Stochastic Cellular Automata Solutions to the Density Classification Problem - When randomness helps computing

Nazim Fatès

Stochastic Cellular Automata Solutions to the Density Classification Problem – When randomness helps computing *

Nazim Fatès
INRIA Nancy – Grand Est, LORIA, Nancy Université
nazim.fates@loria.fr
Campus scientifique, BP 239
54 506 Vandœuvre-lès-Nancy, France

February 14, 2012

Abstract

In the density classification problem, a binary cellular automaton (CA) should decide whether an initial configuration contains more 0s or more 1s. The answer is given when all cells of the CA agree on a given state. This problem is known for having no exact solution in the case of binary deterministic one-dimensional CA.

We investigate how randomness in CA may help us solve the problem. We analyse the behaviour of stochastic CA rules that perform the density classification task. We show that describing stochastic rules as a "blend" of deterministic rules allows us to derive quantitative results on the classification time and the classification time of previously studied rules.

We introduce a new rule whose effect is to spread defects and to wash them out. This stochastic rule solves the problem with an arbitrary precision, that is, its quality of classification can be made arbitrarily high, though at the price of an increase of the convergence time. We experimentally demonstrate that this rule exhibits good scaling properties and that it attains qualities of classification never reached so far.

*Extended version of "Stochastic Cellular Automata Solve the Density Classification Problem with an Arbitrary Precision", Proceedings of STACS 2011, Dortmund, Germany. The final publication will appear in the TOCS journal and made available at www.springerlink.com

Introduction

The density classification problem has a surprisingly simple formulation: how does a dynamical system decide what is the majority state in its initial configuration? The problem is trivial in most "classical" computing frameworks (*e.g.*, Turing machines) but the challenge here is to perform the task with the following constraints:

1. *autonomy:* There is no external "head" or operator to perform the computation. Instead the system should perform the calculus autonomously, allowing only changes in the state of its components, the *cells*.

2. *consensus:* The end of the computation is attained when the system reaches a stable consensus in which all the cells agree on a given state.

3. *locality:* Each cell has only a partial view of the system. Typically, a cell can see only the other cells that are located within a given range of perception.

4. *spatial and temporal uniformity:* All cells obey the same law, which remains fixed over time.

5. *simplicity:* The state of the cells is binary.

This problem, also known as the *majority problem*, has been mostly studied in the context of one-dimensional binary cellular automata (CA): the cells are arranged in a ring and cells change their state synchronously, depending on their own state and the states of the cells at a fixed distance, called the *radius* of the neighbourhood. It is of course possible to generalise the problem to other similar systems such as higher-dimension CA [ASB09] or networks [DGT06, B09].

This inverse problem has attracted a considerable amount of research since its formulation by Packard [Pac88]. Note that originally, the goal was to search for non-ergodic rules rules that would be robust to noise in the sense that they would suppress errors and return to equilibrium if their evolution was perturbed [GKL87]. The difficulty of finding a solution comes from the impossibility to centralise the information or to use any classical counting technique. Instead, the convergence to a uniform state should be obtained by using only *local* decisions, which might in some cases contradict the global trend of the system. Moreover, as the structure of CA is homogeneous in space and time, there can be no specialisation of the cells for a partial computation. Solving the problem efficiently requires to find the right balance between deciding locally with a short-range view and following other cells' decision to attain a global consensus.

The quest for efficient rules has been conducted on two main directions: man-designed rules and rules obtained with large space exploration techniques such as genetic algorithms (*e.g.*, [MCH94]). The Gacs-Kurdymov-Levin (GKL) rule, which was originally designed in the purpose of resisting small amounts of noise [GKL87, dSM92], proved to be a good candidate ($\sim$80% of the initial

conditions well-classified on rings of 149 cells) and remained unsurpassed for a long time. In 1995, after observing that outperforming this rule was difficult, Land and Belew issued a key result: no perfect (deterministic) density classifier that uses only two states exist [LB95]. However, this did not stop the search for efficient CA as nothing was known about how well a rule could perform. The search for rules with an increasing quality has been carried on until now, with genetic algorithms as the main investigation tool (see *e.g.* [dOBO06, OMdCF09] and references therein).

On the other hand, various modifications to the classical problem were proposed, allowing one to solve the problem exactly. For instance, Capcarrere *et al.* proposed to modify the output specification of the problem to find a solution that classifies the density perfectly [CST96]. Fukś showed that running two CA rules, namely the "traffic" and "majority" rules, successively would also provide an acceptable solution [Fas97]. Note that the new stochastic rule that we propose also combines these two elementary rules, although in a different way. Later, Martins and Oliveira discovered various couples and triples of rules that solve the problem when applied sequentially and for a fixed number of steps [MdO05] that depends on lattice size. Some authors also proposed to embed a memory in the cells to enhance the abilities of the rules [ASB09, SB09].

However, all of these solutions break at least one of the constraints mentioned above. The use of stochastic (or probabilistic)[1] CA is an interesting alternative that complies with these constraints. Indeed, in stochastic CA, the only modification to the CA structure is that the outcome of the local transitions of the cells is no longer deterministic: it is specified by a probability to update to a given state. The use of randomness to solve the problem was first proposed by Fukś who exhibited a rule which acts as a "stochastic copy" of the state of the neighbouring cells [Fas02]. However, this mechanism generates no force that drives the system towards its goal; the convergence is mainly attained with a random drift of the density (see Sec. 2). Recently, Schüle *et al.* proposed a stochastic rule that implements a local majority calculus [SOS09]. This allows the system to converge to its goal more efficiently, but the convergence still remains bounded by some intrinsic limitations (see Sec. 3).

We propose to follow this path and present a new stochastic rule that solves the density classification problem with an arbitrary precision, that is, for any ring size, the probability of success of the classification can be set arbitrarily close to 1 (Sec. 4). This answers negatively to whether there exists an upper bound on the success rate one can reach without extending the radius of the neighbourhood of the rules.

The idea is to use randomness to solve the dilemma between the local majority decisions and the propagation of a consensus state. A trade-off is obtained by tuning a single parameter that weights two well-known deterministic rules, namely the majority rule and the "traffic" rule. We show that the probability of making a good classification approaches 1 as the probability to apply the ma-

[1] Both terms 'stochastic' and 'probabilistic' CA are found in literature. We prefer to employ the former as etymologically the Greek word 'stochos' implies the idea of goal, aim, target or expectation.

jority rule gets close to 0. This increase of quality however comes at the expense of a longer convergence time. We perform numerical simulations and show that good classification rates can be attained within a "reasonable" computational time.

1 Formalisation of the Problem

In this section, we define the deterministic Elementary Cellular Automata and their stochastic counterpart. We introduce the main notations for studying our problem.

1.1 Elementary Cellular Automata

Let $\mathcal{L} = \mathbb{Z}/n\mathbb{Z}$ represent a set of n cells arranged in a ring. Each cell can hold a state in $\{0, 1\}$ and we call a *configuration* the state of the system at a given time ; the configuration space is $\mathcal{E}_n = \{0, 1\}^{\mathcal{L}}$, it is finite and we have $|\mathcal{E}_n| = 2^n$. We denote by $|x|_P$ the number of occurrences of a pattern P in x. The *density* $\rho(x)$ of a configuration $x \in \mathcal{E}_n$ is the ratio of 1s in this configuration: $\rho(x) = |x|_1/n$. We denote by $\mathbf{0} = 0^{\mathcal{L}}$ and $\mathbf{1} = 1^{\mathcal{L}}$ the two special uniform configurations. For $q \in \{0, 1\}$, a configuration x is a *q-archipelago* if all the cells in state q are isolated, that is, if x does not contain two adjacent cells in state q.

In all the following, we assume that n is odd. This will prevent us from dealing with configurations that have an equal number of 0s and 1s.

An *Elementary Cellular Automaton* (ECA) is a one-dimensional binary CA with nearest neighbour topology, defined by its *local transition rule*, a function $\phi : \{0, 1\}^3 \to \{0, 1\}$ that specifies how to update a cell using only nearest-neighbour information. For a given ring size n, the *global transition rule* $\Phi : \mathcal{E}_n \to \mathcal{E}_n$ associated to ϕ is the function that maps a configuration x^t to a configuration x^{t+1} such that:

$$\forall c \in \mathcal{L},\ x_c^{t+1} = \phi(x_{c-1}^t, x_c^t, x_{c+1}^t)\,.$$

A *stochastic Elementary Cellular Automaton* (sECA) is also defined by a local transition rule, but the next state of a cell is known only with a given probability. In the binary case, we define $f : \{0, 1\}^3 \to [0, 1]$ where $f(x, y, z)$ is probability that the cell updates to state 1 given that its neighbourhood has the state (x, y, z). The *global transition rule* F associated to the local function f is the function that assigns to a random configuration x^t the random configuration x^{t+1} characterised[2] by:

$$\forall c \in \mathcal{L},\ x_c^{t+1} = \mathcal{B}_c^t\left(f(x_{c-1}^t, x_c^t, x_{c+1}^t)\right) \tag{1}$$

where x_c^t denotes the random variable that is given by observing the state of cell c at time t and where $(\mathcal{B}_c^t)_{c\in\mathcal{L},t\in\mathbb{N}}$ is a sequence of independent Bernoulli random

[2]Note that defining rigorously the sequence of random variables x^t obtained from F would require to introduce advanced tools from the probability theory.

variables, that is, $\mathcal{B}_c^t(p)$ is a random variable that equals to $\mathtt{1}$ with probability p and $\mathtt{0}$ with probability $1-p$.

1.2 Density Classifiers

We say that a configuration x is a *fixed point* for the global function F if we have $F(x) = x$ with probability 1 and we say that F is a *(density) classifier* if $\mathbf{0}$ and $\mathbf{1}$ are its two only fixed points.

For a classifier $\mathcal{C}$, let $T(x)$ be the random variable that takes its values in $\mathbb{N} \cup \infty$ defined as:

$$T(x) = \min\left\{t : x^t \in \{\mathbf{0}, \mathbf{1}\}\right\} .$$

We say that $\mathcal{C}$ *correctly classifies* a configuration x if $T(x)$ is almost surely finite and if $x^{T(x)} = \mathbf{1}$ for $\rho(x) > 1/2$ and $x^{T(x)} = \mathbf{0}$ for $\rho(x) < 1/2$. The *probability of good classification* $\mathsf{G}(x)$ of a configuration x is the probability that $\mathcal{C}$ correctly classifies x.

To evaluate quantitatively the quality of a classifier, we need to choose a distribution of the initial configurations. Various such distributions are found in literature, often without an explicit mention, and this is why one may read different quality evaluations for the same rule (for instance compare the results given for the GKL rule: 82% in Ref. [CST96] and 97.8% in Ref. [LB95]). In order to avoid ambiguities, we re-define here the three main distributions of initial configurations that have been used by authors:

(a) The *binomial* distribution μ_{b} is obtained by choosing a configuration uniformly in $\mathcal{E}_n$.

(b) The *d-uniform* distribution μ_{d} is obtained by choosing an initial probability p uniformly in $[0, 1]$ and then building a configuration by assigning to each cell a probability p to be in state $\mathtt{1}$ and a probability $1-p$ to be state $\mathtt{0}$.

(c) The $\mathtt{1}$*-uniform* distribution $\mu_{\mathtt{1}}$ is obtained by choosing a number k uniformly in $\{0, \ldots, n\}$ and then by choosing uniformly a configuration in the set of configurations of $\mathcal{E}_n$ that contain exactly k ones.

Formally,

$$\forall x \in \mathcal{E}_n, \quad \mu_{\mathrm{b}}(x) = \frac{1}{2^n} \quad ; \quad \mu_{\mathrm{d}}(x) = \int_0^1 p^k (1-p)^{n-k} dp \quad ; \quad \mu_{\mathtt{1}}(x) = \frac{1}{n+1} \cdot \frac{1}{\binom{n}{k}}$$

where $k = |x|_{\mathtt{1}}$ is the number of $\mathtt{1}$s in x.

Proposition 1. *The d-uniform distribution μ_{d} and the $\mathtt{1}$-uniform distribution $\mu_{\mathtt{1}}$ are equal.*

Let us show the equality:

$$\forall x \in \mathcal{E}_n, \mu_{\mathrm{d}}(x) = \int_0^1 \rho^k (1-\rho)^{n-k} d\rho = \frac{1}{n+1} \cdot \frac{1}{\binom{n}{k}} = \mu_{\mathtt{1}}(x) .$$

Proof. A first technique to prove the result is to remark that $\mu_{\mathrm{d}}(x)$ corresponds to the calculus of the Beta function $B(k+1, n-k+1)$. This function is defined on real numbers and it was studied by Euler and Legendre. Its analytical form is generally obtained by applying a change of variable and using trigonometric relations.

We now propose a different proof, which relies on more combinatorial arguments. For $i \in \{0, \ldots, n\}$, let us introduce $M_i = \int_0^1 \binom{n}{i} p^i (1-p)^{n-i} dp$. By noting that:

$$1^n = (p + (1-p))^n = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

we find that $\sum_{i=0}^{n} M_i = \int_0^1 1 \cdot dp = 1$. On the other hand, using the integration by parts of M_i gives:

$$\begin{aligned} M_i = & \quad \tbinom{n}{i} \cdot \int_0^1 p^i (1-p)^{n-i} dp \\ = & \quad \tbinom{n}{i} . \left\{ \left[\tfrac{1}{i+1} p^{i+1} (1-p)^{n-i} \right]_0^1 + \int_0^1 \tfrac{n-i}{i+1} p^{i+1} (1-p)^{n-(i+1)} dp \right\} \end{aligned}$$

As the first term of the sum equals 0 and $\binom{n}{i} \cdot \frac{n-i}{i+1} = \binom{n}{i+1}$, we obtain $M_i = M_{i+1}$, which gives $M_i = \frac{1}{n+1}$ and confirms $\mu_{\mathrm{d}} = \mu_1$.

□ □

We can now define the *quality* Q of a classifier $\mathcal{C}$ for a given a ring size n and an initial distribution μ on $\mathcal{E}_n$:

$$Q(n) = \sum_{x \in \mathcal{E}_n} \mathsf{G}(x) \cdot \mu(x) \, .$$

We denote by Q_{b} the *quality*, obtained with the distribution μ_{b}, and Q_{d} the *d-uniform quality* obtained with the distribution μ_{d}. In most cases, we will have $Q_{\mathrm{b}} < Q_{\mathrm{d}}$ as for a large n, most initial configurations of $\mathcal{E}_n$ have a density close to $1/2$ and are generally more difficult to classify. The d-uniform distribution avoids putting too much weight on the "difficult" configurations by assigning an equal chance to appear to all the initial densities.

Similarly, we define the average classification time with regard to distribution μ as

$$T_\mu = \sum_{x \in \mathcal{E}_n} \mathbb{E}\big\{ \, T(x) \, \big\} \mu(x) \, .$$

We denote by T_{b} and T_{d} the average classification time obtained with the μ_{b} and μ_{d} distributions, respectively. As for most classifiers we have $T_{\mathrm{d}} < T_{\mathrm{b}}$, we are only interested in estimating T_{b}.

Table 1: Table of the 8 active transitions and their associated letters. The transition code of an ECA is the sequence of letters of its active transitions.

A	B	C	D	E	F	G	H
000 1	001 1	100 1	101 1	010 0	011 0	110 0	111 0

1.3 Structure of the sECA space

The space of sECA can be described as an eight-dimensional hypercube with the 256 ECA in its corners. This can be perceived intuitively if we see sECA rules as vectors, to which we apply the operations of addition and multiplication. More formally, taking k sECA $f_1, \ldots, f_k$ and $w_1, \ldots, w_k$ real numbers in $[0, 1]$ such that $\sum_{i=1}^{k} w_i = 1$, the barycenter of the sECA (f_i) with weights w_i is the sECA g defined with:

$$\forall x, y, z \in \{0, 1\},\ g(x, y, z) = \sum_{i=0}^{k} w_i\, f_i(x, y, z)\,.$$

A potential advantage of using a barycenter is to re-interpret the dynamics of the sECA g: it is equivalent to the the dynamics of a process where, for each cell independently, the rule f_i would be applied with probability w_i.

Another way to define an sECA is to use a simple combination of other sECA. In this case, the functions (f_i) and the weights w_i should be such as, for all triples (x, yz), the sum $\sum_{i=1}^{k} w_i f(x, yz)$ is in $[0, 1]$.

The most intuitive basis of the sECA space is formed by the 8 (deterministic) ECA that have only one transition that leads to 1: the coordinates correspond to the values $f(x, y, z)$. Alternatively, one may express any sECA as a combination of the 8 ECA that have only one *active* transition, that is, only one change of state in their transition table. Such ECA are labelled A, B, ..., H according to the notation introduced in Ref. [FMST06] and summed up in Tab. 1. Formally, for every sECA f, there exists a 8-tuple $(p_{\mathrm{A}}, p_{\mathrm{B}}, \ldots, p_{\mathrm{H}}) \in [0, 1]^8$ such that: $f = p_{\mathrm{A}} \cdot \mathsf{A} + p_{\mathrm{B}} \cdot \mathsf{B} + \cdots + p_{\mathrm{H}} \cdot \mathsf{H}$. We denote this relationship by $f = [p_{\mathrm{A}}, p_{\mathrm{B}}, \ldots, p_{\mathrm{H}}]_{\mathrm{T}}$, where the subscript T stands for (active) *transitions*.

This basis presents many advantages for studying the random evolution of configurations (see Ref. [FMST06]), as the eight weights code for how much "change" is assigned to each transition. The group of symmetries of a rule can easily be obtained: the left-right symmetry permutes p_{B} and p_{C}, and p_{F} and p_{G}, whereas the 0-1 symmetry permutes p_{A} and p_{H}, p_{B} and p_{G}, etc.

The main advantage of this *transition code* is to allows us to easily write the conservation laws of a stochastic CA and thus to estimate some aspects of its global behaviour. To do this analysis, we will write $a(x) = |x|_{000}$, $b(x) = |x|_{001}$, $\ldots$, $h(x) = |x|_{111}$ (see Tab. 1) and drop the argument x when there is no ambiguity. The following equalities hold [FMST06]:

$$b + d = e + f = c + d = e + g \quad ; \quad b = c \quad ; \quad f = g\,. \tag{2}$$

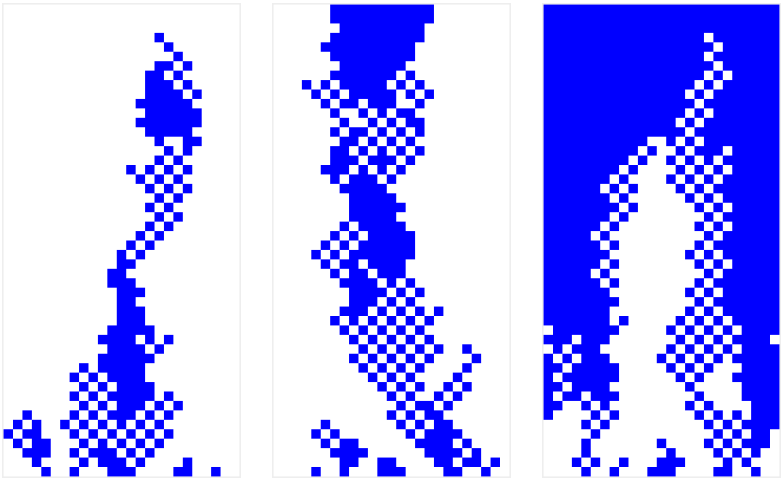


Figure 1: Space-time diagrams showing the evolution of Fuks' classifier $\mathbf{C}_{\mathrm{F}}$ with $n = 25$, $p = 0.48$ and the same initial condition of density $8/25 = 0.32$. Time goes from bottom to top; white cells are 0-cells and blue cells are 1-cells. (left) good classification; (middle) uncertain evolution of the system; (right) bad classification.

Starting from a simple example, we now detail how to use these tools to analyse the behaviour of an sECA.

2 Fukś Density Classifier

To start examining how stochastic CA solve the density classification problem, let us first consider the probabilistic density classifier proposed by Fukś [Fas02]. For $p \in (0, 1/2]$, the local rule $\mathbf{C}_{\mathrm{F}}$ is defined with the following transition table:

(x, y, z)	000	001	010	011	100	101	110	111
$\mathbf{C}_{\mathrm{F}}(x, y, z)$	0	p	$1-2p$	$1-p$	p	$2p$	$1-p$	1

For any ring size n, this rule is a density classifier as $\mathbf{0}$ and $\mathbf{1}$ are its only fixed points. With the transition code of Sec. 1.3, we write:

$$\begin{aligned} \mathbf{C}_{\mathrm{F}} &= [0, p, p, 2p, 2p, p, p, 0]_{\mathrm{T}} \\ &= p \cdot \mathsf{BDEG} + p \cdot \mathsf{CDEF} \end{aligned}$$

where the rules[3] $\mathsf{BDEG}(\mathbf{170})$ and $\mathsf{CDEF}(\mathbf{240})$ are the left and right shift respectively. This means that Fukś' rule can be interpreted as applying, for each cell

[3] We give the "classical" rule code into parenthesis; it is obtained by converting the sequence of 8 bits of the transition table (000 to 111) to the corresponding decimal number.

independently: (a) a left shift with probability p, (b) a right shift with probability p, and (c) staying in the same state with probability $1-2p$ (see Fig. 1). We also note that this rule is invariant under both the left-right and the `0-1` permutations (as $p_{\mathrm{B}} = p_{\mathrm{C}} = p_{\mathrm{F}} = p_{\mathrm{G}}$, $p_{\mathrm{A}} = p_{\mathrm{H}}$ and $p_{\mathrm{D}} = p_{\mathrm{E}}$).

Theorem 1. *For the classifier $\boldsymbol{C}_{\mathrm{F}}$ set with $p \in (0, 1/2]$,*

$$\forall x \in \mathcal{E}_n,\ \mathsf{G}(x) = \max\{\rho(x), 1-\rho(x)\} \quad \textit{and} \quad \mathbb{E}\{\, T(x)\,\} \leq \frac{|x|_1\,(n-|x|_1)}{p}\,.$$

The relationship on $\mathsf{G}(x)$ was observed experimentally with simulations and partially explained by combinatorial arguments [Fas02]. As for the classification time of the system, no predictive law was given. We now propose a proof that uses the analytical tools developed for asynchronous ECAs [FMST06] and completes the results established by Fukś. The proof stands on the following lemma:

Lemma 1. *Let (X_t) be a sequence of random variables that represents the evolution of a Markov process on $\{0, \ldots, m\}$ where m is any integer. Let $T(x) = \min\{t : X_t = 0$ or $X_t = m\}$ be the absorbing time of the process.*

If X_t and $\Delta X_{t+1} = X_{t+1} - X_t$ verify that:

- *the stochastic process (X_t) is a martingale, that is, for the filtration $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$, $\mathbb{E}\{\, \Delta X_{t+1} \,|\mathcal{F}_t\,\} = 0$,*
- *for $t < T$, $\mathbb{E}\{\, \Delta X_{t+1}^2 \,|\mathcal{F}_t\,\} > v$,*

then:

$$\Pr\{X_T = m\} = \frac{q}{m}$$

and T is almost surely finite and obeys:

$$\mathbb{E}\{T(x)\} \leq \frac{q(m-q)}{v} \leq \frac{m^2}{4v}$$

where $q = \mathbb{E}\{X_0\}$.

Proof. First, let us show that T is almost surely finite. We remark that, as long as the process has not converged, ΔX_{t+1} has a null expectation and a non zero variance. There exists a constant p such that $t < T \implies \Pr\{\Delta X_{t+1} \leq 1\} > p$, that is, the probability that X_t decreases by at least one is non zero. For a given $t < T$, if we look m steps ahead, the probability that X_t makes m successive decreases or attains the fixed point 0 before is then non-zero, say α. As (X_t) is a Markov process, this statement can be repeated for any t; the time of convergence to a fixed point is thus upper-bounded by the geometric law of parameter α.

As T is almost surely finite and (X_t) is bounded, T is a stopping time and we can apply Doob's optional stopping theorem on the martingale (X_t):

$$\mathbb{E}\{\, X_T\,\} = \mathbb{E}\{\, X_0\,\} = q\,.$$

On the other hand we have:

$$\mathbb{E}\{\, X_T \,\} = 0 \cdot \Pr\{X_T = 0\} + m \cdot \Pr\{X_T = m\}\,,$$

which, by combining the two equations gives us: $\Pr\{X_T = m\} = q/m$.

To calculate an upper bound on T, let us consider the following process: $Y_t = X_t^2 - v \cdot t$. For $t < T$, we have:

$$\begin{aligned}\mathbb{E}\{\, Y_{t+1} - Y_t \,|\mathcal{F}_t \,\} &= \mathbb{E}\{\, X_{t+1}^2 - X_t^2 \,|\mathcal{F}_t \,\} - v \\ &= \mathbb{E}\{\, (\Delta X_{t+1} + X_t)^2 - X_t^2 \,|\mathcal{F}_t \,\} - v \\ &= \mathbb{E}\{\, \Delta X_{t+1}^2 \,|\mathcal{F}_t \,\} + 2 \cdot \mathbb{E}\{\, \Delta X_{t+1} \cdot X_t \,|\mathcal{F}_t \,\} - v\,.\end{aligned}$$

By definition of the conditional expectation, we have: $\mathbb{E}\{\, \Delta X_{t+1} \cdot X_t \,|\mathcal{F}_t \,\} = \mathbb{E}\{\, \Delta X_{t+1} \,|\mathcal{F}_t \,\} \cdot X_t = 0\,.$

By using the hypothesis of the Lemma: $t < T \implies \mathbb{E}\{\, \Delta X_{t+1}^2 \,|\mathcal{F}_t \,\} \geq v$, we obtain: $\mathbb{E}\{\, Y_{t+1} - Y_t \,|\mathcal{F}_t \,\} \geq 0$, that is, (Y_t) is a submartingale[4]. Again, noting that $\mathbb{E}\{\, T \,\} < \infty$ and that (Y_t) is a process with finite increments, we apply the Optional Stopping Theorem: $\mathbb{E}\{\, Y_T \,\} \geq \mathbb{E}\{\, Y_0 \,\}$, which reads: $\mathbb{E}\{\, X_T^2 - v \cdot T \,\} \geq q^2$ or $\mathbb{E}\{\, T \,\} \leq \frac{\mathbb{E}\{\, X_T^2 \,\} - q^2}{v}$. To calculate $\mathbb{E}\{\, X_T^2 \,\}$, we write:

$$\mathbb{E}\{\, X_T^2 \,\} = 0^2 \cdot \Pr\{X_T = 0\} + m^2 \cdot \Pr\{X_T = m\} = m \cdot q$$

which gives us the expected result on $\mathbb{E}\{\, T \,\}$.

□ □

We can now prove Theorem 1.

Proof. We simply show that Lemma 1 applies to $X_t = |x^t|_1$. Denoting by $\tilde{A}$, $\tilde{B}$, the cells where transitions A, B, ... apply, and given that transitions B, C, D (resp. E, F, G) increase (resp. decrease) ΔX_{t+1} by 1, we write:

$$\Delta X_{t+1} = \sum_{c \in \tilde{B}, \tilde{C}} \mathcal{B}_c^t(p) + \sum_{c \in \tilde{D}} \mathcal{B}_c^t(2p) - \sum_{c \in \tilde{E}} \mathcal{B}_c^t(2p) - \sum_{c \in \tilde{F}, \tilde{G}} \mathcal{B}_c^t(p) \qquad (3)$$

where $(\mathcal{B}_c^t)$ is the sequence of the Bernoulli random variables of Eq. (1). This gives:

$$\begin{aligned}\mathbb{E}\{\, \Delta X_{t+1} \,|\mathcal{F}_t \,\} &= p \cdot b + p \cdot c + 2p \cdot d - 2p \cdot e - p \cdot f - p \cdot g \\ &= p \cdot (b + d - e - f) + p \cdot (c + d - e - g)\end{aligned}$$

Using Eq. (2), we obtain $\mathbb{E}\{\, \Delta X_{t+1} \,|\mathcal{F}_t \,\} = 0$.

Let us now assume that $t < T$, that is, x^t is not a fixed point. Using the independence of the random variables of Eq. (3) and $\mathrm{var}\{\, \mathcal{B}(p) \,\} = p(1-p)$, we obtain:

$$\begin{aligned}\mathbb{E}\{\, \Delta X_{t+1}^2 \,|\mathcal{F}_t \,\} &= (b + c + f + g) \cdot p(1-p) + (d + e) \cdot 2p(1-2p) \\ &= p \cdot [(s_1 + 2s_2) - (s_1 + 4s_2) \cdot p]\end{aligned}$$

[4]Rigorously, one needs to use the "stopped" process $Y_t = X_{t \wedge T}^2 - v \cdot (t \wedge T)$.

where $s_1 = b + c + f + g$ and $s_2 = d + e$. Using Eq. (2) and noting that the value of n is odd, we remark that there exists at least one 00 or 11 pattern in the configuration and that $s_1 = b + c + f + g \geq 2$. From $p \leq 1/2$, we obtain $(s_1 + 2s_2) - (s_1 + 4s_2) \cdot p \geq 1$ and $\mathbb{E}\{\, \Delta X_{t+1}^2 \,|\mathcal{F}_t\,\} \geq p$.

Lemma 1 thus applies by taking $v = p$ and $m = n$. The probability that the process stops on $X_T = n$, that is, on the fixed point $\mathbf{1}$, is equal to the initial density $\rho(x) = |x|_1/n$. From this result, we derive that for any configuration x the probability of good classification is $\mathsf{G}(x) = \max\{\rho(x), 1 - \rho(x)\}$ and obtain the upper-bound on $\mathbb{E}\{\, T(x) \,\}$.

□ □

Theorem 2. *The d-uniform quality of $\boldsymbol{C}_{\mathrm{F}}$ is $Q_{\mathrm{d}}(n) = 3/4 + 1/4n$. The quality of $\boldsymbol{C}_{\mathrm{F}}$ is*

$$Q_{\mathrm{b}}(n) = \frac{1}{2} + \frac{1}{2^n} \cdot \binom{n-1}{\frac{n-1}{2}}.$$

The average classification time verifies $T_{\mathrm{b}} \in \mathcal{O}(n^2)$.

Proof. The result on Q_{d} is obtained with an integration or a sum (use μ_{d} or μ_1); the result on T_{b} is obtained with a simple upper-bound. Let us now calculate the quality Q_{b}:

$$Q_{\mathrm{b}}(n) = \frac{1}{2^n} \sum_{x \in \mathcal{L}} \max\{\rho(x), 1 - \rho(x)\}.$$

We group the configurations with equal number of 1s:

$$Q_{\mathrm{b}}(n) = \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} \max\{\frac{i}{n}, 1 - \frac{i}{n}\}.$$

For $n = 2k + 1$, we decompose the sum into:

$$Q_{\mathrm{b}}(n) = \frac{1}{2^n} \left(\sum_{i=0}^{k} \binom{n}{i} \frac{n-i}{n} + \sum_{i=k+1}^{n} \binom{n}{i} \frac{i}{n} \right).$$

Using the symmetry $\binom{n}{i} = \binom{n}{n-i}$, we obtain:

$$Q_{\mathrm{b}}(n) = \frac{1}{2^n} \cdot 2 \cdot \sum_{i=k+1}^{n} \binom{n}{i} \frac{i}{n} \quad = \frac{1}{2^{n-1}} \sum_{i=0}^{k} \binom{n-1}{i}.$$

To calculate the sum, we write $n - 1 = 2k$ and:

$$\sum_{i=0}^{k} \binom{2k}{i} = \sum_{i=0}^{k} \binom{2k}{2k-i} = \sum_{i=k}^{2k} \binom{2k}{i}.$$

This gives:

$$2. \sum_{i=0}^{k} \binom{2k}{i} = \sum_{i=0}^{k} \binom{2k}{i} + \binom{2k}{k} + \sum_{i=k+1}^{2k} \binom{2k}{i} = 2^{2k} + \binom{2k}{k}.$$

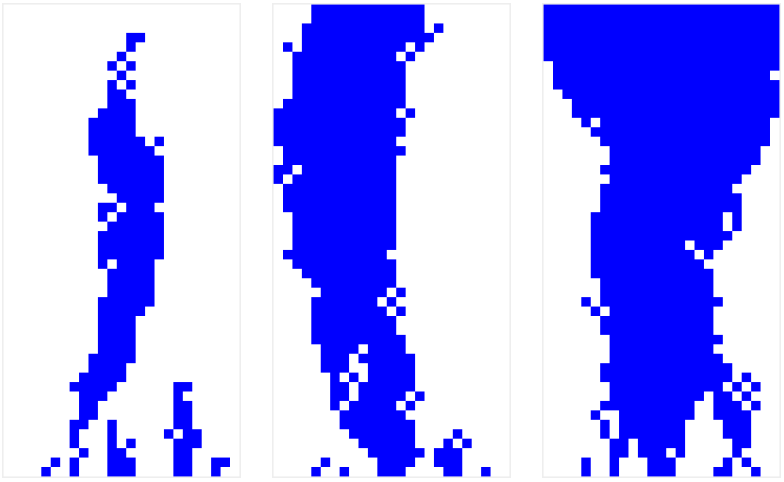


Figure 2: Space-time diagrams showing the evolution of Schuele classifier $\mathbf{C}_{\mathrm{S}}$ with $n = 25$, $\varepsilon = 0.75$ and the same initial condition of density $8/25 = 0.32$. (left) good classification; (middle) uncertain evolution of the system; (right) bad classification.

We then obtain:

$$Q_{\mathrm{b}}(n) = \frac{1}{2^{2k}} \frac{2^{2k} + \binom{2k}{k}}{2} = \frac{1}{2} + \frac{\binom{n-1}{(n-1)/2}}{2^n} .$$

□ □

This formula explains why the quality of classification of $\mathbf{C}_{\mathrm{F}}$ quickly decreases as the ring size n increases. For instance for $n = 49$, we have: $Q_{\mathrm{b}}(n) = 0.557$, that is, the gain of using $\mathbf{C}_{\mathrm{F}}$ compared to a random guess is less than 6%. For the reference value $n = 149$, the gain drops down to 3.3% (see Fig. 4 p. 19). It is interesting to note that the setting of p does not have any influence on the quality, but there exists a value p which minimises the classification time.

3 Schüle Density Classifier

We now consider the probabilistic density classifier proposed by Schüle *et al.* [SOS09]. It was designed to improve the convergence of the system towards a fixed point. The idea of the authors is to "blend" the majority rule, which has good properties for the classification problem, with the XOR rule, which introduces some noise to make the boundaries between regions of 0s and 1s unstable. Formally, for $\varepsilon \in (0, 1)$, the local rule $\mathbf{C}_{\mathrm{S}}$ is defined with the following transitions:

(x, y, z)	000	001	010	011	100	101	110	111
$\mathbf{C}_{\mathrm{S}}(x, y, z)$	0	$1-\varepsilon$	$1-\varepsilon$	ε	$1-\varepsilon$	ε	ε	1

.

This rule is a density classifier as **0** and **1** are its only fixed points. With the transition code of Sec. 1.3, we write:

$$\begin{aligned}\mathbf{C}_{\mathrm{S}} &= [0, 1-\varepsilon, 1-\varepsilon, \varepsilon, \varepsilon, 1-\varepsilon, 1-\varepsilon, 0]_{\mathrm{T}} \\ &= (1-\varepsilon)\cdot \mathsf{BCFG} + \varepsilon \cdot \mathsf{DE}\end{aligned}$$

where rule $\mathsf{BCFG}(\mathbf{150})$ is the rule that implements a XOR function with three neighbours and $\mathsf{DE}(\mathbf{232})$ is the majority rule. This means that Schüle's rule can be interpreted as applying for each cell independently: (a) a XOR with probability $1-\varepsilon$ (b) a majority with probability ε (see Fig. 2). This rule is invariant under both the left-right and the 0-1 symmetries (as we have: $p_{\mathrm{B}} = p_{\mathrm{C}} = p_{\mathrm{F}} = p_{\mathrm{G}}$, $p_{\mathrm{A}} = p_{\mathrm{H}}$ and $p_{\mathrm{D}} = p_{\mathrm{E}}$).

3.1 Exact results on Schüle's rule

Theorem 3. *For the classifier $\boldsymbol{C}_{\mathrm{S}}$ set with $\varepsilon = 2/3$,*

$$\forall x \in \mathcal{E}_n,\ \mathsf{G}(x) = \max\{\rho(x), 1-\rho(x)\} \quad \textit{and} \quad \mathbb{E}\{\, T(x)\,\} \leq \frac{|x|_1(n-|x|_1)}{\epsilon(1-\epsilon)}\,.$$

As a consequence, it has the same quality as Fukś rule and its classification time T_{b} has the same scaling order: $T_{\mathrm{b}} \in \mathcal{O}(n^2)$.

The relationship on $\mathsf{G}(x)$ was proved under the mean-field approximation [SOS09]. We now propose to re-derive this result without approximation.

Proof. Let us take again $X_t = |x^t|_1$ and show that Lemma 1 applies to (X_t). Recall that we denote by $\tilde{A}$, $\tilde{B}$,... the cells where transitions A, B ,... apply. We have:

$$\Delta X_{t+1} = \sum_{c\in\tilde{B},\tilde{C}} \mathcal{B}_c^t(1-\varepsilon) + \sum_{c\in\tilde{D}} \mathcal{B}_c^t(\varepsilon) - \sum_{c\in\tilde{E}} \mathcal{B}_c^t(\varepsilon) - \sum_{c\in\tilde{F},\tilde{G}} \mathcal{B}_c^t(1-\varepsilon)\,,$$

where $(\mathcal{B}_c^t)$ is the sequence of Bernoulli random variables of Eq. (1). This gives:

$$\begin{aligned}\mathbb{E}\{\, \Delta X_{t+1}\,|\mathcal{F}_t\,\} &= (1-\varepsilon)(b+c-f-g) + \varepsilon(d-e) \\ &= (1-\varepsilon)\cdot(b+c-d+e-f-g) + d - e\,.\end{aligned}$$

Using Eq. (2), we obtain:

$$\mathbb{E}\{\, \Delta X_{t+1}\,|\mathcal{F}_t\,\} = (3\varepsilon-2)(d-e)\,, \tag{4}$$

which leads to $\mathbb{E}\{\, \Delta X_{t+1}\,\} = 0$ for $\varepsilon = 2/3$.

On the other hand, let us assume that $t < T$, that is, x^t is not a fixed point. This implies that $b+c+d+e+f+g \geq 1$. As $\mathrm{var}\{\, \mathcal{B}(\varepsilon)\,\} = \mathrm{var}\{\, \mathcal{B}(1-\varepsilon)\,\} = \varepsilon.(1-\varepsilon)$, we obtain:

$$\begin{aligned}\mathbb{E}\{\, \Delta X_{t+1}^2\,|\mathcal{F}_t\,\} &= (b+c+d+e+f+g)\cdot\varepsilon(1-\varepsilon) \\ &\geq \varepsilon(1-\varepsilon)\,.\end{aligned}$$

Lemma 1 thus applies by taking $v = \varepsilon(1-\varepsilon)$ and $m = n$. Consequently, we find that the probability that the process stops on the fixed point $\mathbf{1}$ (given by $X_T = n$) is equal to $X_0/n = \rho(x)$, which gives the value of $\mathsf{G}(x)$ and the upper-bound on $\mathbb{E}\{\, T(x) \,\}$.

□ □

3.2 General behaviour of Schüle's rule

Equation (4) also allows us to understand the general behaviour of the classifier $\mathbf{C}_{\mathrm{S}}$ for $\varepsilon \neq \frac{2}{3}$. Informally, let us consider a configuration x with a density close to 1. For such a configuration, we most likely have more isolated `0`s than isolated `1`s, that is, $d - e > 0$ and the sign of ΔX_{t+1} is the same as $3\epsilon - 2$. As for such configurations, we want the density to *increase*, we see that setting $\epsilon > 2/3$ drives the system more rapidly towards its goal. This also explains why for $\epsilon < 2/3$, it was no longer possible to observe the system convergence within "reasonable" simulation times. In fact, as observed by Schüle *et al.* [SOS09], the system is then in a metastable state: although the classification time is almost surely finite, the system is always attracted towards a density 1/2. Last, but not least, we think that for $\epsilon > 2/3$, only isolated `0`s or `1`s of the *initial* configuration contribute to driving the system to its goal. This leads us to formulate the following statement:

Proposition 2. *For the classifier $\boldsymbol{C}_{\mathrm{S}}$ set with $\varepsilon > 2/3$, the quality of classification $Q_{\mathrm{b}}(n)$ is bounded. More precisely:*

$$\forall x \in \mathcal{E}_n : |x|_{\mathtt{010}} = |x|_{\mathtt{101}} = 0, \mathsf{G}(x) = \max\{\rho(x), 1-\rho(x)\}$$

and

$$\forall x \in \mathcal{E}_n, \mathsf{G}(x) \leq \max\{\rho^*(x), 1-\rho^*(x)\}$$

where $\rho^*(x)$ is the density attained by an asymptotic evolution of x under the majority rule.

Intuitively, let us consider a configuration x^t that has no isolated `0` or `1`. Equation (4) then reads $\mathbb{E}\{\, \Delta X_{t+1} \,|\mathcal{F}_t\,\} = 0$. At time $t+1$, isolated `0`s or `1`s may appear (if transitions B and F, or C and G, are applied simultaneously) but there is an equal chance to make an isolated `0` than an isolated `1`. As this analysis can be repeated recursively, this shows that once the first isolated `0`s or `1`s have disappeared (with the effect of the majority rule), the boundary of the homogeneous regions randomly drift until they merge. As a consequence, the probability of good classification of a configuration with no isolated `0` or `1` is the same as for Fukś classifier. This behaviour implies that the average convergence times scales as n^2, which is confirmed experimentally (see Fig. 4).

We now present a rule that does not suffer from such limitations.

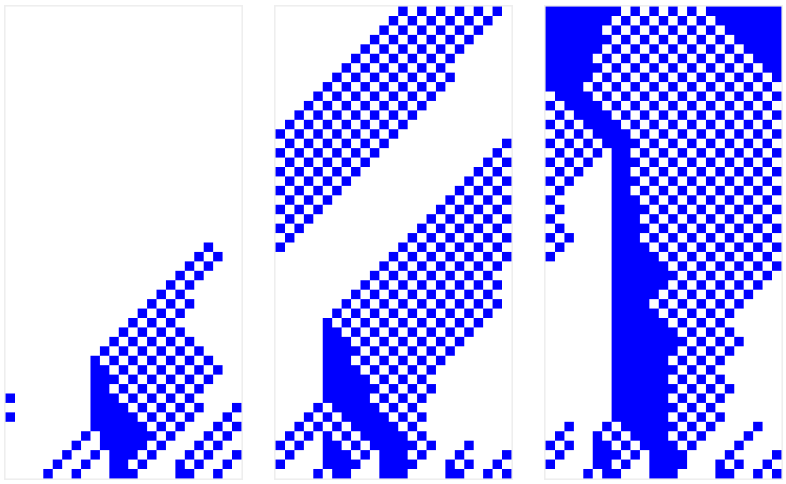


Figure 3: Space-time diagrams showing the evolution of traffic-majority classifier $\mathbf{C}_{\mathrm{TM}}$ with $n = 25$, $\eta = 0.30$ and the same initial condition of density $8/25 = 0.32$. (left) good classification; (middle) good classification will happen with proba 1; (right) bad classification will happen with proba 1.

4 A New Rule for Density Classification

For $\eta \in (0, 1]$, let us consider the stochastic rule $\mathbf{C}_{\mathrm{TM}}$ defined with:

(x, y, z)	000	001	010	011	100	101	110	111
$\mathbf{C}_{\mathrm{TM}}(x, y, z)$	0	0	0	1	$1-\eta$	1	η	1

With the transition code, we write:

$$\begin{aligned} \mathbf{C}_{\mathrm{TM}} &= [0, 0, 1-\eta, 1, 1, 0, 1-\eta, 0]_{\mathrm{T}} \\ &= \eta \cdot \mathsf{DE} + (1-\eta) \cdot \mathsf{CDEG}\,. \end{aligned}$$

For $\eta \in (0, 1)$, the effect of the rule is the same as applying, for each cell and at each time step, the so-called "traffic" rule (see below) CDEG(**184**) with probability $1 - \eta$ and the majority rule DE(**232**) with probability η. We thus call this rule the *traffic-majority* rule.

4.1 General behaviour of the rule

The traffic rule is a well-known elementary rule that has the property to be number conserving, that is, the number of 1s is conserved as the system evolves (see *e.g.*, [BF02]). Observing the evolution of this rule, we see that a 1 with a 0 at its right moves to right while a 0 with a 1 at its left is moved to the left. So everything happens as if the 1s were cars that tried to go to the right, possibly

meeting traffic jams. These jams resorb by going in the inverse directions of the cars (when possible).

Combining this rule with applications of the majority creates a positive effect: a 1 followed by two 0 might simply disappear (see Fig. 3). So, if we start from a configuration with less 1s than 0s, all happens as if the 1s, which can be considered as defects, will spread on the lattice, and then progressively disappear by the effect of the majority rule. Moreover, even though the system is stochastic, there exists an infinity of configurations that can be classified with no error.

Lemma 2. *An archipelago is well-classified with probability 1. A configuration x that is q-archipelago with k cells in state q has an average convergence time $\mathbb{E}\{\, T(x) \,\} \leq k/\eta$.*

Proof. The proof is simple and relies on two observations.

First, let us note that the successor of a q-archipelago is a q-archipelago. To see why this holds, without loss of generality, let us assume that x is a 1-archipelago. By the application of the rule, all 1s are isolated and will be transformed into 0s. As for the 0s that will be transformed into 1s, they are necessarily preceded by a 1, which implies that no such two 0s can be found next to another in x.

Second, we remark that the number of 1s in x^t is a non-increasing function of t. Indeed, at each time step, each isolated 1 can "disappear" if transition C is not applied, which happens with probability $\eta > 0$. As a result, all the 1s will eventually disappear and the system will attain the fixed point $\mathbf{0}$, which corresponds to a good classification as we have $\rho(x) < 1/2$. As long as the system has not reached a fixed point, there exists a probability greater than η to decrease the number of "islands" of the archipelago by 1; so if the archipelago is made of k islands, its disappearing average time is upper-bounded by k/η.

□ □

Interestingly, we experimentally observed that the traffic-majority rule seems to have a bifurcation point at $\eta = 1/2$: for $\eta > 1/2$, the qualitative behaviour of the rule is the same as Fukś or Schüle's rule: isolated 0s and 1s quickly disappear, then boundaries between homogeneous regions follow a random walk until they meet and annihilate. On the contrary, for $\eta < 1/2$, the system behaves "correctly": if for instance we start from a density lower than $1/2$, we observe that regions of consecutive 1s are progressively transformed into regions of 01 patterns, and then, that these patterns dissolve from right to left due to the mechanism explained above.

Fig. 4 shows the values of Q_{b} and T_{b} estimated by numerical simulations. We can observe that the quality rapidly increases to high values, even when keeping the average convergence time to a few thousand steps. In particular for $\eta < 1\%$, the quality goes above the symbolic rate of 90%, which, to our knowledge, had not been reached for one-dimensional systems (see *e.g.* [dOBO06, OMdCF09]). Another major point regards the classification time of $\mathbf{C}_{\mathrm{TM}}$: for $n \leq 300$ and

$\eta \leq 0.1$, it is experimentally determined as varying linearly (or quasi-linearly) with the ring size n (see Fig. 5).

The general analytical estimation of the quality of $\mathbf{C}_{\mathrm{TM}}$ and its time of convergence is more complex than for Fukś and Schüle classifiers. However, some analytical can be obtained when the rate of application of the majority rule η, which represents the "randomness" of the system, becomes small.

4.2 Behaviour of the rule for a small amount of randomness

For a configuration x, we say that an archipelago a is *appropriate* if it has the same majority state as x, that is, if $\rho(x) < 1/2$ and $\rho(a) < 1/2$ or if $\rho(x) > 1/2$ and $\rho(a) > 1/2$. The second important property of the rule is that any configuration is transformed into an "appropriate" archipelago with a probability that approaches 1 as $\eta \to 0$.

Lemma 3. *For every $p \in [0, 1)$, there exists a setting η of the classifier $\boldsymbol{C}_{\mathrm{TM}}$ such that* every *configuration $x \in \mathcal{E}_n$ will evolve to an appropriate archipelago with a probability greater than p.*

Proof. The idea of the proof is to compare the evolution of the traffic-majority rule $\mathbf{C}_{\mathrm{TM}}$ and the Traffic rule. First, we claim that, starting from any configuration x if these two rules evolve identically during the $T = \lceil n/2 \rceil$ steps, the system reaches the appropriate archipelago (for x). This relies on the fact that (a) the Traffic rule is number conserving and (b) it always evolves to an archipelago in at most T steps (see *e.g.*, Ref. [CST96] Lemma 4).

It is then sufficient to take η such that the probability that a difference between the two rules occurs is lower than p to prove the lemma. Looking at the transition tables the traffic rule and $\mathbf{C}_{\mathrm{TM}}$, we see that such differences only occur if a transition $\mathtt{100} \to \mathtt{0}$ or $\mathtt{110} \to \mathtt{1}$ occur, which happens with probability η. Moreover, the number of occurrence of such transitions can be grossly upper-bounded by nT, which gives $\eta < p^{1/nT}$ to ensure that the probability that a difference occurs is lower than p.

□ □

However, it should be noted that the upper-bound established in the proof above does not give the correct order of the probability that a non-appropriate archipelago is reached (which implies fatally an error of classification). Indeed, if for instance the initial condition x has a density $\rho(x) < 1/2$, the transitions $\mathtt{100} \to \mathtt{0}$ are "beneficial" to the evolution and only the transitions $\mathtt{100} \to \mathtt{1}$ can lead to a "bad" state the density of which is higher than $1/2$. Moreover, we can remark that (a) such "bad" transitions have to be more numerous than the difference between 0s and 1s in x and (b) a transition $\mathtt{100} \to \mathtt{1}$ can only concern a 0 that is always preceded by a 11 pattern before its "transformation". We leave as an open question the quantitative estimation of these effects and now to state our main result.

Theorem 4. *For a quality $q \in [0,1)$, there exists a setting η of the classifier $\boldsymbol{C}_{\mathrm{TM}}$ such that $\forall x \in \mathcal{E}_n$, $\mathsf{G}(x) \geq q$. As a consequence, for $\eta \to 0$ the quality verifies $Q_{\mathrm{b}}(n) \to 1$.*

Proof. We simply combine the two previous lemmas: for η small enough, the system evolves to an "appropriate" archipelago (Lemma 3); it will then progressively "drift" towards the appropriate fixed point (Lemma 2).

□ □

The drawback of setting η small is the increase of the time taken to classify a configuration. We experimentally verified that the system possesses an optimal convergence time. For $n = 149$, it is obtained for $\eta \sim 0.30$ for which value the classification time is as low as $T_{\mathrm{b}} \sim 250$ steps (see Fig. 4 p. 19). We leave the question open as to how to calculate this optimal value analytically.

Conclusion

This paper presented an analysis of three stochastic cellular automata that solve the density classification problem. We studied two known rules and presented a new rule which, for a given ring size, solves the problem with an arbitrary precision. This illustrates how, in some cases, randomness is not an obstacle to overcome but a useful means to achieve a computation that would otherwise be impossible in a strict deterministic computing framework. One may see an analogy with biological organisms, which *need* a certain degree of randomness to maintain their state and functions.

The use of analytical techniques originally developed for asynchronous cellular automata allowed us to gain a better understanding of the behaviour of the stochastic rules. Numerical simulations showed that for the standard reference value $n = 149$, the traffic-majority rule obtains a quality of classification over 90%, a result never attained so far. This shows that the construction is not only "theoretical" but also leads to an effective classification method. However, some questions remain unsolved, among which:

Question 1. The results also suggest that the quality of classification cannot be taken as the unique criterion for evaluating the classifiers. As there exists some trade-off between quality and time to classify, we ask how to estimate numerically the best compromise between these two constraints. Schüle proposed to use the parameter Q/T [SOS09]; one may also consider the quantity $(Q - 1/2)/T$ by noting that a purely random rule is useless. There are also other possibilities as to consider an average quantity of information gained at each time step, etc.

Question 2. Are there are other couples (or t-uples) of rules that can be "blended" to obtain similar or even better results? Note that the "blending"

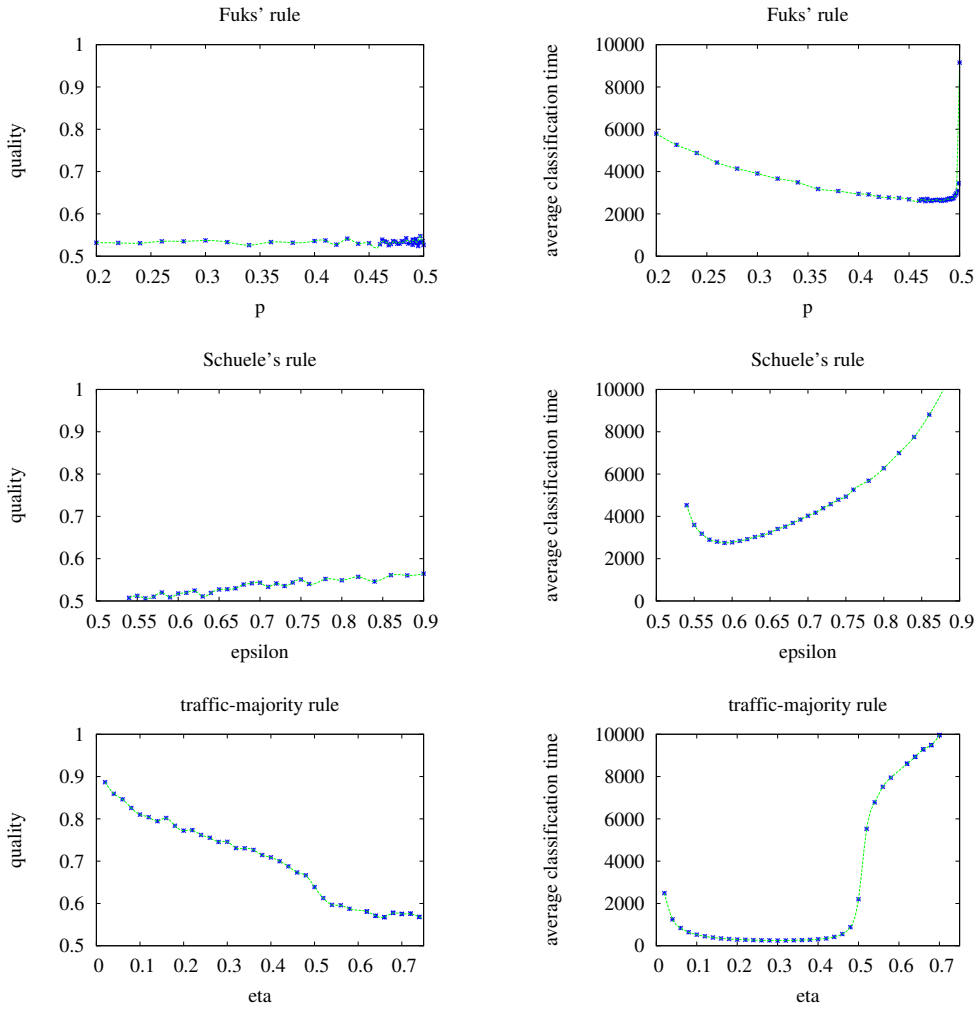
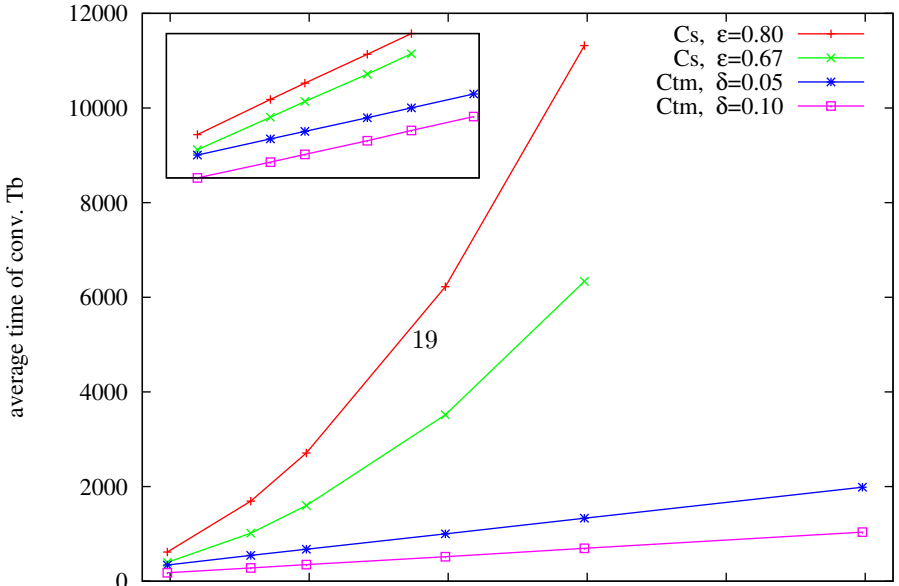


Figure 4: Quality of classification Q_{b} and convergence time T_{b} for a ring size $n = 149$; averages obtained with 10 000 samples. (top) Fukś rule; (middle) Schüle's rule ; (bottom) traffic-majority rule.



technique could also be used with rules that have a larger radius, and on lattices with two or more dimensions.

Question 3. It was shown that for a small neighbourhood (radius 1), one may construct a rule that performs arbitrarily well for any ring size. However, what can we say if the rule is fixed and if we evaluate its quality over any ring size? In particular, does there exist a rule whose quality for a large ring size does not approach 1/2 (pure randomness)? We conjecture that such a rule does *not* exist for the Binomial distribution (but we saw that there are solutions for other initial distributions such as the d-uniform distribution).

Question 4. Last, we ask how to generalise the problem to infinite lattices. In a recent work, Marcovici *et al.* obtained various results for CA where the cells are initialised by independent Bernoulli random variables of parameter p [BFMM11]. However, the question as to whether the traffic-majority rule solves the problem for $\mathbb{Z}$ remains an open problem.

acknowledgements

The author expresses his sincere gratitude to the two anonymous referees and to his collaborators for interesting remarks and suggestions which resulted from a careful reading of the manuscript.

References

[ASB09] Ramón Alonso-Sanz and Larry Bull. A very effective density classifier two-dimensional cellular automaton with memory. *Journal of Physics A*, 42(48):485101, 2009.

[BÓ9] Florence Bénézit. *Distributed average consensus for wireless sensor networks*. PhD thesis, EPFL, Lausanne, 2009.

[BF02] Nino Boccara and Henryk Fukś. Number-conserving cellular automaton rules. *Fundamenta Informaticae*, 52(1-3):1–13, 2002.

[BFMM11] Ana Busic, Nazim Fatès, Jean Mairesse, and Irène Marcovici. Density classification on infinite lattices and trees. arXiv:1111.4582 - Short version to appear in the procceedings of LATIN 2012, LNCS series, 2011.

[CST96] Mathieu S. Capcarrere, Moshe Sipper, and Marco Tomassini. Two-state, $r = 1$ cellular automaton that classifies density. *Phys. Rev. Lett.*, 77(24):4969–4971, 1996.

[DGT06] Christian Darabos, Mario Giacobini, and Marco Tomassini. Scale-free automata networks are not robust in a collective computational task. In Samira El Yacoubi, Bastien Chopard, and Stefania Bandini, editors, *Cellular Automata*, volume 4173 of *Lecture Notes in Computer Science*, pages 512–521. Springer Berlin / Heidelberg, 2006.

[dOBO06] Pedro P.B. de Oliveira, José C. Bortot, and Gina M.B. Oliveira. The best currently known class of dynamically equivalent cellular automata rules for density classification. *Neurocomputing*, 70(1-3):35 – 43, 2006.

[dSM92] Paula Gonzaga de Sá and Christian Maes. The Gacs-Kurdyumov-Levin automaton revisited. *Journal of Statistical Physics*, 67:507–522, 1992.

[Fas97] Henryk Fukś. Solution of the density classification problem with two cellular automata rules. *Physical Review E*, 55(3):R2081–R2084, Mar 1997.

[Fas02] Henryk Fukś. Nondeterministic density classification with diffusive probabilistic cellular automata. *Physical Review E*, 66(6):066106, 2002.

[FMST06] Nazim Fatès, Michel Morvan, Nicolas Schabanel, and Eric Thierry. Fully asynchronous behavior of double-quiescent elementary cellular automata. *Theoretical Computer Science*, 362:1–16, 2006.

[GKL87] Peter Gács, Georgii L. Kurdiumov, and Leonid A. Levin. One-dimensional homogeneous media dissolving finite islands. *Problemy Peredachi Informatsii*, 14:92–96, 1987.

[LB95] Mark Land and Richard K. Belew. No perfect two-state cellular automata for density classification exists. *Physical Review Letters*, 74(25):5148–5150, 1995.

[MCH94] Melanie Mitchell, James P. Crutchfield, and Peter T. Hraber. Evolving cellular automata to perform computations: Mechanisms and impediments. *Physica D*, 75:361–391, 1994.

[MdO05] Claudio L.M. Martins and Pedro P.B. de Oliveira. Evolving sequential combinations of elementary cellular automata rules. In Mathieu S. Capcarrere, Alex A. Freitas, Peter J. Bentley, Colin G. Johnson, and Jon Timmis, editors, *Advances in Artificial Life*, volume 3630 of *Lecture Notes in Computer Science*, pages 461–470. Springer Berlin Heidelberg, 2005.

[OMdCF09] Gina M. B. Oliveira, Luiz G. A. Martins, Laura B. de Carvalho, and Enrique Fynn. Some investigations about synchronization and

density classification tasks in one-dimensional and two-dimensional cellular automata rule spaces. *Electronic Notes in Theoretical Computer Science*, 252:121–142, 2009.

[Pac88] Norman H. Packard. *Dynamic Patterns in Complex Systems*, chapter Adaptation toward the edge of chaos, pages 293 – 301. World Scientific, Singapore, 1988.

[SB09] Christopher Stone and Larry Bull. Evolution of cellular automata with memory: The density classification task. *BioSystems*, 97(2):108–116, 2009.

[SOS09] Martin Schüle, Thomas Ott, and Ruedi Stoop. Computing with probabilistic cellular automata. In *ICANN '09: Proceedings of the 19th International Conference on Artificial Neural Networks*, pages 525–533, Berlin, Heidelberg, 2009. Springer-Verlag.