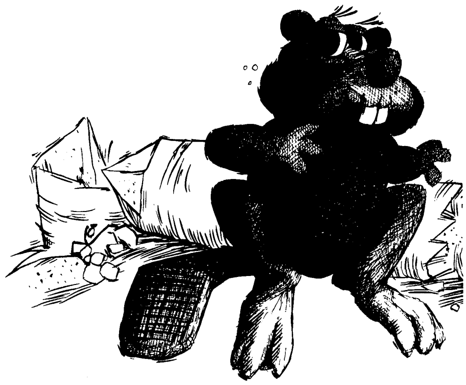
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



On the Limit Sets of Cellular Automata

Karel Culik II
Jan Pachl
Sheng Yu

Research Report
CS-87-47

August 10, 1987

On the Limit Sets of Cellular Automata *

Karel Culik II, Jan Pachl and Sheng Yu

Department of Computer Science
University of Waterloo
Waterloo, Ontario
Canada N2L 3G1

(August 10, 1987)

Abstract

The limit sets of cellular automata, defined by Wolfram, play an important role in applications of cellular automata to complex systems. We prove a number of results on limit sets, considering both finite and infinite configurations of cellular automata. We are mainly concerned with testing membership and (essential) emptiness of limit sets for linear and two-dimensional cellular automata. In our proofs we use results on finite recognizability of sets of biinfinite words and topological properties of product spaces.

*Research supported by the Natural Sciences and Engineering Research Council of Canada grants A7403 and A0952.

1 Introduction

Cellular automata (CA) are used to model complex natural systems containing large numbers of simple identical components with local interactions [12,13,14]. An important role in these models is played by the limit sets of CA. The limit set of a CA consists of those configurations that might occur after arbitrarily many computation steps of the automaton.

Mathematical investigation of CA limit sets has been initiated in [6]. In the present paper we prove the undecidability of the problem whether the limit set of a given kD CA is a singleton for $k \geq 2$. We improve the proof of the main result from [6], namely that the CA limit language is generally nonrecursive, and prove a number of additional results. The methodology of this work is of interest. Besides automata theoretical techniques we also consider, as Hurd did [6], the product topology on the space of configurations; we use the fact that this topological space is compact.

In Section 2 we define cellular automata and their limit sets. In Section 3 we endow the set of states of a CA cell with the discrete topology and observe that the space of all the configurations of a CA with the product topology is compact by Tychonoff's theorem. We then use compactness to prove several properties of the limit sets, including nonemptiness. We also use Baire's category theorem to derive a classification of cellular automata.

In the following section we combine topological techniques and the properties of regular sets of biinfinite words (see appendix) to prove more results on the limit sets of linear automata.

In Section 5 we use the undecidability of the tiling problem to prove that for $k \geq 2$ it is undecidable whether the limit set of a given k-dimensional cellular automaton is a singleton.

A configuration of a CA is called finite if all but finite (non zero) number of cells are quiescent. In Section 6 we study the limit sets of finite configurations, i.e. the intersection of the limit set with the set of finite configurations. Our main result here is that it is undecidable whether the limit set of finite configurations is empty even for linear CA. The set of all finite subwords of the configurations in a limit set is called a limit language. We show that the membership problem for CA limit languages, i.e. whether a given string is in the limit language of a given linear CA, is undecidable. Using the existence of a universal CA [1], we prove that there exists a CA whose limit language is not recursive. Similar results were proved by Hurd [6]. In section 6 it is explained how our results relate to those in [6].

There are simple regular languages that are not CA limit languages. On the other hand, since the complement of any CA limit language is recursively enumerable, a limit language is recursive if and only if it is recursively enumerable. Thus the result mentioned above implies that not all limit languages are recursively enumerable [6]. Moreover, there is no obvious effective translation between the description of a recursive limit set by its CA and the description by a Turing machine.

In the appendix we study biinfinite words, $\omega\omega$-finite automata and $\omega\omega$-regular sets; our definitions are slightly different from but equivalent to those in [9]. We prove that $\omega\omega$-regular sets are closed under $\omega\omega$-finite transduction.

2 Cellular automata, basic definitions

Let Z be the set of integers, and Z^k the set of k-tuples of integers. A *cellular automaton*, abbreviated CA (or, more specifically, a *k-dimensional cellular automaton*, kD CA) is an infinite array, indexed by Z^k, of *cells*. Each cell is identified by its *location* $I \in Z^k$.

At any time, each cell has a *state*, which belongs to a finite set S. The dynamic behavior of the CA is determined by a rule that describes the state of each cell at time $t+1$ as a function of the states of some neighboring cells at time t. The rule is invariant with respect to translations (shifts) of Z^k.

Formally, a cellular automaton is a quadruple $A = (k, S, N, f)$, where $k \geq 1$ is the dimension, S is the finite set of states, N is the *neighborhood*, and f is the *local function* of A. The dimension k is an integer, $k \geq 1$. The (relative) neighborhood N is a sequence $(I_1, I_2, \ldots, I_h)$ of relative locations $I_j \in Z^k$, $1 \leq j \leq h$. The local function is a total function $f : S^h \to S$.

A *configuration* c of the CA is a function $c : Z^k \to S$, which assigns a state in S to each cell of the CA. The set of configurations is denoted S^{Z^k}. The local function f is extended to the *global function*

$$G_f : S^{Z^k} \to S^{Z^k}$$

of the set of configurations into itself. By definition, for $c_1, c_2 \in S^{Z^k}$,

$$G_f(c_1) = c_2$$

if and only if

$$c_2(I) = f(c_1(I + I_1), c_1(I + I_2), \ldots, c_1(I + I_h))$$

for all $I \in Z^k$.

The function G_f describes the dynamic behavior of the CA: The CA moves from the configuration c at time t to the configuration $G_f(c)$ at time $t+1$. The state of the cell I at time $t+1$ depends only on the states of the cells in the neighborhood $(I+I_1, I+I_2, \ldots, I+I_h)$ at time t. Notice that besides being locally defined, the global function G_f is total and translation-invariant.

Example 1 Let $A = (1, S, N, f)$ be a linear CA, where $S = \{0, 1\}$, $N = (-2, -1, 0, 1, 2)$, and

$$f(x_1, x_2, x_3, x_4, x_5) = \begin{cases} 1 & \text{if } x_1 + x_2 + x_3 + x_4 + x_5 = 4; \\ 0 & \text{otherwise.} \end{cases}$$

If c is a configuration consisting of all 1's and c' is a configuration consisting of all 0's, then $G_f(c) = c'$. □

For $c \in S^{Z^k}$, the sequence $(c, G_f(c), G_f{}^2(c), G_f{}^3(c), \ldots)$ is called the *orbit of c*.

Frequently, a state $\tilde{q}$ with the property

$$f(\tilde{q}, \tilde{q}, \ldots, \tilde{q}) = \tilde{q}$$

is distinguished and called the *quiescent state*. In a CA, there may be more than one state with the above property, but at most one of them is distinguished as the quiescent state. The configuration with all cells in the quiescent state is called the *quiescent configuration*, denoted by $\tilde{Q}$.

Let $A = (k, S, N, f)$ be a CA. Define

$$\begin{aligned} \Omega^{(0)} &= S^{Z^k}, && \text{and} \\ \Omega^{(i)} &= G_f(\Omega^{(i-1)}) && \text{for } i \geq 1 . \end{aligned}$$

Then

$$\Omega = \bigcap_{i=0}^{\infty} \Omega^{(i)}$$

is called the *limit set* of A.

Define

$$\Phi = \{ c \in S^{Z^k} \mid G_f(c) = c \}$$

(this is the set of "fixed points" of G_f). Obviously $\Phi \subseteq \Omega$.

3 The product topology on configurations

The configuration space S^{Z^k} is a product of infinitely many finite sets S. When S is endowed with the discrete topology, the *product topology* on S^{Z^k} is compact by Tychonoff's theorem ([7], Theorem 5.13). A subbasis of open sets for the product topology consists of all sets of the form

$$\{ c \in S^{Z^k} \mid c(i) = a \} , \tag{1}$$

where $i \in Z^k$ and $a \in S$. A subset of S^{Z^k} is open if and only if it is a union of finite intersections of sets of the form (1). It is easy to show that the global function G_f defined in the previous section is continuous from S^{Z^k} to S^{Z^k}. (Thus the pair (S^{Z^k}, G_f) is a classical dynamical system, in the sense of [3].)

Theorem 1 *The limit set Ω is non-empty.*

Proof Since G_f is continuous, each $\Omega^{(i)}$, $i \geq 0$, is a continuous image of the compact space S^{Z^k}. Hence $\Omega^{(i)}$ are non-empty compact subsets of S^{Z^k}, and $\Omega^{(0)} \supseteq \Omega^{(1)} \supseteq \Omega^{(2)} \supseteq \ldots$. Therefore the intersection $\Omega = \bigcap_{i=0}^{\infty} \Omega^{(i)}$ is non-empty. □

The theorem has also an easy non-topological proof:

An alternative proof Let $c : Z^k \to S$ be a constant function (i.e. there is $a \in S$ such that $c(I) = a$ for all $I \in Z^k$). In the orbit $(c, c_1, c_2, \ldots)$ of c, each c_j is a constant function. Since the set S is finite, there are only finitely many constant functions from Z to S, and thus there exists m such that $c_m = c_j$ for infinitely many j. Hence $c_m \in \Omega^{(i)}$ for all $i \geq 0$, and $c_m \in \Omega$. $\square$

For some CA, the limit set Ω contains only one configuration. In particular, for a CA with a special quiescent state $\tilde{q}$, it is possible that the limit set contains only the quiescent configuration $\tilde{Q}$. It is an open question whether the problem $\Omega \stackrel{?}{=} \{\tilde{Q}\}$ is decidable for $k = 1$.

Now we are going to use Baire's category theorem to classify cellular automata by the limit behavior of their orbits. A subset of a topological space is called a G_δ set if it is the intersection of a countable family of open sets.

Theorem 2 *Let C be a closed translation-invariant subset of S^{Z^k}. Exactly one of these two conditions is true:*

(i) There exists an integer $i \geq 0$ such that $G_f{}^i(S^{Z^k}) \subseteq C$.

(ii) There exists a dense G_δ set $D \subseteq S^{Z^k}$ such that

$$C \cap \bigcup_{i=0}^{\infty} G_f{}^i(D) = \emptyset.$$

Proof For $i = 0, 1, 2, \ldots$, let

$$F_i = \{\, c \in S^{Z^k} \mid G_f{}^i(c) \in C \,\}.$$

The sets F_i are closed and translation-invariant. Let

$$D = S^{Z^k} - \bigcup_{i=0}^{\infty} F_i \ .$$

Thus D is a G_δ set in S^{Z^k} and

$$C \cap \bigcup_{i=0}^{\infty} {G_f}^i(D) = \emptyset.$$

If D is dense in S^{Z^k} then condition (ii) holds.

If D is not dense then $\bigcup_{i=0}^{\infty} F_i$ contains a nonempty open subset E of S^{Z^k}. The set E is locally compact; therefore, by Baire's theorem ([7], Theorem 6.34), there exists i such that F_i contains a nonempty open subset of E, which is an open subset of S^{Z^k}. Since F_i is translation-invariant, it follows that F_i contains a nonempty open *translation-invariant* subset of S^{Z^k}. However, every nonempty open translation-invariant subset of S^{Z^k} is dense in S^{Z^k}. Since F_i is closed, it follows that $F_i = S^{Z^k}$, and therefore (i) holds. □

In the notation of section 2, $\Omega^{(i)} = {G_f}^i(S^{Z^k})$. The set $\{\tilde{Q}\}$ (the singleton set containing only the quiescent configuration) and the limit set Ω are closed and translation-invariant. Thus we obtain two corollaries:

Corollary 1 *If $\Omega \neq \{\tilde{Q}\}$ then there exists a configuration whose orbit does not contain $\tilde{Q}$.*

Proof If there exists i such that $\Omega^{(i)} \subseteq \{\tilde{Q}\}$ then $\Omega = \{\tilde{Q}\}$. Therefore, by Theorem 2, if $\Omega \neq \{\tilde{Q}\}$ then condition (ii) holds with $C = \{\tilde{Q}\}$. If $c \in D$ then the orbit of c does not meet $\{\tilde{Q}\}$. □

Corollary 2 *For each CA, exactly one of these two conditions is true:*

(i) There exists an integer $i \geq 0$ such that $\Omega^{(i)} = \Omega$.

(ii) There exists a dense G_δ set $D \subseteq S^{Z^k}$ such that

$$\Omega \cap \bigcup_{i=0}^{\infty} {G_f}^i(D) = \emptyset.$$

□

It is easy to find CA satisfying condition (i) in Corollary 2. For instance, (i) holds whenever G_f is surjective (because in that case $\Omega = \Omega^{(i)} = S^{Z^k}$ for every i). On the other hand, the CA in the following example does not satisfy (i) (and therefore it satisfies (ii)).

Example 2 Let $A = (1, S, N, f)$ be a linear CA such that $S = \{0, 1\}$, $N = (-1, 0, 1)$, and

$$f(a_{-1}, a_0, a_1) = \begin{cases} 1 & \text{if } a_{-1} = a_0 = a_1 = 1; \\ 0 & \text{otherwise.} \end{cases}$$

In this example,

$$\Omega = \{\, {}^{\omega}1^{\omega} \,\} \cup \{\, {}^{\omega}01^{n}0^{\omega} \mid n = 0, 1, 2, \dots \}.$$

However,

$$G_f{}^{i}({}^{\omega}01^{2i+1}01^{2i+1}0^{\omega}) = {}^{\omega}010^{2i+1}10^{\omega}$$

and therefore $\Omega^{(i)} \neq \Omega$ for every i, which means that condition (i) in Corollary 2 does not hold. □

Now we prove that condition (ii) in Corollary 2 never holds when $\Omega = \{\tilde{Q}\}$.

Theorem 3 $\Omega = \{\tilde{Q}\}$ *if and only if there exists an integer* $i \geq 0$ *such that* $\Omega^{(i)} = \{\tilde{Q}\}$.

Proof The *if* part is trivially true. To prove the *only if* part, assume that $\Omega = \{\tilde{Q}\}$. Choose one cell $I_0 \in Z^k$, and define

$$C = \{\, c \in S^{Z^k} \mid c(I_0) \neq \tilde{q} \,\}\,.$$

Then C is a closed set and

$$\bigcap_{i=0}^{\infty}\left(\Omega^{(i)}\cap C\right)=\{\tilde{Q}\}\cap C=\emptyset\,.$$

By compactness, $\Omega^{(i)}\cap C=\emptyset$ for some i. Since $\Omega^{(i)}$ is translation-invariant,

$$\Omega^{(i)}\cap\{\,c\in S^{Z^k}\;\mid c(I)\neq\tilde{q}\,\}=\emptyset$$

for every $I\in Z^k$. Hence $\Omega^{(i)}=\{\tilde{Q}\}$. □

Theorem 3 yields a semi-procedure for demonstrating that Ω contains only the quiescent configuration. To define the semi-procedure, we extend the global function G_f to operate on partial configurations: If $W\subseteq Z^k$, define

$$N^{-1}(W)=\{\,I\in Z^k\mid I+I_j\in W\;\text{ for }\;1\leq j\leq h\,\}$$

and for a function $c_1:W\to S$ define

$$G_f(c_1)=c_2$$

where $c_2:N^{-1}(W)\to S$ is such that

$$c_2(I)=f(c_1(I+I_1),c_1(I+I_2),\ldots,c_1(I+I_h))$$

for all $I\in N^{-1}(W)$.

For $r\geq 0$, define the k-dimensional interval W_r to be the product of one-dimensional intervals $[-r,r]$; that is,

$$W_r=\{\,(i_1,i_2,\ldots,i_k)\in Z^k\mid -r\leq i_j\leq r\;\text{ for }\;j=1,2,\ldots,k\,\}$$

Denote by I_0 the origin in Z^k, i.e. the k-tuple of zeros.

Corollary 3 *Let $r\geq 0$ be such that $N\subseteq W_r$. Then $\Omega=\{\tilde{Q}\}$ if and only if there exists an integer $i\geq 0$ such that for every function $c:W_{ir}\to S$ the function ${G_f}^i(c)$ maps I_0 to $\tilde{q}$.*

Proof The corollary follows from the theorem and from this observation, which can be proved by induction in i: If $c : W_{ir} \to S$ and $c' : Z^k \to S$ agree on W_{ir}, then ${G_f}^i(c)$ and ${G_f}^i(c')$ agree at I_0. □

The following semi-procedure determines that $\Omega = \{\tilde{Q}\}$: Let $A = (k, S, N, f)$ be the given CA. Find r such that $N \subseteq W_r$. For $i = 1, 2, \ldots$, generate all functions $c : W_{ir} \to S$, and for each such c compute the value of ${G_f}^i(c)$ at I_0. Stop when, for some i, all the values are $\tilde{q}$. This is only a semi-procedure because it never halts when $\Omega \neq \{\tilde{Q}\}$.

We conjecture that the problem $\Omega \stackrel{?}{=} \{\tilde{Q}\}$ is decidable for *linear* CA. In section 5 we show that the problem is undecidable for 2D CA, and therefore also for kD CA when $k \geq 2$. In the remainder of this section we show that the same problem for Φ (the set of fixed points of G_f, defined at the end of section 2) is decidable for linear CA, although (as will be proved in Theorem 9) it is undecidable for dimensions $k \geq 2$.

A configuration $c : Z \to S$ is called *periodic* if there is $m > 0$ (called a *period of c*) such that $c(j + m) = c(j)$ for every $j \in Z$.

Lemma 1 *For a 1-dimensional CA $(1, S, N, f)$, let $r > 0$ be such that $N \subseteq [-r, r]$, and let n be the cardinality of S. If $\Phi \neq \{\tilde{Q}\}$ then there exists a periodic configuration $c \in \Phi$, $c \neq \tilde{Q}$, with period at most n^{2r+1}.*

Proof There are n^{2r+1} different partial configurations $d : [-r, r] \to S$.

Choose any $c' \in \Phi - \{\tilde{Q}\}$. Among the restrictions of c' to the intervals $[j, j + 2r]$, $j = 0, \ldots, n^{2r+1}$, at least two must be identical (modulo a shift). Suppose that the two are the restrictions of c' to the intervals $[j_1, j_1 + 2r]$ and $[j_2, j_2 + 2r]$, $0 \leq j_1 < j_2 \leq n^{2r+1}$. Define c to be the (unique) configuration that is equal to c' on the interval $[j_1, j_2 + 2r]$ and is periodical with the period $j_2 - j_1$. □

Theorem 4 *The problem $\Phi \stackrel{?}{=} \{\tilde{Q}\}$ is decidable for $k = 1$.*

Proof In view of Lemma 1, the following algorithm decides whether $\Phi = \{\tilde{Q}\}$: Generate all partial configurations $c : [-r, n^{2r+1}+r] \to S$, and for each such c check whether there exists an integer m, $0 < m \leq n^{2r+1}$, such that $G_f(c)(j) = c(j)$ for $0 \leq j \leq n^{2r+1}$, $c(j) = c(j+m)$ for $-r \leq j \leq r$, and $c(0) \neq \tilde{q}$. If there is at least one c for which the test is positive then $\Phi \neq \{\tilde{Q}\}$. Otherwise $\Phi = \{\tilde{Q}\}$. □

4 The limit sets of linear cellular automata

In this section we assume that $A = (1, S, N, f)$; that is, A is a linear CA.

We treat a configuration of A as a biinfinite word over the alphabet S. With every set of configurations we associate a set of finite words (strings) over S, as follows: For a biinfinite word $c \in S^Z$, define (as in [6])

$$L[c] = \{\ w \in S^* \mid w \text{ is a finite subword of } c\},$$

and, for $C \subseteq S^Z$, define

$$L[C] = \bigcup_{c \in C} L[c].$$

$L[C]$ is called the *language of* C. If Ω is the limit set of the CA A, then we call $L[\Omega]$ the *limit language of* A.

The next theorem gives an alternative definition of the limit language $L[\Omega]$.

Theorem 5 $$L[\Omega] = \textstyle\bigcap_{i=0}^{\infty} L[\Omega^{(i)}].$$

Proof Since $\Omega \subseteq \Omega^{(i)}$ for every i, it follows that $L[\Omega] \subseteq \bigcap_i L[\Omega^{(i)}]$. To prove the opposite inclusion, choose any $w \in \bigcap_i L[\Omega^{(i)}]$. Let j be the length of w.

Define

$$C = \{ c \in S^{Z} \mid c(1)c(2)\ldots c(j) = w \}.$$

Then C is a closed set and $C \cap \Omega^{(i)} \neq \emptyset$ for every i, by the choice of w and the translation-invariance of $\Omega^{(i)}$. By compactness,

$$\Omega \cap C = \bigcap_{i=0}^{\infty} \left(\Omega^{(i)} \cap C\right) \neq \emptyset\,,$$

which means that w is a subword of some $c \in \Omega$, hence $w \in L[\Omega]$. □

Wolfram [13] shows that the set $L[\Omega^{(i)}]$ is regular for each $i \geq 0$. This result can be proved using the fact that regular sets are closed under GSM mappings [5]. Indeed for each local function f it is easy to construct the GSM T_f that maps each word w in $L[\Omega^{(i)}]$ with $|w| \geq r$, where r is the span of the neighborhood, into the successor string of length $|w| - r$ in $L[\Omega^{(i+1)}]$. The set $L[\Omega^{(0)}] = S^*$ is regular, therefore for each $i > 0$, $L[\Omega^{(i+1)}] = T_f(L[\Omega^{(i)}])$ is regular as well. We omit the details of this proof. However, we obtain the same result again in Corollary 4, using a more general approach.

A natural extension of finite automata (FA), called $\omega\omega$-FA, is described in the appendix. The set of biinfinite words recognized by an $\omega\omega$-FA M is denoted $B(M)$; every set of this form is called an *$\omega\omega$-regular set.* Every $\omega\omega$-regular set is translation-invariant. The family of $\omega\omega$-regular sets is closed under $\omega\omega$-finite transductions, union, and intersection. Details are given in the appendix.

Example 3 Let X be the set of all the biinfinite words over $\{a, b\}$ that have a prime number of a's. Let Y be the set of all biinfinite words over $\{a, b\}$. Then $L[X] = L[Y] = \{a, b\}^*$. □

In the above example, X and Y have the same set of finite subwords although $X \neq Y$. Clearly, Y can be recognized by an $\omega\omega$-FA, but X cannot. The example shows that the finite subwords of biinfinite words do not always capture the characteristics of the biinfinite words themselves. This suggests that it is useful to study directly the properties of biinfinite words as well as their relations with finite subwords.

It can be verified that the global function G_f of a CA is a $\omega\omega$-finite transduction on biinfinite words. From Theorem A.1 of the appendix, which states that $\omega\omega$-regular sets are closed under $\omega\omega$-finite transductions, we obtain the following theorem which formalizes comments made in Section 2 of [13].

Theorem 6 *For any integer $i \geq 0$, $\Omega^{(i)}$ is $\omega\omega$-regular.* □

In view of Theorem A.2 in the appendix, the following is a direct consequence of Theorem 6.

Corollary 4 *For any integer $i \geq 0$, $L[\Omega^{(i)}]$ is regular.* □

Although the sets $L[\Omega^{(i)}]$ are regular, Hurd [6] shows that $L[\Omega]$ can be non-regular. As a corollary of the next theorem, we shall show that the set Ω is $\omega\omega$-regular if and only if the set $L[\Omega]$ is regular.

Theorem 7 *If $C \subseteq S^Z$ is translation-invariant then the set $\{ c \in S^Z \mid L[c] \subseteq L[C] \}$ is the closure of C in the product topology.*

Proof Let $D = \{ c \in S^Z \mid L[c] \subseteq L[C] \}$. The complement of D in S^Z is open. Indeed, if $c' \notin D$ then $c'(i) \ldots c'(j) \notin L[C]$ for some $i, j \in Z, i \leq j$. In that case the set

$$\{ c \in S^Z \mid c(i) \ldots c(j) = c'(i) \ldots c'(j) \},$$

which is a neighborhood of c' in the product topology, does not intersect D.

Since D is closed and $C \subseteq D$, it follows that the closure $\bar{C}$ of C is a subset of D. To prove that $D \subseteq \bar{C}$, choose any $d \in D$. Then for every $j \geq 0$ the word $d(-j)\ldots d(j)$ is a subword of some $c_j \in C$. Since C is translation invariant, we can choose c_j so that $d(-j)\ldots d(j) = c_j(-j)\ldots c_j(j)$. But then d is the limit of the sequence $(c_j | j = 0, 1, \ldots)$ in the product topology, which proves that $d \in \bar{C}$. □

Corollary 5 *Let $C \subseteq S^Z$ be a translation-invariant closed set. Then C is $\omega\omega$-regular if and only if $L[C]$ is regular.*

Proof Apply Theorems 7, A.2 and A.3. □

Corollary 6 *A configuration $c \in S^Z$ belongs to the limit set Ω if and only if $L[c] \subseteq L[\Omega]$.*

Proof By Theorem 7, $\Omega = \{ c \in S^Z \mid L[c] \subseteq L[\Omega] \}$. □

5 The limit sets of 2D cellular automata

In contrast to the conjecture we made for linear CA in Section 3, we are now going to show that it is undecidable whether or not the limit set of a given 2D CA consists of the quiescent configuration only. Consequently, the same problem for kD CA limit sets is undecidable for any $k \geq 2$. The proof is based on a well-known deep result, the undecidability of the tiling problem. In the following, we first give a brief description of the tiling problem and then prove our results. Readers who are interested in the details of the tiling problem are referred to [10].

We are given a set of tiles with colored edges. The tiles are squares of equal size. In the set, there are finitely many different types of squares and there are infinitely many squares for each type. The tiles are to be used to tile the entire plane, without rotating any tiles. In a valid tiling, each pair of abutting edges have the same color. The tiling problem is to decide whether or not a given set of squares can tile the entire plane.

The tiling problem was raised by Wang [11], and proved to be undecidable five years later by Berger [2]. Robinson gave a very readable proof in [10].

Theorem 8 *It is recursively undecidable whether or not the limit set Ω of a given 2D CA consists of $\tilde{Q}$ only.*

Idea of the proof Each type of tiles in the tiling problem corresponds to a state in a 2D CA. The state set of the CA consists of the states representing all types of squares, and a quiescent state. The neighborhood of a cell contains the cell itself and its four neighboring cells. At each computation step, every cell checks if it has correct neighbors in the sense of tiling. If it does, it remains in the same state. If it doesn't, it changes to the quiescent state. Moreover, the quiescent state spreads to its neighbors. The given tiles admit a valid tiling of the plane if and only if the limit set of the CA does not consist of the quiescent configuration only. Therefore, the undecidability of our problem is implied by the undecidability of the tiling problem.

Proof of Theorem 8 We show that the tiling problem can be transformed into our problem. We are given a set of tiles,

$$T = \{ \, (l_i, r_i, u_i, d_i) \mid 1 \leq i \leq n \, \},$$

where l_i, r_i, u_i, and d_i denote the colors of the left, right, upper, and lower edges, respectively. In the following, we use $l(t), r(t), u(t)$, and $d(t)$ to denote the colors of the four edges of a tile t; that is, $t = (l(t), r(t), u(t), d(t))$. We construct a CA $A = (2, Q, N, f)$ where

$Q = T \cup \{\tilde{Q}\}$;

$N = ((0,0), (-1,0), (1,0), (0,1), (0,-1))$;

and

$$f(t_o, t_l, t_r, t_u, t_d) = \begin{cases} t_o & \text{if } l(t_o) = r(t_l), r(t_o) = l(t_r), \\ & \quad u(t_o) = d(t_u), \text{ and } d(t_o) = u(t_d); \\ \tilde{q} & \text{otherwise.} \end{cases}$$

Now we show that there is a valid tiling of the plane if and only if the limit set of A is not $\{\tilde{Q}\}$, where $\tilde{Q}$ denotes the quiescent configuration. If the plane can be tiled with the given tiles, then all the valid tilings are configurations in the limit set of A. If the plane cannot be tiled with the given tiles, then there is an integer $i \geq 1$ such that the square of size i cannot be tiled (this follows from König's infinity lemma – cf. pp. 381-383 in [8]). By the definition of f, $G_f{}^i(c) = \tilde{Q}$ for all $c \in S^{Z^2}$. This implies that $\Omega^{(i)} = \{\tilde{Q}\}$, and the limit set Ω is equal to $\{\tilde{Q}\}$.

Since the tiling problem is undecidable, the problem $\Omega \stackrel{?}{=} \{\tilde{Q}\}$ for 2D CA is also undecidable. □

Corollary 7 *It is recursively undecidable whether or not the limit set of a given kD CA consists of the quiescent configuration only, for any $k \geq 2$.*

Proof Define the local function such that only two dimensions are actually effective. A cell remains in the same non-quiescent state if its four neighbors in two specific dimensions satisfy the rule of tiling. □

The same technique does not work for linear CA because the tiling problem in one dimension is trivially decidable.

Observe that, for the given set of tiles and the cellular automaton constructed in the proof of Theorem 8, there is a valid tiling of the plane if and only if the set $\Phi \subseteq S^{Z^k}$ (defined at the end of section 3) contains some configuration different from the quiescent configuration $\tilde{Q}$. Thus the proof of Theorem 8 also proves the following result.

Theorem 9 *For $k \geq 2$ it is recursively undecidable whether $\Phi = \{\tilde{Q}\}$.* □

6 Limit sets of finite configurations

A configuration is *finite* if the number of nonquiescent cells is finite but not zero. Let $\mathcal{F}$ denote the set of all finite configurations of a CA A. We define the *limit set of finite configurations* of A as

$$\Omega_F = \Omega \cap \mathcal{F}.$$

In this section, we show that, given an arbitrary CA, it is undecidable whether Ω_F is empty. The difficulty in transforming the Turing machine halting problem into this problem is that CA do not distinguish input symbols from working symbols. Note also that the limit set of finite configurations may be nonempty even if every finite configuration eventually becomes quiescent.

Example 4 Let $A = (1, S, N, f)$ be the linear CA defined in Example 2.

That is, $S = \{0, 1\}$, $N = (-1, 0, 1)$, and

$$f(a_{-1}, a_0, a_1) = \begin{cases} 1 & \text{if } a_{-1} = a_0 = a_1 = 1; \\ 0 & \text{otherwise.} \end{cases}$$

In this example, either "0" or "1" can be distinguished as the quiescent state. If "0" is the quiescent state, then the limit set of finite configurations is the set of all configurations that have exactly one substring of the form "011..10". If "1" is the quiescent state, then the limit set of finite configurations is empty. □

Theorem 10 *Given a CA, it is undecidable whether $\Omega_F = \emptyset$.*

Proof In this proof we consider linear CA. The result can be easily extended to multidimensional CA.

We prove the undecidability of the problem by reduction from the halting problem for Turing machines on the blank tape. For any Turing machine M we construct a CA A such that the limit set of A contains a finite configuration if and only if M (starting with the blank tape) never halts.

Given a Turing machine M operating on a one-way infinite tape, we construct a CA $A = (1, S, N, f)$ with $N = (-1, 0, 1)$ as follows. Besides the quiescent state $\tilde{q}$, S consists of *blue states, a yellow state*, and *auxiliary states*. Each blue state has two components. The first component contains a tape symbol of M. The second component contains a marker which shows whether the cell is to the left or to the right of the head of M, or directly under the head; in the last case the marker encodes also the state of M. Any instantaneous description of M is encoded as a sequence of blue states (in fact, any instantaneous description of M is encoded by infinitely many sequences of blue states — they differ from each other in the number of trailing blanks). The set of auxiliary states consists of the left boundary state l, the right boundary states r_0 and r_1, the left shift state s_l, the right shift state s_r, and the destroyer state d.

We will simulate computations of M by evolutions of A on so called *valid segments*. A *segment* is a finite nonempty sequence of nonquiescent states, and a segment is *valid* if it is of the form

$$l\ b_1\ b_2\ \dots\ b_m\ s\ y^n\ r$$

where $m \geq 0$, $n \geq 0$, and

(i) $b_1, b_2, \dots, b_m$ is a sequence of blue states which is a prefix of an encoding of an instantaneous description of M;

(ii) y is the yellow state;

(iii) $s \in \{s_l, s_r\}$.

(iv) $r \in \{r_0, r_1\}$.

The validity of a segment can be checked locally in one step: One can define *valid neighborhoods* so that a segment is valid if and only if each neighborhood intersecting the segment is valid.

The function f in $A = (1, S, N, f)$ is defined in accordance with the following principles:

(a) The destroyer state d spreads at the full speed (one cell at a time) in both directions.

(b) Each invalid neighborhood generates the destroyer state d.

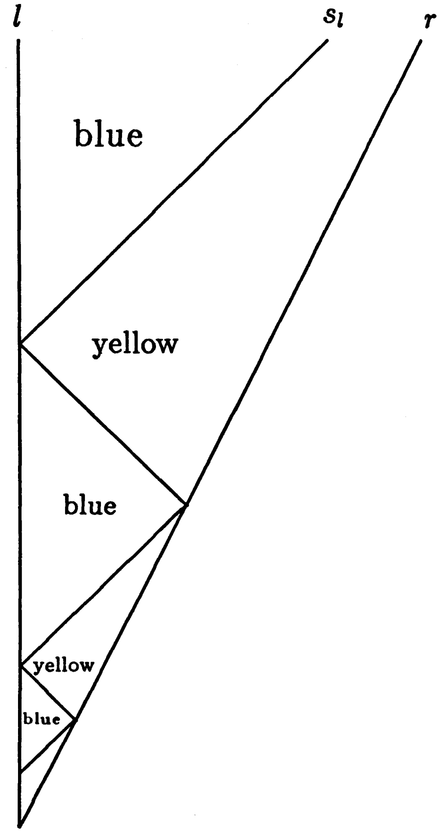


Figure 1: A computation of CA A

(c) As long as a valid segment

$$\alpha = l\ b_1\ b_2\ \ldots\ b_m\ s\ y\ y\ \ldots\ y\ r$$

does not encounter spreading destroyers, it evolves as follows (see Figure 1, where time increases from top to bottom):

- If the blue states of α encode a non-halting instantaneous description of M then the blue states of its successor β encode the next instantaneous description in the computation of M. If M halts, then a destroyer d is produced.

- In β, the left boundary l is at the same location as in α. If the right boundary $r = r_0$, then it changes to r_1 and stays at the same location; if $r = r_1$, then it changes to r_0 and moves one cell to the left. In the second case, the rightmost yellow state disappears.
- If $s = s_l$ ($s = s_r$) then s moves one cell to the left (right) provided its left (right) neighbor is not l (r). If s_l (s_r) meets l (r), then s_l (s_r) changes to s_r (s_l) and starts to move to the right (left). The s_l/s_r state continues crossing between the two boundaries l and r until it finally vanishes when l and r meet, as shown in Figure 1. Depending on the value of s ($s = s_l$ or $s = s_r$), the state b_m (if $s = s_l$) or the leftmost yellow state y (if $s = s_r$) disappears.
- Whenever s_r leaves l and a blue state appears between them, this blue state encodes the initial instantaneous description of M with the blank tape.
- When l and r meet (as at the bottom of Figure 1), both are replaced by $\tilde{q}$.

We are going to prove that the limit set of A contains a finite configuration if and only if M starting on the blank tape never halts.

Observe that if M never halts when started on the blank tape, then for every instantaneous description of M that occurs in some computation of M (starting on the blank tape) there exists a configuration $c \in \Omega_F$ such that c contains exactly one valid segment and the blue states of the segment encode the instantaneous description. In particular, if M does not halt then $\Omega_F \neq \emptyset$.

Thus it remains to be proved that if M halts then $\Omega_F = \emptyset$. First we prove several claims.

Claim 1 If M halts, then no finite configuration that contains a valid segment is in the limit set.

Proof : For $n \geq 0$, let $\alpha_n = l\, s_r\, y^n\, r_0$ be a valid segment. By the rules for f, a valid segment can only evolve from a valid segment, and if a valid segment β is contained in a configuration in Ω_F then there are infinitely many values of n for which α_n evolves into β. If M, when started on the blank tape, halts in t steps and $n > 3(t+1)/2$, then in $t+1$ steps α_n evolves into the valid segment $l\, b_1\, b_2\, \ldots\, b_{t+1}\, s_r\, y^{n-(t+1)-\lfloor t/2 \rfloor}\, r$ whose blue states encode the halting instantaneous description. Hence in $t+2$ steps α_n evolves into a segment which contains at least one d. Therefore for each valid segment β there are only finitely many n for which α_n evolves into β. It follows that no configuration that contains a valid segment is in Ω_F.

Claim 2 If a finite configuration c contains an invalid neighborhood and does not contain any destroyer state d, then there is no c' such that $c = G_f(c')$.

Proof : If $c = G_f(c')$ and c' contains no invalid neighborhood and no encoding of halting instantaneous description then c contains no invalid neighborhood; but if c' contains an invalid neighborhood or an encoding of halting instantaneous description then c contains d.

Claim 3 If a finite configuration c contains an invalid neighborhood, then $c \notin \Omega_F$.

Proof : Since the result of Claim 2, it suffices to consider the case when c has destroys (d's). By the rules for f, if $c = G_f(c')$ then the number of

d's in c' is smaller than the number of d's in c. Therefore there exists an integer $n \geq 0$ such that every configuration in $G_f^{-n}(c)$ contains an invalid neighborhood and no d. By Claim 2, $G_f^{-(n+1)}(c) = \emptyset$.

By Claims 1 and 3, if M halts then $\Omega_F = \emptyset$. □

The following theorem was proved by Hurd (Theorem 4 in [6]).

Theorem 11 *Given a CA $A = (1, S, N, f)$ and a string $w \in S^*$, it is undecidable whether w is in the limit language of A.*

Proof: For the cellular automaton A constructed in the proof of Theorem 10, the string lr is in the limit language if and only if $\Omega_F \neq \emptyset$. □

Our next result (Corollary 8) has been stated in [6] as a direct consequence of a theorem equivalent to our Theorem 11. However, we feel that Corollary 8 does not *immediately* follow from Theorem 11. In order to find a nonrecursive limit language, one must show that for *one particular* CA A it is undecidable whether a given string is in the limit language of A.

Corollary 8 *There exists a linear cellular automaton such that its limit language is not recursive.*

Proof The collolary follows from Theorem 11 with the help of a universal CA. A universal CA is given in [1], where any CA is simulated by encoding its local function in the states of the universal CA. Given a CA A and a string w of A, there is a string w' of the universal CA such that w' encodes both A and w. Now, the problem of whether w is in the limit language of A is transformed to the problem of whether w' (the encoding of A and w) is in the limit language of the universal CA. Since the former is undecidable, the latter is undecidable, too. □

Appendix

A Biinfinite Words and $\omega\omega$-Regular Sets

Biinfinite words and their regularity (finite recognizability) have been studied by Nivat and Perrin [9]. Here we use slightly different but equivalent definitions which are more convenient for our purpose.

A biinfinite word c is a mapping $Z \to S$, where Z is the set of all integers and S is a finite alphabet. The symbol $c(j)$, $j \in Z$, denotes the j-th letter of c.

An *$\omega\omega$-finite automaton* ($\omega\omega$-FA) M is a quintuple (Q, S, δ, Q_L, Q_R),

where

Q is the finite set of state;

S is the input alphabet;

δ is the transition function;

$Q_L \subseteq Q$ is the set of left (accepting) states; and

$Q_R \subseteq Q$ is the set of right (accepting) states.

A biinfinite word c is said to be recognized by M if there is a mapping $Z \to Q$, i.e. a biinfinite sequence of states

$$\cdots q_{-2}, q_{-1}, q_0, q_1, q_2, \cdots$$

such that, for all $j \in Z$,

(1) $\delta(q_j, c(j)) = q_{j+1}$; and

(2) there exist $m, n \in Z$, $m \le j \le n$, such that $q_m \in Q_L$ and $q_n \in Q_R$.

In other words, c is said to be recognized by M if there is a biinfinite computation of M on input c such that there is a left state appearing arbitrarily early, and there is a right state appearing arbitrarily late in the computation. Such a computation will be called an *accepting computation.*

The set of biinfinite words recognized by M is denoted $B(M)$. The sets of the form $B(M)$ for some $\omega\omega$-FA M are called *$\omega\omega$-regular.*

If a biinfinite word c is recognized by an $\omega\omega$-FA, then all the translations of c are recognized by the same $\omega\omega$-FA. Therefore every $\omega\omega$-regular set is translation-invariant.

Example A.1 Let $M = (Q, S, \delta, Q_L, Q_R)$ be an $\omega\omega$-FA, where $Q = \{0,1\}$, $S = \{a,b\}$, $Q_L = \{0\}$, $Q_R = \{1\}$, and δ is given in Figure 2.

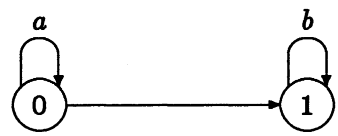


Figure 2:

The set of biinfinite words recognized by M is the set of all words which have infinitely many a's followed by infinitely many b's, i.e. $\{{}^{\omega}a\, b^{\omega}\}$.

□

A finite automaton operating on finite or one-way infinite words is a special case of an $\omega\omega$-FA, in the following sense. With a special blank symbol, a one-way infinite word is represented as a biinfinite word with infinitely many blanks on the left end. A finite word is represented as a biinfinite word with infinitely many blanks added at both ends. Then for every finite automaton there is an $\omega\omega$-FA that recognizes the same set of finite and one-way infinite words.

In an $\omega\omega$-FA, a left (right) state that is not in a cycle can be changed into a non-left (non-right) state without affecting the set of biinfinite words recognized by the $\omega\omega$-FA. A state which cannot be reached from any left state or from which no right state can be reached is useless – it does not contribute to the recognition of any biinfinite word. We say that an $\omega\omega$-FA is *reduced* if it satisfies the following conditions:

(i) every left state is in a cycle;

(ii) every right state is in a cycle;

(iii) every state can be reached from some left state;

(iv) from every state some right state can be reached;

For any given $\omega\omega$-FA we can construct a reduced one that recognizes the same set of biinfinite words.

An *$\omega\omega$-finite transducer* is a 6-tuple $(P, S, S', \rho, P_L, P_R)$

where

P is the finite set of state;
S is the input alphabet;
S' is the output alphabet;
ρ is the transition function;
$P_L \subseteq P$ is the set of left (accepting) states; and
$P_R \subseteq P$ is the set of right (accepting) states.

TO BE DEFINED: A biinfinite word d is an output on input c under the $\omega\omega$-finite transduction.

Theorem A.1 *The family of $\omega\omega$-regular sets is closed under $\omega\omega$-finite transduction.*

Proof Let $C = B(M)$ for some $\omega\omega$-FA $M = (Q, S, \delta, Q_L, Q_R)$. Let $T = (P, S, S', \rho, P_L, P_R)$ be a $\omega\omega$-finite transducer. We shall construct an $\omega\omega$-FA M' such that $T(C) = B(M')$.

Define $\pi_L : \{0,1,2\} \times Q \times P \to \{0,1,2\}$ by

$$\begin{aligned}
\pi_L(2,q,p) &= 0 \\
\pi_L(0,q,p) &= 0 \text{ if } q \notin Q_L \\
&= 1 \text{ if } q \in Q_L \\
\pi_L(1,q,p) &= 1 \text{ if } p \notin P_L \\
&= 2 \text{ if } p \in P_L
\end{aligned}$$

Similarly, $\pi_R : \{0,1,2\} \times Q \times P \to \{0,1,2\}$ is defined by

$$\begin{aligned}
\pi_R(2,q,p) &= 0 \\
\pi_R(0,q,p) &= 0 \text{ if } q \notin Q_R \\
&= 1 \text{ if } q \in Q_R \\
\pi_R(1,q,p) &= 1 \text{ if } p \notin P_R \\
&= 2 \text{ if } p \in P_R
\end{aligned}$$

The $\omega\omega$-FA M' simulates simultaneous execution of M and T; the states of M' have also two additional components whose purpose is to remember the passage through left and right states of M and T. Define $M' =$

$(Q', S', \delta', Q'_L, Q'_R)$, where

$$\begin{aligned} Q' &= \{0,1,2\} \times \{0,1,2\} \times Q \times P \\ Q'_L &= \{2\} \times \{0,1,2\} \times Q \times P \\ Q'_R &= \{0,1,2\} \times \{2\} \times Q \times P \end{aligned}$$

and

$$\delta'((i_1, j_1, q_1, p_1),\ x') = (i_2, j_2, q_2, p_2)$$

if $x' \in S'^*$ and there exists $a \in S \cup \{\lambda\}$ such that

$$\delta(q_1, a) = q_2$$
$$\rho(p_1, a) = (p_2, x')$$
$$\pi_L(i_1, q_1, p_1) = i_2$$
$$\pi_R(i_1, q_1, p_1) = j_2$$

Obviously, we could add finitely many additional states to Q' and replace the definition of $\delta'(q',\ x')$ by one written in terms of $\delta'(q',\ a')$, $a' \in S'$. Accepting executions of M' are in one-to-one correspondence with simultaneous accepting executions of M and T. Thus it is easy to check that $T(C) = B(M')$. □

Using the same technique, we can prove that $\omega\omega$-regular sets are closed under intersection. (A similar result for deterministic ω-regular sets is proved in [4].)

We conclude with two results about the connection between the $\omega\omega$-regularity of a set C of biinfinite words and the regularity of the set $L[C]$. Recall from section 4 that, for a biinfinite word $c \in S^{\mathbf{Z}}$,

$$L[c] = \{\ w \in S^* \mid w \text{ is a finite subword of } c\},$$

and, for $C \subseteq S^Z$,

$$L[C] = \bigcup_{c \in C} L[c].$$

Theorem A.2 *Let C be a set of biinfinite words. If C is $\omega\omega$-regular, then $L[C]$ is a regular language.*

Proof Let $M = (Q, S, \delta, Q_L, Q_R)$ be a reduced $\omega\omega$-FA such that $C = B(M)$. We modify M to produce a FA M' accepting $L[C]$, as follows. We add a new start state that has a λ-transition to every state of M, and make every state of M a final state. Since M is reduced, every word accepted by M' can be extended to a biinfinite word in C. By the construction of M', every subword of a biinfinite word recognized by M is accepted by M'. Therefore, $L[C]$ is the language accepted by M', and thus is regular.

□

Theorem A.3 *If $R \subseteq S^*$ is a regular set then the set $\{ c \in S^Z \mid L[c] \subseteq R \}$ is $\omega\omega$-regular.*

Proof Let $C = \{ c \in S^Z \mid L[c] \subseteq R \}$. We can assume, without loss of generality, that every subword of every word in R is in R. Thus there is a finite automaton M that accepts R and such that every state in M is final. Let $M = (Q, S, \delta, q_0, F)$, where q_0 is the start state and $F = Q$ is the set of final states. Assume that every state in Q can be reached from the start state. Define an $\omega\omega$-FA M_g by $M_g = (Q, S, \delta, Q_L, Q_R)$ where $Q_L = Q_R = Q$. We are going to show that $C = B(M_g)$.

To show that $B(M_g) \subseteq C$, choose any $c \in B(M_g)$ and any finite subword w of c. Then there exist states $q_1, q_2 \in Q$ such that $q_1 w \vdash_\delta^* q_2$. Since every state in Q can be reached from q_0, we have $q_0 x \vdash_\delta^* q_1$ for some $x \in S^*$. It follows that $xw \in R$, and therefore also $w \in R$, because R contains every subword of every word in R. We conclude that $L[c] \subseteq R$, and $c \in C$.

To show that $C \subseteq B(M_g)$, choose any $c \in C$. We use the infinity lemma (p. 383 in [8]) to prove that there is a biinfinite path in M_g labeled by c. We form an oriented tree in which all the finite paths in M_g labeled by the words $c(-j)\dots c(j)$, $j \geq 0$, are vertices. A path π (of length $2j+3$) labeled by $c(-j-1)\dots c(j+1)$ is a son of a path π' (of length $2j+1$) labeled by $c(-j)\dots c(j)$ if π is a concatenation of one transition in M_g followed by π' followed by one transition in M_g. All paths (of length 1) labeled by $c(0)$ are sons of a special root element. The oriented tree is infinite, and every vertex has finite degree. By the infinity lemma there is an infinite path from the root in the tree. Thus there is an infinite sequence of finite paths in M_g labeled by finite subwords of c; each path in the sequence extends its predecessor at both ends. Therefore there is an infinite path in M_g labeled by c. Since $Q = Q_L = Q_R$, it follows that $c \in B(M_g)$. $\square$

References

[1] J. Albert and K. Culik II, "A Simple Universal Cellular Automaton and its One-Way and Totalistic Version", *Complex Systems 1* (1987) 1-16.

[2] R. Berger, "The Undecidability of the Domino Problem", *Mem. Amer. Math. Soc. 66* (1966).

[3] J. R. Brown, *Ergodic Theory and Topological Dynamics*, Academic Press (1976).

[4] K. Culik II and J. Pachl, "Equivalence Problems for Mappings on Infinite Strings", *Information and Control 49* (1981), 52-63.

[5] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley (1979).

[6] L. P. Hurd, "Formal Language Characterizations of Cellular Automaton Limit Sets", *Complex Systems 1* (1987) 69-80.

[7] J. L. Kelley, *General Topology*, American Book Company (1955).

[8] D. E. Knuth, *The Art of Computer Programming, Vol. I (second edition)* Addison-Wesley (1973).

[9] M. Nivat and D. Perrin, "Ensembles Reconnaissables de Mots Biinfinis", *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing* (1982) 47-59.

[10] R. M. Robinson, "Undecidability and Nonperiodicity for Tiling of the Plane", *Inventiones Mathematicae 12* (1971) 177-209.

[11] H. Wang, "Proving Theorems by Pattern Recognition II", *Bell System Tech. J. 40* (1961) 1-41.

[12] S. Wolfram, "Universality and Complexity in Cellular Automata", *Physica 10D* (1984) 1-35.

[13] S. Wolfram, "Computation Theory of Cellular Automata", *Communications in Mathematical Physics 96* (1984) 15-57.

[14] S. Wolfram, "Twenty Problems in the Theory of Cellular Automata", *Physica Scripta T9* (1985) 170-183.