Physica D 261 (2013) 70–80

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

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

Raimundo Briceño [a], Pablo Moisset de Espanés [b], Axel Osses [a,b], Ivan Rapaport [a,b,*]

[a] Departamento de Ingeniería Matemática, Universidad de Chile, Chile
[b] Centro de Modelamiento Matemático (UMI 2807 CNRS), Universidad de Chile, Chile

HIGHLIGHTS

- We solve the density classification problem.
- This is one of the most studied inverse problems in cellular automata.
- We solve this problem with a CA inspired by two mechanisms that are ubiquitous in nature.
- These mechanisms are diffusion and nonlinear sigmoidal response.
- Our solution works in any dimension, for an arbitrary number of cells, and any critical density.

ARTICLE INFO

Article history:
Received 27 November 2012
Received in revised form
14 May 2013
Accepted 2 July 2013
Available online 9 July 2013
Communicated by. Thomas Wanner

Keywords:
Cellular automata
Density classification
Local averaging and saturation

ABSTRACT

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 0s and 1s, it converges to the all-1 fixed point configuration if the fraction of 1s 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 inspired by two mechanisms that are ubiquitous in nature: diffusion and nonlinear sigmoidal response. This CA, which is different from the classical ones because it has many states, has a success ratio of 100%, and works for any system size, any dimension, and any critical density.

1. Introduction

Cellular automata (CAs) are discrete dynamical systems. They were introduced by John von Neumann [1] after a suggestion of Stanislaw Ulam [2]. Here, we consider finite CAs: more precisely, n^d cells arranged uniformly spaced in the d-dimensional torus and following a local rule identical in 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 0s and 1s, it converges to the all-1 fixed point configuration if the fraction of 1s in x^0 is greater than ρ_c, and it converges the all-0 fixed point configuration otherwise. The number $0 < \rho_c < 1$ denotes the *critical density*.

The problem was first formulated for dimension $d = 1$ (a ring) and critical density $\rho_c = 1/2$ [3]. The best-known two-state CA for tackling this instance of the density classification problem is called GKL [4,5]. Its original purpose was to resist small amounts of noise. The performance of GKL was very good but not perfect. In fact, an impossibility result was proved: there is no perfect density classifier with two states [6].

The impossibility of finding perfect classifiers led many researchers to use different evolutionary computation approaches to evolve good *approximate solutions* [7–11]. But, in order to obtain perfect density classifiers, researchers were forced to modify the original problem. One idea was to change the output specifications [12]. Another idea was to allow the existence of more than one local rule [13,14] or to embed a memory on the cells [15,16]. A very subtle and interesting relaxation of the original problem is related to determinism. In fact, following other works [17,18], Fatès designed a two-state stochastic CA that solves the density classification problem with arbitrary precision [19].

* Corresponding author at: Departamento de Ingeniería Matemática, Universidad de Chile, Chile. Tel.: +56 2 978 44 70.
E-mail address: rapaport@dim.uchile.cl (I. Rapaport).

0167-2789/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physd.2013.07.002

The idea of the present paper is to use a continuous approach for solving the density classification problem *deterministically*. 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 transitions and front propagation in spatial ecology [20], physiology [21], and chemistry and physics [22]. 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 [23,24]).

Two parameters characterize CA Φ: the amount of nonlinearity σ (or amplification factor) and the number of states s (or discretization factor). The main result of this paper is that—given arbitrary parameters n, d, and ρ_c—there exist σ and s such that the large diffusion and small amplification CA $\Phi(\sigma, s)$ solves the density classification problem with a success ratio of 100%.

CA Φ, besides solving the density classification problem perfectly in any dimension d for an arbitrary number of cells n^d, and with any critical density ρ_c, allows an intuitive interpretation together with a deep theoretical analysis. It also maintains the same classification properties for a wide range of different averages and nonlinear amplifications.

The theoretical result concerning the existence of CA Φ does not give us any indication about the critical values of s and σ (even simply as a function of n, with ρ_c and d fixed). Therefore, in Section 6, we implement Φ and we compare it not only with GKL but also against a variant of the elementary CA Rule 184 [25]. This CA, that we denote by 184*, is particularly interesting because it is similar to CA Φ in two senses: (1) it is hand designed, and (2) cells are augmented with much more memory (note that the only way to increase the memory capabilities of a finite-state machine is by augmenting its number of states).

We run simulations using a fixed set of extremely hard instances (initial configurations). These initial configurations are generated by randomly permuting m 1s and $n - m$ 0s. It turns out that CA Φ with $\sigma = 0.007$ (and before discretizing the set of states, i.e., $s \sim 200\,000$), classifies all instances except a few cases when $m = 75$ and $n = 149$. By contrast, both GKL and 184* have a success ratio slightly above 50% (which can be interpreted as random success). Note that, when the initial configurations are generated with a uniform independent probability law (a binomial law of parameters $n = 149$ and $p = 1/2$), the observed success ratio of GKL and 184* is ∼80%.

2. 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-(von Neumann) neighborhood* of v_0 is $\mathcal{N}_{r,d}(v_0) = \{v \in [n]^d : |v - v_0| \leq r\}$, 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 by $N_{r,d}$. 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 by Ψ, transforms a configuration $x^k \in [0, 1]^N$ into a new configuration $x^{k+1} \in [0, 1]^N$ by applying 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 $x \in [0, 1]^N$, its mean value $\bar{x}$ corresponds to the *density*. Therefore, the value $\bar{x}^0$ denotes the fraction of 1s in the initial configuration x^0. Given $\rho_c \in (0, 1)$ and $\varepsilon > 0$, we introduce the sets

$$\mathcal{X}_{\rho_c}(-\varepsilon) = \{x \in [0, 1]^N | \, \bar{x} < \rho_c - \varepsilon\},$$
$$\mathcal{X}_{\rho_c}(+\varepsilon) = \{x \in [0, 1]^N | \, \bar{x} > \rho_c + \varepsilon\}.$$

Definition 1 (*Generalized Density Classification Problem*)**.** Given $\rho_c \in (0, 1)$ 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 0s if $\bar{x}^0$ is less than $\rho_c - \varepsilon$ and converges to the configuration of only 1s if $\bar{x}^0$ is greater than $\rho_c + \varepsilon$. That is,

$$\forall x^0 \in \mathcal{X}_{\rho_c}(-\varepsilon), \quad \lim_{k\to\infty} x^k = [0 \cdots 0]^T \quad \text{and}$$
$$\forall x^0 \in \mathcal{X}_{\rho_c}(+\varepsilon), \quad \lim_{k\to\infty} x^k = [1 \cdots 1]^T.$$

Remark 1. 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 the following senses.

1. In our definition, the initial configuration is arbitrary (not restricted to 0s and 1s).
2. The dimension d, the number of cells N, the radius r, and the critical density ρ_c are also arbitrary.
3. The definition of convergence as a limit when $k \to \infty$ is useful because it can be applied for CAs having either infinite or finite number of states. When the set of states is finite, the convergence definition implies that all cells reach state 0 or 1 in a finite number of steps. In the next section, we are going to introduce a CA with states in [0, 1]. Later, in Section 5, we are going to discretize it in order to obtain a standard CA. The discretized version will have a finite number of states.

Remark 2. In the classical density classification problem, initial configurations are restricted to 0s and 1s. Therefore, if we prove that some CA Ψ solves the problem with accuracy $\varepsilon < \frac{1}{2n}$, then we will be proving that Ψ solves the problem for every initial configuration in $\{0, 1\}^n$. In fact, suppose without loss of generality that the critical density $\rho_c = \frac{2l+1}{2n}$. In other words, if the initial configuration has at most l 1s, then Ψ must converge to the all-0 configuration. Otherwise, if it has at least $l + 1$ 1s, then it must converge to the all-1 configuration. Obviously, if Ψ solves the problem with accuracy $\varepsilon < \frac{1}{2n}$, then it solves the problem for every initial configuration in $\{0, 1\}^n$.

3. 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 (see Fig. 1). More precisely, given a critical density $\rho_c \in (0, 1)$, the idea is to build a local rule based on a discrete version of the following equation:

$$\frac{\partial u}{\partial t} - \nu \Delta u = \gamma \, b_{\rho_c}(u), \tag{1}$$

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_c} is some suitable *bistable function*. In this paper (readers can see a discussion about other choices in Appendix A) we choose the cubic polynomial:

$$b_{\rho_c}(u) = u(1 - u)(u - \rho_c).$$

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_c, repulsor).

3.1. 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 Eq. (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 Eq. (1) by finite differences is

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

where the sums 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:

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

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 the previous relation can be rewritten as

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

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

$$f(x) = x + \sigma\, b_{\rho_c}(x),$$

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, Eqs. (2) and (3) are equivalent. In fact, it is easy to see that vector u converges to a constant vector in Eq. (2) if and only if it converges to the *same* constant vector in Eq. (3).

The two-dimensional case with radius 1 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 $\overline{u}_{i,j}^k = \frac{1}{5}(u_{i-1,j}^k + u_{i+1,j}^k + u_{i,j}^k + u_{i,j-1}^k + u_{i,j+1}^k)$.

Note that previous CAs, defined by Eq. (3) in one dimension and by Eq. (4) in two dimensions, correspond to discrete approximations of Eq. (1) that satisfy (even strictly) the corresponding Courant–Friedrich–Lewy stability condition $\frac{\nu\Delta t}{h^2} = \frac{1}{3} < \frac{1}{2}$ in dimension 1 or $\frac{\nu\Delta t}{h^2} = \frac{1}{5} < \frac{1}{4}$ in dimension 2. This guarantees convergence of the corresponding CAs to the continuous equation as h and Δt tend to zero, but only in the case when $\sigma = 0$ [26]. This convergence result cannot 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 4.

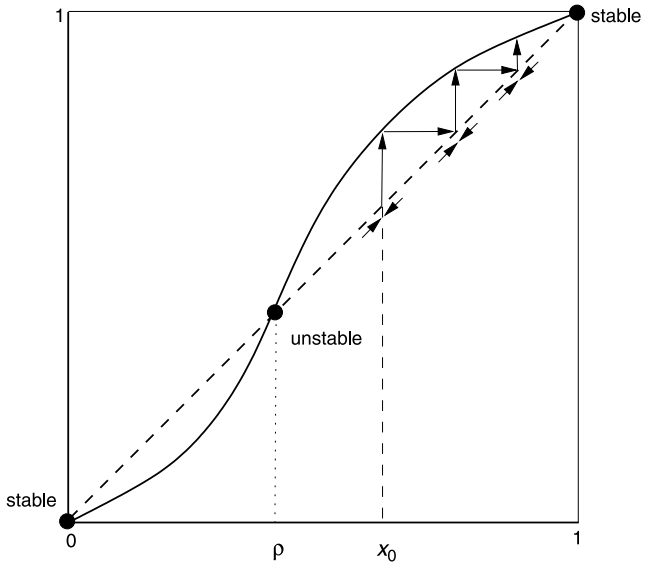

Fig. 1. Schematic view of Φ. Given an initial state, the CA corresponds to a rule obtained by first averaging neighbors (arrows towards the diagonal) and then applying a nonlinear amplification function f (arrows towards the curve). In the figure the system is converging to the all-1 vector configuration.

3.2. 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:

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

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

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

Recall that ν and γ are respectively the diffusion and amplification parameters of Eq. (1), the bistable heat equation. 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 sufficiently small. We impose the following:

$$0 < \sigma < \min\left\{\frac{1}{\rho_c}, \frac{1}{1-\rho_c}\right\}. \tag{7}$$

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 4. Moreover, the values σ that solve the problem are typically much smaller than the upper bound, as our experiments of Section 6 show.

We say that CA Φ solves the density classification problem if, regardless of how close the average of the initial configurations is to ρ_c, there is a range of values for the nonlinearity σ that guarantees convergence to the correct answer. Formally, we have the following.

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 $x^0 \in \mathcal{X}_{\rho_c}(-\varepsilon) \cup \mathcal{X}_{\rho_c}(+\varepsilon)$ classification succeeds.

4. Mathematical analysis of CA Φ

Eq. (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 is zero otherwise. Obviously, if $x^k \in [0,1]^N$ is a configuration, then $\bar{x}_i^k = (Ax^k)_i$.

Example 1. For instance, in the case when $d = 1, n = 8$, 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 1. 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 $M^T DM$, 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

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

The next identity is obtained by considering 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 Eq. (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 a case, the average of the initial configuration is preserved in time. Moreover, from (8), $\lim_{k\to\infty} x^k = \lim_{k\to\infty} A^k x_0 = [\bar{x}_0 \cdots \bar{x}_0]^T$. 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 $f_\sigma(1) = 1$, and then we have, by continuity, that f_σ is a one to one map from $[0,1]$ onto $[0,1]$ (see Fig. 1).

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

Remark 3. The property we just proved can be generalized. Suppose that not all the components of x^0 are equal, but $\rho_c < x_i^0 \leq 1$ for all i. In this case, it is clear that $\rho_c < \bar{x}_i^0 \leq 1$ for all i. Therefore, $\rho_c < \bar{x}_i^0 < f_\sigma(\bar{x}_i^0) = x_i^1 \leq 1$ 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 $0 \leq x_i^0 < \rho_c$ is analogous. Therefore, if $\rho_c < x_i^0 \leq 1$ for all i then $x^k \to [1 \cdots 1]^T$, and if $0 \leq x_i^0 < \rho_c$ 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, $[\rho_c \cdots \rho_c]^T$ is an unstable equilibrium point.

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

Remark 4. If we allow an initial configuration x^0 such that $x_{i_1}^0 < \rho_c < x_{i_2}^0$ 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:

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

where the blocks of consecutive 1s and $\frac{1}{3}$s have the same size. Suppose that $d = 1, r = 1$, and $\rho_c = 1/2$. 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, $\bar{x}^1 < \bar{x}^0$.

Let $\varepsilon > 0$. Now we prove that CA Φ solves the density classification problem with accuracy ε. We split the proof into three parts.

1. Because of Eq. (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 functions involved 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. In Appendix B, we show that, for every $k \in \mathbb{N}$ and for every i,
$$\Phi_\sigma^k(x)_i \geq \min_{i'} \Phi_0^k(x)_{i'} - \frac{\sigma k \rho_c}{4}. \tag{10}$$
Similarly, using that $b_{\rho_c}(\cdot) \leq \frac{(1-\rho_c)}{4}$, we can prove that, for every $k \in \mathbb{N}$ and for every i,
$$\Phi_\sigma^k(x)_i \leq \max_{i'} \Phi_0^k(x)_{i'} + \frac{\sigma k(1-\rho_c)}{4}. \tag{11}$$

Combining the result of Appendix B with Eq. (11), and now considering the particular case $k = k_0$, for all i,

$$\min_{i'} \Phi_0^{k_0}(x)_{i'} - \frac{\sigma k_0 \rho_c}{4} \le \Phi_\sigma^k(x)_i \le \max_{i'} \Phi_0^{k_0}(x)_{i'} + \frac{\sigma k_0 (1-\rho_c)}{4}.$$

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

$$\bar{x}^0 - \frac{\varepsilon}{3} - \frac{\sigma k_0 \rho_c}{4} \le \Phi_\sigma^k(x)_i \le \bar{x}^0 + \frac{\varepsilon}{3} + \frac{\sigma k_0 (1-\rho_c)}{4}.$$

Thus

$$\left|\bar{x}^0 - \Phi_\sigma^{k_0}(x)_i\right| \le \frac{\varepsilon}{3} + \frac{\sigma k_0 \max(\rho_c, 1-\rho_c)}{4}.$$

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

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

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

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

Now consider the case of $x^0 \in \mathcal{X}_{\rho_c}(+\varepsilon)$, which means that $\bar{x}^0 > \rho_c + \varepsilon$. This, combined with Eq. (12), implies that $\Phi_\sigma^{k_0}(x)_i > \rho_c$. By Remark 3, we can conclude that $\Phi_\sigma^k(x^0) \to [1 \cdots 1]^T$. The case $x^0 \in \mathcal{X}_{\rho_c}(-\varepsilon)$ is similar, and hence is omitted.

5. 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 Fig. 2) as

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

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

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

We are going to show that the CA defined in Eq. (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 C) that

$$\left|b_{\rho_c}(x \pm \delta) - b_{\rho_c}(x)\right| \le \delta.$$

Using this fact, provided that s is sufficiently large, the CA defined in Eq. (13) behaves like the one of Definition 2. More precisely (see Appendix D),

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

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

Let k_0 be as before. We know from the previous section that

$$\max_{x^0 \in Q_s(\mathcal{X}_{\rho_c}(+\varepsilon))} \left\| \frac{1}{N} U x^0 - \Phi_\sigma^{k_0}(x^0) \right\|_\infty \le \frac{2\varepsilon}{3}.$$

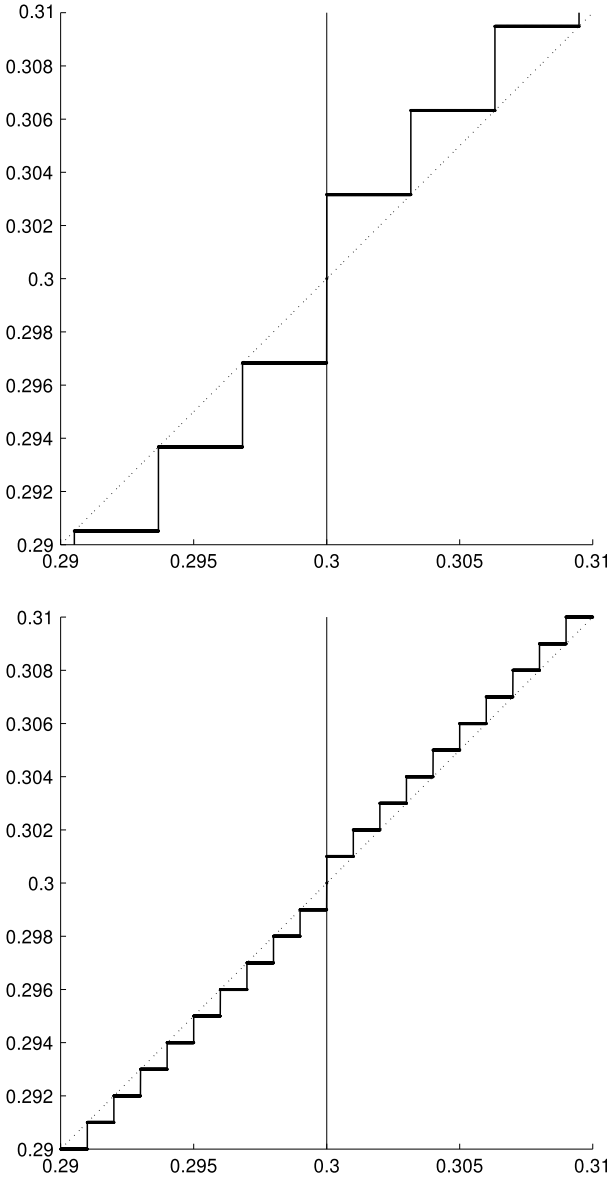

Fig. 2. Quantization function $Q_s(x)$ of the number of states near $\rho_c = 0.3$ for $s = 10^{2.5}$ and $s = 10^3$.

If s is sufficiently large 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,

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

It follows that

$$\max_{x^0 \in Q_s(\mathcal{X}_{\rho_c}(+\varepsilon))} \left\| \frac{1}{N} U x^0 - (Q_s \circ \Phi_\sigma)^{k_0}(x^0) \right\|_\infty < \underbrace{\frac{2\varepsilon}{3} + \frac{\varepsilon}{3} - \frac{1}{s}}_{\varepsilon - \frac{1}{s}}.$$

Finally, for the case $\bar{x}^0 \ge \rho_c + \varepsilon$, we have that

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

Since Q_s rounds up the values above ρ_c, 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 3, we have $(Q_s \circ \Phi_\sigma)^k(x^{k_0}) \to [1 \cdots 1]^T$, and we conclude. The other case, where $\overline{x}^0 \leq \rho_c - \varepsilon$, is analogous, and we have $(Q_s \circ \Phi_\sigma)^k(x^{k_0}) \to [0 \cdots 0]^T$.

Remark 5. 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:

$$\left[0, \frac{1}{s}, \frac{1}{s}, \ldots, \frac{m-1}{s}, \frac{m-1}{s}, \frac{m}{s}, \frac{m}{s}, \frac{m-1}{s}, \frac{m-1}{s}, \ldots, \frac{1}{s}, \frac{1}{s}, 0\right]^T,$$

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

$$\overline{x}^0_{2i} = \frac{1}{3}\left(\frac{i-1}{s} + \frac{i}{s} + \frac{i}{s}\right) = \frac{i}{s} - \frac{1}{3s},$$

$$\overline{x}^0_{2i+1} = \frac{1}{3}\left(\frac{i}{s} + \frac{i}{s} + \frac{i+1}{s}\right) = \frac{i}{s} + \frac{1}{3s}$$

and

$$f_\sigma(x^0)_{2i} = \left(\frac{i}{s} - \frac{1}{3s}\right) + \sigma b_{\rho_c}\left(\frac{i}{s} - \frac{1}{3s}\right),$$

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

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

Remark 7. By considering Remark 2, we conclude that our CA Φ solves the density classification problem for every initial configuration in $\{0, 1\}^n$ (provided that the parameters σ and s are fine tuned).

6. 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 [4]. GKL is one of the best-known density classifiers when $\rho_c = 1/2$ and $n = 149$. 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 probability 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.

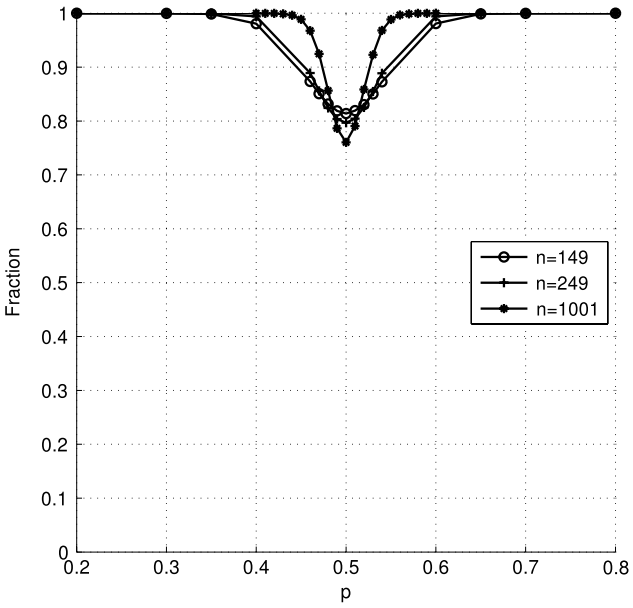

Fig. 3. GKL success ratio for $n = 149$, $n = 249$, and $n = 1001$.

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.

For analyzing the scalability of GKL, we explored the success ratio for different values of n. Fig. 3 describes the results for $n = 149$, $n = 249$, and $n = 1001$. This figure shows, 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:

$$\overline{x}^k_i = \frac{x^k_{i-3} + x^k_{i-2} + x^k_{i-1} + x^k_i + x^k_{i+1} + x^k_{i+2} + x^k_{i+3}}{7},$$

$$x^{k+1}_i = f_\sigma(\overline{x}^k_i).$$

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 ($p = 1/2$). 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 these initial configurations, until one of the following conditions was satisfied.

1. Either all the x_i are above $3/4 = \rho_c + 1/4$ or all of them are below $1/4 = \rho_c - 1/4$. This is a realistic surrogate for convergence to 1 or 0, based on Remark 3. 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.

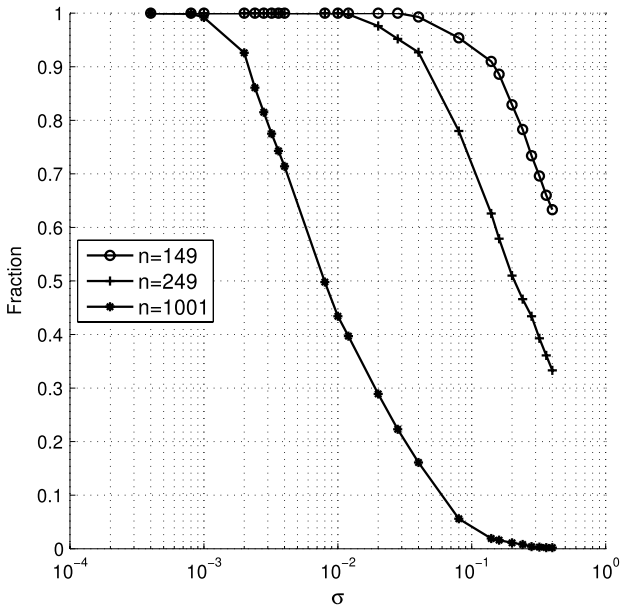

Fig. 4. Breakdown of the success ratio for $n = 149$, $n = 249$, and $n = 1001$.

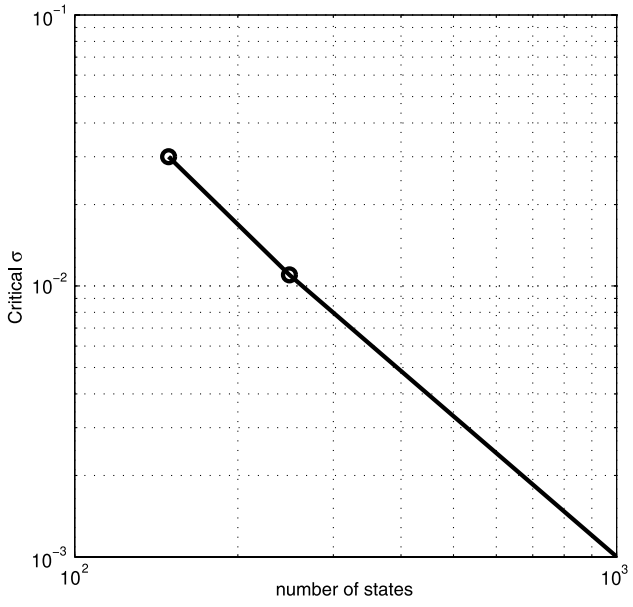

Fig. 5. Critical σ as a function of n.

Table 1
Comparison between GKL and Φ (for $p = 1/2$).

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	1 000
1001	500	30 000

Table 2
The local function R_{184}.

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

If the system converged to a constant, we tested whether Φ reached the "correct" fixed point. In Fig. 4, we plot the ratio of well-classified instances (as a function of σ).

In order to study the scalability of Φ, we repeated the simulations with $n = 249$ and $n = 1001$ (also presented in Fig. 4). 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 (Fig. 5).

We can see in Table 1 the comparison between the performances of GKL and Φ. We already knew that below some threshold σ^* the success ratio of Φ would be 100%. Nevertheless, the convergence time increases a lot (with respect to GKL). There 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.

Stone and Bull [25] studied a variant of the elementary CA Rule 184. They augmented the state of each cell with a real number. We call the resulting CA 184*. Since CA Φ also uses more memory (more number of states), it is interesting to compare the behavior of our method against CA 184*. In a notation consistent with ours, the rule is described 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 is the configuration of 0s and 1s 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 [25], 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 in Table 2.

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 1s and $n - m$ 0s. We chose values of m close to $n/2$.

This allowed us to analyze the behavior of the system with more precision. The variance of the binomial distribution results from populating an array using a fair coin in the usual way, and the averages result from cases of different difficulties. Therefore, for each pair m, n that 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 Fig. 6.

Fig. 6 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 184* had a success ratio slightly above 1/2. This contrasts with the ∼80% ratio observed when independent probabilities are used in the population process.

We tested larger instances of the problem. Fig. 7 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 they 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 ∼ 80% of the hardest

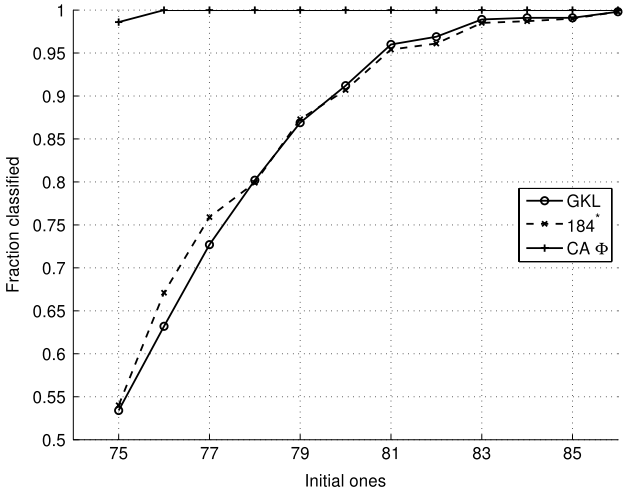

Fig. 6. Success ratios for $n = 149$.

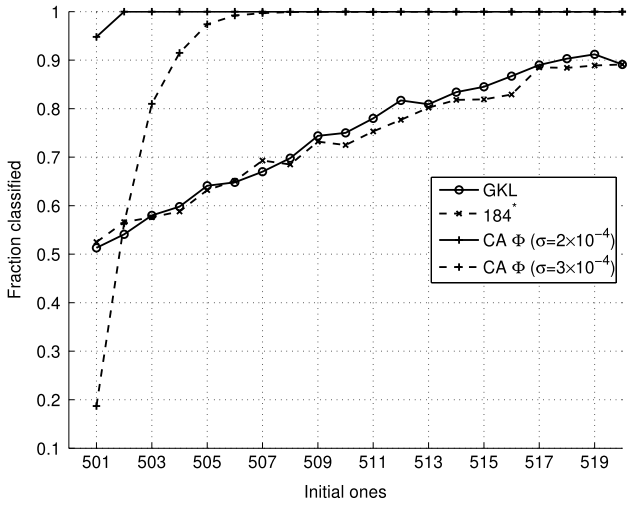

Fig. 7. Success ratios for $n = 1001$.

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.

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

In Fig. 8, we present the results for $s = 128$ ($q = 2.1$), $s = 2000$ ($q = 3.3$) and $s = 200\,000$ ($q = 5.3$), obtaining in this last case a result similar to the "almost continuous" case of Fig. 4.

If s is too small, regardless of σ, the success ratio is low. 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 Fig. 8, for $s = 2000$.

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 configuration: uniformly random and strip shape, both with initial mean near the critical threshold $\rho_c = 1/2$. In all the simulations we considered $\sigma = 0.05$ and

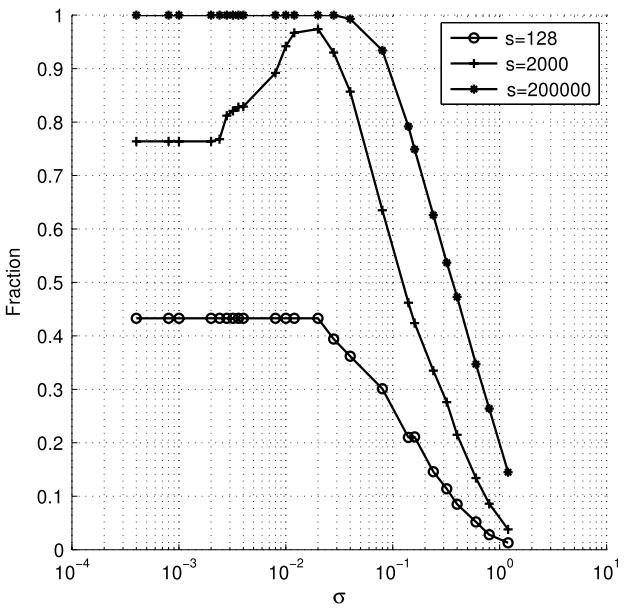

Fig. 8. Breakdown of the success ratio for $n = 149$ and $s = 128, 2000, 200\,000$.

$s = 10^4$. With these parameters we obtained a success ratio of 100% (in more than 200 random trials). In Fig. 9 we show two particular runs.

7. Conclusions and perspectives

The most important advantage of the large diffusion and small amplification CA Φ we present in this work is its success ratio of 100%. This can be achieved by tuning two parameters: the amplification factor σ and the number of states s. Other important advantages of Φ are the following.

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

Generalized classification. The critical density ρ_c can be arbitrary.

Analogy with continuous model. The fact that Φ was originated from a partial differential equation bistable model allows us to gain theoretical and physical insight.

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

In order to achieve a 100% success ratio, we need small values for σ and large values for s. Small values of σ imply large convergence time. Therefore, it seems that the price one has to pay for achieving a 100% success ratio is somehow related to the two most relevant resources in computer systems: time and memory (also known as space).

In other words, the main question this paper leaves open is a complexity question. More precisely, let ρ_c be the critical density, d the dimension, n^d the size of the system, $s(\rho_c, n, d,)$ the number of states, $t(\rho_c, n, d)$ the convergence time, and χ the success ratio. The general question is the following: what are the critical values of $s(\rho_c, n, d,)$ and $t(\rho_c, n, d)$ for which there exists a CA that solves the density classification problem with a success ratio of χ? This question has been partially studied for $\chi = 100\%$, $d = 1, 2$, and $s = 2$.

Acknowledgments

This work has been partially supported by programs Basal-CMM, Fondecyt 1130061 (I.R.) and Fondecyt 1110290 (A.O.).

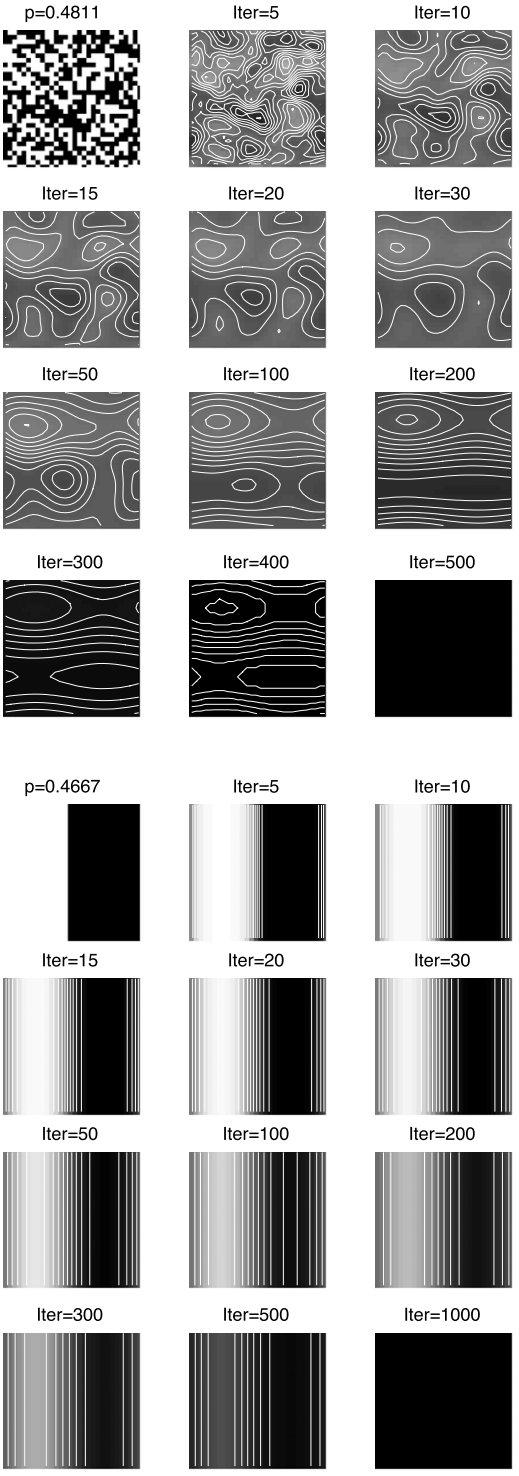

Fig. 9. Two-dimensional CA Φ with radius $r = 1$. Random and stripe type initial configurations with initial mean 0.4811 in the first case and 0.4667 in the second case (i.e., slightly more 0s than 1s). Both systems should converge to the all-0 fixed point configuration (black pixels).

Appendix A

Our choice of a cubic polynomial as a sigmoid function f_σ was somewhat arbitrary, but this choice seems not to be critical for the results of this paper. For example, take $\rho_c = 1/2$, 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 can be proved easily using a Taylor approximation. These are all desirable properties for our nonlinear amplification function. In fact, most proofs in 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 $x \simeq \rho_c = 1/2$, 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 $\rho_c = 1/2$.

Appendix B

To prove that

$$\Phi^k_\sigma(x)_i \geq \min_{i'} \Phi^k_0(x)_{i'} - \frac{\sigma k \rho_c}{4},$$

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

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

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

$$\begin{aligned}\Phi^{k+1}_\sigma(x)_i &= \overline{\Phi^k_\sigma(x)}_i + \sigma b_{\rho_c}(\overline{\Phi^k_\sigma(x)}_i)\\ &\geq \min_{i'} \Phi^k_\sigma(x)_{i'} - \frac{\sigma \rho_c}{4}\\ &\geq \left(\min_{i'} \Phi^k_0(x)_{i'} - \frac{\sigma k \rho_c}{4}\right) - \frac{\sigma \rho_c}{4}\\ &= \min_{i'} \Phi^k_0(x)_{i'} - \frac{\sigma(k+1)\rho_c}{4}.\end{aligned}$$

Appendix C

Given δ sufficiently small and $0 < \rho_c < 1$, we are going to analyze the difference,

$$\left|b_{\rho_c}(x + \delta) - b_{\rho_c}(x)\right|.$$

By a Taylor expansion, we have that

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

where

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

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

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

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

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

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

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

Note that

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

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

$$b''_{\rho_c}(x^*) = -6\left(\frac{(1+\rho_c)}{3} - \frac{\delta}{2}\right) + 2(1+\rho_c) = 3\delta.$$

Then

$$\Psi_\delta(x) = b'_{\rho_c}(x) + \frac{1}{2} b''_{\rho_c}(x)\delta + \frac{1}{6} b'''_{\rho_c}(x)\delta^2.$$

$$\begin{aligned}\Psi_\delta(0) &= b'_{\rho_c}(0) + \frac{1}{2} b''_{\rho_c}(0)\delta + \frac{1}{6} b'''_{\rho_c}(0)\delta^2 \\ &= -\rho_c + (1+\rho_c)\delta - \delta^2.\end{aligned}$$

$$\begin{aligned}\Psi_\delta(1) &= b'_{\rho_c}(1) + \frac{1}{2} b''_{\rho_c}(1)\delta + \frac{1}{6} b'''_{\rho_c}(1)\delta^2 \\ &= -(1-\rho_c) + (-2+\rho_c)\delta - \delta^2.\end{aligned}$$

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

Therefore,

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

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

$$3|\delta| + \delta^2 < 1 - \max\left\{\rho_c, (1-\rho_c), \frac{1}{3}(\rho_c^2 - \rho_c + 1)\right\}. \tag{C.1}$$

Appendix D

To prove that

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

we proceed by induction. For $k = 1$,

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

The case $k+1$ goes as follows:

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

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):

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

In addition, in the amplified case, the previous bounds are correct if s is sufficiently large and k is not so big. More precisely, when (see Appendix C, Eq. (C.1))

$$\begin{aligned}&3\frac{(1+\sigma)^k - 1}{s\sigma} + \left(\frac{(1+\sigma)^k - 1}{s\sigma}\right)^2 \\ &\quad < 1 - \max\left\{\rho_c, (1-\rho_c), \frac{1}{3}(\rho_c^2 - \rho_c + 1)\right\}.\end{aligned}$$

References

[1] J. von Neumann, A.W. Burks, Theory of Self-Reproducing Automata, University of Illinois Press, Urbana, 1966.
[2] S. Ulam, Random processes and transformations, in: Proceedings of the International Congress on Mathematics, Vol. 2, 1952, pp. 264–275.
[3] N.H. Packard, Dynamic Patterns in Complex Systems, in: Chapter Adaptation Toward the Edge of Chaos, World Scientific, Singapore, 1988, pp. 293–301.
[4] P. Gács, G.L. Kurdyumov, L.A. Levin, One dimensional uniform arrays that wash out finite islands, Problemy Peredachi Informatsii 14 (1978) 92–98 (in Russian).
[5] P. Gonzaga de Sá, C. Maes, The Gács-Kurdyumov-Levin automaton revisited, J. Stat. Phys. 67 (1992) 507–522.
[6] M. Land, R.K. Belew, No perfect two-state cellular automata for density classification exists, Phys. Rev. Lett. 74 (25) (1995) 5148–5150.
[7] H. Juillé, J.B. Pollack, Coevolving the ideal trainer: application to the discovery of cellular automata rules, in: Genetic Programming 1998: Proceedings of the Third Annual Conference, Morgan Kaufmann, San Francisco, 1998.
[8] M. Mitchell, J.P. Crutchfield, P.T. Hraber, Evolving cellular automata to perform computations: mechanisms and impediments, Physica D 75 (1994) 361–391.
[9] P.P.B. de Oliveira, J.C. Bortot, G.M.B. Oliveira, The best currently known class of dynamically equivalent cellular automata rules for density classification, Neurocomputing 70 (1–3) (2006) 35–43.
[10] S. Verel, P. Collard, M. Tomassini, L. Vanneschi, Fitness landscape of the cellular automata majority problem: view from the Olympus, Theoret. Comput. Sci. 378 (1) (2007) 54–77.
[11] D. Wolz, P.P.B. de Oliveira, Very effective evolutionary techniques for searching cellular automata rule spaces, J. Cell. Autom. 3 (2008) 289–312.
[12] M.S. Capcarrere, M. Sipper, M. Tomassini, Two-state, $r = 1$ cellular automaton that classifies density, Phys. Rev. Lett. 77 (24) (1996) 4969–4971.
[13] H. Fukś, Solution of the density classification problem with two cellular automata rules, Phys. Rev. E 55 (3) (1997) R2081–R2084.
[14] C.L.M. Martins, P.P.B. de Oliveira, Evolving sequential combinations of elementary cellular automata rules, in: Advances in Artificial Life, in: Lecture Notes in Computer Science, vol. 3630, Springer, Berlin, Heidelberg, 2005, pp. 461–470.

[15] R. Alonso-Sanz, L. Bull, A very effective density classifier two-dimensional cellular automaton with memory, J. Phys. A 42 (48) (2009) 485101.
[16] C. Stone, L. Bull, Evolution of cellular automata with memory: the density classification task, Biosystems 97 (2) (2009) 108–116.
[17] H. Fukś, Nondeterministic density classification with diffusive probabilistic cellular automata, Phys. Rev. E 66 (6) (2002) 066106.
[18] M. Schüle, T. Ott, R. Stoop, Computing with probabilistic cellular automata, in: ICANN 09: Proceedings of the 19th International Conference on Artificial Neural Networks, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 525–533.
[19] N. Fatès, Stochastic Cellular automata solve the density classification problem with an arbitrary precision, in: Proceedings of the 28th International Symposium on Theoretical Aspects of Computer Science, STACS 2011, 2011, pp. 284–295.
[20] R.S. Cantrell, C. Cosner, Spatial Ecology via Reaction–Diffusion Equations, John Wiley & Sons, Sussex, 2003.
[21] J.P. Keener, J. Sneyd, Mathematical Physiology, Springer, New York, 1998.
[22] J. Xin, Front propagation in heterogeneous media, SIAM Rev. 42 (2) (2000) 161–230.
[23] S.-N. Chow, Lattice dynamical systems, in: Lecture Notes in Mathematics, vol. 1822, 2003, pp. 1–102.
[24] C.M. Elliott, S. Zheng, On the Cahn–Hilliard equation, Arch. Ration. Mech. Anal. 96 (1986) 339.
[25] C. Stone, L. Bull, Solving the density classification task, in: Using Cellular Automaton 184 with Memory, in: Complex Systems, vol. 18, Complex Systems Publications Inc., 2009, pp. 229–344.
[26] R. Courant, K. Friedrichs, H. Lewy, On the partial difference equations of mathematical physics, IBM J. Res. Dev. 11 (2) (1928) 215–234.