Evolution Complexity of the Elementary Cellular Automaton Rule 18

Zhi-Song Jiang[*] [†]
Hui-Min Xie[*]

[*]*Department of Mathematics,*
Suzhou University, Suzhou, China 215006

[†]*Physical School,*
East China University of Science and Technology,
Shanghai, China 200237

Cellular automata are classes of mathematical systems characterized by discreteness (in space, time, and state values), determinism, and local interaction. Using symbolic dynamical theory, we coarse-grain the temporal evolution orbits of cellular automata. By means of formal languages and automata theory, we study the evolution complexity of the elementary cellular automaton with local rule number 18 and prove that its width 1-evolution language is regular, but for every $n \geq 2$ its width n-evolution language is not context free but context sensitive.

1. Introduction

Cellular automata (CAs) are classes of mathematical systems consisting of a regular lattice of sites and characterized by discreteness (in space, time, and state values), determinism, and local interaction. CAs have been widely used to model a variety of dynamical systems in physics, biology, chemistry, and computer science [1]. Despite their apparent simplicity, CAs can display a rich and complex evolution. The exact determination of their temporal evolution is in general very hard, if not impossible. In particular, many properties of the temporal evolution of CAs are undecidable [2–4].

A one-dimensional CA consists of a double infinite line of sites whose values are taken from an alphabet, that is, a finite set of symbols $A_k = \{0, 1, \ldots, k-1\}$. The symbols of each site update synchronously according to a function of the values of the neighboring sites at the previous time step. The general form of a one-dimensional CA is given by

$$f : A_k^{2r+1} \longrightarrow A_k,$$

$$x_i^{t+1} = f(x_{i-r}^t, \ldots, x_i^t, \ldots, x_{i+r}^t),$$

where x_i^t denotes the value of site i at time t, f represents the local rule defining the automaton, and r is a nonnegative integer specifying the radius of the rule. Therefore, f can induce a function $F : A_k^Z \longrightarrow A_k^Z$,

$$(F(x))_i = f(x_{i-r}x_{i-r+1}\dots x_i \dots x_{i+r})$$

where $x = \cdots x_{-2}x_{-1}x_0x_1x_2\cdots \in A_k^Z$ is a double infinite symbol sequence. We call x the *configuration* and F the *global rule* of the CA. The simplest CAs are those with alphabet $k = 2$ and $r = 1$, and named by Wolfram elementary CAs (ECAs) [5, 6].

In the early 1920s, M. Morse first succeeded in using symbolic dynamics to study mathematical problems [7, 8]. After that, this method which is later called by physicists *coarse-graining* was applied in ergodic theory, differential dynamical systems, and other fields by many researchers. Gradually it became an important way of studying dynamical systems. A famous example is the study of Smale's horseshoe [9, 10]. Another typical example is coarse-graining of unimodal maps, which proved to be rather successful [11–14]. However, much information will be lost during the course of coarse-graining. Therefore a suitable coarse-graining for a system is quite important. This mainly depends on our aims and the real system. If the lost information is not important for the aims focused on, the coarse-graining method can help remove the useless information and more easily grasp the core of the problem. Otherwise, the system will probably become too simple and the results which are drawn through this method will be quite trivial.

Similar to unimodal maps, CAs can also be coarse-grained. In the following we let $A = A_2 = \{0, 1\}$ be the alphabet set and A^Z denote the configuration set. Denote $x = \cdots x_{-2}x_{-1}xx_1x_2\cdots \in A^Z$. First we divide the configuration set into two disjoint clopen (closed and open) sets:

$$A_0 = \{x_0 = 0 \mid x \in A^Z\}, \quad A_1 = \{x_0 = 1 \mid x \in A^Z\}.$$

The orbit $(x, F(x), F^2(x), \dots)$ is coarse-grained into a binary sequence $a_0a_1\dots a_i\dots$, where

$$a_i = \begin{cases} 0, & \text{if } F^i(x) \in A_0; \\ 1, & \text{if } F^i(x) \in A_1. \end{cases}$$

In a general way, we first let $A^n = \{\alpha_0\alpha_1\dots\alpha_{n-1} \mid \alpha_i \in A, 0 \le i \le n-1\}$ and every $\alpha_0\alpha_1\dots\alpha_{n-1} \in A^n$ is regarded as a new symbol. So there are 2^n different symbols in A^n which, of course, is viewed as a new alphabet. Then we divide A^Z into the following 2^n disjoint clopen sets:

$$A_{\alpha_0\alpha_1\dots\alpha_{n-1}} = \{x_0x_1\dots x_{n-1} = \alpha_0\alpha_1\dots\alpha_{n-1} \mid x \in A^Z\}$$

where

$$\alpha_0\alpha_1\dots\alpha_{n-1} \in A^n.$$

Then the orbit $(x, F(x), F^2(x), \ldots)$ is coarse-grained into $a_0 a_1 \ldots a_i \ldots$, where

$$a_i = \alpha_0 \alpha_1 \ldots \alpha_{n-1}, \text{ if } F^i(x) \in A_{\alpha_0 \alpha_1 \ldots \alpha_{n-1}}.$$

Then we may define a function T_n as follows:

$$T_n : x \longmapsto a_0 a_1 \ldots a_i \ldots .$$

The domain of T_n is A^Z and $T_n(x)$ is a single infinite sequence over the alphabet A^n.

Definition 1. Given a CA with local rule f, let $S_n = \{T_n(c) \mid c \in A^Z\}$ and $E_n = \{u \in (A^n)^* \mid u$ is a finite-length substring of y, where $y \in S_n\}$. We call the element S_n a *width n-evolution sequence* (or simply n-evolution sequence) and E_n a *width n-evolution language* (or simply n-evolution language) generated by the CA.

Here the notation $(A^n)^*$ in Definition 1 is defined as a set of strings where every string consists of zero and more symbols of A^n. From Definition 1 we know S_n consists of one-way infinite sequences, such as $(0 + 10)^\infty$, which are defined as

$$(0 + 10)^\infty = \{a_1 a_2 \ldots a_k \ldots \mid a_k \in \{0, 10\}, k \geq 1\}.$$

E_n is a formal language consisting of all the substrings of S_n.

From another point of view, these coarse-graining sequences are exactly the observation windows or evolution sequences which are put forward by Gilman, Kůrka, and others [15–17]. The width of observation windows is the above-mentioned n [15].

Starting from studying the mathematical models of natural languages, N. Chomsky put forward four levels of language hierarchy, that is, the Chomsky hierarchy, according to the complexity of their generating grammar: regular languages, context free languages (CFLs), context sensitive languages (CSLs), and recursively enumerable languages (RELs). Their complexity and scope increase successively [18].

Definition 2. The grammatical complexity of the evolution language generated by a CA is called the *evolution complexity* of the CA.

In this paper we use formal language theory to study the grammatical complexity of evolution sequences (or simply, the evolution complexity of a CA). As a matter of fact, Jen has studied the aperiodicity of 1-evolution sequences of some one-dimensional CAs [19]. Many experts, including Gilman, Kůrka, and Maass, have done some meaningful work on evolution complexity, and have obtained many interesting results [16, 17, 20]. Gilman has proved the following proposition in [16].

Proposition 1. Every evolution language of a CA is always context sensitive.

Thus only three levels of the Chomsky hierarchy have to be considered: regular language, context free language, and context sensitive language. It is not trivial to prove that the evolution language generated by a CA is irregular. Gilman has given a concrete example of a CA which is not elementary and proved that its 1-evolution language is neither regular nor context free [20]. But the explanation is not clear in [20] and no rigorous or satisfying proof is provided (see appendix B for it). But for ECA, as we know, there is still no example in which the irregularity of an evolution language is rigorously proved. On the other hand, is it true that 1-evolution, 2-evolution, ..., n-evolution, ... are at the same grammatical level? The evolution complexity of ECA 18 explains this point clearly (the indexing rule for ECA can be found in [5, 6]). Though ECA is a class of CA with simple rules, some ECAs can display chaotic behaviors [5, 6]. The ECA 18 is a typical one which is studied by many experts from different points of view[19, 21–23]. In this paper we also consider ECA 18 and mainly prove the following two theorems.

Theorem 1. For ECA 18, E_1 is a regular language.

Theorem 2. For ECA 18, E_n is not a CFL but a CSL ($n \geq 2$).

We also prove incidentally that, for any general CA and any n, the level of its $(n+1)$-evolution language is not lower than the level of its n-evolution language in the Chomsky hierarchy.

The organization of the paper is as follows. Section 2 provides some new notions and presents two important propositions which will be needed to prove the two theorems. Section 3 gives the proofs of the main results. Section 4 gives some useful lemmas and proves the two propositions in section 2. Further discussions are made in section 5 in which several conjectures and open problems are proposed. Appendix A proves Lemma 4.11, whose proof is too technical and too long to be included in the text. Appendix B will provide a detailed proof for a nonECA put forth by Gilman whose evolution languages are also not CFLs.

2. Definitions and propositions

In this section, we first state some notions and symbols that are used later and then give the two important propositions.

We use f as the local rule of the CA and extend its domain to A_k^* as follows:

$$c_1c_2\ldots c_m \mapsto \begin{cases} \epsilon, & (m \leq 2r); \\ f(c_1c_2\ldots c_{2r+1})f(c_2c_3\ldots c_{2r+2})\ldots \\ \quad f(c_{m-2r}c_{m-2r+1}\ldots c_m), & (m \geq 2r+1) \end{cases}$$

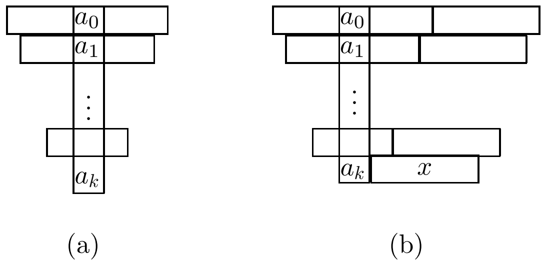

Figure 1. (a) Define $CP_n(a_0a_1 \ldots a_k)$. (b) Define $CRP_n(a_0a_1 \ldots a_k, x)$.

where r is the radius of the local rule, $c_i \in A_k (1 \leq i \leq m)$, and ϵ is the empty string containing no symbol. Let $c = c_1c_2 \ldots c_m$, then the string $c_m \ldots c_2c_1$ is called the *mirror* of string c. For $c = c_1c_2 \ldots c_m$, the operator π is defined as follows:

$$\pi c = c_2 \ldots c_m, \quad c\pi = c_1c_2 \ldots c_{m-1}.$$

$|c| = m$ denotes that the length of c is m. So $|\epsilon| = 0$. We also need the regular expression [14, 18]. The following two definitions are important in this paper (see Figure 1).

Definition 3. Let $n > 0$, $|a_i| = n (0 \leq i \leq m)$ be in A^n. We define the *center-restriction preimage* $CP_n(a_0a_1 \ldots a_m)$ as

$$\{s \in A^* \mid \pi^j f^{m-j}(s)\pi^j = a_{m-j}, 0 \leq j \leq m\}.$$

Definition 4. Let $n > 0$, $|a_i| = n (0 \leq i \leq m)$ be in A^n. Define the *center-right-terminal preimage* $CRP_n(a_0a_1 \ldots a_m, x)$ as

$$\{s \in A^* \mid \pi^j f^{m-j}(s)\pi^{j+l} = a_{m-j},\ f^m(s) = a_m x,\ 0 \leq j \leq m\},$$

where $x \in A^*$, $l = |x|$.

Clearly, when $x = \epsilon$, $CRP_n(a_0a_1 \ldots a_m, x) = CP_n(a_0a_1 \ldots a_m)$. If $CP_n(a_0a_1 \ldots a_m) \neq \emptyset$, then the string $a_0a_1 \ldots a_m \in (A^n)^*$ can appear in the evolution of the CA, that is, $a_0a_1 \ldots a_m \in E_n$ and *vice versa*. If $s \in CP_n(a_0a_1 \ldots a_m)$, we say s can generate the string $a_0a_1 \ldots a_m$. In appendix A, we give other similar notions so as to prove the important Lemma 4.11. In this paper, we mainly consider ECA 18 whose local rule is defined as follows:

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

We can see that the rule of ECA 18 satisfies

$$(1)\ f(000) = 0; \quad (2)\ f(x_{-1}xx_1) = f(x_1xx_{-1}).$$

In [5, 6], Wolfram named a CA, whose local rule satisfies the above two conditions, a *legal* CA. Hence the ECA of rule 18 is a legal CA. In the evolution of the ECA of rule 18, the strings $10^{2m}1 (m \geq 0)$ play an important role. Many experts give them some special names such as kinks, particles, defects, and irregular blocks [19, 21, 24–26]. In this paper we call them *defects*. $\#(x)$ is used to denote the number of defects in x. For example, $\#(1001011) = 2$ and $\#(10011) = 2$. If the initial configuration does not contain any defect, then its temporal evolution with local rule 18 is equal to its temporal evolution with local rule 90. Therefore, in a sense, the evolution of ECA 18 with an initial configuration that contains defects represents its characteristic behaviors. In order to quickly prove Theorems 1 and 2, we list the following important propositions whose proofs are given in section 4.

Proposition 2. For ECA 18, $S_1 = (0 + 10)^\infty$.

Proposition 3. Let $a = 11, b = 00$, and $m > 0$, then $b^l a b^{2m} a \in E_2$ if and only if $l \leq 2^{k+1} - 2$. Here, k is determined as follows:

$$2^k \leq m < 2^{k+1}.$$

3. Proofs of the two theorems

In this section, we prove the two theorems. The main tools in our proof are Propositions 2 and 3 which are proved later in section 4.

3.1 Proof of Theorem 1

Using the regular expressions, by Proposition 2, E_1 can be written as

$$E_1 = (0 + 10)^*(\epsilon + 1).$$

Therefore, E_1 is regular. ∎

3.2 Proof of Theorem 2

The following two propositions are useful results taken from [18].

Proposition 4. (1) If L is a CFL and R is a regular set, then $L \cap R$ is a CFL. (2) Each family of the regular languages: CFLs, CSLs, and RELs is closed under homomorphism and inverse homomorphism.

Proposition 5. (Pumping Lemma) Let L be any CFL. Then there is a constant N, depending only on L, such that if s is in L and $|s| \geq N$, then we may write $s = uvwxy$ such that

1. $|vx| \geq 1$
2. $|vwx| \leq N$
3. for all $i \geq 0$, $uv^i wx^i y$ is in L.

The following proposition holds for any CA.

Proposition 6. For every n, E_n is not more complex than E_{n+1} in the Chomsky hierarchy.

Proof. We define a homomorphism $R : A^{n+1} \to A^n$ as follows:

$$R : x_0 x_1 \ldots x_n \to x_0 x_1 \ldots x_{n-1}$$

where $x_i \in A (0 \leq i \leq n)$. It is easy to show that the following equation holds:

$$R(E_{n+1}) = E_n.$$

By Proposition 4, the results hold. ■

Theoretically, it is possible to determine its exact grammatical level if we can solve the membership problem of a language over an alphabet, that is, the necessary and sufficient condition for a string belonging to the language. However, it is difficult to solve the membership of E_2 completely, that is, we may not decide completely whether a string over A^2 is in E_2 or not. But it turns out that in order to decide a language's level in the Chomsky hierarchy it is not necessary to solve the membership problem of E_2 completely. In the following discussion we will only decide the membership of a suitable subset of $(A^2)^*$, that is, for a string in this subset, we are able to judge whether the string is in E_2 or not.

We know the alphabet A^2 has four symbols: $00, 01, 10$, and 11. Therefore, every string in E_2 contains at most four different symbols. Practically, it is difficult to deal with all kinds of strings in E_2. So we focus on a particular subset of E_2 that only contains two symbols: 00 and 11 which are denoted by b and a respectively. From Proposition 3, we can decide whether a string in a particular subset of $(00 + 11)^*$ is in E_2 or not. Using this proposition, we can prove Theorem 2. But we still need some preliminaries.

We define a homomorphism *hom* starting from the simple rules:

$$hom : \underline{a} \longrightarrow a, \quad \underline{b} \longrightarrow bb.$$

Let

$$L = hom^{-1}(E_2 \cap b^* a (bb)^* a).$$

According to Proposition 4, we know that the grammatical complexity of L is the same as that of E_2. Therefore, if it is CS, it suffices to know the grammatical complexity of L. Using the symbols $\underline{a}$ and $\underline{b}$, we restate Proposition 3 as follows.

Proposition 7. Let $l, m \geq 0$ then

$$\underline{b}^l \underline{a} \underline{b}^m \underline{a} \in L \Longleftrightarrow l \leq 2^k - 1,$$

where k satisfies

$$2^k \le m < 2^{k+1}.$$

Now we begin to prove Theorem 2. First, we prove that L is not a CFL. Otherwise, by the pumping lemma (Proposition 5), there exists an $N > 0$, say $N > 2$, such that if s is in L, and $|s| \ge N$, then we may write $s = uvwxy$ such that

1. $|vx| \ge 1$
2. $|vwx| \le N$
3. for all $i \ge 0$, uv^iwx^iy is in L.

Now we take $s = \underline{b}^{2^N-1}\underline{ab}^{2^N}\underline{a} \in L$. From the definition of L, we know every string in L contains only two $\underline{a}$s. Therefore, by 3, neither v nor x contains $\underline{a}$. Hence we can be certain that

$$v = \underline{b}^l \text{ and } x = \underline{b}^t,$$

where, by 1 and 2, $1 \le l + t \le N$.

(a) If both v and x are substrings of $\underline{b}^{2^N-1}\underline{a}$ which is the prefix of s, then by 3, we have

$$uv^2wx^2y \in L \Longrightarrow \underline{b}^{2^N-1+l+t}\underline{ab}^{2^N}\underline{a} \in L.$$

This is in contradiction with Proposition 7.

(b) If both v and x are substrings of $\underline{ab}^{2^N}\underline{a}$ which is the suffix of s, then by 3 we have

$$uwy \in L \Longrightarrow \underline{b}^{2^N-1}\underline{ab}^{2^N-l-t}\underline{a}.$$

This is also in contradiction with Proposition 7.

(c) If v is the substring of $\underline{b}^{2^N-1}\underline{a}$ and x is the substring of $\underline{ab}^{2^N}\underline{a}$, which means w contains $\underline{a}$, then we may suppose $t > 0$. (In fact, if $t = 0$, then $l > 0$, similar to (a), we can obtain a contradiction.) Then by 3, $uwy \in L \Longrightarrow \underline{b}^{2^N-1-l}\underline{ab}^{2^N-t}\underline{a} \in L$. By Proposition 7, $2^N - 1 - l \le 2^k - 1$ where k satisfies $2^k \le 2^N - t < 2^{k+1}$. Hence we have $k \le N - 1$ and then $2^N - 1 - l \le 2^{N-1} - 1$. That is to say $l \ge 2^{N-1}$. By 2, $l \le N$, then $N \ge 2^{N-1}$. This is impossible for $N > 2$. Therefore, L is not a CFL. By Proposition 4, E_2 is also not a CFL. By Proposition 6 and 1, $E_n (n \ge 2)$ are all not CFLs but CSLs. ∎

In the remainder of this paper (including appendix A) we give the proofs of Propositions 2 and 3 which are essential to get the main results in this work, that is, Theorems 1 and 2.

4. Proofs of the two propositions

In this section, we present some lemmas in order to prove Propositions 2 and 3.

4.1 Proof of Proposition 2

In order to prove Proposition 2 we need to establish some lemmas concerning the properties of the width 1-evolution strings.

Lemma 4.1. $CP_1(11) = \emptyset$.

Proof. By the local rule f, for any $a_1, a_2 \in A$, $f(a_1 1 a_2) = 0$. Therefore, $CP_1(11) = \emptyset$. ■

Lemma 4.2. If $x, y \in A^*$ and $f(x) = y$, then $\#(x) \geq \#(y)$.

Proof. This can be obtained simply since each preimage of the defect has at least one defect [19, 21, 22, 24, 25]. ■

Remark. Furthermore, if $f(x) = y$, then we have the following additional results.

(a) If 11 is not a substring of x, then $\#(x) = \#(y)$.

(b) If 11 is either a prefix or suffix of x, then $\#(x) > \#(y)$.

Lemma 4.3. If $x, y \in A^*$, $f(x) = y$ and the suffix of y is a defect, then for each y_1 which satisfies $\#(y) = \#(yy_1)$, there exists an x_1 such that $f(xx_1) = yy_1$ and $\#(xx_1) = \#(x)$.

Proof. Because the string y must have the symbol 1 as its suffix, then either 00 or 01 is the suffix of x. (Because 1 has only two possible preimages: 100 and 001.) First we may suppose $|y_1|$ is even. Then we have $\#(yy_1) = \#(y)$ if and only if $y_1 \in (00 + 01)^*$ and $\#(xx_1) = \#(x)$ if and only if $x_1 \in (00 + 01)^*$. We will prove the existence of x_1.

Note that $f(0000) = 00$, $f(0001) = 01$, $f(0100) = 01$, and $f(0101) = 00$. We now use the simplified form for the strings in $(00 + 01)^*$. Let 0 stand for 00, 1 for 01, then we have $f(00) = 0$, $f(01) = 1$, $f(10) = 1$, and $f(11) = 0$. We can also write

$$f(a_1 a_2) = a_1 + a_2 \pmod 2$$

where $a_1, a_2 \in \{0, 1\}$ are the simplified forms.

Let $b_0 b_1 \ldots b_{m-1}$ be the simplified form of y_1 and $a_1 a_2 \ldots a_m$ be the simplified form of x_1. Let a_0 be the simplified form of the suffix of x with length 2, then

$$f(a_0 a_1 \ldots a_m) = b_0 b_1 \ldots b_{m-1}.$$

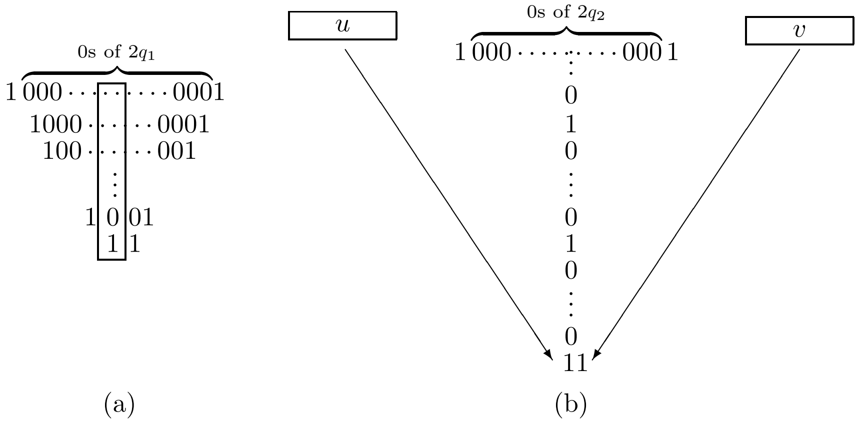

Figure 2. (a) In the rectangular box is an evolution string $0^{q_1}1$. (b) Define u and v. πx (in fact $x\pi$ also) can generate the string $0^{q_2}10^{q_3}1\ldots0^{q_m}1$.

That is,

$$a_i + a_{i+1} = b_i \ (i = 0, 1, \ldots, m-1).$$

Whether there exists an x_1 which satisfies the lemma or not is up to whether the above equations have solutions or not. It is easy to see that the solutions are

$$a_i = a_0 + \sum_{j=0}^{i-1} b_j \ (i = 0, 1, \ldots m).$$

If $|y_1|$ is odd, it must end with 0, hence we may let $y_1' = y_1 1$, and have x_1', such that $f(xx1') = yy_1'$ and $\#(xx_1') = \#(x)$. Then the string that we need can be obtained by removing the last symbol of x_1'. ∎

Lemma 4.4. If $q_i \in \mathbf{N}$ for $i = 1, 2, \ldots m$, then $CP_1(0^{q_1}10^{q_2}1\ldots0^{q_m}1) \neq \emptyset$.

Proof. We can obtain this conclusion by proving that there exists at least one string $x \in CRP_1(0^{q_1}10^{q_2}1\ldots0^{q_m}1, 1)$ such that x has only one defect: $10^{2q_1}1$.

We will use induction in m. When $m = 1$, $10^{2q_1}1 \in CRP_1(0^{q_1}1, 1)$, and the lemma is true. See Figure 2(a). Supposing the lemma is true for $m = k$, then for the case of $m = k + 1$, by the inductive hypothesis there exists a string x with only one defect $10^{2q_2}1$ in $CRP_1(0^{q_2}10^{q_3}1\ldots0^{q_{k+1}}1, 1)$. Let $x = u10^{2q_2}1v$, where the substring $u1$ and $1v$ do not contain any defect. See Figure 2(b). If q_2 is odd then $f(00(10)^{(q_2-1)/2}11(01)^{(q_2+1)/2}00) = 10^{2q_2}1$. By Lemma 4.3, there exist u_1, v_1 which satisfy

$$\#(u_1 00(10)^{(q_2-1)/2}11(01)^{(q_2+1)/2}00v_1) = 1$$

such that

$$f(u_1 00(10)^{(q_2-1)/2} 11(01)^{(q_2+1)/2} 00v_1) = x.$$

Hence we obtain $y = u_1 00(10)^{(q_2-1)/2} 11(01)^{(q_2+1)/2} 00v_1$.

If q_2 is even then $f(00(10)^{q_2/2} 11(01)^{q_2/2} 00) = 10^{2q_2}1$. By Lemma 4.3, there exist u_1', v_1' which satisfy $\#(u_1' 00(10)^{q_2/2} 11(01)^{q_2/2} 00v_1') = 1$ such that

$$f(u_1' 00(10)^{q_2/2} 11(01)^{q_2/2} 00v_1') = x.$$

Then we have $y = u_1' 00(10)^{q_2/2} 11(01)^{q_2/2} 00v_1'$.

So $CRP_1(10^{q_2}10^{q_3}1 \ldots 0^{q_{k+1}}1, 1)$ has an element y which contains only one defect 11. Therefore, $CRP_1(0^{q_1}10^{q_2}1 \ldots 0^{q_{k+1}}1, 1)$ has an element which contains only one defect $10^{2q_1}1$.

Thus the inductive proof is completed. ■

Corollary. For all the $q_i \in \mathbf{N}$, $i = 1, 2, \ldots, m$, we have

$$CP_1(10^{q_1}10^{q_2}1 \ldots 0^{q_m}1) \neq \emptyset.$$

Proof of Proposition 2. This result can be easily obtained by Lemma 4.1, Lemma 4.4, and its corollary. ■

4.2 Proof of Proposition 3

Now we focus on proving Proposition 3 and turn to discuss the width 2-evolution strings.

Lemma 4.5. $CP_2(1111) = \emptyset$, $CP_2(0000) = 0000$, and $CP_2((00)^n 11) = 10^{2n}1 (n \geq 1)$.

Proof. It is easy to obtain these equations by the local rule. ■

Remark. Although $CP_2(x)$ is a set where every string can generate the evolution string $x \in (A^2)^*$, if there exists only one string y in $CP_2(x)$, then we will simply write $CP_2(x) = y$. This is done for the other set equations in the remainder of this paper.

Lemma 4.6. If $\#(1x) = 1$ then there exists an $m \in \mathbf{N}$ such that

$$CRP_2((00)^m 11, x) = \emptyset.$$

Proof. Assume the contrary: if it is not true for every m then

$$CRP_2((00)^m 11, x) \neq \emptyset.$$

However it is clear that if $y \in CRP_2((00)^m 11, x)$, then $1(00)^m 1$ is its prefix. Let a_m be one element of $CRP_2((00)^m 11, x)$ and denote $a_m = 1(00)^m b_m$, where b_m has 1 as its prefix and $|b_m| = |x| + 1$. See Figure 3. We also have $f(00b_{m+1}) = b_m$. Since all the lengths of b_m are the same,

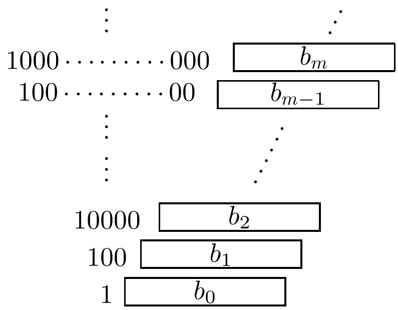

Figure 3. Define $\{b_m\}_{m\geq 0}$. If there are infinite b_m, then $\{b_m\}_{m\geq 0}$ should be periodic.

there will exist i and l, $i \neq l$, such that $b_i = b_l$. Therefore $\{b_m\}_{m\geq 0}$ is eventual periodic. But from the relation of $f(00b_{m+1}) = b_m$, $\{b_m\}_{m\geq 0}$ must be periodic. Let q be its period. We might as well suppose that some suffix of x is a defect. By Lemma 4.2, we know that each b_m has only one defect. Next we will observe the rule of the inverse floating of a defect contained in b_m as m increases. First we state what is meant by "inverse floating." In inverse evolution, the change of the position of 1, which is the rightmost symbol in a defect, is called *inverse floating*. For example, if $f(100001) = 1001$, we say that defect 1001's inverse floating is 1. If $f(001100) = 1001$, we say that defect 1001's inverse floating is -1.

It is easy to obtain the following fact: defect $10^{2m}1$'s inverse floating value set is $\{-2m+1, -2m+2, \ldots, -1, 1\}$. When its inverse floating value is 1, its preimage is $10^{2m+2}1 \ldots\ldots (*)$.

Now we will observe the inverse floating of defects in $\{b_m\}_{m\geq 0}$. For the sake of the periodicity of $\{b_m\}_{m\geq 0}$, we have $b_0 = b_q$ and the defect in b_0 has floated q positions in q steps. Therefore, in every step the value of the inverse floating of a defect can only be 1. Again using the periodicity, the value of the inverse floating of a defect in each b_m is 1. By (*), we know, after enough steps, all b_m are $0^{|x|}1$. But this is in contradiction with $\#(b_m) = 1$. ■

Using the result of Lemma 4.6, we know that the following definition is well-defined.

Definition 5. $h(x) = \max\{i \mid CRP_2((00)^i 11, x) \neq \emptyset, i \geq 0\}$, where $\#(1x) = 1$.

Lemma 4.7. For every even number $m, h((01)^m 1) = 0$.

Proof. Noticing $f^{-1}(11(01)^m 1) = \emptyset$ then $CRP_2(0011, (01)^m 1) = \emptyset$, therefore we know the lemma is correct [23]. ■

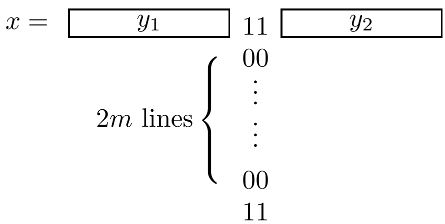

Figure 4. Define y_1 and y_2. $y_1 11 y_2$ can generate the evolution string $11(00)^{2m}11 \in A^{2*}$.

Lemma 4.8. If y is a nonempty prefix of x and $\#(x) = \#(y)$, then $h(x) = h(y)$.

Proof. Because y is the prefix of x, $h(x) \leq h(y)$. Let $h(y) = m$, then $CRP_2((00)^m 11, y) \neq \emptyset$. So the result holds on condition that there exists s_1 such that $f^m(ss_1) = 11x$, where s is in $CRP_2((00)^k 11, y)$. Using the mathematical inductive method and Lemma 4.3, we can easily obtain this lemma. ■

Lemma 4.9. For each integer $m > 0$, there exists an $l > 0$, such that $CP_2((00)^l 11(00)^{2m}11) = \emptyset$.

Proof. For every $x \in CP_2(11(00)^{2m}11)$, we can write $x = y_1 11 y_2$ as in Figure 4. We say either y_1 or y_2 has at least one defect. In fact, if this is not true, x should be $00(10)^l 11(01)^l 00$, where l is up to the length of y_1 (because both πy_1 and $y_2 \pi$ cannot contain 00 or 11). But $f(00(10)^l 11(01)^l 00) = 1(00)^{2l+1}1 \notin CP_2((00)^{2m}11)$. Furthermore, by symmetry, both y_1 and y_2 have at least one defect.

If this lemma is not true, then for every $l \geq 1$ we have

$$CRP_2((00)^l 11, y_2) \neq \emptyset.$$

But this contradicts Lemma 4.6. ■

From Definition 3 we know $CP_2((00)^l 11(00)^{2m}11) \neq \emptyset$ if and only if $(00)^l 11(00)^{2m}11 \in E_2$. The function $g(m)$ defined below tells us which l can make the string $(00)^l 11(00)^{2m}11 \in (A^2)^*$ be in E_2.

Definition 6. $g(m) = \max\{l \mid CP_2((00)^l 11(00)^{2m}11) \neq \emptyset, l \geq 0\}$.

From the definition we can obtain:

$$(00)^l 11(00)^{2m}11 \in E_2 \text{ iff } l \leq g(m).$$

Therefore, the function $g(m)$ is rather important in order to prove Proposition 3. In the remainder of this paper we calculate $g(m)$ exactly. The first step is to obtain the following formula.

Lemma 4.10. $g(m) = \max\{h(x) \mid x = (01)^l 1, l = 0, 1, \ldots, m-1\}$.

Proof. Every element $x = x_{-1}11x_1$ in $CP_2(11(00)^{2m}11)$ has the mirror symmetry, therefore we only need to consider $h(x_1)$. It is easy to know that $\#(x_1) > 0$ by Lemma 4.2. Denote

$$D = \{x_1 \mid x = x_{-1}11x_1, x \in CP_2(11(00)^{2m}11)\}.$$

Then we have $g(m) = \max\{h(x_1) \mid x_1 \in D\}$.

If $\#(1x_1) = 1, x_1 \in D$, then x_1 can be written as

$$(01)^{t_1}1(01)^{t_2}00,$$

where $t_1 + t_2 = m - 1$.

We denote

$$D_1 = \{x \mid \#(1x) = 1, x \in D\}$$
$$D' = \{(01)^{t_1}1 \mid x = (01)^{t_1}1\ldots, x \in D\}.$$

Clearly, $D_1 \subset D$. So we have

$$\begin{aligned} g(m) &= \max\{h(x) \mid x \in D\} \le \max\{h(x) \mid x \in D'\} \\ &= \max\{h(x) \mid x \in D_1\} \le \max\{h(x) \mid x \in D\}. \end{aligned}$$

Therefore, we have

$$g(m) = \max\{h(x) \mid x \in D'\}.$$

Since D' also equals $\{x \mid x = (01)^t1, t = 0, 1, \ldots, m-1\}$. ■

Remark. Using the method of this lemma, we can obtain the following result:

$$h((00)^m1) = \max\{h(x) \mid x = (01)^t1, 0 \le t \le m-1\} + 1.$$

Lemma 4.11. The following results hold with respect to each integer $q \ge 0$ and $l \ge 0$.

1. $CRP_2((00)^{2^q-1}11, (01)^{2^q-1}0^l1) = 10^{2^{q+1}-2}10^{2^{q+1}+l-2}1$, where $l \ge 0$.
2. If $0 \le t < 2^q - 1$, then $h((01)^t1) \le h((01)^{t+2^q}1) < 2^q - 1$.
3. $h((01)^{2^q-1}1) = 2^{q+1} - 2$.

This lemma is important, but due to its length the proof is presented in appendix A. This lemma is enough to help us know the function $g(m) = 2^{k+1} - 2$ where k satisfies $2^k \le m < 2^{k+1}$. If it is proved, then we complete the proof of Proposition 3.

Proof of Proposition 3. By Lemma 4.10, $g(m) = \max\{h(x) \mid x = (01)^l1, l = 0, 1, \ldots, m-1\}$. By Lemma 4.11, if $0 \le t < 2^{l+1} - 1$, then $t = 2^l - 1, h((01)^t1)$ is maximal. Therefore, when $2^l \le m < 2^{l+1}$, $g(m) = h((01)^{2^l-1}1) = 2^{l+1} - 2$. ■

5. Further discussions

For unimodal maps, using two symbols (not including the critical point) we can well reflect the dynamical behaviors [12–14]. As for CAs, how many symbols do we need? We guess, for a general CA, we need 2^{2r} symbols. Here r is the radius of the local rule. This means the width of the observation window is $2r$. An example at hand is the ECA 18. From the above discussions, we know the complexities of E_1 and E_2 of ECA 18 are not in the same level of the Chomsky hierarchy. But $E_2, E_3, \ldots$ are in the same level. In a sense, we can say E_2 is a representation for the evolution complexity of ECA 18. So a width 2 observation window is enough to show the complexity of all evolution languages. That means we need $2^2 = 4$ symbols. Using formal language theory, we state the conjecture as follows.

Conjecture 1. E_{2r} and $E_{2r+m} (m > 0)$ are in the same level of the Chomsky hierarchy, where r is the radius of the local rule.

For unimodal maps there is the following open problem [13, 14].

Open Problem 1. If the language L generated by a unimodal map is a CFL, then it is a regular language.

In [13], many circumstances are discussed and the results therein strongly suggest that Open Problem 1 may indeed be true.

It is strange to find that, for CAs, there is a similar conjecture [20].

Open Problem 2. If the evolution language L of a CA is a CFL, then it is a regular language.

The two open problems seem to say that the language families which are generated by dynamical coarse-graining orbits do not contain proper CFLs.

There are two ways of studying CAs in language theory. One way is to discuss the complexity for its evolution language, the other way is to dicuss the complexity for its limit language [5, 6]. Gilman has proved that for every CA and every integer n, E_n is a CSL. But the limit language can be a REL or not a REL [27]. Is it true that for every CA and every n, E_n is less complex than L_Ω? Here the notation L_Ω is the limit language which consists of all the substrings of the limit set Ω, where Ω is defined as follows:

$$\Omega = \bigcap_{j=0}^{\infty} F^j(A_k^Z).$$

The following example proposed by Gilman in [16] denies this point.

We still take the alphabet set $A = \{0, 1\}$ and the global rule is defined as follows:

$$(F(x))_i = x_{i+1}x_{i+2}.$$

Then we obtain a CA whose radius $r = 2$, hence it is not elementary. We have the following proposition [16, 17, 20].

Proposition 8. For the above CA, the evolution language E_n $(n > 0)$ is not a CFL.

But, obviously, the limit language is $0^*1^*0^*$. So it is regular. Proposition 8 can be seen in [17, 20] but no rigorous proof is provided. Appendix B gives a clear proof.

But for a great majority of CAs, we still think that if their evolution languages are not regular, so are the limit languages. Particularly, for ECA 18, we have another conjecture.

Conjecture 2. The limit language of ECA 18 is not a CFL.

However, up to now, the irregularity of the limit language for ECA 18 is still not rigorously proved, though many experts think it is so [5, 22].

Acknowledgments

We thank Y. Wang, Y. L. Cao, J. Luo, and W. X. Qin for useful discussions. We are also grateful to Y. X. Shao for much useful help. This work is supported by the Special Funds for Major State Basic Research Projects.

Appendix

A. The proof of Lemma 4.11

In order to prove Lemma 4.11 we need some additional notions.

A *block* is an ordered arrangement of some strings $x_1, x_2, \ldots, x_m$ with certain structures, where m is called the *thickness* of the block.

In Figure 5, we can see some examples of blocks with thickness 5. A *sub-block* is a part of a block. In Figure 5, we call A an *evolution block* which, in general, satisfies

$$f(x_i) = x_{i+1} (i = 1, 2, \ldots, m-1)$$

and is denoted by $EB(x_1, m)$. When $|x_m| < 2r + 1$, the thickness m is determined by x_1 uniquely, and we simply write $EB(x_1, m) = EB(x_1)$. We call B in Figure 5 a *right-skew-evolution block* (RSB) which, in general, satisfies

$$f(y_i x_i) = x_{i+1} \ (i = 1, 2, \ldots, m-1)$$

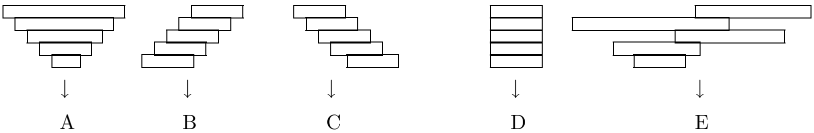

Figure 5. Some examples of blocks: A is an evolution block, B is a right-skew-evolution block, C is a left-skew-evolution block, D is a rectangle block, and E is an irregular block.

where $|y_i| = 2r$. We denote RSB as $RSB(x_1, x_2, \ldots, x_m)$. Similarly, we can define a *left-skew-evolution block* (LSB) and denote it by $LSB(x_1, x_2, \ldots, x_m)$. See C in Figure 5. If $|x_1| = |x_2| = \cdots = |x_m| = n$, $RSB(x_1, x_2, \ldots, x_m)$ is also denoted as $RSB_n(x_1 x_2 \ldots x_m)$, and $LSB(x_1, x_2, \ldots, x_m)$ is also denoted as $LSB_n(x_1 x_2 \ldots x_m)$. If $|x_1| = |x_2| = \cdots = |x_m| = n$, like D in Figure 5, we call it a *rectangle block* and denote it, in general, by $RB(x_1, x_2, \ldots, x_m)$ or $RB_n(x_1 x_2 \ldots x_m)$. If block A is made up of $x_1, x_2, \ldots, x_m$, then $MR_n(A)$ ($ML_n(A)$) is a sub-block of A, which is made up of the suffix (prefix) of $x_i (1 \leq i \leq m)$ with length n. We call E in Figure 5 an *irregular block*.

In the following we introduce some set functions. Let $a_1, a_2, \ldots, a_m$, $x, y, s \in A^*$ and $|a_1| = |a_2| = \cdots = |a_m| = n$. We can easily prove that the center-restriction preimage $CP_n(a_1 a_2 \ldots a_m)$ is equal to

$$\{s \mid RB_n(a_1 a_2 \ldots a_m) \text{ is the sub-block of } EB(s), \text{ and } f^{m-1}(s) = a_m\},$$

and that the center-right-terminal preimage $CRP_n(a_1 a_2 \ldots a_m, x)$ is equal to

$$\{s \mid RB_n(a_1 a_2 \cdots a_m) \text{ is the sub-block of } EB(s), \text{ and } f^{m-1}(s) = a_m x\}.$$

Now we define other useful notions.

Definition 7. Define the *right-restriction preimage* $RP_n(a_1 a_2 \ldots a_m)$ as

$$\{s \mid RSB_n(a_1 a_2 \ldots a_m) \text{ is the sub-block of } EB(s), \text{ and } f^{m-1}(s) = a_m\}.$$

Definition 8. Define the *right-restriction-right-terminal preimage* $RRP_n(a_1 a_2 \ldots a_m, x)$ as

$$\{s \mid RSB_n(a_1 a_2 \ldots a_m) \text{ is the sub-block of } EB(s), \text{ and } f^{m-1}(s) = a_m x\}.$$

Definition 9. Define the *center-restriction-double-terminal preimage* $CDP_n(a_1 a_2 \ldots a_m, x, y)$ as

$$\{s \mid RB_n((a_1 a_2 \cdots a_m) \text{ is the sub-block of } EB(s), \text{ and } f^{m-1}(s) = x a_m y\}.$$

Finally, we state several important propositions which are self-evident.

Proposition 9. Let $|a_1| = |a_2| = \cdots = |a_m|, |b_1| = |b_2| = \cdots = |b_m|$. If

$$CP_n(a_1a_2 \ldots a_m) = RP_n(b_1b_2 \ldots b_m)$$

then for every x we have

$$CRP_n(a_1a_2 \ldots a_m, x) = RRP_n(b_1b_2 \ldots b_m, x).$$

Proposition 10. Let $|a_1| = |a_2| = \cdots = |a_m| = |b| = |c| = 2r, x \in A^*$. Then

$$MR_n(EB(RRP_{2r}(a_1a_2 \ldots a_m b, x), m + 1)) = \\ MR_n(EB(RRP_{2r}(a_1a_2 \ldots a_m c, x), m + 1)),$$

where $n = |x|$.

Now we state Lemma 4.11 again and start to prove it.

Lemma 4.11. The following results hold with respect to each integer $q \geq 1$.

1. $CRP_2((00)^{2^q-1}11, (01)^{2^q-1}0^l1) = 10^{2^{q+1}-2}10^{2^{q+1}+l-2}1$, where $l \geq 0$.
2. If $0 \leq t < 2^q - 1$, then $h((01)^t1) \leq h((01)^{t+2^q}1) < 2^q - 1$.
3. $h((01)^{2^q-1}1) = 2^{q+1} - 2$.

Proof. First we note that $CP_2((00)^n11) = RP_2((01)^n11)$. By Proposition 9,

$$CRP_2((00)^n11, x) = RRP_2((01)^n11, x).$$

Therefore, claim 1 equals

$$RRP_2((01)^{2^q-1}11,\ (01)^{2^q-1}0^l1) = 10^{2^{q+1}-2}10^{2^{q+1}+l-2}1.$$

We will use the mathematical inductive method on q. In the remainder of this proof the notation $(1)_2$ stands for claim 1 when $q = 2$. Similarly, the notation $(2)_m$, $(3)_m$ stands for claims 2 and 3 when $q = m$.

Step one: Let $q = 1$.

$(1)_1$ is $CRP_2(0011, 010^l1) = 10010^{l+2}1$. It is easy.

For $(2)_1$, by Lemma 4.7, it is naturally right.

For $(3)_1$, we notice the following facts:

$$CRP_2(0011, 011) = 1001001,$$
$$CDP_2(0000, 1, 1001) = 100001100,$$
$$CDP_2(0000, 10, 01100) = \emptyset.$$

Therefore $h(011) = 2$.

Step two: Suppose that claims 1 through 3 are true for $q \leq k$. We will prove the case when $q = k + 1$.

$$2^k \text{ lines} \left\{ \begin{array}{l} 1\overbrace{00\cdots\cdots00}^{\text{0s of } 2^{k+1}-2}1\overbrace{000\cdots\cdots\cdots\cdots00}^{\text{0s of } 2^{k+1}-1}1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad 100001\cdots\cdots\cdots\cdots\cdots\cdots001 \\ \quad\; 10010001\cdots\cdots\cdots\cdots10001 \\ \quad\;\; 11\underbrace{0101\cdots\cdots\cdots\cdots0101}_{\text{01s of } 2^k} \end{array}\right.$$

Figure 6. Applying the inductive hypothesis where $l = 1$.

$$2^k \text{ lines} \left\{ \begin{array}{l} 1\overbrace{00\cdots\cdots00}^{\text{0s of } 2^{k+1}-2}1\overbrace{000\cdots\cdots\cdots\cdots00}^{\text{0s of } 2^{k+1}-1}1\overbrace{00\cdots\cdots\cdots\cdots\cdots\cdots0}^{\text{0s of } 2^{k+1}+l-2}1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad 100001\cdots\cdots\cdots\cdots\cdots\cdots0010\cdots\cdots\cdots\cdots\cdots\cdots001 \\ \quad\; 10010001\cdots\cdots\cdots\cdots100010\cdots\cdots100\cdots\cdots\cdots01 \\ \quad\;\; 11\underbrace{0101\cdots\cdots\cdots\cdots0101}_{\text{01s of } 2^k}\quad\underbrace{01\cdots\cdots01}_{\text{01s of } 2^k-1}\underbrace{00\cdots\cdots0}_{\text{0s of } l}1 \end{array}\right.$$

Figure 7. Applying the inductive hypothesis again.

For $(1)_{k+1}$, by the inductive hypothesis $(1)_k$, we have Figure 6. Hence

$$MR_2(EB(CP_2((00)^{2^k-1}11))) = RSB_2((01)^{2^k-1}11),$$
$$MR_2(EB(CRP_2((00)^{2^k-1}11))) = RSB_2((01)^{2^k}).$$

By Proposition 10 and the inductive hypothesis $(1)_k$, we have Figure 7. That is,

$$CRP_2((00)^{2^k-1}11,\ (01)^{2^{k+1}-1}0^l1) = 10^{2^{k+1}-2}10^{2^{k+1}-1}10^{2^{k+1}+l-2}1.$$

We will prove

$$CDP_2(0000,\ 10^{2^k-2},\ 0^{2^k-2}10^{2^{k+1}-1}10^{2^{k+1}+l-2}1) = 10^{2^{k+1}}1(01)^{2^k-1}0^{2^{k+1}+l}1,$$

which is equivalent to proving that y, which is defined in Figure 8, does not contain the string 11.

In fact, if it is not right, y contains at least one defect. So πy has a prefix $(01)^t1$ ($0 \le t \le 2^k-2$). By $(2)_k$, $h(\pi y) \le h((01)^t) < 2^k-1$. But on the other hand, y must satisfy $h(\pi y) \ge 2^k - 1$, which is impossible. Then we have Figure 9. Again using the inductive hypothesis $(1)_k$, we can prove $(1)_{k+1}$. For $(2)_{k+1}$, let $h((01)^t1) = T$, by $(2)_k$ and $(3)_k$, $T \le 2^{k+1}-1$. There exists a string y with length $2t+1$, such that $f^T(1(00)^T1y) = 11(01)^t1$. Let $A = MR_{2t+1}(EB(1(00)^T1y, T))$. See Figure 10.

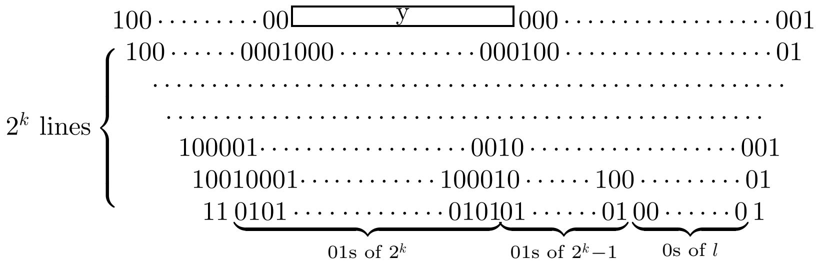

Figure 8. Define a string y.

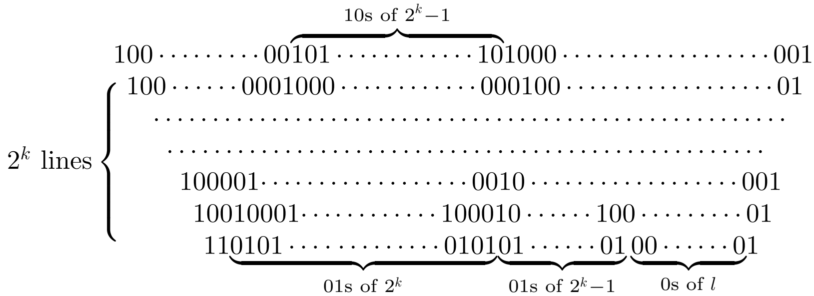

Figure 9. $y = (10)^{2^k-1}1$ and then use the inductive hypothesis again.

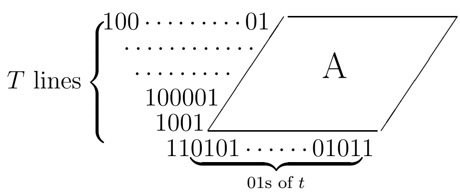

Figure 10. Define the right-skew-evolution block A.

By $(1)_{k+1}, f^{2^{k+1}-1}(10^{2^{k+2}-2}10^{2^{k+2}-1}1) = 11(01)^{2^{k+1}}$. So we have Figure 11, where $B = A$. Then $h((01)^{2^{k+1}+t}1) \geq T$.

If $h((01)^{2^{k+1}+t}1) \geq 2^{k+1} - 1$, then we have Figure 12. Also by Proposition 10, we have Figure 13 where $D = C$.

Then $h((01)^t 1) \geq 2^{k+1} - 1$, which is a contradiction, so $(2)_{k+1}$ holds.

For $(3)_{k+1}$, by $(1)_{k+1}$, we have Figure 14. By $(1)_{k+1}$, we have $h((01)^{2^{k+1}}1) = 2^{k+1} - 1 + h((00)^{2^{k+1}-2}1)$:

$$h((00)^{2^{k+1}-2}1) = 1 + \max\{h(\pi z) \mid f(00z00) = 10^{2^{k+1}-2}1\}.$$

Noting the remark of Lemma 4.10, it follows that

$$h((00)^{2^{k+1}-2}1) = 1 + \max\{h(x) \mid x = (01)^t 1, t = 0, 1, \ldots, 2^{k+1} - 3\}.$$

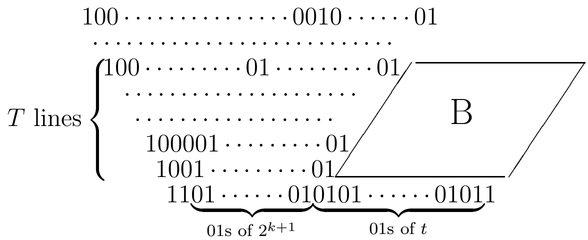

Figure 11. Define the right-skew-evolution block B=A.

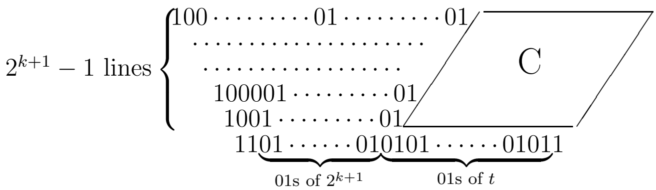

Figure 12. Define the right-skew-evolution block C.

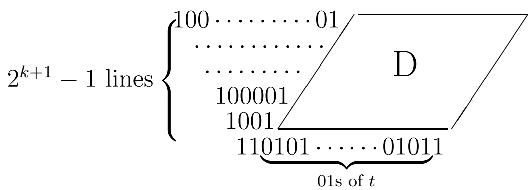

Figure 13. Define the right-skew-evolution block D=C.

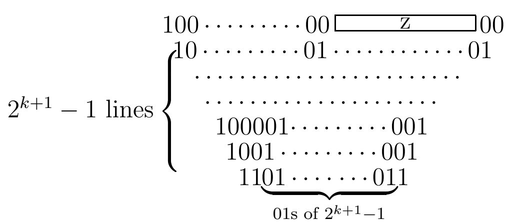

Figure 14. Define the string z.

By $(2)_k$ and $(3)_k$, we have

$$h((00)^{2^{k+1}-2}1) = 1 + 2^{k+1} - 2 = 2^{k+1} - 1.$$

Hence

$$h((01)^{2^{k+1}-1}1) = 2^{k+1} - 1 + 2^{k+1} - 1 = 2^{k+2} - 2.$$

So $(3)_{k+1}$ holds.

Up to now, we complete the second step and complete the proof of this lemma. ∎

B. Proof of Proposition 8

We first present some useful lemmas and also use f as its local rule. From the definition of the CA: $(F(x))_i = x_{i+1}x_{i+2}$, we can know a symbol is 1 if and only if its upper-right two consecutive symbols are both 1s in temporal evolution. Extending this result, we can obtain the following lemma.

Lemma B.1.

1. $f^{-1}(1) = * * *11$.
2. (2) $f^{-n}(1) = \overbrace{*\cdots*}^{3n}1^{n+1}$.
3. $f^{-1}(101) = \emptyset$.
4. $f^{-n}(10^n1) = \emptyset$, where $*$ stands for 0 or 1.

Proof. All of these can be verified directly. ∎

Lemma B.2. $CP_1(1^n) = \overbrace{*\cdots*}^{2n-2}1^{2n-1}$, where each $*$ stands for 0 or 1.

Proof. This can be obtained by using Lemma B.1 repeatedly. ∎

Lemma B.3. Let k and $l > 0$, then

$$1^k0^l1 \in E_1 \Longleftrightarrow k \leq l.$$

Proof. This is equivalent to proving the two equations

$$CP_1(1^k0^k1) \neq \emptyset; \quad CP_1(1^{k+1}0^k1) = \emptyset.$$

We need to prove that there exists a string in $CP_1(1^k0^k1)$ and every string in $CP_1(1^k0^k1)$ contains a substring 101. In fact, every string in $CP_1(10^k1)$ has a substring $10^m1 (1 \leq m \leq k)$. But every string in $f^{-j}(10^m1)$ has a substring $10^q1 (1 \leq q \leq m-j)$, where $0 < j < m$. Therefore, every string in $CP_1(1^j0^k1)$ has a substring $10^q1 (1 \leq q \leq m-j+1)$. So every string in $CP_1(1^k0^k1)$ contains a substring 101. Clearly, the string $0^{4k}1^{2k-1}01^{2k+1}$ which contains the substring 101 is in $CP_1(1^k0^k1)$. Then by Lemma B.1, $CP_1(1^{k+1}0^k1)$ must be empty. ∎

Using this lemma, we can obtain the irregularity of E_1 by applying the Myhill–Nerode theorem [18]. But the proof is omitted because we directly prove a stronger result: E_1 is not a CFL. Similar to the above proof, we can also obtain every string in $CP_1(0^j1^k0^l1)$ that has a substring 101, where $l = j + k$. Then we have Lemmas B.4 and B.5.

Lemma B.4. Let j, k, and $l > 0$, then

$$0^j 1^k 0^l 1 \in E_1 \Longleftrightarrow j + k \leq l.$$

Proof. Similar to Lemma B.3. ■

Lemma B.5. Let i, j, k, and $l > 0$, then

$$1^i 0^j 1^k 0^l 1 \in E_1 \Longleftrightarrow i \leq j \text{ and } i + j + k \leq l.$$

Proof. Similar to Lemma B.4. ■

Proof of Proposition 8. Now we can prove Proposition 8. Similar to the proof of Proposition 3, let

$$L = 1^+ 0^+ 1^+ 0^+ 1 \cap E_1.$$

First we prove L is not a CFL. Otherwise, by the Ogden lemma [28], there exists a constant $N > 0$, such that for each string $s \in L, |s| > N$, if any N or more positions in s are designated as distinguished, then there exists a decomposition of s by $s = uvwxy$ which satisfies:

1. either each of u, v, w contains distinguished positions, or each of w, x, y contains distinguished positions;
2. vwx contains at most k distinguished positions;
3. $uv^i wx^i y \in L (i \geq 0)$.

Now we take $s = 1^N 0^N 1^N 0^{3N} 1 \in L$, and designate the prefix 1^N as our N distinguished positions:

$$s = \overbrace{\underbrace{111\cdots\cdots\cdots\cdots111}_{\text{1s of } N}}^{\text{distinguished positions}}\underbrace{000\cdots\cdots000}_{\text{0s of } N}\underbrace{111\cdots\cdots111}_{\text{1s of } N}\underbrace{000\cdots\cdots\cdots000}_{\text{0s of } 3N}1.$$

By the definition of L, we know that v and x are in $0^* \cup 1^*$. By 2, w must contain distinguished positions, therefore v must be in 1^*. Let $v = 1^{t_1}$ and $|x| = t_2$. By 1, $t_1 + t_2 \geq 1$. Then $uv^2 wx^2 y \in L$ must be the following four possibilities:

- $1^{N+t_1+t_2} 0^N 1^N 0^{3N} 1$
- $1^{N+t_1} 0^{N+t_2} 1^N 0^{3N} 1$
- $1^{N+t_1} 0^N 1^{N+t_2} 0^{3N} 1$
- $1^{N+t_1} 0^N 1^N 0^{3N+t_2} 1$

but all four of these possibilities contradict Lemma B.5. (In fact, the first three are trivial, and the last one, $1^{N+t_1} 0^N 1^N 0^{3N+t_2} 1 \in L$, implies $t_1 = 0$. But neither v nor x contain distinguished positions.)

Therefore, L does not satisfy the Ogden lemma. So L is not a CFL. By Proposition 4, E_1 is also not a CFL. Hence, for all $n > 0$, E_n are not CFLs but CSLs. ■

Through careful and lengthy discussions, we can obtain a full image of E_1.

Proposition B.1. Let $n_0, n_1, \ldots, n_{2k} > 0$, then we have

$$1^{n_1} 0^{n_2} 1^{n_3} 0^{n_4} \ldots 0^{n_{2k}} 1 \in E_1 \iff \sum_{i=1}^{j} n_i \leq n_{j+1} (j = 1, 3, \ldots 2k-1)$$

$$0^{n_0} 1^{n_1} 0^{n_2} 1^{n_3} \cdots 0^{n_{2k}} 1 \in E_1 \iff \sum_{i=0}^{j} n_i \leq n_{j+1} (j = 1, 3, \ldots 2k-1).$$

References

[1] L. Lam, *Introduction to Nonlinear Physics* (Springer-Verlag, New York, 1997).

[2] K. Culik, Y. Pachl, and S. Yu, "On the Limit Set of CA," *SIAM J*, **18** (1989) 831–842.

[3] K. Culik and S. Yu, "Undecidability of CA Classification Schemes," *Complex Systems*, **2**(2) (1988) 177–190.

[4] J. Kari, "Rice's Theorem for the Limit Sets of Cellular Automata," *Theoretical Computer Science*, **127** (1984) 229–254.

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

[6] S. Wolfram, *Theory and Application of Cellular Automata*, (World Scientific, Singapore, 1986).

[7] M. Morse, "Recurrent Geodesics on a Surface of Negative Curvatures," *Transactions of the AMS*, **22** (1921) 84–100.

[8] M. Morse, "A One-to-one Representation of Geodesics on a Surface of Negative Curvature," *American Journal of Mathematics*, **43** (1921) 33–51.

[9] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, second edition (Addison-Wesley, Redwood City, 1989).

[10] S. Smale, "Differentiable Dynamical Systems," *Bulletins of the American Mathematical Society*, **73** (1967) 747–781.

[11] P. Collet and J. P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Boston, 1980).

[12] B. L. Hao, "Symbolic Dynamics and Characterization of Complexity," *Physica D*, **51** (1991) 161–176.

[13] Y. Wang, L. Yang, and H. M. Xie, "Complexity of Unimodal Maps with Kneading Sequences," *Nonlinearity*, **12** (1999) 1151–1176.

[14] H. M. Xie, *Grammatical Complexity and One-dimensional Dynamical Systems* (World Scientific, Singapore, 1996).

[15] M. Delorme and J. Mazoyer, *Cellular Automata: A Parallel Model* (Kluwer Academic Publishers, Dordrecht/Boston/London, 1999).

[16] R. H. Gilman, "Notes on Cellular Automata," preprint, 1998.

[17] P. Kůrka, "Languages, Equicontinuity and Attractors in Cellular Automata," *Ergodic Theory & Dynamic Systems*, **17** (1997) 417–433.

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

[19] E. Jen, "Aperiodicity in One-dimensional Cellular Automata," *Physica D*, **45** (1990) 3–18.

[20] F. Blanchard, P. Kůrka, and A. Maass, "Topological and Measure-theoretic Properties of One-dimensional Cellular Automata," *Physica D*, **103** (1997) 86–99.

[21] P. Grassberger, "New Mechanism for Deterministic Diffusion," *Physical Review A*, **28**(6) (1983) 3666–3667.

[22] J. E. Hanson and J. P. Crutchfield, "The Attractor-basin Portrait of a Cellular Automaton," *Journal of Statistical Physics*, **66** (1992) 1415–1463.

[23] O. Martin, A. M. Odlyzko, and S. Wolfram, "Algebraic Properties of Cellular Automata," *Communications in Mathematical Physics*, **93** (1984) 219–258.

[24] N. Boccara, J. Nasser, and M. Roger, "Particlelike Structures and their Interactions in Spatiotemporal Patterns Generated by One-dimensional Deterministic Cellular-automaton Rules," *Physical Review A*, **44**(2) (1991) 866–875.

[25] N. Boccara, J. Nasser, and M. Roger, "Ideal Patterns and Defects of Some Families of One-dimensional Deterministic Cellular Automata," *Complex Dynamics*, (1993) 113–124.

[26] K. Eloranta and E. Nummelin, "The Kink of Cellular Automaton Rule 18 Performs a Random Walk," *Journal of Statistical Physics*, **69**(5–6) (1992) 1131–1136.

[27] E. Goles, A. Maass, and S. Martinez, "On the Limit Set of Some Universal Cellular Automata," *Theoretical Computer Science*, **110** (1993) 53–78.

[28] W. Ogden, "A Helpful Result for Proving Inherent Ambiguity," *Math System Theory*, **2** (1968) 191–194.