TUCS

Charalampos Zinoviadis

Dimension sensitive properties of cellular automata and subshifts of finite type

Turku Centre *for* Computer Science

TUCS Technical Report
No 977, May 2010

Dimension sensitive properties of cellular automata and subshifts of finite type

Charalampos Zinoviadis

University of Turku, Department of Mathematics

This work has been supported by the Academy of Finland grant 131558.

TUCS Technical Report

No 977, May 2010

Abstract

The topics of this thesis are dimension-sensitive properties of cellular automata and subshifts of finite type. Cellular automata are a well known model of parallel computation capturing the notions of locality and homogeneity in space and time, while subshifts of finite type can be used to study the geometric aspects of computation. Both cellular automata and subshifts of finite type are examples of symbolic dynamical systems. One-dimensional cellular automata and subshifts of finite type have been studied extensively and their properties are well understood. However, trying to generalize these properties to higher-dimensional cellular automata or subshifts of finite type is often impossible.

In particular, we are interested in recursion theoretic and dynamical system properties that are dimension sensitive. Approaching the subject through both points of view, we find out that many theorems holding in one-dimension are no longer true in the multidimensional cases, or are only true under much sronger conditions. For example, the decidability status of many problems turns from decidable in the one-dimensional case to undecidable when we go over to the multidimensional cases. In addition, topological entropy cannot give as satisfactory a classification for multidimensional symbolic systems as it gives for one-dimensional systems.

One of the aims of this thesis is to serve as a first reading for everyone that would like to know about this fascinating new area of symbolic dynamics. For this reason, we have included dimension sensitive properties originating from various points of view, and in many cases we have tried to give an intuitive explanation of what causes the difference between the one-dimensional and the mutlidimensional cases.

Keywords: Cellular automata, subshifts of finite type, multidimensional symbolic dynamics, topological entropy, Nivat's conjecture

TUCS Laboratory
Fundim

Contents

1 Introduction

The topic of this master thesis are dimension sensitive properties of cellular automata (CAs) and subshifts of finite type (SFTs) or, in other words, theorems that are valid for in the one-dimensional case but not in two dimensions. Our object is to give a modest presentation of the results on this topic obtained up to now and a smooth introduction for those interested to know more about the motivating problems and the special techniques of this rapid-growing area of symbolic dynamics.

CAs are a model of massive parallel computation. Equivalently, they can be seen as dynamical systems acting on a particular zero-dimensional compact space. CAs were introduced by von Neumann in 1966, see [vNB66]. A simple model of two-dimensional (2-D) SFTs called Wang tiles was introduced by Hao Wang in 1961, see [Wan61]. Wang was interested in the recursion theoretic aspect of this model and did not consider their geometric and dynamical properties. Also, the correspondence between Wang tiles and one-dimensional (1-D) SFTs was not immediately noticed, since the latter originated from a totally different point of view, namely as topological models of Markov chains.

Although von Neumann originally defined 2-D CAs, that is CAs that work on configurations of the plane, for many years only the 1-D case was studied, where configurations are defined over the line. Quite a lot of research was done on the 1-D dimensional case and a lot of theorems regarding their properties were proved, see [Kur03]. The question about whether these results also hold in the 2-D case was either not posed at all or considered trivial. It was not until 1993, see [She93], that Shereshevsky proved that, in contrast to the 1-D case, 2-D CAs cannot be positively expansive, hence establishing that the dynamical properties of 2-D CAs and 1-D CAs can be strikingly different. Then, in 1994, Kari proved that it is undecidable whether a 2-D CA is injective, see [Kar94], so that the computational properties of 2-D and 1-D CAs can also differ a lot. After that, many people have tried to explore what is the exact nature of these differences and at present our knowledge about dimension sensitive properties of CAs is much broader.

The history of dimension sensitive properties of SFTs has been slightly different. Namely, as soon as it was realized that Wang tiles can be seen as 2-D SFTs, we already had two very important dimension sensitive properties at hand: the existence of an aperiodic 2-D SFT and the undecidability of the non-emptiness problem for 2-D SFTs, see [Rob71, Ber72]. In 1994, Burton and Steif proved something equally surprising: there exist irreducible 2-D SFTs with multiple measures of maximal entropy, see [BS94]. The fact that an irreducible 1-D SFT has a unique measure of maximal entropy is known since 1964, see [Par64]. Following, that there has been a constant flow of results regarding the topological entropy of 2-D SFTs and its connections with the factoring relation between 2-D SFTs. We will state and prove some of these in this thesis. Finally, in 2009, Hochman presented an impressive construction saying that 2-D SFTs can be "almost everything" which changed totally the kind of questions we would like to pose for multi-dimensional SFTs, see [Hoc09].

The layout of the thesis is the following:

In Chapter 1, we define Wang tiles in the classical sense and give Robinson's construction of an aperiodic tile set and sketch the proof or the undecidability of the Domino problem Our exposition essentially follows the seminal paper of Robinson [Rob71].

In Chapter 2, we introduce CAs in the usual way and state that they can be seen as the shift-invariant transformations of the configuration space. Then, we prove the undecidability of the following problems for 2-D CAs: reversibility, surjectivity, number-decreasingness and openness. The proofs of the first three are based on the papers [Kar94, Dur93, BDFK05], while the undecidability of openness is a new result. Then, we turn our attention to topological entropy and give a result of Lakshtanov and Langvagen [LL04] stating that every 2-D CA that has a spaceship has infinite topological entropy. Then, we describe briefly Meyerovitch's construction from [Mey08] of a 2-D Ca with non-zero, finite topological entropy. Several other properties of the topological entropy of 2-D CA are mentioned without a proof. In regards to dimension sensitive dynamical properties, we prove Shereshevsky's classical result of the non-existence of positively expansive CAs found in [She93], and also give an example due to Kari of a 2-D non-sensitive CA without any points of equicontinuity. A similar example was independently given By Sablik and Theyssier in [ST08].

In Chapter 3, we define SFTs in the mathematically rigorous way, as subsystems of the full shift dynamical system, and prove a few results about 2-D SFTs and the factors of 2-D SFTs, called 2-D sofic shifts. These results, due to Desai [Des05, Des09] state that every 2-D SFT of positive entropy has a family of SFT subsystems with entropy dense in $[0, h(X)]$, and that the same is true if we substitute every occurrence of the term ”SFT” with ”sofic shift”. Then, we define a strong mixing condition for 2-D SFTs and prove that 2-D SFTs satisfying this mixing condition factor onto lower entropy full shifts and that their automorphism group is contains a copy of every finite group. This last result was first proved by Ward in [War91], although our proof employs techniques developed by Desai.

Finally, in Chapter 4 we give a short exposition of open problems in the topic of multidimensional CAs and SFTs. We also give part of the proof of the best result we currently have for Nivat's conjecture. Namely, Quas and Zamboni in [QZ04] have proved that if the complexity function of a 2-D configuration ξ satisfies $p_\xi(n_1, n_2) \leq \frac{1}{16}n_1n_2$, for some $n_1, n_2 \geq 2$, then ξ is periodic.

2 Wang tiles and the Tiling problem

Wang tiles are unit squares with colored edges. We will usually represent colours as letters. A finite set S of tiles is called a *tile set*. A *configuration* with tiles from the tile set S is a function $f\colon \mathbb{Z}^2 \to S$. Intuitively, a configuration is a way to fill the plane with unit squares from S, where abutting squares are put side-to-side. Notice that we are not allowed to rotate the tiles. A configuration f is *valid at point* $(x, y) \in \mathbb{Z}^2$ if the edges of the tile $f(x, y)$ have the same color as the abutting edges of its neighboring tiles, i.e. if the upper edge of $f(x, y)$ has the same color as the lower edge of $f(x, y+1)$, the right edge of $f(x, y)$ has the same color as the left edge of $f(x + 1, y)$ etc.

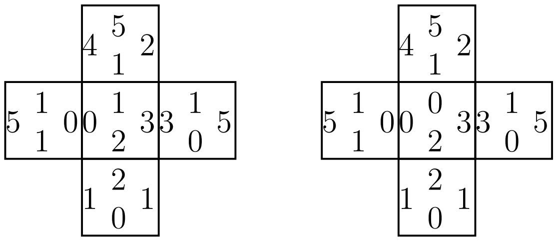


Figure 1: A point where the tiling conditions are satified and a point where they are not.

A configuration f is called a *(valid) tiling* if it is valid at all points $(x, y) \in \mathbb{Z}^2$. We also use the expression that S *admits* the tiling f. A configuration f is *(one-way) periodic* with *period* $(a, b) \in \mathbb{Z}^2$ if $f(x, y) = f(x + a, y + b)$, for every $(x, y) \in \mathbb{Z}^2$. A configuration is called *two-way periodic* if it has two linearly independent periods. If a valid tiling has no periods, it is called a *non-periodic* tiling. A tile set is called *aperiodic* if it admits a tiling, but all the tilings that it admits are non-periodic.

Wang tiles were introduced by logician Wang in 1961 [Wan61]. He discovered them while investigating the decidability of the satisfiability problem of a certain class of first-order formulas. His investigations led him to the following decision problem, which is known as the *Tiling problem*:

Does a given tile set S admit a valid tiling?

He proved that this problem would be decidable if there existed no aperiodic tile sets, and went on to conjecture that this is indeed the case. Berger showed that this is not true: he explicitly constructed an aperiodic tile set and then used it to prove the undecidability of the Tiling problem.

Most of the material of this chapter is based on Jarkko Kari's lecture notes for the class ”Tilings and patterns” taught during the winter semester of 2008 in the University of Turku.

2.1 Compactness principle

In this section, we will give a fundamental result about tile sets. In the form that will be of most use for us in this chapter, it states that if a tile set can tile validly arbitrarily large squares, then it can also tile validly the whole plane. This is extremely useful, since we can apply it to show that a tile set admits a valid tiling, without having to describe explicitly how this tiling actually looks like. We first prove a more general result which states that a certain topological space that will be introduced in the next chapter is compact.

Let S be a tile set and consider an infinite sequence $c_1, c_2, \ldots$ of configurations. We say that the sequence *converges* to $c \in S^{\mathbb{Z}^2}$ and that c is its *limit*, if for every $(x, y) \in \mathbb{Z}^2$ there exists some $k \geq 0$ such that $c_i(x, y) = c(x, y)$ for every $i \geq k$. In other words, if we choose an arbitrary position of the plane and check the sequence of the tiles appearing in this position, then from some moment on we will always see the same tile. It is obvious that if a sequnce has a limit, then this limit is unique. We will denote it by $\underset{i\to\infty}{lim} c_i$.

A *subsequence* of a sequence $c_1, c_2, \ldots$ is another sequence $c_{i_1}, c_{i_2}, \ldots$ where $i_1 < i_2 < \ldots$ is an increasing sequence of numbers. Clearly, if a sequence of configurations converges to $c \in S^{\mathbb{Z}^2}$, then all of its subsequences also converge to the same configuration.

We are now ready to state and prove the **compactness principle**:

Proposition 1. *Every sequence of configurations has a converging subsequence.*

Proof. Let $c_1, c_2, \ldots$ be an arbitrary sequence of configurations and consider an enumeration $\vec{r_1}, \vec{r_2}, \ldots$ of $\mathbb{Z}^2$. We will show that there exists a subsequence $c_{i_1}, c_{i_2}, \ldots$ such that for every $n \geq 1$, if $j \geq n$ then $c_{i_j}(\vec{r_n}) = c_{i_n}(\vec{r_n})$, i.e. the tiles in the n'th position of the plane are fixed from the n'th moment on. Let us define the indices $i_1, i_2, \ldots$ inductively as follows:

- i_1 is the smallest index such that infinitely many elements of the sequence $c_1, c_2, \ldots$ agree with c_{i_1} in position $\vec{r_1}$. Since there exist only a finite number of tiles which can appear in position $\vec{r_1}$, at least one of them is repeated an infinite number of times, hence i_1 is well-defined. Let $I_1 = \{j : j \geq 1 \text{ and } c_j(\vec{r_1}) = c_{i_1}(\vec{r_1})\}$ be the set of all indices such that the corresponding element of the sequence agrees with c_{i_1} in position $\vec{r_1}$. It is clear by the definition of c_{i_1} that I_1 is an infinite set.

- Let us suppose that $i_1, i_2, \ldots, i_{k-1}$ have already been defined and that the corresponding sets $I_1, I_2, \ldots, I_{k-1}$ are all infinite and such that $I_1 \supseteq I_2 \supseteq \ldots \supseteq I_{k-1}$. We choose i_k to be the smallest integer in I_{k-1} that satisfies the following conditions:

  1. $i_k > i_{k-1}$
  2. There exist infinitely many indices $j \in I_{k-1}$ which agree with c_{i_k} in position $\vec{r_k}$.

Since the set I_{k-1} is infinite, there exists some number which satisfies both conditions. Also, we define I_k to be the set $I_k = \{j \in I_{k-1} \colon c_j(\vec{r_k}) = c_{i_k}(\vec{r_k})\}$. According to the definiton of i_k, I_k is an infinite set and obviously $I_k \subseteq I_{k-1}$. So, the induction can go on.

Now, it is easy to prove that the sequence $c_{i_1}, c_{i_2}, \ldots$ is converging: Let $\vec{r_n}$ be any position in the plane and let $j \geq n$. Then, $i_j \in I_j \subseteq I_n$, which means that $c_{i_n}(\vec{r_n}) = c_{i_j}(\vec{r_n})$, which is exactly what we wanted to prove. □

Actually, the argument used in the proof of Proposition 1 is nothing but a application of Knig's principle.

Corollary 1. *Let S be a tile set. If for every finite $F \subset \mathbb{Z}^2$ there exists a configuration that is valid at every $(x, y) \in F$, then S admits a valid tiling.*

Proof. Let $\vec{r_1}, \vec{r_2}, \ldots$ be an enumeration of $\mathbb{Z}^2$ and let $F_n = \{\vec{r_1}, \vec{r_2}, \ldots, \vec{r_n}\}$, for each $n \geq 1$. According to the hypothesis, for each $n \geq 1$ there exists a configuration c_n that is valid at all positions of F_n. By the compactness principle, we know that the sequence $c_1, c_2, \ldots$ has a converging subsequence $c_{i_1}, c_{i_2}, \ldots$. Let $c \in S^{\mathbb{Z}^2}$ be its limit. Then, c is a valid tiling.

Indeed, let $\vec{r_k}$ be any position of the plane. Since the sequence $F_1 \subseteq F_2 \subseteq \ldots$ is increasing and every element of $\mathbb{Z}^2$ belongs to some set of the sequence, there exists $n_1 \geq 1$ such that for every $n \geq n_1$, F_n contains $\vec{r_k}$ and all of its neighbors. By convergence of the sequence $c_{i_1}, c_{i_2}, \ldots$, there exists some $n_2 \geq 1$ such that for every $n \geq n_2$, c and c_{i_n} agree on $\vec{r_k}$ and all of its neighbors. If we consider some $n \geq \max\{n_1, n_2\}$, we can see that c agrees with c_{i_n} on $\vec{r_k}$ and all of its neighbors, and also that c_{i_n} tiles correctly at $\vec{r_k}$. Therefore, c tiles correctly at $\vec{r_k}$, and since $\vec{r_k}$ was an arbitrary position of the plane, c is a valid tiling. $\square$

Remark 1. *From the above corollary, we can conclude that if a tile set S can tile validly arbitrarily large squares, then it can tile validly the whole plane.*

For more properties of tilings, see [BDJ08, CD04, Dur99].

2.2 An aperiodic tile set

In this section, we are going to describe an aperiodic tile set that we use later to prove that the Tiling problem is undecidable. Before doing that, we need one more fundamental result about Wang tile sets.

Proposition 2. *If a tile set S admits a one-way periodic tiling, then it also admits a two-way periodic tiling.*

Proof. Let $f \in S^{\mathbb{Z}^2}$ be a one-way periodic tiling and $(a, b) \neq (0, 0)$ be a period of f. Without loss of generality, we can suppose that $b \geq 0$. Consider a horizontal stripe of height b extracted from f, i.e. the tiles $f(x, y)$ where $x \in \mathbb{Z}$ and $1 \leq y \leq b$. The sequence of colors on the bottom edge of this stripe is the same as the sequence of colors on its top edge with a horizontal offset a.

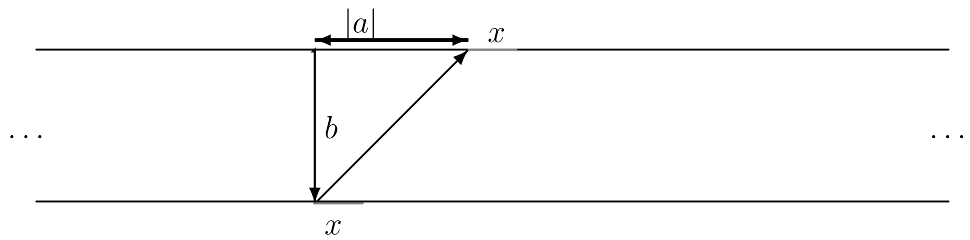


Figure 2: A horizontal stripe of height b extracted from f.

Let us define for every $j \in \mathbb{Z}$ the rectangle of width $|a|$ extracted from this stripe at position $(j, 1)$ as $R_j = \{f(x, y) : j \leq x \leq j + |a|; 1 \leq y \leq b\}$. Since the tile set S is finite, there is only a finite number of such rectangles. Therefore, some rectangle R is repeated infinitely many times, i.e. $R_{j_1} = R_{j_2} = R$, for some $j_1, j_2 \in \mathbb{Z}$. We can also assume that $R_{j_1} \cap R_{j_2} = \emptyset$. A periodic stripe of height b can now be constructed by repeating periodically the pattern between positions j_1 and j_2. We know that the sequence of colors on the bottom edge of this stripe is again identical on its top edge with a horizontal offset a.

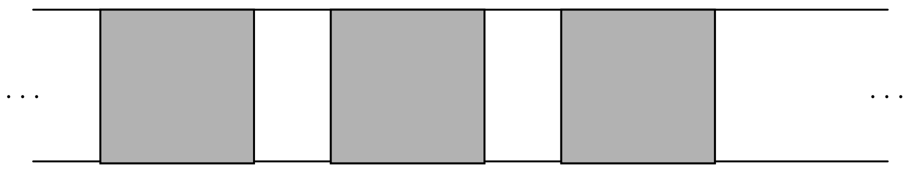


Figure 3: A periodic stripe of height b.

A two-way periodic tiling can now be obtained by stacking copies of this stripe on top of each other with the horizontal offset a. This tiling has periods (a, b) and $(j_2 - j_1, 0)$, which are obviously linearly independent.

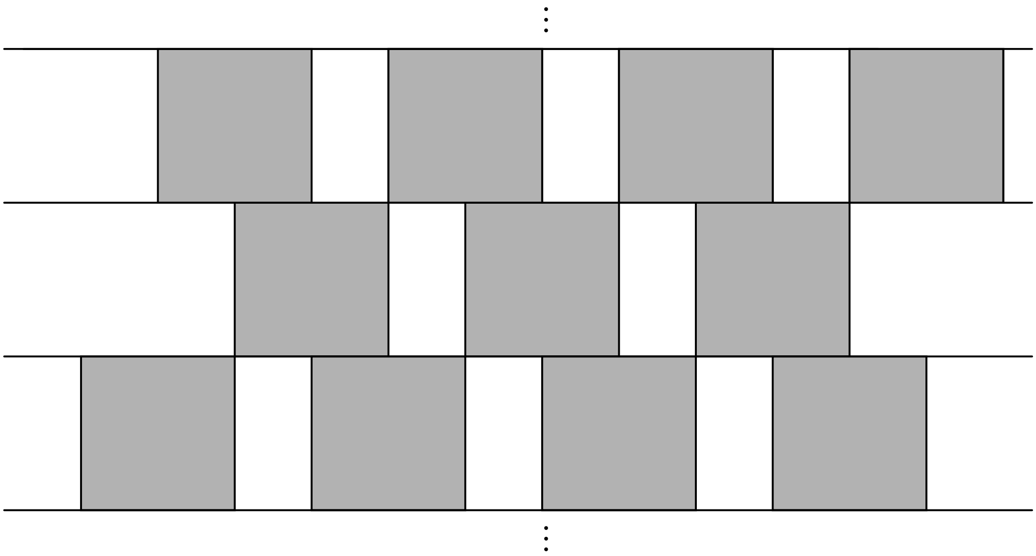


Figure 4: A two-way periodic tiling.

$\square$

Remember that when we defined aperiodic tile sets, we did not specify whether they should accept no one-way or two-way periodic tiling. The next corollary shows that this makes no real difference.

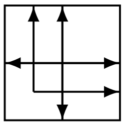
Corollary 2. *A tiling set admits a two-way periodic tiling if and only if it admits a tiling with a horizontal period.*

Proof. Obviously, if a tile set S admits a two-way periodic tiling f, then this same tiling also has a horizontal period. Indeed, let (a, b) and (c, d) be two linearly independent periods of f. Then $d(a, b) - b(c, d) = (ad - bc, 0)$ is a horizontal period for f.

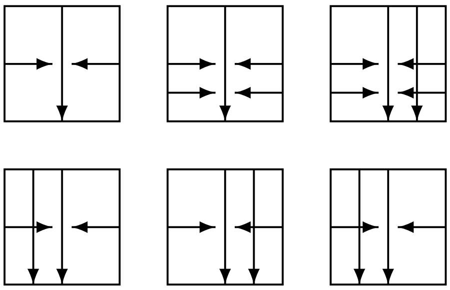
On the other hand, suppose that S admits the one-way periodic tiling f. According to Proposition 2, S also admits a two-way periodic tiling g. Now, this g has a horizontal period. $\square$

We are now ready to give the fundamental construction of **Robinson's aperiodic tile set**[Rob71]. In doing so, we will use arrows to describe the tiling conditions instead of colors. The rule is that arrow heads must meet arrow tails and every arrow tail must be met by some arrow head.

Robinson's tile set consists of tiles



which are called crosses, and tiles



called arms. All tiles can be rotated, so every tile comes in four different orientations. Hence, the total number of such tiles is 28. The following terminology will be used extensively:

- Every tile has *central arrows* at the centers of all four sides, and possibly some *side arrows*.

- A cross is said to *face* the direction of its side arrows. For example, the cross that is drawn in the figure above is facing north-east, or, in other words, it is a NE-cross.

- The central arrow that runs through an arm is called the *principal arrow* of the arm, and the direction of the principal arrow is called the *orientation* of the arm. For example, all arrows in the figure above are drawn in the north-to-south orientation.

An important thing to notice about arms is that if there are side arrows perpendicular to the principal arrow, then they are on the side of the head of the arrow.

We would like to enforce a cross in the intersection of every other row and column. In order to achieve that, we use a construction known as superposition or forming ”sandwich tiles”, which consists in taking the cartesian product of the tiles we have defined up to now with some other tile set. Intuitively, this means that we are superimposing a new layer over every configuration, and the tiles in the second layer come from the new tile set. Then, in a valid tiling with the extended tile set, we demand that in both layers the tiling be valid. More specifically, we take the cartesian product with the *parity tiles*

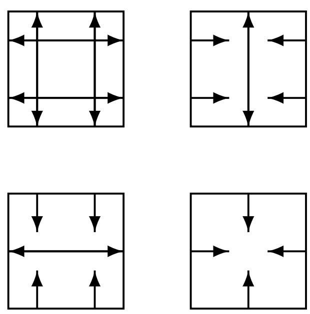


Figure 5: The parity tiles.

and allow the first parity tile to be ”sandwiched” only with crosses. Since the only way to tile the plane with the parity tiles is by alternating the tiles on even and odd rows and columns, the first parity tile will appear in every other row and column. Therefore, we have guaranteed the existence of a cross in every other row and column. By convention, we can assume that every odd-odd position of the plane contains a cross.

Note that since a cross doesn't have any incoming arrows, between two crosses there can only be an arm and there are two possible choices for the orientation of this arm, as it cannot point towards any of the crosses. Therefore, in valid tilings, the second parity tile is only paired with north-to-south or south-to-north arms and the third parity tile can only be paired with east-to-west or west-to-east arms. The fourth parity tile can be paired both with crosses and every type of arms. Hence, in our final tile set, we include only these tiles for a total of 4+12+12+28=56 different tiles. This is Robinson's tile set.

Before proving that it is aperiodic, we need a specific kind of pattern called a $(2^n - 1)$-square which plays a special role in the study of Robinson's and Robinson's-like tile sets.

$(2^n - 1)$-squares are defined inductively as follows:

- A 1-square is a cross in an odd-odd position.
- A $(2^{n+1} - 1)$-square consists of a cross in an even-even position, sequences of arms radiating out of the center, and four $(2^n - 1)$-squares facing each other in the four quadrants. A $(2^n - 1)$-square is said to have the orientation of its cross.

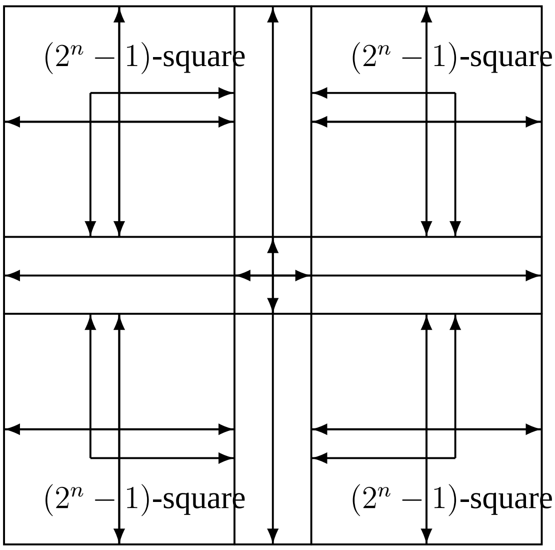


Figure 6: Inductive definition of a $(2^{n+1} - 1)$-square.

Note that for every $n \geq 1$ there exists four different $(2^n - 1)$-squares, as the orientation of the central square uniquely determines the rest of the tiles.

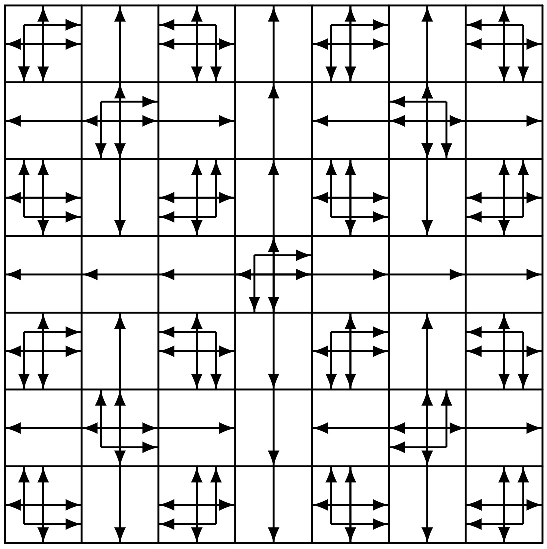


Figure 7: The 7-square facing south-east where side arrows are not drawn.

Inductively, one can prove the following facts about $(2^n - 1)$-squares:

1. The tiling is valid inside the square.
2. All edges on the border of the square have arrow heads pointing out, so all neighbors of $(2^n - 1)$-squares are forced to be arms.
3. The only side arrows on the edges of the border are in the middle of the borders in the directions where the center cross of the square is facing.
4. The tiles at the four corners of the square are crosses in odd-odd positions.

Since Robinson's tile set can tile validly arbitrarily large squares, it can also tile the whole plane. Let us consider an arbitrary valid tiling of the plane by Robinson's tiles. Using induction on n we will prove that every cross in an odd-odd position belongs to a unique $(2^n - 1)$-square, for every $n \geq 1$.

- The case $n = 1$ is trivial, since by definition 1-squares are exactly the crosses in odd-odd positions.
- Suppose that the claim is true for $n - 1$, and let us consider an arbitrary cross C in an odd-odd position. By the inductive hypothesis, it belongs to a unique $(2^{n-1} - 1)$-square. There are four possibilities for the orientation of

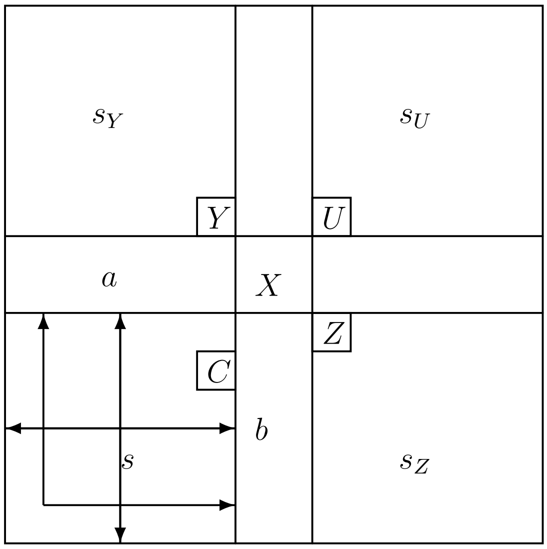


Figure 8: The inductive step of our argument.

this square, but all the cases are symmetric. Let us assume that the square is facing north and east.

We claim that the tile X in the Figure 1.7 must be a cross. Indeed, suppose on the contrary that it is an arm. Since arms have incoming arrows in all but one of their sides, one of the neighbors of X in region a or b is an arm directed towards X. Without loss of generality, we can suppose that this neighbor is in the a region. Repeating this argument, one concludes that all tiles in region a are arms. Let us now examine what happens in the center of the region a. The tile in this position must be an arm with its side arrows on the tail's part. As we have previously noted, this is impossible since arrows perpendicular to the principal arrow can only appear towards the head of the principal arrow. Hence, our assumption that there is an arm in X is wrong, and X must be a cross.

Now, let us consider the tile Y. It is in an odd-odd position, so it must be a cross. According to the induction hypothesis, it belongs to a unique $(2^n - 1)$-square s_Y. This square cannot overlap with the square s, since then the tiles in the overlap region would belong to two different $(2^n - 1)$-squares, which contradicts the induction hypothesis. Also, the tiles in the region a cannot belong in s_Y, since they would be part of the bottom line of s_Y. However, the tiles in the bottom line of every $(2^n - 1)$-square belong to odd lines, while region a is part of an even line. Finally the tile north of X cannot be in s_Y, since X is a cross, and we know that all edge neighbors of $(2^n - 1)$-squares are arrows. Therefore, Y has to be in the south-east corner of s_Y, as in the picture above. Similarly, U and Z are forced to be in the south-west and north-west corner of the squares s_U and s_Z, respectively. The tiles between these four squares are forced to be sequences of arms radiating from X, and the side arrows in the middle of regions a and b force the center crosses of s_Y and s_Z to face the squares s and s_U. Hence, the squares s, s_Y, s_U, s_Z and the tiles between them form a $(2^n - 1)$-square that contains tile C.

The uniquness of this $(2^n - 1)$-square is obvious, since in the reasoning above the orientation of the $(2^{n-1} - 1)$-square containg C uniquely determines the location and the orientation of the $(2^n - 1)$-square.

We are now in position to prove that:

Theorem 1. *Robinson's tile set is aperiodic.*

Proof. We have already noted that Robinson's tile set admits a valid tiling of the whole plane. Let us assume that it admits a periodic tiling. According to Corollary 2, it also admits some tiling f with a horizontal period. As every tiling contains $(2^n - 1)$-squares for every $n \geq 1$, there exist crosses in even-even positions, namely the centers of the $(2^n - 1)$-squares, followed horizontally by arbitrarily large sequences of arms. Hence, f cannot have a horizontal period, which is a contradiction. $\square$

2.3 Undecidability of the Tiling problem

In this section, we are going to prove that the Tiling problem is undecidable. In order to do this we will show how an arbitrary Turing machine can be simulated by tiles, thus reducing the Tiling problem to the Halting problem of Turing machines. However, this construction, introduced by Wang in [Wan61], faces some fundamental problems. In order to fix them we will use Robinson's tile set. Let us first agree on some terminology about Turing machines.

It has long been a mathematical folklore, explicitly expressed in the Church-Turing thesis, that Turing machines are a formal definition for the intuitive idea of an algorithm and that they can perform every operation that modern, super-powerful computers can. The Turing machines that we will use here work on a *two-way infinite tape* and the *tape alphabet* has a distinguished *blank symbol* b. The fundamental undecidable problem we are going to consider is whether a

Turing machine halts when it is started on the all-blank tape. A good introduction to Turing machines and undecidability results is contained in [HMU06]

Formally, we define a *Turing machine* (TM) as a sixtuple

$$M = (S, \Gamma, s_o, s_h, \delta, b)$$

where S and Γ are finite sets called the *state set* and the *tape alphabet* respectively, s_0 and s_h are distinguished states called the initial and the halting state, respectively, $b \in \Gamma$ is the blank symbol and $\delta\colon S \times \Gamma \to S \times \Gamma \times \{L, R\}$ is the *transition function*

At every time step, the processor of the TM is in a specific state and position of the input tape, reading the letter in that position. Depending on its state and the letter currently scanned, it takes the following actions; it changes its state, replaces the letter in the current position of the input tape by a new one, and, finally, moves to the left or right.

Formally, we define a *configuration* of the TM as an element of $S \times \Gamma^{\mathbb{Z}} \times \mathbb{Z}$. Configuration (q, f, i) means that the TM is in state q, the content of the tape is the bi-infinite sequence $f\colon \mathbb{Z} \to \Gamma$ and the processor is in position i. We say that the configuration (q, f, i) is *transformed in one step* to configuration (p, g, j), where

$$\delta(q, f(i)) = (p, y, d)$$
$$g(k) = \begin{cases} y & \text{if } k = i, \\ f(k) & \text{otherwise.} \end{cases}$$
$$\begin{cases} j = i + 1 & \text{if } d = R \\ j = i - 1 & \text{if } d = L \end{cases}$$

We denote this move by $(q, f, i) \vdash (p, g, j)$. Let $\overset{*}{\vdash}$ be the reflexive and transitive closure of $\vdash$. We say that the TM *halts* if $(s_0, \mathbf{b}, 0) \overset{*}{\vdash} (s_h, f, i)$, where $\mathbf{b}$ has the blank b in every position of the input tape, and f and i are arbitrary. The **Halting problem** consists in deciding if an arbitrary TM halts. By Turing's classical result, we know that:

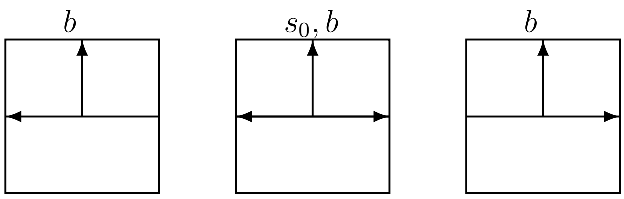
Proposition 3. *The halting problem is undecidable, i.e. there is no algorithm to solve it.*

Remark 2. *The careful reader might have noticed that we have not defined what an algorithm is, so Proposition 3 is not well-defined mathematically. However, one can think of algorithms as programs written in any strong enough propgramming language, for example PYTHON. Then, Proposition 3 states that there is no program written in PYTHON which takes as an input the description of an arbitrary TM and returns "yes" if the TM halts and "no" if it doesn't.*

Next, we are going to describe how Wang tiles can simulate TMs. Although this might seem a little absurd, since TMs are dynamic objects and Wang tiles are static ones, we will see that it is possible by using the vertical direction to represent time. The colors of the tiles we are going to use are arrows, state symbols of the TM, tape symbols of the TM and combinations of these. In neighboring tiles, arrow heads must meet arrow tails and labels must be the same.

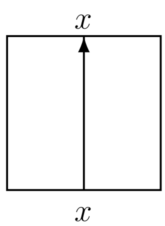
Let $M = (S, \Gamma, s_0, s_h, \delta, b)$ be a TM. The corresponding Wang tile set that simulates M contains:

1. Three *starting tiles* to represent the blank tape.

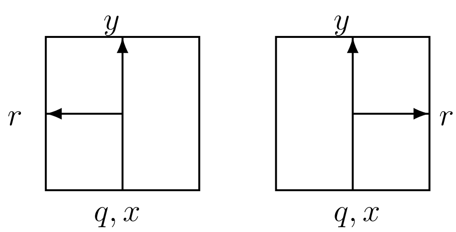


   The tile in the middle is called the *seed tile*.

2. For every tape letter $x \in \Gamma$ the *alphabet tile*

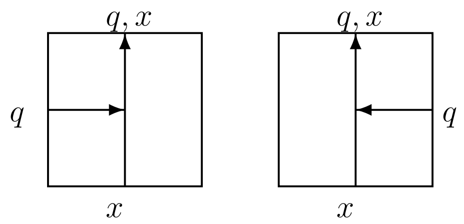


3. For every non-halting state $q \in S \backslash h$ and tape symbol $x \in \Gamma$ one *action tile*

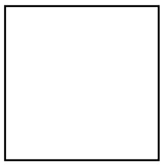


where the tile in the left is included in the tile set if $\delta(q, x) = (r, y, L)$ and the tile in the right is included if $\delta(q, x) = (r, y, R)$.

4. For every non-halting state $q \in S \backslash h$ and tape symbol $x \in \Gamma$ the two *merging tiles*



5. The *blank tile*



Lemma 1. *Let $M = (S, \Gamma, s_0, s_h, \delta, b)$ be a TM and let P_M be the above constructed tile set. Then, M does not halt if and only if P_M admits a valid tiling of the plane that contains at least one occurrence of the seed tile.*

Proof. Suppose that M does not halt. Then, it is possible to make a valid tiling that contains exactly one occurrence of the seed tile. Indeed, let us place the seed tile at position $(0, 0)$. Those tiles that lie on the same horizontal line as the seed tile are forced to be starting tiles. The labels of the top edges of the tiles in this line can be viewed as a representation of the initial configuration of M. The tile above the seed tile is forced to be an action tile. Also, a merging tile is also placed in the right place so that the labels of the top edges of the tiles in the second line represent the configuration of M after one step, as in figure 1.9.

Inductively, the labels of the top edges of the tiles on the $(n+1)$'th line represent the configuration of M after n steps. Since the TM does not halt, this procedure can be continued indefinitely to fill up the upper half of the plane. In the lower half, we can just place the blank tile. This is a valid tiling that contains the seed tile.

For the other direction, assume that M halts. Since there does not exist a merging tile for the halting state, the above procedure will be impossible to continue when the halting state is reached, so there does not exist a valid tiling which contains an occurrence of the seed tile. □

Therefore, given an arbitrary TM M we can algorithmically construct a tile set P_M such that M does not halt if and only if P_M admits a tiling satisfying a certain innocent-looking condition (at least one appearance of the seed tile). This, however, does not imply the undecidability of the Tiling problem, since P_M always admits some tilings, for example the all-blank tiling, or a tiling where in every column there is a fixed alphabet tile. One might think (or hope) that if we do some minor adjustments to the tile set, we can guarantee the existence of the seed tile in every valid tiling of P_M. Unfortunately, by a simple application of the compactness principle, this is not possible if P_M can tile arbitrarily large squares without using the seed tile and, therefore, the seed tile has to be enforced in every $n \times n$-square, for some $n \geq 1$. This seems to contradict the fact that we want to represent arbitrarily large computations of M.

Berger's classical construction gives a way to bypass this problem. The idea is to partition the plane into disjoint rectangles with unbounded size. In every such rectangle we can represent a finite but arbitrarily large part of the computation of M. Then, if M halts, some big enough rectangle will contain the whole computation of M, hence the tiling will not be able to be extended in this rectangle. If, on the other hand, M does not halt, we can fill arbitrarily large squares with initial parts of the computation of M, hence there will be a valid tiling.

The real challenge is to implement this partition of the plane into disjoint and unboundedly large squares with tiles, hence finite conditions. Berger did this with

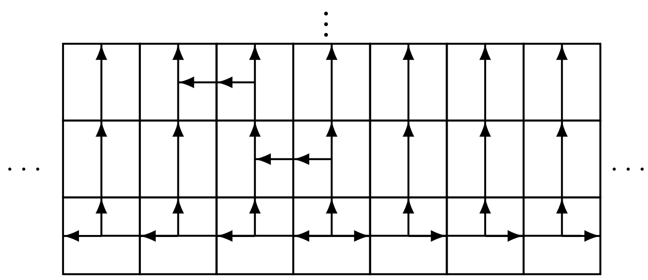


Figure 9: The simulation of M by the tile set P_M.

a tile set of some thousands of tiles. Later, Robinson gave a simpler proof using Robinson's tile set which is significantly smaller. Details can be found in the seminal papers [Ber72, Rob71]. Another intuitively more transparent and easier-to-use way to do this is to use substitution systems defining almost-odometers and a theorem of Mozes which states that every two-dimensional (2-D from now on) substitution system can be implemented with a tile set [Moz89, Hoc09]. However, mathematically this is equivalent to Robinson's construction since Mozes's theorem is proved by a more lengthy and ellaborate application of Robinson's arguments.

In this thesis, we will not show any of the two ways because both of them are very lengthy, but instead we have given references that direct the interested reader to the original articles.

Theorem 2. *The Tiling problem is undecidable.*

2.4 1-D case

The two main theorems stated in the previous sections were the existence of an aperiodic tile set and the undecidability of the Tiling problem. Since this thesis is about dimension-sensitive properties and differences between the one-dimensional and the multidimensional case, it is just a small logical step to assume that for 1-D "tile sets" things are different. In this section, we will briefly explain why this is the case. No proofs will be given, but rather an informal description and references to other documents.

Let S be a finite set called the *alphabet*. A *configuration* with *letters* from the alphabet S is a function $f \colon \mathbb{Z} \to S$. We say that a configuration $c \in S^{\mathbb{Z}}$ is *periodic*, if there exists some $n \geq 1$ such that $c(i+n) = c(i)$, for every $i \in \mathbb{Z}$. Instead of colors like in the 2-D case, we have a subset $X \subseteq S \times S$, that specifies which adjacencies are forbidden in the configurations. Obviously, we can also define Wang tile sets with such a set of forbidden adjacencies, although we must deal separately with the horizontal and the vertical directions. The (1-D *subshift*) S_X defined by X is the set of all configurations such that no forbidden adjacencies appear in it,

$$S_X = \{c \in S^{\mathbb{Z}} : c(i)c(i+1) \notin X, \text{ for every } i \in \mathbb{Z}\}$$

A similar definition can be given for the 2-D case. These objects are examples of so-called *subshifts of finite type* (or SFTs) and they will be the main topic of Chapter 3. Then, the results of the previous chapters can be understood to say that there exists a 2-D non-empty SFT with no periodic configurations and that the emptiness problem for 2-D SFTs is undecidable. On the other hand,

Theorem 3. *Every non-empty 1-D SFT contains a periodic point. It is decidable whether a 1-D SFT is non-empty.*

Proof. See [LM95]. The basic idea is that the elements of 1-D SFTs can be represented as infinite paths on finite directed graphs, and that a finite graph has an infinite path if and only if it has a cycle. Furthermore, a cycle corresponds to a periodic point and it is certainly decidable if a finite directed graph has a cycle. Both of the claims are now obvious. □

The book [LM95] has an excellent exposition of 1-D SFTs.

3 Cellular automata

3.1 Preliminaries

Cellular automata (CA from now on) are defined formally as quadruples $A = (d, S, N, f)$, where

- $d \geq 1$ is the *dimension* of A,
- S is a finite set called the *state set*,
- $N = (\vec{n_1}, \vec{n_2}, \dots, \vec{n_m})$, where $\vec{n_i} \in \mathbb{Z}^d$ and $\vec{n_i} \neq \vec{n_j}$ for $i \neq j$ is the *neighborhood vector*, and
- $f\colon S^m \to S$ is the *local function*.

Here, d defines the dimension of the configurations on which A will work. For example, if $d = 1$ then the space on which A acts is $S^{\mathbb{Z}}$. The elements of the neighborhood vector specify the (ordered) relative locations of the neighbors of a cell: the neighbors of cell $\vec{n}$ are the cells $\vec{n} + N = \{\vec{n} + \vec{n_1}, \vec{n} + \vec{n_2}, \dots, \vec{n} + \vec{n_m}\}$. The smallest natural number r such that $N \subseteq [-r, r]^d$ is called the *radius* of A. Intuitively, the radius of a CA is how far away a cell has to look in order to determine its state in the next time step.

In every time step, the local rule f is used to change a configuration c to another one c' in the following way:

$$c'(\vec{n}) = f(c(\vec{n} + N)) = f(c(\vec{n} + \vec{n_1}), c(\vec{n} + \vec{n_2}), \dots, c(\vec{n} + \vec{n_m}))$$

The transformation $c \mapsto c'$ defines a global function

$$F\colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d},$$

the *transition function* of the CA. This is our main object of study. In fact, when we talk about a CA, we will often refer only to its transition function.

For example, the 1-D left-shift σ is a 1-D CA with neighborhood $N = \{1\}$ and local rule the identity function $id\colon S \to S$. The transition function is thus defined as $\sigma(c)(n) = c(n + 1)$. Similarly, in 2-D , we define the shift-map $\sigma_{\vec{u}}$ for every $\vec{u} \in \mathbb{Z}^2$ as the CA with neighborhood $N = \{\vec{u}\}$ and local rule the identity function. The transition function satisfies $\sigma_{\vec{u}}(c)(\vec{n}) = c(\vec{n} + \vec{u})$. In the same way, we can define the shifts for every dimension $d \geq 1$.

Two widely used 2-D neighborhoods are the so-called *Moore* neighborhood $M = \{(x,y)\colon |x| \leq 1 \text{ and } |y| \leq 1\}$ and the *von Neumann* neighborhood $vN = \{(x,y)\colon |x| + |y| \leq 1\}$. Look at the following figure for a geometrical representation:

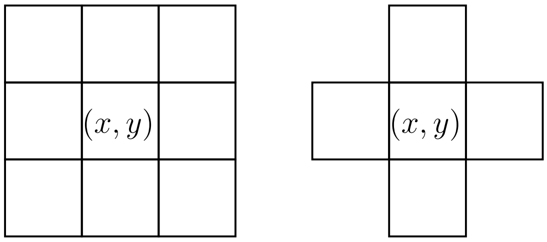


Figure 10: The Moore neighborhood of cell (x,y) on the left and the von-Neumann neighborhood on the right.

Let $G\colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ be a CA with local function g and suppose that there exists some state $q \in S$ such that $g(q,q,\ldots,q) = q$. State q is called a *quiscent state* of G. A configuration $c \in S^{\mathbb{Z}^d}$ is called *finite* if the set $supp(c) = \{\vec{n}\colon c(\vec{n}) \neq q\}$ is finite. Let S_F denote the set of all finite configurations with state set S. From the definition of a quiscent state it is obvious that if $c \in S_F$, then $G(c) \in S_F$. Hence, in the case where G has a quiscent state, we denote by

$$G_F\colon S_F \to S_F$$

the restriction of G to finite configurations.

Let S_P denote the set of totally periodic configurations of $\mathbb{Z}^d$ (a d-D configuration is called totally periodic if it has d linearly independent periods). A similar definition can be given for the restriction of G to S_P. In fact, for totally periodic configurations we do nott even have to assume the existence of a quiscent state, since if $c \in S_P$, then always $G(c) \in S_P$. We denote by

$$G_P\colon S_P \to S_P$$

the restriction of G on totally periodic configurations.

For our purposes, it is useful to imagine the di integer lattice $\mathbb{Z}^d$ to be equipped with the infinite norm:

$$|\vec{r}_\infty| = \max_{1 \leq i \leq d} r_i, \text{ for every } \vec{r} = (r_1, r_2, \ldots, r_d) \in \mathbb{Z}^d.$$

It turns out that by equipping $S^{\mathbb{Z}^d}$ with a naturally defined metric, we can turn it into a compact topological space that is particularly suited for the study of CAs.

Let $c, e \in S^{\mathbb{Z}^d}$. If $c \neq e$, then $d(c,e) = 2^{-k}$, where $k = \min\{|\vec{r}|_\infty : c(\vec{r}) \neq e(\vec{r})\}$.

There are at least two ways of describing the topology generated by this metric. We will use the simplest and most useful one: Let $D \subseteq \mathbb{Z}^d$ be a finite set and $p\colon D \to S$ be a function that assigns states only to the cells in D. The *cylinder*

$$Cyl(D,p) = \{c \in S^{\mathbb{Z}^d} : c(\vec{n}) = p(\vec{n}), \text{ for every } \vec{n} \in D\}$$

is defined as the set of all configurations that agree with p inside D. Let

$$\mathcal{B} = \{Cyl(D,p)\colon D \subseteq \mathbb{Z}^d \text{ is finite, and } p\colon D \to S\}.$$

It can be proven that $\mathcal{B}$ satisfies the conditions for being a base of a topology. Let $\mathcal{T}$ be the topology created by $\mathcal{B}$. The complement of a cylinder is a union of cylinders, namely

$$S^{\mathbb{Z}^d} \setminus Cyl(D,p) = \bigcup Cyl(D,p')$$

where the union is taken over all functions $p'\colon D \to S, p' \neq p$. Therefore, cylinders are clopen and since there is only a countable number of them, $\mathcal{T}$ has a countable clopen base. Even more useful is the following fact:

Proposition 4. $(S^{\mathbb{Z}^d}, \mathcal{T})$ *is a compact topological space.*

Proof. This is just a restatement of the compactness principle (Proposition 1). □

The following proposition is a fundamental topological characterization of CAs known as *Hedlund's theorem*:

Proposition 5. *A function* $G\colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ *is a CA if and only if it is continuous and it commutes with the shifts of* $\mathbb{Z}^d$.

For a proof, see [Hed69]. By the way, the reason for calling a result known as Hedlund's theorem a Proposition is that throughout this thesis only results that concern dimension-sensitive properties are termed Theorems, while Hedlund's theorem is valid in every dimension.

As known, a compact metric space with a continuous function is called a *dynamical system*. It is very fruitful to see CAs as dynamical systems and study them using ideas from this subject. Let (X, f) be a dynamical system and $x \in X$. Let $B_\delta(c) = \{e \in X : d(c, e) < \delta\}$ denote the open ball with center c and radius δ. We say that x is an *equicontinuity point* of f if

$$\forall\varepsilon, \exists\delta, \forall e \in B_\delta(c), \forall n \geq 1\colon d(G^n(c), G^n(e)) < \varepsilon.$$

f is called *sensitive* if

$$\exists\varepsilon, \forall c \in X, \forall\delta, \exists e \in B_\delta(c), \exists n \geq 1\colon d(G^n(c), G^n(e)) > \varepsilon.$$

Finally, f is *positively expansive* if

$$\exists\varepsilon\colon c \neq e \Rightarrow \exists n \geq 0\colon d(G^n(c), G^n(e)) > \varepsilon.$$

In Chapter 2.4 we will see how these notions are interpreted for CAs.

Finally, let G be a CA with a quiscent state. A finite configuration $\alpha \in S_F$ is called a *spaceship* if there exist $p \geq 1$ and $\vec{v} \in \mathbb{Z}^d$ such that $G^p(\alpha) = \sigma_{\vec{v}}(\alpha)$. Number p is called the *period* of the spaceship and $\vec{v}$ is the *displacement vector*.

3.2 Undecidability questions

In this section, we will examine dimension-sensitive undecidability questions concerning CAs. As a matter of fact, ”most” questions concerning 2-D CAs are undecidable, unless there is some trivial algorithm for them. For example, given the local rule of a CA, we can trivially decide if it has a quiscent state. On the other hand, for 1-D CAs there exist algorithms for a large class of natural algorithmic questions. More specifically, we will see that it is undecidable whether a 2-D CAs is reversible, surjective, open or number-decreasing. In the proofs, we will use some results whose proofs will be omitted. Hints and references will be given instead.

3.2.1 Reversibility

A CA is called *reversible* if it is bijective and its inverse function is also a CA. The following facts about reversible CAs are true in every dimension:

Lemma 2. *Every bijective CA is reversible.*

Proof. This can be proved using a straightforward compactness argument. See [Hed69]. □

Lemma 3. *Every injective CA is bijective.*

Proof. One can easily prove the following implications:

$$G \text{ injective } \Rightarrow G_P \text{ injective } \Rightarrow G_P \text{ surjective } \Rightarrow G \text{ surjective }.$$

$\square$

Proposition 6. *A CA is reversible if and only if it is injective.*

Proof. By definition, if a CA is reversible, it is also injective. The other direction is an immediate consequence of the previous two Lemmas. $\square$

In order to prove that it is undecidable if an arbitrary 2-D CA is reversible, we need a specific tile set with a property called the plane-filling property. Let us start with some definitions:

Directed tiles are normal Wang tiles to which a *follower vector* $\vec{f} \in \mathbb{Z}^2$ is associated. A *directed tile set* is a set of directed tiles, i.e. a pair (S, F), where S is a Wang tile set and $F\colon S \to \mathbb{Z}^2$ is a function that assigns a follower vector to every tile. From now on, we will refer to a directed tile set using only its ”base” tile set S. Let $c \in S^{\mathbb{Z}^2}$ be a configuration, which is not necessarily a valid tiling, and let $\vec{p} \in \mathbb{Z}^2$ be a position of the plane. The notion of validness of c in position $\vec{p}$ is the same as in the undirected case, which means that we do not care about follower vectors when we consider whether c is valid in $\vec{p}$ or not. The *follower of $\vec{p}$ in c* is the position $\vec{p} + F(c(\vec{p}))$. In other words, the follower is the cell to which the follower vector of $\vec{p}$ is pointing to. Notice that in different configurations the same position might have different followers. However, we will usually talk about the follower of a position, assuming that the configuration to which we are referring is fixed. Also, observe that the notions of follower position and validness are independent. Otherwise stated, in a configuration $c \in S^{\mathbb{Z}^2}$ every position has a follower, not only those positions where c is valid. In the tile set we are going to use, the follower of every position is one of the four adjacent positions, that is $F(a) \in \{(\pm 1, 0), (0, \pm 1)\}$, for every $a \in S$.

A sequence $\vec{p_1}, \vec{p_2}, \ldots, \vec{p_k}$, where every $\vec{p_i} \in \mathbb{Z}^2$ is called a *path* if $\vec{p}_{i+1}$ is the follower of $\vec{p_i}$, for every $i = 1, 2, \ldots, k-1$. The notions of *one-way infinite path* and *two-way infinite path* are defined analogously in the obvious way.

A directed tile set S is set to have the *plane-filling property* if it satisfies the following conditions:

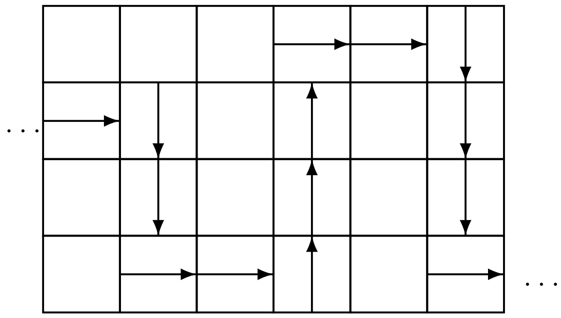


Figure 11: A path in a configuration.

1. S admits a valid tiling of the plane.
2. For every configuration $c \in S^{\mathbb{Z}^2}$, only two different types of infinite paths are defined:
   (a) There exists a tile on the path where the configuration is not valid, or
   (b) the path covers arbitrarily large squares.

Therefore, if the tiling conditions are not violated on any position of the path, then for every $n \geq 1$, there exist an $n \times n$ square each tile of which is visited by the path and, hence the path must be infinite. Notice, also, that this condition does not claim anything about the validity of the whole configuration. As long as the configuration is valid on the path, arbitrarily large squares are visited. This does not prevent tiling errors from occuring outside the path.

The above definitions would have absolutely no meaning if not for the following Proposition, whose proof will not be given. One can find it in [Kar94].

Proposition 7. *There exists a directed tile set with the plane-filling property.*

Let D be this directed tile set. Using it, we will prove that:

Theorem 4. *[Kar94] It is undecidable if an arbitrary* 2*-D CA is reversible.*

Proof. We are going to reduce the Tiling problem to the Reversibility problem of 2-D CA. Given an arbitrary tile set T, we algorithmically construct the following 2-D CA G:

The state set is $S = D \times T \times \{0, 1\}$. Therefore, the CA is working on configurations consisting of three different layers. In the first layer, there exist tiles from the fixed tile set D; in the second layer there exist tiles from the arbitrary tile set T, and on the third one there are the bits 0 and 1. The von Neumann neighborhood is used. The local rule only updates the bit components of a position $\vec{p}$ as follows:

- If either the T-layer or the D-layer contains a tiling error at $\vec{p}$, then the bit is not changed, but
- if the tiling is valid in both the D- and T-layers in $\vec{p}$, then the bit is changed by performing modulo 2 addition with the bit of the follower of $\vec{p}$.

Let us now prove that this CA is not injective (we know that this is equivalent to it not being reversible) if and only if T admits a valid tiling. This reduces the Tiling problem to the problem of deciding if a given 2-D CA is reversible and, hence, completes the proof of the Theorem.

Suppose that T admits a valid tiling t. Consider two configurations c_0 and c_1 where the D-components contain the same valid D-tiling, the T-components contain t and in c_0 all the bits are equal to 0, while in c_1 all the bits are equal to 1. Since the both the D- and T-layers are valid everywhere, modulo 2 addition is perofrmed in every position of the plane, hence

$$G(c_0) = G(c_1) = c_0,$$

and G is not injective.

Suppose, then, that G is not injective. This means that there exist two different configuration t_1 and t_2 such that $G(t_1) = G(t_2)$. The tile components are not changed by G and thus they must be identical in t_0 and t_1, so there is a position $\vec{p_1}$ where they have different bits. Since after the application of G these bits become identical the configurations are valid in both layers and the bits of the follower $\vec{p_2}$ of $\vec{p_1}$ must be different. Repeating this reasoning, we obtain an infinite path $\vec{p_1}, \vec{p_2}, \vec{p_3}, \ldots$ such that the tiling conditions are satisfied in both the D- and T-components at all positions of the path. Since D has the plane-filling property, this infinite path covers arbitrarily large squares. Therefore, T can also tile validly arbitrarily large squares, which means that it admits a valid tiling of the whole plane. $\square$

In fact, we have proved something slightly stronger, namely that the Reversibility problem is undecidable even for 2-D CA with the von Neumann neighborhood.

3.2.2 Openness

Let $G\colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ be a CA. Then, G is called *open*, if the following is true: If $W \subseteq S^{\mathbb{Z}^d}$ is open, then $G(W)$ is also open.

By tickling a little bit the construction of the proof in the previous section and examining carefully what kind of tilings the directed tile set D admits, we can prove that it is undecidable whether a given 2-D CA is open. For this, we also need the notion of left-permutiveness.

Let F be a 1-D CA with state set P, neighborhood $\{0,1\}$ and local rule f. We say that F is *left-permutive* if for every $b, c \in P$, there exists a unique $a \in P$ such that $f(a,b) = c$. This also means that $a_1 \neq a_2$ and $f(a_1,b_1) = f(a_2,b_2)$, then $b_1 \neq b_2$.

We denote by $a^\infty.b^\infty$ a configuration $c \in S^{\mathbb{Z}}$ such that $c(i) = a$ for all $i < 0$ and $c(i) = b$ for all $i \geq 0$.

Lemma 4. *There exists a left-permutive* 1*-D CA with state set* $\{0,1,2\}$ *which is not open.*

Proof. Let $F\colon S^{\mathbb{Z}} \to S^{\mathbb{Z}}$ be the 1-D CA with the neighborhood $\{0,1\}$ and the following local rule:

$x \backslash y$	0	1	2
0	0	1	1
1	1	0	2
2	2	2	0

F is left-permutive as every column of its transition matrix is a permutation of the state set. However, as $F(0^\infty.1^\infty) = F(0^\infty.2^\infty) = 0^\infty 1.0^\infty$, F is not right-closing. Since a 1-D CA is open if and only if it is both right- and left-closing (see [Kur03]), we conlcude that F is not open. $\square$

The following lemma is not necessary for the proof of Theorem 5, but it certainly makes its exposition more clear.

Lemma 5. *[Mey08] Let D be the tile set of Proposition 7. There exists a valid tiling d of D where all positions of the plane belong to the same path, i.e. they* *form a two-way infinite, non-intersecting, plane-filling path. In addition, every valid tiling $c \in D^{\mathbb{Z}^2}$ has at most four different infinite paths.*

Let us now change a little bit the construction of the previous subsection. Given an arbitrary tile set T, we construct a CA H as follows:

The state set is now $S = D \times T \times \{0, 1, 2\}$, the neighborhood is the von Neumann neighborhood, and the local rule only changes the bit components and does this only when the tiling is valid in both of the tile components. However, instead of performing modulo 2 addition, the new CA uses the local rule of F from Lemma 4 to determine the new bit. We claim that H is open if and only if it is reversible if and only if it T does not admit a valid tiling.

Indeed, if it is reversible, it is of course also open.

Suppose, on the other hand, that H is not reversible but it is open. Note that in the proof of Theorem 4, we used only the left-permutiveness of modulo 2 addition in order to conclude that T admits a valid tiling. Since f is also left-permutive, exactly the same reasoning applied in this case shows that T admits a valid tiling t. Let d be the tiling from Lemma 5 and consider the set $B = d \times t \times \{0, 1, 2\}^{\mathbb{Z}^2}$. Since in d there exists only one path and both d and t are valid everywhere, the restriction of H on B is in some sense the same as the 1-D CA F. Indeed, $H(d, t, c) = (d, t, F(c))$, where the 2-D configuration c is interpreted as the 1-D configuration obtained by following the unique path in d. Using the same interpetation, we have that $H^{-1}(B) = B$ and $H(d, t, W) = (d, t, F(W))$, for every $W \subseteq \{0, 1, 2\}^{\mathbb{Z}^2}$. According to Theorem 1,page 116 in [Kur66], if $W \subseteq \{0, 1, 2\}^{\mathbb{Z}^2}$ is open, then $H(d, t, W) = (d, t, F(W))$ is also open in the relative topology. In addition, since it is a basic topological fact that all projections are open, $F(W)$ is also open. But this means that F is an open CA, which is a contradiction. Therefore, H is not open.

We have proved the following equivalences:

$$H \text{ reversible } \Leftrightarrow H \text{ open } \Leftrightarrow T \text{ does not admit a valid tiling}$$

From this we can finally conclude that:

Theorem 5. *It is undecidable if an arbitrary 2-D CA is open.*

We have actually proved something stronger, namely that the class of reversible CA is recursively inseparable from the class of non-open CAs.

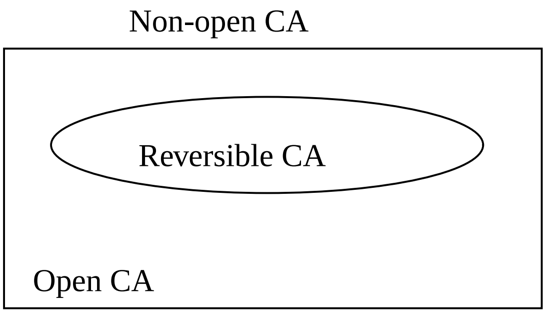


Figure 12: Reversible CAs are recursively inseparable from non-open CAs.

3.2.3 Surjectivity

In this subsection, we are going to prove that it is undecidable if an arbitrary 2-D CA is surjective. In order to do this, we first have to talk about another undecidable problem concerning tiles, the Finite Tiling problem, and in addition introduce a specific tile set with a special property. The contents of this subsection are based on Jarkko Kari's lecture notes for the class "Cellular Automata" taught during the spring semester of 2009 in the University of Turku.

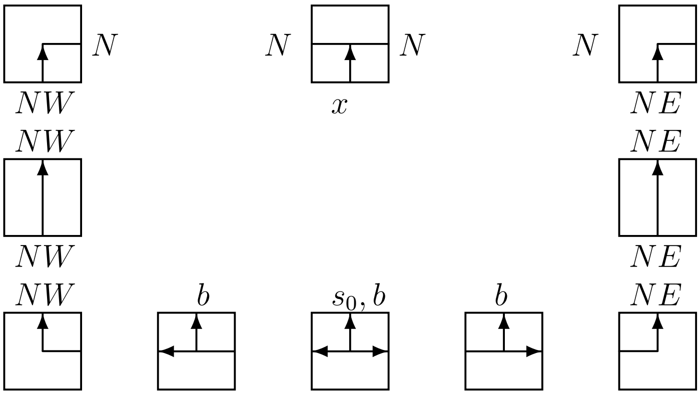
In the Finite Tiling problem we are given a tile set T together with a specified tile $b \in T$, the *blank tile*. The blank tile has the same color in all of its sides. A *finite tiling* is a valid tiling where only a finite number of tiles are non-blank. A finite tiling where all tiles are blank is called *trivial*. The *Finite Tiling problem* consists in deciding whether an arbitrary tile set T with a blank tile b admits a valid, finite, non-trivial tiling. It is a simple excercise, no aperiodic tile sets or ingenious simulating methods needed, to prove that:

Proposition 8. *The Finite Tiling problem is undecidable.*

Proof. Given an arbitrary TM $M = (S, \Gamma, s_0, s_h, \delta, b)$ we know from the first chapter how to construct a tile set P_M with a specified tile, the seed tile, such that M does not halt if and only if P_M admits a valid tiling with at least one appearance of the seed tile. We will know modify a little that construction to create a new tile set Q_M such that M halts if and only if Q_M admits a valid, finite, non-trivial tiling. Notice that in the first case negative instances of the Halting problem are associated to positive instances of the Tiling problem, while in the second one positive instances of the Halting problem are associated to positive instances of the Finite Tiling problem.

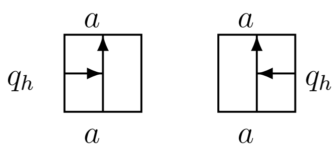
In addition to the machine tiles, that is *alphabet, action* and *merging tiles*, the tile set Q_M contains the blank tile and some more tiles to initiate and terminate the computation of M. Namely,

1. The following *boundary tiles*:



   The tile in the middle of the bottom line is called the *initialization tile*, while the tile in the middle of the upper line is constructed for every $x \in \Gamma$.

2. For every $x \in \Gamma$, the following *halting tiles*:



One can now prove that Q_M admits a valid, finite, non-trivial tiling if and only if M halts. We will give a sketchy proof of this.

Suppose, first, that M halts. Then, we can enclose its computation in a finite rectangle and put the blank tile outside from this rectangle. Hence, in this case a valid, finite, non-trivial tiling exists. Notice how the halting tiles are use to make the head of M disappear and then the upper boundary tiles "absorb" all the letters.

Suppose, on the other hand, that a valid, finite, non-trivial tiling exists. Then, some upper boundary tile must have been used, since boundary tiles are the only tiles that can have the blank tile as a northern neighbor. Using the hypothesis of finiteness, we can see that a rectangle is enforced, inside which the computation of M takes place. Since the tiling is valid, the machine head of M must have disappeared at some moment, which means that M halts. $\square$

In order to prove our main undecidability result, we also need the fixed tile set of Figure 13:

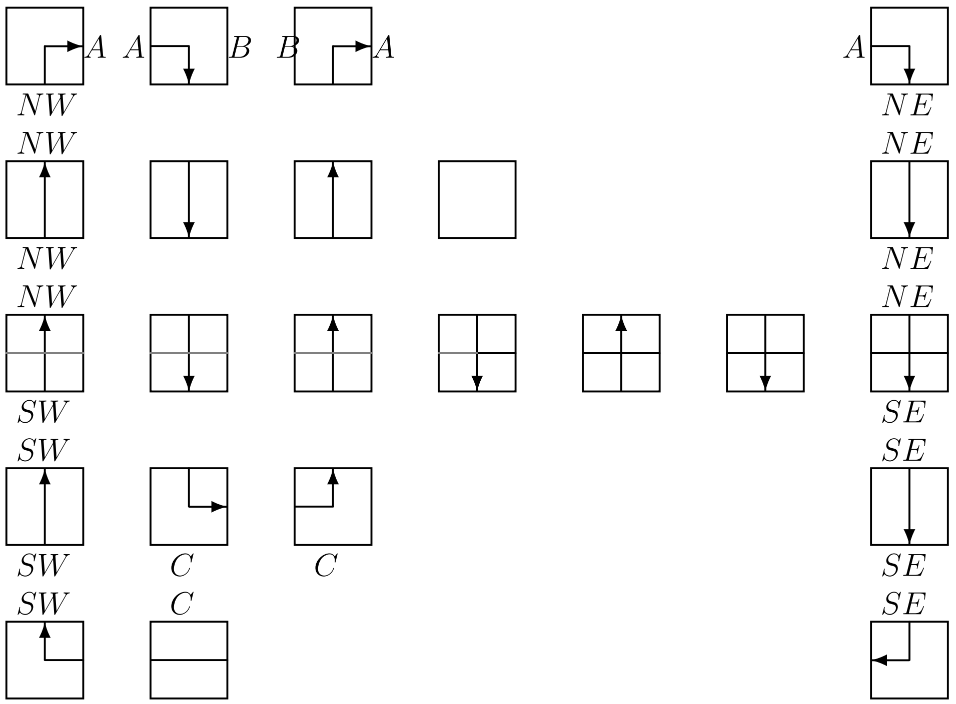


Figure 13: The tile set E.

We denote this tile set by E. The tile without any arrows or lines is called the *blank tile*. All the other tiles have a unique incoming and outgoing arrow. In valid tilings, incoming and outgoing arrows of adjacent tiles must match. The non-blank tiles are considered to be directed. The follower of a tile is the neighbor pointed by the outgoing arrow on the tile.

The tile where the light and dark grey horizontal thick lines meet is called the *cross*, and it will play a special role in the forthcoming proof of Theorem 6. Finally, a *rectangular loop* is a valid tiling of a rectangle with tiles from E, where the follower path forms a loop that visits every tile of the rectangle and the outside border of the rectangle is colored blank. By inspecting the tile set E, it is easy to see that there exist rectangular loops of size $2n \times m$, for every $n \geq 2$ and $m \geq 3$. Every rectangular loop contains a unique cross which can be placed in any position in the interior of the rectangle.

Figure 14 shows a rectangular loop of size 10×5, where the labels have been omitted:

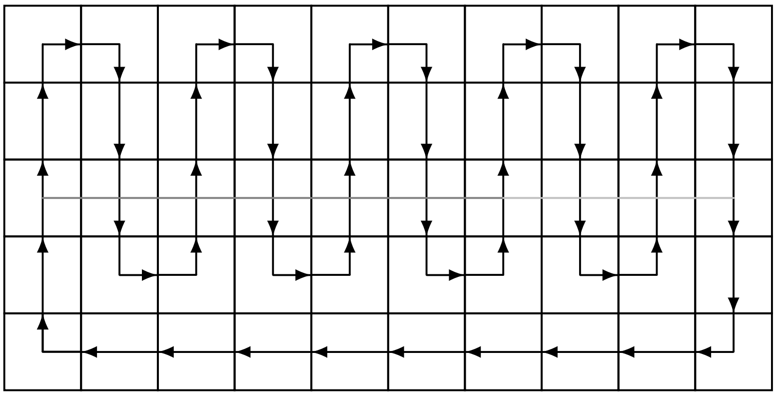


Figure 14: A rectangular loop of size 10×5.

The tile set E has the following convenient property:

Lemma 6. *Let $t \in E^{\mathbb{Z}^2}$ be a configuration, and $\vec{p_1}, \vec{p_2}, \ldots$ a path in t such that t is valid at $\vec{p_i}$, for all $i \geq 1$. If the path covers only a finite number of different cells, then the cells on the path form a rectangular loop.*

Proof. We note that in a loop there must exist at least one tile pointing to every direction. Therefore, the existence of at least one left arrow is guaranteed. After that, everything follows from the hypothesis of finiteness. $\square$

We will also need the following basic result, which is one of the first results about the mathematical properties of CAs:

Proposition 9 (Garden of Eden theorem)**.** *[Moo62, Myh63] Let G be a CA with a quiscent state. Then, G is surjective if and only if G_F is injective.*

We are now ready to prove the main theorem of this subsection:

Theorem 6. *It is undecidable if an arbitrary 2-D CA is surjective.*

Proof. We will reduce the Finite Tiling problemn to the 2-D Surjectivitry problem. Let b and c be the blank and the cross of the tile set E, respectively. Given an arbitrary tile set T with blank tile B, we can algorithmically construct the following 2-D CA G:

The state set contains triplets $(e, t, x) \in E \times T \times \{0, 1\}$ under the following constraints:

- if $e = c$, then $t \neq B$, and

- if $e = b$ or e is any tile containing SW,SE,NW,NE,A,B or C as a label, then $t = B$.

In other words, the cross is always paired with a non-blank tile while the tiles on the boundary of a rectangular loop are paired only with the blank tile. The triplet $(b, B, 0)$ is the quiscent state of G.

As in the proof of the undecidability of the 2-D Reversibility problem, the local rule will only update the bit of a triplet leaving the tiles in the first two layers unchanged. This is done in the following way: Let (e, t, x) be the current state:

- if $e = b$, then the state is not changed,
- if $e \neq b$, then G checks the validity of the tiling in the first two layers. If a tiling error occurs in either layer, the state is not changed. If the tiling is valid in both layers, the bit is updated by performing addition modulo 2 with the bit in the follower position.

Let us prove that this G is not surjective if and only if T admits a valid, finite, non-trivial tiling t:

Assume that T admits a valid, finite, non-trivial tiling t. Consider two finite configurations c_0 and c_1 whose T-components contain t and the E-components have the same rectangular loop that contains all non-blank tiles of t. In c_0 every bit is equal to 0, while in c_1 the bits inside the rectangular loop are equal to 1 and all the bits outside of the loop are equal to 0. The local rule updates only the bits in the rectangles in such a way that

$$G(c_0) = G(c_1) = c_0.$$

Since c_0 and c_1 are finite configurations, G_F is not injective. According to the Garden of Eden theorem, G is not surjective.

Assume, then, that G is not surjective. According to the Garden of Eden theorem, there exist two different finite configurations e_0 and e_1 such that $G(e_0) = G(e_1)$. Since G updates only the bit components, the T- and E- components of e_0 and e_1 are identical. There exists a cell $\vec{p_1}$ where e_0 and e_1 have different bits. Since these bits become identical in the next step, the tiling conditions must be satisfied in position $\vec{p_1}$ and the bits must also be different in the follower position. Repeating this argument, there exists an infinite sequence $\vec{p_1}, \vec{p_2}, \vec{p_3}, \ldots$ such that the E-tiling is valid at $\vec{p_i}$ for all $i \geq 1$. Moreover, e_0 and e_1 have different bits at each position $\vec{p_i}$. Since the configurations e_0 and e_1 are finite, the path visits only a finite number of different positions. According to Lemma 6, it forms a rectangular loop. In addition, the tiling in the T-components is also valid. Because of the constraints on the allowed triplets, the tiles in the T-component on the boundary of the rectangle are all the blank B, while the tile in the position of the cross is a non-blank tile. Therefore, there exists a valid tiling of a rectangle with blank boundary and a non-blank tile in the interior. This is equivalent to the existence of a valid, finite, non-trivial tiling. □

3.2.4 Number-decreasing CA

The problems proved to be undecidable for 2-D CAs in the previous subsections concern general set-theoretic (reversibility, surjectivity) or topological (openness) properties that can be posed for any dynamical system. On the other hand, number-decreasingness, which will be the subject of this subsection is a combinatorial property that can only be defined for CA. It is dimension-sensitive, too, and, at the time when this thesis is being written, it is the only dimension-sensitive undecidability property with a purely CA-theoretic flavor.

Let $A = (d, S, N, f)$ be a CA with state set $S = \{0, 1, \ldots, n-1\}$, for some $n \geq 1$ which has 0 as a quiscent state. G is called *number-decreasing* if

$$\sum_{\vec{n} \in \mathbb{Z}^d} G(c)(\vec{n}) \leq \sum_{\vec{n} \in \mathbb{Z}^d} c(\vec{n}),$$

for every finite configuration $c \in S_F$. Both sums are well-defined, since 0 is a quiscent state and c is a finite configuration.

The intuitive idea behind the formulation is that number-decreasing CAs can be used to model systems where a measurable quantity, the number of particles or mass for example, does not increase during the evolution of the system.

In [BDFK05], it is proved that it is decidable if a 1-D CA is number-decreasing, and a combinatorial, physics-oriented characterization of those local rules that give rise to number-decreasing CAs is given. It is also proved that if a CA is strictly number-decreasing, i.e. for at least one finite configuration the above inequality is strict, then it is not surjective. The proof of the following theorem is from the same paper as well:

Theorem 7. *[BDFK05] It is undecidable if an arbitrary* 2*-D CA is number-decreasing.*

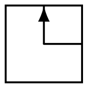
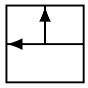
Proof. Given an arbitrary TM M, we will algorithmically construct a 2-D CA A_M such that M halts if and only if A_M is not number-decreasing. Hence, the Tiling problem can be reduced to the Number-decreasingness problem, and thus the Number-decreasingness problem is undecidable. Let M be an arbitrary TM. We construct the tile set Q_M from the proof of Proposition 8. Recall that this tile set admits a valid, finite, non-trivial tiling if and only if M halts. Let us now define the CA A_M:

The state set of A_M is $S = \{0, 1, , \ldots, n-1\}$, where $n \geq 1$ is the number of tiles in Q_M. We establish a one-to-one correspondence ϕ between S and T_M such that 0 is mapped to the blank tile and 1 is mapped to the initialization tile. Using ϕ we can identify S and Q_M and talk about tiling conditions between the elements of S. The neighborhood of A_M is the standard von-Neumann neighborhood and the local rule updates the state of the cell as follows:

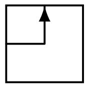
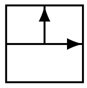
If a non-blank (i.e. non-0) tile sees a tiling error in its von Neumann neighborhood, it decreases its value by 1. If an initialization tile (i.e. a cell with value 1) does not see a tiling error, it increases its value by 1. Otherwise, the state of the cell is left unchanged. Let us prove that M halts if and only if A_M is not number-decreasing:

Assume, first, that M halts. Then, Q_M admits a valid, finite, non-trivial tiling f where the initialization tile is used at least once. Using ϕ^{-1}, f is mapped to a finite configuration $c\colon \mathbb{Z}^2 \to S$. Since f is valid, so is c and thus no cell sees a tiling error. Therefore, no cell decreases its value, while the cell in position 1 (where there is an initialization tile in f) increases its value by 1. So, the number-decreasing inequality is not satisfied for c and A_M is not number-decreasing.

On the other hand, assume that M does not halt. Let $c\colon \mathbb{Z}^2 \to S$ be a finite configuration. Our aim is to associate every initialization tile in c to a unique tiling error. Let $\vec{p}$ be a position at state 1. We define inductively the *domain* $D_{\vec{p}}$ of $\vec{p}$ as follows:

- $\vec{p} \in D_{\vec{p}}$.
- If in position $\vec{q}$ there is a tile NW  or b  and its right neighbor is in

$D_{\vec{p}}$, then $\vec{q} \in D_{\vec{p}}$.

- Similarly, if in position $\vec{q}$ there is a tile NE  or b  and its left neighbor is in $D_{\vec{p}}$, then $\vec{q} \in D_{\vec{p}}$.

- If in $\vec{q}$ there is an alphabet, action, merging or halting tile or any other type of boundary tile and its southern neighbor is in $D_{\vec{p}}$, then $\vec{q}$ is in $D_{\vec{p}}$, too.

- These are the only positions that belong to $D_{\vec{p}}$.

It is easy to prove that if $\vec{p}$ and $\vec{r}$ are two different positions in state 1, then $D_{\vec{p}} \cap D_{\vec{r}} = \emptyset$. In addition, every $D_{\vec{p}}$ contains a tiling error. Indeed, if some $D_{\vec{p}}$ did not contain a tiling error, then its domain would represent a halting computation of M, which contradicts the hypothesis that M does not halt.

We have thus associated every cell in state 1 to a unique tiling error. This means that the number of tiling errors is greater than or equal to the number of initialization tiles. By the definition of the local rule of A_M, this means that A_M is number-decreasing. $\square$

3.2.5 1-D case

All the above decision problems are decidable in 1-D . The reason for this is quite similar to that of the previous chapter: namely, we can represent 1-D CAs using finite directed graphs, and CA decidability questions are translated as some trivially decidable graph-theoretic decidability questions. For example, to decide if a given 1-D CA is injective, all one has to do is to construct a certain directed graph (the pair graph) and then examine if there is a cycle in this graph that contains a vertex outside from a specified subset of the vertices.

Notice that for different undecidability problems one might have to build different graphs. Here, we will exhibit how to construct the most basic directed graph that can be associated to a 1-D CA and use it to give an algorithm for deciding if a 1-D CA is number-decreasing.

Let $m \geq 1$. We construct the *de-Bruijn graph* of width m. This is a directed graph with vertex set $V = S^{m-1}$, edge set $E = S^m$ with the edge $s_1 s_2 \ldots s_m$ leaving from vertex $s_1 s_2 \ldots s_{m-1}$ and reaching vertex $s_2 s_3 \ldots s_m$. Figure 3.2.5 shows the de-Bruijn graph of width 3 for the set $S = \{0, 1\}$:

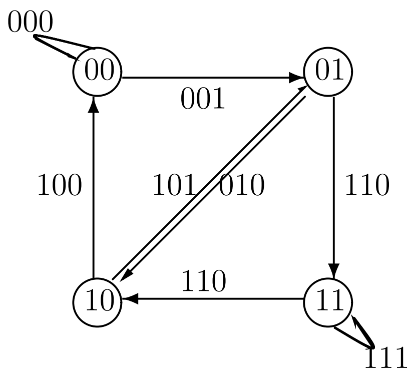


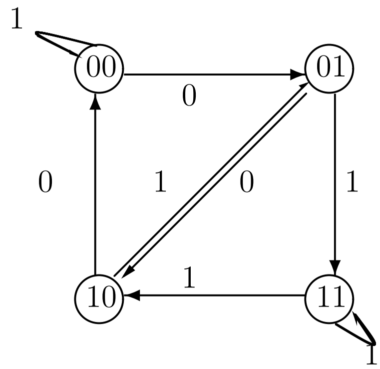
Figure 15: The de-Bruijn graph of order 3 for the set $S = \{0, 1\}$.

There exists a natural one-to-one correspondence between two-way infinite paths in the de-Bruijn graph and configurations in $S^{\mathbb{Z}}$: for every two-way infinite path $p\colon \mathbb{Z} \to E$ there is a corresponding element $c_p \in S^{\mathbb{Z}}$ such that

$$p(i) = c_p(i)c_p(i+1)\ldots c_p(i+m-1).$$

This correspondence is well defined because of the overlap of the edges of the de-Bruijn graphs. One can also easily prove that it is bijective. For every configuration $c \in S_{\mathbb{Z}}$ one can define a two-way infinite path $p_c\colon \mathbb{Z} \to E$ such that $p_{c_p} = p$ and $c_{p_c} = c$.

Now, let G be a 1-D CA defined by the local function g. Without loss of generality we can suppose that its neighborhood is a contiguous segment of $m \geq 1$ cells. We construct the de-Bruijn graph of width m as above, but in this case we also add labels to the edges: The label of the edge $s_1s_2\ldots s_m$ is $g(s_1s_2\ldots s_m)$. This directed graph is called the ***de-Bruijn graph of G***. it contains almost all of the information concerning G. Notice, however, that it does not capture the relative positioning of the neighborhood, i.e. the CA G and $\sigma \circ G$ have the same de-Bruijn graphs. However, since σ is an isomorphism of $S^{\mathbb{Z}}$, most of the properties that are of interest are invariant under the shifts. For example, for pr $\in$ {reversible, surjective, open, number-decreasing}, G is pr if and only if $\sigma \circ G$ is. Below we give an example of the de-Bruijn graph of a CA:



The two-way infinite paths of this directed graph give us two sequences: the first one is the sequence c_p defined earlier and the other one is the sequence g_p of the labels of the edges. Because the lables are the outputs of the local function, g_p is, a possibly translated version of, $G(c_p)$, i.e

$$g_p = \sigma^k(G(c_p)), \text{ for some } k \in \mathbb{Z}.$$

Different versions of the above graph are needed to attack different problems. However, in order to provide an algorithm for number-decreasingness we do not need any more technical machinery:

Theorem 8. *[BDFK05] A* 1*-D CA is number-decreasing if and only if for every simple (aka not self intersecting) circle of its de-Bruijn graph the value of the name of the circle is greater than or equal to the sum of the labels of the edges of the circle.*

Proof. Let G be a 1-D CA. For the sake of simplicity, we will assume that the range of G is 3 and the state set is $S = \{0, 1\}$. All of the following arguments can be adapted to be valid in the general case.

We construct the de-Bruijn graph of G as described above. To make clear what the name of a circle means, let $000 \rightarrow 001 \rightarrow 010 \rightarrow 000$ be a circle. Its name is 0001 (we pick the first bit from every vertex) and its labels are $f(000), f(001), f(010), f(100)$.

Suppose, first, that the value of the name of every simple circle is greater than or equal to the sum of its lables. Let $c \in S^{\mathbb{Z}}$ be a finite configuration. Then, the two-way infinite path p_c is actually a finite circle, since p_c stays almost always in vertex 000 both in the left and in the right part of the path. Since every circle can be decomposed into disjoint simple circles and the number-decreasing condition is additive and is satisfied in all of the simple circles, G is number-decreasing.

If, on the other hand, there exists a simple circle for which the above is not true, then we can construct a finite configuration that does not satisfy the number-decreasing condition as follows:

Let z be this "bad" circle. Every vertex of the graph can be reached by 000 with a path of length at most two. Let p be the shortest path from 000 to any vertex of z. The finite configuration corresponding to the path $(000)^{\infty}pz^5p^r(000)^{\infty}$ violates the number-decreasing condition, where p^r is the reverse path of p. To see this, consider that both p and p^r are paths of length at most two, so their joint contribution to the difference $\sum\limits_{x\in\mathbb{Z}} G(c)(x) - \sum\limits_{x\in\mathbb{Z}} c(x)$ is at least -4. In addition, since z is a "bad" circle, every copy of z adds at least 1 to this difference. There are five copies of z, hence

$$\sum_{x\in\mathbb{Z}} G(c)(x) - \sum_{x\in\mathbb{Z}} c(x) \geq 1$$

So, G is not number-decreasing.

Since there exists only a finite number of different simple circles in a finite directed graph and we can find them all algorithmically, Number-decreasingness is decidable in 1-D . $\square$

For proofs of the decidability of injectiveness, surjectiveness and openness of 1-D CA see [AP72, Wil91, Nas78, Sut91]

3.3 Topological entropy properties

In regards to topological entropy, the difference between the 1-D case and the multidimensional case is pretty straightforward: In 1-D every CA has finite entropy which is bounded by a factor that depends only on the radius of the CA and its number of states, while the topological entropy of large classes of 2-D CAs is infinite. All in all, this could be the context of this whole section if it were not for some interesting theorems and some even more interesting constructions that shed more light on what makes multidimensional CAs have such a different behaviour than 1-D CAs. As a matter of fact, topological entropy of multidimensional CAs is a more complex subject than topological entropy of 1-D CAs so that most of the theorems cannot even be stated in the 1-D case.

In the theory of dynamical systems, topological entropy (with the term "topological" to distinguish it from measure-theoretic entropy) is a measure of the com-

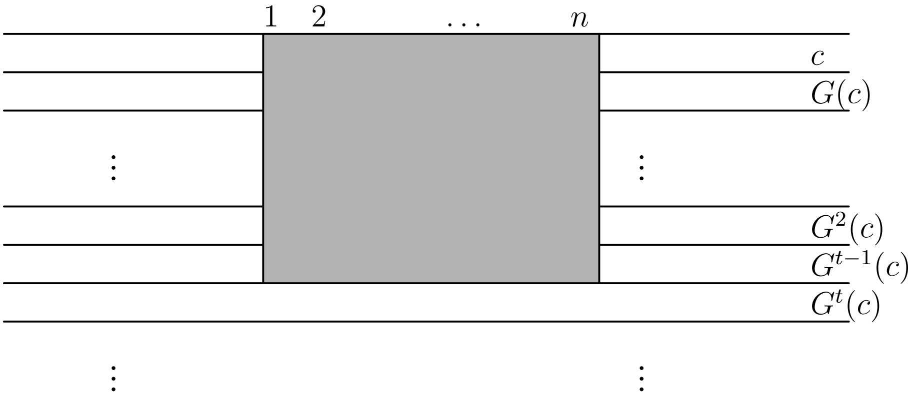


Figure 16: $W(n,t)$ consists of all the elements that can be taken in the above way.

plexity of a dynamical system. It has some nice categorical properties and is a useful invariant of conjugacy. For a detailed, deep introduction with a lot of insights about the intuitive meaning behind the formal definitions, see [Wal75].

In the context of CAs, topological entropy, like most of the analytical properties defined for general dynamical systems, has an equivalent combinatorial description. Suppose that G is a 1-D CA. Let $I_n = \{1, 2, \dots, n\}$ be the segment of length n. Let

$$W(n,t) = \{(c_{I_n}, G(c)_{I_n}, \dots, G^{t-1}(c)_{I_n}) \colon c \in S^{\mathbb{Z}^2}\}.$$

Intuitively, $W(n,t)$ is the set of different orbits distinguished by an observer that can observe the system only fot t time steps and only in n consecutive positions. Then, the *topological entropy* of G is defined as:

$$h(G) = \lim_{n\to\infty} \lim_{t\to\infty} \frac{\log |W(n,t)|}{t}.$$

For a proof of the existence of this limit, see [Kur03, Wal75].

A similar definition can be given for the topological entropy of 2-D CAs and, generally, d-D CAs. The only difference in the definition is that instead of using the segments I_n of length n, we use the rectangles $R_n = \{1, 2, \dots, n\}^2$ or, in the general case, the d-dimensional hypercube of side n.

As stated previously, the topological entropy of every 1-D CA is finite. This can be seen with the following argument: Let G be a 1-D CA and let r be its radius. If $c, e \in S^{\mathbb{Z}}$ and $c_{[-r+1,n+r]} = e_{[-r+1,n+r]}$, then $G(c)_{[1,n]} = G(e)_{[1,n]}$. Similarly, since a segment of length $n + 2r(t-1)$ uniquely determines the evolution of the central n cells for the first t time steps, we can prove that

$$|W(n,t)| \le |S|^{n+2r(t-1)}, \text{ for every } n, t \ge 1.$$

Therefore,

$$h(G) = \lim_{n\to\infty}\lim_{t\to\infty} \frac{\log |W(n,t)|}{t} \le \lim_{n\to\infty}\lim_{t\to\infty} \frac{\log |S|^{n+2r(t-1)}}{t} \le 2r\log|S|.$$

However, the same argument cannot be applied for 2-D CAs. The reason is that the size of the boundary of the rectangle R_n tends to infinity when n goes to infinity, while for 1-D CAs the size of the boundary of I_n is constant for all $n \ge 1$. The same idea will be used when we prove that there do not exist any positively expansive 2-D CAs. As a matter of fact, even the 2-D shifts have infinite topological entropy. This can be seen either by proving directly that for the horizontal shift $\sigma_{(1,0)}$, for example, $|W(n,t)| = |S|^{n^2+n(t-1)}$, or can be termed a corollary of the following much more general theorem of Lakshtanov and Langvagen:

Theorem 9. *[LL04] If a 2-D CA has a spaceship, then its topological entropy is infinite.*

Proof. Let G be a 2-D CA with a spaceship and quiscent state q. (The existence of a quiscent state is not necessary for the proof of the Theorem, but it is a convention to accept that a quiscent state exists when talking about spaceships.) Let us recall that a spaceship with period $p \ge 1$ and displacement vector $\vec{u} \ne \vec{0} \in \mathbb{Z}^2$ is a non-trivial finite configuration α such that $G^p(\alpha) = \sigma_{\vec{u}}(\alpha)$. We will show that a 2-D CA with a spaceship can realize systems with arbitrarily large topological entropy.

Let S be the state set of G. Consider the 1-D right-shift σ of $X = \{0,1\}^{\mathbb{Z}}$, where $\sigma(x)_i = x_{i+1}$. It can be easily proved that $h(\sigma) = \log 2$. Let α be a spaceship of G with period $p \ge 1$ and displacement vector $\vec{u}$. We can find a number $l \ge 1$ such that the square $K = [-l,l] \times [-l,l]$ contains all the non-quiscent states of $\alpha, G(\alpha), \ldots, G^{p-1}(\alpha)$ and all of its neighbors. If we choose $m = 2l + 2$, then we have that

$$(K + mi\vec{u}) \cap (K + mi'\vec{u}) = \emptyset, \text{ whenever } i \ne i'.$$

Indeed, if $(K + mi\vec{u}) \cap (K + mi'\vec{u}) \ne \emptyset$, then there exist $\vec{k_1}, \vec{k_2} \in K$ such that

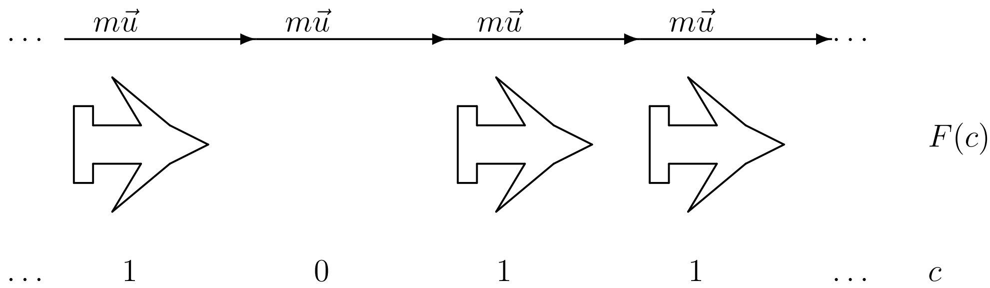


Figure 17: A rather informal presentation of F.

$$
\begin{aligned}
& \vec{k_1} + mi\vec{u} = \vec{k_2} + mi'\vec{u} \\
\Rightarrow\ & \vec{k_1} - \vec{k_2} = m(i'-i)\vec{u} \\
\Rightarrow\ & |\vec{k_1} - \vec{k_2}|_\infty = |m(i'-i)\vec{u}|_\infty \\
\Rightarrow\ & |\vec{k_1} - \vec{k_2}|_\infty \geq m|\vec{u}|_\infty \geq m \\
\Rightarrow\ & |\vec{k_1} - \vec{k_2}|_\infty \geq 2l+2
\end{aligned}
$$

which is impossible since $\vec{k_1}$ and $\vec{k_2}$ belong in a square with diameter $2l$.

Now, let us describe a function $F\colon X \to S^{\mathbb{Z}}$. Configuration $c \in X$ will be sent to $F(c) = \sum_{i\colon c_i=1} \sigma_{mi\vec{u}}(\alpha)$. The summation on the right hand side means joining the supports of the summed configurations. This is well-defined, since these supports are disjoint according to the definition of m. Intuitively, a 1 is coded as a spaceship and a 0 is coded as the absence of the spaceship. We also leave enough empty space between them so as they evolve independently.

We claim that F embeds the 1-D right-shift σ into G^{mp}. Indeed, it is obvious that F is injective, and a moment's thought can convince us that it is also continuous. We will now show that it is a map between the dynamical systems. We claim that $G^{mp} \circ F = F \circ \sigma$. In fact,

$$
\begin{aligned}
& G^{mp} \circ F(c) = G^{mp}(F(c)) = \\
= \ & G^{mp}\Big(\sum_{i\colon c_i=1} \sigma_{mi\vec{u}}(\alpha)\Big) \overset{(1)}{=} \sum_{i\colon c_i=1} \sigma_{mi\vec{u}}(G^{mp})(\alpha) \overset{(2)}{=} \\
= \ & \sum_{i\colon c_i=1} \sigma_{mi\vec{u}}(\sigma_{m\vec{u}}(\alpha)) = \sum_{i\colon c_i=1} \sigma_{m(i+1)\vec{u}}(\alpha) = F \circ \sigma(c)
\end{aligned}
$$

where equality (1) is true because m was chosen so as spaceships evolve independently and equality (2) is true because α is a spaceship. Therefore, our claim is true and this proves that the right-shift is a subsystem of G^{mp}. This means that $h(G^{mp}) \geq h(\sigma) = \log 2$.

Now, using the extra space offered by the second dimension, we can simulate the product of the right-shift with itself. Indeed, if $(c^1, c^2) \in X \times X$ is a pair of configurations, then we can define:

$$F(c) = \sum_{i:\ c_i=1} \sigma_{mi\vec{u}+\vec{v}}(\alpha) + \sum_{i:\ c_i=1} \sigma_{mi\vec{u}+2\vec{v}}(\alpha)$$

where $\vec{v}$ and $\vec{u}$ are linearly independent and m is chosen such that if $(i,j) \neq (i',j')$, then

$$(K + m(i\vec{u} + j\vec{v})) \cap (K + m(i'\vec{u} + j'\vec{v})) = \emptyset.$$

Exactly as in the previous case, we can prove that this is an embedding of $\sigma \times \sigma$ into G^{mp}, for the reason that spaceships put on different $\vec{v}$-levels do not interact. Hence, $h(G) \geq h(\sigma \times \sigma) = 2\log 2$. Generally, the same argument shows that

$$h(G^{mp}) \geq n \log 2,\ \text{for every } n \geq 1.$$

This means that $h(G^{mp}) = \infty$, and since $h(G^{mp}) = mph(G)$, we can finally conclude that $h(G) = \infty$. $\square$

There is another large class of 2-D CA with infinite topological entropy:

Theorem 10. *[DMM03] The topological entropy of a* 2*-D sensitive linear CA is infinite.*

Proof. We will not give the proof here. The main idea is that 2-D sensitive linear CA satisfy a certain kind of 2-D -permutivity. $\square$

On the other hand, there exist trivial examples of 2-D CA with zero topological entropy. For example, periodic or equicontinuous CA. For some time, it was not known whether these two case exhaust the class of 2-D CA, i.e. it was not know whether there exists a 2-D CA with finite, positive entropy, until Meyerovitch proved that:

Theorem 11. *[Mey08] There exists a* 2*-D CA with finite, positive topological entropy.*

Proof. We will give an informal, non-technical exposition of the ideas used in the proof. A full proof can be found in [Mey08].

The essence of the proof of Theorem 9 is that we can fit in an infinite number of lines in the plane. In every line, we can simulate a 1-D CA, thus giving rise to a CA of infinite topological entropy. If, in some way, we could partition the plane into a finite number of infinite lines, then we would only have a finite number of 1-D CA and so guarantee finite topological entropy.

The way to do this is to use Kari's tile set D from Proposition 7. We construct a CA G with state set $S = D \times \{0, 1\}$. The neighborhood of G is the von Neumann neighborhood and the local rule is very similar to the local rule of Theorem 4:

- if the matching is valid in position $\vec{p}$, then G updates the bit of $\vec{p}$ by performing modulo 2 addition with the bit of the follower position.
- If the matching is not valid in position $\vec{p}$, then the bit remains as it is.
- The tile components never change.

Therefore, given a configuration $c \in S^{\mathbb{Z}^2}$, the tile layer divides the plane into some number of finite and infinite paths. There could be an infinite number of finite paths, but we know from Lemma 5 that there exist at most four (one-way or two-way) infinite paths. Based on this, we can prove that G has finite, positive topological entropy. This is because in every finite path the action of G is periodic and, hence, has entropy 0, while in the infinite paths it has positive entropy h. Therefore, the entropy of G is bounded from above by $4h$. □

Recently, M. Hochman gave a complete characterization of the class of numbers which are the topological entropies of d-dimensional CA, where $d \geq 3$. The proof of this characterization, which is certainly outside the scope of this thesis, can be found in [Hoc09].

Theorem 12. *Let $d \geq 3$. Then, the entropies of d-D CA are exactly the non-negative numbers that are the $\liminf$ of a recursive sequence of numbers.*

Also, E. Lakshtanov and E. Langvagen have introduced for every $d \geq 1$ another notion of entropy of CA that coincides with topological entropy when $d = 1$, and is always finite when $d \geq 1$. Whether this is a more ”appropriate” notion of entropy for multidimensional CA is still out for the jury to tell. Details can be found in [LL05].

3.4 Dynamical system properties

According to Hedlund's Theorem, CAs are exactly the continuous, translation invariant transformations of $S^{\mathbb{Z}^d}$, where $S^{\mathbb{Z}^d}$ is given the topology generated by the cylinders $Cyl(D,p) = \{e \in S^{\mathbb{Z}^d} \colon e(\vec{x}) = p(\vec{x}),\ \text{for every } \vec{x} \in D\}$, with $D \subseteq \mathbb{Z}^d$ finite and $p\colon D \to S$. Hence, they can be viewed as dynamical systems, allowing us to define and study properties such as equicontinuity or chaos for CAs. The reason to do this is double: on one hand, we get more insight and better classification schemes for CAs, and on the other hand, CAs often provide nice examples or counterexamples of dynamical systems. This is partially due to the fact that almost all of the analytical definitions of general dynamical systems theory have simple equivalent combinatorial definitions for CAs. For example, the usual epsilon-delta definitions of equicontinuity, sensitivity and positive expansiveness translate to the following:

Let $G\colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ be a CA. Configuration $c \in S^{\mathbb{Z}^d}$ is an *equicontinuity point* if and only if for every finite observation window $E \subseteq S^{\mathbb{Z}^d}$ there exists a finite domain $D \subseteq S^{\mathbb{Z}^d}$ such that

$$e \in Cyl(D,c) \Rightarrow G^n(e) \in Cyl(E, G^n(c)),\ \text{for every } n \geq 1.$$

Recall that G is called *equicontinuous* if every configuration is an equicontinuity point.

Similarly, G is *sensitive* if and only if there exists a finite observation window $E \subseteq S^{\mathbb{Z}^d}$ such that for every configuration c and every finite domain D, there exists $e \in Cyl(D,c)$ such that

$$G^n(e) \notin Cyl(E, G^n(c)),\ \text{for some } n \geq 1.$$

Finally, G is *positively expansive* if and only if there exists a finite observation window $E \subseteq S^{\mathbb{Z}^d}$ such that for any distinct configurations $c \neq e$

$$G^n(e)(\vec{x}) \neq G^n(c)(\vec{x}), \text{ for some } n \geq 0 \text{ and } \vec{x} \in E.$$

One can prove that in every dimension it is true that a CA is equicontinuous if and only if it is ultimately periodic, that a surjective CA is equicontinuous if and only if it is periodic and that a sensitive CA does not have any equicontinuity points. In addition, in 1-D there exists the notion of *blocking word* that provides the following classification of 1-D CA:

Theorem 13. *Let G be a 1-D CA. Then, exactly one of the following two alternatives holds:*

1. *The set of equicontinuity points of G is dense. This is equivalent to the existence of a large enough blocking word.*

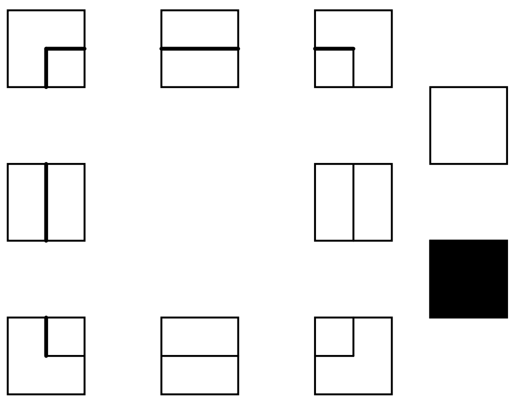
2. *G is sensitive. This happens if and only if G does not have any equicontinuity points.*

A blocking word is a finite segment that does not allow information to pass through it. For a strict definition and a proof of Theorem 13, see [Kur03]. The main idea is that the line is simply connected, so by removing a point, or a finite contiguous segment, we disconnect it and don not allow the flow of information between the two components. Unfortunately, in order to disconnect the plane, we have to remove a line, which is an infinite object and, hence cannot be recognized by CA. Indeed, the following 2-D unpublished counterexamples to Theorem 13 due to Kari exploit this idea. Other counterexamples have been given by Theyssier and Sablik, see [ST08].

Theorem 14. 1. *There exists a 2-D CA with no equicontinuity points that is not sensitive.*

2. *There exists a 2-D CA with a non-empty, non-dense set of equicontinuity points.*

Proof. 1. Let G be a CA over the state set $C \times \{0, 1\}$, where C is the state of the *control layer*, containing the following ten tiles:



The black tile is called *active*, while the other nine control states are called *inactive*. A *blank rectangle* is a finite rectangle of blanks surrounded by a contiguous boundary.

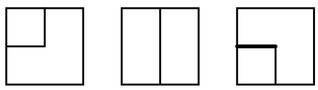
It is obvious that there exists blank rectangles of size $m \times n$, for every $m, n \geq 2$.

The control layer evolves independently from the bit layer according to the following rule: Let $\vec{p} \in \mathbb{Z}^2$. If the Moore neighborhood of $\vec{p}$ is consistent with it being part of a blank rectangle surrounded by black and white cells (no boundary tiles allowed), then the state of the control layer does not change. Otherwise, it changes to black.

Note that if any of the cells outside from the rectangle contains a boundary tile, then a black tile is created after one time-step and, eventually, the whole rectangle gets converted to black. Also, by construction, a blank rectangle surrounded by black and white cells never gets changed.

The essential property of this local rule is that if a position $\vec{p}$ contains a black, then in the next time step either its northern neighbor or its western neighbor also contains a black. This is not immediately obvious, so some kind of proof is needed:

Suppose that the western neighbor of $\vec{p}$ does not become a black. This means that it is part of a blank rectangle surrounded by black and white tiles. So, it must on the right border of this rectangle and, hence in $\vec{p}$ there is one of the three following tiles:



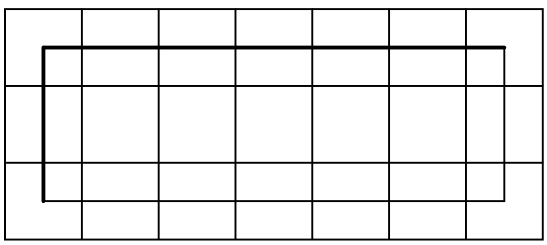
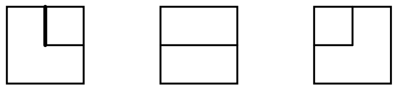


Figure 18: A blank rectangle of size 7×3.

Similarly, if the northern neighbor of $\vec{p}$ does not get turned into a black, then it must lie on the lower border of a blank rectangle, hence it has to be one of the following tiles:



If neither of them gets turned into a black, then a tile of the first group will be a neighbor with a tile from the second group, and since they cannot be a part of a blank rectangle, they both change to black. This contradiction proves our claim.

We are now ready to describe how the bits are updated: If a cell is inactive, then its bit does not change. If a cell is active, then:

- If its western neighbor is active, then it copies the bit of this neighbor, else
- if its northern neighbor is active, then it copies the bit of this neighbor, else
- the bit remains unchanged.

G is not sensitive: Indeed, given any finite observation window E, we can construct a blank rectangle R surrounded by black and white tiles such that $R \supseteq E$. Consider, then, a configuration $c \in (C \times \{0,1\})^{\mathbb{Z}^2}$ with R on its control layer and arbitrary bits. Then, changes made outside from R never propagate in R, a fortiori neither in E. This means that for every finite observation window E there exists a finite domain $D(= R)$ and a configuration c such that if $e \in Cyl(c, D)$, then $G^n(e) \in Cyl(G^n(c), E)$. Therefore, G is not sensitive.

However, G does not have any equicontinuity points, neither. Indeed, let $c \in S^{\mathbb{Z}^2}$ be an arbitrary configuration. Either the control layer of c is completely white, in which case c is not an equicontinuity point since every black tile will eventually spread everywhere in the plane, or $G(c)$ has black in some position $\vec{n}$ of the control layer. In this case, let $E = \{\vec{n}\}$. We will show that no matter how large a square domain $D \subseteq S^{\mathbb{Z}^2}$ is chosen, changes made to c outside from D will eventually propagate into E.

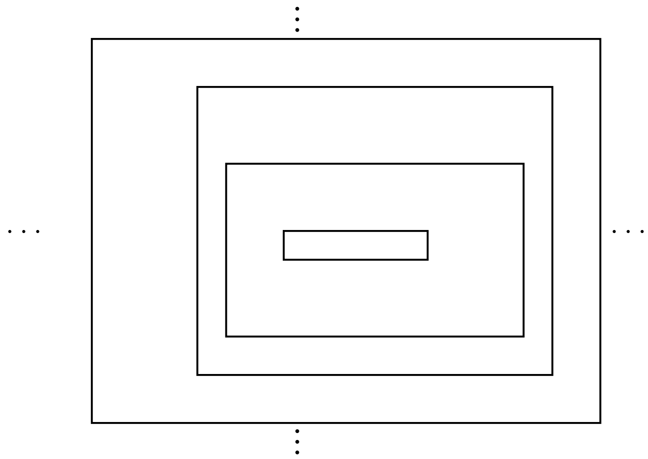


Figure 19: An equicontinuity point for F.

For any square domain $D \subseteq S^{\mathbb{Z}^2}$, consider the two configurations c_0 and c_1 that agree with c inside D and have black control layer outside D. In c_0 all bits outside D are equal to 0, while in c_1 they are equal to 1. According to the fundamental property of G proven before, in the control layers (which are always the same for c_0 and c_1) there will eventually appear a path of black cells connected $\vec{n}$ to the upper or left halfplane outside D. Then, in c_0 the bit 0 will get transmitted to position $\vec{n}$ and in c_1 the 1 will get transmitted to position $\vec{n}$. Hence, for some $t \geq 1$ either $G^t(c_0)(\vec{n})$ or $G^t(c_1)(\vec{n})$ is different than $G^t(c)(\vec{n})$. This means that c is not an equicontinuity point.

2. In order to construct a 2-D CA F that has a non-empty, non-dense set of equicontinuity points, we have to change a little bit the previous construction. The only difference is that F turns a white cell to black if and only if it has a black neighbor. F has equicontinuity points: any configuration of infinitely many nested blank rectangles is an equicontinuity point. Indeed, let $e \in S^{\mathbb{Z}^2}$ be such a configuration and let $E \subseteq S^{\mathbb{Z}^2}$ be a finite observation window. We can find a blank rectangle R such that $R \supseteq E$. Then, changes made in e outside of R will never propagate into E, since the boundary of R prevents them from penetrating.

   However, if a configuration has a black cell, then it is not an equicontinuity point. This can be proved in the same way as for G.

$\square$

There also exists a dimension-sensitive property concerning positive expansiveness. The following theorem was first proved by Shereshevsky in a much more general setting.

Theorem 15. *[She93] For $d \geq 2$, no d-D CA is positively expansive.*

Proof. Suppose that G is a d-dimensional positively expansive CA and let E be a finite observation window confirming positive expansiveness. Since, obviously, every set containing E is also a witness of positive expansiveness, we can suppose that E is a $(2k+1)^d$ hypercube centered around the origin. We denote by D_i the hypercube of size $(2i+1)^d$ centered around the origin.

Now, let us show that there exists some $n \geq 1$ such that for any configurations $c, e \in S^{\mathbb{Z}^d}$ if $G^t(c)(\vec{x}) = G^t(e)(\vec{x})$ for every $0 \leq t \leq n$ and $\vec{x} \in E = D_k$, then $c(\vec{y}) = e(\vec{y})$, for every $\vec{y} \in D_{k+1}$.

Suppose the contrary: For every $n \geq 1$ there exist configurations $e_n, c_n \in S^{\mathbb{Z}^d}$ such that $G^t(c_n)(\vec{x}) = G^t(e_n)(\vec{x})$ for every $0 \leq t \leq n$ and $\vec{x} \in D_k$, but $c(\vec{y}) \neq e(\vec{y})$ for some $\vec{y} \in D_{k+1}$. Let c, e be the limit of converging subsequences of these sequences. Since D_{k+1} is a finite set, $c(\vec{y}) \neq e(\vec{y})$ for some $\vec{y} \in D_{k+1}$. However, $G^t(c)(\vec{x}) = G^t(e)(\vec{x})$ for all $\vec{x} \in D_k = E$ and all $t \geq 0$, which contradicts positive expansiveness. Therefore, our initial claim is true.

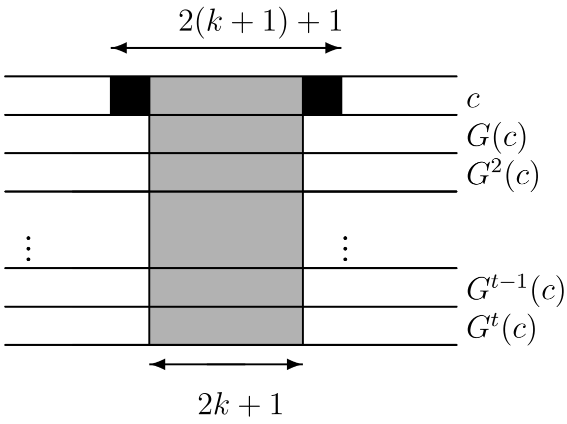


Figure 20: The states $G^t(c)(\vec{x})$ for $\vec{x} \in E$
and $0 \leq t \leq n$ determine $c(\vec{y})$ for $\vec{y} \in D_{k+1.}$

This means that the states $G^t(c)(\vec{x})$ for $\vec{x} \in E$ and $0 \leq t \leq n$ uniquely determine $c(\vec{y})$ for $y \in \vec{D}_{k+1}$. Inductively, we can easily prove that for any $j \geq 0$, the states $G^t(c)(\vec{x})$ for $\vec{x} \in D_{k+j}$ and $0 \leq t \leq n$ uniquely determine $c(\vec{y})$ for $\vec{y} \in D_{k+j+1}$.

In addition, for every $j \geq 1$ the states $G^t(c)(\vec{x})$ for $\vec{x} \in E$ and $0 \leq t \leq jn$ uniquely determine $c(\vec{y})$ for $\vec{y} \in D_{k+j}$. This can be seen in figure 21:

A contradiction now arises from the fact that for $d \geq 2$ the volume of D_{k+j} grows faster than the size of $D_k \times \{0, 1, 2, \ldots, jn\}$ as j increases. Indeed, for sufficiently large j

$$\mid D_{k+j} \mid = (2k + 2j + 1)^d > (2k + 1)^d (jn + 1) = \mid D_k \times \{0, 1, 2, \ldots, jn\} \mid .$$

This means that for some sufficiently large j, there exist two configurations c, e that are not identical in D_{k+j} although $G^t(c)(\vec{x}) = G^t(e)(\vec{x})$ for every $0 \leq t \leq jn$ and $\vec{x} \in D_k$. This is a contradiction, hence our initial assumption that there exists a 2-D positively expansive CA must be wrong.

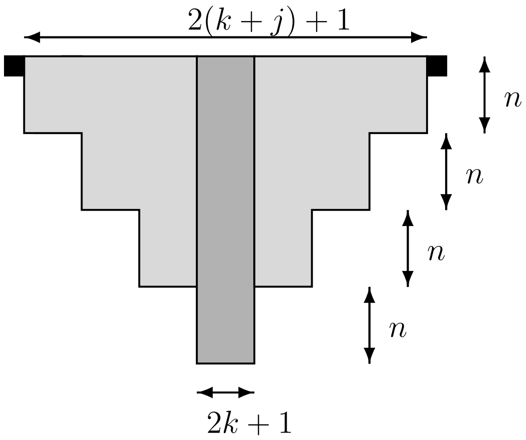


Figure 21: The states $G^t(c)(\vec{x})$ for $\vec{x} \in E$ and $0 \leq t \leq jn$ uniquely determine $c(\vec{y})$ for $\vec{y} \in D_{k+j}$.

$\square$

Finally, as a last testimony of the usefulness of Wang tiles in CA problems, we will use aperiodic tile sets to give examples of 2-D CAs with behavior that is not possible in the 1-D case. For example, it is known that if G is a 1-D CA and G_P is injective, then G is injective, too. The following example shows that this is not always true in the 2-D case:

Consider the CA G constructed in the proof of Theorem 11. G is not injective: Indeed, if t is any valid tiling with Kari's tile set, consider the configurations c_0 and c_1 defined in the usual way. Then, $G(c_1) = G(c_0)$.

However, G is injective on periodic configurations. Suppose that it is not. Then, there exist two periodic configurations c, e such that $G(c) = G(e)$. Configurations c and e have the same tiling components. Consider a position $\vec{p_1}$ where their bits are different. Using the standard argument concerning Kari's tile set, we conclude that their exists an infinite path $\vec{p_1}, \vec{p_2}, \vec{p_3}, \ldots$ where the tiling is valid. According to the plane-filling property, this path must cover arbitrarily large squares. However, c and e are periodic so every path they define is also periodic and it is impossible for a periodic path to cover arbitrarily large squares. This contradiction proves our claim.

Using any aperiodic tile set, we can also construct examples of 2-D CAs with fixed points but without any periodic fixed points or with a periodic point that has a pre-image but does not have any periodic pre-image. Both of these are impossible in 1-D .

4 Subshifts of Finite Type

4.1 Preliminaries

In this chapter we will make a more sytematic study of the d-D *full shift* and its *subshifts*. The book [LM95] is, and will probably be for quite a long time to come, the standard reference for the 1-D case. The following definitions are an easy generalization to d-D of the corresponding definitions for the 1-D case.

If S is a finite *alphabet* of n *letters*, then the space $(S^{\mathbb{Z}^d}, \mathcal{T})$ is a compact topological space, where $\mathcal{T}$ is the topology defined in Section 2.1. We can define a $\mathbb{Z}^d$-action σ on $S^{\mathbb{Z}^d}$ as follows: To every $\vec{v} \in \mathbb{Z}^d$ we assign the *shift-map* $\sigma_{\vec{v}} \colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ defined as

$$\sigma_{\vec{v}}(c)(\vec{x}) = c(\vec{x} + \vec{v}), \text{ for every } c \in S^{\mathbb{Z}^d} \text{ and } \vec{x} \in \mathbb{Z}^d.$$

As noted in the previous chapter, every shift is a continuous function, hence the pair $((S^{\mathbb{Z}^d}, \mathcal{T}), \sigma)$ is a dynamical system. We call it the d-D *full shift on n letters*. A *subshift* X is a subsystem of the full shift, aka a closed, shift-invariant subset of $S^{\mathbb{Z}^d}$. The restriction σ_X of σ defines a new dynamical system with base set X. When there is no danger of misunderstanding, we drop the subscript and denote σ_X as σ, too.

Luckily enough, there exists a simple characterization for the subshifts of $S^{\mathbb{Z}^d}$. Recall that a pattern is a pair (D, p) where D is a finite subset of $\mathbb{Z}^d$ and $p \colon D \to S$ assigns values to the cells of D. We say that (D, p) *appears* in $c \in S^{\mathbb{Z}^d}$ at position $\vec{x}$ if $c_{\vec{x}+D} = p$. Let $F = \{(D_1, p_1), (D_2, p_2), (D_3, p_3), \ldots\}$ be a, finite or infinite, set of *forbidden* patterns. The subshift S_F defined by F is the set of all configurations where no pattern (D_i, p_i) appears at no position of the plane, namely

$$S_F = \{c \in S^{\mathbb{Z}^d} : c_{\vec{x}+D_i} \neq p_i, \text{ for every } \vec{x} \in \mathbb{Z}^d \text{ and every } i = 1, 2, 3, \ldots\}.$$

It can be easily proven that for every set of patterns F, S_F is indeed a closed, shift-invariant subset of $S^{\mathbb{Z}^d}$, therefore a subshift. What is more interesting is that *every* subshift can be defined as a S_F for some suitable set of F of forbidden patterns. Notice that, as a rule, there are a lot of sets of forbidden patterns that define the same subshift.

If a subshift X can be defined using a finite set of forbidden patterns, then it is called a *subshift of finite type* (SFT from now on). SFTs will be our main object of study in the subsequent sections, so some observations are in order:

- By enlarging the domain of the forbidden patterns, we can suppose that all of them have the same domain and that this domain is the d-D hypercube $[-m, m]^d$, for some $m \geq 1$. In this case, we can check if a configuration belongs to X by moving a window of size $[-m, m]^d$ around every position of the plane and checking if the pattern we see in the window is forbidden. Equivalently, we can define X with its set of *allowed* patterns of size $[-m, m]^d$.

- As noted in the first chapter, the set of valid tilings admitted by a Wang tile set is a SFT. Also, every SFT is conjugate, in the sense that will be defined shortly, to a d-D Wang tile set.

We denote by $R_N(X)$ the set of *allowable* patterns of X with domain a hypercube of size $2N + 1$ centered in the origin:

$$R_N(X) = \{(D, p) : D = [-N, N]^d \text{ and no forbidden pattern appears in } p\}$$

Let $X \subseteq S^{\mathbb{Z}^d}$ be an arbitrary subshift (not necessarily of finite type) and let $f\colon R_N(X) \to T$ be a function mapping blocks of size N appearing in X to letters of T, where T is an arbitrary finite alphabet. Then, f defines a *block code* $F\colon X \to T^{\mathbb{Z}^d}$ in exactly the same way as CA are defined. Namely,

$$F(c)(\vec{x}) = f(c(\vec{x} + [-N, N]^d)), \text{ for every } c \in X \text{ and } \vec{x} \in \mathbb{Z}^d.$$

Exactly as in the case of CA, we can prove that block maps are continuous and commute with the corresponding shift actions., that is $\sigma \circ F = F \circ \sigma_X$. An analog of Proposition 6 is also true: Every shift-commuting, continuous transformation $F\colon X \to T^{\mathbb{Z}^d}$ is a block map.

If a block map $F\colon X \to Y$ is surjective, then it is called a *factor map*. If F is bijective and its inverse is also a block map, then it is a *conjugacy*.

A subshift Y is called a *sofic shift* if it is a factor of a SFT, that is if there exists a SFT X and a surjective block map $F\colon X \to Y$. We say that X is a *cover* of Y.

Topological entropy of subshifts has the following nice combinatorial description: If X is a subshift and $B_N(X)$ is the set of allowable patterns with domain $[0, N-1]^d$, then

$$h(X) = \lim_{N \to \infty} \frac{\log(|B_N(X)|)}{N^d}.$$

For the full shift we have that $|B_N(S^{\mathbb{Z}^d})| = |S|^{N^d}$, since no pattern is forbidden, hence $h(S^{\mathbb{Z}^d}) = \log(|S|)$.

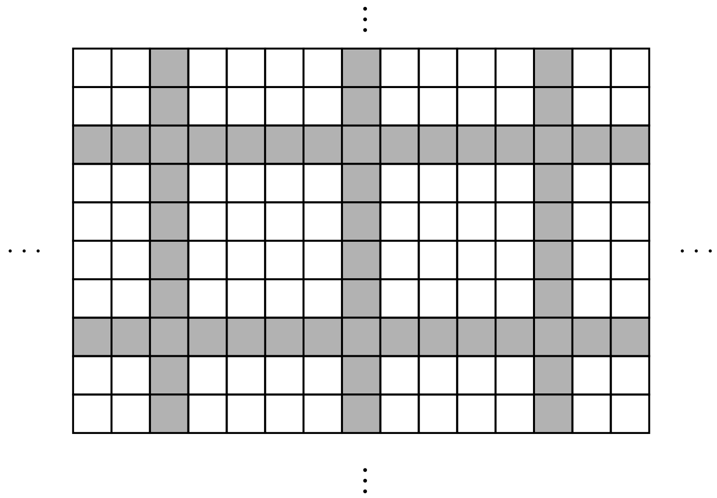
Finally, let us state that for reasons of simplicity, all of the proofs in this chapter will be given for the 2-D case. However, they can be easily generalized in the higher-dimensional cases.

4.2 SFTs and their subshifts

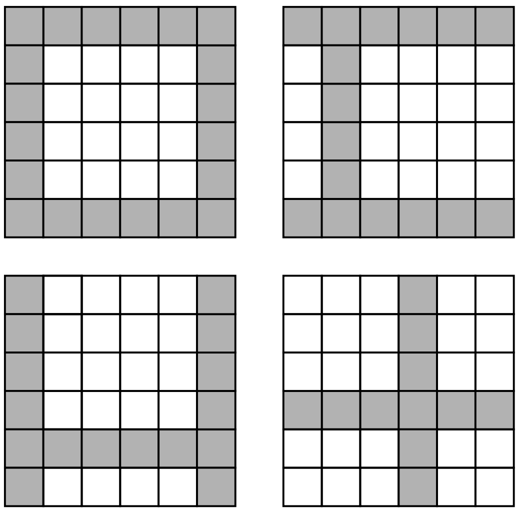
In this section we are going to talk about SFTs and the entropies of their subsystems. First of all, we prove that in every dimension the entropies of the SFT subshifts of a SFT X with positive topological entropy $h(X) > 0$ are dense in $[0, h(X)]$. We will then turn our attention to sofic shifts and ask the same question about them: Are the entropies of the SFT subsystems of a sofic shift S with positive topological entropy $h(S) > 0$ dense in $[0, h(S)]$? We will prove a partial result, namely that the entropies of the sofic subsystems of S are dense. For $d \geq 2$ this is indeed the best we can hope for, since there exist d-D sofic shifts with arbitrarily large topological entropy whose only SFT subsystem is a fixed point. Most of the material of this section is based on Desai's paper [Des05].

Let X be a 2-D SFT. As noted before, we can assume that X is a one-step SFT, in other words the set of valid tilings of a Wang tile set. For every $N \geq 2$, we construct another equal entropy SFT $Y_{(N)}$ that factors onto X as follows:

The alphabet of $Y_{(N)}$ is the disjoint union of two copies of the alphabet of X. The tiles from the first copy are colored white while the tiles from the second copy are colored gray. Let $\pi\colon A_{Y_{(N)}} \to A_X$ be the projection that forgets the colors. A configuration $c \in A_{Y_{(N)}}^{\mathbb{Z}^2}$ belongs to $Y_{(N)}$ if there is a $\vec{n} \in \mathbb{Z}^2$ such that $c(\vec{z})$ is colored gray if and only if $\vec{z} \in \vec{n} + ((N\mathbb{Z} \times \mathbb{Z}) \cup (\mathbb{Z} \times N\mathbb{Z}))$, and if by removing the colors we obtain a point in X. The elements of $Y_{(N)}$ are thus points of X "sandwiched" with a grid-like pattern looking like this:



We can examine whether a configuration $c \in A_{Y_{(N)}}^{\mathbb{Z}^2}$ is in $Y_{(N)}$ using a $(N+1) \times (N+1)$ observation window. In every point, in the first layer we must see a pattern from $B_{N+1}(X)$ and in the second layer a pattern of one of the following forms:



Hence, $Y_{(N)}$ can be defined with its set of allowable $(N+1)$-blocks so it is a SFT. Let $\Pi_{(N)} \colon Y_{(N)} \to X$ be the map induced by π. $\Pi_{(N)}$ is obviously onto and N^2-to-1 since there are N^2 ways to place the gridlike background on an element of X.

The following three Lemmas are valid in every dimension.

Lemma 7. *[MS01] Let $X, Y \subseteq S^{\mathbb{Z}^d}$ be subshifts. If $\Pi \colon X \to Y$ is finite-to-one, then $h(\Pi(X')) = h(X')$, for every subsystem $X' \subseteq X$.*

Lemma 8. *If X is a SFT and $X_1 \supseteq X_2 \supseteq X_3 \supseteq \ldots$ is a decreasing sequence of SFTs such that $\bigcap_{i=1}^{\infty} X_i \subseteq X$, then there exist a $k \geq 1$ such that $X_n \subseteq X$ for every $n \geq k$.*

Proof. Let $p_1, p_2, \ldots, p_r$ be the forbidden patterns of X. Since $\bigcap_{i=1}^{\infty} X_i \subseteq X$, for every $1 \leq m \leq r$ there exists a k_m such that p_m is forbidden in X_{k_m}. Choose $k = \max\limits_{1 \leq m \leq r} k_m$. Then, all of the patterns $p_1, p_2, \ldots, p_r$ are forbidden in X_k, hence $X_k \subseteq X$. Since the sequence of SFTs is decreasing, $X_n \subseteq X$ for every $n \geq k$. $\square$

Lemma 9. *Let X be a SFT and $X' \subseteq X$ be a subshift of X. Then, there exists a decreasing sequence of SFTs $X_1 \supseteq X_2 \supseteq X_3 \supseteq \ldots$ such that $X' = \bigcap_{i=1}^{\infty} X_i$.*

Proof. We know that X' can be defined with a, possibly infinite, set of forbidden patterns. Let $F = \{p_1, p_2, \ldots\}$ be this set and let X_k be the SFT with forbidden patterns $F_k = \{p_1, p_2, \ldots, p_k\}$. Obviously, $X' = \bigcap_{i=1}^{\infty} X_k$. Notice that X' does not need to be a SFT. $\square$

Combining the two previous Lemmas, we have the following Proposition:

Proposition 10. *Let X be a SFT and $X' \subseteq X$ be a subsystem of X. Then, there exists a SFT subsystem of X with topological entropy arbitrarily close to the topological entropy of X'.*

Proof. From Lemma 9, we know that there exists a decreasing sequence of SFTs $X_1 \supseteq X_2 \supseteq X_3 \supseteq \ldots$ such that $\bigcap_{i=1}^{\infty} X_i = X'$. Since $X' \subseteq X$, by Lemma 8 we know that there exists some $k \geq 1$ such that $X_n \subseteq X$ for every $n \geq k$. Also, since the entropy function is upper semi-continuous on subshifts, we have that $\lim\limits_{i \to \infty} h(X_i) = h(X')$. Hence there exists some SFT subsystem of X with topological entropy arbitrarily close to $h(X')$. $\square$

Proposition 10 implies that if X is a SFT with $h(X) > 0$ and if there exists a family of arbitrary subshifts of X whose topological entropies are dense in $[0, h(X)]$, then there also exists a family of SFT subshifts of X with the same property. In the following, we will prove that there exists a family of sofic shifts whose entropies are dense in $[0, h(X)]$ and, thus, we will have proved the following Proposition:

Proposition 11. *Let X be a SFT. Then, there exists a family of SFT subshifts of X whose entropies are dense in* $[0, h(X)]$.

Proof. As noted, it suffices to prove the existence of a family of sofic subshifts of X whose entropies are dense in $[0, h(X)]$. We set $\alpha = |A_X|$. In order to prove the existence of such a family of sofic subsystems, let us go back to the SFTs $Y_{(N)}$ constructed earlier.

Let us choose $l \geq 1$ such that $\log l - \log(l-1) < \varepsilon$. Since $h(X) > 0$, for all sufficiently large N there is at least one border of a block in $B_{N+1}(X)$ whose interior can be completed to an allowed $(N+1)$-block of X in more than l ways. Let us choose N large enough so that the following inequalities are also satisfied:

$$\frac{16N}{(N+1)^2}\log\alpha < \frac{\varepsilon}{3}$$
$$\frac{4}{(N+1)^2}\log l < \frac{\varepsilon}{3}$$
$$\frac{2}{(N+1)^2}\log N < \frac{\varepsilon}{3}$$

For this N construct $Y_{(N)}$ and let $\Pi_{(N)} \colon Y_{(N)} \to X$ be as in the construction above. Let us denote by $B^0_{kN+1}(Y_{(N)})$ the set of $(kN+1)$-blocks allowed in $Y_{(N)}$ where border symbols of the square are colored gray. Then,

$$B^0_{kN+1}(Y_{(N)}) = B_{kN+1}(X) \text{ and } B_{kN+1}(Y_{(N)}) = N^2 B^0_{kN+1}(Y_{(N)}),$$

since there are N^2 for the position of the grid on a configuration of $Y_{(N)}$.

We will now construct a decreasing sequence of SFT subsystems of $Y_{(N)}$ such that their entropies are ϵ-dense in $[0, h(Y_{(N)})]$. Let $Y_0 = Y_{(N)}$ and define Y_i inductively from Y_{i-1} by disallowing one block of $B^0_{N+1}(Y_{i-1})$ whose gray border has more than l allowable interiors. According to the choice of N the first step of the induction is valid. This procedure will eventually terminate, since there are only a finite number of blocks in $B^0_{N+1}(Y_{(N)})$, giving us the last element Y_M of the sequence. This means that for every $B \in B^0_{N+1}(Y_M)$, the outer boundary ∂B of B can be completed to a $(N+1)$-valid block in Y_M in at most l ways. Obviously, $Y_0 \supseteq Y_1 \supseteq \ldots \supseteq Y_M$.

We prove first that $h(Y_{i-1}) - h(Y_i) < \epsilon$ for every $i = 1, 2, \ldots, m$:

Indeed, every gray border ∂B of some $B \in B^0_{N+1}(Y_{i-1})$ is also a border of some block in $B^0_{N+1}(Y_i)$. Thus, the number of allowable interiors of ∂B in Y_i is at least $\frac{l-1}{l}\times$ the number of allowable interiors of ∂B in Y_{i-1}.

Given a positive integer $k \geq 1$ fix the gray symbols in an element of $B^0_{kN+1}(Y_i)$. Considering in how many ways this can be completed into an allowable block of Y_i we have that:

$$\begin{array}{rl} & |\text{ways to complete into a block of } Y_i| \geq \\ \geq & |(\frac{l-1}{l})^{k^2}\text{ways to complete into a block of } Y_{i-1}|. \end{array}$$

This means that $|B^0_{kN+1}(Y_i)| \geq (\frac{l-1}{l})^{k^2}|B^0_{kN+1}(Y_{i-1})|$ hence that

$$\begin{aligned} |B_{kN+1}(Y_i)| &\geq |B^0_{kN+1}(Y_i)| \\ &\geq (\frac{l-1}{l})^{k^2}|B^0_{kN+1}(Y_{i-1})| \\ &\geq (\frac{l-1}{l})^{k^2}|B_{(k-1)N+1}(Y_{i-1})|, \end{aligned}$$

and therefore

$$\begin{aligned} h(Y_i) &= \lim_{k\to\infty}(\frac{1}{kN+1})^2\log|B_{kN+1}(Y_i)| \\ &\geq \lim_{k\to\infty}[(\frac{1}{kN+1})^2\log(\frac{l-1}{l})^{k^2} + \frac{1}{(kN+1)^2}\log|B_{(k-1)N+1}(Y_{i-1})|] \\ &= \frac{1}{N^2}\log\frac{l-1}{l} + h(Y_{i-1}) \\ &> h(Y_{i-1}) - \varepsilon. \end{aligned}$$

This proves our claim.

Next, we go on to prove that $h(Y_M) < \epsilon$:

Indeed, there are at most α^{4N} possible boundaries for configurations in $B^0_{N+1}(Y_M)$, since the perimeter of the $(N+1)\times(N+1)$ square is $4N$. For every one of these possible perimeters, (some of them might not even be allowed in Y_M) there are at most l ways to complete them into a valid pattern in $B^0_{N+1}(Y_M)$. Hence $|B^0_{N+1}(Y_M)| \leq \alpha^{4N}l$. As every $B \in B_{N+1}(Y_M)$ touches at most four blocks of $B^0_{N+1}(Y_M)$, we have that

$$|B_{N+1}(Y_M)| \leq \alpha^{16N} l^4 N^2.$$

Hence,

$$\begin{aligned} h(Y_M) &\leq \frac{1}{(N+1)^2}\log|B_{N+1}(Y_M)| \\ &\leq \frac{1}{(N+1)^2}\log\alpha^{16N} + \frac{1}{(N+1)^2}\log l^4 + \frac{1}{(N+1)^2}\log N^2 \\ &< \varepsilon. \end{aligned}$$

Therefore, the topological entropies $h(Y_0), h(Y_1), \ldots, h(Y_M)$ are ϵ-dense in $[0, h(Y_{(N)})]$.

Consider, now, the sofic subshifts $X_0, X_1, \ldots, X_M$ of X defined as $X_i = \Pi_{(N)}(Y_i)$ for every $0 \leq i \leq M$. $\Pi_{(N)}$ is finite-to-one, hence Lemma 7 says that $h(X_i) = h(Y_i)$ for every $0 \leq i \leq M$, and since $h(X) = h(Y_{(N)})$, the topological entropies of $X_0, X_1, \ldots, X_M$ are ϵ-dense in $[0, h(X)]$. This ends the proof of Proposition 11. □

Next, let us turn our attention to sofic shifts. By definition, if T is a sofic shift, there exists a SFT S and a shift-commuting surjection $\pi\colon S \to T$. S is called a *cover* of T. Naturally, $h(S) \geq h(T)$. The question is: How close to $h(T)$ can $h(S)$ be? In a mathematically more rigorous way, we can ask if every sofic shift has an equal entropy cover. In the 1-D case, this is true, see [CP75]. In the d-D case the best result that is known at the time of writing is:

Theorem 16. *Let T be a d-D sofic shift and $\varepsilon > 0$. Then, T has a SFT cover S with factor map $\pi\colon S \to T$ such that $h(S) < h(T)+\varepsilon$. Moreover, $h(S') < h(\pi(S'))+\varepsilon$ for every subshift $S' \subseteq S^{\mathbb{Z}^d}$ of S.*

Proof. Since T is sofic, there exists a SFT R and a shift-commuting surjection $\phi\colon R \to T$. We can suppose that R and T have disjoint alphabets, that R is a one-step SFT and that ϕ is a 1-block code. For every $N \geq 1$ we construct the following SFT $S_{(N)}$:

The alphabet of $S_{(N)}$ is the union of the alphabets of T and R. For every allowable $(N+1)$-block of R, replace the interior letters by their images through ϕ and leave the boundary letters as they are. Let $B^0_{N+1}(S_{(N)})$ denote the set of all such blocks. A configuration $c \in A^{\mathbb{Z}^2}_{S_{(N)}}$ belongs to $S_{(N)}$ if there exists $\vec{n} \in \mathbb{Z}^2$

such that $c(\vec{z})$ is a letter from A_R if and only if $\vec{z} \in \vec{n} + (N\mathbb{Z} \times \mathbb{Z}) \cup (N\mathbb{Z} \times \mathbb{Z})$ and, in addition, the pattern of every $(N+1) \times (N+1)$ grid square belongs to $B^0_{N+1}(S_{(N)})$. Therefore, in every $c \in S_{(N)}$ there exists an orthogonal grid of size $(N+1) \times (N+1)$ with letters from A_R and the rest of the letters belong to A_T.

Clearly, $S_{(N)}$ is a SFT. Define $\pi\colon S_{(N)} \to T$ to be the 1-block map that sends letters of A_R to their images under ϕ and leaves the letters of A_T unchanged. We claim that the image of π is contained in T and that π is surjective:

For any block $B \in B^0_{N+1}(S_{(N)})$ there exists a block $B' \in B_{N+1}(R)$ with the same boundary such that when we change the interior symbols of B' with their images under ϕ we take B. Make such a choice for every $B \in B^0_{N+1}(S_{(N)})$. Let $c \in A_{(N)}^{\mathbb{Z}^d}$ be an element of $S_{(N)}$. Replace every $(N+1)$-block of the gridlike structure with the corresponding block B'. Since R is a one-step SFT, this defines an element $\psi(c)$ of R and $\pi(c) = \phi(\psi(c))$. Hence, the image of π is contained in T.

Next, let $z \in T$. Since ϕ is surjective, there exists some $y \in R$ with $\phi(y) = z$. Given this y, leave the symbols in $(N\mathbb{Z} \times \mathbb{Z}) \cup (N\mathbb{Z} \times \mathbb{Z})$ unchanged and replace the rest of them with their images through ϕ. It can be immediately checked that this defines a point c in $S_{(N)}$ and that $\pi(c) = \phi(y) = z$. Hence, π is also onto and, therefore, a factor map.

Finally, we show that $h(S_{(N)}) < h(T) + \varepsilon$ for large N:

Let $\alpha = |A_R|$. There are at most α^{4N} possible borders for blocks in $B^0_{N+1}(S_{(N)})$ and each of them has at most $|B_N(T)|$ interiors. Therefore, $|B^0_{N+1}(S_{(N)})| \leq \alpha^{4N}|B_N(T)|$ and since there are N^2 possible locations for the grid,

$$|B_{N+1}(S_{(N)})| \leq \alpha^{4N}|B_N(T)|N^2.$$

Thus, we have the following inequalities:

$$\begin{aligned} h(S_{(N)}) &\leq \frac{1}{(N+1)^2}\log|B_{N+1}(S_{(N)})| \\ &\leq \frac{4N}{(N+1)^2}\log\alpha + \frac{1}{(N+1)^2}|B_N(T)| + \log\frac{N^2}{(N+1)^2} \\ &< h(T) + \varepsilon, \text{ for large values of N.} \end{aligned}$$

The second claim of the Theorem can be proved similarly. □

We note again that it is still not known whether every multi-dimensional sofic shift has a SFT cover of equal entropy.

Theorem 17. *Let T be a d-D sofic shift with $h(T) > 0$. Then, there exists a family of sofic subshifts of T with dense topological entropies in $[0, h(T)]$. However, for every $d \geq 2$, there are d-D sofic shifts with arbitrarily large topological entropy whose only SFT subshift is a fixed point.*

Proof. For the proof of the first claim, let T be a d-D subshift and $\varepsilon > 0$. Construct the SFT cover S of T as in Theorem 16. By Proposition 11 we know that there exists a family $(S_i)_{i\in\mathbb{N}}$ of SFT subshifts of S with topological entropies dense in $[0, h(S)]$. The images of these subshifts under π are sofic subshifts of T and their entropies are ε-close to the entropies of the members of $(S_i)_{i\in\mathbb{N}}$. As ε can be arbitrarily small, we conclude our claim.

For the second claim of the Theorem, see [BPS10]. In that paper, there is explicitly constructed an example of a d-D sofic shift with entropy M, where M can be arbitrarily large, whose only SFT subsystem is a fixed point. □

We state again that the 1-D case is drastically different. Namely, if T is a 1-D sofic shift with positive topological entropy $h(T)$, then there always exists a family of SFT subsystems of T with entropies dense in $[0, h(T)]$.

4.3 Factoring onto the full shift

In the theory of dynamical systems, a frequently asked question is whether a dynamical system factors onto another and if there is some set of necessary and sufficient conditions for the existence of a factor map between classes of dynamical systems. In this chapter, we will deal with this question in the following context: When does a d-D SFT factor onto the d-D full shift on n letters $X_{d,n}$?

Obviously, a necessary condition is that $h(X) \geq h(X_{d,n}) = \log n$. Is it also sufficient? For 1-D SFTs, the answer is yes, see [LM95]. For d-D SFTs, it is not. This has been proven separately for the cases $h(X) = \log n$ in [BS09] and $h(X) > \log n$ in [BPS10]. We will describe neither of these counterexamples here, since the first one is based on results of ergodic theory and the second one is prohibitively long. However, as it usually happens in the theory of d-D SFTs, we will prove that with the additional assumption of a ”strong”-mixing property, it is

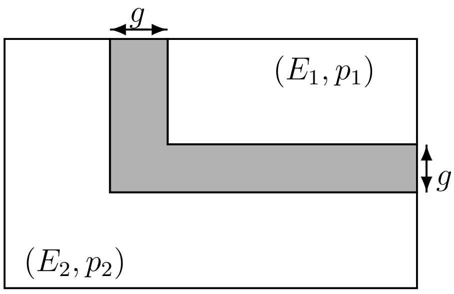


Figure 22: Gluing allowable patterns in a corner-gluing SFT.

true that X factors on $X_{d,n}$ if and only if $h(X) \geq \log n$. Most of the material of this section is based on Desai's article [Des09]

Let $\vec{k}$ be a non-negative vector. We denote by $R_{\vec{k}}$ the rectangle $\{(x, y) : 0 \leq x < k_1$ and $0 \leq y < k_2\}$.

Let us first introduce our notion of "strong"-mixing: If X is a d-D SFT, then X is called *corner-gluing*[JM05], if there exists a $g > 0$ such that given any allowable patterns p_1 and p_2 on domains $E_1 = R_{\vec{k}}+\vec{k'}-\vec{k}$ and $E_2 = R_{\vec{k'}}\backslash(R_{\vec{k}+g\vec{c}}+(\vec{k'}-\vec{k}-g\vec{c}))$, where $\vec{k}, \vec{k'} \in \mathbb{N}^d$, $\vec{c} = (1, 1, \dots, 1)$ and $\vec{k'} > \vec{k} + g\vec{c}$, there exists a configuration $x \in X$ with $x|_{E_1} = p_1$ and $x|_{E_2} = p_2$. g is called the *gluing constant*. Look at figure 22 for a graphical explanation:

The grey-colored area is called the *gluing band*. The condition of the corner-gluing property can also be expressed by saying that (E_1, p_1) and (E_2, p_2) are g*-separated*. We also say that p_1 and p_2 are *glued together* or that p_1 is glued to p_2 when there is no ambiguity about the domains of p_1 and p_2. From now on, we will often refer to patterns using only their domains, when there is no danger of confusion. Also, once the formal definition has been written down explicitly once, we will never use it again. Instead, we will rather talk with and about pictures, like in the following lemma:

Lemma 10. *Let $X \subseteq A^{\mathbb{Z}^d}$ be a corner-gluing SFT and R_1,R_2 be rectangular patterns. If $d_\infty(R_1, R_2) \geq g$, then R_1 and R_2 can be glued together.*

Proof. First of all, extend the domain of R_1 to a corner configuration C_1 and extend R_2 to C_2 so as to align it with C_1. Then, C_1 and C_2 are also g-separated, so they can be glued together. Hence, R_1 and R_2 can also be glued together.

□

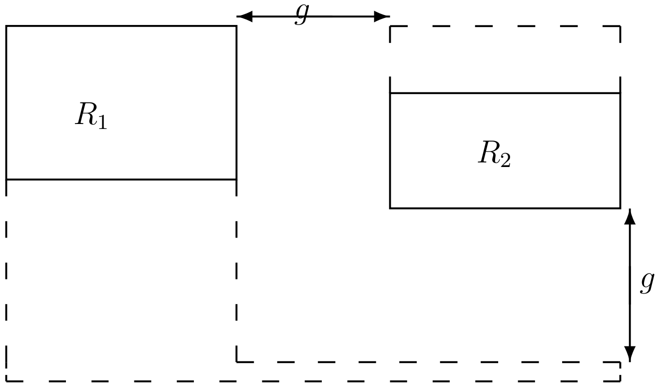


Figure 23: Gluing together g-separated rectangles.

Remark 3. *What Lemma 10 actually says is that every corner-gluing* 2*-D SFT is also* block-gluing. *For the definition of block-gluingness and the description of a whole hierarchy of mixing conditions for* 2*-D SFTs, see [BPS10]. Without giving further details, we just mention that in* 2*-D , we can define at least four non-equivalent different mixing conditions while the* 1*-D analogue of all these mixing conditions is equivalent to normal mixing, see also [LM95].*

Before proving the main Theorem of this section, we still need one more definition: Let $R \subseteq \mathbb{Z}^d$ and $\vec{v} \in \mathbb{Z}^d$,$\vec{v} \neq \vec{0}$. A pattern p on R is called $\vec{v}$*-periodic* if whenever $\vec{w}$ and $\vec{w} + \vec{v}$ belong to R, we have that $p(\vec{w}) = p(\vec{w} + \vec{v})$.

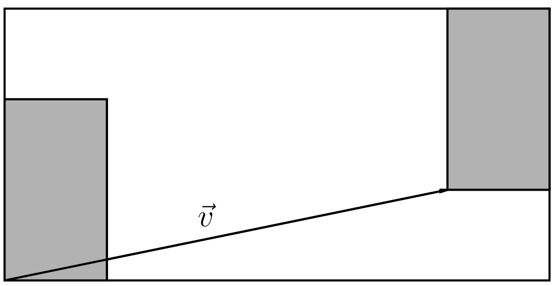


Figure 24: A $\vec{v}$-periodic pattern.

Lemma 11. *Let* $X \subseteq A^{\mathbb{Z}^d}$ *be a corner-gluing SFT with* $h(X) > 0$ *and gluing constant* g*. Then, for every* $f, c \in \mathbb{N}$ *and allowed square* $F \in B_f(X)$*, there exists some square* $M \in B(X)$ *with* F *appearing in each one of its four corners such that* M *is not* $\vec{v}$*-periodic whenever* $|\vec{v}|_\infty < c$*.*

Proof. Let $Q_0 \in B_c(X)$ be any admissible square of side c and $\vec{v_1}, \vec{v_2}, \dots, \vec{v_m}$ be the periodicity vectors of Q_0 with $|\vec{v}|_\infty < c$. We will construct a sequence $Q_0, Q_1, \dots, Q_m$ of squares such that for $i = 1, 2, \dots, m$:

1. Q_i is an extension of Q_{i-1}, and

2. Q_i is not $\vec{v_i}$-periodic.

Then, as one can easily see, Q_m is not $\vec{v}$-periodic for any vector $\vec{v}$ with $|\vec{v}|_\infty < c$.

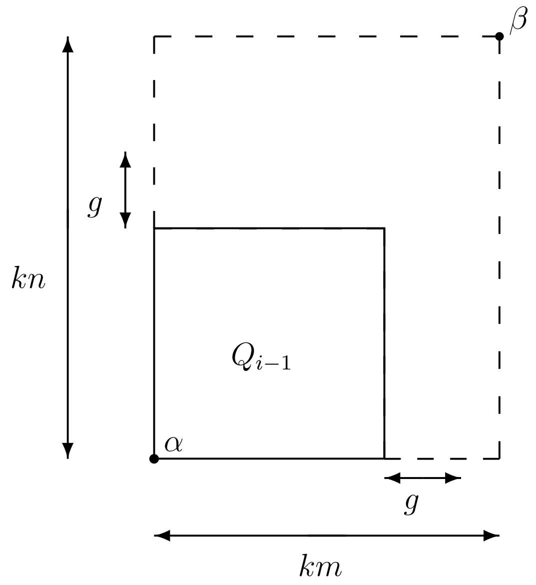


Figure 25: Constructing Y_i from Q_{i-1}.

Suppose that we have already constructed $Q_{i-1} \in B_l(X)$ for some $l \in \mathbb{N}$ and that $v_i = (m, n)$, where $m, n \geq 0$. Let α be the letter appearing in the lower left corner of Q_{i-1}. Since $h(X) > 0$, there is a letter $\beta \neq \alpha$ appearing in an element of X. Consider β as a 1×1 rectangle and place it in position (km, kn), where $k \in \mathbb{N}$ is such that $km, kn > l + g$. Since X is corner-gluing, we can glue β to Q_{i-1}. Let Y_i be the rectangle thus obtained. Y_i is not $\vec{v_i}$-periodic as this would imply that $\alpha = \beta$. Extend Y_i to an allowed square in order to get Q_i. Then, Q_i is not $\vec{v_i}$-periodic, too.

We can treat the cases $m \geq 0, n < 0$ or $m < 0, n \geq 0$ or $m < 0, n < 0$ similarly. The only thing that changes is the placement of the letter β.

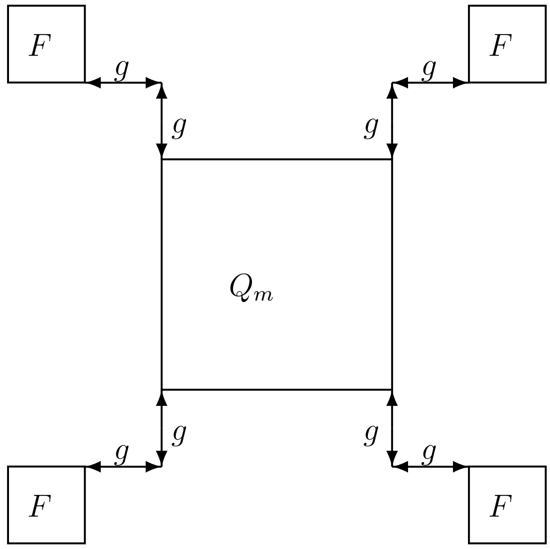


Figure 26: Constructing M from Q_m.

Having constructed $Q_m \in B_r(X)$ for some $r \in \mathbb{N}$, it is easy to extend it to a square pattern M with F appearing in every one of its corners. Just place four copies of F around Q_m as in the following figure and glue them one-by-one to Q_m according to Lemma 10.

□

Theorem 18. *Let X be a d-D corner-gluing SFT with $h(X) > \log n$. Then, X factors onto the d-D n-full shift $X_{d,n}$.*

Proof. By Theorem 11, we know that there exists a proper subsystem Y of X with $h(Y) > \log n$. Since $Y \subsetneq X$, there exists some square pattern $F \in B(X)$ which is forbidden for Y. Let f be its size. Construct M using F as in Lemma 11, and for $c = 2(f + g)$. Let the side length of M be m.

Let us consider the pattern L appearing in Figure 3.6, where $l \in \mathbb{N}$ and $G \in B(Y)$ can be any rectangular pattern of size $m \times l$ (or $l \times m$), $l \in \mathbb{N}$, and the grey regions are the necessary gluing strips needed to hold the patterns together.

Notice that, since $G \in B(Y)$, F does not appear in G. The upper right corner of L, (marked with a black dot in the figure above) is called the *inside corner* of L.

Next, for every $G' \in B_l(Y)$, we construct the *follower pattern* with *central block* G' of Figure 3.7.

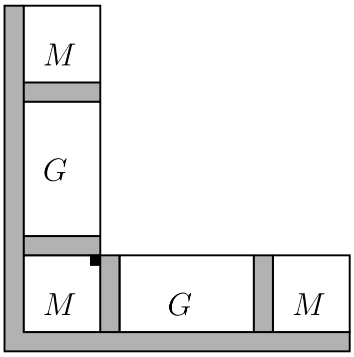


Figure 27: A L-shaped configuration and its inside corner.

As before, $G \in B(Y)$ can be any allowed block of the appropriate size and the grey regions are gluing strips. Hence, there there does not exist a unique follower pattern with central block G', but rather a whole family of them obtained by letting the blocks G vary among all the allowable blocks of Y. However, we are going to abuse the terminology and talk about *the* follower pattern with central block G', since we don't care about the patterns G. Since X is corner-gluing, for every $G' \in B_l(Y)$, there is a choice for the gluing strips such that the pattern of figure 3.8 with central block G' is allowed in X.

We will refer to such patterns as $D_1, D_2, \ldots$. Don't forget that the blocks labeled G in the previous picture are not necessarily the same. They are just some blocks allowed in Y. Let $J = l + 2g + m$. Then, every follower pattern is a member of $B_J(X)$. Since for every $G' \in B_l(Y)$ there exists a follower pattern with central pattern G', we deduce that the number of follower patterns is at least $B_l(Y)$. In addition, as $h(Y) > 0$ we have that

$$
\begin{aligned}
& \lim_{l\to\infty} \left(\frac{\log |B_l(Y)|}{l^2}\right) > \log N \Rightarrow \\
\Rightarrow \quad & \lim_{l\to\infty} \left(\frac{(l+2g+m)^2}{l^2} \frac{\log |B_l(Y)|}{(l+2g+m)^2}\right) > \log N \Rightarrow \\
\Rightarrow \quad & \frac{\log |B_l(Y)|}{(l+2g+m)^2}) > \log N, \text{ for } l \text{ large enough } \Rightarrow \\
\Rightarrow \quad & B_l(Y) > N^{J^2}, \text{ for } l \text{ large enough.}
\end{aligned}
$$

For such a fixed, large enough l, we partition its followers into N^{J^2} different sets $F_1, F_2, \ldots, F_{N^{J^2}}$ according to their central block. (To be honest, this is not

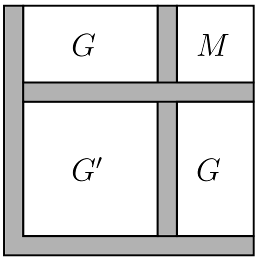


Figure 28: The *follower pattern* with *central block* G'.

a partition since there are more than N^{J^2} followers, so some of them will be left over, although this really doesn't matter.)

We claim that if D_1 and D_2 occur at different places in a configuration $x \in X$, then their follower parts do not intersect.

Indeed, let D_1 and D_2 occur in different places in x and such that their follower parts intersect. We can assume that the lower left corner of D_1 lies on the origin of the plane. Let $\vec{v}$ be the coordinate of the lower left corner of D_2. Then, $|\vec{v}_\infty| > c = 2(f+g)$, since , by construction, M is not $\vec{u}$-periodic for any $\vec{u}$ with $|\vec{u}|_\infty < c$. Similarly, $|\vec{v}_\infty| \leq J - c$, since otherwise the lower left corner M of D_2 would intersect too much with some M of D_1, or the follower parts of D_1 or D_2 would not intersect. Now, since M was constructed with a F in every corner and $c = 2(f+g)$, we conclude that at least one subblock F of D_2 occurs entirely into some G or in G' of D_1. However, this is impossible, since F is forbidden in Y and G, G' are chosen from $B(Y)$. Hence, the follower patterns of D_1 and D_2 cannot overlap.

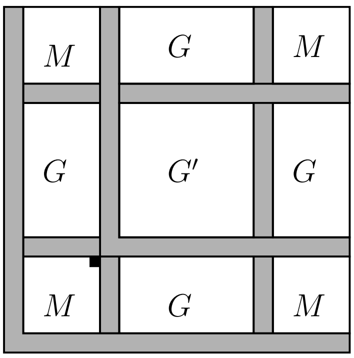
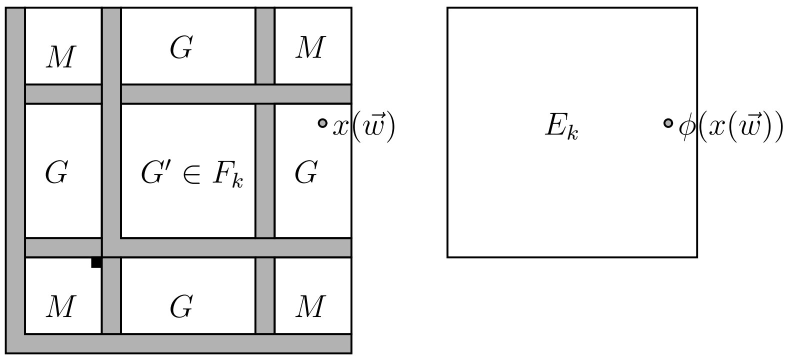


Figure 29: Generic form of a D pattern.

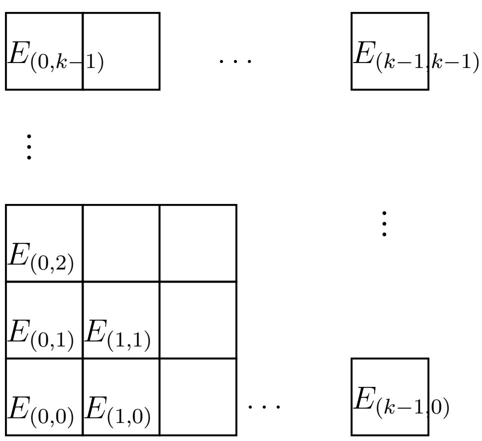
We are now ready to describe the factor map ϕ between X and $X_{2,n}$. Let $E_1, E_2, \ldots, E_{N^{J^2}}$ be all the $(J \times J)$-patterns in alphabet $\{0, 1, \ldots, n-1\}$. We define the obvious bijection $\pi\colon \{F_1, F_2, \ldots, F_{NJ^2}\} \to \{E_1, E_2, \ldots, E_{NJ^2}\}$. Let $x \in X$ be an arbitrary configuration and $\vec{w}$ a position of the plane. First of all, the local rule of ϕ checks if $\vec{w}$ belongs in the follower pattern of some D (this can be checked with a window of size $2(J+g+m)$). If it doesn't, then $\phi(x(\vec{w})) = 0$. Otherwise, thanks to the previous claim, we know that there exists a unique D such that $\vec{w}$ belongs to its follower part. Therefore, there exist unique $\vec{u}$ and $\vec{z}$ such that $\vec{w} = \vec{u} + \vec{z}$, the inside corner of D lies on $\vec{u}$ and $0 < z_i \leq J$, for $i = 1, 2$. Then, if the follower part of D belongs to F_k, we map $x(\vec{w})$ to $E_k(\vec{z})$. This can be seen more clearly in the following figure:



It is obvious that in this way, the whole follower part of D gets mapped to E_k.

By the previous discussion, ϕ is well-defined. It is also continuous and shift-commuting, as it is defined with a local rule. We only have to show that it is surjective:

In order to do that, it suffices to show that square patterns of arbitrarily large size have preimages. Let $E \in B_{kJ}(X_{2,n})$ be a square of size kJ, where $k \in \mathbb{N}$. E can be decomposed as follows:



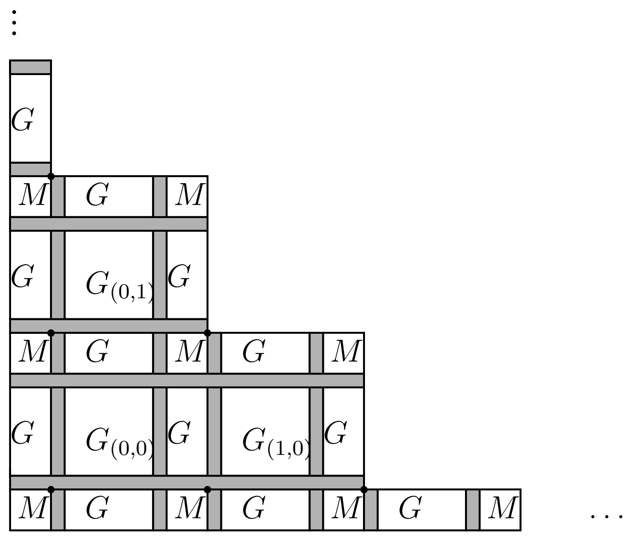


Figure 30: Constructing a preimage of E.

We can construct a preimage of E in X inductively as follows: Starting with an L-shaped configuration of length $kJ + m + g$, we focus on the inside corner thus defined. If $E_{(0,0)} = E_i$, for some $1 \leq i \leq N^{J^2}$, then we place at that corner a follower pattern from F_i. In this way, we have defined two new inside corners, where we will place the appropriate follower patterns for $E_{(1,0)}$ and $E_{(0,1)}$, and so on. In the end, we have a pre-image of E, (together with a band of 0's of thickness $g+m$ on the left and lower edge, but this is not a problem). Hence, ϕ is surjective and X factors onto the 2-D n-full shift.

$\square$

4.4 Automorphism groups of d-D SFTs

The last subject in our brief exposition of multidimensional SFTs is the automorphism group of a d-D SFT. Recall that by an *automorphism* of a SFT X we mean a homeomorphism of X commuting with the shifts. Since these obviously form a group under composition, the term automorphism group is justified. We denote this group by $aut(X)$. When $X = A^{\mathbb{Z}^d}$ is a full shift, then $aut(X)$ is just the set of reversible d-D CA with alphabet A. Generally, Proposition 6 can be generalized to say that $F\colon X \to X$ is an automorphism if and only if it is a bijective block map, that is there exists some $n \geq 1$ and some $f\colon B_{[-n,n]^d} \to A$ such that $F(c)(\vec{x}) = f(c(\vec{x} + [-n,n]^d)$, for every $c \in A^{\mathbb{Z}^d}$ and every $\vec{x} \in \mathbb{Z}^d$. It follows

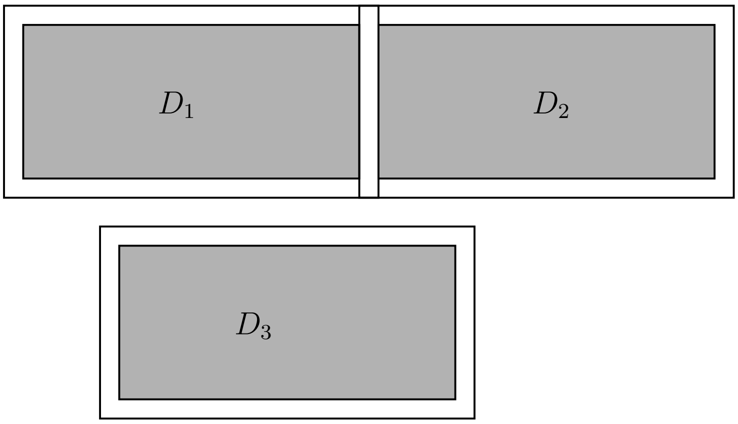


Figure 31: A marker in the special case when R is a rectangle and S is its interior.

that $aut(X)$ is countable.

Usually, when examining the automorphism group of a SFT we assume some kind of mixing property, since without such an assumption it is very difficult to obtain reasonable results. For example, Hedlund in [Hed69] proved that the automorphism group of a 1-D full shift contains a copy of every countable direct sum of finite groups (this means that for every such group, there exists a subgroup of $aut(X)$ isomorphic to it). Later, it was proved that if X is a mixing 1-D SFT, then the same is true for $aut(X)$, and $aut(X)$ contains the free group on a countable number of generators. For results on the automorphism group of 1-D SFTs, see [BLR88, KR91].

Much less work has been done and less results are known for the d-D case, which is of most interest for us. We are going to present a weaker result for d-D SFTs, assuming the corner-gluing property. Namely, we are going to prove that:

Theorem 19. *[War91] Let X be a corner-gluing SFT. Then, $aut(X)$ contains a copy of every finite group.*

Before proving Theorem 19, let us note that recently M. Hochman proved the same result under the much weaker assumption of positive entropy. However, the proof is too long to be included in this thesis. Almost all of the arguments dealing with automorphism groups of SFTs use the notion of markers. In Hochman's proof, the existence of a marker is proved with a non-constructive argument, while our argument is constructive.

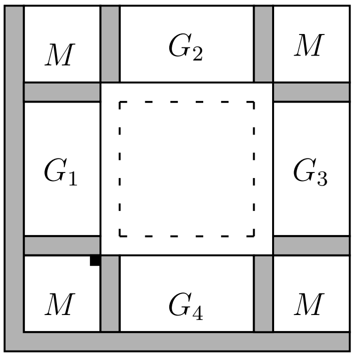


Figure 32: The marker M. The blocks G_i are some random but fixed blocks of $Y \subset X$ of the appropriate size.

Proof. First of all, let us define formally what a marker is. Let $R, S \subseteq \mathbb{Z}^d, R \cap S = \emptyset$, and $D \subseteq A^S$ be a set of patterns with domain S. $M \in A^R$ is called a *marker* for D if it has the following *trivial overlap property*: if $c \in A^{\mathbb{Z}^d}$ satisfies $c|_S \in D, c|_R = M$ and $c|_{S+\vec{n}} \in D, c|_{R+\vec{n}} = M$, then $S \cap (S + \vec{n}) = \emptyset$. Hence, we can isolate the occurences of patterns from D using M.

If M is a marker for D, we can embed a copy of $Sym(D)$ in $aut(X)$, where $Sym(D)$ denotes the symmetry group of D: For $\tau \in Sym(D)$, define $\alpha_\tau \in aut(X)$ as follows. If $c \in X$ has $c|_{S+\vec{n}} = D_1 \in D$ and $c|_{R+\vec{n}} = M$, then $\alpha_\tau(c)|_{S+\vec{n}} = \tau(D_1)$. Otherwise, α_τ leaves the symbols of c unchanged. Therefore, α_τ acts by applying τ to patterns from D that are marked with M. Since M is a marker for D, α_τ is well-defined and it is obvious that it is an automorphism of X and that the correspondence $\tau \rightarrow \alpha_\tau$ embeds $Sym(D)$ into $aut(X)$.

If we can construct markers for sets D of arbitrarily large number of elements, then, according to the previous observation, every finite symmetry group can be embedded into $aut(X)$. Since every finite group can be embedded into a finite symmetry group (Cayley's theorem), we conclude that $aut(X)$ contains a copy of every finite group, see [War91].

Luckily enough, this has already been done: namely if we define M as in Figure 32 and let D be the set of all blocks $G' \in B_l(X)$ that can appear as a central block of a follower, then M is marker for D. Indeed, if M occurs in two different positions of the plane, then the follower parts do not intersect, therefore a fortiori, neither the central blocks intersect. In addition, since $h(X) > 0$ the number of blocks in $B_l(X)$ can be arbitrarily large.

5 Open problems

5.1 Nivat's conjecture

Let S be a finite alphabet with at least two letters. A 1-*D word* is an element $c \in S^{\mathbb{Z}}$. Word c is called *periodic* if there exists some $m \geq 1$ such that $c(x+m) = c(x)$, for every $x \in \mathbb{Z}$. For every $n \geq 1$ we define $p_c(n)$ to be the number of different words of length n appearing in c. Formally,

$$p_c(n) = |\{x \in S^n : \text{ there exists } i \in \mathbb{Z} \text{ such that } c(i + [0, n]) = x\}|.$$

We call $p_c(n)$ the *complexity function* of c. The following Theorem characterizes the periodic words in terms of their complexity function.

Theorem 20. *[Morse-Hedlund Theorem] c is periodic if and only if there exists some $n \geq 1$ with $p_c(n) \leq n$.*

A proof can be found in [Lot02].

Similarly, let $\xi \in S^{\mathbb{Z}^2}$ be a 2-*D word*. ξ is called *periodic* with period $\vec{m} \in \mathbb{Z}^2$ if $\xi(\vec{x} + \vec{m}) = \xi(\vec{x})$, for every $\vec{x} \in \mathbb{Z}^2$. For $n_1, n_2 \geq 1$, let $N_\xi(n_1, n_2)$ be the number of different rectangular patterns of size $n_1 \times n_2$ appearing in ξ. Formally, let $B(n_1, n_2) = \{(i, j) : 0 \leq i < n_1 \text{ and } 0 \leq j < n_2\}$ be the rectangle of size $n_1 \times n_2$ with lower left corner in the origin. Then, $N_\xi(n_1, n_2) = |\{x \in S^{n_1 \times n_2} : \text{ there exists } (i, j) \in \mathbb{Z}^2 \text{ such that } \xi((i, j) + B(n_1, n_2)) = x\}|$. A natural question is whether Theorem 20 is true also in the 2-D setting, namely whether it is true that a 2-D word is periodic if and only if there exists n_1, n_2 such that $N_\xi(n_1, n_2) \leq n_1 n_2$. The following example from [BV00] shows that the best we can hope for is the "if direction":

There exists a periodic 2-D word ξ such that $N_\xi(n_1, n_2) = |S|^{n_1+n_2-1}$, for every $n_1, n_2 \geq 1$. Indeed, let $c \in S^{\mathbb{Z}}$ be a 1-D word that contains every finite word as a subword. For the existence of such a word, see Example 9.4 in [RS97]. Let us define $\xi \in S^{\mathbb{Z}^2}$ by $\xi(i, j) = c(i+j)$. Then, $\xi(i+1, j-1) = c(i+1+j-1) = c(i + j) = \xi(i, j)$ hence ξ is $(1, -1)$ periodic. This means that the letters of a rectangle are uniquely determined by the letters on the bottom and right edges of the rectangle, and that these in turn are uniquely determined by the subword of c of length $n_1 + n_2 - 1$ observed in the bottom edge of the rectangle as in

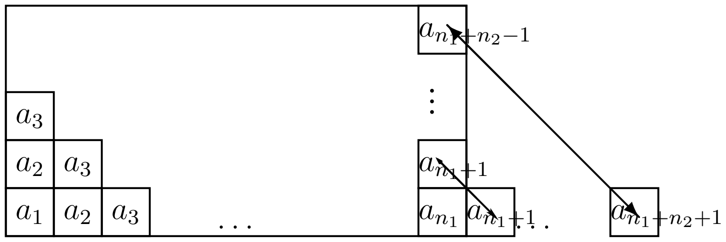


Figure 33: A periodic word with many rectangles appearing in it.

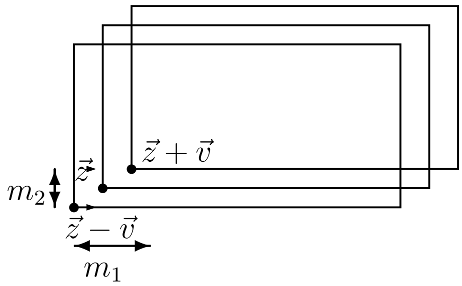


Figure 34: A graphical statement of lemma 12.

figure 4.1. Since all words of lentgh $n_1 + n_2 - 1$ appear in c, we conclude that $N(n_1, n_2) = |S|^{n_1+n_2-1}$.

A similar construction can be given in any dimension $d \geq 3$. In addition, for $d \geq 3$, even the ”if implication” is false. There exists an aperiodic $3{-}D$ word $\phi \in S^{\mathbb{Z}^3}$ with $N_\phi(n_1, n_2, n_3) \leq n_1 n_2 n_3$, for every $n_1, n_2, n_3 \geq 1$, see [ST00]. The ”if direction” for the 2-D case is still open and is known as *Nivat’s conjecture*.

Conjecture 1 (Nivat’s conjecture)**.** *Let $\xi \in S^{\mathbb{Z}^2}$. If there exist $n_1, n_2 \geq 1$ with $N_\xi(n_1, n_2) \leq n_1 n_2$, then ξ is periodic.*

Observe that if Nivat’s conjecture is true, as most people believe it is, then it will be the first dimension-sensitive property where the differnce is noticed when going from 2-D to 3-D and not from 1-D to 2-D . In the rest of this section, we will give part of the proof of the best currently known result towards a positive answer for Nivat’s conjecture. For other results pointing towards a positive answer, see [EKM03, ST02]. Of course, we should always bear in mind that Nivat’s conjecture could as well be false.

Theorem 21. *[QZ04] Let $\xi \in S^{\mathbb{Z}^2}$. If there exist $n_1, n_2 \geq 1$ with $N_\xi(n_1, n_2) \leq \frac{1}{16} n_1 n_2$, then ξ is periodic.*

From now on, we will always assume that $\xi \in S^{\mathbb{Z}^2}$ satisfies the hypothesis of Theorem 21. For vectors $\vec{v} = (v_1, v_2), \vec{z} = (z_1, z_2) \in \mathbb{Z}^2$ we say that $|\vec{v}| < \vec{z}$ if and only if $|v_1| < z_1$ and $|v_2| < z_2$. Finally, we set $m_i = \lfloor \frac{n_i}{4} \rfloor$ and $l_i = n_i - m_i$ for $i = 1, 2$ and $\vec{m} = (m_1, m_2)$. The proof is based on the following observation:

Lemma 12. *For every $\vec{z} = (z_1, z_2) \in \mathbb{Z}^2$, there exists a vector $\vec{v} = (v_1, v_2)$ with $|\vec{v}| \leq \vec{m}$ such that $\xi(\vec{z} + B(l_1, l_2)) = \xi(\vec{z} + \vec{v} + B(l_1, l_2)) = \xi(\vec{z} - \vec{v} + B(l_1, l_2))$.*

Proof. Let us consider the $n_1 \times n_2$ rectangles $\xi(\vec{y} + B(n_1, n_2))$ appearing in ξ in positions $\vec{y}$ with $z_i - m_i \leq y_i \leq z_i$. The number of these rectangles is $(m_1 + 1)(m_2 + 1) > \frac{n_1 n_2}{16} \geq N_\xi(n_1, n_2)$, therefore two of them must be equal. Let $\xi(\vec{y} + B(n_1, n_2)) = \xi(\vec{y'} + B(n_1, n_2))$ and $\vec{v} = \vec{y} - \vec{y'}$.

If $\vec{x} \in \vec{z} + B(l_1, l_2)$, then $\vec{x} \in \vec{y} + B(n_1, n_2) \cap \vec{y'} + B(n_1, n_2)$. Also, since $\xi(\vec{y} + B(n_1, n_2)) = \xi(\vec{y} - \vec{v} + B(n_1, n_2))$, we conclude that $\xi(\vec{x} - \vec{v}) = \xi(\vec{x})$. Hence, $\xi(\vec{z} + B(l_1, l_2)) = \xi(\vec{z} - \vec{v} + B(l_1, l_2))$.

In a similar way, we can prove that $\xi(\vec{z} + B(l_1, l_2)) = \xi(\vec{z} + \vec{v} + B(l_1, l_2))$. $\square$

We have shown that for every $\vec{z} \in \mathbb{Z}^2$, there exists some vector $\vec{v} \in \mathbb{Z}^2$ with $|\vec{v}| \leq \vec{m}$ such that the $l_1 \times l_2$ pattern that appears in position $\vec{z}$ is the same as the $l_1 \times l_2$ patterns that appear in positions $\vec{z} \pm \vec{v}$. We call $\vec{v}$ an *actual translation vector* for this specific appearance of $\xi(\vec{z} + B(n_1, n_2))$.

We say that $\vec{u} \in \mathbb{Z}^2$ is a *potential translation vector* for a $l_1 \times l_2$ pattern B if $|\vec{u}| \leq \vec{m}$, and in addition $B(\vec{x}) = B(\vec{x} + \vec{u})$, whenever $\vec{x}$ and $\vec{x} + \vec{u}$ belong to B. Using the terminology of Chapter 3, we could rephrase the second condition by saying that $\vec{u}$ is a period of B.

The following observations about actual and potential translation vectors should clarify these notions a little bit:

- The notion of an actual translation vector refers to a specific appearance of a $l_1 \times l_2$ pattern in ξ, while the notion of potential translation vector does not.

- For every $l_1 \times l_2$ pattern B in ξ, the actual translation vector of each specific appearance of B is one of its potential translation vectors. For different appearances of the same pattern, we might have different actual translation vectors (hence the terms potential and actual).

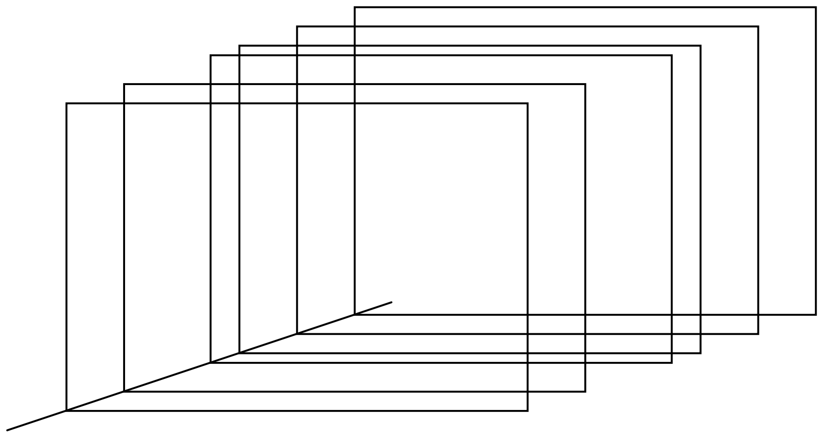


Figure 35: Copies of B along the line $L_{\vec{h}}$.

- Of course, there might be more than one actual translation vector for a specific appearance of a $l_1 \times l_2$ pattern. This is not a problem. We are only interested in the fact that there exists at least one.

The proof now is divided in the analysis of two cases. Either there exists a $l_1 \times l_2$ pattern all of whose translation vectors are collinear or every $l_1 \times l_2$ pattern has at least two non-collinear potential translation vectors. We are only going to treat the first case which, admittedly, is easier. For the complete proof, see [QZ04].

So, let us assume that B is a $l_1 \times l_2$ rectangle all of whose potential translation vectors are collinear. Let $\vec{h} = (h_1, h_2)$ be a potential translation vector. Withour loss of generality, we can suppose that the coordinates of $\vec{h}$ are positive and that $\frac{h_2}{h_1} \leq \frac{l_2}{l_1}$. In fact, if both h_1 and h_2 are negative, then replace $\vec{h}$ by $-\vec{h}$. If one of the coordinates of $\vec{h}$ is negative, then reflect around the appropriate axis to get a 2-D word ξ' that is periodic if and only if ξ is periodic and whose potential translation vector is positive. Finally, if $\frac{h_2}{h_1} > \frac{l_2}{l_1}$, then reflect around the main diagonal. Therefore, we may assume that $\vec{h}$ is a positive vector lying under the main diagonal of B.

Now, let us consider a specific occurence of B in ξ. We denote by $L_{\vec{h}}$ the line that goes through the origin and has slope $\vec{h}$. Using an a appropriate shift, we can suppose that B appears in the origin. Since all of the potential translation vetors of B are collinear, the actual translation vector for this appeareance of B is $t\vec{h}$, where t is rational number and $|t\vec{h}| \leq \vec{m}$. Therefore, B appears also in position

$\pm t\vec{h}$. Similarly, the actual translation vector for the appearance of B in position $t\vec{h}$ is equal to $s\vec{h}$ for some suitable s, hence B appears also in position $(t+s)\vec{h}$. Inductively, we can see that B is placed all over $L_{\vec{h}}$, possibly irregularly spaced, but in such a way that if B appears in some position of $L_{\vec{h}}$, then there is another appearance of B on $L_{\vec{h}}$ within horizontal distance at most m_1.

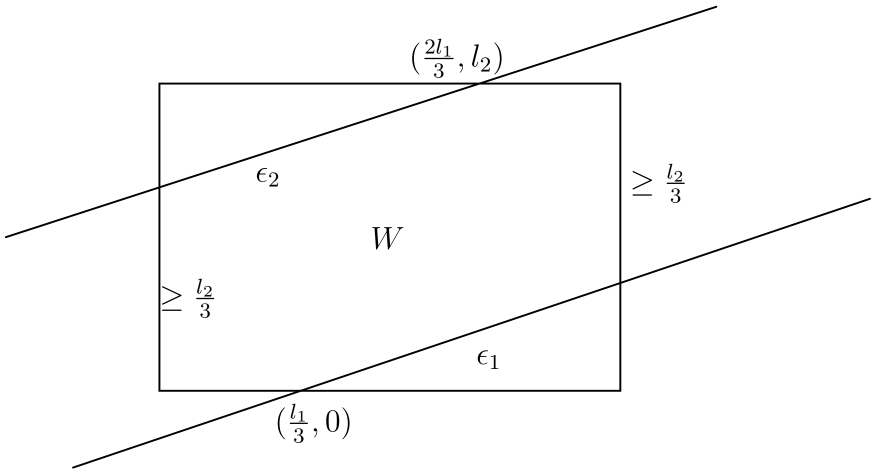


Figure 36: The band region W defined by the lines ϵ_1 and ϵ_2.

Next, let us consider the line ϵ_1 which goes through point $(\frac{l_1}{3}, 0)$ and has slope $\frac{h_2}{h_1}$, and the line ϵ_2 which goes through $(\frac{2l_1}{3}, l_2)$ and has the same slope. Let W be the infinite band consisting of all the points lying between these two lines. Writing down explicitly the equations of ϵ_1 and ϵ_2 in analytic form and using some basic analytical geometry and the fact that $\frac{h_2}{h_1} \leq \frac{l_2}{l_1}$, we can show that ϵ_1 intersects the right edge of the rectangle more than $\frac{l_2}{3}$ away from the upper right corner, and that ϵ_2 intersects the left edge of the rectangle more than $\frac{l_2}{3}$ away from the origin. Hence, if $\vec{u} = (u_1, u_2)$ and the points $\vec{z}$ and $\vec{z} + \vec{u}$ lie on opposite sides of the band, then $|u_1| > m_1$ or $|u_2| > m_2$.

Let $\vec{x} \in W$ be a lattice point. By definition of W, $\vec{x} + \vec{h}$ also belongs to W. As we have already observed, the appearances of B along $L_{\vec{h}}$ are spaced out at horizontal intervals of at most m_1. This means that there exists a copy of B that contains $\vec{x}$ and lies at a horizontal distance of at most m_1 to the left of $\vec{x}$. Then, $\vec{x} + \vec{h}$ also belongs to this copy of B as can be seen in Figure 4.5. Since $\vec{h}$ is a period of B, we conclude that $\xi(\vec{x}) = \xi(\vec{x} + \vec{h})$. Since $\vec{x} \in W$ was arbitrary, we conclude that ξ is periodic on W.

Therefore, ξ is periodic with period $\vec{h}$ on an infinite band with the property that if $\vec{z}$ and $\vec{z}+\vec{u}$ lie on opposite parts of the band, then $|u_1| > m_1$ or $|u_2| > m_2$. We will now show that every infinite band satisfying these properties can be extended to a larger infinite band satisfying the same properties on which ξ is periodic with a bounded period. Notice, also, that although our initial assumption was that every potential vector of B is parallel to $\vec{h}$, the existence of W follows from the weaker assumption that there exists an appearance of B such that for each appearance of B on $L_{\vec{h}}$ all of its actual translation vectors are parallel to $\vec{h}$.

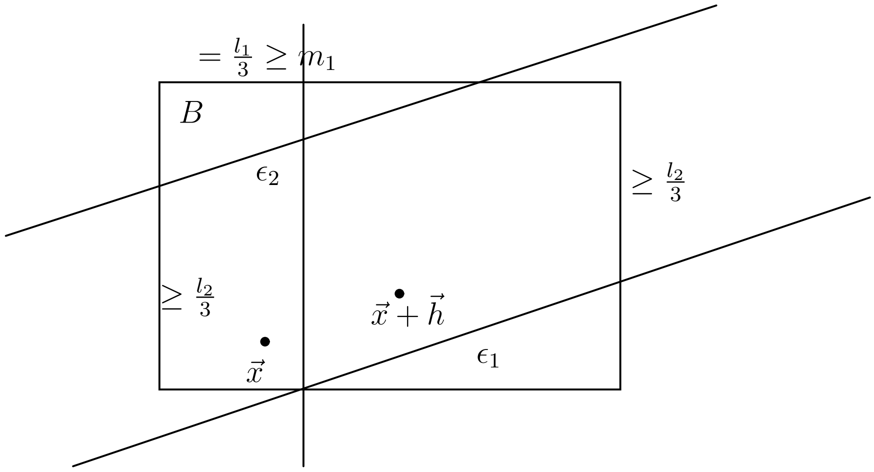


Figure 37: ξ is periodic on the infinite band W.

Let Y be any infinite band on which ξ is periodic and such that if $\vec{z}$ and $\vec{z}+\vec{u}$ lie on opposite parts of the band, then $|u_1| > m_1$ or $|u_2| > m_2$. Consider the closest line L' wich is parallel to $\vec{h}$ but is not contained in Y. Without loss of generality, we can assume that L' lies below Y. If L' is periodic with period $\vec{h}$, then we can definitely adjoin it to Y to obtain a wider band that satisfies the same conditions. If, on the other hand, there exists some $\vec{z} \in L'$ with $\xi(\vec{z}) \neq \xi(\vec{z}+\vec{h})$, let us consider the $l_1 \times l_2$ pattern C appearing in $\vec{z}$. We claim that for each appearance of C on L', every actual translation vector of C is parallel to $\vec{h}$. Indeed, let us suppose that for some appearance of C on L', there exists an actual translation vector $\vec{y} = (y_1, y_2)$ of this appearance of C that is not parallel to $\vec{h}$. Let us suppose that $y_2 > 0$. Then, $\vec{z}+\vec{y}$ and $\vec{z}+\vec{h}+\vec{y}$ both belong in Y. Indeed, $\vec{z}$ and $\vec{z}+\vec{h}$ lie on one side of the band and, since $|\vec{y}| \leq \vec{m}$, our assumption about Y indicates that $\vec{z}+\vec{y}$ and $\vec{z}+\vec{h}+\vec{y}$ cannot lie on the other side. Hence, they must be in Y. But then, Lemma 12 and the fact that Y is periodic with period $\vec{h}$ imply that

$$\xi(\vec{z}) = \xi(\vec{z} + \vec{y}) = \xi(\vec{z} + \vec{y} + \vec{h}) = \xi(\vec{z} + \vec{h}),$$

which is a contradiction with our initial assumption. If $y_2 < 0$, then the same argument works by replacing $\vec{z}+\vec{y}$ and $\vec{z}+\vec{h}+\vec{y}$ by $\vec{z}-\vec{y}$ and $\vec{z}+\vec{h}-\vec{y}$, respectively.

Therefore, we have proved that for every appearance of C on L', every actual translation vector is parralel to $\vec{h}$. This means that L' is contained in an infinite band on which ξ is periodic with period $n\vec{h}$, for some integer $n \geq 1$ with $n\vec{h} \leq \vec{m}$. Since there are finitely many such n, we may take their least common multiple M. Then, ξ is periodic on $W \cup L'$ with period $M\vec{h}$ and M is independent from Y, L' and C.

$M\vec{h}$ is a period of ξ. Indeed, for every $\vec{x} \in \mathbb{Z}^2$, we can extend W as many times as needed in order to obtain a band that contains $\vec{x}$ on which ξ is periodic with period $M\vec{h}$. Then, $\xi(\vec{x}) = \xi(\vec{x} + M\vec{h})$.

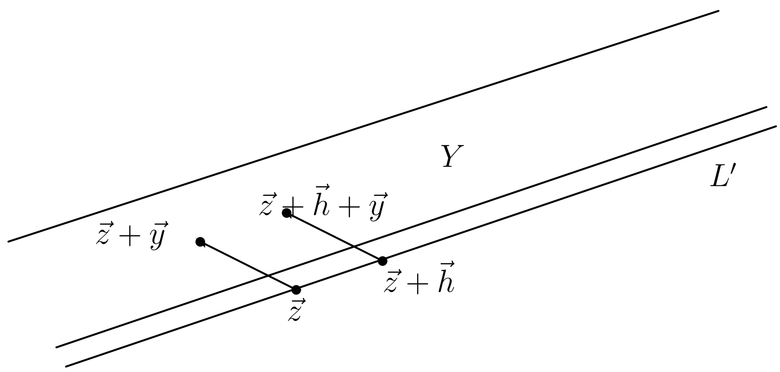


Figure 38: Expanding the infinite band Y.

5.2 Other open problems

5.2.1 Periodic points for d-D CA

Question 1. *Does a surjective, almost equicontinuous d-D CA have dense periodic points?*

In 1-D , the answer is positive, see [BT00] . However, the proof uses the characterization of almost equicontinuous CAs in terms of blocking words, so it cannot be generalized to greater dimensions.

5.2.2 Entropy of 1-D and 2-D CA

In section 2.3, we gave a recursion theoretic characterization of the class of numbers that can be the topological entropy of a d-D CA, for $d \geq 3$. Is this characterization also true in the 1-D and 2-D case? What else can we say about the topological entropy of 1-D and 2-D CA?

Question 2. *Is it true that a number is the topological entropy of a 1-D or 2-D CA if and only if it is the $\liminf$ of a recursive sequence of numbers?*

We note that the topological entropy of CAs is uncomputable even in the 1-D case, see [HKC92].

5.2.3 Openness and number of preimages

Theorem 22. *Let $G\colon S^{\mathbb{Z}} \to S^{\mathbb{Z}}$ be a 1-D CA. The following are equivalent:*

1. *G is open.*
2. *There exists $p \geq 1$ such that $|G^{-1}(x)| = p$ for every $x \in S^{\mathbb{Z}}$.*
3. *There exist continuous functions $f_1, f_2, \dots, f_p\colon S^{\mathbb{Z}} \to S^{\mathbb{Z}}$ such that $G^{-1}(x) = \{f_1(x), f_2(x), \dots, f_p(x)\}$ and $f_i(x) \neq f_j(x)$ for every $x \in S^{\mathbb{Z}}$ and every $i \neq j$.*

The set of functions $\{f_1, f_2, \dots, f_p\}$ is called a *cross-section*.

For a proof, see [Kur03, Hed69]. Again, the proof of this theorem makes extensive use of notions and results that are exclusive to the 1-D case. Can Theorem 22 be generalized to higher dimensions? The answer is negative as the following example shows.

Let F_1 be an open, non-injective 1-D CA, for example the CA induced by the *xor* function. Then, according to Theorem 22, every point $x \in S^{\mathbb{Z}}$ has exactly p preimages, where $p > 1$. Let F be the 2-D CA that works by applying F_1 on every line of the plane independently. Obviously, F is open and $|G^{-1}(c)| = \infty$ for every $c \in S^{\mathbb{Z}^2}$.

Hence, in greater dimensions we cannot have a result as strong as Theorem 22. However, the following questions deserve some attention:

Question 3. *Does there exists an open d-D CA $G\colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ and points $c_1, c_2 \in S^{\mathbb{Z}^d}$ such that $|G^{-1}(c_1)| = k_1$, $|G^{-1}(c_2)| = k_2$ and $k_1 \neq k_2$?*

Question 4. *Does there exists an open d-D CA $G\colon S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ and points $c_1, c_2 \in S^{\mathbb{Z}^d}$ such that $|G^{-1}(c_1)| < \infty$ and $|G^{-1}(c_2)| = \infty$?*

5.2.4 Decidability of positive expansiveness

Question 5. *Given a 1-D CA G, can we algorithmically decide whether G is (positively) expansive?*

Stricly speaking, this question does not concern dimension-sensitive properties of CAs. However, since the decidability status of almost every other dynamic property has been settled and expansive CAs seem to be a peculiarity of 1-D, this problem fits in this list.

For an investigation on the general properties of positively expansive CAs, see [BM97, Nas02, Kur97].

5.2.5 Relations between G, G_F and G_P

Recall the definition of G_F and G_P given in Chapter 2.

Question 6.

1. *If G_P is injective, is G_F surjective?*
2. *If G_F is surjective, is G_P also surjective?*
3. *If G is surjective, is G_P also surjective?*

If G is a 1-D CA, the answer to all questions is positive, see [Kar05]. For higher-dimensional CAs, the answer is currently unkwown.

5.2.6 Sofic shifts and entropy of their covers

In section 3.2 we proved that for every d-D sofic shift T and every $\epsilon > 0$ there exists a SFT cover S of T with $h(S) < h(T) + \varepsilon$.

Question 7. *Does every d-D sofic shift have a cover of equal entropy?*

The answer for the 1-D case is positive, see [CP75].

5.2.7 A problem on direct products of SFTs

Conjecture 2. *Let $X, Y \subseteq S^{\mathbb{Z}^d}$ be subshifts. If $X \times Y$ is conjugate to a full shift, then some powers of X and Y are conjugate to full shifts, too.*

The above conjecture arose in the study of reversible CAs. For $d = 1$ it is true. If it is true for higher dimensions too, then every d-D reversible CAs can be described as a composition of block permutations, see [Kar96].

References

[AP72] S. Amoroso and Y.N. Patt. Decision procedure for surjectivity and injectivity of parallel maps for tessellation structures. *J. Comput. System Sci.*, 6:448–464, 1972.

[BDFK05] Vincent Bernardi, Bruno Durand, Enrico Formenti, and Jarkko Kari. A new dimension sensitive property for cellular automata. *Theoretical Computer Science*, 345:235–247, 2005.

[BDJ08] Alexis Ballier, Bruno Durand, and Emmanuel Jeandal. Structural aspects of tilings. In *25th International Symposium on Theoretical Aspects of Computer Science (STACS 2008)*, volume 1, pages 61–72. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2008.

[Ber72] Robert Berger. The undecidability of the domino problem. *Memoirs of the AMS*, 66:1–72, 1972.

[BLR88] Mike Boyle, Douglas Lind, and Daniel Rudolph. The automorphism group of a shift of finite type. *Transactions of the AMS*, 306:71–114, 1988.

[BM97] F. Blanchard and A. Maass. Dynamical properties of expansive one-sided cellular automata. *Israel journal of mathematics*, 99:149–174, 1997.

[BPS10] Mike Boyle, Ronnie Pavlov, and Miachael H. Scraudner. Multidimensional sofic shifts without separation and their factors. *Transactions of the AMS*, to appear, 2010.

[BS94] Robert Burton and Jeffrey E. Steif. Nonuniqueness of measures of maximal entropy for subshifts of finite type. *Ergodic Theory and Dynamical Systems*, 14:213–235, 1994.

[BS09] Mike Boyle and Michael Schraudner. $\mathbb{Z}^d$ shifts of finite type without equal entropy full shift factors. *Journal of Difference Equations and Applications*, 15:47–52, 2009.

[BT00] F. Blanchard and P. Tisseur. Some properties of cellular automata with equicontinuity points. *Probabilites et Statistiques*, 36:569–582, 2000.

[BV00] V. Berhte and L. Vuillon. Tilings and rotations on the torus: a two-dimensional generalization of sturmian sequences. *Discrete Mathematics*, 233:27–53, 2000.

[CD04] Julien Cervelle and Bruno Durand. Tilings: recursivity and regularity. *Theoretical Computer Science*, 310:469–477, 2004.

[CP75] Ethan M. Coven and Michael E. Paul. Sofic systems. *Israel Journal of Mathematics*, 20:165–177, 1975.

[Des05] Angela Desai. Subsystem entropy for $\mathbb{Z}^d$ sofic shifts. *Indagationes Mathematicae*, 17:353–359, 2005.

[Des09] Angela Desai. A class of $\mathbb{Z}^d$ shifts of finite type which factors onto lower entropy full shifts. *Proceedings of the AMS*, 137:2613–2621, 2009.

[DMM03] Michele D'amico, Giovanni Manzini, and Luciano Margara. On computing the entropy of cellular automata. *Theoretical Computer Science*, 290:1629–1646, 2003.

[Dur93] Bruno Durand. Undecidability of the surjectivity problem for 2-D cellular automata: A simplified proof. *Lecture Notes on Computer Science, Fundamentals of Computation Theory*, 710:204–211, 1993.

[Dur99] Bruno Durand. Tilings and quasiperiodicity. *Theoretical Computer Science*, 221:61–75, 1999.

[EKM03] C. Epifanio, M. Koskas, and F. Mignosi. On a conjecture on bidimensional words. *Theoretical Computer Science*, 299:123–150, 2003.

[Hed69] Gustav A. Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Theory of Computing Systems*, 3:320–375, 1969.

[HKC92] Lyman P. Hurd, Jarkko Kari, and Karel Culik. The topological entropy of cellular automata is uncomputable. *Ergodic Theory and Dynamical Systems*, 12:255–265, 1992.

[HMU06] John E. Hopcroft, Rajeev Motvani, and Jeffrey D. Ullman. *Introduction to automata theory, languages and computation*. Addison-Wesley, 2006.

[Hoc09] Michael Hochman. On the dynamics and recursive properties of multidimensional symbolic systems. *Inventiones Mathematicae*, 176:131–167, 2009.

[JM05] Aimee Johnson and Kathleen Madden. Factoring higher-dimensional shifts of finite type onto the full shift. *Ergodic theory and dynamical systems*, 25:811–822, 2005.

[Kar94] Jarkko Kari. Reversibility and surjectivity problems of cellular automata. *J. Comput. Syst. Sci.*, 48(1):149–182, 1994.

[Kar96] Jarkko Kari. Representation of reversible cellular automata with block permutations. *Therory of Computing Systems*, 29:47–61, 1996.

[Kar05] Jarkko Kari. Theory of cellular automata: a survey. *Theoretical computer science*, 334:3–33, 2005.

[KR91] K.H. Kim and F.W. Roush. On the automorphism groups of subshifts. *Pure Math Appl Ser. B*, 1:203–230, 1991.

[Kur66] Kazimierz Kuratowski. *Topology- Volume 1*. Academic Press, 1966.

[Kur97] Petr Kurka. Languages, equicontinuity and attractors in cellular automata. *Ergodic Theory and Dynamical Systems*, 17:417–433, 1997.

[Kur03] Petr Kurka. *Topological and symbolic dynamics*. Cours Specialises 11. Societe Mathematique de France, 2003.

[LL04] E.L. Lakshtanov and E.S. Langvagen. Criterion of infinite topological entropy for multidimensional cellular automata. *Problems of Information Transmition*, 40:165–167, 2004.

[LL05] E.L. Lakshtanov and E.S. Langvagen. Entropy of multidimensional cellular automata. *Problems of Information Transmition*, 42:43–51, 2005.

[LM95] Douglas A. Lind and Brian Marcus. *Introduction to symbolic dynamics and coding*. Cambridge University Press, 1995.

[Lot02] M. Lothaire. *Algebraic combinatorics on words*. Cambridge University Press, 2002.

[Mey08] Tom Meyerovitch. Finite entropy for multidimensional cellular automata. *Ergodic Theory and Dynamical Systems*, 28:1243–1260, 2008.

[Moo62] E.F. Moore. Machine models of self-reproduction. *Proc. Sympos. Applied Math.*, 14:17–34, 1962.

[Moz89] Shahar Mozes. Tilings, substitution systems and dynamical systems generated by them. *Hebrew University Magnes Press*, 53:139–186, 1989.

[MS01] Ronald Meester and Jeffrey E. Steif. Higher-dimensional subshifts of finite type, factor maps and measures of maximal entropy. *Pacific journal of mathematics*, 2:497–510, 2001.

[Myh63] J. Myhill. The converse of Moore's Garden-of-Eden theorem. *Proceedings of AMS*, 14:685–686, 1963.

[Nas78] Masakazu Nasu. Local maps inducing surjective global maps of one-dimensional tesselation automata. *Math. Systems Theory*, 11:327–351, 1978.

[Nas02] Masakazu Nasu. The dynamics of expansive invertible onesided cellular automata. *Transactions of AMS*, 354:4067–4084, 2002.

[Par64] W. Parry. Intrinsic markov chains. *Transactions of the AMS*, 112:55–65, 1964.

[QZ04] Anthony Quas and Luca Zamboni. Periodicity and local complexity. *Theoretical computer science*, 319:229–240, 2004.

[Rob71] Raphael M. Robinson. Undecidability and nonperiodicity for tilings of the plane. *Inventiones Mathematicae*, 12:177–209, 1971.

[RS97] Grzegorz Rozenberg and Arto Salomaa. *Handbook of formal languages*. Springer, 1997.

[She93] Mark A. Shereshevsky. Expansiveness, entropy and polynomial growth for groups acting on subshifts by automorphisms. *Indagationes Mathematicae*, 4:203–210, 1993.

[ST00] J.W. Sander and R. Tijdeman. The complexity of functions on lattices. *Theoretical computer science*, 246:195–225, 2000.

[ST02] J.W. Sander and R. Tijdeman. The rectangle complexity of functions on two-dimensional lattices. *Theoretical Computer Science*, 270:857–863, 2002.

[ST08] Mathieu Sablik and Guillaume Theyssier. Topological dynamics of 2d cellular automata. *Lecture Notes in Computer Science*, 5028:523–532, 2008.

[Sut91] Klaus Sutner. De Bruijn graphs and linear cellular automata. *Complex Systems*, 5:19–30, 1991.

[vNB66] John von Neumann and Arthur W. Burks. *Theory of self-reproducing automata*. University of Illinois Press, 1966.

[Wal75] Peter Walters. *An introduction to ergodic theory*. Graduate texts in mathematics. Springer, 1975.

[Wan61] Hao Wang. Proving theorems by pattern recognition ii. *Bell System Technical Journal*, 40:1–41, 1961.

[War91] Thomas Ward. On the autmorphism groups of subshifts. *Pure Math Appl Ser. B*, 1:203–230, 1991.

[Wil91] Stephen J. Wilson. Decision procedures for openness and local injectivity. *Complex Systems*, 5:497–508, 1991.

TURKU

CENTRE *for*

COMPUTER

SCIENCE

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Information Technologies

Turku School of Economics

- Institute of Information Systems Sciences

ISBN 978-952-12-2434-8
ISSN 1239-1891