Physica D 432 (2022) 133074

Contents lists available at ScienceDirect

Physica D

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

Review

Progress, gaps and obstacles in the classification of cellular automata

Milan Vispoel *, Aisling J. Daly, Jan M. Baetens

KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, Ghent, 9000, Belgium

ARTICLE INFO

Article history:
Received 9 August 2021
Received in revised form 5 October 2021
Accepted 18 October 2021
Available online 17 November 2021
Communicated by N. Fates

Keywords:
Cellular automata
Classification

ABSTRACT

Classification is one of the most important problems in the theory of CA. For instance, assessing how CA design choices impact the generated dynamics requires an overview of the types of dynamics that can occur. Nevertheless, an overview and critical comparison that includes works from the last five years is currently lacking. This paper provides such a structured overview, where firstly classifications based on the CA's rule table are considered, and secondly classifications based on the space–time pattern generated by a CA.

The review indicates that most currently available classification schemes are limited to elementary cellular automata and highlights the existing dichotomy in CA research: theoretical research focusing on analytical results from topological dynamics and the theory of computation on the one hand, experimental research focusing on the statistical properties of the simulated space–time patterns on the other hand.

Contents

* Corresponding author.
E-mail address: milan.vispoel@ugent.be (M. Vispoel).

https://doi.org/10.1016/j.physd.2021.133074

1. Introduction

Classification is one of the most important problems in the theory of cellular automata (CA) [1]. It entails the study of the features, similarities and differences between the behavior of different CA and how they can be grouped ultimately into classes. Such classifications play an important role by allowing one to study and make generalizations about discrete, homogeneous groups of objects [2]. Without consensus on the classification of objects within a certain field of research, knowledge accumulation and meta-analysis are impeded [3]. For instance, assessing how CA design choices impact the generated dynamics requires an overview of the types of dynamics that can occur. Furthermore, a good classification facilitates the search for 'interesting' rules in a CA rule space, which is an important focus in CA theory [4,5].

Yet, a review of the most important CA classifications and recent advances therein is lacking. This review aims to give such an overview. The most important classification schemes as well as their strengths, weaknesses and connections will be critically reviewed and mutually compared.

2. Preliminaries

2.1. Definition

An N-dimensional k-state cellular automaton consists of an infinite N-dimensional regular lattice with discrete state variables at each site, which can take on k possible values [6]. The site states are updated synchronously in discrete time steps using the states of the neighborhood sites at the previous time step. The CA rule is the dynamical system's evolution rule and implements the local dynamics by defining the way in which each possible neighborhood configuration is mapped to any of the k possible values a site can take.

Below we give a more formal definition that will be used throughout this paper.

Definition 2.1.1 (*Cellular Automaton*)**.** A cellular automaton $\mathcal{C}$ can be represented as a sextuple

$$\mathcal{C} = \langle \mathcal{T}, S, s, s_0, N, \Phi \rangle ,$$

where

(i) $\mathcal{T}$ is a countably infinite regular tessellation of an n-dimensional Euclidean space $\mathbb{R}^n$, consisting of sites c_i, $i \in \mathbb{Z}$.
(ii) S is a finite set of k states, often $S \subset \mathbb{Z}$.
(iii) The output function $s : \mathcal{T} \times \mathbb{Z} \to S$ yields the state value of site c_i at the tth discrete time step, i.e. $s(c_i, t)$.
(iv) The function $s_0 : \mathcal{T} \to S$ assigns to every site c_i an initial state, i.e. $s(c_i, 0) = s_0(c_i)$.
(v) The neighborhood function $N : \mathcal{T} \to \bigcup_{p=1}^{\infty} \mathcal{T}^p$ maps every site c_i to a finite sequence $N(c_i) = \left(c_{i_j}\right)_{j=1}^{|N|}$, consisting of distinct cells. $|N|$ denotes the size of the neighborhood and is independent of c_i.
(vi) The function $\phi : S^{|N|} \to S$ governs the dynamics of site c_i, i.e.

$$s(c_i, t+1) = \phi\big(\tilde{s}(N(c_i), t)\big),$$

where $\tilde{s}(N(c_i), t) = \left(s(c_{i_j}, t)\right)_{j=1}^{|N(c_i)|}$. ϕ is often called the *local update rule*.

The space of all possible configurations of a D-dimensional CA is denoted by $S^{\mathbb{Z}^D}$. The local update rule ϕ induces a global transition function $\Phi : S^{\mathbb{Z}^D} \to S^{\mathbb{Z}^D}$ on $S^{\mathbb{Z}^D}$.

Each of the definition's components are now discussed in more detail.

2.2. Neighborhood and boundary conditions

The neighborhood function N constructs a subset $\mathcal{N}_i \subset \mathcal{T}$ that contains those $c_j \in \mathcal{T}$ that are considered c_i's neighbors. Common choices in the two-dimensional case are the von Neumann neighborhood and the Moore neighborhood. The former contains c_i and its immediately adjacent neighbors, whereas the latter contains c_i and all of its immediately and diagonally adjacent neighbors. More extend neighborhoods, possibly shaped differently, can be defined as well [7–9].

In theory, the CA tessellation is infinite. When simulation the CA, however, the tessellation is necessarily finite. Denoting the number of sites in the finite tessellation of a D-dimensional k-state CA as $|\mathcal{T}|$, the number of possible CA configurations is also finite and equal to $k^{|\mathcal{T}|}$. Throughout this paper, periodic boundary conditions will be used as those tend to have a limited effect on the behavior of the CA [10].

2.3. The state space S

The set $S = \{S_1, S_2, \ldots, S_k\}$ contains all k possible states that a site can take. Often addition modulo-k is used on S, in which case the elements of S are often labeled by elements from the integers up to $k-1$: $S \equiv \mathbb{Z}_{k-1} = \{0, 1, \ldots, k-1\}$.

A state $q \in S$ is called quiescent whenever a neighborhood consisting solely of q-values is mapped to q, i.e. $\phi(q, q, \ldots, q) = q$, where ϕ is the local update rule.

2.4. Local update rule

The transition function is often called the CA "update rule". Given the number of states k and the size of the neighborhood $|N|$, the set of all possible CA rules is denoted by $\mathcal{D}^k_{|N|}$. It has cardinality $k^{k^{|N|}}$, since the number of possible neighborhood configurations is $k^{|N|}$. A CA rule is often specified by its rule table, obtained by listing all possible neighborhood configurations along with their corresponding update values. For one-dimensional k-state CA, the neighborhoods can be considered as $|N|$-digit base-k numbers. Usually these neighborhoods are listed in descending order of their base-k value. $\mathcal{D}^k_{|N|}$ denotes the rule space of a one-dimensional CA unless stated otherwise.

A well-studied CA family are elementary cellular automata (ECA), for which $k = 2$ and $|N| = 3$. For ECA, the rule space is $\mathcal{D}^2_3$ with cardinality $2^{2^3} = 256$. Any rule in $\mathcal{D}^2_3$ is specified by an eight-bit string. Typically, Wolfram's numbering system is used [11]. Given their importance in CA theory, we list below the most important types of CA rules.

2.4.1. Totalistic CA

In totalistic CA rules, the state at site i only depends on the sum (modulo k) of all the sites in its neighborhood, so

$$s(c_i, t+1) = \phi_i\left(\tilde{s}(N(c_i), t)\right) = \Omega\left(\sigma_i\right), \tag{1}$$

where $\sigma_i = \sum_{j=1}^{|N|} s(c_{i_j}, t)$. Alternatively, if the site value depends on both the sum of all states in its neighborhood and its own state (e.g. Conway's Game of Life [12]), the CA rule is called outer-totalistic [11].

2.4.2. Legal CA rules

Usually, the number of valid CA rules decreases significantly by placing some restrictions on the possible rules. CA rules satisfying these restrictions are called legal rules. As a first possible restriction, only rules that have a quiescent state are considered legal. When $S \subset \mathbb{Z}$, the quiescent state is customarily chosen to be 0. Quiescence gives rise to the useful concept of so-called 'finite configurations'. A CA configuration C is finite when all but a finite number of sites in C are quiescent.

As a second restriction, only rules that are symmetric under reflection and rotations are considered legal. This means that neighborhood configurations that are equivalent under rotation and reflection of the CA lattice will map to the same state. Of the 256 ECA, only 32 fulfill the quiescence and symmetry conditions. This review will deal mainly, but not exclusively, with legal CA rules since the dynamics generated by illegal rules is generally not physically meaningful.

2.4.3. Simple, complex and additive CA rules

When examining the evolution of a CA from an initial configuration that contains a single non-zero site (a single seed), different behavioral classes can be defined [11]. Certain CA maintain this initial non-zero value forever without change, while other CA cause the initial value to propagate uniformly to nearby sites. CA rules that exhibit these kinds of behavior are called simple. On the other hand, complex CA rules generate non-trivial self-similar patterns when evolved from a single seed.

Additionally, some CA rules have the property of additive superposition, which means that

$$s(c_i, 0) = a(c_i, 0) \oplus b(c_i, 0) \Longleftrightarrow s(c_i, t) = a(c_i, t) \oplus b(c_i, t), \tag{2}$$

for all c_i and $t > 0$. This superposition differs from the linearity that is found in some continuous systems, since the addition $\oplus$ is not the regular addition defined on $\mathbb{R}$, but addition modulo k. CA satisfying this additive property are often easier to analyze [11]. For binary rules, additivity can be proven on the basis of their rule table entries [13]. In this way, it can be shown that $k^{|N|}$ rules of the $k^{k^{|N|}}$ possible rules are additive. In the case of ECA, this yields only eight additive rules: the simple rules 0, 170, 204 and 240 and the complex rules 60, 102, 90 and 150. All of the latter are dynamically equivalent to either rule 90 or rule 150.

2.4.4. Equivalent CA rules

Since the states in S are merely labels, relabeling S does not yield a dynamically different CA rule. Similarly, rules that can be transformed into each other under reflection or rotation of the CA lattice are also considered equivalent. When considering all possible relabelings and reflections (and combinations of both) the number of non-equivalent ECA rules drops to 88 of the original 256.

In addition, one can define other transformations on a CA rule space that preserve certain dynamical or language theoretical properties. An overview of such transformations and the equivalence relations they induce can be found in [14].

2.4.5. Reversible CA rules

In general, CA time evolution is irreversible: a configuration can have multiple predecessors. Yet, it is possible to construct reversible rules [15] and the degree of irreversibility depends on the CA family.

The number of predecessors of a certain configuration is called the *in-degree* of the configuration. Configurations that have no predecessor are called *Garden of Eden configurations*. They can only occur as initial configurations. The density of Garden of Eden states in the space of all possible CA configurations is called the *G-density* [4]. It is therefore a useful measure for the CA irreversibility.

2.5. Review scope

Extensive simulation led Wolfram to propose the first CA classification in 1984, based on a visual inspection of the evolved space–time patterns [16]. Wolfram's classification is qualitative in nature and lacks a formal definition. This prompted Culik and Yu to propose a formal definition of Wolfram's classes and show that such a classification is in fact undecidable [17]. This result may be generalized to any classification based on the evolved space–time patterns, as the long-term behavior of a CA is generally undecidable.

Section 3 describes Wolfram's classification in detail because it was at the basis of many subsequent works. The properties of each class are described as well as the most important extensions and variations of this classification that will be used throughout this review. Wolfram's classification is rather qualitative in nature. In order to quantitatively characterize a CA rule, a summarizing parameter is needed. We therefore review in subsequent sections other classifications and assess their merits and shortcomings.

A good CA classification divides a rule space into partitions with meaningfully related properties. In general, there are two types of CA classifications. The first type is based on the CA's rule table. Such a rule-table classification is often called *genotypic*. This is the topic of Section 4.

The second type is based an inspection of the space–time patterns the CA generates, often called a *phenotypic* classification. In this case, there are again two options. A space–time classification based on an inspection of the patterns generated from a single initial configuration is called a local classification. Local parameters are discussed in Section 5. Local parameters generally depend on the initial configuration, and so they are not uniquely determined for a certain CA rule. Often they are be computed as a function of a certain parameter that characterizes the statistical properties of the initial configuration [18].

On the other hand, space–time classifications that are based on an inspection of the CA's behavior over all possible initial configurations, are called global classifications. Global parameters are discussed in Section 6.

The distinction between rule-table and space–time classifications is meaningful, as CA with very similar rule tables may behave quite differently, while CA with very different rule tables may behave (practically) identically [19].

Throughout this review, it is kept in mind that since the long-term behavior of CA is undecidable, any CA classification is ultimately undecidable. This undecidability implies that no single parameter can be used to discriminate between all classes. Furthermore, when deciding which class certain rules belong to based on a parameter, one must choose a set of thresholds. This makes the classification dependent on the rather subjective choice of thresholds. It is therefore important to consider the relative advantages and disadvantages of different classifications for various purposes.

Similar reviews of CA classification schemes have been conducted in the past, but so far these have been less comprehensive as they generally focus on a theoretical description of their global properties [20,21]. Other reviews attempt to be more comprehensive, but limit themselves to providing an overview of how the space of ECA rules is subdivided by the different classifications, without providing a motivation for them or subjecting them to a critical comparison [22].

3. CA behavioral classes

3.1. Wolfram's behavioral classes

One of the most important CA classifications is due to Wolfram [16]. It was the first attempt to classify CA according to their behavioral features. It bears significant historical importance and still serves as the basis for most currently employed classifications. Unless stated otherwise, the term 'class' will denote Wolfram's classes in this paper. Wolfram conjectured that all CA can be assigned to one of four behavioral classes. His classification is based on the qualitative features of the space–time patterns. The first three classes have analogues in continuous dynamical systems.

- **W1**: CA evolution leads to a constant homogeneous final state (homogeneous limit points).
- **W2**: CA evolution leads to a constant inhomogeneous final state or periodic structures (inhomogeneous limit points and limit cycles respectively).
- **W3**: CA evolution leads to chaotic structures (strange attractors).
- **W4**: CA evolution leads to long transients and complex structures. This class supports CA rules with exceptional properties such as computational universality. There is no direct analogue to continuous dynamical systems, but there are some similarities with soliton waves and Prigogine's definition of far-from-equilibrium dissipative structures [23–25].

An alternative formulation of Wolfram's classification is based on the growth rate of a pattern generated from a finite initial configuration [26], while yet another formulation of the classification looks at the effects of small changes in the initial configuration on the evolved configurations [26].

3.1.1. Class I

Class I CA rules yield a homogeneous state after a finite number of time steps (Fig. 1a and b). For nearly all initial configurations, the CA evolves to a limit point in the space of all possible CA configurations $S^{\mathbb{Z}}$. This means that these CA are highly irreversible since any information contained in the initial configuration is lost. The fraction of initial configurations that deviate from this approaches zero as the lattice size grows.

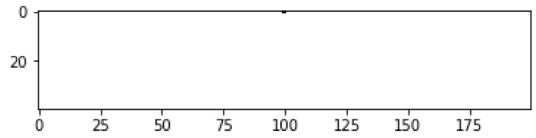
3.1.2. Class II

The space–time diagrams generated by Class II CA rules contain simple stable or periodic structures for nearly all initial configurations (Fig. 1c and d). The information contained in the initial state is not completely lost, since the density of certain sequences in the initial configuration determines the statistical properties of the final configuration. This means that these CA effectively act as information filters. Small (local) changes to the initial configuration have only a local effect. There exist initial configurations that yield unbounded growth, but the fraction of those tends to zero as the lattice size tends to infinity.

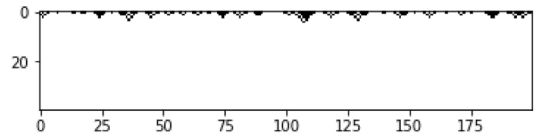
3.1.3. Class III

Class III CA rules yield chaotic, aperiodic structures for nearly all initial configurations (Fig. 1). An equilibrium state of the statistical properties of the CA is reached exponentially with time. In general, these statistical properties are the same for all initial configurations. However, some CA rules result in defects consisting of a group of sites that deviate from the statistical properties of the equilibrium state. The number of such defects decreases with time as $t^{-1/2}$ [16], which implies that the transients for these CA rules will be significantly longer.

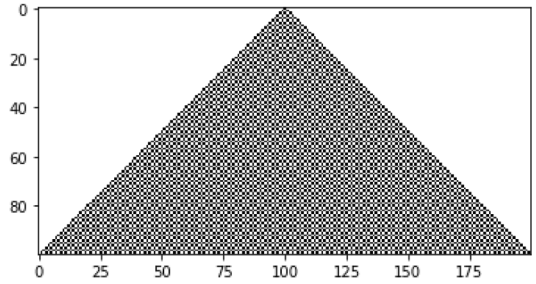
Typical for one-dimensional Class III CA is the appearance of triangular clearings of all sizes in the generated space–time pattern from a disordered configuration, and analogous patterns for higher dimensions (Fig. 1e and f). Evolution from a single non-zero site typically yields self-similar fractal structures in the large time limit. Both behaviors indicate spatial correlations, which signifies the appearance of self-organization during the CA evolution.



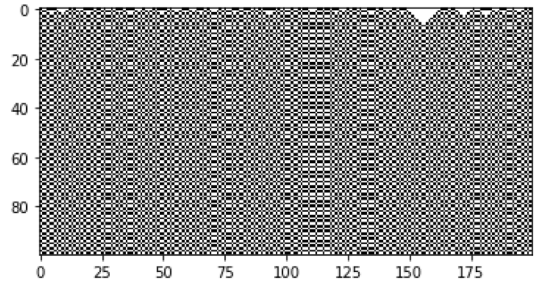
(a) Class I ECA (rule 160) evolving from a single seed.



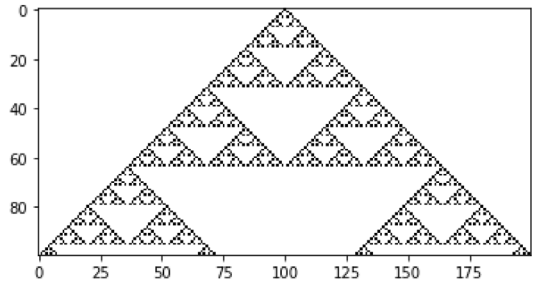
(b) Class III ECA (rule 160) evolving from a random initial configuration.



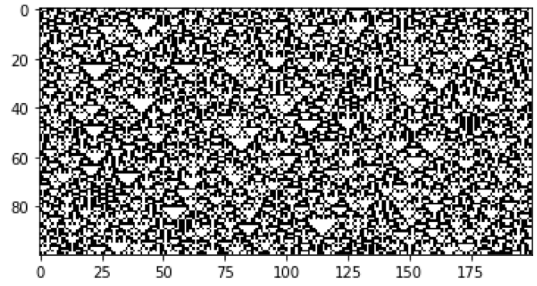
(c) Class II ECA (rule 50) evolving from a single seed.



(d) Class II ECA (rule 50) evolving from a random initial configuration.

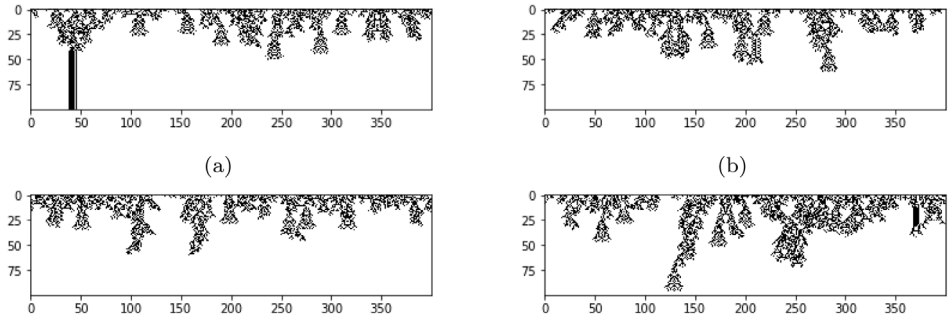


(e) Class III ECA (rule 90) evolving from a single seed.



(f) Class III ECA (rule 90) evolving from a random initial configuration.

Fig. 1. Space–time patterns generated by a Class I, II and III ECA evolving from both a single seed (a, c, e), and a random initial configuration (b, d, f).



(a) (b) (c) (d)

Fig. 2. Four examples of space–time patterns generated by a Class IV CA (totalistic rule 20, $k = 2$, $r = 2$) evolving from 4 different random initial configurations.

3.1.4. Class IV

Class IV CA rules yield particularly interesting space–time patterns for certain initial configurations. Stable and propagating structures can be identified in these space–time patterns (Fig. 2). Class IV is characterized by the emergence of interacting long-lived structures during CA evolution, which is sometimes called the emergence of "glider dynamics" [4].

A CA can be regarded as a purely computational construct [27]. Some Class IV CA rules have the unique property of being computationally universal [28,29], i.e. Turing-complete. Unfortunately, this fact places important limits on the predictions that can be made regarding the behavior of the CA. More precisely, as the halting problem states that it is impossible to decide whether an arbitrary program will run forever or finish running, given its description and input data, the time evolution of a computationally universal CA is undecidable. This means that, given an initial and final pattern, there exists no algorithm that can determine whether the final pattern will ever appear in the time evolution of the initial pattern, when evolving according to a computationally universal CA rule. This indicates that there is a certain level of unpredictability inherent to CA that belong to Class IV, even when only looking at their statistical properties. To further complicate matters, there exists no general purpose algorithm to determine whether an arbitrary CA is computationally universal. Still, for a specific CA, one can attempt to identify structures supported by the CA that are necessary for it to be computationally universal

[28,29]. Some authors have claimed that some CA from Class I, II or III might support universal computation [22].

The behavior of Class IV rules exhibits a mixture of stability (Class I and II) with respect to some local configurations and instability (Class III) with respect to others, which provides some Class IV rules with the necessary requirements to both store and transmit information, yielding computational universality [18]. The fact that Class IV rules lie in a transition region explains why these rules are relatively rare and the set of all Class IV rules has measure zero in $\mathcal{D}^k_{|N|}$.

In the case of ECA, there are four equivalence classes exhibiting Class IV behavior. These are {110, 124, 137, 193}, {54, 147}, {106, 120, 169, 225} and {41, 97, 107, 121}, from which rule 110 has been proven to support universal computation [28]. Whether or not rule 54 and its equivalents support universal computation remains an open problem [30]. More interesting Class IV rules, such as Conway's computationally universal Game of Life, can be found in larger rule spaces [12]. Note that since finite CA have a finite number of possible configurations, they cannot support universal computation since they are only capable of simulating a subset of all possible computable functions.

3.2. Variations of Wolfram's classification

Wolfram's classification was the first attempt at identifying universality classes for CA behavior. The more recent qualitative behavioral CA classifications are variations of Wolfram's classification.

3.2.1. Li–packard classification

With the goal of better distinguishing between locally and globally chaotic behavior, Li and Packard extended Wolfram's classification by splitting Class II into three separate classes [31]:

- **LP1**: CA evolution leads to a homogeneous final state.
- **LP2**: CA evolution leads to a fixed final state that remains invariant after applying the CA rule once.
- **LP3**: CA evolution leads to a two-cycle final state that remains invariant after applying the CA rule twice.
- **LP4**: CA evolution leads to an L-cycle final state that remains invariant after applying the CA rule L times.
- **LP5**: CA evolution leads to chaotic structures.
- **LP6**: CA evolution leads to long transients and complex structures.

A simplified version of this classification by merging LP3 and LP4 and dropping LP6 consists of four behavioral classes [32]:

- **SLP1**: CA evolution leads to a homogeneous final state.
- **SLP2**: CA evolution leads to a fixed final state that remains invariant after applying the CA rule once.
- **SLP3**: CA evolution leads to a periodic final state.
- **SLP4**: CA evolution leads to chaotic structures.

As a more detailed classification, the Li–Packard classification is often used instead of Wolfram's classification in recent works [10]. In addition, the Li–Packard classification is preferred when attempting to classify CA rules using machine learning approaches [33,34].

3.2.2. Order and chaos

A meaningful classification of CA is often obtained by distinguishing CA rules leading to order from those leading to chaos [27]. The former ultimately leads to a constant or periodic final state, corresponding to Class I and II in Wolfram's classification. Chaotic behavior indicates that CA evolution leads to chaotic structures, corresponding to Class III in Wolfram's classification. CA in the transition region between order and chaos induce time evolutions leading to complex interacting structures, corresponding to Class IV in Wolfram's classification.

3.2.3. Culik–Yu classification

Culik and Yu show that in general, it is undecidable which class a CA belongs to, even for Class I and Class II CA [17]. This is partially due to the dependence of the space–time pattern on the initial configuration [35,36]. Furthermore, a full behavioral characterization requires knowledge of all configurations that will occur during the CA's time evolution, which is impossible due to the halting problem [17]. For example, Mazoyer et al. showed that it is undecidable whether a one-dimensional CA will evolve to the same fixed point for every finite initial configuration [37]. Because no single classification scheme correctly predicts the CA behavior for all initial configurations, any formalization of Wolfram's classes will classify a CA by its most complicated behavior [38].

The classification by Culik and Yu is defined hierarchically since each class contains the previous one as a subset [17]. Often, Decidability properties of CA time evolution are employed by formalizations of Wolfram's classification, such as the one of Culik–Yu, as follows

- **CY1**: CA evolution ultimately leads to a homogeneous final state.
- **CY2**: CA evolution ultimately leads to a periodic final state.
- **CY3**: CA for which it is decidable whether any configuration x will ever evolve to the configuration y, for all possible configurations x and y.
- **CY4**: All CA.

Wolfram's classification is useful for identifying universality classes in CA behavior and it will take a central role throughout this paper, yet the undecidability associated with the Culik–Yu classification indicates that it should be kept in mind that there is a significant amount of fuzziness associated with any kind of CA classification scheme.

4. Rule table classification

In this section some of the most important parameters that can be derived from the CA rule table will be discussed. This table is of crucial importance for analyzing CA behavior, since it describes the CA dynamics [39,40]. In this way, the average CA behavior over all possible initial configurations can be predicted, which is often called *dynamics forecasting*. As mentioned in Section 1, predicting the long-term behavior of CA is undecidable [17]. This means that forecasting the dynamics of a CA for an arbitrary initial configuration can only ever be done approximately. Conversely, CA with certain behavioral characteristics can be obtained by constructing CA rules whose rule table parameters fall within a certain range. Rule table parameters also have the advantage of being low effort computations.

4.1. $\mathcal{D}^k_{|N|}$ As a hypercubic space

Rule table parameters are useful because they impose a structure on $\mathcal{D}^k_{|N|}$, the set of all possible rules. These should define a metric that partitions $\mathcal{D}^k_{|N|}$ into regions characterized by a similar dynamics. Before moving on to such parameters, first a different way to impose a structure on $\mathcal{D}^k_{|N|}$ is discussed. Any rule in $\mathcal{D}^k_{|N|}$ can be written as a string of length $k^{|N|}$: $(t_{k^{|N|}-1} \dots t_1 t_0)$, with $t_i \in S$. The set of all such strings is a subset of a $k^{|N|}$-dimensional grid. This subset has the same size along each of the $k^{|N|}$ dimensions of the grid, so it can be identified with the $k - k^{|N|}$-dimensional hypercube [32]. This notation means that the hypercube is $k^{|N|}$-dimensional, where each edge is subdivided in $k-1$ equal parts. In the case of binary CA, the subset simply becomes the $k^{|N|}$-dimensional unit hypercube or the $k^{|N|}$-hypercube. Every vertex on the hypercube therefore represents a CA rule. The Hamming distance defines a natural metric on the $k - k^{|N|}$-dimensional

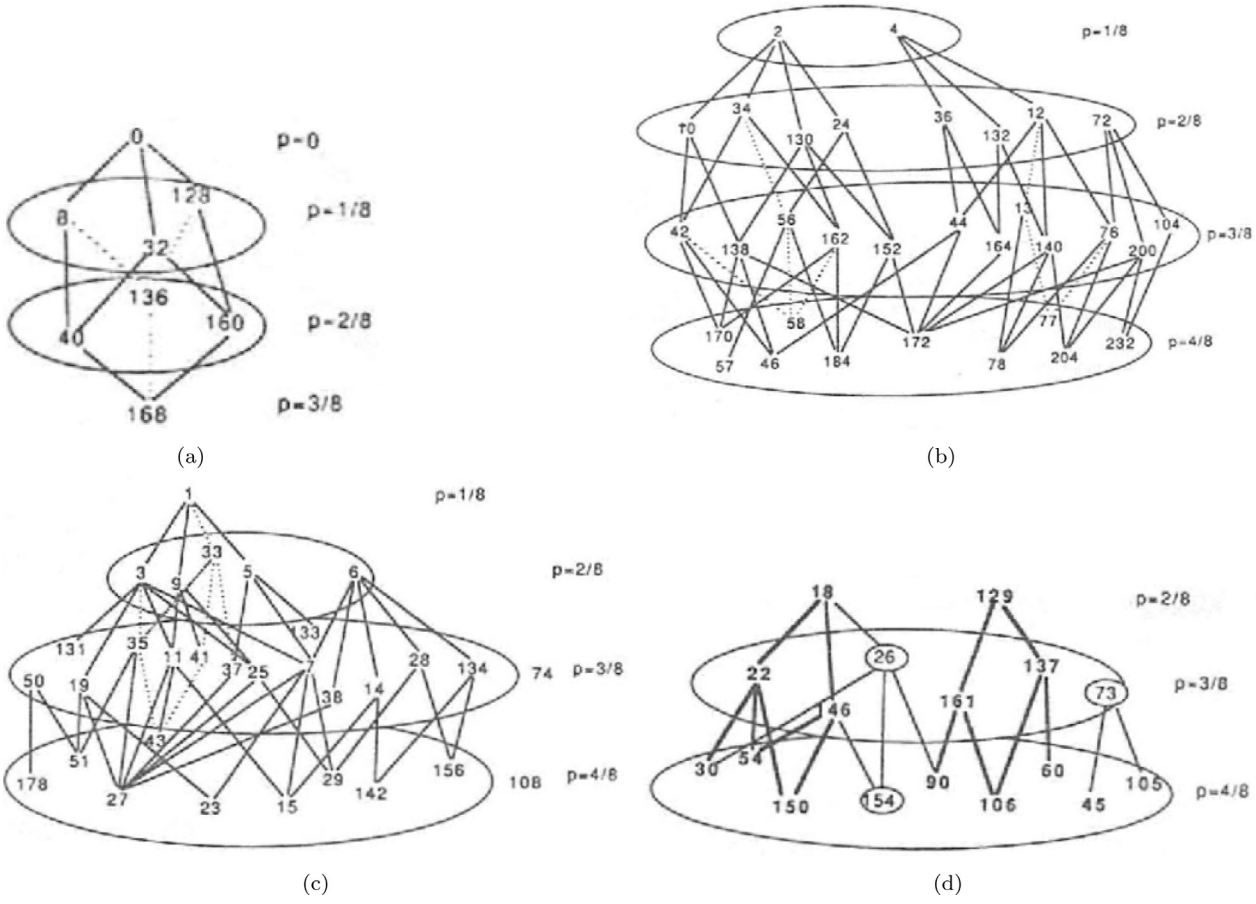


Fig. 3. 88 independent ECA rules along with their connecting edges, grouped by (a) homogeneous limit point rules, (b) inhomogeneous limit point rules, (c) periodic rules and (d) chaotic rules [32].
Source: Courtesy of Wentian Li and Norman Packard.

hypercube. When sticking to rules from different equivalence classes, the hypercubic space reduces to the *folded hypercubic space* [32]. Elements from the lattice point group and elements from the set of possible relabelings of S provide two different ways to fold the original hypercubic space so that equivalent rules are folded onto each other.

As an example, the elementary rule space $\mathcal{D}_3^2$ corresponds to the 2–8-hypercube or the eight-bit unit hypercube. Every ECA rule $(t_7t_6t_5t_4t_3t_2t_1t_0)$ specifies eight coordinates defining a vertex on the hypercube. Fig. 3 shows all 88 independent ECA rules along with their connecting edges. Here the rules are grouped into four classes: homogeneous limit point rules (Class I), inhomogeneous limit point rules (Class II), periodic rules (Class II) and chaotic rules (Class III). Wolfram's Class II is split in two classes since it contains the majority of the rules, while Class IV rules are grouped in whichever class they have the most connecting edges with, since there are only three independent Class IV ECA rules. Within each class, rules are grouped on according to the fraction of neighborhoods mapping to 1, denoted by p.[1]

The homogeneous limit point rules form a three-dimensional cube (Fig. 3a). The inhomogeneous limit point rules are divided in two sets, connected only by rule 172 and rule 44 (Fig. 3b). The periodic rules (Fig. 3c) form a highly interconnected cluster, with the level where $p = 3/8$ is the most populated. Periodic rules 74 and 108 are disconnected from all other periodic rules. The chaotic rules (Fig. 3d) are also divided into two groups, connected only by rule 90. Rule 73 spawns a small third group completely disconnected from the first two.

It is conjectured that the conclusions for $\mathcal{D}_3^2$, can be generalized to arbitrary spaces $\mathcal{D}_{|N|}^k$ [32], but the size of such rule spaces hinders a visual representation. Nevertheless, it is clear that the Hamming distance yields a metric that successfully partitions $\mathcal{D}_3^2$ in different behavioral classes, so an extension of this approach to more general rule spaces may yield interesting results.

4.2. Langton's parameter

One of the most established rule table parameters is Langton's parameter, denoted by λ and defined for a certain quiescent state $q \in S$ by

$$\lambda = \frac{k^{|N|} - n_q}{k^{|N|}}, \tag{3}$$

where n_q is the number of neighborhood configurations that map to q [27]. Clearly, λ lies in the unit interval.

It gives the fraction of neighborhood configurations that do not map to the quiescent state. If $q = 0$, λ measures the activity in the CA rule table. Hence, λ divides $\mathcal{D}_{|N|}^k$ into $k^{|N|} + 1$ subsets:

$$\mathcal{D}_{|N|}^k = \bigcup_{\lambda=0}^{1} P_\lambda, \tag{4}$$

where P_λ is the subset of $\mathcal{D}_{|N|}^k$ containing all rules whose Langton parameter equals λ, which has cardinality $\binom{k^{|N|}}{\lambda k^{|N|}}$.

[1] Note that p is equal to Langton's parameter in the binary case.

When $\lambda = 0$, all neighborhood configurations map to the zero state and there is no activity in the CA. Conversely, when $\lambda = 1 - 1/k$, the activity is maximal and a complex space–time pattern tends to occur. When $\lambda > 1-1/k$, more ordered behavior occurs once more. The $\lambda = 1 - 1/k$ line is called the *symmetry line*. A similar reasoning holds for $1 - 1/k < \lambda \leq 1$. For a fixed k and $|N|$, a systematic traversal of $\mathcal{D}^k_{|N|}$ can be accomplished by varying λ from 0 to 1. Langton uses this method to traverse CA rule spaces with $k \geq 4$ and $|N| \geq 5$ [27]. His results indicate that λ partitions $\mathcal{D}^k_{|N|}$ in subsets that roughly correspond to Wolfram's behavioral classes (Fig. 4) [27].

For values of λ between approximately 0 and 0.20, most neighborhood configurations map to the zero state, implying low activity and a homogeneous final state that is reached after a short transient. This corresponds to Wolfram's Class I. The transient length increases with λ. For values of λ lying roughly between 0.20 and 0.55, increased dynamical activity is possible. This yields periodic structures after a transient, corresponding to Wolfram's Class II behavior. Both the period and transient length increase with λ. When λ approaches 0.50, the transient length increases dramatically. When λ reaches 0.55, the transient behavior effectively becomes the steady state.

For values of λ lying roughly between 0.55 and $1-k^{-|N|}$, most neighborhood configurations will map to a non-zero state, yielding unbounded growth. There is significant dynamical activity that leads to chaotic structures, corresponding to Wolfram's Class III behavior. The length of the transients now decreases with λ to such an extent that when λ reaches $(1 - 1/k) = 0.75$, chaotic behavior is reached after a single time step. When $\lambda > 0.75$, the symmetry line is crossed and more ordered behavior is reached once more. Between 0.75 and 1, the same process as described above occurs in reverse. This indicates that as λ is varied between 0 and $(1 - 1/k)$, a phase transition between ordered (Class I and II) and chaotic (Class III) behavior takes place.

The fact that the transient length diverges and reaches a maximum near the phase transition region often occurs near phase transitions and is called critical slowing down [41]. In this sense, complex structures generated by Class IV rules are an example of a critical phenomenon occurring during a second order phase transition [42]. Essentially, λ can be understood as the degree of non-linearity in the CA, which explains the transitions between Wolfram's classes as bifurcations [18].

It is clear that although λ is a very basic parameter, it reveals a great deal about the structure of the CA rule space. It also provides a scalar index for $\mathcal{D}^k_{|N|}$ that can be used to search for the exceptional Class IV rules. Nevertheless, it should be noted that λ distinguishes ordered from chaotic behavior only for relatively large values of k and N, so ECA are poorly classified by λ [16]. Besides, rules in the same equivalence class can have different λ values upon relabeling the states in S, which is counter-intuitive. Nevertheless, λ bears significant historical importance in the theory of CA. Furthermore, it is often associated with the idea that computation occurs at the edge of chaos and as such, it is used when attempting to design complex systems that are capable of computation [43–46].

4.3. Mean-field parameters

Due to its simplicity, Langton's parameter is a very coarse approximation of the dynamical behavior of CA. Mean-field parameters provide a similar but more detailed way to characterize CA behavior. Again, the mean-field parameters are defined for a certain quiescent state $q \in S$, i.e.

$$M_i = \sum_{\substack{c_1 c_2 \ldots c_{|N|} \text{ having} \\ i \text{ non-}q \text{ values}}} \left[\phi(c_1 c_2 \ldots c_{|N|}) \neq q\right], \tag{5}$$

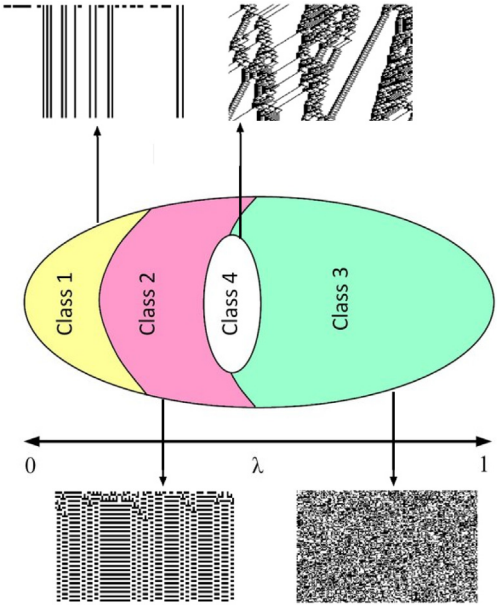


Fig. 4. Langton's parameter λ divides $\mathcal{D}^k_{|N|}$ into subspaces roughly corresponding to Wolfram's behavioral classes [27].

where the summation goes over all neighborhoods that contain i non-q values. From this definition, it is clear that Langton's parameter is the average of the mean-field parameters, that is

$$\lambda = \frac{\sum_{i=0}^{|N|} M_i}{k^{|N|}}. \tag{6}$$

The mean-field parameter M_i gives the number of neighborhood configurations that contain i values differing from q, while at the same time mapping to a non-quiescent state themselves, with $i \in \{0, 1, \ldots, |N|\}$ and $M_i \in \{0, 1, \ldots, \binom{|N|}{i}\}$. The set of mean-field parameters $\{M_0, M_1, \ldots, M_{|N|}\}$ is called the mean-field cluster [32].

The mean-field cluster can take $\prod_{i=0}^{|N|} \binom{|N|}{i}$ possible values. This implies that the mean-field parameters divide $\mathcal{D}^k_{|N|}$ into $\prod_{i=0}^{|N|} \binom{|N|}{i}$ subsets:

$$\mathcal{D}^k_{|N|} = \bigcup_{\{M_0, M_1, \ldots, M_{|N|}\}} P_{M_0, M_1, \ldots, M_{|N|}}, \tag{7}$$

where $P_{M_0, M_1, \ldots, M_{|N|}}$ is the subset of $\mathcal{D}^k_{|N|}$ containing all rules with mean-field cluster $\{M_0, M_1, \ldots, M_{|N|}\}$. This subset has cardinality $\prod_{i=0}^{|N|} \binom{\binom{|N|}{i}}{M_i}$.

In the case of ECA, this yields 64 mean-field clusters. Complementation of an ECA rule with mean-field cluster $\{M_0, M_1, M_2, M_3\}$ yields an equivalent ECA rule with mean-field cluster $\{1-M_3, 3-M_2, 3-M_1, 1-M_0\}$ [32]. These clusters are equivalent since the underlying rules have equivalent dynamics, which reduces the number of independent clusters to 36. CA in the same mean-field cluster will generally demonstrate similar behavior.

When characterizing ECA by their mean-field cluster, the details of which neighborhoods map to 1 are ignored, so that we only retain from the rule table how many neighborhoods containing a certain number of ones will map to 1. Langton's parameter ignores even more details, only retaining how many neighborhoods map to 1, and as such being less informative than the mean-field cluster.

The size of the mean-field cluster scales linearly with $|N| + 1$, while the size of the rule table scales exponentially with $k^{|N|}$. The mean-field cluster provides a multi-dimensional index for $\mathcal{D}^k_{|N|}$, which allows for a similar but more accurate partition of $\mathcal{D}^k_{|N|}$ into

subspaces with behavioral features similar to those discussed in Section 4.2 [32].

Li and Packard used the mean-field parameters to partition the ECA rule space $\mathcal{D}_2^3$ and identify transitions between Wolfram's classes similar to bifurcations in the theory of dynamical systems [32]. Furthermore their partitioning highlighted the importance of the mean-field parameters M_3 and M_0, which correspond to the first and the last entry of the rule table respectively. This explains the peculiar inter-class transitions when changing the first bit, which occur when considering the ECA rule space as a hypercubic space (see Section 4.1).

Applying a similar analysis to more general rule spaces might provide interesting insights, but since the cardinality of $\mathcal{D}_{|N|}^k$ increases exponentially with k and $|N|$, a complete characterization of the entire rule space would become impractical, instead requiring a description of the rule space in terms of some statistical properties. Nevertheless, as a set of uncorrelated parameters, the mean-field parameters are useful, for instance when it comes to training a neural network to forecast the dynamics of a CA based on its rule table [33,34].

4.4. Z-reverse parameter

The Z-reverse parameter was defined by Wuensche [47]. When using an algorithm to sequentially construct the predecessor of a certain configuration, Z is defined as the probability that the next site value of the predecessor that is filled in is deterministic. A low Z value therefore indicates a high number of predecessors for any arbitrary configuration, yielding highly irreversible and ordered behavior. On the other hand, a high Z value indicates a low number of predecessors for any arbitrary configuration, yielding a low degree of irreversibility and chaotic behavior. The Z-parameter is connected to Langton's parameter ($\lambda \geq Z$ [33]) and fulfills a similar role, with both parameters varying between 0 (ordered behavior) and 1 (chaotic behavior). In addition, the Z-reverse parameter has proven useful when investigating CA-based encryption methods since such methods generally depend on reversing chaotic CA rules [48,49].

4.5. μ-sensitivity

The so-called μ-sensitivity is a rule table parameter that has been defined so far only for binary CA as the average number of changes in the output of the state transition function ϕ by changing the state of just one site in the neighborhood [39]:

$$\mu = \frac{1}{|N|\,k^{|N|}} \sum_{c_1 c_2 \ldots c_{|N|}} \sum_{q-1}^{|N|} \frac{\partial \phi}{\partial c_q}. \tag{8}$$

Here $\dfrac{\partial \phi}{\partial c_q}$ is the Boolean derivative of ϕ, given by [50]

$$\frac{\partial \phi}{\partial c_q} = \begin{cases} 1, & \text{if } \phi(c_1 \ldots c_q \ldots c_{|N|}) = \phi(c_1 \ldots \bar{c}_q \ldots c_{|N|}) \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

In this way, μ varies between 0 and 1/2. Similarly to Langton's parameter and the Z-reverse parameter, a low μ value indicates ordered behavior while a high μ value indicates chaotic behavior [39]. This is to be expected since chaos is associated with a sensitive dependence on the initial conditions, which is precisely μ tries to quantify. Yet, as opposed to he Langton parameter, μ does not distinguish between the central bits and the bits near the edges of the rule table, which often play a critical role when it comes to forecasting CA dynamics, as revealed by analyzing $\mathcal{D}_3^2$ using the mean-field parameters (see Section 4.3).

Note that in addition to forming the basis for the definition of μ, the concept of Boolean derivatives on CA provides a useful tool in the dynamical systems approach to studying CA, as it allows for the definition of Lyapunov exponents in a CA context while also facilitating the control of CA [51,52].

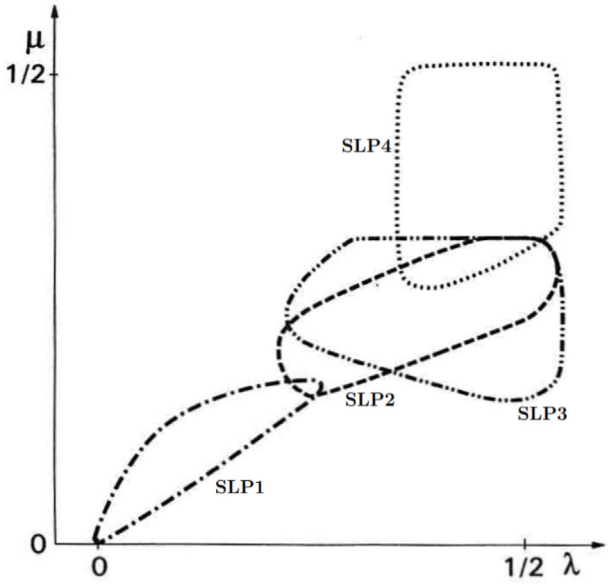


Fig. 5. Partition of $\mathcal{D}_3^2$ by λ and μ [53].

Binder used λ and μ to partition $\mathcal{D}_3^2$ (Fig. 5) [53]. This diagram shows the simplified Li–Packard classes (see Section 3.2.1) in (λ, μ)-space, considering only rules obeying the quiescence condition. Furthermore, since the fraction of exceptional rules such as the identity rule becomes negligible for larger rule spaces, such rules are removed from the diagram as well. The diagram indicates that each class forms a connected and convex set in the (λ, μ)-space [53]. Furthermore μ facilitates the distinction of SLP2 and SLP3 from SLP4, a task where λ sometimes fails.

Finally, it should be noted that λ is likely not the best choice to consider alongside μ for such a diagram, since both parameters seem to encode partially the same information regarding the rule table. Such a correlation between different rule table parameters is important to keep in mind when considering them jointly. This is confirmed by the correlation analysis of Weinert for the most relevant rule table parameters [54].

Finally, it should be noted that μ is defined for binary CA, but an extension to k-state CA can be envisaged and might yield interesting insights regarding the classification of such CA.

4.6. The obstruction parameter

Recall from Section 2.4.3 that a binary CA rule is additive if its rule table satisfies

$$x_i + x_j = x_{i+j}, \tag{10}$$

where x_i denotes the ith entry in the rule table, the addition is modulo-2 and $i+j$ is obtained by converting both i and j to binary numbers, adding them site-wise modulo-2 and converting the result back to decimal form. Bearing this in mind, the obstruction parameter Θ is now defined for binary CA as [55]

$$\Theta = \frac{1}{2^{2|N|}} \sum_{i,j=0}^{2^{|N|}-1} x_i + x_j + x_{i+j}. \tag{11}$$

Hence, it gives the fraction of pairs of neighborhoods on which the CA rule is additive. In this sense it is a measure of the degree of additivity of a CA rule. In the case of ECA, Θ can take on five

different values and therefore divides $\mathcal{D}_3^2$ in five subspaces characterized by Θ-values of 0 (additive rules), 9/32, 5/16, 3/8 and 21/32. Once more Θ is defined solely for binary CA. Furthermore, while Θ characterizes the degree of non-additivity of a CA rule, it distinguishes poorly between the dynamical regimes in $\mathcal{D}_3^2$, as Class I, II and III all contain rules that are additive [39,40].

4.7. Kolmogorov complexity

A central concept in algorithmic information theory is Kolmogorov complexity. It is denoted by $\boldsymbol{K}_\phi(y|x)$ and defined as

$$\boldsymbol{K}_\phi(y|x) = \min\{l(p), \phi\langle p, x\rangle = y\}, \tag{12}$$

where $\boldsymbol{K}_\phi(y|x)$ gives the length of the shortest program p providing y as output when given x as input, and run on a universal Turing machine ϕ [56]. Usually x is the empty string. Furthermore, ϕ is usually dropped from the notation as switching from one Turing-universal programming language to another will alter the Kolmogorov complexity by at most an additive constant independent of y. In this case, the Kolmogorov complexity of y can simply be denoted by $\boldsymbol{K}(y)$. In the context of CA, a rule in $\mathcal{D}_{|N|}^k$ constitutes a string of length $k^{|N|}$ over an alphabet Σ.

The Kolmogorov complexity is an interesting measure that is in many ways similar to entropy. Complexity in the rule table is reflected in the space–time patterns that the rule tends to generate. A string (or CA rule) is Kolmogorov random when the shortest program generating it is at least as long as the string itself, since this indicates a complete lack of patterns or structure in the string. Langton found that the most chaotic CA occur near $\lambda = 1 - 1/k$ [27], while it was indeed shown that random rules occur near (but not precisely at) $\lambda = 1 - 1/k$ [57]. Furthermore, it has been shown that the complexity of the coarse-grained space–time patterns of a CA reaches a maximum when the Kolmogorov complexity of the rule is maximal [58].

In contrast to the other parameters discussed in this section, the Kolmogorov complexity is not computable, but is approximable from the definition above. Since the program-size complexity of a string is the ultimate compressed version of that string, compression algorithms such as Lempel–Ziv based algorithms are used to estimate the Kolmogorov complexity [59]. The length of the resulting compressed string yields an upper bound for the Kolmogorov complexity [60].

Compared to the other parameters, the Kolmogorov complexity has the advantage that it is suitable for arbitrarily large rule spaces, yet very few studies of the Kolmogorov complexity of CA rule tables have been conducted. In addition, the Kolmogorov complexity has proven useful to arrive at certain theoretical results regarding the topological dynamics of CA [61].

4.8. Comparison of rule table parameters

De Oliveira et al. proposed eight desirable properties that a rule table parameter ought to have when used to forecast CA dynamics [39]:

i. Rules from the same equivalence class should have the same parameter value.
ii. No distinction should be made between the states in a parameter definition.
iii. The parameter should define a metric that partitions the rule space into regions characterized by a similar dynamic behavior.
iv. Any parameter definition should be amenable to generalization for any rule space.
v. The focus in the parameter definition should be a measure of the rule's capacity to characterize some type of change. The rule probability to exhibit a specific type of dynamic behavior is associated with its capacity to operate on an initial configuration, taking the CA to a more orderly or more chaotic behavior. This capacity is directly related to the ability (or inability) of the rule to entail some kind of temporal or spatial change perceived in the patterns generated by the CA.
vi. The parameter should have the ability to support a relative discrimination between dynamic regimes.
vii. The parameter should take into account all the cells that compose the neighborhood.
viii. The parameter should have sufficient granularity.

Table 1
Reviewed rule table parameters versus desired properties of rule table parameters according to De Oliveira et al. [39].

Parameter	i	ii	iii	iv	v	vi	vii	viii
λ				x		x		x
mean-field		x	x	x	x		x	x
Z-reverse	x	x	x	x	x		x	
μ	x	x	x	x	x	x	x	x
Θ		x		x			x	
$\boldsymbol{K}$	x	x	x	x	x	x	x	x

Table 1 summarizes the properties satisfied by the rule table parameters discussed in this section. Clearly λ and θ perform the worst in this sense, whereas μ and $\boldsymbol{K}$ perform best. Yet $\boldsymbol{K}$ has the drawback of not being exactly computable. The drawbacks of λ are most evident for low k and N, as mentioned in Section 4.2. Even though λ is historically significant, there exist more informative parameters that facilitate forecasting CA dynamics. The mean-field parameters improve upon λ, as they have the additional properties (ii), (iii), (v) and (vii).

New rule table parameters have been constructed specifically to satisfy all of the above criteria. In this way, De Oliveira et al. defined the neighborhood dominance, activity propagation and absolute activity parameters [40]. Yet, most of them do not provide any additional insights for what concerns the classification of CA, but together with the parameters listed in Table 1 they provide a large set of parameters characterizing a certain CA rule. Together, these can serve as the input vector to a machine learning algorithm designed to automatically classify CA in certain behavioral classes. Such a machine learning-based classification has proven to be one of the most efficient ways to classify ECA based on their rule table [33,34]. However, to this day this approach has not been investigated in the context of more general rule spaces.

5. Local properties

5.1. Space–time parameters and complexity

The parameters discussed in Section 4 often require a large leap of intuition to go from the simple CA rule table to the resulting space–time pattern. In contrast, the approaches reviewed in this and the following section are often more intuitive since they are inferred from the evolved space–time patterns.

Applying a CA rule to an initial configuration results in a specific space–time pattern. This can be characterized by a *local* parameter, so named since it generally depends on the initial configuration. Alternatively, a CA rule can be characterized by a parameter derived from *all* possible space–time patterns that a CA rule can generate. Such a parameter is called a *global* parameter since it characterizes the average CA behavior over all possible

initial configurations, similarly to the rule table parameters from Section 4. Local parameters will be discussed in this section, while global parameters are the focus of Section 6.

Since space–time parameters are derived from observed CA behavior, they are not suitable for forecasting CA dynamics, unlike rule table parameters, but merely give a means to characterize and classify CA. It is interesting to instead consider space–time parameters as a function of rule table parameters, as this allows the identification of bifurcation-like phenomena in the rule space $\mathcal{D}^k_{|N|}$. Indeed, drastic changes in CA behavior can be detected by sudden changes in those space–time parameters. In a certain sense, a bifurcation analysis was already performed in Section 4. There, qualitative changes in CA behavior were investigated by varying a given rule table parameter. In this section, space–time parameters will be used to perform a similar but more quantitative and detailed analysis.

Many of the parameters discussed here can be interpreted as measures of the complexity of a space–time pattern. Complexity is a ubiquitous concept in many areas of science [62], though an exact definition is lacking. Most complexity measures quantify complexity by considering the amount of 'randomness' in a system. Some of the most important complexity measures quantifying randomness are the Rényi entropy, Kolmogorov complexity [56], logical depth [63], effective information content [64], thermodynamic depth [65], effective measure complexity [66] and regular language complexity [67]. A different approach starts by recognizing that neither completely ordered nor completely random systems exhibit an interesting structure. Genuine complexity in the sense of 'interestingness' lies in the transition region between ordered and random systems, as was already mentioned in Section 4. There it was shown that complex Class IV rules generally occur in the phase transition region between Class II and Class III. Some of the most important measures that aim to capture complexity in the sense of 'interestingness' are the effective complexity [68], statistical complexity [69], and predictive information [70].

5.2. Evolution from simple, finite and random seeds

As local parameters depend on the initial configuration, it is necessary to work with standardized initial configurations when working with classifications based on local parameters. Regarding the choice of the initial configuration when computing local parameters, there are three options. The first involves a single non-zero (i.e. non-quiescent) site at the center of the lattice [11], referred to as a *simple seed* and discussed in Section 5.2.1. The second option uses a finite array of non-zero sites at the center of the lattice, referred to as a *finite seed*, discussed in Section 5.2.2. The third option uses random configurations, where the probability that a certain site has value $S_i \in S$ is $1/k$. Such a configuration is called a *random seed*, discussed in Section 5.2.3. Having standardized initial configurations facilitates the comparison of the space–time patterns generated by different CA rules, which in turn permits the classification of CA based on local parameters.

5.2.1. Evolution from a simple seed

In this case, there are $k-1$ options for the possible initial configuration.

Evolution from a simple seed makes it possible to categorize every ECA in one of five classes [71]:

- **Evolution to zero (0)**: After a single time step, a homogeneous zero configuration is reached (Fig. 6a).
- **Finite growth (F)**: The non-zero site evolves into a periodic pattern of finite size. In the case of non-symmetric rules, such a pattern shifts uniformly left or right on the lattice (Fig. 6b).
- **Periodic patterns (P)**: The non-zero site evolves into a periodic pattern of infinite size (Fig. 6c).
- **Sierpiński patterns (S)**: The full space–time pattern constitutes the fractal Sierpiński triangle (Fig. 6d).
- **Complex patterns (C)**: No simple structure is discernible in this case. Instead, evolution of the initial condition leads to the emergence of complex structures that are difficult to summarize (Fig. 6e).

There is some agreement between these classes and Wolfram's classes (Section 3): 0 with Class I, F and P with Class II, S with Class III and C with Class IV. This agreement is not complete since Wolfram's classification is based on the behavior of CA starting from random initial configurations [11].

Letting $h(t) \in S^{\mathbb{Z}}$ denote the configuration at time t that evolved from a simple seed, the *space-filling ratio* can be defined as [71]:

$$\zeta[h(t)] = \frac{\#_1[h(t)]}{2t+1},$$

where $\#_1[h(t)]$ denotes the number of 1's in $h(t)$. The space-filling ratio has a maximal value of 1.

5.2.2. Evolution from a finite seed

The space–time patterns evolved from a finite seed by the same ECA rules from Fig. 6 are shown in Fig. 7. None of the patterns are qualitatively different when compared to the evolution from a simple seed, aside from rule 36. In addition, the pattern exhibits self-similarity for features that are large compared to the length of the finite seed [11].

Braga [72] proposed an ECA classification based on the growth rate of patterns emerging from finite seeds as follows:

- **B1**: The limit configuration of a CA evolving from a finite seed has length zero.
- **B2**: The limit configuration of a CA evolving from a finite seed has a constant length.
- **B3**: The limit configuration of a CA evolving from a finite seed is constantly growing in length.

For ECA, this classification is decidable from the rule table, so the long-term behavior of the ECA can be predicted on this basis.

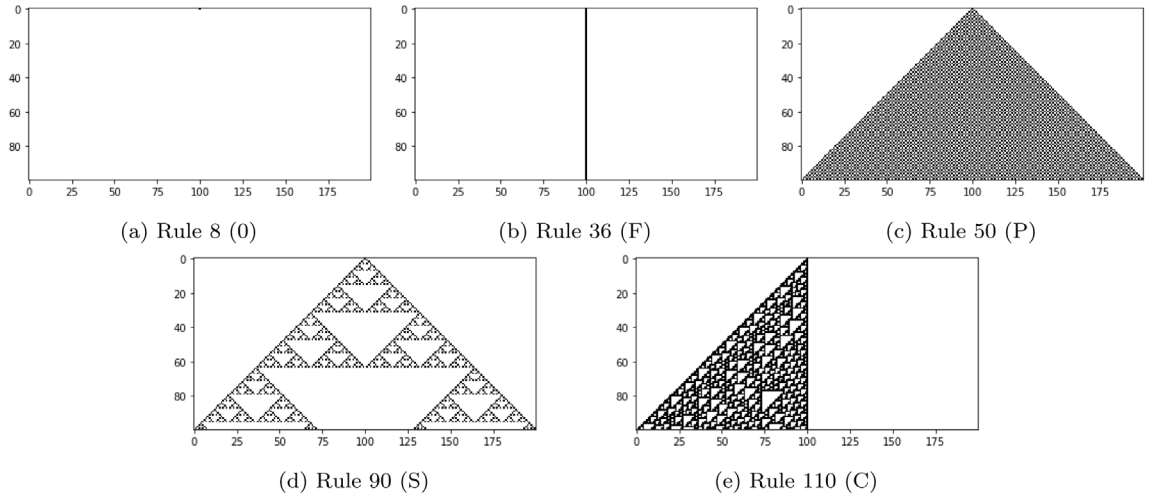
5.2.3. Evolution from a random seed

Fig. 8 shows the space–time patterns evolved from a random seed for each of the five rules whose simple seed pattern was also shown in Fig. 6, together with the random seed space–time pattern of rule 122. Here, it is interesting to look at the spectrum of emerging triangles for Class III rules which obey the following distribution for large n [11]:

$$T(n) = l^{-n}, \tag{13}$$

where $T(n)$ denotes the frequency with which triangles of base length n occur. It turns out that $l = 4/3$ for non-additive Class III rules (e.g. rule 122) and $l = 2$ for additive Class III rules (e.g. rule 90). Furthermore, l is independent from the probability that a certain site will have value 1 in the initial configuration. A universal triangle spectrum of the form l^{-n} indicates that triangles of all sizes occur, but there is no self-similarity.

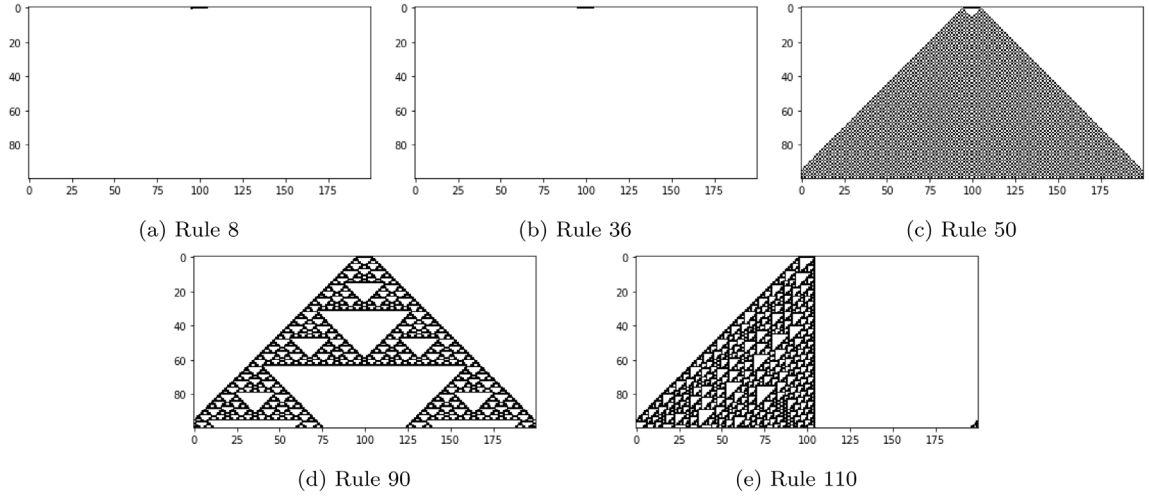
The triangle density $T_{(i)}(n)$ considers the entire space–time pattern. Individual configurations can be characterized by their sequence density $Q_{(i)}(n)$, defined as the density of sequences of length n, all with value i [11]. Whenever a random fluctuation gives rise to a sequence of n sites with value 0, the quiescence condition $000 \to 0$ causes a triangular clearing with base n to appear. This yields the following relationship between the sequence



(a) Rule 8 (0) (b) Rule 36 (F) (c) Rule 50 (P)

(d) Rule 90 (S) (e) Rule 110 (C)

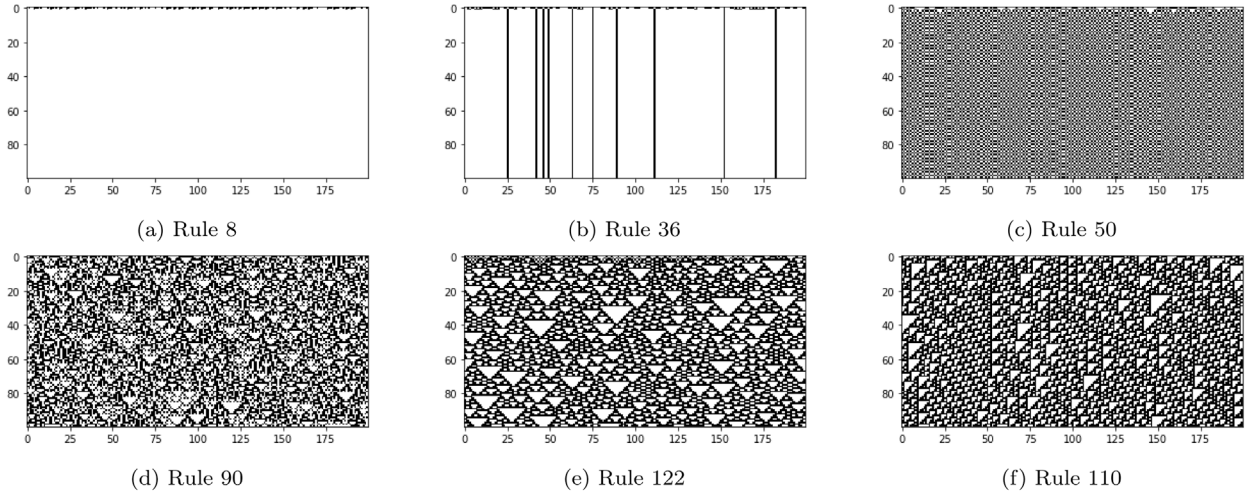
Fig. 6. Space–time patterns for some ECA rules evolved from a simple seed.



(a) Rule 8 (b) Rule 36 (c) Rule 50

(d) Rule 90 (e) Rule 110

Fig. 7. Space–time patterns for some ECA rules generated from a finite seed of length 10.



(a) Rule 8 (b) Rule 36 (c) Rule 50

(d) Rule 90 (e) Rule 122 (f) Rule 110

Fig. 8. Space–time patterns for some ECA rules generated from a random seed.

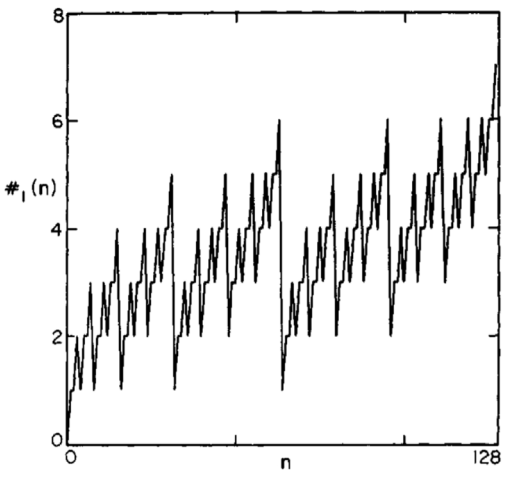


Fig. 9. Graph of $\#_1(n)$ for all ECA n up to rule 128 [11].

density and the triangle density in a space–time pattern [11]:

$$Q_{(0)}(n) = \sum_{j=n}^{\infty} \frac{2\,T_{(0)}(j)}{j}. \tag{14}$$

For large t, $Q_{(0)}(n)$ has a similar exponential form as $T_{(0)}(n)$, which is again largely independent of the density of zeros in the initial configuration.

5.3. Mean-field theory

Mean-field theory provides a way to study CA configurations by drastically reducing the number of degrees of freedom in a certain CA configuration. This is done by taking an average over all degrees of freedom. Given a CA rule $\phi \in \mathcal{D}^k_{|N|}$ and site values $S = \{S_1, S_2, \ldots, S_k\}$, let $p(N, t)$ denote the probability that any three neighboring sites in the CA configuration at time t will constitute the neighborhood N, while $\#_{S_i}(\phi)$ and $\#_{S_i}(N)$ denote the number of S_i-entries in the rule table of ϕ and the neighborhood N respectively. The graph of $\#_1(n)$ for all ECA n up to rule 128 (Fig. 9) shows significant drops when the last couple of digits in the rule table of the ECA are 1 except for the very last digit.

A CA configuration can be characterized by its average density of a certain S_i value, i.e. the probability that a certain site in the configuration will have value S_i. Then, $\rho_{S_i}(t)$ denotes the density of S_i-values in the CA configuration occurring at time t. The probability of observing a certain neighborhood N at time t is

$$p(N, t) = \prod_{i=1}^{k} (\rho_{S_i}(t))^{\#_{S_i}(N)}, \tag{15}$$

which reduces to

$$p(N, t) = (\rho_1(t))^{\#_1(N)}(1 - \rho_1(t))^{\#_0(N)}. \tag{16}$$

In this approach, CA configurations are considered completely random, characterized by their densities $\rho_{S_i}(t)$. However, CA configurations are not random since the CA time evolution gives rise to spatial correlations, which implies that a neighborhood N is more or less likely than expected on the basis of $\rho_{S_i}(t)$. The mean-field approximation neglects these spatial correlations, so that the density of S_i-values at a certain time step can be determined by

$$\rho_{S_i}(t+1) = \sum_{\substack{N\text{'s having} \\ \text{update value } S_i}} p(N, t) \tag{17}$$

$$= \sum_{\substack{N\text{'s having} \\ \text{update value } S_i}} \prod_{i=1}^{k} (\rho_{S_i}(t))^{\#_{S_i}(N)}. \tag{18}$$

Given an initial configuration characterized by $(k-1)$ densities $\rho_{S_i}(0)$, this yields a set of $(k-1)$ recursive equations to calculate the *mean-field approximation*. This approach is similar to how the rule table based mean-field parameters discussed in Section 4.3 characterize each neighborhood by the density of non-quiescent site values, ignoring the specific neighborhood structure.

In the case of ECA, Eq. (18) reduces to:

$$\rho_1(t+1) = \sum_{\substack{N\text{'s having} \\ \text{update value } S_i}} (\rho_1(t))^{\#_1(N)}(1 - \rho_1(t))^{\#_0(N)}. \tag{19}$$

In the equilibrium limit, this yields $\rho_1(\infty) = 0$ for legal Class I rules. Class II rules will often have $\rho_1(\infty)$ proportional to the density of a certain sequence in the initial configuration, indicating they act as a filter [11].

Class III and IV rules are more complicated. In general, these rules yield equilibrium densities that are independent of the initial densities [11]. The equilibrium condition $\rho_{S_i}(t+1) = \rho_{S_i}(t)$ together with Eq. (18) allows to compute this equilibrium density:

$$\rho_1 = \sum_{\substack{N\text{'s having} \\ \text{update value } S_i}} \rho_1^{\#_1(N)}(1 - \rho_1)^{\#_0(N)}. \tag{20}$$

The resulting $(k-1)$ equations in $\{\rho_{S_1}, \rho_{S_2}, \ldots, \rho_{S_{k-1}}\}$ are called the *master equations*. In the case of ECA, a single polynomial equation in ρ_1 is obtained:

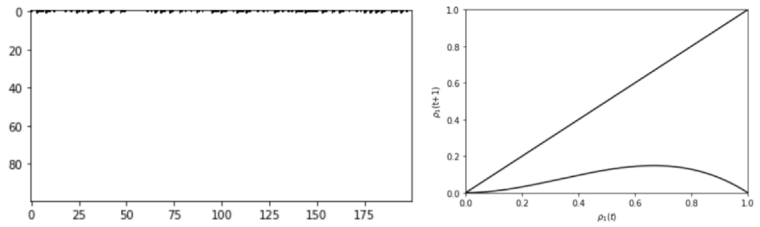
$$\rho_1 = \sum_{\substack{N\text{'s having} \\ \text{update value } S_i}} \rho_1^{\#_1(N)}(1 - \rho_1)^{\#_0(N)}. \tag{21}$$

Fig. 10 shows the space–time patterns of five ECA rules starting from five different random seeds, along with their corresponding mean-field curves. The mean-field curves show the left-hand side (the diagonal) and the right-hand side of Eq. ((21)). The intersection of these lines indicates the occurrence of a fixed point, where we can find the equilibrium density. Since $\partial\rho_1/\partial t = \partial/\partial t\,(\rho_1(t+1) - \rho_1(t))$, the fixed point is stable, marginally stable or unstable when the derivative of the polynomial at the fixed point is smaller than 1, equal to 1 or greater than 1 respectively.

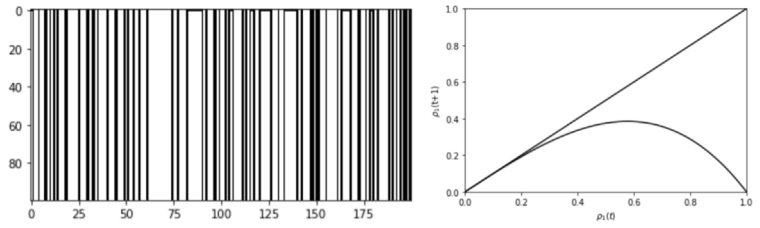
A number of interesting observations can be made from Fig. 10. CA evolving to a homogeneous fixed point have a single stable fixed point at the origin. The derivative of the polynomial at the origin is zero, indicating that any excitation will eventually die out. CA evolving to an inhomogeneous fixed point have a single marginally stable fixed point at the origin. The derivative of the polynomial at the origin is one, indicating an excitation will neither grow nor decay. The mean-field approximation is not suitable for determining the equilibrium density in this case, since this density will generally depend on the initial configuration.

Periodic and chaotic CA have an unstable fixed point at the origin and a stable non-zero fixed point. Between Li–Packard classes LP2 and LP3, a bifurcation has occurred. The fixed point at the origin is now unstable, since a single non-zero site in the initial configuration will multiply in the subsequent time steps. The stable fixed point provides a good estimate of the equilibrium density. ECA rule 110 has an unstable fixed point at the origin and a stable non-zero fixed point. However, complex ECA rules do not illustrate the typical mean-field curve of Class IV rules, due to their simplicity.

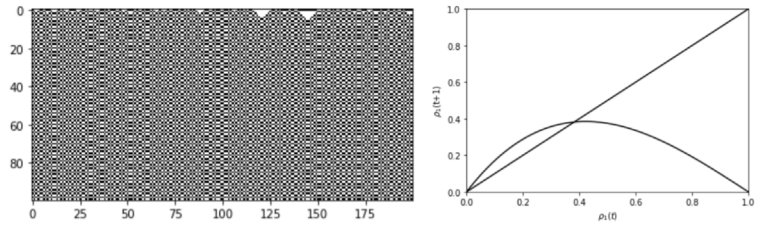
Fig. 11 shows the mean-field curve of the Game of Life, which provides a more representative example for Class IV rules. A stable fixed point occurs at the origin. An unstable fixed point occurs



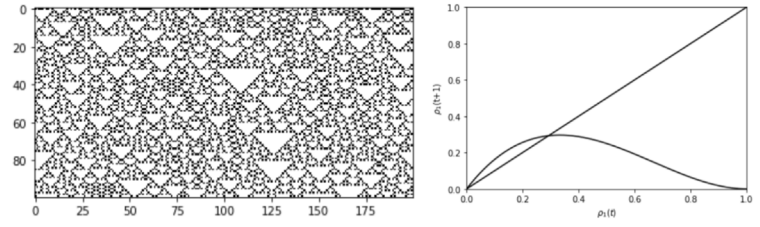
(a) Space-time pattern from a random initial configuration and mean-field curve of Class I ECA rule 8.



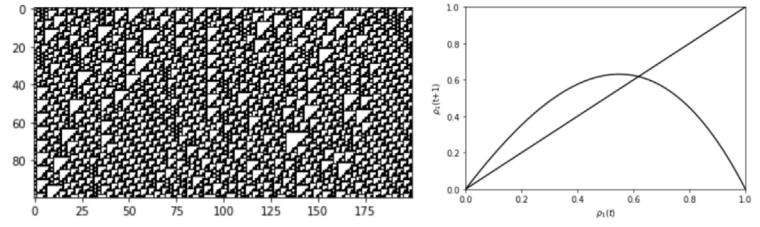
(b) Class II ECA rule 76.



(c) Class II ECA rule 50.



(d) Class III ECA rule 18.



(e) Class IV ECA rule 110.

Fig. 10. Space–time patterns (left) and corresponding mean-field curves (right) for some ECA.

near $\rho_1 = 0.2$, which indicates that complex non-equilibrium behavior tends to occur around this point. Finally, a stable fixed point occurs near $\rho_1 = 0.37$. However, not that in practice, an equilibrium is never observed for the Game of Life.

5.4. Local structure theory

Gutowitz introduced the local structure theory of CA, which is a generalization of the mean-field theory approach presented in Section 5.3 [19,73]. It has been used to study spatially explicit models in sociology [74], evolutionary game theory [75,76], asynchronous CA theory [77], and ecology [78].

Local structure theory of order n describes a CA configuration by specifying the density (or probability) of each of the k^n possible blocks of length n in the configuration. The first order approximation involving $\rho_0(t)$ and $\rho_1(t)$ is equivalent to the mean-field theory approximation. The second-order approximation involves $\rho_{00}(t)$, $\rho_{01}(t)$, $\rho_{10}(t)$ and $\rho_{11}(t)$ and so on. Each CA configuration is specified by a total of $k^n - 1$ independent densities/probabilities. The probability that a certain neighborhood occurs and the update equation for densities are given by Eq. (15) and (18).

For the second-order approximation, the probability a certain block with size $|N| + 1$ occurs is:

$$p(N = S_1 \dots S_{|N|+1}, t) = \frac{\prod_{i=1}^{|N|} \rho_{S_i S_{i+1}}(t)}{\prod_{i=1}^{|N|} \rho_{S_i}(t)}, \tag{22}$$

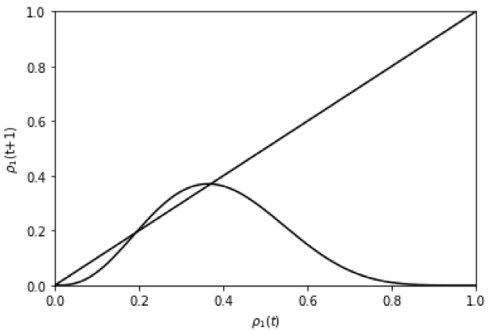


Fig. 11. Mean-field curve of the Game of Life [12].

where a single block probability in the second-order approximation is given by $\rho_{S_i}(t) = \sum_{S_j} \rho_{S_iS_j}(t)$. The update equations for the densities now become

$$\rho_{S_iS_j}(t+1) = \sum_{\substack{N\text{'s having} \\ \text{update value } S_iS_j}} p(N,t). \tag{23}$$

The mean-field approximation ignores spatial correlations. Since self-organization is an essential feature of CA, this is a significant limitation. The second-order approximation is a first step towards taking these spatial correlations into account. Besides, higher-order approximations account for correlations over a longer distance with an increasing accuracy [73]. If the differences between two CA can only be resolved on the basis of spatial correlations beyond a distance of n sites, then these CA are considered to be in the same class of order n [19].

5.5. Two-point correlation function

Local structure theory refines mean-field theory by accounting for spatial correlations. The two-point correlation function, which is complementary to the mean-field approximation provides a simple measure of these correlations [11]:

$$C^{(2)}(r,t) = \langle s(c_i,t)s(c_{i+r},t)\rangle - \langle s(c_i,t)\rangle\langle s(c_{i+r},t)\rangle, \tag{24}$$

where the average is taken over all lattice sites. In the case of binary CA, a slightly different definition is often used:

$$C^{(2)}(r,t) = \langle s_{(b)}(c_i,t)s_{(b)}(c_{i+r},t)\rangle - \langle s_{(b)}(c_i,t)\rangle\langle s_{(b)}(c_{i+r},t)\rangle, \tag{25}$$

where $s_{(b)}(c_i,t) = 1$ if $s(c_i,t) = 1$, and $s_{(b)}(c_i,t) = -1$ if $s(c_i,t) = 0$.

The emergence of spatial correlations during CA time evolution indicates self-organization, which occurs in irreversible systems when they settle around an attractor. This means that, in general, the two-point correlation function will be maximized for ordered CA since they are highly irreversible. On the other hand, smaller (but non-zero) values occur for chaotic CA because they are only slightly irreversible.

5.6. Difference patterns

It is interesting to investigate how the space–time pattern evolved from a random seed changes when a single site is changed in the initial configuration. Subtracting (modulo-k) the corresponding site values of one space–time pattern from the other yields the difference pattern $\Delta s(c_i,t) = s_0(c_i,t) - s_0^*(c_i,t)$, where $\boldsymbol{s}_0$ denotes the initial configuration, $\boldsymbol{s}_0^*$ the perturbed initial configuration, Δs the difference pattern and subtraction is modulo-k. Equivalently, in vector notation, we have $\Delta\boldsymbol{s}(t) = \boldsymbol{s}(t) - \boldsymbol{s}^*(t)$.

The single site that is perturbed in the initial configuration is referred to as a *defect*. The difference pattern indicates how such a defect propagates. The rate at which this happens is called the difference pattern spreading rate. Due to the local nature of CA dynamics, the difference pattern constitutes a widening 'defect cone', meaning that the maximal difference pattern spreading rate is $|N| - 1$ sites per time step.

For binary one-dimensional CA, the time evolution of the difference pattern can be written recursively using the Boolean Jacobian matrix $\boldsymbol{J}(\boldsymbol{s}(t),t)$ [51]:

$$\Delta\boldsymbol{s}(t+1) = \boldsymbol{J}(\boldsymbol{s}(t),t) * \Delta\boldsymbol{s}(t), \tag{26}$$

where $*$ denotes the usual matrix multiplication but where the regular summation is replaced by summation modulo-2. $\boldsymbol{J}(\boldsymbol{s}(t),t)$ contains the Boolean partial derivatives of the transition function $\phi : S^{|N|} \to S$, i.e.:

$$\left[\boldsymbol{J}(\boldsymbol{s}(t),t)\right]_{ij} = \frac{\partial s(c_i,t+1)}{\partial s(c_j,t)} \tag{27}$$

$$= \begin{cases} 1, & \text{if } \phi\left(\tilde{s}(N(c_i),t)\right) = \phi\left(\tilde{s}_j(N(c_i),t)\right) \\ 0, & \text{otherwise,} \end{cases} \tag{28}$$

where $\tilde{s}_j(N(c_i),t)$ is the set obtained by replacing $s(c_j,t)$ by $\bar{s}(c_j,t)$ in the set $\tilde{s}(N(c_i),t)$.

Whether or not a change at site c_j propagates to site c_i depends on the entire neighborhood $N(c_i)$, which is why $\boldsymbol{J}(\boldsymbol{s}(t),t)$ depends on the configuration $\boldsymbol{s}(t)$. Given the nature of the Boolean derivatives, this approach applies to binary CA only.

Fig. 12 shows the difference patterns of ECA rules from each of Wolfram's classes when a single site value in the center of the random seed is perturbed. The difference pattern spreading rate for Class I and Class II CA is zero, indicating a weak sensitive dependence on the initial configuration. In this case, there is no defect cone. On the contrary, Class III CA often yield a difference pattern spreading rate that is maximal, and hence a strong sensitive dependence on initial conditions, which is typical for chaotic systems [79]. Finally, Class IV CA yield a highly variable difference pattern spreading rate that fluctuates between both extreme values (Fig. 12).

The sum of all elements of the difference pattern at a certain time step yields the Hamming distance $H(t)$ between two configurations that evolved from initial configurations differing only in a single site:

$$H(t) = d_H\left(\boldsymbol{s}(t),\boldsymbol{s}^*(t)\right) \tag{29}$$

$$= \sum_i \Delta s(c_i,t). \tag{30}$$

Due to the local interactions there is a clear upper bound on this distance. For one-dimensional CA it holds that $H(t) \leq (|N|-1)t + 1$ for all t. As expected, Class I and Class II rules are stable to small perturbations while Class III is unstable. Finally, Class IV rules yield highly irregular and unpredictable behavior in response to perturbations, with $H(t)$ being significantly lower than for Class III rules.

5.7. Lyapunov exponents

Lyapunov exponents are an important tool in the theory of continuous dynamical systems [80,81], but as a consequence of their discreteness the extension to CA is not straightforward. Given an n-dimensional continuous dynamical system with evolution equation $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$, two trajectories that are initially separated by an infinitesimally small separation $\boldsymbol{\delta}_0$ will diverge at a rate

$$|\boldsymbol{\delta}(t)| \approx e^{\Lambda t}|\boldsymbol{\delta}_0|, \tag{31}$$

where Λ is the Lyapunov exponent. The rate of separation of two infinitesimally close trajectories generally depends on the

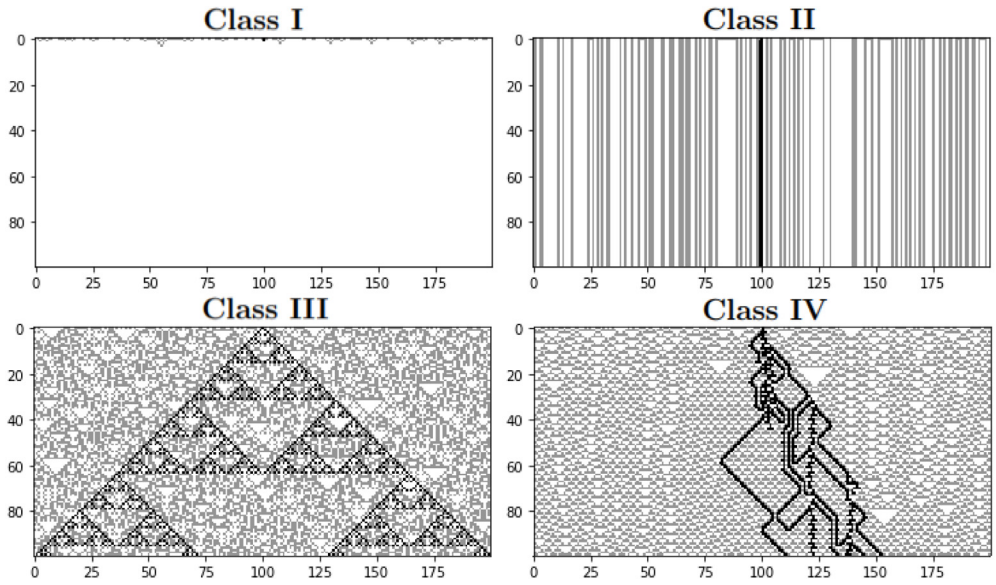


Fig. 12. Difference patterns for ECA rules 32, 76, 90 and 54 respectively, evolving from a random seed perturbed at its center.

orientation of the initial separation vector $\boldsymbol{\delta}_0$. The phase-space is n-dimensional, so there are n different orientations of $\boldsymbol{\delta}_0$, yielding a spectrum of n different Lyapunov exponents $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_n\}$. The largest Lyapunov exponent is called the maximal Lyapunov exponent (MLE). A randomly chosen $\boldsymbol{\delta}_0$ will virtually always have at least some component along the direction associated with the MLE. This means that any two infinitesimally close trajectories will diverge at an exponential rate when the MLE is positive, as the effect of the positive MLE will dominate over the effects of the negative Lyapunov exponents.

5.7.1. Lyapunov exponents of CA

The definition of Lyapunov exponents relies on tools from differential calculus, which are not directly applicable to CA since the state space $S^{\mathbb{Z}}$ is fully discrete. Consequently, two separate viewpoints emerged for what concerns Lyapunov exponents of binary one-dimensional CA.

The first viewpoint was suggested by Wolfram, who defined them empirically by looking at the average propagation velocity to the left and right of the difference pattern [16]. This idea was later formalized for binary one-dimensional CA by Shereshevsky [82], and further refined and extended to higher-dimensional CA by other authors [83–85]. This approach yields the left and right Lyapunov exponents $\Lambda^+(x)$ and $\Lambda^-(x)$.

The second viewpoint is more similar to the familiar notion of Lyapunov exponents as the exponential divergence rate of nearby trajectories. Denoting the Hamming distance between two configurations whose initial configuration differs by a single site by $H(t)$, we already know from Section 5.6 that this distance grows at most linearly with time, it is not a suitable metric to define exponential divergence in S^N.

Bagnoli et al. resolved this issue for binary one-dimensional CA by considering the number of ways in which defects can propagate to a certain time step, instead of simply considering the number of defects at a certain time step [51]. They define the damage vector $\boldsymbol{n}(t)$ as the vector whose ith entry contains the number of ways in which the initial defect can propagate to site c_i in t time steps. The evolution of $\boldsymbol{n}(t)$ then becomes:

$$\boldsymbol{n}(t+1) = \boldsymbol{J}(\boldsymbol{s}(t), t)\boldsymbol{n}(t), \tag{32}$$

where the regular matrix multiplication takes place. It is clear that the total number of ways in which defects can propagate to time step t, given by the sum of all entries in $\boldsymbol{n}(t)$ and denoted by $\epsilon(t)$, can grow exponentially with time.

Considering $\epsilon(t)$ instead of $H(t)$ provides a way to meaningfully define the exponential divergence of trajectories in $S^{\mathbb{Z}}$. This can be used to define Lyapunov exponents in the context of CA in a similar way as for continuous dynamical systems. The sign of such a Lyapunov exponent induces a useful CA classification.

5.7.2. Lyapunov profiles of ECA

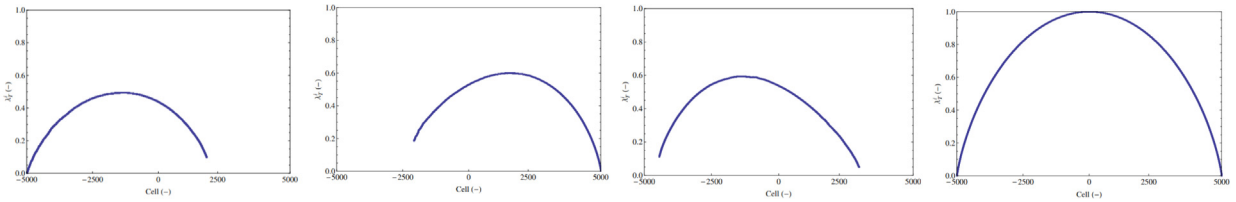
The MLE as defined by Bagnoli et al. [51] discards the information regarding the distribution of defects in $\boldsymbol{n}(t)$ [86]. A more detailed approach considers a one-dimensional finite CA of size N as an N-dimensional dynamical system and introduces a finite-time spectrum of N Lyapunov exponents as follows [86]:

$$\boldsymbol{\Lambda}(T) = \frac{1}{T}\log(\boldsymbol{n}(T)), \tag{33}$$

where the log is applied element-wise. The elements of $\boldsymbol{\Lambda}(T)$ may be understood as the time-averaged exponential rates by which the number of defects grows in the CA sites. Now the MLE can be defined as the largest element of $\boldsymbol{\Lambda}(T)$ [86].

In addition to perturbing a single site in the initial configuration, one can also introduce a topological perturbation by slightly altering the lattice topology. This yields the topological Lyapunov exponents of the CA [87]

The normalized Lyapunov profiles of different ECA rules are shown in Fig. 13. Class III rules 90 and 150 have positive λ-values across the entire spectrum, which indicates that the defect cone widens at a maximum speed. This is not the case for rule 30, which has $\Lambda_i(T) > 0$ for $i \gtrapprox -2300$ and $\Lambda_i(T) = -\infty$ for $i \lessapprox -2300$. This indicates that rule 30 has an asymmetric defect cone propagating to the right at maximum speed but propagating to the left at a lower speed. This yields an MLE that does not occur at the site where the initial defect was introduced. Rule 90 has a smaller MLE than rule 150, indicating that the intensity is lower for rule 90. The normalized Lyapunov profiles of Class IV rules 106 and 110 are asymmetric and show one or two sharp phase transitions. Class II rules generally allow a defect to propagate only within a small region, yielding $\Lambda_i(T) = -\infty$ almost everywhere except in a few sites. However, certain exceptional Class II rules (6, 25 and 38) can give rise to non-trivial profiles [86]. Rule 25 is the only Class II rule having a non-zero profile across the entire spectrum. Other Class II rules yield a profile that is smaller and more narrow than that of Class III and IV rules.



(a) Rule 6 (b) Rule 30 (c) Rule 110 (d) Rule 150

Fig. 13. Normalized Lyapunov profiles of some ECA rules [86].

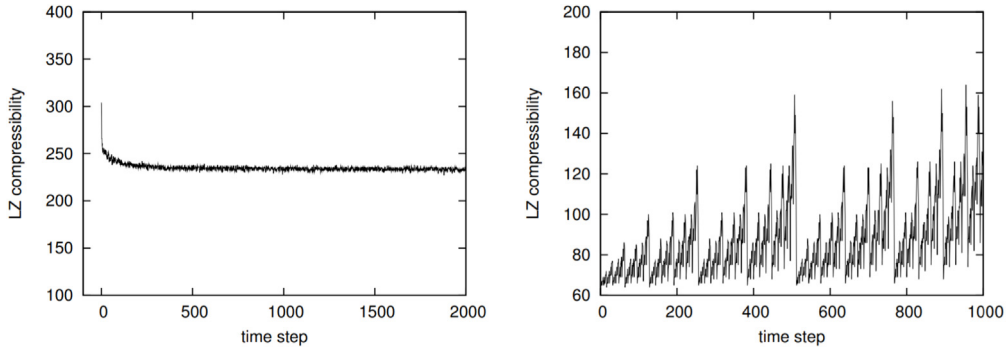


Fig. 14. Compressibility of Class III ECA rule 146 starting from a random seed (left) and from a simple seed (right) [88].

5.8. Kolmogorov complexity

In Section 4.7, we introduced Kolmogorov complexity as a compression-based measure to characterize the rule table, but it can also be used to characterize the complexity of CA configurations. In this setting, a string of infinite length over the alphabet Σ or in the case of CA, a string of length k^N, is compressed. The complexity of a CA configuration $s(\cdot, t) \in S^{\mathbb{Z}}$ is denoted by $\boldsymbol{K}(s(\cdot, t))$. The compressibility ratio $\gamma_{CR}(s(\cdot, t))$ of a configuration c is then defined as $\boldsymbol{K}(s(\cdot, t))/l(s(\cdot, t))$, where $l(s(\cdot, t))$ is the uncompressed length of the configuration [88].

The resulting graphs of γ_{CR} versus time are often very irregular (Fig. 14). For this reason, the Kolmogorov complexity is often estimated for the set of all configurations occurring in a space–time pattern up to time t, denoted by $\{s(\cdot, t)\}$. This is denoted by $\boldsymbol{K}(\{s(\cdot, t)\})$ and is simply the Kolmogorov complexity of the concatenation of all configurations in $\{s(\cdot, t)\}$.

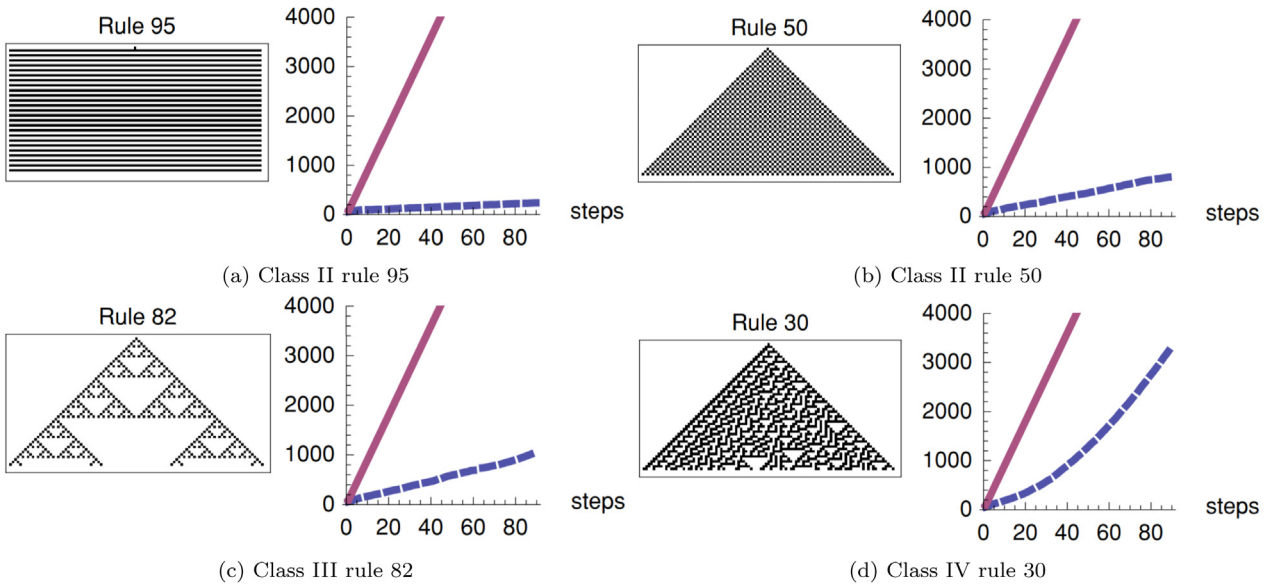
Compressibility has been successfully used to identify turning points in the time evolution of biological systems [89], locate interesting rules in large rule spaces [5], and study artificial life [90]. In addition, Zenil established an ECA classification scheme based on a lossless compression algorithm that is a combination of the LZ algorithm and Huffman coding [91]. The resulting algorithm, independent of CPU type, operating system, file system, and character set, has an efficiency comparable to the Lempel–Ziv–Welch (LZW) algorithm [92]. The LZW algorithm is currently the best general purpose compression algorithm available; it forms the basis for the gzip file format. Zenil's classification considers the Kolmogorov complexity of finite CA of size N evolving from a simple seed.

Fig. 15 shows $\boldsymbol{K}(\{s(., t)\})$ and $l(\{s(., t)\})$ as a function of t for some ECA rules. For Class II rules 95 and 50 and Class III rule 82, the compressed length is far below the uncompressed length and grows linearly with time. For Class III rule 30, the uncompressed length starts to approach the compressed length, indicating the appearance of significant complexity during the CA's time evolution.

Using the above method, Zenil ranked all ECA according to their Kolmogorov complexity $\boldsymbol{K}(\{s(., t)\})$, generated from a simple seed after 200 time steps [91] (Fig. 16). The first group of rules in his classification consists of the 13 ECA with the largest compression length: rules 30, 45, 73, 75, 86, 89, 101, 110, 124, 135, 137, 149 and 193. The second group consists of all remaining rules. The first group can be divided into equivalence classes which, as should be expected, yield roughly the same compression length: Class III equivalence classes {45, 75, 89, 101} and {30, 86, 135, 149} and Class IV equivalence class {110, 124, 137, 193}. Rule 109 has the highest compression length in the second group. The second group can be split into the rules lying below, on, or above the line representing a compression length of 1550. In the simplified Li–Packard classification (Section 3.2.1), such a divide corresponds to SLP1, SLP2 and SLP3/SLP4 respectively. In general, Zenil concluded that the strings of configurations of Class I and II CA are highly compressible while Class III and IV CA yield strings whose compressed lengths asymptotically converge to their uncompressed lengths [93]. Zenil successfully applied the same approach to 3-state nearest neighbor one-dimensional CA [91]. Classifying CA rules by their $\boldsymbol{K}(\{s(., t)\})$ therefore provides a classification that is similar to Wolfram's classification.

It is useful to investigate how the initial configuration affects the complexity of the resulting space–time pattern [91]. Class III rules 22 and 109 are characterized by a sequence of phase transitions when the initial configuration is varied. In contrast, Class III rules 30 and 45 have a consistently high compression length without exhibiting any discernible phase transitions. One way to quantify the sensitivity to the initial configuration uses the average of the absolute value of the finite difference approximation of the derivative of graphs containing the compressed length of the string of configurations versus the initial configuration. This is the so-called *characteristic exponent* c_n^t of CA, where n denotes the number of initial configurations included in the average, and t denotes the time step.

This characteristic exponent fulfills a role similar to the Lyapunov exponents discussed in Section 5.7, except that the divergence of CA trajectories is now measured using a metric based on the compression length.



(a) Class II rule 95 (b) Class II rule 50

(c) Class III rule 82 (d) Class IV rule 30

Fig. 15. Length of the compressed (dotted) and uncompressed (solid) string of configurations as a function of time steps for four ECA rules evolved from a simple seed [91]. The x-axis shows the number of time steps while the y-axis shows the length of the (un)compressed string of configurations.

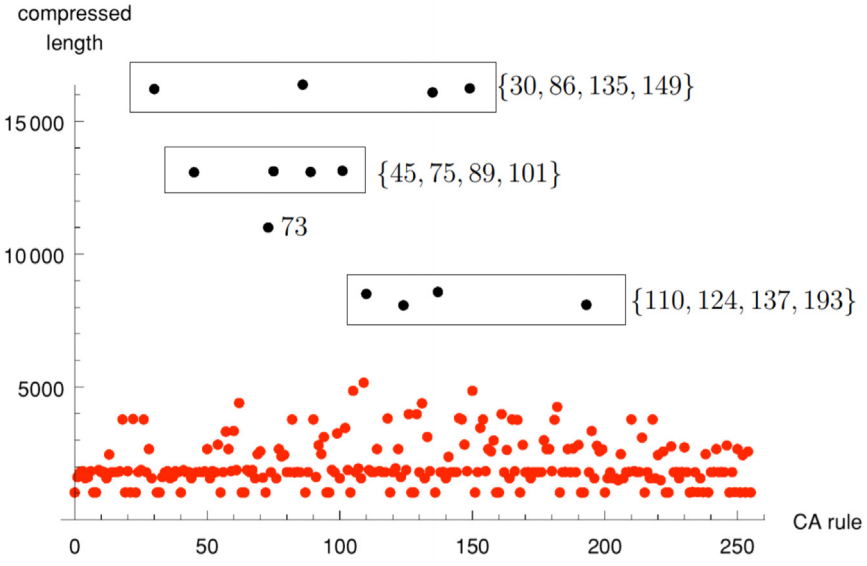


Fig. 16. Partitioning ECA in two groups according to their Kolmogorov complexity $\boldsymbol{K}(\{s(.,t)\})$ [91].

Essentially, the compressibility ratio $\gamma_{CR}(s(\cdot,t))$ can be used to define Wolfram's classification [93]:

- **W1**: $\{s(\cdot,t)\}$ has an asymptotic compressibility ratio equal to 0.
- **W2**: $\{s(\cdot,t)\}$ has a compressibility ratio smaller than 1/2.
- **W3**: $\{s(\cdot,t)\}$ has a compressibility ratio of 1.
- **W4**: $\{s(\cdot,t)\}$ has an asymptotic compressibility ratio equal to 1.

This definition provides a formal but uncomputable way to determine the Wolfram class of a certain CA.

5.9. Spectral properties

Investigating the properties of $s(\cdot,t)$ for $t = 0, 1, \ldots, T-1$ in the frequency domain can reveal interesting CA properties. Since $s(\cdot,t)$ is finite and discrete, the discrete Fourier transform (DFT) can be used to obtain its representation in the frequency domain $\hat{s}(\cdot,f)$:

$$\hat{s}(\cdot,f) = \frac{1}{T}\sum_{t=0}^{T-1} s(\cdot,t)\exp\left(-i\frac{2\pi tf}{T}\right), \tag{34}$$

where $f = 0, 1, \ldots, T-1$. Ninagawa determined the spectral properties of different ECA rules by investigating the power density of a CA, defined as

$$S(f) = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{s}(\cdot,f)\right|^2, \tag{35}$$

where N is the lattice size [94]. The power $S(f)$ at frequency f can be intuitively interpreted as the strength of the vibration with period T/f in the evolution of observation length T. From Fig. 17, which shows the power spectra of different ECA rules, the following conclusions can be drawn.

Class I: The power density is low at all frequencies. The steady-state behavior does not contribute to the power spectrum since it consists of zero configurations. The short transients are virtually random and lead to a low constant power density. If a Class I rule leads to a steady-state of non-zero homogeneous configurations, the power spectrum will exhibit a peak at $f = 0$.

Class II: The short transients again lead to a low constant power density at all frequencies. However, here the steady-state behavior is non-zero and periodic. This causes a peak in the power spectrum at $f = T/P$, with P the period of the steady-state behavior.

Class III: The power density is relatively high and constant at all frequencies. Such a white noise power spectrum indicates a lack of correlations and order in the temporal sequences of the space–time pattern. The space–time pattern is causally determined by a deterministic rule, but the resulting time evolution at a single site is virtually orderless. The white noise power spectrum has its origin in the randomness induced by the Class III rule itself, not the randomness inherent in the initial configuration.

Class IV: For low frequencies, the power spectrum follows a power law. The dashed lines in Figs. 17(h) and 17(i) were obtained by a least square fitting to the power spectrum for low frequencies by $\ln S(f) = \alpha + \beta \ln f$. This indicates that the power spectrum is proportional to f^{β}, with f close to -1. A time series with this type of spectrum is called $1/f$ noise [95]. Since Class IV CA lead to unpredictable space–time patterns which depend on the initial configuration, the spectrum depends on the initial configuration as well. Still, a $1/f$ shape can always be discerned. A $1/f$ power spectrum has also been found in other Class IV CA, most notably in the Game of Life [96].

1/f noise has been observed in an exceptionally diverse number of physical and biological systems [95]. However, a universal theory explaining its ubiquitous occurrence is currently lacking. As mentioned throughout this review, Class IV CA lie in the phase transition region between ordered and chaotic behavior: gliders and other interesting structures that interact in a complex and chaotic fashion are embedded in an ordered background. The space–time pattern consists of alternating periodic phases in which gliders are shifting in a periodic background without colliding, followed by a sudden 'burst' that is caused by colliding gliders [94]. Such a burst causes a streak of temporal correlations at certain sites. This phenomenon is sometimes called 'intermittency' or 'intermittent chaos' [97].

A $1/f$ spectrum indicates that the probability of a temporal correlation with correlation length τ will decrease with τ. Manneville studied the relationship between $1/f$ noise and intermittency in continuous dynamical systems. He concluded that the $1/f$ spectrum suggests that such spectra can originate from dynamics deterministic in the short term though unpredictable in the long term [98]. This is precisely what characterizes Class IV CA. It should also be noted that $1/f$ noise is commonly associated with self-organized criticality [99], which occurs when a system organizes itself around a critical configuration, and which also occurs during the evolution of Class IV CA.

It is clear that the power spectral analysis of CA yields insights that could shed light on the relationship between the computational and dynamical properties of physical systems. The above analysis is limited to ECA. A spectral analysis of larger rule spaces could be useful.

6. Global properties

In Section 5 we discussed the most relevant local parameters. Here, we turn to global parameters instead. Those are derived from the set of all possible CA configurations that a certain CA rule can generate at a certain time step, denoted as $\Omega(t)$. Since it is either impossible (infinite CA) or computationally expensive (finite CA) to cover all possible initial configurations, global parameters will typically be based on theoretical results or extensive simulation, as compared to their local and rule-table counterparts. Doing so, one however gets a more comprehensive understanding of a CA.

6.1. Attractor structure

One way of providing such a comprehensive overview is by describing a CA's attracting set.

6.1.1. Infinite CA

The concept of an *attractor* has also been used to study CA. The notion of an attractor of a global CA rule Φ in the space of all CA configurations $S^{\mathbb{Z}}$ was defined by Hurley using the following objects [100]:

- The *ω-limit set* of $x \in S^{\mathbb{Z}}$ is the set of limit points of the set of evolved configurations $\{\Phi^n(x)\}_{n\in\mathbb{Z}}$ and is denoted by $\omega(x, \Phi)$. The ω-limit set of a subset U of $S^{\mathbb{Z}}$ is the union of the ω-limit sets of all configurations in U and is denoted by $\omega(U, \Phi)$.
- A non-empty closed set $\mathcal{A} \subseteq S^{\mathbb{Z}}$ is an *attractor* of Φ if and only if $\mathcal{A} = \omega(U, \Phi)$ where $\Phi(U) \subseteq U$, i.e. U is a Φ-invariant set.[2]
- A non-empty set $\mathcal{A}_q \subseteq S^{\mathbb{Z}}$ is a *quasi-attractor* if it is a countable intersection of attractors, but is not itself an attractor.
- $\mathcal{A}_m \subseteq S^{\mathbb{Z}}$ is a *minimal attractor* (minimal quasi-attractor) if it does not contain any attractor (quasi-attractor) as a proper subset.
- The *basin of attraction* of an attractor $\mathcal{A}$ is defined as $\mathcal{B} = \{x \in X : \omega(x, \Phi) \subseteq \mathcal{A}\}$. The basin of attraction of $\mathcal{A}$ is the subset of $S^{\mathbb{Z}}$ containing all configurations that will eventually settle on $\mathcal{A}$.

Using these definitions, Kůrka proved that every CA falls into one of the following classes [20]:

- **A1**: There exists a pair of disjoint attractors.
- **A2**: There exists a unique minimal quasi-attractor.
- **A3**: There exists a unique minimal attractor different from $\omega(S^{\mathbb{Z}}, \Phi)$.
- **A4**: There exists a unique attractor $\omega(S^{\mathbb{Z}}, \Phi) \neq S^{\mathbb{Z}}$.
- **A5**: There exists a unique attractor $S^{\mathbb{Z}}$.

Note that both Hurley's definition of an attractor and Kůrka's attractor classification are applicable to any dynamical system, not just CA. The classifications by Kůrka and Wolfram are linked as follows: Class I and II CA belong to A2 ∪ A3, Class III CA belong to A4 ∪ A5 and Class IV CA belong to A1. As with all CA classifications, it is in general undecidable which attractor class it belongs to.

6.1.2. Finite CA

For finite CA of size N, the space of all possible configurations S^N is finite with cardinality k^N. This makes a complete overview of the attractor structure possible. Furthermore, since the number of possible configurations is finite, these CA are ultimately periodic which guarantees that the dynamics will eventually settle on a periodic attractor. Finite CA allow for a complete enumeration of all configurations in S^N along with their interconnections determining the structure of the attractors and the basins of attraction. This allows for a less abstract approach.

[2] If U is not invariant under Φ, then U may contain two configurations that lead to a different attractor. If any two configurations in U are to lead to the same attractor, then one of the configurations must eventually evolve to the other under successive applications of Φ, yielding a Φ-invariant set.

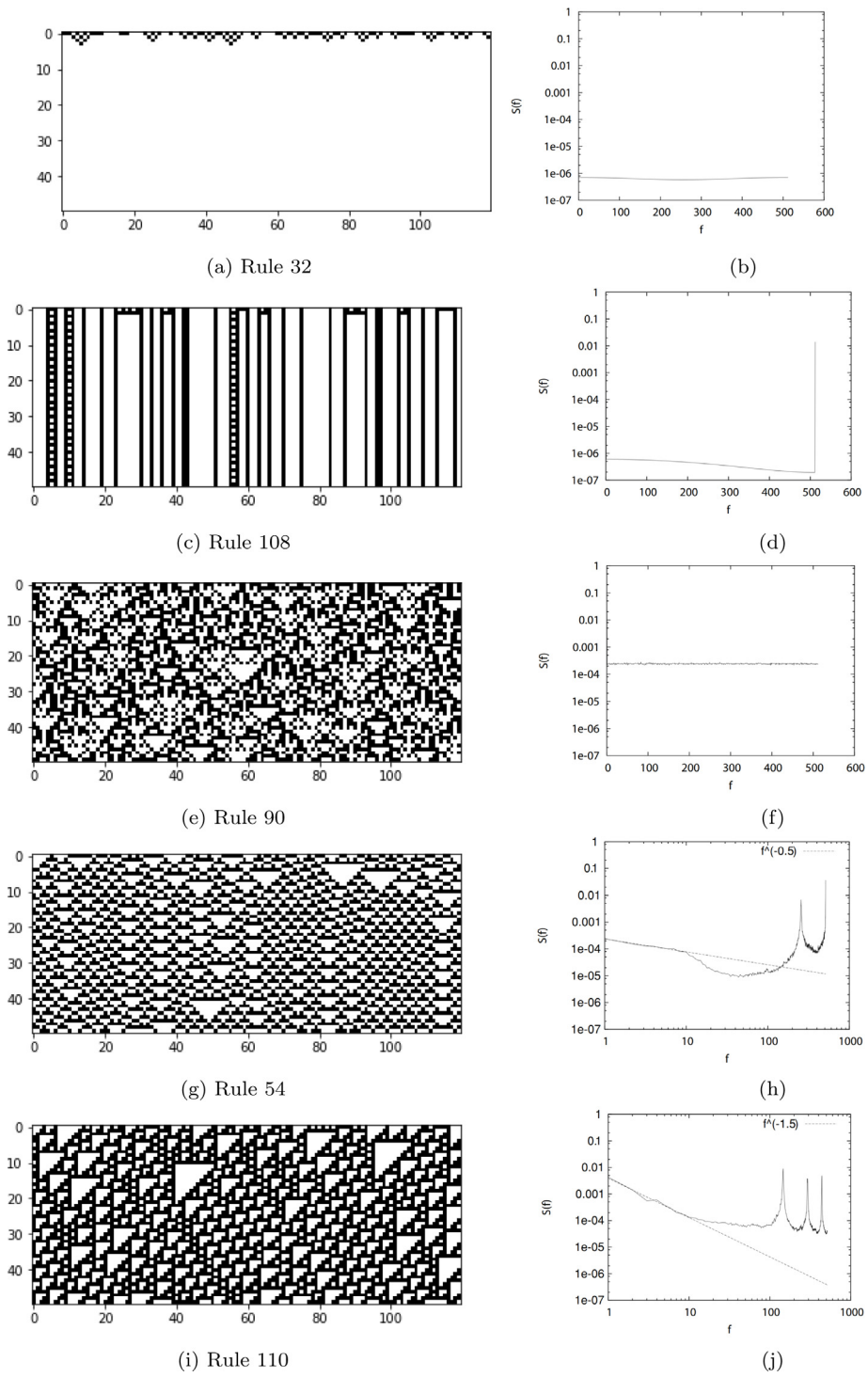


Fig. 17. Space–time patterns (a, c, e, g, i) and corresponding power spectrum (b, d, f, h, j) of ECA rules from different Wolfram classes [94].

An attractor along with the entire transient tree leading to the attractor state is called a *basin of attraction*. The leaves of the tree are the Garden of Eden configurations and the branches are the transients leading to the attractor. The collection of all basins of attraction is called the *basin of attraction field*. This is nothing more than the space of configurations along with all the possible interconnections specified by the CA rule. This means that for a given rule, there are three kinds of CA configurations in $S^{\mathbb{Z}}$ (or S^N for finite CA) that can occur during the evolution, namely Garden of Eden, cyclic or transient configurations, where Garden of Eden configurations constitute a subset of the transient configurations. In the most extreme case, all configurations are Garden of Eden configurations except for one (e.g. ECA rule 0). For irreversible rules, the fraction of all configurations that are Garden of Eden configurations tends to zero at a rate λ^N, where $\lambda = 0.78$ for non-additive rules [101]. If the CA rule is reversible, there are no Garden of Eden configurations. Cyclic configurations constitute the attractor. If the CA rule is reversible, all configurations are cyclic.

The convergence of configurations towards an attractor is characteristic of dissipative dynamical systems [102]; CA generally fall under this category since the vast majority of CA rules

are irreversible. A high degree of convergence is characterized by bushy transient trees and a high G-density. This points to highly irreversible, ordered dynamics. Alternatively, a low degree of convergence is characterized by sparsely branched transient trees with a low G-density. This points to slightly irreversible, chaotic dynamics [4]. These observations led both Wuensche and Kaneko to similar conclusions regarding the typical attractor basins of finite CA for each of Wolfram's classes [103,104]:

- **Class I**: Very short transients, a small number of attractors with short periods (mainly point attractors), very high in-degree, very high leaf density (very ordered dynamics).
- **Class II**: Very short transients, a high number of short periodic attractors (but also point attractors), high in-degree, very high leaf density. The number of attractors grows with lattice size as $e^{\alpha N}$, with $\alpha \approx 0.2 - 0.4$.
- **Class III**: Very long transients, a few long periodic attractors, low in-degree, low leaf density (chaotic dynamics). The number of attractors changes irregularly with N.
- **Class IV**: Moderate transients, moderate-length periodic attractors, moderate in-degree, very moderate leaf density (complex dynamics). The number of attractors increases exponentially yet irregularly with N.

Clearly, the attractor structure of a CA provides a visual overview of the CA's properties such as the G-density, transient length, and behavioral class. To this end, Wuensche developed the Discrete Dynamics Lab simulation software [105]. Besides, the resulting basin of attraction offers an interesting dynamical systems theory view on CA issues related to the theory of computation [106–108].

6.2. Topological dynamics

Wolfram's classes each have analogous classes in the theory of continuous dynamical systems. Kůrka defined a classification inspired by general concepts from dynamical systems theory that is applicable to both CA and continuous dynamical systems [20]. This classification is based on the topological concepts of equicontinuity, sensitivity, transitivity, and positive expansivity. Given a metric space $\mathcal{M}$, a map f on $\mathcal{M}$ and a metric d on $\mathcal{M}$, these concepts are defined as follows;

- f is *equicontinuous* at $x \in \mathcal{M}$ if

$$\forall \epsilon > 0, \exists \delta > 0 : \forall y \in B_\delta(x), \forall n \in \mathbb{Z}, d\left(f^n(x), f^n(y)\right) < \epsilon, \tag{36}$$

  where $B_\delta(x) = \{y \in \mathcal{M}; d(x,y) < \delta\}$ is an open ball with radius δ centered at x. Equicontinuity of f at x indicates that the orbits of x and all points in a sufficiently small neighborhood of x will remain arbitrarily close to each other.
- f is *sensitive* to initial conditions if

$$\forall x \in \mathcal{M}, \forall \epsilon > 0, \exists \delta > 0 : \forall y \in B_\delta(x), \forall n \in \mathbb{Z}, d\left(f^n(x), f^n(y)\right) \geq \epsilon. \tag{37}$$

  Sensitivity means that for every $x \in \mathcal{M}$, there exists at least one $y \in \mathcal{M}$ arbitrarily close to it such that the orbits of x and y will eventually become separated.
- f is *transitive* if

$$\forall U, V \subseteq \mathcal{M} \text{ where } U, V \text{ are open}, \exists n \in \mathbb{Z} \text{ such that } f^n(U) \cap V \neq \emptyset. \tag{38}$$

  Transitivity means that for any two points $x, y \in \mathcal{M}$, there is in any arbitrarily small neighborhood of x a point whose orbit will eventually cross any arbitrarily small neighborhood of y. Transitivity implies that the system cannot be split into two independent subsystems [57].
- f is *positively expansive* if

$$\forall \epsilon > 0, \forall x, y \in \mathcal{M} \text{ where } x \neq y, \exists n \in \mathbb{Z} : d\left(f^n(x), f^n(y)\right) \geq \epsilon. \tag{39}$$

  Positive expansivity means that the orbits of any two distinct points in $\mathcal{M}$ will eventually diverge, so positive expansivity is a stronger than sensitivity.

These concepts can be applied to CA, in which case $\mathcal{M}$ is the space of all possible CA configurations $S^{\mathbb{Z}}$, $f = \Phi$ is the global update rule and the distance between two configurations $C_1 = \{s_1(c_i, t)\}_{i \in \mathbb{Z}}$ and $C_2 = \{s_2(c_i, t)\}_{i \in \mathbb{Z}}$ is given by

$$d(C_1, C_2) = \sum_{i=-\infty}^{+\infty} \frac{\delta(s_1(c_i, t), s_2(c_i, t))}{k^{|i|}}, \tag{40}$$

where δ is the Kronecker delta function. This distance function induces the standard topology on $S^{\mathbb{Z}}$. Now Kůrka's classification is defined as follows [20]:

- **K1**: All points (i.e. all configurations in $S^{\mathbb{Z}}$) are equicontinuous under Φ. This implies that the limiting configuration of the CA is reached in a finite number of time steps.
- **K2**: Some, but not all points are equicontinuous under Φ.
- **K3**: Φ is sensitive but not positively expansive.
- **K4**: Φ is positively expansive.

For arbitrary CA, we have Kůrka's classification and Wolfram's agree as follows: Class I and II CA belong to K1 ∪ K2, Class III CA belong to K3 ∪ K4 and Class IV CA belong to K2 ∪ K3. In some papers, K3 is further subdivided [109–111]:

- **K3a** Φ is sensitive but not transitive.
- **K3b** Φ is transitive but not positively expansive.

It is again undecidable to which class an arbitrary CA rule belongs, because whether a certain CA rule has a topological property like equicontinuity can generally not be determined without considering the time evolution of all possible initial configurations of the CA. However, in the case of ECA, the classification is decidable [112]. In the work of Schüle and Stoop [112], it was found that sensitivity is likely a condition for computational universality. This agrees with the previously discussed idea that complex dynamics is to be found in the transition region between order and chaos (Section 4.2), which corresponds to K3 in Kůrka's classification [20].

6.3. Formal language theory

Formal language theory provides a computational approach to study the attractors generated by the CA evolution. When CA time evolution is regarded as a purely computational process, the initial configuration constitutes the information to be processed. The structure of the output can then be described using the mathematical theory of formal languages. Formal language theory provides a useful analytical tool to study CA theory and its applications [113–116].

6.3.1. Formal languages and CA

The set $\Omega(t)$ of all possible CA configurations generated at time step t by a certain CA rule constitutes a formal language over the alphabet $\Sigma = \{0, 1 \ldots, k-1\}$ [26]. As an example, Fig. 18 shows three state transition graphs corresponding to regular languages over {0, 1}, corresponding to binary CA. Fig. 18(a) represents the set of all possible sequences of zeros and ones,

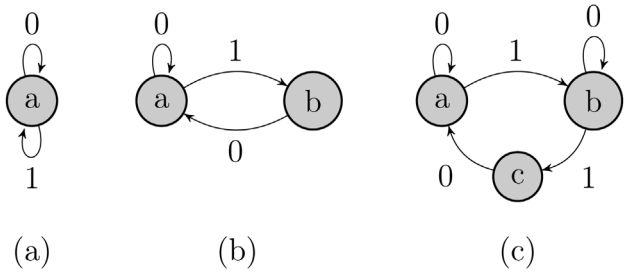


Fig. 18. Examples of state transition graphs of regular languages over the alphabet $\Sigma = \{0, 1\}$.

Fig. 18(b) excludes any sequence that contains pairs of adjacent ones, and Fig. 18(c) represents the set of sequences in which an even number of isolated ones appear between every 0110 block. The increasing complexity in these regular languages is reflected by the increase in the number of nodes. Letting α^* denote an arbitrary number of repetitions of the string α, the regular languages shown in Fig. 18 can be represented by $((0^*)(1^*))^*$, $(0^*(10)^*)^*$ and $(0(0^*)1(0^*)1^*$ respectively. The non-terminals a, b and c hold no special significance and can be omitted in the state transition diagram.

The so-called Chomsky hierarchy is a hierarchy of formal languages based on the properties of the grammars required to generate the languages and characterized by the memory needed to implement the rules of the grammar [117]:

- **Type 0**: A recursively enumerable grammar. There are no restrictions on the grammar, which means that all grammars are type-0 grammar. They require an indefinitely large memory to implement.
- **Type 1**: A context-sensitive grammar. The required memory is proportional to the input word length.
- **Type 2**: A context-free grammar. The required memory is arranged in a stack, with a fixed number of elements available at a given time.
- **Type 3**: A regular grammar. Only finite memory is required, which means regular languages can be recognized by a finite automaton.

Any set $\Omega(t)$ of all possible CA configurations generated at time step t generated after a finite number of time steps forms a regular language [26]. The same is not necessarily true for the limiting configuration $\Omega(t \to \infty)$, which can be a non-regular language [20]. A CA is called regular if $\Omega(t \to \infty)$ is regular. Culik described this result from a topological point of view [118]. Gilman proved that all CA are at most context-sensitive and gives an example of a CA that is neither regular nor context-free [119]. For ECA, it has been shown that rules 22 and 94 are non-regular [120,121], while rule 122 is not context-free [122].

The number of nodes in the minimal state transition graph that generates $\Omega(t)$, denoted by $\Xi(t)$, is a measure for the complexity of $\Omega(t)$. In a sense, $\Xi(t)$ yields the shortest specification of $\Omega(t)$ in terms of regular languages. For Class I and II CA, $\Xi(t)$ stays constant or increases at most linearly with time. For Class III and IV CA, $\Xi(t)$ increases rapidly with time, eventually leading to a non-regular limit set $\Omega(t \to \infty)$ [26].

6.3.2. The finite time set $\Omega(t)$

Construction. Here, we will consider the construction of the minimal grammar of the regular language $\Omega(t)$.

Consider for example ECA rule 76 which has the following rule table entries

$$111 \to 0,\ 110 \to 1,\ 101 \to 0,\ 100 \to 0,\ 011 \to 1,\ 010 \to 1,$$
$$001 \to 0,\ 000 \to 0. \tag{41}$$

Fig. 19(a) shows the corresponding De Bruijn graph [123]. Each possible path through the graph corresponds to a certain initial configuration in $\Omega(0)$. The sequence of arc values corresponding to this path represents the updated configuration in $\Omega(1)$. In this way, the entire set $\Omega(1)$ is represented by the De Bruijn graph of the ECA rule. Note that, due to the irreversibility of rule 76, multiple paths through the graph may correspond to the same updated configuration.

A simplified state transition graph representing the same regular language can be obtained from the De Bruijn graph by finding pairs of nodes where the arcs with the same values are directed to the same nodes, and then combining these pairs of nodes into a single node [124] (Fig. 19(b)).

The simplified De Bruijn diagram and its simplified versions are non-deterministic finite automata (NDFA), since there are nodes with more than a single outgoing arc carrying the same symbol. Otherwise, we have a deterministic finite automaton (DFA). Often, it is convenient to have a DFA that is equivalent to the NDFA because DFA are more efficiently executable [125]. The former can always be found by the power set construction method [26]. The resulting DFA resulting can be simplified by joining equivalent nodes, which yields the minimal DFA corresponding to the minimal grammar that generates $\Omega(1)$.

Any path through the minimal DFA yields a word in the regular language $\Omega(1)$ [26], i.e. $\Omega(1) = ((0^*)1(0 \vee 10))$, which indicates that partial neighborhoods in a configuration in $\Omega(1)$ can overlap to form a full neighborhood. With this information, the De Bruijn diagram for $\Omega(2)$ can be constructed, after which the steps outlined above are followed to construct the minimal DFA. Repeating this process shows that $\Omega(t)$ constitutes a regular language for finite values of t and allows for determining its structure. The above discussion is easily generalized to arbitrary one-dimensional CA. Generally, the number of nodes in the De Bruijn diagram of $\Omega(t)$ is $k^{(|N|-1)t}$, while DFA size can be as large as $2^{k^{(|N|-1)t}} - 1$ but is usually much smaller. On the other hand, a generalization to higher dimensional CA is not straightforward, as the De Bruijn diagrams cannot be used in this case [67,126].

Properties. Any path through the minimal DFA with length greater than $\Xi(t)$ contains a cycle, since it must necessarily traverse at least one arc more than once. Moreover, $\Xi(t)$ is a non-decreasing function of time because for Class I and II CA, $\Xi(t)$ either tends to a constant value after only a few time steps or increases linearly with time. For Class III and IV CA, $\Xi(t)$ increases rapidly with time. The upper bound on $\Xi(t)$ is $2^{k^{(|N|-1)t}} - 1$ [26].

Many useful properties of $\Omega(t)$ can be derived from the adjacency matrix M of its minimal DFA. The (i, j) element of M^X gives the number of paths of length X from node i to node j. The sum of all elements of M^X is denoted by $N_\Sigma(X)$ and yields the total number of different paths of length X through the minimal DFA. Every such path corresponds to a sequence of length X that can occur in the regular language the DFA represents.

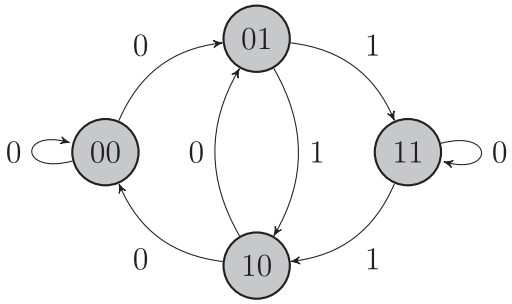
For large values of X, we have the following approximation

$$N_\Sigma(X) \approx \mathrm{Tr}[M^X] = \sum_i \lambda_i^X \approx \lambda_{\max}^X, \tag{42}$$

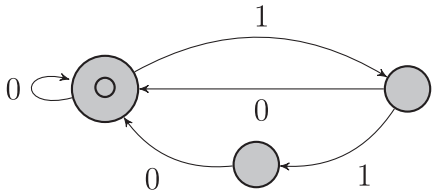
where λ_i are the eigenvalues of M and $\lambda_{\max}$ is the largest value. In the case of ECA rule 76, this yields $\lambda_{\max} \approx 1.839$.

The set $\Omega(t \to \infty)$ constitutes a Cantor space with dimension $S^{(x)} = \lim_{X\to\infty} \frac{1}{X} \log_k N_\Sigma(x)$. If $\Omega(t \to \infty)$ constitutes a regular language, this reduces to $S^{(x)} = \log_k \lambda_{\max}$.

Let $L(t)$ denote the length of the shortest block that is excluded from $\Omega(t)$. For irreversible CA, $\Xi(t)$ provides an upper bound on $L(t)$. In general, $L(t)$ increases (not necessarily monotonically) with time, which means that progressively more distinct blocks are excluded from $\Omega(t)$.



(a) De Bruijn graph corresponding to $\Omega(1)$ of ECA rule 76.



(b) Minimal DFA corresponding to $\Omega(1)$ of ECA rule 76. A node containing a circle indicates the start symbol.

Fig. 19. De Bruijn graph and minimal DFA for ECA rule 76.

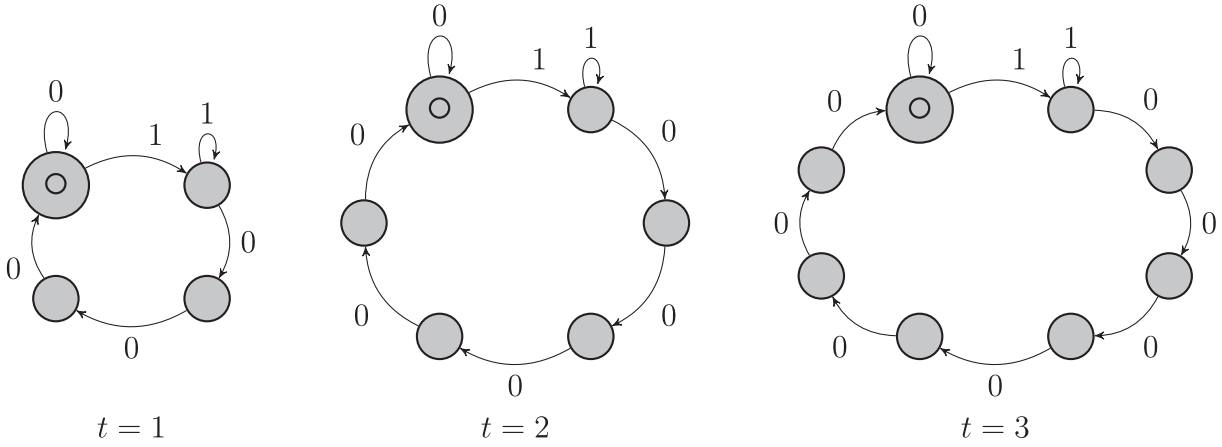


Fig. 20. Minimal DFA corresponding to $\Omega(t)$ generated by ECA rule 128 during the first three time steps.

Class I and II. There are two options regarding the size of the minimal DFA. In the first case, once the initial transients have passed, the minimal DFA, and consequently $\Omega(t)$, are invariant under the global transition function Φ or some finite iteration of it [26]. Generally, the number of configurations in $\Omega(t)$ is infinite. In the second case, the size of the minimal DFA increases linearly with time, while it maintains its overall structure. Besides, $L(t)$ increases linearly with time. The dimension $S^{(x)}$ quickly tends to zero for large t. More specifically for Class I and II CA, $\Xi(t)$ increases as some polynomial function of time [26]. An example of this behavior is given by ECA rule 128, shown in Fig. 20.

Class III and IV. Class III and IV CA generate $\Omega(t)$ sets that correspond to a minimal DFA whose structure becomes increasingly complex as the number of time steps increases (Fig. 21).

Many Class III rules yield $\Xi(t)$ that grows on average more rapidly than any polynomial with time [26].

The limit set $\Omega(t \to \infty)$). The limit sets of Class I and II CA constitute regular languages. However, for Class III and IV CA, the regular language complexity of $\Omega(t)$ increases steadily with time, yielding non-regular languages in the infinite time limit. Limit sets of Class III CA generally constitute context-sensitive languages while limit sets of Class IV CA seem to constitute recursively enumerable languages [26].

The properties of regular languages can generally be found by a finite number of computations. This is not the case for more general languages such as context-free or recursively enumerable languages. The most efficient way to determine the outcome of such CA is achieved by simply simulating the CA itself. There is no possible shortcut to determine the outcome (see Section 3.1.4). This implies that computation of the limiting configuration for Class III and IV CA would require an infinite amount of computations, as opposed to Class I and II CA, for which a shortcut does exist [26].

The computational approach to study CA highlights the undecidability associated with many questions regarding CA behavior. Many properties of CA limit sets have been investigated, where it was found that all non-trivial properties of CA limit sets are undecidable [118,127,128]. This is related to Rice's theorem from computability theory, which states that all non-trivial properties of a program's behavior are undecidable [129]. The above discussion on formal language theory in a CA context concerns itself with the limit set of one-dimensional CA starting from a random seed. One can also look at the limit set of higher dimensional CA [130] or the limit set of CA for other sets of initial configurations [131]. In addition, another approach was formulated by Li, who provided analytical and experimental results regarding the spectra of regular languages generated by CA [132].

6.3.3. Language-based CA classification

Kůrka formulated the following CA classification based on the complexity of the generated languages [20]:

- **L1**: $\Omega(t \to \infty)$ is bounded periodic
- **L2**: $\Omega(t \to \infty)$ is regular but not bounded periodic
- **L3**: $\Omega(t \to \infty)$ is non-regular

6.4. D-spectrum

The mean-field approach (Section 5.3) led Fates [133] to propose a classification based on the so-called *D-spectrum* parameter.

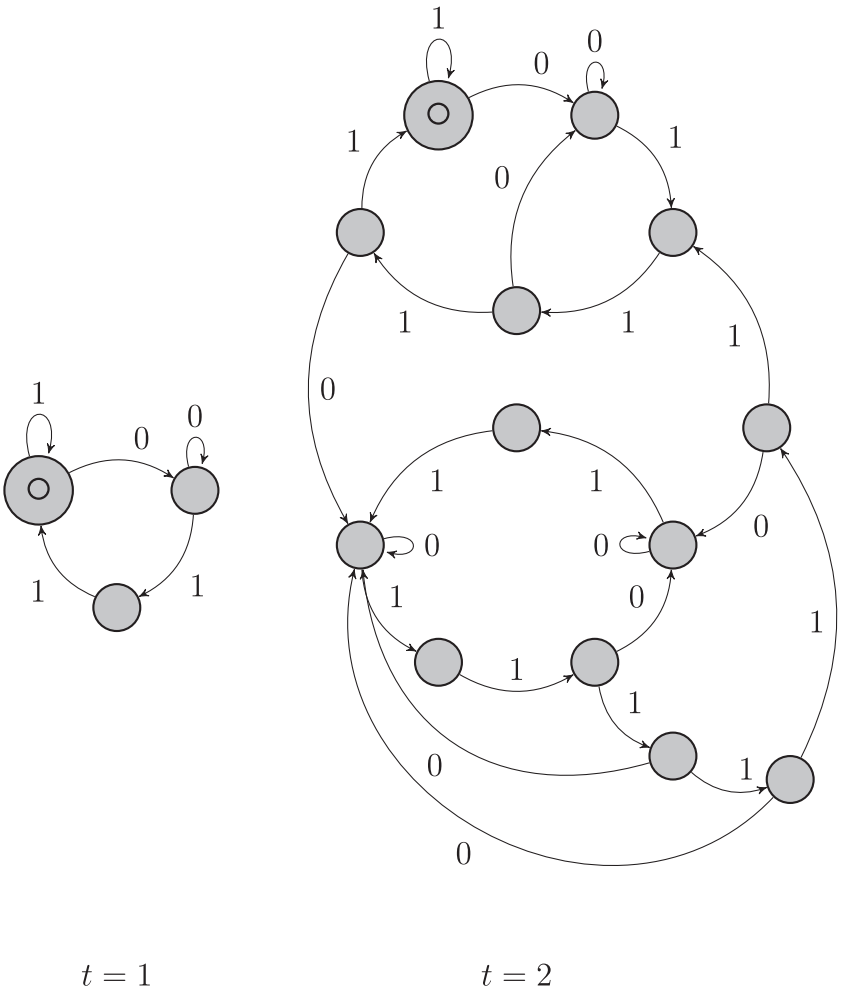


Fig. 21. Minimal DFA corresponding to $\Omega(t)$ generated by ECA rule 126 during the first two time steps.

Letting $P \subset S^N$ denote the set of configurations that appear in the periodic part of the orbit of a finite CA of size N, Fates [133] defined the D-spectrum as:

$$\tau = |\#_1(x), x \in P|. \tag{43}$$

The D-spectrum equals the number of different densities that occur in the periodic part of the orbit of the CA, and induces the following (global) CA classification [133]:

- **P**: a CA is periodic if there exists a constant K_p such that $\tau \in [1, K_p]$ for all initial configurations.
- **C**: a CA is chaotic if there exists a constant K_c such that $\tau \in [K_c, \infty]$ for all initial configurations.
- **H**: this class contains CA that are neither periodic nor chaotic.

Here it should be noted that K_p and K_c have a fixed value for all the rules in a given rule space. Using a numerical estimation of the D-spectra for all ECA rules and comparing the resulting (global) classification to Wolfram's (local) classification, it was found that Class P contains rules from Class I and II, Class C contains rules from Class III and Class H contains all Class IV rules in addition to some rules from classes I, II and III [133].

6.5. Rényi entropy

Entropy is a useful concept from statistical physics and information theory that can be extended in many ways to CA [16]. A generalization of Shannon's definition of entropy is provided by the Rényi entropy of order α. Given a discrete random variable X with possible outcomes $\{x_1, \ldots, x_n\}$ and corresponding probabilities $P(X = x_i) = p_i$, the Rényi entropy of order α, denoted by $S_\alpha(X)$, is defined as

$$S_\alpha(X) = \frac{1}{1-\alpha} \log_k \left(\sum_{i=1}^{n} p_i^\alpha \right), \tag{44}$$

where $\alpha \geq 0$ and $\alpha \neq 1$.

6.5.1. Global Rényi entropy and CA

There are several ways to define Rényi entropy in the framework of CA, depending on how the discrete variable X is defined, whether finite or infinite CA are considered, whether spatial sequences, temporal sequences or spatio-temporal patches are considered, and whether the entropy is defined globally or locally. Here, we first discuss the former, while the latter is the topic of 6.5.3.

We consider the probability $P(\Omega)$ that a certain configuration in $\Omega(t)$ occurs. For a finite CA of size N, an equiprobable ensemble of initial configurations consists of all 2^N possible initial configurations, where each configuration occurs with equal probability. The evolution of reversible CA preserves such an equiprobable ensemble. However, most CA rules are irreversible (dissipative), implying ensembles that are not equiprobable. The properties of the more probable configurations will then tend to dominate the statistical averages over the ensemble [11,101].

The Rényi entropy can be defined for finite CA of size N by using $X = \Omega(t)$ in Eq. (44) to arrive at

$$S_\alpha(t) = \frac{1}{N}\frac{1}{1-\alpha}\log_k\left(\sum_{C\in\Omega(t)}(P(C,t))^\alpha\right). \tag{45}$$

In the case of infinite CA, the same arguments apply, but the number of possible configurations in $\Omega(t)$ is infinite. One then defines the block Rényi entropy by using $X = S^{|B|}$:

$$S_\alpha(|B|,t) = \frac{1}{|B|}\frac{1}{1-\alpha}\log_k\left(\sum_{B\in S^{|B|}}(P(B,t))^\alpha\right), \tag{46}$$

where $P(B \in S^{|B|}, t)$ is the frequency with which the block B occurs in the configurations in $\Omega(t)$. Now, the Rényi entropy for infinite CA is defined as the asymptotic limit for large blocks of the block Rényi entropy, i.e.

$$S_\alpha(t) = \lim_{|B|\to\infty} S_\alpha(|B|,t). \tag{47}$$

Since the spatial correlation in a single configuration quickly tends to zero as the distance between sites increases, $S_\alpha(|B|,t)$ quickly converges to its asymptotic value for $X \to \infty$. This means that the block Rényi entropy is a very good approximation for the Rényi entropy for sufficiently large $|B|$.

For $\alpha = 0$, the Rényi entropy corresponds to the spatial set entropy that considers the configurations that can be reached at time t, regardless of the probability that they occur. In the large time limit, the configurations occurring with a non-zero probability in the statistical ensemble are typically configurations lying on the attractor of the CA. This means that $S_0(t \to \infty)$ gives the fraction of configurations that is cyclic. The quantity $S_0(t \to \infty)$ is called the spatial set dimension and is denoted by d_0. Similarly to the theory of continuous dynamical systems, d_0 yields the fractal dimension of the CA attractor. If the attractor is simply the set of all possible configurations (which is the case for reversible CA), $d_0 = 1$. When the number of configurations on the attractor is finite, then $d_0 = 0$. For Class III rules, the attractor is usually some fractal subspace of the space of all configurations, so $0 < d_0 < 1$.

For $\alpha = 1$, the Rényi entropy corresponds to the spatial Shannon entropy or spatial measure entropy. It considers the probabilities of all reachable configurations at time t. It yields the average information per site of a configuration at time t, so $S_1(t \to \infty)$ yields the average information per site for configurations lying on the attractor. $S_1(t \to \infty)$ is called the spatial measure dimension and is denoted by d_1. This quantity is similar to the fractal dimension, but also takes into account the relative frequency with which the different points on the attractor are visited. The spatial measure and set dimensions are related through the following inequality: $0 \le d_1 \le d_0 \le 1$.

6.5.2. Temporal entropies

So far, a statistical ensemble was obtained by fixing the time t, and considering all CA configurations that occurred at that time along with their corresponding probabilities. Alternatively, a statistical ensemble can be obtained by fixing the lattice position i, and looking at all possible temporal sequences that can occur at that site along with their corresponding probabilities. The definitions of the temporal Rényi entropy $S_\alpha^{(t)}(T,i)$ for finite blocks with length T are similar to that of the spatial Rényi entropy:

$$S_\alpha^{(t)}(T,i) = \frac{1}{T}\frac{1}{1-\alpha}\log_k\left(\sum_{B\in S^T}(P(B,i))^\alpha\right), \tag{48}$$

where $P(B,i)$ denotes the probability that the temporal sequence B occurs at site i. For $\alpha = 0$ and $\alpha = 1$ the temporal set dimension $d_0^{(t)}$ and temporal Shannon dimension $d_1^{(t)}$ are obtained in the $T \to \infty$ limit.

Table 2
CA spatial (d_1) and temporal ($d_1^{(t)}$) Shannon dimensions for Wolfram classes I–III.

Wolfram class	d_1	$d_1^{(t)}$
Class I	0	0
Class II	>0	0
Class III	>0	> 0

Spatial and temporal entropies of CA are generally quite different [16] (Table 2). Moreover, Class IV CA are generally too unpredictable to define meaningful dimensions [101].

When looking at the typical Rényi entropies during the CA time evolution as a function of the Langton parameter (Section 4.2, it becomes clear that there is a sharp transition from ordered to chaotic behavior, which shows characteristics of a first order phase transition. However, there are exceptional cases where the transition occurs relatively smoothly, indicating a second order phase transition [18]. The rules in such a transition region are typically Class IV rules.

6.5.3. Local entropy

Local Rényi entropy can be defined for individual configurations as the block entropy (Eq. (46)). Such an entropy parameter is inherently local, but will be discussed here in order to prevent a scattered discussion of entropy. For $|B| = |N|$, the statistical ensemble of blocks consists of all possible neighborhood configurations along with the frequency with which they occur in the configuration. In this case, the block entropy is called the input entropy since the frequency with which different neighborhood configurations occur is also the frequency with which the different entries in the rule table are visited.

Fig. 22 shows how the input entropy varies during the time evolution of four different ECA rules evolving from a random seed. For Class I and II rules, the input entropy quickly drops to a zero and non-zero value respectively. For Class III rules, the input entropy remains high during the entire time evolution, reflecting the chaotic and almost random nature of Class III CA. For Class IV rules, the input entropy is high and shows a lot of variation due to the complex nature of their space–time pattern. In fact, the variance of the input entropy has been suggested as a measure to detect Class IV rules [4].

6.6. Computational complexity

Computational problems associated with the long-term behavior of infinite CA are notoriously hard. Bar a few pathological exceptions for one-dimensional CA, they are all undecidable [134]. Nevertheless, when considering only the finite time behavior and finite strings (subconfigurations) instead of infinite configurations, the associated computational problems become decidable. Naturally, this means that problems in the

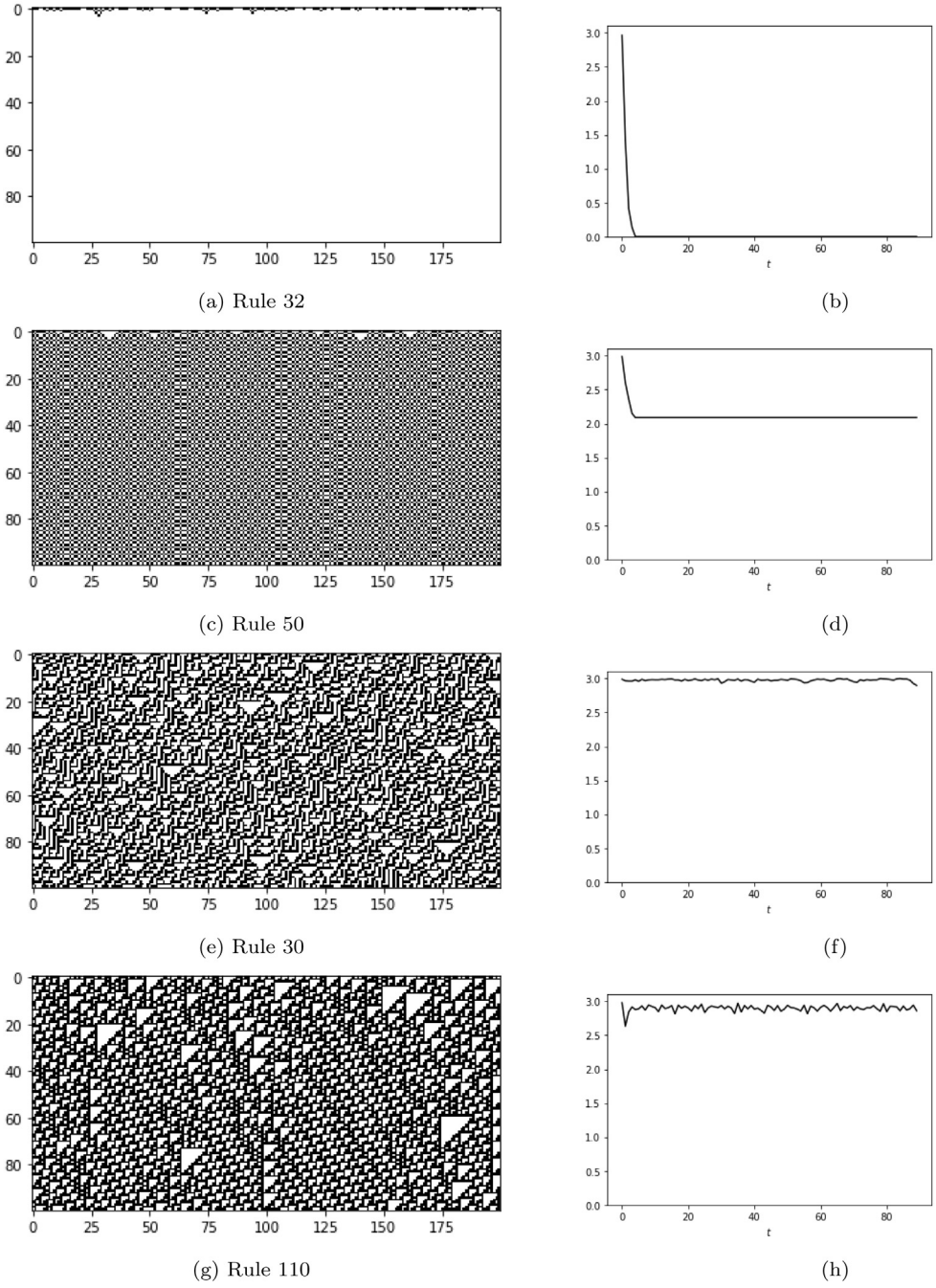


Fig. 22. Space–time patterns (a, c, e, g) and corresponding input entropy (b, d, f, h) of ECA rules from different Wolfram classes.

context of finite CA are also decidable. The computational complexity of a certain CA problem generally strongly depends on the dimension and rule [135]. Moreover, it has been proven that there exist CA for which the following decision problems are NP-complete [135]:

- Given a string of length K over S, is there a predecessor configuration that could have led to it in K time steps?
- Given a configuration that contains a string of length K over S, will the string occur after K time steps?
- Given a temporal sequence of K states in a given site, is there a string over S that could have led to it in K time steps?

Instead of considering finite blocks of length K, one can also consider problems related to entire CA configurations. In the case of infinite CA, it is possible to determine arbitrarily large finite segments of the evolved configuration $\Phi(c)$ using only finite segments of a configuration $c \in S^{\mathbb{Z}}$ [21]. This provides a way to model the computational complexity of infinite CA.

When it comes to CA classification, the computational complexity of the configuration reachability problem (CREP) deserves special attention. It addresses a central question in the theory of CA [136]: given a CA, an initial configuration X and a final configuration Y, will X ever evolve to Y? As the infinite time behavior of infinite CA is generally undecidable, generally the CREP will be undecidable as well. However, as finite CA ultimately yield periodic behavior, their infinite time behavior is not undecidable. In this case, the CREP will generally be a PSPACE problem [67]. However, there are certain CA rules for which the CREP can be solved more quickly. This makes the CREP an interesting problem,

Table 3
Summary of the discussed local and global parameters over the different Li–Packard classes.

	LP1	LP2	LP3/LP4	LP6	LP5
Local					
Simple seed evolution	Evolution to zero	Finite growth	Periodic patterns	Complex patterns	Sierpinski patterns
Mean-field curves	Stable fixed point at the origin	Marginally stable fixed point at the origin	Unstable fixed point at the origin and a stable non-zero fixed point	Two or more fixed points	Unstable fixed point at the origin and a stable non-zero fixed point
Two-point correlation	High			Intermediate	Low
Difference pattern spreading rate	Zero			Highly variable	Maximal
Lyapunov exponent	Negative			Positive	
Compressibility ratio $\gamma_{CR}(s(\cdot,t))$	$\gamma_{CR}(s(\cdot,t)) \to 0$	$\gamma_{CR}(s(\cdot,t)) < 1/2$	$\gamma_{CR}(s(\cdot,t)) < 1/2$	$\gamma_{CR}(s(\cdot,t)) \to 1$	$\gamma_{CR}(s(\cdot,t)) = 1$
Spectrum $S(f)$	Low at all frequencies	Low at all frequencies, with a peak at $f = 0$	Low at all frequencies, with a peak at $f = T/P$	1/f shape for low frequencies	High at all frequencies (white noise)
Local entropy evolution	Zero/constant	Intermediate/constant		High/variable	High/constant
Global					
Attractor transient length	Short			Intermediate	Long
Attractor type	Point attractor		Short periodic attractor	Intermediate-length periodic attractor	Long periodic attractor
Attractor in-degree/leaf-density	High			Moderate	Low
Minimal DFA size	Increases at most linearly with time. Structure remains invariant.			Increases faster than linearly with time. Structure changes drastically.	
Spatial entropy	Low			Intermediate	High
D-spectrum τ	$\tau \in [1, K_P]$			$\tau \notin [1, K_P] \cup [K_C, \infty]$	$\tau \in [K_C, \infty]$

since its computational complexity induces a CA classification of CA rules based on predictability [136]:

- When the CREP of a certain CA rule is a P problem, the rule is called predictable.
- When the CREP of a certain CA rule is an NP problem, the rule is called weakly predictable. This holds for additive CA.
- Rules that are not weakly predictable are called unpredictable. These rules are computationally irreducible.

In this sense, predictable and weakly predictable CA are considered to be computationally reducible systems, as there does exist some shortcut to determine the evolved configuration at an arbitrary time step that is faster than simply evolving the CA itself.

7. Conclusions

The aim of this review was to give an overview of the most common approaches to classify CA, which is one of the main open problems in CA research [1]. The reviewed parameters and classification schemes originate from many different fields, involving concepts from non-linear science, statistical physics, information theory, computation theory and topological dynamics, highlighting the interdisciplinary character inherent to CA research.

We distinguished between parameters based on the rule table and those based on the space–time patterns, with the latter being further subdivided into local and global parameters. We framed our review along those three lines because of their different nature, both in terms of the tools and methods required, as well as their applicability and generality. There is no single all-purpose CA classification scheme. Each approach has its merits and suits specific needs, depending on the family of CA under consideration and the research context. Table 3 summarizes how the local and global parameters characterize different behavioral classes. The table places LP6 between LP4 and LP5 in order to highlight how complex CA occur as a transition region between order and chaos, an idea central throughout this review that is applicable to any dynamical system, not just CA [137]. Note that the entries regarding attractor properties pertain to finite CA, not Kůrka's attractor classification which applies to infinite CA.

Due to their simplicity and limited size of the rule space, as such facilitating a comprehensive study of all possible rules, ECA are the most studied family of CA. Consequently, the scope of many approaches such as the two-point correlation function and Lyapunov exponents is limited to this CA family. Moreover, some approaches (e.g. the global classification based on the D-spectrum and local classifications based on the Kolmogorov complexity and spectral properties of the space–time pattern) are applicable to arbitrary CA, but have only been investigated and applied to ECA. Restricting to ECA is convenient, yet significantly reduces the applicability of the proposed classifications in real-world scenarios, as CA that occur in real-world applications are generally more complex than ECA [138,139]. Furthermore, more complex CA can exhibit behavioral features which cannot occur in ECA due to their simplicity, so conclusions drawn based on an investigation of the ECA rule space cannot be extended to arbitrary CA without deliberation.

Furthermore, this review highlights a clear dichotomy in CA research. On the one hand, there is theoretical research that focuses on analytical results from the theory of computation and topological dynamics, while on the other hand there is a vast literature on simulation-based research focusing on the statistical properties of the simulated space–time patterns. In the context of classification, the former is typically related to global parameters while the latter uses local parameters. There is a significant gap between both lines of research and cross-fertilization of ideas is rare, yet is essential to bridge the gap between the theory and phenomenology of CA. To this end, comparing classifications based on local and global parameters may be a starting point.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We acknowledge the support from the Special Research Fund of Ghent University, Belgium (Grant N° 01Z02016).

References

[1] S. Wolfram, Twenty problems in the theory of cellular automata, Phys. Scr. T9 (1985) 170–183, http://dx.doi.org/10.1088/0031-8949/1985/t9/029.

[2] P. Rich, The organizational taxonomy: Definition and design, Acad. Manag. Rev. 17 (1992) 758–781, http://dx.doi.org/10.5465/AMR.1992.4279068.

[3] J. Hollenbeck, B. Beersma, M. Schouten, Beyond team types and taxonomies: A dimensional scaling conceptualization for team description, Acad. Manag. Rev. 37 (2012) 82–106, http://dx.doi.org/10.5465/amr.2010.0181.

[4] A. Wuensche, Classifying cellular automata automatically: Finding gliders, filtering, and relating space-time patterns, attractor basins, and the Z parameter, Complexity (1999) 47–66.

[5] N. Ahmed, W. Teahan, Using compression to find interesting one-dimensional cellular automata, Complex Intell. Syst. 6 (2019) 123–146, http://dx.doi.org/10.1007/s40747-019-00121-7.

[6] J.V. Neumann, A.W. Burks, Theory of Self-Reproducing Automata, University of Illinois Press, USA, 1966.

[7] J.M. Baetens, B. De Baets, Phenomenological study of irregular cellular automata based on Lyapunov exponents and Jacobians, Chaos 20 (3) (2010) http://dx.doi.org/10.1063/1.3460362, URL http://aip.scitation.org/doi/10.1063/1.3460362.

[8] B. Carter, Cellular automata on triangular, pentagonal and hexagonal tessellations, in: R. Meyers (Ed.), Encyclopedia of Complexity and Systems Sciences, Springer, Heidelberg, Germany, 2009.

[9] T. Toffoli, N. Margolus, Cellular Automata Machines: A New Environment for Modeling, MIT Press, Cambridge, MA, USA, 1987.

[10] B. LuValle, The effects of boundary conditions on cellular automata, Complex Syst. 28 (2019) 97–124, http://dx.doi.org/10.25088/ComplexSystems.28.1.97.

[11] S. Wolfram, Statistical mechanics of cellular automata, Rev. Modern Phys. 55 (1983) 601–644, http://dx.doi.org/10.1103/RevModPhys.55.601, URL https://link.aps.org/doi/10.1103/RevModPhys.55.601.

[12] M. Gardner, Mathematical Games - The fantastic combinations of John Conway's new solitaire game "life", Sci. Am. (1970) 120–123.

[13] B.H. Voorhees, Computational Analysis of One-Dimensional Cellular Automata, World Scientific, 1995, http://dx.doi.org/10.1142/2712, arXiv:https://www.worldscientific.com/doi/pdf/10.1142/2712, URL https://www.worldscientific.com/doi/abs/10.1142/2712.

[14] G. Cattaneo, E. Formenti, L. Margara, G. Mauri, Transformations of the one-dimensional cellular automata rule space, Parallel Comput. 23 (11) (1997) 1593–1611, http://dx.doi.org/10.1016/S0167-8191(97)00076-8.

[15] J.S.T. Mora, M. Gonzalez-Hernandez, H. McIntosh, S. Chapa, Construction of reversible cellular automata by amalgamations and permutations of states, J. Cell. Autom. 4 (2009) 311–322.

[16] S. Wolfram, Universality and complexity in cellular automata, Physica D (1984) 1–35.

[17] K. Culik, S. Yu, Undecidability of CA classification schemes, Complex Syst. (1988) 177–190.

[18] W. Li, N. Packard, C. Langton, Transition phenomena in cellular automata rule space, Physica D 45 (1) (1990) 77–94, http://dx.doi.org/10.1016/0167-2789(90)90175-O.

[19] H.A. Gutowitz, A hierarchical classification of cellular automata, Physica D 45 (1–3) (1990) 136–156, http://dx.doi.org/10.1016/0167-2789(90)90179-S.

[20] P.K.A. rka, Languages, equicontinuity and attractors in cellular automata, Ergodic Theory Dynam. Systems 17 (2) (1997) 417–433, http://dx.doi.org/10.1017/S014338579706985X.

[21] K. Sutner, Computational classification of cellular automata, Int. J. Gen. Syst. 41 (6) (2012) 595–607, http://dx.doi.org/10.1080/03081079.2012.695899.

[22] G. Martinez, J.S.T. Mora, H. Zenil, Computation and universality: Class IV versus class III cellular automata, J. Cell. Autom. (2013) 393–430.

[23] I. Prigogine, I. Stengers, Order Out of Chaos, Heinemann, New York, 1984.

[24] J. Park, K. Steiglitz, W. Thurston, Soliton-like behavior in automata, Physica D 19 (1986) 423–432, http://dx.doi.org/10.1016/0167-2789(86)90068-0.

[25] G. Piccinini, Computation in physical systems, in: E. Zalta (Ed.), Stanford Encyclopedia of Philosophy, 2010.

[26] S. Wolfram, Computation theory of cellular automata, Comm. Math. Phys. 96 (1) (1984) 15–57, URL https://projecteuclid.org:443/euclid.cmp/1103941718.

[27] C. Langton, Computation at the edge of chaos: Phase transitions and emergent computation, Physica D 42 (1) (1990) 12–37, http://dx.doi.org/10.1016/0167-2789(90)90064-V, URL http://www.sciencedirect.com/science/article/pii/016727899090064V.

[28] M. Cook, Universality in elementary cellular automata, Complex Syst. 15 (2004) 1–40.

[29] P. Rendell, Turing universality of the game of life, in: A. Adamatzky (Ed.), Collision-Based Computing, Springer London, London, 2002, pp. 513–539, http://dx.doi.org/10.1007/978-1-4471-0129-1_18.

[30] J. Kari, Theory of cellular automata: A survey, Theoret. Comput. Sci. 334 (1) (2005) 3–33, http://dx.doi.org/10.1016/j.tcs.2004.11.021.

[31] W. Li, Phenomenology of nonlocal cellular automata, J. Stat. Phys. 68 (1992) 829–882.

[32] W. Li, N. Packard, The structure of the elementary cellular automata rule space, Complex Syst. 4 (1990) 281–297.

[33] D. Kunkle, Automatic Classification of One-Dimensional Cellular Automata, 2003.

[34] J. Gorodkin, A. Srensen, O. Winther, Neural networks and cellular automata complexity, Complex Syst. 7 (2001) 1–23.

[35] A. Dhar, P. Lakdawala, G. Mandal, S. Wadia, Role of initial conditions in the classification of the rule space of cellular automata dynamics, Phys. Rev. E 51 (1995) 3032–3037.

[36] F. Ohi, Chaotic properties of the elementary cellular automaton rule 40 in Wolfram's class I, Complex Syst. 17 (2007) 73–82.

[37] J. Mazoyer, I. Rapaport, Global fixed point attractors of circular cellular automata and periodic tilings of the plane: Undecidability results, Discrete Math. 199 (1) (1999) 103–122, http://dx.doi.org/10.1016/S0012-365X(98)00203-9.

[38] J. Baldwin, S. Shelah, On the classifiability of cellular automata, Theoret. Comput. Sci. 230 (1) (2000) 117–129.

[39] G. Oliveira, P. Oliveira, N. Omar, Guidelines for dynamics-based parameterization of one-dimensional cellular automata rule spaces, Complexity 6 (2000) 63–71.

[40] G. Oliveira, P.D. Oliveira, N. Omar, Definition and application of a five-parameter characterization of one-dimensional cellular automata rule space, Artif. Life 7 (2001) 277–301, http://dx.doi.org/10.1162/106454601753238645.

[41] M. Scheffer, J. Bascompte, W. Brock, V. Brovkin, S. Carpenter, V. Dakos, H. Held, E. van Nes, M. Rietkerk, G. Sugihara, Early-warning signals for critical transitions, Nature 461 (7260) (2009) 53—59, http://dx.doi.org/10.1038/nature08227.

[42] H. Chaté, P. Manneville, Criticality in cellular automata, Physica D 45 (1) (1990) 122–135, http://dx.doi.org/10.1016/0167-2789(90)90178-R, URL http://www.sciencedirect.com/science/article/pii/016727899090178R.

[43] S. Carbajal, D. Corne, A. Conty, Parallelizing Automatic Induction of Langton Parameter with Genetic Programming, 2006, pp. 540–544.

[44] H. Wang, G. Yan, Z. Cai, Emergent computing of one-dimensional DNA cellular automata based on langton parameter, J. Jilin Univ. Sci. Ed. 47 (2009) 937–940.

[45] X. Li, X. Li, X. Li, H. Qiu, Multi-agent fare optimization model of two modes problem and its analysis based on edge of chaos, Physica A 469 (2016) 405–419, http://dx.doi.org/10.1016/j.physa.2016.11.022.

[46] S. Raiesdana, M.H. Golpayegani, A. Nasrabadi, Complexity evolution in epileptic seizure, in: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, pp. 4110–4113, http://dx.doi.org/10.1109/IEMBS.2008.4650113.

[47] A. Wuensche, M. Lesser, The global dynamics of cellular automata : an atlas of basin of attraction fields of one-dimensional cellular automata, 1992.

[48] A. A. Wuensche, Cellular automata encryption: the reverse algorithm, Z-parameter and chain-rules, Parallel Process. Lett. 19 (2009) 283–297, http://dx.doi.org/10.1142/S0129626409000225.

[49] A. Wuensche, Complex and chaotic dynamics, basins of attraction, and memory in discrete networks, Acta Phys. Pol. B 3 (2010) 463–478.

[50] G. Vichniac, Boolean derivatives on cellular automata, Physica D 45 (1–3) (1990) 63–74.

[51] F. Bagnoli, R. Rechtman, S. Ruffo, Damage spreading and Lyapunov exponents in cellular automata, Phys. Lett. A 172 (1992) 34–38, http://dx.doi.org/10.1016/0375-9601(92)90185-O, URL http://www.sciencedirect.com/science/article/pii/037596019290185O.

[52] F. Bagnoli, R. Rechtman, E.S. Yacoubi, Control of cellular automata, Phys. Rev. Lett. E 86 (2012) 066201, http://dx.doi.org/10.1103/PhysRevE.86.066201, URL https://link.aps.org/doi/10.1103/PhysRevE.86.066201.

[53] P. Binder, A phase diagram for elementary cellular automata, Complex Syst. 7 (1993) 241–247.

[54] W. Weinert, H. Lopes, Evaluation of dynamic behavior forecasting parameters in the process of transition rule induction of unidimensional cellular automata, Bio Syst. 99 (1) (2010) 6—16, http://dx.doi.org/10.1016/j.biosystems.2009.08.002.

[55] B. Voorhees, Some parameters characterizing cellular automata rules, Complex Syst. 11 (1997) 373–386.

[56] A.N. Kolmogorov, On tables of random numbers, Sankhya A 25 (4) (1963) 369–376, URL http://www.jstor.org/stable/25049284.

[57] J. Dubacq, B. Durand, E. Formenti, Kolmogorov complexity and cellular automata classification, Theoret. Comput. Sci. 259 (1) (2001) 271–285, http://dx.doi.org/10.1016/S0304-3975(00)00012-8, URL http://www.sciencedirect.com/science/article/pii/S0304397500000128.

[58] N. Israeli, N. Goldenfeld, Coarse-graining of cellular automata, emergence, and the predictability of complex systems, Phys. Rev. E 73 (2006) 026203, http://dx.doi.org/10.1103/PhysRevE.73.026203.

[59] J. Ziv, A. Lempel, A universal algorithm for sequential data compression, IEEE Trans. Inform. Theory 23 (3) (1977) 337–343, http://dx.doi.org/10.1109/TIT.1977.1055714.

[60] K. Sayood (Ed.), Introduction to Data Compression, Fifth Edition, M. Kaufmann, 2018, p. v.

[61] F. Blanchard, J. Cervelle, E. Formenti, Some results about the chaotic behavior of cellular automata, Theoret. Comput. Sci. 349 (3) (2005) 318–336, http://dx.doi.org/10.1016/j.tcs.2005.06.038.

[62] N.F. Johnson, Simply Complexity: A Clear Guide to Complexity Theory, 2007.

[63] C.H. Bennett, Logical depth and physical complexity, in: A Half-Century Survey on the Universal Turing Machine, Oxford University Press Inc., USA, 1988, pp. 227–257.

[64] S. Wolfram, Origins of randomness in physical systems, Phys. Rev. Lett. 55 (1985) 449–452, http://dx.doi.org/10.1103/PhysRevLett.55.449.

[65] S. Lloyd, H. Pagels, Complexity as thermodynamic depth, Ann. Physics 188 (1) (1988) 186–213, http://dx.doi.org/10.1016/0003-4916(88)90094-2, URL https://www.sciencedirect.com/science/article/pii/0003491688900942.

[66] P. P. Grassberger, Toward a quantitative theory of self-generated complexity, Internat. J. Theoret. Phys. 25 (1986) 907–938, http://dx.doi.org/10.1007/BF00668821.

[67] J. Hopcroft, R. Motwani, J. Ullman, Introduction To Automata Theory, Languages, and Computation, third ed., Addison-Wesley Longman Publishing Co., Inc., USA, 2006.

[68] M. Gell-Mann, What is complexity? in: Complexity and Industrial Clusters, Physica-Verlag HD, Heidelberg, 2002, pp. 13–24.

[69] J. Crutchfield, K. Young, Inferring statistical complexity, Phys. Rev. Lett. 63 (1989) 105–108, http://dx.doi.org/10.1103/PhysRevLett.63.105.

[70] W. Bialek, N. Tishby, Predictive information, 1999, arXiv:1512.03696.

[71] M. Redeker, A. Adamatzky, G. Martinez, Expressiveness of elementary cellular automata, Internat. J. Modern Phys. C 24 (2012) 1–14, http://dx.doi.org/10.1142/S0129183113500101.

[72] G. Braga, G. Cattaneo, P. Flocchini, C.Q. Vogliotti, Pattern growth in elementary cellular automata, Theoret. Comput. Sci. 145 (1) (1995) 1–26.

[73] H.A. Gutowitz, J.D. Victor, B.W. Knight, Local structure theory for cellular automata, Physica D 28 (1) (1987) 18–48, http://dx.doi.org/10.1016/0167-2789(87)90120-5, URL http://www.sciencedirect.com/science/article/pii/0167278987901205.

[74] N. Boccara, H. Fuks, Modeling diffusion of innovations with probabilistic cellular automata, in: M. Delorme, J. Mazoyer (Eds.), Cellular Automata: A Parallel Model, Springer Netherlands, Dordrecht, 1999, pp. 263–277, http://dx.doi.org/10.1007/978-94-015-9153-9_9.

[75] G. Szabó, J. Vukov, A. Szolnoki, Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices, Phys. Rev. E 72 (2005) http://dx.doi.org/10.1103/PhysRevE.72.047107.

[76] G. Szabó, G. Fáth, Evolutionary games on graphs, Phys. Rep. 446 (4) (2007) 97–216, http://dx.doi.org/10.1016/j.physrep.2007.04.004.

[77] H. Fukś, N. Fatès, Local structure approximation as a predictor of second order phase transitions in asynchronous cellular automata, Nat. Comput. 14 (2015) 507–522, http://dx.doi.org/10.1007/s11047-015-9521-6.

[78] D. Hiebeler, Stochastic spatial models: From simulations to mean field and local structure approximations, J. Theoret. Biol. 187 (3) (1997) 307–319, http://dx.doi.org/10.1006/jtbi.1997.0422.

[79] D. Ruelle, Sensitive dependence on initial condition and turbulent behavior of dynamical systems, Turbulence, Strange Attractors and Chaos, pp. 175–184, http://dx.doi.org/10.1142/9789812833709_0008.

[80] J. Vallejo, M. Sanjuán, Predictability of Chaotic Dynamics: A Finite-Time Lyapunov Exponents Approach, 2019, http://dx.doi.org/10.1007/978-3-030-28630-9.

[81] L. Arnold, Lyapunov exponents of nonlinear stochastic systems, in: F. Ziegler, G. Schuëller (Eds.), Nonlinear Stochastic Dynamic Engineering Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 1988, pp. 181–201.

[82] M. Shereshevsky, Lyapunov exponents for one-dimensional cellular automata, J. Nonlinear Sci. 2 (1992) 1–8.

[83] M. Courbage, B. Kaminski, Space-time directional Lyapunov exponents for cellular automata, J. Stat. Phys. 124 (2006) 1499–1509, http://dx.doi.org/10.1007/s10955-006-9172-1.

[84] P. Tisseur, Cellular automata and Lyapunov exponents, Nonlinearity 13 (5) (2000) 1547–1560, http://dx.doi.org/10.1088/0951-7715/13/5/308.

[85] F. Blanchard, P. Tisseur, Entropy rate of higher-dimensional cellular automata, ArXiv: Dyn. Syst. (2012).

[86] J. Baetens, J. Gravner, Introducing Lyapunov profiles of cellular automata, J. Cell. Autom. 13 (2015) 267–286.

[87] J. Baetens, B. De Baets, On the topological sensitivity of cellular automata, Chaos 21 (2) (2011) 023108, http://dx.doi.org/10.1063/1.3535581.

[88] S. Ninagawa, A. Adamatzky, Classifying elementary cellular automata using compressibility, diversity and sensitivity measures, Internat. J. Modern Phys. C 25 (2014) 1350098–1350109, http://dx.doi.org/10.1142/S0129183113500988.

[89] A. Adamatzky, J. Jones, On using compressibility to detect when slime mould completed computation, Complexity 21 (2015) 162–175, http://dx.doi.org/10.1002/cplx.21645.

[90] H. Zenil, F. Soler-Toscano, J. Delahaye, N. Gauvrit, Two-dimensional Kolmogorov complexity and an empirical validation of the coding theorem method by compressibility, PeerJ Comput. Sci. 1 (2015) e23, http://dx.doi.org/10.7717/peerj-cs.23.

[91] H. Zenil, Compression-based investigation of the dynamical properties of cellular automata and other systems, Comput. Res. Reposit. - CORR 19 (2009) 1–28, http://dx.doi.org/10.25088/ComplexSystems.19.1.1.

[92] P. Deutsch, RFC1951: DEFLATE Compressed Data Format Specification Version 1.3, Tech. Rep., RFC Editor, USA, 1996.

[93] H. Zenil, E. Villarreal-Zapata, Asymptotic behaviour and ratios of complexity in cellular automata, Int. J. Bifurcation Chaos (2013) 1350159.

[94] S. Ninagawa, Power spectral analysis of elementary cellular automata, Complex Syst. 17 (2008) 399–411.

[95] M.S. Keshner, 1/f noise, Proc. IEEE 70 (3) (1982) 212–218, http://dx.doi.org/10.1109/PROC.1982.12282.

[96] S. Ninagawa, M. Yoneda, S. Hirose, 1/f fluctuation in the "Game of Life", Physica D 118 (1998) 49–52.

[97] Y. Pomeau, P. Manneville, Intermittent transition to turbulence in dissipative dynamical systems, Comm. Math. Phys. 74 (1980) 189–197, http://dx.doi.org/10.1007/BF01197757.

[98] P. Manneville, Intermittency, self-similarity and 1/f spectrum in dissipative dynamical systems, J. Phys. 41 (11) (1980) 1235–1243, http://dx.doi.org/10.1051/jphys:0198000410110123500, URL https://hal.archives-ouvertes.fr/jpa-00208950.

[99] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: An explanation of the 1/f noise, Phys. Rev. Lett. 59 (1987) 381–384, http://dx.doi.org/10.1103/PhysRevLett.59.381.

[100] M. Hurley, Attractors in cellular automata, Ergodic Theory Dynam. Systems 10 (1) (1990) 131–140, http://dx.doi.org/10.1017/S0143385700005435.

[101] A. Ilachinski, Cellular Automata: A Discrete Universe, World Scientific, 2001.

[102] S. Strogatz, Nonlinear Dynamics and Chaos: With Applications To Physics, Biology, Chemistry and Engineering, Westview Press, 2000.

[103] K. Kaneko, Collapse of Tori and Genesis of Chaos in Dissipative Systems, World Scientific, USA, 1986.

[104] A. Wuensche, M. Lesser, M. Lesser, Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata, in: Santa Fe Institute studies in the sciences of complexity: Reference volumes, Avalon Publishing, 1992.

[105] A. Wuensche, Discrete dynamics lab, Artif. Life Models Softw. 1 (2009) 215, http://dx.doi.org/10.1007/978-1-84882-285-6_8.

[106] J. Hanson, J. Crutchfield, The attractor-basin portrait of a cellular automaton, J. Stat. Phys. 66 (1992) 1415–1462, http://dx.doi.org/10.1007/BF01054429.

[107] L. Milea, M. Teodorescu, V. Grosu, O. Oltu, Cellular Automata Encryption Method: Description, Evaluation and Tests, 2010, pp. 121–124.

[108] J. Soto, A. Wuensche, Logical universality from a minimal two-dimensional glider gun, Complex Syst. 27 (2018) 1–17, http://dx.doi.org/10.25088/ComplexSystems.27.1.1.

[109] A. Dennunzio, P.D. Lena, E. Formenti, L. Margara, Periodic orbits and dynamical complexity in cellular automata, Fund. Inform. 126 (2013) 183–199, http://dx.doi.org/10.3233/FI-2013-877.

[110] A. Dennunzio, E. Formenti, M. Weiss, Multidimensional cellular automata: Closing property, quasi-expansivity, and (un)decidability issues, Theoret. Comput. Sci. 516 (2013) 40–59, http://dx.doi.org/10.1016/j.tcs.2013.11.005.

[111] A. Dennunzio, P. di Lena, L. Margara, Strictly temporally periodic points in cellular automata, in: E. Formenti (Ed.), Proceedings 18th International Workshop on Cellular Automata and Discrete Complex Systems and 3rd International Symposium JournÉEs Automates Cellulaires, AUTOMATA & JAC 2012, la Marana, Corsica, September 19-21, 2012, in: EPTCS, vol. 90, 2012, pp. 225–235, http://dx.doi.org/10.4204/EPTCS.90.18.

[112] M. Schüle, R. Stoop, A full computation-relevant topological dynamics classification of elementary cellular automata, Chaos (Woodbury, N.Y.) 22 (2012) 043143, http://dx.doi.org/10.1063/1.4771662.

[113] L. Hurd, The Application of Formal Language Theory to the Dynamical Behavior of Cellular Automata, (Ph.D. thesis), Princeton University, USA, 1988, Order No. GAX88-19588.

[114] A. Smith, Two-Dimensional Formal Languages and Pattern Recognition by Cellular Automata, 1971, pp. 144–152, http://dx.doi.org/10.1109/SWAT.1971.29.
[115] M. Kutrib, A. Malcher, Cellular automata with sparse communication, 411, 2009, pp. 34–43, http://dx.doi.org/10.1007/978-3-642-02979-0_7,
[116] J. Lee, K. Imai, K. Morita, Simulation of one-dimensional cellular automata by uniquely parallel parsable grammars, Theoret. Comput. Sci. 304 (2003) 185–200, http://dx.doi.org/10.1016/S0304-3975(03)00083-5.
[117] N. Chomsky, Three models for the description of language, J. Symbolic Logic 23 (1) (1956) 71–72, http://dx.doi.org/10.2307/2964524.
[118] K. Culik, J. Pachl, S. Yu, On the limit sets of cellular automata, SIAM J. Comput. 18 (4) (1989) 831–842, http://dx.doi.org/10.1137/0218057.
[119] R. Gilman, Notes on cellular automata, Preprint (1988).
[120] H. Xie, The complexity of limit languages of cellular automata: An example, J. Syst. Sci. Complex. 14 (2001) 17–30.
[121] J. Zhisong, W. Yi, Complexity of limit language of the elementary cellular automaton of rule 22, Appl. Math. 20 (2005) 268–276.
[122] Z. Jiang, A complexity analysis of the elementary cellular automaton of rule 122, Chin. Sci. Bull. 46 (2001) 600–603.
[123] N. de Bruijn, A combinatorial problem, Proc. Sect. Sci. Koninklijke Nederlandse Akademie Van Wetenschappen Te Amsterdam 49 (7) (1946) 758–764.
[124] A. Church, Edward F. Moore. Gedanken-experiments on sequential machines. Automata studies , edited by C. E. Shannon and J. McCarthy, Annals of Mathematics studies no. 34, litho-printed, Princeton University Press, Princeton1956, pp. 129–153, J. Symbolic Logic 23 (1958) 60.
[125] F. Beckman, Mathematical Foundations of Programming, Addison-Wesley Longman Publishing Co., Inc., USA, 1980.
[126] H. McIntosh, Wolfram's class IV automata and a good life, Physica D 45 (1) (1990) 105–121, http://dx.doi.org/10.1016/0167-2789(90)90177-Q.
[127] L. Hurd, Formal language characterizations of cellular automaton limit sets, Complex Syst. 1 (1987) 69–80.
[128] J. Kari, Reversibility and surjectivity problems of cellular automata, J. Comput. System Sci. 48 (1) (1994) 149–182, http://dx.doi.org/10.1016/S0022-0000(05)80025-X, URL http://www.sciencedirect.com/science/article/pii/S002200000580025X.
[129] R. Péter, H. g. Rice. Classes of recursively enumerable sets and their decision problems, J. Symbolic Logic 19 (2) (1954) 121–122, http://dx.doi.org/10.2307/2268870.
[130] M. Delacourt, B.H. de Menibus, Characterisation of limit measures of higher-dimensional cellular automata, 2017, arXiv:1512.03696.
[131] P. Kurka, A. Maass, Limit sets of cellular automata associated to probability measures, J. Stat. Phys. 100 (2000) 1031–1047, http://dx.doi.org/10.1023/A:1018706923831.
[132] W. Li, Power spectra of regular languages and cellular automata, Complex Syst. 1 (2000) 107–130.
[133] N. Fatès, Experimental study of elementary cellular automata dynamics using the density parameter, in: Discrete Models for Complex Systems, 2003, pp. 155–166.
[134] K. Sutner, Cellular automata and intermediate reachability problems, Fund. Inform. 52 (2002) 249–256.
[135] F. Green, NP-Complete problems in cellular automata, Complex Syst. 1 (1987) 453–474.
[136] K. Sutner, The computational complexity of cellular automata, in: Fundamentals of Computation Theory, Springer Berlin Heidelberg, Berlin, Heidelberg, 1989, pp. 451–459.
[137] R. Lewin, Complexity: Life at the Edge of Chaos, Maxwell Macmillan International, 1993.
[138] B. Pfeifer, K.G. Kugler, M.M. Tejada, C. Baumgartner, M. Seger, M. Osl, M. Netzer, M. Handler, A. Dander, M. Wurz, A. Graber, B. Tilg, A cellular automaton framework for infectious disease spread simulation, Open Med. Inf. J. 2 (2008) 70–81.
[139] M. Markus, B. Hess, Isotropic cellular automaton for modelling excitable media, Nature 347 (1990) 56–58.