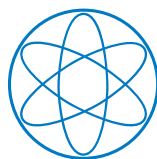# DEPARTMENT OF PHYSICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

PhD Thesis

Group of Complex Biophysics

Cellular Automata, from Theory to Applications in Biological Pattern Formation

Author: Mareike Elena Bojer
First examiner: Prof. Dr. Ulrich Gerland
Second examiner: Prof. Dr. Karen Alim
Submission Date: 05.04.2022

Fakultät für Physik
Technische Universität München

Group of Complex Biophysics

Cellular Automata, from Theory to Applications in Biological Pattern Formation

Mareike Elena Bojer

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
        Prof. Dr. Friedrich C. Simmel
Prüfende der Dissertation:
        1. Prof. Dr. Ulrich Gerland
        2. Prof. Dr. Karen Alim

Die Dissertation wurde am 05.04.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 19.07.2022 angenommen.

Abstract

The development of an organism is a collective effort of a growing number of cells, which subsequently form precise patterns. Cells coordinate by processing local interactions, and global cues to output changes in gene expression. However, the underlying mechanisms leading to robust patterning of tissue, despite the presence of noise inherent to biological processes, are only partially understood. The predominate modeling approach is using continuous differential equations. Motivated by the intrinsic cellularity of the system, we here explore cellular automata (CA) as a discrete dynamical framework. CA are discrete in a threefold sense: in state space, physical space, and time. They are capable of producing patterns of arbitrary complexity from simple local input signals, with simple update rules. Here, we show that an extended, stochastic cellular-automata framework is well suited for modeling pattern-formation processes, by studying three examples of increasing biological specificity.

We first aim for a solid background on cellular-automata theory (Part I) to then turn to modeling of biological processes (Part II). Therefore, we begin with a discussion of schemes to classify CA based on their dynamical behavior evolving from a single, an ensemble of, or all possible initial conditions. In this context, we investigate the relation between information-theoretic quantities characterizing the dynamical behavior of 2D CA rules.

In Part II, we focus on biologically inspired CA rules, and how they can mimic pattern formation in tissues: (1) For formation of a gene-expression boundary, we find that adding a local signal reduces boundary 'fuzziness', even if the additional signal is noisier than the global signal 'gradient'. (2) Forming the rows of bristles in Drosophila can be qualitatively reproduced in our minimal stochastic CA model, if we include the experimentally observed global prepattern to guide the self-organized dot formation. (3) For the intestine's stationary epithelial cell-type distribution, we turn away from a model with minimal number of parameters, and optimize cell-type update probabilities with experimental data of stationary cell-type distributions. This model 'learns' that cells do not 'de-differentiate', and is able to correctly infer the order of magnitude of timescales in the system. Common ingredients all three models require to qualitatively reproduce the — quite distinct — patterns are noise, and the combination of a local with a global signal. Noise breaks symmetries of our simplified models. The interplay of local and global signaling, however, is a consistent theme in developmental biology.

Although it turns out that we do not directly benefit from our theoretical study of cellular automata in Part I for the biology inspired models in Part II, we could at least sketch some ideas for the CA-classification question. Part II promotes an extended, stochastic CA modeling framework for developmental pattern formation that can generate hypotheses about the mode of cell-cell communication, range of signals, and timescales in the system.

Zusammenfassung

Die Entwicklung eines Organismus ist eine kollektive Anstrengung einer wachsenden Zahl von Zellen, die iterativ präzise Muster bilden. Zellen koordinieren sich, indem sie lokale Interaktionen und globale Signale verarbeiten, um Änderungen in der Genexpression zu steuern. Allerdings sind die zugrunde liegenden Mechanismen, die zu einer robusten Gewebemusterung führen, trotz des Vorhandenseins von Stochastizität, die biologischen Prozessen innewohnt, sind nur teilweise verstanden. Der vorherrschende Modellierungsansatz verwendet kontinuierliche Differentialgleichungen. Motiviert durch die intrinsische Zellularität des Systems untersuchen wir hier zelluläre Automaten (CA) als Klasse diskreter dynamischer Modelle. CA sind im dreifachen Sinne diskret: im Zustandsraum, im physikalischen Raum und in der Zeit. Sie sind in der Lage, aus einfachen, lokalen Eingangssignalen mit einfachen Regeln Mustern beliebiger Komplexität zu erzeugen. Hier zeigen wir, dass ein erweiterter, stochastischer CA Ansatz sich gut eignet zur Modellierung von Musterbildungsprozessen, indem wir drei Beispiele zunehmender biologischer Spezifität untersuchen.

Wir streben zunächst einen soliden Hintergrund zur Theorie zellulärer Automaten (Teil I) an, um uns dann der Modellierung biologischer Prozesse zu zuwenden (Teil II). Daher beginnen wir mit einer Diskussion der Schemata zur Klassifizierung von CA basierend auf ihrem dynamischen Verhalten, das sich entwickelt aus einer einzigen, einem Ensemble von, oder allen möglichen Anfangsbedingungen. In diesem Zusammenhang untersuchen wir Relationen zwischen informationstheoretischen Größen, die das dynamische Verhalten von 2D-CA-Regeln charakterisieren.

In Teil II konzentrieren wir uns auf biologisch inspirierte CA-Regeln und wie sie Musterbildung in Geweben nachbilden können: (1) Zur Bildung einer Genexpressionsgrenze finden wir, dass das Hinzufügen eines lokalen Signals die 'Unschärfe' der Grenze reduziert, selbst wenn das zusätzliche Signal verrauschter ist, als der globale Signal 'Gradient'. (2) Bildung der Borstenreihen bei Drosophila kann in unserem minimalen stochastischen CA-Modell qualitativ reproduziert werden, wenn wir das experimentell beobachtete globale Mustervorlage miteinbeziehen zur Koordination der selbstorganisierten Punktbildung. (3) Für die stationäre Epithelzelltypenverteilung des Darms wenden wir uns ab von Modellen mit einer minimalen Anzahl an Parametern und optimieren Zelltyp-Wahrscheinlichkeiten mit experimentellen Daten stationärer Zelltypverteilungen. Dieses Modell 'lernt' das Zellen nicht 'entdifferenzieren' und ist in der Lage, korrekt auf die Größenordnung von Zeitskalen im System zu schließen. Gemeinsame Zutaten, die alle drei Modelle benötigen um qualitativ die sehr verschiedenen Muster zu reproduzieren sind Stochastizität und die Kombination eines lokalen mit einem globalen Signal. Stochastizität bricht Symmetrien unserer vereinfachten Modelle. Das Zusammenspiel der lokalen und globalen Signale scheint hingegen ein durchgängiges Thema in der Entwicklungsbiologie zu sein. Obwohl sich herausstellt, dass wir nicht direkt von unserem theoretischen Studium der zellulären Automaten in Teil I für die von der Biologie inspirierten Modelle in Teil II profitieren, könnten wir zumindest einige Ideen für

die CA-Klassifizierungsfrage skizzieren. Teil II wirbt für einen erweiterten, stochastischen CA-Modellierungsansatz zur Bildung von Mustern in der Entwicklungsbiologie, der Hypothesen über die Art der Zell-Zell-Kommunikation, Reichweite von Signalen, und Zeitskalen im System generieren kann.

Contents

Symbols

CA	cellular automaton or cellular automata
ECA	elementary cellular automata, i.e. 1D and $k = 2$
c_ν	CA state at position ν
k	Number of different CA states
Σ	CA state space
C	CA-network state, or CA grid configuration
B	Subset of C, or block of cells
ϕ	CA update function (CA rule)
$\mathcal{N}_\nu$	neighborhood of position ν
r	range of a 1D CA
L	length of a 1D or 2D grid
λ	fraction of non-zero entries in the rule table (Langton's λ)
RTE	rule-table entropy
H_1	single-site entropy
$\langle H_1 \rangle$	time-averaged single-site entropy
Δ	difference-pattern size
$\langle \Delta \rangle$	ensemble-averaged difference-pattern size
$\sqrt{\Delta}/t$	speed of growth of $\sqrt{\langle \Delta \rangle}(t)$ for $t < \frac{L}{2}$
$t \to \infty \sqrt{\Delta}$	saturation value of $\sqrt{\langle \Delta \rangle}$
$I(x; y)$	mutual information of x and y
$I(c(t); c(t+1))$	local information conservation
KC	Kolmogroff complexity estimate
S	synergy
max, mean, median, min	maximum, mean, median, minimum of an observable
$\mathcal{P}$	cycle length
deg	in-degree
GoE	number of garden of Eden states
$\mathcal{T}$	transient length
$|BoA|$	attractor basin size
No. BoA	number of separate attractor basins in the field
ξ_ν	noise at position ν

1. Introduction and Outline

Every living organism consists of cells — and is thus 'discretized' by nature. Nevertheless, many pattern formation processes during organism development are modeled in a continuous framework, e.g., as reaction-diffusion systems introduced by Alan Turing (Turing 1952) and reviewed, e.g., in (Kondo and Miura 2010). A common drawback of the predominant modeling approach, tailoring differential equations to living systems, is the large number of parameters (Koopmans and Youk 2021). In this work, we explore cellular automata (CA) as a discrete modeling framework for pattern formation in biological tissue.

Cellular Automata (CA) are a class of discrete, dynamical systems that update each of their cells' states with respect to their neighboring cells. CA are discrete in a threefold sense — discrete in space (cellularized), discrete in time, and discrete in state space. In the 1940s, the concept of CA was first implied by John von Neumann studying self-replicating systems (Neumann 1966), and put into mathematical context by Stanislaw Ulam (Ulam 1952), working on crystal growth. It gained greater general popularity with John Conways 'Game of Life' in the 1970s (Gardner 1970), a very simple rule[1] that was shown to be capable of universal computation.

Successful examples of the cellular automata framework in development are modeling of labyrinthine patterned lizard skin scales (Manukyan, Montandon, Fofonjka, et al. 2017), liver lobule injury recovery (Adhyapok, Fu, Sluka, et al. 2021), nervous-tissue development (Lehotzky and Zupanc 2019), and insect eye photoreceptor patterning (Ebadi, Perry, Short, et al. 2018), further discussed in Sec. 3.1.

Cellular Automata definition A cellular automata (network) is defined on a d-dimensional grid, often of finite length L with periodic boundary conditions. Each 'cell' of the grid is assigned a state c_v in a finite state space $\Sigma = \{0, 1, \ldots, k-1\} \subset \mathbb{N}_0$, $c_v \in \Sigma$. The states of a grid are updated synchronously at discrete time steps, by a local update function, which is *identical* for all cells, called the cellular automaton rule ϕ_v,

$$\phi_v : \Sigma^{|\mathcal{N}(v)|} \to \Sigma$$
$$\phi_v\left(\{c_\mu(t)|\mu \in \mathcal{N}(v)\}\right) = c_v(t+1),$$

with $\mathcal{N}(v)$ denoting the neighborhood of cell v, including the cell itself. In one dimension, a range-r neighborhood includes all cells that are maximally r sites to left or right of cell v. On a 2D square grid, common choices are the Von-Neumann neighborhood, denoting the upper, lower, left, and right neighbor of a cell, or the Moore neighborhood, containing the Von

---

[1]On a 2D grid with a Moore neighborhood and two possible states - alive or dead, (i) any cell with less than two live neighbors dies of loneliness, (ii) any live cell with two or three neighbors remains alive, (iii) any live cell with more than three live neighbors dies of overpopulation and (iv) any dead cell with three live neighbors becomes a live cell, as if by reproduction.

Neumann neighborhood and extending it to the four diagonal cells. As the number of states and the neighborhood are finite, we can list all input configurations and the corresponding outputs in a table that uniquely specifies the CA rule, its *rule table*.

Common extensions to this notion of cellular automata are asynchronous updates (Fatès 2014; Nehaniv 2003), and probabilistic update rules as used in the second part of this thesis (Part II). More precisely, we work with two types of probabilistic (or stochastic) CAs: For the first type, the state of a cell $c_v^*(t)$ is a random variable, consisting of a deterministic state $c_v(t) \in \Sigma$ and additive noise $\xi_v(t)$. The update rule processes a realization to output a deterministic value $c_v(t+1) \in \Sigma$. For the second type, the states are in Σ, but the update rule assigns to each input a tuple of update probabilities, one probability for each state in Σ. The state at the next time step is drawn from this distribution. In Part II, we also extend the framework to accept one external space- and time-dependent input $x_v(t) \in X \subset \mathbb{R}$ such that

$$\phi_v^X : \Sigma^{|\mathcal{N}(v)|} \times X \to \Sigma$$
$$\phi_v^X \left(\{ c_\mu(t) | \mu \in \mathcal{N}(v) \}, x_v(t) \right) = c_v(t+1) .$$

Cellular automata are arguably the simplest class of systems that realizes the full spectrum of dynamical behavior — from approaching a simple fixed point, from any initial condition, up to universal computation (Wuensche 1998).

Outline In the first part of this thesis, we aim to characterize the space of potentially biologically interesting cellular automata. We are interested in general statements about what types of patterns can be generated by which types of rules, how the patterning process depends on initial conditions, and how many time steps the process takes. Therefore, we review and take further steps towards a classification of 2D cellular automata. This unsolved problem is more than 50 years old and has been strongly marked by Stephen Wolfram's work Vispoel, Daly, and Baetens 2021, introducing four *heuristic* classes based on 'typical' rule behavior from random initial conditions (Wolfram 1984a; Wolfram 2002). We divided the classification ideas into 'rule table based', 'local', and 'global' approaches, following a recent review (Vispoel, Daly, and Baetens 2021). 'Local' means with respect to a single or, here, an ensemble of, random initial conditions, whereas 'global' considers all initial conditions. We focus on information-theoretic quantities for 2D cellular automata, such as the entropy of a generated pattern, mutual information for local information conservation, difference-pattern spreading, and the amount of synergistic information processing between cells. For 1D cellular automata of fixed length, we study the whole basin-of-attraction field.

In the second part, we take the opposite approach, and explicitly focus on biologically in-spired cellular automata rules, and how they can aid pattern formation observed in tissues. We model patterning processes of increasing biological specificity, starting from gene-expression boundary formation, *Drosophila* bristle-prosecutor formation, and the stationary small-intestine epithelium cell-type distribution. All three examples have in common that using a *stochastic* cellular automata with extension to one long-range signal was a successful modeling strategy to robustly form the observed patterns. This might be surprising as those patterns are qualita-tively quite different — a sharp boundary, evenly spaced dots and the stationary but dynamic small intestine cell type distribution.

Part I.

First part: Characterization of Cellular Automata

2. Characterization Schemes of Cellular Automata

Our first approach towards gaining insight from cellular-automata modeling of biological pattern-formation phenomena is trying to characterize the space of potentially interesting cellular automata. Can we make some general statements about which patterns can be generated by which type of rules? How does the patterning process depend on initial conditions? How long does it take?

Most biological tissues are at least two dimensional thus we want to get a better understanding of the 2D cellular-automata rule space. A 2D cellular automaton in a Von-Neumann neighborhood is uniquely defined by its rule table listing the update value to each of the k^5 neighborhood configurations. For many biological systems, it can be argued that a cell treats information from its neighbors, e.g., in the form of signaling molecules, independent of their orientation. For cellular automata this translates to the constraint of *outer-totalistic* update functions. The cell's state at the next time step only depends on its own state and how often each possible state $c \in \Sigma$ is present among the neighboring cells. It does not matter, which orientation the neighboring cells have, e.g. if it is the upper left or the upper right neighbor that is in a particular state. Mostly, we consider outer-totalistic rules in this chapter. A constraint enforced throughout this chapter is the existence of a 'quiescent state'. A state $q \in \Sigma$ is *quiescent*, if a cell c_v with a neighborhood in state q, $\{c_\mu = q\}_{\mu \in \mathcal{N}(v)}$, is again mapped to state q in the next time step, $c_v(t+1) = \phi\left(\{c_\mu(t) = q\}_{\mu \in \mathcal{N}(v)}\right) = q$. We will often take state '0' to be quiescent. Rules that have a quiescent state and give the same result, when rotating or reflecting their neighborhood, are called *legal* rules. The dynamics generated by illegal rules is in general considered not to be 'physically meaningful' (C. G. Langton 1990). Of the 256 ECA, only 32 are legal, for instance. However, 2D rule spaces are in general extremely large, containing k^{k^5} rules, thus imposing constraints is common practice in order to reduce the number of rules.

A rules' behavior, meaning its 'typical' space-time traces, can *roughly* be characterized as evolving to (Class I) fixed, homogeneous states, (Class II) simple, separated periodic structures, (Class III) chaotic, aperiodic patterns or (Class IV) complex patterns of localized structures. These heuristic categories are called Wolfram Classes, after Stephen Wolfram, who classified all 256 1D two state ('elementary') cellular automata accordingly (Wolfram 1984a). However, this characterization is not exact it the sense that not every rule belongs to one particular class as the class of its behavior can be different for different initial initial conditions (Culik II and Yu 1988). As some cellular automata are capable of universal computation, there are initial conditions for which the actual behavior is undecidable (Halting problem). Also, Wolfram's heuristic classification procedure is not feasible for an arbitrary 2D k-state rule in practice, as it would require labeling time traces for a significant amount of initial conditions for a

time long enough to allow transition effects to die out. The concept of a 'transient' is not well defined, either — e.g., for complex rules the Halting problem implies, that we can never say if a rule has actually transitioned to its stationary state. Also, for chaotic rules, it is far from obvious to decide, if they have reached 'their typical dynamics' or not.

In this chapter, we want to explore different approaches to characterize 2D rules. Following a current review (Vispoel, Daly, and Baetens 2021), we structured those by, first, introducing rule-table based parameters Sec. 2.1, second, studying measures that are determined from simulations starting from a single up to few initial conditions of same type ('local parameters') Sec. 2.2 and, third, measures based on simulations from all possible initial conditions ('global parameters') Sec. 2.3. We will explore, among others, Langton's parameter in Sec. 2.1, time averaged single-site entropy, Kolmogoroff complexity, difference pattern spreading and local memory in Sec. 2.2, and a rule's capacity for synergistic information processing in Sec. 2.3. Our explorations are aimed to find parameters from the definition of a rule, or parameters that can be obtained by little computational effort, that allow for conclusions about either the rules 'class' or any of the information theoretic properties determining their dynamic behavior, such as entropy of the pattern, local information conservation and difference pattern spreading.

2.1. Rule-table based Parameters and derived rule-sampling Schemes

2.1.1. Langton's λ Parameter

At first, we want to discuss work by (C. G. Langton 1990) that inspired our explorations. Langton designed two related paths in rule space to construct a set of rules of k states that are parameterized by a single variable λ: a 'rule table walk through' and a 'random table'. He characterized legal rules along these paths with respect to their single-site entropy, memory and transition time.

λ denotes the fraction of neighborhood configurations that are not mapped to the quiescent state, commonly chosen to be '0', but are randomly assigned any of the remaining $k-1$ possible values. A rule with small λ is expected to evolve to a stationary pattern (Wolfram Class I behavior for most initial conditions), because nearly all neighborhood configurations map to the quiescent state (C. G. Langton 1990). For larger λ rules, non-trivial limit cycles occur. At some λ, different for each table-walk-through, the non-zero rule table entries 'percolate' and more complex patterns emerge. For $\lambda = 1 - \frac{1}{k}$ typically the most chaotic patterns are created. As λ increases from 0 to $\lambda = 1 - \frac{1}{k}$ there consequently is a transition between periodic and chaotic rules. This transition, however is not sharp, in the sense that there is no unique λ at which the transition happens for independent table-walk-throughs. It is hypothesized (H. Gutowitz and C. Langton 1995) that rules with complex behavior can be found at this transition regime: rules capable of universal computation need to allow for long transients, memory and information transmission. All those criteria have been shown to occur in the transition regime (C. G. Langton 1990).

Within the space of rules with quiescent states, a rule table-walk-through samples an ordered set of rules with roughly increasing 'complexity'. The first rule has only one non-zero entry (ignoring the quiescency condition for a moment), which is a number randomly chosen

among the remaining $k-1$ options. The probability of those $k-1$ options is taken to be uniform. For the second rule, another entry in the rule vector is randomly chosen and assigned a random state different from zero. We fill up the set of rules in this way until every neighborhood configuration (except the quiescent state) is mapped to a non-zero state. See Fig. 2.1 for an example in 1D with only two possible states for better illustration. Langton et al. give an example of typical sequence of behavior observed for a rule table walk through for $k = 4$ states and a neighborhood of 5 cells (including the central cell) (C. G. Langton 1990):

- $\lambda \approx 0$: The initial grid is mapped to a quiescent grid of zeros within one time step

- $\lambda \approx 0.1$: After 1-5 time steps, a homogeneous fixed point is reached

- $\lambda \approx 0.25$: Transient lengths increase, about three different outcomes can occur (depending on the initial condition): a homogeneous fixed state, a heterogeneous fixed state, or a periodic state

- $\lambda \approx 0.5$: Transients can be over 10'000 time steps, structures with a periodicity of more than 1000 time steps appear.

- $\lambda \approx 0.55$: The steady state is essentially transient behavior and long-time dynamics is chaotic

- $\lambda \approx 0.6$: Transients decrease with further increasing λ and long-time behavior is chaotic

- $\lambda \approx 0.75$: Behavior is essentially random after a single time step; State of maximal disorder.

Numerical simulations in (W. Li, Packard, and C. G. Langton 1990) generally suggest that (i) transients grow exponentially with λ near the transition from 'typically periodic' to 'typically chaotic' behavior and (ii) lattice size has a stronger influence on the typical behavior of a rule, the closer it is to λ of this transition-regime value.

For the random-table rule set, for each λ we randomly select a fraction λ of configurations (except the quiescent configuration). For each of those configurations, we uniformly choose a non-zero output, from the remaining $k-1$ states. By both procedures we obtain a set of rules with increasing λ. If not explicitly stated otherwise, we use the random-table method, as it does not introduce correlations between the rules.

In the limit of infinitely many states, a rule with λ close to one approaches an annihilated system, i.e., in every time step the next update for a cell is essentially given by drawing a random number, because the number of possible neighborhoods is so large that they don't repeat. Langton's λ can be seen as the zeroth order of a hierarchy of parametrization schemes, named 'local structure theory' and developed by Gutowitz (H. A. Gutowitz, Victor, and Knight 1987).

Rule-table Entropy

We here propose a modification of λ, the 'rule-table entropy' RTE, i.e., the entropy of different states within a rule vector, based on how often each state $c \in \{0, \dots, k-1\}$ occurs, instead of

Neighborh. config.s			λ = 1/8	1/4	3/8	1/2	5/8	2/3	7/8
c_i	c_{i-1}	c_{i+1}	rule vec.	rule vec.	rule vec.	rule vec.	rule vec.	rule vec.	rule vec.
0	0	0	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0	0	1	0	0	0	1	1	1	1
0	1	0	0	0	0	0	0	1	1
0	1	1	0	1	0	1	1	0	1
1	0	0	0	0	1	0	1	1	1
1	0	1	0	0	1	0	0	1	1
1	1	0	0	0	0	1	1	1	1
1	1	1	**1**	**1**	**1**	**1**	**1**	**1**	**1**

random-table method

Neighborh. config.s			λ = 1/8	1/4	3/8	1/2	5/8	2/3	7/8
c_i	c_{i-1}	c_{i+1}	rule vec.	rule vec.	rule vec.	rule vec.	rule vec.	rule vec.	rule vec.
0	0	0	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0	0	1	0	0	0	0	1	1	1
0	1	0	0	0	0	0	0	0	1
0	1	1	0	1	1	1	1	1	1
1	0	0	0	0	1	1	1	1	1
1	0	1	0	0	0	0	0	1	1
1	1	0	0	0	0	1	1	1	1
1	1	1	**1**	**1**	**1**	**1**	**1**	**1**	**1**

table-walk-through method

Figure 2.1.: Sketch of Langton's rule-sampling routines at the example of the 1D, two-state rules. Each rule is defined by its rule table, which consists of all possible neighborhood configurations and a rule vector. Each neighborhood configuration, comprising the central cell's state c_i, its left and right neighbor, c_{i-1} and c_{i+1}, is assigned an output. The tuple of all outputs forms the rule 'vector'. *Left:* Table-walk-through method: Start with a rule vector containing only zeros, except from the quiescent configurations (printed bold). It has the smallest possible fraction of non-zero sites, $\lambda = \frac{1}{\text{len(rule vector)}}$. For the next larger λ, randomly select one of the non-quiescent configurations and change the according rule-vector entry to a non-zero state, by uniformly drawing from the other $k-1$ options (here only '1'). *Right*: The random-table method: For any λ, randomly choose a corresponding fraction of non-quiescent zeros in the rule table, and change them to non-zero sites, again by uniformly drawing from the other $k-1$ options. Rules sampled by a random-table method are independent of each other, while rules in one table-walk-through are correlated.

the fraction of non-zero entries λ,

$$RTE(\phi) \equiv -\sum_{c=0}^{k-1} p_c(\phi) \log_k \left(p_c(\phi) \right) , \tag{2.1}$$

with p_c the fraction of rule vector entries of rule ϕ in state c. Rules of the same equivalence class have the same *RTE* value. This measure ensures that a rule with many, but equal non-zero values has a lower value than a rule with many, but different non-zero entries. For example, in case of $k = 8$ and $\lambda = 1 - \frac{1}{8}$ a rule generated with any of the two of Langton's sampling schemes is probably very chaotic, but it could in principle also contain only ones for 7/8th of the rule table, and with that be rather simple as suggested by its rule table entropy value. One example for $k = 3$, where the large λ value results from a single predominant non-zero state in the rule vector can be found in the Appendix, Fig. A.1. For a large number k of states, sampling the same state more often than the quiescent state is rare, thus in this case we will use λ or *RTE*, which ever better fits the question at hand.

2.1.2. Other Rule-table based Parameters

This short introduction of further rule-table based parameters is for a better overview and follows (Vispoel, Daly, and Baetens 2021).

Z-reverse parameter reflects how much we know about the preimage, when we know the current state. It is introduced by Wuensche (Wuensche 1998) in the context of reverse constructing precursor configurations of grid states. Each iteration of the reconstruction algorithm takes into account a larger part of a neighborhood configuration and therefore gives a better approximation of a rule's Z value. In contrast to λ, Z-reverse has the same value for rules of the same equivalence class, i.e. left-right or color interchanged versions of a rule (G. Oliveira, P. Oliveira, and Omar 2000).

μ sensitivity is defined as the average number of changes in the output of the update rule ϕ when changing the state c of a single site of the neighborhood $\mathcal{N}$ (G. Oliveira, P. Oliveira, and Omar 2000).

$$\mu := \frac{1}{|\mathcal{N}|k^{|\mathcal{N}|}} \sum_{\Sigma^{|\mathcal{N}|}} \sum_{\nu=1}^{|\mathcal{N}|} \frac{\partial \phi}{\partial c_\nu} \tag{2.2}$$

with $\frac{\partial \phi}{\partial c_\nu}$ the Boolean derivative of ϕ, and $\Sigma^{|\mathcal{N}|}$ denoting all neighborhood configurations. A low μ value indicates ordered behavior, while a high μ value suggests chaotic behavior, similar to Langton's λ and Wuensches Z-reverse parameter. However, μ is capable of separating ECA's with a periodic steady state more clearly from chaotic ECAs in comparison to Langton's λ (G. Oliveira, P. Oliveira, and Omar 2000).

2.2. Local Parameters

Having explored parameters based on the rule table, we now want to characterize the dynamical behavior of cellular automata. First, we want to do so for single initial conditions or an ensemble of initial conditions of the same type, mostly random initial conditions. The corresponding parameters are 'local' in this sense, in contrast to the 'global' parameters, studied in the next section, that depend on the full set of initial conditions.

2.2.1. Pattern Entropy as a Measure of CA Complexity

In order to characterize a CA generated pattern, we study the *B*-block entropy $H(B, t)$ of how often a block of cells *B* occurs in the pattern *C*,

$$H(B, t) := \frac{1}{|B|} \sum_{b \in \Sigma^{|B|}} P(B = b, t) \log_k(P(B = b, t)). \tag{2.3}$$

$P(B = b, t)$ denotes the frequency at which a specific block configuration $B = b$ occurs in the pattern *C* at time *t*, and Σ the set of possible states. In the simplest case, *B* consists of a single cell and we define
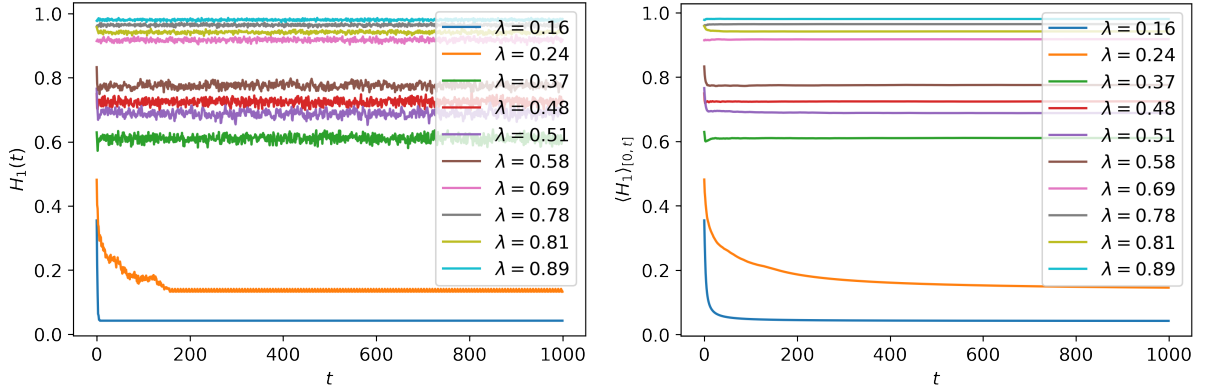
$$H_1 \equiv H(|B| = 1). \tag{2.4}$$

Figure 2.2.: *Left:* Single-site entropy time traces H_1 for increasing λs for outer-totalistic rules. *Right:* H_1 time traces averaged within the window $[1, t]$. For 1000 time steps — our default choice of simulation time — the entropies have converged for all tested λs. The $k = 8$ state rules are generated via the random-table method and run on a random initial square grid of length $L = 64$.
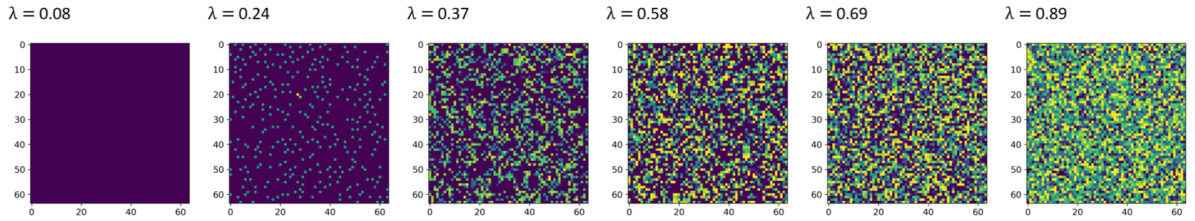

Figure 2.3.: Exemplary grid configurations at $t = 1000$ for some of the rules of Fig. 2.2.

Another natural choice would be the neighborhood form, e.g., the 5 Von-Neumann cells. A block can have any shape, such as three cells arranged in a 'L'.

We will first focus on H_1 and later compare to $|B| > 1$-site entropy results.

Convergence of Entropy in Time   When we use 'entropy' as an observable later in this section, we usually (implicitly) mean the time-average entropy, in order to approximate the long time behavior, after transients have settled. More precisely, we use entropy averaged over a fixed simulation time $t_{\max}$,

$$\langle H(B,t) \rangle \equiv \frac{1}{t_{\max}} \sum_{t=1}^{t_{\max}} H(B,t) \,. \tag{2.5}$$

Therefore, we want to start with explicitly looking at entropy time traces. For a parameter choice commonly used in this section (i.e. $k = 8$ states, $L^2 = 64^2$ cells, $t_{\max} = 1000$), the left plot of Fig. 2.2 shows the single-site entropy time traces for a set of different outer-totalistic rules with increasing λ. The right plot shows the time average of those entropies up to time t. We observe that the average entropy very quickly converges for all, but one rule. For this rule, with an intermediate λ of 0.24, it takes less than 500 time steps. This observation qualitatively matches our expectation, as Wolfram Class I or Class II behavior, common for small λ, and
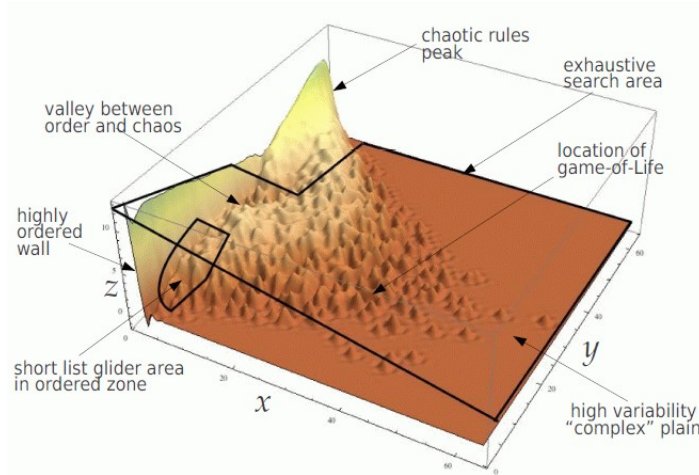
Figure 2.4.: Histogram of randomly sampled rules. On the y-axis the mean entropy, taken over a span of time-steps is plotted and on the y-axis the corresponding standard deviation. We see a smooth 'valley' between 'ordered' and 'chaotic' rules, rather than a sharp transition. Taken from Andy Wuensche, discrete dynamics lab *ddl* (Wuensche 2021).

chaotic behavior, common for large λ, are both characterized by short transients. However, for intermediate λ, there is a higher temporal variation in the time trace.

Wuensche suggested to use this qualitative difference in the variances to differentiate between complex and non-complex rules (Wuensche 1998). However, he could only find a continuous transition in variance between this different behavioral classes, see Fig. 2.4.

Time averaged single-site Entropy   Langton (C. G. Langton 1990) discusses the single-site entropy for a random-table (see Fig. 2.5) and table-walk-through (Fig. 2.6). For 'legal' general rules, Langton observed that $\langle H_1 \rangle$ has a maximum for $\lambda = 1 - \frac{1}{k}$, as the rule table becomes more regular again for higher λ values. This result is also confirmed in Fig. 2.5(*Left*), showing data for three different k (for outer-totalistic rules). If we assume that a rule acts on a random grid at reach time step, then we make a mean-field estimation (MF) of the single-site entropy:

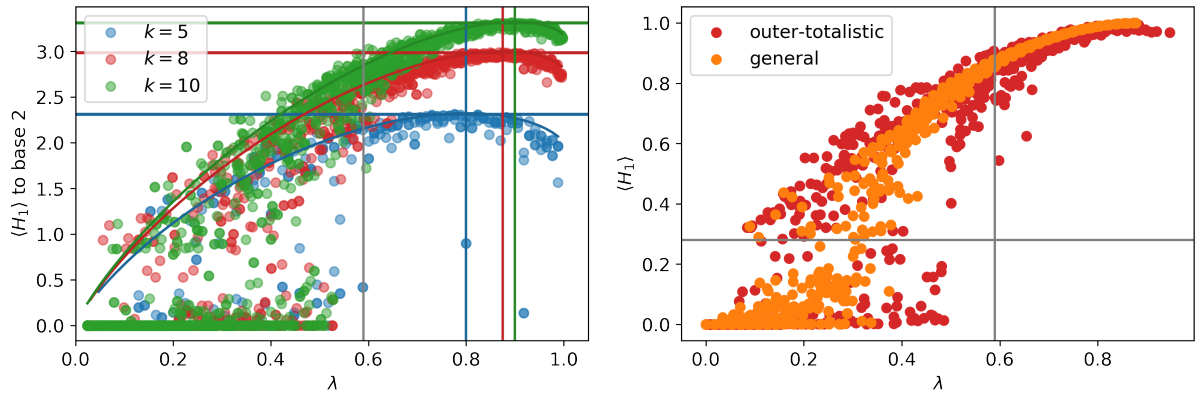$$\text{MF } H_1 = -(1-\lambda)\log_k(1-\lambda) - \lambda\log_k\left(\frac{\lambda}{(k-1)}\right) \tag{2.6}$$

Figure 2.5.: Time-averaged single-site entropies $\langle H_1 \rangle$ of random-table sampled rules starting from a random initial 64×64 grid, with time average taken over 1000 steps. *Left:* Entropies of outer-totalistic rules of $k = 5$, 8 or 10 states. Here, $\langle H_1 \rangle$ is to the base 2. Horizontal lines indicate the maximum entropy possible. Vertical lines mark $\lambda = 1 - \frac{1}{k}$, for which we expect the most chaotic rules. Indeed, those have maximal $\langle H_1 \rangle$. The gray line shows the percolation threshold of the Von-Neumann neighborhood, $\lambda_P = 0.59$. The colored curved lines are the mean-field approximations MF H_1 (Eq. 2.6). *Right:* In red we depicted outer-totalistic, in orange general $k = 8$ rules. General rules do take the relative location of their neighbors into account. The gray horizontal lines marks the entropy of one of the simplest chaotic rules. The vertical gray line corresponds to the percolation threshold. Almost all randomly sampled rules with $\lambda > \lambda_P$ are expected to have an entropy larger than the horizontal line.

For $k \to \infty$ this mean field estimate equals λ. As a CA rule in general does form patterns, we only expect this approximation to be relevant for chaotic rules. The curved lines in the Fig. 2.5 show this mean-field estimation for three different values of k. Langton further observed that there is a 'gap' of single-site entropy values below 0.28, see Fig. 2.5. This value corresponds to one of the simplest chaotic rules: The rule realizes only two states and essentially maps a neighborhood that contains mostly zeros and one non-zero state to that non-zero state. Moreover, observe that for a λ of 0.59 only very few $k = 5$ rules have a single-site entropy below 0.28. This value of λ equals the static-site percolation threshold for the Von-Neumann neighborhood. It is consistent that the non-zero rule-table entries contribute in the dynamics, starting from random initial grids, and not just die out, if the fraction of neighborhood configurations that are mapped to non-zero states, reaches this value.

In Fig. 2.5(*Right*) results for general and outer-totalistic rules are shown. Naively, we would expect that the spread within the outer-totalistic rules should be smaller than the spread within the general rules, as outer-totalistic rules are subset of general rules. However, we observe the opposite. The reason is Langton's sampling scheme: In fact, for sufficiently large k it is highly unprobable that a general rule, sampled with Langton's random-table method, is outer-totalistic. Also, the sampling scheme works best for large rule vectors, as a larger sample size can better represent the underlying distribution, i.e., the uniform distribution of the non-zero $k - 1$ states. The number of different input configurations, and consequently, the sample size, is much smaller for an outer-totalistic than for a (legal) general rule.

For rules generated by a rule table-walk-through, we show $\langle H_1 \rangle$ in Fig. 2.6. For small λ, rules have zero entropy. For some table-walk-throughs cyclic rules with a slightly higher entropy occur, probably with a 'non-trivial' periodicity of at least two time steps. A 'trivially' cyclic rule would have a periodicity of one, i.e. a fixed point. At a λ value different for each table-walk-through, we observe a sharp increase in entropy. The lower plot in Fig. 2.6 shows that rules have indeed change their behavior from cyclic (including the ones with trivial limit cycle) to non-cyclic (or 'chaotic') at this entropy jump. On the right, the same data is plotted against the rule-table entropy RTE. We observe that for high RTE values, $\langle H_1 \rangle$ traces converge to the RTE value.
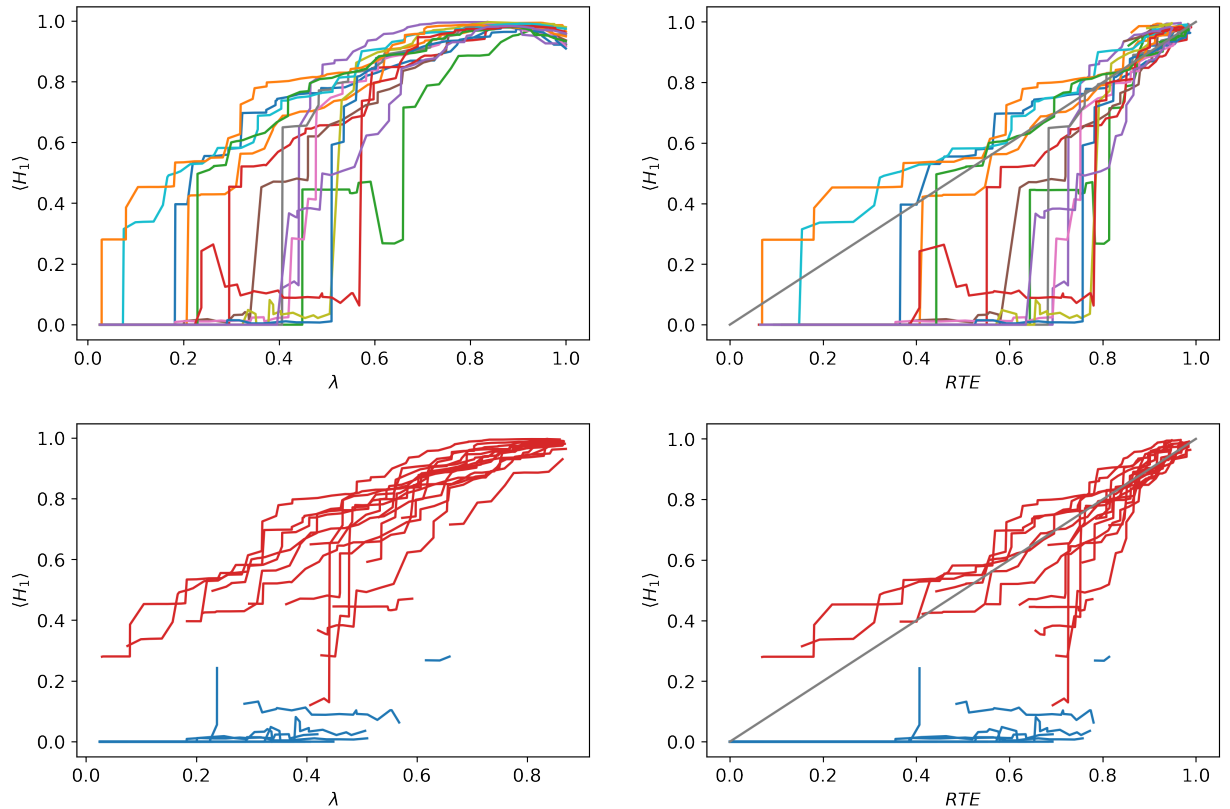
Figure 2.6.: Time-averaged single-site entropies $\langle H_1 \rangle$ for a set of 15 independently sampled rule table-walk-throughs. Each table-walk through comprises 63, $k = 8$-state outer-totalistic rules. Rules start from a random initial square grid of length $L = 64$. *Upper left:* Independent table-walk-throughs are shown in different colors. There is no unique λ where the jump in entropy occurs. *Lower left:* Same data, but colored in blue for rules that did reach a fixed point or limit cycle within the simulation time of 1000 time steps, and in red, for those that did not. The jump in entropy within each rule table-walk through seems correlates with the transition from cyclic to chaotic behavior. *Right:* Same data as the left plots, but plotted against *RTE*. We observe that for rules with large *RTE* rules converge to the diagonal, plotted in gray.

Block Expansion of Entropy

In order to better capture the spatial complexity of the pattern Olivier Martin[1] suggested to expand the single-site entropy to blocks of multiple cells in order to study the convergence of this quantity w.r.t increasing block sizes $|B|$ (October 2018, personal correspondence). We tested blocks of up to 13 cells, enumerated as follows: 'C' means the central cell, subsequent 0/1/2/3 refers to its upper, left, lower or right neighbor. Those are all cells in the Von Neumann neighborhood of 'C'. 'D' refers to cells diagonal to the central cell, i.e., a subsequent 0/1/2/3 means C's upper-right / lower-right / lower-left / upper-left diagonal neighbor. That covers all cells within in the square surrounding the central cell (Moore neighborhood). With 'L0123' we denote the upper, right, lower and left cell of this C0123D0123 square.

Fig. 2.7 depicts the mean B-block entropy deviations from the single-site entropy in dependence of *RTE* for different blocks. The larger the deviation of the block from the single-site entropy is, the higher are dependencies within cells of the block. If the deviation equals zero, there are no spatial correlations with the considered block configurations. Thus, we expect the blocks, consisting of the central cell and a direct neighbor, to show a higher deviation than the blocks, consisting of central cell and a diagonal neighbor. As the tested rules are outer-totalistic, we do not expect the orientation of cell towards each other to matter, e.g., results for block 'C0' should equal results for 'C1'. Surprisingly, we find that deviations from block to single-site entropy are small, not exceeding 2.5%. Examples of snapshots for three of the rules and according B-block frequency distributions are shown in the Appendix A.1. Subjectively, the patterns look much more interesting than what the small deviation of the B-block entropies from the single-site entropy would suggest. Consequently, we wonder, if the simulations have actually converged, as larger blocks require larger grid sizes for sufficient statistics. For blocks up to $|B| = 3$ convergences is demonstrated by blocks 'C0' and 'C2' as well as 'C02' and 'C03' agreeing with each other.
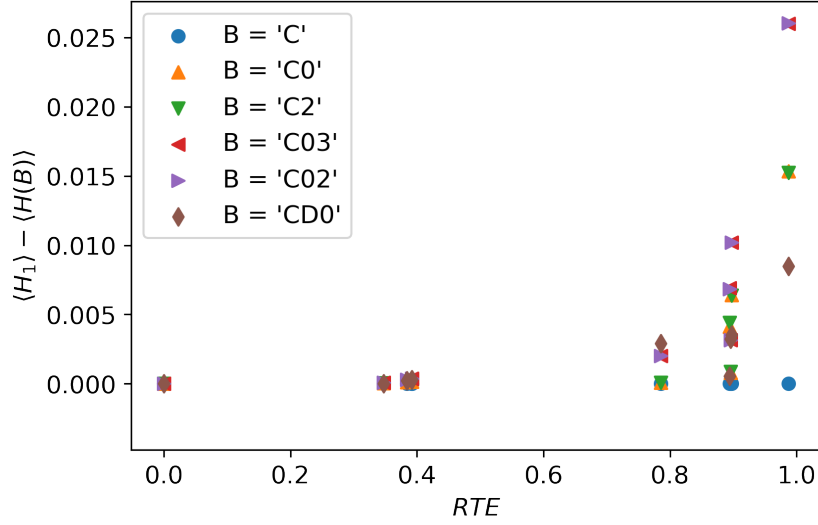

Figure 2.7.: On the y-axis we have the difference between the time-averaged single-site entropy from time-averaged entropies of larger blocks *B*. Different blocks are color-coded as follows: 'C': Central cell (blue), 'C0': Central cell and the one above (orange), 'C2' center and lower neighbor (green), 'CD0' center and lower right corner (brown), 'C03': Center, upper and right neighbor (red), 'C02': Center, upper and left neighbor (purple). Results are for outer-totalistic $k = 3$ rules sampled with Langton's random table-method for $t_{max} = 100$ and run on a square grid of length $L = 200$ from random initial conditions. We observe that simulations have converged as results for blocks equivalent under outer-totalistic rules coincide. Further, the maximal deviation from the block entropy to the corresponding single-site entropy is small, i.e. less than 3% of the single-site entropy.

To get a better idea of the order of magnitude of the grid size needed for convergence of the block entropies for larger block sizes, we studied $H(B)$ of the random initial grids, see Fig. 2.8. This can be seen as $H(B)$ generated by the 'identity rule', as the identity does not correlate sites and should yield the maximal entropy of one, when starting from random initial conditions. Consequently any deviation from one is due to a lack of convergence.

Olivier Martin (April to May 2019, email correspondence) pointed out the following estimation for this lack of convergence in dependence of the block size $|B|$, which we could confirm numerically in Fig. 2.8:

$$\langle H(|B|)_{\text{identity}} \rangle \approx 1 - \frac{1}{L^2} \frac{k^{|B|} - 1}{2|B| \log(k)} \tag{2.7}$$

The larger the block size $|B|$, the larger is the expected deviation. Moreover, it scales inversely with the number of cells in the grid. For the parameters used in Fig. 2.7, i.e. $k = 3$, $L = 200^2$, this implies the $|B| = 3$ entropy of the identity rule would show a deviation on the order of 10^{-4}. We also observe, that the smallest grid size such that in principle all different configurations could occur, $|B|^k$, is a good order of magnitude estimate for the minimum grid size needed. At about $|B|^k$ the deviation $1 - \langle H(B) \rangle$ changes to exponential decay, with respect to the number of cells in the grid.
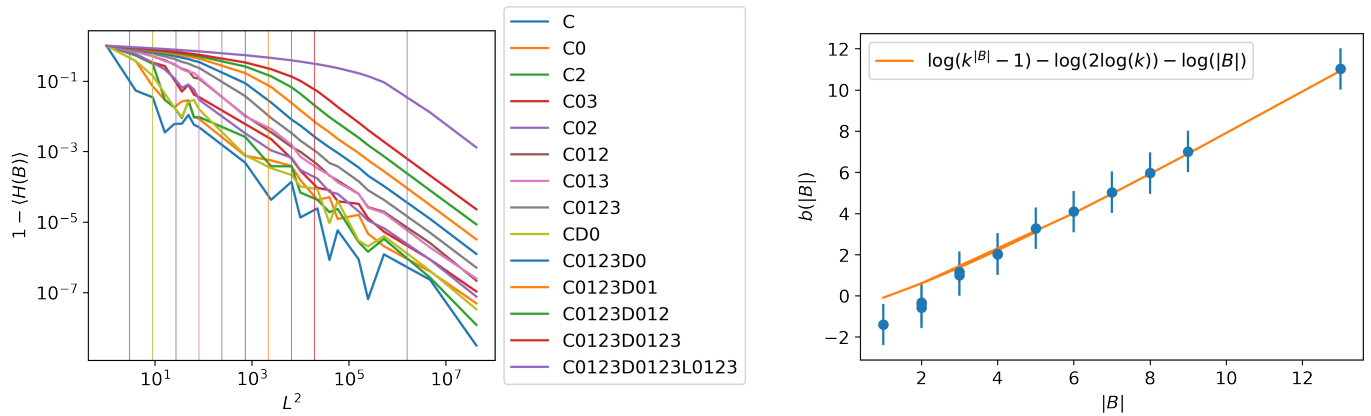
Figure 2.8.: *Left:* For the identity rule with $k = 3$ states, we show the deviation of the block entropy from the single-site entropy (equaling 1) for increasing random initial grids and depict different blocks in different colors: 'C' means the central cell, subsequent 0/1/2/3 refers to its upper, left, lower or right neighbor. Those are all cells in the Von Neumann neighborhood of 'C'. 'D' refers to cells diagonal to the central cell, i.e., a subsequent 0/1/2/3 means C's upper-right / lower-right / lower-left / upper-left diagonal neighbor. With 'L0123' we denote the upper, right, lower and left cell of this C0123D0123 square. Vertical lines mark $|B|^k$. *Right:* Blue dots depict the intercept $b(|B|)$ for different block sizes $|B|$, when performing a linear fit on the $\log(1 - \langle H(B) \rangle)$ curves in the left plot for sufficiently large L, thus $b(|B|) = L^2 \log(1 - \langle H(B) \rangle)$. The orange line shows the theoretical estimate Eq. 2.7.

2.2.2. Kolmogroff Complexity Approximation

The Kolmogoroff complexity is an abstract measure of the complexity of a string s, which, e.g., encodes the spatial pattern produced by a cellular automaton. More formally, given an arbitrary universal programming language, then the Kolmorgoroff Complexity KC of string s is the length of the shortest program that outputs s when run on a Turing machine plus an unknown constant (independent of s and specific to the language) (Kolmogorov 1963). It can be shown that KC is not a computable function, i.e. there is no program that takes a string s and returns $KC(s)$ as output, only upper bounds can be provided (Vitányi 2020). The probably most straightforward version of computing an upper bound to a string s is to use a lossless compression algorithm.

Hector Zenil (Zenil 2010) characterized the ECA by measuring the compression ratio of a pattern's compressed size and its original size, averaged over time. Inspired by his work, we compress the 2D patterns generated by randomly sampled outer-totalistic rules by a lossless compression algorithm, zip. In Fig. 2.9 we find though, that the result of compressing each grid at a particular time step, and averaging the results over time, yields qualitatively the same result as $\langle H_1 \rangle$. One problem could be our choice of compression algorithm as it is designed for strings, i.e. 1D arrays, not 2D pictures as our grids. Zenil has further put forth related approaches to classify elementary cellular automata with Kolmogorov complexity, that we have not tested for our 2D problem. He introduced a characteristic exponent similar in the spirit to a Lyapunov coefficient for cellular automata, but with respect to variations of the compressed patterns (Zenil 2010), (Zenil and Villarreal-Zapata 2013), and a block decomposition approach of the 2D space time pattern (Zenil, Soler-Toscano, Delahaye, and Gauvrit 2015).

However, for any fixed choice of compression algorithm, the question remains, if we are actually measuring some sort of complexity or rather testing the performance of the concrete algorithm.
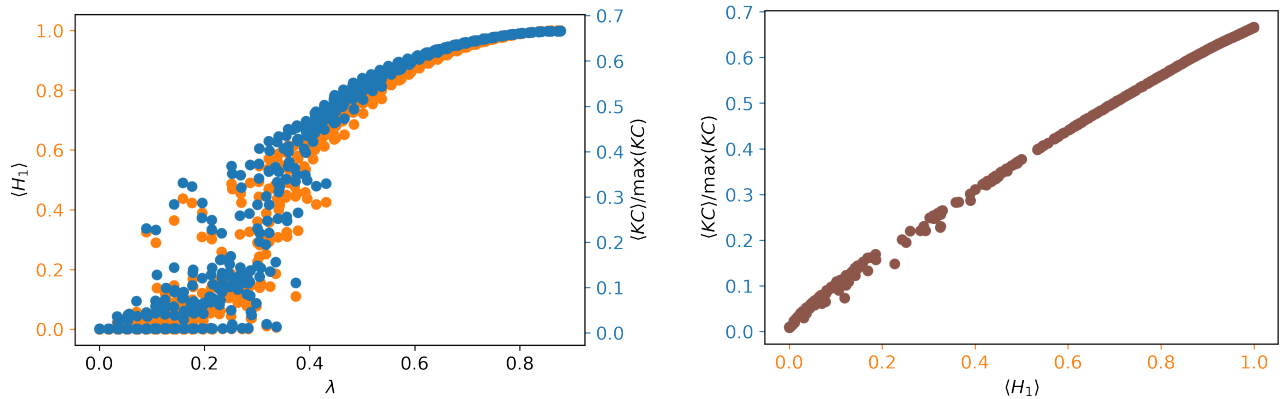
Figure 2.9.: For 2D $k = 8$ state CAs sampled with Langton's random-table method starting from a random initial square grid of length $L = 64$: *Left*: We show the time-averaged single-site entropy $\langle H_1 \rangle$ in orange and the time-averaged, normalized, Kolmogoroff-complexity estimate in blue. max KC is taken to be KC of an 64×64 grid with cell states uniformly drawn from $\{0, \ldots, k-1\}$. *Right*: Shows the Kolmogoroff complexity estimate KC plotted versus the single-site entropy of the same data. Both quantities are averaged over 1000 time steps. The measures appear to be highly correlated.

2.2.3. Difference-Pattern Spreading

Another way to classify cellular automata is with respect to their stability under small perturbations of initial conditions. Generally, this is done by examining the dynamics of difference spreading patterns in analogy to Lyapunov exponents in continuous-state dynamical systems. However, in contrast to continuous-state systems, there is no notion of closeness for initial conditions. Different approaches exist, put forth in (W. Li, Packard, and C. G. Langton 1990), (Wolfram 1984a), (Bagnoli, Rechtman, and Ruffo 1992) and (Zenil 2010). Bagnoli et al. determined the number of ways in which a defect can propagate to a certain time step. Essentially, each defect is propagated in a different replica of the simulation, such that defects cannot annihilate. Consequently, the number of replicas can grow exponentially in time. Other approaches are based on the number of defect cells at a particular time step, which can increase at most linearly for 1D CA. Hector Zenil used the compressed difference between patterns at each time step, generated by 'neighboring' initial condition pairs, and examined the slope in time for the ECA (Zenil 2010). 'Neighboring' in this setting is with respect to one site difference in the initial grid's 'Gray Codes'[2].

The definition of difference-pattern spreading by Li et al. (W. Li, Packard, and C. G. Langton 1990) is for range r cellular automata and essentially measures the speed γ of the *fastest* spread of a defect (different from our definition below). They study the speed at which the outmost non-zero state of the difference pattern moves from its initial position towards the boundaries

---

[2]Gray Code is an ordering of the binary system such that two successive values differ in only one binary digit. E.g., the two binary initial grids representing decimal values '1' and '2' in normal binary code are '0...001' and '0...010', thus their Hamming distance equals two sites. In Gray code, '1' and '2' are represented by '0...001' and '0...011', differing only in a single site.

of the grid. Li, Packard and Langton (W. Li, Packard, and C. G. Langton 1990) estimated an average (w.r.t. initial conditions) value of the difference-pattern spreading rate γ in terms of λ for 1D $k = 2$ range r rules: For a two-state rule, a fraction of λ of the $2r + 1$ neighborhood configurations map to 1 and the other $1 - \lambda$ configurations to 0. If we randomly choose two $2r + 1$-sized blocks from the set of all possible blocks, the probability that they map to the same symbol is

$$P_{\text{same}} = \lambda^2 + (1 - \lambda)^2 \,.$$

The probability that the left spreading rate γ_{left} equals its upper bound r is given by the probability, that the blocks $(c_{-r}, c_{-(r-1)}, \ldots, c_{r-1}, 1)$ and $(c_{-r}, c_{-(r-1)}, \ldots, c_{r-1}, 0)$ map to different symbols, $1 - P_{\text{same}}$, in the first time step. Implicitly, Li et al. (W. Li, Packard, and C. G. Langton 1990) make the assumption that the blocks at the outmost edge of the difference spread are drawn randomly from the set of all possible blocks. However, for a rule with small λ, most parts of the grid will be in state 0, not random, unlike the initial condition. Consequently, we expect the estimation to fail for small λ. On the contrary, if the rule table rather is random, assuming a uniformly random grid for $t > 1$ might be a useful approximation. In this line of argument, the probability that the spreading rate to the left equals $r - 1$ is given by the probability that the blocks stated above map to the same symbol, multiplied with the probability that blocks $(c_{-r}, c_{-(r-1)}, \ldots, 1, c_{r-1})$ and $(c_{-r}, c_{-(r-1)}, \ldots, 0, c_{r-1})$ map to different symbols. To summarize:

$$P(\gamma_{\text{left}} > r) = 0$$
$$P(\gamma_{\text{left}} = r) = 1 - P_{\text{same}}$$
$$P(\gamma_{\text{left}} = r - i) = (r - i)P_{\text{same}}^i(1 - P_{\text{same}})$$
$$P(\gamma_{\text{left}} = 0) = 0 \,.$$

With that, the average left spreading rate is

$$\gamma_{\text{left}} = \sum_{i=0}^{r}(r - i)P_{\text{same}}^i(1 - P_{\text{same}}) \tag{2.8}$$

$$= r - \frac{P_{\text{same}}}{1 - P_{\text{same}}} + \frac{P_{\text{same}}^{r+1}}{1 - P_{\text{same}}} \tag{2.9}$$

$$= r - \frac{1}{2}\frac{2\lambda^2 - 2\lambda + 1}{\lambda(1 - \lambda)} + \frac{1}{2}\frac{(2\lambda^2 - 2\lambda + 1)^{r+1}}{\lambda(1 - \lambda)} \,, \tag{2.10}$$

where the last term can be dropped for larger r (in this case the result matches (W. Li, Packard, and C. G. Langton 1990)). For example, for a $\lambda = 0.5$ ECA, i.e., one that is expected to be about as chaotic as possible, we obtain $\gamma_{\text{left}} = 1 - 1 + 0.5 = 0.5$. This result is reasonable, as for perfect chaos we expect the outmost cell to be in the same state with probability 0.5. Further, for a k-state $r = 1$ cellular automata it follows that

$$\gamma_{\text{left}} = 1 - P_{\text{same}} \,,$$

meaning that the left-difference spreading rate is equal to the probability of sampling cells in a different state.

We can extend the work by Li et al. (W. Li, Packard, and C. G. Langton 1990) to a *k* state cellular automata as follows: Essentially, we have to modify the expression of the probability of evolving two configurations to the same state,

$$P_{\text{same}} = \left(\frac{\lambda}{k-1} \right)^2 + (1 - \lambda)^2$$

and substitute it in Eq. 2.9. Note that this expression, and also the remaining derivation, should hold true for the 2D outer-totalistic cellular automata, if we consider spread only in one direction of the grid. Moreover, for large *k*, allowing for a better sampling of the rules, it follows that $P_{\text{same}} \to 0$. Thus we can expand Eq. 2.9 and obtain

$$\gamma_{\text{left}} \approx r - 1 + 2\lambda - \lambda^2 \,. \tag{2.11}$$

The version of difference-pattern spreading-rate used for 2D outer-totalistic CA in the following is inspired by Stephen Wolfram's approach (Wolfram 1984a).

For the 2D CA we apply a simplified procedure: Run each rule twice, for the second run, change a single site and measure the difference-pattern size Δ. We define Δ as the Hamming distance between the pattern of the primary run and the defect run. We average Δ over an ensemble of independent initial conditions, $\langle \Delta \rangle$ and study $\sqrt{\langle \Delta \rangle}$s time trace (Fig. 2.10). From Fig. 2.10 we see that the difference-pattern curves $\sqrt{\langle \Delta \rangle}$ can be characterized by their initial slope, $\sqrt{\Delta}/t$, and their final saturation value, $t \to \infty \sqrt{\Delta}$ (note, that we omitted the ensemble average in the notations for conciseness).
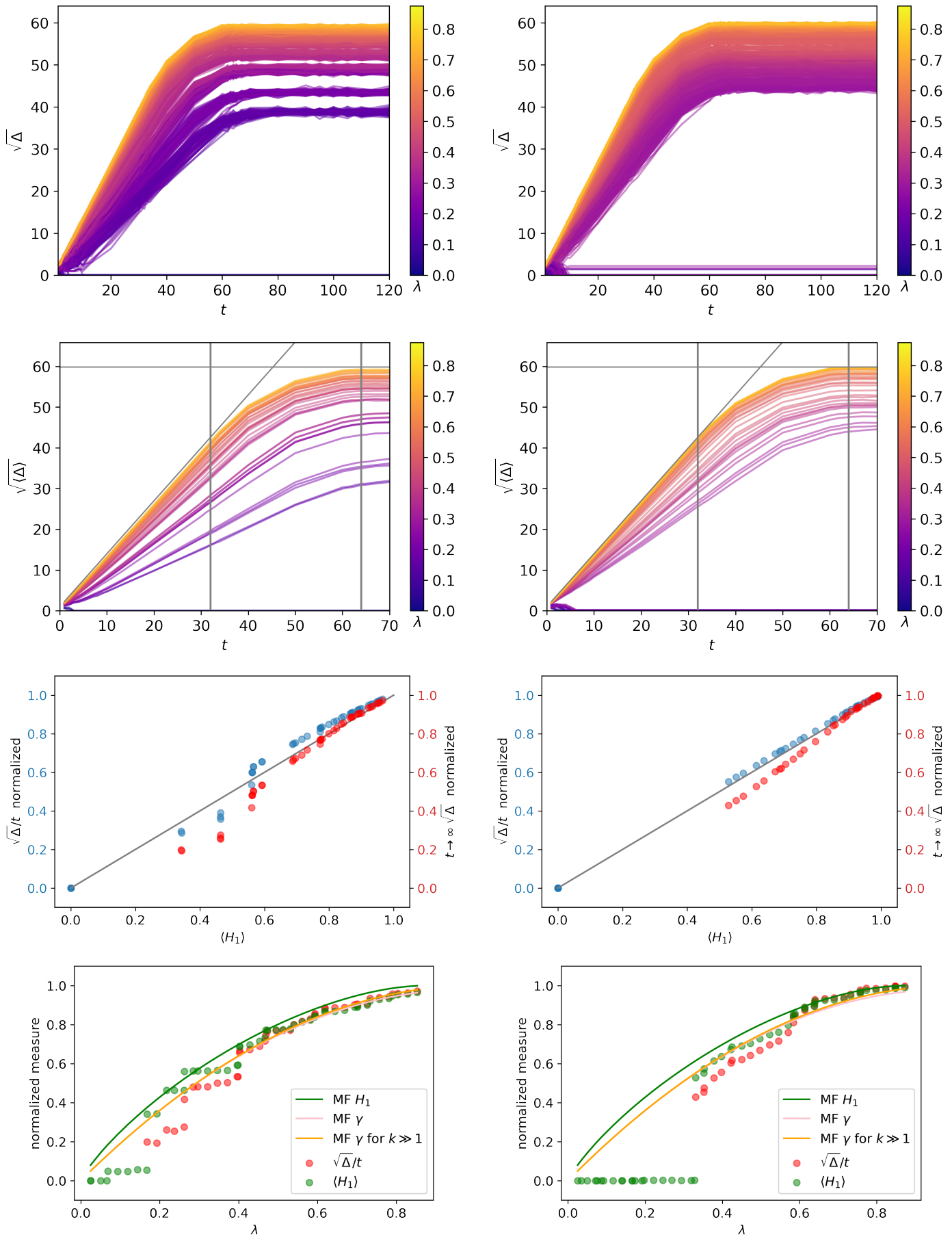
Figure 2.10.: Difference-spreading measures for the first two of four independent table-walk-throughs (each in a separate column) with 55 outer-totalistic $k = 8$ rules of increasing λ. Each rule is run twice, starting from one of 50 independent random initial square grids of length $L = 64$, with the second run initially differing in a single site. The first row shows the square root of the size of the difference pattern $\sqrt{\Delta}$, measured as the number of differing sites, over time, with λ color-coded. The second row shows the ensemble average thereof, including analytical estimates of a random rules slope (≈ 1.3), saturation value ($\frac{7}{8}L^2$), start ($L/2$) and end (L) of transition region between growth and saturation, as gray lines.
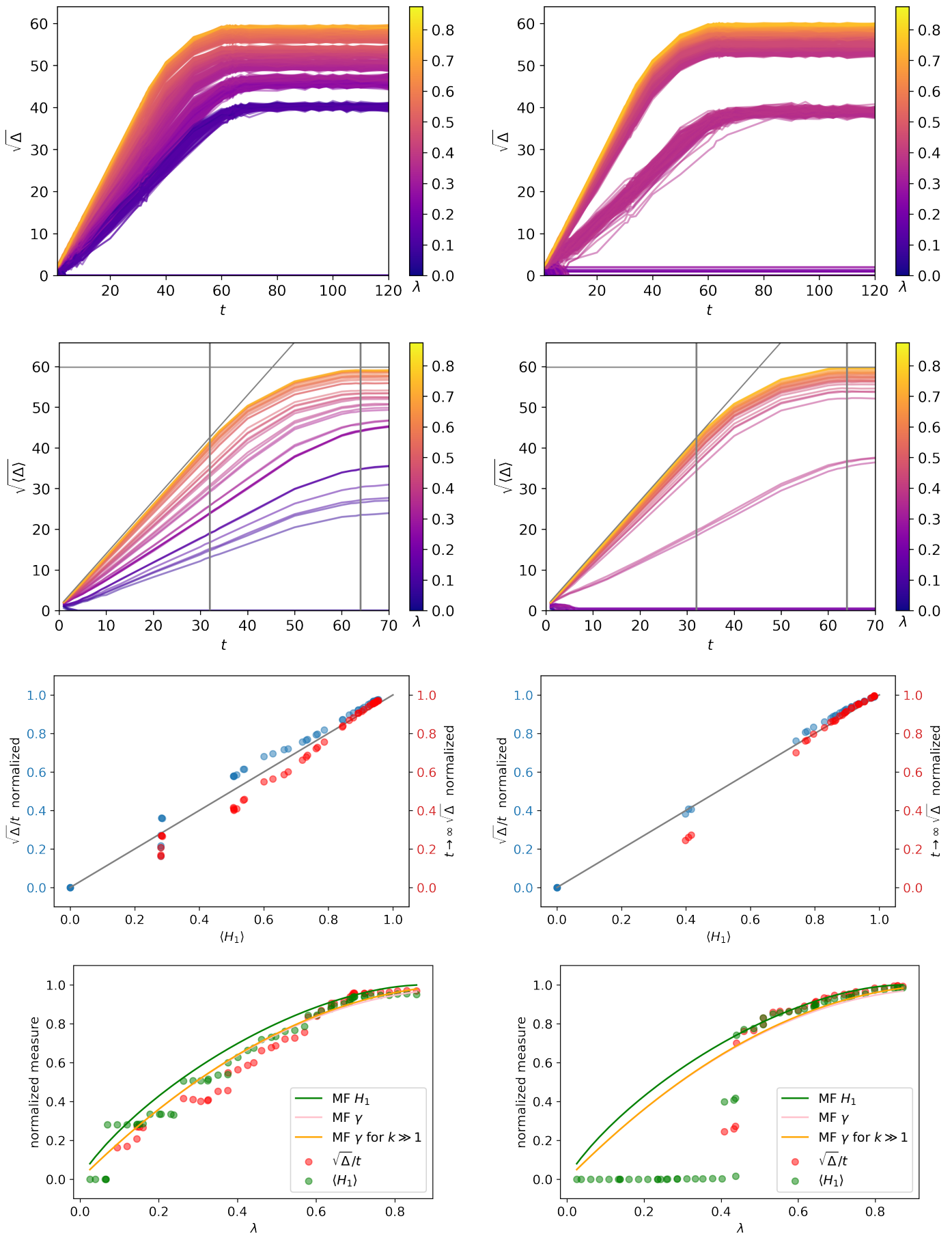
Figure 2.10.: Difference-spreading measures for the last two of four independent table-walk-throughs. To each rule's $\sqrt{\langle \Delta \rangle}$ curve, we fitted a line in the initial linear regime to obtain its difference-pattern spreading velocity $\sqrt{\Delta}/t$ and a second line in the final, constant regime for its difference-pattern saturation value $t \to \infty \sqrt{\Delta}$. The third row shows $t \to \infty \sqrt{\Delta}$ and $\sqrt{\Delta}/t$ of the ensembles with respect to the measured single-site entropies $\langle H_1 \rangle$, all rescaled by their maximal value. In the last row, the rescaled $\sqrt{\Delta}/t$ and $\langle H_1 \rangle$ values are plotted for direct comparison with their corresponding mean-field estimates.

As an example, and for proper rescaling of our results, let us consider what we expect for a 'random' rule — i.e. a stochastic rule that at each time step randomly, with respect to a uniform distribution, updates each cell state independent of its or its neighbors previous state (annealed limit). To calculate $\sqrt{\Delta}/t$ and $t \to \infty \sqrt{\Delta}$, we need the maximal area, that is affected at time t, times the probability, that a site has changed its state. The region at time t that can be maximally affected by a single-site change in the initial grid is given by $1 + \sum_{i=1}^{t} 4i = 2t^2 + 2t + 1$, with $+1$ due to the initial single defect. The probability for a site to change state under the random rule is given by $\frac{k-1}{k}$, with k the number of states a cell can have. With that, the expected difference spreading speed $\sqrt{\Delta}/t$ can be approximated to be

$$\sqrt{\Delta}/t \approx \frac{\sqrt{\frac{2t^2(k-1)}{k}}}{t} \approx 1.3 \tag{2.12}$$

for our $k = 8$ state space. A line with this slope is plotted in Fig. 2.10 and yields a tight upper bound. Moreover, the saturation value of the difference pattern $t \to \infty \sqrt{\Delta}$ on an $L \times L$ lattice is expected to equal $\frac{k-1}{k}L^2$ sites, which corresponds to the horizontal gray line in Fig. 2.10, second column. The two vertical gray lines mark the expected time, the defect spread hits the horizontal/vertical boundary (at $t = \frac{L}{2}$) and the corner (at $t = L$), respectively. These lines agree well with the qualitative change of the difference pattern size from linear to slowing down and from slowing down to saturated. Due to the tightness of the upper bound, we can conclude that a deterministic cellular automata rule with maximal λ behaves similarly to a stochastic random rule with regard to difference-pattern spreading.

How do $\sqrt{\Delta}/t$ and $t \to \infty \sqrt{\Delta}$, our measures of the sensitivity of a CA rule to small changes in initial conditions, relate to the single-site entropy $\langle H_1 \rangle$ of this CA? Fig. 2.10, second column, suggests that both measures are roughly proportional to $\langle H_1 \rangle$. In the last row of Fig. 2.10, we show our rescaled estimate Eq. 2.11 of the mean-field difference-spreading rate MF γ, together with the entropy estimate, Eq. 2.6. As γ refers to the maximum spread, it is rescaled by the expected fraction of sites in the 'wrong' state, $\frac{k-1}{k}$. We observe that the rescaled estimates MF γ and MF H_1 can actually describe the rescaled difference-pattern slopes $\sqrt{\Delta}/t$ (rescaled by their maximum value given by the stochastic rule derived before) and $\langle H_1 \rangle$ well for sufficiently large λ. The similarity of these two estimates MF γ and MF H_1 explains, why the difference-spreading rate $\sqrt{\Delta}/t$ seems to be equivalent to $\langle H_1 \rangle$. Can we analytically find an relationship between $\langle H_1 \rangle$ and $\sqrt{\Delta}/t$ by comparing their mean-field estimates? The mean-field curves for different number of states k seem to approach each other for increasing k up to an intermediate value, on the order of $k = 10$, see Fig. 2.11. However, for very large k, e.g., $k = 100$, they diverge again, and for $k = 1000$, we can see their $\lim k \to \infty$ limit, $\gamma \to 2\lambda - \lambda^2$ (Eq. 2.11 and MF $\langle H_1 \rangle \to \lambda$). Thus, single-site entropy and difference-pattern spreading rate are not equivalent, but for many practical purposes — i.e., intermediate values of k — results are so similar, that we don't consider them as different characteristics of a rule.

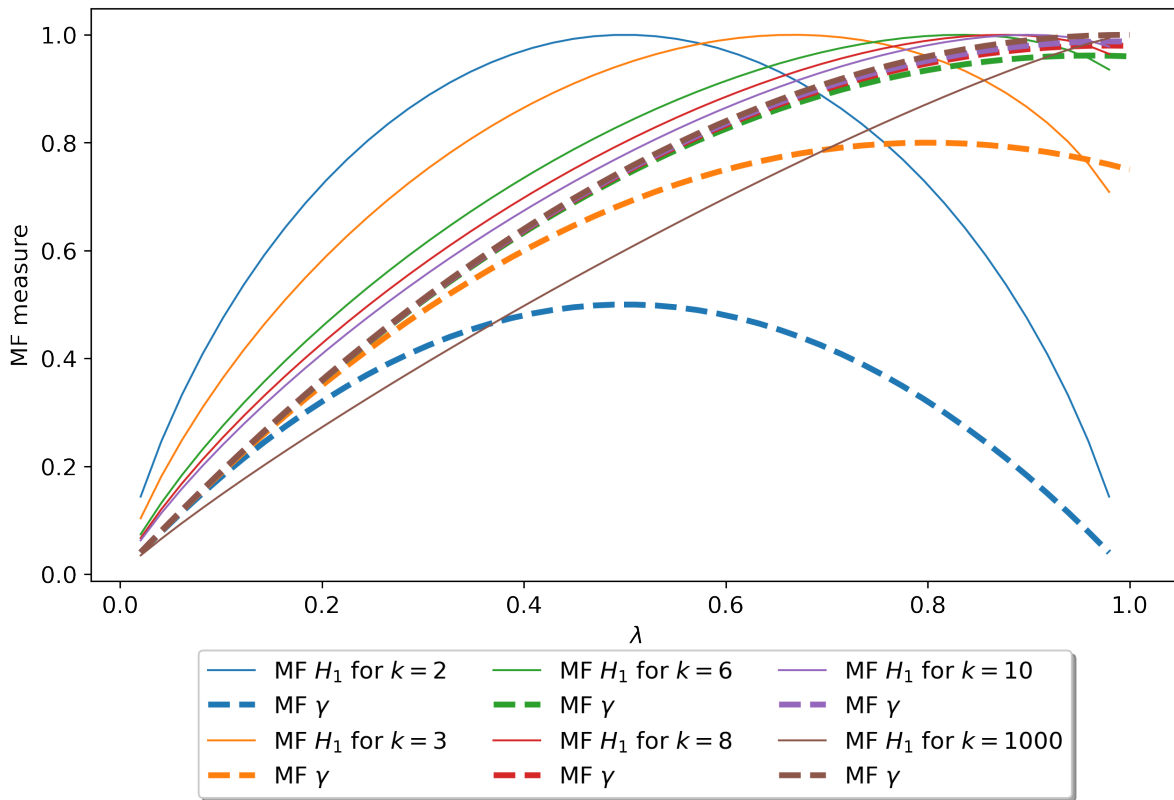## 2.2.4. Local Memory as a Measure of CA Complexity

Figure 2.11.: Comparison of the mean-field estimates for single-site entropy MF $\langle H_1 \rangle$ and difference spreading rate MF γ for different k values. We find that both measures agree rather well only for larger number of states, such as $k = 8$. For $k \to \infty$ MF $\langle H_1 \rangle$ approaches λ.

Another information theoretic measure to quantify a rule's complexity, already studied by Langton (C. G. Langton 1990), is its memory in the sense of local information conservation. It is given as the mutual information of a cell c at time t and its future state at $t + 1$. As with entropy, we will use time averaged local memory,

$$\langle I(c(t); c(t+1)) \rangle = \langle H(c(t+1)) - H(c(t+1)|c(t)) \rangle_t, \qquad (2.13)$$

with H denoting entropy of the according probability distributions. The probabilities for each state, or each tuple of states at time t and time $t + 1$, are calculated from the frequencies of both states within a grid at time t and $t + 1$. For rules with very small λ that produce a fixed pattern, the mutual information equals zero. We expect that a highly chaotic rule has low mutual information, as the state of a cell at the next time step is nearly independent of its state at t in the limit of infinitely many states $k \to \infty$. For a general rule, this is what we find, see Fig. 2.12 in agreement with (C. G. Langton 1990).

However, for outer-totalistic rules, local information conservation does hardly decrease for chaotic rules with increasing λ. Generally local memory is much higher than for typical general rules, see Fig. 2.12. There seem to be too many dependencies induced in the outer-totalistic rule table such that mutual information results of a cell and its future state are not exactly meaningful. Also, there is no clear distinction between cyclic and non-cyclic rules, as in case of general rules studied in (C. G. Langton 1990) as shown in Fig. 2.13.
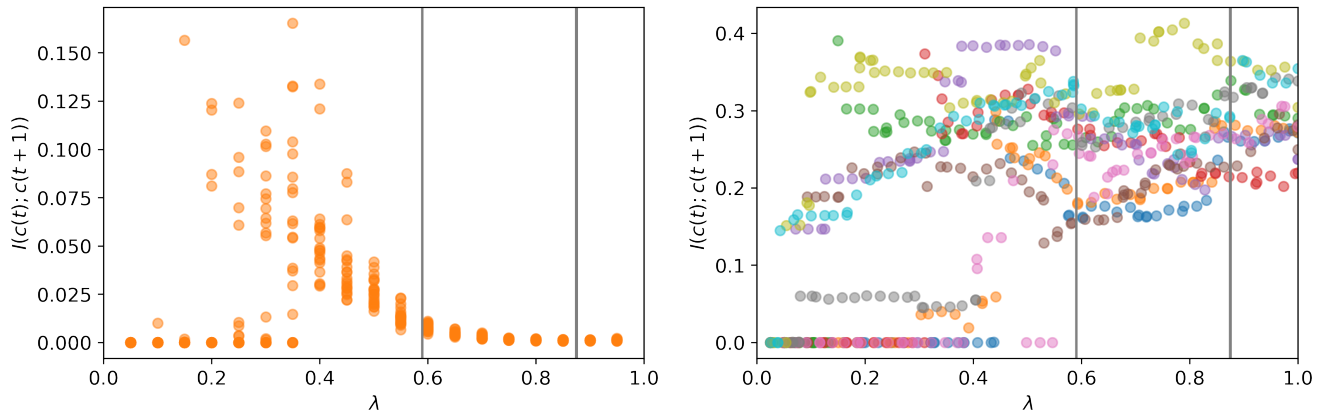
Figure 2.12.: Local information conservation in terms of time averaged mutual information $\langle I(c(t); c(t+1)) \rangle$. *Left:* Shown for general $k = 8$ state rules sampled with Langton's random-table method, and run on a $L = 64$ square grid from random initial conditions for 1000 time steps. The gray lines mark the percolation threshold $\lambda_P = 0.59$ and $1 - \frac{1}{k}$, for reference. We expect rules fixed point rules to have zero $\langle I(c(t); c(t+1)) \rangle$, and small λ, and equivalently for chaotic rules, which have a large λ. *Right:* Shown is $\langle I(c(t); c(t+1)) \rangle$ for outer-totalistic rules, sampled in 10 independent rule table-walk-throughs (different colors). The outer-totalistic $k = 8$ rules are run from random initial conditions for 1000 time steps on a $L = 64$ square grid. Apparently, temporal correlation for outer-totalistic rules are higher than for typical general rules, also they do not decay for more chaotic rules.
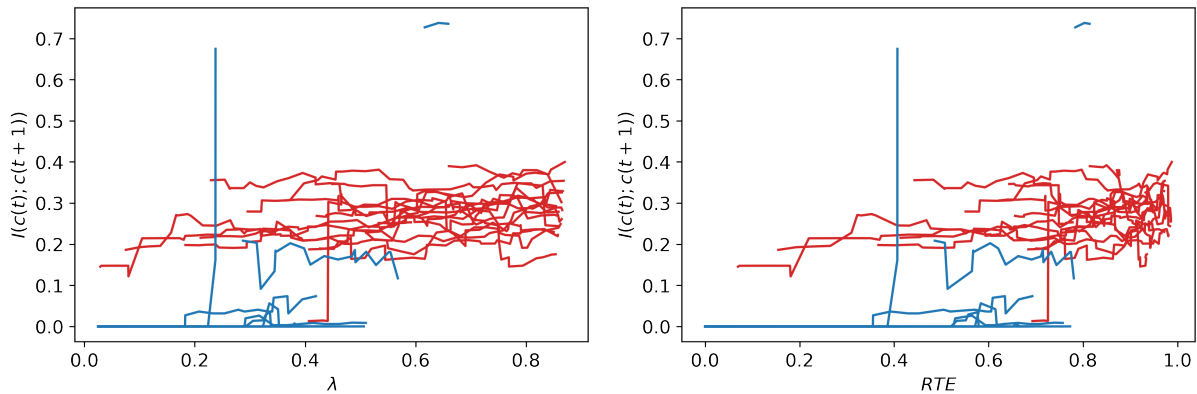
Figure 2.13.: Time averaged mutual information $\langle I(c(t); c(t+1)) \rangle$ for 15 independent rule table-walk-throughs of outer-totalistic $k = 8$ rules, run a random initial square grids of length $L = 64$ for 1000 time steps. Each walk through is divided into a blue line, depicting cyclic rules, and a red line, depicting non-cyclic rules. *Left:* $\langle I(c(t); c(t+1)) \rangle$ is plotted versus λ; *Right:* $\langle I(c(t); c(t+1)) \rangle$ plotted versus *RTE*. In contrast to the case of general rules, $\langle I(c(t); c(t+1)) \rangle$ does not appear to be very informative for outer-totalistic rules.

2.2.5. Cycle Length

The length of the limit cycle, $\mathcal{P}$ does not exist for chaotic rules, but among the ordered rules it can serve as a measure of how 'complex' their dynamics is. When sampling rules according to Langton's table-walk-through method with increasing λ, we first expect to obtain rules with $\mathcal{P}=1$, then rules with roughly increasing cycle length, and for an even larger λ a 'jump' to the simulation time, i.e. to not cyclic. $\mathcal{P}$ results, together with the corresponding entropies, are shown in Fig. 2.14. We hardly observe any spread in $\langle H_1 \rangle$ starting from independent random initial conditions. However, we do observe that the cycle length can strongly depend on the initial condition for intermediate *RTE* rules. For one of the rule-table walk throughs, it seems that we 'have just not simulated long enough', as some rules with λ values between cyclic rules of $\mathcal{P} \approx L$ have not converged to a limit cycle within the simulation time. However, according to Langton, for rules right at the entropy jump, the transition time is expected to diverge (C. G. Langton 1990).
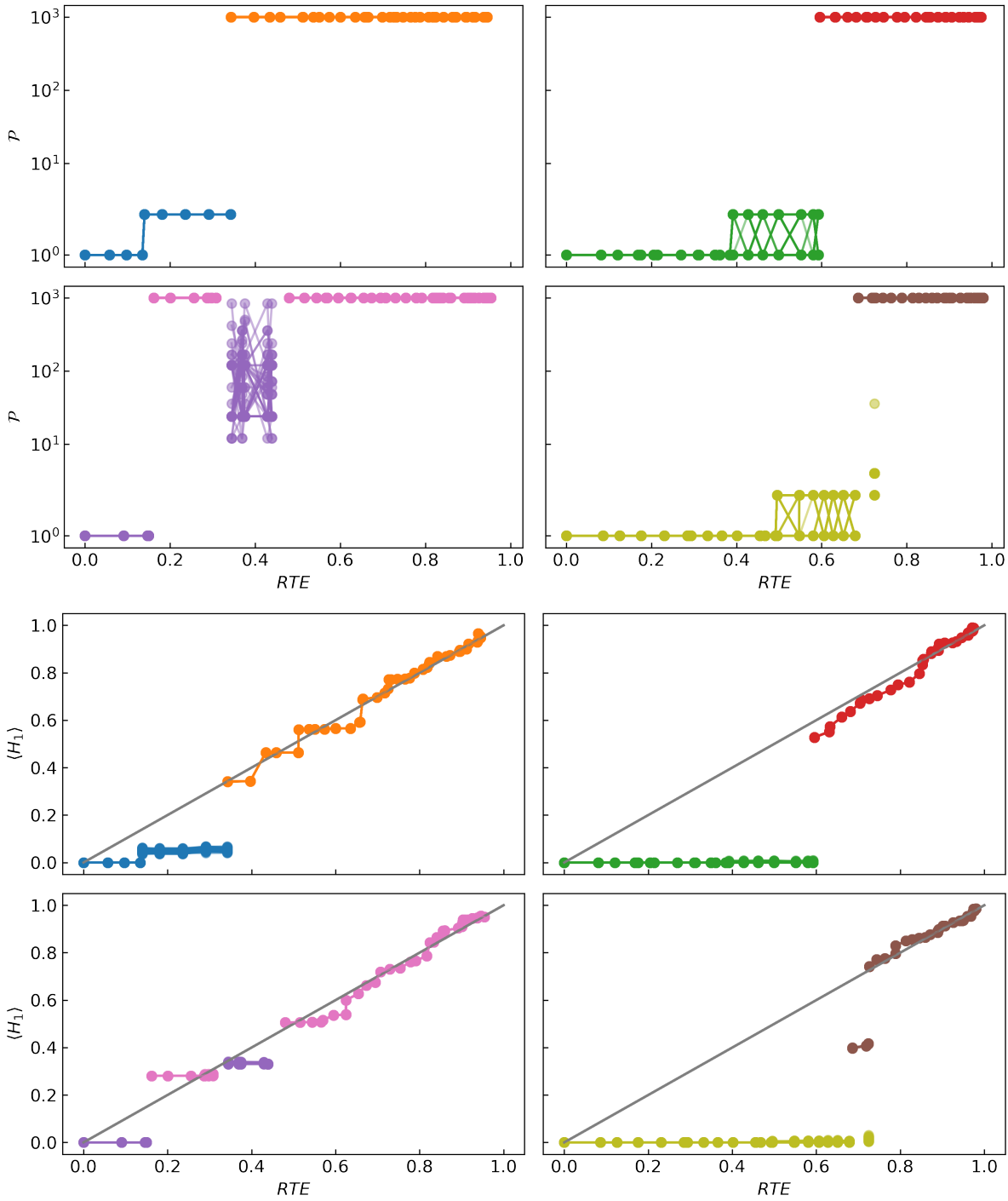
Figure 2.14.: 4 different table-walk-through sampled outer-totalistic $k = 8$ rule sets (same data as Fig. 2.10), each shown in a separate plot versus RTE. Each rule is run on 50 independent random initial conditions for 1000 time steps on a 64×64 square grid. Cyclic and non-cyclic rules of a table-walk-through are shown in different colors. *Upper:* Limit cycle lengths $\mathcal{P}$ are plotted for each of the 50 independent runs of each rule. We observe, e.g. in the lower-left plot that $\mathcal{P}$ strongly depends on the actual initial condition, varying between two orders of magnitude. In the lower-right $\mathcal{P}$ plot, there is one rule that has a finite cycle length for some initial conditions, but does not transition to a limit cycle for others. *Lower:* Single-cell entropies $\langle H_1 \rangle$ of the same simulations. $\langle H_1 \rangle$ of runs starting from independent initial conditions hardly deviate.

2.2.6. Other local Parameters

The presentation of further local parameters in this section is based on (Vispoel, Daly, and Baetens 2021).

Mean field theory   can determine the equilibrium densities ρ_c^{eq} of cell states $c \in \Sigma$, in some cases. It neglects the spatial correlations induced by a CA rule, which make certain neighborhood configurations $N \in \Sigma^{|\mathcal{N}|}$ more or less likely than expected by the density of each state c in the grid $C(t)$ at time t, $\rho_c(t)$. Using this mean field approximation, we can estimate the density of cells in state c as

$$\rho_c(t+1) = \sum_{N \in \phi^{-1}(c)} p(N,t),$$

$$\text{with} \quad p(N,t) = \prod_{c'=0}^{k-1} \left(\rho_c'(t)\right)^{\#c'(N)},$$

with $\#c'(N)$ denoting the number of neighbors in state c' in the neighborhood configuration N.

In equilibrium, we have $\rho_c(t+1) = \rho_c(t) =: \rho_1^{\text{eq}}$, which yields a set of $k-1$ so called master equations. For the elementary cellular automata, this translates to

$$\rho_1^{\text{eq}} = \sum_{N \in \phi^{-1}(c)} (\rho_1^{\text{eq}})^{\#_1(N)}(1 - \rho_1^{\text{eq}})^{\#_1(N)}. \tag{2.14}$$

To 'solve' this equation, we plot both sides against ρ_1^{eq} on the x-axis, consequently the first line is simply the diagonal. Stephen Wolfram observed the following properties of the resulting intersections, that correspond to the systems' fixed points (Wolfram 1984a): CA evolving to a homogeneous fixed point have a single stable fixed point at $\rho_1 = 0$, which implies that excitations die out. For CA evolving to an inhomogeneous stationary pattern, the fixed point at the origin is marginally stable, indicating that excitations will neither decay nor grow. The densities will depend on the initial configuration, thus the mean-field approximation cannot predict them. Periodic and chaotic CA have an unstable fixed point at the origin, as a small number of non-zero states in the initial grid is increased by the CA. The stable non-zero fixed point gives a good estimate of the equilibrium densities. Apparently, complex ECA rules are 'too simple' to generate a typical Class IV mean-field curve with a stable fixed point at the origin and for a larger ρ_1 value, as well as an unstable fixed point in between.

Two-point correlation functions   are complementary to the mean-field approximation. As the induction of spatial correlations by a CA indicates self-organization in the sense of settling to an attractor, the two-point spatial correlation function

$$Corr^{(2)}(r,t) := \langle c_i(t)c_{i+r}(t)\rangle_{\text{grid}} - \langle c_i(t)\rangle_{\text{grid}}\langle c_{i+r}(t)\rangle_{\text{grid}} \tag{2.15}$$

will be maximized for periodic CA and take small, but generally non-zero, values for chaotic cellular automata (Wolfram 1984a).

Power spectral properties   For elementary cellular automata, Ninagawa explored the frequency distribution of each grid configuration $C \in \{0,1\}^L$ of L sites for $f = 0, 1, \ldots T$ by applying a discrete Fourier transform of the CA time trace, starting from a given initial condition (Ninagawa 2008),

$$\hat{C}(f) = \frac{1}{T} \sum_{t=0}^{T-1} C(t) \exp\left(-i\frac{2\pi t f}{T}\right) . \tag{2.16}$$

He investigated the resulting power spectral density

$$PSD(f) := \frac{1}{L} \sum_{i=1}^{L} |\hat{C}_i(f)| . \tag{2.17}$$

In case of a time trace starting from a random initial condition with quiescent steady state, the power density is low at all frequencies as short transients are virtually random. If a rule has a non-zero homogeneous steady sate, the *PSD* will have a peak at $f = 0$. A cyclic steady state of period $\mathcal{P}$ leads to a peak at $f = \frac{T}{\mathcal{P}}$. In case of (very) chaotic behavior, the *PSD* is non-zero and relatively homogeneous for all frequencies, i.e., displaying a white noise spectrum. For complex rules, and a random initial condition, the power spectrum follows a power law. However, checking *PSD* plots for all ECA on Stephen Wolfram's simple program atlas[3] (Wolfram 2022), suggests that this spectrum is neither sufficient nor necessary for Class IV rules, given a particular random initial condition.

2.3. Global Parameters

Now we turn to 'complexity' measures based on all possible initial conditions.

2.3.1. Synergy

We have seen that entropy is an insightful information-theoretic measures to characterize the dynamics of a CA rule. Another promising information-theoretic measure, which combines entropy and mutual information, is the notion of *synergy S*. This approach was put forth in (Quax, Chliamovitch, Dupuis, et al. 2018). In information theory, synergy in the sources of information means, that multiple sources together provide more information, than each source alone. It is therefore often referred to as information of the 'whole minus sum of parts'. If we think of a cellular automaton as an information-processing device that takes the initial condition as an input, and outputs a time series of grid configurations, synergy measures the rule's capability to process the initial grid, not just each cell-state input separately.

More formally, we define the synergy S of a CA at time t with respect to a subset O of the initial grid $C(t = 0) \equiv C^0$ as

$$S(t) \equiv I(c_i^t; O) - \sum_{c_j^0 \in O} I(c_i^t; c_j^0) \ \text{ with } O \subseteq C^0 , \tag{2.18}$$

---

[3]website link

with $c_i^t \equiv c_i(t)$ for conciseness. As all cells are the same, the index i could actually be dropped. However, we keep it to indicate that c_i^t means a single fixed cell. We simply refer to synergy with respect to the full initial grid, $O = C^0$, as 'synergy'. By definition, it is the mutual information of a fixed cell i at time t, and the full initial grid (='whole'), minus the sum of the mutual information of each single cell $c_j(0)$ of the initial grid, and the current cell state (='sum of parts'). In this case, the first mutual information quantity, $I(c_i^t; C^0)$, reduces to $H(c_i^t) - H(c_i^t|C^0) = H(c_i^t)$, as the initial condition fully predicts the state of c_i^t for deterministic cellular automata. The probabilities, that are required for determining the mutual information quantities, are obtained by running the CA of interest for t time steps, from all possible initial conditions, and counting the occurrences of each c_i^t and c_j^0 realization. Considering all initial conditions implies, that all neighborhood configurations at $t = 0$ are equally probable, thus $S(t = 1)$ can be calculated directly from the rule table. As in 1D, a cell c_i^t can actually only be influenced by $c_{i-t}^0, \ldots, c_{i+t}^0$ cells, thus the number of initial conditions is computationally feasible for small times t (and small k).

Let us look at some ECA examples. Reversible (bijective) rules have zero synergy, which is directly clear from their algebraic or logic representation, as they can only depend on a single input

- rule 170: $(c_{i-1}, c_i, c_{i+1}) \to c_{i+1}$ and

- rule 15: $(c_{i-1}, c_i, c_{i+1}) \to \text{not}(c_{i-1})$.

Chaotic rules with maximal synergy $S = 1$ are, for instance,

- rule 90: $(c_{i-1}, c_i, c_{i+1}) \to (c_{i-1} + c_{i+1}) \mod 2$ and

- rule 150: $(c_{i-1}, c_i, c_{i+1}) \to (c_{i-1} + c_i + c_{i+1}) \mod 2$.

In those cases, knowing only a single cell state of c_i's neighborhood does not give any information about the outcome c_i^{t+1}. Interestingly, all of those rules are 'controllable', i.e., can be steered to any pattern, if one can control the left and right grid-boundary cell at each time step, (Ramalho, Kremser, Wu, and Gerland 2021).

Synergy for ECA

For ECA (Quax, Chliamovitch, Dupuis, et al. 2018) showed by a machine-learning approach that synergy is the single most predictive feature, of a set of information-theoretic measures $\vec{\psi}(\phi)$, to determine a rule ϕ's Wolfram class $Class(\phi)$. In fact, synergy and a single other quantity, roughly described as 'longest-range information transfer', are sufficient to predict the Wolfram class, when not distinguishing between Wolfram Class I and II. The set of information measures tested includes mutual information I_ϕ of any subset O of cells of the initial grid C^0 and the current cell state c_i^t,

$$\vec{\psi}^t(\phi) \equiv \left(I_\phi(c_i^t; O) \mid O \in 2^{C^0} \right) , \tag{2.19}$$

where 2^{C^0} is the power set notation for all subsets of cells of the initial grid. The span of $\vec{\psi}(\phi)$ includes local 'memory' $I_\phi(c_i^t; c_i^0)$, information 'transfer' $I_\phi(c_i^t; c_{i-t}^0) + I_\phi(c_i^t; c_{i+t}^0)$ and information synergy $I_\phi(c_i^t; \{c_{i-t}^0, \ldots, c_{i+t}^0\}) - \sum_{n=-t}^{t} I_\phi(c_i^t; c_{i+n}^0)$.

Let us study the results for the first time step, to gain some intuition. At $t = 1$, only the cells c^0_{i-1}, c^0_i and c^0_{i+1} can contribute, thus

$$O \in \left\{ \{c^0_i\}, \{c^0_{i-1}\}, \{c^0_{i+1}\}, \{c^0_{i-1}, c^0_i\}, \{c^0_i, c^0_{i+1}\}, \{c^0_{i-1}, c^0_{i+1}\}, \{c^0_{i-1}, c^0_i, c^0_{i+1}\} \right\} .$$

We can simplify the notation in 1D, by fixing the order of cells within the set, and using '1' if a cell at that position is within the initial grid's subset O, or '0', if not. With that, we can write the feature vector at, e.g. $t = 1$, as follows:

$$\vec{\psi}^{t=1}(\phi) = \left(I_\phi 100, I_\phi 010, I_\phi 100, I_\phi 110, I_\phi 011, I_\phi 101, I_\phi 111 \right) ,$$

where, e.g., '110' corresponds to $\left(c^t_i; \{c^0_{i-1}, c^0_i\} \right)$.

For $t = 1$, Quax' and Langton's notion of 'memory', respectively 'local information conservation' are similar, as both measure the mutual information of a cell at the current and previous time step. The notions differ, because Quax uses an ensemble of initial conditions, while Langton evaluates memory by averaging over many time steps, after the rule has reached its 'typical' behavior.

Quax et al. formalize the prediction problem as follows: First, consider the rule number R of ECA rule ϕ as a uniformly distributed random stochastic variable, i.e., $P(R) = \frac{1}{256}$. The (normalized) predictive power is then defined as

$$\frac{I(Class(R); \vec{\psi}^t(\phi))}{H(Class(R))} ,$$

with $H(Class(R))$ the entropy of the rule classes. This normalization is chosen, because the number of rules within in a Wolfram Class differs strongly between the Classes. We are interested in the single feature that has maximal individual predictive power,

$$\psi^t_1 \equiv \arg\max_{\psi \in \vec{\chi}^t} \frac{I(Class(R); \psi^t)}{H(Class(R))} .$$

Therefore we extend the feature set to all synergy quantities to $\vec{\varphi}^t$,

$$\vec{\chi}^t(R) \equiv \left(\vec{\psi}^t, \vec{\varphi}^t \right) , \tag{2.20}$$

with the set of synergies $\vec{\varphi}^t$ defined as,

$$\vec{\varphi}^t \equiv \left(I_\phi(c^0_i; O) - \sum_{c_j \in O} I_\phi(c^t_i; c^0_j) ; O \in 2^{C^0} \right) .$$

For example, at $t = 1$ we have $\vec{\varphi}^{t=1} = \left\{ S_\phi 110, S_\phi 011, S_\phi 101, S_\phi 111 \right\}$.

Quax et al. showed that synergy (with respect to the full initial grid) is the most predictive feature: At $t = 1$, synergy alone has a predictive power of 0.37, with 1.0 implying perfect prediction. When adding a second time step ($t = 2$), synergy increases its predictive power to 0.9, and slightly further for three time steps. Already at $t = 3$, two information features allow for uniquely identifying the behavior class: Synergy together with the mutual information

quantity $I_\phi 1001001$, a combination of mutual information and longest-range information transfer, achieve a predictive power of 1.0.

Fig. 2.15 shows for $t = 1, 2, 3$ and 7 synergy $S(t)$ plotted versus rule number. The non equivalent 88 ECA rules are colored according to their Wolfram Classes. For perfect prediction, we expect to see a clear separation between the different classes, except for Class I and Class II. At $t = 1$, this is not the case, as expected. We further observe that there are five rules with $S = 1$: those are, the rule that maps everything to zero, and the four reversible rules (rule 15, rule 51, rule 170 and rule 204), discussed before. Likewise, the four data points with maximal synergy correspond to the two chaotic rules mentioned (rule 90 and rule 150), and their relatives[4], rule 60 and rule 105. For $t = 2$, separation between different Wolfram Classes indeed seems to increase, and slightly further improves at $t = 3$. $t = 7$ is plotted to show that running the simulation longer does not further improve separation of rules, consistent with (Quax, Chliamovitch, Dupuis, et al. 2018). This fast temporal convergence implies that all 2^{2t+1} initial conditions can be tested, as $t = 3$ suffices.

Synergy for 2D outer-totalistic CA

In contrast to the ECA, there is no common empirical classification such as the Wolfram classification, that we could use to directly check, how synergy S performs in characterizing 2D outer-totalistic CA rules. Thus, we will have to restrict ourselves to the time-averaged single-site entropy $\langle H_1 \rangle$ of those rules. Fig. 2.16 shows $\langle H_1 \rangle$ for $k = 3$ rules plotted for increasing λ, $S(t = 1)$ calculated form the rule table, and $S(t = 3)$. From those three options, $S(t = 3)$ seems to correlate best with $\langle H_1 \rangle$. Nonetheless, this correlation is not convincing for rules with small $\langle H_1 \rangle$. We further observe that $\langle H_1 \rangle$ versus RTE looks very similar to $\langle H_1 \rangle$ versus $S(t = 1)$.

In Fig. 2.17, we thus check $\langle H_1 \rangle$, $S(t = 1)$ calculated from the rule table, and RTE of outer-totalistic and typical general rules. For typical general $k = 8$ rules, we observe that $S(t = 1)$ is highly correlated to the rule-table entropy RTE. For outer-totalistic rules, the deviation between RTE and $S(t = 1)$ theory is larger, and grows with RTE.

In contrast to 1D, where we can check how well we can reproduce the Wolfram classification for increasing time, it is not clear, after how many time steps t synergy gives the best prediction. In Fig. 2.18 we plotted synergy for $t = 1, 2, 3$ over the rule-table entropy RTE for $k = 3$ rules. As expected, for $t = 0$ the synergy is zero for all rules, but for larger times there is no clear order. For most, but not for all rules $S(t = 2)$ is closer to $S(t = 3)$ and $S(t = 3) \leq S(t = 2)$. Fig. 2.18 shows the synergy of outer-totalistic $k = 3$ rules for up to 3 time steps, including its values at time $t = 1$, as calculated from the rule table. The good agreement of these theoretical values with the measured $S(t = 1)$ values suggests, that the number of initial conditions is sufficiently large. As the outer-totalistic rules only radially transmit information, similarly to 1D, we hypothesize that $S(t = 3)$ should suffice to separate rules with qualitatively different dynamic behavior.

---

[4]rule 60: $(c_{i-1}, c_i, c_{i+1}) \to (c_{i-1} + c_i) \mod 2$, and rule 105: $(c_{i-1}, c_i, c_{i+1}) \to (1 + c_{i-1} + c_i + c_{i+1}) \mod 2$
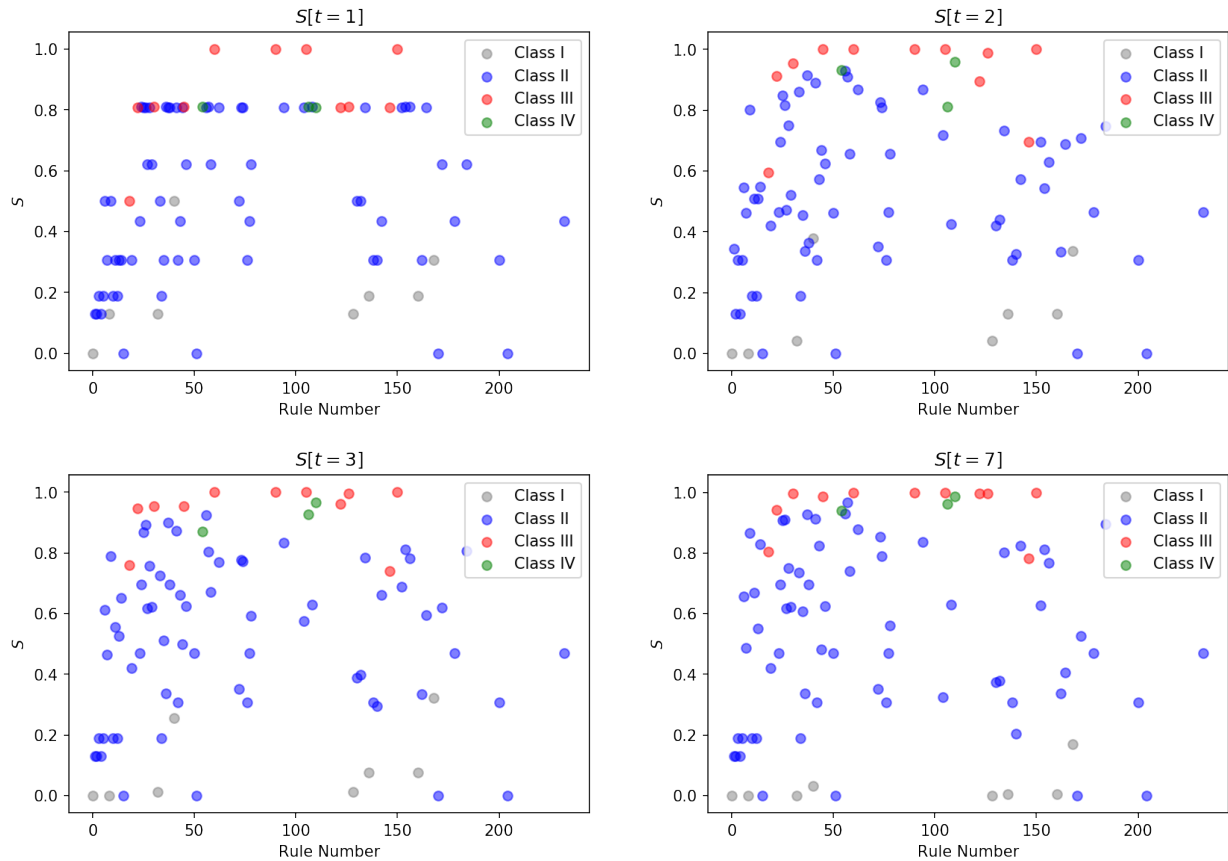
Figure 2.15.: Synergy *S* with respect to the whole initial grid is plotted for each of the 88 non-equivalent elementary cellular automata colored according to their Wolfram class for times $t = 1, 2, 3$ and 7. Separation of different Wolfram Classes increases from $t = 1$ to $t = 3$, but does not further improve for longer times ($t = 7$).

Figure 2.16.: Time-averaged single-site entropies $\langle H_1 \rangle$ of 100 outer-totalistic $k = 3$ rules, sampled with Langton's random-table method and run on a square grid of length $L = 64$ are shown. Each rule is run for 1000 time steps from three independent random initial grids (No. 0, No. 1, No. 2). Synergy values are determined from running each rule for up to 3 time steps on 10^5 random initial grids. *Upper left*: $\langle H_1 \rangle$ is plotted against *RTE*. We hardly observe any deviation of $\langle H_1 \rangle$ between the three independent runs. *Upper right*: $\langle H_1 \rangle$ plotted against S at $t = 1$, as calculated from the rule table ('theory'). *Bottom:* $\langle H_1 \rangle$ plotted against S at $t = 3$.
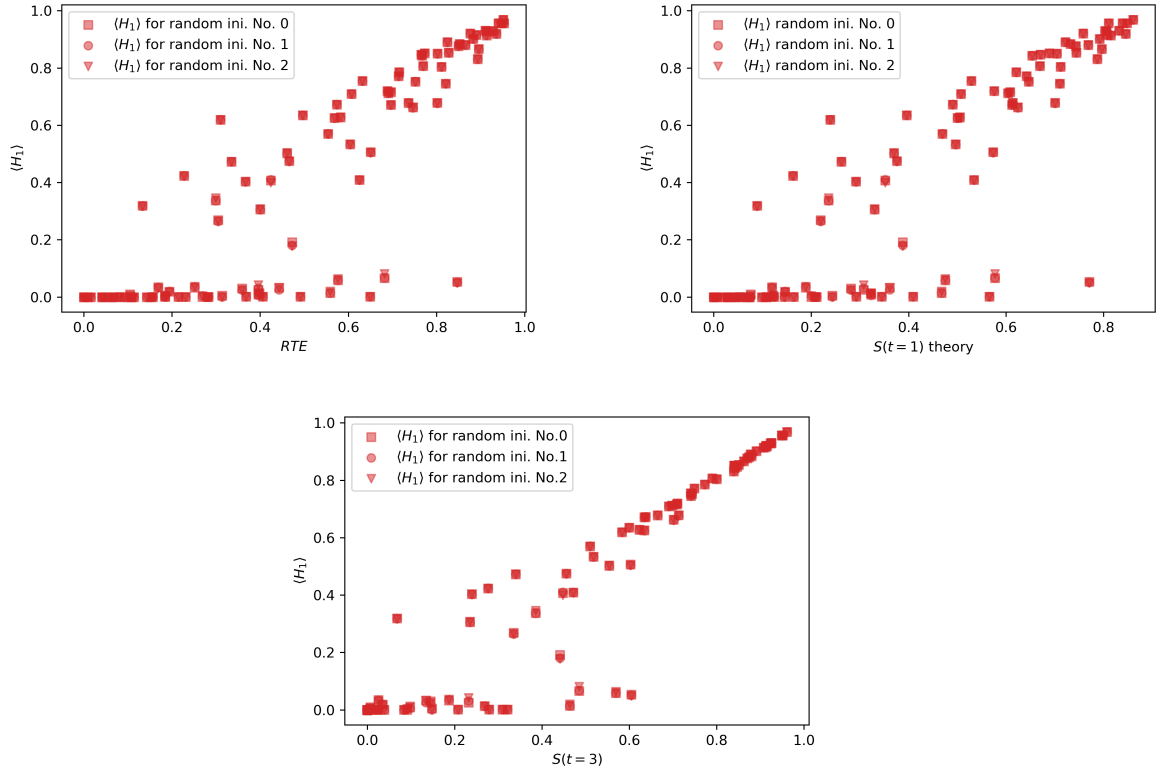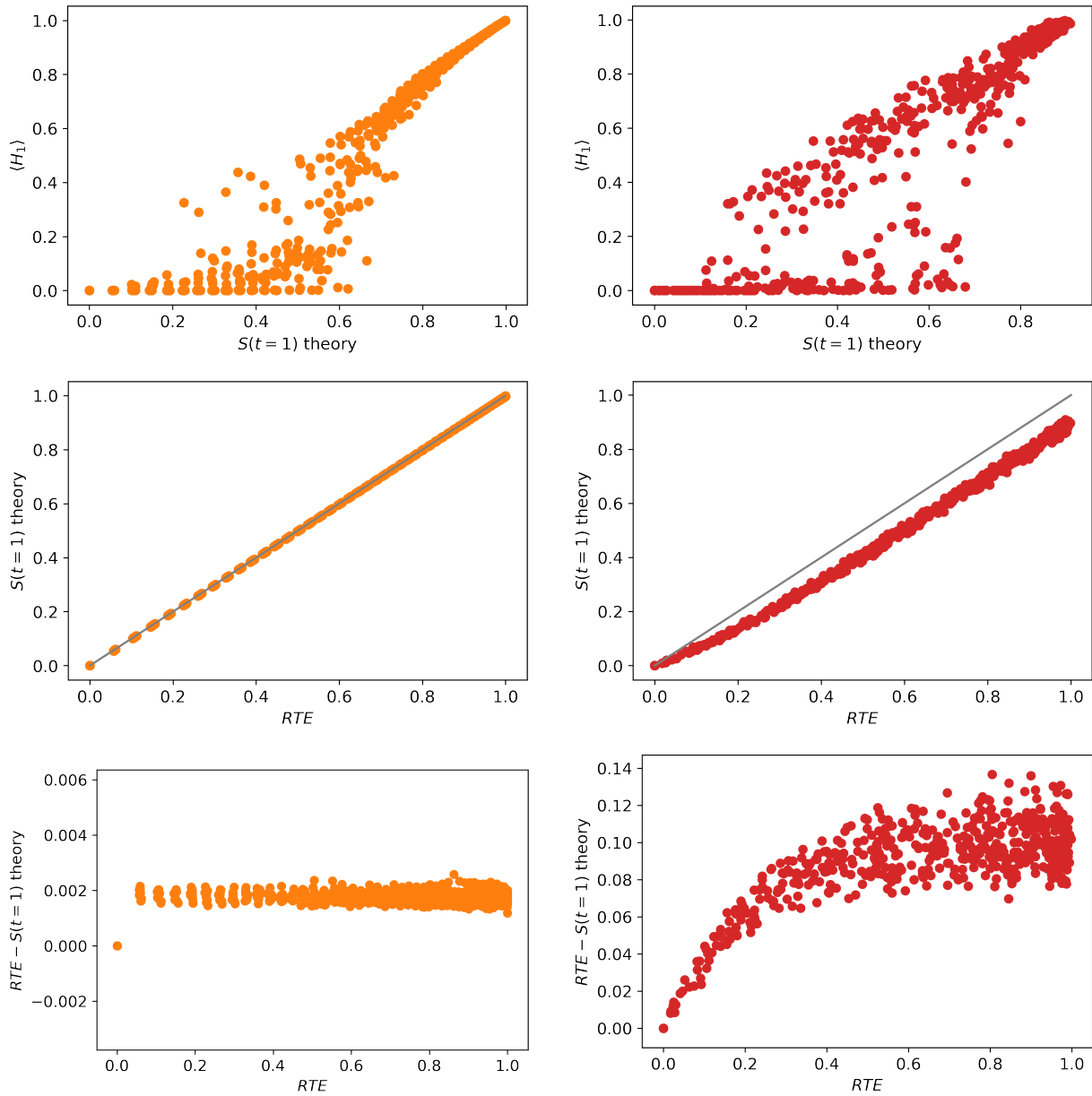
Figure 2.17.: Time-averaged single-site entropies $\langle H_1 \rangle$ of 550 $k = 8$ rules sampled with Langton's random-table method, and run on an initial random square grid of length $L = 64$ for 1000 time steps. The *left* plots show results for general (orange), and the *right* plots for outer-totalistic rules (orange). *Top*: $\langle H_1 \rangle$ is plotted versus the $S(t = 1)$ theory value, which is calculated directly from the rule table. The *center* plot confirms our observation that the top plot is quite similar to $\langle H_1 \rangle$ versus *RTE* plots. In the *bottom*, we can show that the difference between *RTE* and $S(t = 1)$ theory is indeed quite small. On the *right*, we can see, that for outer-totalistic rules, the deviation between *RTE* and $S(t = 1)$ theory is growing with *RTE*, and larger, in general.

Figure 2.18.: Synergies for outer-totalistic $k = 3$ state rules. Same data as in Fig. 2.16 is used. We observe that the measured $S(t = 1)$ values agree well with the ones calculated from the rule table ('theory'). However, we do not observe that the rules' synergy values converge in time.
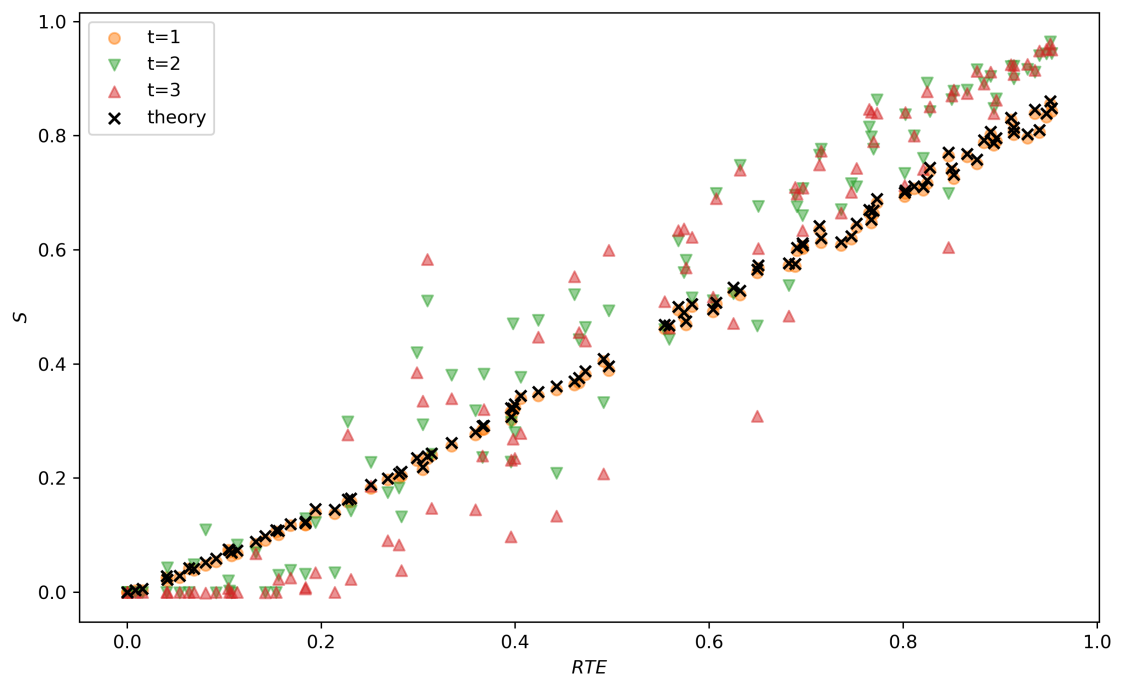
2.3.2. Attractor-based Classification

Finite Cellular Automata

We first summarize results from (Wuensche, M. Lesser, and M. J. Lesser 1992) and (Wuensche 1998), before reporting on our numerical checks of those statements.

For any finite CA of length L it is possible to divide its grids state space Σ^{L^d} into basins of attraction (with d denoting the dimension), see Fig. 2.19. The basin-of-attraction field, '*BoA*', can be characterized by the following properties:

- max *deg*, the maximum in-degree, respectively number of pre-image states (direct predecessors) any grid configuration can have,

- *GoE*, the number of 'garden-of-Eden' grid states, i.e., grid states, that cannot be reached by the CA for $t > 0$,

- max $\mathcal{P}$, the maximum period of the attractor cycles, respectively the cycle length,

- max $\mathcal{T}$, the maximum transient-tree length,

- No. *BoA*, the number of separate basins in the field, respectively the number of different limit cycles.

Max $\mathcal{P}$, max $\mathcal{T}$ and No. *BoA* will vary together, as prevalence of one will reduce margin for the other two. The max *deg* reflects the *GoE*. The in-degree *deg* strongly determines the number of possible ways, by that grid states can be connected to basins, which partition the grid-state space. If *deg* diverges exponentially with system size L^d, most states are in pre-images, implying that the scope for $\mathcal{P}$, $\mathcal{T}$ and No. *BoA* is limited to remain either fixed, or grow arithmetically with L^d and vice versa.

According to (Wuensche 1998) an approximate relation between an elementary cellular automatons' Wolfram Class and its basin-of-attraction field can be drawn as follows:

Class I The maximum pre-image degree *deg* diverges exponentially with L, as rules have a large proportion of neighborhoods mapping to the same state. Most states are locked into pre-images of branches rooted at point attractors, i.p. the quiescent state configuration. The few states outside those basins of attraction only have enough scope for attractors with short transients $\mathcal{T}$ and periods $\mathcal{P}$.

Class II Max *deg* grows exponentially with L, but slower than for Class I rules, thus slightly longer cycles $\mathcal{P}$ and transients $\mathcal{T}$ are possible.

Class IV Max *deg* and some combination of max $\mathcal{P}$, $\mathcal{T}$, and No. *BoA* is finely balanced and grows by 'some intermediate function' with L, such as $(k^L)^{1/k}$. Enough scope for moderately long cycles and transients remains.

Class III Max *deg* is fixed irrespective of L. So some combination of max $\mathcal{P}$, $\mathcal{T}$, and No. *BoA* will grow exponentially with L.
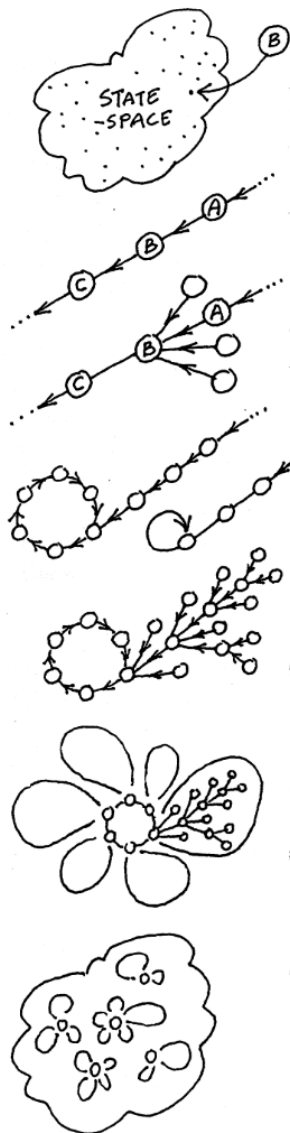
Figure 2.19.: *From top to bottom*: (1) For a *k*-state CA on a finite grid of size L^d we have a state space of k^{L^d} different grid configurations. 'B' denotes one of them. (2) Part of a trajectory in state-space generated by the update rule: 'C' is a successor of 'B' and 'A' is a pre-image of 'B'. (3) State 'B' may have other pre-images besides 'A'. Their number is the 'in-degree'. Pre-image states may again have pre-images. If not, they are called 'garden-of-Eden' states or 'leaves'. (4) As the system is deterministic, finite and discrete, when simulating long enough (with an upper bound given by the Poincare recurrence time), any trajectory has to enter a state it has been in before, and with that an attractor cycle. The period of the attractor is the number of states in its cycle. The trajectory leading to the attractor is a transient. (5) Sub tree 'transient tree' rooted on one of the attractor states including all transients from all Garden of Eden states in its iterated pre-image. (6) The tree containing the transient trees for each attractor state is called 'basin of attraction'. (7) The basin of attraction field contains all basins of attractions in the state space. It is a partition of the state space. Taken from (Wuensche 1998).

An atlas depicting the basin-of-attraction fields for each ECA up to a length of $L = 15$ can be found in (Wuensche, M. Lesser, and M. J. Lesser 1992) or constructed by Wuensches' 'discrete dynamics lab' (Wuensche 2021).

We tested the statements above, with Stephan Kremser[5] providing data for the observables characterizing the attractor space of each ECA rule, up to a grid length of $L = 16$ sites, generated by a simulation routine adapted from (Ramalho, Kremser, Wu, and Gerland 2021). For the maximal in-degree, max *deg*, we observe in Fig. 2.20, that indeed, it exponentially increases with L for Class I rules, and also for Class II rules, to a smaller extend, as stated by (Wuensche, M. Lesser, and M. J. Lesser 1992) — except from the reversible Class II rules. This exception is obvious, as a rule can only be reversible if *deg* is exactly one for all options. However, there are Class III rules (i.e. 125, 22, 146 and 122) for which max *deg* also exponentially increases with L. When checking the corresponding cycles explicitly, it turns out that these cycles are the trivial fixed points — i.e., the all-zeros or all-ones state. The number of Garden-of-Eden states *GoE* increases exponentially with system size for all rules, but the reversible ones. The maximal attractor size max $\mathcal{P}$ is constant and for most Class I and Class II rules, and non-monotonously increases for most Class III and Class IV rules. The maximum transient length max $\mathcal{T}$ does not exceed 100 time steps (on a grid with a maximal length of $L = 16$) for Class I and Class II rules, and non-monotonously increases for Class III and Class IV rules. The number of separate attractors max No. *BoA* exponentially increases with system size for Class II rules. For Class I rules it can also be constant. For Class IV and Class III rules, it is non-monotonously increasing, except for rule 90.
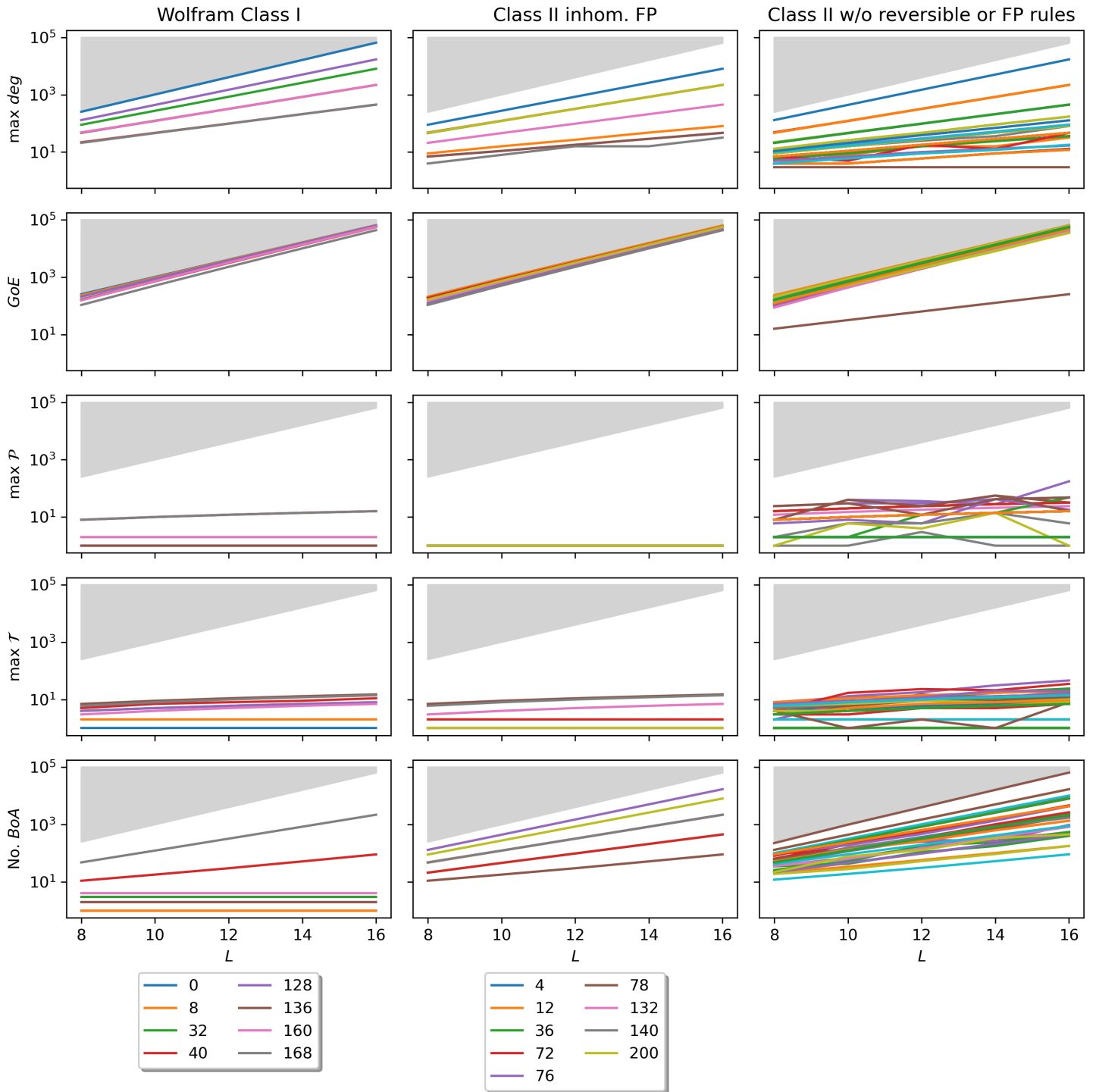
Figure 2.20.: Observables characterizing the attractor field are plotted for increasing lengths all 88 non-equivalent ECA rules grouped w.r.t. to their Wolfram Class. Note that the lines interpolate between even length values as for a two state system, many rules showed a periodicity of 2. Also, we omitted values smaller than $L = 8$, as related studies showed, that observable curves begin to converge from $L = 8$ onward (Ramalho, Kremser, Wu, and Gerland 2021). These omissions are for better visibility of the results. Also, labels for Class II rules, without rules not realizing a proper cyclic attractor (proper in the sense of periodicity > 1), and without reversible rules are omitted, as these are 53 in number (i.e. 1, 2, 3, 5, 6, 7, 9, 10, 11, 13, 14, 19, 23, 24, 25, 26, 27, 28, 29, 33, 34, 35, 37, 38, 41, 42, 43, 44, 46, 50, 56, 57, 58, 62, 73, 74, 77, 94, 104, 108, 130, 134, 138, 142, 152, 154, 156, 162, 164, 172, 178, 184, 232). The lower edge of the shaded area corresponds to the maximal number of states, 2^L.
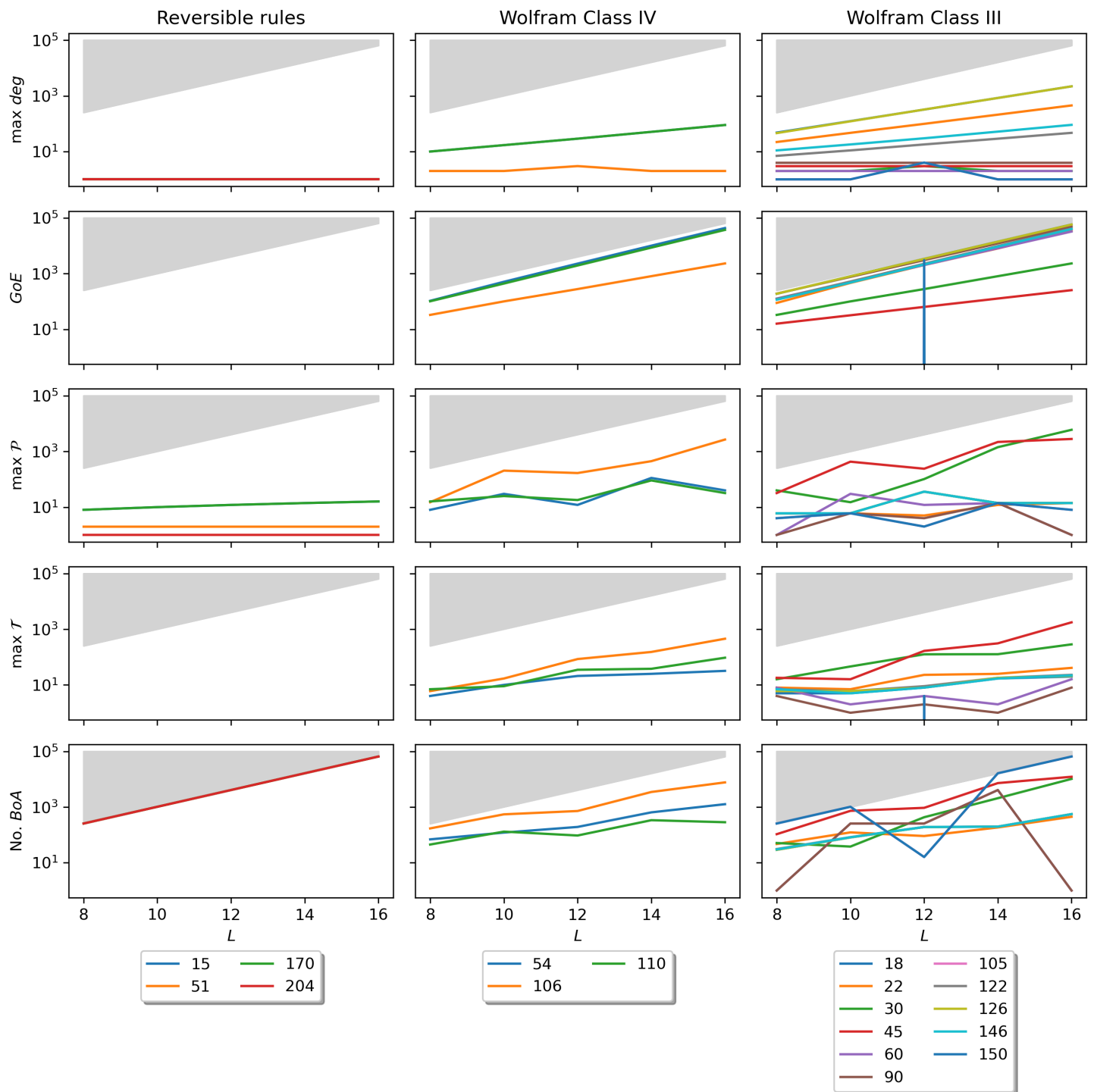
Figure 2.20.: We can roughly confirm some, but not all of the statements in (Wuensche 1998), as discussed in the main text. Most notably, max *deg* is apparently not fixed for all Wolfram Class III rules, e.g. max *deg* increases exponentially for 'chaotic' rule 126, 22, 146 and 122. However, we can confirm that max deg increases exponentially for most Wolfram Class I and II rules, naturally except the reversible ones.
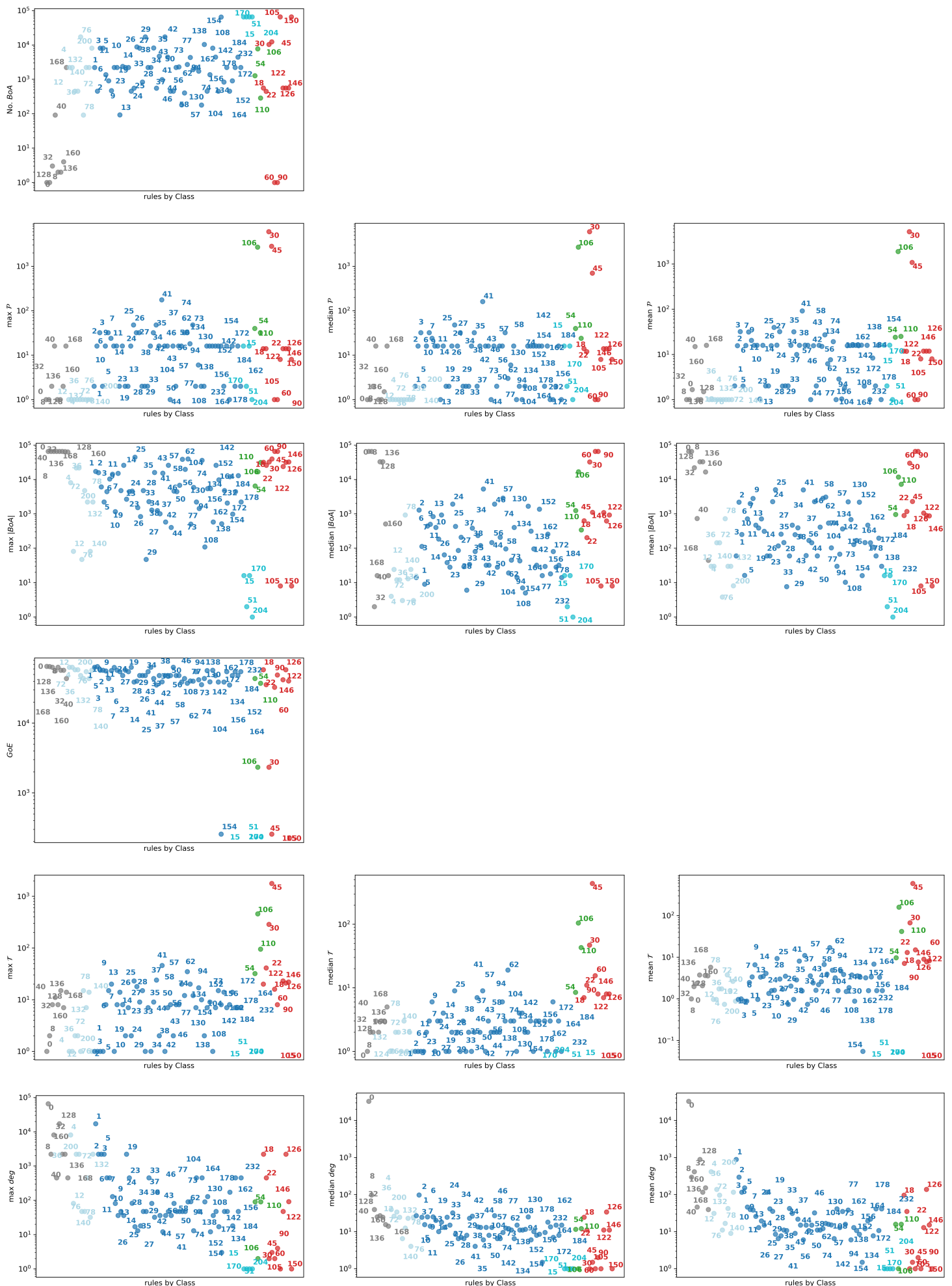
Figure 2.21.: For a fixed length $L = 16$ observables, characterizing the attractor field, are plotted for each rule, manually grouped according to its class: Class I (grey), Class II with only (inhomogeneous) fixed point stationary states (light blue), Class II without the rules just mentioned, and without reversible rules (blue), reversible rules (cyan), Wolfram Class IV 'complex' rules (green) and Wolfram Class III 'chaotic' rules (red).

In a sense as a cross section of Fig. 2.20 for $L = 16$, we show the same observables for each rule, grouped by its Class, in Fig. 2.21. To better characterize the distributions of the observables (except from No. *BoA* and *GoE*), we added plots for the observables median and mean. For the cycle length the maximum, median and mean results are very similar. The minimal cycle length is either one or two (not shown). The maximum and median basin of attraction sizes No. *BoA* are qualitatively different, i.e. a rule with a larger max No. *BoA* than another rule can have a lower median No. *BoA* than this rule. Median and mean are different for Class I rules, otherwise quite similar. The maximal and median in-degree behave similarly (except for the zero rule), also median and mean are rather similar. Also, for the transient length, maximum, median and mean results vary similarly from rule to rule. For the number of Garden-of Eden-states there is not much spread in the rules except from the reversible rules and six 'outliers', 154, 105, 150, 45 and 106, respectively, 30.

As the observables are highly correlated, we performed a principal component analysis (PCA) on the normalized logarithm of the data, to construct an orthogonal basis with eigenvectors in the direction of maximal variance increase within the data, see Fig. 2.23. The PCA loads, i.e., the observables plotted with respect to the first two principal components are shown in Fig. 2.22. The first PCA component PC 1 can explain for 35.5% of the variance in the data, the second for 31.7%. We observe that for the cycle length $\mathcal{P}$, the transient length $\mathcal{T}$, and the in-degree *deg* results for maximum, median and mean contribute very similarly to the principal components, in agreement with our observations from Fig. 2.21. To PC 1, $\mathcal{P}$ and $\mathcal{T}$ have the largest contribution. For PC 2, it is the No. *BoA*, followed by *deg*, *GoE* and min, mean, median and maximum of the basin of attraction size distribution, $|BoA|$. $|BoA|$ is an observable we added to Wuensches list of observables, characterizing the attractor-basin field.

We then used a k-means clustering for $n = 2$ up to 9 components (only $n \in \{2, 4, 6\}$ shown). However, already visually from Fig. 2.23, its is clear that no version even roughly reproduces a separation of (i) Wolfram Class I and Class II, from Class III and Class IV rules ($n = 2$), or (ii) the Wolfram Classes ($n = 4$) or (iii) additionally splitting off inhomogeneous fixed point rules and reversible rules from Class II ($n = 6$).
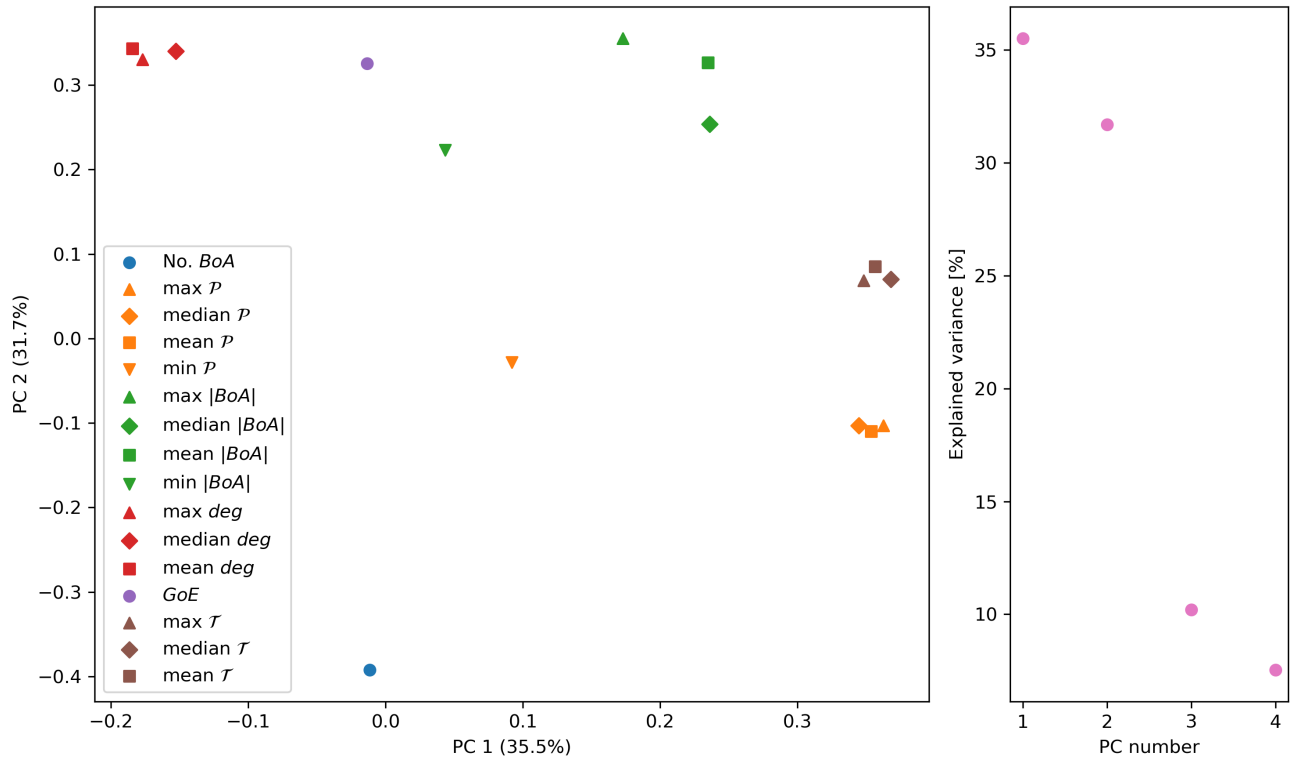
Figure 2.22.: PCA loads (*right*) and explained variance (*left*) of the PCA analysis of a number of observables, listed in the legend, characterizing the basin-of-attraction field for ECA on an $L = 16$ grid. *Left*: Maximum, median and mean of observable have a similar contribution to the principal components (CP). The minima contribute less. There is no single observable that essentially forms a principal component. *Right*: The explained variance drops significantly from the second to the third principal component, suggesting the first two are sufficient for our analysis.
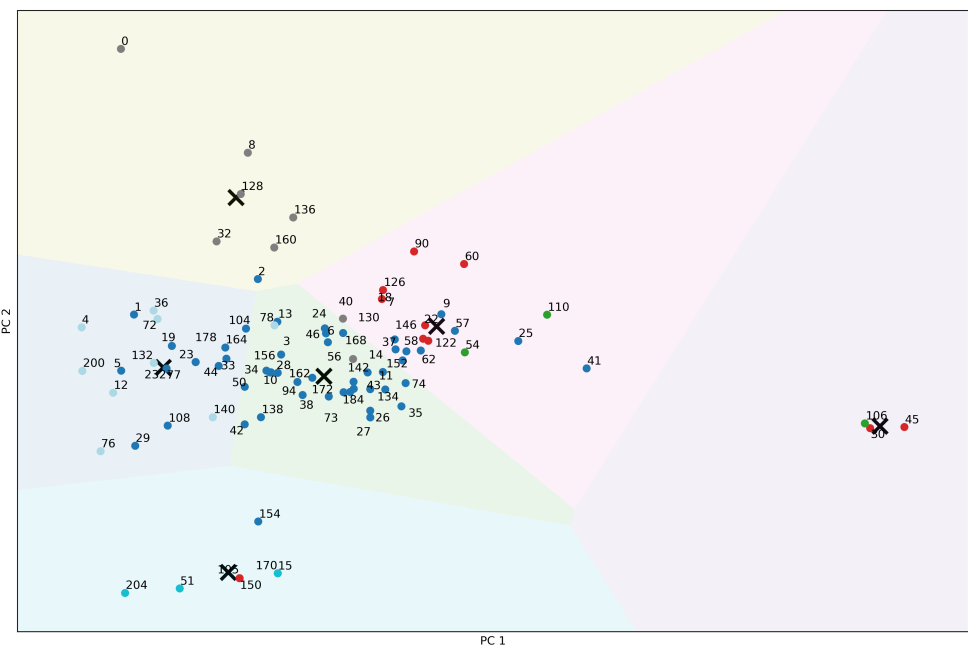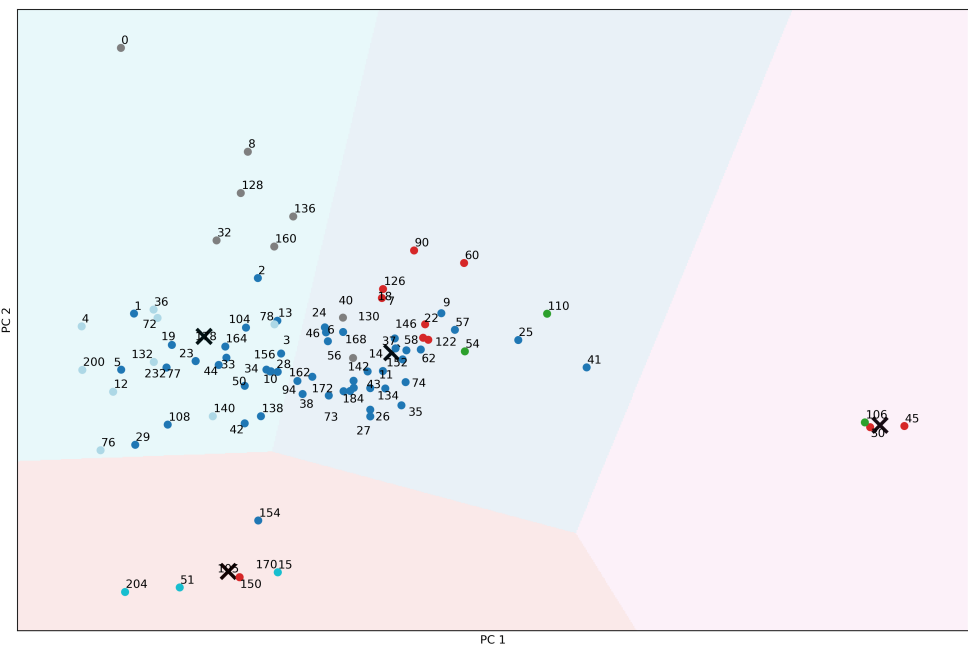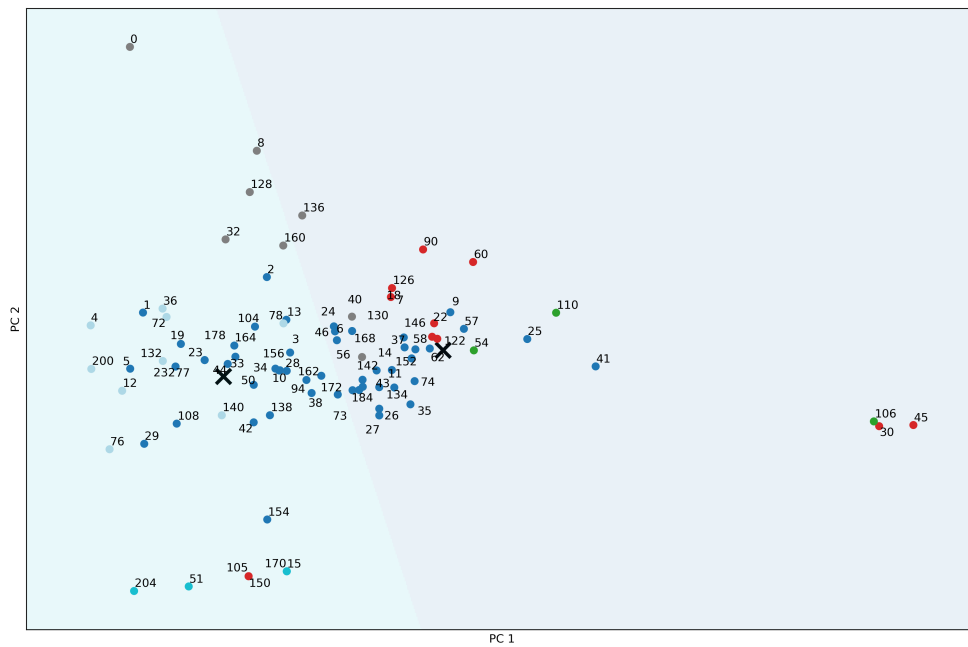
Figure 2.23.: K-mean results for $n = 2, 4, 6$ components for PCA reduced data (PC 1 and PC 2). Black crosses depict the center of each component. Class I (grey), Class II with only (inhomogeneous) fixed point stationary states (light blue), Class II without the rules just mentioned, and without reversible rules (blue), reversible rules (cyan), Wolfram Class IV rules (green), and Wolfram Class III rules (red).
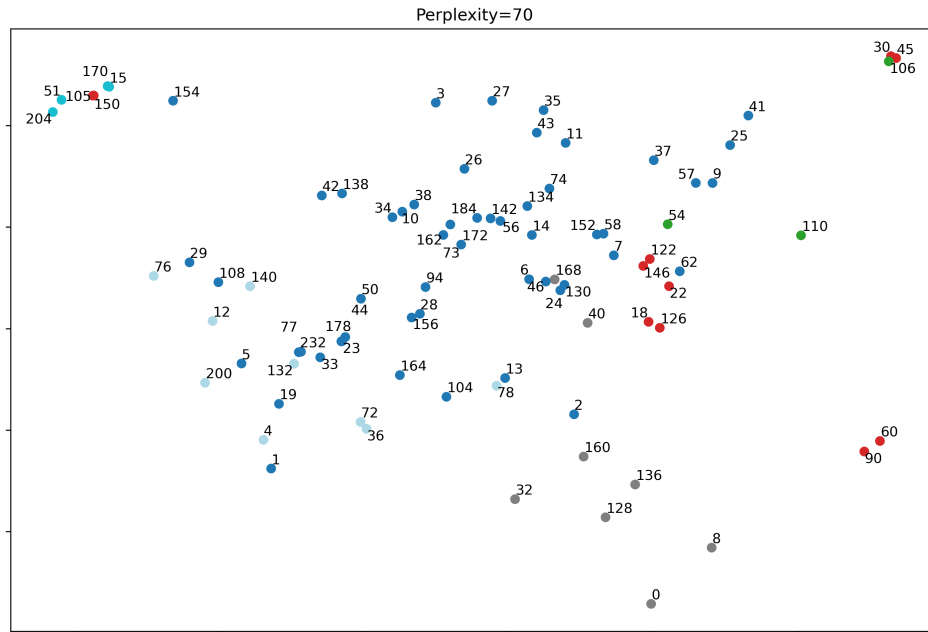
Figure 2.24.: t-SNE representation of the attractor data underlying Fig. 2.21 with a perplexity of 70, random initialization and 5000 time steps. This representation agrees well with the PCA results shown in Fig. 2.23 and also suggest, that there is no clear separation of the rules into classes with respect to the basin-of-attractor field observables tested.

Using a t-distributed stochastic neighbor embedding 't-SNE' representation gives very similar results, see Fig. 2.24. This is non-trivial, because the t-distributed stochastic neighbor embedding does not rely on linear correlations, but tries to conserve closeness between data points when mapping them to 2D. Apparently the Wolfram Classification does not properly reflect the finite length dynamic behavior. However, the scaling of the attractor-space observables with system length might be more insightful. It would be interesting to repeat the analysis of this chapter with scaling properties obtained from data including larger grids.

Fractal dimension is another commonly used measure to describe the attractor-basin field (Wolfram 1984a),(Vispoel, Daly, and Baetens 2021): The block entropy $H(B, t)$ as defined in Eq. 2.3 can be seen as the $\alpha \to 1$-limit of the the Renyi entropy $H_\alpha(B, t) = \frac{1}{1-\alpha} \log \left(\sum_{B \in \Sigma^{|B|}} P(B, t) \right)$. For $\alpha = 0$ the Renyi entropy measures the uncertainty of how many configurations can be reached at time t, regardless of how probable each configuration is. For large times, counted configurations lie on the attractor of the CA. In case of a finite grid, $H_{\alpha=0}(t \to \infty) \equiv d_0$ gives the fraction of configurations that is cyclic, and with that the fractal

dimension of the CA attractor. If the attractor is the set of all possible configurations, then $d_0 = 1$ for reversible, or maximally chaotic CAs. In general, for chaotic rules the attractor is some fractal subspace of all configurations, i.e. $0 < d_0 \leq 1$. If there is a finite number of configurations on the attractor, as common for Wolfram's Class I and Class II rules, we have $d_0 = 0$.

Infinite Cellular Automata

For infinite cellular automata, Hurley (Hurley 1990) has put forth an attractor based classification of cellular automata which is uncomputable in general, but gives insight into the spectrum of possible dynamics.

First, we need to actually define the notion of a (minimal) attractor or quasi attractor.

Let (X, ϕ) be a dynamical system. The ω-limit of a subset $U \subseteq X$ is $\omega(U) = \bigcap_{n>0} \overline{\bigcup_{m>n} \phi^m(U)}$ with $\overline{\cdot}$ denoting closure. A (non-empty) set $Y \subseteq X$ is an *attractor* if there exists an open set U such that $Y = \omega(U)$ (and $\phi(\overline{U}) \subseteq U$). A (non-empty) set is a *quasi-attractor* if it is a countable intersection of attractors, but itself not an attractor. Y is a *minimal attractor* (minimal quasi-attractor) if it does not contain any attractor (quasi-attractor) as a proper subset. Hurley proved that any cellular automaton rule ϕ satisfies exactly one of the following cases (Hurley 1990):

Case 1 There exists a unique minimal attractor of ϕ.

Case 2 There exists a unique minimal quasi-attractor Q of ϕ.

Case 3 There exists a pair of disjoint attractors. In this case ϕ as uncountably many minimal quasi-attractors.

The zero-map CA, $(c_{i-1}, c_i, c_{i+1}) \mapsto 0$, is in Case 1, as well as any rule ϕ with a finite number of attractors. The identity rule belongs to Case 3.

This classification is further refined in (Kůrka 1997), see Appendix A.2. Note that it is applicable to any discrete dynamical system.

2.3.3. Other global Parameters

This short introduction is based on (Vispoel, Daly, and Baetens 2021) and included here to give a better overview of what has been done already.

Topological properties of infinite CA time traces were used by Kurka to build a classification scheme (Kůrka 1997). By translating the notions of equicontinuity and expansivity from continuous to discrete dynamical systems, he could construct three classes and relate them to his attractor based classification, as well as the Wolfram Classes.

Formal language theory methods can also be applied to 1D cellular automata. A formal language can be characterized by the memory needed to implement the rules of its grammar. The set of all possible CA configurations generated after a finite number of time steps forms a regular language (Wolfram 1984b). This set can be represented by a De Bruijn graph for 1D

cellular automata for short times only. Consequently, formal language theory methods are rather limited for studying cellular automata.

2.4. Conclusions of Part I

Here, we reviewed and further explored characterization schemes of cellular automata. Although the focus is on 2D CA rules with a 'large' number of states $k \geq 8$, we also investigated the opposite limit, the two state 1D CAs (elementary cellular automata, ECA). Whereas it is possible to study all 256 ECAs, we need to sample rules in 2D, in particular for a larger number of states k from the k^{k^5} options. Therefore, we applied one of Langton's sampling schemes, essentially generating a set of rules of increasing 'chaoticity'. They start with a rule table containing only zeros and subsequently randomly increase the fraction of non-zero values, uniformly drawn from the remaining $k - 1$ states. Rules sampled with any of Langton's methods will be referred to as *typical* CA rules here.

We first studied rule table based quantities, in particular Langton's λ, the fraction of non-zero entries in the rule table, and our version of it ('rule table entropy'). Both are only insightful for a large number of states (or large neighborhoods). Then, we turned to 'local measures', which are observables of the CA dynamics starting from a single or an ensemble of random initial conditions. Among the local measures, we found that the time averaged single-site entropy is a quite robust option. For a typical $k = 8$ rule, differences in single-site entropy between runs starting from independent random initial conditions are quite small in our simulations, compared to the maximum entropy of 1, see Fig. 2.7. Our simulations of typical $k = 3$ rules indicated that deviations of the single-site entropy from its extension to larger blocks of sites are small. Further, we observed a strong correlation of the single-site entropy of typical $k = 8$ rules with the time averaged Kolmogorov complexity approximation of the generated patterns, performed by a *string* compression algorithm, see Fig. 2.9. This is probably not generally true, e.g. not for the ECAs[6]. For typical 2D rules with on the order of ten states, single-site entropy also correlates well with difference pattern spreading slope and saturation value, however this is apparently not true for a smaller or larger number of states, see Fig. 2.11. Moreover, the single-site entropy is computationally more efficient than the other local measures discussed in this section. These observations suggest that, if we want to get a first impression of the complexity of patterns generated by a given 2D rule, starting from random initial conditions, the single-site entropy is a good starting point.

Generally, the behavior of a CA rule is strongly dependent on the initial condition — in the most extreme case, the Turing complete CAs, any computation can be performed by tuning the initial conditions accordingly. For many elementary cellular automata, the time evolution can be either cyclic or chaotic, depending on the type of initial conditions. For typical $k = 8$ rules, starting from random grids, we have seen that cycle lengths are initial condition dependent, for instance. Thus, to properly characterize a rule, an approach based on 'all' initial conditions, a global approach, as studied in the previous section, appears favorable. We focus on two

---

[6]ECA simulations by Stephan Kremser comparing different 'complexity' measures to find controllable rules in the sense of (Ramalho, Kremser, Wu, and Gerland 2021) showed that the spatially averaged Kolmorgorov complexity estimate outperforms single-site entropy (personal correspondence, January 2022; planned to be published in his dissertation).

global approaches, 'synergy' between a fixed cell's state and the initial grid as proposed by Quax et al. (Quax, Chliamovitch, Dupuis, et al. 2018) and an attractor based classification by Wuensche (Wuensche 1997). Quax et al. showed that synergy after two time steps calculated for an ECA rule run on all initial conditions can serve as a predictor of the rule's Wolfram Class to some extend. We discussed first experiments with synergy for typical 2D rules. A recipe how Quax work could potentially be extended to 2D for a 'finite time' classification is outlined in the Appendix A.3. Wuensche constructed the attractor space of ECA rules up to a grid length of $L = 15$ (Wuensche, M. Lesser, and M. J. Lesser 1992), as each rule eventually evolves to its limit cycle, latest at the Poincare time. We tested and relativized some of his statements (Wuensche 1998) that draw connections between the scaling of properties of the attractor space with system size and the Wolfram Classification. Further, we performed a principal component analysis of the attractor based observables for fixed grid length and found good agreement with a t-distributed stochastic neighbor embedding visualization of the same data. However, we could hardly find clusters of rules in these representations, also rules of different Wolfram Classes did not separate. It would be interesting to repeat this analysis with scaling properties of the attractor-based observables, obtained from computationally expensive data including larger grids.

However, both approaches only seem to be feasible for small rule spaces on finite grids, small enough that the number of possible grid configurations is still computationally feasible. For infinite grids there exists an abstract attractor based classification. Hurley and Kurka (Hurley 1990; Kůrka 1997) showed that each CA rule, in fact any discrete dynamical system, can be uniquely assigned a class characterizing its attractor-basin field. However, it is generally not possible to infer the class from the rule. It would be exciting if Wuensches finite length approach for increasing lengths could be connected to the limiting behavior studied by Hurley and Kurka.

To conclude our classification survey, despite many efforts in this field, there is still no convincing classification of the cellular automata rule dynamics. The mostly cited one is Wolfram's heuristic classification. However, it is also commonly accepted that Wolfram's classification is initial condition depended, undecidable for some rules, requires a lot of (subjective) studying of rule patterns and is therefore hard to extend to 2D CAs.

So, what have we learned about our initial questions? — Which types of patterns can be generated by a rule? How long does it take? How robust are those patterns? For very small grids and a small number of states, those questions can be answered by Wuensches attractor space construction. For a sufficiently large number of states k in 2D, Langton's λ compared to $\lambda_c = 1 - 1/k$ can give a rough idea of the patterns and transients to be expected of a typical rule: short transients and fixed point patterns for $\lambda \approx 0$, cyclic patterns and increasing transients for larger λ, long transients and either cyclic or chaotic patterns for $\lambda \leq 0.59$ (the percolation threshold of the Von Neumann neighborhood), shorter transients and chaotic patterns for $0.59 < \lambda \leq \lambda_c$. Also the difference-pattern spreading speed, characterizing how sensitive the CA evolution is to small changes in random initial conditions, can be estimated with λ — for sufficiently large λ values.

'Types of patterns' in the theory of elementary cellular automata context often refer to categories like fixed point, cyclic, complex or chaotic, with a focus on complex evolutions. Those complex rules are hypothesized to be related to long transients and sensitivity in initial

conditions (C. G. Langton 1990). However, in biological systems, we are rarely interested in long transients and high sensitivity of the patterning process. In a developmental biology context the listed 'pattern types' seem very crude — what kind of fixed point patterns are possible? Stripes, dots, clusters? Of which patterns do cyclic states consist? How robust are those cycles? In this sense of the question, we have however not gained new insights.

Part II.

Second part: Interplay of Local and Global signals in Developmental Biology

3. Introduction of Part II: Biologically motivated CA Logics

Having started our journey from the pure theory of cellular automata, gaining an overview of their (lack of) meaningful characterization, we now take the opposite approach — studying cellular automata motivated from developmental biology. Each biological cell in a two- or one-dimensional tissue is modeled as a stochastic cellular automaton. A CA's state either corresponds to the 'cell type' or to the gene-expression state (simplified to 'On' or 'Off') of a single gene. Common properties of developmental systems are (1) stochasticity, (2) ordered final patterns, and (3) local and global signaling. The following discussion of those three properties focuses on biological examples and assumes familiarity with signaling pathways in developmental biology. For an introduction to this topic, see Sec. 3.2.

(1) Noise is ubiquitous in cellular biology, from signal molecule production, signal transmission, to readout. For example, signal molecule production appears to be subject to stochastic variations in molecule synthesis and secretion (Raser and O'Shea 2005; Bollenbach, Pantazis, Kicheva, et al. 2008). Transport of long-range signaling molecules by diffusion is a stochastic process itself, and might moreover be hindered by barriers (Restrepo, Zartman, and Basler 2014; Stapornwongkul and J. P. Vincent 2021). At the signal-uptake stage, cell-cell variability in the number of receptors for the signaling molecule, binding of molecules to receptors, and receptor occupancy are additional sources of noise (Jaeger, Irons, and N. Monk 2008; Bollenbach, Pantazis, Kicheva, et al. 2008; Colman-Lerner, Gordon, E. Serra, et al. 2005). Finally, activation of the signaling pathways, and induced gene expression regulation are noisy processes (Raser and O'Shea 2005). We therefore extend our explorations to *stochastic* cellular automata.

(2) Examples of patterns are gene-expression boundaries, stripes, dots, or more irregular but stationary, patterns. Cell-type or gene-expression boundaries are well studied in the *Drosophila* wing disc such as the anterior-posterior, and dorsal-ventral boundary formation (Aliee, Röper, Landsberg, et al. 2012; Michel, Aliee, Rudolf, et al. 2016). *Drosophila* wing-vein formation (Bier 2000; De Celis 2003; Crozatier, Glise, and A. Vincent 2004; Blair 2007; Sprinzak, Lakhanpal, Lebon, et al. 2010b), and gap-gene patterns in the early *Drosophila* embryo (Gregor, Tank, Wieschaus, and Bialek 2007) are examples of stripe formation. Dots in the form of sensory organ precursors are singled out during *Drosophila*-bristle formation (Corson, Couturier, Rouault, et al. 2017), an example we will study in more detail in Sec. 5.1. Non static, but stationary patterns include the small-intestine epithelial cell-type pattern (Buske, Galle, Barker, et al. 2011; Kai, Trw, Dongen, and Parsy 2021; Takahashi and Shiraishi 2020; Zhou, Ramachandran, Mansouri, and Dailey 2018; Howitt, Lavoie, Michaud, et al. 2017; Gassler 2017; Mah, Yan, and Kuo 2016), that we will model in Sec. 5.2. We observe that patterns are static or cyclic, thus cellular automata rules with chaotic or complex behavior are of less biological interest.

(3) 'Global signaling' includes long range morphogens, such as Dpp (Bollenbach, Pantazis,

Kicheva, et al. 2008) and bicoid (Gregor, Tank, Wieschaus, and Bialek 2007), electric potentials (Levin 2021) or stress on the tissue (Aliee, Röper, Landsberg, et al. 2012). With 'local signaling' we refer to any information exchange between spatially neighboring cells, be it via diffusive, small molecules, or ions channels, mechanical cues or interaction of membrane-bound proteins. Suggested 'communication codes' include (i) sensing of signal identity, (ii) sensing of signal concentration, (iii) combination of different inputs, and (iv) dynamics (P. Li and Elowitz 2019). We will give a more detailed overview of chemical signaling during development in Sec. 3.2. This large variety of signal processing modes in this wider sense justifies the generality a CA modeling approach. Also, it motivates the extension of CA models to process an external input from a global signal.

We begin with the arguably simplest class of rules that we can motivate biologically — CA rules inspired by sigmoidal gene-expression regulation (Alon 2007), idealized to a single-threshold rule. Our minimal boundary formation CA uses only this single parameter. The next step is a rule with two thresholds, depending on the state of a cell, used in our *Drosophila* bristle-formation model. For the small-intestine model, we allow for the full rule space, and try to derive the update rule by comparison to experimental data.

The knowledge gap we are motivated by is that in developmental biology cells process signals from a small number of different pathways, but collectively perform a broad variety of robust patterning processes (P. Li and Elowitz 2019). On the other hand, cellular automata process few different input components, but can, in principle, generate any pattern. Thus we here explore cellular-automata models of biological cells.

Part II of this thesis is organized as follows: After outlining other cellular modeling approaches in developmental biology, and a short overview of chemical signaling pathways, we discuss three biological pattern formation systems modeled in a CA framework as named above. We start with single-threshold logics to study the example of gene expression boundary formation in Chap. 4, and continue with two threshold logics for *Drosophila* bristle formation (Sec. 5.1). Finally, we give an example where potential update rules, summarizing differentiation, and cell movement, are inferred from a stationary spatial cell-type distribution: our model the small intestine epithelium (Sec. 5.2).

3.1. Other CA Models for developmental Biology

A great advantage of CA models is their ability to bridge scales between the behavior of individual, interacting cells, and the development of tissues, requiring only a minimal set of parameters (Lehotzky and Zupanc 2019; Koopmans and Youk 2021). The models can then be used to generate hypotheses to be tested in experiments. In the following, we give some examples of CA models in developmental biology. For a review of current agent-based models that are not cellular automata, see Ref. (Shaebani, Wysocki, Winkler, et al. 2020).

The maybe most visual cellular automata are the skin-scale colors of the ocellated lizard (Manukyan, Montandon, Fofonjka, et al. 2017). Their labyrinthine pattern can be modeled by a quasi-hexagonal stochastic CA. Different to the focus of this work, skin scales consist of multiple biological cells.

Adhyapok, Fu, Sluka et al. developed a stochastic 1D and 2D CA model of liver lobule injury from experimental data with three different cell types — healthy, stressed, dead — and update

rules modeling proliferation, cell death, and conversion from healthy to stressed (Adhyapok, Fu, Sluka, et al. 2021). They distinguished time scales for each process, and performed a parameter sweep (always keeping one parameter fixed). From those phase diagrams, they could make predictions about evolution to tissue damage, recovery, and potential divergent fates.

In nervous tissue development, Lehotzky and Zupanc modeled three systems: enteric nervous system formation, neurosphere growths, and neutral fate specification (Lehotzky and Zupanc 2019). Their stochastic CA model included mitosis, cell movement, and differentiation after mitosis.

A stochastic 2D CA model of two types of photoreceptors (yellow and pale) in the fly eye that could reproduce the qualitative distributions, i.e., random vs. alternating stripes of these two receptors in the *Drosophila* and *Dolichopodidae* retinas, is put forth in Ref. (Ebadi, Perry, Short, et al. 2018). The model used a threshold CA with five parameters. Rows of retina cells are patterned subsequently, thus, the CA only performs a single update step.

Nissen, Perera, Gonzalesz, et al. have modeled mammalian blastocyst generation in an agent-based model including differential adhesion with four rules, the second one corresponding to a CA (Nissen, Perera, Gonzalez, et al. 2017): 'Switch fate if too many (next-nearest) neighbors are in the same state', resulting in differentiation of the inner cell mass to primitive endoderm and epiblast cells in a salt-and-pepper pattern, which was rearranged by subsequent rules. Those rules were qualitatively confirmed by experiments.

These examples cover very different phenomenons, and have the stochastic approach in common. However, given the broad applicability of CA modeling to stem-cell differentiation, the small number of studies using this approach is rather surprising (Lehotzky and Zupanc 2019). We will add three more examples in the following, after giving a short introduction on chemical signaling.

3.2. Chemical signaling Pathways in developmental Biology

Chemical signaling pathways take external signals as input, and output, often switch like, gene-expression changes within a cell. External signals, called ligands, bind to receptors in the receiving cell membrane, and thereby trigger a cascade of biochemical reactions within the cell, that result in transcription factors down-, or up-regulating transcription of a set of genes. For chemical signaling during development, more than 17 kinds of signal transduction pathways are known. However, only five of them dominate in early embryonic development of animals: Notch, Transforming Growth Factor beta (TGF-β including BMP and Dpp), Wnt, Hedgehog (HH) and Fibroblast growth factor (FGF) families (Gerhart 1999). An essential difference between pathways is their communication range: Long-ranged via the circulatory system — 'endocrine', traveling a few cell widths — 'paracrine', or between neighboring cells — 'juxtacrine'. If cells can sense their own signal, it is an 'autocrine' pathway. Hormones are an example for endocrine signaling. TGF-β, Wnt, HH and FGF pathways are paracrine signaling pathways. Their ligands are well suited to form concentration gradients that can encode positional information, and are then called 'morphogens'. Juxtacrine signaling, e.g., the Notch pathways, requires direct contact between signaling cells.

The pathways allow for a large variety of distinct functions in different contexts. A number

of 'communication codes' have been suggested (P. Li and Elowitz 2019) to enable their multipurpose function: (i) Ligand identity: Different ligands for the same receptor can induce different differentiation outputs, e.g., Dll1 and Dll2 in the Notch pathway. Ligand specificity can be modulated, e.g., by co-receptor enzymes. (ii) Ligand concentration: Not only absolute concentration levels of the signaling molecule can be measured, but also fold changes, in some examples. Fold-change measurements are more robust to cell-to-cell variability (Adler and Alon 2018). (iii) Combination of different inputs: The bone-morphogenetic protein (BMP) pathway is an example where different ligands, or different receptors can interact in an additive, ratiometric, or imbalance-detection mode (P. Li and Elowitz 2019). An example of an orthogonal pathway is sonic hedgehog (SSH), a HH pathway, and BMP in the developing neural tube. The two long range morphogens form anti-parallel dorsal-ventral gradients and encode at least ten different cell fates along this axis (Zagorski, Tabata, Brandenberg, et al. 2017). At the transcription-factor stage of the signal processing, known 'logics' of combining factors include AND, OR, NAND and SUM (Buchler, Gerland, and Hwa 2003; Alon 2007). In synthetic biology, possibilities are manifold, see (Bashor, Patel, Choubey, et al. 2019), and (Toda, Frankel, and Lim 2019; Chen and Elowitz 2021) for recent reviews. (iv) Dynamics: Also amplitude, duration, and modulation of the morphogen gradient can contribute to the signal. For SHH signaling, evidence has been presented that both the amplitude and duration determine neural-progenitor cell fates in chick and mice (Dessaud, Yang, K. Hill, et al. 2007). Further, a transient Delta prepattern in *Drosophila* bristle formation was observed by Corson et al. (Corson, Couturier, Rouault, et al. 2017).

After having gained an impression of the variety of modes in signaling pathways we focus on the probably most well studied example: Delta-Notch signaling. This is the pathway we will revisit in the *Drosophila*-bristle and small-intestine model.

A short Introduction to Delta-Notch Signaling

The Delta-Notch signaling is a highly conserved signaling pathway among animals (Artavanis-Tsakonas, Rand, and Lake 1999), and is used in diverse functions such as control of cell differentiation, proliferation, and migration during development (Boareto 2020; Artavanis-Tsakonas, Rand, and Lake 1999). By regulating cell differentiation it contributes to create qualitatively different patterns, such as dots during *Drosophila* bristle formation (Corson, Couturier, Rouault, et al. 2017) or sharp boundaries for *Drosophila* wing vein patterns (De Celis 2003).

Delta and Notch are transmembrane protein families acting as ligand (Delta) and receptor (Notch) of the signaling pathway. It has been shown (Sprinzak, Lakhanpal, Lebon, et al. 2010a) that (i) Delta transactivates Notch in neighboring cells, and (ii) cis-inhibits Notch in its own cell. Moreover, evidence is presented in Ref. (Sprinzak, Lakhanpal, Lebon, et al. 2010a) that (iii) Delta is also cis-inhibited by Notch, i.e., the cis-inactivation is mutual. Also, (iv) Notch can *transcriptionally* down-regulate cis-Delta (De Celis and Bray 1997; Huppert, Jacobsen, and Muskavitch 1997). In order to better understand all those interactions, we provide a more formal version of the above statements, based on (Sprinzak, Lakhanpal, Lebon, et al. 2010b).

Generally, Delta D_i and Notch N_i in a cell i are produced at rate $\beta_D(x)$, (possibly depending

on position x), and β_N, respectively, and degraded at rate γ_D and γ_N:

$$\dot{D}_i = \beta_D(x) - \gamma_D D_i + other$$
$$\dot{N}_i = \beta_N - \gamma_N N_i + other\,,$$

where '$+other$' is a placeholder for remaining terms, that we will subsequently introduce. (i) During intercellular signaling, Delta D_j in cell j binds to N_i in a different cell i, leading to the release of the Notch intracellular domain S_i, $\dot{S}_i = N_i D_j / k_t + other$, and degradation of its extracellular domain $\dot{N} = -D_j N_i / k_t + other$, with k_t the transactivation strength. (ii) Delta D_i cis-inhibits Notch N_i in its own cell at cis-inhibition strength k_c, $\dot{N}_i = -D_i N_i / k_c + other$. (iii) Notch also cis-inhibits Delta, i.e. $\dot{D}_i = -D_i N_i / k_c + other$. Moreover, (iv) Notch N_i can, by transcription of a 'reporter' R_i, down-regulate the rate of Delta production in cell i, modeled as a repressive Hill function $f_R(R_i; \beta_D, m, k_D R) = \beta_D \frac{k_{DR}^m}{k_{DR}^m + R_i^m}$ with Hill coefficient m and capacity $k_D R$. Transcription of the reporter R_i ('Notch response') is modeled as Hill type activation function f_A with Hill coefficient p and capacity k_{RS} and is degraded at rate γ_R. Putting it all together for a grid of cells, and denoting the summed X_j contribution of cells j that are neighbors of cell i by $\langle X_j \rangle_i$, with $X_j = D_j$ or N_j, (i.e. $\langle X_j \rangle_i \equiv \sum_{\langle i,j \rangle} X_j$), we obtain:

$$\dot{N}_i = \beta_N - \gamma_N N_i - N_i \frac{D_i}{k_c} - N_i \frac{\langle D_j \rangle_i}{k_t}$$
$$\dot{D}_i = f_R(R_i; \beta_D, m, k_{DR}) - \gamma_D D_i - N_i \frac{D_i}{k_c} - \langle N_j \rangle_i \frac{D_i}{k_t}$$
$$\dot{S}_i = N_i \frac{\langle D_j \rangle_i}{k_t} - \gamma_S S_i$$
$$\dot{R}_i = f_A(S_i; \beta_R, p, k_{RS})\,.$$

This model can reproduce the following experimental findings (Sprinzak, Lakhanpal, Lebon, et al. 2010a): A graded reporter response R of Notch activation by Delta transactivation, a sharp trans-Delta independent inhibition of Notch response R by cis-Delta and lateral inhibition, even without cooperativity (in contrast to the 'classic' model by Collier et al. (Collier, N. A. Monk, Maini, and Lewis 1996)).

However, the above model of Delta *Dll1* and NOTCH1 interaction is just one piece of the puzzle. Within the mammalian pathway alone, there are four different Notch receptors interacting in a promiscuous fashion with multiple Delta ligands, including *Dll1* and *Dll4* (P. Li and Elowitz 2019). For *Dll1* and *Dll4* it has been shown that they can have opposite effect on cell fate, either promoting or inhibiting myogenesis in neural crest cells of chick embryos (Nandagopal, Santat, LeBon, et al. 2018). Also it is still not quite understood how in the presence of intrinsic noise (Yaron, Cordova, and Sprinzak 2014), the Notch pathway can generate qualitatively different patterns such as dots and stripes (Sprinzak, Lakhanpal, LeBon, et al. 2011). So, despite its importance, not all modes of Notch signaling are known yet (Boareto 2020).

4. Boundary Formation

This chapter is based on and uses parts of the manuscript: "Robust boundary formation in a morphogen gradient via cell-cell signaling", by M. Bojer, S. Kremser, U. Gerland, *which has been submitted for publication.*

4.1. Introduction

The formation of sharp boundaries between different tissues at their precisely determined positions is fundamental in the development of an embryo. Groups of cells with distinct functions often must be kept physically separated. Also, boundary cells act as organizing centers for subsequent patterning processes in many cases (Dahmann, Oates, and Brand 2011). In the presence of noise inherent to biological processes it is not fully understood how the observed precision is achieved (Exelby, Herrera-Delgado, Perez, et al. 2021; Lander 2013; Gregor, Tank, Wieschaus, and Bialek 2007; Bollenbach, Pantazis, Kicheva, et al. 2008; Jaeger, Irons, and N. Monk 2008) Here, we investigate the potential contributions and optimal design of local signaling as a correction mechanism to ensure a sharp and straight, correctly positioned boundary.

Common ancestor of many boundary formation models is the 'French Flag' model (historically 'French Flag problem') (Wolpert 1969). The basic idea is that a morphogen gradient activates cell-fate-determining genes subject to different thresholds and thereby patterns the tissue.

However, each process — morphogen production (Bollenbach, Pantazis, Kicheva, et al. 2008; Raser and O'Shea 2005), morphogen transport (Restrepo, Zartman, and Basler 2014; Staporn-wongkul and J. P. Vincent 2021), morphogen uptake (Jaeger, Irons, and N. Monk 2008; Bollenbach, Pantazis, Kicheva, et al. 2008; Colman-Lerner, Gordon, E. Serra, et al. 2005) and signaling (Raser and O'Shea 2005) — is subject to noise.

On the other hand, short range signaling is commonly observed during embryogenesis, such as short range morphogens, Delta-Notch signaling and ion transport via gap junctions or ion channels (Levin 2021). To our knowledge, the potential of short range signaling for boundary formation has not been systematically explored yet. Lander does mention spatial pooling, but in a system where the cell's state, once set, cannot change anymore with morphogen concentration (Lander 2013). Also, in the well studied model system for morphogenesis, *Drosophila*, (Gregor, Tank, Wieschaus, and Bialek 2007) and (He, Saunders, Wen, et al. 2010; Erdmann, Howard, and Ten Wolde 2009; Okabe-Oho, Murakami, Oho, and Sasai 2009) employed the the idea of spatial averaging to explain the observed high precision in expression of the (gap) gene, Hunchback: Hunchback is induced in a concentration dependent manner by the morphogen Bicoid during early Drosophila syncytia stage. The suggested mechanism consists of nuclei-nuclei communication via a morphogen of comparable diffusion constant as Bicoid.

Our aim here is to identify underlying principles, rather than to restrict ourselves to a specific biological model system. To this end (Hillenbrand, Gerland, and Tkačik 2016) studied the positional information given by a morphogen gradient combined with a local interaction between neighboring cells that could be either repulsive or attractive in one dimension. Their model is based on an equilibrium Ising model with external field. It thus neglects any dynamics during boundary formation and it assumes a specific signal integration logic.

A conceptual and systematic study of how local signaling can optimally be incorporated as a correction mechanism for the global morphogen signal still seems to be missing. Here, we investigate which signal processing logic meets best the following four key properties of boundary formation, motivated by morphogenesis, in each noise regime: (i) Reduction in boundary fuzziness, to ensure a sharp separation of different tissues. (ii) A short transition time to stationary boundary position. Fast development generally is beneficial for the survival of the embryo, e.g. to escape predators. (iii) A tuneable boundary position, as required, if the same logic shall allow for observed variations of the same pattern in related species. (iv) Scaling with system size, to conserve pattern proportions among embryos of different sizes. Therefore we put forth a minimal model of tissue boundary formation based on signaling and neglect cell migration, proliferation, cell death and cell shape change.

Assuming that the morphogen signal (global signal) is accompanied by a signal of short range (local signal) we want to explore how to optimally combine these two signals with respect to the criteria (i)-(iv). We focus on signal integration mechanisms that are parameterized by a minimal number of variables, the morphogen signal gradient slope m and a threshold a for the signal. The three candidates consist of signal integration by a SUM, AND or OR logic.

These logics can also be motivated by common input functions to gene transcription (Mayo, Setty, Shavit, et al. 2006; Buchler, Gerland, and Hwa 2003; Bolouri and Davidson 2002): either both signals, or their products within the signal processing pathway, can occupy the same promoter — a regulation scheme modeled by the SUM logic, such as in (Kalir and Alon 2004). Or there is a different promoter for each signal such that either both have to be occupied to switch on gene transcription, corresponding to the AND logic, or occupation of one is sufficient implying an OR logic.

In order to exclude effects independent or not essential to the local signaling, we constructed a minimal model where the state of a cell is reduced to one gene either switched On or switched Off and the tissue is reduced to a regular quadratic 2D grid of fixed shape.

The model presented here is not limited to regulation of a cell's transcriptional state by chemical signaling. It also applies to bioelectrical signaling during embryogeneis for patterning processes involving a long ranged electric gradient (Levin 2021). Cell-cell signaling is then performed via ion channels and gap junctions. The models' simplicity also allows us to systematically test the performance of the logics combining the long and short range signal for all morphogen signal slopes m and thresholds a over a large range of noise levels on both signals.

In the following we show that (i) correction by using information from neighbors outperforms the pure gradient mechanism for nearly all noise regimes although it adds an additional source of noise to the system. Among the correction mechanisms, the SUM logic performs best for sufficiently large noises, but convergence to the correctly positioned boundary takes longer for lower noise levels. The AND and OR logic perform equivalently. (ii) Correction

a)

b)

[GRAD] if   **global signal** $> a$:

[SUM] if (**local signal**   **+**   **global signal**) $> a$:

[AND] if (**local signal** $> a^{AND}$) **AND** (**global signal** $> a$):

[OR] if (**local signal** $> a^{OR}$)   **OR** (**global signal** $> a$):

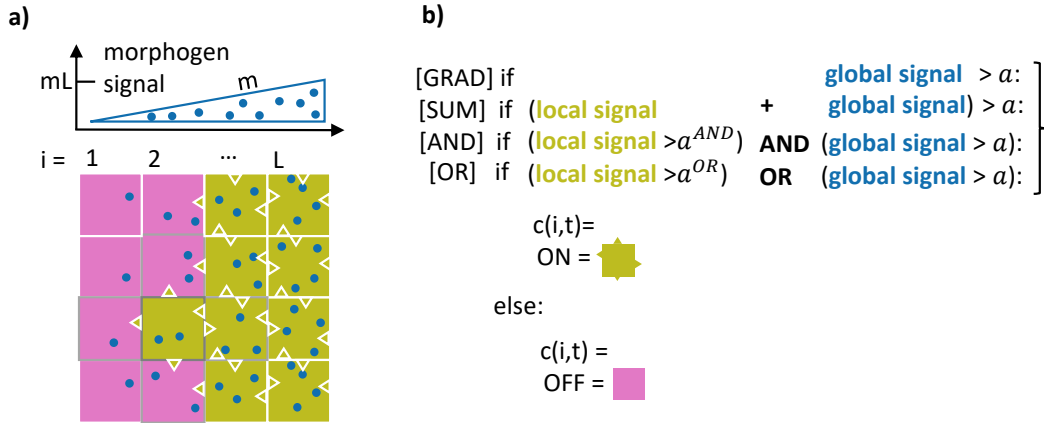$c(i,t)=$
ON =

else:

$c(i,t) =$
OFF =

Figure 4.1.: (a) Sketch of the minimal model. The morphogen signal is represented by blue dots and nearest neighbor interaction by green triangles. A green cell is in state On and signals this to its direct neighbors, a pink cell is Off, i.e. not signaling. Periodic boundary conditions in y direction are depicted as arrows below the grid. The left boundary at $i = 0$ is fixed to Off cells, while the right boundary at $i = L + 1$ is fixed to On cells. (b) Summary of the three logics: (SUM) Given a cell at position i at time i in state $c(i,t)$: If the local signal including Gaussian noise plus the global signal including noise exceed a global threshold a the cell state c at the next time step $t + 1$ is On, else Off. (AND/OR) If the local signal exceeds a local threshold AND/OR the global signal exceeds the global signal, $c(t + 1) =$On. For example, the green cell highlighted by a dark gray frame in (a) senses the local signal from it's upper, lower, left and right neighbor (light gray frames) canceling to one Off state plus local noise as well as a morphogen concentration of $2m$ plus global noise.

mechanisms strongly differ with respect to the time it takes to reach the stationary boundary position for a given noise level. The stationary boundary can be tuned to any position within the grid by varying the global threshold (iii) and does scale linearly with system size (iv) for all logics for a fixed morphogen slope. As boundary formation is a fundamental patterning mechanism, insights about the potential of a local signal as a correction mechanism might find applications in synthetic biology for constructing various patterns.

4.2. Model and Observables

Model

We consider a square grid of $L \times L$ cells, with cylindrical boundary conditions, see Fig. 4.1. The state of a cell is reduced to either 'On' or 'Off', $c = +\frac{1}{2}$ or $-\frac{1}{2}$. Along the axis of the gradient (index i) the boundary conditions of our grid are fixed to Off on the left ($i = 0$) and On on the right side of the grid ($i = L$). In the perpendicular direction (index j) we apply periodic boundary conditions. The cell state is updated according to a signal processing rule

ϕ, which we also refer to as signal integration rule. The state can change at discrete time steps. The rule processes two different signals: the global signal at the cells position (i, j), $s_{ij}^G(t)$, representing the morphogen gradient, and a local signal, $s_{ij}^L(t)$, encoding the state of the cells within c_{ij}'s neighborhood,

$$c_{ij}(t+1) = \phi \left[s_{ij}^G(t), s_{ij}^L(t) \right] . \tag{4.1}$$

A Boolean logic is the common simplification of the biologically observed Hill type regulation, as the sigmoidal form becomes a sharp threshold in the limit of large Hill coefficients (Bolouri and Davidson 2002).

Global signal   The stochastic global signal at a cell with index (i, j) is given as

$$s_{ij}^G(t) = m\,i + \xi_{ij}^G(t), \tag{4.2}$$

with m the morphogen gradient slope at i and ξ^G additive Gaussian white noise with mean zero and standard deviation σ^G.

   During embryogenesis, morphogen molecule concentration is commonly assumed to be exponentially decaying within the tissue. Inspired by Ref. (Hillenbrand, Gerland, and Tkačik 2016), we interpret the logarithm of the molecule concentration as the actual signal (Weber-Fechner law), resulting in a linear morphogen gradient signal.

French Flag mechanism   We refer to a signal processing rule that only depends on the global signal and compares it to a global threshold a as *pure gradient* rule $\phi^{\text{GRAD}}\left[s^G\right]$. It is the analogue of the French Flag mechanism (Wolpert 1969). More precisely

$$\phi^{\text{GRAD}}\left[s_{ij}^G(t)\right] := \Theta\left[s_{ij}^G(t) - a\right] - \frac{1}{2}, \tag{4.3}$$

with Θ denoting the Heaviside step function with convention $\Theta\left[0\right] = 0$. The state of a cell at position i at $t+1$ is $+\frac{1}{2}$ if the global signal exceeds the global threshold a and $-\frac{1}{2}$ else.

Local signal   The signal processing rules with correction ability additionally make use of a local signal s^L that stems from nearest neighbor cells communicating their state. We conservatively assume that the central cell cannot sense from which neighbor the signal came from and thus define s^L to be the sum of these signals

$$s_{ij}^L(t) = \sum_{(k,l)\,\in\,\text{neighbors}(i,j)} c_{kl}(t) \quad + \xi_{ij}^L(t). \tag{4.4}$$

'neighbors' refers to the upper, lower, left and right neighbor (von Neumann's neighborhood). ξ^L is chosen to be Gaussian white noise with a mean of zero and standard deviation σ^L.

Correction mechanisms   To implement a correction mechanism, each cell needs to combine the two noisy signals s^L and s^G. It is by no means clear how this combination is optimally

performed. Straight forwardly, we can add up both signals and compare the result to the global threshold *a*. We will refer to this procedure as SUM rule ϕ^{SUM},

$$\phi^{\text{SUM}}\left[s_{ij}^L(t), s_{ij}^G(t)\right] = \Theta\left[s_{ij}^L(t) + s_{ij}^G(t) - a\right] - \frac{1}{2}. \tag{4.5}$$

Note that the contribution of the local signal to the full signal can take any value by rescaling *m* and *a* simultaneously.

Alternatively, both signals could be processed separately and the results combined by an AND or OR rule. SUM, AND and OR rule are not the only options to process two signals. Others are XOR and PROD, but we can argue that they are not suited for the boundary formation problem as we modeled it.

Let us start from an all Off grid with an XOR logic. In the next time step without noise it would form the correct boundary. In the consecutive update step, all On cells except those at the boundary would switch Off though, as each is subject to a local neighbor signal greater than any (sensible) local threshold value. Consequently the boundary would not be stable.

The product rule PROD in the presence of noise reads

  if   (global signal(i) $+\xi^G$)· (local signal(i,t) + ξ^L) > a^2:   which implies that the noise would
       cell(i,t+1) = On

be multiplied by the signal. Consequently, we expect this rule to perform poorly in the presence of sufficiently large noise.

Processing the local signal separately requires an additional threshold, a^{AND} respectively a^{OR},

$$\phi^{\text{AND}}\left[s_{ij}^L(t), s_{ij}^G(t)\right] = \Theta\left[s_{ij}^L(t) - a^{\text{AND}}\right]\Theta\left[s_{ij}^G(t) - a\right] - \frac{1}{2},$$

$$\phi^{\text{OR}}\left[s_{ij}^L(t), s_{ij}^G(t)\right] = \Theta\left[s_{ij}^L(t) - a^{\text{OR}}\right] + \Theta\left[s_{ij}^G(t) - a\right]$$

$$- \Theta\left[s_{ij}^L(t) - a^{\text{OR}}\right]\Theta\left[s_{ij}^G(t) - a\right] - \frac{1}{2}.$$

In the following paragraph we show that it makes sense to choose $a^{\text{AND}} = -1$ and $a^{\text{OR}} = 1$.

a_{local} **optimization**   We want the AND and OR rule to be able to produce a boundary from an arbitrary initial grid for all noise levels, equivalently to the pure gradient mechanism. Here, we argue that the initial grids all Off and all On are sufficient to fix the additional local thresholds a^{AND} and a^{OR}.

Let us consider an all Off initial grid. At the right border, $i = L$, the global signal exceeds the global threshold *a* (otherwise the pure gradient rule could not form a non-trivial boundary either). For the AND rule to exit the initial condition, we need a^{AND} to be smaller or equal than the local signal $s_{ij}^L(t) = -1 + \xi_{ij}^L(t)$. Thus, we need $a^{\text{AND}} \leq -1$. Similarly, for an all On initial grid it follows that $a_l^{\text{OR}} \geq 1$. The second condition to determine the optimal local threshold comes from demanding that it stabilizes a straight boundary. To this end, consider a straight boundary, implying that the global signal is close to *a*, but with one On cell left of the boundary. For the AND rule, we want $a^{\text{AND}} \geq -1$ in order to switch Off the defect cell. Taken together this suggests choosing $a^{\text{AND}} = -1$. The opposite scenario, one Off cell right of

the boundary, yields $a^{\text{AND}} < 1$, which is well satisfied by our choice. Equivalent reasoning yields $a^{\text{OR}} = 1$.

We confirmed these analytic arguments numerically for an exemplary small and large noise level, see Fig. 4.2 and Fig. 4.3. We observe that the fuzziness decreases with increasing a^{AND} values. In the small noise example, $a^{\text{AND}} = -1$ is the largest a^{AND} value such that the AND rule forms a non-trivial boundary (i.e. $\frac{\mathcal{B}}{L} \neq 1$ or 0) independent of starting from an all On grid (red line) or an all Off grid (green line). If we drop the initial grid independence condition, e.g., if it suffices that the AND rule only patterns when starting from an all On initial grid, the optimal choice of a^{AND} would depend on the magnitude of deviation of boundary position from the one generated by pure gradient rule (dashed gray line) we are willing to accept. This is deviation is particularly pronounced for large noise as shown in Fig. 4.3. Results for the OR rule are depicted in the right column and findings are analogous.

For a choice of local thresholds such that $a^{\text{AND}} = -a^{\text{OR}}$ we observe in simulations, and will show a direct 'particle-hole' correspondence between the AND and the OR rule.
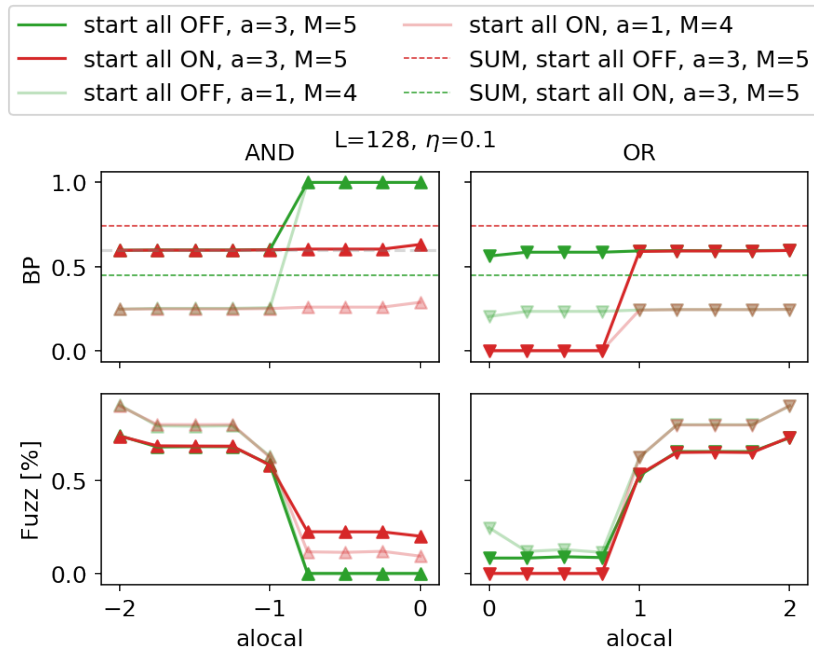
Figure 4.2.: *Left column*: Relative boundary position BP and fuzziness Fuzz with respect to the local threshold a^{AND} for two different a and $M = mL$ combinations, one plotted full saturation, one light. Green (red) triangles show results for an initial all Off (all On) grid, interpolated by solid lines. The dashed gray line shows the pure gradient BP. The first column shows that the all Off initial condition can only be exited for $a^{\text{AND}} \leq -1$. The second row shows the fuzziness decrease with increasing a^{AND}. *Right column*: Analogous results for the OR rule. A small noise level of $0.1 = \eta = \sqrt{3}\sigma$ is used.
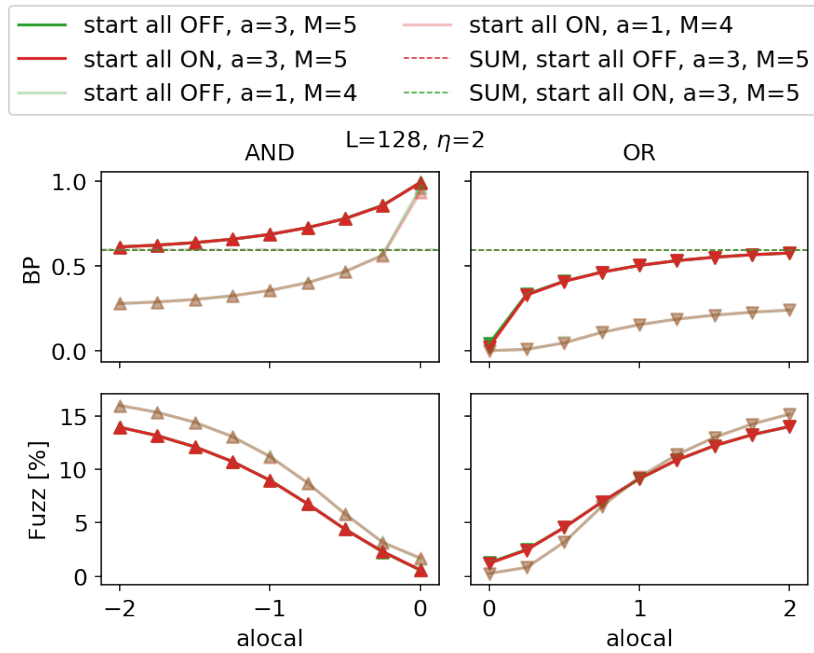

Figure 4.3.: Same quantities as in Fig. 4.2, but for a large noise level. Here, both initial conditions yield the same boundary position independent of a^{AND} choice, but the boundary position deviates strongly from the pure gradient value.

Relationship of AND and OR rule for $a^{\text{AND}} = -a^{\text{OR}}$ The choice of $a^{\text{AND}} = -a^{\text{OR}}$ implies close correspondence of the AND and the OR rule in the stationary state. For an infinite grid we have

$$\langle c_{ij}^{\text{AND}} \rangle = -\langle c_{\tilde{i}j}^{\text{OR}} \rangle \quad \text{with} \quad \tilde{i} = 2\frac{a}{m} - i,$$

where $\langle \rangle$ denotes an ensemble average. Visually speaking $\tilde{i}$ is the mirror reflection of i at the zero transition of the morphogen gradient minus its threshold at $\frac{a}{m}$. For a finite grid, this relation still holds true for parameter combinations a, m such that the boundary is distant from edges of the grid. Then we can assume that cells not covered by the $\tilde{i}$ index, which are cells close to the grid-boundaries, do not change their state.

We can derive the above relation as follows:

$$c_{i,j}^{\text{AND}} = \phi^{\text{AND}} \left(s_{ij}^L(t), s_{ij}^G(t) \right)$$
$$= \Theta \left(s_{ij}^L(t) - a^{\text{AND}} \right) \Theta \left(s_{ij}^G(t) - a \right) - \frac{1}{2},$$

whereas

$$-c_{\tilde{i},j}^{\text{OR}} = -\phi^{\text{OR}} \left(s_{\tilde{i}j}^L, s_{\tilde{i},j}^G \right)$$
$$= -\left\{ 1 - \left[1 - \Theta \left(s_{ij}^L(t) - a^{\text{OR}} \right) \right] \right.$$
$$\left. \left[1 - \Theta \left(s_{\tilde{i},j}^G(t) - a \right) \right] \right\} + \frac{1}{2}$$
$$= \Theta \left(-s_{ij}^L(t) + a^{\text{OR}} \right) \Theta \left(-s_{\tilde{i},j}^G(t) + a \right) - \frac{1}{2},$$

using that $1 - \Theta(x) = \Theta(-x)$.

Now observe that

$$\Theta \left(-s_{\tilde{i},j}^G(t) + a \right) = \Theta \left(-s_{\tilde{i},j}^G(t) + a \right)$$
$$= \Theta \left(-\tilde{i}m - \xi_{\tilde{i},j}^G + a \right)$$
$$= \Theta \left(im - \xi_{ij}^G(t) - a \right),$$

where we can neglect the sign change for $\xi_{ij}^G(t)$ as it is symmetric around its zero mean. Thus, the contribution by the global signal is by construction of $\tilde{i}$ the same as in the case of the AND rule.

In the stationary state, we have on ensemble average that

$$\left\langle -\sum_{(k,l) \in \mathcal{V}(\tilde{i},j)} c_{k,l} \right\rangle = \left\langle \sum_{(k,l) \in \mathcal{V}(i,j)} c_{k,l} \right\rangle$$

as the global signal is mirror antisymmetric with respect to the vertical $i = i_c$ line and the local signal is independent of the absolute position.
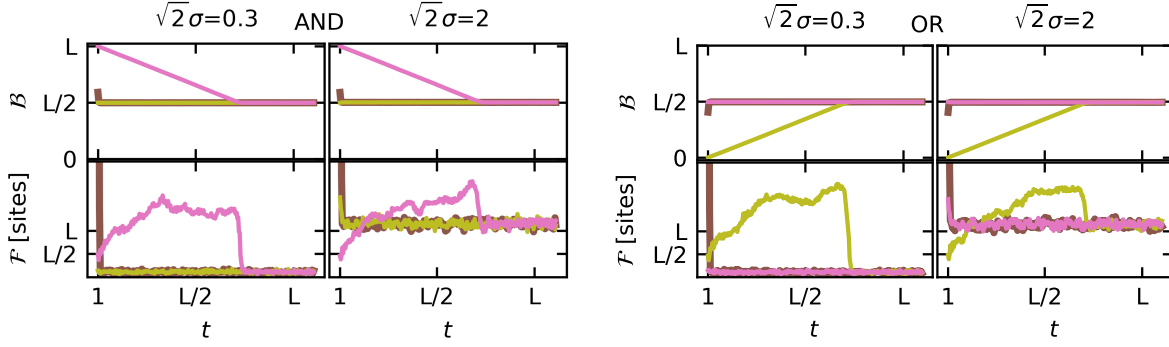
Figure 4.4.: Exemplary time traces from simulation with same parameters as in the main text. Green, pink and brown lines are for all Off, all On and random initial conditions, respectively.

Thus,

$$
\left\langle \Theta\left(-s^L_{\tilde{i},j}(t) + a^{\mathrm{OR}}\right) \right\rangle
$$

$$
= \left\langle \Theta\left(- \sum_{(k,l)\in\mathcal{V}(\tilde{i},j)} c_{k,l} - \xi^L_{\tilde{i},j} + a^{\mathrm{OR}}\right) \right\rangle
$$

$$
= \left\langle \Theta\left(\sum_{(k,l)\in\mathcal{V}(i,j)} c_{k,l} + \xi^L_{ij}(t) - a^{\mathrm{AND}}\right) \right\rangle .
$$

Inserting those observations yields

$$
\langle c^{\mathrm{AND}}_{ij} \rangle \approx -\langle c^{\mathrm{OR}}_{ij} \rangle, \quad \text{with} \quad \tilde{i} = 2\frac{a}{m} - i. \tag{4.6}
$$

Exemplary time traces from simulation confirming this relation are shown in Fig. 4.4.

Transforming and reducing the parameter set   The parameters characterizing the model are the grid length L, morphogen gradient slope m, global threshold a and the standard deviations of the global and local noise σ^G and σ^L. In order to arrive at a description in more natural parameters, we transform m, a to m, $\frac{a}{m}$ as $\frac{a}{m}$ corresponds to the spatial position where the morphogen signal equals its global threshold. Also, we transform the independent noises to a total noise and the relative contributions. The total noise is defined as $\xi := \xi^L + \xi^G$ with standard deviation σ and local to total noise ratio α defined as

$$
\sigma^2 := \left(\sigma^G\right)^2 + \left(\sigma^L\right)^2, \quad \text{and } \alpha := \frac{\left(\sigma^L\right)^2}{\sigma^2}. \tag{4.7}
$$

If not stated otherwise, we set the local to total noise ratio α to $\alpha = \frac{2}{mL+2}$ as 2 is the maximal deterministic local signal and mL the maximal global signal.
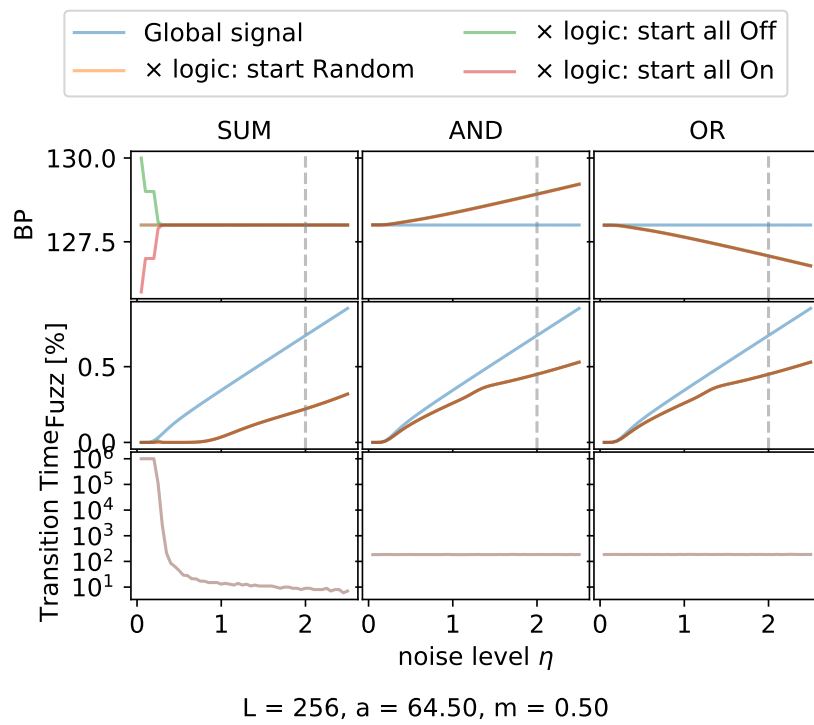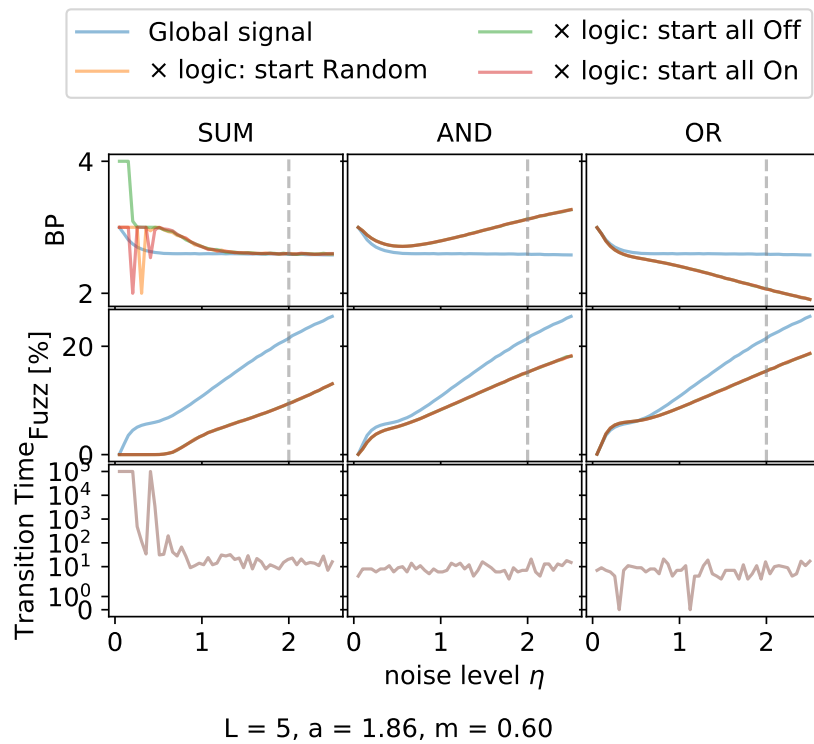
Figure 4.5.: *Upper*: Overview plot for asynchronous update, *Lower*: Same for synchronous update.

Simulation scheme For simplicity and computational efficiency, we chose the dynamics to consist of synchronous updates of the complete grid at equidistant, discrete time steps.

   To make sure that our stationary state results for the boundary position and fuzziness do not depend on the exact updating procedure, we compared the overview plot produced by the synchronous update scheme to that of a random update scheme in Fig. 4.5. One time step in the random update scheme corresponds to L^2 times drawing a cell from the grid at random and updating it according to the pure gradient, SUM, AND or OR rule. From Fig. 4.5, we observe that the results qualitatively agree for all rules. Convergence for small noises is different, which makes sense as an asynchronous update could help to exit the metastable boundary position more quickly. From a computational perspective the synchronous update scheme is by far more efficient than the asynchronous one.

Observables

We are interested in a correctly positioned boundary between different cell types that is straight and sharp. In Fig. 4.6 we can observe in the first row for a small noise level an exemplary grid for each rule as well as a representation for large noise in the second row. We observe that all four rules construct the boundary at a similar position. To make this notion quantitative, we define the boundary position $\mathcal{B}(t)$ of a grid at time t to be the number of cells in Off state,

$$\mathcal{B}(t) := \frac{1}{L} \sum_{i,j} \delta\left(c_{ij}, -\frac{1}{2}\right), \tag{4.8}$$

with δ the Kronecker Delta. This boundary position definition is related to the magnetization in Ising models and circumvents problems of other measures. One complication are 'holes' in the grid. Otherwise we would have to decide to either ignore them or introduce a left (most) and right (most) boundary and combining these in an arguable way. Another common definition that elegantly deals with 'holes' is fitting a tanh and using the x-value of it's zero crossing as the boundary position. Here, we want to work with grids ranging from 2^3 up to 2^{13} cells. For the small grids, fitting a tanh gives poor results.

The other boundary property we want to study is its' fuzziness. In Fig. 4.6, for small noise levels, the SUM rule produces a sharp boundary, the AND, OR and GRAD rule show one to two misplaced cells. For a larger noise level, also the boundary produced by the SUM rule shows defects. The number of defects by the AND and OR rule is similar, although they seem to occur at different sites of the boundary. One might even note that the number of cells in the wrong state for the AND and OR rule is smaller than for the pure gradient. A definition of boundary fuzziness is in order to discuss these heuristics.

The fuzziness $\mathcal{F}(t)$ of a grid at time t is defined as the number of sites in the wrong state with respect to the boundary position, rounded to its closest integer, in percent of the total number of cells,

$$\mathcal{F}(t) := \frac{1}{L^2} \left(\sum_{i<\mathcal{B},j} \delta\left(c_{ij}, \frac{1}{2}\right) + \sum_{i>\mathcal{B},j} \delta\left(c_{ij}, -\frac{1}{2}\right) \right). \tag{4.9}$$

Note that this definition combines two notions characterizing the quality of a boundary, its roughness and its softness. Given a unique boundary line, i.e., a grid configuration without holes, the roughness quantifies the boundaries' deviation from a straight line. In case
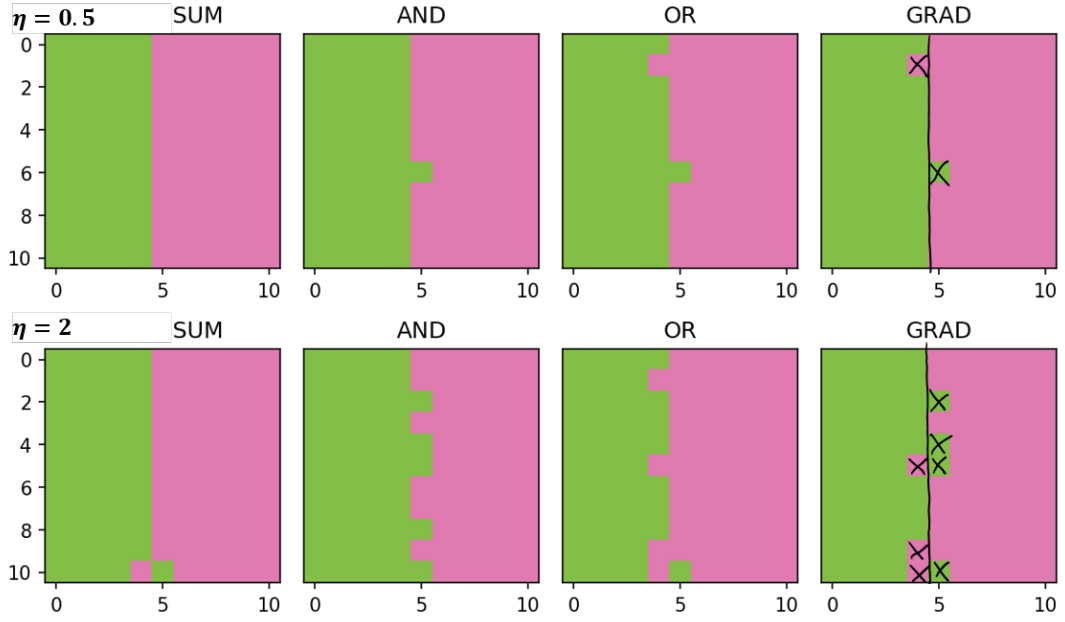
Figure 4.6.: The first row shows an exemplary steady state grid for a small noise level of $\sqrt{2}\sigma = \eta = 0.5$ for each rule, the second row for a large $\sqrt{2}\sigma = \eta = 2$. $L = 11$, $m = 1$, $a = 5.5$.

of frequent holes, the softness measures the width of the holey region that constitutes the boundary. In the system presented here, holes do occur, but are rare, thus we don't account for them separately. The chosen definition of boundary fuzziness counts both types of boundary errors equivalently. In Fig. 4.6 the instantaneous fuzziness $\mathcal{F}(t)$ is visualized in terms of cells in the wrong state as crossed out cells, for the pure gradient rule. For a random grid, the fuzziness reaches it's maximum value of 50%.

The ensemble average of the time it takes for a system to reach its initial condition independent state is taken to be the transition time $\mathcal{T}$. Consequently the transition time is a mean first passage time. As we are only interested in a rough estimate, we will use the maximum of the first passage time from two runs starting at an initial grid of purely Off cells and purely On cells as an approximation. The ensemble averages of the boundary position $\langle \mathcal{B} \rangle$ and the fuzziness $\langle \mathcal{F} \rangle$ are approximated by their time averages in the stationary state.

4.3. Results

For a boundary established by a global signal in the form of a gradient with slope m and local signaling between neighboring cells, we want to measure the dependence of the boundary position $\mathcal{B}$ and its fuzziness $\mathcal{F}$ on the total noise. The total noise with standard deviation σ sums up independent Gaussian noise on the global and the local signal.

4.3.1. Kinetics of approaching the stationary State

We start our investigation of the correction mechanisms SUM, AND and OR by studying the boundary position $\mathcal{B}(t)$ and fuzziness $\mathcal{F}(t)$ as a function of time using a synchronous update scheme of the whole grid. Towards that end, we consider an arbitrary but fixed threshold a, morphogen slope m and grid length L, here chosen such that the boundary position of the pure gradient mechanism is in the middle of the grid.

AND and OR rule We characterize the evolution under the AND rule, plotted in Fig. 4.7a) for three different initial conditions: a random initial grid, a grid of all cells in state Off ('all Off'), and a grid of all cells in state On ('all On'). The first column shows for an exemplary low noise level the boundary position in the first row and the fuzziness in the second row in dependence of time.

Starting from an all Off grid the boundary position $\mathcal{B}$ moves roughly one cell per time step, until it reaches its stationary value, see Fig. 4.7a), first row. The fuzziness $\mathcal{F}$ increases until it drops sharply when the stationary boundary position has been reached. Note that already for low noise the fuzziness time trace remains wiggly for all times implying that the stationary boundary is fuzzy. Remarkably, for larger noise the boundary position moves at the same rate. Only the stationary boundary is more fuzzy compared to the boundary in the low noise regime, see right column of Fig. 4.7a). Consequently, the transition time to the initial condition independent, stationary, boundary position does not depend on the noise level. We also observe that the transition time is on the order of magnitude of the stationary boundary position. The last row of Fig. 4.9 confirms this independence more generally.

An intuitive picture is the following: By definition, the AND rule only allows cells to switch on if they have at least one On neighbor and the gradient signal exceeds its threshold a. The second condition is satisfied for all cells to the right of position $i = \lfloor \frac{a}{m} \rfloor$ (up to noise on the global signal of mean zero). When starting from an all Off grid, the first condition implies that only cells at the right boundary can switch on due to cells at $i = L + 1$ being On (fixed grid-boundary condition choice). Each boundary cell can move at most one cell forward per time step, see Fig. 4.7i) and ii). Quantitatively, a cell with exactly one On neighbor switches on with probability $1/2$ independent of the total noise level, as

$$
\phi^{\text{AND}}\left(s^L, s^G\right) + \frac{1}{2} = \Theta\left(s_{ij}^L(t) + 1\right) \Theta\left(s_{ij}^G(t) - a\right)
$$

$$
\approx \Theta\left(\sum_{j \in \mathcal{V}(i)} c_j(t) + \xi_{ij}^L(t) + 1\right) \cdot 1
$$

$$
= \Theta\left(\xi_{ij}^L(t)\right) .
$$

In words, the condition of the global signal exceeding its threshold is nearly always fulfilled and with one On neighbor the deterministic local neighbor signal equals it threshold, thus the sign of the local noise with mean zero determines the cells state at the next time step. Further, a cell with more than one On neighbor switches on with a probability close to one. Note that our choice of grid-boundary conditions, $c_{0,j} = 0$ and $c_{L+1,j} = 1$ for all rows j, does
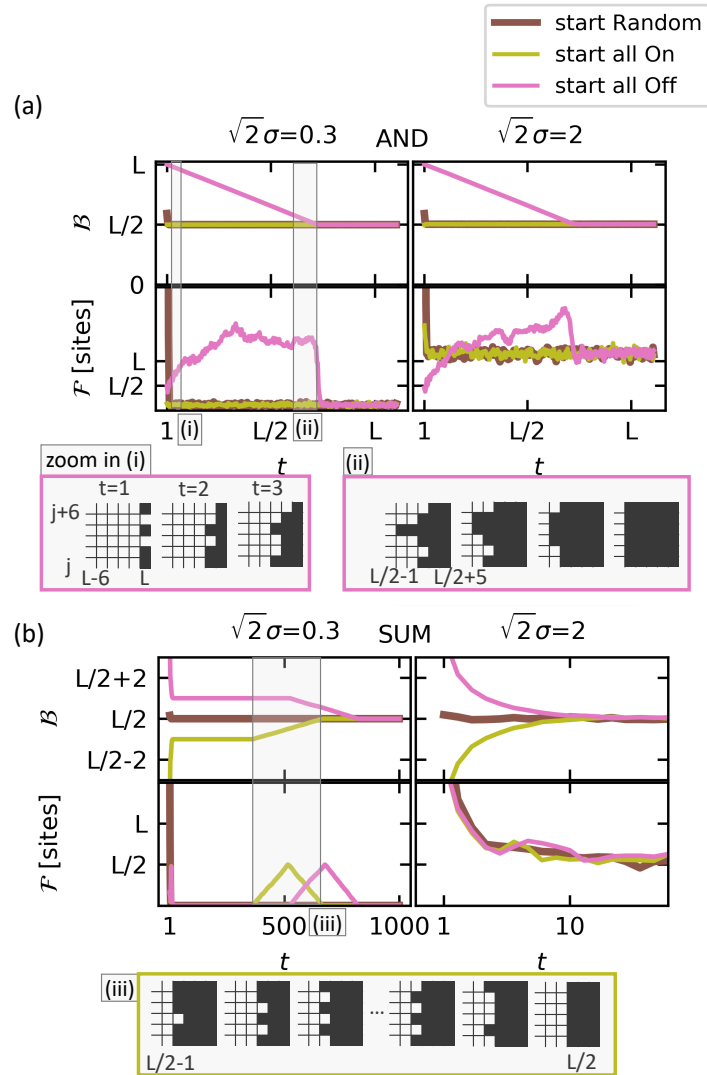
Figure 4.7.: Exemplary boundary position $\mathcal{B}$ and fuzziness $\mathcal{F}$ (in terms of number of wrong cells) time traces starting from different initial conditions at $t = 0$, depicted in different colors, for the AND rule in the upper panel (a) and the SUM rule in the lower panel (b) subject to a small (first column) and a large (second column) noise value. Note that for the AND rule time is measured in terms of grid length. The grid length $L = 256$ is even and $a = 64.5$ and $m = 0.5$ are chosen such that the stable boundary is in the grid center, only $t \geq 1$ are shown for better visibility. Inserts (i)-(iii) provide a zoom in of 6×6 sites to the grid, for few time steps. For the AND rule, starting from an all Off initial grid, zoom in (i) sketches the first three time steps, whilst zoom in (ii) provides a potential time trace of reaching the stable boundary position, which is $L/2$ here. For the SUM rule, zoom in (iii) sketches boundary destabilization and transition to a straight and sharp boundary.

not substantially simplify the patterning task for the rules. The AND rule (and also the OR and SUM rule as we will see later) cannot just shift the sharp cell state boundary at L to its stationary position. Even if we had only one cell in On state at $i = L + 1$ the rules could still establish a boundary at the center of the grid. However, this choice of grid-boundary would artificially destabilize the correct stationary pattern due to local signaling, as the correct pattern requires that cells at $i = L$ are On.

The OR rule by definition can only switch off one cell width at a time when starting from an initial grid of On cells. Consequently, the time traces of the OR rule qualitatively correspond to the ones of the AND rule with On-Off inverted initial conditions, see Fig. 4.4. This was to be expected from the AND-OR relationship, Eq. 4.6.

SUM rule   Fig. 4.7b) shows evolution under the SUM rule and exhibits qualitatively different dynamics. The right column displays the large noise regime. We see that already within 20 time steps the different boundary position traces have converged. We also note that the boundary is fuzzy in contrast to the low noise regime. For a low noise level, as depicted in the left column of Fig. 4.7b), the dynamics is more complex. Whilst for all initial conditions, the boundary quickly reaches a position close to its stationary value, full convergence is very slow. Thus, in contrast to the AND and OR rule the transition time strongly depends on the noise level. Starting from any initial condition, the boundary reaches a position close to its stationary value within few time steps. Movement towards its final position happens when noise induces a seed at the boundary linearly spreading until all cells within the same column have switched state. As sketched in Fig. 4.7iii), one seed induces a switch of both of its neighbors in the next time step and so forth until the complete column of former boundary cells has switched state, in case of an odd grid height. Then the boundary remains straight until the next seed occurs. This intuition for sufficiently small noise of the 'boundary jump process' we further substantiated by resolving the process in time for an ensemble of runs Fig. 4.8. In case of an even grid length, the pattern with every second boundary cell switched on corresponds to a metastable state. The waiting time distribution for the next seed to destabilize the metastable boundary is strongly noise level dependent. In fact, the transition time until the stationary state at $\lfloor \frac{a}{m} \rfloor$ is reached can be approximated by

$$\mathcal{T}(\sigma) \approx \frac{12}{L} \exp\left(\frac{1}{2\sigma^2}\right)$$

for sufficiently small σ, as we show in the following. This functional form clearly shows the non-linear dependence of the transition time on the noise level. As expected, the transition time is inversely proportional to the system length as a seed is more likely the more sites in the column next to the boundary noise is acting on.

Quantitative take on transition time of SUM rule   To get a better understanding of SUM rule dynamics, we quantitatively work out the transition time for (very) small noise. The scaling with noise level $\sqrt{3}\sigma$ of the resulting transition time approximation will moreover explain, why we cannot just simulate long enough get the SUM rules stationary state behavior in this noise regime.

The probability density of the first passage time (FPT) from boundary position i to $i + 1$

depends on the probability of a seed for the jump, $P(\text{seed})$, as follows

$$P(\text{FPT of } i \to i-1 = t) = (1 - P(\text{seed}))^{t-1} P(\text{seed})$$

$$= \left(\frac{1}{1 - P(\text{seed})}\right)^{-t} \frac{P(\text{seed})}{1 - P(\text{seed})}$$

$$\approx \lambda \exp(-\lambda t)$$

with $\lambda = P(\text{seed})$ and using $P(\text{seed}) \ll 1$. The approximation to an exponential distribution is also confirmed numerically, see 4.8 and Fig. 4.8.

Consequently, the mean first passage time (MFPT) of the boundary jump from i to $i-1$ is given by λ^{-1}. $P(\text{seed})$ is the probability of a single cell at the boundary to switch to the wrong state, which we call a 'defect'. For an On defect at the left side of the boundary it is given by grid height L times the probability that any of the boundary cells switches state. The probability of a single cell to switch equals the cumulative Gaussian total noise distribution with standard deviation σ for noise realizations exceeding the threshold a reduced by the local signal s^L and global signal $s^G = im$ contribution. A cell left of the boundary has three Off neighbors and one On neighbor, thus $s^L = -1$. With that

$$P(\text{seed}) = L \int\limits_{a+1-im}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{1}{2}\left(\frac{n}{\sigma}\right)^2\right) dn$$

$$= \frac{L}{2} \text{erfc}\left(\frac{1}{\sqrt{2}} \frac{a+1-im}{\sigma}\right). \tag{4.10}$$

We can read off that the mean first passage time rapidly increases with decreasing i thus to get an estimation of the transition time towards the stationary boundary position, we can neglect all previous boundary jumps

$$\mathcal{T}(\sigma) \approx \text{MFPT}(i_c + 1 \to i_c)$$

$$\approx \frac{2}{L} \text{erfc}^{-1}\left(\frac{1}{\sqrt{2}} \frac{1}{\sigma}\right)$$

$$\approx \frac{12}{L} \exp\left(\frac{3}{2} \frac{1}{\sqrt{3}\sigma^2}\right)$$

where in the last step we have used an approximation of the complementary error function valid for small σ. We immediately see that simulating the stationary state of very small σ is not feasible.
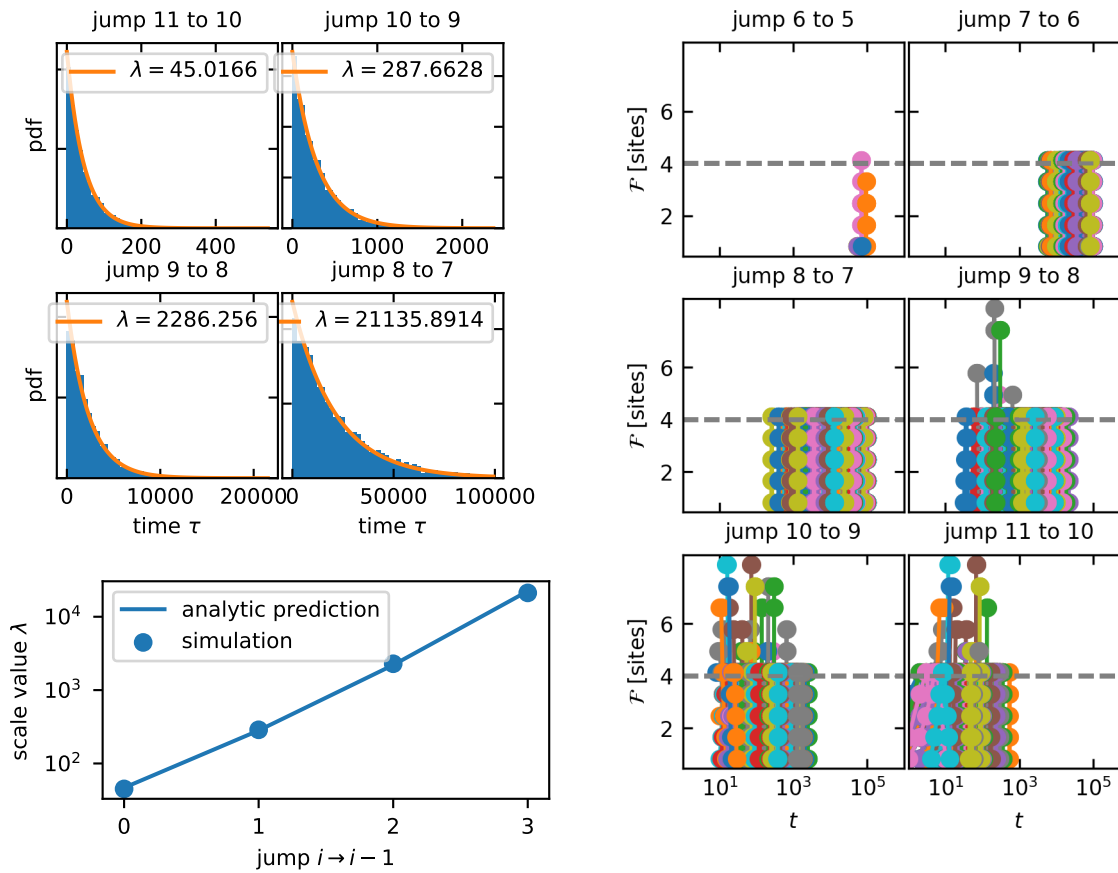
Figure 4.8.: Shown are boundary jumps occurring within 10^5 simulation time steps in an ensemble of 5000 SUM rule runs. Results are for an $L = 11$ square grid and a noise level of $\sqrt{3}\sigma = 0.3$. *Upper left:* Plotted is the waiting time distribution τ until the next boundary jump for the ensemble of 5000 runs. In orange the fit result for the distribution is shown, $P(\tau, i) \propto \exp(-\lambda(i)\tau)$, with i denoting the boundary position. *Lower left:* Comparison of the simulation results for the scale $\lambda(i)$ of each boundary jump to the analytic prediction, $\lambda^{-1} = \frac{L}{2}\text{erfc}\left(\frac{1}{\sqrt{2}}\frac{a+1-im}{\sigma}\right)$. Jumps are numbered in order of their occurrence starting from 0, i.e. jump 0 corresponds to $i = 11 \rightarrow i = 10$, jump 1 to $i = 10 \rightarrow i = 9$, ect. *Right:* The fuzziness time trace in number of cells is plotted in a separate panel for each boundary jump. At $t = 0$ we start from an all Off grid in the last panel and a boundary position at $L = 11$ and observe that the fuzziness on average within the first 50 time steps reaches its maximum at 4, before it drops again to zero, implying that a new sharp boundary at $i = 10$ has been established. Different colors indicate different runs. The full run time is limited to 10^5 time steps, consequently in the jump 7 to 6 and 6 to 5 panel there are significantly fewer runs plotted, as for most simulations those jumps haven't happened within the time simulated.

4.3.2. Characterizing the stationary State's dependence on the Noise Level
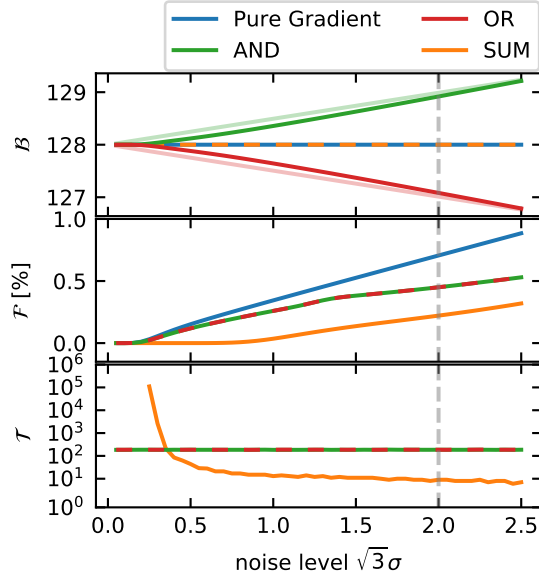

Figure 4.9.: Overview of different stationary state behavior of the pure gradient, SUM, AND and OR rule depending on noise level regime, exemplary for $L = 256$, $a = 64.5$, $m = 0.5$. On the x-axis, we use *noise level* $\sqrt{3}\sigma$ instead of σ for better intuition of the strength of the total noise: more than 90% of all noise realizations are within the interval $\left[-\sqrt{3}\sigma, +\sqrt{3}\sigma\right]$. Also, a uniform distribution within $\left[-\sqrt{3}\sigma, +\sqrt{3}\sigma\right]$ has variance σ^2. In the first row, the time averaged boundary position $\mathcal{B}$ for each rule is shown, where dashes indicate that the lines overlap. The light green and red lines show the analytical estimate of the boundary position for the AND and OR rule. In the second row, time averaged fuzziness $\mathcal{F}$ in % of the total number of grid sites is plotted. The last row shows the transition time until the initial condition independent state was reached, where a value of 10^6 implies that it has not converged within the simulation time.

Fig. 4.9 shows the characteristic stationary state properties of the three correction mechanisms SUM, AND and OR as a function of the total noises' standard deviation σ, for an exemplary threshold a and morphogen signal slope m choice. For direct comparison, the results without correction mechanism (pure gradient) are plotted in blue. Each row shows a different observable — the time averaged boundary position $\mathcal{B}$, fuzziness $\mathcal{F}$ and, in the last column, the transition time $\mathcal{T}$. For simplicity, we here use the maximal number of time steps until reaching the stationary state from an all On and an all Off grid as a measure of $\mathcal{T}$. We will discuss the different phenomenologies starting from low noise levels and ending with large noise levels.

Small noise levels In the last row of Fig. 4.9 we observe that the SUM rule results have not converged to stationary state within the simulation time of 10^6 time steps in the regime of very low noise levels. This is expected from the previous transition time discussion. Loosely speaking, noise is needed to forget the initial grid state. In contrast, the AND and OR rule's transition time scales linearly with the boundary position, irrespective of the noise level as discussed in Subsec. 4.3.1. Turning to the second row of Fig. 4.9, we note that the boundary fuzziness for the SUM rule is remarkably close to zero, for small yet sufficiently large noise levels to allow for convergence of the SUM rule pattern. An effectively non-fuzzy regime is not observed for other rules. In the first row, we observe that the SUM's boundary position agrees well with the pure gradient rules' for zero noise, whilst the AND and OR rules' boundary position slightly deviate up to a cell width. In the following we derive analytic approximations for the stationary boundary positions of the different logics.

Condition for stationary boundary position We have two conditions to be satisfied such that the stationary boundary position is at cell index $i = i_c$: The probability to destabilize a boundary at $i_c - 1$ by an Off-in-On defect has to be smaller or equal to the probability destabilizing a boundary at i_c by an On-in-Off defect. For a sketch, see the first grid of Fig. 4.7, inset (iii) and (iv), respectively. For the right-hand side of i_c we can formulate the conditions as:

$$(i) \qquad \qquad \mathrm{P(OffInOn}\,|b = i_c - 1) > \mathrm{P(OnInOff}\,|b = i_c)$$

$$(ii) \qquad \qquad \mathrm{P(OffInOn}\,|b = i_c) < \mathrm{P(OnInOff}\,|b = i_c + 1)$$

with, e.g., $\mathrm{P(OffInOn}\,|b = i_c - 1)$ denoting the probability of an Off-in-On defect, if the sharp boundary position is at $b = i_c + 1$.

Stationary boundary position for the SUM rule For the SUM rule the individual probabilities are given by

$$\mathrm{P(OnInOff}\,|b = i) = \mathrm{P}(s^G(i) + s^L(i) \geq a)$$

$$= \mathrm{P}\left(mi + \sum_{(k,l) \in \mathcal{V}(i,j)} c_{k,l} + \xi \geq a \right)$$

$$= \mathrm{P}\left(\xi \geq a + 1 - mi \right),$$

with $\mathcal{V}(i,j)$ the Von Neumann neighborhood of the cell at (i,j), i.e., it's upper, lower, right and left neighbor and $\xi = \xi^L + \xi^G$ the total noise. We at first used that with a straight boundary at i a cell at (i,j) (j arbitrary) has one On and three Off neighbors.
Further,

$$\mathrm{P(OffInOn}\,|b = i) = \mathrm{P}(s^G(i+1) + s^L(i+1) < a)$$

$$= \mathrm{P}\left(m(i+1) + \sum_{(k,l) \in \mathcal{V}(i+1,j)} c_{k,l} + \xi < a \right)$$

$$= \mathrm{P}\left(\xi < a - m(i+1) - 1 \right).$$

Inserting both into conditions (i) and (ii) and that ξ has the same distribution as $-\xi$ gives

$$
\begin{aligned}
(i) \quad & \mathrm{P}(\xi < a - mi_c - 1) > \mathrm{P}(\xi \leq -(a - mi_c) - 1) \\
& \Rightarrow \quad a - mi_c \geq 0 , \\
(ii) \quad & \mathrm{P}(\xi < a - m(i_c + 1) - 1) < \\
& \mathrm{P}(\xi \leq -(a - m(i_c + 1)) - 1) \\
& \Rightarrow \quad a - m(i_c + 1) < 0 .
\end{aligned}
$$

As $i_c \in \mathbb{N}$, these two inequalities are satisfied by

$$
i_c^{\mathrm{SUM}} = \left\lfloor \frac{a}{m} \right\rfloor = i_c^{\mathrm{Grad}} . \tag{4.11}
$$

The stationary boundary will scale with system size in the same way as for the pure boundary formation by gradient mechanism for zero noise.
Note that in the stationary state, implying that $i = i_c$, the probabilities for an On-In-Off defect and an Off-In-On defect only depend on the morphogen slope m and the deviation of $\frac{a}{m}$ from its subsequent integer:

$$
\begin{aligned}
\mathrm{P}(\mathrm{OffInOn} \,|\, b = i_c) &= \mathrm{P}\left(\xi < a - m(i_c + 1) - 1\right) \\
&= \mathrm{P}\left(\xi < m\left(\frac{a}{m} - \left\lfloor \frac{a}{m} \right\rfloor + 1\right) - 1\right)
\end{aligned}
$$

and equivalently for $\mathrm{P}(\mathrm{OnInOff} \,|\, b = i_c)$.

Stationary boundary position for the AND and OR rule Let us consider the AND rule. The individual probabilities are given by

$$
\begin{aligned}
&\mathrm{P}(\mathrm{OnInOff} \,|\, b = i) \\
&= \mathrm{P}\left(s_{ij}^G(t) + \xi_{ij}^G(t) > a\right) \mathrm{P}\left(\sum_{(k,l) \in \mathcal{V}(i,j)} c_{k,l} + \xi_{ij}^L(t) \geq -1\right) \\
&= \mathrm{P}(s_{ij}^G(t) + \xi_{ij}^G(t) > a)\mathrm{P}\left(\xi_{ij}^L(t) \geq 0\right) \\
&= \mathrm{P}(s_{ij}^G(t) + \xi_{ij}^G(t) > a)\frac{1}{2} ,
\end{aligned}
$$

where we at first used that with a straight boundary at i a cell at i has one On and three Off neighbors. Then we observed that the for non-zero noise the probability of the local noise to exceed its mean 0 is 1/2, independent of its precise distribution (as long as it is symmetric).

Further,

$$
\begin{aligned}
&P(\text{OffInOn} \,|b = i) \\
&= 1 - P(s^G_{i+1,j} + \xi^G_{ij}(t) > a) \\
&\quad \cdot P\left(\sum_{(k,l) \in \mathcal{V}(i+1,j)} c_{k,l} + \xi^L_{ij}(t) \geq -1 \right) \\
&= 1 - P(s^G_{i+1,j} + \xi^G_{ij}(t) > a) P\left(\xi^L_{ij}(t) \geq -2 \right) \\
&\approx 1 - P(s^G_{i+1,j}) + \xi^G_{ij}(t) > a)
\end{aligned}
$$

Here we used that for an Off-in-On defect given the boundary is at i, we need to consider a cell at $i+1$, which consequently has one Off and three On neighbors.
For the noise level regime, $\sqrt{3}\sigma \in [0, 2.5]$ that we consider in this paper, $P\left(\xi^L \geq -2\right) \approx 1$ is a good approximation.
With that follows from condition (i)

$$
P\left(s^G_{i_c^{\text{AND}},j} + \xi^G_{ij}(t) > a \right) < \frac{2}{3},
$$

which is easily satisfied for $i_c^{\text{Grad}} = \left\lfloor \frac{a}{m} \right\rfloor$ as $s^G(i_c^{\text{Grad}}) \approx 0$.
From condition (ii) follows

$$
P\left(s^G_{i_c^{\text{AND}}+1,j} + \xi^G_{ij}(t) > a \right) \geq \frac{2}{3},
$$

which consequently determines i_c^{AND}. For the Gaussian noise distribution with mean zero and standard deviation $\sigma^G = \sigma\sqrt{1 - \frac{2}{2+mL}}$, it follows

$$
\begin{aligned}
i_c^{\text{AND}} &= \left\lfloor \frac{a}{m} \right\rfloor - \frac{\sqrt{3}\sigma}{m}\sqrt{1 - \frac{2}{2+mL}} \left(\sqrt{\frac{2}{3}}\text{erfc}^{-1}\left(\frac{4}{3}\right) \right) \\
&\approx \left\lfloor \frac{a}{m} \right\rfloor + 0.25 \frac{\sqrt{3}\sigma}{m}\sqrt{1 - \frac{2}{2+mL}} \,. \tag{4.12}
\end{aligned}
$$

We see that $i_c^{\text{AND}} \approx i_c^{\text{Grad}} + \frac{\sqrt{3}\sigma}{4m}$ for $mL \gg 2$ which agrees nicely with simulation results shown in Fig. 4.9.
For the OR rule, it follows

$$
i_c^{\text{OR}} \approx \left\lfloor \frac{a}{m} \right\rfloor - 0.25 \frac{\sqrt{3}\sigma}{m}\sqrt{1 - \frac{2}{2+mL}} , \tag{4.13}
$$

respectively by the AND-OR equivalence established in Eq. 4.6.

From this calculation we can also see that the probabilities for an On-in-Off defect and an Off-in-On defect at i_c depend only on m and $\frac{a}{m} - \left\lfloor \frac{a}{m} \right\rfloor$. These findings suggest that the fuzziness in stationary state only depends on the deviation of $\frac{a}{m}$ to the next integer value. This is also confirmed by numeric results. Intuitively, the morphogen changes at the same rate everywhere in the system and the local interaction is independent of the position per definition.

Intermediate noise levels For larger noise levels the AND and the OR rule qualitatively exhibit the same behavior as for low noise, in contrast to the SUM rule. In the second row of Fig. 4.9 we observe a rapid increase in fuzziness for the SUM rule. Time averaged fuzziness seems to arise from alternating between time intervals of a straight, sharp boundary and time intervals with a disturbance, seeded by a single defect cell, that grows and shrinks for some time before it decays. The probability for a seed is highly non-linearly, but smoothly, increasing with noise level as discussed in the previous section.

Large noise levels In the regime of large noise we see in the second row in Fig. 4.9 that the SUM rule yields a less fuzzy boundary than the AND and OR rule, which behave similarly. Indeed, all correction mechanisms outperform the pure gradient rule. We will show that this result is robust for all *a*, *m* parameter combinations determining the three rules in subsection 4.3.4, for all grid lengths (Subsec. 4.3.3) and for a surprisingly large range of local to total noise ratios α (Subsec. 4.3.5).

The boundary position $\mathcal{B}$ resulting from the SUM rule agrees well with the boundary position from the pure gradient as analytically deduced. Although the boundary positions from the AND and OR rules deviate linearly with the standard deviation of the total noise σ, we will show in Subsec. 4.3.3 that both nevertheless scale linearly with system size.

4.3.3. All rules conserve scaling of the Boundary Position with System Size

In embryogenesis, proportions commonly remain the same irrespective of different embryo or compartment sizes (see Inomata 2017 for a review). The pure gradient mechanism also exhibits this scaling behavior, provided that the maximal morphogen concentration and the threshold remain constant. In our model these conserved proportions translate to a fixed fraction of On to Off cells within grids of different size for a fixed maximal morphogen signal mL. In Fig. 4.9 we have observed that the stable boundary position formed by the AND and OR rule deviates from the position established by the pure gradient rule. Thus we need to investigate if also the AND and OR rule ensure this property, just with a different fraction. For an exemplary parameter set, $a = 2$, $mL = 8$ at a large noise level $\sqrt{3}\sigma = 2$, we can see in Fig. 4.10 that this is indeed the case. The reason that AND (OR) rules' boundary position tends to larger (smaller) $\mathcal{B}$ values is that it discourages (encourages) On cells. As the slope m becomes smaller and smaller (mL fixed) the regime around the boundary in which this effect plays a role increases linearly with system length. This leads to a constant boundary position $\mathcal{B}$ over grid length L ratio. The deviation of this value from the pure gradient rules' result can be approximated by Eq. 4.12 and Eq. 4.13, respectively.

For small grid sizes, we see that the relative boundary position of any rule has not yet converged to its large grid limit. The reason why the boundary position and fuzziness for small lengths deviate from the values for large grids is of technical nature, as our parameter choice yields a boundary at $\lfloor \frac{a}{m} \rfloor = \frac{a}{m} = \frac{1}{4}L$. For the pure gradient rule, for instance, this choice implies

$$c_{i_c,j} = \Theta(mi_c + \xi_{ij}^G(t) \geq a) - \frac{1}{2} = \Theta(\xi_{ij}^G(t) \geq 0) - \frac{1}{2},$$

thus on average the cell at i_c is switched On every second time step due to noise. In contrast,

for $c_{(i_c+1,j)}$ these parameters more stably yield On as

$$c_{i_c+1,j} = \Theta(mi_{c+1} + \xi^G_{i_c+1,j} \geq a) - \frac{1}{2} = \Theta(\tilde{\xi}^G \geq -m) - \frac{1}{2}.$$

Thus defect cells at the left of the boundary are more common than at its right.

Fig. 4.10 shows that the fraction of sites in the wrong state converges to a stable value for large system lengths. Consequently, the improved boundary sharpness is not only a finite grid size effect. The observation that the SUM rule performs best, the AND and OR rule not as well, but better than pure gradient also holds true for all tested grid lengths.
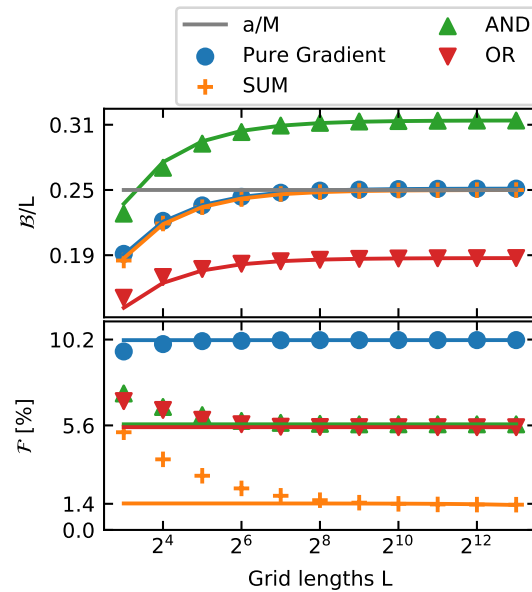
Figure 4.10.: (a) Relative time averaged boundary position $\mathcal{B}$ for all three rules and pure gradient for increasing grid lengths L for a fixed noise level of $\sqrt{3}\sigma = 2$, $a = 2$, $mL = 8$, and all Off initial grid. (b) Time averaged boundary fuzziness $\mathcal{F}$ in percent of the number of cells in the grid. For sufficiently large grids the relative fuzziness converges to a constant value that is smallest for the SUM rule, larger for the AND and OR rule and highest for the pure gradient rule.

4.3.4. Systematic Exploration of the Parameter Space

Figure 4.11.: For a fixed noise level $\sqrt{3}\sigma = 2$ and grid length $L = 256$ the fuzziness $\mathcal{F}$ fraction of the boundary formed by a correction rule (SUM, AND, OR) and the boundary fuzziness due to pure gradient is shown in dependence of the gradient slope m. Single markers correspond to the average fuzziness fraction over all a values and are interpolated by a solid line. The sha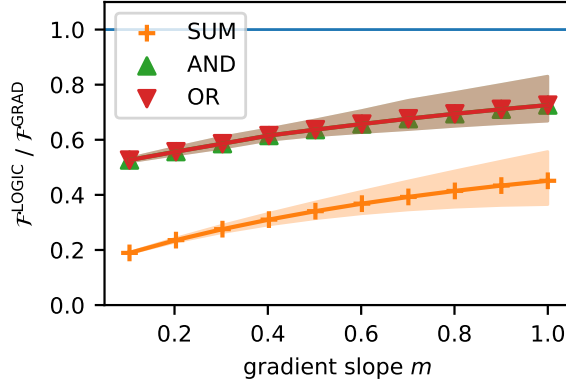ded area is restricted by the interpolation of the minimal (lower edge) and maximal (upper edge) fuzziness fraction with respect to all threshold a values.

We want to know, how much smoother the boundary established by the correction mechanisms is compared to the boundary established by the pure gradient rule. Until now, we have shown results for isolated points in the parameter space. Now, we want to study the whole parameter space spanned by threshold a and morphogen signal slope m. It is equivalent to varying m and $\frac{a}{m}$ from $\lfloor \frac{a}{m} \rfloor$ to $\lceil \frac{a}{m} \rceil$ as discussed in Subsec. 4.3.2.

Let us fix a high noise level, $\sqrt{3}\sigma = 2$. We measure the boundary smoothing capability of a correction rule in terms of the ratio of the boundary fuzziness resulting from a correction mechanism to the boundary fuzziness caused by the pure gradient rule, $\mathcal{F}^{\text{LOGIC}}(\sqrt{3}\sigma = 2)/\mathcal{F}^{\text{GRAD}}(\sqrt{3}\sigma = 2)$. In order to show the full range of fuzziness ratios caused by varying a and m in Fig. 4.11, we choose for each m to display the ratios' minimal and maximal value for any a (interpolated by the lower and upper edge of the shaded area), as well as the ratios' average over a (data points interpolated by sold line).

We observe that $\mathcal{F}$ ratios are below one and that the SUM ratio is smallest for all morphogen slopes m. This implies that all three rules perform better for every threshold a and gradient slope m than the pure gradient rule with respect to fuzziness reduction. The performance gap between the pure gradient and the correction rules is smaller for larger morphogen slopes, thus the most conservative choice for m is $m = 1.0$. Intuitively, this is because the steepest gradient provides the largest signal differences between neighboring cells in i direction.
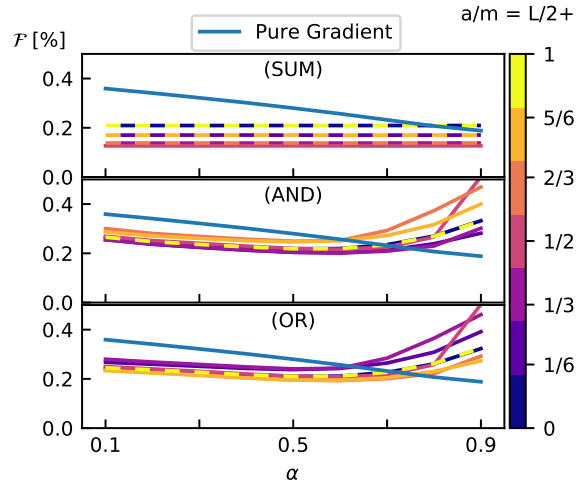
4.3.5. Variation of local to total Noise Ratio

Figure 4.12.: Fuzziness $\mathcal{F}$ at $\sqrt{3}\sigma = 2$ in dependence of the local to total noise variance, α for $m = 1.0$, one row for each correction mechanism. The blue line depicts the pure gradient performance, while the differently colored lines correspond to the correction mechanism result for different a/m combinations, $L = 256$.

The magnitude and ratio of local to global noise representing a variety of different processes as mentioned in the introduction is not known. Still, for a fixed maximum morphogen signal mL, we have only considered one particular ratio of the local noise ξ^L to total noise $\xi^L + \xi^G$ ratio α, $\alpha = \frac{(\sigma^L)^2}{(\sigma^L)^2 + (\sigma^G)^2} = \frac{2}{2+mL}$ so far. Intuitively, the pure gradient rule should perform better than the correction mechanisms for large α. By definition, the pure gradient rule only experiences noise on the global signal, which approaches zero for the local to total noise ratio α approaching one. The correction mechanisms, however, process an additional highly error-prone signal, the local signal. To test this intuition, we vary the local to total noise ratio α between 0.1 and 0.9, see Fig. 4.12 for all different $\frac{a}{m}$ combinations with $m = 1.0$, depicted in different colors. The maximal slope $m = 1.0$ is chosen to ensure the best relative pure gradient performance, as discussed in Sec. 4.3.4.

Per definition, the SUM rule (first column) is independent of α. The AND and OR rule perform worse relative to the SUM rule the larger α is. Surprisingly, the α value, where pure gradient (blue line) performs equally well than the correction mechanisms (other colors) is well beyond one half. This implies that a correction mechanism can sharpen the boundary, even if it is subject to more than twice as much noise as the pure gradient rule.

4.4. Discussion

Precise boundary formation is a remarkable phenomenon in developing systems and aspiring goal in synthetic systems. Here, our objective was not to study any specific system, but to systematically explore how different logical couplings of cell-cell communication with a gradient

signal can aid boundary formation. A comparison to the Ising model and Hillenbrandt et al.'s work (Hillenbrand, Gerland, and Tkačik 2016) can be found in the appendix B.1.

In a minimal model of boundary formation consisting of cells that are either signaling (On) or inactive (Off), we studied three different correction mechanisms using this nearest neighbor interaction in addition to a global (grid spanning) signal gradient. Those three rules either sum up both signals (SUM) and subsequently compare the result to a global threshold *a* or compare each signal to a local and global threshold separately (AND, OR). Consequently, for a signal processing cell to switch to or maintain an On state, in case of the SUM rule the sum of both signals has to exceed the global threshold. The AND rule requires both signals to exceed their respective thresholds independently, while the OR rule requires only one signal to exceed its threshold. We examined which rule performs best in which regime of total noise, consisting of additive Gaussian noise on the local and the global signal. As motivated in the introduction, performance is measured in terms of: (i) Reduction in boundary fuzziness, (ii) short transition time to stationary boundary position, (iii) position tuneable by threshold and (iv) scaling with system size.

We found that (i) the SUM rule achieves the strongest fuzziness reduction, while the AND and OR rule yield comparably less reduction. Only if the noise on the global signal is much smaller than the noise on the local signal, the pure gradient ensures a sharper separation of cells with different gene expression states. (iii) However, transition to stationary state takes more time for any correction mechanism than for the pure gradient rule, which establishes the stationary boundary within one time step. The SUM rule generates a boundary position that deviates by few sites within one time step for sufficiently steep morphogen slopes. Exact convergence strongly depends on the noise level — the smaller, the slower. Qualitatively different, convergence of the AND and OR rule does not depend on the noise level, but on the initial state of the grid. In the best case it happens within one time step, in the worst case in order of grid length *L* time steps. A short transition time is desirable in development, as boundary cells often act at organizing cells for the next patterning process Dahmann, Oates, and Brand 2011. Also, fast morphogenesis is favorable to protect against predators. (ii) The boundary position can be tuned by changing the global threshold value *a* in a similar manner than for the pure gradient mechanism. This is programability is of biological significance, as the threshold *a* was motivated by the binding affinity of the signals' transcription factor to the promoter of the gene, that is switched on. Possibly, variations of same theme among related species, such as a stripe that differs in width, can be explained as variations of the binding affinity. (iv) For fixed threshold value *a* and morphogen signal slope *m* the SUM rule yields the same boundary position as the pure gradient, while the AND and OR rules' boundary positions deviate by a small amount. Nevertheless, the boundary position established by any rule scales linearly with system size. The scaling property ensures compatibility with the observation that embryos of the same species differ (slightly) in size, but pattern ratios are often conserved Inomata 2017.

What is the underlying reason that the SUM rule outperforms the AND and OR rule in terms of fuzziness reduction? We believe this is because the SUM rule first averages the involved noises allowing them to cancel, before applying the non-linearity in the form of comparing to a threshold. The deterministic signal is summed up, while the Gaussian noise's standard deviations only add in quadrature. For the AND and OR rule it is the other way round:

they threshold each signal separately before combining the pieces of information. However, there are biological problems where this signal processing scheme — first thresholding, then combining — is actually optimal, for example in case of a rare signal. Rod cells in retina specialized to detecting very dim light first threshold before transmitting information to their common bipolar cell, that consecutively sums up those digitized signals Field and Rieke 2002. If the bipolar cell would first sum the signal of its numerous rod cells, the simultaneously summed up noise would make it likely for the bipolar cells to confuse total darkness (no photon) with dim light (three or more photons) Bialek 2012. Consistent with the system presented here, it has been found that only those rod cells specialized for dim light detection process signal by thresholding before summing up.

In a synthetic setup it is feasible to experimentally test our predictions as all necessary components have already been designed. On one hand, tuning of local interactions has been successfully realized in multicellular systems, for example, AND-like, OR-like and in between, graded, SUM-like, regulatory behavior in yeast Bashor, Patel, Choubey, et al. 2019. A remarkably customizable signal processing scheme in the form of a synthetic Notch pathway is presented in Ref. Toda, Frankel, and Lim 2019, and also reviewed in Ref. Chen and Elowitz 2021 among other protein based synthetic circuits in eukaryotic cells. On the other hand, tuneable processing of synthetic gradients has been demonstrated, such as toggle switch processing of a signal gradient in E.coli Barbier, Perez-Carrasco, and Schaerli 2020. A setup close to developmental biology was constructed within Drosophila wing primordia Stapornwongkul, De Gennes, Cocconi, et al. 2020. A combination of a synthetic gradient and a signal processing pathway that can in principle be customized to the rules discussed in this paper, is the synthetic GFP morphogen that regulates target gene expression by a synNotch circuit Toda, Mckeithan, Hakkinen, and Lopez 2020.

Remarkably, the study presented in Ref. Stapornwongkul, De Gennes, Cocconi, et al. 2020 raises the question of which additional mechanisms are required for sharp boundary formation as the explored setups did not give rise to sharp boundaries Barkai and Shilo 2020. The toolboxes listed above might be able to test whether the three different rules studied in this paper are candidates for the actual boundary correction necessary to ensure the exact results observed in nature. Also, insights about the potential of local signaling as a correction mechanism might find applications in synthetic biology more generally, as boundary and stripe formation are fundamental tasks in patterning and morphogenesis. To better meet such applications, the presented model can be trivially extended to stripe generation by introducing extra thresholds accordingly. It can be further generalized or modified, by including more complex rules or smoothing the hard threshold via Hill-type functions, by considering irregular and dynamic cell grids, or by including additional mechanisms such as cell sorting.

Outlook: Possible implications for Synthetic Systems modeling biological boundary formation and further research

In a synthetic setup it is feasible to experimentally test our predictions as all necessary components have already been designed. On one hand, tuning of local interactions has been successfully realized in multicellular systems, for example, AND-like, OR-like and in between, graded, SUM-like, regulatory behavior in yeast (Bashor, Patel, Choubey, et al. 2019). A

remarkably customizable signal processing scheme in the form of a synthetic Notch pathway is presented in (Toda, Frankel, and Lim 2019), and also reviewed in (Chen and Elowitz 2021) among other protein based synthetic circuits in eukaryotic cells. On the other hand, tuneable processing of synthetic gradients has been demonstrated, such as toggle switch processing of a signal gradient in E.coli (Barbier, Perez-Carrasco, and Schaerli 2020). A setup close to developmental biology was constructed by (Stapornwongkul, De Gennes, Cocconi, et al. 2020) within Drosophila wing primordia. A combination of a synthetic gradient and a signal processing pathway that can in principle be customized to the logics discussed in this paper, is the synthetic GFP morphogen that regulates target gene expression by a synNotch circuit (Toda, Mckeithan, Hakkinen, and Lopez 2020).

Remarkably, (Stapornwongkul, De Gennes, Cocconi, et al. 2020) does pose the question which additional mechanisms are required for sharp boundary formation as the explored setups did not give rise to sharp boundaries (Barkai and Shilo 2020). These toolboxes might be able to test whether the three different logics studied in this paper are candidates for the actual boundary correction necessary to ensure the exact results observed in nature. Also, insights about the potential of local signaling as a correction mechanism might find applications in synthetic biology more generally, as boundary formation is a fundamental patterning mechanism.

To better meet such applications, the presented model could be generalized or modified in interesting ways, as for instance by including more complex logics or smoothing the hard threshold by modeling actual Hill type dynamics, by modeling irregular grids and cell number as well as form changes, or by including additional correction mechanisms such as cell sorting.

5. Drosophila Bristle and Intestine Epithelium cell-type Pattern

In the following two sections we will briefly present an extended CA modeling approach to the formation of bristles for *Drosophila* and the small intestine homeostasis. A discussion of all three biological systems follows in Chap. 6.

5.1. Dots in Drosophila Bristle Formation

Here, we want to sketch a 'proof of principle' that simple threshold CA rules combined with a global gradient can not only generate of sharp boundaries or stripes in the presence of noise, but also to produce regularly spaced rows of dots. The biological system we are referring to is the generation of sensory organ precursor (SOP) cell patterns at the *Drosophila* thorax that will mature to bristles, thoroughly studied by Corson et al. (Corson, Couturier, Rouault, et al. 2017).

Corson et al. added fluorescent reporters of proneural factors, Achaete (Ac) and Scute (Sc), as well as of an early-response Notch factor. So they could observe the development of evenly spaced bristle rows one to five, called position '*Pos*-1' to '*Pos*-5' here (see Fig. 1 and supplementary Fig. 2 and Fig. 3 in (Corson, Couturier, Rouault, et al. 2017) for *Drosophila*-thorax snapshots at different time points, and spatially resolved Delta, Ac and Sc levels, that provide evidence for, and visualize the following paragraph):

About *2.5h after pupal formation* (APF), i.e., before onset of proneural activity, Delta is expressed as a smooth bimodal concentration profile, peaking at *Pos*-1 and *Pos*-5 with a minimum at the center (i.e. *Pos*-3). Notch activity is maximal at intermediate Delta levels, i.e., at *Pos*-2 and *Pos*-4, presumably as a result of cis-inhibition of Notch competing with trans-activation.

At about *6h APF*, proneural genes Ac and Sc inducing Delta activity are beginning to be expressed at *Pos*-1, *Pos*-3 and *Pos*-5.

7h APF the proneural stripes at *Pos*-1 and *Pos*-5 are refined via Notch mediated lateral inhibition. Cells in the stripe at position 3 begin to express Ac and Sc. The Delta gradient at *Pos*-1 and *Pos*-5 has decreased, while increasing at *Pos*-3.

8h APF the stripe at *Pos*-3 got refined. Expression of Ac and Sc has started at *Pos*-2 and *Pos*-4. The Delta gradient at *Pos*-1, *Pos*-3 and *Pos*-5 has decreased, while it began to peak at *Pos*-2 and *Pos*-4.

12h APF all stripes of proneural activity have been resolved into regularly spaced SOPs and any other cell has become an epidermis cell. The average distance between SOPs exceeds two cells, suggesting that Notch signaling also affects next nearest neighbors, e.g. via filopodia. Also, cells do not divide during the patterning process and rarely move.

Inspired by this research, we wondered, which extensions of a two-state CA model are

sufficient for a minimal model of the sensory-organ precursor patterning process. Is it sufficient to assume a nearest-neighbor range for the local signal? What happens without the decaying long-range signal gradient? We therefore chose a bottom up approach. Concretely, our cellular-automata system runs on a 2D grid with a dimension of 13×21 cells, and periodic boundary conditions. Each grid cell is assigned a cellular automaton that can be either in SOP state (red or '1'), or in epidermis state (green or '0'). The CA synchronously update their state. In the most simplistic model version, we started our exploration with, the update function only the takes state of it's four direct neighbors, and its own state, as an input. Essentially, it compares the signal, corresponding to its neighbors' summarized states, to a threshold a_0 or a_1, depending on its own state:

```
signal = 0
for neighbor in neighbor_state_list:
        signal += neighbor


if cell_state == 0:
        if signal <= a_0:
                newState = 1
        else:
                newState = 0
elif cellstate == 1:
        if signal <= a_1:
                newState = 1
        else:
                newState = 0
```

We started the simulation from a grid with all cells in epidermis state '0' (green). At every consecutive time step, the whole grid flips state, see Fig. 5.1(a). In order to break the symmetry, we introduce noise by adding a uniformly distributed random number r, $r \in [-\eta/2, +\eta/2]$, different for each cell at each time step, to the signal. This noise can be seen to represent a readout error of the cell. The 'noise level' η, determining the length of the interval a random number is drawn from, is used as a measure of the strength of the noise. Within few time steps, the dynamics changes from 'the majority of cells switch state every time step' to 'only few cells change state', and cells are arranged in a salt-and-pepper pattern (Fig. 5.1(b)).
In the biological system however, SOP cells do not change back anymore, thus we at least aim for a stationary state that is a fixed point. Also, SOP cells are arranged in orderly lines of equally spaced dots with an average distance larger than one neighbor. Consequently, we extended the neighborhood to also include the diagonal cells, i.e., to a Moore neighborhood, consisting of $\mathcal{N} = 9$ cells. The simulation results are shown for two different noise levels, $\eta = 3$ in the upper row of Fig. 5.1(c) and for $\eta = 1$ in the lower row. For the larger noise level, the system does not run into a fixed point within 100 time steps. For $\eta = 1$ we observe a fixed point with SOP cells being roughly homogeneously distributed within the grid. However, the SOPs are not arranged in lines, and the pattering process requires quite some switching back and forth between cell fates. Thus we included a representation of the initial Delta gradient, by adding to the signal one of the simplest non-monotonic functions, a piece wise linear V-shape with a maximum value of $M(t)$ and a minimum of 0, see Fig. 5.1(d), second

row. We assumed that the local, inhibitive dynamics would be much faster than the decay of the graded prepattern, and consequently decreased the maximum value of the external signal by a small amount only when the grid was in the same state at two consecutive time steps. In Fig. 5.1(d) we can observe that switching back from SOP fate to epidermis fate occurs rarely. The final fixed point pattern shows SOP cells in neatly organized rows, except from the outermost ones, due to periodic boundary conditions.
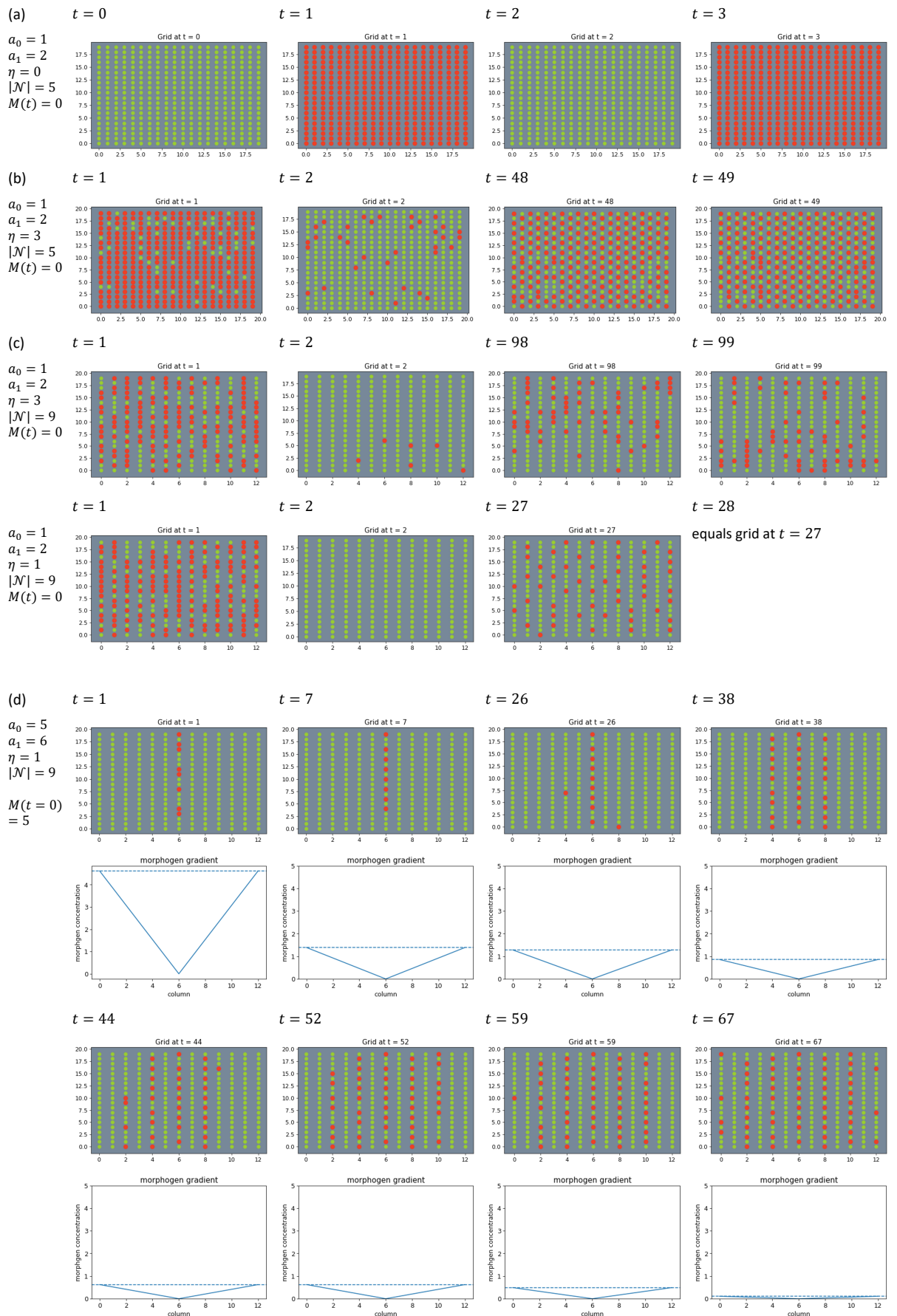
Figure 5.1.: Exemplary snapshots of the dynamics generated by different stages of the bottom-up model.

Figure 5.1.: *continued* (a) Snapshots of the dynamics for the first model without noise $\eta = 0$, with a Von Neumann neighborhood $\mathcal{N}$ and without external signal $M(t) = 0$. (b) Snapshots for the second model extending the first one by including noise, $\eta = 3$. (c) Snapshots for the third model extending the second model to a Moore neighborhood $\mathcal{N} = 9$ for two different noise levels: $\eta = 3$ in the first and $\eta = 1$ in the second row. (d) Snapshots for extending the model to an additional external V-shaped signal (second row) with a maximal value $M(t = 0) = 5$.

We conclude that, when adding noise, and a non-monotonous global signal decaying in time, we could qualitatively reconstruct aspects of the *Drosophila* patterning process, in the sense of generating a pattern of roughly evenly spaced SOP cells, organized in rows, in a process that consecutively patterns each row. However, we do not claim to have modeled the actual underlying signal mechanism: Delta-Notch signaling has many different aspects as outlined in Sec. 3.2, but we only used a simplistic inhibitive two-threshold rule. Nevertheless, it would be interesting to further test, if we can fully eliminate SOP to epidermis back transitions, by introducing an intermediate state that can either switch back to epidermis, or mature to an SOP state in the next time step. Also, simulation on a hexagonal grid with a next nearest cell neighborhood would probably give biologically more realistic results. Implementing both aspects might be sufficient to compare to the actual data in (Corson, Couturier, Rouault, et al. 2017).

5.2. **Homeostatis of the small Intestine Epithilium cell-type Distribution**

The following section is based on Max Knicker's work, documented in Ref. (Knicker 2021). He also kindly provided all figures. Max Knicker was jointly supervised by Stephan Kremser and the author, as well as by Ulrich Gerland.

The small intestine epithelium consists of crypts and villi to enlarge its surface for taking up nutrients. At the base of the crypts, stem cells divide and differentiate primarily to paneth, goblet, or enterocyte cells (Haber, Biton, Rogel, et al. 2017). As they are small in numbers (Haber, Biton, Rogel, et al. 2017), we decided to neglect enteroendocrine and tuft cells. All differentiated cells, except paneth cells, travel in a 'conveyor belt' like fashion to the tip of the villus, were they get aborted into the lumen (Chin, D. R. Hill, Aurora, and Spence 2017). Paneth cells move into the opposite direction (Gassler 2017). It is known that the differentiation frequencies of the cell types are influenced by an external, long-range Wnt signal gradient and by Notch-mediated lateral inhibition (Takahashi and Shiraishi 2020). We chose to model the small intestine crypts as a 1D stochastic CA system. A cell can 'update' to four different cell types — stem cell, enterocyte, paneth or goblet cell. Update probabilities of a cell depend on the cell types of its left and right neighbor, as well as the local concentration of the Wnt gradient (reduced to be either 'On' or 'Off'), see Fig. 5.2. Note that this update effectively summarizes conveyor-belt movement towards the villus and differentiation. We thus obtain a model with $3 \times 2 \times 4^3$ update probabilities as parameters. This large number can be reduced by biologically motivated assumptions such as, (i) goblet and enterocytes in the Wnt 'On' regime are essentially noise, i.e., the probability to update to a goblet or enterocyte is equal 'noise' n, $n \ll 1$, and vice versa for stem and paneth cells in the Wnt 'Off' regime. (ii) A cell cannot sense the direction of a signal, i.e. update probabilities for neighborhoods should be equal to those with the left and the right neighbor interchanged. Depending on how many additional assumptions were used, the automata system has 384, 128 or 80 free parameters. In order to model stem-cell differentiation and paneth-cell movement, we add two additional parameters p_{divide} and p_{move}. An exemplary run from an initial grid with only stem cells is shown in Fig. 5.2.

In order to obtain biologically reasonable values for the free parameters, we use experimental data from Ref. (Chwalinski and Potten 1989) and Ref. (Paulus, Loeffler, Zeidler, et al. 1993) showing a spatial probability distribution of how probable a certain cell type at each position is, see Fig. 5.3. Parameters are optimized with a particle-swarm algorithm, using the Canberra distance of the stationary CA probability distribution and this data. More precisely, first, a set of ('swarm' of) either 250 or 150 parameter sets ('particles') is chosen. Then the CA simulation is run 20 times, starting from an 'all stem cells' initial condition, for each parameter set for 100 time steps, in order to extract a stationary cell-type probability distribution. As a loss function, to update the parameter set, we decided for the Canberra distance between those distributions and the experimental data. Subsequently, the parameters, i.e., update probabilities, are changed to reduce the loss function, following the particle-swarm-algorithm: essentially, each particle 'moves' to its new coordinate (i.e. probability-value set) by a velocity, determined by it's own smallest loss value obtained so far, the swams' smallest loss value,
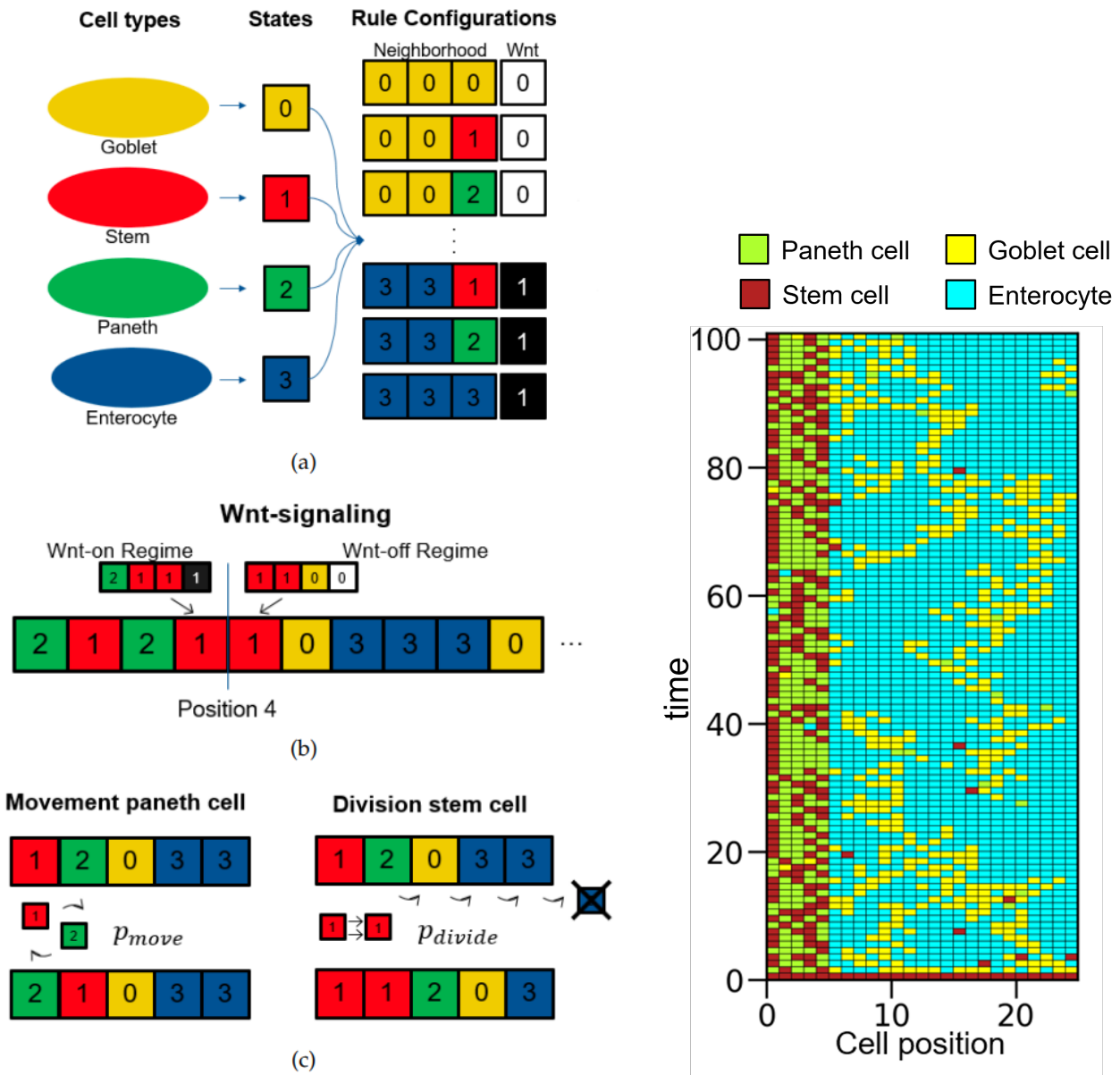
Figure 5.2.: *Left:* Schematic model description. Each cell of the 1D grid is assigned one of four cell states (goblet, stem, paneth or enterocyte). *(a)* The state of a cell is updated by an extended probabilistic cellular automaton. The CA is determined by the update probabilities it assigns to each of the cell states, given the neighborhood configuration and the presence or absence of the external signal (Wnt). *(b)* The external signal is On for positions 0 to 4 and Off elsewere. *(c)* In addition, backwards movement of paneth cells is explicitly modeled by the probability p_{move} and stem cell division by p_{divide}. *Right:* Exemplary simulation run with an optimized set of update probabilities.
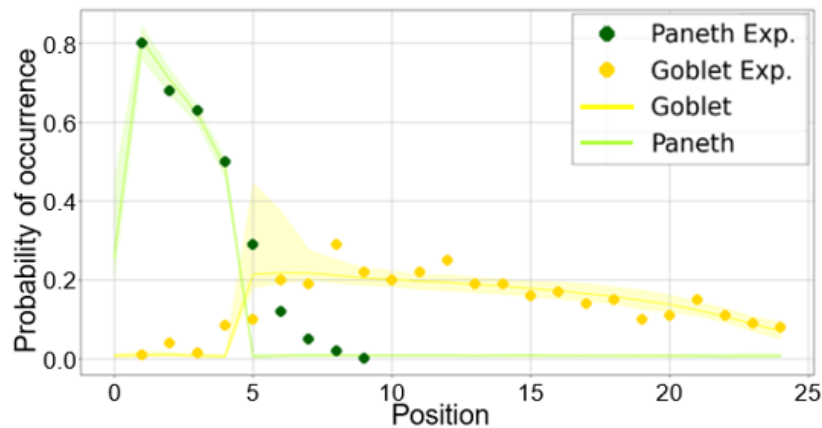
Figure 5.3.: Dots display experimental results, i.e., the spatial distribution frequency of paneth (green, (Chwalinski and Potten 1989)) and goblet (yellow, (Paulus, Loeffler, Zeidler, et al. 1993)) in the crypts of the mouse' small intestine. Shaded areas interpolate numerical results from the 100 update-probability sets resulting in smallest deviation of the generated patterns from the experimental data. The discrepancy at position 5 stems from our simplification of the Wnt gradient to switch from On to Off there.

and the direction it has been moving in before. This step is repeated 600 times. The results of 44 such optimization runs are displayed as a t-SNE representation in Fig. 5.4 for the largest parameter set.

The first panel shows the runs colored by the number of their run. Clusters of different colors do not mix, thus either the exploration of the particle swarm algorithm is not sufficient (despite parameter tuning) or the loss landscape is rather flat with many similarly deep local minima. The second panel shows that the Canberra distances of the 100 best solutions of each run are quite similar between the runs. In the last panel, each rule's entropy, defined as $\sum_{i=0}^{128} \left(-\sum_{k=0}^{3} p_{k,j} \log_2 p_{k,j} \right)$ is shown as a measure of its bias.

As a global minimum could not be found, we analyzed the ensemble of the 100 rules with the lowest Canberra distance, see Fig. 5.5 for results in the Wnt On regime. The upper row shows that in all those models paneth cells move about 10 to 25 times as fast than stem cells divide. As stem-cell division occurs roughly once a day (Schepers, Vries, Van Den Born, et al. 2011), and paneth cells move about one position per hour (Qiu, Roberts, and Potten 1994), this result is of the same order of magnitude. The remaining results shown are probabilities of becoming a stem cell for different neighborhood configurations, sorted with decreasing occurrence frequency. For example, the configuration listed in the 1st segment occurs most often during the simulation, the ones in the 2nd segment similarly often, and less frequent than the configuration in the first segment. All the rules of the ensemble have in common that the probability of becoming a stem cell ('1') is close to zero, if the central cell, and the left neighbor are also paneth cells ('2') ([2,2,2], [2,2,1]), except from the configuration with an enterocyte as a right neighbor. For this by far less frequent configuration, there is a wide spread in the probability of becoming a stem cell. These observations match the biological fact that differentiated cells rarely, if at all, reverse back to stem cells. Probabilities of updating to
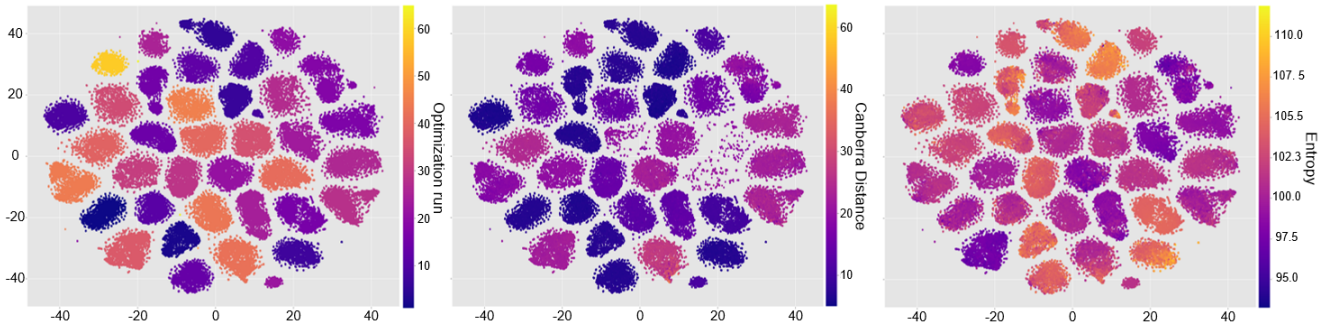
Figure 5.4.: t-SNE representation of 128'157 solutions of 44 particle swarm optimization runs in the Wnt On regime. *Left:* Different runs are plotted in different colors. *Center:* Solutions are colored with respect to the Canberra distance of the ensemble averaged stationary cell type distribution they generated and the experimental distributions. The ensemble average is over 20 duplicates and the stationary distribution was determined from the last 60 out of 100 time steps. *Right:* Solutions are colored by their entropy value, $\sum_{i=0}^{128} \left(-\sum_{k=0}^{3} p_{k,j} \log_2 p_{k,j} \right)$.
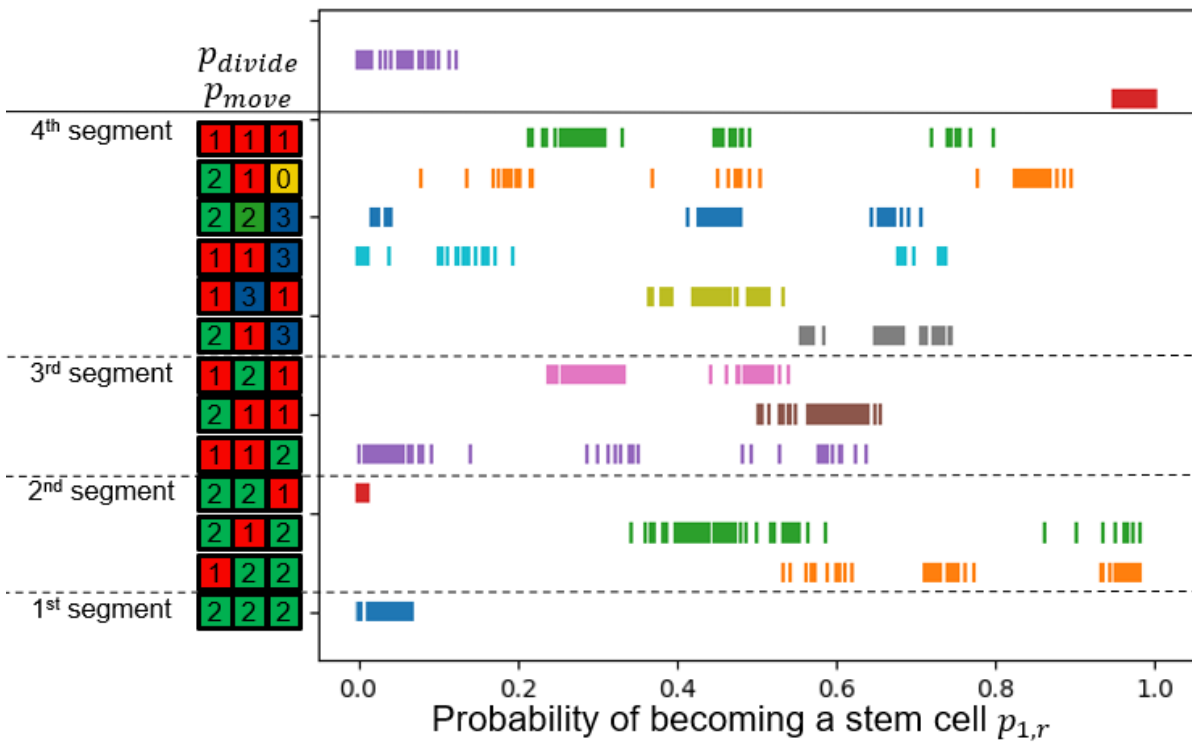
Figure 5.5.: Update probabilities for the 100 solutions of the 128'157 of Fig. 5.4 with lowest distance to experimental data in the Wnt On regime are shown. In the upper part, p_{divide} and p_{move} values are plotted. In the lower part, the probability of becoming a stem cell in the next time step is on the x-axis and on the y-axis different neighborhood configurations, ordered by how often they occurred during the simulation. 1$^{\text{st}}$ segment thereby refers to the most abundant configurations, ect.

a stem cell are greater than zero for input configurations with a stem cell as a left neighbor, and a differentiated cell in the center (i.e. [1,2,2], [1,2,1], [1,3,1]). However, this does not imply that cells 'de-differentiate', as cell movement to the right is modeled by the same update rule. For future work we suggest to allow for one more parameter, by introducing a cell-moving probability, for all but the paneth cells, such that cellular automata updates purely model cell differentiation instead of mixing time scales of differentiation and cell movement towards the tip.

Also, using additional data for optimizing the update probabilities would be beneficial, as in the current simulation up to 386 parameters where optimized with an input of 33 experimental data points. Therefore, we have unsuccessfully[1] tried to extract a spatial profile from small-intestine crypt single-cell RNA sequencing data (Haber, Biton, Rogel, et al. 2017) with the NovoSpaRc algorithm (Nitzan, Karaiskos, Friedman, and Rajewsky 2019). This algorithm essentially minimizes the difference of, the distance distribution of all pairwise cell distances in a chosen grid, and the pairwise distance between cells in gene-expression space.

A general limitation when comparing to experimental data is that the small intestine crypt is neither 1D, nor of the same size, even within the same mouse. Also cell type distributions are generally subject to many other factors than chemical signaling, e.g., nutrition dependent.

Despite the discussed shortcomings and the exploratory character of this work, the minimal extended CA model with parameters optimized by comparing to the *stationary* state frequency distributions of paneth and goblet cells could qualitatively reproduce two aspects of the patterning *dynamics*: the time scale ratio of stem cell division to movement of paneth cells and that differentiated cells generally do not become stem cells again.

---

[1]The algorithms performs best if additional marker genes are provided, which were not available in our case. Our results with single-cell spatial resolution are neither consistent among the two 'unbiased' data sets used, nor with the spatial cell-type distributes of (Paulus, Loeffler, Zeidler, et al. 1993) and (Chwalinski and Potten 1989). These two issues might stem from a different aim of the small-intestine study, which was to identify new cell types, not to faithfully represent distributions of known ones. Further, we observed that the algorithm clustered cells of the same type. This behavior was to be expected to some extend, as the algorithm is based on the assumption that cells close in gene expression space are also close in physical space. However, the small intestine is highly dynamic, and different cell-types often alternate spatially. Excluding the cell type marker genes from the gene expression profiles and feeding them to the algorithm only partly resolved the clustering issue.

6. Discussion of Part II

In Part II, we have studied three different developmental-biology systems, modeled in a minimal, extended CA framework — boundary formation, *Drosophila* bristle formation, and the small-intestine epithelial cell-type distribution. In the boundary formation system, the extended CA rules combine a noisy external morphogen with a noisy local signal, encoding the neighboring cell states. We find that first summing up both signals, and then comparing to a common threshold (SUM rule), outperforms an AND or OR logic combination of both signals, that first compare to separate thresholds. Even if noise on the local signal is significantly higher compared to noise on the global signal, all three correction mechanisms reduce boundary 'fuzziness'. For a more thorough discussion of this example see Sec. 4.4.

In the *Drosophila* bristle-formation example, we found that qualitatively reproducing the biological patterning process was possible if we included noise, used a larger neighborhood, and added a decaying global signal.

The small-intestine epithelium system was different in the sense that we allowed for the full range of parameters, rather than a minimal set: Each neighborhood configuration is assigned three probabilities of updating to one of four possible cell types. Ensemble averages of stationary-state 1D grid configurations generated by these sets of parameters are then compared to experimental data of stationary cell-type distributions. We optimize the CA parameters by comparing the CA generated and experimental cell-type distribution, but each particle-swarm optimization run found a different local minimum. One reason might be that the number of free parameters significantly exceeds the number of experimental data points. Nevertheless, studying the 100 'best' solutions allows us to draw order-of-magnitude conclusions of the ratio of paneth-cell movement to differentiation speed, that matches biological findings. We also observed that the model has apparently 'learned' that differentiated cells generally do not 'de-differentiate' back to stem cells.

All three examples have in common that noise and a global signal are crucial ingredients to the model. Due to the idealizations the model introduces — synchronous updates, symmetric grids, discrete cell states — noise is needed to break artificially introduced symmetries. In the boundary-formation model, noise destabilized the metastable, intermediate boundary positions, and thereby aided quick convergence to its stationary position, when established via the SUM rule. For the *Drosophila* bristle formation noise broke the spatial symmetry and ensured that we do not observe flickering behavior, despite the inhibitive nature of the rule. In the small intestine, we start from a 1D grid of just stem cells. If we had no noise, they could only update to one other cell type. The role of noise for increasing cellular diversity, has indeed been observed in developmental processes. For example photoreceptors sensitive to different wavelenghts are distributed in stochastic patterns, in *Drosophila* and human eyes (Ebadi, Perry, Short, et al. 2018).

The interplay of local and global signal as a successful patterning strategy seems to be a common motif in developmental systems. Morales, Raspopovic and Macron see 'embryonic

development as a guided self-organization process, where patterning and morphogenesis are controlled by a combination of exogenous signals and endogenous self-organization' (Morales, Raspopovic, and Marcon 2021). They substantiate this notion by numerous examples, many of them from embryonoids and organoids. For embryonic development, they, inter alia, list: digit formation, where a self-organized Turing mechanism creates periodic stripes that are aligned by an external Fgf gradient (Raspopovic, Marcon, Russo, and Sharpe 2014); formation of the primitive streak in mouse embryo, where self-organized anterior-posterior symmetry breaking in the epiblast is modulated by external inhibitors Dkk and Lefty (Yamamoto, Saijoh, Perea-Gomez, et al. 2004); and cells in the early stage zebrafish embyro, where dorso-ventral patterning is guided by the marginal and drosal organizer (Schier 2001). Schweisguth and Corson state that purely self-organized patterns are by essence variable, but for a reproducible outcome, such as digits in mammals, or a gradient with an invariant orientation, like Nodal in zebrafish, it requires particular initial or boundary conditions that may arise from the geometry of the tissue or pre-existing positional cues (Schweisguth and Corson 2019). They further point out that 'there is much evidence to suggest that cell-cell interactions can refine the interpretation of morphogen gradients, or even that the receiving cells contribute to shaping the gradient' and mention the example of the vertebrate neural tube patterned by Sonic Hedgehog.

Another point, we wish to discuss at the example of the small-intestine model, is the potential of CA modeling, when optimized with actual data. In the small-intestine model, we could determine candidate ensembles of update probabilities by comparing to stationary, spatial data of paneth and goblet cells. With more data, e.g., spatial distributions of stem cells and enterocytes, the ensembles might be reduced to a single optimal update-probability set. The field of single-cell data, including spatially resolved methods, multiplex transcriptomics, and temporally resolved single-cell methods, is quickly progressing, see (Mayr, D. Serra, and Liberali 2019) for a recent review. Nature Methods, for instance, has selected spatially resolved transcriptomics to be the 'Method of the Year' 2020 (Marx 2021). Thus, we are optimistic that data to compare the CA-generated cell distributions to will not be a bottleneck in the near future.

Assuming that we can find a global minimum, corresponding to a unique set of update probabilities, what do we actually model? The probabilities determine how probable it is for a cell at a certain position, of a certain type, to change to another type in the next time step, given the state of its neighboring cells, and an external gradient. First of all, these updates combine cell differentiation with cell movement, processes that could be disentangled by introducing an additional movement probability. Nevertheless, differentiation probabilities of a cell, in addition to chemical signaling, also depend on other factors, such as cell cycle, circadian rhythm, nutrients, mechanics and microenvironments, such as the extracellular matrix, and cell crowding (Mayr, D. Serra, and Liberali 2019). We can either use an extended CA model to infer 'effective' update probabilities, as parameters for all those processes are difficult to obtain. Or, we can extend it further to include more factors, and maybe make statements about how strongly each factor actually influences differentiation. Also, we can adjust time scales to average out some factors, such as cell cycle effects. In any case, we need to be careful, which factors are present in the data, and what the CA update probabilities consequently mean.

7. Resumee

The purpose of this work is to explore cellular automata models for pattern formation in biological tissues. In the first part, our aim is to characterize the dynamical behavior, or 'phenotype', of two-dimensional, outer-totalistic cellular automata, and to relate it to their 'genotype', their rule table (which uniquely defines a cellular-automaton rule). Therefore, we studied information-theoretic quantities to characterize the range of dynamical behaviors of the cellular-automata rules. In this context, we reviewed classification attempts based on their dynamical behavior resulting from single, an ensemble of, or all possible initial conditions. We have found that the single-site entropy of a pattern is a comparably robust information-theoretic observable that might aid as a pragmatic proxy of a rule's 'complexity'. However, we could not identify any rule-table based quantity that can predict a rule's ensemble averaged entropy.

In the second part, we explored how biologically inspired 'extended' stochastic cellular automata rules, extended to processing an additional global signal, can model pattern formation observed in tissues. More precisely, these extended cellular automata combine cell-cell communication with long-range morphogen input signals. Concretely, we modeled patterning processes of increasing biological specificity, starting from gene-expression boundary formation, *Drosophila* bristle-prosecutor formation and the stationary small-intestine epithelium cell-type distribution. In all three examples a long-range signal was experimentally observed, although the patterns are qualitatively different. With the biologically motivated cellular-automata-rule families we chose, the global signal was necessary for fast and robust formation of the observed pattern. We also observed that in our minimal models noise implicitly aided the pattern formation, by breaking spatial or cell-type symmetries. For a more detailed discussion of the interplay of local and global signaling and the role of noise, see Chap. 6.

Our hope to gain some general insights about cellular automata in Part I, in order to apply them to concrete biological pattern-formation problems in the second part has not substantiated. On a purely theoretical level, characterization naturally aims at distinguishing 'complex' cellular automata from chaotic or periodic ones. Indeed, a part of the problem is to define what 'complex', 'chaotic' and 'periodic' behavior for those discrete dynamical systems actually is. Complex CAs, capable of universal computation, attract most attention (C. G. Langton 1990; Wuensche, M. Lesser, and M. J. Lesser 1992; Wolfram 1984a). However, these CAs are the ones that we would not expect in a biological system: their space-time traces strongly depend on initial conditions, and their transients to 'typical behavior' can be of undetermined length due to Turing's halting problem. Nevertheless, meaningful and decidable classification of cellular automata is a highly interesting problem, unsolved for more than 50 years and we could outline some ideas on how to further tackle it.

Finally, we are excited to see that quite different biological pattern-formation processes can indeed be phrased in a CA-based framework — possibly by extending it to, e.g., cell movement, division, apoptosis. In case of boundary formation, we could make statements on which of the

elementary signal-processing logics works best, in the sense of creating a sharp boundary at the correct position that scales with system size. In particular, boundary fuzziness was reduced by adding a local signal of the neighboring cell's states to the morphogen concentration — even when noise on the local signal was twice as strong as noise on the morphogen. In the small-intestine epithelium example, the CA model allowed order-of-magnitude statements about the relation of different time scales in the system. In principle, differentiation probabilities for the distinct cell types, based on their neighbors and longer-ranged morphogen signals, could be extracted with the CA modeling pipeline. The optimal CA rule could allow for conclusions about the involved type of cell-cell communication: inductive, inhibitive or more complex. That requires more complete cell-type, or higher spatially resolved gene-expression data, subject of a quickly evolving research field. The potential of the modeling framework, to infer time scales and direct-neighbor communication, in principle extends to any biological system that — is effectively one or two dimensional and of fixed size, for which the global signal(s) are known qualitatively, and spatially resolved cell-type data is available.

Appendices

A. Appendix for Part I

A.1. Additional plots for m-site Entropy expansion
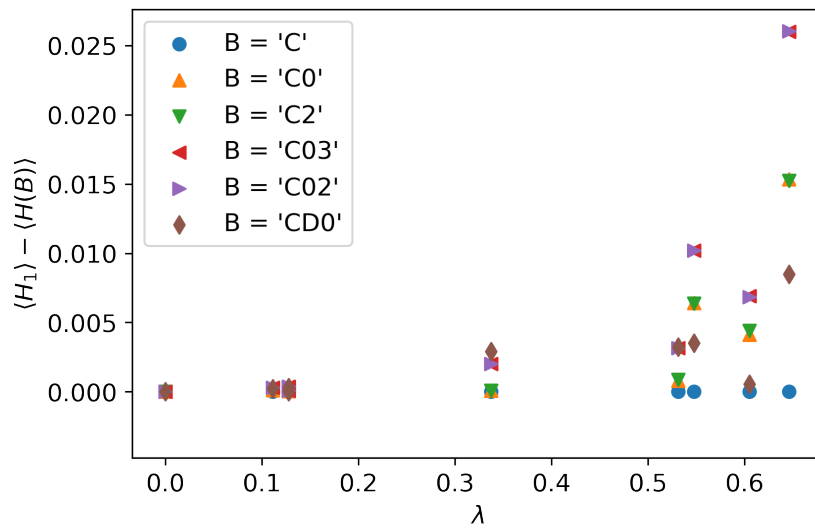

Figure A.1.: Same data as main-text Fig. 2.7, here plotted versus λ. Snapshots for the rules with $\lambda = 0.55$, 0.60, 0.65 are shown below. This random-table sampling has actually created a $k = 3$ rule, with a single non-zero state being more frequent than the quiescent state, i.e. the rule with $\lambda = 0.6$. Such 'artifacts' that can be circumvented by the *RTE*.

Figure A.2.: For a randomly sampled outer-totalistic rule with $\lambda \approx 0.55$ and $k = 3$ states, we show exemplary grid snapshots (first column), spatial autocorrelation of the grid (second column), and histograms for different block sizes (third to fifth column), at $t = 5$ (upper row), $t = 10$ (middle row) and $t = 75$ (bottom row). The rule starts from are random initial grid of 200×200 sites. Although we can observe structures of more than one site, the difference in frequency between different blocks is rather small.
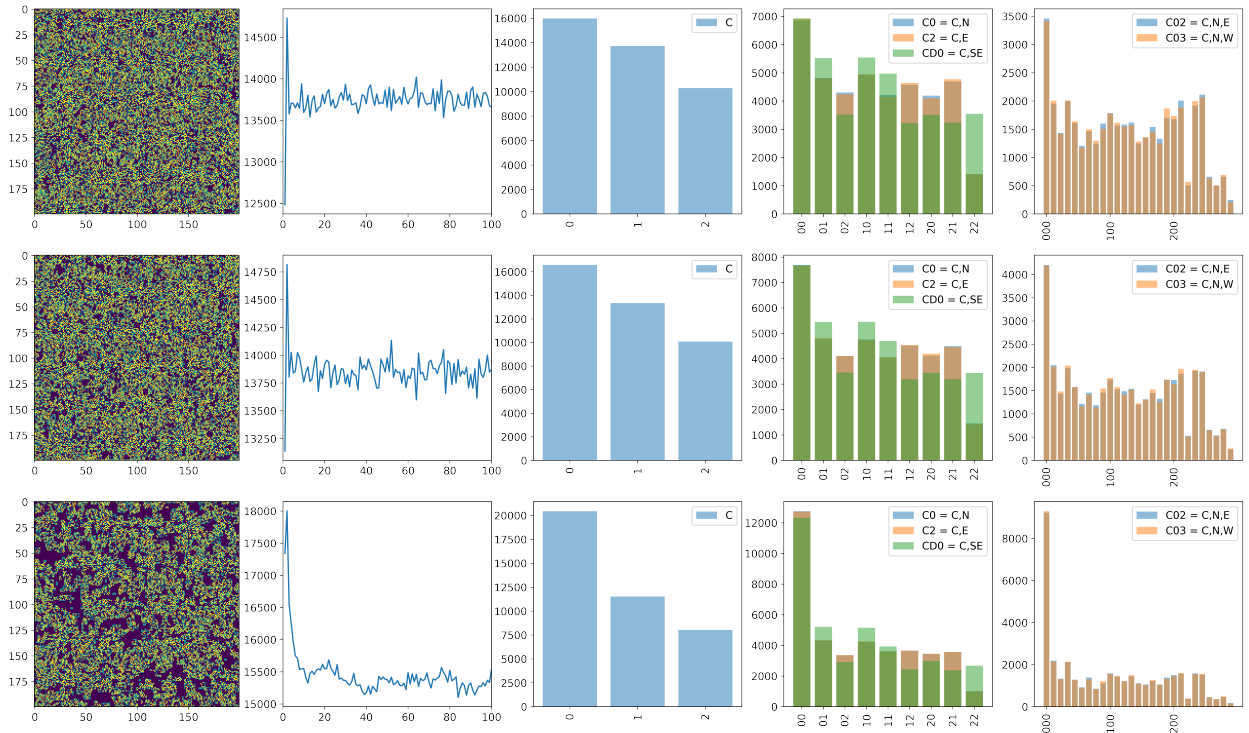

Figure A.3.: Same as Fig. A.2 for a randomly sampled rule with $\lambda \approx 0.65$.
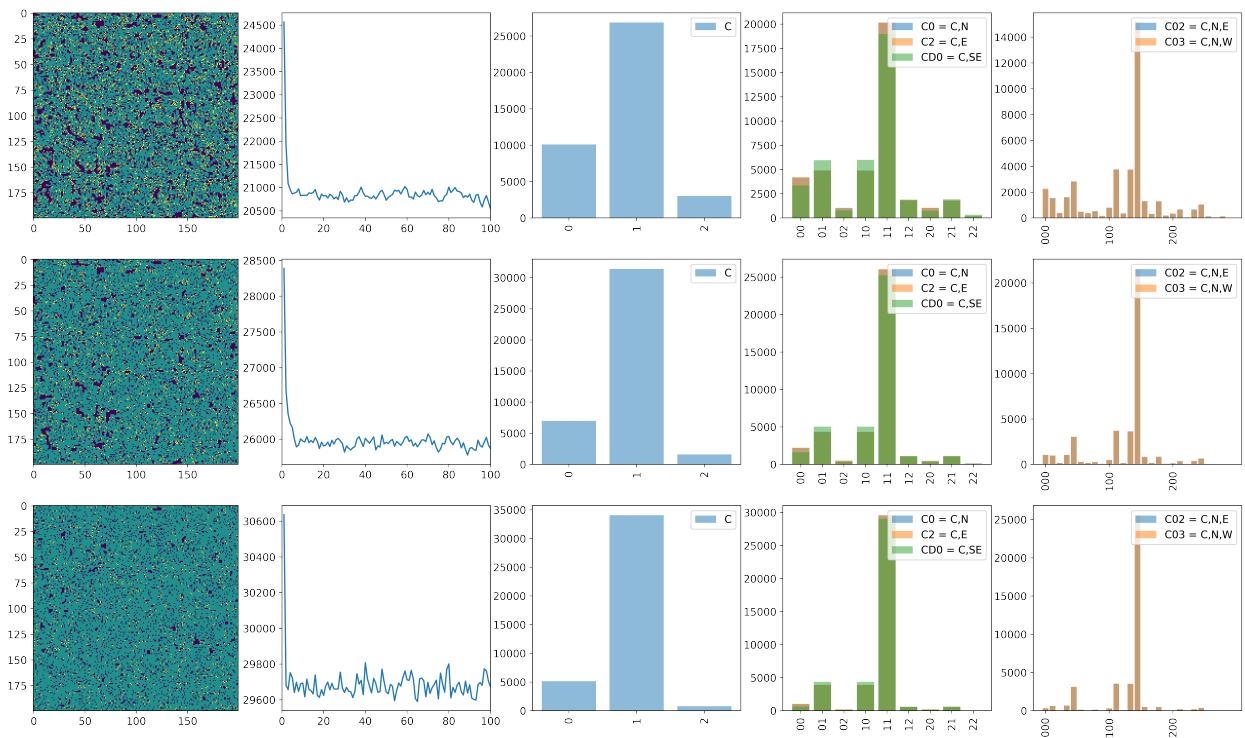
Figure A.4.: $\lambda \approx 0.60$

A.2. Kurkas attractor-based classification of infinite CA

Every dynamical system (X, r) falls in exactly one of the following classes:

A1 There exists a pair of disjoint attractors

A2 There exists a unique minimal quasi-attractor Q of r

A3 There exists a unique minimal attractor different from $\omega(X)$

A4 There exists a unique attractor $\omega(X) \neq X$

A5 There exists a unique attractor X

According to (Kůrka 1997) (and cited by (Vispoel, Daly, and Baetens 2021)) Wolfram Class I and II CA's belong to $A2 \cup A3$, Class III CA belong to $A4 \cup A5$ and Class IV CA belong to A1.

However, the original work of Hurley states that 'a possible description of the relation of Wolfram's classes to [his classification stated above] is [...] Class 4: Case 2'. Note that Hurley's Case 3 corresponds to Kurka's Class A1. Moreover, Hurley gives an example for a rule in his Case 3, the identity rule, which obviously does not belong to Wolfram's complex cellular automata.

A.3. Extension of Quax work to 2D

An interesting extension of Quax work would be to check if it generalizes to outer-totalistic 2D Cellular Automata. In contrast to the elementary cellular automata there is no heuristic classification to measure the Synergies' mutual information with. However, we can use a finite-time classification as a proxy. Parameters to decide on are the maximal simulation time $t_{\max}$, the grid size L and the number of initial conditions N_{ini} and the number of states k.

Choosing $k = 2$ In order not having to rely on a sampling scheme for rules, let us start with only two states ($k = 2$). Thus there are 10 different neighborhood configurations: either 0, 1, 2, 3, or 4 neighbors in state 1 times the two options for the central cell. Consequently we have $2^{10} = 1024$ rules. A number still computationally feasible.
A small k obviously also strongly decreases the number of possible initial conditions.

Choosing the grid length L A somewhat natural choice of grid size is one such that each neighborhood configuration can in principle occur at least once. Also some measures we discussed, like the transition time, depend on change of number of neighborhood configurations. As there are only 10, we have $L \geq 4$. Also, the initial grids should be large enough to avoid feedback due to periodic boundary conditions, when measuring the synergy after t time steps. Quarx has shown that for the ECA $t = 3$ is sufficient to reach maximal predictive power. For 2D Cellular Automata, information can also only spread one cell per time step. Also, the outer-totalistic Cellular Automata are 1D in the sense that they don't use any directional information. Thus we expect $t = 3$ to give reliable results also for our 2D system. Consequently, we choose a grid size of $L = 2 \times 3 + 1 = 7$. After obtaining results with those choices, we need to check, if these are robust with respect to larger grid sizes.

Choosing the number of initial conditions $N_{\mathbf{ini}}$ For the choices of grid size $L = 7$ and $k = 2$, it is not computationally feasible to sample all initial conditions: $k^{L^2} = 2^{49}$. For the smallest grid that would allow for every neighborhood configuration to occur at least once, i.e. $L = 4$, we would have $2^{16} = 65'536$ initial configurations. We consequently start with testing $N_{\mathrm{ini}} = 2^{16}$ random initial conditions.

Choosing $t_{\mathbf{max}}$ The maximal simulation time has to be sufficiently small to be computationally feasible given the above choices and large enough to allow transient effects to die out.

Finite time characterization of the rules For a given initial condition, a rule is classified as follows:

- Type I rules (fixed point rules): stationary pattern reached within $t_{\max}$ time steps

- Type II rules (cyclic rules): cyclic sequence of patterns reached within $t_{\max}$ time steps

- Type III (somewhat chaotic rules): The rule is neither Type I nor Type II and no new neighborhood configurations are found after some time $t < t_{\max}$, as this is a necessary condition for having converged to typical behavior. The simplest chaotic rule realizes only neighborhood conditions with two different states

- Type IV (highly chaotic rules): all possible, i.e. $2^5 = 32$ neighborhood configurations have been realized (not counting the initial grid)

- Undetermined

An obvious question is, how to characterize rules that are in one class for some initial conditions and in another for others. We could actually just use a vector of three entries to characterize a rule: the first entry describes the fraction of initial conditions that displayed 'Type I' behavior, the second 'Type II' behavior and the third 'Type III' behavior. With this characterization, we could also study, if there actually exist rules that show Type I and Type III behavior, but no Type II behavior - a result that would not really fit in the rule space picture Langton is drawing. If such rules don't exist, we could pragmatically also average over the class values: '1' for Type I, '2' for Type II, '3' for 'Undetermined' and '4' for 'Clearly chaotic'. With this approach we might repeat (Quax, Chliamovitch, Dupuis, et al. 2017) procedure to measure the mutual information between our finite time classification and the set of information theoretic quantities including Synergy.

B. Appendix for Part II

B.1. Comparison of our boundary-formation model to other models

Comparison to the Ising model

The Ising model fundamentally is an equilibrium model, whereas the CA model also describes dynamics. To nevertheless compare both models, we choose Monte Carlo Updates for Ising and random updates (i.e. only one cell at a time) for the cellular automaton.

For all possible grid configurations we need to compare transition probabilities. As the interaction is local and only one site is changed at a time, it is sufficient to compare $P^{\text{Ising}}_{-1 \to 1}$ and $P^{\text{CA}}_{-1 \to 1}$ for a configuration s such that $s_k(t) = -1$ as well as $P^{\text{Ising}}_{1 \to -1}$ and $P^{\text{CA}}_{1 \to -1}$ for $s_k(t) = 1$. $P_{-1 \to -1}$ and $P_{1 \to 1}$ follow as one minus the checked ones.

Concretely we can pose the following question: if it is possible to find a noise distribution for the cellular automaton s.t. the Ising MC-update probability $P^{\text{Ising}}_{-1 \to 1} = P^{\text{CA}}_{-1 \to 1}$, can $P^{\text{Ising}}_{1 \to -1} = P^{\text{CA}}_{1 \to -1}$ also be fulfilled?

For the Ising model with an MC updating scheme: Choose a random site k in state s_k:
If $r \propto \text{Uni[0,1]} < P = \text{Min}\left[1, \exp\left(-\frac{\Delta E}{T}\right)\right]$: flip state,
with k_B set to 1, and Uni[0,1] referring to the uniform distribution within the Interval [0, 1] and

$$
\begin{aligned}
\Delta E &= \Delta H \left(s_k(t+1), s_k(t)\right) \\
&= -J \sum_{j \in \mathcal{N}(k)} \left(s_k(t+1) - s_k(t)\right) s_j(t) \,,
\end{aligned}
$$

with

$$
H = -J \sum_{\langle i,j \rangle} s_i s_j - \sum_i (g x_i + c) s_i \,.
$$

For the CA model choose a random site k
If signal(k) + noise$(n, k) > a$: remain respectively switch ON,
with signal$(k) = \sum_{j \in \mathcal{V}(k)} s_j + km$
$\Leftrightarrow$ if noise$(n, k) > -\sum_{j \in \mathcal{V}(k)} s_j + km$: $s_k(t+1) = 1$

A comparison to the Ising model with MC updates suggests that if there exists a probability distribution noise(n, k) such that Ising corresponds to CA then it also exists for 0 external field and $J = 1$ and vice versa.

Let us start with the actual comparison.

Assume $s_k = -1$, then for the Ising model

$$P^{\text{Ising}}_{-1 \to 1} = \begin{cases} \exp\left(-\dfrac{\Delta E}{T}\right) & \text{for} \quad \Delta E > 0 \\[2mm] = 1 & \text{for} \quad \Delta E < 0 \end{cases}$$

$$= \begin{cases} \exp\left(-\dfrac{-2\sum_{j\in\mathcal{V}(i)} s_j}{T}\right) & \text{for} \quad -\displaystyle\sum_{j\in\mathcal{V}(i)} s_j > 0 \\[4mm] = 1 & \text{for} \quad -\displaystyle\sum_{j\in\mathcal{V}(i)} s_j < 0 \end{cases}$$

In case of the CA noise(n,k) is a random number r, $r \propto f(r)$ implying

$$P^{\text{CA}}_{-1 \to 1} = P\left(r > -\sum_{j\in\mathcal{V}(i)} s_j\right) = \int\limits_{-\sum_{j\in\mathcal{V}(i)} s_j}^{\infty} r f(r) \mathrm{d}r$$

This form suggests the choice of an exponential distribution:

$$P^{\text{CA}}_{-1 \to 1} = \begin{cases} \exp\left(-\dfrac{-2\sum_{j\in\mathcal{V}(i)} s_j}{\sigma}\right) & \text{for} \quad -\displaystyle\sum_{j\in\mathcal{V}(i)} s_j > 0 \\[4mm] = 1 & \text{for} \quad -\displaystyle\sum_{j\in\mathcal{V}(i)} s_j < 0 \end{cases}$$

Consequently $P^{\text{CA}}_{-1\to 1} = P^{\text{Ising}}_{-1\to 1}$. But as $P^{\text{CA}}_{-1\to 1} = P^{\text{CA}}_{1\to 1}$ and $P^{\text{Ising}}_{1\to 1} = 1 - P^{\text{Ising}}_{-1\to 1}$ the transition probabilities $P^{\text{CA}}_{1\to 1}$ and $P^{\text{Ising}}_{1\to 1}$ cannot coincide.

Comparison to Hillenbrandt et al.

The linear form of the morphogen gradient signal is due to interpreting the signal to be the logarithm of the actual morphogen molecule concentration, which is commonly assumed to be approximately exponentially decaying. This choice of interpretation is inspired by (Hillenbrand, Gerland, and Tkačik 2016)s work showing a one to one correspondence between gene regulation by a Hill type thermodynamic model and a Boltzmann distribution, if the logarithm of the exponential morphogen profile contributes to the Hamiltonian as a linear external field, as we will discuss below. Note that the additive noise N_g on the global signal consequently maps to fluctuations proportional to the morphogen molecule concentration representing extrinsic noise on the morphogen molecule concentration.

First, we explain the biological Hill-type gene regulation model we want to map to, then we explain how this is achieved in a simplified version of (Hillenbrand, Gerland, and Tkačik 2016)s work. I.e. the thermodyamic Hill-type gene regulation is mapped to a Bolzmann distribution representing internal noise in the system. Finally we compare this work to the system studied in the main text with the result that both models share the same external noise, but our system has no internal noise in that sense.

Thermodynamic Hill-type gene regulation For simplicity let use a 1 dim grid and denote by $\sigma(x) = \pm 1$ the state of a cell at position x in order to comply with notation used in (Hillenbrand, Gerland, and Tkačik 2016). In the Hill type thermodynamic model of gene regulation, we have

$$P(\sigma(x) = 1) = \frac{c(x)^{\tilde{n}}}{c(x)^{\tilde{n}} + \tilde{K}^{\tilde{n}}}, \tag{B.1}$$

with $c(x)$ the morphogen concentration at x, $\tilde{n}$ the number of binding sites to the promoter that have to be occupied simultaneously in order to trigger gene expression and $\tilde{K}^{\tilde{n}} = \frac{k_{\text{off}}}{k_{\text{on}}}$ the promoter affinity [Alon 2019, An Introduction to Systems Biology Design, A.3].

(Hillenbrand, Gerland, and Tkačik 2016) model For an equilibrium thermodynamic system in an external field represented by the morphogen concentration and nearest neighbor interaction, we have

$$Q_\theta(\vec{\sigma}) = \frac{1}{Z_\theta} \exp\left(-\beta H_\theta(\vec{\sigma})\right), \tag{B.2}$$

with θ a set of parameters and

$$H_\theta(\vec{\sigma}) = -\sum_{x=1}^{N} h(m(x))\sigma(x) - J \sum_{x=1}^{N-1} \sigma(x)\sigma(x+1), \tag{B.3}$$

and $h(x) = n(m(x) - E)$.

Now comes the crucial step: if we assume that $m(x) = \log(c(x))$, then 'there is an exact mathematical relation between the probability that the patterning gene is ON, $P(\sigma(x) = 1)$ and the Hill-type thermodynamic model of regulation for a gene σ. [...] it is easy to show that the parameters n and E relate the bias $h(x)$ to the morphogen signal $m(x)$ in our model up to a multiplicative factor, to $\tilde{n}$ and $\log(\tilde{K})$.' (Hillenbrand, Gerland, and Tkačik 2016).

For $J = 0$ it follows that

$$P_{\text{Ising, J=0}}(\sigma(x) = 1) = \frac{\exp\left(\beta h(m(x))\right)}{\exp\left(\beta h(m(x))\right) + \exp\left(-\beta h(m(x))\right)}$$

$$\text{using the Ansatz } h(m(x)) = n\left(\log(c(x)) - E\right)$$

$$= \frac{e^{-\beta n E}c(x)^{\beta n}}{e^{-\beta n E}c(x)^{\beta n} + e^{\beta n E}c(x)^{-\beta n}}$$

$$= \frac{c(x)^{2n\beta}}{c(x)^{2n\beta} + e^{2n\beta E}}.$$

Setting $\tilde{n} = 2n\beta$ and $\log(K) = E$, it follows that $P_{\text{Ising, J=0}} = P_{\text{Hill}}$.

Note that this mapping has not taken the external noise on the gradient into account, i.e. that the morphogen concentration is subject to multiplicative Gaussian noise. This noise is consequently an additive Gaussian noise on the signal, $\log(m(x))$, caused by the morphogen $m(x)$.

Comparison to Morphogen Only model discussed in the main text For the cellular automata though, i.e. the Gradient only version without correction mechanisms, this external noise is the only noise we consider. The internal noise is neglected and we assume a sharp threshold rule corresponding to Hill type regulation in the limit $\tilde{n} \to \infty$.

For the CA we have

$$P(\sigma(x) = 1) = \int\limits_{a-mx}^{\infty} \mathcal{N} \left(n; \mu = 0, \text{Var} = \frac{\eta^2}{3} \right) dn . \tag{B.4}$$

This expression is in general different from a Boltzmann distribution.

Bibliography

Adhyapok, P., X. Fu, J. P. Sluka, S. G. Clendenon, V. D. Sluka, Z. Wang, K. Dunn, J. E. Klaunig, and J. A. Glazier (2021). "A computational model of liver tissue damage and repair." In: *PLoS ONE* 15.12 December, pp. 1–26. ISSN: 19326203. DOI: 10.1371/journal.pone.0243451.

Adler, M. and U. Alon (2018). "Fold-change detection in biological systems." In: *Current Opinion in Systems Biology* 8, pp. 81–89. ISSN: 24523100. DOI: 10.1016/j.coisb.2017.12.005.

Aliee, M., J. C. Röper, K. P. Landsberg, C. Pentzold, T. J. Widmann, F. Jülicher, and C. Dahmann (2012). "Physical mechanisms shaping the Drosophila dorsoventral compartment boundary." In: *Current Biology* 22.11, pp. 967–976. ISSN: 09609822. DOI: 10.1016/j.cub.2012.03.070.

Alon, U. (2007). "Network motifs: Theory and experimental approaches." In: *Nature Reviews Genetics* 8.6, pp. 450–461. ISSN: 14710056. DOI: 10.1038/nrg2102.

Artavanis-Tsakonas, S., M. D. Rand, and R. J. Lake (1999). "Notch signaling: Cell fate control and signal integration in development." In: *Science* 284.5415, pp. 770–776. ISSN: 00368075. DOI: 10.1126/science.284.5415.770.

Bagnoli, F., R. Rechtman, and S. Ruffo (1992). "Damage spreading and Lyapunov exponents in cellular automata." In: *Physics Letters A* 172.1-2, pp. 34–38. ISSN: 03759601. DOI: 10.1016/0375-9601(92)90185-0.

Barbier, I., R. Perez-Carrasco, and Y. Schaerli (2020). "Controlling spatiotemporal pattern formation in a concentration gradient with a synthetic toggle switch." In: *Molecular Systems Biology* 16.6, pp. 1–15. ISSN: 1744-4292. DOI: 10.15252/msb.20199361.

Barkai, N. and B.-Z. Shilo (2020). "Reconstituting tissue patterning." In: *Science (New York, N.Y.)* 370.6514, pp. 292–293. ISSN: 1095-9203. DOI: 10.1126/science.abe4217.

Bashor, C. J., N. Patel, S. Choubey, A. Beyzavi, J. Kondev, J. J. Collins, and A. S. Khalil (2019). "Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies." In: *Science (New York, N.Y.)* 597.May, pp. 593–597. DOI: 10.1126/science.aau8287.

Bialek, W. (2012). *Biophysics: Searching for principles*. New Jersey: Princeton University Press, p. 640. ISBN: 9780691138916.

Bier, E. (2000). "Drawing lines in the Drosophila wing: Initiation of wing vein development." In: *Current Opinion in Genetics and Development* 10.4, pp. 393–398. ISSN: 0959437X. DOI: 10.1016/S0959-437X(00)00102-7.

Blair, S. S. (2007). "Wing vein patterning in Drosophila and the analysis of intercellular signaling." In: *Annual Review of Cell and Developmental Biology* 23, pp. 293–319. ISSN: 10810706. DOI: 10.1146/annurev.cellbio.23.090506.123606.

Boareto, M. (2020). "Patterning via local cell-cell interactions in developing systems." In: *Developmental Biology* 460.1, pp. 77–85. ISSN: 1095564X. DOI: 10.1016/j.ydbio.2019.12.008.

Bollenbach, T., P. Pantazis, A. Kicheva, C. Bökel, M. González-Gaitán, and F. Jülicher (2008). "Precision of the Dpp gradient." In: *Development* 135.6, pp. 1137–1146. ISSN: 09501991. DOI: 10.1242/dev.012062.

Bolouri, H. and E. H. Davidson (2002). "Modeling transcriptional regulatory networks." In: *BioEssays* 24.12, pp. 1118–1129. ISSN: 02659247. DOI: 10.1002/bies.10189.

Buchler, N. E., U. Gerland, and T. Hwa (2003). "On schemes of combinatorial transcription logic." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.9, pp. 5136–5141. ISSN: 00278424. DOI: 10.1073/pnas.0930314100.

Buske, P., J. Galle, N. Barker, G. Aust, H. Clevers, and M. Loeffler (2011). "A comprehensive model of the spatio-temporal stem cell and tissue organisation in the intestinal crypt." In: *PLoS Computational Biology* 7.1. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1001045.

Chen, Z. and M. B. Elowitz (2021). "Programmable protein circuit design." In: *Cell*, pp. 1–18. ISSN: 0092-8674. DOI: 10.1016/j.cell.2021.03.007.

Chin, A. M., D. R. Hill, M. Aurora, and J. R. Spence (2017). "Morphogenesis and maturation of the embryonic and postnatal intestine." In: *Seminars in Cell and Developmental Biology* 66, pp. 81–93. ISSN: 10963634. DOI: 10.1016/j.semcdb.2017.01.011.

Chwalinski, S. and C. S. Potten (1989). "Crypt base columnar cells in ileum of BDF1 Male Mice—their numbers and some features of their proliferation." In: *American Journal of Anatomy* 186.4, pp. 397–406. ISSN: 15530795. DOI: 10.1002/aja.1001860409.

Collier, J. R., N. A. Monk, P. K. Maini, and J. H. Lewis (1996). "Pattern formation by lateral inhibition with feedback: A mathematical model of delta-notch intercellular signalling." In: *Journal of Theoretical Biology* 183.4, pp. 429–446. ISSN: 00225193. DOI: 10.1006/jtbi.1996.0233.

Colman-Lerner, A., A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce, and R. Brent (2005). "Regulated cell-to-cell variation in a cell-fate decision system." In: *Nature* 437.7059, pp. 699–706. ISSN: 14764687. DOI: 10.1038/nature03998.

Corson, F., L. Couturier, H. Rouault, K. Mazouni, and F. Schweisguth (2017). "Self-organized Notch dynamics generate stereotyped sensory organ patterns in Drosophila." In: *Science* 356.6337, eaai7407. ISSN: 0036-8075. DOI: 10.1126/science.aai7407.

Crozatier, M., B. Glise, and A. Vincent (2004). "Patterns in evolution: Veins of the Drosophila wing." In: *Trends in Genetics* 20.10, pp. 498–505. ISSN: 01689525. DOI: 10.1016/j.tig.2004.07.013.

Culik II, K. and S. Yu (1988). "Undecidability of CA classification schemes." In: *Complex Systems* 2.2, pp. 177–190. ISSN: 0891-2513.

Dahmann, C., A. C. Oates, and M. Brand (2011). "Boundary formation and maintenance in tissue development." In: *Nature Reviews Genetics* 12.1, pp. 43–55. ISSN: 14710056. DOI: 10.1038/nrg2902.

De Celis, J. F. (2003). "Pattern formation in the Drosophila wing: The development of the veins." In: *BioEssays* 25.5, pp. 443–451. ISSN: 02659247. DOI: 10.1002/bies.10258.

De Celis, J. F. and S. Bray (1997). "Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the Drosophila wing." In: *Development* 124.17, pp. 3241–3251. ISSN: 09501991. DOI: 10.1242/dev.124.17.3241.

Dessaud, E., L. L. Yang, K. Hill, B. Cox, F. Ulloa, A. Ribeiro, A. Mynett, B. G. Novitch, and J. Briscoe (2007). "Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism." In: *Nature* 450.7170, pp. 717–720. ISSN: 14764687. DOI: 10.1038/nature06347.

Ebadi, H., M. Perry, K. Short, K. Klemm, C. Desplan, P. F. Stadler, and A. Mehta (2018). "Patterning the insect eye: From stochastic to deterministic mechanisms." In: *PLoS Computational Biology* 14.11, pp. 1–15. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1006363. arXiv: 1704.02379.

Erdmann, T., M. Howard, and P. R. Ten Wolde (2009). "Role of spatial averaging in the precision of gene expression patterns." In: *Physical Review Letters* 103.25, pp. 2–5. ISSN: 00319007. DOI: 10.1103/PhysRevLett.103.258101.

Exelby, K., E. Herrera-Delgado, L. G. Perez, R. Perez-Carrasco, A. Sagner, V. Metzis, P. Sollich, and J. Briscoe (2021). "Precision of tissue patterning is controlled by dynamical properties of gene regulatory networks." In: *Development (Cambridge)* 148.4. ISSN: 14779129. DOI: 10.1242/dev.197566.

Fatès, N. (2014). "A guided tour of asynchronous cellular automata." In: *Journal of Cellular Automata* 9.5-6, pp. 387–416. ISSN: 15575977.

Field, G. D. and F. Rieke (2002). "Nonlinear Signal Transfer from Mouse Rods to Bipolar Cells and Implications for Visual Sensitivity." In: *Neuron* 34, pp. 773–785.

Gardner, M. (1970). "The fantastic combinations of John Conway's new solitaire game 'life'." In: *Scientific American* 223, pp. 120–123.

Gassler, N. (2017). "Paneth cells in intestinal physiology and pathophysiology." In: *World Journal of Gastrointestinal Pathophysiology* 8.4, pp. 150–160. ISSN: 2150-5330. DOI: 10.4291/wjgp.v8.i4.150.

Gerhart, J. (1999). "1998 Warkany Lecture: Signaling pathways in development." In: *Teratology* 60.4, pp. 226–239. ISSN: 00403709. DOI: 10.1002/(SICI)1096-9926(199910)60.

Gregor, T., D. W. Tank, E. F. Wieschaus, and W. Bialek (2007). "Probing the Limits to Positional Information." In: *Cell* 130.1, pp. 153–164. ISSN: 00928674. DOI: 10.1016/j.cell.2007.05.025.

Gutowitz, H. and C. Langton (1995). "Mean field theory of the edge of chaos." In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 929, pp. 52–64. ISSN: 16113349. DOI: 10.1007/3-540-59496-5_288.

Gutowitz, H. A., J. D. Victor, and B. W. Knight (1987). "Local structure theory for cellular automata." In: *Physica D: Nonlinear Phenomena* 28.1-2, pp. 18–48. ISSN: 01672789. DOI: 10.1016/0167-2789(87)90120-5.

Haber, A. L., M. Biton, N. Rogel, R. H. Herbst, K. Shekhar, C. Smillie, G. Burgin, T. M. Delorey, M. R. Howitt, Y. Katz, I. Tirosh, S. Beyaz, D. Dionne, M. Zhang, R. Raychowdhury, W. S. Garrett, O. Rozenblatt-Rosen, H. N. Shi, O. Yilmaz, R. J. Xavier, and A. Regev (2017). "A single-cell survey of the small intestinal epithelium." In: *Nature* 551.7680, pp. 333–339. ISSN: 14764687. DOI: 10.1038/nature24489.

He, F., T. E. Saunders, Y. Wen, D. Cheung, R. Jiao, P. R. T. Wolde, M. Howard, and J. Ma (2010). "Shaping a morphogen gradient for positional precision." In: *Biophysical Journal* 99.3, pp. 697–707. ISSN: 15420086. DOI: 10.1016/j.bpj.2010.04.073.

Hillenbrand, P., U. Gerland, and G. Tkačik (2016). "Beyond the French Flag Model: Exploiting Spatial and Gene Regulatory Interactions for Positional Information." In: *PloS one* 11.9, e0163628.

Howitt, M. R., S. Lavoie, M. Michaud, A. M. Blum, S. V. Tran, V. Weinstock, C. A. Gallini, K. Redding, R. F. Margolskee, C. Lisa, D. Artis, and W. S. Garrett (2017). "Tuft cells, taste-

chemosensory cells, orchestrate parasite type 2 immunity in the gut." In: 351.6279, pp. 1329–1333. DOI: 10.1126/science.aaf1648.Tuft.

Huppert, S. S., T. L. Jacobsen, and M. A. Muskavitch (1997). "Feedback regulation is central to Delta-Notch signalling required for Drosophila wing vein morphogenesis." In: *Development* 124.17, pp. 3283–3291. ISSN: 09501991. DOI: 10.1242/dev.124.17.3283.

Hurley, M. (1990). "Attractors in cellular automata." In: *Ergodic Theory and Dynamical Systems* 10.1, pp. 131–140. ISSN: 14694417. DOI: 10.1017/S0143385700005435.

Inomata, H. (2017). "Scaling of pattern formations and morphogen gradients." In: *Development Growth and Differentiation* 59.1, pp. 41–51. ISSN: 1440169X. DOI: 10.1111/dgd.12337.

Jaeger, J., D. Irons, and N. Monk (2008). "Regulative feedback in pattern formation: Towards a general relativistic theory of positional information." In: *Development* 135.19, pp. 3175–3183. ISSN: 09501991. DOI: 10.1242/dev.018697.

Kai, G., O. Trw, V. S. Dongen, and S. K. Parsy (2021). "Title: Cells of the human intestinal tract mapped across space and time." In: *bioRxiv* 597.September, p. 2021.04.07.438755.

Kalir, S. and U. Alon (2004). "Using a quantitative blueprint to reprogram the dynamics of the flagella gene network." In: *Cell* 117.6, pp. 713–720. ISSN: 00928674. DOI: 10.1016/j.cell.2004.05.010.

Knicker, M. S. (2021). "A Minimal Model for Tissue Homeostasis in the Small Intestine Based on Probabilistic Cellular Automata." Master Thesis. Technical University Munich.

Kolmogorov, A. N. (1963). "On tables of random numbers." In: *Sankhyā Ser.* A.25, pp. 369–376.

Kondo, S. and T. Miura (2010). "Reaction-Diffusion Model as a Framwork for Understanding Biological Pattern Formation." In: *Science (New York, N.Y.)* 329.September, pp. 1616–1620.

Koopmans, L. and H. Youk (2021). "Predictive landscapes hidden beneath biological cellular automata." In: *Journal of Biological Physics* 47.4, pp. 355–369. ISSN: 15730689. DOI: 10.1007/s10867-021-09592-7. arXiv: 2105.01999.

Kůrka, P. (1997). "Languages, equicontinuity and attractors in cellular automata." In: *Ergodic Theory and Dynamical Systems* 17.2, pp. 417–433. ISSN: 01433857. DOI: 10.1017/S014338579706985X.

Lander, A. D. (2013). "How cells know where they are." In: *Science* 339.6122, pp. 923–927. ISSN: 10959203. DOI: 10.1126/science.1224186.

Langton, C. G. (1990). "Computation at the edge of chaos: Phase transitions and emergent computation." In: *Physica D: Nonlinear Phenomena* 42.1-3, pp. 12–37. ISSN: 01672789. DOI: 10.1016/0167-2789(90)90064-V. arXiv: 9306003 [adap-org].

Lehotzky, D. and G. K. Zupanc (2019). "Cellular Automata Modeling of Stem-Cell-Driven Development of Tissue in the Nervous System." In: *Developmental Neurobiology* 79, pp. 497–517. DOI: 10.1002/dneu.22686.

Levin, M. (2021). "Bioelectric signaling: Reprogrammable circuits underlying embryogenesis, regeneration, and cancer." In: *Cell* 184.8, pp. 1971–1989. ISSN: 10974172. DOI: 10.1016/j.cell.2021.02.034.

Li, P. and M. B. Elowitz (2019). "Communication codes in developmental signaling pathways." In: *Development (Cambridge)* 146.12, pp. 1–12. ISSN: 14779129. DOI: 10.1242/dev.170977.

Li, W., N. H. Packard, and C. G. Langton (1990). "Transition phenomena in cellular automata rule space." In: *Physica D* 45.1-3, pp. 77–94. DOI: 0167-2789.

Mah, A. T., K. S. Yan, and C. J. Kuo (2016). "Wnt pathway regulation of intestinal stem cells." In: *Journal of Physiology* 594.17, pp. 4837–4847. ISSN: 14697793. DOI: 10.1113/JP271754.

Manukyan, L., S. A. Montandon, A. Fofonjka, S. Smirnov, and M. C. Milinkovitch (Apr. 2017). "A living mesoscopic cellular automaton made of skin scales." In: *Nature* 544.7649, pp. 173–179. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature22031.

Marx, V. (2021). "Method of the Year: spatially resolved transcriptomics." In: *Nature Methods* 18.1, pp. 9–14. ISSN: 15487105. DOI: 10.1038/s41592-020-01033-y.

Mayo, A. E., Y. Setty, S. Shavit, A. Zaslaver, and U. Alon (2006). "Plasticity of the cis-regulatory input function of a gene." In: *PLoS Biology* 4.4, pp. 555–561. ISSN: 15457885. DOI: 10.1371/journal.pbio.0040045.

Mayr, U., D. Serra, and P. Liberali (2019). "Exploring single cells in space and time during tissue development, homeostasis and regeneration." In: *Development (Cambridge)* 146.12. ISSN: 14779129. DOI: 10.1242/dev.176727.

Michel, M., M. Aliee, K. Rudolf, L. Bialas, F. Jülicher, and C. Dahmann (2016). "The selector gene apterous and notch are required to locally increase mechanical cell bond tension at the Drosophila dorsoventral compartment boundary." In: *PLoS ONE* 11.8, pp. 1–18. ISSN: 19326203. DOI: 10.1371/journal.pone.0161668.

Morales, J. S., J. Raspopovic, and L. Marcon (2021). "From embryos to embryoids: How external signals and self-organization drive embryonic development." In: *Stem Cell Reports* 16.5, pp. 1039–1050. ISSN: 22136711. DOI: 10.1016/j.stemcr.2021.03.026.

Nandagopal, N., L. A. Santat, L. LeBon, D. Sprinzak, M. E. Bronner, and M. B. Elowitz (2018). "Dynamic Ligand Discrimination in the Notch Signaling Pathway." In: *Cell* 172.4, 869–880.e19. ISSN: 10974172. DOI: 10.1016/j.cell.2018.01.002.

Nehaniv, C. L. (2003). "Asynchronous automata networks can emulate any synchronous automata network." In: *Journal of Algebra* December, pp. 1–21. ISSN: 0218-1967; 1793-6500/e. DOI: 10.1142/S0218196704002043.

Neumann, J. von (1966). "Theory of self-reproducing automata." In: *Theory of self-reproducing automata*. University of Illinois Press. Chap. 4, pp. 63–87.

Ninagawa, S. (2008). "Power Spectral Analysis of Elementary Cellular Automata." In: *Complex Systems* 17.4, pp. 399–411.

Nissen, S. B., M. Perera, J. M. Gonzalez, S. M. Morgani, M. H. Jensen, K. Sneppen, J. M. Brickman, and A. Trusina (2017). "Four simple rules that are sufficient to generate the mammalian blastocyst." In: *PLoS Biology* 15.7, pp. 1–30. ISSN: 15457885. DOI: 10.1371/journal.pbio.2000737.

Nitzan, M., N. Karaiskos, N. Friedman, and N. Rajewsky (2019). "Gene expression cartography." In: *Nature* February, pp. 1–6. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1773-3.

Okabe-Oho, Y., H. Murakami, S. Oho, and M. Sasai (2009). "Stable, precise, and reproducible patterning of Bicoid and hunchback molecules in the early Drosophila embryo." In: *PLoS Computational Biology* 5.8. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1000486.

Oliveira, G., P. Oliveira, and N. Omar (2000). "Guidelines for dynamics-based parameterization of one-dimensional cellular automata rule spaces." In: *Complexity* 6, pp. 63–71.

Paulus, U., M. Loeffler, J. Zeidler, G. Owen, and C. S. Potten (1993). "The differentiation and lineage development of goblet cells in the murine small intestinal crypt: Experimental

and modelling studies." In: *Journal of Cell Science* 106.2, pp. 473–483. ISSN: 00219533. DOI: 10.1242/jcs.106.2.473.

Qiu, J. M., S. A. Roberts, and C. S. Potten (1994). "Cell migration in the small and large bowel shows a strong circadian rhythm." In: *Epithelial Cell Biol.* 3.4, pp. 137–48.

Quax, R., G. Chliamovitch, A. Dupuis, J.-L. Falcone, B. Chopard, A. G. Hoekstra, and P. M. A. Sloot (Sept. 2017). "Information processing features can detect behavioral regimes of dynamical systems." In: *ArXiv e-prints*, pp. 1–31. arXiv: 1709.09447.

– (2018). "Information processing features can detect behavioral regimes of dynamical systems." In: *Complexity* 2018. ISSN: 1076-2787. DOI: https://doi.org/10.1155/2018/6047846. arXiv: 1709.09447.

Ramalho, T., S. Kremser, H. Wu, and U. Gerland (2021). "Programmable pattern formation in cellular systems with local signaling." In: *Communications Physics* 4.1, pp. 1–11. ISSN: 23993650. DOI: 10.1038/s42005-021-00639-8.

Raser, J. M. and E. O'Shea (2005). "Noise in Gene Expression." In: *Science* 309.5743, pp. 2010–2014. DOI: 10.1126/science.1105891.

Raspopovic, J., L. Marcon, L. Russo, and J. Sharpe (2014). "Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients." In: *Science* 345.6196, pp. 566–570. ISSN: 10959203. DOI: 10.1126/science.1252960.

Restrepo, S., J. J. Zartman, and K. Basler (2014). "Coordination of patterning and growth by the morphogen DPP." In: *Current Biology* 24.6, R245–R255. ISSN: 09609822. DOI: 10.1016/j.cub.2014.01.055.

Schepers, A. G., R. Vries, M. Van Den Born, M. Van De Wetering, and H. Clevers (2011). "Lgr5 intestinal stem cells have high telomerase activity and randomly segregate their chromosomes." In: *EMBO Journal* 30.6, pp. 1104–1109. ISSN: 02614189. DOI: 10.1038/emboj.2011.26.

Schier, A. F. (2001). "Axis formation and patterning in zebrafish." In: *Current Opinion in Genetics and Development* 11.4, pp. 393–404. ISSN: 0959437X. DOI: 10.1016/S0959-437X(00)00209-4.

Schweisguth, F. and F. Corson (2019). "Self-Organization in Pattern Formation." In: *Developmental Cell* 49.5, pp. 659–677. ISSN: 18781551. DOI: 10.1016/j.devcel.2019.05.019.

Shaebani, M. R., A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger (2020). "Computational models for active matter." In: *Nature Reviews Physics* 2.April, pp. 181–199. ISSN: 2522-5820. DOI: 10.1038/s42254-020-0152-1. arXiv: 1910.02528.

Sprinzak, D., A. Lakhanpal, L. LeBon, J. Garcia-Ojalvo, and M. B. Elowitz (2011). "Mutual inactivation of Notch receptors and ligands facilitates developmental patterning." In: *PLoS Computational Biology* 7.6. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1002069.

Sprinzak, D., A. Lakhanpal, L. Lebon, L. A. Santat, M. E. Fontes, G. A. Anderson, J. Garcia-Ojalvo, and M. B. Elowitz (2010a). "Cis-interactions between Notch and Delta generate mutually exclusive signalling states." In: *Nature* 465.7294, pp. 86–90. ISSN: 00280836. DOI: 10.1038/nature08959.

– (2010b). "Supplement: Cis-interactions between Notch and Delta generate mutually exclusive signalling states." In: *Nature* 465.7294, pp. 86–90. ISSN: 00280836. DOI: 10.1038/nature08959.

Stapornwongkul, K. S., M. De Gennes, L. Cocconi, G. Salbreux, and J.-P. Vincent (2020). "Patterning and growth control in vivo by an engineered GFP gradient." In: *Science* in press, pp. 321–327.

Stapornwongkul, K. S. and J. P. Vincent (2021). "Generation of extracellular morphogen gradients: the case for diffusion." In: *Nature Reviews Genetics* 22.6, pp. 393–411. ISSN: 14710064. DOI: 10.1038/s41576-021-00342-y.

Takahashi, T. and A. Shiraishi (2020). "Stem Cell Signaling Pathways in the Small Intestine." In: *International Journal of Molecular Sciences* 21, pp. 1–18. DOI: 10.3390/ijms21062032.

Toda, S., N. W. Frankel, and W. A. Lim (2019). "Engineering cell – cell communication networks : programming multicellular behaviors." In: *Current Opinion in Chemical Biology* 52, pp. 31–38. ISSN: 1367-5931. DOI: 10.1016/j.cbpa.2019.04.020.

Toda, S., W. L. Mckeithan, T. J. Hakkinen, and P. Lopez (2020). "Engineering synthetic morphogen systems that can program multicellular patterning." In: *Science* 331.October, pp. 327–331. DOI: 10.1126/science.abc0033.

Turing, A. M. (1952). "The chemical basis of morphogenesis." In: *Bulletin of Mathematical Biology* 52.1-2, pp. 153–197. ISSN: 00928240. DOI: 10.1007/BF02459572.

Ulam, S. (1952). "Random processes and transformations." In: *Proc. Int. Congr. Math.* 2, pp. 264–275.

Vispoel, M., A. J. Daly, and J. M. Baetens (2021). "Progress, gaps and obstacles in the classification of cellular automata." In: *Physica D: Nonlinear Phenomena*, p. 133074. ISSN: 01672789. DOI: 10.1016/j.physd.2021.133074.

Vitányi, P. M. (2020). "How incomputable is Kolmogorov complexity?" In: *Entropy* 22.4, pp. 1–6. ISSN: 10994300. DOI: 10.3390/E22040408. arXiv: 2002.07674.

Wolfram, S. (1984a). "Universality and Complexity in Cellular Automata, Cellular Automata and Complexity." In: *Physica D* 10D.January, pp. 1–35. ISSN: 01672789.

Wolfram, S. (1984b). "Mathematical Physics Computation Theory of Cellular Automata." In: *Commun. Math. Phys* 96, pp. 15–57.

– (2002). *A new kind of science*. Wolfram Media, Inc. ISBN: 1579550088.

– (2022). *ECA: Ensemble properties: Power spectrum*.

Wolpert, L. (1969). "Positional information and the spatial pattern of cellular differentiation." In: *Journal of Theoretical Biology* 25.1, pp. 1–47. ISSN: 10958541. DOI: 10.1016/S0022-5193(69)80016-0.

Wuensche, A. (1997). *Attractor basins of discrete networks*.

– (1998). "Classifying cellular automata automatically." In: *Complexity* 0526.September, pp. 1–26. DOI: 10.1002/(SICI)1099-0526(199901/02)4.

– (2021). *DDLab*.

Wuensche, A., M. Lesser, and M. J. Lesser (1992). *Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata*.

Yamamoto, M., Y. Saijoh, A. Perea-Gomez, W. Shawiot, R. R. Behringer, S. L. Ang, H. Hamada, and C. Meno (2004). "Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo." In: *Nature* 428.6981, pp. 387–392. ISSN: 00280836. DOI: 10.1038/nature02418.

Yaron, T., Y. Cordova, and D. Sprinzak (2014). "Juxtacrine signaling is inherently noisy." In: *Biophysical Journal* 107.10, pp. 2417–2424. ISSN: 15420086. DOI: 10.1016/j.bpj.2014.10.006.

Zagorski, M., Y. Tabata, N. Brandenberg, M. P. Lutolf, T. Bollenbach, J. Briscoe, and A. Kicheva (2017). "Decoding of position in the developing neural tube from antiparallel morphogen gradients." In: *Science* 356.June, pp. 1379–1383. DOI: 10.1126/science.aam5887.

Zenil, H. (2010). "Compression-based investigation of the dynamical properties of cellular automata and other systems." In: *Complex Systems*, p. 28. arXiv: 0910.4042.

Zenil, H., F. Soler-Toscano, J.-P. Delahaye, and N. Gauvrit (Sept. 2015). "Two-dimensional Kolmogorov complexity and an empirical validation of the Coding theorem method by compressibility." In: *PeerJ Computer Science* 1.:e23. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.23.

Zenil, H. and E. Villarreal-Zapata (2013). "Asymptotic behaviour and ratios of complexity in cellular automata." In: *International Journal of Bifurcation and Chaos* 23.9. DOI: 10.1142/S0218127413501599. arXiv: arXiv:1304.2816v4.

Zhou, W., D. Ramachandran, A. Mansouri, and M. J. Dailey (2018). "Glucose stimulates intestinal epithelial crypt proliferation by modulating cellular energy metabolism." In: *Journal of Cellular Physiology* 233.4, pp. 3465–3475. ISSN: 10974652. DOI: 10.1002/jcp.26199.

Acknowledgments

There are many people who have contributed to this thesis in various ways and the following acknowledgments are not ordered in a meaningful way.

First of all, I am very grateful to Uli Gerland for providing the conditions and the environment I could freely live out my curiosity, where ever I wanted and without pressure. His door was always open, be it for scientific discussions or organizational matters. Many thanks to Karen Alim for volunteering as the second examiner to evaluate this thesis and the examination chairperson.

I am very happy to have been part of the QBM graduate school, the financial support, and broad scientific exchange it provided, including all the talks, retreats and conferences, as well as its members.

I am deeply grateful to my colleague Stephan Kremser, for regular scientific discussions up to CA attractor data and a lot of proof reading.

Thanks to Max Knicker, our former Master student for sharing his figures for this thesis.

I am grateful to Olivier Martin for long discussion on extensions of Langton's work, as well as to Gasper Tkacik and Nicolas Gompel for discussions on the boundary formation project.

Further, many thanks to Cesar, Hamid and Bernhard for scientific discussions on the boundary formation and the classification problem, and the rest of the past and present Gerland group.

There is a too long list of thinks I want thank Leo for, including: Listening to scientific issues and details I needed to reflect on; For his ideas on the 'particle hole symmetry' intuition; Supporting me to speed up my code, and helping me with any IT issues I had on the way; For critically reading through this thesis.

I am deeply grateful to my parents and my sister, Sile, for their continuous encouragement, trust and support.