Stochastic Cellular Automata Solve the Density Classification Problem with an Arbitrary Precision

Nazim A. Fatès

Stochastic Cellular Automata Solve the Density Classification Problem with an Arbitrary Precision

Nazim Fatès
INRIA Nancy – Grand Est, LORIA, Nancy Université
nazim.fates@loria.fr

January 7, 2011

Abstract

The density classification problem consists in using a binary cellular automaton (CA) to decide whether an initial configuration contains more 0s or 1s. This problem is known for having no exact solution in the case of binary, deterministic, one-dimensional CA. Stochastic cellular automata have been studied as an alternative for solving the problem. This paper is aimed at presenting techniques to analyse the behaviour of stochastic CA rules, seen as a "blend" of deterministic CA rules. Using analytical calculations and numerical simulations, we analyse two previously studied rules and present a new rule. We estimate their quality of classification and their average time of classification. We show that the new rule solves the problem with an arbitrary precision. From a practical point of view, this rule is effective and exhibits a high quality of classification, even when the simulation time is kept small.

Introduction

The density classification problem is one of the most studied inverse problems in the field of cellular automata. Informally, it requires that a binary cellular automaton — or more generally a discrete dynamical system — decides whether an initial binary string contains more 0s or more 1s. In its classical formulation, the cells are arranged in a ring and each cell can only read its own state and the states of the neighbouring cells. The challenge is to design a behaviour of the cells that drives the system to a uniform state, that consists of all 1s if the initial configuration contained more 1s and all 0s otherwise. In short, the convergence of the cellular automaton should decide whether the initial density of 1s was greater or lower than 1/2.

Although the task looks trivial, it has attracted a considerable amount of research since its formulation by Packard [13]. 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, that is, by using an information that is limited to the close neighbours of a cell. Moreover, as CA are homogeneous by nature (the cells obey the same law), 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.*, [11]). The Gacs-Kurdymov-Levin (GKL) rule, which was originally designed in the purpose of resisting small amounts of noise [14, 5], 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 [9]. However, this did not stop the search for efficient CA as nothing was known about how well a rule could perform. In particular, it was asked whether an upper bound on a rule quality would exist. 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.* [4, 12] 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 [3]. Fukś showed that running two CA rules successively would also provide an acceptable solution [7]. This issue was further explored by Martins and Oliveira, who discovered various couples and triples of rules that solve the problem when applied sequentially and for a given number of steps [10]. Some authors also proposed to embed a memory in the cells, which is another method for enhancing the abilities of the rules [1, 16].

However, all of these solutions break the original specification of the problem, where the cells have only two states and obey a homogeneous rule. The use of stochastic (or probabilistic)[1] CA is an interesting alternative that complies with these two conditions. 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. This research path was opened by Fukś who exhibited a rule which acts as a "stochastic copy" of the state of the neighbouring cells [8]. 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 [15]. This allows the system to converge to its goal more efficiently, but the convergence rates still remain bounded by some intrinsic limitations (see

[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.

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, with a probability of success arbitrarily close to 1. This result answers negatively the open question to whether there exists an upper bound on the success rate one can reach. 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, η, 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 η is set closer to 0. To evaluate the "practical" use of our rule, we perform numerical simulations. Results show that this rule attains qualities of classification that have been out of reach so far.

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, *i.e.*, 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 series of independent Bernoulli random variables, *i.e.*, $\mathcal{B}_c^t(p)$ is a random variable that equals to `1` with probability p and `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 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 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 requires 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. [3] and 97.8% in Ref. [9]). 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 `1` and a probability $1-p$ to be state `0`.

(c) The `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.

[2]Note that defining rigorously the series of random variables x^t obtained from F would require to introduce advanced tools from the probability theory. In particular, one should define a space of realisation Ω and always consider probability measures on Ω and for all $\omega \in \Omega$, define the random variables with respect to the configurations $x^t(\omega) \in \mathcal{E}_n$. For the sake of simplicity, and as it is frequently done, the parameter ω is omitted and the random variables are defined only with regard to their probability of realisation on Ω.

Formally,

$$\forall x \in \mathcal{E}_n, \quad \mu_{\mathrm{b}}(x) = \frac{1}{2^n} \quad ; \quad \mu_{\mathrm{d}}(x) = \int_0^1 p^k (1-p)^{n-k} dp \quad ; \quad \mu_1(x) = \frac{1}{n+1} \cdot \frac{1}{\binom{n}{k}}$$

where $k = |x|_1$ is the number of 1s in x.

Proposition 1. *The d-uniform distribution μ_{d} and the 1-uniform distribution μ_1 are equivalent.*

This equality can be established by identifying $\mu_{\mathrm{d}}(x)$ to the values of the so-called 'Beta function' or 'Euler's integral of first kind'. A direct proof is given in Appendix A.

Given a ring size n, a distribution μ on $\mathcal{E}_n$, the *quality* Q of a classifier $\mathcal{C}$ is defined by:

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

In this paper, we evaluate the performance of a classifier using a binomial quality Q_{b}, defined with the distribution μ_{b} and a d-uniform quality Q_{d}, defined with the distribution μ_{d}. Intuitively, we see that for most classifiers, we will have $Q_{\mathrm{d}} > Q_{\mathrm{b}}$. Indeed, when we take the binomial distribution, as n grows to infinity, most initial configurations of $\mathcal{E}_n$ have a density close to $1/2$ and are generally more difficult to classify that configuration with densities close to 0 or 1. The d-uniform distribution avoids this difficulty 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}.

1.3 Structure of the sECA space

Obviously, the classical deterministic ECA are particular sECA with a local rule that takes its values in $\{0, 1\}$. 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 points of a hypercube, 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 $\sum_{i=0}^{k} w_i = 1$, the barycenter of the sECA (f_i) with weights w_i is the sECA g defined with:

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

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

As a consequence, one may choose to express an sECA as a barycenter of other ECA. The most intuitive basis of the sECA space is formed by the 8 ECA that have only one transition that leads to 1: the coordinates correspond to the values $f(x, y, z)$. Equally, one may express a sECA as a barycenter of the 8 (deterministic) ECA that have only one *active* transition, *i.e.*, only one change of state in their transition table. Such ECA are labelled A, B, ..., H according to the notation introduced in Ref. [6] 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. [6]). For instance, 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.

This *transition code* also allows us to easily write the conservation laws of a stochastic CA and to estimate some aspects of its global behaviour. To do this analysis, we write $a(x) = |x|_{000}$, $b(x) = |x|_{001}$, ..., $h(x) = |x|_{111}$ (see Tab. 1) and drop the argument x when there is no ambiguity. The following equalities hold [6]:

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

We now detail how to use these tools to analyse the behaviour of an sECA.

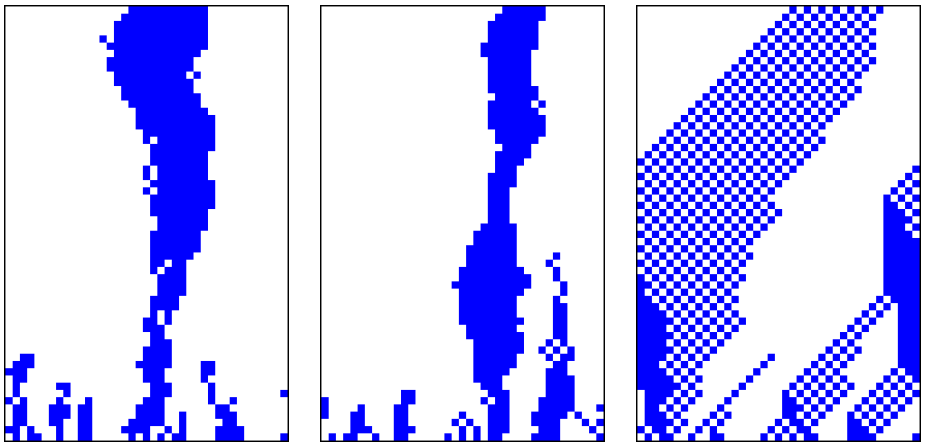
2 Fukś Density Classifier

To start examining how stochastic CA solve the density classification problem, let us first consider the probabilistic density classifier proposed by Fukś [8]. For $p \in (0, 1/2]$, the local rule $\mathbf{C}_1$ is defined with the following transition table:

(x, y, z)	000	001	010	011	100	101	110	111
$f(x, y, z)$	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:

$$\begin{aligned} \mathbf{C}_1 &= [0, p, p, 2p, 2p, p, p, 0]_{\mathrm{T}} \\ &= p \cdot \mathsf{BDEG} + p \cdot \mathsf{CDEF} \end{aligned}$$



$\mathbf{C}_1, p = 0.25$ $\mathbf{C}_2, \epsilon = 0.8$ $\mathbf{C}_3, \eta = 0.25$

Figure 1: Space-time diagrams showing the evolution of the $\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3$ classifiers with $n = 39$, and same initial density ~ 0.4. Time goes from bottom to top ; white cells are 0-cells and blue cells are 1-cells. (left & middle) evolution will most probably end with a good classification (**0**); (right) evolution will with end with a good classification with probability 1 (an archipelago has been reached).

where the rules[3] BDEG(**170**) and CDEF(**240**) are the left and right shift respectively. This means that Fuks' rule can be interpreted as applying, for each cell 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 $\boldsymbol{C}_1$ set with $p \in (0, 1/2]$,*

$$\forall x \in \mathcal{E}_n,\ \mathsf{G}(x) = \max\{\rho(x), 1 - \rho(x)\} \quad \textit{and} \quad T_{\mathrm{b}} \leq \frac{1}{4p} \cdot n^2$$

The relationship on $\mathsf{G}(x)$ was observed experimentally with simulations and partially explained by combinatorial arguments [8]. 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 [6] and completes the results established by Fukś. The proof stands on the following lemma:

Lemma 1. *For a sequence of random variables $(x^t)_{t\in\mathbb{N}}$ that describes the evolution of a stochastic CA with the initial condition $x \in \mathcal{E}_n$, let M be a mapping $M : \mathcal{E}_n \to \{0, \ldots, m\}$ where m is any integer, and let (X_t) be the sequence of*

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

random variables defined by $\forall t, X_t = M(x^t)$. *If* X_t *and* $\Delta X_{t+1} = X_{t+1} - X_t$ *verify that:*

- *the stochastic process* (X_t) *is a martingale on* $\{0, \ldots, m\}$, *that is, for a filtration* $\mathcal{F}_t$ *adapted to* (X_t), $\mathbb{E}\{\, \Delta X_{t+1} \,|\mathcal{F}_t \,\} = 0$,
- $X_t \in \{1, \ldots, m-1\} \implies \mathrm{var}\{\, \Delta X_{t+1} \,\} > v$,

then:

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

and the absorbing time of the process $T(x) = \min\{t : X_t = 0 \text{ or } X_t = m\}$ *is finite and obeys:*

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

where $q = \mathbb{E}\{X_0\} = M(x)$.

Sketch. A similar lemma was formulated for studying asynchronous CA [6]. The main elements of its proof are: (1) to note that T is a stopping time, (2) to use the Optional Stopping Time theorem to calculate $\mathbb{E}\{X_T\}$, (3) to note that the process $Y_t = X_t^2 - v \cdot t$ is a submartingale and use again the Optional Stopping Time theorem. (See Appendix B). □

Proof of Theorem 1. We simply take $X_t = |x^t|_1$ and show that Lemma 1 applies to X_t. We write:

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

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

Second, we assume that $X_t \in \{1, \ldots, n-1\}$. It implies that $x^t \notin \{\mathbf{0}, \mathbf{1}\}$, that is, x^t is not a fixed point. 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 series of the Bernoulli random variables of Eq. (1). Using the independence of these variables and $\mathrm{var}\{\, \mathcal{B}(p) \,\} = p(1-p)$, Eq. (3) gives:

$$\begin{aligned} \mathrm{var}\{\, \Delta X_{t+1} \,\} &= (b + c + f + g) \cdot p(1-p) + (d + e) \cdot 2p(1-2p) \\ &= p \cdot [(s_1 + 2s_2) - (s_1 + 4s_2) \cdot p] \end{aligned}$$

with $s_1 = b + c + f + g$ and $s_2 = d + e$. Using Eq. (2) and noting that the value of n is odd, we remark that there exists a 00 or 11 pattern 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 $\mathrm{var}\{\, \Delta X_{t+1} \,\} \geq p$.

Lemma 1 thus applies by taking $v = p$ and $m = n$. Finally, we find that the probability that the process stops on $X_T = n$, that is, on the fixed point $\mathbf{1}$, is equal to the initial density $\rho(x) = |x|_1/n$. We also find that :

$$\forall x \in \mathcal{E}_n,\, T(x) \leq \frac{|x|_1\,(n - |x|_1)}{p} \quad \text{and} \quad T_{\rm b} \leq \frac{n^2}{4p}$$

□

From this result, we derive that the probability of good classification of any configuration x is equal to $\mathsf{G}(x) = \max\{\rho(x), 1-\rho(x)\}$. The d-uniform quality of $\mathbf{C}_1$ is thus equal to $Q_{\rm d}(n) = 3/4$ (obtained y a simple integration). For $n = 2k+1$, the binomial quality of $\mathbf{C}_1$ is equal to: (see Appendix C) $Q_{\rm b}(n) = 1/2 + \binom{2k}{k}/2^{2k+1}$. This formula explains why the quality of classification of $\mathbf{C}_1$ quickly decreases as the ring size n increases. For instance for $n = 49$, we have: $Q_{\rm b}(n) = 0.557$, that is, the gain of using $\mathbf{C}_1$ compared to a random guess is less than 6%. For the reference value $n = 149$, the gain drops down to 3.3% (see Tab. 2 p. 13).

3 Schüle Density Classifier

We now consider the probabilistic density classifier proposed by Schüle et al [15]. It was designed to improve the convergence of the system towards a fixed point. For $\varepsilon \in (0, 1]$, the local rule $\mathbf{C}_2$ is defined with the following transitions:

(x, y, z)	000	001	010	011	100	101	110	111
$f(x, y, z)$	0	$1-\varepsilon$	$1-\varepsilon$	ε	$1-\varepsilon$	ε	ε	1

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:

$$\begin{aligned}\mathbf{C}_2 &= [0, 1-\varepsilon, 1-\varepsilon, \varepsilon, \varepsilon, 1-\varepsilon, 1-\varepsilon, 0]_{\rm T} \\ &= (1-\varepsilon)\cdot \mathsf{BCFG} + \varepsilon \cdot \mathsf{DE}\end{aligned}$$

where rule $\mathsf{BCFG}(\mathbf{150})$ is the rule that implements a XOR function with three neighbours and DE 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. 1). This rule is invariant under both the left-right and the 0-1 symmetries (as we have: $p_{\rm B} = p_{\rm C} = p_{\rm F} = p_{\rm G}$, $p_{\rm A} = p_{\rm H}$ and $p_{\rm D} = p_{\rm E}$).

Theorem 2. *For the classifier $\boldsymbol{C}_2$, for $\varepsilon = 2/3$,*

$$\forall x \in \mathcal{E}_n,\, \mathsf{G}(x) = \max\,\{\rho(x), 1-\rho(x)\} \quad \textit{and} \quad T_{\rm b} \leq 9/2 \cdot n^2$$

The relationship on $\mathsf{G}(x)$ was proved under the mean-field approximation [15]. We now propose to re-derive this result more directly.

Proof. Let us take $X_t = |x^t|_1$ and show that Lemma 1 applies to X_t. We have:

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

Using Eq. (2), we obtain:

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

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

Let us now assume that $X_t \in \{1,\ldots,n-1\}$. This implies that x^t is not a fixed point and that $b+c+d+e+f+g \geq 1$. 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 series of Bernoulli random variables of Eq. (1). This results in:

$$\begin{aligned}\mathrm{var}\big\{\,\Delta X_{t+1}\,\big\} &= (b+c+d+e+f+g)\cdot\varepsilon(1-\varepsilon)\\ &\geq \varepsilon(1-\varepsilon)\end{aligned}$$

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 $X_0/n = \rho(x)$ and that $\forall x \in \mathcal{E}_n$, $T(x) \leq \frac{|x|_1(n-|x|_1)}{\epsilon(1-\epsilon)}$, which implies

$$T_{\mathrm{b}} \leq \frac{n^2}{4\epsilon(1-\epsilon)} \leq \frac{9n^2}{2}$$

□

Equation (4) also allows us to understand the general behaviour of Schüle's classifier $\mathbf{C}_2$ 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 0s than isolated 1s, 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 and al. [15], the system is then in a metastable state: although the classification time is finite, the system is always attracted towards a density 1/2. Last, but not least, we think that for $\epsilon > 2/3$, only isolated 0s or 1s 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 $\boldsymbol{C}_2$ set with $\varepsilon > 2/3$, the quality of classification $Q_{\mathrm{b}}(n)$ is bounded. More precisely:*

$$\forall\epsilon > 2/3, \forall x \in \mathcal{E}_n : |x|_{010} = |x|_{101} = 0, \mathsf{G}(x) = \max\{\rho(x), 1-\rho(x)\}$$

and

$$\forall \epsilon > 2/3, \forall x \in \mathcal{E}_n, \mathsf{G}(x) \leq \max\{\rho^*(x), 1-\rho^*(x)\}$$

where $\rho^*(x) = (\Phi^{\infty}_{\mathrm{MAJ}}(x))$ *is the density attained by an asymptotic evolution of* x *under the majority rule.*

Theses hypotheses are partially confirmed by numerical simulations (see Tab. 2 and Appendix E)). We also verified experimentally that for $\epsilon \to 1$, the quality approaches an asymptotic limit while the average classification time diverges. We leave a rigorous proof this statement for future work (see Appendix D) and now present a rule that does not suffer from such limitations.

4 A New Rule for Density Classification

For $\eta \in (0, 1]$, let us consider the following sECA:

(x, y, z)	000	001	010	011	100	101	110	111
$f(x, y, z)$	0	0	0	1	$1-\eta$	1	η	1

With the transition code, this writes:

$$\begin{aligned} \mathbf{C}_3 &= [0, 0, 1-\eta, 1, 1, 0, 1-\eta, 0]_{\mathrm{T}} \\ &= \eta.\mathsf{DE} + (1-\eta).\mathsf{CDEG} \end{aligned}$$

For $\eta = 0$ we have $\mathsf{CDEG}(\mathbf{184})$, which is a well-known rule, often called the "traffic" rule. This rule is number conserving, *i.e.*, the number of 1s is conserved as the system evolves (see *e.g.*, [2]). Observing the evolution of the 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). For $\eta = 1$, we have the majority rule $\mathsf{DE}(\mathbf{232})$. For $\eta \in (0, 1)$, the effect of the rule is the same as applying, for each cell and at each time step, the traffic rule with probability $1-\eta$ and the majority rule with probability η (see Fig. 1). This combination of rules has a surprising property: although 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.*

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. Let us denote by y a potential successor of x. Let C be the 1-cells in y: $C = \{c \in \mathcal{L} : y_c = 1\}$. If we look in x at the local predecessor pattern of a cell $c \in C$, we have $(x_{c-1}, x_c, x_{c+1}) \in \{100, 101, 011, 111\}$ by examining the transition function of $\mathbf{C}_3$, and, as x is a 1-archipelago, $(x_{c-1}, x_c, x_{c+1}) \in \{100, 101\}$. As these two patterns do not overlap, it is not possible to have two successive cells of $\mathcal{L}$ contained in C and y is a 1-archipelago.

Second, we remark that the number of 1s of x^t is a non-increasing function of t. 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 $\rho(x) < 1/2$. □

The second interesting property of $\mathbf{C}_3$ is its ability to make any configuration evolve to an archipelago with a probability that can be made as large as wanted.

Lemma 3. *For every $p \in [0, 1)$, there exists a setting η of the classifier $\boldsymbol{C}_3$ such that for every configuration $x \in \mathcal{E}_n$, the probability to evolve to an archipelago x_A such that $d(x_A) = d(x)$ is greater than p.*

Proof. The proof relies on the well-known property of the traffic rule to evolve to an archipelago in at most $n/2$ steps. Let us denote by Φ the global transition function of CDEG and write $y^t = \Phi^t(x)$, that is, (y^t) is the series of configurations obtained with x as an initial condition. From the properties of the traffic rule, we have that $\rho(y^t) = \rho(x)$ and that $y^{\lceil n/2 \rceil}$ is an archipelago[4] (see *e.g.*, Ref. [3] Lemma 4).

For a given p and given n, without loss of generality, let us consider a configuration x such that $\rho(x) < 1/2$. Let us now evaluate the probability that rule $\mathbf{C}_3$ does *not* behave like the traffic rule in the first $T = \lceil n/2 \rceil$ steps. Formally let $D^t = card\{c \in \mathcal{L} : x_c^t \neq y_c^t\}$.

Comparing the transition rules of CDEG and $\mathbf{C}_3$, we see that differences in the evolution of the two rules can only occur for cells where transitions C and G apply, that is, cells that have a 100 and 110 neighbourhood. As we have $b = c$ and $f = g$, and $b + f + g + c \leq n$, we write $b + f \leq \lceil n/2 \rceil$. For such cells, differences of evolution occur with a probability η, which implies that, at each time step, the probability $p_{\text{diff}} = \Pr\{D^t > 0\}$ that the evolution of $\mathbf{C}_3$ and CDEG differ on T steps is upper-bounded by: $p_{\text{diff}} \leq \eta^{T \cdot \lceil n/2 \rceil} \leq \eta^{T^2}$. The probability $P_{\text{eq}} = \Pr\{D^1 = 0, \ldots, D^T = 0\}$ that the two rules evolve identically on T steps is thus greater than or equal to $1 - p_{\text{diff}}$ and we find that it is sufficient to set: $\eta < 1 - p^{\frac{1}{T^2}}$ to guarantee that $P_{\text{eq}} > p$, *i.e.*, that the probability to reach a 1-archipelago is greater than p. As the traffic rule is number-conserving, the archipelago has the same density as the initial configuration. □

This inequality shows that, by taking η small enough, the probability that a configuration x with $\rho(x) < 1/2$ evolves to a 1-archipelago can be made arbitrarily small. This allows us to state our main result.

Theorem 3. *For all $p \in [0, 1)$, there exists a setting η of the classifier $\boldsymbol{C}_3$ such that $\forall x \in \mathcal{E}_n$, $\mathsf{G}(x) \geq p$. As a consequence, $\forall n \in 2\mathbb{N} + 1$, setting $\eta \to 0$ implies $Q_{\text{b}}(n) \to 1$.*

Proof. Combining the two previous lemmas to prove the theorem is straightforward: for η small enough, the system evolves to an archipelago that has the same

[4] As remarked by an anonymous referee, $\lfloor n/2 \rfloor$ steps should be sufficient.

Table 2: Results for $n = 149$; averages on 10 000 samples, the values 53.3 and 75.0 are calculated.

model	setting	Q_{b} (in%)	Q_{d} (in%)	T_{b}
$\mathbf{C}_1$	$p = 0.25$	53.3	75.0	4638
$\mathbf{C}_1$	$p = 0.48$	53.3	75.0	2652
$\mathbf{C}_1$	$p = 0.5$	53.3	75.0	8985
$\mathbf{C}_2$	$\epsilon = 0.7$	54.0	80.1	4061
$\mathbf{C}_2$	$\epsilon = 0.8$	55.1	83.8	6223
$\mathbf{C}_2$	$\epsilon = 0.9$	56.6	85.8	11887
$\mathbf{C}_3$	$\eta = 0.1$	82.4	98.1	517
$\mathbf{C}_3$	$\eta = 0.01$	91.0	99.1	4950
$\mathbf{C}_3$	$\eta = 0.005$	93.4	99.3	9981

density as the initial condition (Lemma 3). It is then necessarily well-classified as it will progressively "drift" towards the appropriate fixed point (Lemma 2). However, we remark that the time taken to reach the fixed point increases as η decreases. □

The analytical estimation of the quality of $\mathbf{C}_3$ and its time of convergence is more complex than for Fukś and Schüle classifiers. Table 2 shows the values of Q_{b}, Q_{d} 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, has not been yet reached for one-dimensional systems (see *e.g.* [4, 12]). Another major point regards the classification time of $\mathbf{C}_3$: 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 Appendix E).

Conclusion

This paper presented an analysis of three parametrised stochastic cellular automata that solve the density classification problem. We first re-examined the behaviour of two previously studied rules using the analytical techniques originally developed for asynchronous cellular automata. This allowed us to analytically estimate their quality of classification and average convergence time.

We then presented a new parametrised rule that solves this problem with an arbitrary precision. This shows that no upper bound exists on how well a rule can perform, at least for stochastic cellular automata. Numerical simulations were performed to estimate the behaviour of the rules for various settings. It was observed that for the standard reference value $n = 149$, quality of classification of over 90% could be attained, with an average convergence time of the order of a few thousand steps.

Far giving a definitive answer to the problem, the existence of this solution suggests that the quality of classification cannot be taken as the unique criterion for evaluating the classifiers. Instead, it is some trade-off between quality and time to classify that has now to be searched for. We are particularly interested in knowing whether there are other couples (or t-uples) of rules that can be "blended" to obtain similar or even better results. The "blending" technique could also be used with rules that have a larger radius, and on lattices with two or more dimensions.

References

[1] Ramón Alonso-Sanz and Larry Bull, *A very effective density classifier two-dimensional cellular automaton with memory*, Journal of Physics A **42** (2009), no. 48, 485101.

[2] Nino Boccara and Henryk Fukś, *Number-conserving cellular automaton rules*, Fundamenta Informaticae **52** (2002), no. 1-3, 1–13.

[3] Mathieu S. Capcarrere, Moshe Sipper, and Marco Tomassini, *Two-state, $r = 1$ cellular automaton that classifies density*, Phys. Rev. Lett. **77** (1996), no. 24, 4969–4971.

[4] 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** (2006), no. 1-3, 35 – 43.

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

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

[7] Henryk Fukś, *Solution of the density classification problem with two cellular automata rules*, Physical Review E **55** (1997), no. 3, R2081–R2084.

[8] ———, *Nondeterministic density classification with diffusive probabilistic cellular automata*, Physical Review E **66** (2002), no. 6, 066106.

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

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

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

[12] 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** (2009), 121–142.

[13] Norman H. Packard, *Dynamic patterns in complex systems*, ch. Adaptation toward the edge of chaos, pp. 293 – 301, World Scientific, Singapore, 1988.

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

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

[16] Christopher Stone and Larry Bull, *Evolution of cellular automata with memory: The density classification task*, BioSystems **97** (2009), no. 2, 108–116.

Acknowledgement

The author wishes to express his gratitude to I. Markovici and L. Gerin for their careful reading of the manuscript.

A Equivalence of μ_{d} and μ_{k}

Let us show the equality:

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

Proof. The first technique is to remark that $\mu_{\mathrm{d}}(x)$ corresponds to the calculus of the Beta function $\mathcal{B}(k+1, n-k+1)$, which is defined on real numbers. The analytical form of the function was studied by Euler and Legendre and it is generally obtained by applying a change of variable and using trigonometric relations. We propose here another technique that exploits the fact that we want to evaluate this function only on integer numbers. 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:

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

we find that

$$\sum_{i=0}^{n} M_i = \int_0^1 1 dp = 1$$

On the other hand, using the integration by parts of M_i gives

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

By noting that the first term is equal to 0 and that $\binom{n}{i} \cdot \frac{n-i}{i+1} = \binom{n}{i+1}$, we obtain the equality $M_i = M_{i+1}$, which gives $M_i = \frac{1}{n+1}$ and confirms the result $\mu_{\mathrm{d}} = \mu_{\mathrm{k}}$. □

B Convergence Lemma 1

Proof. First, let us show that T is finite. We remark that, as ΔX_{t+1} is a process which expectancy is 0 and whose variance is greater than v, 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 all $t < T$, 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. The time of convergence to a fixed point is thus upper-bounded by the time of first success of a geometric law of parameter α, it is thus finite.

As T is finite, we apply Doob's optional stopping theorem on the martingale (X_t) and write

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

On the other hand we have:

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

Combining the two equations gives us : $\Pr\{X_T = n\} = q/m$.

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:

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

We also have :

$$\mathrm{var}\{\, \Delta X_{t+1} \,\} = \mathbb{E}\{\, \Delta X_{t+1}^2 \,\} - \mathbb{E}\{\, \Delta X_{t+1} \,\}^2 = \mathbb{E}\{\, \Delta X_{t+1}^2 \,\}$$

and

$$\mathbb{E}\{\, \Delta X_{t+1} \cdot X_t \,|\mathcal{F}_t \,\} = \mathbb{E}\{\, \Delta X_{t+1} \,|\mathcal{F}_t \,\} \cdot X_t = 0$$

Finally, using the Lemma hypothesis $\mathrm{var}\{\, \Delta X_{t+1} \,\} \geq v$, we obtain: $\mathbb{E}\{\, Y_{t+1} - Y_t \,\} \geq 0$, that is, (Y_t) is a submartingale. Again, applying the Optional Stopping Theorem gives:

$$\mathbb{E}\{\, Y_T \,\} \geq \mathbb{E}\{\, Y_0 \,\}$$

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:

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

which gives us the expected result: $\mathbb{E}\{\, T \,\} \leq \frac{q(m-q)}{v}$ □

C d-binomial quality of $\mathbf{C}_1$

Proof. We take an odd ring size $n = 2k + 1$. By definition of the d-binomial quality Q_{b}:

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

Grouping the configurations with equal number of 1s gives:

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

which we decompose into:

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

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

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

We then obtain:

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

□

D Justification of Proposition 2

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 above).

E Scaling law of T_{b}

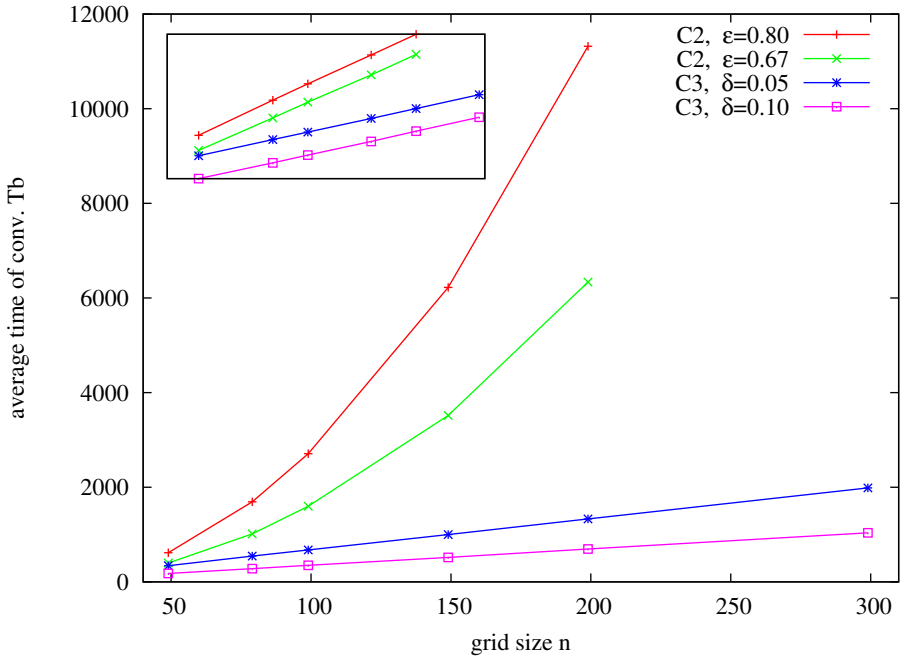


Figure 2: Average classification time T_{b} as a function of ring size n for different settings of $\mathbf{C}_2$ and $\mathbf{C}_3$. inset: same data with logscale