

Target-Oriented Self-Structuring in Classifying Cellular Automata

CHRISTOPHER AUER^{1*}, PATRICK WÜCHNER^{1†}, HERMANN DE MEER^{1‡}

*Faculty of Informatics and Mathematics, University of Passau
94032 Passau, Germany*

Some cellular automata are able to solve classification problems on their initial configuration by building globally visible structures. However, no formal measures exist yet for describing or detecting this behavior in general. The lack of such formal methods often leads to quite observer-dependent discussions of emergent computation. In this paper, we propose the measures of *target orientation* and *self-structuring* that allow to formally evaluate a cellular automaton's ability to solve a classification problem by emergent computation. By the means of these measures, globally emerging patterns can be recognized and their contribution to the solution of the classification problem can be judged in an observer-independent way.

Key words: One-dimensional cellular automata; Classification problem; Emergent computation; Formal measures

1 INTRODUCTION

The computational aspects of cellular automata (CAs) have been studied in various publications. The author of [7] presents a short survey and investigates the four different viewpoints on the computational power of CAs:

*email: christopher.auer@uni-passau.de

†email: patrick.wuechner@uni-passau.de

‡email: demeer@fim.uni-passau.de

The first one considers CAs as abstract computing machines, e.g., as a finite automaton or a Turing machine (see, e.g., [11]), and studies the formal languages they accept and their space and time complexity. The second viewpoint tries to develop special initial configurations for the CAs which can then be proven to be as powerful as certain models of computation like Turing machines. For instance, in Conway's Game of Life (cf. [3]), structures are developed, e.g., gliders and glider guns, that are used to produce logic gates by which it is possible to implement universal computation (cf. [22]). The third viewpoint studies the computational mechanics of CAs that is concerned with emergent properties, i.e., spatial and temporal regularities and their dynamics that are visible from a global perspective. Computational mechanics discusses *how* a CA computes but not *what* it computes (cf. [7, 10, 13, 23]). The fourth viewpoint, called *emergent computation* (cf. [7]), investigates *what* a CA computes by its emergent properties.

In this paper, we take the latter viewpoint: We study whether a CA solves a classification problem (CP) on its initial configuration (IC) by producing global structures, i.e., by emergent computation. Here, we are concerned with one-dimensional CAs which consist of cells that are placed on a circle. At each time step, each cell is in a state from a finite set of states. The successor state of a cell is determined by a transition function which takes into account the cell's state and the states of the cell's neighbors. The states of the cells at time step 0 are determined by the CA's IC, which can be seen as the input to the CA. We assume a randomly chosen IC.

In the simplest case, there are just two states (0 and 1) of each cell. An interesting aspect of such two-state CAs is the ability to classify the IC's density of cells in State 1. This problem is called *density CP* (or *majority problem*) and has been studied in various scientific publications, e.g., in [1, 7, 8, 10, 12, 18, 21, 25]. Let ρ be the number of State-1 cells in the IC divided by the number of cells in the CA, then some CAs are able to decide whether $\rho < 1/2$, $\rho = 1/2$, or $\rho > 1/2$. We denote this density CP by $\mathbf{L}_{1/2}$. The existence of CAs being able to solve the density CP is remarkable since the cells' transition rules only take the immediate vicinity of each cell into account, while the density class is a global aspect of the IC.

Different perspectives on how a CA can solve the density CP $\mathbf{L}_{1/2}$ exist, where a common property of these perspectives is that the density class of the IC is communicated to each cell of the CA. For instance, one approach has been studied in [10] by defining that all cells have to converge to State 0 if $\rho < 1/2$ and to State 1 otherwise. The authors of [10] use a genetic algorithm to find CAs that solve $\mathbf{L}_{1/2}$ in the described manner. The results show that

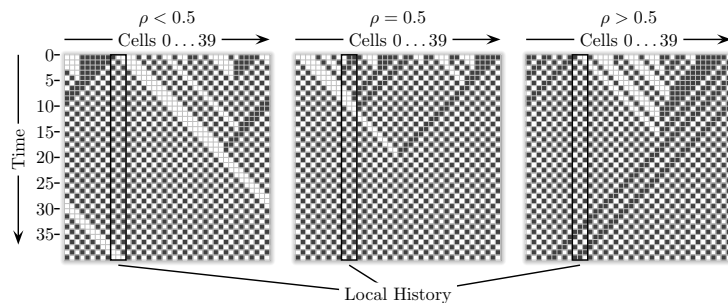


FIGURE 1
Progress of Rule 184 for three different initial configurations (space-time diagrams).

each found CA solves $\mathbf{L}_{1/2}$ only with a high probability. Indeed, as shown in [18], a two-state CA that solves $\mathbf{L}_{1/2}$ in this way cannot exist if only one transition rule is allowed. However, by the consecutive application of two different transition rules, a suitable CA can be found: In [12], such a CA is constructed using Rule 184 at first and then Rule 232, both given in Wolfram’s notation (cf. [25, p. 419]).

Capcarrère et al. show in [8] how Rule 184 solves $\mathbf{L}_{1/2}$ on its own when investigated from a different perspective: Figure 1 illustrates the behavior of a Rule-184 CA for three randomly chosen ICs belonging to the density classes $\rho < 1/2$, $\rho = 1/2$, and $\rho > 1/2$. The progress is displayed using space-time diagrams in which the cells in State 0 are drawn white and cells in State 1 in black. Each row shows the state of the 40 cells and time runs from top to bottom. The first row is the IC of the CA.

Figure 1 shows how the CA builds up and maintains a structure that reflects the density class of the IC: a checkerboard pattern, i.e., an alternating sequence of black and white cells, that is disturbed by adjacent white cells if $\rho < 1/2$, disturbed by adjacent black cells if $\rho > 1/2$, and contains no disturbances if $\rho = 1/2$. In [8], it is shown that Rule 184 always solves $\mathbf{L}_{1/2}$ with respect to this behavior.

In [7], the ability of a CA to classify its IC by building a globally visible, i.e., emergent, structure is called *emergent computation*. According to [7], it is “a matter of visual efficiency” whether a CA can classify its IC by emergent computation, i.e., it strongly depends on the ability of an observer to recognize the globally visible structure and how it relates to the solution

of the density CP. However, this would rule out the existence of a formal and observer-independent criterion for arbitrary CAs to detect and describe emergent computation, as observed with Rule 184.

A formal approach developed in [6], further studied in [10, 13, 17, 23] and [4, p. 194], that describes how CAs can classify their IC is the *particle-based description*. Homogeneous areas are considered as the medium by means of which particles, i.e., disturbances of these homogeneous areas, can propagate. The left-most diagram in Fig. 1 shows such a particle of three white cells after time step 18 that is continuously propagating from the left to the right side on the homogeneous checkerboard pattern. If two particles collide, they interact by extinguishing or modifying each other. According to [6], particles carry information that is exchanged whenever particles interact. For instance, in the space-time diagram for $\rho = 1/2$ in Fig. 1, a white and a black particle interact at time step 18. Informally, the white particle carries the information “ $\rho < 1/2$ at Cells 0 . . . 19 of the IC” and the black particle “ $\rho > 1/2$ at Cells 20 . . . 39 of the IC”. The extinction of both particles combines both pieces of information to “ $\rho = 1/2$ in the IC”. The particle-based description is a high-level description that is helpful to understand *how*, for instance, a CA with Rule 184 solves the density CP $L_{1/2}$. It, however, presumes that the particles are identified, either manually (cf. [6]) or automatically (cf. [23]). Furthermore, the description provides no criteria to decide *whether* a CA classifies its IC by the help of the particles it produces.

In this paper, we generalize the density CP to the classification of ICs with respect to arbitrary equivalence classes on the set of possible ICs. Studying such general CPs gives deeper insights into the computational power of CAs: Any function f that maps an IC to a value, corresponding to a particular property of the IC, implies a CP in which each equivalence class corresponds to all ICs that map to the same value of f . Hence, knowing the equivalence class enables to determine the value of f . In this sense, CAs that solve the CP also compute the function f on the IC. To the best of the authors’ knowledge, still no formalism exists that describes the ability of a CA to solve a general CP by using emergent computation, i.e., by a process of building a global structure.

Two important observations can be made when investigating a Rule-184 CA. In Fig. 1, the sequence of states of Cell 10 and its left- and right-side neighbor is framed by a black rectangle. We call this sequence the Cell 10’s *local history*. The local history incorporates the succession of states of Cell 10 and the states of the cells that locally determine the successor state of Cell 10. The first observation is that the globally visible structure that is built up by

the CA to solve $L_{1/2}$ is also visible in the local history of Cell 10. As the second observation, the density class of the IC can be derived from the local history of Cell 10. These observations similarly apply to all cells in the space-time diagrams of Fig. 1. Hence, the information about the density class is communicated to all cells.

The generalization of these two observations leads to the main contribution of this paper: We introduce two formal measures that describe to which extent a CA builds up a structure in the cells' local history by which it is possible to classify the IC. The first measure, called the *degree of target orientation*, evaluates to which extent it is possible to classify the IC when only the local history of a cell is given. A high average degree of target orientation then indicates that the IC's equivalence class can be derived from an arbitrary cell's local history, i.e., the equivalence class is distributed among all cells. Hence, an external observer does not have to scan the entire CA to classify the IC. We then define that the CA *solves* the CP.

The second measure, called the *degree of self-structuring*, describes to which extent the IC's equivalence class influences the local histories of the cells. A high degree of self-structuring indicates that the CA does not distribute superfluous information about the IC. By combining both measures, it is then possible to detect and analyze a CA's ability to solve a CP on its IC by using a process we call *target-oriented self-structuring*. In this sense, the two degrees help to characterize emergent computation in CAs in an observer-independent way.

Both measures utilize Shannon's information entropy. The approach of using the entropy to characterize emergent properties is also proposed in, e.g., [14, 16, 15, 2, 23, 19]. Recently, in [14, 16, 15], entropy is used to define the level of autonomy, emergence, adaptivity, homogeneity, resilience, and self-organization of a system of interacting entities. The measures proposed in [14, 16, 23, 19] evaluate emergent properties that are *intrinsic* to a system, i.e., independently of the task the system ought to fulfill. In contrast, the measures presented in this paper evaluate a CA's emergent properties *with respect to* a given CP. In [2], we define the system entities' degree of global-state awareness which is, in the case of CAs, related to the degree of target-orientation presented in this paper. However, [2] does not include degree of self-structuring and is focusing on the engineering of technical self-organizing systems.

The remainder of this paper is structured as follows: Section 2 introduces basic notation, gives a primer to Shannon's information entropy, which we use in our definitions, introduces this paper's model of one-dimensional CAs,

and defines the concept of a cell's local history and the CP. Section 3 introduces the degrees of target orientation and self-structuring and gives further insights into the relationship between the two degrees. Section 4 applies our definitions to five examples of CPs. Finally, Sec. 5 provides a summary and gives an outlook to future work.

2 PRELIMINARIES

2.1 Entropy of Information

In this paper, a random variable is denoted by a capital letter X , the bold-face letter \mathbf{X} denotes the range of X , and $x \in \mathbf{X}$ its realization. $P[X = x]$ is the probability that X takes the value x . We are only concerned with discrete random variables with a finite range, i.e., $\#\mathbf{X} < \infty$, where $\#\mathbf{X}$ is the number of elements in \mathbf{X} .

The entropy $H[X]$ of a discrete random variable X is defined as (cf. [9, 24]):

$$H[X] = - \sum_{x \in \mathbf{X}} P[X = x] \cdot \log_2 P[X = x],$$

and measures the uncertainty of the outcome of X : The higher $H[X]$, the less certain is the prediction of the outcome of X . A uniform distribution leads to $H[X] = \log_2 \#\mathbf{X}$, which is maximal. $P[X = x] = 1$ for exactly one $x \in \mathbf{X}$ yields $H[X] = 0$, which is minimal.

$H[X]$ can also be interpreted as the mean amount of information (measured in bits) needed to specify the outcome of X or, equivalently, the mean amount of information provided by the outcome of X .

If X and Y are random variables with ranges \mathbf{X} and \mathbf{Y} , respectively, the conditional entropy $H[X|Y]$ measures the remaining entropy of X if the outcome of Y is known (cf. [24]):

$$H[X|Y] = H[X, Y] - H[Y], \quad (1)$$

where $H[X, Y]$ is the entropy of the joint random variable (X, Y) .

The following two observations, taken from [9], are crucial to understand the concepts presented in this paper. First,

$$H[X|Y] = H[X] \iff X \text{ and } Y \text{ are independent}, \quad (2)$$

i.e., the knowledge of the outcome of Y gives no information on the outcome of X iff X and Y are independent random variables. Second, at the other

extreme, the outcome of X can be predicted certainly by knowing Y iff X is a non-stochastic function of Y :

$$H[X|Y] = 0 \iff X = f(Y). \quad (3)$$

2.2 Model Description

In this paper, we focus on one-dimensional two-state CAs for the purpose of clarity and conciseness. However, all concepts presented in this section and Sec. 3 can readily be applied to more general CA classes.

A one-dimensional two-state CA is defined by the tuple (n, R_{r_l, r_r}) and denoted by $\mathcal{C}_{R_{r_l, r_r}}^n$: The first element of the tuple is the number $n \in \mathbb{N}$ of cells, where the set of cells is denoted by $\mathbb{N} = \{0, \dots, n-1\}$. The cells are placed on a one-dimensional torus (circle), i.e., each cell i has two immediate neighbor cells $(i+1 \bmod n)$ and $(i-1 \bmod n)$, called left- and right-side neighbor of cell i , respectively. In the following, the indices of the cells are calculated modulo- n and we omit “mod n ” entirely.

The second element of the tuple is R_{r_l, r_r} , where the integer $R \in \mathbb{N}_0$ defines the transition function ϕ of the automaton’s cells in Wolfram’s notation (cf. [25, p. 419]). Each cell can take one of two states of the state space $\mathcal{S} = \{0, 1\}$. To calculate the successor state of cell i , the states of cell i , its $r_l \in \mathbb{N}_0$ left-side neighbors, and its $r_r \in \mathbb{N}_0$ right-side neighbors are taken into account, i.e., $\phi : \{0, 1\}^{r_l+1+r_r} \rightarrow \{0, 1\}$. ϕ is assumed to be identical for all cells. However, the ideas presented in the following also apply when cells have individual transition functions.

If $r_l = r_r = 1$, we omit the subscripts in “ R_{r_l, r_r} ”. We omit the superscript n in “ $\mathcal{C}_{R_{r_l, r_r}}^n$ ” if the corresponding statement refers to one-dimensional two-state CAs of arbitrary finite size. We also omit the subscript R_{r_l, r_r} if the corresponding statement refers to one-dimensional two-state CAs with arbitrary transition functions.

The state of a CA \mathcal{C} at time step $t \in \mathbb{N}_0$ is denoted by the configuration $\gamma_t \in \Gamma$, where $\Gamma = \mathcal{S}^n$ is the state space of the CA and γ_t is an n -dimensional vector which defines the state of each cell. The state of cell i ($i \in \mathbb{N}$) at time step $(t+1)$ is denoted by $\gamma_{t+1, i}$ and can be calculated by:

$$\gamma_{t+1, i} = \phi(\gamma_{t, i-r_l}, \gamma_{t, i-r_l+1}, \dots, \gamma_{t, i}, \dots, \gamma_{t, i+r_r-1}, \gamma_{t, i+r_r}),$$

where γ_0 is the IC.

The CA starts to evolve with a random IC Γ_0 with realization γ_0 and range $\Gamma_0 \subseteq \Gamma$. The CA’s random configuration at time step $t \in \mathbb{N}_0$ is denoted by Γ_t and the corresponding random state of cell i by $\Gamma_{t, i}$.

The state of a cell, the states of the cell's r_l left-side, and the cell's r_r right-side neighbors constitute the *local configuration* of the cell. We call the sequence of a cell's local configurations until time step t the cell's *local history* $\overleftarrow{\Gamma}_{t,i} = (\Gamma_{t',i-r_l}, \Gamma_{t',i-r_l+1}, \dots, \Gamma_{t',i}, \dots, \Gamma_{t',i+r_r-1}, \Gamma_{t',i+r_r})_{t'=0\dots t}$. The range of $\overleftarrow{\Gamma}_{t,i}$ is denoted by $\overleftarrow{\Gamma}_{t,i}$ and its realization by $\overleftarrow{\gamma}_{t,i}$.

We generalize the density CP, as discussed in Sec. 1, to general CPs as follows: Let \mathbf{L} be a non-trivial partition of Γ_0 into equivalence classes denoted by $l \in \mathbf{L}$, i.e., the union of all $l \in \mathbf{L}$ equals Γ_0 , all $l \in \mathbf{L}$ are disjoint, non-empty, and \mathbf{L} contains at least two classes. $\delta : \Gamma_0 \rightarrow \mathbf{L}$ is a function which maps any IC γ_0 to its corresponding equivalence class l in \mathbf{L} , i.e.,

$$\forall l \in \mathbf{L} : \forall \gamma_0 \in \Gamma_0 : \delta(\gamma_0) = l \iff \gamma_0 \in l.$$

By applying the function δ to the random variable Γ_0 , the random variable $L = \delta(\Gamma_0)$ with realization $l \in \mathbf{L}$ is obtained. L is the random variable of the equivalence class of the IC and it inherits its probability distribution from Γ_0 . The *classification problem* corresponding to the partition \mathbf{L} (or, in short, “CP \mathbf{L} ”) is the problem of determining the outcome of L , that is, the equivalence class $l \in \mathbf{L}$ of the IC.

3 TARGET-ORIENTED SELF-STRUCTURING

We now introduce two measures, namely, the degree of target orientation (DOTO; Sec. 3.1) and the degree of self-structuring (DOSS; Sec. 3.2). The DOTO measures to which extent a CA solves a CP and the DOSS to which extent the IC's equivalence class determines the structure in the local histories of the cells.

3.1 Degree of Target Orientation

To measure the uncertainty of the IC's equivalence class L when the outcome of $\overleftarrow{\Gamma}_{t,i}$ is known, the conditional entropy (cf. Sec. 2.1) can be used:

Definition 1. *The degree of target orientation $\tau_{t,i}(\mathcal{C}, \mathbf{L})$ of \mathcal{C} observed at cell i ($i \in \mathbf{N}$) at time step t is defined as:*

$$\tau_{t,i}(\mathcal{C}, \mathbf{L}) = 1 - \frac{H[L | \overleftarrow{\Gamma}_{t,i}]}{H[L]}, \quad (4)$$

where $0/0 := 0$. The average degree of target orientation $\tau(\mathcal{C}, \mathbf{L})$ of \mathcal{C} is then defined as the limiting value of $\tau_{t,i}(\mathcal{C}, L)$ for $t \rightarrow \infty$ averaged over all cells:

$$\tau(\mathcal{C}, \mathbf{L}) = \lim_{t \rightarrow \infty} \frac{1}{n} \sum_{i \in \mathbf{N}} \tau_{t,i}(\mathcal{C}, \mathbf{L}). \quad (5)$$

Note that $H[L]$ is the entropy of L if no additional information is given and, consequently, $H[L|\overleftarrow{\Gamma}_{t,i}] \leq H[L]$. Hence, $H[L|\overleftarrow{\Gamma}_{t,i}]/H[L]$ lies in $[0, 1]$ and so do $\tau_{t,i}(\mathcal{C}, \mathbf{L})$ and $\tau(\mathcal{C}, \mathbf{L})$.

With the denominator $H[L]$ in Eq. (4) being 0, $H[L|\overleftarrow{\Gamma}_{t,i}]$ also takes the value 0 and, consequently, $P[L = l] = 1$ for exactly one $l \in \mathbf{L}$, i.e., the outcome of L can certainly be predicted. Hence, we define $0/0 := 0$ in Eq. (4), which yields $\tau_{t,i}(\mathcal{C}, \mathbf{L}) = 1$.

$\tau_{t,i}(\mathcal{C}, \mathbf{L}) = 1$ iff $H[L|\overleftarrow{\Gamma}_{t,i}] = 0$. This is the case iff L is a (non-stochastic) function of $\overleftarrow{\Gamma}_{t,i}$, i.e., $L = f(\overleftarrow{\Gamma}_{t,i})$ (Eq. (3)). By applying function f to the local history of cell i , the outcome of L can be predicted. Hence, we define that the CA *solves* \mathbf{L} iff $\tau_{t,i}(\mathcal{C}, \mathbf{L}) = 1$. Note that a DOTO of 1 only proves the existence of f without explicitly constructing f . In [2, Sec. 5], an engineering approach is given how to construct such an f .

At the other extreme, $\tau_{t,i}(\mathcal{C}, \mathbf{L}) = 0$ holds iff $H[L|\overleftarrow{\Gamma}_{t,i}] = H[L]$ or, equivalently, $\overleftarrow{\Gamma}_{t,i}$ and L are independent random variables (Eq. (2)). Hence, knowing $\overleftarrow{\Gamma}_{t,i}$ provides no information on L at all.

All intermediate values of $\tau_{t,i}(\mathcal{C}, \mathbf{L}) \in (0, 1)$ can be interpreted similarly: Any value of $\tau_{t,i}(\mathcal{C}, \mathbf{L})$ close to 1 means a high certainty when predicting L if $\overleftarrow{\Gamma}_{t,i}$ is known. Hence, a function f exists that maps the majority but not all local histories to the correct equivalence class of the IC. A low value of $\tau_{t,i}(\mathcal{C}, \mathbf{L})$ implies that the knowledge of $\overleftarrow{\Gamma}_{t,i}$ provides only little information about L .

Note that a high average DOTO $\tau(\mathcal{C}, \mathbf{L})$ indicates that the information about L is distributed among most cells since all cells equally contribute to the average DOTO of the CA: If L can be predicted from the local history of each cell, the necessary information to predict L is distributed throughout the whole CA.

Moreover, $\tau_{t,i}(\mathcal{C}, \mathbf{L})$ is non-decreasing over time since information is accumulated in the cell's local history. Since $\tau_{t,i}(\mathcal{C}, \mathbf{L})$ is bounded from above by 1 for all $t \in \mathbb{N}$, the limiting value for $t \rightarrow \infty$ exists and so does the limiting value in Eq. (5).

3.2 Degree of Self-Structuring

So far, by using the DOTO, we can only measure to which extent a CA solves a CP. In order to evaluate to which degree a CA uses self-structuring to obtain its solution, we complement our definition of the DOTO by introducing the DOSS, i.e., the degree of self-structuring.

An important implication of Def. 1 is the following: There exists a CA whose cells are placed on a one-dimensional torus and with a DOTO of 1

for every CP: CA \mathcal{C}_{240} uses Rule 240 that simply moves the state of the left-side neighbor to the current cell, i.e., all cells continuously shift the IC to the right. By this mechanism, the IC is mapped to the local history of each cell and, consequently, L is a function of the local history and $\tau(\mathcal{C}_{240}, \mathbf{L}) = 1$ for any CP \mathbf{L} .

As an example, take the density CP $\mathbf{L}_{1/2}$: Both \mathcal{C}_{184} and \mathcal{C}_{240} solve this CP by distributing the information about the IC's density class among all cells. However, in contrast to \mathcal{C}_{240} , \mathcal{C}_{184} simplifies the local history such that superfluous information that is not needed to specify the density class is minimal. In this sense, \mathcal{C}_{184} solves the density CP more economically than \mathcal{C}_{240} , i.e., the function f that maps $\overleftarrow{\Gamma}_{t,i}$ to L has to extract less information from the local history. This is also shown more formally in Secs. 3.3 and 4. Let $i \in \mathbf{N}$ be a cell of the CA and $\overleftarrow{\Gamma}_{t,i}$ be the local history of cell i , then the amount of superfluous information in the local history of cell i is minimal if the IC's equivalence class L has major impact on the local history $\overleftarrow{\Gamma}_{t,i}$. Similarly to Def. 1, the impact of L on $\overleftarrow{\Gamma}_{t,i}$ can be measured by the conditional entropy of $\overleftarrow{\Gamma}_{t,i}$ when L is known, which leads to the definition of the DOSS:

Definition 2. *The degree of self-structuring $\sigma_{t,i}(\mathcal{C}, \mathbf{L})$ of \mathcal{C} observable at cell i ($i \in \mathbf{N}$) at time step t is defined as follows:*

$$\sigma_{t,i}(\mathcal{C}, \mathbf{L}) = 1 - \frac{\mathrm{H}[\overleftarrow{\Gamma}_{t,i}|L]}{\mathrm{H}[\overleftarrow{\Gamma}_{t,i}]}, \quad (6)$$

where $0/0 := 0$. The average degree of self-structuring $\sigma(\mathcal{C}, \mathbf{L})$ of \mathcal{C} is then defined as the average of $\sigma_{t,i}(\mathcal{C}, L)$ for $t \rightarrow \infty$:

$$\sigma(\mathcal{C}, \mathbf{L}) = \lim_{t \rightarrow \infty} \frac{1}{n} \sum_{i \in \mathbf{N}} \sigma_{t,i}(\mathcal{C}, \mathbf{L}). \quad (7)$$

In contrast to the DOTO, the DOSS is non-increasing over time since the information provided by L is constant while the amount of information accumulated in $\overleftarrow{\Gamma}_{t,i}$ is non-decreasing. Also note that a high average DOSS implies that the impact of L on the local history can be seen at all cells and, consequently, L determines to a large extent the state history of the whole CA.

3.3 The Relationship between the Proposed Measures

In this section, we give further insights into the relationship between the DOTO and the DOSS, defined in Secs. 3.1 and 3.2, respectively.

Note that the DOTO (Def. 1) measures to which extent L is a function of $\overleftarrow{\Gamma}_{t,i}$, whereas the DOSS (Def. 2) measure the converse, i.e., how much $\overleftarrow{\Gamma}_{t,i}$ is a function of L . This symmetry becomes more apparent by the following relationship:

$$\frac{\tau_{t,i}(\mathcal{C}, \mathbf{L})}{\sigma_{t,i}(\mathcal{C}, \mathbf{L})} = \frac{H[\overleftarrow{\Gamma}_{t,i}]}{H[L]}. \quad (8)$$

This equation follows immediately from Eq. (1). The right side of Eq. (8) relates the amount of information provided by the local history $\overleftarrow{\Gamma}_{t,i}$ to the amount of information that is needed to specify L .

As an example, suppose a DOTO with value 1, then L can always be derived from the local history. Equation (8) simplifies to:

$$\sigma_{t,i}(\mathcal{C}, \mathbf{L}) = \frac{H[L]}{H[\overleftarrow{\Gamma}_{t,i}]}.$$

Since $\sigma_{t,i}(\mathcal{C}, \mathbf{L}) \leq 1$, it follows that $H[\overleftarrow{\Gamma}_{t,i}] \geq H[L]$. In other words, the information provided by the local history is the least amount that is needed to specify L plus additional information that is irrelevant for determining L . The smaller this surplus, the larger the DOSS. Hence, the DOSS can also be interpreted as a measure of how economically the classification problem is solved: CAs that maximize the DOSS and achieve a DOTO of 1 distribute the information about L with a minimal amount of superfluous information.

At the other extreme, let the DOSS be 1. In this case, $\overleftarrow{\Gamma}_{t,i}$ is a function of L . Then, the IC's equivalence class completely determines the local history, that is, the CA strongly structures itself with respect to L . It follows that:

$$\tau_{t,i}(\mathcal{C}, \mathbf{L}) = \frac{H[\overleftarrow{\Gamma}_{t,i}]}{H[L]},$$

and, consequently, $H[\overleftarrow{\Gamma}_{t,i}] \leq H[L]$. Assuming that $H[\overleftarrow{\Gamma}_{t,i}] < H[L]$, this corresponds to the case where the structure built up by the CA is too simple to solve the classification problem: Since the local history is a function of L and $H[\overleftarrow{\Gamma}_{t,i}] < H[L]$, at least two different equivalence classes $l, l' \in \mathbf{L}$ map to the same local history which makes these classes indistinguishable. In general, the information provided by the local history is then not enough to determine the IC's equivalence class, i.e., there is a deficiency of information to determine L . CAs that maximize the DOTO minimize this deficiency.

This discussion shows that the DOTO and DOSS describe two different aspects of the CA. The combination of both describe a CA's ability to solve a

CP by a process of self-structuring. Whether the emphasis should be put on the DOTO, the DOSS, or equally on both depends on the specific application. For instance, CAs that optimize a trade-off between the DOTO and DOSS maximize the following weighted sum:

$$\alpha\tau(\mathcal{C}, \mathbf{L}) + (1 - \alpha)\sigma(\mathcal{C}, \mathbf{L}),$$

where parameter $\alpha \in [0, 1]$ allows to control the relative importance of the DOTO. For instance, if both degrees are considered equally important, then α can be set to $1/2$.

4 EXAMPLES

In the following sections, we calculate the average DOSS and DOTO for different CAs applied to solve five distinct CPs and select those CAs that achieve an average DOTO of 1 and maximize the average DOSS. As discussed in Sec. 3.3, by this selection mechanism, we can find those CAs that distribute the information about the IC's equivalence class throughout the whole CA with a minimum amount of information that is irrelevant to solve the CP. In this sense, the classification problem is solved most economically. In the remainder of this section, we refer to the average DOTO and average DOSS as DOTO and DOSS, respectively.

For each CP \mathbf{L} , we choose the following probability distribution for the IC Γ_0 : *

$$\forall l \in \mathbf{L} : P[L = l] = \frac{1}{\#\mathbf{L}} \wedge \forall \gamma_0 \in l : P[\Gamma_0 = \gamma_0 | L = l] = \frac{1}{\#l}. \quad (9)$$

This probability distribution corresponds to the worst-case scenario: Any equivalence class $l \in \mathbf{L}$ of the CP is equally probable, maximizing $H[L]$. Similarly, each IC $\gamma_0 \in l$, given that $L = l$, is equally probable. If a CA achieves an average DOTO of 1 when the probability distribution of Eq. (9) is applied, then it also solves the CP for any other probability distribution of Γ_0 . Informally, the reason for this is the following: With Eq. (9), all ICs in Γ_0 have a non-zero probability, i.e., Γ_0 has full support, and L is a non-stochastic function of $\bar{\Gamma}_{i,t}$ for all $i \in \mathbf{N}$ and a sufficiently large t . Then, L is also a non-stochastic function if Γ_0 has a different probability distribution with potentially non-full support.

To determine the DOTO and DOSS for a CA of size n , we run the CA for every IC in Γ_0 until the limit cycle is reached to ensure that $\tau_{t,i}(\mathcal{C}, \mathbf{L})$ and

*Note that any equivalence class $l \in \mathbf{L}$ is a set $l \subseteq \Gamma_0$.

$\sigma_{t,i}(\mathcal{C}, \mathbf{L})$ are constant. The results, i.e., the local histories of the cells and the IC's equivalence class, are then weighed with the corresponding probabilities given in Eq. (9) to calculate the entropies needed for the Eqs. (4) and (6). This calculation is only tractable for small sizes n of the CA. For larger n , the DOTO and DOSS can be approximated by running the CA for a fixed number of randomly generated ICs according to Eq. (9).

4.1 1/2-Density Classification

As an example, for a symmetric neighborhood of $r_l = r_r = 1$, size $n = 12$, and the density classification problem $\mathbf{L}_{1/2}$, among the $2^{2^3} = 256$ rules four rules that achieve a DOTO of 1 can be found, namely, Rules 170, 184, 226, and 240. For the DOSS we obtain the following values:

$$\begin{aligned}\sigma(\mathcal{C}_{184}^{12}, \mathbf{L}_{1/2}) &= \sigma(\mathcal{C}_{226}^{12}, \mathbf{L}_{1/2}) = 0.3328, \\ \sigma(\mathcal{C}_{240}^{12}, \mathbf{L}_{1/2}) &= \sigma(\mathcal{C}_{170}^{12}, \mathbf{L}_{1/2}) = 0.2717.\end{aligned}$$

\mathcal{C}_{184}^{12} and \mathcal{C}_{226}^{12} maximize the DOSS. Rule 226 is the mirrored version of Rule 184, i.e., in a CA with Rule 226, white cells are propagating from right to left, while in a Rule-184 CA, they are propagating from left to right (see also Fig. 1). Consequently, they both solve $\mathbf{L}_{0.5}$ equivalently. Rule 240, which continuously shifts the IC to the right side, was already discussed in Sec. 3.2. Rule 170 is the mirrored version of Rule 240 and their DOSS is lower than the DOSS of Rules 184 and 226. These results are in accordance with the discussion in Secs. 3.2 and 3.3 stating that \mathcal{C}_{184}^{12} solves $\mathbf{L}_{1/2}$ more economically than \mathcal{C}_{240}^{12} .

4.2 Even-Odd Distance Classification

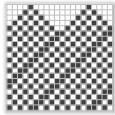

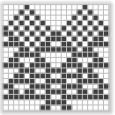
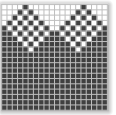
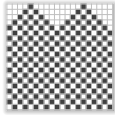
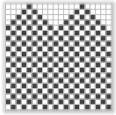

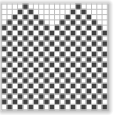
We now consider the following CP: The set of possible ICs consists of all configurations in which exactly two cells are in State 1 and the CA has to determine whether these two cells have odd or even distance, presuming a CA with even size. The set of possible ICs is defined as:

$$\mathbf{\Gamma}_0 = \{\gamma_0 \in \mathcal{S}^n \mid \sum_{i \in \mathbf{N}} \gamma_{0,i} = 2\},$$

and the classification problem as $\mathbf{L}_{o/e} = \{l_o, l_e\}$, where the two cells in State 1 have odd distance in class l_o and even distance in class l_e . For this CP, we restrict ourselves to transition functions with a symmetric neighborhood of radius 1, i.e., $r_l = r_r = 1$, of which there are 256.

We calculated the DOTO and the DOSS for all 256 possible transition rules for CAs of size $n = 8, 10, \dots, 20$. Of these rules, 95 achieve a DOTO

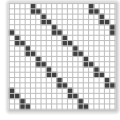
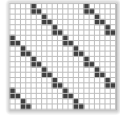
TABLE 1
Space-time diagrams of the rules with a DOTO of 1 and maximum DOSS with respect to $\mathbf{L}_{o/e}$.

Rule	58, 186	114, 242	122	250
$L = l_e$				
$L = l_o$				

of 1, where six rules solve the CP also by maximizing the DOSS. These six rules are listed in Tab. 1, where exemplary space-time diagrams are given for the two classes of $\mathbf{L}_{o/e}$. All six rules solve $\mathbf{L}_{o/e}$ in a similar fashion: Starting from the two black cells, two cones of a checkerboard pattern are extending in both directions. If the two cells have an odd distance in the IC, i.e., $L = l_o$, then all rules continue the checkerboard pattern when the two cones meet. If $L = l_e$, two adjacent black cells can be observed at the points, where the cones meet. This information of two adjacent black cells is then propagated throughout the whole CA: Rules 58 and 186 (the first row and column of Tab. 1) move this information periodically to the left side, Rules 114 and 242 (first row, second column) to the right side, and Rules 122 and 250 (third row) in both directions. This propagation of information is comparable to the particle-based description as described in Sec. 1. Note that the binary representations of Rules 58 and 186 differ only in the most significant bit, which defines the mapping of three adjacent black cells. Since, this pattern of three adjacent cells never occurs with Rules 58 and 186, both rules produce exactly the same behavior. The same is true for Rules 114 and 242.

For size $n = 20$, the six rules of Tab. 1 achieve a DOSS with a value of 0.1548 which is maximal. In comparison, Rule 148 achieves a DOTO of 1 but a DOSS of only 0.1351. The investigation of the space-time diagrams of Rule 148 (see Tab. 2) shows that Rule 148 solves $\mathbf{L}_{o/e}$ very similar to Rule 240, which is discussed in Sec. 3.2: Rule 148 continuously shifts the two black cells to the right side, where, in comparison to Rule 240, an addi-

TABLE 2
Space-time diagrams of Rules 148 with a DOTO of 1 and a DOSS that is not maximal with respect to $\mathbf{L}_{o/e}$.

Rule	$L = l_e$	$L = l_o$
148		

tional oscillation between two adjacent and a single black cell is involved at each time step. Hence, Rule 148 effectively maps the whole IC to the local history of each cell. In contrast, the rules in Tab. 1 distribute considerable less information, i.e., two adjacent black cells or the checkerboard pattern. It can be seen that the DOSS reflects how economically a CA solves the CP, by measuring the intensity of the IC's equivalence class' impact on the local history.

4.3 1/3-Density Classification

Let ρ again be the density of State-1 cells in the IC. Then, the density CP $\mathbf{L}_{1/3}$ is defined by seven classes on $\Gamma_0 = S^n$ which correspond to:




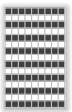


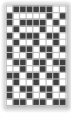
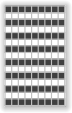
$$\rho = 0, \quad \rho = \frac{k}{3}, \quad \text{and} \quad \frac{k-1}{3} < \rho < \frac{k}{3},$$

for $k = 1, 2, 3$.

For this CP, we allow an asymmetric neighborhood with $r_l = 2$ and $r_r = 1$. Again, we calculated the DOTO and DOSS for all $2^{2^4} = 65536$ rules and size $n = 9$. Among these rules, 706 achieve a DOTO of 1 and two rules (12605_{2,1} and 17267_{2,1}) maximize the DOSS. Table 3 shows space-time diagrams for both rules in the cases $\rho = 0$, $\rho = 1/3$, and $\rho = 2/3$. Both rules expose an oscillating behavior: Homogeneous areas of black cells are mapped to white cells and vice versa. Although the space-time diagrams converge to regular patterns in Tab. 3, it is hard to recognize how the rules solve $\mathbf{L}_{1/3}$ due to the oscillation.

To better understand the behavior of Rules 12605_{2,1} and 17267_{2,1}, we suppress the oscillation by inverting the states of all cells at all odd time steps $t = 1, 3, 5, 7, \dots$ in the space-time diagrams. The result of this filtering process is displayed in Tab. 4. For the case $\rho = 1/3$, both rules separate all black

TABLE 3
 Space-time diagrams of Rules 12605_{2,1} and 17267_{2,1} that solves $\mathbf{L}_{1/3}$ and maximize the DOSS.

Rule	ρ			
	0	1/3	2/3	1
12605 _{2,1}				
17267 _{2,1}				

by white cells until all black cells are separated by exactly two white cells. The resulting pattern, i.e., the information “all black cells can be separated by exactly two white cells”, is then continuously shifted through the CA, indicating $\rho = 1/3$ at all cells. Similarly, for $0 < \rho < 1/3$, the information “there is at least one pair of black cells that is separated by more than two white cells” is propagated within the CA. Most interestingly, for the cases $\rho = 2/3$ and $2/3 < \rho < 1$, both rules simply swap the roles of black and white. Whenever it is neither possible to separate all black by two white cells nor all white by two black cells, it follows that $1/3 < \rho < 2/3$ as displayed in Tab. 4. Table 5 shows two space-time diagrams of a CA with 90 cells and Rule 12605_{2,1} for two exemplary ICs with $\rho = 1/3$ and $\rho = 2/3$, respectively.

Consequently, by these results, the DOTO and the DOSS detect target-oriented self-structuring in CAs even if it involves oscillation.

Among the rules that achieve a DOTO of 1 is also Rule 47288_{2,1} which is equivalent to Rule 184. Indeed, it is possible to derive the IC’s density class with respect to $\mathbf{L}_{1/3}$ from the local history generated by \mathcal{C}_{184} : In the limit cycle, the number of cells that disturb the checkerboard pattern can be counted and added to $\lfloor n/2 \rfloor$. The result is the IC’s number of cells in State s , where $s = 0$ if $\rho < 1/2$ and $s = 1$ if $\rho > 1/2$, from which the respective density class of the IC can be derived. In comparison, the information distributed by Rules 12605_{2,1} and 17267_{2,1} is only the information whether all black/white cells can be separated by exactly two white/black cells, which is less infor-

TABLE 4
 Filtered space-time diagrams of Rule 12605_{2,1} and 17267_{2,1}, where the states are inverted at all odd time-steps.

Rule	ρ				
	(0, 1/3)	1/3	(1/3, 2/3)	2/3	(2/3, 1)
12605 _{2,1}					
17267 _{2,1}					

TABLE 5
 Filtered space-time diagrams of Rule 12605_{2,1} for a CA with 90 cells.

$\rho = 1/3$	$\rho = 2/3$

mation to extract from the local history of a cell. Again, this is reflected by the DOSS which has a value of 0.4183 for Rules 12605_{2,1} and 17267_{2,1} and a value of 0.4060 for Rule 184 for size $n = 9$.

4.4 1/4-Density Classification

We now investigate the density CP $\mathbf{L}_{1/4}$ that consists of nine equivalence classes corresponding to:

$$\rho = 0, \quad \rho = \frac{k}{4}, \quad \text{and} \quad \frac{k-1}{4} < \rho < \frac{k}{4},$$

for $k = 1, 2, 3, 4$.

For this example, we calculated the DOTO and DOSS for transition functions with a symmetric neighborhood of $r_l = r_r = 2$ for sizes $n = 8$ and $n = 12$. Since there are $2^{2^5} \approx 4 \cdot 10^9$ different transition rules with $r_l = r_r = 2$, which is too many for an exhaustive calculation, we restrict ourselves to the 428 *number-conserving rules*, i.e., those transition functions for which the number of cells in State 0 (and State 1) stays constant for all time steps (cf. [5]). A list of all number-conserving rules with $r_l = r_r = 2$ can be found in the WWW,[†] while the theoretical background is given in [20].

Among these 428 rules, 296 rules achieve a DOTO of 1 and two rules maximize the DOSS ($R_{2,2}^{(1)} := 3202581216_{2,2}$ and $R_{2,2}^{(2)} := 4173215874_{2,2}$). Space-time diagrams for the cases $\rho = k/4$ ($k = 1, 2, 3$) are shown for both rules and size $n = 12$ in Tab. 6.

Rule $R_{2,2}^{(1)}$ is the mirrored version of Rule $R_{2,2}^{(2)}$ (see Tab. 6). Consequently, they solve $\mathbf{L}_{1/4}$ similarly. To indicate $\rho = 1/4$, all black cells are separated by exactly three white cells (see Tab. 6). For $\rho = 3/4$, the same pattern emerges only with black and white in swapped roles. If $\rho = 1/2$, all black cells are separated by exactly one white cell, which results in a pattern of vertical stripes. No exact separation by one or three cells is possible for $\rho = 1/3$ and $\rho = 2/3$. Table 7 shows two space-time diagrams of a CA with 40 cells and Rule $R_{2,2}^{(1)}$ for three exemplary ICs with $\rho = 1/4$, $\rho = 1/2$, and $\rho = 3/4$, respectively.

These results, together with the results of Sec. 4.3 and the discussion of Rule 184 in Sec. 1, indicate a general principle of how a CA can solve a density CP by trying to separate all State-0 (State-1) cells by a fixed number of State-1 (State-0) cells. By the help of the DOTO and the DOSS, we are able to systematically find such transition rules even if their behavior involves oscillation (see also Sec. 4.3).

[†]<http://www.dim.uchile.cl/~anmoreir/ncca/mywork.html> (last accessed: November 13, 2009)

TABLE 6
 Space-time diagrams of Rules $R_{2,2}^{(1)} = 3202581216_{2,2}$ and $R_{2,2}^{(2)} = 4173215874_{2,2}$
 that solve $\mathbf{L}_{1/4}$ and maximize the DOSS.

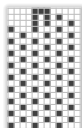
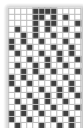
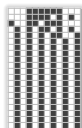
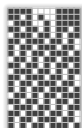
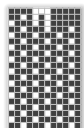
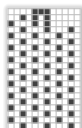
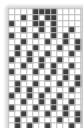

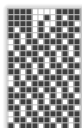
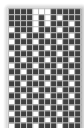
Rule	ρ				
	1/4	1/3	1/2	2/3	3/4
$R_{2,2}^{(1)}$					
$R_{2,2}^{(2)}$					

TABLE 7
 Space-time diagrams of Rule $R_{2,2}^{(1)} = 3202581216_{2,2}$ for a CA with size 40.

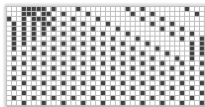

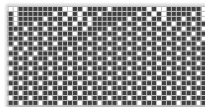
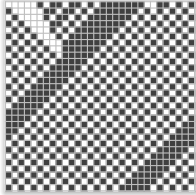
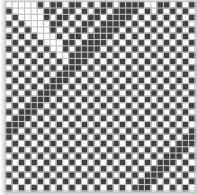
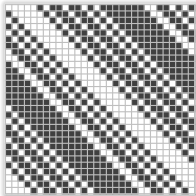
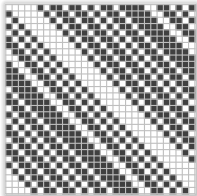
$\rho = 1/4$	$\rho = 1/2$	$\rho = 2/3$
		

TABLE 8
Space-time diagrams of Rules 184 and 240.

	$L = l_{\mathbb{P}}$ 17 black cells	$L = l_{\overline{\mathbb{P}}}$ 16 black cells
184		
240		

4.5 Prime Number Classification

In this final example, we investigate the 256 rules with $r_l = r_r = 1$ with respect to the prime-number CP: The CA has to decide whether there is a prime number of black cells in the IC. Formally, the prime-number CP $\mathbf{L}_{\mathbb{P}}$ is defined as $\mathbf{L}_{\mathbb{P}} := \{l_{\mathbb{P}}, l_{\overline{\mathbb{P}}}\}$, with:

$$l_{\mathbb{P}} = \{\gamma_0 \in \Gamma^n \mid \sum_{i \in \mathbb{N}} \gamma_{0,i} \in \mathbb{P}\} \quad \text{and} \quad l_{\overline{\mathbb{P}}} = \Gamma^n \setminus l_{\mathbb{P}},$$

where \mathbb{P} is the set of prime numbers.

We calculated the DOTO and DOSS for all rules for sizes $n = 6, 7, \dots, 11$. Among these rules, 30 achieve a DOTO of 1 and Rules 184 and its mirrored version Rule 226 maximize the DOSS with a value of 0.11 (for $n = 11$). The remaining 28 rules achieve a DOSS of 0.09 (for $n = 11$). Among these 28 rules is Rule 240, which periodically shifts the IC to the right (see also Sec. 3.2).

Table 8 shows space-time diagrams of a CA with size $n = 30$ with Rules 184 and 240 for two exemplary ICs with a prime and non-prime number of black cells. When looking at the space-time diagrams of Rule 184 in Tab. 8,

there seems to be no discrimination between the prime and non-prime case in the structure built up in the space-time diagram. However, the amount of information that has to be extracted from the local history in a Rule-184 CA to solve L_p is indeed less compared to Rule 240. This is due to the fact that the number of black cells can be extracted from the local history of a Rule-184 CA by counting the number of cells that disturb the checkerboard pattern (comparable to Sec. 4.3). By knowing the number of black cells, it is then possible to determine whether this number of black cells is prime or not. In comparison, in a Rule 240 CA, all cells have to be investigated to obtain the number of black cells.

This analysis shows again that CAs that achieve a DOTO of 1 and maximize the DOSS minimize the amount of superfluous information distributed within the CA.

5 CONCLUSION AND FUTURE WORK

In this paper, we propose formal measures of *target orientation* and *self-structuring* in finite one-dimensional two-state cellular automata, which are both based on Shannon's information entropy. The measures allow to formally evaluate a cellular automaton's ability to classify its initial configuration and the degree to which the cellular automaton builds a structure in the local history of its cells for this purpose. Hence, both measures are able to capture behavioral aspects of emergent computation in an observer-independent way. The relationship between the two measures reveals that cellular automata that use a process of self-structuring to solve a CP distribute information among all cells in an economic way. The measures are exemplarily applied to find cellular automaton rules that solve various density classification problems, the even-odd distance, and prime-number classification problem. The results show that both measures indeed characterize emergent computation in one-dimensional cellular automata.

Our future research directions include the application of the presented concepts to other classification problems (e.g., identification of all initial configurations with odd/even number of State-1 cells) and other classes of cellular automata, e.g., with a higher dimension or more than two states. Moreover, we plan to use our proposed measures as fitness functions for genetic algorithms employed to find cellular automata that solve further classification problems. However, this application needs further discussion of the scalability of our approach. Therefore, we are trying to find approximations for our measures that can be applied with less effort, e.g., by inspecting the transi-

tion rules directly and without exhaustive simulation of the cellular automata. These approximations will also be helpful for discussing cellular automata of infinite size. Finally, we aim at applying our findings in the development of new engineering approaches for self-organizing systems.

6 ACKNOWLEDGEMENTS

We wish to thank Prof. Franz-Josef Brandenburg for the discussions and his valuable suggestions. This research is partially supported by the SOCIONICAL project (IP, FP7 Call 3, ICT-2007-3-231288), by the ResumeNet project (STREP, FP7 Call 2, ICT-2007-2-224619), and by the EuroNF Network of Excellence (IST, FP7 Call 1, ICT-2007-1-216366).

REFERENCES

- [1] David Andre, Forrest H. Bennet, III, and John R. Koza. (1996). Discovery by genetic programming of a cellular automata rule that is better than any known rule for the majority classification problem. In J.R. Koza, D.E. Goldberg, D.B. Fogel, and R.L. Riolo, editors, *Proc. of the First Annual Conference on Genetic Programming*, pages 3–11, Cambridge, MA. MIT Press.
- [2] Christopher Auer, Patrick Wüchner, and Hermann de Meer. (December 2009). The degree of global-state awareness in self-organizing systems. In *Proc. 4th International Workshop on Self-Organizing Systems (IWSOS '09)*, Lecture Notes in Computer Science. Springer Verlag, Heidelberg.
- [3] Elwyn R. Berlekamp, John H. Conway, and Richard K. Guy. (1982). *Winning Ways for your Mathematical Plays*, volume 2. Academic Press.
- [4] Nino Boccara. (November 2003). *Modeling Complex Systems*. Graduate Texts in Contemporary Physics. Springer Verlag.
- [5] Nino Boccara and Henryk Fukś. (2002). Number-conserving cellular automaton rules. *Fundamenta Informaticae*, 52(1-3):1–13.
- [6] Nino Boccara, J. Nasser, and Michel Roger. (July 1991). Particlelike structures and their interactions in spatiotemporal patterns generated by one-dimensional deterministic cellular-automaton rules. *Phys. Rev. A*, 44(2):866–875.
- [7] Mathieu S. Capcarrère. (2002). Emergent computation in CA: A matter of visual efficiency? In *Proc. of the Fourth International Conference on Complex Systems*.
- [8] Mathieu S. Capcarrère, Moshe Sipper, and Marco Tomassini. (December 1996). Two-state, $r = 1$ cellular automaton that classifies density. *Physical Review Letters*, 77(24):4969–4971.
- [9] Thomas M. Cover and Joy A. Thomas. (1991). *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA.
- [10] James P. Crutchfield and Melanie Mitchell. (1995). The evolution of emergent computation. In *Proc. of the National Academy of Sciences, USA*, pages 10742–10746.
- [11] Karel Culik, Lyman P. Hurd, and Sheng Yu. (1990). Computation theoretic aspects of cellular automata. *Physica D*, 45(1-3):357–378.

- [12] Henryk Fukś. (March 1997). Solution of the density classification problem with two cellular automata rules. *Physical Review E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 55(3):R2081–R2084.
- [13] James E. Hanson and James P. Crutchfield. (October 1995). Computational mechanics of cellular automata: An example. Working Papers 95-10-095, Santa Fe Institute.
- [14] Richard Holzer and Hermann de Meer. (September 2008). On modeling of self-organizing systems. In *Proc. of Autonomics 2008*, Turin, Italy.
- [15] Richard Holzer and Hermann de Meer. (December 2009). Quantitative modeling of self-organizing properties. In Bernhard Plattner, Thrasyvoulos Spyropoulos, and Karin Anna Hummel, editors, *Proc. of the 4th International Workshop of Self-Organizing Systems (IWSOS '09)*, LNCS, Zurich, Switzerland. Springer Verlag.
- [16] Richard Holzer, Hermann de Meer, and Christian Bettstetter. (December 2008). On autonomy and emergence in self-organizing systems. In *Proc. of IWSOS'08, Vienna, Austria*, LNCS. Springer Verlag, Heidelberg.
- [17] Wim Hordijk, James P. Crutchfield, and Melanie Mitchell. (1996). Embedded-particle computation in evolved cellular automata. In *Physics and Computation 96 (Pre-Proceedings)*, New England Complex Systems Institute, pages 153–158.
- [18] Mark Land and Richard K. Belew. (June 1995). No perfect two-state cellular automata for density classification exists. *Physical Review Letters*, 74(25):5148–5150.
- [19] Moez Mnif and Christian Müller-Schloer. (July 2006). Quantitative emergence. In *Proc. of the 2006 IEEE Mountain Workshop on Adaptive and Learning Systems (SMCals 2006)*, pages 78–84, Piscataway, NJ, USA. IEEE.
- [20] Andrés Moreira, Nino Boccara, and Eric Goles. (2004). On conservative and monotone one-dimensional cellular automata and their particle representation. *Theoretical Computer Science*, 325(2):285–316.
- [21] Gina M. B. Oliveira, José C. Bortot, and Pedro P. B. de Oliveira. (2003). Multiobjective evolutionary search for one-dimensional cellular automata in the density classification task. In *Proc. of ICAL 2003*, pages 202–206, Cambridge, MA, USA. MIT Press.
- [22] Paul Rendell. (2002). Turing universality of the game of life. *Collision-Based Computing*, Andrew Adamatzky (Ed.), pages 513–539.
- [23] Cosma Rohilla Shalizi. (2001). *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata*. PhD thesis, University of Wisconsin.
- [24] Claude Elwood Shannon. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27.
- [25] Stephen Wolfram. (1986). *Cellular Automata and Complexity: Collected Papers*. Westview Press.