $\tau X = X$, then clearly $P_n(\tau) = 2^n$ for all n. However, if $\tau X \neq X$, one can show [11] that $h(\sigma,\tau X) < h(\sigma, X) = \log 2$. This means in particular that $P_n(\tau)$ increases like $\exp[nh(\sigma,\tau X)]$, so $P_n(\tau)/2^n \to 0$ exponentially fast.

4. Exact results for additive cellular automata

Let τ denote the additive rule (2.2). Using the ergodic theorem for commuting transformations together with Fourier analysis, we show that the statistical regularities observed in [2] can be proved in an exact form. Indeed, all statistical quantities can be explicitly computed this way.

Since each coordinate of X is an element of the finite group $\mathbb{Z}_2 = \{0,1\}$, X itself is a compact group. Additivity of τ means that τ is a group homomorphism. Furthermore, τ is a four-to-one homomorphism of X onto itself. Therefore [12] it preserves Haar measure $\mu_{\mathrm{H}} = \mu_{1/2}$. Throughout this section, "almost everywhere" refers to μ_{H} unless otherwise indicated.

The character group G of X is the countable abelian group whose elements have the form $g = \{g_i\}$, $g_i\in\mathbb{Z}_2$ with all but a finite number of g_i equal to 0. The value of a character g on a group element x is computed by

$$g(x) = \prod_{i=-\infty}^{\infty} (1-2x_i)^{g_i}\,,$$

where there are only a finite number of terms in the product different from 1. The dual homomorphism of G induced by τ, which we also call τ, is given by $(\tau g)_i = g_{i-1} + g_{i+1}$.

For $g \neq 0$, $\{\tau^n g\}$ is infinite, implying τ is ergodic for μ_H [12]. Thus σ and τ induces a jointly ergodic action of $\mathbb{Z} \times \mathbb{N}$, given by the correspondence $(i, n) \leftrightarrow \sigma^i \tau^n$. Under this correspondence, an initial configuration $x \in X$ generates a pattern $f_1(\sigma^i \tau^n x)$ of 0's and 1's indexed by $\mathbb{Z} \times \mathbb{N}$.

Let E be a finite subset of $\mathbb{Z} \times \mathbb{N}$, and $P : E \to \mathbb{Z}_2$ assign a pattern of 0's and 1's to E. Put $f_P(x) = 1$ if $f_1(\sigma^i \tau^n x) = P(i, n)$ for all $(i, n) \in E$, and 0 otherwise.

Define the frequency of the pattern P in the τ-evolution of x to be

$$v_P(x) = \lim_{N\to\infty} \frac{1}{(2N+1)(N+1)} \sum_{i=-N}^{N} \sum_{n=0}^{N} f_P(\sigma^i \tau^n x),$$

provided the limit exists. An application of the ergodic theory for commuting measure-preserving transformation [13, VIII. 6.9] implies that $v_P(x)$ exists for almost every x and can be computed from f_P.

Theorem 1. For almost every $x \in X$ the frequency $v_P(x)$ of the pattern P in the τ-evolution of x exists and has the value

$$v_P(x) = \int_X f_P \, d\mu_H .$$

Thus the statistical patterns in the evolution of a configuration can be exactly computed. In particular, let $Q(k)$ denote the frequency of the block $[10^k 1]$ in the evolution of x. By Theorem 1, $Q(k) = (\frac{1}{4})2^{-k}$, a result observed in approximate form in [2].

As another application of theorem 1, let P denote the pattern of a triangle of 0's with base length k. Then P is described as the occurrence of $[10^k 1]$ in τx together with at least one 1 occurring in x above a 0. If f_P is the corresponding function, it is easy to compute that the density $T(k)$ of triangles of base k has the value

$$T(k) = v_P(x) = \int f_P \, d\mu_H = \left(\frac{3}{16}\right) 2^{-k}, \tag{4.1}$$

again observed in approximate form in [2].

The evolution of an initial distribution μ of configurations is given by the sequence of measures $\tau^n \mu$. In particular, the study of $\tau^n \mu_p$ reflects the evolution from a disordered state with density p of 1's. We show that although $\tau^n \mu_p$ does not itself converge if $p \neq \frac{1}{2}$, the averages converge to μ_H.

We shall use the machinery of Fourier analysis of the compact group X. The Fourier transform of a measure μ on X is given by

$$\hat{\mu}(g) = \int_X g(x) \, d\mu(x) \qquad (g \in G).$$

It is easy to show that $\mu_n \to \mu$ exactly when $\hat{\mu}_n(g) \to \hat{\mu}(g)$ for every $g \in G$. Since $(\tau^n \mu)\hat{}\ (g) = \hat{\mu}(\tau^n g)$, we have that

$$\tau^n \mu \to \mu_\infty \quad \text{if and only if} \quad \hat{\mu}(\tau^n g) \to \hat{\mu}_\infty(g)$$

$$\text{for all } g \in G. \tag{4.2}$$

Thus the behavior of $\tau^n \mu$ depends on the behavior of $\hat{\mu}$, which can be computed in particular cases. We give two examples.

Theorem 2. (i) If $p \neq \frac{1}{2}$, then $\tau^n \mu_p$ does not converge.

(ii) For $0 < p < 1$,

$$\frac{1}{N} \sum_{n=1}^{N} \tau^n \mu_p \to \mu_H, \quad \text{as } N \to \infty .$$

Proof. We first calculate $\hat{\mu}_p$. For $g = \{g_i\} \in G$, let $S(g) = \{i : g_i = 1\}$ be the support of g, and put $r(g) = |S(g)|$, where $|E|$ denotes the cardinality of a set E. Then

$$\hat{\mu}_p(g) = \int_X g(x)\,\mathrm{d}\mu_p(x)$$
$$= \sum_{\substack{b_i = 0,1 \\ i \in S(g)}} \left(\prod_{i \in S(g)} (1 - 2b_i) \right) \mu_p\{x: x_i = b_i\}$$
$$= \sum_{k=0}^{r(g)} (-1)^k \binom{r(g)}{k} p^k (1-p)^{r(g)-k}$$
$$= (1-2p)^{r(g)}. \tag{4.3}$$

Let $g(i)$ be the element of G with a 1 in the ith coordinate and 0's elsewhere. Then $\tau 2^n g(i) = g(i - 2^n) + g(i + 2^n)$, so $r(\tau 2^n g(0)) = 2$ and $r(\tau^{2^n+2} g(0)) = 4$ for every n. Thus by (4.3)

$$\hat{\mu}_p(\tau^{2^n} g(0)) = (1-2p)^2,$$

while

$$\hat{\mu}_p(\tau^{2^n+2} g(0)) = (1-2p)^4.$$

This shows that if $p \neq \frac{1}{2}$, then $\hat{\mu}_p(\tau^n g(0))$ does not converge as $n \to \infty$. Using (4.2) we obtain (i).

To show (ii), we begin with the following fact, which can be easily deduced from properties of the binomial coefficients modulo 2. For $g \neq 0$ and every $M > 0$,

$$\frac{1}{N} |\{n: r(\tau^n g) \leqslant M\}| \to 0, \quad \text{as } N \to \infty.$$

Then for $g \neq 0$ and $M > 0$, using (4.3) we have

$$\left| \frac{1}{N} \sum_{n=1}^{N} \hat{\mu}_p(\tau^n g) \right| \leqslant \frac{1}{N} \sum_{n=1}^{N} |1 - 2p|^{r(\tau^n g)}$$
$$\leqslant \frac{1}{N} |\{n: r(\tau^n g) \leqslant M\}| + |1-2p|^M.$$

This implies that

$$\limsup_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} \hat{\mu}_p(\tau^n g) \right| \leqslant |1-2p|^M,$$

for all $M > 0$, and hence that

$$\frac{1}{N} \sum_{n=1}^{N} \hat{\mu}_p(\tau^n g) \to 0 = \hat{\mu}_{\mathrm{H}}(g) \qquad (g \neq 0).$$

Since $\hat{\mu}_p(0) = \hat{\mu}_{\mathrm{H}}(0) = 1$, another application of (4.2) establishes (ii).

The next result shows that if the Haar initial distribution is distorted by a continuous function, the distorted distribution will evolve under τ to the Haar distribution.

Theorem 3. If f is a continuous real-valued function on X with Haar integral 1 and $\mathrm{d}\mu = f\,\mathrm{d}\mu_{\mathrm{H}}$, then $\tau^n \mu \to \tau^n \mu_{\mathrm{H}}$.

Proof. We first compute $\hat{\mu}$. Let $g \neq 0$, and let $S(g)$ have the same meaning as in the proof of theorem 2. Put $u(g) = \max\{|i|: i \in S(g)\}$. Let $j \in S(g)$ be such that $|j| = u(g)$. Define $\psi: X \to X$ by $(\psi x)_i = x_i + \delta_{ij}$, so ψ switches 0 and 1 in the jth coordinate. Clearly, ψ preserves μ_{H}. Put $E = \{x \in X: x_j = 0\}$, so E and ψE are disjoint with union X. Then

$$\hat{\mu}(g) = \int_{E \cup \psi E} g(x) f(x)\,\mathrm{d}\mu_{\mathrm{H}}(x)$$
$$= \int_E \{g(x)f(x) + g(\psi x) f(\psi x)\}\,\mathrm{d}\mu_{\mathrm{H}}(x).$$

Since $j \in S(g)$, it follows that $g(\psi x) = -g(x)$. Therefore

$$|\hat{\mu}(g)| \leqslant \int_E |f(x) - f(\psi x)|\,\mathrm{d}\mu_{\mathrm{H}}(x).$$

Now $d(x, \psi x) = 2^{-u(g)}$, so if we define $\mathrm{var}(f, \delta) = \sup\{|f(x) - f(y)|: \ d(x,y) \leqslant \delta\}$, we have

$$|\hat{\mu}(g)| \leqslant \mathrm{var}(f, 2^{-u(g)}).$$

Note that since f is continuous, $\mathrm{var}(f, \delta) \to 0$ as $\delta \to 0$. Now for $g \neq 0$, $u(\tau^n g) = u(g) + n$, so

$$|\hat{\mu}(\tau^n g)| \leqslant \mathrm{var}(f, 2^{-u(g)-n}) \to 0 = \hat{\mu}_H(g), \quad \text{as} \quad n \to \infty .$$

Since $\int f \, d\mu_H = 1$, $\hat{\mu}(\tau^n 0) = 1 = \hat{\mu}_H(0)$, and an appeal to (4.2) completes the proof.

We point out that theorem 3 generalizes to an arbitrary ergodic automorphism τ of a compact abelian group X as follows. Let f be a continuous real-valued function on X, and μ_H be Haar measure on X. Put $d\mu = f \, d\mu_H$. Then $\tau^n \mu \to \mu(X)\mu_H$. The proof uses ergodicity of τ to show that for any nonzero character g on X, $\tau^n g$ ultimately escapes a given finite set of characters, together with a version of the Riemann–Lebesgue lemma for X.

Finally, consider cellular automata whose state space S is a finite group. Additive rules such as (2.2) are clearly endomorphisms of the compact group $S^{\mathbb{Z}} = X$, and the dynamical and statistical structure of their evolution can be completely analysed using the tools indicated here.

5. A nonlinear example

In this section we will discuss the map $\tau: X \to X$ defined by $(\tau x)_i = 1$ if $[x_{i-1} x_i x_{i+1}] = [100]$ or $[001]$, and $(\tau x)_i = 0$ otherwise. This is Rule 18 of [2] and [3]. We show that the natural space on which τ acts is τX, and that the restriction of σ to τX is a finitely-described object called a sofic system. Using the theory of sofic systems, we prove that the restriction of σ to τX has entropy log 1.7549. We then formulate some specific conjectures about the evolution of this cellular automata suggested in part by the numerical work of Grassberger [3].

There are two ways to give a finite description of certain σ-invariant compact subsets Y of X. The first is to specify a finite list of forbidden blocks $\{B_1, \dots, B_r\}$, and define Y to be the set of configurations in which no block B_j appears. For example, if [11] is the only forbidden block, then Y consists of all sequences in X with isolated 1's. Such a Y is called a subshift of finite type or a topological Markov shift [14]. It is effectively described by a 0–1 matrix of allowed transitions, from which the entropy and dynamical behavior can be easily derived.

A more general kind of finite description is one which characterizes allowed sequences as those that can be checked by a finite-state automaton. An example is the set of configurations with an even number of 0's between consecutive occurrences of 1's. No finite list of forbidden blocks can specify this set, so this is not a topological Markov shift. Weiss [8] introduced such systems in 1973 and called them "sofic" from the Hebrew word for "finite." He showed that sofic systems are exactly those obtained by taking continuous images of topological Markov shifts. Later Coven and Paul [15] showed every sofic system is a finite-to-one continuous image of a topological Markov shift, and so properties of the sofic system are closely paralleled by such a Markov cover.

Theorem 4. For the nonlinear rule τ defined above, let $Y = \tau X$. The restriction σ_Y of σ to Y is a sofic system of entropy $\log \lambda$, where $\lambda \approx 1.7549$ is the largest root of $\lambda^3 - 2\lambda^2 + \lambda - 1$.

Proof. It is clear from the form of τ that [111] does not occur in τX. Let $M = [0110]$. We claim that τX is the set of configurations such that between consecutive occurrences of the "marker" M there are an even number of isolated 1's.

Let x be an allowed configuration. First observe that M can only come from $\tilde{M} = [1001]$ under τ. Thus occurrences of M in x determine occurrences of $\tilde{M}$ in any $\tilde{x}$ with $\tau \tilde{x} = x$. A simple parity argument shows that if $x \in \tau X$, the number of isolated 1's between markers must be even. Thus it suffices to show that any block of the form $B = [M 0^{n_1} 1 0^{n_2} 1 \dots 1 0^{n_r} M]$, where r is odd, is the image of a block of the allowed type. Let $\tilde{B} = [\tilde{M} B_1 B_2 \dots B_r \tilde{M}]$, where we construct the B_j as follows. Each

$$B_{2k} = [10^{n_{2k}+2}1] .$$

Each B_{2k+1} has length $n_{2k+1} - 1$, and can be filled in according to the parity of n_{2k+1} to yield a $\tilde{B}$ of the allowed type with $\tau\tilde{B} = B$. This shows that τX has the required form.

M. Boyle has constructed a finite-to-one Markov cover for σ_Y as follows. The idea is to use blocks in Y of length 3 for states, with 3-blocks not in a marker having two "flavors," odd or even according to whether there have been an odd or an even number of isolated 1's since the last occurrence of the marker M. Then M can be entered again only from the even-flavored blocks.

Specifically, let $S = \{[110], [000]_e, [001]_e, [010]_e, [011]_e, [100]_e, [101]_e, [000]_o, [001]_o, [101]_o, [100]_o, [010]_o\}$. Here [110] represents the marker M, and subscripts "e" and "o" represent parity. Let

$$C = \begin{bmatrix} 0&0&0&0&0&1&0&0&0&0&0&1 \\ 0&1&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0&0&0&1 \\ 1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0&0&1&0 \\ 0&0&1&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&1&0&1 \end{bmatrix} .$$

Thus C reflects allowed transitions, with the parity subscript changed if a new 1 is added. Define

$$\Sigma_C = \{w \in S^{\mathbb{Z}} \colon C_{w_i w_{i+1}} = 1 , \quad \text{for all } i\} ,$$

and $(\sigma_C w)_i = w_{i+1}$. For $[abc]$ in S, subscripted or not, put $\pi_0([abc]) = b$. Then π_0 yields a continuous covering $\pi\colon \Sigma_C \to Y$ such that $\pi\sigma_C = \sigma_Y\pi$. It happens that π is one-to-one on the subset of Σ_C of those w with some occurrence of the marker [110]. This means that π is one-to-one over the set in Y of points containing a marker. Marker-free configurations are characterized as those with just isolated 1's, and this forms the topological Markov shift Z of those sequences with [11] forbidden. Over Z, π is at most two-to-one.

The entropy of σ_C is the logarithm of the spectral radius of C [14]. Since the characteristic polynomial of C is $\lambda^9(\lambda^3 - 2\lambda^2 + \lambda - 1)$, $h(\sigma_C) = \log \lambda$, where $\lambda \approx 1.7549$ is the largest root. Entropy is preserved under π, so $h(\sigma_Y) = \log \lambda$ as well. It is interesting to note that on the set Z over which π is not one-to-one, $h(\sigma_Z) = \log(1 + \sqrt{5}/2) \approx \log 1.6180$.

Let $\beta_n(Y)$ denote the number of allowed blocks in Y of length n. Since the Markov cover $\pi\colon \Sigma_C \to Y$ is at most two-to-one, $\beta_n(Y)$ is within two absolute constants times $\beta_n(\Sigma_C)$. Now $\beta_n(\Sigma_C)$ is the sum of the entries in C^n, and therefore is within two absolute constants times λ^n, where λ is as in theorem 4. We therefore draw the following conclusion:

Corollary. There are positive absolute constants c_0 and c_1 such that $c_0\lambda^n < \beta_n(Y) < c_1\lambda^n$, where $\lambda \approx 1.7549$ is the largest root of $\lambda^3 - 2\lambda^2 + \lambda - 1$.

Although the sofic system Y is the natural stage on which τ acts, numerical work indicates that a thin subset of Y is statistically much more significant.

Let $A_0 = \{x \in X \colon x_{2i} = 0 \text{ for all } i\}$, $A_1 = \{x \in X\colon x_{2i+1} = 0 \text{ for all } i\}$, and $A = A_0 \cup A_1$. Clearly, $A \subset Y$. Also, $\tau\colon A_0 \to A_1$, $\tau\colon A_1 \to A_0$, and the restriction of τ to A is the additive rule of section 4. Thus τ evolves on A according to a rule known in considerable detail. A very stimulating observation of Grassberger [3] is that under the evolution of τ on Y, typical points are attracted to A most of the time.

To make this precise, for $y \in Y$ define a "kink" in y to occur between sites i and $i + 1$ if there is an n for which

$$[y_{i-n} \ldots y_i y_{i+1} \ldots y_{i+n+1}] = [10^{2n}1] .$$

Then configurations in Y can be decomposed into alternating regions of parity consistent with A_0 or with A_1, separated by kinks. Grassberger observed that under τ, kinks are never created, but that neighboring kinks can move together and annihilate each other. Thus $\tau^n y$ tends to have increasing domains of pure type A_0 or A_1, separated by kinks which decrease in density as $n \to \infty$. If kinks move according to a random walk, the kink density of $\tau^n y$ should be about $c/\sqrt{n}$. However, recurrence of random walks means that at a fixed site, a kink would move through 0 infinitely often. Grassberger's numerical work on this problem gives good evidence that the random walk model is accurate, even producing the value of the "diffusion constant" c. These observations can be formulated as follows.

On A_i put the measure α_i which independently assigns probability $\frac{1}{2}$ to seeing a 0 or 1 at sites $2j + i - 1$ for all j. Then τ preserves the average $\alpha = (\alpha_0 + \alpha_1)/2$ on A, and this system is essentially that of section 4. Some additional numerical work has indicated that reasonable measures on Y evolve towards α under τ. However, A is not an attractor for τ in the usual sense, since kinks apparently recur infinitely often. This would mean that $\tau^n y$ typically moves away from A infinitely often, but so rarely that statistical features are not disturbed.

The mechanism for kink movement, annihilation, and the resulting diffusion of kinks appears to be mathematically complicated. A very simple model problem of this behavior was proposed a decade ago by Erdös and Ney [4]. Suppose there is a particle at each site in $\mathbb{Z}$, and these particles undergo independent simple random walks, with annihilation if particles cross paths or land in the same site. Erdös and Ney conjectured, but could not prove, that the probability that the origin is visited equals 1. This problem was solved later by Adelman [5]. Kinks in Y move much more irregularly, so the exact mathematical description of kink movement and annihilation represents a challenging problem.

Based on computational evidence, the following conjectures seem quite likely to be true. To state the first, let $\operatorname{dist}(x, A) = \inf\{d(x, y): y \in A\}$.

Conjecture 1. For μ_H-almost every $x \in X$,

$$\limsup_{n\to\infty} \operatorname{dist}(\tau^n x, A) > 0\,,$$

while for every $\epsilon > 0$,

$$\frac{1}{N}\left|\{n: 1 \leqslant n \leqslant N, \operatorname{dist}(\tau^n x, A) > \epsilon\}\right| \to 0\,.$$

Conjecture 2. $\tau^n \mu_H \to \alpha$.

Conjecture 3. For μ_H-almost every $x \in X$, the density of kinks in $\tau^n x$ exists and equals $(8\pi D n)^{-1/2}$, where the diffusion coefficient D has the approximate value $\frac{1}{2}$.

We remark that were conjecture 2 true, one could deduce from the ergodic theorem that the density $T(k)$ of triangles of 0's with base k should behave as $T(k) \sim (\sqrt{2})^{-k}$. This is about the same as the rate $(4/3)^{-k}$ observed experimentally in [2].

Acknowledgements

The author wishes to thank the organizers of the Workshop on Cellular Automata for their generous hospitality during a stimulating conference. He also gratefully acknowledges the support of National Science Foundation grant MCS 8201542.

References

[1] J. von Neumann, Theory of Self-Reproducing Automata, A.W. Burks, ed. (Univ. of Ill. Press, Urbana, 1966).

[2] S. Wolfram, Statistical mechanics of cellular automata, Rev. Mod. Physics 55 (1983) 601–644.

[3] P. Grassberger, A new mechanism for deterministic diffusion, Univ. of Wuppertal preprint.
[4] P. Erdös and P. Ney, Ann. of Prob. 2 (1974) 828.
[5] O. Adelman, Ann. Inst. H. Poincaré XII (1976) 193.
[6] S.J. Willson, Math. Systems Theory 9 (1975) 132.
[7] D. Lind, Israel J. Math. 28 (1977) 205.
[8] B. Weiss, Monatsh. Math. 77 (1973) 462.
[9] G. Hedlund, Math. Systems Theory 3 (1969) 320.
[10] J. Lamperti, Probability (Benjamin, New York, 1966).
[11] E. Coven and M. Paul, Math. Systems Theory 8 (1974) 167.
[12] P. Halmos, Bull. Amer. Math. Soc. 49 (1943) 619.
[13] N. Dunford and J. Schwartz, Linear Operators I (Wiley, New York, 1964).
[14] M. Denker et al., Ergodic Theory of Compact Spaces (Springer, New York, 1976).
[15] E. Coven and M. Paul, Israel J. Math. 20 (1975) 165.