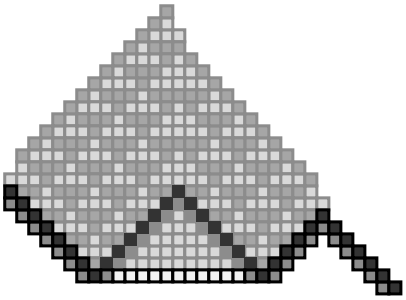
Markus Redeker
Faculty of Environment and Technology

Flexible Time and Ether
in One-dimensional Cellular Automata

Thesis submitted in partial fulfilment of the requirements
of the University of the West of England, Bristol
for the degree of Doctor of Philosophy



December 2013

Abstract

A one-dimensional cellular automaton is an infinite row of identical machines—the cells—which depend for their behaviour only on the states of their direct neighbours.

This thesis introduces a new way to think about one-dimensional cellular automata. The formalism of Flexible Time allows one to unify the states of of a finite number of cells into a single object, even if they occur at different times. This gives greater flexibility to handle the structures that occur in the development of a cellular automaton. Flexible Time makes it possible to calculate in an algebraic way the fate of a finite number of cells.

In the first part of this thesis the formalism is developed in detail. Then it is applied to a specific problem of one-dimensional cellular automata, namely ether formation. The so-called ether is a periodic pattern of cells that occurs in some cellular automata: It arises from almost all randomly chosen initial configurations, and why this happens is not clear. For one of these cellular automata, the elementary cellular automaton with rule code 54, ether formation is expressed in the formalism of Flexible Time.

Then a partial result about ether formation is proved: There is a certain fragment of the ether that arises with probability 1 from every random initial configuration, and it is then propagated with probability 1 to any later time. The persistence of the ether fragment is a strong argument that the ether under Rule 54 indeed arises from almost all input configurations. The result only requires that the states of the cells are chosen independently and with equal probability distributions, and that all cell states can occur. This is not yet a full proof of ether formation, but it is derived by formal means, not just by computer simulations.

Contents

List of Symbols

Sets and Functions

Sequences

Space-time

Processes

List of Figures

List of Tables

Chapter 1

Introduction

Cellular automata came into being as an almost brutally simplified model of information processing in a physical medium.

John von Neumann invented cellular automata (together with Stanislaw Ulam) as part of his work on self-reproducing systems [31, p. 42]. He needed a simplified physical universe in which he could construct a model which captures the essential properties of self-reproduction in a biological organism. This model universe had to be simple enough that a single person could reason about it while using only paper and pencil and no mechanical aid. It also had to consist of simple components in order to make sure that self-reproduction was a property of the simulated organism and not already built into the physics of its universe. These requirements lead to several simplifications. The first one is that only discrete parameters could be used, especially no real numbers. Time in a cellular automaton therefore runs in discrete steps, like the ticks of a clock. Space is reduced to a rectangular grid. It consists of the points of the n-dimensional grid $\mathbb{Z}^n$: we speak then of an n-dimensional cellular automaton. The second simplification concerns the interior of the universe. It must be possible to describe the self-reproducing organism completely with a finite number of symbols. The world simulated by the automaton consists of objects at the lattice points, which are called cells. A cell can be in one of several states, and there could be only finitely many of them if it was possible to write down a configuration of the automaton. The cells are thought as small information-processing machines, representing atoms, electrical components or possibly biological cells in a tissue. For the purpose of von Neumann, the simulated universe of the cellular automaton would contain only a finite number of cells that simulated the self-reproducing system. All other cells were in a special, quiet state, which stood for a kind of vacuum that remained unchanged. Activity was always caused by cells in other states.

The physics of this model universe is one of local interaction. The purpose of the model universes of cellular automata is the simulation of an object of moderate size. Therefore the world view of cellular automata is based on Newtonian physics. It especially ignores General Relativity. The physical laws

in a cellular automaton are then the same at every point and for every time step. In order to make it possible that they can be described completely, they also have to allow a finite description. Therefore the state of a cell in the next time step cannot depend on the states of all the cells in the automaton. This leads to the idea that the state of a cell in the next time step should only depend on the states of the cells in its direct neighbourhood. Such a neighbourhood contains only the cells at an Euclidean distance less than or equal to a given constant r. (Note that with this definition a cell is always part of its neighbourhood.) The number of cells in a neighbourhood is then always finite. The number of combinations of states that these cells can have together is also finite: therefore the behaviour of each cell can be described by a finite table. It maps each state of the neighbourhood of a cell to the state of the cell in the next time step. The neighbourhoods of all points in $\mathbb{Z}^n$, and therefore those of the cells in the cellular automaton, look the same: therefore it is possible to specify the behaviour of the whole system of cells with a single finite rule. The number r is called the radius of the cellular automaton. A beneficial side effect is that no point of $\mathbb{Z}^n$ is special, as it is in Newtonian physics. Another side effect is that signal transmission in the universe of cellular automata always has a finite speed. This is what John von Neumann did, and he started to develop a self-reproducing system in a specific two-dimensional cellular automaton.

A second important step in the history of cellular automata was the invention of the "Game of Life" by John H. Conway in 1970 [31, p. 49–53]. It is a two-dimensional cellular automaton with two states and an especially simple rule and became soon very popular. This was the time when computers with a graphics display started to become accessible to many people. A development that is important in the context of the present work is that some of them did run "Life" systematically with initial configurations that were chosen at random. The usual method is to chose a probability p and then let the computer initialise independently every cell with probability p in state 1 and with probability $1 - p$ in state 0. State 0 is the quiet state in Life. Then after several time steps of evolution* stable patterns often emerge. Some of them stay unchanged, others oscillate with a period of 2, seldom more, time steps, and a few move through the two-dimensional cellular space.

This set a precedent, and random initial configurations have become a standard tool that is used when one wants to get an overview of the behaviour of an unknown cellular automaton. So when in 1983 Stephen Wolfram [59] wanted to survey the possible behaviours of cellular automata, he too tested them on random initial configurations. The plan to survey the possible behaviours of cellular automata had a side effect that influences cellular automata theory until now. Wolfram worked with one-dimensional cellular automata, and in order to have a subset of manageable size, he chose the set of automata with two states and radius 1, the simplest class at all from which one can expect nontrivial behaviour. Wolfram called them "elementary cellular automata".

*The word "evolution" means different things in different contexts. Here I use it in the wider sense of "development over time", not in the narrow sense of Darwinian evolution.

There is also research on cellular automata where the subject is the Darwinian evolution of transition rules for a specific purpose. (See e. g. Mitchell, Crutchfield and Hraber [46] or the review by Mitchell [45].) Nevertheless the use of "evolution" in the wider sense is also established in cellular automata theory. The word "evolution" has more specific associations than e. g. "development", therefore I use it here.

There are 256 of them, but if one views automata as equivalent if they differ only by an interchange of the states 0 and 1 or of left and right, only 88 types of behaviour remain. (See Li and Packard [33, p. 284].) There has been earlier research on elementary cellular automata, but Wolfram gave this class a name and introduced a system of code numbers with which one can refer to their rules, and they have stayed in the centre of research since then.

The choice of one-dimensional automata is also in another sense advantageous, since their behaviour can be easily displayed in a two-dimensional diagram with one space and one time dimension. This makes communication about their behaviour much more direct than that about two-dimensional automata.

In contrast to John von Neumann's rule and "Life", the set of elementary cellular automata contains rules in which there is no state that can be considered quiet. With some of the rules, the cellular automata stay chaotic when started from a random initial configuration. But with others, a similar phenomenon as with "Life" occurs. After some time, particles appear and move or stay on a simple background, but here the background does not consist of a region of cells all in the same state. Instead, the regions between the particles consist of a spatially periodic pattern. After several steps of evolution of the cellular automaton, the same pattern occurs again. When the evolution of the cellular automaton is displayed as a space-time diagram, the background looks like a wallpaper pattern. Nowadays, such a periodic background pattern that arises from almost every random initial configuration is often called an "ether", in analogy to the ether concept of pre-relativistic physics. There, as in cellular automata, the ether is a background in which particles and signals move.

Examples. While this thesis focuses on ether formation under Rule 54, we will now take a larger perspective and look for examples of ether formation among the elementary cellular automata in general.

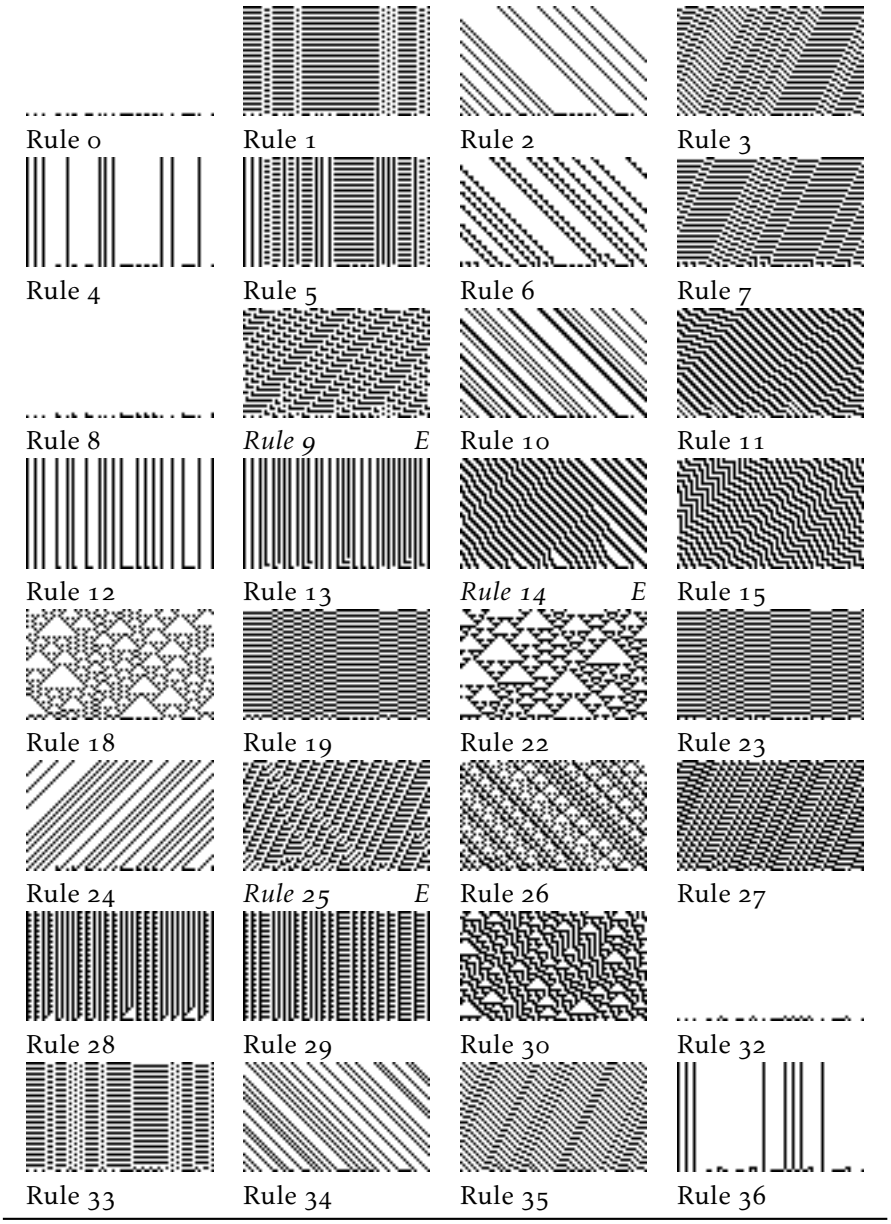
For this we need a criterion that tells us whether a cellular automaton has an ether. To my knowledge there is however no general definition in the literature under which conditions a cellular automaton has an ether. As a working condition for the following small survey, we will use the following definition, based on Rule 54:

> If the evolution diagram of a typical random initial configuration contains regions that are periodic in space and time, and these regions grow over time, then this cellular automaton has an ether.[†]

With it we can check empirically for all elementary cellular automata whether they have an ether. This is done in Table 1.1–1.3. The diagrams in this figure show the evolution of a random initial configuration under all equivalence classes of elementary cellular automata. All evolution diagrams use the same initial configuration, one in which each cell is with probability $\frac{1}{2}$ in state 0 and with probability $\frac{1}{2}$ in state 1. For this probability distribution the behaviour of the cellular automaton is unchanged when state 0 is exchanged with 1 in its

[†]Thus an ether may be an uniform pattern consisting of a single cell, like the quiet state under some rules. I do not exclude it because one of the properties of the ether is that it is the background on which particles move. In this aspect, an uniform ether is not different from other ethers. Not excluding the uniform case also makes the definition simpler.

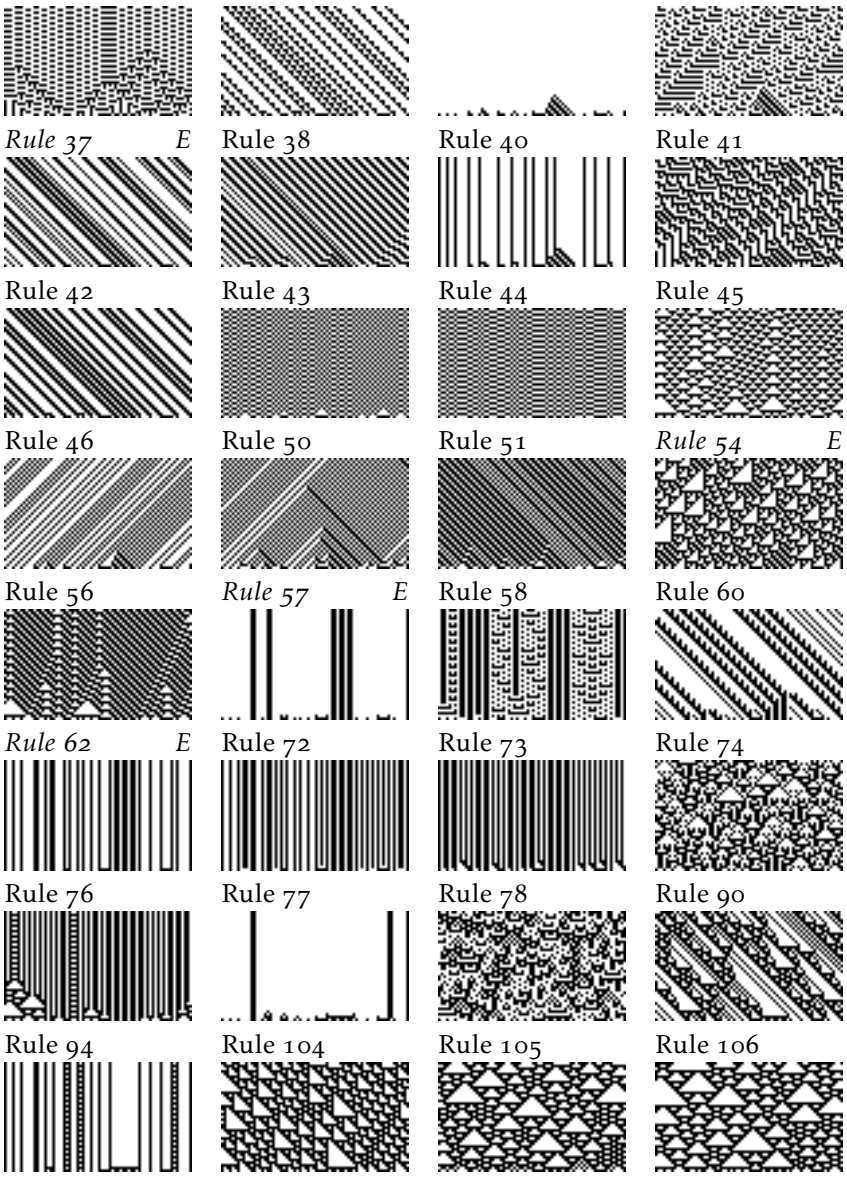
Table 1.1: A survey of elementary cellular automata (Part 1).



transition rule, or left with right. We will therefore show only the evolution of one of these related rules, namely that with the lowest code number.

For some of the rules, e. g. Rule 110, a larger evolution diagram than that one shown here is needed to clarify whether they have an ether. But if we do this, we see that only the cellular automata with the code numbers 9, 14, 25, 37, 54, 57, 62, 110, 142 and 184 (and those equivalent to them) support an ether. Their evolution diagrams are marked in Table 1.1–1.3 with an *E*.

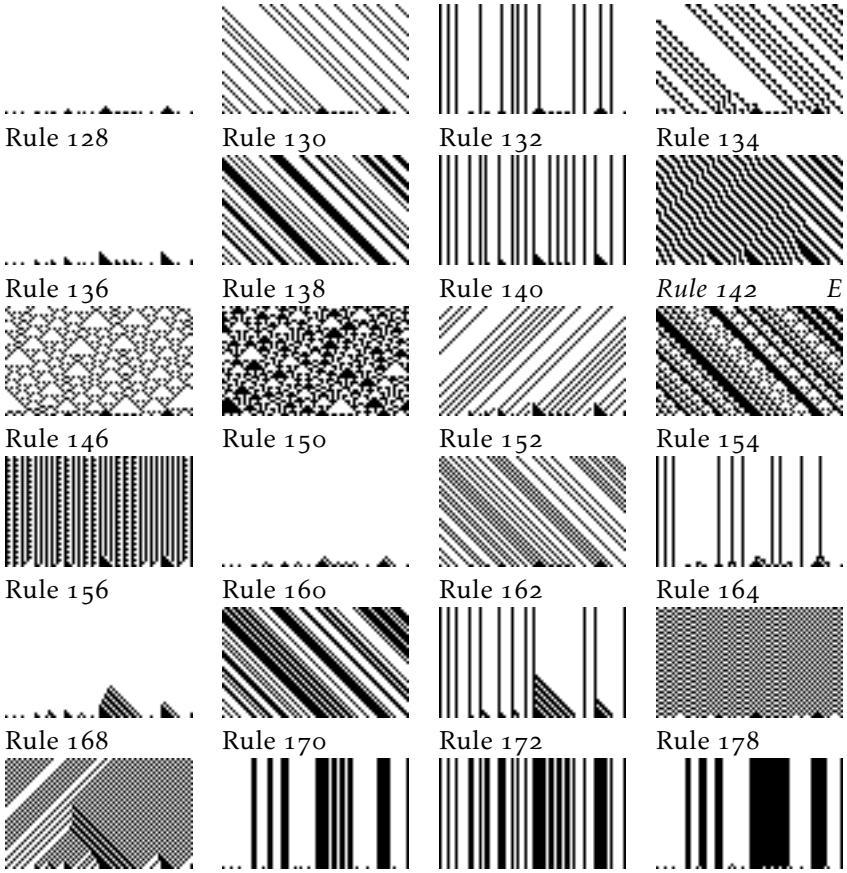
Table 1.2: A survey of elementary cellular automata (Part 2).



The Problem of Ether Formation. The formation of an ether in some cellular automata has been mentioned by some authors as an open question, but to my knowledge no one has published a solution. Bruno Martin [34, p. 3] writes,

> "By observing space-time diagrams of the rule 54 on a random configuration, we always see a kind of background with space and time period 4 [...]. The background apparition is usually very fast (less than ten iterations are enough) and still mysterious, we have no explanation of this phenomenon."

Table 1.3: A survey of elementary cellular automata (Part 3).



As a repeating pattern, the ether is the simplest structure in a cellular automaton that can grow to an unlimited size. The problem of ether formation is therefore the simplest question about an important form of self-organisation in cellular automata, namely the autonomous emergence of large structures.

Self-organisation and the construction of complex patterns and machines in a cellular automaton have in common that they involve the synchronised behaviour of many cells. To understand them one first needs a language in which one can express the behaviour of a large number of interacting cells over time. In an earlier publication [51] I have called this language "Flexible Time".

1.1 Flexible Time

Motivation. A basic idea behind the formalism of Flexible Time is that it generalises the way in which a finite part of a configuration of a cellular automaton is written. For a one-dimensional cellular automaton, which is an infinite sequence of cells, a configuration is simply an infinite sequence of cell states. A single cell state is usually written as a symbol, often a number,

and then one naturally writes a finite region of the cellular automaton as a string of symbols. Then a sequence like 011101110111 is a possible content of a region in a cellular automaton for which the state set contains the values 0 and 1. Mathematics has already developed notations and theories to work with such strings of symbols efficiently. One simple notational device is the use of exponents to express repetition, which allows us to express the previous string as $(0111)^3$. There is also a whole theory of formal languages to handle such strings. I want to be able to use such methods.

Another ingredient is selective interest. A natural way to understand a complex system is to decompose it into subsystems and first try to understand them first. However, a cellular automaton is a model of a physical (or biological or information processing, *&ct.*) system. In such a system there are lots of processes that start and end independently of each other—lots of organisms that are born and die at any moment in time, or lots of tasks that are completed independently. There is no global synchronisation. In a world in which information travels with finite speed, the starting time of a process is only influenced by the processes in its direct neighbourhood. And if it consists itself of subprocesses, which are also loosely coupled, then they may finish at different times, and it makes no sense to speak of the "end time" of the main process of which they are parts. A notation for cellular automata that allows one to focus on the behaviour of arbitrary subprocess has to take that into account.

The Formalism of Flexible Time. My idea to solve to the questions implied here is influenced by the concepts of Relativity theory. We will now give up the thought that there must be a globally determined time. Instead, when a complex process consists of subprocesses that end at different times in different places, then the end times of the subprocesses together form the end time of the process. They are, all together, viewed as a single moment in time. With this concept of time it is no longer necessary to work with configurations of infinite size. We will instead specify the content of a finite region of the cellular automaton, namely that where a specific process starts, and are then able to compute the generalised end time of this process and the states of the cells at this time. There is a mathematical object that specifies the location of some cells, both in space and in time, together with their states. I call it a *situation*, and it is a generalisation of the finite sequence of cell states that I mentioned before. Situations have in common with cell state sequences that they are formally strings of symbols; therefore the familiar concepts for words in a formal language can be applied to them.

There is another kind of mathematical object, the *reaction*, which we will use to specify such a process. It is a pair of situations, one for the beginning and one for the end of the process. The set of all reactions for a cellular automaton provides the same information as the transition rule.

Reactions and situations are thus an alternative means to reason about the behaviour of a cellular automaton. Together they form the formalism of Flexible Time. We have thus the following (approximate) equivalences:

Flexible Time	Global Time
situation	configuration
reaction	step in evolution

1.2 Aims and Methods

The intent of this thesis is to construct a general framework in which one can solve questions about collective behaviour in one-dimensional cellular automata. It should be a language that is adaptable to a wide range of questions. This way I want to provide a step forward towards the solution of the last one among Stephen Wolfram's "Twenty Problems in the Theory of Cellular Automata" [61],

> What higher-level descriptions of information processing in cellular automata can be given?

My approach to address Wolfram's question is to invent a language in which one can describe more easily the components of a cellular automaton's information processing system. For a concrete problem this allows to create a vocabulary of space-time patterns and their interactions. With this language, increasingly larger structures could be described and understood, until one would understand the behaviour of a large and complex information-processing system.

In its current state the language is not so powerful. It does however allow to express with situations the patterns that one can see in space-time diagrams more or less directly, and then to express and prove general theorems about them.

The development of such a tool is easier with a concrete application in mind. Therefore, a second aim of this work is to find an explanation why there is an ether in the elementary cellular automaton with rule code 54.

I have chosen ether formation because it is a case of self-organisation and therefore interesting in its own right. Furthermore it is the simplest case of self-organisation in cellular automata that I am aware of. Rule 54 has a relatively simple ether that arises early in the evolution of a random initial configuration. Nevertheless this cellular automaton is far from trivial: With Rule 54 one can e. g. compute arbitrary Boolean functions [36]. This makes it more probable that the results about ether formation in Rule 54 and the methods to derive them carry over to other interesting cellular automata.

The thesis is a kind of sequel to my article [51]: There, a description of Flexible Time in the context of a specific cellular automaton was given, but its general theory was missing. Here I provide a theoretical justification for the formalism, together with a study of ether formation for a specific cellular automaton. A more general theory of structure formation in one-dimensional cellular automata is something I would like to be able to do at a later time.

Requirements on the Theory. At a very basic level it has always been difficult to express the phenomena in a cellular automaton in an understandable way. This is true especially if emphasis is placed on concrete interactions, like the collision of two particles. All authors use pictures in some way. This becomes

however difficult once larger structures are involved and the details of such diagrams would become smaller and smaller.

The aim of this work is therefore to construct a formalism that expresses the behaviour of a large mass of cells in a cellular automaton. We need to express the evolution of a cellular automaton in a way that corresponds to the structures that one can see in the two-dimensional space-time diagrams, and which is at the same way able to represent arbitrarily large structures. My goal was to find a language in which one can describe the behaviour of cellular automata in an algebraic notation—a kind of cellular automata evolution program that is run by the human brain. The formalism is therefore completely algebraic. No pictures are necessary to specify details. They are however still useful for clarification and to get ideas.

It is an important requirement that the formalism is not bound to a specific moment in time. A pattern that we see in a space-time diagram usually extends over a longer period of time. We see it as a two-dimensional form, and pieces that appear to us as connected may belong to different times. A formalism that restricts us to snapshots of the cellular automaton at specific moments in time can not display this. I have therefore developed a formalism that allows us to jump forward and backward in time.

Two Kinds of Mathematics. This work is also an attempt to support a return, after years in which computer experiments dominated, to the idea that the behaviour of cells in a cellular automaton should be something that can be comprehended with the help of pencil and paper alone. I will call these methods here "traditional mathematics".

The methods of computer experiments and of traditional mathematics have different aims and result in different kinds of understanding. The result of a computer experiment is knowledge about a single case. A result of traditional mathematics is a theorem about an infinity of cases. One may say that, since the majority of the questions asked in scientific research are about "all cases" of a certain kind, traditional mathematics is the only way to answer them. However, the requirement of traditional mathematics always to work with an infinity of cases at once is a severe restriction. In contrast, computer experiments can be done even in cases where there is not enough understanding of the question to apply the methods of traditional mathematics.

The restricted nature of traditional mathematics also holds a great promise. If it works, traditional mathematics has results that automatically apply to a great range of cases. This is because the very restrictedness of its methods makes sure that its results have only a small number of preconditions. This automatism suggests a method that combines some of the advantages of both approaches: If one takes a phenomenon that has been found empirically in a small number of cases and finds a proof for it, then it will automatically tell us something about an infinity of other cases. This is what I do here with the ether in Rule 54.

Previous Work. This work is an extension of the ideas presented in [51], where the formalism of Flexible Time was introduced for the case of Rule 110. An application of the formalism to Rule 54 was presented in [52], where it was

also shown how the ether and particles were represented with situations and reactions.

Thanks. I want to thank Andrew Adamatzky, Genaro J. Martínez and Lars Immisch for their help in getting me to Bristol.

I also want to thank my thesis supervisors Andrew Adamatzky, Rob Laister and Tony Solomonides for general support and for comments on earlier stages of this work. The comments on this thesis during the various examinations also proved extremely helpful.

1.3 Notation

Since the purpose of this thesis is to introduce a new mathematical language, I have to introduce many new words and notations. When reading the thesis, it may be difficult to keep an overview of all these concepts. All newly introduced words are therefore listed in the index at the end, and on page v there is a list of all symbolic notations used in the text, together with short explanations.

The new concepts themselves are introduced and explained step by step in the following chapters. But first we have to clarify the notations for some basic and well-known concepts which are written differently by different authors.

Sets and Functions. The set of positive integers is $\mathbb{N} = \{1, 2, 3, \dots\}$, and the set of non-negative integers is $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

If A and B are two sets, then B^A is the set of functions from A to B. Therefore we can say of a function $f : A \to B$ that $f \in B^A$. The set A is then called the *domain* of f, and we will write it dom f. Thus for the specific function f just mentioned we have dom $f = A$.

There are two conventions in use for the symbol of set inclusion, $\subset$. They differ in the case where the two sets to be compared are equal. Here we will use the convention that $A \subset B$ means that A is a proper subset of B. If we want to include the case that $A = B$, we write $A \subseteq B$.

A useful property of functions between sets is monotonity. Let F be a function that maps subsets of A to subsets of B. Then we will say that F is *monotone* if $F(a) \subseteq F(a')$ whenever $a \subseteq a'$. Similarly, a property P of subsets of a set A is monotone if, when $a \subseteq A$ has property P and a' is a set with $a \subseteq a' \subseteq A$, then a' has property P.

Let now F be a monotone function F that maps subsets of A to subsets of A. If it also has the property that $a \subseteq F(a)$ and $F(F(a)) = F(a)$ for all $a \subseteq A$, then F is called a *closure operator*. (The last two concepts are taken from order theory [11, p. 145].)

Sequences. We will work very often with finite sequences of arbitrary objects.

Let A be a set. An *A-sequence* of length ℓ is then an ℓ-tuple of elements of A. As it is usual in formal language theory, we may write a sequence as a formal product of its elements. So if $a = (\alpha_1, \dots, \alpha_\ell)$ is an element of A^ℓ, then it can also be written as $\alpha_1 \dots \alpha_\ell$. This automatically leads to the notion of a product of A-sequences, defined by concatenation. If $b = \beta_1 \dots \beta_m \in A^m$ is

another A-sequence, then their product is

$$ab = \alpha_1 \dots \alpha_\ell \beta_1 \dots \beta_m \in A^{\ell+m}. \tag{1.1}$$

It is easier to work with sequences if one does not always have to refer to its elements. We therefore introduce now a small arithmetic for sequences, beginning with the product just defined. It introduces a semigroup structure in the set of all sequences, therefore it is natural to introduce an *empty sequence*. It is written λ and will be used a lot, albeit mostly under another name. Then, since it is useful to have a product of an A-sequence and an element of A, we identify A with A^1, the set of 1-tuples. There is also A^0, the set that only contains λ.

With these notations we can introduce a name for the set of all A-sequences. It is called the *Kleene closure* [26, p. 29] of A and has the algebraic structure of a monoid,

$$A^* = \bigcup_{\ell \geq 0} A^\ell. \tag{1.2}$$

We will also use other notations and notions that are related to products, without making much fuss about it. One example for this is the use of exponents, another the concept of the *decomposition* of a sequence: If $a \in A^*$ and there are $b, c \in A^*$ with $a = bc$, then we will speak of this equation as the decomposition of a into b and c. We will use this as a way to introduce the variables b and c without explicitly mentioning that they are elements of A^*. Note also that if e. g. $a = bc$ and of the two factors of a only b is known, then this already determines c. We will use this as a way to introduce c.

A concept that we have already used implicitly is the *length* of a sequence. We will now introduce a notation for it: if $a \in A^\ell$, then its length is $|a| = \ell$. We will often use the fact that the length of a product ab is $|ab| = |a| + |b|$.

Sequences and Functions. The identification of A-sequences with tuples makes another simplification possible. A function $f\colon A^n \to B$ can be viewed as taking n parameters from the set A and mapping them to an element of the set B. In this case we will encounter expressions of the form $f(\alpha_1, \dots, \alpha_n)$, where $\alpha_1, \dots, \alpha_n$ are elements of A. But with the definitions above, f is also a function that maps A-sequences of length n to B. Then we can use expressions of the form $f(a)$ instead, with an $a \in A^n$, for example with $a = \alpha_1 \dots \alpha_n$.

This simplification becomes especially useful if f is the transition function of a one-dimensional cellular automaton. It is especially convenient if a is a product of sequences, say $a = bc$ with $b \in A^\ell$ and $c \in A^{n-\ell}$. We can then write terms like $f(bc)$ instead of much more voluminous expressions like $f(\beta_1, \dots, \beta_\ell, \gamma_1, \dots, \gamma_{n-\ell})$.

Chapter 2

Background

In this chapter I describe in greater detail how this thesis relates to cellular automata research in general. I also describe the relation of this work to other kinds of research that inspired it and how they influenced it.

2.1 Structures in Cellular Automata

There are several approaches in use by with which researchers try to get an understanding of the space-time structures that occur in the evolution of one-dimensional cellular automata. To give an overview I will now describe some of these works.

Besides the projects that are directly concerned with pattern formation I will also describe research that has the description of patterns as its main theme.

Turing's Work. The ancestor of all mathematical research about pattern formation is certainly Alan Turing's paper on the chemical basis of morphogenesis [57]. From the viewpoint of cellular automata, there are may similarities: Turing worked with a ring of cells that have only knowledge of their direct neighbours, he stressed the necessity of a randomised initial configuration, and he found that his setup created periodic patterns. He even did a computational (but not computer) experiment. On the other hand, his time parameter was continuous, and the state of his cells was characterised by two or three continuous parameter. The view that cellular automata are a good model to study pattern formation still had to wait for some time.

Triangles. For any research on pattern formation it is useful to find a kind of structure that occurs in many cellular automata. This enhances the probability that the results of the research are applicable to many kinds of automata. Different approaches on pattern formation can therefore be classified by the kinds of patterns on which they concentrate.

Triangular structures appear in many one-dimensional cellular automata when they are run from random initial configurations. It is therefore natural to use them as the building block for the description of more complex structures.

There is one such approach that uses triangles as building blocks for larger structures [37, 43]. It currently concentrates on Rule 110. This is one of the cellular automata that have been studied in great detail. It has a very complex behaviour and became even more interesting after Matthew Cook had proved its support for universal computation [6]. The triangles in Rule 110 are the building blocks of larger structures. They are therefore represented by "tiles", which are subsets of the two-dimensional plane. The development of the cellular automaton can therefore be understood by a covering of the two-dimensional plane without a gap. From the work with these tiles one can therefore derive the possible periodic patterns in a cellular automaton, especially candidates for the ether and for particles [40].

Tilings. A tiling approach somewhat similar to this is used by Ollinger and Richard [48, 53] to express the interactions of particles under Rule 110. It uses this approach to express the behaviour of the cellular automaton in terms of particles and collisions. There are "tiles" which represent pieces of the ether, others which represent the movement of a particle over a finite amount of time, and others that represent the collision of two or more particles. A tiling of the two-dimensional plane that corresponds to the space-time diagram of a cellular automaton is then represented in an abstract form by "a planar map whose vertices are labeled by collisions and edges by particles" [53, Definition 1.3]. These graphs are then used to represent complex interactions between particles, especially by Richard [53] to understand Rule 110. The method is however applicable to cellular automata in general.

Replicating Patterns. Another specialised approach to express the large-scale structure for a specific class of cellular automata concerns those rules which support *replication*. This class is a subclass of those rules that have a quiet state. In them one can look at localised patterns that consist of a finite number of cells in non-quiet states, while all the other cells are quiet. Replication then occurs in rules under which a small localised pattern in an initial configuration later reappears as several copies. These too then replicate, and the evolution of such a pattern in a one-dimensional cellular automaton generates a fractal-like structure, a generalisation of the Sierpiǹski triangle. Gravner and Griffeath [21] give a formal definition for replication in one-dimensional automata and then search among other things for replicating patterns under Rule 22. In another paper, by Gravner, Gliner and Pelfrey [20], several transition rules are investigated for their replicating patterns.

Domains and Defects. In several articles the configuration of the cellular automaton is decomposed into regions with a regular structure and defects between them [5, 27, 29, 30]. When the initial configuration is chosen at random, the defects usually take a random walk. From time to time two of them collide and annihilate each other, which enlarges the regular regions. This way the state of the cellular automaton becomes more ordered over time, a phenomenon that has some similarities with ether formation.

There is a more theoretical view of these phenomena, in a paper by Eloranta [15], which yields rigorous results in a simpler case. In it, the set of states of the cellular automaton is divided into two subsets, S and T, such that the next state of a cell the neighbours of which are all elements of S is another element of S, and the same is true for T. The author then investigates the behaviour of the boundary between a region of S cells and a region of T cells in which the cell states were chosen at random. He finds that the boundary moves either deterministically with maximal speed or it is a random walk, and that it is possible to give explicit, albeit complicated expressions for the speed of the walk.

A similar pattern of self-organisation occurs in cellular automata with an ether and particles, as in Rule 54 [4] and Rule 110 [32]. In these automata an ether forms that is disrupted by particles; the particles move and collide and sometimes destroy each other. While the transition rule of these automata is deterministic, the number of particles behaves nevertheless in these automata as if the collisions and decays occurred at random. In both papers a power law is found by the computer simulations. In the paper by Li and Nordahl [32] it concerns the dependence of the density of particles over time, while Boccara, Nasser and Roger [4] measure the density of a specific particle.

How does one define particles and background? Mostly it was obvious to the researchers, but there are systematic approaches. The method of "computational mechanics" by Crutchfield and Hanson [7, 8, 23] is a systematic approach that allows, among other things, to divide the configuration of the cellular automaton into regular *domains* and the *domain walls* between them. A domain wall may move, therefore particles count as domain walls. A domain is in the simplest case a spatially periodic pattern that is preserved by the transition rule, so the ether counts as a domain. There also exist more complex domains, and the authors have found a way to identify them mechanically by a program. Then it is possible to create another finite automaton that classifies the cells as belonging either to a domain or one of the domain wall. This allows to show simplified pictures of the often very complicated evolution diagrams.

Further research in this direction has been done by Marcus Pivato [49, 50]. Here, too, the aim is to divide the cellular evolution into different regions with different behaviour, again in the form of patterns and defects, but with finer subdivisions.

Grouping and Supercells. Another method to describe large-scale structures simply ignores the structures that arise in the evolution of the cellular automaton. It uses "grouping" operations for the classification of cellular automata [13, 14, 34, 42, 47]. In it the cells of the automaton are arranged in blocks of n cells, and one then considers the cellular automaton that consists of these "supercells". One also considers transition rules that aggregate several's time steps into one. This way one can establish equivalences between automata and introduce a partial order between them in terms of the complicatedness of their behaviour. Among these works the most elaborate is the work of Delorme, Mazoyer, Ollinger and Theyssier [13, 14]. In it the authors give a formal definition for the generalised grouping operations and then prove theorems about them in an abstract way. They also define three concrete grouping operations, find some equivalence classes of one-dimensional automata under

these operations and prove how they are related in the partial order defined by the grouping operation.

Global Behaviour. All this work with local structures in cellular evolution has also as its goal the understanding of cellular automata and to classify them by their behaviour.

The first approach of this kind that found greater resonance was Wolfram's [60] classification. It divides the cellular automata into four classes according to the behaviour they show when starting from a random initial configuration—in other words, by their ability for self-organisation. However, this classification scheme is not decidable, as Culik and Yu [9] showed.

Another point is that only four classes provide only a very small amount of information about the cellular automata—especially because the automata with nontrivial behaviour end up in only two of them. For this and other reasons, the business of finding classification schemes for cellular automata is still going on actively. A recent survey [35] lists 18 different classification schemes, just for the elementary cellular automata.

2.2 Physics as Metaphor and Model

As we have seen in the introduction, a cellular automaton can be understood as the simulation of a physical system. The nature of this system is however the subject of some confusion: Is it Newtonian or is it relativistic—and what role does such an old-fashioned concept as the ether play?

Newtonian and Relativistic Physics. Since a cellular automaton is only a rough approximation to a physical system, we have a certain amount of freedom in our interpretation. We can choose what kind of physics our cellular automaton should resemble. The formalism of Flexible Time is an attempt to bring a relativistic interpretation into the cellular automata, which have before mostly interpreted in a Newtonian fashion.

A sign of the Newtonian viewpoint is the existence of an universal clock. In the usual formulation, a one-dimensional cellular automaton consists of an infinite line of cells, and they evolve in discrete time steps. Time passes therefore at every point in the same way.

The central point of Relativity, on the other hand, is the finite maximal speed with which signals can propagate. In a cellular automaton we also have a finite maximal speed: It is given by the radius of the transition rule. If the transition rule has radius r, then the state of a cell can influence in the next time step only the cells at most r places to the left or the right. The analogy has been known for a long time: In the context of the Game of Life, this maximal speed has already been called by J. H. Conway the "speed of light" [17, p. 217]. We can use the analogy to let the cellular automaton play the role of the universe of Special Relativity.

We can take this analogy a step further. As in Relativity, when there is no global concept of time, causality becomes important. For cellular automata, causality can become the question, "If I change the state of one cell in the initial configuration, which cells change their state in later time steps?" This has been asked e. g. by Wolfram [61, p. 171]. In this thesis, the dual question

becomes important, "If the states of only a finite number of cells are known in a cellular automaton, the states of which other cells can be determined from this knowledge?" This question will lead to the concept of the closure in Definition 3.10.

We can maintain the standpoint that the set of all cells at a given time is not such an important concept. After all, each cell knows only about a finite number of its direct neighbours. The concept of a configuration, consisting of all cells at a time step, is therefore nothing which one is forced to use. We have, as in Relativity, a freedom to choose which events we consider as occurring at the same time. In Relativity, they form a "space-like" set. In Flexible Time, we will speak of achronal situations. We only have the requirement that the events that can influence each other causally cannot be part of the same time slice.* Then we have the flexibility that allows us to follow more easily the structures that occur in the evolution of a cellular automaton.

Space-time. Another important concept that became popular through Relativity is that of space-time. We will used heavily the freedom that it provides. The space-time viewpoint for cellular automata is actually quite old. An early example occurs in Konrad Zuse's article about Calculating Space [62]. Here, in Figures 9 and 10, the author uses a mode of display in which events from different times are displayed together. This way the movement of a particle can be shown, even though it extends over several time steps. This is however an informal use of a flexible time; I have not seen diagrams of the same style elsewhere.

There is however an example where events from different times occur naturally during a computation of a cellular automaton. William Gosper [18] uses such a scheme to compute the evolution under the Game of Life (or another two-dimensional cellular automaton) in a faster way. One could view his scheme as a form of Flexible Time in two dimensions—albeit one in which all situations are based on squares with an edge length that is a power of 2. This work was an important inspiration for me.

The Ether and Other Muddled Metaphors. There still remains the question which role the old-fashioned concept of the ether plays in such highly modern physics.

A part of the answer is that the word is already in use: The name "ether" has apparently been introduced by Matthew Cook [6] for the regular pattern in Rule 110, and it has been used by other authors too.

We can however take the concept of the ether a bit more seriously, as the physicists of the 19th century did. For them, the ether formed a background on which signals travel. The ether was however specifically invented to support the transmission of waves, for which there is no analogue in the context of cellular automata. We do have particles that move in the ether, but there is no ether at the place where the particle is located. To have a true ether, we would need an analogy to ether vibrations as they were thought to occur in the physical ether. To my knowledge, nobody has attempted such an analogy. I

*This requirement is broken a little bit in achronal situations, but it is correct in the large scale.

therefore believe that we should take the analogy to the ether—in contrast to that with Relativity—not too seriously.

If one keeps this in mind, even an inexact metaphor can serve as support for the intuition and help to find names for the phenomena that occur in cellular automata. In case of the ether, I have done so in Chapter 8, where I speak of "pure" and "disturbed" ether (the disturbances being the particles), thus using exactly that analogy I just have rejected as being not exact.

Another incongruent use of physical metaphors is the use of the word "particle". It is nowadays a common word for a localised structure in a cellular automaton that moves with a constant speed. The name is especially used for a localised structure that stays in its place, like the static structures that occur under Rule 54 [4]. One of the earliest uses must be again Zuse [62], who explicitly set out to simulate physical particles with cellular automata.

The particle metaphor is nowadays used by many authors (and also in this work), but it is not a faithful image of, say, elementary particles, or Newtonian idealised point particles. Among the features that are generally missing are an analogy to mass or impulse, or to any kind of conservation theorems. What remains is a kind of "topological" image of physical particles, in which the particles move in straight lines and interact only when they collide, but there are no general laws about that what is the result of the collision. Once again this is a metaphor that should not be followed too far.

2.3 Relation to Logic and Language

As the subject of this thesis is the construction of a language for easier mathematical reasoning, I have to name other projects that are related to mathematical languages and their construction. Most of them have provided context or direction for this project.

Combinatory Logic. The structure of the resulting reaction system has some similarities with the systems used in Combinatory Logic [10, 24]. One of the motivation that lead to the introduction of Combinatory Logic was the analysis of the substitution process in formulas [10, p. 1]. The formal process with which a term is substituted for a variable in a formula is quite complicated, especially if the formula may contain free and bound variables. In Combinatory Logic, the substitution process is decomposed into elementary steps, which consists of purely textual substitutions. (See e. g. [24, p. 7].)

This concentration on elementary, textual substitutions in Combinatory Logic served as a model for the development of Flexible Time. Especially the concepts of applying a reaction (Definition 4.9) and confluence (Theorem 6.22) have their similarities in Combinatory Logic.

Development of a Language. There are many predecessors for the idea to develop a language that helps us to think more efficiently about cellular automata. Among the first, and certainly the most illustrious, was Gottfried Wilhelm Leibniz with his project of a "universal characteristic". This was to be an ideal language and a general symbolic method or both, because there is a certain ambiguity in Leibniz's writings [55, p. 226–227]. In the first interpretation, the language should consist of "signs which process a determinate content and

exactly correspond in their structure to the analysis of thought" [55, p. 227], in the second it would be a symbolic calculus, an "instrument to reason" [55, p. 230].

In this second aspect, Leibniz's work is widely seen as a predecessor to formal logic. (A rare concrete example for his thoughts about formal reasoning looks to modern eyes like a formalisation of set inclusion or propositional calculus [12, p. 18].) The first aspect emphasises the idea the *signs* of the language should correspond to the concepts of thought in a simple way. This is an aspect that is mostly ignored in the theory of formal systems, but not by Leibniz: His symbolism for integration was clearly designed with the intention in mind to find symbols that aid thought. The resulting mathematical language has always been seen as widely superior to Newton's version [12, p. 12].[†]

So far the current thesis could be seen as a part of Leibniz's project, but there are differences. First, Leibniz had viewed the signs of his universal characteristic as the most primitive concepts, and believed that they could be found once and for all. Second, he imagined his universal language as something complete, encompassing all human knowledge. A growing language, intended for a small subuniverse of mathematics, would not be his intention.

In order to find a model for this kind of project we need to look into a direction that at first seems to be completely unrelated: the construction of languages for the communication with extraterrestrials. There is a program outlined by Lancelot Hogben [25] on how to establish communication with an extraterrestrial civilisation via radio signals. To establish a means of communication, and a common vocabulary of concepts, "lessons" are sent out to the extraterrestrials, starting with numbers and arithmetic and then building up on this base increasingly complex concepts. Hogben's paper is only a sketch of such a program. The most elaborate implementation is certainly Hans Freudenthal's "Lincos" [16], in which he introduces step by step the concepts for mathematics, time, basic human behaviour and elementary physics.

But this description is misleading in one point: Freudenthal's primary interest was to create a logical language that was actually usable for communication, and in order to do this he used interstellar communication as an example problem. This is then the point where a project like Freudenthal's becomes a model for works like this thesis. We have here namely an example for a language that grows step by step from examples, which is never complete, and which at every step of its development can only access a limited set of concepts. It also sets an example by requiring a concrete example to let the language grow.

Influence on this Work. As Freudenthal needed a communication problem to develop a language for communication, we will need a self-organisation problem to develop a language about self-organisation and structures.

Another lesson from Freudenthal's work is to let the language evolve step by step, from simple to complex concepts. For this thesis this means that at the beginning the concepts are quite general and are valid for every one-dimensional cellular automaton. Step by step, by the amount that we learn about the theory and its abilities, the range of the definitions and theorems

[†]However, to my knowledge, Leibniz seems not to have understood the formalism of calculus as a part of his project of finding an universal characteristic.

becomes more restricted, but in exchange they become more powerful, until finally they offer insight into ether formation under Rule 54.

Chapter 3

Cellular Evolution

This chapter starts with the definition of the basic terms that are needed to speak about one-dimensional automata, and then introduces concepts that captures their development over time. It finishes with theorems about that what can be said about the development of a cellular automaton when only a part of its cells is known.

3.1 One-dimensional Cellular Automata

Imagine the cellular automaton as a physical object.

It consists of an infinite row of *cells*. The positions of the cells are integers. The cells are simple machines with a finite amount of memory, and they are all equal. Two cells may differ only by the content of their memory. The possible states of a cell are elements of the finite *state set* Σ. There is a function

$$c\colon \mathbb{Z} \to \Sigma \tag{3.1}$$

that maps the position of a cell to its state. Such a function is called here a *configuration* of the cellular automaton. The set of all configurations of a cellular automaton is therefore the set $\Sigma^{\mathbb{Z}}$.

Cellular automata evolve over time. Time for cellular automata is discrete and the time coordinate takes integer values. We speak of *time steps*. The behaviour of a cellular automaton is given by a *local transition rule* φ with *radius* r,

$$\varphi\colon \Sigma^{2r+1} \to \Sigma\,. \tag{3.2}$$

It maps the neighbourhood of a cell to the state of it one time step later. To see in which way, we need a notion for the collection of all configurations of a cellular automaton at all time steps. I call such a collection the *evolution sequence* of the cellular automaton.* It is an infinite sequence $(c_t)_{t\geq 0}$ of configurations. We

*In the theory of dynamical systems this is often called the *orbit*. But the definition of this term seems to vary between authors. For Alligood, Sauer and Yorke [2, p. 5], the orbit is a set, while for Strogatz [56, p. 348] it is a sequence. However, in the case of cellular automata a sequence is the more natural choice.

will only consider evolution sequences that belong to a specific local transition rule φ. In such an evolution the configuration c_0 must be specified in advance: it is the *initial configuration* of the sequence. Each later configuration c_t, with $t \geq 1$, depends by the *global transition rule*

$$c_t(x) = \varphi(c_{t-1}(x-r), \ldots, c_{t-1}(x+r)) \qquad \text{for all } x \in \mathbb{Z}. \tag{3.3}$$

on its predecessor configuration c_{t-1}.

Conventions. We can specify a one-dimensional cellular automaton completely by specifying Σ, r and φ. Since we will work almost always with one specific cellular automaton at a time, we will from now on keep Σ, φ and r fixed and not refer to it in most of the notation.

Furthermore, since the subject of this thesis are one-dimensional automata, we will from now on in most cases omit the adjective "one-dimensional".

Radius Invariance. It is possible that two different local transition rules lead to the same global rule. A pair of such equivalent transition rules is easy to construct: For a given local transition rule φ with radius r, let $\varphi' \colon \Sigma^{2r'+1} \to \Sigma$ be a rule with radius $r' > r$ such that

$$\varphi'(\sigma_{-r'}, \ldots, \sigma_{r'}) = \varphi(\sigma_{-r}, \ldots, \sigma_r) \qquad \text{for all } \sigma_{-r'}, \ldots, \sigma_{r'} \in \Sigma. \tag{3.4}$$

Then φ', which ignores the states of the additional cells, has the same global transition rule as φ. We call such a φ' the *extension* of φ to the radius r'. It is easy to see that if two local transition rules lead to the same global rule, then one must be the extension of the other one.

The centre of our interest in a cellular automaton is the behaviour of its cells, not its local transition rule. Therefore we will view here cellular automata with the same global transition rule as equivalent, since they have the same evolutions. Nevertheless the local transition rule provides an easy way to specify the properties of a cellular automaton. So we will use it, but we will require that the properties and functions defined for cellular automata are invariant of the radius of its local transition rule, in the following sense:

***Definition* 3.1** (Radius Invariance)**.** A property of a cellular automaton is *radius-invariant* when it is true for a local transition rule φ if and only if it is true for all its extensions.

What is then the radius of the cellular automaton itself? We will here allow that a cellular automaton has more than one radius: A number r is a *radius* for a cellular automaton if there exists a local transition rule with radius r that generates its global transition function.

3.2 Cellular Processes

At this point, our only tool to analyse the concrete behaviour of a cellular automaton—i. e. when its initial configuration c_0 is given—is its evolution sequence $(c_t)_{t\geq 0}$. But this is for many applications not enough. It requires the knowledge of infinitely many cell states, which is too much when our interest is only on the development of a specific localised pattern.

A better way of formalisation for cellular evolution is inspired by the way cellular evolution is usually shown in pictures.

Space-time Diagrams. The evolution of a cellular automaton is in general shown by a diagram like that in Figure 3.1. It is a rectangular array in which

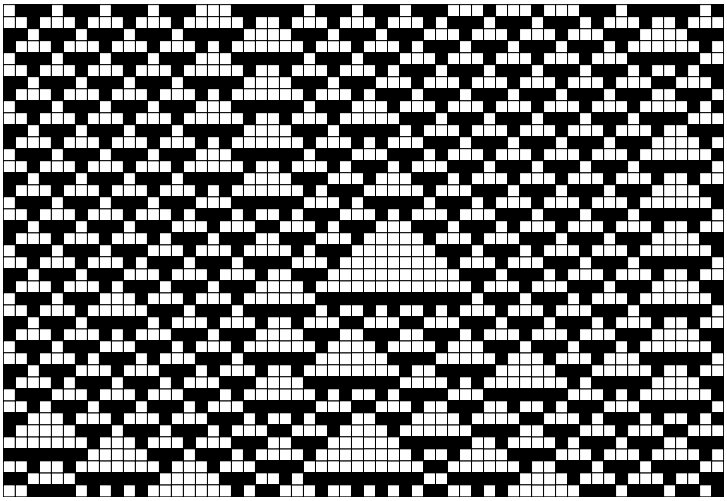


Figure 3.1: Space-time diagram of an evolution under Rule 54.

the states of the cells are shown by squares of different colours. In this picture, which shows the evolution of an elementary cellular automaton, the colours are white for cells in state 0 and black for cells in state 1. This colour convention is kept—sometimes in a modified way—in all the other pictures of cellular evolution that occur in this thesis.

The place of the square in the diagram specifies its place in space and time. The x-coordinate of a square determines its location in space, and its horizontal position marks the point in time to which it refers. In this thesis I use the physicist's convention for space-time diagrams in which time runs upward.[†] Each row in such a diagram is then part of a configuration, and the partial image of configuration c_{t+1} is directly above that of c_t.

Sometimes we will not need a whole rectangle and draw therefore only a subset of its squares.

Explanation of Figure 3.1. The diagram in Figure 3.1 then displays the evolution of a random initial configuration under Rule 54. Since it displays a the evolution of an infinite line of cells, there is no wraparound and, in contrast to many other space-time diagrams, the leftmost cell in each time step is *not* the right neighbour of its rightmost cell.

[†]There is also a strong tradition to draw the diagrams with time running downward. It probably has its origin in the time when one-dimensional cellular were simulated in a computer and then printed with a line printer, with one letter for every cell state and the cells of one time step in a line. Then the natural way to display the evolution is to print them in the way they are computed, and in the resulting diagram time runs downward. (For an example see Wolfram [59].)

For some reason, the physicist's convention is also the preferred convention used by French authors on cellular automata.

Events and Space-time. We will now introduce a mathematical equivalent to the space-time diagrams, namely the concept of cellular processes. It will be able to display any behaviour of the cells and is not restricted to the case that the cells follow a transitions rule.

The first step in defining a mathematical analog of an space-time diagram is to find a representation for a single square. A square has a position and a colour, and the colour represents a state. The position has a space and a time component, and both are integers. Therefore the following definition is reasonable.

***Definition* 3.2** (Cellular Event)**.** A *cellular event* is a pair $(p, \sigma) \in \mathbb{Z}^2 \times \Sigma$.

It consists of a *position* p and a *state* σ. The first component of p is its time coordinate and the second its space coordinate. The pair (p, σ) will be usually written $[p]\sigma$.

With this definition, negative time values are explicitly allowed. We will need them later, when situations are introduced in Definition 4.1, because for them it is natural to refer to events with negative time values. Viewed from our current standpoint, this gives us the choice to let the evolution of the cellular automaton begin at an arbitrary time step, not just at time 0.

The name "event" has been taken from Relativity theory. There it stands for a point of four-dimensional space-time (see Wald [58, p. 4]). I have here extended it to mean "space + time + physical conditions at this point", first because I have not found another word for this idea, and second because it then harmonises with the use of "event" in everyday language.

Using another word from Relativity [58], we will call the position p of an event $[p]\sigma$ a *space-time point*. The convention that the first component of a space-time point is the time and the second the space coordinate is also from Relativity. It extends in a natural way to all other cases where an element of $\mathbb{Z}^2$ is used for the same purpose in the context of cellular automata. I use here concepts from Relativity because Relativity theory has already well-developed concepts to treat space and time in an unified way.

With the notations

$$p_T = t \qquad \text{and} \qquad p_X = x \tag{3.5}$$

we will refer to the components of a space-time point $p \in \mathbb{Z}^2$ with $p = (t, x)$. The use of capital letters for this purpose in unusual, but lower case t is already used as time variable and occurs also as index letter.

As a kind of inverse to the component notation we will need the *unit vectors* of space-time,

$$T = (1, 0) \qquad \text{and} \qquad X = (0, 1), \tag{3.6}$$

especially to refer to differences between space-time points in a more abstract way.

A final set of conventions refer to the "point" component of an event $[p]\sigma$. If $p = (t, x)$, we may write $[p]$ as $[t, x]$. If $t = 0$, we may abbreviate $[t, x]$ to $[x]$. We also write $[p][q]$ for $[p + q]$. I list them here only for reasons of completeness. They will become useful later in the context of situations.

Processes. Next we need a mathematical object that resembles a whole space-time diagram. There are two possibilities. We may interpret the diagram as the picture of a function that maps a space-time point to a cell state. Then the correct way to represent a space-time diagram would be a map from a subset of $\mathbb{Z}^2$ to Σ. We will however also need to express unions and intersections of cellular processes, something that is easier to express if a cellular process were a set of events. Therefore I will introduce now a concept that intends to unite the good properties of functions and sets.‡

Let A and B be two sets. I call a set $F \subseteq A \times B$ *function-like* if there is a set $D \subseteq A$ and a function $f\colon D \to B$ such that

$$F = \{\, (a, f(a))\colon a \in D \,\}. \tag{3.7}$$

In notation we will treat function-like sets like functions. The term $F(a)$ stands for the element $b \in B$ for which $(a, b) \in F$. There is exactly one such b because F is function-like. The *domain* of F is the set

$$\operatorname{dom} F = \{\, a\colon \exists b \in B\colon (a, b) \in F \,\}. \tag{3.8}$$

For the F of equation (3.7) we have $\operatorname{dom} F = D$. The *restriction* of F to a set $A' \subseteq A$ is the set

$$F|_{A'} = \{\, (a, b) \in F\colon a \in A' \,\}, \tag{3.9}$$

which is also function-like.

Then we can define cellular processes as a special kind of function-like sets. Together with the cellular processes we define also a short notation for the subset of all events at a certain time.

***Definition* 3.3** (Cellular Process)**.** A *cellular process* is a function-like set of cellular events. The set of all cellular processes is called $\mathscr{P}$.

If $\pi \in \mathscr{P}$ and $t \in \mathbb{Z}$, then its restriction to events at time t, its *time slice*, is the cellular process

$$\pi^{(t)} = \{\, ([t, x]\sigma \in \pi\colon x \in \mathbb{Z} \,\}. \tag{3.10}$$

Compatibility. Next we consider the set-theoretic operations for cellular processes. Here we must know whether the result of a set-theoretic operation applied to one or more cellular processes is again a cellular process.

This is no problem with subset formation and intersection: since the subset of a cellular process is again a cellular process, the intersection of two processes is a process too. The only exception is the union of cellular processes. It is not always a function-like set.

An exception may occur when two cellular processes $\pi, \theta \in \mathscr{P}$ have domains that overlap in a point p. It is then possible that there are events $[p]\sigma \in \pi$ and $[p]\tau \in \theta$ with $\sigma \neq \tau$. Then the set $\pi \cup \theta$ exists, but it is no longer function-like. If it were, there would be a function $f\colon \operatorname{dom}\pi \cup \operatorname{dom}\theta \to \Sigma$ with $f(p) = \sigma$ and $f(p) = \tau$ at the same time, which is impossible.

If this does not happen, we say that π and θ are compatible:

‡There is a viewpoint in mathematics that functions *are* sets, but it is apparently not shared by everyone. Therefore I do here the unification explicitly. (I had used the other approach in my previous paper [51].)

***Definition* 3.4** (Compatibility)**.** Two cellular processes $\pi, \theta \in \mathscr{P}$ are *compatible* if

$$\pi(p) = \theta(p) \qquad \text{for all } p \in \operatorname{dom}\pi \cap \operatorname{dom}\theta. \tag{3.11}$$

We write this as π comp θ.

The question of compatibility plays an important role in this theory of cellular processes. In the rest of this text we must check very often whether a certain construction is possible, and if it is not, the cause is almost always incompatibility.

3.3 Evolution

We will now define what it means when a cellular process follows a transition rule. The construction that will be defined at the end must generalise the way in which an evolution sequence depends on its initial configuration. This is because currently the only way the behaviour of a cellular automaton is formally defined is via the evolution sequence.

As an intermediate step and to verify later the definition, we will translate now evolution sequences and configurations into the language of cellular processes. For this, let $(c_t)_{t\geq 0}$ be an evolution sequence of a cellular automaton. Then there exists a cellular process

$$\gamma = \{\, [t, x]c_t(x) \colon t, x \in \mathbb{Z} \,\} \tag{3.12}$$

that contains all the information in $(c_t)_{t\geq 0}$. The information of every configuration c_t in the sequence is contained in the time slice $\gamma^{(t)} = \{\, [t, x]c_t(x) \colon x \in \mathbb{Z} \,\}$ of γ. Our task is then to find a construction that, among other things, extends the initial time slice $\gamma^{(0)}$ to the whole process γ, in the same way as the global transition rule (3.3) extends the initial configuration c_0 to $(c_t)_{t\geq 0}$.

We will call this construction the *closure* of a process, because it will turn out to be a closure operator as defined on page 10.

The closure will be defined in two steps. First we consider the case of a single event. Given a process π and a point p, what does it mean that we can reconstruct the state of the event at p from π? If this is the case, we say that the event at p is determined by π. What this means exactly will be described in Definition 3.8.

As a second step we consider the events that are determined by π, together with the events that are determined by π and them, and so on: together they form the closure of π. It will turn out that not every process has a closure. The result of the second step is Definition 3.10.[§]

Determined Events. Let $p = (t, x)$ be a space-time point. Given a cellular process $\pi \in \mathscr{P}$ and a transition rule φ of radius r, what could be the state of the event at p?

We will answer this question first for the case of $\pi = \gamma$, with γ as in (3.12). In γ, the states of an event at time $t > 0$ depend, by the global transition rule,

[§]The construction introduced here has some similarity with the use of "tiling constraints" to specify the space-time pattern of the cell states in a one-dimensional cellular automaton by Ollinger and Richard [48, p. 4].

on the events at time $t-1$. We will then say that the events of $\gamma \setminus \gamma^{(0)}$ are *determined* by γ. From this we will now distill concepts that tell us how a transition rule φ acts on a cellular process. The first goal is then to express the global transition rule for evolution sequences in a form that is meaningful for processes like γ.

The point $p = (t, x)$ is the coordinate of the cell at position x and time t. The state of a cell at time t depends on the states of the cells in its neighbourhood at time $t-1$. So we must consider the neighbourhood of the point $p - T = (t-1, x)$ to compute the state of the event at p.

Neighbourhoods. For easier notation we will now first describe the neighbourhood of the point p instead of that of $p - T$. We begin with the neighbourhood of a cell as a set of space-time points, without reference to a cellular process. Since the transition rule has radius r, the cell in the cellular automaton at position x has a neighbourhood that consists of the cells at positions $x - r$, $\ldots$, $x + r$. At a time t, these cells are located at the space-time points $(t, x - r)$, $\ldots$, $(t, x + r)$. The central cell itself is located at (t, x), or p. Therefore we can say that the neighbourhood of the point p is consists of the points $p - rX, \ldots$, $p + rX$. To refer to it we introduce the following definition.

***Definition* 3.5** (Neighbourhood Domain)**.** Let $p \in \mathbb{Z}^2$ and $r \in \mathbb{N}_0$. The *neighbourhood domain* of p with radius r is the set

$$N(p, r) = \{p - rX, \ldots, p + rX\}. \tag{3.13}$$

Next we must find an expression for the states of those events in γ that are located at the points of $N(p, r)$. We need them not just as a set, but also in their natural order. Therefore we express them as the cellular process $\nu(p, w)$, defined below. In the same way that we can write $[p]\sigma \in \gamma$ to express the fact that in the process γ the event at point p has state σ, we will write $\nu(p, \omega_{-r} \ldots \omega_r) \subseteq \gamma$ to express the fact that in γ the events at the points $p - rX$, $\ldots$, $p + rX$ have, respectively, the states $\omega_{-r}, \ldots, \omega_r$.

***Definition* 3.6** (Neighbourhood Process)**.** Let $w = \omega_{-r} \ldots \omega_r \in \Sigma^{2r+1}$. The *neighbourhood process* for w at p is the cellular process

$$\nu(p, w) = \{[p - rX]\omega_{-r}, \ldots, [p + rX]\omega_r\}. \tag{3.14}$$

The Transition Rule. We now return to the computation of the state of γ at p. With neighbourhood processes we can express the global transition rule (3.3) for evolution sequences in a new way for cellular processes like for γ. A direct translation of (3.3) uses the fact that $\gamma(t, x) = c_t(x)$ for all $t \in \mathbb{N}_0$ and $x \in \mathbb{Z}$. We now replace all terms like $c_t(x)$ with terms of the form $\gamma(t, x)$ and get the formula

$$\gamma(t, x) = \varphi(\gamma(t-1, x-r), \ldots, \gamma(t-1, x+r)), \tag{3.15}$$

which is valid for all $t > 0$ and $x \in \mathbb{Z}$. With neighbourhood processes this becomes the condition,

$$\text{if} \quad \nu((t-1, x), \omega_{-r} \ldots \omega_r) \subseteq \gamma, \quad \text{then} \quad \gamma(t, x) = \varphi(\omega_{-r}, \ldots, \omega_r), \tag{3.16}$$

which can be shortened by using $p = (t, x)$ again and by setting $w = \omega_{-r} \dots \omega_r$. Then it becomes the requirement that

$$\text{if} \quad w \in \Sigma^{2r+1} \quad \text{and} \quad \nu(p - T, w) \subseteq \gamma, \quad \text{then} \quad \gamma(p) = \varphi(w). \tag{3.17}$$

This is is a formulation of the global transition rule for the process γ. It is valid for all $p \in \mathbb{Z}^2$ with $p_T > 0$.

Arbitrary Processes. Now we return to an arbitrary process π. We ask which state we should expect for the event at (t, x), given the information in π. To do this we will view π as a window into the evolution of a cellular automaton that follows rule φ.

If $w = \omega_{-r} \dots \omega_r$ and $\nu((t-1, x), w) \subseteq \pi$, then at the time $t - 1$ the states of the cells in the neighbourhood of the cell at x are $\omega_{-r}, \dots, \omega_r$. At time t, the state of the cell at x must then be $\varphi(w)$. This is then the expected state for the event at (t, x).

When however $N(p, r) \not\subseteq \operatorname{dom} \pi$, then we have not enough information about the evolution to find the state for p in this way. Instead we can find a set of possible states: If there is a process $\pi' \supseteq \pi$ such that $\nu(p - T, w) \subseteq \pi'$, then $\nu(p - T, w)$ is a possible neighbourhood for $p - T$, and $\varphi(w)$ is a possible state for the event at p. The set of possible states for the event at p is therefore

$$\{\, \varphi(w) \colon w \in \Sigma^{2r+1}, \exists \pi' \subseteq \mathscr{P} \colon \nu(p - T, w) \subseteq \pi' \supseteq \pi \,\}. \tag{3.18}$$

However, as stated here this definition involves an infinite number of processes π', which is bad for actual computations. We avoid this by choosing only those π' that contain only as many additional points that $\nu(p - T, w) \subseteq \pi$. Then $\nu(p, -T, w)$ is compatible with π (Figure 3.2). This then leads to the following definition, in which π' does no longer occur explicitly.

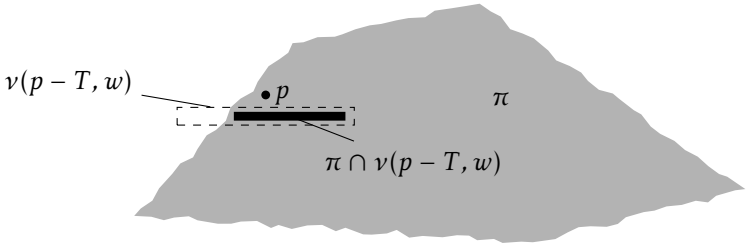


Figure 3.2: Determining the possible states for the event at p.

***Definition* 3.7** (Set of Possible States)**.** Let $\pi \in \mathscr{P}$ be a cellular process and φ be a transition rule for Σ of radius r. The *set of possible states* for the event at p is

$$S(p, \pi) = \{\, \varphi(w) \colon w \in \Sigma^{2r+1}, \nu(p - T, w) \operatorname{comp} \pi \,\}. \tag{3.19}$$

If there is only one possible state for the event at p, then it is determined. This is expressed in the following definition.

***Definition* 3.8** (Determined Events)**.** Let $\varphi \colon \Sigma^{2r+1} \to \Sigma$ be a transition rule and $\pi \in \mathscr{P}$ a process. If

$$S(p, \pi) = \{\sigma\}, \tag{3.20}$$

then both the event $[p]\sigma$ and the point p are *determined* by π. The set of all events that are determined by π is $\Delta\pi$.

Non-constant Transition Rules. Note that if φ is a constant function, then every point of $\mathbb{Z}^2$ is determined. We will therefore restrict the following definitions and theorems to non-constant transition functions, in order to avoid this unintuitive property.

The following lemma summarises useful properties of determinedness that are only true for a transition rule that is non-constant. The statements of this lemma are also the reason why the set $N(p, r)$ gets a special name.

***Lemma* 3.9** (Events Determined by a Time Slice)**.** *Let φ be a non-constant transition rule for Σ and $\pi \in \mathscr{P}$ a process. Then:*

1. *A point p is determined by π if and only if it is determined by $\pi|_{N(p-T,r)}$.*
2. *If p is determined by π, the set $\pi|_{N(p-T,r)}$ is nonempty.*
3. *For every time t, the set $\Delta\pi^{(t)}$ consists only of events at time $t+1$.*

The first statement of this lemma expresses again the fact that determinedness is a local property and relies only on a finite number of events.

The second statement is about causality. For it we need to have the view of a cellular automaton as a physical system governed by the "physics" φ. Then the events of $\pi|_{N(p-T,r)}$ can be understood as the "cause" of the event at p. The second statement of the theorem then states that an event at time t which is determined by π is always caused by an event at time $t-1$. It is also a statement about the maximal speed with which information is transmitted: the state of a cell at time t can only be caused by the cells at most r positions to its left or right.

The third statement is tailored for its use in connection with the closure of a process, which will be defined next.

Proof of the lemma. For the proof of the first statement we note that for every $w \in \Sigma^{2r+1}$ the domain of the neighbourhood process $\nu(p, w)$ is $N(p, r)$. Therefore the set $S(p, \pi)$ does actually depend only on $\pi|_{N(p-T,r)}$. The knowledge of this part of π is therefore also enough to find out whether p is determined.

To show the second statement we prove its converse. Assume that $\pi|_{N(p-T,r)}$ is empty. Then every neighbourhood process $\nu(p-T, w)$ in (3.19) is compatible to π. Since φ is non-constant, the set $S(p, \pi)$ has more than one element. Therefore the point p is then not determined by π.

The third statement then follows from the second. □

The Closure. Now we can extend the global transition rule (3.17) from γ to arbitrary cellular processes. Similar to the way an evolution sequence is generated by always computing the configuration for time t from the configuration for time $t-1$, the closure of a cellular process is created from time slices, each of them depending on the previous one, that are finally put together. There is however no direct analog to the initial configuration.

To understand what is meant with the closure of a cellular process, imagine that cells are multicoloured lights that can be switched on or off. If a light is switched on, it has one of a finite set of colours. A cellular automaton is then

an infinite line of such lights, and the colours represent the states of its cells. A cellular process π is then a certain light pattern, a rule when to switch the lights on and with which colours. The closure of π is another light pattern, where the lights are switched on not only when it is required by π but also depending on the lights that were switched on at the previous time step. When the lights switched on at the previous time step determine the state of a cell in the current time step, then this cell is also switched on, besides the cells that are required by π to be switched on. A conflict between these two rules is possible: It can happen that the light pattern prescribes one colour for a cell at a certain time and π describes another. Then we will say that for this π the closure does not exist.

To be practical this procedure must have a starting time. We will require that there was a time when no event of π happened; otherwise π will have no closure. In the following definition we therefore call a cellular process π *quiet before* t_0 if $\pi^{(t)} = \emptyset$ for all $t < t_0$.

***Definition* 3.10** (Closure)**.** Let φ be a non-constant transition rule for Σ and $\pi \in \mathscr{P}$ be a cellular process that is quiet before t_0.

The *closure of* π *at time* t under φ is the process $\mathrm{cl}^{(t)}\,\pi$. It exists always when $t \le t_0$. When $t > t_0$, the process $\mathrm{cl}^{(t)}\,\pi$ exists if $\mathrm{cl}^{(t-1)}\,\pi$ exists and $\pi^{(t)}$ is compatible with $\Delta\,\mathrm{cl}^{(t-1)}\,\pi$. It is defined by the recursion

$$\mathrm{cl}^{(t)}\,\pi = \begin{cases} \pi^{(t)} & \text{if } t \le t_0, \\ \pi^{(t)} \cup \Delta\,\mathrm{cl}^{(t-1)}\,\pi & \text{if } t > t_0. \end{cases} \tag{3.21}$$

The *closure of* π under φ exists if all time slices $\mathrm{cl}^{(t)}\,\pi$ exist. It is the process

$$\mathrm{cl}\,\pi = \bigcup_{t\in\mathbb{Z}} \mathrm{cl}^{(t)}\,\pi\,. \tag{3.22}$$

An incidental result of equation (3.22) is that always $(\mathrm{cl}\,\pi)^{(t)} = \mathrm{cl}^{(t)}\,\pi$. This is the way the operator $\mathrm{cl}^{(t)}$ fits into the formalism of time slices of Definition 3.3.

There is one act of choice in this definition. If a process π is quiet before t_0, then it is also quiet before any time $t_0' < t_0$. Therefore one can also compute the closure of π using t_0' as starting time. But this has no influence on $\mathrm{cl}\,\pi$. To see this, assume that ψ is the closure of π as computed with t_0 and ψ' the closure of π as computed with t_0'. Then for $t \le t_0$, the first case of (3.21) applies to the computation of ψ, and we have $\psi^{(t)} = \pi^{(t)} = \emptyset$. For $t \le t_0'$ it also applies to the computation of ψ', so we have then $\psi'^{(t)} = \pi^{(t)} = \emptyset$. For $t_0' < t \le t_0$, the second case of (3.21) applies and we can see by induction that then $\psi'^{(t)} = \emptyset$: If $\psi'^{(t-1)} = \emptyset$, then

$$\psi'^{(t)} = \pi^{(t)} \cup \Delta\psi'^{(t-1)} = \emptyset \cup \Delta\emptyset = \emptyset\,. \tag{3.23}$$

Therefore for $t \le t_0$, $\psi^{(t)} = \psi'^{(t)}$. At later times, the second part of (3.21) comes into play for both ψ and ψ' to construct the next time slice. Therefore for $t > t_0$ the time slice $\psi^{(t)}$ exists if and only if $\psi'^{(t)}$ exists, and when one of them exists, then $\psi^{(t)} = \psi'^{(t)}$. This shows then that ψ exists if and only ψ' exists, and if they exist, they are equal. In other words, the choice of t_0 has no influence on the closure.

Most properties of the closure are proved by inductions in the style of this proof. In it one can also see why the definition of the closure was restricted to non-constant transition rules: This restriction ensured in (3.23) that $\Delta\emptyset = \emptyset$ and thus removed unnecessary complexity.

3.4 Properties of the Closure

We now will prove some properties of the closure that either become useful later or will provide insight about this concept.

Shift Invariance. Like the physical laws which they imitate, the laws of a cellular automaton are independent of an absolute location in space and time. In the next chapter we will use this property for a simplification of the formalism; at this point we are mainly concerned with a way to express it for cellular automata.

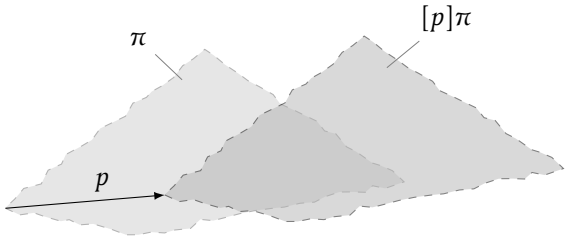


Figure 3.3: A shifted cellular process.

We can express it by the concept of a space-time shifted process (Figure 3.3). The notation $[p]\pi$ has been chosen in harmony to the other uses of square brackets in this text.

***Definition* 3.11** (Space-time Shift)**.** Let $\pi \in \mathscr{P}$ be a process and $p \in \mathbb{Z}^2$ be a space-time point. We write for the copy of π that is *shifted* by p,

$$[p]\pi = \{\, [p+q]\sigma : [q]\sigma \in \pi \,\}. \tag{3.24}$$

A property of cellular processes is *shift-invariant* if it is true for a process π if and only if it is true for $[p]\pi$. A function $F\colon \mathscr{P} \to \mathscr{P}$ between cellular processes is shift-invariant if $F([p]\pi) = [p]F(\pi)$ for all p and π.

Then we can say that Δ and cl are two shift-invariant functions. In a more informal way we will also say that determinedness is a shift-invariant notion, meaning that π determines $[p]\sigma$ if and only if $[q]\pi$ determines $[q+p]\sigma$. I do not give here a formal proof for these facts: they can easily be verified from their definitions, by checking that they use only differences between space-time points and no absolute coordinates.

Radius Invariance. As requested at the beginning of this chapter, the concepts introduced here are radius-invariant. Neither the closure nor the set of determined events of a process are dependent of the radius of the transition rule. To verify this, the following lemma is sufficient, since the closure is

defined with help of the set of determined events, and determined events are defined with help of the set of possible states for a point.

In the proof we have to distinguish between the sets $S(p, \pi)$ for the transition rules φ and φ; we will therefore write the two sets of possible states as $S_\varphi(p, \pi)$ and $S_{\varphi'}(p, \pi)$.

***Lemma* 3.12** (Possible States are Radius Invariant)**.** *Let $\varphi\colon \Sigma^{2r+1} \to \Sigma$ be a transition rule and $\varphi'\colon \Sigma^{2r'+1} \to \Sigma$ an extension of φ with radius $r' > r$.*

Let $\pi \in \mathscr{P}$ be a cellular process and $p \in \mathbb{Z}^2$. Then $S_\varphi(p, \pi) = S_{\varphi'}(p, \pi)$.

Proof. Assume that $\sigma \in S_\varphi(p, \pi)$. Then there are states $\omega_{-r}, \ldots, \omega_r \in \Sigma$ such that $\varphi(\omega_{-r}, \ldots, \omega_r) = \sigma$ and $\nu(p - T, \omega_{-r} \ldots \omega_r)$ comp π. Now choose the cell states $\omega_{-r'}, \ldots, \omega_{-r-1}$ and $\omega_{r+1}, \ldots \omega_{r'} \in \Sigma$ in the following way: If $p - T + iX \in \operatorname{dom} \pi$, then $\omega_i = \pi(p - T + iX)$; otherwise ω_i is arbitrary. Then $\nu(p - T, \omega_{-r'} \ldots \omega_{r'})$ is compatible with π. Since φ' is an extension of φ, we have also $\varphi'(\omega_{-r'} \ldots \omega_{r'}) = \varphi(\omega_{-r}, \ldots, \omega_r) = \sigma$. Therefore $\sigma \in S_{\varphi'}(p, \pi)$, which in turn proves that $S_\varphi(p, \pi) \subseteq S_{\varphi'}(p, \pi)$.

Assume that $\sigma \in S_{\varphi'}(p, \pi)$. This means that there are states $\omega_{-r'}, \ldots, \omega_{r'} \in \Sigma$ such that $\varphi'(\omega_{-r'}, \ldots, \omega_{r'}) = \sigma$ and $\nu(p - T, \omega_{-r'} \ldots \omega_{r'})$ comp π. Then $\varphi(\omega_{-r}, \ldots, \omega_r) = \sigma$ because φ' is an extension of φ, and $\nu(p - T, \omega_{-r} \ldots \omega_r)$ is compatible with π because $\nu(p-T, \omega_{-r} \ldots \omega_r)$ is a subset of $\nu(p-T, \omega_{-r'} \ldots \omega_{r'})$. This then proves that $\sigma \in S_\varphi(p, \pi)$, and therefore that $S_{\varphi'}(p, \pi) \subseteq S_\varphi(p, \pi)$. □

Monotony. We return for a moment to the view of a cellular process as a partial description for an evolution of a cellular automaton. Finding the closure can then be seen as reconstructing an evolution from incomplete information. More information should then result in a larger reconstruction. So we will expect that the closure of the superset of a process is a superset of its closure, or, in other words, that the closure operator defines a monotone function. This property is used very often.

Its proof begins with the proof of the same property for determinateness.

***Lemma* 3.13** (Determinateness is Monotone)**.** *Let $\pi \subseteq \psi \in \mathscr{P}$ be two processes and φ be a non-constant transition rule. Then $\Delta\pi \subseteq \Delta\psi$.*

Proof. Let $[p]\sigma$ be determined by π. Then $S(p, \pi) = \{\sigma\}$. Since $\psi \supseteq \pi$, the requirement that $\nu(p - T, w)$ comp ψ is a stronger restriction on w than the requirement that $\nu(p - T, w)$ comp π. So we must have $S(p, \psi) \subseteq S(p, \pi)$ and therefore $S(p, \psi) \subseteq \{p\}$. On the other hand, $S(p, \psi)$ has at least one element. So $[p]\sigma$ is determined also by ψ. □

All properties of the closure involve questions of its existence, therefore also this one. The theorem below expresses the intuitive notion that more requirements on the behaviour of a cellular automaton make it more likely that they are inconsistent and cannot be satisfied by the evolution of a cellular automaton.

***Theorem* 3.14** (Closure is Monotone)**.** *Let $\pi \subseteq \psi \in \mathscr{P}$ be two processes and φ be a non-constant transition rule.*

If cl ψ *exists, then* cl π *exists and* cl $\pi \subseteq$ cl ψ.

Proof. Assume that $\mathrm{cl}\,\psi$ exists. We will say that the theorem is *true for time t* if $\mathrm{cl}^{(t)}\,\pi$ exists and $\mathrm{cl}^{(t)}\,\pi \subseteq \mathrm{cl}^{(t)}\,\psi$.

Since $\mathrm{cl}\,\psi$ exists, there must be a $t_0 \in \mathbb{Z}$ such that ψ is quiet before t_0. Then π is quiet before t_0 too, because $\pi \subseteq \psi$. Therefore $\mathrm{cl}^{(t)}\,\pi = \pi^{(t)} \subseteq \psi^{(t)} = \mathrm{cl}^{(t)}\,\psi$ for all $t \leq t_0$. So the theorem is true for every $t \leq t_0$.

Let now $t > t_0$ and assume that the theorem is true for $t-1$. Then $\mathrm{cl}^{(t-1)}\,\pi$ exists. Because $\mathrm{cl}^{(t-1)}\,\pi \subseteq \mathrm{cl}^{(t-1)}\,\psi$, we have $\Delta\,\mathrm{cl}^{(t-1)}\,\pi \subseteq \Delta\,\mathrm{cl}^{(t-1)}\,\psi$ with Lemma 3.13. Because $\pi^{(t)} \subseteq \psi^{(t)}$, we have

$$\pi^{(t)} \cup \Delta\,\mathrm{cl}^{(t-1)}\,\pi \subseteq \psi^{(t)} \cup \Delta\,\mathrm{cl}^{(t-1)}\,\psi, \tag{3.25}$$

but only as an inclusion between sets of events. We have not yet proved that these sets are cellular processes. The right side of (3.25) is however the cellular process $\mathrm{cl}^{(t)}\,\psi$, and therefore the left side, as its subset, must also be a process. This then means that $\pi^{(t)}$ is compatible with $\Delta\,\mathrm{cl}^{(t-1)}\,\pi$ and that therefore $\mathrm{cl}^{(t)}\,\pi$ exists. Since the left side of (3.25) is then $\mathrm{cl}^{(t)}\,\pi$, while its right side is $\mathrm{cl}^{(t)}\,\psi$, we have proved that $\mathrm{cl}^{(t)}\,\pi \subseteq \mathrm{cl}^{(t)}\,\psi$. Therefore the theorem is true for time t if it is true for time $t-1$ when $t > t_0$.

So we have shown by induction that the theorem is true for all times t, and therefore true in general. □

Closure and Evolution Sequence. As the final task of this chapter we now verify what we required of the closure at the beginning, when we motivated its construction: The process γ of (3.12), the translation of the evolution sequence $(c_t)_{t\geq 0}$, is the closure of the cellular process for its initial configuration, $\gamma^{(0)}$.

We will prove a bit more: The following lemma shows that for every time $t \geq 0$, the time slice $\gamma^{(t)}$ determines the following time slice $\gamma^{(t+1)}$ in the same way that the configuration c_t determines the following configuration c_{t+1}. The lemma is then the analog of the transition rule (3.3) for cellular processes.

***Lemma* 3.15** (Global Transition Rule for Processes)**.** *Let φ be a non-constant transition rule and let γ be as in* (3.12)*. Then for all $t \geq 0$,*

$$\gamma^{(t+1)} = \Delta\gamma^{(t)}\,. \tag{3.26}$$

Proof. By Lemma 3.9, the only events that can possibly be determined by $\gamma^{(t)}$ have a time coordinate of $t+1$. It only remains to prove that $\gamma^{(t)}$ determines all events $[t+1, x]c_{t+1}(x)$ with $x \in \mathbb{Z}$.

Let $p = (t+1, x)$. To know whether this point is determined we have to find $S(p, \gamma^{(t)})$.

Because $\mathrm{dom}\,\nu(p-T, w) \subset \mathrm{dom}\,\gamma^{(t)}$ for all $w \in \Sigma^{2r+1}$, the process $\nu(p-T, w)$ is compatible with $\gamma^{(t)}$ if and only if it is a subset of $\gamma^{(t)}$. So we must find all $w \in \Sigma^{2r+1}$ that satisfy

$$\nu(p-T, w) = \{[t, x-r]c_t(x-r), \ldots, [t, x+r]c_t(x+r)\}, \tag{3.27}$$

where r is the radius of φ. This equation has one solution, $w = c_t(x-r)\ldots c_t(x+r)$.

So the set $S(p, \gamma^{(t)})$ has exactly one element, which means that the event at p is determined. Its state is $\varphi(w)$, which is equal to $c_{t+1}(t+1, x)$ by the global transition rule (3.3). This proves the lemma. □

Using this lemma we can then easily see that $\mathrm{cl}^{(t)}\,\gamma^{(0)} = \gamma^{(t)}$ for $t \geq 0$ and that $\mathrm{cl}^{(t)}\,\gamma^{(0)}$ is empty for $t < 0$. This then shows that γ is indeed generated by its initial time slice $\gamma^{(0)}$.

3.5 Summary

In this chapter we have formalised the concept of the evolution of a cellular automaton in a new way, in order to be able to understand the fate of a localised arrangement of cells. The goal was to have a notation that treats events at different times on an equal footing.

The starting point was the description of cellular automata with configurations and evolution sequences. This is a natural way to understand the cellular automaton as a machine that evolves over time. The transition rule then is a description of the law that governs the behaviour of the cells.

This method to describe the behaviour of cellular automata was then decomposed into its components. The configurations became sets of cellular events. The transition rule was expressed in a radius-invariant way as the function that generates the set of determined events for a process. The evolution sequence became the closure of a process.

The concept of closure helps us to express how information propagates in a cellular automaton. We have seen that the closure operator is monotone, which will help us to reason in an abstract way about cellular processes. We have seen how the evolution of a configuration is expressed with cellular processes. We have introduced the concept of radius-invariance. The closure operator, as the new form of the transition rule, is radius-invariant even for a cellular process of finite size. We have therefore extended the concept of the initial configuration to an arbitrary set of cells at arbitrary times.

Some cellular processes however have no closure and it is not yet clear how to construct processes that have a closure. This question will be answered in Chapter 6 with the concept of achronal situations.

Chapter 4

Reaction Systems

One of the goals of this thesis is to find a way in which we can express the laws of large-scale behaviour in a cellular automaton. A "law of large-scale behaviour" is here any statement that involves an arbitrarily large number of cells. The triangles below in Figure 4.1 are an intuitive example: One knows that the exact number of white cells in the triangle's base does not matter. It could become arbitrarily large, and the same kind of triangular shape would result. We need to express this kind of intuitive law—and much more complex laws—in a formal way.*

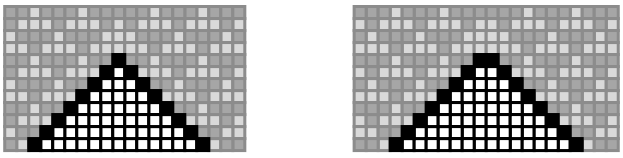


Figure 4.1: Triangles as computations under Rule 54.

In this chapter we will define a formalism with which we can express instances of such large-scale laws in terms of the input and result of a computation. The computation is then represented by the ordered pair of input and output; intermediate steps are ignored.†

*The specific law that is expressed in Figure 4.1 is expressed in Table 8.1. Other examples are the laws of ether formation, like Lemma 8.16.

†The concepts introduced here rely to a great extent on the *virtual state machine* (VSM) introduced by Christopher G. Langton [28]. In contrast to the VSM, which is understood by Langton as a long-lived entity, a reaction (defined below) always refers to a finite time span. One could then understand a reaction as the description of a single computational step in the existence of a VSM.

4.1 Situations

We will take here the viewpoint that we build the computer inside the universe represented by the cellular automaton. The transition rule is kept fixed; it represents the physics of that world.

In a computer, we distinguish between the data and the computing machine. In the cellular automaton, the data and the machine are both cell states. A very simple example for the way data and computation interact are the triangular structures found under Rule 54 and other cellular automata (Figure 4.1). Here we may view the cell configuration at the initial time as the input; it consists of two cells in state 1 that surround a sequence of zeros. The highlighted cells in the figure are then the computation initiated by this input. How long the computation lasts depends directly on the number of zeros in the input. We can view therefore it as a kind of counter, or a loop that counts downward. Another viewpoint, since the triangles have a different shape depending on whether the number of zeros is odd or even, is to understand the triangles as programs that test for parity.

When we now look at these triangular processes in terms of the cells involved, we see that the computation takes a different amount of time at different locations. It lasts longer at the center of the triangle than at its margins. Since we have viewed the initial interval of cells as the input of the computation, we will take the other sides of the triangle as its result. It then consists roughly of the black cells at the boundaries of the triangles, together with their direct neighbours. Later, in Figure 8.2, the input and result of this kind of computation will be shown explicitly.

So the formalism of Flexible Time makes no difference between the data and the machine and represents them in a single mathematical object. In a similar vein, input and result of a computation are the same kind of mathematical object. This is so because it must be possible that the result of one computation is the input of another computation.

Properties of Situations. The mathematical objects that represent the input or result of a cellular computation are called here *situations*; they are defined below. But before we can write down the formal definition, we will collect the properties a situation must have.

(a) A situation specifies a finite cellular process. To specify the input or the result of a computation in a cellular automaton means to express requirements on the states of cells: at a specified time a cell must be in a specified state. Put together, these requirements are the events of a cellular process. It is a finite process because the input and output of a computation always have a finite number of bits.

(b) A situation may specify events at different times. With the finite speed at which signals travel in a cellular automaton, exact synchronisation of the components in the computer is difficult, and different parts of a computation may end at a different time. We must therefore allow that the events of a situation belong to different times.

(c) A situation specifies a sequence of events. The cells in a one-dimensional cellular automaton have a natural order from left to right. Situations generalise finite sets of adjacent cells together with their states. Therefore it is good for

our intuition if the events in a situation also form a sequence. It especially allows us to use the formalism of finite sequences for situations.

I will now sketch the way in which we get a meaningful left-to-right arrangement for the events of a situation, starting from the order in a set of adjacent cells. Take a situation that specifies the states of such a set of cells and use it as the input of a computation. In general, the output situation of this computation specifies events at different times. They have no longer a natural left-to-right order, but if the duration of the computation is short enough, there remains an approximate order. As information travels with finite speed in a cellular automaton, each event in the result has only a few events in the input that are near enough to have caused it. So we can arrange the events in the output approximately from left to right with help of the input events. Later we will make this idea precise; we will then have a correspondence between the order of the input of any computation and that of its output.

(d) A situation has a size vector. This vector is an analog to the length $|u|$ of a finite sequence u. Imagine that the sequence u consists of letters written on grid paper, with the $|u|$ letters of u on the squares numbered $0, \ldots, |u|-1$. Now consider uv, the product of u with some other sequence v. In it, when it is written down the same way, the letters of v will occupy the positions $|u|, \ldots, |u|+|v|-1$, instead of $0, \ldots, |v|-1$, as it would have been if v alone were written. So the length of a sequence u marks the point behind u, and this point is the starting point of the second factor in a product involving u. The same happens with situations, except that its size vector is an element of $\mathbb{Z}^2$ and that it can be chosen arbitrarily. We then can define a notion of product for situations similar to that for sequences; the arbitrariness of the size vector makes it possible that there are gaps between the event sequences of the factors.

The following is then a situation: a sequence of cellular events together with a size vector.

***Definition* 4.1** (Situations)**.** A *situation of length* ℓ with states in Σ is a pair

$$((e_0, \ldots, e_{\ell-1}), p) \in (\mathbb{Z}^2 \times \Sigma)^\ell \times \mathbb{Z}^2 \tag{4.1}$$

for which the set $\{ e_i : 0 \le i < \ell \}$ is a cellular process. The set of all situations of length ℓ is $\mathscr{S}_\ell$. A *situation* in general is an element of the set

$$\mathscr{S} = \bigcup_{\ell \ge 0} \mathscr{S}_\ell . \tag{4.2}$$

The *event sequence* of the situation $((e_0, \ldots, e_{\ell-1}), p)$ is then the tuple of events, $(e_0, \ldots, e_{\ell-1})$, and its size vector is the point p. A third property of the situation that we required before, namely a cellular process associated to it, is introduced in the following definition. It also specifies a notation for the size vector that does not require to spell out a situation as a pair.

***Definition* 4.2** (Components of a Situation)**.** Let $a = ((e_0, \ldots, e_{\ell-1}), p)$ be a situation. The *process* of a is the cellular process

$$\mathrm{pr}(a) = \{ e_i : 0 \le i < \ell \} . \tag{4.3}$$

For p, the *size vector* of a, we write $\delta(a)$.

The symbol for the size vector, δ, should remind of another way to view a situation. We can view a situation a as having a left end at the coordinate origin an a left end at $\delta(a)$; then $\delta(a)$ is the difference between the two ends of a situation. Therefore the symbol.

With the notations of Definition 4.2 we can already express a convenient shorthand notation. Let a and b be two situations. The *process of b, shifted by a* is then

$$\mathrm{pr}_a(b) = [\delta(a)]\mathrm{pr}(b)\,. \tag{4.4}$$

It will become useful once we have defined the product of situations.

Path Notation. Let a be a situation of length ℓ. If we want to express it in full detail, we currently have to write it in the form

$$a = (([p_0]\alpha_0, \ldots, [p_{\ell-1}]\alpha_{\ell-1}), p_\ell), \tag{4.5}$$

with $p_i \in \mathbb{Z}^2$ and $\alpha_i \in \Sigma$ for all i. The use of expressions of this kind for longer calculations and proofs would however soon become quite cumbersome. Therefore we now introduce a shorter form. It will fulfil the remaining requirement on situations and provide a way to treat situations in the same manner as finite sequences. In the full-developed formalism we will then refer to the properties of a situation a only with help of the new notation, the process $\mathrm{pr}(a)$ and the size vector $\delta(a)$ and no longer refer to terms of the form (4.5) directly.

The new notation uses a relative notation for the locations of the events in a situation.

***Definition* 4.3** (Path Notation)**.** Let a be a situation written in the form (4.5). Let $\tilde{p}_0 = p_0$ and $\tilde{p}_i = p_i - (p_{i-1} + X)$ for $i > 0$. The *long path notation* for a is then

$$a = [\tilde{p}_0]\alpha_0[\tilde{p}_1]\alpha_1 \ldots [\tilde{p}_{\ell-1}]\alpha_{\ell-1}[\tilde{p}_\ell], \tag{4.6}$$

The terms $[\tilde{p}_i]$ are the *displacements* of a.

The *short path notation* of a is similar to this, but all terms $[\tilde{p}_i]$ with $\tilde{p}_i = (0, 0)$ are removed from it. An exception is the case of $\ell = 0$: a situation $a = [\tilde{p}_0]$ cannot be shortened. It is by definition already in short path notation.

One can understand the path notation as the description of a writing process. In it, symbols for the cell states are written into a square grid similar to a space-time diagram. After writing a symbol into the square at point p_{i-1} the cursor is at $p_{i-1} + X$. The displacement $[\tilde{p}_i]$ is then the amount of *extra* movement before the next symbol can be written down. This explains the occurrence of the unit vector X in the definition of $\tilde{p}_i$ above. It also explains the formula

$$p_i = \sum_{j=0}^{i} \tilde{p}_j + jX \qquad \text{for } i = 0, \ldots, \ell \tag{4.7}$$

that converts the displacements of the long path notations back into absolute positions. (Note that $p_0, \ldots, p_{\ell-1}$ are locations of events, while p_ℓ is the size vector!)

The abbreviations for $[\tilde{p}_i]$ that were defined before in the context of cellular events are also allowed for situations. So we can write $[(t, x)]$ as $[t, x]$, and $[0, x]$

as $[x]$. Sometimes we will use a "mixed" path notation, with not all $[0]$-terms omitted.

As an example of how this works, let us look at a cellular automaton with state set $\Sigma = \{0, 1\}$. We assume that at time $t = 0$ the cells at position $x = 0, 1, 2$ and 3 are in the states 1, 0, 0 and 1. We now want to express this information with a situation. For this we start with a cellular process. We know the cellular events $[0, 0]1$, $[0, 1]0$, $[0, 2]0$ and $[0, 3]1$. Using the abbreviation convention for the positions of cellular events, we can write the process that contains them as

$$\pi = \{[0]1, [1]0, [2]0, [3]1\}. \tag{4.8}$$

A natural way to write π as a situation in the form (4.5) is

$$a = (([0]1, [1]0, [2]0, [3]1), (0, 4)). \tag{4.9}$$

In a situation, the events of π must be arranged in a sequence; we have here chosen the most natural one, an arrangement from left to right by their x-positions. For the size vector of the situation we have chosen the point $(0, 4)$, one position to the right of the last event in the event sequence of a. If then a is written in the long path notation (4.6), it becomes

$$a = [0]1[0]0[0]0[0]1[0]. \tag{4.10}$$

We now see that the choice of a in (4.9) was natural: all displacements in the new notation become $[0]$. We can remove them all, and this leads to the short path notation for a, namely 1001. (The similarity between the short path notation and the notation for finite sequences is intended.)

4.2 Products

The path notation leads to a natural definition for the product of two situations. For the concatenation of situations in the following definition we employ the convention that $[p][q] = [p + q]$ that was already introduced for cells: Here is where it becomes useful.

***Definition* 4.4** (Product)**.** Let $a, b \in \mathscr{S}$ be two situations. We get their *product* by concatenating the long path notation for a with the long path notation for b. The product exists if the resulting expression is a situation.

The product of a and b is written ab.

To understand this definition, let

$$a = [\tilde{p}_0]\alpha_0 \dots \alpha_{\ell-1}[\tilde{p}_\ell] \qquad \text{and} \qquad b = [\tilde{q}_0]\beta_0 \dots \beta_{m-1}[\tilde{q}_m] \tag{4.11}$$

be two processes in long path notation. Then their product, if it exists, has the long path notation

$$\begin{aligned} ab &= [\tilde{p}_0]\alpha_0 \dots \alpha_{\ell-1}[\tilde{p}_\ell][\tilde{q}_0]\beta_0 \dots \beta_{m-1}[\tilde{q}_m] \\ &= [\tilde{p}_0]\alpha_0 \dots \alpha_{\ell-1}[\tilde{p}_\ell + \tilde{q}_0]\beta_0 \dots \beta_{m-1}[\tilde{q}_m]. \end{aligned} \tag{4.12}$$

The first line in this equation is that what we get when we simply concatenate the path notations for a and b. The second line is that what we get after applying the convention.

We now return to the pair notation for situations to find the process of ab. We assume that a is as in (4.5), and b is similar, so that we have

$$a = (([p_0]\alpha_0, \dots, [p_{\ell-1}]\alpha_{\ell-1}), p_\ell), \tag{4.13}$$
$$b = (([q_0]\beta_0, \dots, [q_{m-1}]\beta_{m-1}), q_m)\,. \tag{4.14}$$

In the second equation the q_i are related to the $\tilde{q}_i$ in the same way as the p_i to the $\tilde{p}_i$ in Definition 4.3. Then the product of a and b has the form

$$\begin{aligned} ab = (([p_0]\alpha_0, \dots, [p_{\ell-1}]\alpha_{\ell-1}, \\ [p_\ell + q_0]\beta_0, \dots, [p_\ell + q_{m-1}]\beta_{m-1}), p_\ell + q_m)\,. \end{aligned} \tag{4.15}$$

We can then translate this formula into a lemma that describes the product in a more abstract form.

***Lemma* 4.5** (Properties of the Product)**.** *Let $a, b \in \mathscr{S}$ be two situations. If* $\mathrm{pr}(a)$ *is compatible to* $\mathrm{pr}_a(b)$*, then the product ab exists. Its process and size vector are*

$$\mathrm{pr}(ab) = \mathrm{pr}(a) \cup \mathrm{pr}_a(b) \qquad \text{and} \qquad \delta(ab) = \delta(a) + \delta(b)\,. \tag{4.16}$$

Proof. The event sequence in (4.15) contains the events in $\mathrm{pr}(a)$ together with the events in $\mathrm{pr}(b)$, but the latter shifted by the size vector p_ℓ of a. The set of these events is therefore $\mathrm{pr}(a) \cup [\delta(a)]\mathrm{pr}(b)$, or $\mathrm{pr}(a) \cup \mathrm{pr}_a(b)$ in the notation of (4.4). It is a cellular process if $\mathrm{pr}(a)$ is compatible to $\mathrm{pr}_a(b)$, and if this is true, then ab is a situation.

The equations in (4.16) can then be read directly from (4.15). □

The left equation in (4.16) is the *chain rule* for situations. It is the reason why the notation $\mathrm{pr}_a(b)$ was introduced in (4.4).

Together with the product for situations we get the usual notations that are related to it. Among them are exponentiation, Kleene closure and other constructions that were already described for sequences. In contrast to ordinary sequences we must however be careful whether a product actually exists. The set of situations therefore does not form a semigroup. Nevertheless it has a neutral element of multiplication. As we can infer from (4.12), it is the situation $[0]$. We will speak of it as the *empty situation*.

A very important subset of $\mathscr{P}$ is the set $\{\,[0]\sigma\colon \sigma \in \Sigma\,\}^*$. It contains all those situations that, when written down in short path notation, look like elements of Σ^*. Therefore we will introduce no special symbol for them but call this set of situation also Σ^*. It will be always clear from the context which set is meant.

As a means to distinguish the elements of Σ^* from other situations we will use for them, and only for them, the length notation $|\cdot|$ of finite sequences.

Induction Proofs with Situations. The product of situations is also important because it allows induction proofs. Assume e. g. that a property P of situations is "closed under non-empty multiplication": This shall mean that if two non-empty situations $a, b \in \mathscr{S}$ have the property P and ab exist, then it has property P. Then if all situations in Σ^* have property P, all elements of $\mathscr{S}$ have it. This is in fact an induction over the length of the situations.

This specific form of induction requires however that the factors a and b are nonempty, which is not always easy to check. Therefore we will use another, more complex induction principle.

Before we can express it, we have to handle an ambiguity of the notation for situations. We will have to split a situation s into the product of the three terms a, $[p]$ and b. But because of the convention that $[p_1][p_2]$ is equal to $[p_1 + p_2]$, we could also split s into $a[p - q]$, $[q]$ and b, for any $q \in \mathbb{Z}$. So we cannot unambiguously say that the displacement $[p]$ is a factor of the situation s. This problem is solved with the help of the long path notation (4.6): in it the terms $[\tilde{p}_i]$ are unambiguous.

***Definition* 4.6** (Honest Decomposition)**.** Let $s \in \mathscr{S}$ be a situation with the long path notation $s = [p_0]\sigma_0 \dots \sigma_{\ell-1}[p_\ell]$. A decomposition $s = a[p]b$ of s is *honest* if there is an index i such that

$$a = [p_0]\sigma_0 \dots \sigma_{i-1}[0], \qquad p = p_i, \qquad b = [0]\sigma_i \dots \sigma_{\ell-1}[p_\ell]. \tag{4.17}$$

The cases of $a = [0]$ or $b = [0]$ are here explicitly allowed. They refer to the one-sided decompositions $s = [p]b$ and $s = a[p]$.

With this definition we can now describe the new induction principle.

***Theorem* 4.7** (Induction over Displacements)**.** *Let P be a property of the elements of $\mathscr{S}$. Assume that for all $s \in \mathscr{S}$,*

1. *if $s \in \Sigma^*$, then P is true for s,*
2. *if $s = a[p]b$ is a honest decomposition with $p \neq (0, 0)$ and P is true for a and b, then P is true for s.*

Then P is true for all elements of $\mathscr{S}$.

Proof. For this proof we define the *number of nontrivial displacements* in a situation s as the number of displacements $[p_i]$ with $p_i \neq (0, 0)$ that occur in its long path notation. We write this number as $d(s)$. The proof is then an induction over $d(s)$.

If $d(s) = 0$, then $s \in \Sigma^*$, and P is true for s. If $d(s) > 0$, then a term $[p]$ with $p \neq (0, 0)$ occurs in the long path notation of s. So there is a honest decomposition $s = a[p]b$. We then have $d(s) > d(a) + d(b)$, so P is true for a and b by induction. Therefore P is also true for s. □

4.3 Reactions

A reaction represents a computation in a cellular automaton. It consists of two situations that are related to each other by a cellular process. The first situation is the *input* of the computation. Its events start the activity of the cellular automaton; the activity itself is represented by the cellular process that is the closure of the input situation; the *result* of the computation is represented by the second situation of the reaction. Its process must lie completely inside the closure of the input process: This means that all events of the output are determined by the input via the transition rule. Its size vector must be the same as that of the input situation: This will allow us to replace the input of a reaction with its output when the input is part of a larger situation.‡ Otherwise the choice of the second situation is arbitrary.

Reactions were introduced in [51]. They have their name from the arrow with which reactions are written here, because it reminds of chemical reactions.

‡See below at Theorem 4.11 for more details.

***Definition* 4.8** (Reactions)**.** Let φ be a transition rule for Σ and $a, b \in \mathscr{S}$ be two situations. If

$$\mathrm{pr}(b) \subseteq \mathrm{cl}\,\mathrm{pr}(a) \qquad \text{and} \qquad \delta(a) = \delta(b), \tag{4.18}$$

then the pair (a, b) is a *reaction* for φ. For (a, b) we will usually write $a \to b$.

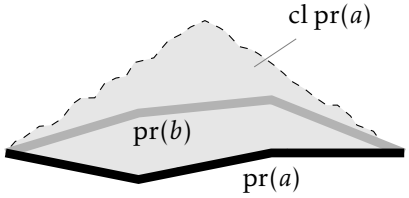


Figure 4.2: The processes involved in the reaction $a \to b$.

Sometimes we will use the more general term of an *abstract reaction*. This is a pair (a, b) with $\delta(a) = \delta(b)$ in which a is compatible with b.

We will also use the formula $a \to b$ as a proposition. Then it expresses the fact that there is a reaction (a, b). It may be a reaction for φ or an abstract reaction, depending on the context.

The processes that belong to this reaction are shown in Figure 4.2. Often, when the situations that are part of a reaction are complex, they will be drawn separately, as in Figure 4.3 below. This diagram is annotated with the names of the situations and not the processes, as in the previous figure: we will choose whichever is appropriate.

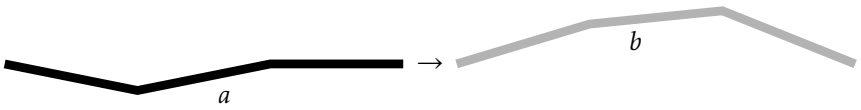


Figure 4.3: Another way to display the reaction of Figure 4.2.

Sets of Reactions. In order to be able to calculate with them, we will now consider reactions that belong to a set. For this let $S \subseteq \mathscr{S}$ be a set of situations and $R \subseteq S \times S$ a set of reactions between its elements. The set S is then the *domain of reaction setdomain* of R.

We use a special notation for reactions that belong to a set. If $(a, b) \in R$, we write this as $a \to_R b$. If R is known from context, we may write it even as $a \to b$. As before, an expression $a \to_R b$ may be used as the proposition. It then means that the pair (a, b) is an element of R.

As a set of pairs, R is a binary relation on S. The reaction sets that we use for the understanding of cellular automata are mainly pre-orders. In order theory, a binary relation is called a *pre-order* if it is transitive and reflexive [11, p. 2]. We recapitulate what this means:

1. R is transitive if $a \to_R c$ whenever $a \to_R b$ and $b \to_R c$.

2. R is reflexive if $a \to_R a$ for all $a \in R$.

In a set of reaction that is a pre-order, reflexivity allows to reconstruct the domain by

$$\operatorname{dom} R = \{\, a \in \mathscr{S} : a \to_R a \,\}. \tag{4.19}$$

Therefore for a set of reactions that is a pre-order it is not necessary to specify the domain separately from the reaction set.

There is another small fact that is useful in its own right: If a set R of reactions is a pre-order, then every situation in its domain has a closure. This is because for every $a \in \operatorname{dom} R$ there is a reaction $a \to_R a$, by the transitivity of R, and the definition of reactions requires that a then has a closure.

Reaction Systems. Transitivity of a reaction set allows to form a chain of reactions, each using as input the result of the previous one, and combine them into a single reaction that computes the result of the last reaction in the chain from the input of the first one. We now introduce another way to create new reactions from old ones, one that is specific to cellular automata. It reflects the local nature of the interactions between the cells.

***Definition* 4.9** (Application of a Reaction)**.** We call the reaction $xay \to xby$, where a, b, x and y are situations, the *application* of $a \to b$ on xay.

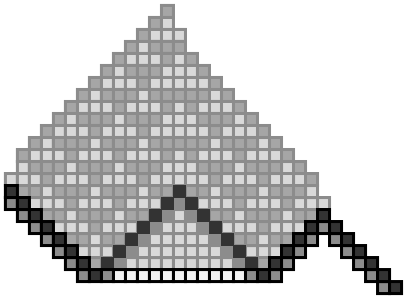


Figure 4.4: Application of a triangle reaction under Rule 54.

One example for the application of the reaction $a \to b$ to the situation xay is shown in Figure 4.4. The events of $\operatorname{pr}(xay)$ are displayed as squares with thick frames, like ▣, and the events of $\operatorname{pr}(xby)$ are displayed in darker colours, like ▣. The events displayed in lighter colours belong to the closure of $\operatorname{pr}(xay)$. In both colour sets, the lighter and the darker, the brighter squares represent cells in state 0 and the darker squares, cells in state 1. One can also see from this diagram that the process $\operatorname{pr}(xby)$ is a subset of the closure of $\operatorname{pr}(xay)$.

The reaction $a \to b$, displayed in the centre of the diagram, is a triangle reaction.[§] The base line of its triangle consists of the squares with frames but in lighter colours at the bottom of the diagram; the events belonging to it are all in state 0. The other two sides of the triangle are part of $\operatorname{pr}_x(b)$; they are shown as squares with frames, but with lightly coloured interior, and they are partially in the states 0 and 1.

[§]See also Figure 8.2. This kind of reaction will be formally defined later, in Definition 8.5.

There is also an operation of *applying a reaction* to a situation: It generates a reaction $xay \to xby$ from a reaction $a \to b$ and a situation xay. We will show now that under reasonable conditions on x, y, a and b application is always possible. This is done in two steps, because the definition of application in the form it was stated above uses too many variables at once. Instead of working with a reaction between situations that consist of three factors, we will first work with reactions between products of two factors.

***Lemma* 4.10** (Parallel Processing)**.** *Let $a \to a'$ and $b \to b'$ be reactions for φ. Assume that* $\operatorname{cl}\operatorname{pr}(ab)$ *exists. Then $ab \to a'b'$ is a reaction for φ.*

Proof. It is clear that $\delta(ab) = \delta(a'b')$. So it remains to prove that $\operatorname{pr}(a'b') \subseteq \operatorname{cl}\operatorname{pr}(ab)$. We have $\operatorname{pr}(a'b') = \operatorname{pr}(a') \cup \operatorname{pr}_{a'}(b')$ by the chain rule (4.16). Therefore the proof of the lemma is complete if we show that $\operatorname{pr}(a') \subseteq \operatorname{cl}\operatorname{pr}(ab)$ and $\operatorname{pr}_{a'}(b') \subseteq \operatorname{cl}\operatorname{pr}(ab)$.

Because $a \to a'$ is a reaction, we have $\operatorname{pr}(a') \subseteq \operatorname{cl}\operatorname{pr}(a)$. Since $\operatorname{pr}(a) \subseteq \operatorname{pr}(ab)$, we have $\operatorname{cl}\operatorname{pr}(a) \subseteq \operatorname{cl}\operatorname{pr}(ab)$ by monotony of the closure (Theorem 3.14). Therefore $\operatorname{pr}(a') \subseteq \operatorname{cl}\operatorname{pr}(ab)$.

Because $b \to b'$ is a reaction, we have $\operatorname{pr}(b') \subseteq \operatorname{cl}\operatorname{pr}(b)$. The closure is shift-invariant, therefore $\operatorname{pr}_{a'}(b') \subseteq \operatorname{cl}\operatorname{pr}_{a'}(b)$. Now $\delta(a) = \delta(a')$ because $a \to a'$ is a reaction, so we have $\operatorname{pr}_{a'}(b') \subseteq \operatorname{cl}\operatorname{pr}_{a}(b)$. Since $\operatorname{pr}_{a}(b) \subseteq \operatorname{pr}(ab)$, we have $\operatorname{cl}\operatorname{pr}_{a}(b) \subseteq \operatorname{cl}\operatorname{pr}(ab)$, again by monotony of the closure. Therefore $\operatorname{pr}_{a'}(b') \subseteq \operatorname{cl}\operatorname{pr}(ab)$. □

In this proof the condition $\delta(a) = \delta(a')$ played a crucial role in keeping the processes $\operatorname{pr}_{a}(b)$ and $\operatorname{pr}_{a'}(b')$ at the same position. This is why it appeared in the definition of reactions.

***Theorem* 4.11** (Applying Creates a Reaction)**.** *Let φ be a transition rule for Σ. If $a \to b$ is a reaction for φ and there are $x, y \in \mathscr{S}$ for which* $\operatorname{cl}(xay)$ *exists, then $xay \to xby$ is a reaction for φ. (Figure 4.5.)*

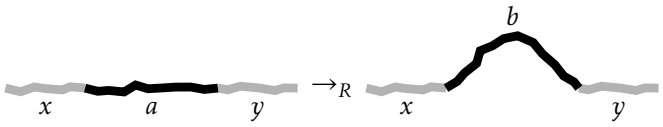


Figure 4.5: Applying $a \to b$ to xay.

Proof. Since $\operatorname{cl}\operatorname{pr}(xay)$ exists, the processes $\operatorname{cl}\operatorname{pr}(x)$ and $\operatorname{cl}\operatorname{pr}_{ax}(y)$ exist by Theorem 3.14. This means that $x \to x$ and $y \to y$ are reactions for φ.

By Lemma 4.10, $xa \to xb$ is a reaction for φ because $x \to x$ and $a \to b$ are, and $xay \to xby$ is a reaction for φ because $xa \to xb$ and $y \to y$ are. □

Now we will introduce a name for the property of a set of reactions that the operation of application in it is freely possible. Note that in its definition there are no explicit restrictions on the situations x and y at the sides of a; there is however the implicit restriction that xay must be an element of R.

***Definition* 4.12** (Closed under Application)**.** Let Σ be a set. Let R be a set of reactions with $\operatorname{dom} R \subseteq \mathcal{S}$.

If for all reactions $a \to_R b$ and for all situations $x, y \in \mathcal{S}$, with $xay \in R$ there is a reaction $xay \to_R xby$, then R is *closed under application*.

Theorem 4.11 also expresses that the reactions in R are local in scope: When R is closed under application, it depends only on the initial situation a and not on the situations x and y around it, whether a reaction $xay \to_R xby$ is possible.

Now we can finally introduce the central concept of this thesis. Reaction systems will serve as a replacement of the evolution sequence defined in Chapter 3 for the understanding of cellular automata.

***Definition* 4.13** (Reaction System)**.** Let φ be a transition rule. A *reaction system* for φ is a set of reactions for φ that is a pre-order and closed under application.

Similarly, an *abstract reaction system* is a set of abstract reactions that is a pre-order and closed under application.

The set of all reactions for a given transition rule is obviously a reaction system, but we will usually need smaller ones. The operation of applying a reaction to a situation will allow us to define a large set of reactions with the help of a small set of local reactions. Therefore we define now how a small set of reactions and a set of situations together generate a reaction system. As it is common with generated sets in mathematics, a large part of the work with generated reaction systems is about deriving properties of the whole system from those of the set of generators.

***Definition* 4.14** (Generated Reaction System)**.** Let $S \subseteq \mathcal{S}$ be a set of situations and G a set of reactions.

Let R be the smallest reaction system with $S \subseteq \operatorname{dom} R$ and $G \subseteq R$ (i. e. no proper subset of R has this property). Then R is the *reaction system generated by* G from S. The set S is the set of *generating situations* for R, and G is the set of *generating reactions* for R.

We are mainly interested in non-abstract reaction systems. An abstract reaction system is usually created from a reaction system for a rule φ in order to have a system that is easier to handle.

With Theorem 4.11 we see that a reaction system for φ can be generated from an arbitrary reaction set G and a set S of situations for which the only requirement is that all its elements must have a closure under φ. If we know this, we can work with the reaction system in a quasi-algebraic way, without referring to the closure again.

4.4 Summary

In this chapter we have introduced situations and reactions. They are, in a manner of speaking, the substantives and basic propositions of the new language. Much effort has been done to establish an intuitive notation for situations.

We have then seen how to construct larger situations from smaller situations by multiplication, and larger reactions from smaller reactions by the concatenation of applications. This made it possible to define a reaction system

in terms of a small number of situations and reactions. A small set of generating reactions then defines the reactions of a large set of situations, in the same way as the local transition rule defines the behaviour of a cellular automaton.

Chapter 5

Interval-preserving Automata

In this and the next chapter we will show how to construct a reaction system for a one-dimensional cellular automata from its transition rule. But since the behaviour of cellular automata varies greatly, we will consider here a subclass for which it is not too complex in a geometrical sense. This subclass of *interval-preserving* rules contains however the complex elementary cellular automata rule 54 and the computationally universal rule 110 [6]; therefore no restriction on the computational complexity of cellular automata is apparent if one restricts one's view to interval-preserving transition rules.

5.1 Intervals

Before we can describe what interval preservation shall mean, we must define intervals and develop a notation for them and their arrangement in space-time. We will develop it first in the context of cellular processes and then, a bit later, for situations.

An interval consists of a finite number of cells that are positioned without a gap. In the space-time viewpoint of cellular processes it is also bound to a specific moment in time. We define intervals together with a notation for their domain.*

***Definition* 5.1** (Intervals)**.** An *interval domain* at time $t \in \mathbb{Z}$ is a set of points of the form

$$I_t(i, j) = \{ (t, x) : i \leq x < j \} \tag{5.1}$$

with $i, j \in \mathbb{Z}$ and $i \leq j$ (Figure 5.1). An *interval process* at time t is a cellular process whose domain is an interval domain.

Thus the set $I_t(i, j)$ stands for the interval domain at time t that reaches from the cell position i to the cell position j (but excludes it). We allow that an

*We do not use a square bracket notation analogous to the notation $[i, j]$ for intervals on the real line: This would lead to too much optical confusion with the other uses of square brackets in this text.

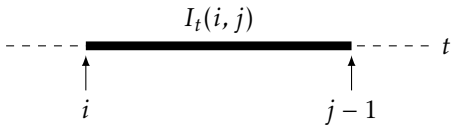


Figure 5.1: An interval process at time t.

interval is empty: This happens if an interval has a domain of the form $I_t(i, j)$ with $i = j$. In this case the time t is no longer determined by the set $I_t(i, j)$. We will then use the convention that the empty set is an interval at any time.

Now consider two compatible intervals that belong to the same time. A nice property of intervals is that there is only a limited number of ways in which these intervals can lie with respect to each other. Three kinds of spatial arrangement are especially important. In the following definition we will introduce notations for them. One arrangement occurs if neither of the two intervals is a proper subset of the other one: then one of them must be at the left of the other one, if we allow overlap. The other two arrangements occur when one of the intervals is the left or the right end of the other interval.

While the three notations below are intended especially for the use with intervals, their definitions are meaningful for any cellular process.

***Definition* 5.2** (Spatial Arrangement of Processes)**.** Let $\pi, \psi \in \mathscr{P}$ be two processes.

1. π is *left of* ψ, written $\pi <> \psi$, if
   (i) for all $p \in \operatorname{dom} \pi$ there is a $\xi \geq 0$ such that $p + \xi X \in \operatorname{dom} \psi$,
   (ii) for all $q \in \operatorname{dom} \psi$ there is a $\xi \geq 0$ such that $q - \xi X \in \operatorname{dom} \pi$, and
   (iii) π is compatible with ψ.
2. ψ is a *left extension* of π, written $\psi \supseteq_L \pi$, if
$$\psi \supseteq \pi \qquad \text{and} \qquad \psi <> \pi\,. \tag{5.2}$$
3. ψ is a *right extension* of π, written $\pi \subseteq_R \psi$, if
$$\pi \subseteq \psi \qquad \text{and} \qquad \pi <> \psi\,. \tag{5.3}$$

The expression "left of" is here used in an inclusive sense, such that always $\pi <> \pi$. The relation $\pi <> \psi$ is always true when π or ψ are empty, and $\psi \supseteq_L \pi$ and $\pi \subseteq_R \psi$ are always true when π is empty. If two nonempty interval processes are related by $<>$, $\supseteq_L$ or $\subseteq_R$, they always occur at the same time.

In the case of two intervals the relations of definition 5.2 have an especially simple form (Figure 5.2). To see this, let π_1 and π_2 be two intervals at time t with $\operatorname{dom} \pi_1 = I_t(i_1, j_1)$ and $\operatorname{dom} \pi_2 = I_t(i_2, j_2)$. Then,

$$\pi_1 <> \pi_2 \quad \text{iff} \quad \pi_1 \operatorname{comp} \pi_2 \quad \text{and} \quad i_1 \leq i_2,\ j_1 \leq j_2, \tag{5.4a}$$
$$\pi_1 \supseteq_L \pi_2 \quad \text{iff} \quad \pi_1 \supseteq \pi_2 \quad \text{and} \quad i_1 = i_2,\ j_1 \leq j_2, \tag{5.4b}$$
$$\pi_1 \subseteq_R \pi_2 \quad \text{iff} \quad \pi_1 \subseteq \pi_2 \quad \text{and} \quad i_1 \leq i_2,\ j_1 = j_2\,. \tag{5.4c}$$

Another connection between the relations in Definition 5.2 is the following lemma. It describes $<>$ in terms of $\subseteq_R$ and $\supseteq_L$.

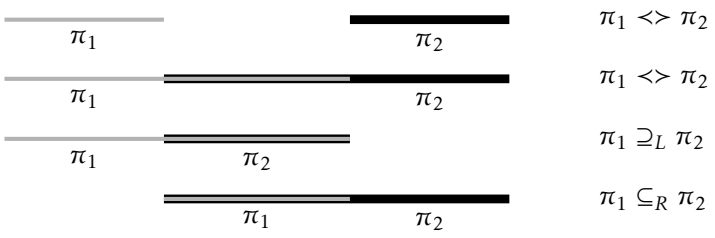


Figure 5.2: Spatial orientation of intervals.

***Lemma* 5.3.** *Let $\pi, \psi \in \mathscr{P}$ be two cellular processes. Then $\pi <> \psi$ is equivalent to $\pi \subseteq_R \pi \cup \psi \supseteq_L \psi$.*

Proof. It can be seen directly from the definition that $\pi \subseteq_R \pi \cup \psi$ is equivalent to $\pi <> \pi \cup \psi$, and that $\pi \cup \psi \supseteq_L \psi$ is equivalent to $\pi \cup \psi <> \psi$. Therefore $\pi \subseteq_R \pi \cup \psi \supseteq_L \psi$ is true if and only if $\pi <> \pi \cup \psi <> \psi$.

Assume $\pi <> \pi \cup \psi <> \psi$: If $p \in \operatorname{dom} \pi$, then there is a $\xi \geq 0$ with $p + \xi X \in \operatorname{dom} \psi$ because $\pi \cup \psi <> \psi$, and if $p \in \operatorname{dom} \psi$, then there is a $\xi \geq 0$ with $p - \xi X \in \operatorname{dom} \pi$ because $\pi <> \pi \cup \psi$; together this shows $\pi <> \psi$.

Assume $\pi <> \psi$: Then π and ψ are compatible, and therefore the process $\pi \cup \psi$ exists. To check whether $\pi <> \pi \cup \psi$ is true, we only have to check that for $p \in \operatorname{dom} \psi$ there is a $\xi \geq 0$ such that $p - \xi X \in \operatorname{dom} \pi$, but that is true because $\pi <> \psi$. The same way we can show that $\pi \cup \psi <> \psi$. Together this proves that $\pi <> \pi \cup \psi <> \psi$. □

Interval Situations. Since a sequence $u \in \Sigma^*$ of cell states is interpreted as a situation, its process $\operatorname{pr}(u)$ is an interval process. More general, the process $\operatorname{pr}([t, x]u)$ has the domain $I_t(x, x + |u|)$. Therefore every interval process can be written as $\operatorname{pr}([p]u)$ with an appropriate $p \in \mathbb{Z}^2$ and $u \in \Sigma^*$. This leads to the following definition for the set of situations that represent interval processes.

***Definition* 5.4** (Interval Situations)**.** An *interval situation* with states in Σ is a situation $[p]u \in \mathscr{S}$ with $p \in \mathbb{Z}^2$ and $u \in \Sigma^*$.

For the following calculation we will need a notation that mirrors the notations for interval processes in the language of situations.

First we introduce a notation for the left and right ends of a situation, in analogy to $\subseteq_R$ and $\supseteq_L$. If a is the left or right end of x, then it is a factor of it; therefore I have chosen symbols for these concepts that remind of division operators.[†]

***Definition* 5.5** (Left and Right Factors)**.** Let a and $x \in \mathscr{S}$ be situations.

If there is a situation x' such that $ax' = x$, then a is a *left factor* of x. We will write this as $a \setminus\setminus x$.

If there is a situation x' such that $x = x'a$, then a is a *right factor* of x. We will write this as $x /\!/ a$.

[†]This notation is also influenced by the alternative notation $m \setminus n$ for "m divides n" by Knuth, Graham and Patashnik [19, p. 102].

Next we need a notation for overlapping situations. Here I have chosen the symbol $\langle b \rangle$, in analogy to the symbol $\diamond$ for the overlapping of two strings that is used by Harold V. McIntosh [44, p. 216].

***Definition* 5.6** (Overlap)**.** Let $b \in \mathscr{S}$ be a situation. Then the displacement $\langle b \rangle = [-\delta(b)]$ is the *overlap operator* for b.

This notation is subject to a convention: We will use $\langle b \rangle$ only for a situation that contains b as a factor. This means that:

1. If $x \langle b \rangle y \in \mathscr{S}$, then $x /\!/ b$ and $b \setminus\!\setminus y$,
2. if $x\langle b \rangle \in \mathscr{S}$, then $x /\!/ b$, and
3. if $\langle b \rangle y \in \mathscr{S}$, then $b \setminus\!\setminus y$.
4. The term $\langle b \rangle$ will never be used at its own to indicate a situation.

With the convention for overlap operators we can state a *chain rule* for situations with overlap, a special case of the chain rule in (4.16). In contrast to the general case, this equation has also a kind of converse, with intersection instead of union:

$$\operatorname{pr}(x \langle b \rangle y) = \operatorname{pr}(x) \cup \operatorname{pr}_{x\langle b \rangle}(y), \tag{5.5a}$$

$$\operatorname{pr}_{x\langle b \rangle}(b) = \operatorname{pr}(x) \cap \operatorname{pr}_{x\langle b \rangle}(y)\,. \tag{5.5b}$$

In almost all cases where we use $\langle b \rangle$, the situation b will be an interval. If the situations x and y in the term $x\langle b \rangle y$ are also intervals, then the relations $<\!\!>$, $\supseteq_L$ and $\subseteq_R$ between their processes can be expressed by the overlap operator.

***Lemma* 5.7** (Overlap and Spatial Arrangement of Intervals)**.** *Let x, y and $b \in \Sigma^*$ be intervals. Then*

$$\operatorname{pr}(x) <\!\!> \operatorname{pr}_{x\langle b \rangle}(y) \quad \textit{iff} \quad x /\!/ b \setminus\!\setminus y, \tag{5.6a}$$

$$\operatorname{pr}(x) \supseteq_L \operatorname{pr}_{x\langle b \rangle}(b) \quad \textit{iff} \quad x /\!/ b, \tag{5.6b}$$

$$\operatorname{pr}(b) \subseteq_R \operatorname{pr}(y) \quad \textit{iff} \quad b \setminus\!\setminus y\,. \tag{5.6c}$$

Proof. These relations can be derived with the help of the equivalences in (5.4).

We will prove (5.6a) first. Assume that $\operatorname{pr}(x) <\!\!> \operatorname{pr}_{x\langle b \rangle}(y)$. Then $\operatorname{pr}(x)$ is compatible with $\operatorname{pr}_{x\langle b \rangle}(y)$, and therefore the union of these processes exists. This union is according to (5.5a) the process $\operatorname{pr}(x \langle b \rangle y)$. Therefore the situation $x \langle b \rangle y$ exists, and this is, according to our convention, equivalent to $x /\!/ b \setminus\!\setminus y$.

For the opposite direction, assume that $x /\!/ b \setminus\!\setminus y$ is true and therefore $x \langle b \rangle y$ exists. We will use here the equivalence (5.4a). We have

$$\operatorname{dom} \operatorname{pr}(x) = I_0(0, |x|), \qquad \operatorname{dom} \operatorname{pr}_{x\langle b \rangle}(y) = I_0(|x| - |b|, |x| - |b| + |y|). \tag{5.7}$$

To apply (5.4a), we have to show that $\operatorname{pr}(x)$ is compatible with $\operatorname{pr}_{x\langle b \rangle}(y)$ and that $0 \le |x| - |b|$ and $|x| \le |x| - |b| + |y|$. The first condition is true because $\operatorname{pr}(x) \cup \operatorname{pr}_{\langle b \rangle}(y)$ exists. The second condition is equivalent to $|x| \ge |b| \le |y|$. It is true because b is, according to the convention for $\langle b \rangle$, the common part of x and y. Therefore $\operatorname{pr}(x) <\!\!> \operatorname{pr}_{x\langle b \rangle}(y)$.

The other two equivalences are special cases of the first one, with $y = b$ or $x = b$ and are proved in a similar way. □

Borrowing an Interval. We will use the overlap operator for making situations and reactions more readable. We will split a situation into an equivalent situation that consists of overlapping parts; then we apply reactions to the parts and put the parts together again.

For this procedure we need a notion of equivalence under which two equivalent situations initiate the same computation. The following definition does this.

***Definition* 5.8** (Equivalent Situations)**.** Let $a, b \in \mathscr{S}$ be two situations. If

$$\mathrm{pr}(a) = \mathrm{pr}(b) \qquad \text{and} \qquad \delta(a) = \delta(b), \tag{5.8}$$

we say that a is *equivalent* to b and write it as $a \sim b$.

With this definition we can express $x \mathbin{/\!/} b$ as $x \sim x \langle b \rangle b$ and $b \mathbin{\backslash\!\backslash} y$ as $y \sim b \langle b \rangle y$. We will use this equivalence from time to time to split situations into overlapping parts.

To see that equivalent situations cause the same reactions we note that if $a \sim b$ and the closure of $\mathrm{pr}(a)$ exists, then there is a reaction $a \to b$, with no other requirements on the transition rule. So if $b \to x$ is a reaction for φ, then $a \to x$ is also a reaction for φ. Equivalence is symmetric, therefore the converse is also true and the set of situations that start from a is the same as the set of reactions that start from b.

The following derivation then illustrates the work with overlapping situations: Assume that the reaction system R contains a reaction $by \to_R by'$ and that there is a situation $x \in \mathrm{dom}\, R$ that ends with b. If also $xy \in \mathrm{dom}\, R$, then there is also a reaction $xy \to_R xy'$. We could prove this by introducing a situation x' such that $x = x'b$ and then applying the reaction $by \to_R by'$ to $x'by$, but there is a notationally shorter way: We will then instead say that $x \mathbin{/\!/} b$ and write the following chain of reactions, withot the need to introduce x'.

$$xy \sim x \langle b \rangle by \to x \langle b \rangle by' \sim xy'. \tag{5.9}$$

With this technique the descriptions of longer chains of reactions become considerably shorter. An example for it occurs in (6.8).

Note that this derivation only shows that $xy \to xy'$ is a reaction for φ, not that it belongs to R. This must be verified separately. Finding a reaction result is however often the more difficult part, especially if the derivation is long.

5.2 Interval Preservation

Like many concepts we need the concept of interval preservation in two forms. One is a global form that applies to a transition rule, the other a localised form that applies to a single interval. The localised form is defined for situations and not processes, because that is the form where we need it in Lemma 5.11.

***Definition* 5.9** (Interval Preservation)**.** A transition rule φ for Σ is *interval-preserving* if for all interval processes $\pi \in \mathscr{P}$ the process $\Delta\pi$ is an interval.

Let $u \in \Sigma^*$ be an interval situation. If $\Delta\mathrm{pr}(u)$ is an interval, then φ is *interval-preserving for* u.

Interval-preserving rules are never constant functions: If φ is constant, then the domain of $\Delta\pi$ is $\mathbb{Z}^2$ for every process π, and it is therefore never an interval. So we do not specify explicitly for interval-preserving rules that they are non-constant.

The simple behaviour of interval-preserving rules, announced at the beginning of this chapter, becomes visible when we look at the closure of an interval process.

***Lemma* 5.10** (Closure of an Interval)**.** *If a process is an interval, then its closure exists under an interval-preserving transition rule and all its time slices are then intervals.*

Proof. Let π be an interval process at time t_0 and φ the transition rule. Then π is quiet before t_0, and we have $\mathrm{cl}^{(t)}\,\pi = \emptyset$ for $t < t_0$ and $\mathrm{cl}^{(t_0)}\,\pi = \pi^{(t_0)}$: these time slices are intervals.

If $t > t_0$, then $\pi^{(t)} = \emptyset$ and equation (3.21) becomes $\mathrm{cl}^{(t)}\,\pi = \Delta\,\mathrm{cl}^{(t-1)}\,\pi$. This means that if $\mathrm{cl}^{(t-1)}\,\pi$ is an interval, then $\mathrm{cl}^{(t)}\,\pi$ is also an interval because φ is interval-preserving. Therefore all $\mathrm{cl}^{(t)}\,\pi$ with $t > t_0$ are intervals by induction, and $\mathrm{cl}\,\pi$ exists. □

We can see an example for such a closure in Figure 5.3. It is part of a larger background process, the same process as in Figure 3.1. The initial interval is shown by the squares ◻■, its closure consists of ◻■■■, and the whole background process of all kinds of squares, ◻■■■■.

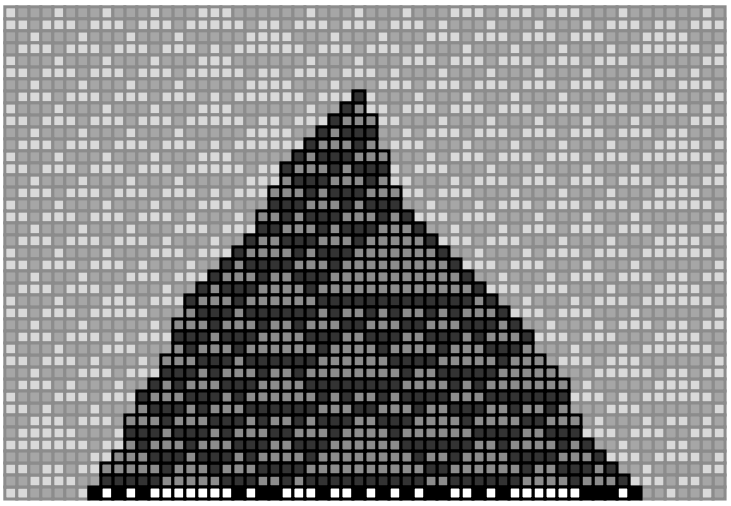
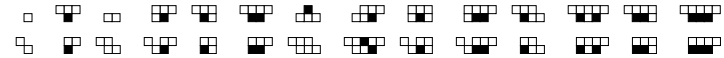
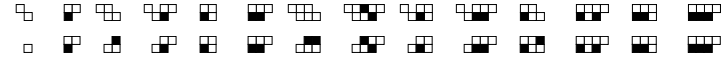
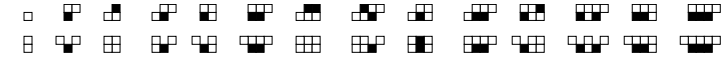
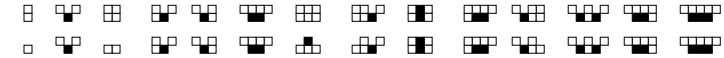
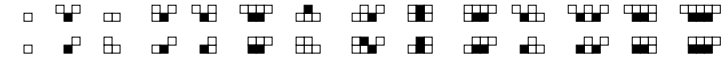
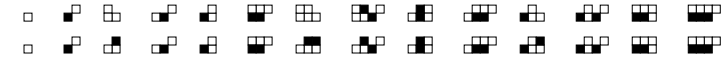
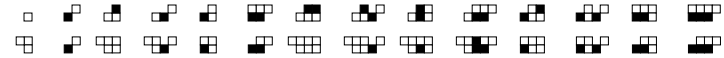
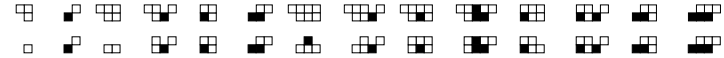
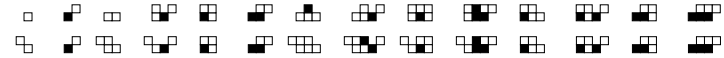
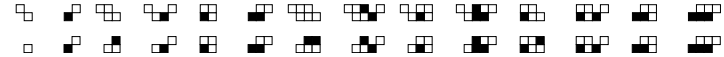
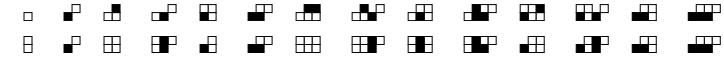
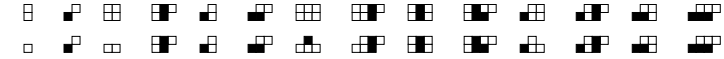
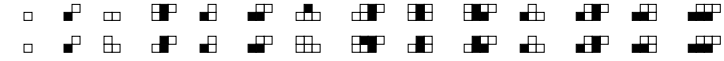
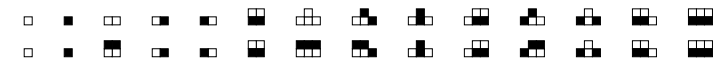
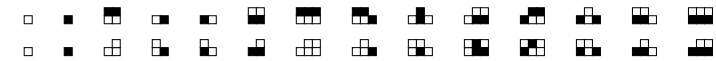
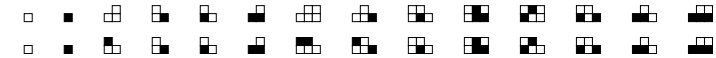
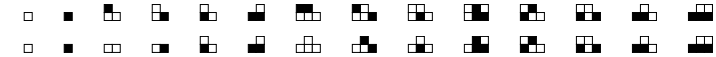
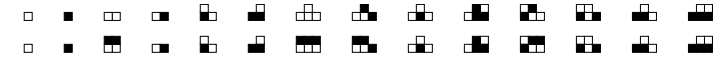
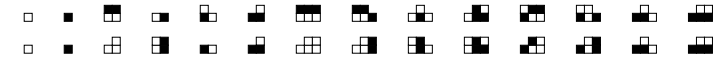
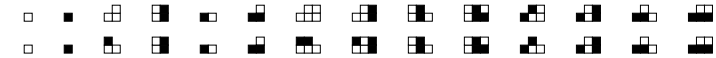
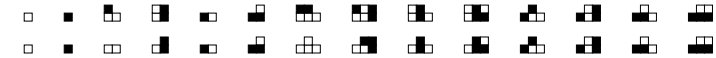
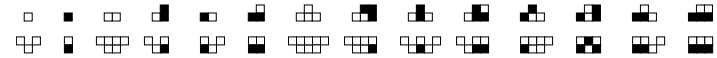
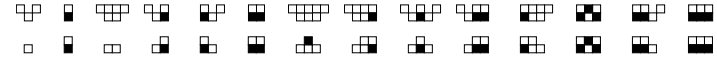
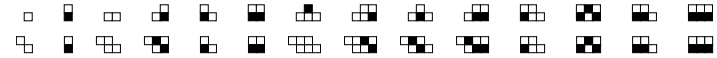
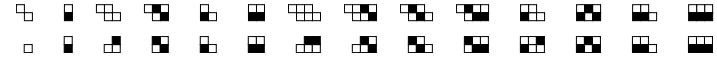
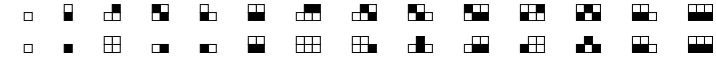
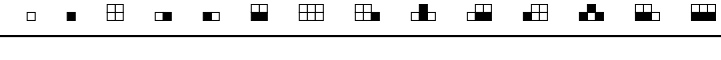


Figure 5.3: The closure of an interval under Rule 54 as part of the evolution of a random initial configuration.

Testing for Interval Preservation. The following theorem shows that it is possible to determine in finite time whether a transition rule is interval-preserving.

Table 5.1: Tests for interval preservation (Part 1).

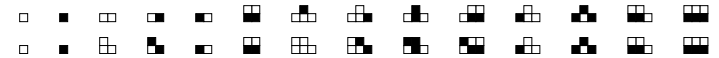
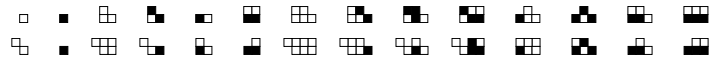
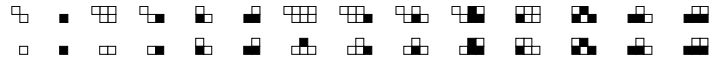
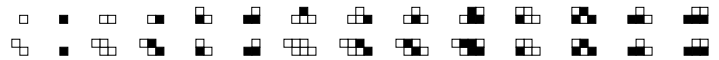
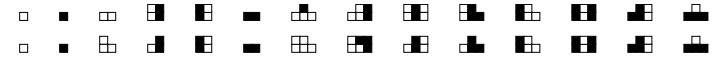
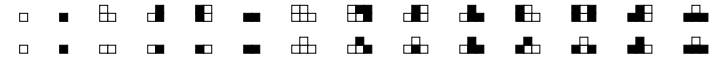
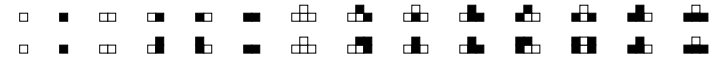
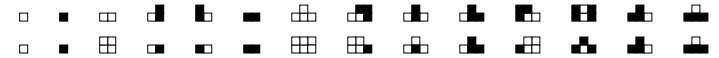
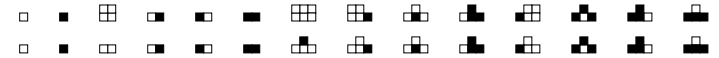
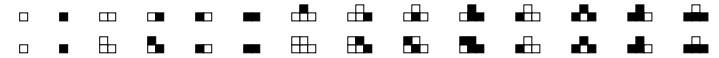
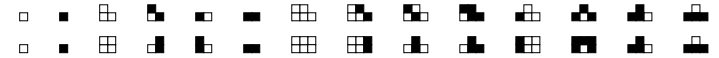
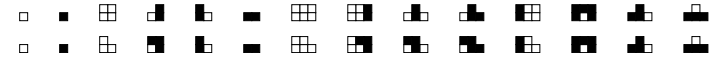
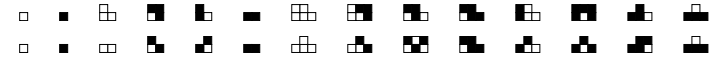
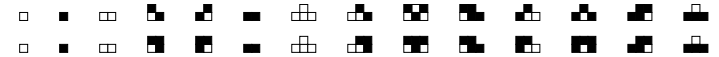
Rule	Evolution	IP
1		
2		
3		
4		no
5		no
6		no
7		no
8		no
9		no
10		no
11		no
12		
13		
14		
15		
18		
19		
22		
23		
24		
25		
26		
27		
28		
29		
30		
32		no
33		
34		
35		
36		

***Theorem* 5.11** (Interval Preservation is Local)**.** *A transition rule φ of radius r is interval-preserving if and only if it is interval-preserving for all $u \in \Sigma^*$ with $|u| \leq 2r + 1$.*

With this theorem we can easily find out which of the elementary cellular automata are interval-preserving. This is done in Table 5.1–5.3. Each row in this table describes one transition rule. It contains a list of evolution diagrams, one for each interval of maximal length 3. If a transition rule is interval-preserving, the top rows of all its diagrams must be intervals.

As before, only one of the maximally four equivalent transition rules is shown. Rule 0 is omitted because it is constant and therefore cannot be interval-preserving. We then find that most of the elementary cellular automata are interval-preserving: Therefore, for better visual recognition only the non-preserving rules are marked with a "no" in the last column. We can then

Table 5.2: Tests for interval preservation (Part 2).

Rule	Evolution	IP
37		
38		
40		no
41		
42		
43		
44		
45		
46		
50		
51		
54		
56		
57		
58		
60		
62		
72		
73		
74		
76		
77		
78		
90		
94		
104		
105		
106		
108		
110		
122		
126		

conclude that all elementary cellular automata except 0, 4–11, 32, 40, 130, 138, 160 and 168 are interval-preserving.

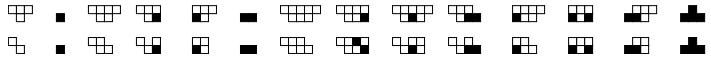
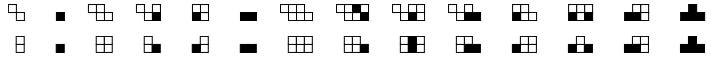
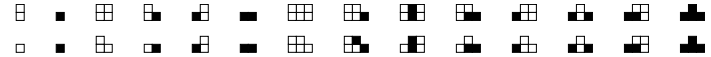
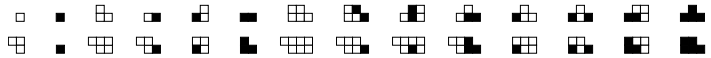
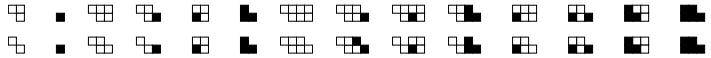
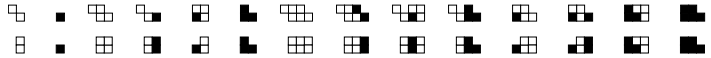
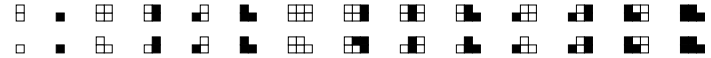
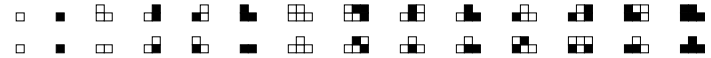
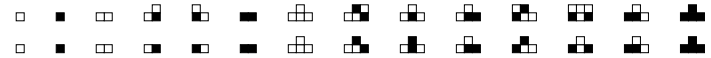
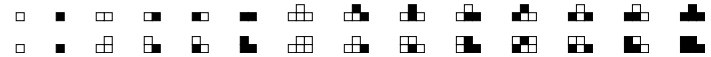
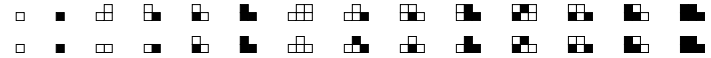
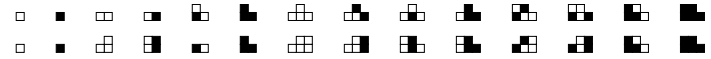
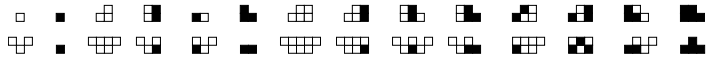
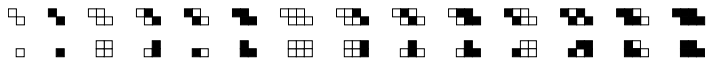
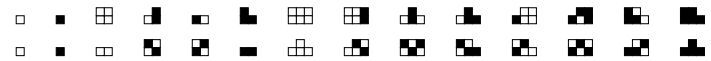
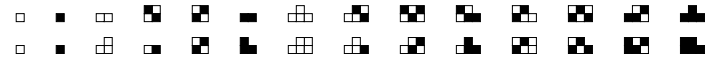
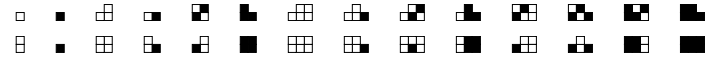
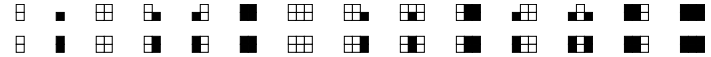
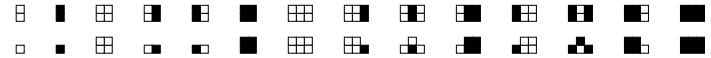
The tables here were generated by a program; a more detailed description how one can check by hand whether a transition rule is interval-preserving appears in Section 7.2.

Proof of Theorem 5.11. We need only to prove that if φ is interval-preserving for all $u \in \Sigma^*$ with $|u| \leq 2r+1$, then it is also interval-preserving for every $v \in \Sigma^*$ with $|v| > 2r+1$.

Let us therefore write $v = v_0 \dots v_{\ell-1} \in \Sigma^\ell$ with $\ell \geq 2r+1$ and let

$$\psi_i = \mathrm{pr}(v)|_{I_0(i-r,i+r)} \tag{5.10}$$

Table 5.3: Tests for interval preservation (Part 3).

Rule	Evolution	IP
128		
130		no
132		
134		
136		
138		no
140		
142		
146		
150		
152		
154		
156		
160		no
162		
164		
168		no
170		
172		
178		
184		
200		
204		
232		

be, for $-r \leq i < \ell + r$, a sequence of maximally $2r + 1$ events of $\mathrm{pr}(v)$, centred at $(0, i)$. Then every point that is determined by $\mathrm{pr}(v)$ is determined by some of the ψ_i, since interval-preserving rules are non-constant. There are three different shapes of ψ_i, depending on i:

$$\psi_i = \begin{cases} \mathrm{pr}([0]v_0 \dots v_{i+r}) & \text{if } -r \leq i < r, \\ \mathrm{pr}([i-r]v_{i-r} \dots v_{i+r}) & \text{if } r \leq i < \ell - r, \\ \mathrm{pr}([i-r]v_{i-r} \dots v_{\ell-1}) & \text{if } \ell - r \leq i < \ell + r. \end{cases} \tag{5.11}$$

Let us call them *left*, *central* and *right* ψ_i (Figure 5.4). If $-r \leq i < r$, then $\psi_i \subseteq \psi_r$, so all events that are determined by a left ψ_i are already determined by ψ_r. If $\ell - r \leq i < \ell + r$, then $\psi_i \subseteq \psi_{\ell-1-r}$, so all events determined by a right ψ_i are already determined by $\psi_{\ell-1-r}$. Therefore all events determined by $\mathrm{pr}(v)$ are determined by at least one of the central ψ_i.

All central ψ_i are intervals of length $2r + 1$. Under a rule with radius r, a central ψ_i therefore determines the point $(1, i)$. Since φ is interval-preserving, the interval determined by ψ_i must therefore contain an event at $(1, i)$. The set of events determined by $\mathrm{pr}(v)$ is a union of such intervals, one for each i with $r \leq i < \ell - r$, and is therefore itself an interval. Therefore φ is also interval-preserving for all u with $|u| \geq 2r + 1$. □

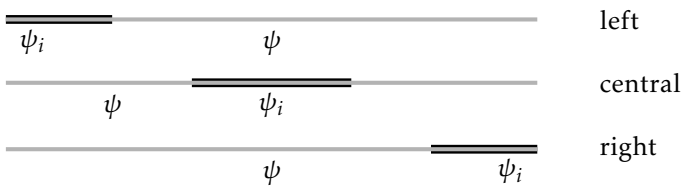


Figure 5.4: The different shapes of ψ_i.

5.3 Separating Intervals

Let now φ be interval-preserving. We will now introduce a method to compute the states of cells at the next time step only for parts of an interval. For this we need as a technical tool a special kind of intervals, the "separating intervals", that serve as a boundary between the different parts of an interval.

***Definition* 5.12** (Separating Intervals)**.** An interval process $\pi \in \mathscr{P}$ is *separating* if for all processes ψ, ψ_1, $\psi_2 \in \mathscr{P}$ with $\psi = \psi_1 \cup \psi_2$ and $\psi_1 \supseteq_L \pi \subseteq_R \psi_2$,

$$\Delta\psi_1 \cup \Delta\psi_2 = \Delta\psi \qquad \text{and} \qquad \Delta\psi_1 \cap \Delta\psi_2 = \Delta\pi\,. \tag{5.12}$$

An interval situation $a \in \mathscr{S}$ is separating if $\mathrm{pr}(a)$ is separating.

Note here that if $\psi_1 \supseteq_L \pi$ and $\pi \subseteq_R \psi_2$, then $\psi_1 \cap \psi_2 = \pi$. So an equivalent formulation of (5.12) is that under the conditions of Definition 5.12,

$$\Delta\psi_1 \cup \Delta\psi_2 = \Delta(\psi_1 \cup \psi_2) \quad \text{and} \quad \Delta\psi_1 \cap \Delta\psi_2 = \Delta(\psi_1 \cap \psi_2)\,. \tag{5.13}$$

With separating intervals we can split an interval process into parts and evolve them independently. They solve the following question: Assume that we have an interval ψ that consists of two parts, ψ_1 and ψ_2, such that $\psi = \psi_1 \cup \psi_2$. (The parts may overlap.) Is it then possible to compute $\Delta\psi_1$ and $\Delta\psi_2$ independently such that their union is $\Delta\psi$, or does something get lost? The answer is: Yes, it is possible, if there is a separating interval π with $\psi_1 \supseteq_L \pi \subseteq_R \psi_2$.

By the first part of (5.12), no element of $\Delta\psi$ is omitted by the separate computation from ψ_1 and ψ_2, while the second part requests that the contents of $\Delta\psi_1$ and $\Delta\psi_2$ are independent of each other. It is actually a restricted form of independence because the process π is subset of ψ_1 and of ψ_2. The elements of $\Delta\pi$ can therefore be computed from both processes, but no other element of $\Delta\psi_1$ can be computed from ψ_2 and no other element of $\Delta\psi_2$ can be computed from ψ_1 because of the second part of (5.12).

In Definition 5.12, the processes ψ_1 and ψ_2 are arranged from left to right. We naturally should then expect that then $\Delta\psi_1$ too is at the left of $\Delta\psi_2$. The following lemma proves this for ψ_1 and ψ_2 that are intervals. It also shows that if $\Delta\psi_1$ and $\Delta\psi_2$ overlap and $\psi_1 <> \psi_2$, then the right boundary of $\Delta\psi_1$ and the left boundary of $\Delta\psi_2$ depend only on π.

***Lemma* 5.13** (Order Preservation for Overlapping Intervals)**.** *Let φ be an interval-preserving transition rule for Σ and let π, ψ_1, $\psi_2 \in \mathscr{P}$ be three intervals with $\psi_1 \supseteq_L \pi \subseteq_R \psi_2$, where π is a separating interval. Then $\Delta\psi_1 <> \Delta\psi_2$.*

Moreover, if $\Delta\psi_1$ and $\Delta\psi_2$ are nonempty and

$$\operatorname{dom}\Delta\psi_1 = I_t(i_1', j_1'), \qquad \operatorname{dom}\Delta\psi_2 = I_t(i_2', j_2'), \tag{5.14}$$

then $i_1' \le i_2' \le j_1' \le j_2'$, and the numbers i_2' and j_1' depend only on π.

The difficulty in the following proof is that $\Delta\pi$ may be empty, so we cannot prove the lemma by proving that $\Delta\psi_1 \supseteq_L \Delta\pi \subseteq_R \Delta\psi_2$. Instead we must work with the interval boundaries.

Proof. We will assume that the four processes ψ_1, ψ_2, $\Delta\psi_1$ and $\Delta\psi_2$ are all nonempty, because otherwise the lemma would be trivially true.

We will now view π as a constant and ψ_1 and ψ_2 as variables. For all the other quantities in the proof we will keep track whether they depend on ψ_1 or ψ_2. Let now

$$\operatorname{dom}\psi_1 = I_t(i_1, j_1), \qquad \operatorname{dom}\psi_2 = I_t(i_2, j_2)\,. \tag{5.15}$$

Then the numbers $i_1, j_1, \ldots, i_2', j_2'$ in (5.15) and (5.14) are uniquely determined by ψ_1 respectively ψ_2.

Because we have $\psi_1 \supseteq_L \pi \subseteq_R \psi_2$, we must have $i_1 \le i_2 \le j_1 \le j_2$. The intersection of ψ_1 and ψ_2 is π, therefore $\operatorname{dom}\pi = I_0(i_2, j_1)$, so i_2 and j_1 are constants. This also means that if ψ_1 is a large set, i_1 must be a small number, and if ψ_2 is large, then j_2 is a large number. The set $\Delta\pi$ may be empty, so we cannot give an expression for $\operatorname{dom}\Delta\pi$ similar to those in (5.14). But the number of elements in $\Delta\pi$ is a well-defined constant, and we will call it ℓ.

We have always

$$\operatorname{dom}\Delta\psi_1 \supseteq I_{t+1}(i_1 - r, j_1 + r), \qquad \operatorname{dom}\Delta\psi_2 \subseteq I_{t+1}(i_2 + r, j_2 - r)\,. \tag{5.16}$$

Assume now that $j_2 > j_1 + 2r$, where r is the radius of φ. Then the rightmost elements of $\Delta\psi_1 \cup \Delta\psi_2$ must belong to $\Delta\psi_2$. This means that $\Delta\psi_1$ is at the left of $\Delta\psi_2$ and they overlap at ℓ events. So we must have $i_1' \le i_2' \le j_1' \le j_2'$ and $i_2' + \ell = j_1'$. If we keep ψ_2 fixed but let ψ_1 vary arbitrarily, the condition $i_2' + \ell = j_1'$ must always be true. This means that j_1' stays the same for all values of ψ_1, and because it only depends on ψ_1, it must be a constant.

The same way, by assuming $i_1 < i_2 - 2r$, we can see that i_2' is a constant.

So we have always $i_2' \le j_1'$, and i_2' and j_1' are constants. For the validity of $\Delta\psi_1 \cap \Delta\psi_2 = \Delta\pi$ it is then necessary that $i_1' \le i_2' \le j_1' \le j_2'$ is true in general, which means that $\psi_1 <> \psi_2$. □

A corollary of this lemma shows what happens if we split π into two intervals.

***Corollary* 5.14** (Separation by Bounded Intervals)**.** *Let φ be an interval-preserving transition rule for Σ.*

Let $\pi_1, \pi_2 \in \mathscr{P}$ be separating intervals and $\psi_1, \psi_2 \in \mathscr{P}$ be intervals with

$$\psi_1 \supseteq_L \pi_1 <> \pi_2 \subseteq_R \psi_2\,. \tag{5.17}$$

Then $\Delta\psi_1 <> \Delta\psi_2$.

Proof. Let π be an arbitrary interval process that reaches from the left end of π_1 to the right end of π_2, such that we have $\pi_1 \subseteq_R \pi \supseteq_L \pi_2$.

Then $\pi \cup \psi_2$ is an interval, and we have $\psi_1 \supseteq_L \pi_1 \subseteq_R \pi \cup \psi_2$. So we can apply Lemma 5.13 and get $\Delta\psi_1 <> \Delta(\pi \cup \psi_2)$. Because π_2 is separating and $\pi \supseteq_L \pi_2 \subseteq_R \psi_2$, we have $\Delta(\pi \cup \psi_2) = \Delta\pi \cup \Delta\psi_2$. We derive from the resulting relation $\Delta\psi_1 <> \Delta\pi \cup \Delta\psi_2$ with Lemma 5.3 the relation $\Delta\psi_1 \subseteq_R \Delta\psi_1 \cup \Delta\pi \cup \Delta\psi_2$, and this leads to $\Delta\psi_1 \subseteq_R \Delta\psi_1 \cup \Delta\psi_2$.

The same way we can also prove $\Delta\psi_1 \cup \Delta\psi_2 \supseteq_L \Delta\psi_2$. These two relations together are equivalent to $\Delta\psi_1 <> \Delta\psi_2$, again by Lemma 5.3. □

The Set of Separating Intervals. All this assumes that separating intervals exist. We need to make that certain and would also like to get an overview about which intervals are separating. This we will do in two steps. First we will prove that being a separating interval is a monotone property: an interval that contains a separating interval as a subset is itself separating. Then we will show that under a rule of radius r, every interval of exactly $2r$ cells is separating. From this we can then conclude that every interval of at least $2r$ cells is separating.

***Lemma* 5.15** (Being Separating is Monotone)**.** *Let φ be an interval-preserving transition rule and let $\pi \subseteq \pi' \in \mathscr{P}$ be two intervals. If π is separating, then π' is separating too.*

Proof. We will first prove the lemma for the case that $\pi \subseteq_R \pi'$. For this, let ψ_1', ψ_2', $\psi' \in \mathscr{P}$ be any processes with $\psi_1' \supseteq_L \pi' \subseteq_R \psi_2'$ and $\psi_1' \cup \psi_2' = \psi'$. Let $\psi_1 = (\psi_1' \setminus \pi') \cup \pi$; then $\psi_1 \supseteq_L \pi \subseteq_R \psi_2'$ and $\psi_1 \cup \psi_2' = \psi'$.

We have $\Delta\psi_1' = \Delta(\psi_1 \cup \pi') = \Delta\psi_1 \cup \Delta\pi'$ because π is separating. Therefore,

$$\begin{aligned}
\Delta\psi_1' \cup \Delta\psi_2' &= \Delta\psi_1 \cup \Delta\pi' \cup \Delta\psi_2' \\
&= \Delta\psi_1 \cup \Delta\psi_2' && \text{because } \Delta\pi' \subseteq \Delta\psi_2' \\
&= \Delta(\psi_1 \cup \psi') && \text{because } \psi_1 \supseteq_L \pi \subseteq_R \psi_2' \\
&= \Delta\psi'. && \text{because } \psi_1 \subseteq \psi
\end{aligned} \tag{5.18}$$

On the other hand,

$$\begin{aligned}
\Delta\psi_1' \cap \Delta\psi_2' &= (\Delta\psi_1 \cup \Delta\pi') \cap \Delta\psi_2' \\
&= (\Delta\psi_1 \cap \Delta\psi_2') \cup \Delta\pi' && \text{because } \Delta\pi' \subseteq \Delta\psi_2' \\
&= \Delta\pi \cup \Delta\pi' && \text{because } \psi_1 \supseteq_L \pi \subseteq_R \psi_2' \\
&= \Delta\pi'. && \text{because } \Delta\pi \subseteq \Delta\pi'
\end{aligned} \tag{5.19}$$

This proves the lemma in the case that $\pi \subseteq_R \pi'$.

The same kind of argument works when $\pi' \supseteq_L \pi$. In the general case we note that if $\pi \subseteq \pi'$, then there is always an interval process π'' such that $\pi \subseteq_R \pi''$ and $\pi' \supseteq_L \pi''$. This reduces the general case to the two other cases. □

***Lemma* 5.16** (Existence of Separating Intervals)**.** *Under a transition rule with radius r, every interval that consists of at least $2r$ events is separating.*

The proof makes use of the fact that determinedness is a local property: if θ is a process, then a point p is determined by θ if and only if it is determined by $\theta|_{N(p-T,r)}$. We have seen this in Lemma 3.9. The set $N(p - T, r)$ is the neighbourhood domain of p for the previous time step, defined in (3.13).

Proof. Let π be the separating interval and let ψ, ψ_1 and ψ_2 be as in Definition 5.12.

We know already that $\Delta\psi_1 \cup \Delta\psi_2 \subseteq \Delta\psi$ and $\Delta\pi \subseteq \Delta\psi_1 \cap \Delta\psi_2$ because Δ is monotone (Lemma 3.13). So it remains to prove

$$\Delta\psi \subseteq \Delta\psi_1 \cup \Delta\psi_2 \qquad \text{and} \qquad \Delta\psi_1 \cap \Delta\psi_2 \subseteq \Delta\pi\,. \tag{5.20}$$

Let now $p \in \operatorname{dom}\Delta\psi$ be an arbitrary point and let N_p stand for $N(p - T, r)$. The proof of (5.20) then relies on the fact that N_p is an interval domain of length $2r + 1$, but $\operatorname{dom}\pi$ has at least $2r$ points. The points of $N_p \setminus \operatorname{dom}\pi$ must therefore be either completely at the left or completely at the right of π. In the first case we have $(\psi_2 \setminus \pi)|_{N_p} = \emptyset$, in the second, $(\psi_1 \setminus \pi)|_{N_p} = \emptyset$. Therefore,

$$\psi_2|_{N_p} = \pi|_{N_p} \qquad \text{or} \qquad \psi_1|_{N_p} = \pi|_{N_p}\,. \tag{5.21}$$

Assume first that $p \in \operatorname{dom}\Delta\psi$. Because $\psi|_{N_p}$ is equal to $\psi_1|_{N_p} \cup \psi_2|_{N_p}$, and with equation (5.21), the process $\psi|_{N_p}$ must either be equal to $\pi|_{N_p} \cup \psi_2|_{N_p} = \psi_2|_{N_p}$ or $\psi_1|_{N_p} \cup \pi|_{N_p} = \psi_1|_{N_p}$ or both. In the first case, $p \in \operatorname{dom}\Delta\psi_1$, in the second case, $p \in \operatorname{dom}\Delta\psi_2$. This proves $\Delta\psi \subseteq \Delta\psi_1 \cup \Delta\psi_2$.

Assume now that $p \in \operatorname{dom}(\Delta\psi_1 \cap \Delta\psi_2)$. Then p depends on $\psi_1|_{N_p}$ and on $\psi_2|_{N_p}$. One of these processes is equal to $\pi|_{N_p}$ by (5.21), therefore $p \in \operatorname{dom}\Delta(\pi|_{N_p})$. This proves $\Delta\psi_1 \cap \Delta\psi_2 \subseteq \Delta\pi$. □

A transition rule may also have separating intervals of less than $2r$ elements. So to get an overview about the separating intervals of a specific transition rule, we should know the set of its *minimal separating intervals*. We will actually need three kinds of minimal intervals, defined below.

***Definition* 5.17** (Minimal Separating Interval)**.** A separating interval process is *left minimal* if no events can be removed from its right side without making it non-separating. It is *right minimal* if no events can be removed from its left side without making it non-separating. It is *minimal* if it is both left minimal and right minimal.

Left and right minimal intervals occur as the boundaries of a separating interval: If ψ is a separating interval, then there are intervals π_1 and π_2 such that $\pi_1 \subseteq_R \psi \supseteq_L \pi_2$. The processes π_1 and π_2 can be viewed as the "left and right end" of ψ. The shortest interval π_1 that is still separating is then a left minimal interval, and the shortest separating interval π_2 is a right minimal interval.

Example: The Elementary Cellular Automata. For the elementary cellular automata there exists a simple test to find the minimal separating intervals. The main reason for this is the small radius of their transition rules. This means, with Lemma 5.16, that we only need to check whether there are intervals of length 0 and 1 that are separating.

The case of a separating interval of length 0 does indeed occur, and it means that the cells never interact. There are two rules that have this property: Rule 51, which lets the cells alternate between the states 0 and 1, and the identity function, Rule 204.

For the other rules we must check whether there are intervals that consist of a single event and are separating. We will call such an interval π and place it at the origin; it consists of the single event $[0, 0]\sigma$.

As we have seen in the proof of Lemma 5.16, we need only to verify that for all cellular processes ψ_1, ψ_2 with $\psi_1 \supseteq_L \pi \subseteq_R \psi_2$ we have $\Delta(\psi_1 \cup \psi_2) \subseteq \Delta\psi_1 \cup \Delta\psi_2$ and $\Delta\psi_1 \cap \Delta\psi_2 \subseteq \Delta\pi$ in order to prove that π is a separating interval.

We will use the convention that at time 0, the cell at position i is in state ξ_i; we have then $\sigma = \xi_0$. For an arbitrary event at time 1 we will write e: its state is η and its location, x, such that we have $e = [1, x]\eta$. Since φ has radius 1, the state of the event e can only depend on ξ_{x-1}, ξ_x and ξ_{x+1}: This will be important in the following derivation.

Let now e be element of $\Delta(\psi_1 \cup \psi_2)$. Then, if $x < 0$, we have $e \in \Delta\psi_1$ and if $x > 1$, we have $e \in \Delta\psi_2$. The remaining case, $x = 0$, is the key to finding a necessary condition for π: If the state of e depends both on events in ψ_1 and ψ_2, then e cannot be an element of $\Delta\psi_1 \cup \Delta\psi_2$, and π cannot be separating. This can only occur when $\varphi(\xi_{-1}, \sigma, \xi_1)$, the state of e, depends on both ξ_{-1} and ξ_1. Therefore a necessary condition for π being separating is that such a dependency does not happen.

We can ensure that by requiring that at least one of the following two equations is true:

$$\forall \xi_{-1} \in \Sigma\colon\ \varphi(\xi_{-1}, \sigma, 0) = \varphi(\xi_{-1}, \sigma, 1), \tag{5.22a}$$

$$\forall \xi_1 \in \Sigma\colon\ \ \varphi(0, \sigma, \xi_1) = \varphi(1, \sigma, \xi_1)\,. \tag{5.22b}$$

Here the first equation means that the cell to the right has no influence on the next state of a cell in state σ, and the second, that the cell at the left has no influence on the next state.

The conditions (5.22) are also sufficient. To show this, we will first prove that $\Delta(\psi_1 \cup \psi_2) \subseteq \Delta\psi_1 \cup \Delta\psi_2$. In the proof we will assume that $e \in \Delta(\psi_1 \cup \psi_2)$. As we have seen before, if $x < 0$, then $e \in \Delta\psi_1$ and if $x > 1$, then $e \in \Delta\psi_2$. In the remaining case of $x = 0$, at least one of the two conditions in (5.22) must be true. If (5.22a) is true, then η depends only on ξ_{-1} and σ, and therefore $e \in \Delta\psi_1$. If (5.22b) is true, then η depends only on σ and ξ_1, and $e \in \Delta\psi_2$. This shows that $\Delta(\psi_1 \cup \psi_2) \subseteq \Delta\psi_1 \cup \Delta\psi_2$.

Next we prove that $\Delta\psi_1 \cap \Delta\psi_2 \subseteq \Delta\pi$. We will assume that $e \in \Delta\psi_1 \cap \Delta\psi_2$. Since φ has radius 1, the only possible values for x are then -1, 0 and 1. In the case of $x = -1$ we use the fact that e is an element of $\Delta\psi_2$. This means that we have $\varphi(\xi_{-2}, \xi_{-1}, \sigma) = \eta$ for all ξ_{-2}, $\xi_{-1} \in \Sigma$ and that therefore $e \in \Delta\pi$. In the case of $x = 0$, the condition $e \in \Delta\psi_1$ requires that (5.22a) is true, and $e \in \Delta\psi_2$ requires that (5.22b) is true. Taken together, the two conditions imply that $\varphi(\xi_{-1}, \sigma, \xi_1) = \eta$ for all ξ_{-1}, $\xi_1 \in \Sigma$ and that we have here again $e \in \Delta\pi$. The case $x = 1$ can be handled in a similar way as $x = -1$. This shows that $\Delta\psi_1 \cap \Delta\psi_2 \subseteq \Delta\pi$.

We therefore have now proved that (5.22) are necessary and sufficient conditions that the interval π is separating.

The tests (5.22) have been done by a program for all interval-preserving elementary cellular automata different from 51 and 204. The results are shown in Table 5.4. If none of the length 1 intervals are separating, then the intervals of length 2 are the minimal separating intervals. If an interval π of length 1 is

Table 5.4: Separating intervals of the interval-preserving elementary cellular automata.

Intervals	Rules
□□ ◨ ◧□ ■	22, 23, 24, 26, 27, 36, 37, 41, 43, 54, 57, 58, 73, 74, 77, 78, 90, 94, 104, 105, 108, 122, 126, 134, 142, 146, 150, 156, 164, 172, 178, 232.
□ ■	25, 38, 42, 56, 72, 76, 106, 110, 128, 132, 140, 162, 200.
■ □□	1, 2, 13, 14, 18, 19, 28, 30, 33, 35, 44, 45, 50, 62, 152, 154.
□ ■	3, 12, 15, 29, 34, 46, 60, 136, 170, 184
[0]	51, 204.

minimally separating, then only those intervals of length 2 are separating that do not contain π. So if, e. g., 0 is a separating interval but 1 is not separating, then the only minimal separating interval is 11. This argument explains the second row of Table 5.4; the other rows are explained similarly.

In Section 7.2 and in the context of Rule 54, we will see a related method to find the minimal separating intervals of a transition rule. It uses the characteristic reactions, which are defined next.

5.4 Characteristic Reactions

We will now introduce the concept of *characteristic reactions* as a way to express the relation between an interval π and the set $\Delta\pi$ in the language of situations and reactions. The characteristic reaction specifies the location and the dimensions of $\Delta\pi$ in relation to π in a kind of shorthand.

We will introduce it in two steps. The first step is for the case that $\Delta\pi$ is nonempty. Then the notion of the location of $\Delta\pi$ has an obvious meaning. If π is separating and it is part of a longer interval, then the cells of $\Delta\pi$ serve as a separator between the events determined by the cells left of π from those determined by the cells right of π. This property is important for proofs about the behaviour of reaction systems, so we would like to have it for all of the separating intervals. We will use it in the second step to extend the notion of the characteristic reaction such that for a separating interval π we can speak of the "location" of $\Delta\pi$ even then when $\Delta\pi$ is empty.

The characteristic reactions themselves will however not become part of the final reaction system; they are only a tool to define it.

Construction of the Characteristic Reactions. We can express the property of a rule φ to be interval-preserving in the following way: If $a \in \Sigma^*$, then there are $i \in \mathbb{Z}$ and $\hat{a} \in \Sigma^*$ such that $\Delta\mathrm{pr}(a) = \mathrm{pr}([1, i]\hat{a})$. The interval situation $\hat{a}$ is then always uniquely determined by a. So we define

***Definition* 5.18** (Determined Interval)**.** Let φ be an interval-preserving transition rule for Σ and let $a \in \Sigma^*$. Then the situation $\hat{a} \in \Sigma^*$ for which there is an $i \in \mathbb{Z}$ such that

$$\Delta\mathrm{pr}(a) = \mathrm{pr}([1, i]\hat{a}), \tag{5.23}$$

is the *determined interval* of a under φ.

The "hat" accent of $\hat{a}$ should remind of the operator Δ.

In contrast to $\hat{a}$, the number i is only then uniquely determined when $\hat{a} \neq [0]$, because only then $\Delta\mathrm{pr}(a) \neq \emptyset$. In this case we can express the relation between $\mathrm{pr}(a)$ and $\Delta\mathrm{pr}(a)$ by a reaction. This definition is important because it specifies the numbers i and j, which will be needed later to define the actual reaction system.

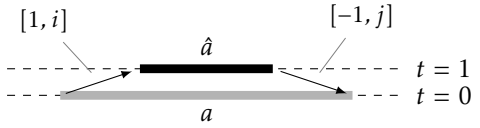


Figure 5.5: The two sides of the characteristic reaction (5.24), overlayed.

***Definition* 5.19** (Characteristic Reactions, Preliminary Form)**.** Let φ be an interval-preserving transition rule for Σ. Let $a \in \Sigma^*$ be an interval with $\hat{a} \neq [0]$ and let $\Delta\mathrm{pr}(a) = \mathrm{pr}([1, i]\hat{a})$. Then the reaction

$$a \to [1, i]\hat{a}[-1, j] \tag{5.24}$$

with $j = |a| - i - |\hat{a}|$, is the *characteristic reaction* for a under φ. (Figure 5.5.)

Now we extend this definition to interval situations a for which $\hat{a}$ can be empty—but in this case a must be a separating interval. We can do this by using other separating intervals as "test functions".

The following lemma is there to show that the final definition of separating intervals extends the preliminary definition.

***Theorem* 5.20** (Intervals as Reactions)**.** *Let φ be an interval-preserving transition rule for Σ and $a \in \Sigma^*$ a separating interval for φ.*

Then there exist $i, j \in \mathbb{Z}$ such that for all $x \in \Sigma^*$ with $\widehat{xa} \neq [0]$ there is an $i' \in \mathbb{Z}$ such that the characteristic reaction for xa is

$$xa \to [1, i']\widehat{xa}[-1, j], \tag{5.25a}$$

and for all $y \in \Sigma^*$ with $\widehat{ay} \neq [0]$ there is a $j' \in \mathbb{Z}$ such that the characteristic reaction for ay is

$$ay \to [1, i]\widehat{ay}[-1, j']. \tag{5.25b}$$

If also $\hat{a} \neq [0]$, then a has the characteristic reaction $a \to [1, i]\hat{a}[-1, j]$, the same as in (5.24).

We will use in this proof that if $u \to [1, i]\hat{u}[-1, j]$ is a characteristic reaction, then $\mathrm{dom}\,\Delta\mathrm{pr}(u) = I_1(i, |u| - j)$: The process $\Delta\mathrm{pr}(u)$ reaches from i cells to the left of $\mathrm{pr}(u)$ to j cells to the right of $\mathrm{pr}(u)$, one time step later.

Proof. To apply Lemma 5.13 we first define processes π, ψ_1 and ψ_2 with $\psi_1 \supseteq_L \pi \subseteq_R \psi_2$, namely

$$\psi_1 = \mathrm{pr}_{[-|x|]}(xa), \qquad \pi = \mathrm{pr}(a), \qquad \psi_2 = \mathrm{pr}(ay)\,. \tag{5.26}$$

In order to verify the inclusion $\psi_1 \supseteq_L \pi$ we note that we have $\mathrm{pr}_{[-|x|]}(xa) = \mathrm{pr}_{[-|x|]}(x) \cup \mathrm{pr}_{[-|x|]x}(a)$ with the chain rule, and $\mathrm{pr}_{[-|x|]x}(a) = \mathrm{pr}(a)$ because $\delta([-|x|]x) = (0,0)$. Therefore $\mathrm{pr}_{[-|x|]}(xa) = \mathrm{pr}(x) \cup \mathrm{pr}(a) \supseteq_L \mathrm{pr}(a)$, which means that $\psi_1 \supseteq_L \pi$. This explains also why the shift of $\mathrm{pr}(xa)$ to the left with $[-|x|]$ is necessary. The other inclusion, $\pi \subseteq_R \psi_2$, can also be verified by an application of the chain rule.

The domains of the processes ψ_1 and ψ_2 are then

$$\mathrm{dom}\,\psi_1 = I_0(-|x|, |a|), \qquad \mathrm{dom}\,\psi_2 = I_0(0, |ay|), \tag{5.27}$$

and the events determined by them are located at

$$\mathrm{dom}\,\Delta\psi_1 = I_1(-|x| + i', |a| - j), \qquad \mathrm{dom}\,\Delta\psi_2 = I_1(i, |ay| - j')\,. \tag{5.28}$$

as we can see from the characteristic reactions (5.25) and from the remark before the proof. For the process ψ_1 we have to keep in mind that it, and therefore also $\Delta\psi_1$, is shifted to the left by $|x|$ cells. Then we know by Lemma 5.13 that the boundaries of the intervals in (5.28) are arranged in the form $-|x| + i' \le i \le |a| - j \le |ay| - j'$ and that i and $|a| - j$ only depend on a. Therefore the variables i and j only depend on a.

Assume now that $\hat{a} \neq [0]$ and therefore $\Delta\pi \neq \emptyset$. Then $\mathrm{dom}\,\pi = I_0(0, |a|)$ and $\mathrm{dom}\,\Delta\pi = I_1(i, |a| - j)$, which means that the characteristic reaction for a (in the preliminary form of Definition 5.19) must be $a \to [1, i]\hat{a}[-1, j]$. □

This theorem then justifies the following definition of the characteristic reaction for separating intervals.

***Definition* 5.21** (Characteristic Reactions, Final Form)**.** Let $a \in \Sigma^*$ be a separating interval situation.

If there are intervals $x, y \in \Sigma^*$ with $\widehat{xa} \neq [0] \neq \widehat{ay}$, and they have the characteristic reactions $xa \to [1, i']\widehat{xa}[-1, j]$ and $ay \to [1, i]\widehat{ay}\,[-1, j']$, then the reaction

$$a \to [1, i]\hat{a}[-1, j] \tag{5.29}$$

is the *characteristic reaction* for a under φ.

The following lemma expresses the monotony of the closure in terms of characteristic reaction. It shows that in the previous definition $\hat{a}$ is always a part of $\widehat{xa}$ and $\widehat{ay}$, and how to recover it.

***Lemma* 5.22** (Reactions of Separating Intervals)**.** *Let $a \in \Sigma^*$ be a separating interval under an interval-preserving transition rule, and let $x, y \in \Sigma^*$.*

Then $\widehat{xa} \mathbin{/\!/} \hat{a}$ and $\hat{a} \mathbin{\backslash\!\backslash} \widehat{ay}$.

Proof. Let φ be the transition rule. We will prove only the first equivalence, the second one is its mirror image and the proof is similar.

Assume that the characteristic reaction of a and xa are

$$a \to [1, i]\hat{a}[-1, j] \qquad \text{and} \qquad xa \to [1, i']\widehat{xa}[-1, j]\,. \tag{5.30}$$

We will show first that $\mathrm{pr}_{x[1,i]}(\hat{a}) \subseteq \mathrm{pr}_{[1,i']}(\widehat{xa})$. We see from (5.30) that $\mathrm{pr}_{[1,i]}(\hat{a}) = \Delta\mathrm{pr}(a)$ and $\Delta\mathrm{pr}(xa) = \mathrm{pr}_{[1,i']}(\widehat{xa})$. With the chain rule we get $\mathrm{pr}_x(a) \subseteq \mathrm{pr}(xa)$,

by the monotony of the Δ operator, $\Delta\mathrm{pr}_x(a) \subseteq \Delta\mathrm{pr}(xa)$, and by putting these relations together we get

$$\mathrm{pr}_{x[1,i]}(\hat{a}) = \Delta\mathrm{pr}_x(a) \subseteq \Delta\mathrm{pr}(xa) = \mathrm{pr}_{[1,i']}(\widehat{xa}). \tag{5.31}$$

Next we show that $\delta(x[1,i]) = \delta([1,i']\widehat{xa}\langle\hat{a}\rangle)$. We see from the characteristic reaction for a that $i = |a| - |\hat{a}| - j$ and from the characteristic reaction for xa that $j = |xa| - i' - |\widehat{xa}|$. Therefore $i = |a| - |\hat{a}| - |xa| + i' + |\widehat{xa}| = i' + |\widehat{xa}| - |x| - |\hat{a}|$, and $\delta(x[1,i]) = (1, |x| + i) = (1, i' + |\widehat{xa}| - |\hat{a}|) = \delta([1,i']\widehat{xa}\langle\hat{a}\rangle)$. This means that

$$\begin{aligned}\mathrm{pr}_{[1,i']}(\widehat{xa}) &= \mathrm{pr}_{[1,i']}(\widehat{xa}) \cup \mathrm{pr}_{x[1,i]}(\hat{a}) \\ &= \mathrm{pr}_{[1,i']}(\widehat{xa}) \cup \mathrm{pr}_{[1,i']\widehat{xa}}(\langle\hat{a}\rangle\hat{a}) = \mathrm{pr}_{[1,i']}(\widehat{xa}\,\langle\hat{a}\rangle\,\hat{a}),\end{aligned} \tag{5.32}$$

which also means that $\widehat{xa} /\!\!/ \hat{a}$. □

5.5 Summary

In this chapter we were concerned with interval behaviour and the left-to-right arrangement of intervals. The goal of this was to find a subset of the transition rules that harmonise with the definition of reaction systems in the previous chapter.

We have seen that interval-preserving transition rules are such a subset. Interval preservation is a useful property because intervals already play an important role in cellular automata: The transition rule is expressed in terms of intervals. Intervals are conceptually simple cellular processes: Their domain can be expressed with three numbers, and their content can be expressed in a natural way as a sequence of cell states. They are therefore easy to express with situations. Interval *preservation* then also puts a limit on the complicatedness of the closure of an interval. Every time slice of it is an interval, so it does not become more complicated over time. The complexity of the behaviour of an interval-preserving cellular automaton is therefore confined to the interior of this closure.

A specific result of this chapter was the usefulness of *separating intervals*. They form the boundaries between different regions in a cellular automaton that do not influence each other in the next time step. This makes them useful for the selective evolution of an initial configuration of cells that we define in the next chapter.

We have learned how to express intervals and their interactions both in terms of processes and of situations. We have seen that it is possible to test for interval preservation of a transition rule in a finite number of steps. It is also possible to find the minimal separating intervals of a transition rule in a finite number of steps. As a technical tool to express the properties of separating intervals we have introduced characteristic reactions. They express which events are determined by a separating interval and where the zones of influence are for the cells at the left and the right of the separating interval.

Chapter 6

A Local Reaction System

Although they do characterise interval-preserving transition rules, characteristic reactions cannot be used as generators of a reaction system without unpleasant side effects.

I mean the following: Let $a_1 \to [1, i_1]\hat{a}_1[-1, j_1]$ and $a_2 \to [1, i_2]\hat{a}_2[-1, j_2]$ be the characteristic reactions of separating intervals. If we apply them in sequence to the situation $a_1 a_2$, then we get

$$\begin{aligned} a_1 a_2 &\to [1, i_1]\hat{a}_1[-1, j_1]a_2 \\ &\to [1, i_1]\hat{a}_1[-1, j_1][1, i_2]\hat{a}_2[-1, j_2] \\ &= [1, i_1]\hat{a}_1[j_1 + i_2]\hat{a}_2[-1, j_2]\,. \end{aligned} \tag{6.1}$$

So, unless $j_1 + i_2 \leq 0$, which is usually not the case, the result of this reaction is not an interval but contains a gap. If we had used the characteristic reaction of $a_1 a_2$ directly, we would have encountered no gap:

$$a_1 a_2 \to [1, i_1]\widehat{a_1 a_2}[-1, j_2]\,. \tag{6.2}$$

(In this reaction the coefficients i_1 and j_2 occur because of Theorem 5.20.)

No characteristic reaction could have recovered the missing piece, even when starting from $[1, i_1]\hat{a}_1[-1, j_1]a_2$: if we had applied a characteristic reaction to $\hat{a}_1$ or parts of it, it would only yield events at time step 2.

But if we had preserved in the first step of (6.1) some events at the right end of a_1, then the gap in the interval could have been avoided. We will now define a new kind of reaction that accomplishes this.

6.1 Reactions for Separating Intervals

The parts of a separating interval that must be preserved in a reaction are the minimal separating intervals of Definition 5.17. We now introduce a notation for the situations that correspond to them.

***Definition* 6.1** (Leftmost and Rightmost Intervals)**.** Let $a \in \mathscr{S}$ be a separating interval.

Let a_L be the shortest separating interval for which $a_L \setminus\!\setminus a$.

Let a_R be the shortest separating interval for which $a \mathbin{/\!/} a_R$.

Then a_L and a_R are the leftmost and *rightmost minimal separating intervals* of a.

If an interval a is minimally separating, then $a_L = a_R = a$. On the other hand, if a is separating, then $(a_L)_R$ and $(a_R)_L$ are minimal separating intervals, but not necessarily identical. Other properties of a_L and a_R are the subject of the following lemma.

***Lemma* 6.2** (Extending a Separating Interval)**.** *Let $a, x \in \Sigma^*$ be interval situations, of which a is separating. Then*

$$(ax)_L = a_L, \qquad (ax)_R = (a_R x)_R, \tag{6.3}$$

$$(xa)_R = a_R, \qquad (xa)_L = (xa_L)_L\,. \tag{6.4}$$

Proof. Only the equations of the first line need to be proved. We get $(ax)_L$ by removing events from the right part of ax as long as the result is still separating. Since a is separating, we can remove x completely. This proves $(ax)_L = a_L$.

The rightmost separating interval in ax of which we know must is $a_R x$ because we do not know whether x is separating. Therefore the rightmost separating part of ax must be the rightmost separating part of $a_R x$. This proves $(ax)_R = (a_R x)_R$. □

The second piece of the definition is a notation for a specific kind of displacement terms.

***Definition* 6.3** (Slope Operators)**.** Let $a \in \Sigma^*$ be a separating interval with characteristic reaction $a \to [1, i]\hat{a}[-1, j]$. Then the *left* and *right slope operators* of a are

$$+_a = [1, i - |a_L|] \qquad \text{and} \qquad -_a = [-1, j - |a_R|]\,. \tag{6.5}$$

The slope operators are defined in this way because then we have

$$\delta(a) = \delta(a_L +_a a -_a a_R) \qquad \text{and} \qquad \Delta\mathrm{pr}(a) = \mathrm{pr}_{a+_a}(\hat{a})\,. \tag{6.6}$$

Taken together these equations imply that $a \to a_L +_a a -_a a_R$ is a reaction for φ. The two conditions can be verified with the help of Definitions 5.19 and 5.21.

The slope operators are no addition operators, and $+_a$ is not the inverse of $-_a$. Their symbols have been chosen to stay in harmony with the operators $\oplus_k$ and $\ominus_k$ that were already introduced in [51]; here they will reappear in Definition 7.2. Other notations that have no relations to addition, like $\uparrow_a$ and $\downarrow_a$, were considered but rejected. In the case of an arrow notation, the main reason was that there is no complete agreement whether the future or the past is "up" (see p. 22) and a notation that is agnostic in this aspect is therefore preferable. Another point is that we will need to distinguish between two kinds of slope operators; the existence of both encircled and not encircled plus and minus symbols in LaTeX is therefore another reason to use them as the notation for slopes.

A System of Interval Reactions. Now we can define the new reaction system. It will later become a part of the "local reaction system" of Definition 6.13, but is conceptually a bit simpler. It is however complex enough to show essential features and motivate the extensions.

For this system let φ be an interval-preserving transition rule for Σ. Let R be the reaction system that is generated by Σ^* and the following reactions, where $u \in \Sigma^*$ may be any separating interval,

$$u \to_R u_L +_u \hat{u} -_u u_R, \tag{6.7a}$$

$$\hat{u} -_u u +_u \hat{u} \to_R \hat{u} \qquad \text{if } u \text{ is minimally separating.} \tag{6.7b}$$

The diagrams for these reactions can be seen in Figure 6.1. The reactions are shown as parts of a larger situation, which is displayed in grey. In the first

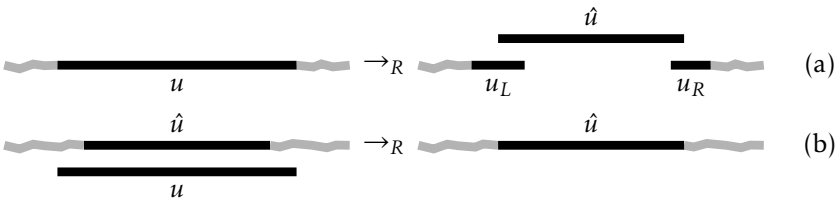


Figure 6.1: A system of interval reactions.

diagram, which shows reaction (6.7a), the interval u is replaced by the interval $\hat{u}$ that is determined by it, but the left and right ends of u are kept for the use in later reactions. The second diagram shows reaction (6.7b). Its left side is somewhat difficult to display: the situation $\hat{u} -_u u +_u \hat{u}$ begins with $\hat{u}$, followed by u, and then, because $\delta(\hat{u} -_u u +_u) = (0, 0)$, the same interval $\hat{u}$ occurs again. The reaction then eliminates the u interval.

To prove that $\delta(\hat{u} -_u u +_u) = (0, 0)$, we only have to notice that u is a minimal separating interval and that therefore $u_L = u_R = u$. Then the first condition in (6.6) becomes $\delta(u +_u \hat{u} -_u u) = \delta(u)$. From this follows $\delta(u +_u \hat{u} -_u) = (0, 0)$, which is equivalent to the assertion.

How it works. With this reaction system we can avoid the problems we had with characteristic reactions. To see how this works, let b be a minimal separating interval. Instead of a_1 and a_2 as before, we now consider the separating intervals $a_1 b$ and $b a_2$, which overlap in b. Because of this we have $(a_1 b)_L = (b a_2)_R = b$ by Lemma 6.2. We can then apply a reaction of the form (6.7a) on $a_1 b$; the result is

$$\begin{aligned} a_1 b &\to_R (a_1 b)_L +_{a_1 b} \widehat{a_1 b} -_{a_1 b} (a_1 b)_R \\ &= (a_1 b)_L +_{a_1} \widehat{a_1 b} -_b b\,. \end{aligned} \tag{6.8}$$

In the same way $b a_2$ reacts to $b +_b \widehat{b a_2} -_{a_2} (b a_2)_R$.

We now evolve the left part of the interval $a_1 b a_2$ first, as we did in (6.1). In the following computation the parts of the formulas that change in the next

step are underlined, to make it more readable. Then we get,

$$\begin{aligned}
\underline{a_1 b a_2} &\to_R (a_1 b)_L +_{a_1} \widehat{a_1 b} -_b \underline{b a_2} \\
&\to_R (a_1 b)_L +_{a_1} \underline{\widehat{a_1 b}} -_b b +_b \underline{\widehat{b a_2}} -_{a_2} (b a_2)_R \\
&\sim (a_1 b)_L +_{a_1} \widehat{a_1 b}\,\langle \hat{b} \rangle\, \underline{\hat{b} -_b b +_b \hat{b}}\,\langle \hat{b} \rangle\, \widehat{b a_2} -_{a_2} (b a_2)_R \\
&\to_R (a_1 b)_L +_b \underline{\widehat{a_1 b}\,\langle \hat{b} \rangle\, \hat{b}}\,\langle \hat{b} \rangle\, \widehat{b a_2} -_{a_2} (b a_2)_R \\
&\sim (a_1 b)_L +_b \widehat{a_1 b}\,\langle \hat{b} \rangle\, \widehat{b a_2} -_{a_2} (b a_2)_R . \qquad (6.9)
\end{aligned}$$

We can then see that $\widehat{a_1 b}\,\langle \hat{b} \rangle\, \widehat{b a_2}$, the part of the reaction result that belongs to time step 1, is now an interval.

If instead we apply rule (6.7a) directly to $a_1 b a_2$, then we get the reaction $a_1 b a_2 \to_R (a_1 b)_L +_b \widehat{a_1 b a_2} -_{a_2} (b a_2)_R$. By comparing its result with the result of the previous computation we see also that $\widehat{a_1 b}\,\langle \hat{b} \rangle\, \widehat{b a_2} \sim \widehat{a_1 b a_2}$.

6.2 Well-Behaved Transition Rules

There were two ideas in the previous section that motivated the jump from characteristic reactions to the reaction system (6.7): It should be possible to reach all elements of the closure of a situation with reactions, and one should be able to do it by applying reactions to this situation in any order. We will make these vague concepts later precise as *covering property* and *confluence* and prove them at the end if this chapter. The proofs however are valid only for a subclass of the interval-preserving transition rules.

This class of *well-behaved* transition rules, which is defined next, is introduced mainly for convenience. It was found by trial and error, trying to exclude special cases that would make proofs and concepts too complex, while keeping the theory applicable for Rules 54 and 110.

***Definition* 6.4** (Well-Behaved)**.** A transition rule φ on Σ is *well-behaved* if

1. φ is interval-preserving,
2. if $\pi \in \mathscr{P}$ is a non-separating interval, then $\Delta\pi = \emptyset$,
3. if $\pi \in \mathscr{P}$ is a minimal separating interval, then $\Delta\pi$ is either a minimal separating interval or non-separating, and
4. the empty interval is not separating.

Condition 2 in this definition is necessary for the proof of Lemma 6.15. It ensures that all reactions that start from intervals and compute new events, i. e. those of the form (6.7a), do indeed start from separating intervals. We do not need to consider very short intervals as special cases. Condition 2 is a completeness property for reaction system (6.7) and for the systems that will be later derived from it.

The condition is always true for elementary cellular automata, but it can become false for radii greater than 1. We will now construct a counterexample for $r = 2$ and $\Sigma = \{0, 1, 2\}$. Its transition rule is

$$\varphi(\sigma_{-2}, \sigma_{-1}, \sigma_0, \sigma_1, \sigma_2) = \begin{cases} 0 & \text{if } \sigma_0 = 0, \\ \max\{\sigma_{-2}, \ldots, \sigma_2\} & \text{otherwise.} \end{cases} \qquad (6.10)$$

Then the interval $\pi = \mathrm{pr}(0)$ provides a contradiction. It is not separating, since the state of the cell at location -1 can influence the next state of the cell at location 1 and *vice versa*, but we also have $\Delta\pi = \mathrm{pr}([1, 0]0)$. It is this crossover influence between cells that Condition 2 prevents.

Condition 3 is necessary in the context of achronal situations (Definition 6.5 below). It concerns situations of the form $a +_a \hat{a}$ or $\hat{b} -_b b$, with a and b minimal separating intervals. These situations arise frequently in reaction system (6.7) and other systems that have a reaction $u \to u_L +_u \hat{u} -_u u_R$. In the result of this reaction, the interval $\hat{u}$ has $\widehat{u_L}$ as its left end and $\widehat{u_R}$ as its right end; so with $a = u_L$ and $b = u_R$ we can say that $u_L +_u \hat{u} -_u u_R$ begins with $a +_a \hat{a}$ and ends with $\hat{b} -_b b$. The condition then ensures that a reaction of the same type as before, when applied to $\hat{u}$, does not destroy $\hat{a}$ and $\hat{b}$. This is because such a reaction, when applied to an interval, leaves its left and right minimal separating intervals intact: The interval $\hat{a}$ is by condition 3 not longer than a minimal separating interval and is therefore part of $\hat{u}_L$, and $\hat{b}$ is for the same reason a part of $\hat{u}_R$.

Here a counterexample occurs with an elementary cellular automaton, Rule 1. We have found in Table 5.4 that the interval 1 is minimally separating for this rule, but $\Delta\mathrm{pr}(1)$ is the interval $\mathrm{pr}([1, -1]000)$, as we can see in Table 5.1. We have thus an interval π for which $\Delta\pi$ consists of three events; such intervals are never minimally separating for elementary cellular automata. The purpose of Condition 3 is to exclude rules with separating intervals that have such excessive influence.

Condition 4 is an intuitively obvious requirement on separating intervals, but it is violated by transition rules of radius 0. These are rules in which the state of a cell does only depend on the state of a single cell at the previous time step. Excluding them from consideration therefore is no loss.

6.3 Achronal Situations

The reaction products in the system (6.7) have a specific form, a generalisation of intervals, for which we will now give a definition. The set is called "achronal" because these situations, like the intervals, consist of events that belong almost to the same time. We think of the events in them as arranged from left to right, not in a temporal sequence.

Achronal situation also have in common with intervals that every achronal situation has a closure and can therefore be the starting point of a reaction. This will be proved later, in Theorem 6.10.

***Definition* 6.5** (Achronal Situations)**.** The set of *achronal situations* for an interval-preserving transition rule φ is the set $\mathscr{A}_\varphi \subset \mathscr{S}$.

It is defined recursively in the following way: A situation $s \in \mathscr{S}$ is an element of $\mathscr{A}_\varphi$ if and only if

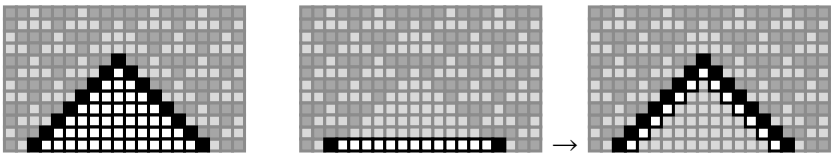
1. $s \in \Sigma^*$, or
2. $s = yb +_b \hat{b}x$, with $yb, \hat{b}x \in \mathscr{A}_\varphi$ and $b \in \Sigma^*$ minimally separating, or
3. $s = x\hat{b} -_b by$, with $x\hat{b}, by \in \mathscr{A}_\varphi$ and $b \in \Sigma^*$ minimally separating.

We will also use two subsets of $\mathscr{A}_\varphi$. The set $\mathscr{A}_{\varphi+}$ consists of those elements of $\mathscr{A}_\varphi$ that are constructed only with the + operators, and the set $\mathscr{A}_{\varphi-}$ consists

of those elements of $\mathscr{A}_\varphi$ that are constructed only with the $-$ operators. These sets are called the *positive* and *negative slopes*.

Similarly, the terms $\hat{b} -_b b$ and $b +_b \hat{b}$ in Definition 6.5 are called *generating slopes*.

Use of Slopes. The positive and negative slopes provide a notation with which we can name the different parts of a situation. Later, in Lemma 6.19, we will see that every situation can react into a situation that is the product of a positive and negative slope. Figure 6.2 is an example. Here the left triangle



(a) Process (b) Reaction

Figure 6.2: A triangle and its slopes under Rule 54.

of Figure 4.1 is expressed as a reaction. It begins with an interval situation and ends with a situation that consists of a positive and a negative slope. (This example will be continued with Figure 8.2.)

The generating slopes are important because the reactions that transform positive generating slopes into positive, or negative generating slopes into negative generating slopes, are among the building blocks for the reaction system associated to a transition rule, which is described in Definition 6.13.

Induction. If we view the recursive construction of the achronal situation as a sequential process, then the intervals are created at its beginning, and every other achronal situation s has either a decomposition $s = yb +_b \hat{b}x$ with yb and $\hat{b}x$ constructed earlier, or a decomposition $s = x\hat{b} -_b by$ with $x\hat{b}$ and by constructed earlier. We have therefore an induction principle for achronal situations:

***Lemma* 6.6** (Achronal Induction)**.** *Let φ be an interval-preserving transition rule. Let $S \subseteq \mathscr{A}_\varphi$ be a set of situations where*

1. $\Sigma^* \subseteq S$,
2. *if $yb, \hat{b}x \in S$, where b is a minimal separating interval, then $yb +_b \hat{b}x \in S$, and*
3. *if $x\hat{b}, by \in S$, where b is a minimal separating interval, then $x\hat{b} -_b by \in S$.*

Then $S = \mathscr{A}_\varphi$. □

This induction principle uses the operators $+_b$ and $-_b$. Situations are however defined in terms of displacements, not in terms of slope operators. Therefore it is not yet clear whether we can, when given an achronal situation, reconstruct the slope operators with which it was constructed. The following lemma shows that the answer is "yes, but it is not completely trivial".

***Lemma* 6.7** (Slope Operators)**.** *Let φ be a well-behaved transition rule and let $s \in \mathscr{A}_\varphi$ be an achronal situation with a honest decomposition $s = x[p]y$ in which $p \neq (0,0)$. Then there is either*

1. *$[p] = +_b$ with $x \mathbin{/\!/} b$, where b is a minimal separating interval, or*
2. *$[p] = -_b$ with $b \mathbin{\backslash\!\backslash} y$, where b is a minimal separating interval, or*
3. *$[p] = +_{b_1} -_{b_2}$ with $x \mathbin{/\!/} b_1$ and $b_2 \mathbin{\backslash\!\backslash} y$, where b_1 and b_2 are minimal separating intervals and $\hat{b}_1 = \hat{b}_2 = [0]$.*

Proof. We see from the definition of $\mathscr{A}_\varphi$ that s can be written as a sequence of cell states and slope operators. In this proof we will call this sequence the *symbol sequence* for s.

In the symbol sequence for s, every one of the symbols either contributes to x, to $[p]$ or to y. The symbols that contribute to $[p]$ can only be slope operators. They form a subsequence of maximal length in the symbol sequence; it is maximal because the decomposition is honest.

If an operator $+_b$ contributes to $[p]$, then the interval b must appear at the left of it in the symbol sequence. Because φ is well-behaved, b is never empty. Therefore $+_b$ can only appear at the left end of the sequence of slope operators that contribute to $[p]$. At its right side it must be followed by $\hat{b}$, but only if $\hat{b} \neq [0]$; in that case $+_b$ is the only factor of $[p]$. For the same reason $-_b$ can only appear at the right end of $[p]$; and if $\hat{b} \neq [0]$, then $-_b$ is the only factor of $[p]$.

Therefore $[p]$ is a product of at most two slope operators in a prescribed order. Since $p \neq (0,0)$, at least one of them must appear. This leads to the three cases of the lemma. □

It is clear that all of these three cases can occur. They can easily be distinguished: we have either $\delta(p)_T = +1, -1$ or 0. Therefore "the number of slope operators"* in a situation is a well-defined concept, and induction over this number is possible. It will be the most common form of induction used in this text.

Achronal Situation Occur Naturally. In the proof of Lemma 6.7 we have seen that every achronal situation can be written as a sequence of elements of Σ together with slope operators: Every $-_b$ must be surrounded by $\hat{b}$ at the left and b at the right and every $+_b$ must be surrounded by b at the left and $\hat{b}$ at the right. Whether a situation is achronal therefore depends only on the terms next to the slope operators. This means that if the two situations $s_1 x$ and $x s_2$ are achronal, their "overlapping product" $s_1 x s_2$ is also achronal.

The converse is not always true, but at least when the common part of the two situations is a separating interval:

***Lemma* 6.8** (Splitting at Separating Intervals)**.** *Let φ be a well-behaved transition rule for Σ. Let $s_1, b, s_2 \in \mathscr{S}$, where b is a separating interval for φ.*

Then if $s_1 b s_2 \in \mathscr{A}_\varphi$, then $s_1 b \in \mathscr{A}_\varphi$ and $b s_2 \in \mathscr{A}_\varphi$.

*More exactly, this number is the minimal number of slope operators with which a situation can be written. Ambiguous cases are possible: if $+_{b_1} = [1,0]$ and $-_{b_2} = [-1,0]$, then $b_1 b_2 = b_1 +_{b_1} -_{b_2} b_2$, and this is in fact an equality, not just an equivalence.

Proof. Let $s = s_1 b s_2$. We perform an induction over the number of slope operators in s.

If $s \in \Sigma^*$, the lemma is obviously true. Otherwise s has at least one slope operator, either $+_a$ or $-_a$.

Assume now that $s = x\hat{a} -_a ay$ is a decomposition of s with $x\hat{a}, ay \in \mathscr{A}_\varphi$, where a is minimally separating. Since b is a nonempty interval, it must be part of either $x\hat{a}$ or of ay.

If b is part of $x\hat{a}$, then there is a situation s_2' such that $x\hat{a} = s_1 b s_2'$. Since $s_1 b s_2'$ has fewer slope operators than s, the induction hypothesis can be applied to it, and therefore $s_1 b \in \mathscr{A}_\varphi$ and $bs_2' \in \mathscr{A}_\varphi$. Because φ is well-behaved, the interval $\hat{a}$ is not longer than a separating interval; therefore bs_2', which contains the separating b interval as its factor, cannot be just a part of $\hat{a}$. So there must be a situation $z \in \mathscr{S}$ such that $bs_2' = z\hat{a}$. Then $bs_2' -_a ay = z\hat{a} -_a ay$ and therefore $bs_2' -_a ay \in \mathscr{A}_\varphi$.

If b is part of ay, then there is a situation s_1' such that $ay = s_1' b s_2$. Then $s_1' b \in \mathscr{A}_\varphi$ and $bs_2 \in \mathscr{A}_\varphi$ by induction. When dividing up ay, the situation a must become a part of $s_1' b$ because it is the leftmost minimal separating interval of this situation and $s_1' b$ contains already the separating interval b: the interval a could not have been cut into pieces. Therefore $x\hat{a} -_a s_1' b \in \mathscr{A}_\varphi$.

So if $s = x\hat{a} -_a ay$, then we could divide s either into $s_1 b$ and $bs_2' -_a ay$ or into $x\hat{a} -_a s_1' b$ and bs_2. If $s = ya +_a \hat{a}x$, then there are similar decompositions for it; they can be found by a mirror image of this argument. □

With the methods developed so far we can now show that achronal situations occur naturally in the reaction system (6.7).

***Lemma* 6.9** (Achronal Domain)**.** *Let R be the interval reaction system of* (6.7)*. Then* dom $R \subseteq \mathscr{A}_\varphi$.

Proof. We will show that the generating reactions (6.7) transform achronal situations into achronal situations. As the initial situations of R are intervals and therefore obviously achronal, this will prove that all situations in dom R are achronal.

Let $xuy \in \mathscr{A}_\varphi$, where u is a separating interval. Then xu_L, u and $u_R y$ are achronal by Lemma 6.8. One can see directly that $u_L +_u \hat{u} -_u u_R$ is achronal. Therefore $xu_L +_u \hat{u} -_u u_R y \in \mathscr{A}_\varphi$, again by Lemma 6.8. This proves that the reaction $u \to_R u_L +_u \hat{u} -_u u_R$ preserves achronality.

Let now $x\hat{u} -_u u +_u \hat{u}y \in \mathscr{A}_\varphi$, where u is a minimal separating interval. Then $x\hat{u}$ and $\hat{u}y$ are achronal by by Lemma 6.8. Therefore we have $x\hat{u}y \in \mathscr{A}_\varphi$. This proves that the reaction $\hat{u} -_u u +_u \hat{u} \to_R \hat{u}$ preserves achronality. □

6.4 Closure

A preference for symmetry now leads to another question: If all reactions results are achronal situations, can we then also extend the set of input situations of the reaction system in (6.7) from Σ^* to $\mathscr{A}_\varphi$? The following theorem shows that this is possible for a subset of the achronal situations.

For this we have to introduce a new concept. It represents the intuitive notion that the events of a situation are arranged approximately from left

to right. To express the concept for a situation s we consider the honest decompositions of s of the form

$$s = s_1[p_1]u_1[q_1]s_2[p_2]u_2[q_2]s_3, \tag{6.11}$$

in which u_1 and u_2 are intervals and $p_1, q_1, p_2, q_2 \neq (0,0)$. We then write $\pi_1 = \mathrm{pr}_{s_1[p_1]}(u_1)$ and $\pi_2 = \mathrm{pr}_{s_1[p_1]u_1[q_1]s_2[p_2]}(u_2)$ for the processes that belong to u_1 and u_2 (Figure 6.3). Now consider the decompositions of the form (6.11)

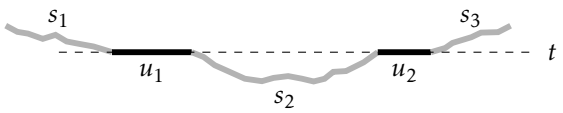


Figure 6.3: The situation s in (6.11). The processes belonging to u_1 and u_2 occur at the same time.

the processes π_1 and π_2 belong to the same time. If for every decomposition of s of this form we have $\pi_1 <> \pi_2$, then the situation s is *ordered*.

***Theorem* 6.10** (Closure of Achronal Situations)**.** *Let φ be a well-behaved transition rule for Σ and $s \in \mathscr{A}_\varphi$ be ordered. Then* $\mathrm{cl}\,\mathrm{pr}(s)$ *exists.*[†]

It is enough if we restrict the proof of the theorem to the case where s is a *balanced* situation. This shall mean that $\delta(s)_T = 0$, that $\mathrm{pr}(s)^{(t)} = \emptyset$ for all $t \geq 1$ and that s is either an interval or there is a decomposition $s = \hat{a} -_a x +_b \hat{b}$ with minimal separating intervals a and b. (See Figure 6.4. Note that $\hat{a}$ or $\hat{b}$ may be empty.) Every situation can be extended to a balanced situation; if that

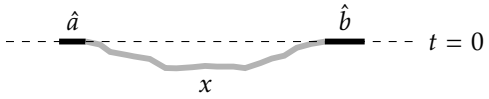


Figure 6.4: A balanced situation.

situation has a closure, then the original situation has a closure too.

First we will however prove the existence of the closure for the simplest nontrivial balanced situations. This result is then used as a stepping stone for the proof of Theorem 6.10.

***Lemma* 6.11** (Closure of Simple Balanced Situation)**.** *Let φ be a well-behaved transition rule for Σ and let $s = \hat{a} -_a u +_b \hat{b}$ be an achronal situation for Σ in which u is an interval and a, b are minimal separating intervals.*

Then $s \to \hat{u}$ is a reaction for φ, and $\mathrm{cl}^{(0)}\,\mathrm{pr}(s) = \mathrm{pr}(\hat{u})$.

A diagram of the processes for s and $\hat{u}$ can be seen in Figure 6.5. The process belonging to $\hat{u}$ overlaps with those of $\hat{a}$ and $\hat{b}$.

[†]It is this theorem for which we need the fact that achronal situations are ordered: A counterexample in the reaction system for Rule 54 is the situation $000 \oplus 01 \oplus_2 1 \ominus_2 10$, written in the notation (7.8). It is not ordered but achronal, and it has no closure.

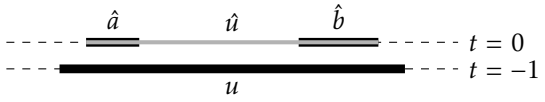


Figure 6.5: The situations in Lemma 6.11.

Proof. To compute the closure of $\mathrm{pr}(s)$ we must express s in the language of cellular processes. Let therefore $\pi = \mathrm{pr}(s)$; its components are then

$$\alpha' = \mathrm{pr}(\hat{a}), \qquad \mu = \mathrm{pr}_{\hat{a}-_a}(u), \qquad \beta' = \mathrm{pr}_{\hat{a}-_a u+_b}(b), \tag{6.12}$$

such that $\pi = \alpha' \cup \mu \cup \beta'$. The process π then consists of the time slices $\pi^{(-1)} = \mu$ and $\pi^{(0)} = \alpha' \cup \beta'$. Since s is an achronal situation, we must have $a \setminus\setminus u /\!/ b$. We will therefore also need names for the end intervals of μ. They are

$$\alpha = \mathrm{pr}_{\hat{a}-_a}(a) \qquad \text{and} \qquad \beta = \mathrm{pr}_{\hat{a}-_a u\langle b\rangle}(b). \tag{6.13}$$

Then we can say that μ begins with α and ends with β, such that we have $\alpha \subseteq_R \mu \supseteq_L \beta$ (Figure 6.6).

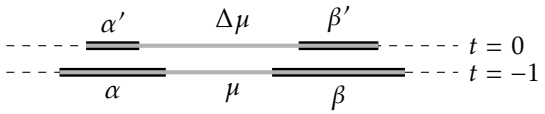


Figure 6.6: The processes related to the situations in Figure 6.5.

The main part of the proof then consists of a computation of the space-time locations of all these processes. For this we let the characteristic reactions of a and b be

$$a \to [1, i]\hat{a}[-1, j'] \qquad \text{and} \qquad b \to [1, i']\hat{b}[-1, j]. \tag{6.14}$$

By Theorem 5.20, we must then have $u \to [1, i]\hat{u}[-1, j]$ as characteristic reaction for u. This is because u begins with a, and therefore its characteristic reaction shares its left displacement term $[1, i]$ with that of a, and u ends with b and therefore its characteristic reaction shares its right displacement term $[-1, j]$ with that of b. This "bounding" of the location of $\hat{u}$ by a and b is the core of the proof. Because of the left-oriented structure of the formalism it however does not become directly visible in the following calculations.

Before we start with the calculations proper, we will determine short expressions for the values of $\delta(\hat{a}-_a)$ and $\delta(\hat{a} -_a u)$, terms that will occur at several places. For the first term we begin with the equation $-_a = [-1, j'-|a|]$, which follows from Definition 6.3. Then we can see that $\delta(\hat{a}-_a) = (0, |\hat{a}|) + (-1, j' - |a|) = (-1, |\hat{a}| + j' - |a|)$. We now use a relation derived from the characteristic reaction for a to simplify that term. The left and the right side of a reaction must have the same size vector, and this means for the characteristic reaction for a that $|a| = i + |\hat{a}| + j'$. Using that we see that $\delta(\hat{a}-_a) = (-1, -i)$. With this result we get an expression for the second size vector, $\delta(\hat{a} -_a u) = (-1, -i) + (0, |u|) = (-1, |u| - i)$. It too can be brought into a form that is more useful later, this

time with the equation $|u| = i + |\hat{u}| + j$ that is derived from the characteristic reaction of u. The result is $(|\hat{u}| + j)$, such that we have

$$\delta(\hat{a}-_a) = (-1,-i) \qquad \text{and} \qquad \delta(\hat{a} -_a u) = (0, |\hat{u}| + j)\,. \tag{6.15}$$

Next we will show that $\alpha' = \Delta\alpha$ and $\beta' = \Delta\beta$. To find a term that expresses $\Delta\alpha$ in terms of situations, we use the characteristic reaction for a. We can read it as saying that the set of events determined by the process $\mathrm{pr}(a)$ is $\mathrm{pr}([1,i]\hat{a})$. (See Definition 5.18.) This is also valid for shifted versions of $\mathrm{pr}(a)$, so the set of events determined by $\alpha = \mathrm{pr}_{\hat{a}-_a}(a)$ must be $\Delta\alpha = \mathrm{pr}_{\hat{a}-_a}([1,i]\hat{a})$. The position of $\hat{a}$ in this term is the sum of two displacements, $\delta(\hat{a}-_a)$ and $(1,i)$. Since $\delta(\hat{a}-_a) + (1,i) = (0,0)$, we have therefore

$$\Delta\alpha = \mathrm{pr}(\hat{a}) = \alpha'\,. \tag{6.16}$$

For the same reason, this time with the characteristic reaction of $\hat{b}$, the set of events determined by $\beta = \mathrm{pr}_{\hat{a}-_a u\langle b\rangle}(\hat{b})$ must be $\Delta\beta = \mathrm{pr}_{\hat{a}-_a u\langle b\rangle}([1,i']\hat{b})$. So we must compute $\delta(\hat{a} -_a u\langle b\rangle) + (1,i')$ to find the position of $\hat{b}$ in this term: Then we get $\delta(\hat{a} -_a u\langle b\rangle) + (1,i') = (-1, |\hat{u}| + j) + (0, -|b|) + (1,i') = (0, |\hat{u}| + j - |b| + i')$. We use the equation $|b| = i' + |\hat{b}| + j$, which is derived from the characteristic reaction for b, to simplify the result of this computation to $(0, |\hat{u}| - |\hat{b}|)$. Therefore we have

$$\Delta\beta = \mathrm{pr}([0, |\hat{u}| - |\hat{b}|]\hat{b})\,. \tag{6.17}$$

To find the location of $\beta' = \mathrm{pr}_{\hat{a}-_a u+_b}(\hat{b})$ we use the fact that $+_b = [1, i' - |b|]$. (See Definition 6.3). Then we can calculate $\delta(\hat{a} -_a u+_b) = (-1, |\hat{u}| + j) + (1, i' - |b|) = (0, |\hat{u}| + j + i' - |b|)$ and simplify the result in the same way as before to $(0, |\hat{u}| - |\hat{b}|)$. Therefore we get

$$\beta' = \mathrm{pr}([0, |\hat{u}| - |\hat{b}|]\hat{b}), \tag{6.18}$$

which shows that $\Delta\beta = \beta'$.

With this data we can compute the time slices of $\mathrm{cl}\,\pi$ and therefore show that π actually has a closure. We have already seen that $\pi^{(-1)} = \mu$, $\pi^{(0)} = \alpha' \cup \beta'$, and that $\pi^{(t)} = \emptyset$ for all other values of t. Therefore, applying the definition (3.21) of the closure we get

$$\mathrm{cl}^{(t)}\,\pi = \pi^{(t)} = \emptyset \qquad \text{for } t < -1, \tag{6.19a}$$

$$\mathrm{cl}^{(-1)}\,\pi = \pi^{(-1)} = \mu, \tag{6.19b}$$

$$\begin{aligned} \mathrm{cl}^{(0)}\,\pi &= \pi^{(0)} \cup \Delta\,\mathrm{cl}^{(-1)}\,\pi \\ &= (\alpha' \cup \beta') \cup \Delta\mu = \Delta\mu, \end{aligned} \tag{6.19c}$$

$$\mathrm{cl}^{(t)}\,\pi = \pi^{(t)} \cup \Delta\,\mathrm{cl}^{(t-1)}\,\pi = \Delta\,\mathrm{cl}^{(t-1)}\,\pi \qquad \text{for } t \geq 1. \tag{6.19d}$$

Only the third equation must be explained. It is true because $\alpha' = \Delta\alpha$ and $\beta' = \Delta\beta$. Since α and β are subsets of μ, we must then have $\alpha' \subseteq \Delta\mu \supseteq \beta'$ by the monotony of the Δ operator. This then proves that $\alpha' \cup \beta'$ is compatible with $\Delta\mu$ and that $\mathrm{cl}^{(0)}\,\pi$ actually exists. Together these equations show that $\mathrm{cl}\,\pi$ exists.

It remains to prove that $s \to \hat{u}$ is a reaction for φ. We have to show that $\delta(s) = \delta(\hat{u})$ and that $\mathrm{pr}(\hat{u}) \subseteq \mathrm{cl}\,\mathrm{pr}(s)$. We know already that $\delta(\hat{a} -_a u+_b) = (0, |\hat{u}| - |\hat{b}|)$; therefore $\delta(s) = \delta(\hat{a} -_a u +_b \hat{b}) = (0, |\hat{u}|) = \delta(\hat{u})$. To prove that $\mathrm{pr}(\hat{u}) \subseteq$

$\operatorname{cl}\operatorname{pr}(s)$ we will now show that $\operatorname{pr}(\hat{u}) = \Delta\mu$. For this we use the fact that the set of events determined by $\mu = \operatorname{pr}_{\hat{a}-_a}(u)$ is the process $\Delta\mu = \operatorname{pr}_{\hat{a}-_a}([-1,-i]u)$. Then, since $\delta(\hat{a}-_a) + (-1,-i) = (0,0)$, we must have $\operatorname{pr}(\hat{u}) = \Delta\mu$. Now we can apply the result of (6.19c) that $\Delta\mu = \operatorname{cl}^{(0)}\pi$ and see that $\operatorname{pr}(\hat{u}) \subseteq \operatorname{cl}\operatorname{pr}(s)$. This then concludes the proof that $s \to \hat{u}$ is a reaction; it also proves that $\operatorname{cl}^{(0)}\pi = \operatorname{pr}(\hat{u})$. □

Proof of Theorem 6.10. As explained above, we restrict our case to balanced intervals.

For a situation a, we will call the first time t for which $\operatorname{pr}(a)^{(t)} \neq \emptyset$ the *starting time* of a. A balanced situation has then a starting time $t \leq 0$, and the balanced situations with starting time 0 are the intervals. Since intervals have a closure, it is therefore enough to show that if $t < 0$ and every balanced situation with starting time $t+1$ has closure, then every situation with starting time t has a closure too.

We do this in the following way. A process s with starting time t_0 has by definition the time slices of the closure $\operatorname{cl}^{(t)}\operatorname{pr}(s)$ for every $t \leq t_0$, with $\operatorname{cl}^{(t_0)}\operatorname{pr}(s) = \operatorname{pr}(s)^{(t_0)}$. We will then show that for every such s exists another situation s' with starting time t_0+1 such that $\operatorname{cl}^{(t_0+1)}\operatorname{pr}(s) = \operatorname{pr}(s')^{(t_0+1)}$. The closure of s' exists by induction, and we have, as before, $\operatorname{cl}^{(t_0+1)}\operatorname{pr}(s') = \operatorname{pr}(s')^{(t_0+1)}$. Therefore,

$$\operatorname{cl}\operatorname{pr}(s) = \bigcup_{t\leq t_0} \operatorname{pr}(s)^{(t)} \cup \bigcup_{t>t_0} \operatorname{cl}\operatorname{pr}(s')^{(t)}, \tag{6.20}$$

so the closure of s exists then.

Now we must isolate in s the factors that contribute to $\operatorname{pr}(s)^{(t_0)}$. For this we will use the decomposition

$$s = s_0 -_{a_1} u_1 +_{b_1} s_1 \dots s_{\ell-1} -_{a_\ell} u_\ell +_{b_\ell} s_\ell\,. \tag{6.21}$$

in which the u_i are the intervals that belong to time t_0. More precisely, we

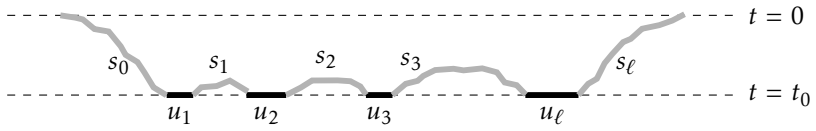


Figure 6.7: The situation s in (6.21).

write $\pi_i = \operatorname{pr}_{s_0\dots s_{i_1}+_{a_i}}(u_i)$ for the process that belongs to u_i and require that $\operatorname{pr}(s)^{(t_0)} = \bigcup_{i=1}^{\ell} \pi_\ell$.

The situations s_i are arbitrary and need not be intervals. Nevertheless, since s is achronal, every situation s_i with $i < \ell$ ends with $\hat{a}_{i+1}$, and every s_i with $i > 0$ begins with $\hat{b}_i$. Therefore

$$s \sim s_0 \langle\hat{a}_1\rangle \underline{\hat{a}_1 -_{a_1} u_1 +_{b_1} \hat{b}_1} \langle\hat{b}_1\rangle s_1 \dots s_{\ell-1} \langle\hat{a}_\ell\rangle \underline{\hat{a}_\ell -_{a_\ell} u_\ell +_{b_\ell} \hat{b}_\ell} \langle\hat{b}_\ell\rangle s_\ell\,. \tag{6.22}$$

We can now apply Lemma 6.11 to the underlined factors in this equation. In the current context it says that $\Delta\pi_i = \operatorname{pr}_{s_0\dots s_{i-1}\langle\hat{a}_i\rangle}(\hat{u}_i)$ for every i. We therefore get the situation s' by replacing the underlined factors in s with the intervals $\hat{u}_i$,

$$s' \sim s_0 \langle\hat{a}_1\rangle \hat{u}_1 \langle\hat{b}_1\rangle s_1 \dots s_{\ell-1} \langle\hat{a}_\ell\rangle \hat{u}_\ell \langle\hat{b}_\ell\rangle s_\ell\,. \tag{6.23}$$

In fact the situation s' is what we get when we resolve the overlaps at the right side of the previous equation. This is always possible because every $\hat{u}_i$ begins with $\hat{a}_i$ and ends with $\hat{b}_i$.

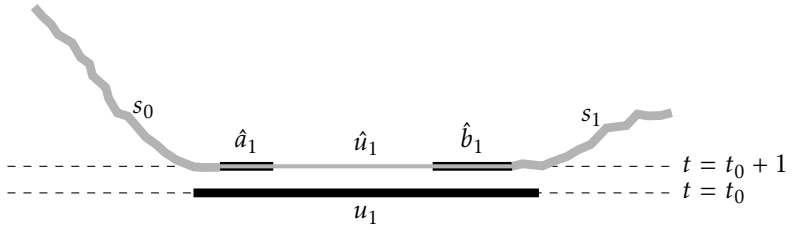


Figure 6.8: The left end of Figure 6.7, with the factors of s and s' overlayed.

We also have to check whether one of the processes $\Delta\pi_i$ intersects with other parts of $\mathrm{pr}(s)^{(t_0+1)}$. But $\Delta\pi_i$ may have a non-empty intersection only with the processes belonging to s_{i-1} and s_i. This is because s is ordered: The intervals of $\mathrm{pr}(s)^{(t_0+1)}$ to the left of $\hat{a}_i$ must all belong to a situation s_k or u_k with $k \leq i-1$. Therefore such an interval may extend to the right at most as far as the right end of $\hat{a}_i$. Similarly, the intervals of $\mathrm{pr}(s)^{(t_0+1)}$ to the right of $\hat{b}_i$ must all belong to a situation s_k or u_k with $i \leq k$ and therefore extend to the left at most to the left end of $\hat{b}_i$. The left-to-right arrangement of the intervals of s is therefore preserved in s', with the intervals $\hat{u}_i$ inserted in the gaps between the s_i. This shows that s' is ordered.

So $\Delta\mathrm{pr}(s)^{(t_0)}$ is compatible with $\mathrm{pr}(s)^{(t_0+1)}$, and $\mathrm{cl}^{(t_0+1)}\,\mathrm{pr}(s)$ exists and is equal to $\mathrm{pr}(s')^{(t_0+1)}$. □

Therefore we can define now a reaction system that has $\mathscr{A}_\varphi$ as its domain.

6.5 The Local Reaction System

The preliminary reaction system (6.7) has the disadvantage that its set of generating reactions is infinite. We cannot specify them in a list in the same way as we can do this with a transition rule.

We will now make the local nature of the interactions in a cellular automaton more visible by decomposing the generating reactions of (6.7) into a finite number of reactions that involve only a finite number of events. The following lemma specifies these reactions and shows how they generate the reactions of (6.7).

***Lemma* 6.12** (Local Generators)**.** *Let R be a reaction system that contains for all separating intervals u and all $\sigma \in \Sigma$ the reactions*

$$u \to_R u_L +_u \hat{u} -_u u_R, \qquad \text{if } u \text{ is minimal,} \tag{6.24a}$$

$$\hat{u} -_u u\sigma \to_R \widehat{u\sigma} -_{u\sigma} (u\sigma)_R, \qquad \text{if } u \text{ is right minimal,} \tag{6.24b}$$

$$\sigma u +_u \hat{u} \to_R (\sigma u)_L +_{\sigma u} \widehat{\sigma u}, \qquad \text{if } u \text{ is left minimal.} \tag{6.24c}$$

Then R contains for all separating intervals v, not just those that are minimally separating, the reaction $v \to_R v_L +_v \hat{v} -_v v_R$.

The diagrams for these reactions are shown in Figure 6.9; it uses the same conventions as Figure 6.1. The intention behind the definitions is that reac-

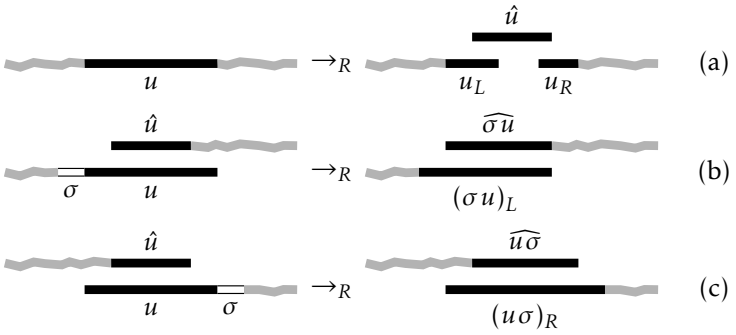


Figure 6.9: Reactions to generate an interval.

tion (6.24a) is used to generate a new interval one time step in the future—but this time one of minimal length—and that (6.24b) and (6.24c) then are used to expand it to the left and the right.

Proof. We prove the lemma by induction over the length of v: A separating interval v is either minimally separating or there exist a separating interval $w \in \Sigma^*$ and a state $\sigma \in \Sigma$ such that $v = w\sigma$ or $v = \sigma w$.

If v is minimal, then there is by (6.24a) a reaction $v \to_R v_L +_v \hat{v} -_v v_R$.

If $v = w\sigma$, then there is by induction a reaction $w \to_R w_L +_w \hat{w} -_w w_R$. We apply it to v and get $v \to_R w_L +_w \hat{w} -_w w_R\sigma$. Now let $x \in \Sigma^*$ such that $w = xw_R$. Since w_R is separating, there is by Lemma 5.22 an $x' \in \Sigma^*$ such that $\widehat{xw_R} = x'\widehat{w_R}$. Therefore

$$v \to_R w_L +_w x'\widehat{w_R} -_w w_R\sigma \tag{6.25}$$

is a reaction in R. Then, since w_R is right minimal and $-_w = -_{w_R}$, we can apply the reaction (6.24c) with $u = w_R$ to the result of (6.25) and get

$$w_L +_w x'\widehat{w_R} -_w w_R\sigma \to_R w_L +_w x'\widehat{w_R\sigma} -_{w_R\sigma} (w_R\sigma)_R\,. \tag{6.26}$$

We must now interpret the result of this reaction in terms of v. We have $w_L = v_L$ and $(w_R\sigma)_R = (w\sigma)_R = v_R$ by Lemma 6.2, which also means that $+_w = +_v$ and $-_{w_R\sigma} = -_v$. For the middle term of the reaction result we apply again Lemma 5.22: Since $w_R\sigma$ is separating and $xw_R\sigma = w\sigma = v$, we must have $x'\widehat{w_R\sigma} = \hat{v}$. Therefore the result of (6.26) is $v_L +_v \hat{v} -_v v_R$. Putting everything together we show this way that $v \to_R v_L +_v \hat{v} -_v v_R$ if $v = w\sigma$.

If $v = w\sigma$, a similar argument can be used. □

***Definition* 6.13** (Local Reaction System)**.** Let φ be a well-behaved transition rule. Let Φ be the reaction system generated by the ordered situations in $\mathscr{A}_\varphi$

and the following reactions, for all separating $u \in \Sigma^*$ and $\sigma \in \Sigma$,

$$u \to_\Phi u +_u \hat{u} -_u u, \qquad \text{if } u \text{ is minimal,} \tag{6.27a}$$

$$\hat{u} -_u u +_u \hat{u} \to_\Phi \hat{u}, \qquad \text{if } u \text{ is minimal,} \tag{6.27b}$$

$$\sigma u +_u \hat{u} \to_\Phi (\sigma u)_L +_{\sigma u} \widehat{\sigma u}, \qquad \text{if } u \text{ is left minimal,} \tag{6.27c}$$

$$\hat{u} -_u u\sigma \to_\Phi \widehat{u\sigma} -_{u\sigma} (u\sigma)_R, \qquad \text{if } u \text{ is right minimal.} \tag{6.27d}$$

This reaction system is called the *local reaction system* for φ.

For completeness, the diagrams for these reactions are also shown, in Figure 6.10.

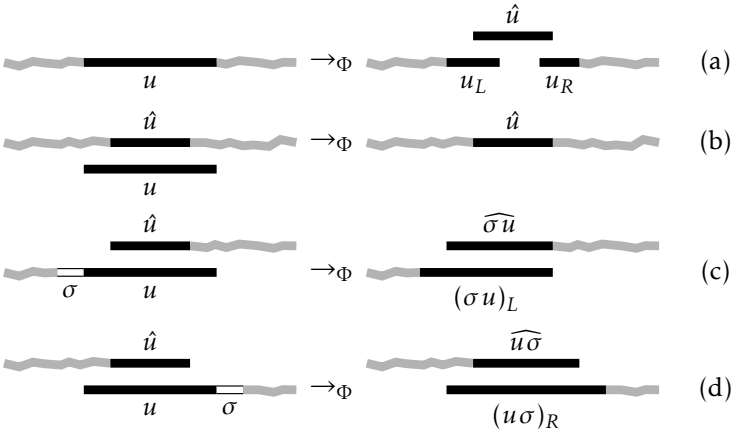


Figure 6.10: Generators of the local reaction system.

In the rest of this chapter we will prove the following properties of the local reaction system.

***Theorem* 6.14** (Properties of Local Reaction Systems)**.** *Let φ be a well-behaved transition rule and Φ the local reaction system for φ. Then Φ has the following properties:*

1. (Covering, Figure 6.11). *If $s \in \operatorname{dom}\Phi$ and $[p]\sigma \in \operatorname{cl}\operatorname{pr}(s)$, then there is a reaction $s \to_\Phi v$ such that $[p]\sigma \in \operatorname{pr}(v)$.*
2. (Confluence, Figure 6.12). *If there are reactions $a \to_\Phi b_1$ and $a \to_\Phi b_2$, then there is a situation $c \in \operatorname{dom}\Phi$ such that $b_1 \to_\Phi c$ and $b_2 \to_\Phi c$.*

Proof. The first property is proved in Theorem 6.17, the second property in Theorem 6.22. □

6.6 Covering

The property that is the subject of this section is a kind of converse to the definition of reactions with help of the closure: Given a reaction system R, do we have

$$\operatorname{cl}\operatorname{pr}(a) = \bigcup \{ \operatorname{pr}(b) : a \to_R b \} \tag{6.28}$$

for a situation $a \in \operatorname{dom} R$? If this is true, then we say that *R covers* the closure of a. If R covers the closure of every $a \in \operatorname{dom} R$, then no information about the cellular automaton gets lost when switching from the work with closures to the work with reaction systems. The most important case is of course the local reaction system Φ.

We will now prove the closure property in a slightly different form, by asking whether a specific event belongs to the closure of a.

The simplest case occurs when the initial situation a itself an interval. We can then express Lemma 5.10 for well-behaved transition rules in terms of reactions.

***Lemma* 6.15** (Intervals are Covering)**.** *Let φ be a well-behaved transition rule. Let R be a reaction system for φ where for every separating interval $a \in \Sigma^*$ with characteristic reaction $a \to [1, i]\hat{a}[-1, j]$ there is a reaction*

$$a \to_R a_+ \hat{a} a_- \tag{6.29}$$

with $\delta(a_+) = (1, i)$ and $\delta(a_-) = (-1, j)$.

Then for every interval $u \in \operatorname{dom} R$ and every event $[p]\sigma \in \operatorname{cl} \operatorname{pr}(u)$ there is a reaction $u \to_R v$ with $[p]\sigma \in \operatorname{pr}(v)$.

Proof. We will prove the following assertion for every $t \geq 0$: If there is a reaction $u \to_R u_+ a u_-$ with $\operatorname{cl}^{(t)} \operatorname{pr}(u) = \operatorname{pr}_{u_+}(a)$, then there is a reaction $u \to_R u'_+ a' u'_-$ with $\operatorname{cl}^{(t+1)} \operatorname{pr}(u) = \operatorname{pr}_{u'_+}(a')$.

Since $cl_\varphi^{(0)} \operatorname{pr}(u) = \operatorname{pr}(u)$, we know then by induction that for all $t \geq 0$ there is a reaction $u \to_R v$ with $\operatorname{cl}^{(t)} \operatorname{pr}(u) \subseteq \operatorname{pr}(v)$, which proves the lemma.

Assume now that $u \to_R u_+ a u_-$ with $\operatorname{cl}^{(t)} \operatorname{pr}(u) = \operatorname{pr}_{u_+}(a)$. If a is separating, then there is a reaction (6.29) for it. Then $\Delta\operatorname{pr}(a) = \operatorname{pr}([1, i]\hat{a})$, and therefore $\operatorname{cl}^{(t+1)} \operatorname{pr}(u) = \Delta(\operatorname{cl}^{(t)} \operatorname{pr}(u)) = \Delta\operatorname{pr}_{u_+}(a) = \operatorname{pr}_{u_+}([1, i]\hat{a})$. So if we set $u'_+ = u_+ a_+$, $a' = \hat{a}$ and $u_- = a_- u_-$, the assertion is true for a separating a.

If a is non-separating, especially empty, then $\Delta\operatorname{pr}(a) = \emptyset$ by assumption. This also means that $\operatorname{cl}^{(t+1)} \operatorname{pr}(u) = \Delta(\operatorname{cl}^{(t)} \operatorname{pr}(u)) = \Delta\operatorname{pr}_{u_+}(a) = \emptyset$. So we may choose $u'_+ = u_+ a$, $a' = [0]$, and $u'_- = u_-$ to fulfil the initial assertion of this proof. □

The local reaction system is then a specific case of the reaction system in the previous lemma, so we get:

***Lemma* 6.16** (Covering)**.** *Let φ be a well-behaved transition rule. Let R be a reaction system for φ which has for every separating interval $a \in \Sigma^*$ a reaction*

$$a \to_R a_L +_a \hat{a} -_a a_R\,. \tag{6.30}$$

Then for every interval $u \in \operatorname{dom} R$ and every event $[p]\sigma \in \operatorname{cl} \operatorname{pr}(u)$ there is a reaction $u \to_R v$ with $[p]\sigma \in \operatorname{pr}(v)$.

Proof. Let $a \to [1, i]\hat{a}[-1, j]$ be the characteristic reaction for a in (6.30). Then $\delta(a_L +_a) = (0, |a_L|) + (1, i - |a_L|) = (1, i)$ and $\delta(-_a a_R) = (-1, j - |a_R|) + (0, |a_R|) = (-1, j)$. Therefore we can apply Lemma 6.15, which finishes the proof. □

With this lemma we can prove covering for the general case (Figure 6.11).

***Theorem* 6.17** (Covering by Achronal Situations)**.** *Let Φ be the local reaction system for a well-behaved transition rule φ and let $s \in \operatorname{dom}\Phi$. Then for all events $[p]\sigma \in \operatorname{cl}\operatorname{pr}(s)$ there is a reaction $s \to_\Phi v$ with $[p]\sigma \in \operatorname{pr}(v)$.*

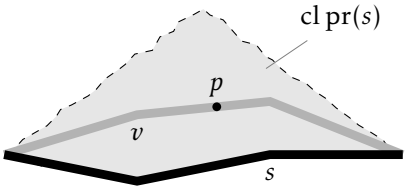


Figure 6.11: A reaction that covers the point p.

Proof. Assume that $a = a_1 u a_2$ with $u \in \Sigma^*$ separating. Then there is a reaction $a \to_\Phi a_1 +_u \hat{u} -_u a_2$, and if $[p]\sigma \notin \operatorname{cl}\operatorname{pr}(a_1 +_u \hat{u} -_u a_2)$, we must have $[p]\sigma \in pr_{a_1}(u)$.

Assume that $a = a_1 \hat{u} -_u u +_u \hat{u} a_2$ with $u \in \Sigma^*$ minimally separating. Then there is a reaction $a \to_\Phi a_1 \hat{u} a_2$, and if $[p]\sigma \notin \operatorname{cl}\operatorname{pr}(a_1 \hat{u} a_2)$, we must have $[p]\sigma \in pr_{a_1 \hat{u} -_u}(u)$.

These two types of reactions generate all reactions in Φ, so we see: If $a \to_\Phi b$ and $[p]\sigma \notin \operatorname{cl}\operatorname{pr}(b)$, then there must be a reaction $a \to_\Phi b'$ with $[p]\sigma \in \operatorname{pr}(b')$.

Now there is for every $a \in \operatorname{dom}\Phi$ a reaction $a \to_\Phi b$ with

$$b = u_1 +_{u_1} \cdots +_{u_{k-1}} u_k +_{u_k} v -_{w_\ell} w_\ell -_{w_{\ell-1}} \cdots -_{w_1} w_1 \,. \tag{6.31}$$

in which v is an interval and the u_i and w_i are minimal separating intervals. This can shown in an analogous way to the proof of Lemma 6.19 below. With this definition we have $\operatorname{cl}\operatorname{pr}(b) = \operatorname{pr}(b) \cup \operatorname{cl}\operatorname{pr}_{u_1 +_{u_1} \cdots +_{u_\ell}}(v)$. This is so because φ is well-behaved and therefore the u_i and v_i provide no additional events to the closure of b.‡

Now v, as an interval, is covering, so there is either $[p]\sigma \in \operatorname{cl}\operatorname{pr}(b)$; then a reaction $v \to_\Phi c$ can be applied to b in order to cover $[p]\sigma$. Or, by the argument outlined above, there is directly a reaction $a \to_\varphi b'$ with $[p]\sigma \in \operatorname{pr}(b')$. In either case $[p]\sigma$ is covered. □

6.7 Confluence

I have borrowed the notion of confluence from the theory of term rewriting systems, especially the lambda calculus [1, p. 4–5]. If a term rewriting system is *confluent* and a term a can be transformed by one rule of that system to a term b_1 and by another rule to a term b_2, then there is a term c in that system that serves as a unifying target for b_1 and b_2: there is a rule that transforms b_1 to c and another rule that transforms b_2 to c.

In the language of reaction systems this means: A reaction system R is confluent if for every pair of reactions $a \to_R b_1$ and $a \to_R b_2$ there is a situation $c \in \operatorname{dom} R$ such that $b_1 \to_R c$ and $b_2 \to_R c$ (Figure 6.12).

‡We have e. g. $\Delta\operatorname{pr}(u_1) \subseteq \operatorname{pr}_{u_1 +_{u_1}}(u_2)$ by the third property of Definition 6.4.

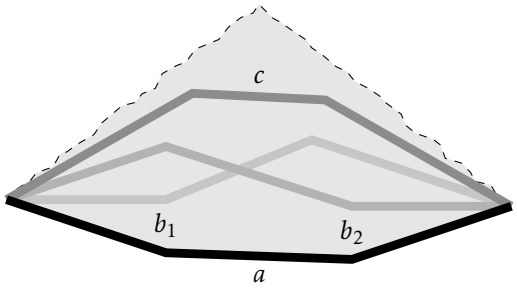


Figure 6.12: Two confluent reactions.

We will now show that this is also true for the local reaction system Φ. The situation c will then have a specific form, which we will specify with help of a subset of Φ.

***Definition* 6.18** (Slope Subsystems)**.** Let $R \subseteq \mathscr{A}_\varphi \times \mathscr{A}_\varphi$ be a reaction system for the transition rule φ. Then the subsystems

$$R_+ = R \cap (\mathscr{A}_{\varphi+} \times \mathscr{A}_{\varphi+}) \qquad \text{and} \qquad R_- = R \cap (\mathscr{A}_{\varphi-} \times \mathscr{A}_{\varphi-}) \tag{6.32}$$

of R are the systems of *positive* and *negative slope reactions*. The elements of $\operatorname{dom} R_-$ and $\operatorname{dom} R_+$ are the *positive* and *negative slopes*.

The slope subsystems of R therefore consist only of reactions between slopes of the same kind. The system R_+ consists of all reactions in R that transform a positive slope into a positive slope, while R_- consists of all the reactions that transform negative slopes into negative slopes. (Note that every interval situation is a negative and a positive slope and that therefore the reactions between intervals belong to both subsystems.)

The following lemma about slope subsystems is also important in its own right.

***Lemma* 6.19** (Slope Decomposition)**.** *Let Φ be a local reaction system and $a \in \Phi$. Then there exist situations $a_+ \in \operatorname{dom} \Phi_+$ and $a_- \in \operatorname{dom} \Phi_-$ such that $a \to_\Phi a_+ a_-$.*

Proof. If a is not the product of an element of $\operatorname{dom} \Phi_+$ with an element of $\operatorname{dom} \Phi_-$, then it must contain a $-_b$ operator left of an $+_c$ operator. Especially there must be a pair of $-_b$ and $+_c$ operator that are separated only by an interval.

This means that there must be a decomposition $a = a_1 \hat{b}_1 -_{b_1} u +_{b_2} \hat{b}_2 a_2$, where u is an interval, b_1 and b_2 are minimal separating intervals and $b_1 = u_L$ and $b_2 = u_R$. Then, since u is a separating interval, there is a reaction $u \to_\Phi b_1 +_{b_1} \hat{u} -_{b_2} b_2$. The application of this reaction to a results in

$$a \to_\Phi a_1 \underline{\hat{b}_1 -_{b_1} b_1 +_{b_1} \hat{u} -_{b_2} b_2} +_{b_2} \hat{b}_2 a_2 \tag{6.33}$$

We concentrate now on the underlined part of the reaction result. Because the result is achronal, there must be an interval u' such that $\hat{u} = \hat{b}_1 u'$. This means that we can apply a reaction of the form (6.27b) to $\hat{b}_1 -_{b_1} b_1 +_{b_1} \hat{u}$: We get then

the reaction $\hat{b}_1 -_{b_1} b_1 +_{b_1} \hat{u} = \hat{b}_1 -_{b_1} b_1 +_{b_1} \hat{b}_1 u' \to_\Phi \hat{b}_1 u' = \hat{u}$. Applying this reaction to the result of (6.33) results in

$$a \to_\Phi a_1 \underline{\hat{u} -_{b_2} b_2 +_{b_2} \hat{b}_2} a_2 \tag{6.34}$$

With the same kind of argument we can show that there is a reaction that replaces the underlined part of this reaction with $\hat{u}$, resulting in $a \to_\Phi a_1 \hat{u} a_2$.

This reaction has removed one $-_{b_1}$ and one $+_{b_2}$ operator from a. Repeating this removes all pairs of $-_b$ and $+_{b'}$ from a. The result is a reaction $a \to_\Phi a_+ a_-$ to a situation of the required form. □

The situation c that makes the two reactions $a \to_\Phi b_1$ and $a \to_\varphi b_2$ confluent will be constructed step by step in an induction proof. The following proposition is a technical lemma that will be used in Lemma 6.21 to transform b_1 and b_2 into successively better approximations of c.

***Lemma* 6.20** (Creation of a Minimal Separating Boundary)**.** *Let $a \in \operatorname{dom}\Phi$, where Φ is a local reaction system. Then at least one of the following cases occurs:*

1. *$\delta(a)_T \le 0$ and there is a reaction $a \to_\Phi a' -_u u$, where u is a right minimal separating interval,*
2. *$\delta(a)_T \ge 0$ and there is a reaction $a \to_\Phi u +_u a'$, where u is a left minimal separating interval, or*
3. *$\delta(a)_T = 0$, there is a reaction $a \to_\Phi v$ to a non-separating interval v, and $\operatorname{cl}^{(t)} \operatorname{pr}(a) = \emptyset$ for all $t \ge 1$.*

Proof. Let $\delta(a)_T \le 0$. Because of Lemma 6.19 there is a reaction $a \to_\Phi a_+ a_-$ with $a_+ \in \operatorname{dom}\Phi_+$ and $a_- \in \operatorname{dom}\Phi_-$.

If $\delta(a_-)_T < 0$, there must be a decomposition $a_- = a'_- -_v vx$, where v is a minimal separating interval and x an interval. Since vx is separating, there is by Lemma 6.12 a reaction $vx \to_\Phi (vx)_L +_{vx} \widehat{vx} -_{vx} (vx)_R$. So we have a reaction $a \to_\Phi a_+ a'_- -_v (vx)_L +_{vx} \widehat{vx} -_{vx} (vx)_R$ and case 1 occurs with $u = (vx)_R$.

If $\delta(a_-)_T = 0$, we must also have $\delta(a_+)_T = 0$ because $\delta(a_+)_T + \delta(a_-)_T = \delta(a)_T \le 0$ while $\delta(a_+) \ge 0$. Then $a_+ a_-$ is an interval. If $\operatorname{cl}^{(1)} \operatorname{pr}(a) \ne \emptyset$, then there must be a reaction $a_+ a_- \to_\Phi (a_+ a_-)_L +_{a_+ a_-} \widehat{a_+ a_-} -_{a_+ a_-} (a_+ a_-)_R$; then case 1 occurs with $u = (a_+ a_-)_R$. If $\operatorname{cl}^{(1)} \operatorname{pr}(a) = \emptyset$, then $a_+ a_-$ must be a non-separating interval because φ is well-behaved; then case 3 occurs with $v = a_+ a_-$.

The case of $\delta(a)_T \ge 0$ is handled in a mirror-symmetric way. □

We will now show a slightly stronger form of confluence, in order to get a good induction proof. In the following lemma we will say that two situations x_1 and x_2 are *equal until time t* if for all $\tau \le t$ we have $\operatorname{pr}(x_1)^{(\tau)} = pr(x_2)^{(\tau)}$.

***Lemma* 6.21** (Approximated Confluence)**.** *Let Φ be a local reaction system.*

If there are reactions $a \to_\Phi b_1$ and $a \to_\Phi b_2$ and the situations b_1 and b_2 are equal before time t, then there are situations $c_1, c_2 \in \operatorname{dom}\Phi$ that are equal until time $t+1$, and reactions $a \to_\Phi c_1$, $b_1 \to_\Phi c_1$, $a \to_\Phi c_2$ and $b_2 \to_\Phi c_2$.

In the following proof, $\mathscr{S}_{+-}$ is the set $\{ a_+ a_- : a_+ \in \Phi_+, a_- \in \Phi_- \}$.

Proof. Let b_1 and b_2 be equal before time t. If both $\mathrm{pr}(b_1)^{(t)}$ and $\mathrm{pr}(b_1)^{(t)}$ are empty, the lemma is trivially true, so we assume from now on that this is not the case.

We know already that $\delta(b_1) = \delta(b_2)$. With Lemma 6.19 we can also assume that b_1 and b_2 are elements of $\mathscr{S}_{+-}$.

If b_1 and b_2 are equal until time t, then there are situations x, y, b_1' and $b_2' \in \mathrm{dom}\,\Phi$ such that $b_1 = xb_1'y$ and $b_2 = xb_2y$, and t is the minimum of $\delta(x)$ and $\delta(xb_1)$. Since $\delta(b_1)$ and $\delta(b_2)$ are equal, $\delta(b_1')$ and $\delta(b_2')$ are equal too and also elements of $\mathscr{S}_{+-}$.

If $\delta(b_1')_T = \delta(b_2')_T > 0$, then there are by Lemma 6.20 two reactions $b_1' \to u_1 +_{u_1} b_1''$ and $b_2' \to u_2 +_{u_2} b_2''$, with intervals u_1 and u_2. We may assume without loss of generality that $|u_1| \leq |u_2|$. Then $\mathrm{pr}_x(u_1) \subseteq \mathrm{cl}\,\mathrm{pr}(a)$ and $\mathrm{pr}_x(u_2) \subseteq \mathrm{cl}\,\mathrm{pr}(a)$, therefore $\mathrm{pr}_x(u_2)|_{\mathrm{dom}\,\mathrm{pr}_x(u_1)} = \mathrm{pr}_x(u_1)$. So $\mathrm{pr}_x(u_1) \subseteq \mathrm{pr}_x(u_2)$, and if $u_1 \neq u_2$, then u_1 is an initial segment of u_2, which is impossible because u_2 is already a right minimal separating interval. So we must have $u_1 = u_2$. Then we can set $c_1 = xu_1 +_{u_1} b_1''y$ and $c_2 = xu_1 +_{u_1} b_2''y$; these situations are equal until time $t+1$. The same kind of argument works if $\delta(b_1') = \delta(b_2') < 0$.

Now we assume that $\delta(b_1')_T = \delta(b_2')_T = 0$. If b_1' and b_2 are intervals, then they must be equal, by an argument similar to that in the previous paragraph. We can then set $c_1 = c_2 = b_2$.

Otherwise, if b_1 is not an interval, it must still be an element of $\mathscr{S}_{+-}$, so there must be a reaction $b_1' \to u_+ +_{u_+} b_1'' -_{u_-} u_-$, with separating intervals u_+ and u_-. But this means that b_2 cannot react to a non-separating interval v: If this were the case, the process $\mathrm{pr}_x(u_+)$ would be a part of $\mathrm{pr}_x(v)$, but $\mathrm{pr}_x(u_+)$ is a separating interval and therefore cannot be part of a non-separating interval. So there must be a reaction $b_2' \to u_+ +_{u_+} b_2'' -_{u_-} u_-$, where the "re-use" of u_+ and u_- can be justified as in the previous paragraphs. In this case we have $c_1 = xu_+ +_{u_+} b_1'' -_{u_-} u_-$ and $c_2 = xu_+ +_{u_+} b_2'' -_{u_-} u_-$.

A similar argument can be used when b_2 is not an interval. This concludes the proof. □

***Theorem* 6.22** (Confluence)**.** *If there are reactions $a \to_\Phi b_1$ and $a \to_\Phi b_2$, then there is a situation $c \in \mathrm{dom}\,\Phi$ such that $b_1 \to_\Phi c$ and $b_2 \to_\Phi c$.*

Proof. We apply the induction steps outlined in Lemma 6.21.

Since b_1 and b_2 are finite, there is certainly a time t_0 such that b_1 and b_2 are equal until t_0. By repeated application of the lemma we get the four sequences of reactions

$$a \to_\Phi c_{1,k}, \qquad b_1 \to_\Phi c_{1,k}, \tag{6.35a}$$
$$a \to_\Phi c_{2,k}, \qquad b_2 \to_\Phi c_{2,k}, \tag{6.35b}$$

with $c_{1,k}$ and $c_{2,k}$ equal until time k. The only remaining question is whether this process stops after a finite number of steps.

To see this this we note that if $\mathrm{pr}(b_1)^{(t)} = \mathrm{pr}(b_1)^{(t)} = \emptyset$ for all time steps $t > t_0$, then the same is true for $c_{1,k}$ and $c_{2,k}$, and for arbitrary k. This can be verified by following the constructions in Lemma 6.19 and Lemma 6.21. □

6.8 Summary

In this chapter we have found a way to construct a reaction system from a transition rule.

Separating intervals played an important role. They allowed us to construct the set of *achronal situations*; and an easily recognisable subset of them, the ordered achronal sets, were shown to have always a closure. We have therefore found a subset of situations that generalises intervals but nevertheless consist of events at different times.

To prove this we had to restrict the set of transition rules a bit further, from interval-preserving to well-behaved rules. I expect that this restriction is only temporary and later may be loosened to allow for an extension of achronal sets to a larger class of transition rules.

For the moment we have nevertheless the definition of a reaction system that is usable for "naturally occurring" transition rules, like Rule 54 in the next chapter. This *local reaction system* was introduced and shown to have useful properties. Since it has the covering property, all information that can be found with help of the closure operator can also be found with reactions. We are therefore no longer dependent on processes to derive results on cellular automata.

Chapter 7

Rule 54

Up to now we have worked with cellular automata only in an abstract way. Now we will introduce a concrete cellular automaton which already has a complex behaviour.

The aim of this chapter is then to demonstrate the concepts of the previous chapters for an elementary cellular automaton, Rule 54. It also shows how structures of intermediate complexity manifest in the context of Flexible Time.

7.1 Elementary Cellular Automata

Rule 54 arises in the context of the *elementary cellular automata*. We have seen in the introduction that they are the one-dimensional cellular automata with radius 1 and $\Sigma = \{0, 1\}$ and that Stephen Wolfram [59] has provided an enumeration scheme for them.

In Wolfram's enumeration scheme we interpret the state set Σ as a set of integers. There is a number s such that $\Sigma = \{0, \dots, s-1\}$, and Σ can be seen as the set of digits for base s. A sequence of such digits is then an integer. Then we can view every state of the neighbourhood of a cell as a number with $2r+1$ digits, the code number for the neighbourhood. If we then enumerate the results of φ applied to every neighbourhood $w \in \Sigma^{2r+1}$ by the code number of w, the transition rule itself is another number under base s, this time with s^{2r+1} digits. This number is the code number for the function φ.

***Definition* 7.1** (Code numbers)**.** Let $\Sigma = \{0, \dots, s-1\} \subseteq \mathbb{N}_0$, $r \in \mathbb{N}$, and let $\varphi\colon \Sigma^{2r+1} \to \Sigma$ be a transition rule. For any sequence $w = \omega_0 \dots \omega_{2r} \in \Sigma^{2r+1}$, let $c(w) = \Sigma_{i=0}^{2r} \omega_i s^i$. Then the *code number* for φ is

$$\sum_{u \in \Sigma^d} \varphi(u) s^{c(u)}. \tag{7.1}$$

Cellular Processes as Diagrams. We need to determine the local reaction system for Rule 54. These computations involve some cellular processes, and the easiest way to write them down is as a rectangular diagram—especially

since these cellular processes will contain events from at most two different time steps.

Such a diagram may have the shape $\begin{smallmatrix} & & \tau & \\ \sigma_0 & \sigma_1 & & \sigma_2 \end{smallmatrix}$. This specific diagram describes a cellular process in which the cells in the states σ_0, σ_1 and σ_2 belong to time step 0 and the cell in state τ belongs to time step 1. In such diagrams the leftmost event in the bottom line always belongs to the space-time point $(0, 0)$, therefore the process can be written in the set notation as $\{[0, 0]\sigma_0, [0, 1]\sigma_1, [0, 2]\sigma_2, [1, 1]\tau\}$.

7.2 Basic Properties of Rule 54

I have chosen Rule 54 because it has some complex behaviour [4, 36], and it is a relatively simple rule in which an ether appears. An example for ether generation is Figure 3.1.

Rule 54 has the transition rule

$$\varphi_{54}(s) = \begin{cases} 1 & \text{for } s \in \{(0,0,1),(1,0,0),(0,1,0),(1,0,1)\}, \\ 0 & \text{otherwise.} \end{cases} \tag{7.2}$$

Note that φ_{54} is symmetric under the interchange of left and right.

Figure 7.1: "Rule icon" for Rule 54.

The rule can be described in a diagram in Figure 7.1. The diagram displays each of the eight possible 3-cell neighbourhoods together with the next state of the central cell. This diagram has been sometimes called the "Rule Icon"*.

The description in (7.2) is for a human reader (in contrast to a computer) difficult to memorise. A simpler description is the following slogan, which can be verified from Figure 7.1.

> "$\varphi_{54}(w) = 1$ if w contains at least one 1, except if the cells in state 1 touch."

Here we say that two cells "touch" if they are direct neighbours. Thus the two cells in state 1 touch in the neighbourhood $(1, 1, 0)$, but not in the neighbourhood $(1, 0, 1)$.

Interval Preservation. Next we must check whether Rule 54 is interval-preserving. To do this, we must test for all intervals $w \in \Sigma^k$ with $k \leq 3$ whether $\Delta\mathrm{pr}(w)$ is an interval under the transition rule φ_{54}. If this is true, then φ_{54} is interval-preserving by Theorem 5.11.

To do this we need a practical way to compute all the events determined by an interval w. Among them are the events that can be found directly by the transition rule, when applied to the intervals of length $2r + 1$, together with those events that are determined by smaller cellular processes. The transition rule implies that an interval w of length $2r + 1$ determines the event $[1, r]\varphi(w)$. We are now interested in all subsets of $\mathrm{pr}(w)$ that already determine the event

*For example by [54, p. 239] and in the "Wolfram Atlas"—see `http://atlas.wolfram.com/01/01/54/` for Rule 54.

$[1, r]\varphi(w)$. Since $\mathrm{pr}(w)$ is an interval, every subset of $\mathrm{pr}(w)$ can be extended to an interval by adding events. Therefore it is enough to search for the interval subsets of $\mathrm{pr}(w)$ that determine the event $[1, r]\varphi(w)$.

These intervals can be found by an application of Definition 3.8: If there is a decomposition $w = xw'y$ such that the value of $\varphi(xw'y)$ is independent of the contents of x and y, then $\mathrm{pr}_x(w')$ already determines $[1, r]\varphi(w)$.[†] This also means that the interval $\mathrm{pr}(w')$ determines the event $[1, r - |x|]\varphi(w)$. With this method we can find all the intervals of length $\leq 2r + 1$ that determine an event under the transition rule φ.

All this can then be expressed by a rule: *If $\varphi(xwy) = \sigma$ for all $x \in \Sigma^k$ and $y \in \Sigma^\ell$, then w determines the event $[1, r - k]\sigma$.* We will now find these reactions for Rule 54. Rule 54 has the following cases where this rule can be applied:

- There are two cases with $k = 1$ and $\ell = 0$, namely $\varphi(001) = \varphi(101) = 1$ and $\varphi(011) = \varphi(111) = 0$. The first equation shows that the interval 01 determines the event $[1, 0]1$ and the second that 11 determines $[1, 0]0$.
- There are two cases with $k = 0$ and $\ell = 1$, namely $\varphi(100) = \varphi(101) = 1$ and $\varphi(110) = \varphi(111) = 0$. The first equation proves that the interval 10 determines the event $[1, 1]1$ and the second that 11 determines $[1, 1]0$.

With these arguments we have found four new rules to find a determined event that belongs to an interval. With them it is now possible to find the next state of the middle cell for all intervals of length 3 that begin with 10 or 11 and for those that end with 01 or 11. There remain the reactions for the neighbourhoods that cannot be shortened in this way, namely 000 and 010. These neighbourhoods determine the events $[1, 1]0$ and $[1, 1]1$, respectively. To the other intervals we can apply one of the four new reactions to get the state of the middle cell one time step later. Therefore we have now the six cases where an interval determines an event,

$$\begin{aligned}
000 \text{ determines } [1, 1]0, &\qquad \begin{smallmatrix} & 0 & \\ 0 & 0 & 0 \end{smallmatrix} && (7.3a)\\
01 \text{ determines } [1, 0]1, &\qquad \begin{smallmatrix} & 1 & \\ 0 & & 1 \end{smallmatrix} && (7.3b)\\
11 \text{ determines } [1, 0]0, &\qquad \begin{smallmatrix} & 0 & \\ 1 & & 1 \end{smallmatrix} && (7.3c)\\
010 \text{ determines } [1, 1]1, &\qquad \begin{smallmatrix} & & 1 & \\ 0 & 1 & & 0 \end{smallmatrix} && (7.3d)\\
10 \text{ determines } [1, 1]1, &\qquad \begin{smallmatrix} & 1 & \\ 1 & & 0 \end{smallmatrix} && (7.3e)\\
11 \text{ determines } [1, 1]0\,. &\qquad \begin{smallmatrix} & 0 & \\ 1 & & 1 \end{smallmatrix} && (7.3f)
\end{aligned}$$

With these rules we can compute the events determined by an interval. At the right they are visualised with space-time diagrams. In them, the bottom line contains the process of w, and on top of it there is the event that is determined by it. Each of these diagrams, when applied to an interval, gives us the identity of one event that is determined by this interval.

No we can prove that φ_{54} is interval-preserving. For this we take all intervals whose length is at most $2r + 1 = 3$ and apply graphically all reactions to them that can be applied. We then get the following diagrams; all of them

[†]This means that the application of the transition rule must lead to the same result $\varphi(x'w'y') = \varphi(w)$ for all intervals $x', y' \in \Sigma^*$ with $|x'| = |x|$ and $|y'| = |y|$.

have $\mathrm{pr}(w)$ as their bottom row and $\Delta\mathrm{pr}(w)$ as the top row.

$$\begin{array}{cccccccccccccccc}
0 & & 1 & & & & & & 0\,0 & & \overset{1}{0\,1} & & \overset{1}{1\,0} & & \overset{0\,0}{1\,1} & \\
\\
\begin{matrix}0\\0\,0\,0\end{matrix} & & \begin{matrix}1\\0\,0\,1\end{matrix} & & \begin{matrix}1\,1\,1\\0\,1\,0\end{matrix} & & \begin{matrix}0\,0\\0\,1\,1\end{matrix} & & \begin{matrix}1\\1\,0\,0\end{matrix} & & \begin{matrix}1\\1\,0\,1\end{matrix} & & \begin{matrix}0\,0\,1\\1\,1\,0\end{matrix} & & \begin{matrix}0\,0\,0\\1\,1\,1\end{matrix}
\end{array} \tag{7.4}$$

They show that φ_{54} is interval-preserving for the intervals of length ≤ 3; by Theorem 5.11 it must then preserve all intervals.

Characteristic Reactions. The diagrams in (7.3) allow us now to write down the characteristic reactions for all those intervals that determine at least one event. Their characteristic reactions are

$$01 \to [1,0]1[-1,1], \tag{7.5a}$$
$$10 \to [1,1]1[-1,0], \tag{7.5b}$$
$$11 \to [1,0]00[-1,0], \tag{7.5c}$$
$$000 \to [1,1]0[-1,1], \tag{7.5d}$$
$$010 \to [1,0]111[-1,0]. \tag{7.5e}$$

Together, these reactions allow us to derive all the events that are determined

Figure 7.2: "Invariant Rule Icon" for Rule 54.

by a given cellular process. We can write them as an alternative form of the rule icon of Figure 7.1, one that in contrast to it does no longer depend explicitly on the radius. This diagram is shown in Figure 7.2.

As we will see later, the minimal separating intervals of Rule 54 are the intervals of length 2. For the first three of them we have just determined the characteristic reactions. The characteristic reaction for the last interval, 00, can now be determined according to Definition 5.21.

The characteristic reaction for the interval $a = 00$ must have the form

$$a \to [1,i]\hat{a}[-1,j], \tag{7.6}$$

and we must now determine i, j and $\hat{a}$. This can be done with help of reaction (7.5d). If we extend a to the left with 0, we get the reaction $0a \to [1,1]0[-1,1]$, therefore we must have $j = 1$. And if we extend a to the right with 0, we get $a0 \to [1,1]0[-1,1]$ and therefore $i = 1$. The only value for $\hat{a}$ for which $\delta(a) = \delta([1,i]\hat{a}[-1,j])$ is $a = [0]$, therefore the characteristic reaction for 00 is

$$00 \to [1,1][-1,1]. \tag{7.7}$$

Next we can see that the intervals of length 1 are not separating. This is because we cannot construct a characteristic reaction for them. For the "interval" consisting of a cell in state 0, we would have $i = 0$ in reaction (7.6) because of the characteristic reaction (7.5a) for 01 but $i = 1$ because of reaction (7.7). To verify that the cell in state 1 does not form a separating interval, note that (7.5a) requires $j = 1$ while (7.5c) requires $j = 0$.

On the other hand, Lemma 5.16 shows that the intervals of length 2 are separating, so they must be the minimal separating intervals. They and their characteristic reactions are shown in Figure 7.3.

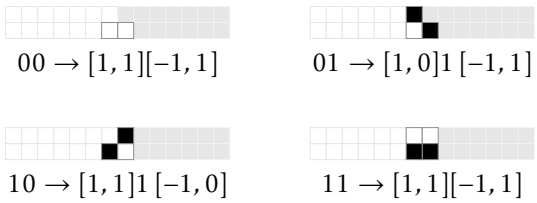


Figure 7.3: Separating intervals and their characteristic reactions for Rule 54.

Table 7.1: Characteristic reactions and generating slopes of Rule 54.

$b \to [1,i]\hat{b}[-1,j]$	$b +_b \hat{b}$	$\hat{b} -_b b$
$00 \to [1,1][-1,1]$	$00\oplus$	$\ominus 00$
$01 \to [1,0]1[-1,1]$	$01 \oplus_2 1$	$1 \ominus 01$
$10 \to [1,1]1[-1,0]$	$10 \oplus 1$	$1 \ominus_2 10$
$11 \to [1,0]00[-1,0]$	$11 \oplus_2 00$	$00 \ominus_2 11$

7.3 The Local Reaction System

With the characteristic reactions for the minimal separating intervals we can now determine the structure of the achronal situations for Rule 54.

Notation. According to Definition 6.5, the set of achronal situations is known when the generating slopes $\hat{b} -_b b$ and $b +_b \hat{b}$ are known for all minimal separating intervals b.

For concrete calculations, the repetition of b in these terms becomes however easily annoying. Therefore we will first introduce another, related, notation. It is a variant of the slope operators of Definition 6.3.

***Definition* 7.2** (Concrete Slope Operators)**.** Let $i \in \mathbb{Z}$. Then we write

$$\ominus_i = [-1, -i] \quad \text{and} \quad \oplus_i = [1, -i]. \tag{7.8}$$

If r is the radius of the transition rule, we will use the abbreviations $\ominus$ for $\ominus_r$ and $\oplus$ for $\oplus_r$.

This notation had been introduced in [51] and was already used in [52].

Achronal Situations. Now we will derive the generating slopes $\hat{b} -_b b$ and $b +_b \hat{b}$ from the characteristic reactions of Figure 7.3. They are listed in Table 7.1. Its first column contains the characteristic reactions, and the other columns contain the generating slopes derived from them.

First we need expressions for the slope operators $+_b$ and $-_b$ in terms of the new operators of Definition 7.2. We assume here, as before, that the characteristic reaction for every minimal separating interval b is $b \to [1,i]\hat{b}[-1,j]$. When we then use the notation of Definition 7.2, the two kinds of slope operators are related by the equations

$$+_b = [1, i - |b|] = \oplus_{|b|-i} \tag{7.9a}$$
$$-_b = [-1, j - |b|] = \ominus_{|b|-j} \tag{7.9b}$$

Here we have used the fact that b is a minimal separating interval and that therefore $b_L = b_R = b$.

With the equations in (7.9a) we now can derive the entries is the second and third column of Table 7.1 from the characteristic reactions in the first columns. This derivation consists of two steps. The first is finding the values of b and $\hat{b}$: They can be read of the characteristic reactions in the first column. The second step consists of finding the slope operators $+_b$ and $-_b$. I will now show this in detail for the second column.

We see from the first column of Table 7.1 that there are two kinds of characteristic reactions, namely those where the reaction product starts with $[1, 1]$ and those where it starts with $[1, 0]$. In the first case there is $i = 1$ and therefore $+_b = \oplus_{2-1} = \oplus$, and that is why in the second column of Table 7.1 the first and the third entry contains a $\oplus$ operator. In the second case there is $i = 0$ and $+_b = \oplus_{2-0} = \oplus_2$, and therefore the last second and fourth entry in the second column in Table 7.1 contain a $\oplus_2$ operator.

The same way we can derive the third column of Table 7.1.

Generating Reactions. Now, to complete the description of the local reaction system for Rule 54, we need to find its generating reactions. They are defined in Definition 6.13 and consist of two subsets: those that are associated to the minimal separating intervals, and those that are found by extending a minimal separating interval to the left or to the right.

(a) The first subset consists of the reactions $b \to_\Phi b +_b \hat{b} -_b b$ and $\hat{b} -_b b +_b \hat{b} \to_\Phi \hat{b}$ for all minimal separating intervals b. We do already know that

$$+_{00} = +_{10} = \oplus, \qquad +_{01} = +_{11} = \oplus_2, \tag{7.10a}$$

$$-_{00} = -_{01} = \ominus, \qquad -_{10} = -_{11} = \ominus_2 \tag{7.10b}$$

and that

$$\widehat{00} = [0], \qquad \widehat{01} = \widehat{10} = 1, \qquad \widehat{11} = 00\,. \tag{7.11}$$

With this we can calculate the reactions of the form $b \to_\Phi b +_b \hat{b} -_b b$ in the following way,

$$00 \to_\Phi 00 +_{00} \widehat{00} -_{00} 00 = 00 \oplus \ominus 00, \tag{7.12a}$$

$$01 \to_\Phi 01 +_{01} \widehat{01} -_{01} 01 = 01 \oplus_2 1 \ominus 01, \tag{7.12b}$$

$$10 \to_\Phi 10 +_{10} \widehat{10} -_{10} 10 = 10 \oplus 1 \ominus_2 10, \tag{7.12c}$$

$$11 \to_\Phi 11 +_{11} \widehat{11} -_{11} 11 = 11 \oplus_2 00 \ominus_2 11, \tag{7.12d}$$

and the reactions of the form $\hat{b} -_b b +_b \hat{b} \to_\Phi \hat{b}$ in the following way,

$$\ominus 00 \oplus = \widehat{00} -_{00} 00 +_{00} \widehat{00} \to_\Phi \widehat{00} = [0], \tag{7.13a}$$

$$1 \ominus 01 \oplus_2 1 = \widehat{01} -_{01} 01 +_{01} \widehat{01} \to_\Phi \widehat{01} = 1, \tag{7.13b}$$

$$1 \ominus_2 10 \oplus 1 = \widehat{10} -_{10} 10 +_{10} \widehat{10} \to_\Phi \widehat{10} = 1, \tag{7.13c}$$

$$00 \ominus_2 11 \oplus_2 00 = \widehat{11} -_{11} 11 +_{11} \widehat{11} \to_\Phi \widehat{11} = 11\,. \tag{7.13d}$$

The results of these two sets of calculations are collected in the left and right bottom fields of Table 7.3, respectively.

(b) The second subset of the reactions in Definition 6.13 consists of reactions of the form $\sigma b +_b \hat{b} \to_\Phi (\sigma b)_L +_{\sigma b} \widehat{\sigma b}$ and $\hat{b} -_b b\sigma \to_\Phi \widehat{b\sigma} -_{b\sigma} (b\sigma)_R$ with $\sigma \in \Sigma$, where b is a left or right minimal separating interval, respectively. We will now concentrate on the second type of reactions, which is sufficient because Rule 54 is symmetric.

To compute the reactions $\hat{b} -_b b\sigma \to_\Phi \widehat{b\sigma} -_{b\sigma} (b\sigma)_R$ for all right minimal separating intervals b we need to know $\widehat{b\sigma}$, $(b\sigma)_R$ and $-_{b\sigma}$ for every right minimal interval b and every $\sigma \in \Sigma$. And for this we first need to know the set of right minimal intervals for Rule 54.

The easiest way to do it is to start in greater generality and to determine the values of b_L and b_R for every separating interval b. In case of Rule 54 this is simple: Since every interval of length 2 is a minimal separating interval, b_L consists of the two leftmost events in b and b_R if the two rightmost events in b. (If b is separating, it must contain a minimal separating interval and is therefore at least two cells long.) For this reason the set of *left* and *right minimal separating intervals* under Rule 54 is both times Σ^2.

Next we need the characteristic reactions for all elements of Σ^3. For them we need, in turn, to know the cells determined by all the intervals in Σ^3. They can be determined from (7.3) and are

$$\begin{array}{cccccccc} {\scriptstyle 0} & {\scriptstyle 1} & {\scriptstyle 1\,1\,1} & {\scriptstyle 1\,0\,0} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0\,0\,1} & {\scriptstyle 0\,0\,0} \\ {\scriptstyle 0\,0\,0} & {\scriptstyle 0\,0\,1} & {\scriptstyle 0\,1\,0} & {\scriptstyle 0\,1\,1} & {\scriptstyle 1\,0\,0} & {\scriptstyle 1\,0\,1} & {\scriptstyle 1\,1\,0} & {\scriptstyle 1\,1\,1} \end{array}. \tag{7.14}$$

These diagrams are also a short description of the connection between b and $\hat{b}$ for all $b \in \Sigma^3$. We can derive the values of $\hat{b}$ for $b \in \Sigma^3$ directly from them,

$$\begin{aligned} &\widehat{000} = 0, \quad \widehat{010} = 111, \quad \widehat{100} = 1, \quad \widehat{110} = 001, \\ &\widehat{001} = 1, \quad \widehat{011} = 100, \quad \widehat{101} = 1, \quad \widehat{111} = 000\,. \end{aligned} \tag{7.15}$$

The values of $\hat{b}$ for all $b \in \Sigma^2$ are already listed in (7.11). To compute $(b\sigma)_R$ and $-_{b\sigma}$ we use the relations

$$(\sigma_1\sigma_2\sigma_3)_R = \sigma_2\sigma_3, \qquad\qquad -_{\sigma_1\sigma_2\sigma_3} = -_{\sigma_2\sigma_3} \tag{7.16}$$

for all σ_1, σ_2, $\sigma_3 \in \Sigma$; for the second relation the values of $-_{\sigma_2\sigma_3}$ can be found in (7.10b). With this data we can now calculate the reactions of the type $\hat{b} -_b b\sigma \to_\Phi \widehat{b\sigma} -_{b\sigma} (b\sigma)_R$ in the following way,

$$\ominus 000 = \widehat{00} -_{000} 000 \to_\Phi \widehat{000} -_{000} (000)_R = 0 \ominus 00, \tag{7.17a}$$

$$\ominus 001 = \widehat{00} -_{001} 001 \to_\Phi \widehat{001} -_{001} (001)_R = 1 \ominus 01, \tag{7.17b}$$

$$1 \ominus 010 = \widehat{01} -_{010} 010 \to_\Phi \widehat{010} -_{010} (010)_R = 111 \ominus_2 10, \tag{7.17c}$$

$$1 \ominus 011 = \widehat{01} -_{011} 011 \to_\Phi \widehat{011} -_{011} (011)_R = 100 \ominus_2 11, \tag{7.17d}$$

$$1 \ominus_2 100 = \widehat{10} -_{100} 100 \to_\Phi \widehat{100} -_{100} (100)_R = 1 \ominus 00, \tag{7.17e}$$

$$1 \ominus_2 101 = \widehat{10} -_{101} 101 \to_\Phi \widehat{101} -_{101} (101)_R = 1 \ominus 01, \tag{7.17f}$$

$$00 \ominus_2 110 = \widehat{11} -_{110} 110 \to_\Phi \widehat{110} -_{110} (110)_R = 001 \ominus_2 10, \tag{7.17g}$$

$$00 \ominus_2 111 = \widehat{11} -_{111} 111 \to_\Phi \widehat{111} -_{111} (111)_R = 000 \ominus_2 11\,. \tag{7.17h}$$

The reactions that are found in this calculation are shown in Table 7.2, both as formulas and as diagrams. Together with the reactions that are their left-to

Table 7.2: Diagrams of the generator reactions for Rule 54.

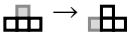
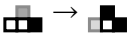
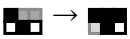
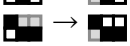
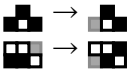
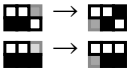
Reactions	Diagrams
$\ominus 000 \to 0 \ominus 00$	
$\ominus 001 \to 1 \ominus 01$	
$1 \ominus 010 \to 111 \ominus_2 10$	
$1 \ominus 011 \to 100 \ominus_2 11$	
$1 \ominus_2 100 \to 1 \ominus 00$	
$1 \ominus_2 101 \to 1 \ominus 01$	
$00 \ominus_2 110 \to 001 \ominus_2 10$	
$00 \ominus_2 111 \to 000 \ominus_2 11$	

Table 7.3: Generator reactions of the local reaction system for Rule 54.

Separating Intervals:	00, 01, 10, 11.	
Generating Slopes:	$\ominus 00$, $1 \ominus 01$, $1 \ominus_2 10$, $00 \ominus_2 11$,	$00\oplus$, $01 \oplus_2 1$, $10 \oplus 1$, $11 \oplus_2 00$.
Reactions:	$\ominus 000 \to 0 \ominus 00$	$000\oplus \to 00 \oplus 0$
	$\ominus 001 \to 1 \ominus 01$	$100\oplus \to 10 \oplus 1$
	$1 \ominus 010 \to 111 \ominus_2 10$	$010 \oplus 1 \to 01 \oplus_2 111$
	$1 \ominus 011 \to 100 \ominus_2 11$	$110 \oplus 1 \to 11 \oplus_2 001$
	$1 \ominus_2 100 \to 1 \ominus 00$	$001 \oplus_2 1 \to 00 \oplus 1$
	$1 \ominus_2 101 \to 1 \ominus 01$	$101 \oplus_2 1 \to 10 \oplus 1$
	$00 \ominus_2 110 \to 001 \ominus_2 10$	$011 \oplus_2 00 \to 01 \oplus_2 100$
	$00 \ominus_2 111 \to 000 \ominus_2 11$	$111 \oplus_2 00 \to 11 \oplus_2 000$
	$00 \to 00 \oplus \ominus 00$	$\ominus 00\oplus \to [0]$
	$01 \to 01 \oplus_2 1 \ominus 01$	$1 \ominus 01 \oplus_2 1 \to 1$
	$10 \to 10 \oplus 1 \ominus_2 10$	$1 \ominus_2 10 \oplus 1 \to 1$
	$11 \to 11 \oplus_2 00 \ominus_2 11$	$00 \ominus_2 11 \oplus_2 00 \to 11$

right mirror images they form the upper part of the "Reactions" section in Table 7.3.

This completes the calculation of the local reaction system for Rule 54. The result is a new form of the transition rule φ_{54}.

7.4 Understanding Rule 54 Better

While φ_{54} describes how a cell's neighbourhood influences its state in the next time step, each reaction in the local reaction system describes the relation between a separating interval π and the interval $\Delta\pi$ that is determined by it. The generating slopes in Table 7.3 describe the relations between the boundaries of π and $\Delta\pi$: The slope $00\oplus$ means that if π begins with 00, then $\Delta\pi$ reaches one cell to the right of the left end of π; the slope $10 \oplus 1$ tells that if π begins with 10, then the left end of $\Delta\pi$ reaches to the same position, but its leftmost event must be in state 1, and so on.

Even this localised knowledge helps us to understand the behaviour of Rule 54 better. For an example we use the task of finding the closure of an interval, something that we had already begun in Figure 5.3. We can now express the closure of an interval with a reaction $u \to_\Phi a_+a_-$, where a_+ is a positive and a_- a negative slope. The reaction system Φ has been constructed in such a way that there is a reaction in which the situations a_+ and a_- form the boundaries of the triangle, which we will now assume. Then the process of a_+ consists of the leftmost separating intervals of each time slice of $\operatorname{cl}\operatorname{pr}(u)$. The analysis of the previous paragraph then helps us to understand better the ragged boundaries of the closure in Figure 5.3.

Now let us add an event to the left side of the interval at the base of Figure 5.3. Then its closure grows too. The kind of growth, and how it depends on the added event, tells us how a change in the initial configuration is propagated to later time steps.

For Rule 54, this is expressed by the reactions in the top half of the "Reactions" section of Table 7.3. The reactions at the left side of the table show what happens when a cell is added to the right, and those at the right side of the table show what happens when a cell is added to the left.

The influence of the added event varies greatly depending on the states of the cells at the end of the original interval. We see from one pair of reactions, $010 \oplus 1 \to 01 \oplus_2 111$ and $110 \oplus 1 \to 11 \oplus_2 001$, that when the left side of the original interval is 00, the added event adds two events in the next time step; on the other hand, another pair, $001 \oplus_2 1 \to 00 \oplus 1$ and $101 \oplus_2 1 \to 10 \oplus 1$, proves that the closure may also stay unchanged. (It is peculiar to Rule 54 that the state of the added event has no influence on the number of cells that are added in the next time step, only on their states.)

We have therefore found for each cellular automaton a specific pattern of influence, described by the generators of the local reaction system.

7.5 Summary

In this chapter we have seen how the local reaction system is computed for a concrete rule.

In calculations with a concrete system, brevity is an advantage and redundancy is annoying. Therefore we used in this chapter the symbols $\ominus_k$ and $\oplus_k$ instead of $-_a$ and $+_a$ for the display of the resulting reaction system. In spite of this the computation of the reaction system may appear to be quite long and complex, with all the explanations given. If one leaves them out, it is however possible to do the whole work described here on a single piece of paper.

Nevertheless the resulting system in Table 7.3 looks somewhat voluminous when compared with the original description of Rule 54 in (7.2). This is caused, among other things, by the requirement that a local reaction system covers the full closure of each of its situations. If we drop this requirement, then we can create for special purposes reaction systems that are easier to describe and more powerful. One of them will be constructed in the next chapter.

The advantage of the large size of Table 7.3 is however that it provides additional information about the way in which information travels in the cellular automaton. This was not directly visible from the transition rule.

7.A Appendix: Rule 110

Here I will give another example and construct the local reaction system for another elementary cellular automaton, Rule 110. The derivation will be much more sketchy, but it should also show how the calculation of a concrete local reaction system can be done in a relatively small space.

Nature of the Rule. We will use as the initial description of the rule an icon similar to that of Figure 7.1. We see especially that Rule 110 is an asymmetric rule, in contrast to Rule 54.

Figure 7.4: "Rule icon" for Rule 110.

To understand Rule 110 better, we will now find a slogan for it, as we had done for Rule 54. This time the slogan becomes especially simple if we take the states 0 and 1 as Boolean values. Then we can write,

"$\varphi_{110}(\sigma_{-1}, \sigma_0, \sigma_1) = \sigma_0$ **xor** σ_1, except that $\varphi_{110}(0, 1, 1) = 1$."

Graphical Evolution. With Figure 7.4 we will now search for the cases where less than three cells determine the state of the central cell in the next time step. From the slogan we know that the value of this cell depends only on the interval consisting of the central cell and its neighbour, except when that interval is 11. For the other cases we can write this as the diagrams $\begin{smallmatrix}0\\0\,0\end{smallmatrix}$, $\begin{smallmatrix}1\\0\,1\end{smallmatrix}$ and $\begin{smallmatrix}1\\1\,0\end{smallmatrix}$, as we have done in (7.3). There is also a fourth case when two adjacent cells determine the next state of a cell: In the interval 01, the next state of the right cell is always 1. Thus we get the following diagrams,

$$\begin{smallmatrix}0\\0\,0\end{smallmatrix} \quad \begin{smallmatrix}1\\0\,1\end{smallmatrix} \quad \begin{smallmatrix}1\\1\,0\end{smallmatrix} \quad \begin{smallmatrix}1\\0\,1\end{smallmatrix}.$$

We can also see from Figure 7.4 that there are no shorter intervals that determine an event.

To get all the diagrams that are needed for a graphical evolution we have to add the diagram for the interval 111, because the next state of its central cell cannot be derived otherwise; then we get

$$\begin{smallmatrix}0\\0\,0\end{smallmatrix} \quad \begin{smallmatrix}1\\0\,1\end{smallmatrix} \quad \begin{smallmatrix}1\\1\,0\end{smallmatrix} \quad \begin{smallmatrix}1\\0\,1\end{smallmatrix} \quad \begin{smallmatrix}0\\1\,1\,1\end{smallmatrix}. \tag{7.18}$$

These diagrams are then enough to find for every cellular process the events that are determined by it.

Minimal Intervals. We have seen in Table 5.2 that Rule 110 is interval-preserving, and from Table 5.4 that its minimal separating intervals are 0 and 11. For the theory we will however need also the intervals that can occur as rightmost or leftmost minimal intervals. There is an additional leftmost minimal interval, 10, and another rightmost minimal interval, 01.

Characteristic Reactions. For these intervals we now determine the events that are determined by them, using the diagrams of (7.18). We get another list of diagrams,

$$\begin{array}{ccccc} 0 & & 1\,1 & & \begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \quad \begin{array}{c} 1 \\ 1\,0 \end{array} \end{array} . \tag{7.19}$$

The characteristic reactions for the minimally separating intervals are then

$$0 \to [1,0][-1,1], \qquad 11 \to [1,1][-1,1], \tag{7.20a}$$

$$01 \to [1,0]11[-1,0], \qquad 10 \to [1,0]1[-1,1]\,. \tag{7.20b}$$

We find the coefficients in the first reaction once we realise that the next state of a cell in state 0 is always the state of the cell at its right. Therefore, the boundary between the cells that are determined by the left side and those determined by the right side is at the left of the cell in state 0, and this is reflected in the reaction. For the second reaction we must note that the next state for the left cell in the interval 11 depends on information at the left, and the state of the right cell depends on information from the right. The last two reactions in (7.20) can be read directly from the diagrams.

Generating Slopes. From the characteristic reactions we derive the generating slopes. For the positive slopes we use the left coefficients of the characteristic reactions that belong to the left minimal separating intervals. We then get the situations $0\oplus$, $11\oplus$ and $01 \oplus_2 1$.

For the negative slopes we use the rightmost minimal separating intervals and the coefficients of the right side of the characteristic reactions and get $\ominus_0 0$, $\ominus 11$ and $11 \ominus_2 01$.

Reactions. The generating reactions can also be derived in a graphical manner, especially the slope reactions. We will give here only a few examples.

One example is the reaction $00\oplus \to 0 \oplus 0$. First we write down the diagram for the input situation, $\begin{array}{cc} & \\ 0 & 0 \end{array}$. This we extend with help of the rule for graphical evolution and get $\begin{array}{ccc} 0 & & \\ 0 & 0 & \end{array}$. We now mark on this diagram the rightmost positive slope. Then we get $\begin{array}{ccc} \boxed{0} & & \\ \boxed{0} & 0 & \end{array}$, and when we translate the sequence of marked events back into a positive slope, it is $0 \oplus 0$.

In a similar way the reaction $010 \oplus_2 1 \to 0 \oplus 11$ is derived. Here the diagram for the initial situation is $\begin{array}{ccc} & 1 & \\ 0 & 1 & 0 \end{array}$. We extend it to $\begin{array}{ccccc} & 1 & & 1 & \\ 0 & & 1 & & 0 \end{array}$ and mark in it the rightmost positive slope. The result is $\begin{array}{ccccc} \boxed{1} & & \boxed{1} & & \\ \boxed{0} & & 1 & & 0 \end{array}$, and the marked events belong to the situation $0 \oplus 11$.

The reactions in the bottom of Table 7.4 are derived more easily algebraically. For the bottom left block in that table we need to find reactions of the form $u \to u +_u \hat{u} -_u u$, as in (6.27a), where u is a minimal separating interval. In the case of the interval 11, we must therefore find generating slopes $u +_u \hat{u}$ and $u +_u \hat{u}$, with $u = 11$. We have already seen that these generating slopes are $11\oplus$ and $\ominus 11$. This leads to the reaction $11 \to 11 \oplus \ominus 11$.

For the bottom right block we need to find in the same way reactions of the form $\hat{u} -_u u +_u \hat{u} \to_\Phi \hat{u}$, as in 6.27b, again for the minimal blocking intervals. In the case of the interval 11 we get the reaction $\ominus 11 \oplus \to [0]$, since $\widehat{11} = [0]$.

Table 7.4: Generator reactions of the local reaction system for Rule 110.

Separating Intervals:		
Rightmost:	0, 01, 11.	
Leftmost:		0, 10 11.
Generating Slopes:	$\ominus_0 0$, $\ominus 11$, $11 \ominus_2 01$,	$0\oplus$, $11\oplus$, $10 \oplus_2 1$.
Reactions:	$\ominus_0 00 \to 0 \ominus_0 0$	$00\oplus \to 0 \oplus 0$
	$\ominus_0 01 \to 11 \ominus_0 01$	$10\oplus \to 10 \oplus_2 1$
	$11 \ominus_2 010 \to 11 \ominus_0 0$	$010 \oplus_2 1 \to 0 \oplus 11$
	$11 \ominus_2 011 \to 11 \ominus 11$	$110 \oplus_2 1 \to 11 \oplus 1$
	$\ominus 110 \to 1 \ominus_0 0$	$011\oplus \to 0 \oplus 11$
	$\ominus 111 \to 0 \ominus 11$	$111\oplus \to 11 \oplus 0$
	$0 \to 0 \oplus \ominus_0 0$	$\ominus_0 0\oplus \to [0]$
	$11 \to 11 \oplus \ominus 11$	$\ominus 11\oplus \to [0]$

Results. All these results are summarised in Table 7.4. It is clearly visible that this local reaction system has a completely different structure than that of Rule 54. But it is not yet clear what exactly this difference means.

Chapter 8

The Ether

Now we will use the reaction system we found in the previous chapter to describe ether formation under Rule 54. As a first step towards this goal we must find a description of the ether in terms of situations and reactions. We will take a more general point of view and explain how periodic structures in a cellular automaton are described in the formalism of Flexible Time.

First we introduce some concepts for simple periodic structures that occur in one-dimensional cellular automata. We will define them in general; then we introduce terms for the special forms they have under Rule 54. At the end of the chapter we will use them to describe why the ether under Rule 54 occurs spontaneously and why it is stable.

8.1 Triangles

As yest we know only the generating reactions of a local reaction system. All of them involve only a small number of cellular events. In order to understand the large-scale behaviour of a cellular automaton we need then to find reactions that involve a larger number of events. And in order to find general laws for the behaviour of the cellular automaton that can be expressed with a theorem or a formula, we need to find sets of situations that all behave in a similar way.

Families of Situations. This section is about the simplest form of such a general law, namely about the evolution of repeated patterns in the initial configuration of a cellular automaton. In case of Rule 54 we are interested in sequences of cells in the same state together with a boundary of a single cell in the opposite state, such as 011110 and 1000001. We can then decompose the initial configurations into such sequences, which overlap at their boundaries. A cell sequence ...100011110... can then be decomposed into 10001 and 011110, and when it is part of an interval situation it can be written as the overlapping product $10001\,\langle 01\rangle\,011110$. The boundary cells in the opposite state are included here because they make the decomposition uniquely determined, and also because the reactions that originate from them lead also to situations

that are useful in the model for ether formation at the end of this chapter. For a different kind of problem another decomposition might be more useful.

In the following definition this kind of decomposition is formalised and generalised, as much as we can do without making it difficult to handle. We especially drop the requirement that the repeated pattern must always be an interval: this would be unimportant for most calculations with situations.

***Definition* 8.1** (Family of Situations)**.** Let R be a reaction system and a, x, $b \in \operatorname{dom} R$ be situations. Then the set

$$\{\, ax^k b \colon k \geq 0 \,\} \tag{8.1}$$

is a *family of situations* in R.

The representation (8.1) is not the only one for a family of situations. For all constants n, $k_0 \in \mathbb{N}_0$ we have the equivalences

$$\{\, ax^k b \colon k \geq k_0 \,\} = \{\, (ax^{k_0})x^k b \colon k \geq 0 \,\}, \tag{8.2}$$

$$\{\, ax^{nk+k_0} b \colon k \geq 0 \,\} = \{\, (ax^{k_0})(x^n)^k b \colon k \geq 0 \,\}, \tag{8.3}$$

therefore the left sides of these equations are also valid representations of situation families. Now we make use of the first of these equations and introduce names for the two families of initial intervals described above:

$$F_0 = \{\, 10^k 1 \colon k \geq 1 \,\} \qquad \text{and} \qquad F_1 = \{\, 01^k 0 \colon k \geq 1 \,\}. \tag{8.4}$$

Layers. Our goal was to find general laws for the evolution of situation families like F_0 and F_1. We will now find the reactions that describe the evolution of the elements of such a situation family over a single time step. Most of these reactions, namely those that start from an element of a situation family that is larger than a certain minimal size, have a similar form.

This general form for a set of reactions is given in the next definition. The elements of these *layer reactions* will be used as building blocks for other reactions that extend over several time steps, therefore their name.

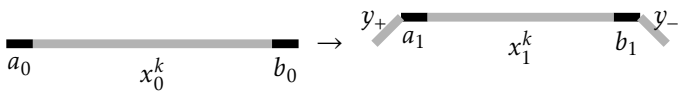


Figure 8.1: A layer reaction.

***Definition* 8.2** (Layer Reactions)**.** Let R be a reaction system and let $A_0 = \{\, a_0 x_0^k b_0 \colon k \geq 0 \,\}$ and $A_1 = \{\, a_1 x_1^k b_1 \colon k \geq 0 \,\}$ be two families of situations in $\operatorname{dom} R$. A set of reactions (Figure 8.1)

$$\{\, a_0 x_0^k b_0 \to_R y_+ a_1 x_1^k b_1 y_- \colon k \geq 0 \,\}, \tag{8.5}$$

with y_+, $y_- \in \operatorname{dom} R$ is then a *family of layer reactions* from A_0 to A_1.

In order to make the definition not unnecessarily specific, the requirement that the situation families A_0 and A_1 consist of intervals is not part of it. In the case that A_0 and A_1 are actually families of intervals, we have usually $y_+ \in \operatorname{dom} R_+$ and $y_- \in \operatorname{dom} R_-$. Then y_+ represents a step into the future and y_- the corresponding movement back to the past.

The following lemmas can be used to construct a family of layer reactions from simpler reactions. We will use them to find layer reactions for the interval families F_0 and F_1.

***Lemma* 8.3** (Repeatable Reactions)**.** *Let R be a reaction system. Then,*

$$\text{if} \quad ax \to_R ya, \quad \text{then} \quad ax^k \to_R y^k a \quad \text{for all } k \geq 0, \tag{8.6a}$$

$$\text{if} \quad xa \to_R ay, \quad \text{then} \quad x^k a \to_R a y^k \quad \text{for all } k \geq 0. \tag{8.6b}$$

Proof. The proof is by induction. The first equation is trivially true for $k = 0$. Now assume that $k \geq 0$ and $ax^k \to_R x^k a$. Then there is a chain of reactions $ax^{k+1} = ax^k x \to_R y^k ax \to_R y^k ya = y^{k+1} a$. Therefore we have by induction $ax^k \to y^k a$ for every $k \geq 0$. The second equation is proved in the same way. □

***Lemma* 8.4** (Generators of Layer Reactions)**.** *Let R be a reaction system which contains the reactions*

$$a \to_R a'c, \qquad cx \to_R yc \qquad \text{and} \qquad cb \to_R b'. \tag{8.7}$$

Then R has the family of layer reactions $\{ ax^k b \to_R a' y^k b' \colon k \geq 0 \}$.

Proof. By Lemma 8.3, applied to $cx \to_R yc$, we have for every $k \geq 0$ a reaction $cx^k \to_R y^k c$. We have then the chain of reactions $ax^k b \to_R a' cx^k b \to_R a' y^k cb \to_R a' y^k b'$, which proves the lemma. □

The layer reactions for the interval families F_0 and F_1 are then

$$L_0 = \{ 10^{k+2}1 \to_\Phi 10 \oplus 10^k 1 \ominus 01 \quad \colon k \geq 0 \}, \tag{8.8a}$$

$$L_1 = \{ 01^k 0 \quad \to_\Phi 01 \oplus_2 10^k 1 \ominus_2 10 \colon k \geq 2 \}. \tag{8.8b}$$

In the terminology of Definition 8.2, the set L_0 is a family of layer reactions from F_0 to $F_0 \cup \{11\}$, and L_1 is a family of layer reactions from F_1 to F_0. The set $F_0 \cup \{11\}$ is indeed a situation family: It is equal to $\{ 10^k 1 \colon k \geq 0 \}$.

For the proof of the first formula we use the reactions

$$10^2 \to_\Phi 10 \oplus 1(\ominus 00), \quad (\ominus 00)0 \to_\Phi 0(\ominus 00), \quad (\ominus 00)1 \to_\Phi 1 \ominus 01. \tag{8.9}$$

The common part $c = \ominus 00$ of these reaction is put in parentheses for better legibility. Then we can see that $a = 10^2$, $a' = 10 \oplus 1$, $x = y = 0$, $b = 1$ and $b' = 1 \ominus 01$ in the terminology of Lemma 8.4. Therefore there is a family of layer reactions $\{ 10^2 0^k 1 \to_\Phi 10 \oplus 10^k 1 \ominus 01 \colon k \geq 0 \}$ under Rule 54, the same as in (8.8a).

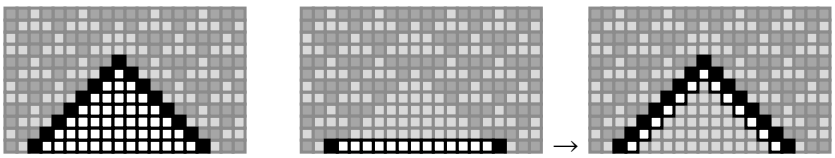
The second formula is derived from the reactions

$$01^2 \to_\Phi 01 \oplus_2 1(00 \ominus_2 11), \quad (00 \ominus_2 11)1 \to_\Phi 0(00 \ominus_2 11),$$
$$(00 \ominus_2 11)0 \to_\Phi 0^2 1 \ominus_2 10. \tag{8.10}$$

The common part of these reactions is once again put into parentheses. We can see from these reactions that $a = 01^2$, $a' = 01 \oplus_2 1$, $x = 1$, $y = b = 0$ and $b' = 0^2 1 \ominus_2 10$ in the terminology of Lemma 8.4. Therefore there is a family of reactions $\{\, 01^2 1^k 0 \to_\Phi 01 \oplus_2 10^k 0^2 1 \ominus_2 10 \colon k \geq 0 \,\}$ under Rule 54, the same as in (8.8b).

Triangles. As a final step in this subproject of finding general laws for the evolution of finitely often repeated patterns in the initial configuration, we now trace their evolution over as many time steps as possible. This is done with *triangle reactions*.

An example for a triangle reaction under Rule 54 is shown in Figure 8.2. It is a copy of Figure 6.2, but now it shows the reaction in the new notation, as $10^{13}1 \to_\Phi (10\oplus)^7 1(\ominus 01)^7$. The evolution diagram at the left side shows the closure of the process for initial situation $10^{13}1$. The two diagrams at the right side show the processes of the input and the result situation of the triangle reaction: they are the edges of the triangle process at the right. If we view the triangle as a temporal process, then its input situation shows two "particles", the interval situations 10 and 01, that are positioned at the boundaries of a sequence of 11 cells in state 0; the result situation shows how the particles move towards each other until they collide. This similarity to the collision of macroscopic particles makes triangle reactions a promising tool for understanding the behaviour of cellular automata.*



(a) Process (b) Reaction: $10^{13}1 \to (10\oplus)^7 1(\ominus 01)^7$

Figure 8.2: Triangle process and triangle reaction under Rule 54.

The following definition of triangle reactions harmonises with the definition of layer reactions. As before, the definition ignores the temporal aspect of the triangle reactions: In the cases that interest us here most, the situations a, x and b are intervals, while y_+ is an element of dom R_+ and y_- of dom R_-.

***Definition* 8.5** (Triangle Reactions)**.** Let R be a reaction system. Let $A = \{\, ax^k b \colon k \geq 0 \,\} \subseteq \operatorname{dom} R$ be a family of situations. A *family of triangle reactions* for A is a set of reactions (Figure 8.3),

$$\{\, ax^k b \to_R y_+^k c y_-^k \colon k \geq 0 \,\}. \tag{8.11}$$

As before with layers, there are many equivalent forms to represent triangle reactions. We are especially interested in the apparently more general form

$$\{\, ax^{kn+i} b \to_R y_+^{k+k_0} c y_-^{k+k_0} \colon k \geq 0 \,\} \tag{8.12}$$

*In the typical behaviour of Rule 54, the structures 10 and 01 are very short-lived [4]. For it and other cellular automata, the word "particle" means therefore generally larger, longer-lived structures. Nevertheless, the interpretation given here looks like a useful generalisation.

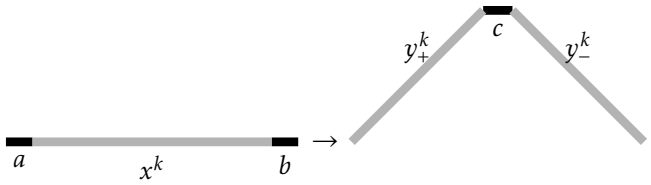


Figure 8.3: Generic form of a triangle reaction.

with n, i and $k_0 \in \mathbb{N}_0$. This set of reactions can be brought into the same form as (8.11) by writing it as $\{(ax^i)(x^n)^k b \to_R y_+^k (y_+^{k_0} c y_-^{k_0}) y_-^k : k \geq 0\}$ and so be shown to be equivalent to it. We can then bring the reaction of Figure 8.2 into the form (8.12) by writing it as $10^{2\times 6+1}1 \to_\Phi (10\oplus)^{6+1}1(\ominus 01)^{6+1}$.

Triangle reactions can be derived from layer reactions. If $A = \{ax^k b : k \geq 1\}$ is a family of situations and L is a family of layer reactions from A to $A \cup \{ab\}$, then there is a family of triangle reactions for A. The following lemma shows why. (The requirement in the lemma that there is a reaction $ab \to_R c$ is no restriction, since one can always set $c = ab$.)

***Lemma* 8.6** (Generators of Triangle Reactions)**.** *Let R be a reaction system. If there is a reaction $ab \to_R c$ and a family of layer reactions*

$$L = \{ax^{k+1} b \to_R y_+ ax^k b y_- : k \geq 0\}, \tag{8.13}$$

then there exists in R a family of triangle reactions

$$\{ax^k b \to_R y_+^k c y_-^k : k \geq 0\}. \tag{8.14}$$

Proof. Let a $k \geq 0$ be given. We can then apply for k times one of the reactions of L, first to the situation $ax^k b$, and then always to the result of the previous reaction. Then we have found a reaction $ax^k b \to_R y_+^k ab y_-^k$. Next we apply the reaction $ab \to_R c$ to the result of this reaction and get $ax^k b \to_R y_+^k c y_-^k$. □

An interesting phenomenon occurs when the layer reaction consumes more that one copy of the repeated pattern x. Then the one family of layer reactions splits into several sets of triangle reactions. This occurs especially in the family of layer reactions L_0 for Rule 54.

***Lemma* 8.7** (Multiple Triangle Reactions)**.** *Let R be a reaction system that contains a family of layer reactions,*

$$L = \{ax^{k+n} b \to_R y_+ ax^k b y_- : k \geq 0\}, \tag{8.15}$$

and the reactions $ax^i b \to_R c_i$ for $i = 0, \ldots, n-1$. Then R contains for every value of i a family of triangle reactions,

$$\{ax^{kn+i} b \to_R y_+^k c_i y_-^k : k \geq 0\}. \tag{8.16}$$

Proof. We will write the terms x^k and x^{k+n} in differing ways, depending on the value of k. If there is a $j \in \mathbb{N}_0$ such that $k = jn + i$ and $i < n$, then $x^k = (x^n)^j$

Table 8.1: Large-scale reactions under Rule 54

Name	Family of reactions
T_{00}	$\{10^{2k}1 \to_\Phi (10\oplus)^k 11(\ominus 01)^k : k \geq 0\}$
T_{01}	$\{10^{2k+1}1 \to_\Phi (10\oplus)^{k+1} 1(\ominus 01)^{k+1} : k \geq 0\}$
L_1	$\{01^k 0 \to_\Phi 01 \oplus_2 10^k 1 \ominus_2 10 : k \geq 2\}$

and $x^{k+n} = x^i(x^n)^j$. Therefore L is the disjoint union of n families of layer reactions,

$$L_i = \{(ax^i)(x^n)^j b \to_R y_+ a(x^n)^j b y_- : j \geq 0\} \qquad \text{for } i = 0, \ldots, n-1. \tag{8.17}$$

Then we can apply the previous lemma to each of the reactions families L_i. We get for every value of i a family of triangle reactions, $\{(ax^i)(x^n)^k b \to_R y_+^k c_i y_-^k : j \geq 0\}$, the same as in (8.16). □

Now we will derive triangles and the triangle reaction for Rule 54. Among the two families of layer reactions only L_0 fulfils the requirement of Lemma 8.7. So we set $L_0 = \{10^{k+2}1 \to_\Phi 10 \oplus 10^k 1 \ominus 01 : k \geq 0\}$ for L and choose as the finishing reactions $ax^i b \to_R c_i$ the two reactions $11 \to_\Phi 11$ and $101 \to_\Phi 10 \oplus 1 \ominus 01$. Then we can apply the lemma and get the following families of triangle reactions,

$$T_{00} = \{10^{2k}1 \to_\Phi (10\oplus)^k 11(\ominus 01)^k : k \geq 0\}, \tag{8.18a}$$

$$T_{01} = \{10^{2k+1}1 \to_\Phi (10\oplus)^k (10 \oplus 1 \ominus 01)(\ominus 01)^k : k \geq 0\}. \tag{8.18b}$$

We can view these two families of reactions as an improved version of the layer reactions L_0, and as their replacement. For L_1 no such replacement has been found, therefore the general laws for Rule 54 that we have found with the methods of this section are T_{00}, T_{01} and L_1. They are listed in Table 8.1.

8.2 What Is the Ether?

Now we can try to understand ether formation under Rule 54. For this we need to express explicitly what the ether is. We will do this first in terms of configurations and then in terms of situations and reactions.

In Figure 3.1 at the beginning we have seen an example of ether formation. In it we can now recognise how, when starting from a random initial configuration, large regions that consist of a regular pattern begin to form. We see that this pattern consists at alternating times of a configuration in which blocks of three cells in state 0 are separated by one cell in state 1, and of a configuration in which three cells in state 1 are separated by one cell in state 0. After two time steps the same patterns arise again, but but shifted to the side by a distance of two cells, such that a true repetition only occurs after 4 time steps. We will speak of the two configurations as the *phases* of the ether, in analogy to the usage of Martínez et al. [40] for Rule 110.

We also see that there are several such ether regions in Figure 3.1. They are separated by larger structures between them, or sometimes just by phase differences. This is a typical phenomenon in ether formation: when starting

from a random configuration one almost never gets a "pure", or empty ether. Instead one gets large regions with a regular pattern that are separated by disturbances. Over time the ether regions coalesce and the distance between the disturbances increases. This has already been noted by Boccara, Nasser and Roger [4]. We will therefore first describe an empty ether and then turn to the more realistic case of an ether with disturbances.

The Empty Ether. A configuration of empty ether then consists of an infinite repetition of one of the patterns $0^3 1$ and $1^3 0$. In contrast to configurations, situations are always finite. We could now represent a finite part of such an ether configuration by a situation of the form $(0^3 1)^k$ or $(1^3 0)^k$, depending on the phase. We choose however the family of situations $\{\, 1(0^3 1)^k \colon k \geq 0 \,\}$ as our standard representative of an ether configuration. The reason for the choice of this phase and for adding a 1 at the left is that then the situation $10^3 1$ is an element of the family and that this situation is the input of one of the triangle reactions of the family T_{01}, namely $10^3 1 \to_\Phi (10\oplus)^2 1(\ominus 01)^2$.

In order to get a notation in which no overlapping situations occur we will write this reaction in a slightly different form,[†]

$$10^3 1 \to_\Phi 1(0 \oplus 1)^2 (\ominus 01)^2 \,. \tag{8.19}$$

The next step to construct a description of the empty ether is then to find reactions similar to (8.19) for all intervals of the form $1(0^3 1)^k$. To do this we need the following lemma, which provides a kind of converse to the reaction (8.19).

***Lemma* 8.8** (Converse Triangle Reaction)**.**

$$1(\ominus 01)^2 (0 \oplus 1)^2 \to_\Phi 10^3 1 \,. \tag{8.20}$$

Proof. First we show that $1 \ominus 010 \oplus 1 \to_\Phi 1^3$. This is true because

$$\underline{1 \ominus 010} \oplus 1 \to_\Phi 111 \underline{\ominus_2 10 \oplus 1} \to_\Phi 111; \tag{8.21}$$

the terms that change in the next reaction step are underlined. We have then also shown that $1(\ominus 01)^2 (0 \oplus 1)^2 \to_\Phi 1 \ominus 01^3 1 \oplus 1$ and only must verify that the reaction product reacts to $10^3 1$:

$$\underline{1 \ominus 011} \underline{10 \oplus 1} \to_\Phi 1 \underline{00 \ominus_2 111} \oplus_2 001 \tag{8.22}$$

$$\to_\Phi 1 \underline{000 \ominus_2 11 \oplus_2 001} \to_\Phi 10001 \,. \tag{8.23}$$

This proves the lemma. □

We now combine (8.20) and (8.19) to a reaction in which the interval $10^3 1$ does no longer occur,

$$1(\ominus 01)^2 (0 \oplus 1)^2 \to_\Phi 1(0 \oplus 1)^2 (\ominus 01)^2 \,. \tag{8.24}$$

[†]The right site of the reaction will be later decomposed into the situations 1, $(0 \oplus 1)^2$ and $(\ominus 01)^2$. These situations do not overlap, even if two of their *processes*, namely $\mathrm{pr}_1((0 \oplus 1)^2)$ and $\mathrm{pr}_{1(0\oplus 1)^2}((\ominus 01)^2)$, do overlap.

The reaction (8.19) and this reaction are the building blocks for the larger reactions that originate from the $1(0^31)^k$ terms. In order to express them better we introduce the abbreviations

$$\varepsilon_- = \ominus 01 \qquad \text{and} \qquad \varepsilon_+ = 0 \oplus 1\,. \tag{8.25}$$

Note that these terms are not achronal. In the formalism they can therefore appear only as factors of situations in $\operatorname{dom}\Phi$, not as terms in their own right.

With the abbreviations of (8.25) we will now summarise the reactions that characterise the empty ether.

***Definition* 8.9** (Basic Ether Reactions)**.** The *basic ether reactions* for Rule 54 are

$$10^31 \to_\Phi 1\varepsilon_+^2\varepsilon_-^2 \qquad \text{and} \qquad 1\varepsilon_+^2\varepsilon_-^2 \to_\Phi 1\varepsilon_+^2\varepsilon_-^2\,. \tag{8.26}$$

The basic ether reactions are shown in Figure 8.4. The cells of the ether situations are shown in black and white, and the background in grey tones is a finite part of the empty ether.

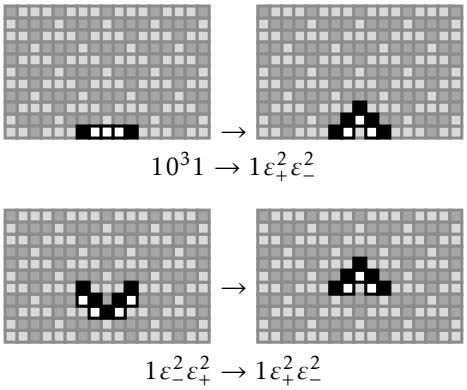


Figure 8.4: Basic ether reactions on a background of empty ether.

From these two reactions we will now derive two sets of reactions that involve larger parts of the empty ether. The set one describes the evolution of larger segments of the initial configuration, and the second one is a description of the evolution of a part of the empty ether.

***Theorem* 8.10** (Ether Reactions)**.**

$$1(0^31)^k \to_\Phi 1\varepsilon_+^{2k}\varepsilon_-^{2k} \qquad \text{for } k \geq 0, \tag{8.27a}$$

$$1\varepsilon_-^{2k}\varepsilon_+^{2\ell} \to_\Phi 1\varepsilon_+^{2\ell}\varepsilon_-^{2k} \qquad \text{for } k, \ell \geq 0. \tag{8.27b}$$

Proof. We need for this proof a reaction that is derived from the reaction at the right side of (8.26) with the help of Lemma 8.3.

In order to apply this lemma to the reaction $1\varepsilon_-^2\varepsilon_+^2 \to_\Phi 1\varepsilon_+^2\varepsilon_-^2$ we need to write it in the form $xa \to_\Phi ay$. We do this by resolving the ε_- terms at the left side of the reaction. Then we get $1 \ominus 01 \ominus 0(1\varepsilon_+^2) \to_\Phi (1\varepsilon_+^2)\varepsilon_-^2$, with the repeated factor a put in parentheses to emphasise it. Now we can apply (8.6b) and get

a reaction $(1 \ominus 01 \ominus 0)^k 1\varepsilon_+^2 \to_\Phi 1\varepsilon_+^2(\varepsilon_-^2)^k$ for every $k \geq 0$. Then we rewrite the left side of this reaction with ε_- terms. The result is

$$1\varepsilon_-^{2k}\varepsilon_+^2 \to_\Phi 1\varepsilon_+^2\varepsilon_-^{2k} \qquad \text{for all } k \geq 0. \tag{8.28}$$

We use it to prove the reactions in (8.27) by induction over k. For both families of reactions the case $k = 0$ is trivially true, and for $k \geq 1$ we get

$$\begin{aligned}
1(0^31)^k &\to_\Phi 1(0^31)^{k-1}\varepsilon_+^2\varepsilon_-^2 \\
&\to_\Phi 1\varepsilon_+^{2(k-1)}\varepsilon_-^{2(k-1)}\varepsilon_+^2\varepsilon_-^2 && \text{by induction,} \\
&\to_\Phi 1\varepsilon_+^{2(k-1)}\varepsilon_+^2\varepsilon_-^{2(k-1)}\varepsilon_-^2 && \text{with (8.28)} \\
&= \varepsilon_+^{2k}\varepsilon_-^{2k}, && \text{(8.29a)} \\
1\varepsilon_-^{2k}\varepsilon_+^{2\ell} &\to_\Phi 1\varepsilon_-^2\varepsilon_+^{2\ell}\varepsilon_-^{2(k-1)} && \text{by induction,} \\
&\to_\Phi 1\varepsilon_+^{2\ell}\varepsilon_-^2\varepsilon_-^{2k-1} && \text{with (8.28).} \\
&= 1\varepsilon_+^{2\ell}\varepsilon_-^{2k}. && \text{(8.29b)}
\end{aligned}$$

This then proves the theorem. □

The reactions of Theorem 8.10 now serve as an inspiration for the way in which one can express the disturbed ether that arises from a random initial configuration in terms of situations and reactions.

Random Initial Situations. We need a method to express with the help of situations the behaviour of random initial configurations. First we define what we mean by a *random initial configuration*. Here we assume that the states of a configuration $c \in \Sigma^{\mathbb{Z}}$ are chosen at random in such a way that the state of every cell is equal to 1 with probability p_1 and that the states of all cells are independent of each other. We will always exclude the trivial cases $p_1 = 0$ and $p_1 = 1$.

In the language of probability theory [22, Chapter 1] we have then performed a *random experiment*. The random choice of a configuration is represented by a *random variable* C with values in $\Sigma^{\mathbb{Z}}$. As it is usual in probability theory, this and other random variables are written in capital letters. The configuration c mentioned above is then a possible *outcome* of the experiment. The probabilities that define the random experiment will however refer to whole sets of outcomes. These sets are called in probability theory "events", but here they will be called *stochastic events*, to avoid confusion with the already existing use of the word "event" in the context of cellular processes.

First we must now specify the probabilities for the outcomes of C and make precise the informal definition given above. We can express the random choice of C by saying that for all $x \in \mathbb{Z}$,

$$P(C(x) = 1) = p_1, \tag{8.30}$$

In a more formal way of speaking, this equation assigns the probability p_1 to the set of outcomes, or stochastic event, $\{\, c \in \Sigma^{\mathbb{Z}} \colon c(x) = 1 \,\}$. Other expressions with random variables are understood in a similar way.

From (8.30) we can derive the probailities for other stochastic events. These are the events that belong to the σ-field [22, p. 19] that is generated by the sets

$\{ c \in \Sigma^{\mathbb{Z}} : c(x) = 1 \}$. We will however not use this σ-field explicitly and work with probabilities in a less formal way.

For the work with situations we can only use finite pieces of the initial configuration C. So we define now the sequence $(U_n)_{n\geq 0}$ of *random intervals*. Each U_n is a random variable with values in Σ^n and defined by the relation

$$U_n(c) = C(-n+1)\dots C(1)C(0)\,. \tag{8.31}$$

With this definition we have a growing sequence of random cellular processes,

$$\cdots \supset \mathrm{pr}([-n]U_n) \supset \cdots \supset \mathrm{pr}([-1]U_1) \supset \mathrm{pr}([0]U_0) = \emptyset, \tag{8.32}$$

that ultimately incorporate all cells of Σ^Z with indices ≤ 0. For each n, the closure of $\mathrm{pr}([-n]U_n)$ then represents a view into the evolution of an element of C. The larger n becomes, the larger the window on C becomes. All these windows contain only events with non-positive space coordinates, but this does not matter, since the probability distribution for C is invariant under left-to-right shifts.

Later we will need to do induction proofs over the length of the intervals U_n. In them we need to express the sequence $(U_n)_{n\geq 0}$ in a recursive way. To do this we now imagine the construction of the U_n as a stochastic process in which the random interval U_{n-1} is extended by an event that is in state 1 with probability p_1 and in state 0 with probability $1 - p_1$. We can then express the values of the $P(U_n)$ in the language of conditional probability with the equations

$$P(U_n = \sigma u \mid U_{n-1} = u) = \begin{cases} p_1 & \text{if } \sigma = 1, \\ (1-p_1) & \text{if } \sigma = 0, \end{cases} \tag{8.33}$$

valid for all $n \geq 1$ and $u \in \Sigma^n$. The starting point of this recursion is the trivial case of $n = 0$, with $P(U_0 = [0]) = 1$.

Ether Fragments. We also need a way to express the fact that the closure of $\mathrm{pr}(U_n)$ contains fragments of the empty ether. To do this we will first consider reactions of the form

$$U_n \to a_+ a_- \tag{8.34}$$

with $a_+ \in \mathrm{dom}\,\Phi_+$ and $a_- \in \mathrm{dom}\,\Phi_-$. They are a generalisation of the triangle reactions (8.27a) for the empty ether. Since however U_n has in general not the form $1(0^3 1)^k$, the situation a_+ will in general not consist entirely of ε_+ terms and a_- not only of ε_- terms. Our goal in this chapter then is to prove that for large n, the closure of $\mathrm{pr}(U_n)$ nevertheless contains pieces of the ether.

We will therefore look for reactions of the form (8.34) in which a_+ contains factors of the form $1\varepsilon_+^2$. The exponent of 2 in ε_+^2 occurs because all the ether reactions in (8.27) involve only even numbers as exponents for ε_+ and ε_-. We will say that an *ether fragment* occurs at time t if

$$a_+ = a'_+ 1 \varepsilon_+^2 a''_+ \qquad \text{and} \qquad \delta(a'_+)_T = t\,. \tag{8.35}$$

The situation $a_+ a_-$ serves here as a probe into the closure of $\mathrm{pr}(U_n)$. We can then use it as a means to express ether formation in such a way that it can be proved with the help of situations and reactions.

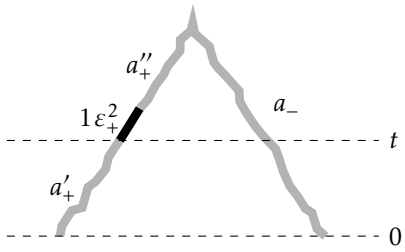


Figure 8.5: Location of the ether fragments of (8.35) in the situation a_+a_-.

We could also have looked for reactions in which a_- contains factors of the form $\varepsilon_+^2 1$. The results would be equivalent because Rule 54 is symmetric under exchange of left and right. The advantage of the definition we use here is that with it we can express more easily the time at which an ether fragment occurs.

Reactions and Probability. Now we will consider arbitrary reactions of the form $U_n \to_\Phi a$ for a given situation a. Whether such a reaction is possible depends on the value of U_n. We will write the probability of such a reaction as $P(U_n \to_\Phi a)$; it is the probability that U_n belongs to the set of situations $u \in \Sigma^n$ for which there is a reaction with result a. So we can write

$$P(U_n \to_\Phi a) = P(\exists u \in \Sigma^n \colon u \to_\Phi a)\,. \tag{8.36}$$

A generalisation of this is the case where a random interval may react to a whole set of possible outcomes. When $\mathscr{A} \subseteq \mathscr{S}$ is a set of situations, we will write

$$P(U_n \to_\Phi \mathscr{A}) = P(\exists u \in \Sigma^n, a \in \mathscr{A} \colon u \to_\Phi a)\,. \tag{8.37}$$

This definition counts the probability that the random situation U_n can react to at least one element of $\mathscr{A}$. So it is meaningful even if there is more than one possible reaction result for a certain value of u. We have therefore always $P(U_n \to_\Phi \mathscr{S}) = 1$.

Notation for Fragments. The definitions of the last paragraphs are aimed at a special kind of situations that are needed in the following proof. We need to express that a certain situation f, the "fragment", occurs at one or more specified time steps. The set of situations that have in common a certain fragment $f \in \mathscr{S}$, beginning at one of the time steps $t_1, \dots, t_n$, is then written

$$\mathscr{F}(f, t_1, \dots, t_n) = \{\, a_+ f a' \colon a_+ \in \operatorname{dom}\Phi_+, a' \in \operatorname{dom}\Phi, \\ \delta(a_+)_T \in \{t_1, \dots t_n\}\,\}\,. \tag{8.38}$$

Among these sets of situations, the most important one is that in which f is an ether fragment; it is called

$$\mathscr{E}(t_1, \dots, t_n) = \mathscr{F}(1\varepsilon_+^2, t_1, \dots, t_n)\,. \tag{8.39}$$

8.3 The Ether Is Inevitable

The following theorem is a way to express the necessity of ether formation in a weak form. It shows that for a random initial configuration, ether fragments can be found arbitrarily far in the future. Ether fragments never vanish totally, and we can view that as a sign (but not a proof) that the ether, too, persists.

***Theorem* 8.11** (Ether Formation)**.** *Let $\varepsilon > 0$ be a probability. The random intervals $(U_k)_{k\geq 0}$ are defined as in* (8.31)*. Then for every $t \in \mathbb{N}$ there is an $n \in \mathbb{N}$ such that*

$$P(U_n \to_\Phi \mathscr{E}(t, t+1, t+2)) \geq 1 - \varepsilon\,. \tag{8.40}$$

Since U_n is part of the random initial configuration C, the theorem shows that every given sequence of three time steps contains with probability 1 the starting time of an ether fragment. Moreover, C consists of infinitely many intervals of length n and these intervals are stochastically independent of each other. Therefore the theorem also shows that during every sequence of three steps time, infinitely many ether fragments begin.

The theorem will now be proved with help of several lemmas. Among them, Lemma 8.12 shows how ether fragments arise from the initial configuration, and Lemma 8.16 shows how they propagate to later times.

***Lemma* 8.12** (Formation of Ether Fragments)**.** *Let $(U_k)_{k\geq 0}$ be the sequence of random intervals from* (8.31)*. Let $n \in \mathbb{N}_0$ and let $u \in \Sigma^n$ be an interval. Then for every $\varepsilon > 0$ there is an integer $m > n$ such that*

$$P(U_m \to_\Phi \mathscr{E}(0) \mid U_n = u) \geq 1 - \varepsilon\,. \tag{8.41}$$

Proof. When extending a random interval U_k by 5 events to U_{k+5}, the probability that U_{k+5} begins with $10^3 1$ is

$$P(U_{k+5} = 10^3 1 U_k) = p_1^2 (1-p_1)^3, \tag{8.42}$$

independent of the value of U_k. We will call this probability $p_{10^3 1}$. Then $p_{10^3 1} > 0$ because we have required earlier that $p_1 \notin \{0, 1\}$.

Let now $k_0 \geq 0$ be an integer. Then the probability that one of the random intervals U_{n+5k} with $k \leq k_0$ begins with $10^3 1$ is $(1 - p_{10^3 1})^{k_0}$. These probabilities are independent of each other because each of them only depends on the states of the cells at the locations $-(n+5k)+1, \ldots, -(n+5k)+5$, and those intervals do not overlap.‡

Since $1 - p_{10^3 1} < 1$, we can find a k_0 such that $(1 - p_{10^3 1})^{k_0} < \varepsilon$. Now we can set $n_0 = n + 5k_0$. Then with a probability greater than $1 - \varepsilon$ there is a $m \leq n_0$ such that U_m begins with $10^3 1$. With such an m we also have a reaction $U_m \to_\Phi 1\varepsilon_+^2\varepsilon_-^2 U_{m-5}$. □

So we see that as n approaches infinity, there is always an ether fragment $1\varepsilon_+^2$ at time 0 that is generated by U_n. Next we will prove that these fragments are almost always copied by other reactions to space-time locations at later times.

‡Note that the intervals U_k grow to the left, as described in (8.31).

For this proof we will need reactions that create a situation with an interval of a given minimal length and at a specified time as a factor. Furthermore, initial and final situation of the reaction must be positive slopes. This means that the required reaction must be a part of the positive slope reaction system Φ_+.

The following lemma then tells us how we can find reactions that create such an interval from a given situation $a_+ \in \operatorname{dom}\Phi_+$. When the conditions of the lemma are met, we will say that every *u enforces v*.

***Lemma* 8.13** (Enforcement of Intervals)**.** *Let $a_+ \in \operatorname{dom}\Phi_+$ be chosen arbitrarily. Then for every $\ell > 0$ there is a $k > 0$ such that for all $u \in \Sigma^*$ with $|u| \geq k$ there is a reaction of the form*

$$u a_+ \to_{\Phi_+} b_+ v \tag{8.43}$$

with $b_+ \in \operatorname{dom}\Phi_+$ and $v \in \Sigma^\ell$.

Proof. We first consider the case where $\ell = 1$ and $\delta(a_+)_T = 1$. Because a_+ is achronal we must then have a decomposition $a_+ = u_1 s_+ u_2$, with u_1, $u_2 \in \Sigma^*$ and s_+ an element of the set of positive generating slopes for Rule 54. They are found in Table 7.1, and we will refer to them as the set

$$S_+ = \{00\oplus, 01 \oplus_2 1, 10 \oplus 1, 11 \oplus_2 00\}. \tag{8.44}$$

The worst possible case for the proof occurs when both u_1 and u_2 are the empty situation. So we may assume without loss of generality that $a_+ \in S_+$. Now, by checking the generating reactions in Table 7.3, we see that as long as $s_+ \neq 01 \oplus_2 1$, they always have the form $\sigma s_+ \to_{\Phi_+} s'_+ v_1$, with $\sigma \in \Sigma$, $v_1 \in \Sigma^*$ and $|v| \geq 1$. The two remaining generating reactions, those with $s_+ = 01 \oplus_2 1$, have the form $\sigma s_+ \to_{\Phi_+} s'_+$, with $s'_+ \neq s_+$. Now we consider the chain of two generating reaction that starts from $\sigma_1\sigma_2 s_+$, with σ_1, $\sigma_2 \in \Sigma$ and v_1, $v_2 \in \Sigma^*$.

$$\sigma_1\sigma_2 s_+ \to_{\Phi_+} \sigma_1 s'_+ v_1 \to_{\Phi_+} s''_+ v_2 v_1\,. \tag{8.45}$$

We have just seen that if $|v_1| = 0$, then $|v_2| = 1$. Therefore $|v_1 v_2| \geq 1$ for all $s_+ \in S_+$.

This means that for $\delta(a_+)_T = 1$ and $\ell = 1$ we always can set $k = 2$. By induction we see then that for $\delta(a_+)_T = 1$ and arbitrary ℓ, a value of $k = 2\ell$ is enough. When we also drop the condition that $\delta(a_+)_T = 1$, a value of $k = 2\ell\delta(a_+)_T$ is enough for the proposition of the lemma. □

The following two reactions are not derived in a completely formal mode; they can instead be derived from the illustrations. Together they will show how an ether fragment causes another ether fragment to occur at a later time.

***Lemma* 8.14** (Destruction of an Ether Fragment)**.** *For every $\sigma \in \Sigma$ there is a reaction of the form*

$$\sigma 1 \varepsilon_+^2 \to_\Phi a_+ 1^3 a'_+ \tag{8.46}$$

with a_+, $a'_+ \in \operatorname{dom}\Phi_+$ and $\delta(a_+)_T \in \{1, 2\}$.

(a) $01\varepsilon_+^2 \to_\Phi 01 \oplus_2 1^3\varepsilon_+$ (b) $11\varepsilon_+^2 \to_\Phi 11 \oplus_2 00 \oplus 1^3$

Figure 8.6: Destruction of ether fragments.

Proof. Depending on the value of σ, one of the following two reactions can be applied to the left side of the reaction (8.46),

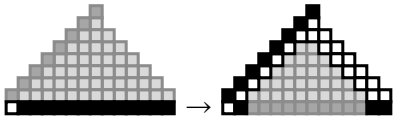
$$01\varepsilon_+^2 \to_\Phi 01 \oplus_2 1^3\varepsilon_+ \qquad \text{or} \qquad 11\varepsilon_+^2 \to_\Phi 11 \oplus_2 00 \oplus 1^3\,. \tag{8.47}$$

An example for these reactions can be seen in Figure 8.6. Now we can set either $a_+ = 01\oplus_2$ and $a'_+ = \varepsilon_+$ or $a'_+ = 11 \oplus_2 00\oplus$ and $a'_+ = [0]$. □

***Lemma* 8.15** (Creation of an Ether Fragment)**.** *For every $j \in \mathbb{N}_0$ there is a reaction of the form*

$$01^{2j+3} \to_\Phi a_+ 1\varepsilon_+^{j+2} a, \tag{8.48}$$

with $a_+ \in \operatorname{dom}\Phi_+$, $a \in \operatorname{dom}\Phi$ and $\delta(a_+)_T = 1$.



$01^{2j+3} \to_\Phi 01 \oplus_2 1\varepsilon_+^{j+2} \ominus_2 10(\ominus 00)^{j+1} \ominus_2 11$

Figure 8.7: Creation of new ether fragments.

Proof. We use the following family of reactions,

$$\{\, 01^{2j+3} \to_\Phi 01 \oplus_2 1\varepsilon_+^{j+2} \ominus_2 10(\ominus 00)^{j+1} \ominus_2 11 \colon j \geq 0 \,\}. \tag{8.49}$$

It can be derived in a way similar to that of the triangle reactions in Table 8.1. An example for these reactions is shown in Figure 8.7. We can then set $b_+ = 01\oplus_2$ and $b = \ominus_2 10(\ominus 00)^{j+1} \ominus_2 11$. □

These two types of reactions then play a role in the following lemma. Its proof uses the fact that the existence of an ether fragment $1\varepsilon_+^2$ at a given time causes the existence of a fragment of another type, which then causes the existence of another ether fragment at still another time. This second type of fragment may be any situation of form 01^{2j+3} with $j \geq 0$. For this kind of fragment we need another set of situations analogous to $\mathscr{E}(t_1, \ldots, t_n)$. We will therefore write

$$\mathscr{E}_1(t_1, \ldots, t_n) = \bigcup_{j\geq 0} \mathscr{F}(01^{2j+3}, t_1, \ldots, t_n). \tag{8.50}$$

for the set of intermediate fragments that occur in the propagation of ether fragments to later times.

***Lemma* 8.16** (Propagation of Ether Fragments)**.** *Let $(U_n)_{n\geq 0}$ be the sequence of random intervals in* (8.31)*. Assume that there is a reaction $u \to_\Phi a$ with $u \in \Sigma^n$ and $a \in \mathscr{E}(t)$. Then for every $\varepsilon > 0$ there is a number $m \geq n$ for which*

$$P(U_m \to_\Phi \mathscr{E}(t+2, t+3) \mid U_n = u) \geq 1 - \varepsilon\,. \tag{8.51}$$

The number m depends only on ε and t, not on a.

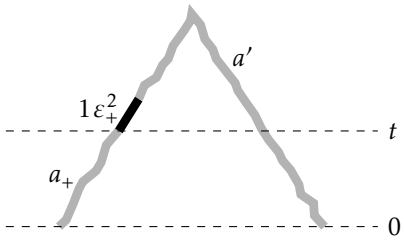


Figure 8.8: Decomposition of a in (8.52).

Proof. Since a is an element of $\mathscr{E}(t)$, it has a representation (Figure 8.8)

$$a = a_+ 1\varepsilon_+^2 a' \tag{8.52}$$

with $a_+ \in \operatorname{dom}\Phi_+$, $a' \in \operatorname{dom}\Phi$ and $\delta(a_+)_T = t$. We now will show that by putting a random interval v at the left of a we will get a situation that reacts to an element of $\mathscr{E}(t+2, t+3)$, with the probability that this happens growing arbitrarily large as the length of v becomes arbitrarily large.

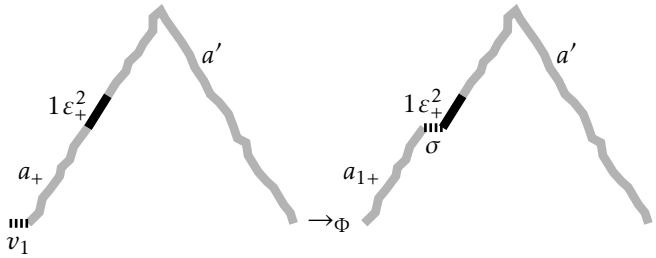


Figure 8.9: Reaction (8.53) puts an event σ to the left of the ether fragment.

We know from Lemma 8.13 that we can find a number m_1 such that for all $v_1 \in \Sigma^{m_1}$ there is a reaction $v_1 a_+ \to_\Phi a_{1+}\sigma$ with $a_{1+} \in \operatorname{dom}\Phi_+$ and $\sigma \in \Sigma$, independent of the value of v_1. The values of a_{1+} and σ depend on the value of u. This also means that for every a_+ there is a reaction (Figure 8.9)

$$v_1 a_+ 1\varepsilon_+^2 a' \to_\Phi a_{1+}\sigma 1\varepsilon_+^2 a'\,. \tag{8.53}$$

This reaction can now be extended to a reaction that modifies the factor $1\varepsilon_+^2$ in the result term. In order to find this reaction we must take both possible values for σ into consideration and need to understand the reactions that start at $01\varepsilon^2$ and $11\varepsilon^2$. From Lemma 8.14 we know that of all $\sigma \in \Sigma$ there is a reaction

$0\sigma\varepsilon_+^2 \to_\Phi x_+ 1^3 a'_{2+}$, with x_+, $a'_{2+} \in \operatorname{dom}\Phi_+$ and $\delta(x_+)_T \in \{1, 2\}$. One of these two reactions can then always be applied to the result of (8.53). The result is a new reaction, $v_1 a_+ 1\varepsilon_+^2 \to_\Phi a_{1+} x_+ 1^3 a'_{2+}$. Now we introduce a new situation, $a_{2+} = a_{1+} x_+$ and have then found that for every $v_1 \in \Sigma^{m_1}$ there is a reaction

$$v_1 a_+ 1\varepsilon_+^2 a' \to_\Phi a_{2+} 1^3 a' \tag{8.54}$$

with $\delta(a_{2+})_T \in \{t+1, t+2\}$.

For the next step of the proof we need to find a reaction that transforms the result of (8.54) into a situation with 10^{2j+3} as factor, where $j \geq 0$. It would be nice if there were a reaction of the form $a_{2+} 1^3 a' \to_\Phi a_{3+} 01^{2j+3} a''$, because then we could apply one the reactions of Lemma 8.15 to the term 01^{2j+3} and get another situation that contains an ether fragment. However we will prove instead a weaker form of this statement: We will show that there is a number $m_2 \geq 0$ such that the probability for a reaction $v_2 a_{2+} 1^3 a' \to_\Phi a_{3+} 01^{2j+3} a''$, with a random interval $v_2 \in \Sigma^{m_2}$, is arbitrarily close to 1. The interval v_2 is then another fragment of the random initial configuration.

We begin with another application of Lemma 8.13. It shows that by choosing an appropriate minimal length for the interval v_2 we can ensure that there is always a reaction $v_2 a_{2+} 1^3 \to_\Phi x_+ w$, with $x_+ \in \operatorname{dom}\Phi_+$ and $w \in \Sigma^*$, in which w is arbitrarily long. We now multiply a' to the right of this reaction and get $v_2 a_{2+} 1^3 a' \to_\Phi x_+ w a'$. The result of this reaction will now written in a different form, depending on the value of w. If w does not consist entirely of events with state 1, we can write it as $w = w' 01^\ell$. Here we must have $\ell \geq 3$, because w is an extension of the interval 1^3. Now we can set $j = \left\lfloor \frac{\ell-3}{2} \right\rfloor$. Then we have either $\ell = 2j+3$ or $\ell = 2j+4$. In the first case we have $w = w' 01^{2j+3}$, and we can set $a_{3+} = x_+ w'$ and $a'' = a'$; in the second case we have $w = w' 01^{2j+3} 1$ and can set $a_{3+} = x_+ w'$ and $a'' = 1a'$. The reaction becomes in both cases $v_2 a_{2+} 1^3 a' \to_\Phi a_{3+} 01^{2j+3} a''$, as required. When we now multiply both sides

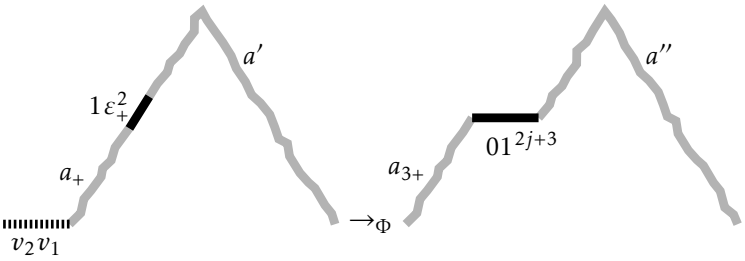


Figure 8.10: Reaction (8.55). The interval 01^{2j+3} occurs later than the ether fragment.

of (8.54) from the right with v_2 and apply the previous reaction to its result, we get (Figure 8.10)

$$v_2 v_1 a = v_2 v_1 a_+ 1\varepsilon_+^2 a' \to_\Phi a_{3+} 01^{2j+3} a''. \tag{8.55}$$

We have in this reaction $\delta(a_{3+})_T \in \{t+1, t+2\}$, because $\delta(a_{3+})_T = \delta(a_{2+})_T$.

The proof for the existence of this reaction is however only valid if w does not consist entirely of ones. Next we must therefore show that the probability

that w contains a zero can be made arbitrarily large by choosing m_2 large enough. For the proof we use the fact that the events of $pr_{x_+}(w)$ in the reaction $v_2 a_{2+} 1^3 \to_\Phi x_+ w$ above must occur either at the time step $t+1$ or $t+2$. We must therefore show that with a high probability there is an event with state 0 at these time steps. For this we use the fact that if the random interval v_2 becomes long enough, the probability that it contains a given interval as its factor becomes arbitrarily close to 1. (This can be proved in a similar way as Theorem 8.12.) In the current proof we use the factor $10^{2t+3}1$ because its closure is a triangle process, and it contains zeros at all time steps from 0 to $t+2$. Therefore, if m_2 is large enough, the probability that a random v_2 contains the interval $10^{2t+3}1$ becomes arbitrarily close to 1, and when v_2 contains such an interval, there is a cell in state 0 in the time steps $t+1$ and $t+2$. Since the interval v_2 is at the left of the situation $a_+ 1^3$, a cell in state 0 must therefore occur in w.

The result of reaction (8.55) is always an element of $\mathscr{E}_1(t+1, t+2)$. We have therefore shown that for every $\varepsilon > 0$ there is a number $m = m_2 + m_1 + n$ such that we have the probability

$$P(U_m \to_\Phi \mathscr{E}_1(t+1, t+2) \mid U_n = u) \geq 1 - \varepsilon\,. \tag{8.56}$$

As a final step we now prove that if there is a reaction $v \to_\Phi b$ with $v \in \Sigma^m$ and $b \in \mathscr{E}_1(t+1, t+2)$, then there is a reaction from b to an element of $\mathscr{E}(t+2, t+3)$. For this we do the same with b as we did before with a and

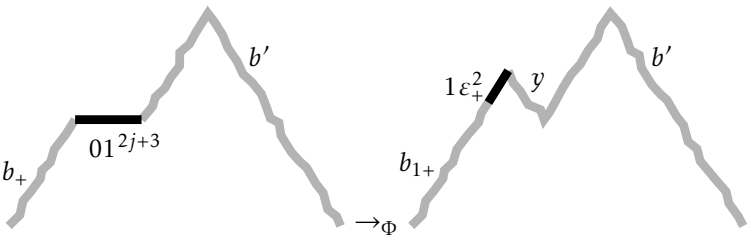


Figure 8.11: Reaction (8.57). A new ether fragment is generated from the interval 01^{2j+3}.

write it as $b = b_+ 01^{2j+3} b'$. We have already seen in Lemma 8.15 that there is a reaction $01^{2j+3} \to_\Phi y_+ 1 \varepsilon_2^{j+2} y$, with $\delta(y_+)_T = 1$. We now set $b_{1+} = b_+ y_+$. Then we have found a reaction (Figure 8.11)

$$b_+ 01^{2j+3} b' \to_\Phi b_{1+} 1 \varepsilon_+^{j+2} y b' \tag{8.57}$$

with $\delta(b_{1+})_T \in \{t+2, t+3\}$, because $\delta(b_{1+})_T = \delta(b_+)_T + 1$.

Therefore the right side of the reaction (8.57) is always an element of $\mathscr{E}(t+2, t+3)$. Putting all these steps together we see therefore that we have found for every $\varepsilon > u$ a probability

$$P(U_m \to_\Phi \mathscr{E}(t+2, t+3) \mid U_n \to_\Phi a) \geq 1 - \varepsilon, \tag{8.58}$$

as stated in the lemma. □

With the reactions in Lemma 8.16 we see that the ether fragments move to the future, and with Lemma 8.13 we see that ether fragments are generated from random initial configurations. Together the lemmas show that ether fragments continue to exist for arbitrarily long times. This is now described in more detail in the following proof.

Proof of Theorem 8.11. The proof is done by induction.

We can see from Lemma 8.12 that the theorem is true for $t = 0$: The lemma shows that there is a $m \geq 0$ such that $P(U_m \to_\Phi \mathscr{E}(0)) \geq 1 - \varepsilon$. Therefore we have $P(U_m \to_\Phi \mathscr{E}(0, 1, 2)) \geq P(U_m \to_\Phi \mathscr{E}(0) \geq 1 - \varepsilon)$.

In the main part of the induction we assume that the theorem is true for a given time $t \geq 0$. With this we mean that for every probability $\varepsilon_1 \geq 0$ there is a number n such that $P(U_n \to_\Phi \mathscr{E}(t, t+1, t+2)) \geq 1 - \varepsilon_1$. We then need to show that for every probability $\varepsilon > 0$ there is a length $m \geq n$ such that

$$P(U_m \to_\Phi \mathscr{E}(t+1, t+2, t+3)) \geq 1 - \varepsilon\,. \tag{8.59}$$

In order to do this we split the reaction $U_m \to_\Phi \mathscr{E}(t+1, t+2, t+3)$ into two subreactions. The first one is derived from the reaction $U_n \to_\Phi \mathscr{E}(t, t+1, t+2)$. This reaction exists by the induction assumption, and we can make its probability arbitrarily high by choosing the right n, but its input interval U_n is by definition shorter than U_m. Nevertheless, a similar reaction starting from U_m does also exist. It will be our first subreaction. In the second subreaction, the result of the first subreaction then reacts to an element of $\mathscr{E}(t+1, t+2, t+3)$, and we will show that we can make its probability as large as we want by making m large enough.

To combine these subreactions we need to split the set of possible input intervals for the first reaction, $U_n \to_\Phi \mathscr{E}(t, t+1, t+2)$, into two disjoint sets. For this we introduce the sets $\mathscr{U}_0$ and $\mathscr{U}_1 \subseteq \Sigma^n$, satisfying the requirement that for every reaction $u \to_\Phi a$ with $u \in \Sigma^n$ and $a \in \mathscr{E}(t, t+1, t+2)$ we have either $u \in \mathscr{U}_0$ and $a \in \mathscr{E}(t)$, or $u \in \mathscr{U}_1$ and $a \in \mathscr{E}(t+1, t+2)$. If u satisfies both conditions, then we may choose arbitrarily $u \in \mathscr{U}_0$ or $u \in \mathscr{U}_1$. The same is true if there is no such reaction $u \to_\Phi a$ for a given u.

Then we can write the probability for the first subreaction as a sum of the probability of two independent stochastic events, namely as

$$P(U_n \to_\Phi \mathscr{E}(t, t+1, t+2)) = P(U_n \in \mathscr{U}_0) + P(U_n \in \mathscr{U}_1)\,. \tag{8.60}$$

The requirements for the reactions from the sets $\mathscr{U}_0$ and $\mathscr{U}_1$ are expressed by the (trivial) probabilities

$$P(U_n \to_\Phi \mathscr{E}(t) \mid U_n \in \mathscr{U}_0) = 1, \tag{8.61a}$$

$$P(U_n \to_\Phi \mathscr{E}(t+1, t+2) \mid U_n \in \mathscr{U}_1) = 1\,. \tag{8.61b}$$

The equations stay true if we replace the leftmost U_n in them with a larger random interval U_m, where $m \geq n$. We will use this now for the second subreaction.

The results (8.61) are the input situations for the second subreaction, so it has two cases as well. We need to show that in both cases there is a reaction into the set $\mathscr{E}(t+1, t+2, t+3)$. If $u \in \mathscr{U}_0$, we use Lemma 8.16. It applies to situation $u \in \Sigma^n$ for which there is a reaction $u \to_\Phi a \in \mathscr{E}(t)$. This property is

of course equivalent to the condition that $u \in \mathscr{U}_0$. The lemma then states that for every $\varepsilon_2 > 0$ there is a number $m \geq n$ such that $P(U_m \to_\Phi \mathscr{E}(t+2, t+3) \mid U_n = u) \geq 1 - \varepsilon_2$. We now collect the probabilities for all $u \in \mathscr{U}_0$ and get the following reaction,

$$P(U_m \to_\Phi \mathscr{E}(t+2, t+3) \mid U_n \in \mathscr{U}_0) \geq 1 - \varepsilon_2 \,. \tag{8.62a}$$

If $u \in \mathscr{U}_1$, then we can use a lengthened form of reaction (8.61b),

$$P(U_m \to_\Phi \mathscr{E}(t+1, t+2) \mid U_n \in \mathscr{U}_1) = 1 \,. \tag{8.62b}$$

The results of both reactions are elements of $\mathscr{E}(t+1, t+2, t+3)$, as was required.

Now we can perform the complete induction step. The following computation begins by splitting the probability for the reaction $U_m \to_\Phi \mathscr{E}(t+1, t+2, t+3)$ into two cases, depending on whether the right end of U_m, i. e. the interval U_n, is an element of $\mathscr{U}_0$ or $\mathscr{U}_1$. Then the equations (8.62) are used to get estimates for these probabilities, and later, at the penultimate step, the two cases are unified again with the help of (8.60).

$$\begin{aligned}
&P(U_m \to_\Phi \mathscr{E}(t+1, t+2, t+3)) \\
&\qquad \geq P(U_m \to_\Phi \mathscr{E}(t+1, t+2, t+3) \mid U_n \in \mathscr{U}_0) P(U_n \in \mathscr{U}_0) \\
&\qquad\quad + P(U_m \to_\Phi \mathscr{E}(t+1, t+2, t+3) \mid U_n \in \mathscr{U}_1) P(U_n \in \mathscr{U}_1) \\
&\qquad \geq P(U_m \to_\Phi \mathscr{E}(t+1, t+2) \mid U_n \in \mathscr{U}_0) P(U_n \in \mathscr{U}_0) \\
&\qquad\quad + P(U_m \to_\Phi \mathscr{E}(t+2, t+3) \mid U_n \in \mathscr{U}_1) P(U_n \in \mathscr{U}_1) \\
&\qquad \geq (1 - \varepsilon_2) P(U_n \in \mathscr{U}_0) + P(U_n \in \mathscr{U}_1) \\
&\qquad \geq (1 - \varepsilon_2)(P(U_n \in \mathscr{U}_0) + P(U_n \in \mathscr{U}_1)) \\
&\qquad = (1 - \varepsilon_2) P(U_n \to_\Phi \mathscr{E}(t, t+1, t+2)) \\
&\qquad \geq (1 - \varepsilon_2)(1 - \varepsilon_1) \,.
\end{aligned} \tag{8.63}$$

Therefore the probability for this reaction can be made greater than $1 - \varepsilon$ by making both ε_1 and ε_2 small enough. □

8.4 Generalisation to Other Rules

While the arguments in this chapter were tailored specifically for the use with Rule 54, it was with the hope that they would lead to ideas that are useful for the understanding of ether formation under other rules, e. g. Rule 110.

To find these new ideas we must generalise the proofs and definitions of this chapter. For some of them it is easy to see how to generalise them, but in other places new ideas are needed. I will now describe the necessary changes in more detail.

First we need a characterisation of the ether in the relevant cellular automaton. The definition of the ether reaction (Definition 8.9) would be generalised to a pair of reactions

$$e_0 v \to_\Phi e_0 e_+ e_- \qquad \text{and} \qquad e_0 e_- e_+ \to_\Phi e_0 e_+ e_-, \tag{8.64}$$

with e_0, $v \in \Sigma^*$, $e_+ \in \operatorname{dom} \Phi_+$ and $e_- \in \operatorname{dom} \Phi_-$. Under Rule 54 we had $e_0 = 1$, $v = 0^3 1$, $e_+ = \varepsilon_+^2$ and $e_- = \varepsilon_-^2$. This can be done in every one-dimensional

cellular automaton with an ether. The situation $e_0 e_+$ then plays the role of the ether fragment.

Then we can construct sets of situations $\mathscr{E}'(t_1, \ldots, t_n) = \mathscr{F}(e_0 e_+, t_1, \ldots, t_n)$, in analogy to the sets $\mathscr{E}(t_1, \ldots, t_n)$ in (8.39). Lemma 8.12, which proves the generation of an ether fragment from a large enough random interval, can be extended to $\mathscr{E}'$: The only requirement for its proof is the existence of a reaction $e_0 v \to \mathscr{E}'(0)$, but this is the first reaction in (8.64). Then, if we have an equivalent to Lemma 8.16, we can prove ether formation in essentially the same way as here, by showing that if there is an ether fragment at time t, there is always a starting time $t' > t$ at which another ether fragment exists with an arbitrarily high probability.

Lemma 8.16, however, is proved with the help of the Lemmas 8.14 and 8.15, which are highly specific to Rule 54. It is not clear whether an equivalent to these lemmas exists for other cellular automata with an ether, and if it exists, how to find it in a systematic way. In Rule 54, they were the result of some experimenting and an already well-developed understanding of the way this rule works. This means that an expert for, say, Rule 110 could find an equivalent to Lemma 8.14 and 8.15 after a similar amount of experimenting, but there is no recipe for a proof of ether formation if the behaviour of a rule is not yet well understood—even if one has already seen that it has an ether.

8.5 Summary

In this chapter we have begun to define concepts with which one can express larger structures in a cellular automaton. There is a small theory of regular structures like triangles. We have seen how one can derive families of reactions.

Then we turned our view to the ether. We first found a way to express the empty ether with situations and reactions. The ether reactions motivated the definition of ether fragments, the situations $1\varepsilon_+^2$. The existence of ether fragments at arbitrary times was defined as a way to express the existence of ether when the initial configuration was chosen at random.

We had to invent an extension of the calculus of Flexible Time that can handle probabilities. This was done in an incomplete way, just enough to get the proofs done. Nevertheless a basic principle did appear: in a random reaction the input had to be a random variable, while the result had to be a set of situations. With these extensions we expressed how ether fragments were generated in the initial configuration and how an ether fragment that was positioned at a given time t caused the existence of an ether fragment at a later time. This then lead to a proof that ether fragments must exist at all time steps when the cellular automaton started from a random initial configuration.

Finally we considered the question whether the proof that we had done for Rule 54 could be generalised to other transition rules. The answer was that it could be only partially, and that filling the gaps would require knowledge about the specific rule involved.

Chapter 9

Conclusions

This work was about finding a way to speak about cellular automata in terms of "traditional mathematics", as I have called it in the introduction. The main results of this thesis are therefore concepts, not theorems.

Reaction Systems. The idea that started my work on cellular automata was that of *situations* and *reaction*, first understood only in an intuitive way and for concrete cellular automata. There was much choice in the way the situations were defined, and it was resolved by trial and error for a specific transition rule. My article [51] was a product of this phase. Already then I tried to find definitions that applied to a large number of cellular automata, even if the work was done for one specific transition rule. This was so because only when a definition applies to a large number of cases, the methods developed for it can also be used to explore the behaviour of unknown cellular automata.

In this work I have therefore searched for principles on which I could base the definition of a reaction system that are valid for a large number of transition rules. It was still possible to exclude certain rules from consideration if their behaviour was difficult to express with situations. One of the first results was the restriction to *interval-preserving* rules. The reason for this decision was that intervals are especially easy to express with situations. We have also seen that the closure of an interval has a simple structure and that the possibilities for their left-to-right arrangement are limited. Only the restriction to interval-preserving rules made the theoretical understanding of cellular processes and their closures possible.

The question which kinds of situations to choose was then resolved by finding the concept of *separating intervals*. A separating interval forms a boundary between the cells left of it and the cells right of it. If two separating intervals form the left and right ends of an interval process π, then the two intervals already determine which points are determined by π, even if the rest of π is unknown. This occurs especially when the events between the separating intervals belong to the closure of a larger process. This then occurs in *achronal situations*: They contain separating intervals at strategic places,

so that it is possible to limit the extension of their closure. Thanks to this construction we have a guarantee that the closure exists at all. The definition of achronal situations and the proof that ordered achronal situations have a closure is another important progress in comparison to [51].

Separating intervals are also interesting in a more general context. They tell about the information transmission in a cellular automaton. The set of separating intervals for a transition rule is a better measure for the speed of information transmission than the radius of the transition rule, another a measure for the speed of information transmission. The radius does however only measure of the maximal possible speed. In contrast to this, the set of separating intervals is an invariant of the cellular automaton that is radius-invariant in the sense of Definition 3.1.

Ether Formation. The sections about ether formation were intended as a reality check for the formalism of Flexible Time. They are a test whether the formalism has been developed far enough to find solutions for a natural-looking question about cellular automata, i. e. a question that did not occur as part of the development of the formalism. The question of ether formation was such a problem.

To become answerable the ether problem had to be reduced to a very simple question. We reduced the question of ether formation from the general case to that of Rule 54, and then characterised the ether by the ether fragments $1\varepsilon_+^2$. The result was a theorem that only showed that such ether fragments exist at arbitrary times when starting from a random initial configuration, not that they become more common with time. The latter fact is clearly visible from computer simulations.

A more positive result is that the proof of the Ether Formation Theorem 8.11 is valid for all probability distributions in which both cells in state 0 and in state 1 can occur in the initial configuration. This is more general than the empirical results, which usually refer to the case that zeros and ones initially occur with equal probability.

Another positive result of the approach with Flexible Time is that it can express the mechanism with which the ether is created and preserved. We have seen that the initial configuration creates with a high probability an ether fragment $1\varepsilon_+^2$, which then causes the creation of an interval 01^{2j+3} at a later time, which in turn reacts to another ether fragment at a still later time: All this was expressed with help of the sets $\mathscr{E}$ and $\mathscr{E}_1$.

The mechanism of Lemma 8.16 then lets us formulate one cause of ether formation under Rule 54. This is the connection of structure formation with loss of information about the initial state. Structure formation in a dynamical system can always be expressed as a case of information loss: Several initial configurations must evolve to the same, more ordered, later configuration. Under Rule 54, information loss occurs during the propagation of ether fragments: In the reactions of Lemma 8.14, the two situations $01\varepsilon_+^2$ and $11\varepsilon_+^2$ react to a situation that contains the term 1^3. This idea has a chance to be also the cause of ether formation in other cellular automata.

The Results in Context. This work has therefore shown that there is an essentially two-dimensional approach to structure formation in cellular auto-

mata and that it can prove nontrivial facts about a cellular automaton. Which place does then the formalism of Flexible Time take among the approaches to understand the behaviour of one-dimensional cellular automata, especially among those described in Section 2.1?

In its two-dimensionality it is similar especially to the work of Ollinger and Richard [48, 53], which was however mainly applied to model a network of colliding particles. While this kind of research is also possible with Flexible Time [52], it has not been done in detail.

In its formalisation of the ether, the approach of this thesis diverges from most other works in that it concentrates directly on fragments of the ether. Most other works, not just Ollinger and Richard, describe in great details the particles that move through the ether. This kind of approach could also become the basis of a proof of ether formation: an analysis how the particles interact and how they gradually destroy each other. (The destruction of gliders does occur under Rule 54 and 110 and was shown experimentally by Boccara *et al.* [4] and by Li and Nordahl [32].) The number of particles and their possible interactions can become however quite large, as it does under Rule 110 [37], and it would be tedious to use them all in a formal proof.

Among the works about the behaviour of random initial configurations, many concentrate on the behaviour of the defects that occur between domains [15, 27]. It has been shown that under many conditions such a defect performs a random walk. In our proof for ether formation we also have a moving "particle", namely the ether fragment, but its position cannot be directly identified from an evolution diagram. Nevertheless it performs a kind of random walk, driven by the states of the cells in the initial configuration. With Flexible Time we have therefore another tool with which one can trace the influence of a random initial configuration to later time step. (The tool however has not yet been developed very far.)

There is also some similarity to the "grouping" approach [13, 14], since Flexible Time also groups several cells to a greater entity. With Flexible Time there is however much more freedom in the choice of the situations, and as a result this approach does not automatically provide tools to put cellular automata into groups according to their behaviour.

Unintentionally, Flexible Time may however lead to new ideas for the classification of cellular automata. We could now classify them by their separating intervals (Table 5.4) or by the pattern of their generating reactions (Table 7.3). What this classification means is not yet clear, but it must have something to do with the information transmission in the cellular automata.

9.1 Ideas for Further Research

Finally I list here a small number of ideas for further research. Their purpose is to extend the system of Flexible Time and also to apply its ideas to other domains.

Separating Intervals. We have defined separating intervals as a purely technical tool. It is not yet clear whether they have an intrinsic meaning, except that they are somewhat related to signal transmission. A possible starting point for further research is therefore the question whether transition rules with the

same set of separating intervals have something in common. Which properties of a transition rule can be derived from knowing its separating intervals?

Another starting point to find out more about separating intervals is the growing body of research about cellular automata with memory [3]. Does the addition of memory change the separating intervals of cellular automata, and if so, in which way?

Explicit Probabilities. Another idea for later work is the search for good explicit probabilities for ether propagation. The reactions only show that the ether fragments survive over time, but they do not give a good estimate about their density. A description of the ether propagation that was more detailed will be needed to get better estimates. With it, there is a chance to find a proof that the density of ether fragments actually grows over time.

Generalising the Way to Find the Ether. Some ideas for this were already outlined in Section 8.4. The current argument for the ether required many *ad hoc* constructions. An example are the ether fragments, which were only found after studying the evolution of configurations under Rule 54 for a long time. There was nothing systematic in their construction. Another example is the structure of the proof for Lemma 8.16. All relied on phenomenology. Nevertheless the current proof may contain the ideas that can be generalised to a more systematic proof. This in turn will of course require a further development of the formalism in order to make it more streamlined and easier to use.

Analysis of Particle Interactions. The ether in Rule 54 is the medium in which particles move. There has been a large amount of research about the particles and their interactions under Rule 54 [4, 34, 36], Rule 110 [6, 37, 38, 39, 40, 41, 48, 53] and other rules [29, 30]. My own earlier paper [52] contains the beginnings of an analysis of the particle interactions under Rule 54. This line of research was left incomplete because the structure of the reaction system for Rule 54 was not yet clear. Now it could be continued, with the hope for a reasonably simple algebra of particle interactions for Rule 54.

A good knowledge of such particle interactions could make another proof for ether formation possible. It has already been noted by Boccara, Nasser and Roger [4] that in the typical evolution of a random initial configuration at a very early time configurations arise that consist of small regions of ether, with particles between them. The particles then interact and slowly destroy each other, such that the ether between them grows. All this has been found in computer simulations, but not proved. A good understanding of particle interactions would therefore allow to understand this process in detail and provide a quantitative estimate for the speed with which the ether grows.

A Generic Case of Self-Organisation. Finally an idea for a larger project. It is inspired by a paper by Boccara and Roger [5], in which the authors describe a whole class of self-organising rules. They are generated from *totalistic* rules, in which the states of the cells are numbers and the next state of a cell only depends on the sum of the states of the cells in the neighbourhood. The authors have found a transformation that transforms an arbitrary totalistic rule into a

rule that shows a certain amount of self-organisation. In it a rule φ of radius r is transformed into a rule φ' of radius nr which works on initial configurations in which every block of n cells have the same state. It is required that if every cell in the initial configuration of φ is expanded to n cells, then the evolution of this configuration under φ' corresponds exactly to the evolution of the original configuration under φ. If both φ and φ' are totalistic, then φ' is uniquely determined by φ. Now if the initial configuration is arbitrary, then in the following time steps in the evolution under φ' the cells begin to organise in blocks of length n, again with defects between the blocks that move randomly and sometimes annihilate.

The understanding of this kind of pattern formation, in the same or a different way as we have done this here for Rule 54, would lead to the understanding of the behaviour of a whole class of cellular automata, not just one. It is therefore very interesting and would also allow the formalism to grow.

Bibliography

[1] S. Abramsky, Dov M. Gabbay, and T. S. E. Maibaum (eds.), *Handbook of logic in computer science, volume 2: Background: Computational structures*, Oxford University Press, 1992. Cited on page 80.

[2] Kathleen T. Alligood, Tim D. Sauer, and James A. Yorke, *Chaos: An introduction into dynamical systems*, Springer Verlag New York, Inc., 1996. Cited on page 20.

[3] Ramon Alonzo-Sanz, *Cellular automata with memory*, Encyclopedia of complexity and systems sciences (R. Meyers, ed.), Springer, 2009, pp. 823–848. Cited on page 120.

[4] N. Boccara, J. Nasser, and M. Roger, *Particlelike structures and their interactions in spatiotemporal patterns generated by one-dimensional deterministic cellular-automaton rules*, Physical Review A **44** (1991), 866–875. Cited on pages 14, 17, 86, 100, 103, 119 and 120.

[5] N. Boccara and M. Roger, *Block transformations of one-dimensional deterministic cellular automaton rules*, Journal of Physics A **24** (1991), no. 8, 1849. Cited on pages 13 and 120.

[6] Matthew Cook, *Universality in elementary cellular automata*, Complex Systems **15** (2004), 1–40. Cited on pages 13, 16, 46 and 120.

[7] James P. Crutchfield, *The calculi of emergence: computation, dynamics and induction*, Physica D **75** (1994), 11–54. Cited on page 14.

[8] James P. Crutchfield and James E. Hanson, *Turbulent pattern bases for cellular automata*, Physica D **69** (1993), no. 3–4, 279 – 301. Cited on page 14.

[9] Karel Culik II and Sheng Yu, *Undecidability of CA classification schemes*, Complex Systems **2** (1988), 177–190. Cited on page 15.

[10] Haskell B. Curry and Robert Feys, *Combinatory logic*, North Holland Publishing Company, Amsterdam, 1958. Cited on page 17.

[11] B. A. Davey and H. A. Priestley, *Introduction to lattices and order*, second ed., Cambridge University Press, 2002. Cited on pages 10 and 41.

[12] Martin Davis, *The universal computer: The road from Leibniz to Turing*, W. W. Norton & Company, Inc., New York, 2000. Cited on page 18.

[13] Marianne Delorme, Jacques Mazoyer, Nicolas Ollinger, and Guillaume Theyssier, *Bulking I: An abstract theory of bulking*, Theoretical Computer Science **412** (2011), no. 30, 3866–3880. Cited on pages 14 and 119.

[14] ———, *Bulking II: Classifications of cellular automata*, Theoretical Computer Science **412** (2011), no. 30, 3881–3905. Cited on pages 14 and 119.

[15] Kari Eloranta, *Random walks in cellular automata*, Nonlinearity **6** (1993), 1025–1036. Cited on pages 14 and 119.

[16] Hans Freudenthal, *Lincos: Design of a language for cosmic intercourse*, vol. 1, North Holland Publishing Company, Amsterdam, 1960. Cited on page 18.

[17] Martin Gardner, *Wheels, life, and other mathematical amusements*, W. H. Freeman and Company, New York and San Francisco, 1983. Cited on page 15.

[18] R. Wm. Gosper, *Exploiting regularities in large cellular spaces*, Physica D **10** (1984), 75–80. Cited on page 16.

[19] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, Addison-Wesley Publishing Company, 1989. Cited on page 48.

[20] Janko Gravner, Genna Gliner, and Mason Pelfrey, *Replication in one-dimensional cellular automata*, Physica D **240** (2011), 1460–1474. Cited on page 13.

[21] Janko Gravner and David Griffeath, *The one-dimensional* Exactly 1 *cellular automaton: Replication, periodicity, and chaos from finite seeds*, Journal of Statistical Physics **142** (2011), 168–200. Cited on page 13.

[22] Geoffrey R. Grimmett and David R. Stirzaker, *Probability and random processes*, Clarendon Press, 1989. Cited on page 105.

[23] James E. Hanson and James P. Crutchfield, *Computational mechanics of cellular automata: An example*, Working Paper 95-10-095, Santa Fe Institute, 1995. Cited on page 14.

[24] J. Roger Hindley and Jonathan P. Seldin, *Introduction to combinators and λ-calculus*, London Mathematical Society Student Texts, 1, Cambridge University Press, Cambridge, Mass., 1986. Cited on page 17.

[25] Lancelot Hogben, *Astraglossa or First steps in celestial syntax*, Journal of the British Interplanetary Society **11** (1952), 258–274. Cited on page 18.

[26] John E. Hopcroft and Jeffrey D. Ullman, *Einführung in die Automatentheorie, formale Sprachen und Komplexitätstheorie*, 3. ed., Addison-Wesley (Deutschland) GmbH, 1993. Cited on page 11.

[27] Erica Jen, *Exact solvability and quasiperiodicity of one-dimensional cellular automata*, Nonlinearity **4** (1990), 251–276. Cited on pages 13 and 119.

[28] Christopher G. Langton, *Studying artificial life with cellular automata*, Physica D **22** (1986), 120–149. Cited on page 34.

[29] Paul-Jean Letourneau, *Particle structure in elementary cellular automaton rule 146*, Complex Systems **19** (2010), 143–156. Cited on pages 13 and 120.

[30] ———, *Persistent structures in elementary cellular automaton rule 146*, Complex Systems **18** (2010), 427–438. Cited on pages 13 and 120.

[31] Steven Levy, *Artificial life: The quest for a new creation*, Penguin Books, 1993. Cited on pages 1 and 2.

[32] Wentian Li and Mats G. Nordahl, *Transient behavior of cellular automata rule 110*, Working Paper 92-03-016, Santa Fe Institute, March 1992. Cited on pages 14 and 119.

[33] Wentian Li and Norman Packard, *The structure of the elementary cellular automata rule space*, Complex Systems **4** (1990), 281–297. Cited on page 3.

[34] Bruno Martin, *A group interpretation of particles generated by one dimensional cellular automaton, 54 Wolfram's rule*, International Journal of Modern Physics C **11** (2000), no. 1, 101–123. Cited on pages 5, 14 and 120.

[35] Genaro J. Martínez, *A note on elementary cellular automata classification*, See `http://adsabs.harvard.edu/abs/2013arXiv1306.5577M`, June 2013. Cited on page 15.

[36] Genaro J. Martínez, Andrew Adamatzky, and Harold V. McIntosh, *Phenomenology of glider collisions in cellular automaton rule 54 and associated logical gates*, Chaos, Fractals and Solitons **28** (2006), 100–111. Cited on pages 8, 86 and 120.

[37] Genaro J. Martínez and Harold V. MacIntosh, *ATLAS: Collisions of gliders like phases of ether in rule 110*, August 2001, `http://uncomp.uwe.ac.uk/genaro/Papers/Papers_on_CA_files/ATLAS/bookcollisions.html`. Cited on pages 13, 119 and 120.

[38] Genaro J. Martínez, Harold V. McIntosh, and Juan Carlos Seck Tuoh Mora, *Production of gliders by collisions in Rule 110*, ECAL (Wolfgang Banzhaf, Thomas Christaller, Peter Dittrich, Jan T. Kim, and Jens Ziegler, eds.), Lecture Notes in Computer Science, vol. 2801, Springer, 2003, pp. 175–182. Cited on page 120.

[39] ———, *Gliders in Rule 110*, International Journal of Unconventional Computing **2** (2006), no. 1, 1–49. Cited on page 120.

[40] Genaro J. Martínez, Harold V. McIntosh, Juan Carlos Seck Tuoh Mora, and Sergio V. Chapa Vergara, *Determining a regular language by glider-based structures called phases f_i_1 in Rule 110*, Journal of Cellular Automata **3** (2008), no. 3, 231–270. Cited on pages 13, 102 and 120.

[41] Genaro J. Martínez, Harold V. McIntosh, Juan Carlos Seck Tuoh Mora, and Sergio V. Chapa Vergara, *Rule 110 objects and other collision-based constructions*, Journal of Cellular Automata **2** (2007), no. 3, 219–242. Cited on page 120.

[42] Jacques Mazoyer and Ivan Rapaport, *Inducing an order on cellular automata by a grouping operation*, Discrete Applied Mathematics **91** (1999), no. 1-3, 177–196. Cited on page 14.

[43] Harold V. McIntosh, *Rule 110 as it relates to the presence of gliders*, May 2001, http://delta.cs.cinvestav.mx/~mcintosh/comun/RULE110W/rule110.pdf. Cited on page 13.

[44] ———, *One-dimensional cellular automata*, Luniver Press, Frome, 2009. Cited on page 49.

[45] Melanie Mitchell, *Cellular automata: A selected review*, Non-Standard Computation (Timo Gramß, Stefan Bornhold, Michael Groß, Melanie Mitchell, and Thomas Pellizari, eds.), Wiley-VCH, 1998, pp. 95–140. Cited on page 2.

[46] Melanie Mitchell, James P. Crutchfield, and Peter T. Hraber, *Evolving cellular automata to perform computations: Mechanisms and impediments*, Physica D **75** (1994), 361–391. Cited on page 2.

[47] Cristopher Moore and Arthur A. Drisko, *Algebraic properties of the block transformation on cellular automata*, Complex Systems **10** (1995), 185–194. Cited on page 14.

[48] Nicolas Ollinger and Gaétan Richard, *Collisions and their catenations: Ultimately periodic tilings of the plane*, IFIP TCS (Giorgio Ausiello, Juhani Karhumäki, Giancarlo Mauri, and C.-H. Luke Ong, eds.), IFIP, vol. 273, Springer, 2008, pp. 229–240. Cited on pages 13, 25, 119 and 120.

[49] Marcus Pivato, *Defect particle kinematics in one-dimensional cellular automata*, Theoretical Computer Science **377** (2007), no. 1-3, 205 – 228. Cited on page 14.

[50] ———, *Spectral domain boundaries in cellular automata*, Fundamenta Informaticae **78** (2007), no. 3, 417–447. Cited on page 14.

[51] Markus Redeker, *Flexible time and the evolution of one-dimensional cellular automata*, Journal of Cellular Automata **5** (2010), no. 4-5, 273–287. Cited on pages 6, 8, 9, 24, 40, 65, 89, 117 and 118.

[52] ———, *Gliders and ether in Rule 54*, Automata 2010: 16th International Workshop on Cellular Automata and Discrete Complex Systems (Nazim Fatès, Jarkko Kari, and Thomas Worsch, eds.), 2010, pp. 299–308. Cited on pages 9, 89, 119 and 120.

[53] Gaétan Richard, *Rule 110: universality and catenations*, JAC 2008, Uzès, France, 2008, http://hal.archives-ouvertes.fr/hal-00273995/en/, pp. 141–180. Cited on pages 13, 119 and 120.

[54] Eric S. Rowland, *Local nested structure in rule 30*, Complex Systems **16** (2006), 239–252. Cited on page 86.

[55] Donald Rutherford, *Philosophy and language in Leibniz*, The Cambridge Companion to Leibniz (Nicholas Jolley, ed.), Cambridge University Press, 2006, pp. 224–269. Cited on pages 17 and 18.

[56] Steven H. Strogatz, *Nonlinear dynamics and chaos*, Perseus Publishing, LLC, Cambridge, Mass., 1994. Cited on page 20.

[57] Alan Turing, *The chemical basis of morphogenenesis*, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **237** (1952), no. 641, 37–72. Cited on page 12.

[58] Robert M. Wald, *General relativity*, The University of Chicago Press, Chicago, 1984. Cited on page 23.

[59] Stephen Wolfram, *Statistical mechanics of cellular automata*, Reviews of Modern Physics **55** (1983), 601–644. Cited on pages 2, 22 and 85.

[60] ———, *Universality and complexity in cellular automata*, Physica D **10** (1984), 1–35. Cited on page 15.

[61] ———, *Twenty problems in the theory of cellular automata*, Physica Scripta **T9** (1985), 170–183. Cited on pages 8 and 15.

[62] Konrad Zuse, *Rechnender Raum*, Elektronische Datenverarbeitung **8** (1967), 336–344. Cited on pages 16 and 17.

Index