Solving the density classification problem with a large diffusion and small amplification cellular automaton

Raimundo Briceño,[1] Pablo Moisset de Espanés,[2] Axel Osses,[1,2] and Ivan Rapaport[1,2]

[1]*Departamento de Ingeniería Matemática, Universidad de Chile*
[2]*Centro de Modelamiento Matemático (UMI 2807 CNRS), Universidad de Chile*

One of the most studied inverse problems in cellular automata (CAs) is the density classification problem. It consists in finding a CA such that, given any initial configuration of 0's and 1's, it converges to the all-1 fixed point configuration if the fraction of 1's is greater than the critical density 1/2 and it converges to the all-0 fixed point configuration otherwise. In this paper we propose an original approach to solve this problem by designing a CA Φ inspired by two mechanisms ubiquitous in nature: diffusion and non-linear sigmoidal response. This solution works for any system size, any dimension and any critical density.

I. INTRODUCTION

Cellular automata (CAs) are discrete dynamical systems. They were introduced by John von Neumann [17] after a suggestion of Stanislaw Ulam [23]. We consider here finite CAs. More precisely, n^d cells arranged uniformly spaced in the d-dimensional torus and following a local rule identical to every cell. This local rule, which specifies how the state of each cell is updated as a function of the states of its neighbor cells, is applied in parallel and in discrete time steps.

One of the most studied inverse problems in CAs is the *density classification problem*. The challenge is to find a CA such that, given any initial configuration x^0 of 0's and 1's, it converges to the all-1 fixed point configuration if the fraction of 1's in x^0 is greater than ρ and it converges the the all-0 fixed point configuration otherwise. The number $0 < \rho < 1$ denotes the *critical density*.

The problem was first formulated for dimension $d = 1$ (a ring) and critical density $\rho = 1/2$ [19]. The best-known two-state CA for tackling this instance of the density classification problem is called GKL and it was designed for a ring of $n = 149$ cells [11, 12] (the original purpose of GKL was to resist small amounts of noise). The performance of GKL was very good but not perfect. In fact, an impossibility result was proved in [14]: there is no perfect density classifier with two states.

The impossibility of finding perfect classifiers led many researchers to use different evolutionary computation approaches to evolve good *approximate solutions* [16, 18, 24]. But in order to obtain perfect density classifiers, researchers were forced to modify the original problem. One idea was to change the output specifications [4]. Another idea was to allow the existence of more than one local rule [9, 15] or to embed a memory on the cells [1, 21]. A very subtle and interesting relaxation of the original problem is related to determinism. In fact, following the works of [10] and [20], Fatès designed a two-state stochastic CA that solves the density classification problem with arbitrary precision [8] and even in any dimension [2].

The idea of the present paper is to use a continuous approach for solving *deterministically* the density classification problem. More precisely, our idea is to use *local averaging* and *saturation*, a process represented by a bistable heat equation. This bistable model, which exhibits two stable critical points (0 and 1), is a particular case of a reaction-diffusion equation widely used for studying phase transition and front propagation in spatial ecology [3], physiology [13], chemistry and physics [25]. The *large diffusion small amplification* CA Φ that we define in this work is a discretization of such bistable nonlinear heat equation (similar approaches have been previously used [5, 7]).

Two parameters characterize CA Φ: the amount of nonlinearity σ (or amplification factor) and the number of states s (or discretization factor). The main theoretical result of this paper is that –given arbitrary parameters n, d and ρ– for every $\varepsilon > 0$, there exist σ and s such that the large diffusion and small amplification CA $\Phi(\sigma, s)$ solves the density classification problem with accuracy ε.

CA Φ presents many advantages with respect to existing solutions: it works in any dimension d for an arbitrary number of cells n^d and with any critical density ρ, it allows an intuitive/physical interpretation together with a deep theoretical analysis and it maintains the same classification properties for a wide range of different averages and nonlinear amplifications.

But the most clear advantage of CA Φ is its adaptive success ratio. In fact, in Section VI we compare Φ not only with GKL but also against a variant of the elementary CA Rule 184 [22]. This CA, that we denote 184^*, is particularly interesting because it is similar to CA Φ in two senses: (1) it is not the result of some evolutionary computation process, (2) cells are provided with a "long term" memory (the size of the memory is the number of states).

The theoretical result concerning the existence of CA Φ does not give us any indication about the critical values of s and σ (as a function of n). So, in order to analyze with more precision the performances of Φ, GKL and 184^*, we run simulations using a fixed set of extremely hard instances (initial configurations). These initial configurations are generated by randomly permuting m 1's and $n-m$ 0's. It turns out that CA Φ classifies all instances except a few cases when $m = 75$ and $n = 149$. On the other hand, both GKL and 184^* have a success ratio slightly above 50% (this contrasts with the $\sim 80\%$ ratio observed when independent probabilities are used in the instance generation process).

It is important to point out that CA Φ exhibits a disadvantage: its low convergence time (compared to other classifiers). This disadvantage is interesting. First, it leaves as an open question the existence of a trade-off between time complexity and success ratio. Second, it shows us why it is often necessary to design discrete models from scratch (instead of discretizing existing continuous models).

II. THE DENSITY CLASSIFICATION PROBLEM

Let $[n]^d = \{0, \ldots, n-1\}^d$ represent a set of n^d cells arranged uniformly spaced in the d-dimensional torus (for instance, $[n]^1$ is the ring, $[n]^2$ is the two-dimensional grid with periodic boundary conditions, etc).

Let $v_0 \in [n]^d$ be a cell and let r be a natural number. The r-*neighborhood* of v_0 is $N_{r,d}(v_0) = \{v \in [n]^d : |v - v_0| \leq r\}$, where the differences are taken modulo n and $|u| = \sum_i |u_i|$. The size of the neighborhood is independent of v_0 and we denote it $N_{r,d}$. For instance, $N_{r,2} = 2r(r+1) + 1$.

A *configuration* $x \in [0,1]^{n^d}$ is an assignment of real numbers (that we call states) to the cells of the lattice. Later in this work we are going to restrict the set of states to a finite one. For simplicity, we write N to denote n^d.

A radius r cellular automaton (CA), that we denote Ψ, transforms a configuration $x^k \in [0,1]^N$ into a new configuration $x^{k+1} \in [0,1]^N$ by appllying in parallel, to all the cells of x^k, its local function $\psi : [0,1]^{N_{r,d}} \to [0,1]$. Hence, by fixing the local function ψ, we fix the CA Ψ.

Given a configuration $x \in [0,1]^N$, we denote by $\overline{x}$ its *mean value* (a real number). Given $\rho \in (0,1)$ and $\varepsilon > 0$ we introduce the sets

$$\begin{aligned}\mathcal{X}_{\rho-\varepsilon} &= \{x \in [0,1]^N \,|\, \overline{x} < \rho - \varepsilon\},\\ \mathcal{X}_{\rho+\varepsilon} &= \{x \in [0,1]^N \,|\, \overline{x} > \rho + \varepsilon\}.\end{aligned}$$

Definition 1 (Generalized density classification problem) *Given $\rho \in (0,1)$ and $\varepsilon > 0$, we say that a CA Ψ solves the density classification problem with accuracy ε if, regardless of the initial configuration x^0, the repeated application of Ψ converges to the configuration of only $0's$ if $\overline{x}^0$ is less than $\rho - \varepsilon$ and converges to the configuration of only $1's$ if $\overline{x}^0$ is greater than $\rho + \varepsilon$. That is:*

$$\begin{aligned}\forall x^0 \in \mathcal{X}_{\rho-\varepsilon},\ \lim_{k\to\infty} x^k &= [0 \cdots 0]^T \quad \textit{and}\\ \forall x^0 \in \mathcal{X}_{\rho+\varepsilon},\ \lim_{k\to\infty} x^k &= [1 \cdots 1]^T.\end{aligned}$$

Note that a configuration of the form $[c \cdots c]^T$ denotes the all-c vector (the letter T stands for transposition). The problem just defined is a generalization of the original density classification problem in more than one sense. First, in our definition the initial configuration is arbitrary (not restricted to 0's and 1's). Also, the dimension d, the number of cells N, the radius r and the critical density ρ are also arbitrary. Note that the GKL local rule was designed for the specific values $d = 1$, $r = 3$, $n = 149$ and $\rho = 1/2$ [11].

III. THE LARGE DIFFUSION AND SMALL AMPLIFICATION CA Φ

In this section we define the large diffusion and small amplification CA Φ. Its local rule ϕ is based on the discretization of a bistable nonlinear heat equation. More precisely, given a critical density $\rho \in (0,1)$, the idea is to build a local rule based on a discrete version of the following equation:

$$u_t - \nu \Delta u = \gamma\, b_\rho(u), \tag{1}$$

where $u(x,t)$ is the state at time $t \geq 0$ of the cell at point x in a domain $\Omega = (0,1)^d$ with periodic boundary conditions. The parameter $\nu > 0$ is a *diffusion coefficient*, $\gamma > 0$ is an *amplification parameter* and b_ρ is some suitable *bistable function*. In this paper (you can see a discussion about other choices in Appendix A) we choose the cubic polynomial:

$$b_\rho(u) = u(1-u)(u-\rho).$$

The resulting nonlinear heat-equation is called the *bistable heat equation*, since it exhibits two stable critical points (0 and 1, attractors) and one unstable critical point (ρ, repulsor).

A. Cases $d = 1$ and $d = 2$

Before presenting the CA that solves the general instance of the density classification problem, let us consider the one-dimensional case $d = 1$ with radius $r = 1$. We can discretize (1) with an explicit finite differences scheme on a uniform lattice of size $h > 0$ defined by $x_i = ih$ and discrete time steps $t_k = k\Delta t$ for some $\Delta t > 0$. Let $u_i^k \approx u(x_i + h/2, t_k)$ denote the corresponding approximate discrete values in each cell of the lattice. An explicit, first order in time and second order in space discretization of (1) by finite differences is:

$$\frac{u_i^{k+1} - u_i^k}{\Delta t} - \nu \frac{u_{i-1}^k - 2u_i^k + u_{i+1}^k}{h^2} = \gamma\, b_\rho(u_i^k),$$

where the sum in the subindices are modulo the size of the lattice (because of the periodic boundary conditions). If we define $\beta = \Delta t / h^2$ and we fix $\nu\beta = 1/3$, then we obtain the particular local rule:

$$u_i^{k+1} = \frac{1}{3}(u_{i-1}^k + u_i^k + u_{i+1}^k) + \frac{\gamma\, h^2}{3\nu}\, b_\rho(u_i^k).$$

If we denote $\sigma = \frac{\gamma\, h^2}{3\nu}$ and

$$\overline{u}_i^k = \frac{1}{3}(u_{i-1}^k + u_i^k + u_{i+1}^k),$$

then previous relation can be rewritten as:

$$u_i^{k+1} = \overline{u}_i^k + \sigma\, b_\rho(u_i^k) \tag{2}$$

Now we are in position to define the local rule ϕ of the large diffusion and small amplification CA Φ for $d = 1$. More precisely, if we define f as

$$f(x) = x + \sigma\, b_\rho(x),$$

then we can write ϕ as:

$$\phi(u_{i-1}^k, u_i^k, u_{i+1}^k) = f(\overline{u}_i^k), \tag{3}$$

In other words, the local rule ϕ is obtained by first averaging and then applying a nonlinear amplification function f. Note that, for our purposes, relations (2) and (3) are equivalent. In fact, it is easy to see that vector u converges to a constant vector in (2) if and only if it converges to the *same* constant vector in (3).

The two-dimensional case with radius one is very similar. Take $\beta = \Delta t/h^2$, fix $\nu\beta = 1/5$ and define $\sigma = \frac{\gamma\, h^2}{5\nu}$. The local update rule is:

$$u_{i,j}^{k+1} = \phi(u_{i-1,j}^k, u_{i+1,j}^k, u_{i,j}^k, u_{i,j-1}^k, u_{i,j+1}^k) = f(\overline{u}_{i,j}^k) \tag{4}$$

where $\overline{u}_{i,j}^k = \frac{1}{5}(u_{i-1,j}^k + u_{i+1,j}^k + u_{i,j}^k + u_{i,j-1}^k + u_{i,j+1}^k)$.

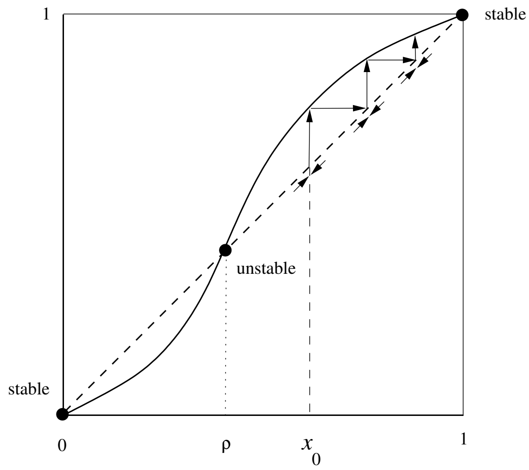

FIG. 1: Schematic view of Φ. Given an initial state, the CA corresponds to a rule obtained by first averaging neighbors (arrows towards the diagonal) and then applying a nonlinear amplification function f (arrows towards the curve). If the amplitude σ of the nonlinear function f is sufficiently small, this moves the system to one of the equilibrium points. In the figure the system is converging to the all-1 vector configuration.

Note that previous CAs, defined by (3) in one dimension and by (4) in two dimensions, correspond to discrete approximations of problem (1) that (even strictly) satisfy the corresponding Courant-Friedrich-Lewy CFL stability condition $\frac{\nu\Delta t}{h^2} = \frac{1}{3} < \frac{1}{2}$ in dimension one or $\frac{\nu\Delta t}{h^2} = \frac{1}{5} < \frac{1}{4}$ in dimension two. This guarantees convergence of the corresponding CAs to the continuous equation as h and Δt tend to zero but only in the case $\sigma = 0$ [6]. This convergence result can not be directly extended to small positive values of σ by a perturbation argument. The reason is that this small perturbation could be arbitrarily amplified (causing instability).

Nevertheless, the properties of the continuous nonlinear heat equation (1) can give us some insight and intuition about the properties of CA Φ for small σ. This is exactly the goal of the theoretical study of Section IV.

B. General case

The neighborhood of cell i is denoted by $\mathcal{N}_i$. Recall that $|\mathcal{N}_i| = N_{r,d}$. We are considering periodic boundary conditions (d-dimensional torus).

Definition 2 (Large diffusion and small amplification CA) *For small values of σ we define the local rule ϕ of the large diffusion and small amplification CA Φ as follows:*

$$\overline{x}_i^k = \frac{1}{N_{r,d}} \sum_{j\in\mathcal{N}_i} x_j^k \quad \textit{(diffusion)} \tag{5}$$

$$x_i^{k+1} = f_\sigma(\overline{x}_i^k) \quad \textit{(amplification)} \tag{6}$$

where $f_\sigma(x) = x + \sigma x(1-x)(x-\rho)$ and $\sigma = \frac{\gamma}{\nu\, n^2 N_{r,d}}$.

Recall that ν and γ are respectively the diffusion and amplification parameters of the bistable heat equation (1). We are going to prove in next section that the CA Φ given by Definition 2 solves the density classification problem for any given accuracy ε, provided that the constant factor σ is small enough. We impose the following:

$$0 < \sigma < \min\{\frac{1}{\rho}, \frac{1}{1-\rho}\}. \tag{7}$$

These bounds guarantee that f_σ is restricted to $[0,1]$ and it is monotonically increasing. Both bounds are strictly required for CA Φ to solve the density classification problem, as we explain in Section IV. Moreover, the values σ that solve the problem are typically much smaller than the upper bound, as our experiments of Section VI show.

In Section V we are going to quantize the values of x_i^k to some number s of discrete values in $[0,1]$. We say that CA Φ solves the density classification problem if, regardless of how close the average of the initial configurations is to ρ, there is a range of values for the nonlinearity σ and discretization s that will guarantee convergence to the correct answer. Formally,

Definition 3 (Solution of accuracy ε) *We say that the large diffusion and small amplification CA Φ solves the density classification problem with accuracy ε if the following property holds. For all $\varepsilon > 0$ there exists $\sigma_0 > 0$ such that for all $\sigma \le \sigma_0$ there exists s_0 such that for all $s \ge s_0$ and for all $x^0 \in \mathcal{X}_{\rho-\varepsilon} \cup \mathcal{X}_{\rho+\varepsilon}$ classification succeeds.*

IV. MATHEMATICAL ANALYSIS OF CA Φ

Equation (5) can be rewritten using matrix notation. This notation turns out to be very useful for analyzing Φ.

The $N \times N$ *averaging matrix* A is given by $A_{i,j} = 1/N_{r,d}$ if $j \in \mathcal{N}_i$ and zero otherwise. Obviously, if $x^k \in [0,1]^N$ is a configuration, then $\overline{x}_i^k = (Ax^k)_i$.

Example 1 *For instance, in the case $d = 1$, $n = 8$ and $r = 2$, the averaging matrix is the following:*

$$A = \frac{1}{5}\begin{bmatrix} \mathbf{1} & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & \mathbf{1} & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & \mathbf{1} & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & \mathbf{1} & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & \mathbf{1} & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & \mathbf{1} & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & \mathbf{1} & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & \mathbf{1} \end{bmatrix}$$

Note that A has a very regular structure. It is doubly stochastic (its entries are non-negative and their sums by rows and columns are always 1) and symmetric. Moreover, A is a primitive matrix. More precisely, there exists $m > 0$ such that $(A^m)_{ij} \neq 0$ for all i, j. The minimum m satisfying this property is called the length path of A. The existence of such m simply follows from the fact that the underlying graph $[n]^d$ is connected. Because of A being symmetric, we know that all of its eigenvalues are real. From the Perron Frobenius Theorem for primitive matrices with non-negative entries we infer some properties. First, $\lambda = 1$ is an eigenvalue of A. Its multiplicity is one. Also, the eigenspace associated with the eigenvalue $\lambda = 1$ is spanned by $[1 \cdots 1]^T$. The absolute value of all the other eigenvalues is strictly less than 1. Finally, A can be decomposed as $M^T D M$, where D is diagonal. The elements on the main diagonal of D are the eigenvalues of A and M is orthonormal. The decomposition of A is unique up to a permutation of rows and columns. Hence, without loss of generality, we can assume that $D_{1,1} = 1$. Note that

$$\lim_{k\to\infty} D^k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Next identity is obtained by recalling that $A^k = M^T D^k M$ and noting that the first column of M is the vector $\frac{1}{\sqrt{N}}[1 \cdots 1]^T$ (because it is in the eigenspace associated with $\lambda = 1$). Therefore,

$$\lim_{k\to\infty} A^k = \frac{1}{N}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{8}$$

Note that (8) has a counterpart for the continuous bistable model. For instance, if we integrate (1) in Ω with $\gamma = 0$ we obtain $\frac{d}{dt}\int_\Omega u = \int_{\partial\Omega} u\frac{du}{dx} = 0$ thanks to the periodic boundary conditions, showing that the mean value is conserved in time.

Parameter σ is the nonlinear amplification factor. The case $\sigma = 0$ corresponds to zero amplification and the corresponding CA only acts by diffusion. In such case, the average of the initial configuration is preserved in time. Moreover, from (8), $\lim_{k\to\infty} x^k = \lim_{k\to\infty} A^k x_0 = [\overline{x}_0 \cdots \overline{x}_0]^T$. Therefore, it would be impossible to solve the density classification without having $\sigma > 0$.

Let $x^0 \in [0,1]^N$ be an initial configuration. It follows that $x^k \in [0,1]^N$ for all k. In fact, since f_σ is strictly increasing, $f_\sigma(0) = 0$ and $f_\sigma(1) = 1$ then we have, by continuity, that f_σ is a one to one map from $[0,1]$ onto $[0,1]$ (see Figure 1).

Suppose that $x^0 = [c \cdots c]^T$ with $c \in (\rho, 1]$. In this case $x^k = [\overline{x}^k \cdots \overline{x}^k]$ with $\overline{x}^k \in (\rho, 1]$. Since $f_\sigma(x) > x$ for $x \in (\rho, 1)$ it follows that $\rho < \overline{x}^k < f_\sigma(\overline{x}^k) = \overline{x}^{k+1} \leq 1$. Therefore, $\overline{x}^k \to 1$ as $k \to \infty$. The case $x^0 = [c \cdots c]^T$ with $c \in [0, \rho)$ is analogous because $f_\sigma(x) < x$ for $x \in (0, \rho)$. In short, if $c < \rho$ then $x^k \to [0 \cdots 0]^T$ and if $c > \rho$ then $x^k \to [1 \cdots 1]^T$.

Remark 1 *The property we just proved can be generalized. Suppose that not all the components of x^0 are the same, but $\rho < x_i^0 \leq 1$ for all i. In this case it is clear that $\rho < \overline{x}_i^0 \leq 1$ for all i. Therefore, $\rho < \overline{x}_i^0 < f_\sigma(\overline{x}_i^0) = x_i^1 \leq 1$ for all i. Inductively, this property is preserved throughout the iterations. Let $c_0 = \min_i x_i^0$. Consider the initial configuration $z^0 = [c_0 \cdots c_0]^T$. We already know that $z^k \to [1 \cdots 1]^T$. Since z^k is dominated by x^k (in every coordinate), we conclude that $x^k \to [1 \cdots 1]^T$. The case $0 \leq x_i^0 < \rho$ is analogous. Therefore, if $\rho < x_i^0 \leq 1$ for all i then $x^k \to [1 \cdots 1]^T$ and if $0 \leq x_i^0 < \rho$ for all i then $x^k \to [0 \cdots 0]^T$.*

Clearly, configurations $[0 \cdots 0]^T$ and $[1 \cdots 1]^T$ are stable equilibrium points. On the other hand, $[\rho \cdots \rho]^T$ is an unstable equilibrium point.

There is also an analogous property for the continuous model (1). Indeed, if $\rho < u(x,0) \leq 1$ for all $x \in \Omega$ then $\rho < u(x,t) \leq 1$ for all $t \geq 0$ and $x \in \Omega$. Let $z^0 = \min_{x\in\overline{\Omega}} u(x,0)$ and let $z(t)$ be the solution of (1) with (constant) initial condition z^0. By a comparison principle we have $\rho < z(t) \leq u(x,t) \leq 1$ for all $x \in \Omega$ and $t \geq 0$. Since $z(t)$ converges to 1 (the only equilibrium point of $z_t = \gamma b_\rho(z)$ greater than ρ is 1) we have that $u(\cdot,t) \to 1$ as $t \to \infty$ uniformly in Ω. Therefore $\rho < u(x,0) \leq 1$ implies that $u(\cdot,t) \to 1$. Analogously, $0 \leq u(x,0) < \rho$ implies that $u(\cdot,t) \to 0$.

Remark 2 *If we allow an initial configuration x^0 such that $x_{i_1}^0 < \rho < x_{i_2}^0$ for some indices $i_1 \neq i_2$, then the dynamics of CA Φ could be non-trivial. More precisely, consider the following initial configuration:*

$$x^0 = \left[\, 1 \ \cdots \ 1 \ \tfrac{1}{3} \ \cdots \ \tfrac{1}{3} \ 1 \ \cdots \ 1 \,\right]^T,$$

where the blocks of consecutive 1's and $\frac{1}{3}$'s have the same size. Suppose $d = 1$, $r = 1$, and $\rho = 1/2$. With these parameters, x^k should converge to $[1 \ldots 1]^T$. As we are going *to see in the proof that follows immediately, this is indeed the case. Nevertheless, it can be checked that there exists a transient period where the global average decreases. For instance* $\overline{x}^1 < \overline{x}^0$.

Let $\varepsilon > 0$. We can prove now that CA Φ solves the density classification problem. We split the proof into three parts.

1. Because of (8), there exists $k_0 = k_0(N, \varepsilon)$ such that for every $k \geq k_0$:

$$\max_{x^0 \in [0,1]^n} \left\| \left(A^k - \frac{1}{N} U \right) x^0 \right\|_\infty \leq \frac{\varepsilon}{3}, \tag{9}$$

where U is the $N \times N$ matrix such that all its entries are 1. This is true because all the involved functions are continuous and $[0,1]^N$ is compact.

2. Recall that $x^k = \Phi^k(x^0)$. In order to include parameter σ in the notation we are going to write $x^k = \Phi_\sigma^k(x^0)$. Therefore, $\Phi_0^k(x^0)$ corresponds to the case where only diffusion is present in the dynamics. Now we are going to show that, for every $k \in \mathbb{N}$ and for every i:

$$\Phi_\sigma^k(x)_i \geq \min_{i'} \Phi_0^k(x)_{i'} - \frac{\sigma k \rho}{4} \tag{10}$$

First observe that, if $x \in [0,1]$, then $-\frac{\rho}{4} \leq b_\rho(x) \leq \frac{(1-\rho)}{4}$. That follows immediately from the definition of b_ρ. Now, by induction, if $k = 1$ we have:

$$\begin{aligned} \Phi_\sigma(x)_i &= \overline{x}_i + \sigma b_\rho(\overline{x}_i) \\ &= \Phi_0(x)_i + \sigma b_\rho(\overline{x}_i) \\ &\geq \min_{i'} \Phi_0(x)_{i'} - \frac{\sigma\rho}{4} \end{aligned}$$

For the inductive step, suppose that the property holds for k. Then:

$$\begin{aligned} \Phi_\sigma^{k+1}(x)_i &= \overline{\Phi_\sigma^k(x)}_i + \sigma b_\rho(\overline{\Phi_\sigma^k(x)}_i) \\ &\geq \min_{i'} \Phi_\sigma^k(x)_{i'} - \frac{\sigma\rho}{4} \\ &\geq \left(\min_{i'} \Phi_0^k(x)_{i'} - \frac{\sigma k \rho}{4} \right) - \frac{\sigma\rho}{4} \\ &= \min_{i'} \Phi_0^k(x)_{i'} - \frac{\sigma(k+1)\rho}{4} \end{aligned}$$

Similarly, using that $b_\rho(\cdot) \leq \frac{(1-\rho)}{4}$, we can prove that for every $k \in \mathbb{N}$ and for every i:

$$\Phi_\sigma^k(x)_i \leq \max_{i'} \Phi_0^k(x)_{i'} + \frac{\sigma k(1-\rho)}{4}. \tag{11}$$

Combining (10) and (11), and now considering the particular case $k = k_0$, for all i

$$\begin{aligned} \min_{i'} \Phi_0^{k_0}(x)_{i'} - \frac{\sigma k_0 \rho}{4} &\leq \Phi_\sigma^k(x)_i \\ &\leq \max_{i'} \Phi_0^{k_0}(x)_{i'} + \frac{\sigma k_0(1-\rho)}{4} \end{aligned}$$

Using (9) we can bracket the values $\Phi_\sigma^k(x)_i$:

$$\begin{aligned} \overline{x}^0 - \frac{\varepsilon}{3} - \frac{\sigma k_0 \rho}{4} &\leq \Phi_\sigma^k(x)_i \\ &\leq \overline{x}^0 + \frac{\varepsilon}{3} + \frac{\sigma k_0(1-\rho)}{4} \end{aligned}$$

Thus:

$$\left| \overline{x}^0 - \Phi_\sigma^{k_0}(x)_i \right| \leq \frac{\varepsilon}{3} + \frac{\sigma k_0 \max(\rho, 1-\rho)}{4}$$

3. We shall see now, that if σ is sufficiently small, then the values for the σ-amplified dynamic at time k_0 are not far away from the original average $\overline{x}^0$. If

$$\frac{\varepsilon}{3} + \frac{\sigma k_0 \max(\rho, 1-\rho)}{4} \leq \frac{2\varepsilon}{3}$$

that is, if $\sigma \leq \frac{4\varepsilon}{3k_0 \max(\rho, 1-\rho)} =: \sigma_0$, it follows that:

$$\left| \overline{x}^0 - \Phi_\sigma^{k_0}(x)_i \right| \leq \frac{2\varepsilon}{3} \tag{12}$$

Now consider the case of $x^0 \in \mathcal{X}_{\rho+\varepsilon}$, which means $\overline{x}^0 > \rho + \varepsilon$. This, combined with (12) implies $\Phi_\sigma^{k_0}(x)_i > \rho$. By Remark 1, we can conclude that $\Phi_\sigma^k(x^0) \to [1 \cdots 1]^T$. The case $x^0 \in \mathcal{X}_{\rho-\varepsilon}$ is similar and hence ommited.

V. QUANTIZATION

Every CA has a finite, well-defined set of states. Therefore, at this point, we need to quantize the values of x_i^k to some number s of discrete values in $[0,1]$. To this end, we define the quantization function (see Figure 2) as

$$Q_s(x) = \begin{cases} \min\left\{1, \frac{\lceil s(x-\rho) \rceil}{s} + \rho\right\}, & \text{if } \rho \leq x \leq 1 \\ \max\left\{0, \frac{\lfloor s(x-\rho) \rfloor}{s} + \rho\right\}, & \text{if } 0 \leq x < \rho \end{cases}$$

By projecting at each iteration the state of each cell we define a new CA as follows:

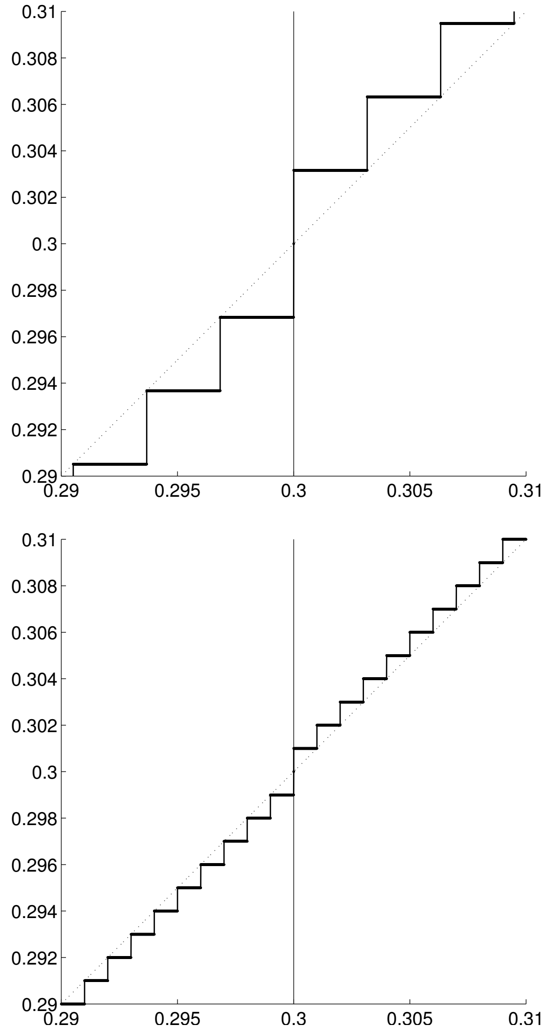

FIG. 2: Quantization function $Q_s(x)$ of the number of states near $\rho = 0.3$ for $s = 10^{2.5}$ and $s = 10^3$.

$$x_i^{k+1} = Q_s(f_\sigma(\overline{x}_i^k)). \tag{13}$$

We are going to show that the CA defined in (13) also solves the density classification problem with accuracy ε. First, if $\delta \in [0, 1]$ is sufficiently small, it can be proven (see Appendix B) that:

$$|b_\rho(x \pm \delta) - b_\rho(x)| \leq \delta.$$

Using this fact, we are going to see, provided that s is large enough, that the CA defined in (13) behaves like the one of Definition 2. More precisely,

$$\left|(Q_s \circ \Phi_\sigma)^k(x)_i - \Phi_\sigma^k(x)_i\right| \leq \frac{(1+\sigma)^k - 1}{s\sigma}.$$

We proceed by induction. For $k = 1$:

$$|(Q_s \circ \Phi_\sigma)(x)_i - \Phi_\sigma(x)_i| \leq \frac{1}{s} = \frac{(1+\sigma)^1 - 1}{s\sigma}.$$

The case $k+1$ goes as follows:

$$\begin{aligned}
&(Q_s \circ \Phi_\sigma)^{k+1}(x)_i \\
&\quad \leq \frac{1}{s} + \Phi_\sigma \circ (Q_s \circ \Phi_\sigma)^k(x)_i \\
&\quad = \frac{1}{s} + \overline{(Q_s \circ \Phi_\sigma)^k(x)}_i + \sigma b_\rho(\overline{(Q_s \circ \Phi_\sigma)^k(x)}_i) \\
&\quad \leq \frac{1}{s} + \overline{\Phi_\sigma^k(x)}_i + \frac{(1+\sigma)^k - 1}{s\sigma} \\
&\qquad + \sigma b_\rho(\overline{\Phi_\sigma^k(x)}_i) + \sigma\frac{(1+\sigma)^k - 1}{s\sigma} \\
&\quad = \Phi_\sigma^{k+1}(x)_i + \frac{(1+\sigma)^{k+1} - 1}{s\sigma}.
\end{aligned}$$

The opposite direction is similar. Note that when $\sigma = 0$, we can recover from the previous bound (making $\sigma \to 0$ and using the L'Hôpital rule) a linear bound for the non-amplified case (which is very tight):

$$\left|(Q_s \circ \Phi_0)^k(x)_i - \Phi_0^k(x)_i\right| \leq \frac{k}{s}.$$

In addition, in the amplified case, the previous bounds are correct if s is large enough and k is not so big. More precisely, if (see the Section B, equation (B1)):

$$\begin{gathered}
3\tfrac{(1+\sigma)^k - 1}{s\sigma} + \left(\tfrac{(1+\sigma)^k - 1}{s\sigma}\right)^2 \\
< \\
1 - \max\{\rho, (1-\rho), \tfrac{1}{3}(\rho^2 - \rho + 1)\}.
\end{gathered}$$

Now we can prove that, if $\frac{1}{s}$ is small enough, every initial condition reaches the correct classification.

Let k_0 be as before. We have:

$$\max_{x^0 \in \mathcal{Q}_s(\mathcal{X}_{\rho+\varepsilon})} \left\|\frac{1}{N}Ux^0 - \Phi_\sigma^{k_0}(x^0)\right\|_\infty \leq \frac{2\varepsilon}{3}.$$

If s is large enough such that $\frac{1}{s} + \frac{(1+\sigma)^{k_0} - 1}{s\sigma} < \frac{\varepsilon}{3}$ (or, equivalently, if $s > \frac{3}{\varepsilon}\left(1 + \frac{(1+\sigma)^{k_0} - 1}{\sigma}\right) =: s_0$) then:

$$\left|(Q_s \circ \Phi_\sigma)^{k_0}(x)_i - \Phi_\sigma^{k_0}(x)_i\right| < \frac{\varepsilon}{3} - \frac{1}{s},$$

Therefore,

$$\max_{x^0 \in \mathcal{Q}_s(\mathcal{X}_{\rho+\varepsilon})} \left\|(Q_s \circ \Phi_\sigma)^{k_0}(x^0) - \Phi_\sigma^{k_0}(x^0)\right\|_\infty < \frac{\varepsilon}{3} - \frac{1}{s}.$$

It follows that

$$\begin{aligned}
&\max_{x^0 \in \mathcal{Q}_s(\mathcal{X}_{\rho+\varepsilon})} \left\|\frac{1}{N}Ux^0 - (Q_s \circ \Phi_\sigma)^{k_0}(x^0)\right\|_\infty \\
&\quad \leq \max_{x^0 \in \mathcal{Q}_s(\mathcal{X}_{\rho+\varepsilon})} \left\|\frac{1}{N}Ux^0 - \Phi_\sigma^{k_0}(x^0)\right\|_\infty \\
&\qquad + \left\|\Phi_\sigma^{k_0}(x^0) - (Q_s \circ \Phi_\sigma)^{k_0}(x^0)\right\|_\infty \\
&\quad < \frac{2\varepsilon}{3} + \frac{\varepsilon}{3} - \frac{1}{s} \\
&\quad = \varepsilon - \frac{1}{s}.
\end{aligned}$$

Finally, we have that

$$x_i^{k_0} := (Q_s \circ \Phi_\sigma)^{k_0}(x^0)_i \geq \overline{x}^0 - \varepsilon + \frac{1}{s} \geq \rho + \frac{1}{s} > \rho.$$

Since Q_s rounds up the values above ρ, it can be easily checked that

$$\Phi_\sigma^k(x^{k_0})_i \leq (Q_s \circ \Phi_\sigma)^k(x^{k_0})_i.$$

Thanks to Remark 1 we have $\Phi_\sigma^k(x^{k_0}) \to [1 \cdots 1]^T$ and we conclude.

Remark 3 *The choice of our quantization function may seem unnecessarily intricate. A simpler quantization function* $\widetilde{Q}_s$ *would be the one which assigns, to any number, the closer value of the* $\frac{1}{s}$*-step staircase. However, we can construct counterexamples where such natural quantization fails to induce the behavior that we want. In fact, consider the case* $d = 1$ *and* $r = 1$ *with the following initial configuration* x^0*:*

$$[0, \tfrac{1}{s}, \tfrac{1}{s}, \cdots, \tfrac{m-1}{s}, \tfrac{m-1}{s}, \tfrac{m}{s}, \tfrac{m}{s}, \tfrac{m-1}{s}, \tfrac{m-1}{s}, \cdots, \tfrac{1}{s}, \tfrac{1}{s}, 0]^T,$$

where s *is the quantization number and* m *is any number between* 1 *and* s*. If* σ *is sufficiently small (*$\sigma < \frac{2}{3s}$*), then:*

$$\begin{aligned} \overline{x}_{2i}^0 &= \frac{1}{3}(\frac{i-1}{s} + \frac{i}{s} + \frac{i}{s}) = \frac{i}{s} - \frac{1}{3s} \\ \overline{x}_{2i+1}^0 &= \frac{1}{3}(\frac{i}{s} + \frac{i}{s} + \frac{i+1}{s}) = \frac{i}{s} + \frac{1}{3s} \end{aligned}$$

and:

$$\begin{aligned} f_\sigma(x^0)_{2i} &= \left(\frac{i}{s} - \frac{1}{3s}\right) + \sigma b_\rho \left(\frac{i}{s} - \frac{1}{3s}\right) \\ f_\sigma(x^0)_{2i+1} &= \left(\frac{i}{s} + \frac{1}{3s}\right) + \sigma b_\rho \left(\frac{i}{s} - \frac{1}{3s}\right). \end{aligned}$$

Therefore, if $\frac{1}{3s} + \frac{\sigma}{4} < \frac{1}{2s}$*, i.e., if* $\sigma < \frac{2}{3s}$*, we would have:*

$$\widetilde{Q}_s(f_\sigma(x^0)) = x^0.$$

In other words, with this choice of quantization $\widetilde{Q}_s$*, the CA would have a non constant valued fixed point.*

Remark 4 *From now on, we are going to assume that the large diffusion and small amplification CA* Φ *is the one defined in (13).*

VI. TESTING CA Φ EMPIRICALLY

We start by comparing the performance of the large diffusion and small amplification CA Φ with a well-known two state, one dimensional, radius $r = 3$ CA called GKL [11]. GKL is one of the best density classifiers for $\rho = 1/2$ and it was designed to classify a maximum number of initial configurations in a ring of $n = 149$ cells. The GKL local rule is the following:

"If the state of a cell is 0, then it takes the majority vote of the first neighbor to its right, the third neighbor to its right, and itself. If the state of the cell is 1, it does so in the opposite direction."

We replicated some simulations of the literature. Initial configurations were generated randomly. More precisely, the initial state for each of the n cells was assigned independently: 1 with probabiity p and 0 with probability $1 - p$. For different values of p we generated 10^5 initial configurations. GKL was applied repeatedly, starting from each of these initial configurations, until one of the following conditions was satisfied:

- GKL reached a fixed point.
- The number of iterations was 10^8.

In all of these experiments, GKL ended up in a fixed point before reaching 10^8 iterations. Then, for each outcome, we checked whether GKL classified or misclassified the initial configuration. The results that show the success ratio appear in Figure 3. These results are for $n = 149$.

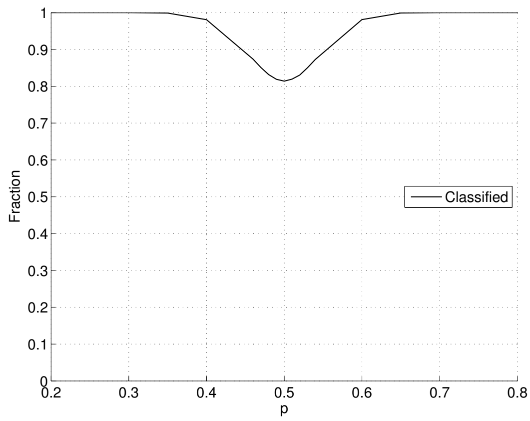

FIG. 3: GKL success ratio for $n = 149$.

For analyzing the scalability of GKL, we explored larger values of n. In Figures 4 and 5 we present the results for $n = 249$ and $n = 1001$. These figures show, as we expected, that the ratio of correctly classified initial configurations decreases as n increases. It is also clear that the hardest instances of the density classification problem are those for which $p = 1/2$. The success ratio of GKL for these hard instances is 82% (in the case n = 149).

Recall that the local rule of the large diffusion and small amplification CA Φ that we considered is the following:

$$\begin{aligned} \overline{x}_i^k &= \frac{x_{i-3}^k + x_{i-2}^k + x_{i-1}^k + x_i^k + x_{i+1}^k + x_{i+2}^k + x_{i+3}^k}{7} \\ x_i^{k+1} &= f_\sigma(\overline{x}_i^k) \end{aligned}$$

To approximate the continuous values, we used standard double precision floating point variables. In order to compare Φ with GKL, we focused on the hardest instances ($p = 1/2$).

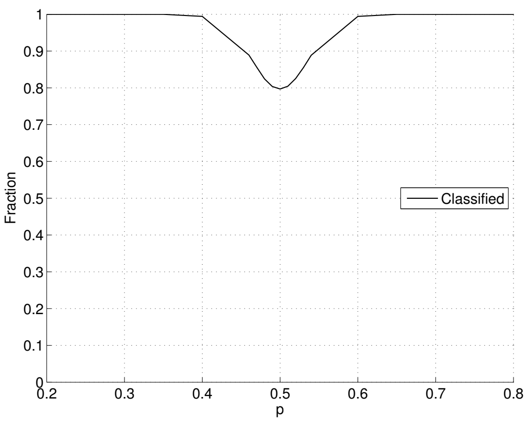

FIG. 4: GKL success ratio for $n = 249$.

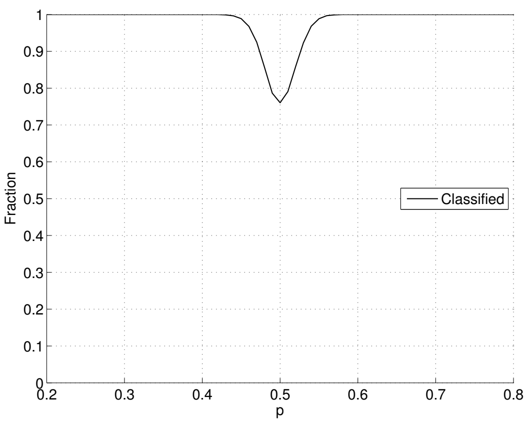

FIG. 5: GKL success ratio for $n = 1001$.

Our existential result does not indicate what the critical values of σ and s are, given n. Also, it does not relate the rate of convergence to the value of σ.

Therefore, we fixed a range $0 < \sigma < 1$. We ran simulations for different values of n and σ. Probability p was always $1/2$. For each pair (n, σ) we ran 1000 simulations using random initial conditions generated in the usual way. CA Φ was applied repeatedly, starting from each of theses initial configurations, until one of the following conditions was satisfied:

1. Either all x_i's are above $3/4 = \rho + 1/4$ or all of them are below $1/4 = \rho - 1/4$. This is a realistic surrogate for convergence to 1 or 0, based on Remark 1. We call this condition *convergence to a constant*.

2. No more progress is detected. That is, the system is at/approaches a fixed point. To detect such a condition, we checked whether $||x^k - x^{k+1}||_1 \leq n \times 10^{-8}$. We call this condition *convergence*.

3. The number of iterations exceeded a threshold, which was chosen to be 2×10^8. This suggests some form of oscillatory behavior although it may not be the case. The bound was chosen by trial and error.

If the sytem converged to a constant, we tested whether Φ reached the "correct" fixed point. In Figure 6, we plot the ratio of well-classified instances (as a function of σ).

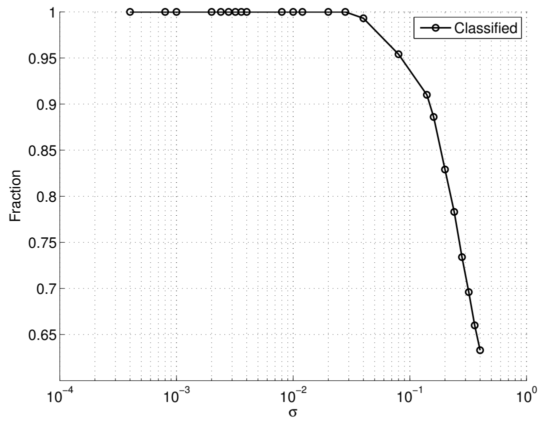

FIG. 6: Breakdown of the success ratio for $n = 149$.

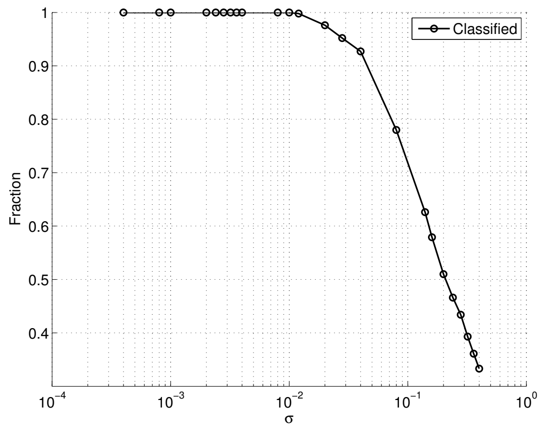

FIG. 7: Breakdown of the success ratio for $n = 249$.

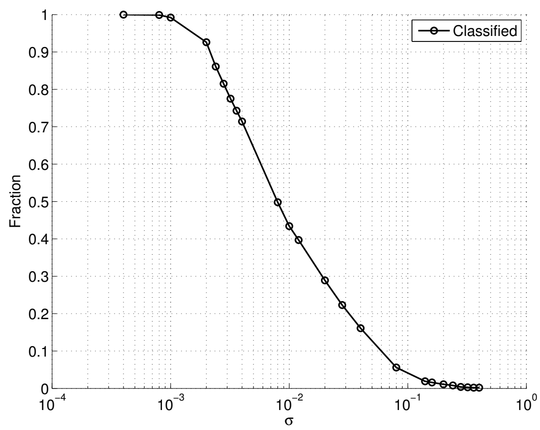

FIG. 8: Breakdown of the success ratio for $n = 1001$.

In order to study the scalability of Φ we repeated the simulations with $n = 249$ (Figure 7) and $n = 1001$ (Figure 8). The estimated values for the critical σ, are 0.030 for $n = 149$, 0.011 for $n = 249$ and 0.001 for $n = 1001$, showing a reciprocal dependency (Figure 9).

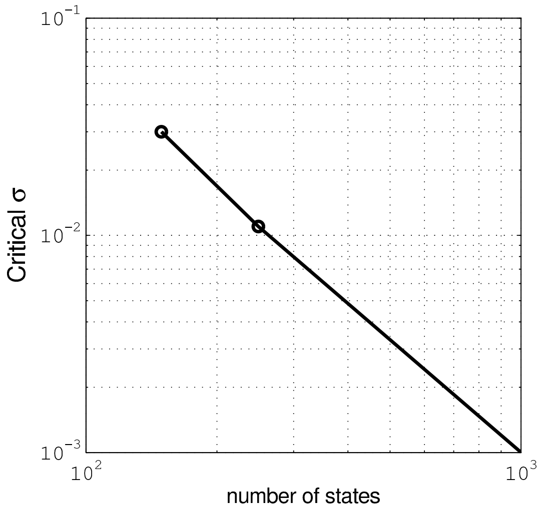

FIG. 9: Critical σ as a function of n.

We can see in Table I the comparison between the performances of GKL and Φ. We already knew that below some threshold σ^* the succes ratio of Φ would be 1. Nevertheless, convergence time increases a lot (with respect to GKL). It seems to be a trade-off between reliability and convergence speed. It may also be interesting to explore the performance degradation of Φ as σ grows.

n	GKL success ratio	Φ success ratio for σ^*
149	0.82	1.0
249	0.80	1.0
1001	0.75	1.0
n	GKL maximal time	Φ maximal time for σ^*
149	80	500
249	140	1000
1001	500	30000

TABLE I: Comparison between GKL and Φ (for $p = 1/2$).

Stone and Bull [22] studied a variant of the elementary CA Rule 184. They augmented the state of each cell with a real number that acts as a "long term" memory. We call the resulting CA 184^*. Since CA Φ also uses memory (number of states), it is interesting to compare the behaviour of our method against CA 184^*. In a notation consistent with ours, the rule is descibed as follows.

The state of cell i consists of two values: $x_i \in \{0, 1\}$ and $m_i \in [0, 1]$. Initially $m_i = 0.5$, for all i, while the configuration of the x_i's is the configuration of 0's and 1's to be classified. To update the state of cell i, we compute:

$$m_i^{k+1} = m_i^k + \beta(x_i^k - m_i^k)$$
$$s_i^k = \begin{cases} 0, & m_i^{k+1} \leq 0.5 \\ 1, & \text{otherwise.} \end{cases}$$
$$x_i^{k+1} = R_{184}(s_{i-1}^k, s_i^k, s_{i+1}^k)$$

β is a positive real parameter. It acts as a learning rate. Following [22], we set $\beta = 0.48$ for all our experiments. Note that, unlike Φ, CA 184^* does not have direct access to the continuous state of the neighboring cells. The function R_{184} is defined as:

s_{i-1}^k	s_i^k	s_{i+1}^k	$R_{184}(s_{i-1}^k, s_i^k, s_{i+1}^k)$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	0
1	1	1	1

In order to compare 184^* with both Φ and GKL, we considered even harder instances than before. In fact, unlike previous experiments, we did not choose the initial configurations by tossing a coin with probability $p = 1/2$ for every cell. Instead, we chose, with uniform probability, a permutation of m 1's and $n - m$ 0's. We chose values of m close to $n/2$.

This allowed us to analyze with more precision the behavior of the system. The variance of the binomial distribution resulting from populating an array using a fair coin in the usual way, averages result from cases of different difficulties. Therefore, for each pair m, n we used in our experiments, we generated 1000 random configurations. We ran our simulations using the same 1000 initial configurations for the three CAs. As we did before, we measured the success ratio. The results are summarized in Figure 10.

Figure 10 describes the cases $n = 149$ and $m = 75, 76, \ldots, 85$. For CA Φ, we chose $\sigma = 0.007$. Note that Φ classified all instances except a few cases when $m = 75$. Both GKL and and 184^* had a success ratio slightly above $1/2$. This contrasts with the $\sim 80\%$ ratio observed when independent probabilities are used in the population process.

We tested larger instances of the problem. Figure 11 describes the cases $n = 1001$ and $m = 501, 502, \ldots, 520$. For CA Φ, we chose $\sigma = 0.0002$ and $\sigma = 0.0003$. These numbers are close to σ^* and allow us to show how the reliability of CA Φ can deteriorate when cases are hard. In fact, we can observe that, when $\sigma = 0.0003$, CA Φ failed to classify about $\sim 80\%$ of the hardest instances. On the other hand, $\sigma = 0.0002$ made the method much more robust.

We also explored a natural question: how does CA Φ behave as we change the number of states s? Note that s is the memory size of each cell.

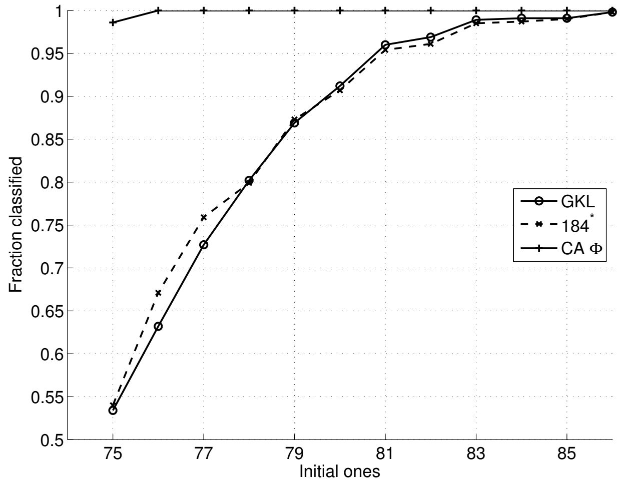

FIG. 10: Success ratios for $n = 149$.

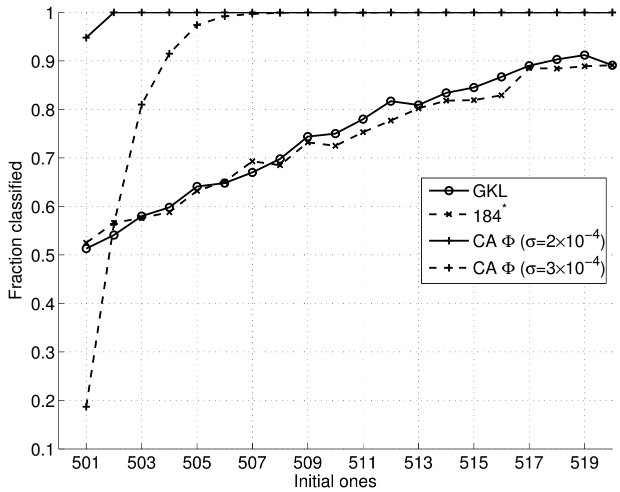

FIG. 11: Success ratios for $n = 1001$.

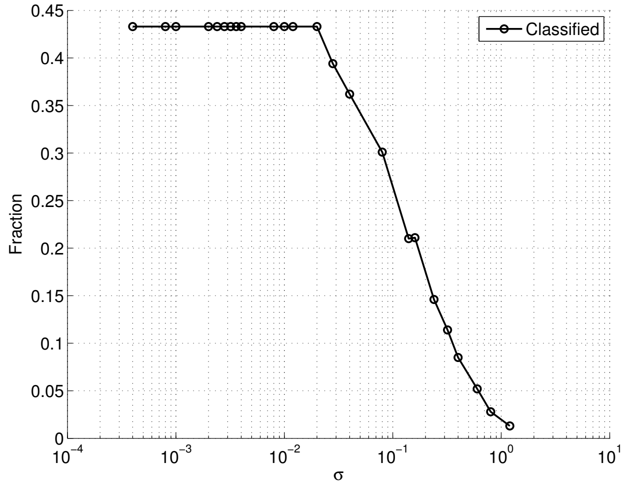

FIG. 12: Breakdown of the success ratio for for $n = 149$ with $s = 128$.

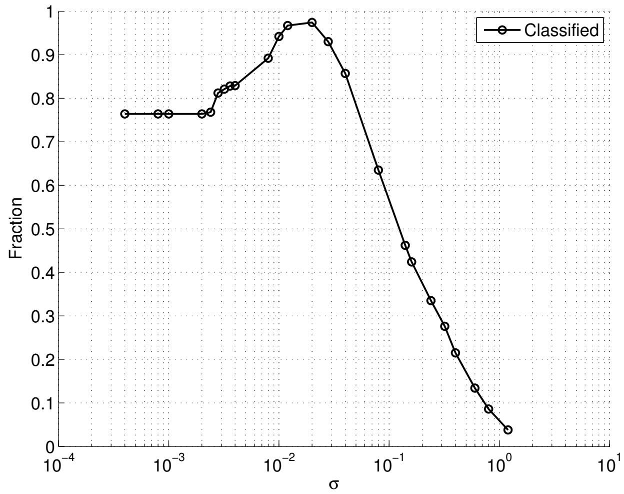

FIG. 13: Breakdown of the success ratio for $n = 149$ with $s = 2000$.

Conceptually, we divide the $[0, 1]$ interval into s subintervals of the same size, centered at ρ (see Figure 2). The cells perform the intermediate computations using exact rational arithmetic and the results are rounded to the center of its closest interval. This is equivalent to use fixed point arithmetic with $q = \log_{10} s$ significative digits.

We present the results for $s = 128$ ($q = 2.1$) in Figure 12, $s = 2000$ ($q = 3.3$) in Figure 13 and $s = 200000$ ($q = 5.3$) in Figure 14, obtaining in this last case a result similar to the "almost continuous" case of Figure 6.

If s is too small, regardless of σ, the success ratio is low (as can be seen in Figures 12 and 13). As s grows, the behavior of the discrete system approaches the behavior of the continuous system. Note that, unlike the continuous case, the reliability of the classification can decrease when decreasing the value of σ, as can be seen in Figure 13.

Finally, in order to illustrate the dimensional scalability of CA Φ, we show some examples of density classification in a two dimensional lattice with periodic boundary conditions. We consider two different types of initial configurations: uniformly random and strip shape, both with initial mean near the critical threshold $\rho = 1/2$. In all the simulations we considered $\sigma = 0.05$ and $s = 10^4$. With these parameters we obtained a success ratio of 100% (in more than 200 random trials). In Figure 15 we show two particular runs.

VII. CONCLUSIONS

We summarize the main advantages of the large diffusion and small amplification CA Φ.

a. Scalability It can be easily modified to work for arbitrary size regular grids in any dimension d.

b. Generalized classification. The critical density ρ can be arbitrary.

c. Large success ratio. By modifying parameters σ and s the CA Φ can be as reliable as needed.

d. Analogy with continuous model. The fact that Φ was originated from a PDE bistable model allows us to capture

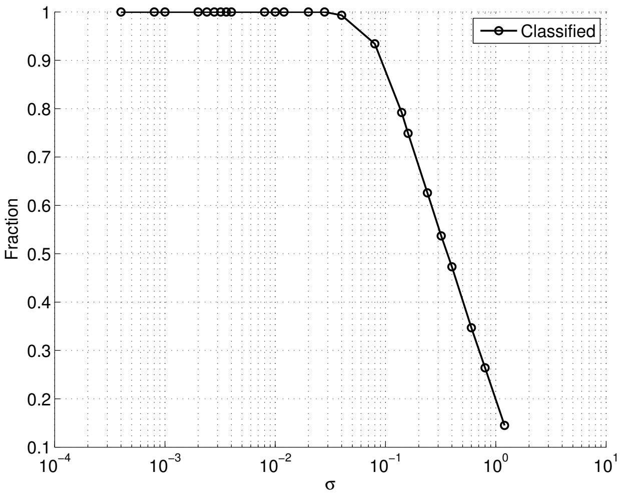

FIG. 14: Breakdown of the success ratio for $n = 149$ with $s = 200000$.

theoretical and physical insight.

e. Robustness. The method maintains the same classification properties for a wide range of different parameters.

Acknowledgments

This work has been partially supported by programs Basal-CMM and Fondecyt 1090156 (R.B., I.R.) and Fondecyt 1110290 (A.O.).

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

[2] A. Bušic, N. Fatès, J. Mairesse and I. Marcovici. *Density classification on infinite lattices and trees.* Proceedings of the 10th Latin American Symposium on Theoretical Informatics (LATIN 2012), Lecture Notes in Computer Science 7256 (2012), 109-120.

[3] R.S. Cantrell and C. Cosner. *Spatial ecology via Reaction-Diffusion equations, John Wiley & Sons, Sussex, 2003.*

[4] M. S. Capcarrere, M. Sipper, and M. Tomassini. *Two-state, r = 1 cellular automaton that classifies density.* Phys. Rev. Lett., 77(24):49694971, 1996.

[5] S-N Chow. *Lattice dynamical systems.* Lecture Notes in Mathematics, 2003, Volume 1822/2003, 1-102.

[6] R. Courant, K. Friedrichs and H. Lewy. *On the partial difference equations of mathematical physics*. IBM Journal of Research and Development 11 (2): 215-234, 1928.

[7] C. M. Elliott and S. Zheng. *On the Cahn-Hilliard equation.* Arch. Rat. Mech. Anal. 96, 339, 1986.

[8] N. Fatès. *Stochastic Cellular automata solve the density classification problem with an arbitrary precision.* Proceedings of the 28th International Symposium on Theoretical Aspects of Computer Science, STACS 2011, 284-295 (2011).

[9] H. Fukś. *Solution of the density classification problem with two cellular automata rules.* Physical Review E, 55(3):R2081R2084, Mar 1997.

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

[11] P. Gács, G .L. Kurdyumov and L.A. Levin. *One dimensional uniform arrays that wash out finite islands (in Russian).* Problemy Peredachi Informatsii 14: 9298 (1978).

[12] P. Gonzaga de Sá and C. Maes. *The Gács-Kurdyumov-Levin automaton revisited.* Journal of Statistical Physics, 67:507522, 1992.

[13] J. P. Keener and J. Sneyd. *Mathematical Physiology, Springer, New York, 1998.*

[14] M. Land and R. K. Belew. *No perfect two-state cellular automata for density classification exists.* Physical Review Letters, 74(25):51485150, 1995.

[15] C. L. M. Martins and P. P. B. de Oliveira. *Evolving sequential combinations of elementary cellular automata rules.* Advances in Artificial Life, volume 3630 of Lecture Notes in Computer Science, pages 461470. Springer Berlin Heidelberg, 2005.

[16] M. Mitchell, J. P. Crutchfield, and P. T. Hraber. *Evolving cellular automata to perform computations: Mechanisms and impediments.* Physica D, 75:361391, 1994.

[17] J. von Neumann. *The general and logical theory of automata* L. A. Jeffress, ed., Cerebral Mechanisms in Behavior The Hixon Symposium, John Wiley & Sons, New York, 1951, pp. 1-31.

[18] P. P. B. de Oliveira, J. C. Bortot, and G. M.B. Oliveira. *The best currently known class of dynamically equivalent cellular automata rules for density classification.* Neuro- computing, 70(1-3):35- 43, 2006.

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

[20] M. Schüle, T. Ott, and R. Stoop. *Computing with probabilistic cellular automata.* In ICANN 09: Proceedings of the 19th International Conference on Artificial Neural Networks, pages 525533, Berlin, Heidelberg, 2009. Springer-Verlag.

[21] C. Stone and L. Bull. *Evolution of cellular automata with memory: The density classification task*. BioSystems, 97(2):108116, 2009.

[22] C. Stone and L. Bull. *Solving the Density Classification Task Using Cellular Automaton 184 with Memory.* Complex Systems 18, pages 229-344, Complex Systems Publications Inc., 2009.

[23] S. Ulam. *Random processes and transformations.* Proceedings of the International Congress on Mathematics, 2 (1952) 264-275.

[24] S. Verel, P. Collard, M. Tomassini and L. Vanneschi. *Fitness landscape of the cellular automata majority problem: View from the Olympus.* Theoretical Computer Science 378, 1 (2007) 54-77.

[25] J. Xin. *Front propagation in heterogeneous media*. SIAM Rev.

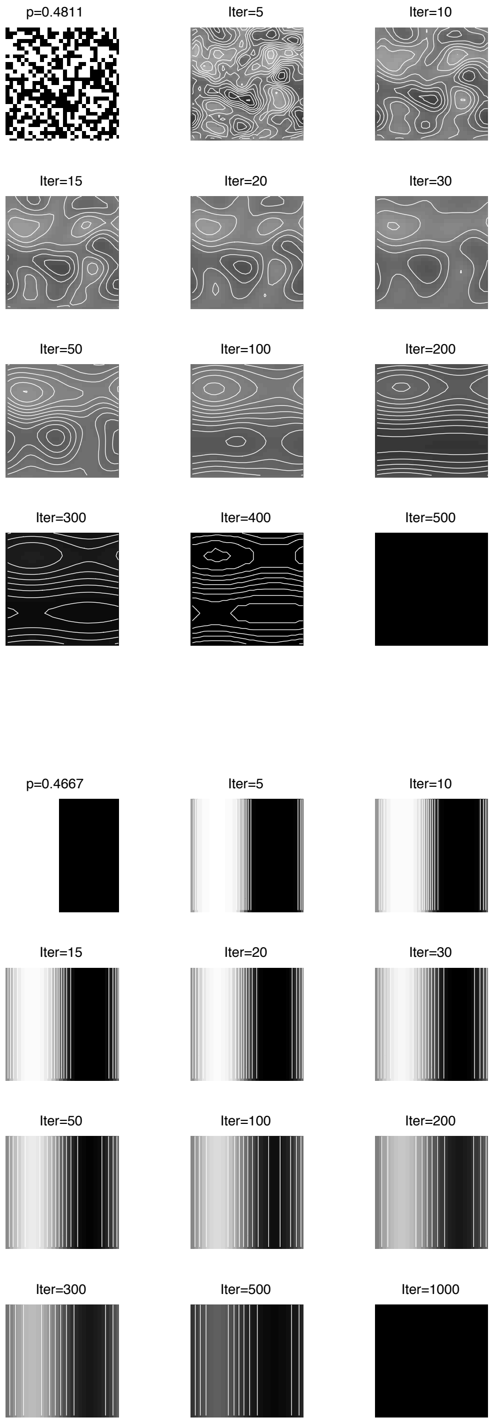

FIG. 15: Two dimensional CA Φ with radius $r = 1$. Random and stripe type initial configurations with initial mean 0.4811 in the first case and 0.4667 in the second case (i.e., slightly more 0's than 1's). Both systems should converge to the all-0 fixed point configuration (black pixels).

42 (2) 161-230, 2000.

Appendix A: Choice of the nonlinear function f_σ

Our choice of a cubic polynomial as a sigmoid function f_σ in was somewhat arbitrary, but this choice seems not to be critical for the results of this paper. For example, take $\rho = 1/2$, and consider the family of functions:

$$g_\alpha(x) = \frac{1}{2}\left(\frac{\tanh(\alpha(2x-1)}{\tanh(\alpha)} + 1\right)$$

characterized by an arbitrary real $\alpha > 0$. Clearly, regardless of α, $g_\alpha(0) = 0$, $g_\alpha(1) = 1$, and $g_\alpha(1/2) = 1/2$, $g'_\alpha(1/2) > 1$. Also, g_α is convex on $[0, 1/2)$ and concave on $(1/2, 1]$. Finally, if $\alpha \to 0$ then g_α resembles the identity, as it can be proved easily using a Taylor approximation. These are all desirable properties for our non-linear amplification function. As a matter of fact, most proofs of this paper can be easily adapted to g_α.

The relation between these two functions is even stronger. In order to achieve a high success ratio we are interested in amplification functions with a small amount of nonlinearity. For the cubic function f_σ that means small σ. For g_α that means small values of α. We are interested in the behavior for values $x \simeq \rho = 1/2$ because they correspond to the cases when the density classification problem is hard to solve. If we approximate $f_\sigma(x)$ at $\alpha = 0, x = 1/2$ using a Taylor polynomial we obtain $f_\sigma(x) \simeq \frac{1}{2} + (x - \frac{1}{2}) + \frac{1}{4}\sigma(x - \frac{1}{2})^3$. Similarly, we approximate $g_\alpha(x) \simeq \frac{1}{2} + (x - \frac{1}{2}) + \frac{1}{3}\alpha^2(x - \frac{1}{2})^3$. Note that if $\sigma = 4\alpha^2/3$ then $f_\sigma(x) \simeq g_a(x)$ at $\alpha \simeq 0, \sigma \simeq 0, x \simeq 1/2$. This establishes an equivalence between the choices of α and σ at least for $\rho = 1/2$.

Appendix B: Bounds on b_ρ

Given δ sufficiently small and $0 < \rho < 1$, we are going to analize the difference:

$$|b_\rho(x+\delta) - b_\rho(x)|\,.$$

By a Taylor expansion, we have that:

$$b_\rho(x+\delta) = b_\rho(x) + b'_\rho(x)\delta + \frac{1}{2}b''_\rho(x)\delta^2 + \frac{1}{6}b'''_\rho(x)\delta^3,$$

where:

$$\begin{aligned} b_\rho(x) &= -x^3 + (1+\rho)x^2 - \rho x \\ b'_\rho(x) &= -3x^2 + 2(1+\rho)x - \rho \\ b''_\rho(x) &= -6x + 2(1+\rho) \\ b'''_\rho(x) &= -6 \end{aligned}$$

Let us define an auxiliary function $\Phi_\delta(x)$ given by:

$$b_\rho(x+\delta) - b_\rho(x) = \underbrace{\delta\left(b'_\rho(x) + \frac{1}{2}b''_\rho(x)\delta + \frac{1}{6}b'''_\rho(x)\delta^2\right)}_{\Phi_\delta(x)}.$$

We are going to study the minima and maxima of $\Phi_\delta(x)$.

$$\begin{aligned}\Phi'_\delta(x) &= b''_\rho(x) + \frac{1}{2}b'''_\rho(x)\delta \\ &= -6x + 2(1+\rho) - 3\delta.\end{aligned}$$

The critical points of $\Phi_\delta(x)$ over $[0,1]$ are 0 and 1. Hence:

$$x^* = \frac{(1+\rho)}{3} - \frac{\delta}{2}.$$

Note that

$$\begin{aligned}cb'_\rho(0) = -\rho, &\quad b''_\rho(0) = 2(1+\rho), \\ b'_\rho(1) = -1+\rho, &\quad b''_\rho(1) = -4+2\rho.\end{aligned}$$

$$\begin{aligned}b'_\rho(x^*) &= -3\left(\frac{(1+\rho)}{3} - \frac{\delta}{2}\right)^2 \\ &\quad +2(1+\rho)\left(\frac{(1+\rho)}{3} - \frac{\delta}{2}\right) - \rho \\ &= \frac{1}{3}(\rho^2-\rho+1) - \frac{3}{4}\delta^2. \\ b''_\rho(x^*) &= -6\left(\frac{(1+\rho)}{3} - \frac{\delta}{2}\right) + 2(1+\rho) = 3\delta.\end{aligned}$$

Then:

$$\begin{aligned}\Phi_\delta(x) &= b'_\rho(x) + \frac{1}{2}b''_\rho(x)\delta + \frac{1}{6}b'''_\rho(x)\delta^2. \\ \Phi_\delta(0) &= b'_\rho(0) + \frac{1}{2}b''_\rho(0)\delta + \frac{1}{6}b'''_\rho(0)\delta^2 \\ &= -\rho + (1+\rho)\delta - \delta^2. \\ \Phi_\delta(1) &= b'_\rho(1) + \frac{1}{2}b''_\rho(1)\delta + \frac{1}{6}b'''_\rho(1)\delta^2 \\ &= -(1-\rho) + (-2+\rho)\delta - \delta^2. \\ \Phi_\delta(x^*) &= b'_\rho(x^*) + \frac{1}{2}b''_\rho(x^*)\delta + \frac{1}{6}b'''_\rho(x^*)\delta^2 \\ &= (\frac{1}{3}(\rho^2-\rho+1) - \frac{3}{4}\delta^2) + \frac{1}{2}(3\delta)\delta + \frac{1}{6}(-6)\delta^2 \\ &= \frac{1}{3}(\rho^2-\rho+1) + \frac{\delta^2}{2}.\end{aligned}$$

Therefore:

$$\begin{aligned}&|b_\rho(x+\delta) - b_\rho(x)| = |\delta||\Phi_\delta(x)| \\ &\quad \leq |\delta| \max\{|\Phi_\delta(0)|, |\Phi_\delta(1)|, |\Phi_\delta(x^*)|\} \\ &\quad \leq |\delta| \max\{\rho + (1+\rho)|\delta| + \delta^2, (1-\rho) + 3|\delta| + \delta^2, \\ &\qquad\qquad \frac{1}{3}(\rho^2-\rho+1) + \delta^2\} \\ &\quad = |\delta|(\max\{\rho, (1-\rho), \frac{1}{3}(\rho^2-\rho+1)\} + 3|\delta| + \delta^2) \\ &\quad < |\delta|.\end{aligned}$$

The last inequality is satisfied when $|\delta|$ is sufficiently small (depending on ρ). More precisely, if:

$$3|\delta| + \delta^2 < 1 - \max\{\rho, (1-\rho), \frac{1}{3}(\rho^2-\rho+1)\}. \quad \text{(B1)}$$