Apparent Entropy of Cellular Automata

Bruno Martin
Laboratoire d'Informatique et du Parallélisme,
École Normale Supérieure de Lyon,
46, allée d'Italie,
69364 Lyon Cedex 07, France

We introduce the notion of apparent entropy on cellular automata that points out how complex some configurations of the space-time diagram may appear to the human eye. We then study, theoretically, if possible, but mainly experimentally through natural examples, the relations between this notion, Wolfram's intuition, and almost everywhere sensitivity to initial conditions.

Introduction

A radius-r one-dimensional cellular automaton (CA) is an infinite sequence of identical finite-state machines (indexed by $\mathbb{Z}$) called *cells*. Each finite-state machine is in a state and these states change simultaneously according to a local transition function: the following state of the machine is related to its own state as well as the states of its $2r$ neighbors. A configuration of an automaton is the function which associates to each cell a state. We can thus define a global transition function from the set of all the configurations to itself which associates the following configuration after one step of computation.

Recently, a lot of articles proposed classifications of CAs [5, 8, 13] but the canonical reference is still Wolfram's empirical classification [14] which has resisted numerous attempts of formalization. Among the latest attempts, some are based on the mathematical definitions of chaos for dynamical systems adapted to CAs thanks to Besicovitch topology [2, 6] and [11] introduces the almost everywhere sensitivity to initial conditions for this topology and compares this notion with information propagation formalization. As in [8], this notion does not really classify the CA but the CA for a measure. This gives, for instance, a tool to understand fluid flow phenomenon modeled by CA: we can now say that the CA is not almost everywhere sensitive to initial conditions (and thus almost everywhere not sensitive) for small fluid speed and almost everywhere sensitive for high fluid speed.

It appears that almost everywhere sensitivity, although close to, does not really match Wolfram's intuitive idea of chaos. To better understand why, we will here introduce a new notion that points out how complex some configurations of the space-time diagram may appear to the human eye. This notion is based on information theory, the idea is to take into account the fact that the human eye cannot detect correlations in very big patterns. We then study, theoretically, if possible, but mainly experimentally, the relations between this notion, Wolfram's intuition, and almost everywhere sensitivity. As we will see, this study is based on Besicovitch topology, Bernoulli measure, and apparent entropy. For simplicity, we only consider one-dimensional CA. However, all the concepts we introduce are topological and they could easily be extended to higher-dimensional CA.

1. Definitions

1.1 Cellular automata

A *radius-r one-dimensional cellular automaton* is a couple (Q, δ) where Q is a finite set of states and $\delta : Q^{2r+1} \longrightarrow Q$ is a transition function. A configuration $c \in Q^{\mathbb{Z}}$ of (Q, δ) is a function from $\mathbb{Z}$ into Q and its global transition function $G_\delta : Q^{\mathbb{Z}} \longrightarrow Q^{\mathbb{Z}}$ is such that $G_\delta(c)(i) = \delta(c(i-r), \ldots, c(i), \ldots, c(i+r))$. An *elementary cellular automaton* (ECA) is a radius-1 two states (usually 0 and 1) one-dimensional CA.

For ECA, we will use Wolfram's notation: they are represented by an integer between 0 and 255 such that the transition function of the CA number i whose writing in base 2 is $i = \overline{a_7 a_6 a_5 a_4 a_3 a_2 a_1 a_0}^2$ satisfies:

$$\begin{array}{ll} \delta_i(0,0,0) = a_0 & \delta_i(1,0,0) = a_4 \\ \delta_i(0,0,1) = a_1 & \delta_i(1,0,1) = a_5 \\ \delta_i(0,1,0) = a_2 & \delta_i(1,1,0) = a_6 \\ \delta_i(0,1,1) = a_3 & \delta_i(1,1,1) = a_7. \end{array}$$

Let us remark that CA with different numbers may have the same behavior by exchanging the states 0 and 1; for instance, $184 = \overline{10111000}^2$ and $226 = \overline{11100010}^2$. If r is a rule number, we will denote $\overline{r}$ the rule after exchanging the states and $\overleftrightarrow{r}$ the rule which has a symmetric behavior (see [4] for more details).

We will discuss the CA $120 = (\{0,1\}, \delta_{120})$, or equivalently, of the rule 120.

In the general definition of additive CA due to Wolfram, an additive CA is a CA that satisfies the superposition principle $(\delta(x+x', y+y', z+z') = \delta(x,y,z) + \delta(x',y',z'))$. These CA are very interesting to choose examples from, because their behaviors obey algebraic rules adapted to a formal study while their space-time diagrams appear complicated. We will here, as in [9, 13], use a strictly more restrictive definition.

Definition 1. We call *additive CA* a one-dimensional CA whose state set is $\mathbb{Z}/n\mathbb{Z}$ and whose transition function is of the form:

$$\delta(x_{-1}, x_0, x_1) = x_0 + x_1 (\mathrm{mod}\ n).$$

1.2 Besicovitch topology

The most natural topology on CA configuration sets is Cantor topology. The problem is that, even if the product topology is shift invariant, the associated distance is not and it focuses on the neighborhood of an (arbitrary) origin while in many applications of CA all the cells have the same importance. This is the reason Formenti suggests Besicovitch topology in the CA context. As noticed in [11], notions we will use are not specific to Besicovitch topology, but are true for a wide class of topologies including the Weyl one, and although there are many ways to extend Besicovitch in higher dimensional grids, all of them are equivalent for our purpose.

Let us define the Besicovitch pseudodistance, which induces a shift invariant topology on the quotient space.

Definition 2. The Besicovitch pseudometric on $Q^{\mathbb{Z}}$ is given by

$$d(c, c') = \limsup_{l \longrightarrow +\infty} \frac{\#\{i \in [-l, l] | x_j \neq y_j\}}{2l + 1} \quad \text{(\# denotes the cardinality).}$$

It is not a distance since obviously the distance between two configurations equal everywhere except on a finite number of cells is null. If we consider two equal configurations except at the cells 2^n, there is an infinite number of differences, but the distance is still null.

Property 1. Supplied with the induced topology, the quotient of $Q^{\mathbb{Z}}$ by the relation $x \sim y \iff d(x, y) = 0$ is metric, path-wise connected, infinite dimensional, complete, neither separable nor locally compact [2]. Furthermore, $x \sim y \Longrightarrow G_\delta(x) \sim G_\delta(y)$ and the transition function of a CA is a continuous map from $Q^{\mathbb{Z}}/\sim$ to itself.

The last property of continuity justifies the attempt to use ergodic theory in this context.

1.3 Measure on the configuration set

Intuitively, a measure describes what is a random element; that is, for our purpose, what is a random configuration. Let Q be a finite alphabet with at least two letters. $Q^* = \cup_{n \geq 1} Q^n$ is the set of finite words on Q. The ith coordinate $x(i)$ of a point $x \in Q^{\mathbb{Z}}$ will also be denoted x_i and $x_{[j,k]} = x_j \ldots x_k \in Q^{k+1-j}$ is the segment of x between indices j and k. The cylinder of $u \in Q^p$ at position $k \in \mathbb{Z}$ is the set $[u]_k = \{x \in Q^{\mathbb{Z}} | x_{[k,k+p-1]} =$

$u\}$. Let σ be the shift toward the left: $\sigma(c)_i = c_{i+1}$ (i.e., the rule number 85).

A Borel probability measure is a nonnegative function μ defined on Borel sets. It is given by its values on cylinders, satisfies $\mu(Q^{\mathbb{Z}}) = 1$, and for every $u \in Q^*$, $k \in \mathbb{Z}$,

$$\sum_{q\in Q} \mu[uq]_k = \mu[u]_k \text{ and } \sum_{q\in Q} \mu[qu]_k = \mu[u]_{k+1}.$$

A measure μ is σ-invariant if $\mu[u]_k$ does not depend on k (and will thus be denoted $\mu[u]$). A σ-invariant measure is σ-ergodic if for every invariant measurable set Y ($\sigma(Y) = Y$), either $\mu(Y) = 0$ or $\mu(Y) = 1$.

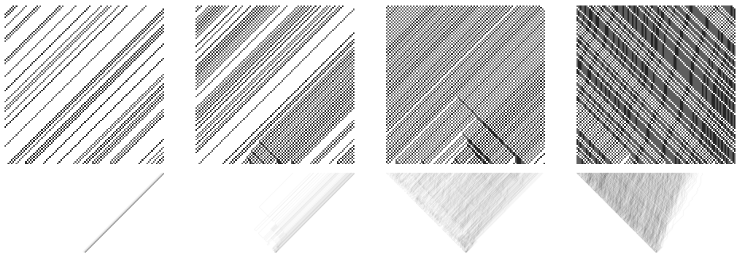
Bernoulli measures are the simplest, the idea is that all the cells are independent, and the probability of each state q is given by a constant p_q. Their simplicity is the reason why we will use them in all our examples while the definitions remain true for other σ-ergodic measures; for instance, Markov measures (with correlations over a finite number of cells) or measures such that the correlation between two states decreases exponentially with their distance.

A Bernoulli measure is defined by a strictly positive probability vector $(p_q)_{q\in Q}$ with $\sum_{q\in Q} p_q = 1$ and if $u = u_0 \ldots u_{n-1} \in Q^n$, $\mu[u_0 \ldots u_{n-1}] = p_{u_0} \ldots p_{u_{n-1}}$.

Let us note the following classical result: *Bernoulli measures are σ-ergodic.*

For 2-states CA, Bernoulli measures will be denoted μ_ρ where $\rho = p_1 = 1 - p_0$ is the probability that a state is 1.

To illustrate Bernoulli measure and introduce the almost everywhere sensitivity to initial conditions, let us consider the example given in [11] that we reuse below. In Figure 1, we see a very simple traffic model (the CA T) based on rule 184 but with two different models of cars. The principle is quite simple, a car stays at the same position if there is a car in front of it and moves toward the right if the next position is free. The system seems "chaotic" when the density ρ of cars is greater than or equal to 0.5 because of the traffic jams, but not "chaotic" otherwise. Saying that this CA is chaotic or not does not make sense since it will depend on its utilization: whether it is used for traffic jam or for fluid traffic simulation. Its average behavior makes no sense if we do not explain what is a "random configuration," that is, which measure we take on its configuration set. If we assume that the cars repartition is initially uniform and that we have the same number of red and blue cars, we will consider the Bernoulli measures μ^*_ρ such that the probability to find a blue car in a cell is $\rho/2$ and equal to the probability to find a red car while the probability that there is no car is $1 - \rho$. The idea is to say that this CA is μ^*_ρ-almost everywhere sensitive to initial conditions when $\rho \geq 1/2$ while it is μ^*_ρ-almost never sensitive to initial conditions



$\rho = 0.2$ $\rho = 0.4$ $\rho = 0.5$ $\rho = 0.7$

Figure 1. The CA T is a very simple traffic model based on rule 184 but with two different models of cars. The system seems chaotic when the density ρ of cars is greater than or equal to 0.5 because of the traffic jams, but not chaotic otherwise. Below the space-time diagrams (time goes toward the top), we show with a gray level the space-time repartition of the average number of alterations induced by the modification of the middle cell.

otherwise. If it is important to take into account the fact that a lot of people take their car at the same time to go to work, other measures allow modeling a nonuniform repartition.

1.4 Dμ-attracting sets

Dμ-attracting sets have been defined in [9], we will use them to point out the homogenization process to periodical configurations, some examples are presented in Figure 2.

Definition 3. A subshift is any subset $\Sigma \subseteq Q^{\mathbb{Z}}$, which is σ-invariant and closed in the Cantor metric. The language $L(\Sigma)$ of a subshift $\Sigma \subseteq Q^{\mathbb{Z}}$, is the set of factors of Σ. A subshift is of finite type (SFT), if there exists a positive integer p called *order*, such that for all $c \in Q^{\mathbb{Z}}$, $c \in \Sigma \iff \forall i \in \mathbb{Z}, c_{[i,i+p-1]} \in L(\Sigma)$.

Definition 4. For a subshift $\Sigma \subseteq Q^{\mathbb{Z}}$ and $x \in Q^{\mathbb{Z}}$, define the density of Σ-defects in x by

$$d_D(x,\Sigma) = \lim_{k\longrightarrow+\infty} \limsup_{l\longrightarrow+\infty} \frac{\#\{i \in [-l,l] \mid x_{[i,i+k-1]} \notin L(\Sigma)\}}{2l+1}.$$

Notation 1. When d is a pseudodistance, we can naturally define the pseudodistance from an element x to a set as the inf of the pseudodistance between x and the elements of the set:

$$d(x,\Sigma) = \inf_{y\in\Sigma} d(x,y).$$

Note that $d_D(x,\Sigma)$ is not associated to any pseudodistance because in the general case $d_D(x,\{y\}) \neq d_D(y,\{x\})$.

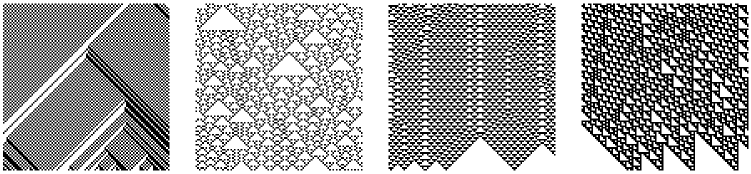


Figure 2. This figure presents the space-time diagram of four ECA, that is, we use the vertical axes to represent their successive configuration (time goes up). $\{(01)^*, (10)^*\}$ is a Dμ-attracting subshift of rule 184 for $\mu_{1/2}$ [9]. [11] conjectures that $\{(1000)^*, (0100)^*, (0010)^*, (0001)^*, (1110)^*, (1101)^*, (1011)^*, (0111)^*\}$ is a Dμ-attracting subshift of rule 54 and that $\{c \in \{0,1\}^{\mathbb{Z}} | \forall i, c(2i) = 0\} \cup \{c \in \{0,1\}^{\mathbb{Z}} | \forall i, c(2i+1) = 0\}$ could be a nonfinite type subshift of rule 18. The latest results of N. Ollinger on rule 110 let us think that its behavior is similar to rule 54 behavior.

Definition 5. Let μ be a σ-ergodic measure, a subshift Σ is Dμ-attracting if

$$\mu(\{c \in Q^{\mathbb{Z}} \mid \lim_{n \to +\infty} d_D(G_\delta^n(c), \Sigma) = 0\}) = 1.$$

Remark 1. As the set $\{c \in Q^{\mathbb{Z}} \mid \lim_{n \to +\infty} d_D(f^n(c), \Sigma) = 0\}$ is shift invariant, its measure is either 0 or 1.

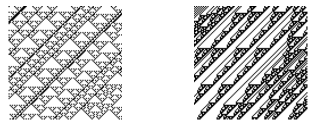
2. μ-almost everywhere sensitivity and apparent entropy

2.1 μ-almost everywhere sensitivity to initial conditions

The definition of almost everywhere sensitivity to initial conditions is not obtained by replacing "for all configurations c" with "for μ-almost all configurations c" in the sensitivity definition, it is a bit more restrictive, so that a CA that is not μ-almost everywhere sensitive, is "μ-almost nowhere" sensitive to initial conditions.

Definition 6. [11] A CA is *μ-almost everywhere sensitive to initial conditions* (for Besicovitch pseudodistance) if there exists M such that for all $\epsilon_0 > 0$, there exists $\epsilon < \epsilon_0$ such that if c and c'' are two μ-random configurations, if e is a μ_ϵ random configuration and if $c' = c\bar{e} + c''e$ is the configuration whose state at a given position is equal to the corresponding state of c when the corresponding state of e is equal to 1 and to the corresponding state of c'' otherwise (see Figure 3), then with probability 1 (for $\mu \times \mu \times \mu_\epsilon$) there exists n such that $d(G_{\delta_{\mathcal{A}}}^n(c), G_{\delta_{\mathcal{A}}}^n(c')) \geq M$.

μ-random configuration												
a	b	c	c	a	c	b	b	b	c	a	b	c
μ-random configuration												
b	c	b	a	a	b	c	a	a	c	c	b	c''
μ_ϵ-random configuration												
0	1	0	0	0	0	1	0	1	0	0	0	e
a	c	c	c	a	c	c	b	a	c	a	b	$c' = c\bar{e} + c''e$



210 120

Figure 3. The configuration $c' = c\bar{e} + c''e$ is the configuration whose state at a given position is equal to the corresponding state of c when the corresponding state of e is equal to 0 and to the corresponding state of c'' otherwise. Remark that, due to the great number law, with probability 1, $d(c, c') = \epsilon d(c, c'') \leq \epsilon$. Rule 210 is μ-almost nowhere sensitive to initial conditions while we guess that rule 120 is μ-almost everywhere sensitive.

Remark 2.

- If c and c'' are two generic configurations for μ and e is a generic configuration for μ_ϵ, it is easy to see that $c' = c\bar{e} + c''e$ is generic for μ. Furthermore, due to the great number law, with probability 1, $d(c, c') = \epsilon d(c, c'') \leq \epsilon$.
- This definition implies that there exists M such that for μ-almost all configurations c and for all $\epsilon > 0$, there exists c' and n with $d(c, c') < \epsilon$ and $d(G^n_{\delta_\mathcal{A}}(c), G^n_{\delta_\mathcal{A}}(c')) \geq M$.
- Usually, when the space is compact, the previous result implies the sensitivity to initial conditions. The main point of the proof is that any compact subset has a nonnull measure, but the Besicovitch topological space $Q^\mathbb{Z}/\sim$ is not (locally) compact and the argument fails.
- The set of configurations 3-uplets (c, c'', e) such that if $c' = c\bar{e} + c''e$ there exists n so that $d(G^n_{\delta_\mathcal{A}}(c), G^n_{\delta_\mathcal{A}}(c')) \geq M$ is obviously shift invariant on $(Q \times Q \times \{0, 1\})^\mathbb{Z}$. As $\mu \times \mu \times \mu_\epsilon$ is σ-ergodic, thus the set measure is either 1 or 0. As a consequence a CA is either μ-almost everywhere sensitive to initial conditions or "μ-almost nowhere sensitive to initial conditions:" for any η there exists ϵ such that if we build c, c' as usual, for any n, $d(G^n_{\delta_\mathcal{A}}(c), G^n_{\delta_\mathcal{A}}(c')) \leq \eta$ almost everywhere.

The μ-almost everywhere sensitivity to initial conditions makes sense because we saw that some CA are not (obviously rule 0 is not) and the additive CA are [11].

2.2 Apparent entropy

By looking at space-time diagrams starting from initial configurations for different densities of 1, Wolfram [14] observed that the rules he termed "chaotic" present a complex behavior and that this behavior does not (asymptotically) depend on the initial way of choosing a random configuration. Our aim here is to give sense to this intuition. The idea is to use information theory to express this. A very first idea would be to consider the amount of information contained in the space-time diagrams of the CA, but obviously, the amount of information of the whole diagram is the same as the amount of information in the initial configuration. This has been pointed out many times (at least in [6, 14]). If we think about information theory in terms of Kolmogorov complexity, it is obvious that a very simple program is able to build a space-time diagram knowing the initial configuration and the transition table (which is finite). A second idea is to consider the evolution of the amount of information of the successive configurations [10, 14]. But again it is easy to prove that the amount of information cannot grow (if $t_1 \leq t_2$ then $S_\mu^{(t_1)}(\mathcal{A}) \geq S_\mu^{(t_2)}(\mathcal{A})$ where $S_\mu^{(t)}(\mathcal{A})$ is the entropy of the configuration after t steps, see the definition below), and contradicts our intuitive idea. Again, this is obvious if we think of Kolmogorov complexity but let us recall the proof for the metric entropy [10].

Definition 7. Let $(\mathcal{A}, \delta)$ be a CA and μ a σ-ergodic measure, the *metric entropy* of its configuration after t computation steps is defined as follows:

$$S_\mu^{(t)}(\mathcal{A}) = \lim_{n\to+\infty} \frac{-\sum_{u\in Q^n} p_u^{(t)} \log(p_u^{(t)})}{n}$$

with the usual convention $0\times\log(0) = 0$ and where p_u is the probability of the pattern u appearing in the configuration c: $p_u^{(t)} = G_\delta^t(\mu)([u]_0)$ where the notation $f(\mu)$ represents the measure defined by $f(\mu)(X) = \mu(f^{-1}(X))$. In mathematical terminology, $S_\mu^{(t)}(\mathcal{A})$ is the metric entropy of σ for the measure $G_\delta^t(\mu)$.

Theorem 1. Let $\mathcal{A}$ be a CA and μ a measure on its configuration set, if $t_1 \leq t_2$ then $S_\mu^{(t_1)}(\mathcal{A}) \geq S_\mu^{(t_2)}(\mathcal{A})$.

Proof. [10] For convenience, let us define

$$S_{\mu,n}^{(t)} = -\sum_{u\in Q^n} p_u^{(t)} \log(p_u^{(t)})$$

where $p_u^{(t)}$ is the probability of the pattern u appearing in the configuration after t steps of computation. The probability of sequences at time t and $t+1$ are related since

$$p_u^{(t+1)} = \sum_{\{v|\{\delta_{\mathcal{A}}^{n+2r\to r}(v)=u\}} p_v^{(t)}$$

which gives the entropy at time $t+1$

$$S_{\mu,n}^{(t+1)} = -\sum_{u\in Q^n} p_u^{(t+1)} \log(p_u^{(t+1)}) \leq S_{\mu,n+2r}^{(t)}.$$

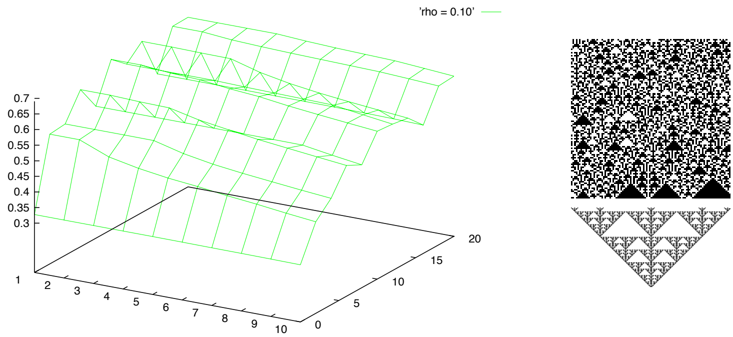


Figure 4. This diagram represents, for the first 20 configurations of the additive rule 150 starting from a random configuration for $\mu_{0.9}$, the ith, $1 \leq i \leq 10$ term of the limit that defines the metric entropy, that is, $-\sum_u p_u \log(p_u)$ when $u \in \{0,1\}^i$ and p_u is the probability of the pattern u of length i. We see that for all i the superior limit is log(2), this means that the probability tends to be the same for all the patterns of length i.

So the increase of measure entropy for one step is

$$\Delta S_\mu^{(t)}(\mathcal{A}) = S_\mu^{(t+1)}(\mathcal{A}) - S_\mu^{(t)}(\mathcal{A}).$$

An explicit calculation gives

$$\begin{aligned}\Delta S_\mu^{(t)}(\mathcal{A}) &= \lim_{n \longrightarrow +\infty} \left(\frac{1}{n} S_{\mu,n}^{(t+1)}(\mathcal{A}) - \frac{1}{n+2r} S_{\mu,n+2r}^{(t)}(\mathcal{A}) \right) \\ &= \lim_{n \longrightarrow +\infty} \left(\frac{1}{n+2r} (S_{\mu,n}^{(t+1)}(\mathcal{A}) - S_{\mu,n+2r}^{(t)}(\mathcal{A})) \right. \\ &\qquad \left. + \left(\frac{1}{n} - \frac{1}{n+2r} \right) S_{\mu,n}^{(t+1)}(\mathcal{A}) \right) \leq 0\end{aligned}$$

since the first term in the limit is negative or zero and the second term goes to zero as $m \longrightarrow +\infty$. Thus measure entropy cannot increase when a deterministic rule is applied. ■

Therefore the entropy cannot grow, but it is possible that the limit that defines it converges more and more slowly in the successive configurations. That means we need to look at bigger and bigger patterns to detect correlations between the states of the configuration. This has been noticed in [10] and represented for an additive one-dimensional CA (Wolfram rule 150). In Figure 4 we can see the correlation on patterns of size 0 to 10 of the first 20 configurations of the evolution of rule 150 on an initial configuration whose density of 1 is 1/10. We see very quickly that there is almost no correlation in the patterns of size less than 10, with some exceptions that are specific to the additive rules at time 2^n.

Note that this observation explains why the CA we want to call chaotic have a complex space-time diagram; or more exactly, complex configurations, most of the time. Actually, when we represent a space-time diagram, we can only represent a finite part of it, so as soon as the size of the correlation is larger than the represented part, we are unable to detect these correlations, and we cannot distinguish it from a random configuration.

To quantify the limit sup of the configuration complexity, we will define the apparent entropy which is obtained by inverting the limits.

Definition 8. Let $\mathcal{A}$ be a CA and μ a Bernoulli measure, we define the *apparent entropy* as follows:

$$S_\mu^a(\mathcal{A}) = \lim_{n\to+\infty} \limsup_{t\to+\infty} \sum_{u\in Q^n} p_u^{(t)} \log(p_u^{(t)}) = \lim_{n\to+\infty} \limsup_{t\to+\infty} S_{\mu,n}^{(t)}$$

where $p_u^{(t)}$ is the probability of the pattern u appearing in the configuration after t computation steps.

Remark 3.

- Note that $\lim_{n\to+\infty} \limsup_{t\to+\infty} S_{\mu,n}^{(t)}$ exists because the sequence $(\limsup_{t\to+\infty} S_{\mu,n}^{(t)})_n$ is subadditive.
- In mathematics, the following equivalent definition of apparent entropy would probably be preferred. Let us consider the topology associated to the simple limit on the set of measures; that is, such that a sequence $(\mu_n)_n$ of measures converges if and only if for any pattern u, the sequence $(\mu_n)_n(u)$ converges, obviously the set of measures is compact (as a product of compacts). Thus, the sequence of measures $(G_\delta^n(\mu))_n$ has adherence point(s). Actually, the apparent entropy is the limit sup of the entropy of the adherence points.

Our definitions have to be linked with Ferrari *et al.* [7] which considers for some additive CA (when the number of states is p^s with p prime) the Cesàro mean of the successive configurations distribution starting from a l-step Markov measure μ

$$\mathcal{M}_\mu = \lim_{M\longrightarrow+\infty} \frac{1}{M} \sum_{m=0}^{M-1} \mu \circ G_\delta^{-m}$$

and proves that it exists and is the product of uniform measures (i.e., it is the measure of maximal entropy). Obviously, this implies that the apparent entropy of additive CA is maximal (i.e., $\log(|Q|)$). Furthermore, it proves that most configurations will look completely random. For this reason defining apparent entropy as the entropy of the Cesàro mean seems more interesting: it represents the complexity of almost all the configurations rather than its limit sup. The problem is that we do not know whether it is defined for any CA and the definition we take is more natural with respect to the notion of almost everywhere sensitivity.

2.3 Apparent entropy: Exact calculation and approximation

2.3.1 Exact calculation of the apparent entropy

Unfortunately, we cannot compute the apparent entropy for most of the rules. There are some exceptions, like the rule for which we can prove that the configuration tends μ-almost surely to be uniform (like Wolfram's rules 0, 1 ...). In this case, the apparent entropy is equal to 0 for any ρ. For the identity or the shifts, obviously, the apparent entropy is equal to the initial configuration entropy. Moreover, we saw that the additive rules have a maximal apparent entropy for any nontrivial Bernoulli measures when $n = p^s$ with p prime (this is a consequence of [7]).

We can also obtain some information about the apparent entropy of a rule when we know a Dμ-attracting set because of the following theorem.

Theorem 2. If Σ is a Dμ-attracting SFT for $\mathcal{A}$, then the apparent entropy $S^a_\mu(\mathcal{A})$ is less than the topological entropy of Σ.

Definition 9. The topological entropy of a subshift Σ is

$$h_\Sigma(\sigma) = \lim_{n \longrightarrow +\infty} \frac{\log(\#\{u \in Q^n \mid \exists c \in \Sigma, c_{[0,n-1]} = u\})}{n}.$$

Lemma 1. Let Σ be a SFT of order p. If $k \geq p$, define

$$d_D^{(k)}(x, \Sigma) = \limsup_{l \longrightarrow +\infty} \frac{\#\{i \in [-l, l] \mid x_{[i,i+k-1]} \notin L(\Sigma)\}}{2l+1}.$$

Then $d_D^{(k)}(x, \Sigma) \leq (k-p) d_D(x, \Sigma)$.

Proof. As Σ is a SFT, $x_{[i,i+k-1]} \in L(\Sigma)$ if and only if $\forall j$, $0 \leq j \leq k-p$, $x_{[j,j+p-1]} \in L(\Sigma)$. Thus for each $i \in [-l, l]$ such that $x_{[i,i+p-1]} \notin L(\Sigma)$, we have $x_{[j,i+k-1]} \notin L(\Sigma)$ for $i-k+p \leq j \leq i$. Thus in the worst case, $\#\{i \in [-l, l] \mid x_{[i,i+k-1]} \notin L(\Sigma)\} \leq (k-p)\#\{i \in [-l, l] \mid x_{[i,i+p-1]} \notin L(\Sigma)\}$. We deduce that $d_D^{(k)}(x, \Sigma) \leq (k-p) d_D(x, \Sigma)$. ■

Proof. (Of Theorem 2.) The proof is a consequence of the fact that the metric entropy is less than or equal to the topological entropy.

By definition,

$$S^a_\mu(\mathcal{A}) = \lim_{n\to+\infty} \limsup_{t\to+\infty} S^{(t)}_{\mu,n} = \lim_{n\to+\infty} \limsup_{t\to+\infty} \sum_{u\in Q^n} p^{(t)}_u \log(p^{(t)}_u)$$

where $p^{(t)}_u = \mu((G^t_\delta)^{-1}([u]))$.

Let us prove that for all $\epsilon > 0$, this limit is less than or equal to $h_\Sigma(\sigma) + f(\epsilon)$ with $f(\epsilon) \longrightarrow_{\epsilon\to 0} 0$.

By definition of $h_\Sigma(\sigma)$, there exists K such that for all $k \geq K$

$$\frac{\log(\#\{u \in Q^k \mid \exists c \in \Sigma, c_{[0,k-1]} = u\})}{k} \leq h_\Sigma(\sigma) + \epsilon.$$

If a SFT Σ is a Dμ-attracting set of $(Q^{\mathbb{Z}}, G_\delta)$, by definition

$$\mu(\{c \in Q^{\mathbb{Z}} \mid \lim_{n\to+\infty} d_D(G^n_\delta(c), \Sigma) = 0\}) = 1$$

thus $\mu(\{c \in Q^{\mathbb{Z}} \mid \lim_{n\to+\infty} d^{(k)}_D(G^n_\delta(c), \Sigma) = 0\}) = 1$ from Lemma 1.

So, for all $\alpha > 0$, there exists M such that for $t \geq M$,

$$\mu(\{c | d^{(k)}_D(G^t_\delta(c), \Sigma) \leq \epsilon\}) \geq 1 - \alpha.$$

As σ is ergodic, the measure of this set is actually 1.

In this case we have

$$\begin{aligned}
S^{(t)}_{\mu,k} &= \frac{-\sum_{u\in Q^k} p^{(t)}_u \log(p^{(t)}_u)}{k} \\
&= \frac{-\sum_{u\in Q^k\cap L(\Sigma)} p^{(t)}_u \log(p^{(t)}_u)}{k} + \frac{-\sum_{u\in Q^k\setminus L(\Sigma)} p^{(t)}_u \log(p^{(t)}_u)}{k} \\
&\leq \frac{\log(\#\{u \in Q^k \mid \exists c \in \Sigma, c_{[0,k-1]} = u\})}{k} \\
&\qquad + \frac{-\sum_{u\in Q^k\setminus L(\Sigma)} p^{(t)}_u \log(p^{(t)}_u)}{k} \\
&\leq h_\Sigma(\sigma) + \epsilon + \frac{-\sum_{u\in Q^k\setminus L(\Sigma)} p^{(t)}_u \log(p^{(t)}_u)}{k}
\end{aligned}$$

now, to deal with the last term, let us denote $\epsilon_t(c) = d^{(k)}_D(G^t_\delta(c), \Sigma)$ and notice that for $t \geq M$ and μ-almost every c, $\epsilon_t(c) \leq \epsilon$.

Our first goal is to prove that $\sum_{u\in Q^k\setminus L(\Sigma)} p^{(t)}_u \leq \epsilon$. Let us define

$$f \left| \begin{array}{ccl} Q^{\mathbb{Z}} & \to & \mathbb{R} \\ c & \mapsto & \begin{array}{l} 1 \text{ if } c_{[0,k-1]} \in Q^k \setminus L(\Sigma) \\ 0 \text{ otherwise} \end{array} \end{array} \right.$$

and $g_t = f \circ G^t_\delta$. We have $\int g_t d\mu = \sum_{u\in Q^k\setminus L(\Sigma)} p^{(t)}_u$. f is measurable (actually continuous) thus g_t is measurable too and we can apply Birkhoff's ergodic theorem with μ and σ: if $S_n(g_t)(c) = 1/(2n+1)\sum_{i=-n}^{n} g_t \circ \sigma^i(c)$, for μ-almost all c, $\lim_{n\to\infty} S_n(g_t)(c) = \int g_t d\mu$

$$\lim_{n\to\infty} S_n(g_t)(c) = \lim_{n\to\infty} \frac{\#\{i \in [-n,n] | G_\delta^t(c)_{[i,i+k-1]} \notin L(\Sigma)\}}{2n+1}$$
$$= d_D^{(k)}(G_\delta^t(c), \Sigma).$$

Thus, for μ-almost all c,

$$\epsilon_t(c) = d_D^{(k)}(G_\delta^t(c), \Sigma) = \int g_t d\mu = \sum_{u \in Q^k \backslash L(\Sigma)} p_u^{(t)} \leq \epsilon.$$

Furthermore, $-\sum p_u^{(t)} \log(p_u^{(t)})$ is maximal when all $p_u^{(t)}$ are equal to $\epsilon / |Q^k \backslash L(\Sigma)|$, thus in the worst case,

$$\begin{aligned} S_{\mu,k}^{(t)} &= \frac{-\sum_{u\in Q^k} p_u^{(t)} \log(p_u^{(t)})}{n} \\ &\leq h_\Sigma(\sigma) + \epsilon + |Q^k \backslash L(\Sigma)| \left(\frac{-1}{k} \frac{\epsilon}{|Q^k \backslash L(\Sigma)|} \log\left(\frac{\epsilon}{|Q^k \backslash L(\Sigma)|} \right) \right) \\ &\leq h_\Sigma(\sigma) + \epsilon + \frac{-\epsilon}{k} \log\left(\frac{\epsilon}{|Q^k \backslash L(\Sigma)|} \right) \\ &\leq h_\Sigma(\sigma) + \epsilon + \frac{-\epsilon}{k} \log\left(\frac{\epsilon}{|Q|^k} \right) \\ &\leq h_\Sigma(\sigma) + \epsilon - \frac{\epsilon}{k} \log(\epsilon) + \epsilon \log(|Q|). \end{aligned}$$

This inequality holds for all $t \geq M$ and for all $k \geq K$, thus, $S_\mu^a(\mathcal{A}) = \lim_n \limsup_t S_{\mu,n}^{(t)} \leq h_\Sigma(\sigma) + \epsilon - \lim_k \epsilon/k \log(\epsilon) + \epsilon\log(|Q|) = h_\Sigma(\sigma) + \epsilon + \epsilon \log(|Q|)$ and, since $\epsilon + \epsilon\log(|Q|) \longrightarrow 0$, $S_\mu^a(\mathcal{A}) \leq h_\Sigma(\sigma)$. ∎

For instance, as noted before (and proved in [9]), the subshift $\{(01)^*, (10)^*\}$ is $B\mu_{1/2}$-attracting for rule 184, thus $S_{\mu_{1/2}}^a(184) = 0$. In addition, if $0 < \rho < 1/2$ (resp. $1/2 < \rho < 1$) then the subshift of the finite type generated by 00 (resp. 11), 01, and 10 is $B\mu_\rho$-attracting for rule 184, thus $S_{\mu_\rho}^a(184) \leq \log(3)/2$.

2.3.2 Approximations of the apparent entropy

To have an idea of the apparent entropy, when we cannot calculate it exactly, we have computed the correlation over patterns of size 10 (the tenth term of the limit that gives the entropy of the configuration) after 100 iterations. The problem is that it is almost impossible to compute the correlation on really large patterns (20 is possible, but 30 is not). By drawing three-dimensional diagrams of the correlation depending on both the size of the patterns and the number of iterations, we can see that it is a good approximation for most of the rules, but not all of them. When more iterations were absolutely necessary (the rule 110, for instance) we have computed 1000 iterations. But we did nothing when the size of the patterns were too small, which mainly occurs for almost chaotic rules (see section 4). The errors on the apparent entropy measures have two consequences: they explain that in our diagrams the apparent entropy always tends to zero when ρ tends to 0 or 1 and in some cases, the apparent entropy may actually be significantly lower than represented in our diagrams (as for rule 184 with $\rho = 1/2$, whose apparent entropy should be zero, and there are also big errors for rules like 54 and 110).

Let us note that to quickly get an idea of the apparent entropy of a rule, we can use the usual compressors on a file that contains the configuration after 100 or 1000 iterations. We did some experiments with gzip and bzip2 (there were no significant differences between both compressors). Furthermore, it appears that the size of the correlation they take into account is less than 10. Anyway, we could not expect better since although they use different algorithms, the compressors cannot take into account bigger patterns because of the same problem of memory. The problem of this method is that we have no idea of the error we have made.

2.4 Relations between μ-almost everywhere sensitivity and apparent entropy

Let us first remark that if a CA has a null apparent entropy then the metric entropy of its successive configurations tends to zero. We do not know if the reverse is true. Actually, in [11], all the CA such that the metric entropy of its successive configurations tends to zero have a Dμ-attracting set of null topological entropy and thus have a null apparent entropy. We conclude that some CA may be μ-almost everywhere sensitive to initial conditions and have a null apparent entropy. Figure 5 summarizes this with some examples in each region.

3. The chaotic rules

We will here say that a CA is "chaotic" for μ when it is μ-almost everywhere sensitive to initial conditions and has a nonnull apparent entropy.

We think that the chaotic rules among ECA consist of the 30 rules represented in Figure 6. We put together the rules that have the same behavior when we change 0 and 1 or left and right. We (arbitrarily) show for each class a space-time diagram of the rule of lower number.

A first observation is that these rules are μ_ρ-chaotic for any $0 < \rho < 1$. We proved that additive rules are chaotic and their apparent entropy was $\log(2)$ for all $0 < \rho < 1$. On the other chaotic rules, we also observe

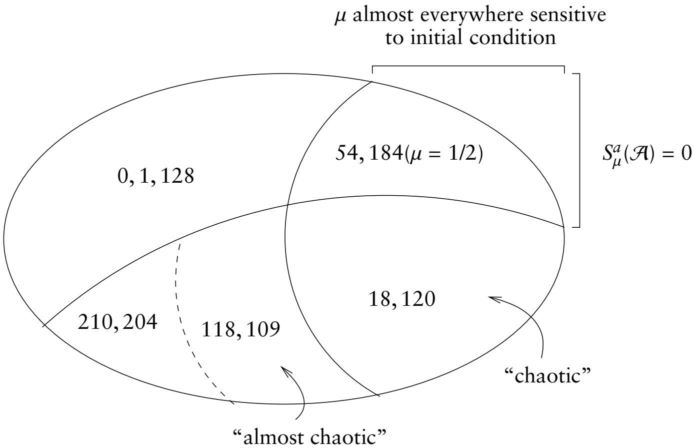
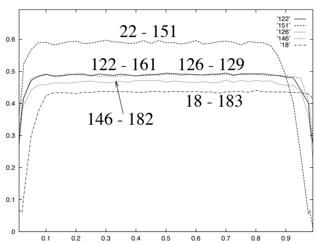


Figure 5. Apparent entropy *versus* μ-almost everywhere sensitivity.

	r	$\overline{r}$	$\overleftrightarrow{r}$	$\overleftrightarrow{r\overline{r}}$
log(2)	150			
	105			
	90	165		
	45	75	101	89
	60	195	102	153
	120	225	106	169
	30	135	86	149
0.59	22	151		
0.49	122	161		
	126	129		
0.47	146	182		
0.43	18	183		



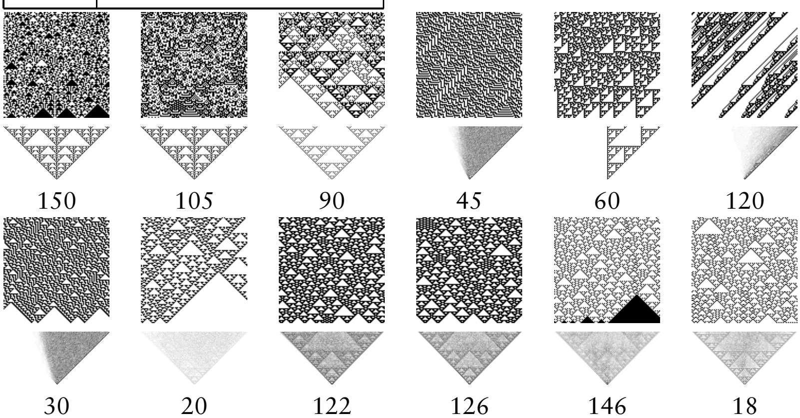


Figure 6. All the chaotic behaviors among 2 states one-dimensional CA.

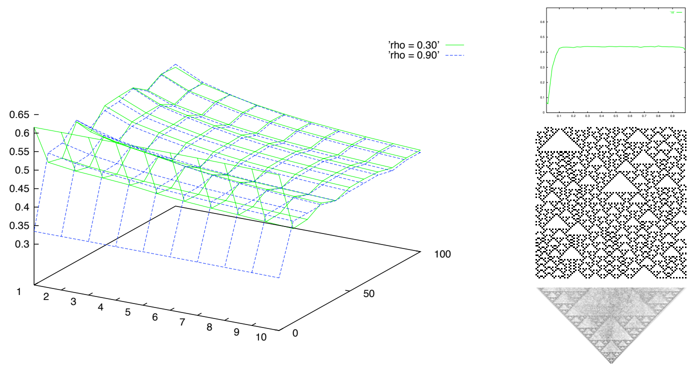


Figure 7. Rule 18 is a chaotic rule which is very likely to have an apparent entropy of log(2)/2 for any nontrivial Bernoulli measure (see the upper right diagram). We see in the left diagram that the correlation of pattern of size less than 10 converges very quickly to the same value starting from two different densities of 1 (0,1 and 0,7). The right diagram represents the apparent entropy of the simple car traffic model T (see [11]) for the measures μ_ρ^* for $0 \leq \rho \leq 1$.

that the apparent entropy is independent of ρ, but it is not necessarily maximum, as it is for rule 18 (see Figure 7). This rule adds the rightmost and leftmost components but only if the center one is equal to 0. The evolution of rule 18 is easy to understand: its space-time diagram can be decomposed into two kinds of regions, those where only even cells have a nonzero state and those where only odd cells have a nonzero state. It is easy to see that these regions are disjoint because the local rule always return 0 when the middle state it depends on is 1. Between these regions, we find particles, as noticed many times (e.g., [3]). The motion of these particles is probably a kind of random walk, anyway, when two particles meet, they annihilate each other and the region that was inside disappears. These events are going to be more and more rare, but anyway, we can imagine that asymptotically, the density of a region tends to zero because a random walker on $\mathbb{Z}$ has a probability 1 to see any cell. In each region, rule 18 behaves like rule 90; so on half of the cells, we can expect that the probability of 1 and 0 tends to 1/2, and hence we may suppose that the apparent entropy of rule 18 is log(2)/2. We see that our experiments give a higher value, probably due to their imprecision as discussed previously. We see that rule 146 (except the fact that it erases long successions of 1 at the very beginning) has the same behavior as rule 18 ($146 = 128 + 18$) so clearly, their apparent entropies are the same although our experiments give different values.

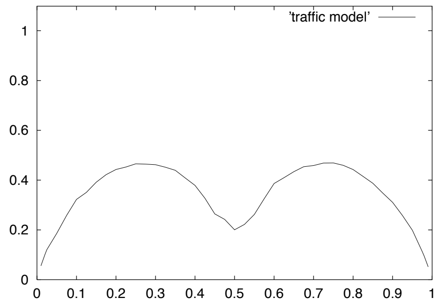


Figure 8. This diagram represents the apparent entropy of the simple car traffic model T (see [11]) for the measures μ_ρ^* for $0 \leq \rho \leq 1$.

We already said that on ECA the chaotic rules are chaotic for all the Bernoulli measures (with $0 < \rho < 1$) and that the apparent entropy for all these measures is the same. This is not always the case. On the one hand, we can build measures so that the evolution of rule 120 on random (for this measure) configurations is a shift. To do this we just have to consider a measure such that there are never two 1s in two successive cells. This is not a Bernoulli measure, but it is a shift invariant one (note that we can find a Bernoulli measure that does this if we consider the CA grouped 2 by 2, see [12, 13] for a definition of the grouping operation). On the other hand, the apparent entropy of the car traffic model T (see Figure 8) is not constant, even on the chaotic part ($\rho \geq 1/2$). Actually, using universal CA that would compute the density of the initial configuration and change its behavior accordingly, we can imagine that the apparent entropy of CA may be almost any arbitrary computable function.

An intuitive idea to provide a definition of "chaos" or "complexity" using apparent entropy would be to consider the CA whose apparent entropy is strictly higher than its initial configuration entropy. However, on the one hand, rule 18 (see Figure 7) is chaotic, while on a wide range of measures its apparent entropy is smaller than its initial configuration entropy. On the other hand, we would not like to call a CA chaotic that makes a xor of the leftmost and the rightmost cell it can access (this is possible by adding states) which has a strictly higher apparent entropy than its initial configuration entropy.

Furthermore, for a given CA and a given measure, the apparent entropy appears to be a good indicator of how complex the CA is. In fact, it is a good way to measure how complex its configuration may look after some computation steps. While the Lyapunov exponent [1] measures at which speed the configuration will appear chaotic, apparent entropy does not depend on the speed but only on how chaotic it asymptotically appears. The upper right diagram of Figure 6 represents the apparent entropy, when it is not maximal, of all the chaotic rules and allows classifying them. We see that, for instance, rule 30 has a smaller Lyapunov exponent than rule 22, but a higher apparent entropy. It seems that the damages of rule 120 give the smallest growth among chaotic CA, actually, in the worst case, they grow as $\alpha \log(n)$ where α is a constant that can be very small. Anyway, a subsequence of its configurations looks asymptotically completely random.

4. Almost chaotic cellular automata

Except the chaotic rules, for some other rules the apparent entropy is close to being constant (see Figure 9). Unfortunately, these CA do not seem to have a specific property that would be a good definition for any state number multidimensional CA. On ECA, there are the nonchaotic CA whose apparent entropy does not tend to 0 when ρ tends to 0 or 1. It seems that all of them (among ECA) present some kind of particles. To present an almost constant apparent entropy, the bound on the average number of modified cells when we change one cell of the initial configuration depends on ρ and is rather huge for very small or very close to 1 values of ρ. This corresponds to the intuitive idea that to have an apparent entropy rather bigger than that of the initial configuration, we have to use a lot of the initial configuration states.

It seems to be difficult to distinguish a chaotic CA, an almost everywhere sensitive CA with a null apparent entropy, and an almost chaotic CA. Actually, we have no convincing experiment that allows deciding where rule 110 is. For rule 118, it is easy, since the damages spreading quickly reaches its bound, it stays constant after about 700 iterations (see Figure 9). Among CA presenting particles in a periodical back-

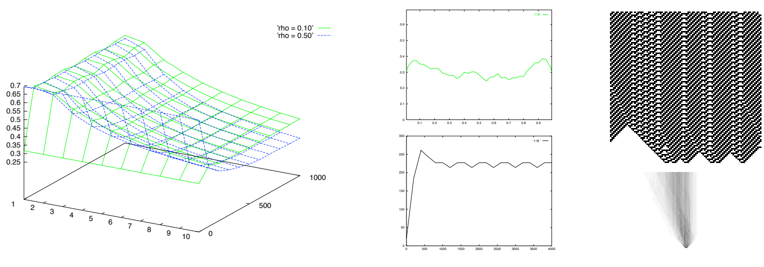


Figure 9. For any $0 < \rho < 1$, rule 118 has bounded damages spreading, thus it is not μ_ρ-almost everywhere sensitive to initial conditions but has an almost constant apparent entropy. The bottom right diagram indicates the average number of damages (multiplied by 20) when we iterate rule 118.

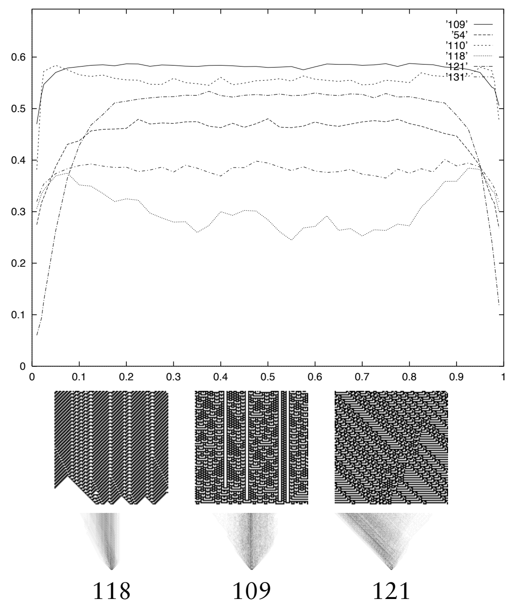


Figure 10. Some almost chaotic behaviors among 2 states one-dimensional CA.

ground, it seems that either all particles disappear (rule 184, $\rho = 1/2$) and then it has a null apparent entropy, or (like rule 118) it remains only one kind of particle randomly placed so that the apparent entropy is nonnull and close to being constant (ultimately, the CA behaves like a shift).

It would be interesting to find a way to split the CA whose damages spreading is bounded but does not tend to 0 in such a way that the almost chaotic rules are not together with the identity. If the behavior of the growing damage of a CA is almost independent of the measure we take, we can separate the CA whose damages spreading is uniformly bounded from the others. This would separate the identity from the almost chaotic CA, but rule 210 would be in the second case, that we did not really expect. Anyway, the fact that Wolfram's class 4 resists our approach was expectable since our approach is exclusively statistic while class 4 definition is based on computability theory.

Conclusion and open questions

In this article, we introduce the notion of apparent entropy. Thanks to this notion, we can better understand why some space-time diagrams that contain the same amount of information may appear more or less complicated to the human eye. The main point is that this tool allows comparing the complexity for the human eye of the space-time diagram with usual notions of chaos. As observed in section 4, the value of the apparent entropy for only one measure is not really significant because we may obtain arbitrary values whether the rule is chaotic or not. However, the values for a set of measures adequately chosen allow finding relations. In the case of ECA, we observed such a relation that confirms Wolfram's intuition. Unfortunately, this relation cannot be easily generalized to more complicated CA.

This work suggests more experiment on rule 110 to better understand its behavior and calculate its apparent entropy. It would also be interesting to study more complex CA since ECA do not present all possible CA behaviors. More theoretically, it may be possible to characterize CA such that the sequence of apparent entropies does not tend to zero for a sequence of measures whose entropy tends to zero.

Acknowledgment

The author is greatly indebted to S. Ruette, B. Allombert, E. Formenti, and A. Maass for helpful conversations and useful comments.

References

[1] F. Bagnoli, R. Rechtman, and S. Ruffo, "Lyapunov Exponents for Cellular Automata," *Lectures on Thermodynamics and Statistical Mechanics*, edited by M. López de Haro and C. Varea (World Scientific, 1994).

[2] F. Blanchard, E. Formenti, and P. Kůrka, "Cellular Automata in the Cantor, Besicovitch, and Weyl Topological Spaces," *Complex Systems*, **11**(2) (1997) 107–123.

[3] N. Boccara, J. Nasser, and M. Roger, "Particle-like Structures and their Interactions in Spatio-temporal Patterns Generated by One-dimensional Deterministic Cellular-automata Rules," *Physical Review A*, **44** (1991) 866–875.

[4] G. Cattaneo, E. Formenti, L. Margara, and G. Mauri, "Transformation of the One-dimensional Cellular Automata Rule Space," *Parallel Computing*, **23** (1997) 1593–1611.

[5] G. Cattaneo, E. Formenti, L. Margara, and G. Mauri, "On the Dynamical Behavior of Chaotic Cellular Automata," *Theoretical Computer Science*, **217** (1999) 31–51.

[6] G. Cattaneo, E. Formenti, L. Margara, and J. Mazoyer, "Shift Invariant Distance on $S^{\mathbb{Z}}$ with Nontrivial Topology," *Proceedings of MFCS'97*, Bratislava, Slovakia, (Springer-Verlag, 1997).

[7] P. A. Ferrari, Alejandro Maass, and Servet Martinez, "Cesàro Mean Distribution of Group Automata Starting from Markov Measures," *preprint*, 2000.

[8] R. H. Gilman, "Classes of Linear Automata," *Ergodic Theory & Dynamical Systems*, 7 (1987) 105–118.

[9] P. Kůrka and A. Maass, "Stability of Subshifts in Cellular Automata," *preprint*, 1999.

[10] K. Lindgren, "Correlations and Random Information in Cellular Automata," *Complex Systems*, **1** (1987) 529–543.

[11] B. Martin, "Damage Spreading and μ-Sensitivity on Cellular Automata," *Proceedings of IFIP TCS2000*, Sendai, Japan, August 2000.

[12] B. Martin, "A Group Interpretation of Particles Generated by One-dimensional Cellular Automaton, 54 Wolfram's Rule," *International Journal of Modern Physics C*, **11**(1) (2000) 101–123.

[13] J. Mazoyer and I. Rapaport, "Inducing an Order on Cellular Automata by a Grouping Operation," *STACS'98*, volume 1373, pages 116–127. Lecture Notes in Computer Science, 1998.

[14] S. Wolfram, "Universality and Complexity in Cellular Automata," *Physica D*, **10** (1984) 1–35.