When are cellular automata random?

J. B. Coe$^{13}$, S. E. Ahnert$^4$, and T. M. A. Fink$^{123}$

1 INSERM U900 and 2 CNRS UMR144, Curie Institute, Paris F-75248, France
3 Ecole des Mines de Paris, ParisTech, Fontainebleau, F-77300 France and
4 Theory of Condensed Matter, Cavendish Laboratory, Cambridge CB3 0HE, UK

A random cellular automaton is one in which a cell’s behaviour is independent of its previous states. Analytical conditions for the existence of random cellular automata are derived and we find that a multitude of non-trivial random cellular automata exist. We develop an indicator variable formalism to further investigate these random automata and confirm analytical results with simulation.

Introduction

A cellular automaton is a collection of cells, each of which is in one of a finite number of states. The cells’ states are updated according to some local rules which take into account the state of the cell and the states of its neighbours. For instance: a two dimensional grid with two states per cell is used in Conway’s Game of Life in which each cell is alive or dead and transition rules are intended to mimic the effects of birth, death from isolation and death from overcrowding [1]. Cellular automata have attracted attention from the physics community as they are an appealing modelling tool for any process driven by local interactions.

Cellular automata have been used to model a wide range of physical phenomena including: traffic flow [2–8]; disease epidemics [9, 10]; stochastic growth [11]; predator-prey dynamics [12, 13]; invasion of populations [14]; earthquakes [15] and dynamics of stock markets [16].

Analytical work has been fruitful in classifying behaviour and studying properties of specific forms. Wolfram exhaustively catalogued and classified behaviour of simple deterministic automata [17, 18]. Phase diagrams and critical exponents have been evaluated for automata with absorbing states [19–23]. Certain probabilistic automata have been shown to fall into the same universality class as directed percolation [24, 25]. Coarse-graining has been shown to predict the emergence of large scale properties even for computationally irreducible systems [26]. Fuku et. al. identified conservative automata in which the number of coloured cells remains constant [27–29].

Classification of rules is of fundamental importance in any broad analytical treatment of cellular automata. Random behaviour is among the most elementary forms of behaviour possible and must be firmly understood if a general theory of structure — deviation from randomness — is to be developed.

In deterministic cellular automata the update rules have no probabilistic component: for a given configuration of cell states the updated cell state is always the same. In probabilistic cellular automata, local rules may have a probabilistic element to them: rather than dictating the state of an updated cell, the rule gives the probability that an updated cell will be in each of a number of states.

The simplest cellular automaton is a line of cells, each of which is in one of two states: coloured or uncoloured. The cells are updated simultaneously; the updated state of a cell depends on its state and the states of its nearest neighbours. There are eight possible configurations for the states of a cell and its nearest neighbours; this gives eight different update rules which we label as $a_i$.

The configurations and rules are numbered according to Fig. 1.

We denote the state of the central cell by $x_c$: when the central cell is coloured $x_c = 1$, when it is uncoloured $x_c = 0$. The left neighbour is denoted by $x_l$ and the right neighbour by $x_r$. Uncoloured cells are denoted by $\bar{x}_l$, $\bar{x}_c$ and $\bar{x}_r$ ($x_l = 1 - x_i$). The state of the central cell after an update, $x'_c$, is given by

$$P(x'_c = 1) = a_0 \bar{x}_l \bar{x}_c \bar{x}_r + a_1 \bar{x}_l \bar{x}_c x_r + a_2 x_l \bar{x}_c \bar{x}_r + a_3 \bar{x}_l x_c \bar{x}_r + a_4 x_l \bar{x}_c x_r + a_5 \bar{x}_l \bar{x}_c x_r + a_6 x_l \bar{x}_c x_r + a_7 \bar{x}_l \bar{x}_c x_r.$$ (1)

In this paper we present the conditions for random cellular automata (which we derive in a later section).

We first solve these conditions over the space of deterministic cellular automata. We find that of the 256 simple cellular automata, 28 of them exhibit random behaviour. In all but two of them, cells are coloured with probability $\frac{1}{2}$. Ten of these random automata are not related by symmetry (reflection or inversion), all of which we illustrate in fig. 2.

Second, we solve the conditions for random behaviour for probabilistic cellular automata. Within the eight-
cubical — the corners of which are the 256 deterministic cellular automata — we find an expression for the density of random cellular automata and find two disjoint five-volumes both of which have volume 0.12. We give examples of random probabilistic cellular automata through illustrations.

In the third section we derive the conditions for randomness by applying an indicator variable formalism to nearest neighbour one-dimensional cellular automata.

We employ indicator variables in the fourth section to study correlation current which manifests itself as a visually apparent flow in many of the random automata. We develop an analytical expression for correlation current and compare predictions with simulation.

In the fifth section we show that the total randomness in random cellular automata, must come from either of two sources: spatial randomness in the initial conditions or randomness injected directly through probabilistic update rules. We derive explicit expressions to calculate these quantities for any random cellular automaton.

**Random cellular automata**

In a random cellular automaton the probability of a cell being coloured is independent of all of its previous states; all cells are taken to be coloured with probability \( p \). There are two sets of criteria on the rules; satisfying either of these produces random behaviour:

\[
\begin{align*}
(1 - p) a_0 + p a_4 &= p, \\
(1 - p) a_2 + p a_6 &= p, \\
(1 - p) a_1 + p a_5 &= p, \\
(1 - p) a_3 + p a_7 &= p;
\end{align*}
\]

\[
\begin{align*}
(1 - p) a_0 + p a_1 &= p, \\
(1 - p) a_2 + p a_3 &= p, \\
(1 - p) a_4 + p a_5 &= p, \\
(1 - p) a_6 + p a_7 &= p.
\end{align*}
\]

Note that these conditions have certain symmetry properties. Switching over left- and right-handed rules turns one set of conditions into the other. State-inversion (where coloured cells and uncoloured cells are switched round) leaves both sets of conditions unchanged.

We examine the behaviour of random automata for both deterministic and probabilistic rules. Since conditions have left-right reflective symmetry, we only look at examples for the second set of conditions, (3).

For a random cellular automaton which satisfies (3), the state of the central cell after an update is given by

\[
P(x'_r = 1) = \Pr(x_r) \left( 1 - \frac{p}{\Pr}\right) x_r \times \left( a_0 \bar{x}_1 \bar{x}_c + a_1 \bar{x}_1 x_c + a_2 \bar{x}_1 x_c + a_3 \bar{x}_1 x_c + a_4 \bar{x}_1 x_c + a_6 \bar{x}_1 x_c \right).
\]

**TABLE I:** The 10 sets of rules which give random behaviour along with their symmetric partners. The rules themselves are shown along with a simplified boolean expression for the updated state of the central cell. Wolfram numbers and Wolfram numbers of symmetric partners are given. The symmetry operations considered are reflection (R), state-inversion (I) and both (R&I). The value of \( p \) for all rule sets is 0.5 except for rule set 170, for which \( p \) can take any value.

**Random deterministic automata**

For deterministic cellular automata all the rules must be either zero or one. We are not interested by trivial solutions in which all cells are coloured \((p = 1)\) or uncoloured \((p = 0)\); this imposes limits on possible choices of \( p \). For (3) this means that if \( a_{2j} = 0 \) then \( a_{2j+1} = 1 \) and if \( a_{2j} = 1 \) then \( a_{2j+1} = 0 \). One choice of rules, where all \( a_{2j} = 0 \), gives random behaviour for arbitrary \( p \); all other choices give \( p = \frac{1}{2} \). There are 16 possible choices of free rules which give random behaviour, though some are related by symmetry.

The left-handed conditions provide an equal number of possible rule sets leading to random behaviour some of which are already given by the right-handed conditions. There are 4 possible choices which satisfy both left- and right-handed conditions for randomness giving a total of 28 distinct random deterministic cellular automata.

Wolfram assigns deterministic automata numbers according to the values of the rules [17]. These Wolfram numbers can be obtained by evaluating \( \sum a_i 2^i \). Of the 28 rule sets which give random behaviour when symmetric partners (either left-right or inversion symmetries) are discounted, there are 10 distinct rule sets. We list these rule sets and their symmetries in table I. Several of these automata belong to Wolfram’s classes 3 and 4 which exhibit the most complex kinds of behaviour [18].

Of the 10 rule sets, some can be qualitatively explained. 170 copies the right site to itself, propagating the initial sequence along space and time. 85 copies the conjugate of the right site to itself. 90 is a NAND gate on the left and right sites, and 102 is a NAND gate in the left and centre sites. 105 gives a coloured cell if its neighbourhood contains one or three zeros. 150 gives a coloured cell if its neighbourhood contains none or two zeros.
Random probabilistic automata

When the rules are no longer constrained to be zero or one we cannot count configurations so instead we consider the density of random states. As pairs of rules are related to one another by the conditions for random behaviour, there are four free rules and a probability of a cell being coloured, \( p \), to choose. If these free parameters were each randomly picked from the interval zero to one it would not necessarily result in random behaviour. The non-free rules must also lie between zero and one which may not be compatible with (2) and (3), the conditions for random behaviour.

Within the five-volume marked out by the free rules and \( p \), there is a density of random configurations given by

\[
dV_5 = \left( \frac{p}{1-p} \right)^4 \, dp, \quad p \leq \frac{1}{2}.
\]

\[
dV_5 = \left( \frac{1-p}{p} \right)^4 \, dp, \quad p > \frac{1}{2}.
\]

Integrating over all possible values of \( p \), we find

\[
V_5 = 17/3 - 8 \ln 2 \simeq 0.12. \tag{6}
\]

Within the five-volume, 12% of possible selections of free rules and colouring probabilities will lead to random behaviour. A second five-volume exists when the other set of conditions is satisfied. Note that the five-volumes that correspond to different sets of conditions are not the same (different rules are free) and have zero intersection in five dimensions.

Satisfying both sets of conditions gives a three-volume within which 23% of randomly picked points correspond to random behaviour.

Imposing left-right symmetry by satisfying both sets of conditions, (2) and (3), yields

\[
P(x' = 1) = 1 - \bar{x}_l \bar{x}_r - x_l x_r \tag{7}
\]

\[
+ (a_0 \bar{x}_l + a_2 x_r)(\bar{x}_r - x_r)(\bar{x}_l - x_l),
\]

which is symmetric under interchange of \( x_l \) and \( x_r \) as expected. In this expression the state of the central cell and its left and right neighbours are all important.

Some illustrative examples, all with \( p = 0.5 \), of random probabilistic cellular automata are shown in Fig. 3. A Java program [31] can be freely downloaded and used to generate figures for probabilistic cellular automata.

Indicator variables

Indicator variables are taken from the theory of random processes and have been used to solve the asymmetric exclusion process [32]. In a powerful formalism in population genetics, indicator variables are used to keep track of genes and correlations between genes [33]. Here we make use of indicator variables to represent the state of a cell and use cancellation properties to simplify expressions for high-order correlations.

Every cell in a one-dimensional cellular automaton can be assigned a unique integer index increasing from left to right. We represent the state of each cell by an indicator variable \( \tau_i \), where \( i \) is the index of the site. The indicator variable takes the value 1 if a site is coloured and 0 if it is not. For notational brevity we define an anti-indicator \( \bar{\tau}_i \) which shows if a cell is uncoloured: \( \bar{\tau}_i = 1 - \tau_i \).

If \( p_i \) is the probability that cell \( i \) is coloured, then the expected value of the indicator variable \( \tau_i \) is \( p_i \). We represent this as

\[
E[\tau_i] = p_i. \tag{8}
\]

Several identities follow from the above definitions and can be used to simplify complicated expressions in which indicator variables appear:

\[
E[\tau_i \tau_j] = E[\tau_i] \tag{9}
\]

\[
E[\tau_i \bar{\tau}_j] = E[\bar{\tau}_j] \tag{10}
\]

\[
E[\tau_i \bar{\tau}_j] = E[\tau_i] - E[\tau_i \tau_j]. \tag{11}
\]

Correlations between indicators and anti-indicators can always be expressed as correlations between indicators by exploiting the identity \( \bar{\tau}_i = 1 - \tau_i \):

\[
E[\bar{\tau}_i \tau_j] = 1 - E[\tau_i] - E[\tau_j] + E[\tau_i \tau_j] \tag{10}
\]

\[
E[\bar{\tau}_i \bar{\tau}_j] = E[\tau_i] - E[\tau_i \tau_j]. \tag{11}
\]
For an unknown configuration of cells, the probability that the cell with index 1 is coloured after an update follows from the definitions of the update rules: each possible neighbourhood can be described by a three-fold product of indicator variables; each neighbourhood has a probability $a_i$ of producing a coloured site after an update. We write the indicator variable for cell $i$ after an update as $\tau'_i$.

$$E[\tau'_i] = a_0E[\tau_0\tau_1\tau_2] + a_1E[\bar{\tau}_0\tau_1\tau_2] + a_2E[\tau_0\bar{\tau}_1\tau_2] + a_3E[\tau_0\tau_1\bar{\tau}_2] + a_4E[\tau_0\bar{\tau}_1\bar{\tau}_2] + a_5E[\bar{\tau}_0\tau_1\bar{\tau}_2] + a_6E[\bar{\tau}_0\bar{\tau}_1\tau_2] + a_7E[\tau_0\tau_1\bar{\tau}_2].$$  

(12)

Without assumptions this cannot be simplified.

**Conditions for random behaviour**

A random configuration is one in which the probability of any given cell being coloured is uniformly and independently distributed through time. As the probability of a cell being coloured can be expressed in terms of neighbouring cells being coloured in the previous time step (12), for a cell to be independent of its previous states requires inter-cell independence at any time step. Single-cell temporal independence requires system wide spatial independence at any timestep. If the cells are coloured with probability $p$, a random configuration can be expressed in terms of the indicator algebra as

$$E[\tau'_i] = p,$$  

(13)

$$E\left[ \prod_{i \in S} \tau'_i \right] = \prod_{i \in S} E[\tau'_i].$$  

(14)

The first condition, (13), can be written in terms of indicator variables and simplified using (14). This results in a polynomial in $p$.

$$p = a_0q + a_1pq + a_2pq^2 + a_3pq^2q + a_4pq^2 + a_5pq^2q + a_7pq^3,$$

(15)

where $q = 1 - p$. While this can be satisfied for many choices of $a_i$ and $p$, the second condition (14) is much stricter and any set of rules satisfying (14) will satisfy this polynomial.

To satisfy (14) no inter-cell correlations can exist for any possible set of cells. For certain choices of $a_i$, larger correlation terms can be expressed in terms of shorter ones: repeating this procedure reduces all correlation terms to products of single cell expectations. This reduction of correlation terms can be performed in one of two ways:

$$E\left[ \prod_{i \in S} \tau_i \right] = E[\tau_0]E\left[ \prod_{i \in S \setminus \{0\}} \tau_i \right],$$

(16)

$$E\left[ \prod_{i \in S} \tau_i \right] = E\left[ \prod_{i \in S \setminus \{N\}} \tau_i \right]E[\tau_N].$$

(17)

where $N$ and 0 have been chosen, without loss of generality, to be the largest and smallest indices in the set $S$. We label $S$ without the index 0 as $S \setminus \{0\}$ and $S$ without the index $N$ as $S \setminus \{N\}$. We refer to removal of the smallest index as left-handed reduction and removal of the largest index as right-handed reduction as the removed index corresponds to the right-most or left-most cell in the set.

For reduction of a correlation, the most restrictive set of conditions exists when the index to be removed from the set is among three neighbouring cells. Adding further cells to the set will not introduce more conditions as there is no overlap between the neighbourhoods of the cell to be removed from the set and any other cells that may be added. As reduction can be performed from the left or right, there are two possible choices

$$E[\tau_0\tau_1\tau_2'] = E[\tau_0']E[E[\tau_1\tau_2]],$$

(18)

or

$$E[\tau_0\tau_1\tau_2'] = E[\tau_0\tau_1']E[\tau_2].$$

(19)

Substituting in indicator variable expressions for all of the $\tau'_i$ into these two expressions leads directly to (2) and (3), the conditions for random behaviour.

**Correlation current**

Some of the random automata we have seen show a flow of some sort. Rather than injecting randomness through probabilistic rules, rule 170 copies the state of an adjacent cell which is guaranteed to be random and independent.
of the history of the updated cell. Randomness requires
that a cell’s state is independent of its previous states
so this copying is allowed. The flow seen in many of
the automata can be quantified by defining a spatio-temporal
relation between an updated cell and its left- or right-sided
neighbour before the update. We define this as
correlation current.

The magnitude of right-handed flow is given by
\[ E[\tau_{\tau_1}' - p^2] = (a_4 - p)(1 - p)^2 + (a_5 - p)p^2(1 - p) + \left( a_6 - p \right)p^2(1 - p) + \left( a_7 - p \right)p^3, \]  
(20)
and left-handed by
\[ E[\tau_{\tau_0}' - p^2] = (a_1 - p)(1 - p)^2 + (a_3 - p)p^2(1 - p) + \left( a_5 - p \right)p^2(1 - p) + \left( a_7 - p \right)p^3. \]  
(21)

Correlation current can be left- or right-handed depending
on which neighbour is considered, and can be
positive or negative. The sign and direction are unconnected,
and a negative right-handed correlation current and a
positive left-handed correlation current are not the same
thing. A negative correlation current is anti-correlation
a positive one is correlation.

Random cellular automata permit right-handed cor-
relation current or left-handed correlation current, but
not both. Satisfying both sets of conditions, (2) and
(3), means that no correlation current exists. Analytical
predictions for correlation current are compared with
simulation results in Fig. 4.

Information flow

The information flow through a channel with input \( Z \)
and output \( Y \) is given by the mutual information \( I(Z; Y) \):
\[ I(Z; Y) = H(Y) - H(Y|Z). \]  
(22)

Where \( H(Z) \) is the Shannon entropy of the random variable
\( Z \) and the conditional entropy \( H(Y|Z) \) is given by
\( H(Y|Z) = H(Z, Y) - H(Z) \) \[34\].

We consider the eight possible cellular automata
neighbourhoods as our \( Z \). We label them in a similar manner
to the rules so that \( z_0 = 000, z_1 = 001, ..., z_7 = 111 \). In
random cellular automata, the probability of one of these
neighbourhoods occurring is given by \( P(z_i) = p^{w_i}q^{3-w_i} \),
where \( w_i \) is the number of coloured cells \( z_i \) and \( q = 1 - p \).

In neighbourhood \( i \), a coloured cell is produced with
probability \( a_i \) and an uncoloured cell with probability
\( 1 - a_i \), which we denote by \( \bar{a}_i \). The information flow
\( F(p, a) \) for a random automaton is given by:
\[ F(p, a) = \sum_i P(z_i) \left( a_i \log_2 \frac{a_i}{p} + \bar{a}_i \log_2 \frac{\bar{a}_i}{q} \right). \]  
(23)

We can write the total entropy of the system as
\[ H_{total} = H_{flow} + H_{rules}, \]  
(24)
where the various entropies are given by:
\[ H_{total} = -p \log_2 p - q \log_2 q, \]  
(25)
\[ H_{flow} = F(p, a), \]  
(26)
\[ H_{rules} = \sum_i -p(z_i)(a_i \log_2 a_i + \bar{a}_i \log_2 \bar{a}_i). \]  
(27)

This relationship tells us how much of the total entropy
\( H_{total} \) at a particular cellular automata site arises from
the probabilistic nature of the set of rules \( H_{rules} \) and
how much is due to information flow from the neighbour-
ning cells \( H_{flow} \).

Conclusion

We have presented analytical conditions for the existence
of random cellular automata with both probabilistic
and deterministic rules. The techniques developed,
in addition to giving a deeper understanding of random
behaviour, can be more broadly applied to analytical
solution of cellular automata models.

Randomness is equivalent to the mean-field approxi-
mation employed by [17] and [19] which was extended to
a broader approximation known as local structure theory
[35]. Within limits, both approximations have had con-
siderable success. Analytical limits on the validity of ap-
proximation will allow more sophisticated approximation
 techniques to be developed and applied appropriately.

The indicator variable approach used is readily gen-
eralisable to larger neighbourhoods and dimensionalities
allowing the techniques employed in this paper to be used
to find conditions for random behaviour in larger, more
complex automata.
Acknowledgements

This work was funded in part by the Defense Advanced Research Projects Agency (DARPA) Fundamental Laws of Biology (FunBio) grant HR 0011-05-1-0057. J.B.C. is supported by FunBio. S.E.A. is supported by a Leverhulme Trust Fellowship.

[30] The eight rule sets used to generate the random probabilistic cellular automata, shown in Fig. 3, listed from $\sigma_7$ to $\sigma_0$ are:
0.833, 0.167, 0.698, 0.302, 0.937, 0.063, 0.968, 0.032;
0.018, 0.982, 0.749, 0.251, 0.982, 0.018, 0.251, 0.749;
0.639, 0.361, 0.639, 0.361, 0.639, 0.361, 0.639, 0.361;
0.886, 0.114, 0.886, 0.114, 0.031, 0.969, 0.031, 0.969;
0.240, 0.760, 0.964, 0.036, 0.240, 0.760, 0.964, 0.036.