SRC TR 86-33

On the Prediction of Local Patterns in Cellular Automata

by

W. John Wilbur, D.J. Lipman, and
S.A. Shamma

ON THE PREDICTION OF LOCAL PATTERNS IN CELLULAR AUTOMATA

by

W. Jphn Wilbur, David J. Lipman

and Shihab A. Shamma

Physica 19D (1986) 397–410
North-Holland, Amsterdam

ON THE PREDICTION OF LOCAL PATTERNS IN CELLULAR AUTOMATA

W. John WILBUR, David J. LIPMAN and Shihab A. SHAMMA*
Mathematical Research Branch, National Institute of Arthritis, Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20205, USA

Received 12 February 1985

The class of deterministic one-dimensional cellular automata studied recently by Wolfram are considered. We represent a state of an automaton as a probability distribution of patterns of a fixed size. In this way information is lost but it is possible to approximate the stepwise action of the automaton by the iteration of an analytic mapping of the set of probability distributions to itself. Such nonlinear analytic mappings generally have nontrivial attractors and in the most interesting cases (Wolfram Class III) these are single points. The point attractors under appropriate circumstances provide good approximations to the frequencies of local patterns generated by the discrete rules from which they were derived. Two appropriate settings for such approximation are transient patterns generated from random starts and patterns generated in a noisy environment. In the case with noise, improvement is found by correction of the analytic mappings for the effects of noise. Examples of both types of approximation are considered.

1. Introduction

Since the pioneering work of von Neumann [7, 8] on cellular automata it has been recognized that such systems provide an interesting and provocative approach to problems of high complexity. As noted by Wolfram [10, 11], cellular automata have been used to model a variety of systems in physics, chemistry, and biology. One of the impediments to progress in this area is the lack of a generally applicable and tractable global theory of cellular automatons which will allow the prediction of "macroscopic" properties or "average" behavior.

In his original investigations von Neumann [7, 8] envisioned a theory of cellular automata which would incorporate elements of the theory of algorithms, information theory, and thermodynamics. Wolfram [10] has recently discussed the "thermodynamic" and statistical aspects of a class of relatively simple cellular automatons. Both local and global properties of states generated by repeated application of the automaton rules are considered. Local properties refer to aspects of a single state or time sequence of states whereas global properties refer to ensembles of states. Under local properties Wolfram considers the densities of zeroes and ones which develop in a state for short or long times and also the correlations which develop between sites in a state during time evolution. It is pointed out however that these are inadequate measures because "Individual configurations appear to contain long sequences of correlated sites, punctuated by disordered regions". This leads to the consideration of the frequency of occurrence of various length runs of constant value (0 or 1) within a state. Such runs correspond to triangular structures within a time sequence of states. Formulas are derived in important cases using the concepts of self-similarity and fractal dimension. A somewhat complementary approach to local structure which is especially useful in the limit of long times has been pioneered by Grassberger [4] who shows many nonadditive rules may be understood as the local operations of

*Presently at: Dept. of Electrical Engineering, Univ. of Maryland, College Park, MD 20742, USA.

0167-2789/86/$03.50 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

additive rules with moving boundaries of "kinks" between them. The boundaries are governed by diffusion laws. In discussing global properties Wolfram assumes the state size, N, large and considers the states of a cellular automaton identified with the 2^N binary numbers or sequences of zeroes and ones of length N. With a periodic boundary condition each elementary automaton rule defines a self mapping of the set of states. Such concepts as the ergodic nature of the mapping, its entropy, and its irreversibility are investigated. By repeated application of the automaton rule one may, for finite N, obtain an equilibrium distribution of states which defines an attractor for the mapping and a set on which it is reversible. Related concepts when $N=\infty$ have been considered by Farmer [2, 3], Waterman [9], Grassberger [5], and Lind [6].

The discussion we present is concerned with the category of local properties. Our objective is to develop a statistical theory to predict the patterns of length k (some fixed k) that may develop when an automaton evolves from some initial state. If the initial state is random the patterns that develop provide evidence of the self-organizing ability of such systems (see Wolfram [10]). Akin to this are the local state patterns that develop when such an automaton evolves in a noisy environment. In both cases one can hope for a tractable statistical theory of patterns. Both cases also appear relevant to biological systems which on the one hand convert relatively disordered environments into ordered structures and on the other persist as ordered structures in the environment in the face of ambient noise. The patterns considered by Wolfram [10] which are either strings of zeroes or strings of ones and which relate to triangular structures are a special case of the general patterns we wish to consider. Our discussion will be limited to the simple one-dimensional automatons considered by Wolfram, but the same methods are applicable in a much broader context. In section 2 we define the analytic mappings which approximate aspects of the local behavior of the automata. It is found that in important cases the fixed point attractors of the mappings may be associated with patterns at various levels. In section 3 examples of the application of this technique are given.

2. The analytic mappings

We restrict our attention to the class of one dimensional automatons considered by Wolfram [10]. Any cell is required to have a state in the set $B=\{0,1\}$ of binary digits. A rule is any mapping R: $B^3 \to B$. Thus there are 2^8 different rules corresponding to the decimal numbers 0 to 255 and following Wolfram [10] we shall employ these numbers to represent the corresponding rules. For any positive integer k a k-tuple is any element of B^k written $(i_1, i_2, \ldots, i_k)$. Corresponding to a given rule R there is a function r defined on the set of all k-tuples, $k \geqq 3$, which maps a k-tuple to a $k-2$ tuple. The function r is defined by

$$r(i_1, i_2, \ldots, i_k) = (R(i_1, i_2, i_3), R(i_2, i_3, i_4), \ldots, R(i_{k-2}, i_{k-1}, i_k)). \qquad (1)$$

For any k-tuple there is a circularizing function c defined by

$$c(i_1, i_2, \ldots, i_k) = (i_k, i_1, i_2, \ldots, i_k, i_1). \qquad (2)$$

To define a cellular automaton we need the rule R (and thus r) and a fixed positive integer N. The cellular automaton $a(R, N)$ is then the composite mapping $r \cdot c$ restricted to act on N-tuples. Any N-tuple is a possible state of $a(R, N)$ in this context and $a(R, N)$ is a self-mapping on the set B^N of states. It is of course possible to have $N=\infty$ with a slight change in notation but this will not be necessary for our purposes.

Now we are interested in local patterns that may occur in a state. Such a local pattern is a k-tuple where k is generally much smaller than N. We shall refer to such k-tuples for small k as k-patterns. Define, for any positive integers m and q,

$$S(m, q) = [(q+m-1) \bmod N] + 1. \qquad (3)$$

Then $S(m, \cdot)$ is a shift operator which will shift the indices of a state, i.e. an N-tuple, m positions on the circle. We say a k-pattern $(i_1, i_2, \ldots, i_k)$ occurs in a state $(j_1, j_2, \ldots, j_n)$ iff (if and only if) there is some $m \geqq 0$ with

$$i_q = j_{S(m,q)}, \quad 1 \leqq q \leqq k. \tag{4}$$

There are 2^k different k-patterns and it will be convenient to define the k-pattern $S(d)$ by

$$S(d) = (i_1, i_2, \ldots, i_k), \\ d - 1 = \sum_{q=0}^{k-1} i_{(q+1)} * 2^q, \tag{5}$$

for each d, $1 \leqq d \leqq 2^k$.

We wish to think of the set of k-patterns $\{S(d)\}_{d=1}^{2^k}$ as the set of states for a Markov process. If an automaton $a(R, N)$ acts on a state we cannot circularize a k-pattern occurring in this state and then apply r to see what k-pattern it will transform to in the subsequent state. To predict accurately its transformation we must know the values in the cells on either side of the k-pattern. For any k-pattern there are four possible extensions which may be written as

$$e(j_1, j_2)(i_1, i_2, \ldots, i_k) = (j_1, i_1, i_2, \ldots, i_k, j_2). \tag{6}$$

If we know exactly which extension, say $e(j_1, j_2)$, a given k-pattern $S(d)$ has in a state then we know $S(d)$ will transform to $r(e(j_i, j_2)S(d))$ after one iteration of $a(R, N)$. If we do not know which extension is correct for $S(d)$ but only that $e(j_1, j_2)$ occurs with probability $q(j_1, j_2, d)$, then $q(j_1, j_2, d)$ is the probability that pattern $S(d)$ undergoes a transition to $r(e(j_1, j_2)S(d))$. Then the matrix A defined by

$$A_{d,d'} = \\ \sum \{ q(j_1, j_2, d) | S(d') = r(e(j_1, j_2)S(d)) \} \tag{7}$$

is the transition matrix for the discrete Markov process which models k-pattern transitions under iteration of the automaton. Strictly speaking the transition probabilities for a given pattern depend in some complicated way on the patterns in the same location for previous states. Thus the process is a stochastic process more complicated than we represent. Our purpose is to ignore this complication and investigate the adequacy of the mathematically much more tractable Markov model. Even for the Markov model the numbers $q(j_1, j_2, d)$ will change with the state so that we are not dealing with a stationary Markov process. Without regard to the precise values of the $q(j_1, j_2, d)$, if we assume they are all nonzero, the maximal transition diagram between patterns is determined. Let us call such a diagram of allowed transitions between patterns a *maximal diagram*. A maximal diagram gives information about transient states, ergodic classes, and the exclusion of certain kinds of transitions. This already provides some insight into how patterns can develop locally in states. Fig. 1 shows several maximal diagrams for Markov processes associated with patterns of size $k = 3$. The diagram for rule 50 shows there is one ergodic class of size 2 which insures the existence of certain alternating patterns in successive states. From the diagram for rule 222 we can deduce that beginning with any state having some 1's, iteration will produce blocks of ones separated at most by isolated zeroes. The rules 122 and 126 are closely related. The maximal diagram for rule 122 shows that all eight states form an ergodic class. The diagram for rule 126 shows that pattern 010 is transient and completely disappears after a single iteration. Of the eight patterns of length three on which rules 122 and 126 are defined, they only differ on 010. Thus after one iteration 126 will produce a state on which 122 and 126 would produce identical results for any number of subsequent iterations. We shall discuss the relation between rules 122 and 126 further in the next section.

We take up now the question of how the $q(j_1, j_2, d)$ are to be determined. If all $q(j_1, j_2, d)$ are set equal to 1/4 then the Markov process is an

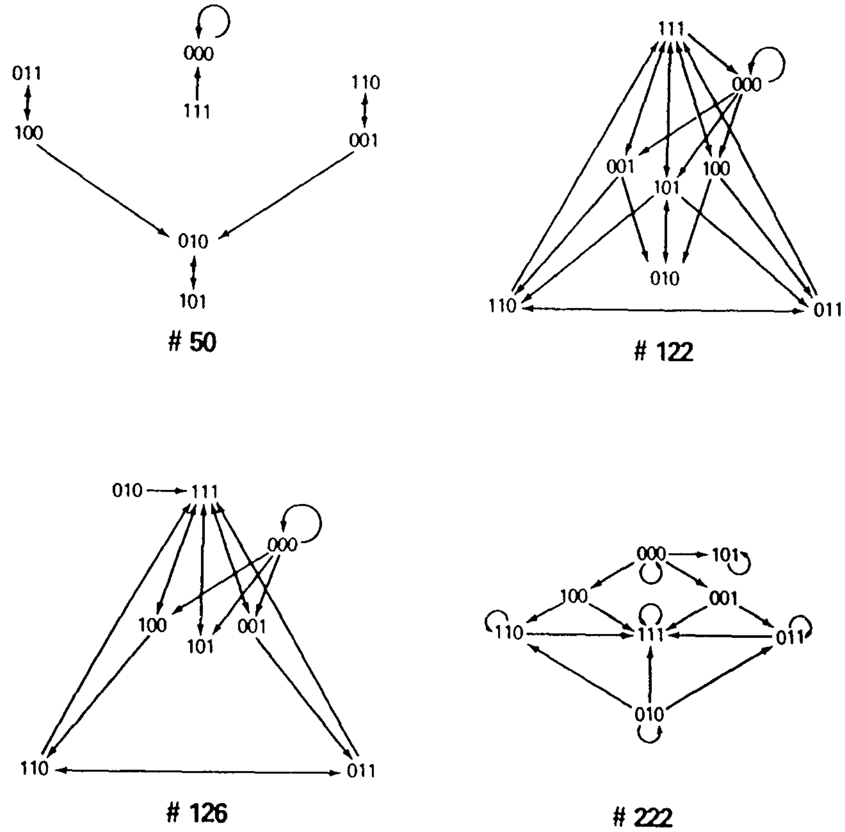

Fig. 1. Maximal diagrams for 3-pattern transitions for the automaton rules 50, 122, 126, and 222. See text for discussion.

accurate description of the pattern transitions that occur when the automaton is applied for one iteration starting with a random state (zero or one has been randomly and independently assigned to each cell with equal probability). After one application of the automaton the state is seldom any longer random; to perfectly predict k-pattern transitions we therefore require knowledge about extensions to $k+2$-patterns. Essentially we require information about $k+2$-patterns at one step to produce information about k-patterns at the next step. Our thesis is that in fact k-pattern information at one step will generally work very well in predicting k-patterns at the next step. We assume that k-pattern information is given in the form of a probability distribution $\{pS(d)\}_{d=1}^{2^k}$ where $pS(d)$ is the probability of the pattern $S(d)$. Then

$$\begin{aligned} p_1(j_1; i_1, i_2, \ldots, i_{k-1}) &= p(j_1, i_1, i_2, \ldots, i_{k-1}) \\ &\quad /(p(0, i_1, i_2, \ldots, i_{k-1}) + p(1, i_1, i_2, \ldots, i_{k-1})), \\ p_r(j_2; i_2, i_3, \ldots, i_k) &= p(i_2, i_3, \ldots, i_k, j_2) \\ &\quad /(p(i_2, i_3, \ldots, i_k, 0) + p(i_2, i_3, \ldots, i_k, 1)), \end{aligned} \tag{8}$$

are, respectively, the conditional probability of extending $(i_1, i_2, \ldots, i_{k-1})$ one cell to the left to $(j_1, i_1, i_2, \ldots, i_{k-1})$ and the conditional probability of extending $(i_2, i_3, \ldots, i_k)$ one cell to the right to $(i_2, i_3, \ldots, i_k, j_2)$. Thus if $S(d) = (i_1, i_2, \ldots, i_k)$ we may set

$$q(j_1, j_2, d) = p_l(j_1; i_1, i_2, \dots, i_{k-1}) \cdot p_r(j_2; i_2, i_3, \dots, i_k). \quad (9)$$

The approximation involved here is that j_1 is predicted on the basis of $i_1, i_2, \dots, i_{k-1}$ and any additional information i_k and j_2 might supply is ignored and a similar statement applies to the prediction of j_2. The additional information noted here can be quantified in terms of standard formulas of information theory (see Billingsly [1]). As such it gives some insight into the goodness of the approximations we are making. It generally decreases with increasing k and examples will be considered in the next section.

The formula (9) defines the matrix A and if $\boldsymbol{P} = \{pS(d)\}_{d=1}^{2^k}$ is the probability distribution vector, then we show the dependence of A on $\boldsymbol{P}$ by writing $A(\boldsymbol{P})$. The transition matrix $A(\boldsymbol{P})$ acting on the vector $\boldsymbol{P}$ gives the distribution vector at the next step, i.e.

$$\boldsymbol{P}' = F(\boldsymbol{P}) = A(\boldsymbol{P})(\boldsymbol{P}) \quad (10)$$

defines the function corresponding to action of the automaton. In order for F to have desirable properties, care must be taken in properly choosing its domain. The difficulty arises because the conditional probabilities defined in (8) can have an indeterminate value, i.e. can be a quotient of zeroes. We arbitrarily introduce a convention to deal with this:

> If the denominator of a conditional probability defined in (8) is zero, then the conditional probability shall assume the value of zero. (11)

Now let H denote the compact subset of 2^k dimensional Euclidean space consisting of all k-pattern probability distribution vectors $\boldsymbol{P}$ which have nonnegative terms and sum to one. The function F has two problems on H. First, at some points in H the image under F is not again in H because its components do not sum to one. This happens when for some k-pattern $S(d)$ with $pS(d) > 0$, $q(j_1, j_2, d) = 0$ for all four possible extensions of $S(d)$. The second difficulty is that F is not continuous on H. This occurs because F maps any point in H all of whose coordinates are nonzero again to a point in H. However, points all of whose coordinates are nonzero can be found arbitrarily close to those points in H which F maps outside of H. Thus F is discontinuous at all such points mapped outside of H. These related problems can be circumvented by choosing a subspace of H which satisfies a simple consistency condition. This condition asserts that given a k-pattern distribution $\boldsymbol{P}$, the two methods of calculating the probability of any $k-1$-pattern from $\boldsymbol{P}$ must yield the same result. This condition is defined by the set of equations

$$p(0, i_1, i_2, \dots, i_{k-1}) + p(1, i_1, i_2, \dots, i_{k-1}) = p(i_1, i_2, \dots, i_{k-1}, 0) + p(i_1, i_2, \dots, i_{k-1}, 1), \quad (12)$$

one for each $k-1$-pattern $(i_1, i_2, \dots, i_{k-1})$. Let HC denote the subset of H which satisfies the equations (12). We shall show that F is a continuous map of HC into itself. We first show that F preserves (12). Let $\boldsymbol{P}$ be in HC and consider the relation

$$p(i_1, \dots, i_k) \cdot p_r(j_2; i_2, \dots, i_k) = p_l(0; i_1, \dots, i_{k-1}) \cdot p(i_1, \dots, i_k) \cdot p_r(j_2; i_2, \dots, i_k) + p_l(1; i_1, \dots, i_{k-1}) \cdot p(i_1, \dots, i_k) \cdot p_r(j_2; i_2, \dots, i_k). \quad (13)$$

This is clearly true if the denominator of the p_l factors appearing on the right is nonzero. If the denominator of the p_l factors is zero then by (12) it is true that $p(i_1, i_2, \dots, i_k) = 0$ so that (13) is satisfied in any case. Thus (13) holds for all distributions satisfying (12). Now suppose we are given a $\boldsymbol{P}$ satisfying (12). We wish to show $\boldsymbol{P}' = F(\boldsymbol{P})$

also satisfies (12). By applying (13) we can write

$$\begin{aligned} p'(0, i_1,\dots, i_{k-1}) + p'(1, i_1,\dots, i_{k-1}) \\ = \sum \{ p(j_1, m_1,\dots, m_{k-1}) \\ * p_r(j_2; m_1,\dots, m_{k-1}) | r(j_1, m_1,\dots, m_{k-1}, j_2) \\ = (i_1,\dots, i_{k-1}) \}, \\ p'(i_1,\dots, i_{k-1}, 0) + p'(i_1,\dots, i_{k-1}, 1) \\ = \sum \{ p_1(j_1; m_1,\dots, m_{k-1}) \\ * p(m_1,\dots, m_{k-1}, j_2) | r(j_1, m_1,\dots, m_{k-1}, j_2) \\ = (i_1,\dots, i_{k-1}) \}. \quad (14) \end{aligned}$$

Here the summands on the right are the same provided $p_r(j_2; m_1,\dots, m_{k-1})$ and $p_1(j_1; m_1,\dots, m_{k-1})$ have the same denominators but that is just what is insured by (12). Thus (12) remains true under the mapping F. Now

$$\begin{aligned} p'(i_1,\dots, i_k) = \sum \{ p_1(j_1; m_1,\dots, m_{k-1}) \\ * p(m_1,\dots, m_k) * p_r(j_2; m_2,\dots, m_k) | \\ re(j_1, j_2)(m_1,\dots, m_k) = (i_1,\dots, i_k) \}. \end{aligned} \quad (15)$$

Thus if the denominator of p_1 or p_r converges to zero in the summand then by (12) $p(m_1,\dots, m_k)$ must also converge to zero. It follows that a limiting indeterminancy in either quotient p_1 or p_r does not lead to indeterminancy in the limit of F and F, based on (11) where necessary, therefore preserves probability and is continuous on HC. Thus F is a continuous mapping of HC into itself. We may now apply the Brower fixed point theorem to F to see that F must have a fixed point in HC. To locate fixed points we have found it generally sufficient to choose a point in HC and iterate the function F. After a few iterations the behavior is usually clear. Oscillations are possible but generally iterations yield a fixed point. In many interesting cases what appears to be a unique stable fixed point, $\boldsymbol{P}_f$, is found. The distribution $\boldsymbol{P}_f$ provides an approximation to the distribution of patterns produced by the automaton in two important cases. First, starting with a randomly determined state and iterating the automaton on this state, one generally passes through a sequence of transients whose distribution of length k patterns come very close to $\boldsymbol{P}_f$ rather early and then move slowly away as longer range correlations ($> k$) gradually build up in the successive states. Second, if a certain level of noise is applied to the state of a cellular automaton after each iteration then long range patterns are broken up and the distribution of length k patterns that develops after a few iterations is approximated by $\boldsymbol{P}_f$. Generally there is a range of noise where the approximation is quite good. For lower noise levels patterns of length greater than k come to predominate and tend to decrease the accuracy of the approximation. For higher levels of noise the patterns of length k are themselves progressively destroyed. This high noise error can be corrected for by a modification of eq. (10).

In order to correct for the effect of noise in eq. (10) we introduce noise as a continuous time Markov process active for unit time between each iteration of the cellular automaton. Specifically we assume that when noise is applied at a level λ to an automaton, then between each iteration each cell of the automaton has its value switched by noise events with an incoming Poisson rate λ. Such noise is to act for a unit time period and is to act on each cell independently of what happens to any other cell. Then if p_0 and p_1 are the probabilities of zero and one in a cell, the flow due to noise between zeroes and ones is governed by

$$\frac{\mathrm{d}p_0}{\mathrm{d}t} = \lambda(p_1 - p_0), \quad \frac{\mathrm{d}p_1}{\mathrm{d}t} = \lambda(p_0 - p_1). \quad (16)$$

When integrated for one unit of time the transformation is

$$\begin{pmatrix} p_0^{T+1} \\ p_1^{T+1} \end{pmatrix} = \begin{pmatrix} \frac{1+\mathrm{e}^{-2\lambda}}{2} & \frac{1-\mathrm{e}^{-2\lambda}}{2} \\ \frac{1-\mathrm{e}^{-2\lambda}}{2} & \frac{1+\mathrm{e}^{-2\lambda}}{2} \end{pmatrix} \begin{pmatrix} p_0^T \\ p_1^T \end{pmatrix}. \quad (17)$$

To apply this transformation to a k-pattern distribution $\boldsymbol{P} = \{ pS(d) \}_{d=1}^{2^k}$ we must for each coordinate-position i, $1 \leq i \leq k$, apply (17) to each pair of patterns which differ only at the ith coordinate. This then produces the effect of the noise level λ acting at the ith coordinate position only and may be represented as a matrix N_i which acts on the vector $\boldsymbol{P}$. To produce the full effect of the noise acting on the k-patterns, all N_i, $1 \leq i \leq k$, must act and the order is unimportant. The result may be represented as a single transition matrix

$$N(\lambda) = N_1 \cdot N_2 \cdot \cdots \cdot N_k. \tag{18}$$

Since $N(\lambda)$ is the transition matrix produced by the noise and it is supposed to be applied once following each iteration of the automaton, the net effect of automaton and noise for each iteration on $\boldsymbol{P}$ may be written

$$\boldsymbol{P}^{T+1} = G(\boldsymbol{P}^T) = N(\lambda) \cdot A(\boldsymbol{P}^T)(\boldsymbol{P}^T). \tag{19}$$

Again we may iterate G, given a particular initial distribution $\boldsymbol{P}$, in an attempt to arrive at a fixed point $\boldsymbol{P}_g$ of G. We generally have found a unique nontrivial fixed point. Such a fixed point then becomes an approximation to the distribution of patterns produced by the automaton running with the same level of noise, after the first few steps. With the correction afforded by $N(\lambda)$ the approximation gets better the higher the level of noise and is only limited by the statistical limitations imposed by the finite size of the state which is being compared with $\boldsymbol{P}_g$.

We close this section with a word about the relation between the functions G and F. Because $N(\lambda)$ is a continuous operator, if $N(\lambda)$ maps HC into HC then G must be continuous and map HC into HC because F has these properties. $N(\lambda)$ is clearly a stochastic matrix. We only need show it preserves (12). We use the fact that $N(\lambda)$ is a product of all the N_j, $1 \leq j \leq k$, in any order. Assume (12) is satisfied. Then for each j, $2 \leq j \leq k$, we let N_j act on the left side of (12) while N_{j-1} acts on the right side. The equations (12) are preserved in successive applications because at each stage (12) insures that equal quantities are being mixed and we know the mixing is in the same proportions for each coordinate. Finally we must let N_1 act on the left of (12) and N_k on the right. Neither of these operations changes the value of the operand in this case because all that is being mixed is already present in the sum. In this way $N(\lambda)$ has acted on both sides of (12) and it remains valid. We conclude that (12) is preserved under an application $N(\lambda)$. Thus G is a continuous map of HC into itself.

Now we can show that the limit of fixed points of G when λ converges downward to zero is a fixed point of F. It is clear from the Brower fixed point theorem that for each $\lambda > 0$, G has at least one fixed point in HC. Let the set of fixed points of G in HC be denoted by U_λ, $\lambda > 0$. Let

$$V_\mu = \cup\{U_\lambda | 0 < \lambda \leq \mu\}$$

and (20)

$$C = \cap\{\overline{V}_\mu | 0 < \mu \leq 1\}.$$

Here $\overline{V}_\mu$ denotes the closure of V_μ and $\overline{V}_\mu$ is always a subset of HC because HC is closed. Because HC is bounded all $\overline{V}_\mu$ are compact. Further, any finite number of $\overline{V}_\mu$ have a nonempty intersection. It follows from the finite intersection property that C is nonempty. We claim that each element of C is a fixed point of F. Any such point $\boldsymbol{P}_c$ is arbitrarily close to a distribution which satisfies

$$\boldsymbol{P} = N(\lambda) \cdot A(\boldsymbol{P})(\boldsymbol{P}) \tag{21}$$

for an arbitrarily small $\lambda > 0$. Because all points are in HC where F is continuous and (21) is equivalent to

$$\boldsymbol{P} = N(\lambda) F(\boldsymbol{P}) \tag{22}$$

we must obtain $\boldsymbol{P}_c = F(\boldsymbol{P}_c)$ as λ goes to zero and $\boldsymbol{P}$ converges to $\boldsymbol{P}_c$.

3. Analysis of examples

While our results have general applicabilility we shall limit our discussion to what Wolfram [10] has

Table I
Survey of the legal rules

Rule	Wolfram Class	Ergodic classes in transition diagram	3-Patterns: Attractors of analytic mappings on HC without noise (0,0,0)	(0,0,1)	(0,1,0)	(0,1,1)	(1,0,0)	(1,0,1)	(1,1,0)	(1,1,1)
0	I	1	Trivial, $p(0,0,0)=1$							
4	II	1,1	2-dimensional manifold of fixed points							
18	III	7	0.327	0.143	0.199	0.044	0.143	0.100	0.044	0
22	III	8	0.285	0.143	0.097	0.090	0.143	0.044	0.090	0.107
32	I	1	Trivial, $p(0,0,0)=1$							
36	II	1	1-dimensional manifold of fixed points							
50	II	2	2-dimen. manifold with 1 fixed pt. & all others period 2							
54	III	8	0.273	0.142	0.094	0.081	0.142	0.032	0.081	0.155
72	II	1	2-dimensional manifold of fixed points							
76	II	1,1,1,1,1	3-dimensional manifold of fixed points							
90	III	8	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
94	II	1	1-dimensional manifold of fixed points							
104	II	1	Trivial, $p(0,0,0)=1$							
108	II	1,1,1	2-dimensional manifold of fixed points							
122	III	8	0.160	0.099	0	0.160	0.099	0.061	0.160	0.259
126	III	7	0.160	0.099	0	0.160	0.099	0.061	0.160	0.259
128	I	1	Trivial, $p(0,0,0)=1$							
132	II	1,1	2-dimensional manifold of fixed points							
146	III	7	0.327	0.143	0.199	0.044	0.143	0.100	0.044	0
150	III	8	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
160	I	1	Trivial, $p(0,0,0)=1$							
164	II	1	1-dimensional manifold of fixed points							
178	II	2	2-dimen. manifold with 1 fixed pt. & all others period 2							
182	III	7	0	0.044	0.100	0.143	0.044	0.199	0.143	0.327
200	II	1,1,1,1	3-dimensional manifold of fixed points							
204	II	identity, each element a class	(identity function)							
218	II	1	1-dimensional manifold of fixed points							
222	II	1,1	2-dimensional manifold of fixed points							
232	II	1,1	2-dimensional manifold of fixed points							
236	II	1,1,1,1	3-dimensional manifold of fixed points							
250	I	1	Trivial, $p(1,1,1)=1$							
254	I	1	Trivial, $p(1,1,1)=1$							

termed "legal rules". A "legal rule" R satisfies

i) $R(0,0,0)=0$;

ii) $R(i_1,i_2,i_3)=R(i_3,i_2,i_1)$.

Condition i) asserts that the all zero states acts as a "stable" ground state for any automaton based on R while condition ii) is a symmetry property removing any directional bias from the automaton. There are thirty-two legal rules and we shall begin with a survey of the legal rules which gives their status according to Wolfram's classification, the sizes of ergodic classes in their 3-pattern maximal transition diagrams, and describes the attractors of their corresponding 3-pattern analytic mappings without noise. This information is contained in table I. The three Wolfram [11] classes relevant to our discussion are defined as follows:

Class I – "evolution leads to a homogeneous state in which, for example, all sites have value 0;"

Class II – "evolution leads to a set of stable or periodic structures that are separated and simple;"

Class III – "evolution leads to a chaotic pattern."

The important observation here is the correlation between Wolfram class, ergodic class size, and the nature of the attractor:

Class I – there is always a single ergodic class of size 1 and the attractor is trivial consisting of the ground state (all zeroes) for rules 0, 32, 104, 128, and 160 and the state of all 1's for rules 250 and 254;

Class II – there are one or more ergodic classes of size 1 or 2 and the attractor is a 1, 2, or 3-dimensional set of fixed points for the analytic mapping;

Class III – there is one ergodic class of 7 or 8 elements and the attractor is a single point which is nontrivial, i.e., does not correspond to a single state of the cellular automaton as it does for class I automatons.

Our main interest is in the class III automatons of which there are nine.

Based on their 3-pattern attractors we can divide them into five groups:

group 1: [18, 146, 182] – 18 and 146 have the same attractor and the mirror image of their common attractor is the attractor for 182;

group 2: [22] – unique single point attractor;

group 3: [54] – unique single point attractor;

group 4: [122, 126] – common attractor;

group 5: [90, 150] – common attractor.

The significance of these groupings lies in the predicted relatedness of local patterns produced by automatons within each group. Rule 18 and 146 of group 1 have the same ergodic class of seven elements which does not include the triplet 111. The two rules disagree only on 111, which is transient, so that given a random starting state the pattern 111 disappears after some finite number of iterations. Thus 18 and 146 not only produce the same local state patterns eventually but they also eventually act identically on states to produce the same dynamic pattern (see Wolfram [10, fig. 8]). This explains the observation of Grassberger [5] that rules 18 and 146 produce the same attractors in the dynamical systems approach. Rules 146 and 182 are related to each other in that each can be thought of as the photographic negative of the other. Thus if 182 were given the photographic negative of the input of 146, 182's output would be the photographic negative of the output of 146. Evidently 182 has the same kind of local structure and constraints as 146. In group 4 rule 122 has an ergodic class one larger than that of 126 but the extra element 010 has a probability of occurrence which decreases exponentially with time as 122 is iterated. Because 122 and 126 agree except on 010 they have the same attractive point and with random initial states produce the same local patterns and the same dynamic sequence after 010 is eliminated (again see Wolfram [10, fig. 8]). Because 010 is progressively eliminated in patterns generated by 122 starting with randomly determined initial states, the element 010 behaves as a transient in this case and its inclusion in the ergodic class as determined from the maximal diagram of fig. 1 is misleading. This illustrates the limitations of the maximal transition diagram representation. The linear rules 90 and 150 of group 5 have the same attractive fixed point. Because all components of the fixed point are equal this suggests that all 3-patterns are equiprobable for these automatons. The same is true for patterns of any finite length. Though the attractors are the same, rules 90 and 150 behave quite differently dynamically.

4. Empirical results

Up to this point we have assumed that the attractive points of the analytic mappings provide reasonable approximations to the pattern frequencies of the corresponding automatons. To justify such an assumption we must justify the approximation employed in constructing the analytic mapping. As discussed in the previous section, for a k-pattern analytic mapping the approximation should be good if the prediction of a cell at position i based on the $k-1$ cell values to its right (left) is only improved by a small amount by including the next two cell values to the right (left) of position $i+k-1$ $(i-k+1)$. We need only consider elements to the right because of symmetry. The improvement in prediction may be calculated in bits of information. Such a calculation may be made in any situation where one wishes to employ the analytic approximation. One situation where we wish to employ the approximation is in the transient generation of patterns by automatons iterated on randomly determined initial states. We have computed the relevant information in bits for $k = 2, \ldots, 8$ for this problem and a representative sample is shown in table II. For each group an automaton was chosen (the results are identical within a group) and started on a randomly constructed (0 and 1 equiprobable) initial state of length 1000 and iterated 500 times and then the tuple counts for various pattern sizes were summed for the next 50 iterations. These counts then provided the data from which the incremental information was computed. The increments in information are generally small though group 3 (rule 54) for k of 2 or 3 is somewhat of an exception. The trend is progressively downward for increasing k. Note in this regard that for a given automaton the sum of all figures for odd or for even k must total less than 1 (i.e., there is at most one bit of information to be determined). Group 5 presents a special case in that all patterns of any given length should be equiprobable and hence incremental information should be zero for all k. A slight departure from zero for larger k simply reveals the effect of the finite number of cells counted. It provides a check on the method and suggests the level of error introduced into the values for other groups because of finite sample size. Because of the special status of group 5 with zero incremental information, the analytic approximation should be perfect with the only error in pattern frequencies arising from the finite lengths of states. No additional analysis of group 5 from this approach is of interest and we shall not consider group 5 further in our discussion.

Table II
The information increment which influences the analytic approximation for k-patterns, $k = 2, \ldots, 8$, for the five groups of class III automatons. See text for the method of calculation employed here.

	Group				
k	1	2	3	4	5
2	0.076	0.023	0.369	0.136	0
3	0.060	0.024	0.410	0.074	0
4	0.034	0.012	0.062	0.057	0
5	0.027	0.014	0.050	0.032	0
6	0.016	0.014	0.051	0.026	0.001
7	0.012	0.012	0.029	0.015	0.002
8	0.008	0.013	0.019	0.012	0.003

The results in table II suggest the analytic mappings should provide a good approximation to the frequency of different patterns in the states of cellular automata. The fixed points are limits and should approximate limiting pattern distributions in the cellular automatons seen after an initial period of iteration. To illustrate this effect we have performed a transient analysis for rules 18, 22, 54, and 122 which are representative of the first four groups. This is shown in fig. 2. For each rule a state of length 1000 was used and 2000 iterations were performed. For each of the first 200 iterates and at multiples of 200 thereafter the sum of the absolute values of the differences between the 3-pattern frequencies predicted by the fixed points of table 1 and the observed 3-pattern frequencies for the automata were computed. The maximal value of such a sum of absolute values is 2. These values were then divided by 2 and converted to percents to represent the percentage of patterns misclassified by the fixed point approximations. Each curve shown represents the average of 100 such transient calculations and the dotted curve is the standard deviation of the 100 different percents computed at the different iteration numbers. The values at zero iterations represent the difference between the fixed point predictions and random starting states which have not been acted on by the automaton. As such these values provide a standard by which to judge the goodness of the approximation. A general pattern of behavior is seen in which the pattern frequencies rapidly approach the fixed point predictions and then move slowly away. The movement away is explained by the development of significant dependencies over distances greater than that encompassed by 3-patterns. Such long range dependencies require time to develop and even when they do the approximation of 3-pattern frequencies by the analytic attractors remains much better than random. The figures in table II were calculated between 500 and 550 iterations so that they represent the situation after long range correlations have become quite well developed. In this regard the relatively large error predicted in table II, group 3 (automaton rule 54), $k = 3$ is confirmed in fig. 2 for rule 54 for 500–550 iterations. While fig. 2 is based on 3-patterns, little change is seen with larger patterns except as pattern size increases, the minima in the curves is seen at progressively higher iteration numbers.

While the transient analysis shows a close relationship between the analytic point attractors and the local pattern frequencies that are produced when an automaton is iterated, it also reveals that long range correlations develop that tend to destroy this correspondence. For a cellular automaton running in a noisy environment such long range correlations tend to be destroyed and the higher the level of noise the shorter the range of correlations that can develop. This suggests we may find the analytic attractors to be even better approximations to the local pattern frequencies produced by cellular automatons operating in a noisy environment. We have found such a correlation by two different methods of approximation.

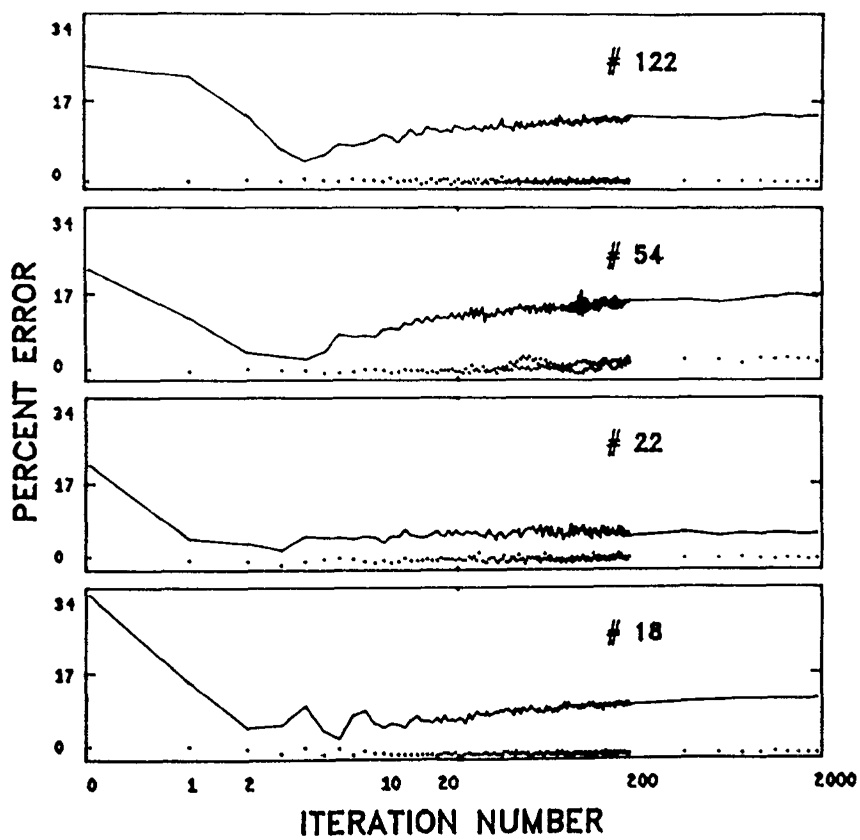

Fig. 2. The deviation of actual 3-pattern frequencies from those predicted by the fixed point attractors for the rules 18, 22, 54, and 122. The average deviation (solid curves) and the standard deviation of deviations (discrete points) are based on a simulation of 100 trials whose details are explained in the text.

The methods are complementary, one working for very low levels of noise and the other for higher levels of noise.

Approximation 1: The fixed point of the analytic 10-pattern mapping is computed and by summing appropriate terms 3-pattern frequencies are computed and used to approximate 3-pattern frequencies of the automaton in an environment with low levels of noise.

Approximation 2: The attractor of the analytic 3-pattern mapping with noise is computed and used to approximate 3-pattern frequencies of the automaton in a noisy environment.

In making the approximations here the cellular automaton is assumed to have run for a sufficient length of time to equilibrate with the effect of the noise in the environment. Curves showing the percent error (computed as in fig. 2) by approximations 1 and 2 (dotted curves) for 3-patterns at different levels of noise for rules 18, 22, 54, and 122 are shown in fig. 3. States of length 2000 were used and each automaton was run for 500 iterations without noise and then for 200 iterations at each noise level and then 3-patterns were counted over five successive iterations to produce the observed 3-pattern frequencies. With low levels of noise relatively long range correlations are possible and the analytic 10-pattern attractor (fixed point) provides for such correlations and allows an accurate prediction of the 3-pattern frequencies in the state of the automaton. As the noise level increases to a λ of between .01 and .05 the longer range correlations begin to be destroyed and approximation 1 degrades. It is in this region that approximation 2, which ignores long correlations, begins to become useful. The only factors preventing the approximation 2 from being perfect are correlations at a longer range than 3-patterns and the finite length of the automaton state. As the noise level increases the patterns of length greater than three are progressively destroyed so that the approximation improves and is only limited by the state length. In our calculations we count five successive states for a total of 10,000 cells to obtain the frequencies of the different 3-patterns. Simulations we have performed show that this introduces a variance due to finite length on the order of 1% so that the approximations are close to the theoretical limit at the extremes in the noise levels. An excellent approximation over all noise levels can be obtained by combining the two methods, i.e., essentially using approximation 1 but with the 10-pattern mapping replaced by the 10-pattern mapping with the appropriate level of noise. This however entails much more calculation (determination of the 10-pattern attractor for each noise level) and also does not allow the illustration of the effect of different factors on the approximation.

We only display results for rules 18, 22, 54, and 122 in fig. 3. However, when noise is applied at any level greater than zero the limiting local pattern frequencies produced by, for example, 18 and 146 become distinct and likewise for any pair of automatons employing different rules. In such a case no pattern is transient and any pattern on which two rules disagree will occur in any state with some finite probability. Of course the difference that develops between a pair like 18 and 146 is greater the greater the noise level applied. This difference is also reflected in the fixed points of the analytic mappings with noise. Another important effect of noise is that the attractors of the analytic mappings with noise are generally single points regardless of the Wolfram class. We have found no exception to this rule. Thus the variety of different types of attractors possible without noise, as exhibited in table I, are excluded by noise. The behavior of the analytic mapping with noise may be viewed as a perturbation of the analytic mapping without noise. Starting at a random initial state and iterating the analytic mapping with low noise one generally finds a rapid convergence to the region close to the attractor without noise. For those automatons which have attractors of dimension 1, 2, or 3 without noise (Wolfram class II) the initial rapid convergence is followed by a period of slow movement approximately across the surface of the attractor finally reaching that unique point which is the attractor for the mapping with noise.

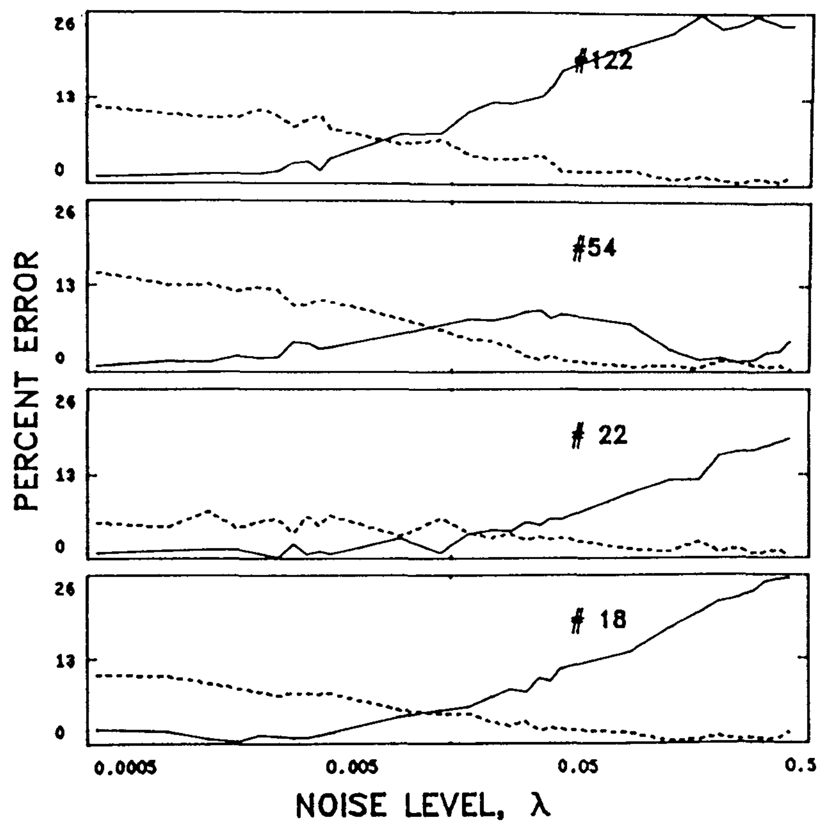

Fig. 3. The deviation of actual 3-pattern frequencies with noise from those predicted by fixed point attractors based on 10-pattern analytic mappings without noise (solid curves) and 3-pattern analytic mappings with noise (short dashed curves). Each point on a curve is an average based on simulation of 100 trials. See text for details.

The period of slow movement may be viewed as a consequence of the noise alone because all the points of the attractor without noise are fixed points of the mapping without noise.

We have developed a statistical theory of local patterns which in some sense fulfills the thermodynamical paradigm. Local pattern frequencies function as "macroscopic" state variables. Such state variables provide a partial description of the state while ignoring the fine structure of interweaving of patterns. The state variables are approximately governed in their evolution by the analytic mappings. This allows the prediction of the general trend of events and the discussion of meaningful equilibrium phenomena. The application of the method to more complicated systems is straightforward. One important limitation on the approach however is the fact that the number of patterns of a given size increases exponentially with the size of the pattern. This quickly leads to enormous numbers of rather modest sized patterns. In one sense this simply reveals the complexity of the problem. On the other hand computation of the analytic mappings can become impractical. One solution to this difficulty may lie in the use of array processors. The possibility of other solutions is a subject for future research.

References

[1] P. Billingsly, Ergodic Theory and Information (Wiley, New York, 1965).
[2] J.D. Farmer, Dimension, fractal measures, and chaotic

dynamics, in: Evolution of Order and Chaos in Physics, Chemistry and Biology, H. Haken, ed., (Springer, Berlin, 1982).
[3] J.D. Farmer, Information dimension and the probabilistic structure of chaos, Z. Naturforsch. 37a (1982b) 1304.
[4] P. Grassberger, A new mechanism for deterministic diffusion, Wuppertal preprint WUB 82-18.
[5] P. Grassberger, Chaos and diffusion in deterministic cellular automata, Physica 10D (1984) 52–58.
[6] D.A. Lind, Application of ergodic theory and sofic systems to cellular automata, Physica 10D (1984) 36–44.
[7] J. von Neumann, The general and logical theory of automata, in: Collected Works, J. von Neumann, A.H. Taub, ed., (1963) 288.
[8] J. von Neumann, Theory of Self-Reproducing Automata, A.W. Burks, ed. (Univ. of Illinois, Urban, 1966).
[9] M.S. Waterman, Some applications of information theory to cellular automata, Physica 10D (1984) 45–51.
[10] S. Wolfram, Statistical mechanics of cellular automata, Reviews of Modern Physics 55 (1983a) 601–644.
[11] S. Wolfram, Cellular automata, Los Alamos Science, Fall 1983 (1983b) 2–21.