



**HAL**  
open science

## Construction of Some Nonautomatic Sequences by Cellular Automata

Irène Marcovici, Thomas Stoll, Pierre-Adrien Tahay

► **To cite this version:**

Irène Marcovici, Thomas Stoll, Pierre-Adrien Tahay. Construction of Some Nonautomatic Sequences by Cellular Automata. AUTOMATA 2018 - 24th International Workshop on Cellular Automata and Discrete Complex Systems, Jun 2018, Ghent, Belgium. pp.113-126, 10.1007/978-3-319-92675-9\_9. hal-01824876

**HAL Id: hal-01824876**

**<https://inria.hal.science/hal-01824876v1>**

Submitted on 27 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Construction of Some Nonautomatic Sequences by Cellular Automata

Irène Marcovici, Thomas Stoll, and Pierre-Adrien Tahay

Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France  
irene.marcovici@univ-lorraine.fr, thomas.stoll@univ-lorraine.fr,  
pierre-adrien.tahay@univ-lorraine.fr

**Abstract.** It is known that if  $p$  is a prime number, the columns of linear CA are  $p$ -automatic sequences and all  $p$ -automatic sequences can be realized by some linear CA with memory. We give some constructions of (nonlinear) CA that realize certain nonautomatic sequences. First, we show through a recoding that from a construction with additional symbols, we can construct a CA using only the symbols occurring in the sequence. This answers a question posed by Rowland and Yassawi. Then, we propose a construction for the characteristic sequence of the integer polynomials, which are nonautomatic sequences by the Minsky–Papert criterion. We also provide a construction based on the indicator of Fibonacci numbers for the Fibonacci word, which is an emblematic nonautomatic sequence.

**Keywords:** cellular automata, automatic sequences, nonautomatic sequences, computability, polynomials, Fibonacci word

## 1 Introduction

A one-dimensional *cellular automaton* (CA) is a dynamical system  $(\mathcal{A}^{\mathbb{Z}}, F)$ , where  $\mathcal{A}$  is a finite set, and where the map  $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$  is defined by a local rule which acts uniformly and synchronously on the configuration space. Precisely, there exists an integer  $r \geq 0$  called the *radius* of the CA, and a *local rule*  $f : \mathcal{A}^{2r+1} \rightarrow \mathcal{A}$  such that  $\forall x \in \mathcal{A}^{\mathbb{Z}}, \forall k \in \mathbb{Z}, F(x)_k = f((x_{k+i})_{-r \leq i \leq r})$ . By the Curtis-Hedlund-Lyndon theorem, a map  $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$  is a CA if and only if it is continuous with respect to the product topology, and it commutes with the shift map  $\sigma$  defined by  $\sigma(x)_k = x_{k-1}, \forall x \in \mathcal{A}^{\mathbb{Z}}, \forall k \in \mathbb{Z}$ .

Let  $\mathcal{A}$  be a finite set containing an element denoted by 0, and let  $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$  be a CA. We say that  $F$  is *0-quiescent* if  $F(0^{\mathbb{Z}}) = 0^{\mathbb{Z}}$ . A configuration  $x \in \mathcal{A}^{\mathbb{Z}}$  is said to be *finite* if the set  $\{k \in \mathbb{Z} : x_k \neq 0\}$  is finite. We denote by  $\mathcal{C}_0(\mathcal{A})$  the set of finite configurations of  $\mathcal{A}^{\mathbb{Z}}$ .

In the present paper, we study the set of sequences of  $\mathcal{A}^{\mathbb{N}}$  that can occur as a column sequence in the space-time diagram of some one-dimensional CA, with a finite initial configuration. Precisely, we study the set  $\mathcal{S}$  of sequences having the form  $(F^n(x)_0)_{n \geq 0}$ , where  $F$  is a 0-quiescent CA on  $\mathcal{A}^{\mathbb{Z}}$  and where  $x \in \mathcal{C}_0(\mathcal{A})$ . Thus,

$$\mathcal{S} = \left\{ (F^n(x)_0)_{n \geq 0} \in \mathcal{A}^{\mathbb{N}} : F \text{ is a 0-quiescent CA on } \mathcal{A}^{\mathbb{Z}} \text{ and } x \in \mathcal{C}_0(\mathcal{A}) \right\}.$$

With the terminology introduced by Mazoyer and Terrier, this corresponds to the notion of *Fischer’s produced* sequences [11, Def. 5]. We will use in particular the following closure property of  $\mathcal{S}$ : if  $u \in \mathcal{S}$  then, for any sequence  $v$  that differs from  $u$  in a finite number of positions only, we also have  $v \in \mathcal{S}$ . Mazoyer and Terrier have also investigated further properties of the set  $\mathcal{S}$  above, and proved that it is stable by different operations. Here, we are particularly interested

in clarifying the relationship to (non)automaticity of the underlying sequences, and we propose effective constructions for some specific families of sequences.

When  $\mathcal{A} = \mathbb{F}_p$ , a CA  $F$  is said to be *linear* if it defines an  $\mathbb{F}_p$ -linear map. Equivalently, this means that the local rule of  $F$  can be written as  $f((x_i)_{-r \leq i \leq r}) = \sum_{i=-r}^r \alpha_i x_i$ , for some  $r \geq 0$  and coefficients  $\alpha_i \in \mathbb{F}_p$ .

A deterministic finite automaton with output (DFAO) is a 6-tuple  $(\mathcal{Q}, \Sigma_k, \delta, s_0, \mathcal{A}, \omega)$  where  $\mathcal{Q}$  is a finite set of *states*,  $\Sigma_k = \{0, 1, \dots, k-1\}$ ,  $s_0 \in \mathcal{Q}$  is the initial state,  $\mathcal{A}$  is a finite alphabet,  $\omega : \mathcal{Q} \rightarrow \mathcal{A}$  is the output function, and  $\delta : \mathcal{Q} \times \Sigma_k \rightarrow \mathcal{Q}$  is the transition function. A sequence  $(u_n)_{n \geq 0}$  of elements in  $\mathcal{A}$  is  $k$ -automatic if there is a DFAO  $(\mathcal{Q}, \Sigma_k, \delta, s_0, \mathcal{A}, \omega)$  such that  $u_n = \omega(\delta(s_0, (n)_k))$  for all  $n \geq 0$ , where  $(n)_k$  is the word constituted by the digits of the standard base- $k$  representation of  $n$ . We refer to the book of Allouche and Shallit [1, Sections 5, 6] for the robustness of the concept of automatic sequences and their various characterizations.

In 1993, Litow and Dumas [10] proved that if  $p$  is a prime number, every column of a linear CA over  $\mathbb{F}_p$  (with a finite initial configuration) is  $p$ -automatic. In 2015, Rowland and Yassawi [14] established a converse statement. Namely, for any  $p$ -automatic sequence, they show how to obtain that sequence as a column of a linear CA with memory, with eventually periodic initial conditions.

Let us now consider nonautomatic sequences. By the results above, they cannot be realized by a linear CA. However, for some nonautomatic sequences having a simple algorithmic description, we can reasonably think that they can be obtained with other kinds of CA. Throughout the present article, we will focus on binary sequences, that is, we will assume that the set of symbols is  $\mathcal{A} = \{0, 1\}$ . Then, a sequence  $u \in \mathcal{A}^{\mathbb{N}}$  is characterized by the set of positions  $k \in \mathbb{N}$  such that  $u_k = 1$ . For a subset  $E$  of  $\mathbb{N}$ , we denote by  $u = \mathbf{1}_E$  the sequence defined by  $u_k = 1 \iff k \in E$ .

Let  $f : \mathbb{N} \rightarrow \mathbb{N}$  be an increasing function and  $\pi_f(x) = \#\{n : f(n) \leq x\}$ . Minsky and Papert [12] showed that if  $\lim_{x \rightarrow \infty} \pi_f(x)/x = 0$  and  $\lim_{n \rightarrow \infty} f(n+1)/f(n) = 1$ , then the sequence  $u = \mathbf{1}_{f(\mathbb{N})}$  is nonautomatic. For example, if  $P \in \mathbb{Q}[X]$  is a polynomial of degree  $d \geq 2$  such that  $P(\mathbb{N}) \subset \mathbb{N}$ , and  $P$  is strictly increasing on  $\mathbb{N}$ , then the Minsky–Papert criterion is satisfied (since  $\pi_P(x)$  is of order  $x^{1/d}$ ), meaning that the sequence  $\mathbf{1}_{P(\mathbb{N})}$  is nonautomatic. Note that the case of nonrecognizability of the squares was known before the work of Minsky and Papert, namely, it follows from work of Büchi [3] concerning second order weak arithmetic. Ritchie [13] gave a short proof of this fact in the case for squares in base-2 expansions (see also [1, Section 5.5, p.166]). However, as shown by Mazoyer and Terrier [11] and by Delacourt, Poupet, Sablik and Theyssier [5], there exists a simple CA giving the characteristic function of the squares in its column, using only three additional symbols corresponding to walls and left and right signals.

Another example of nonautomatic sequence is given by  $u = \mathbf{1}_{\mathcal{P}}$ , where  $\mathcal{P}$  denotes the set of prime numbers. With the help of the prime number theorem, the Minsky–Papert criterion can also be used to prove the nonautomaticity of this sequence. This means that the primes cannot be recognized by a finite automaton. A first intricate construction of that sequence as a column of CA was provided by Fischer [6], then a simpler one was proposed by Korec [8], using only a few signals having different speeds.

In the present article, we describe two constructions of nonautomatic sequences, one that realizes the characteristic function of any polynomial of any arbitrary fixed degree  $d \geq 2$  (Section 3), and one that realizes the Fibonacci word (Section 4) which is an emblematic morphic and nonautomatic word.

Before presenting our constructions, we show that any sequence that can be obtained with the help of auxiliary symbols from a larger set  $\mathcal{B} \supset \mathcal{A}$  can also be obtained with a CA defined on  $\mathcal{A}^{\mathbb{Z}}$  (Prop. 1). Our proof is based on a binary recoding of the elements of  $\mathcal{B}$ , a technique already used for the construction of primes [8]. As a consequence of this result, in order to prove that a sequence belongs to  $\mathcal{S}$ , it is sufficient to provide a CA that uses additional states, and this is what we will do. In particular, this recoding answers a question asked by Rowland and Yassawi,

since it shows that any 3-automatic sequence on a 2-letter alphabet occurs as a column of a (nonlinear) 2-state space-time diagram [14, third question on p. 80].

In contrast with these results, we also give an elementary proof of the fact that there are computable sequences that lie outside of  $\mathcal{S}$ , meaning that the class of computable sequences is “richer” than the set of sequences obtained via CA (Prop. 5). We end with some open questions, asking whether some concrete sequences that are relevant in number theory or in combinatorics can be obtained by a CA or not.

## 2 Binary recoding

Let us recall that for the rest of the article, we will consider a binary set of symbols:  $\mathcal{A} = \{0, 1\}$ .

**Proposition 1.** *Let  $\mathcal{B}$  be a finite set of symbols such that  $\mathcal{B} \supseteq \mathcal{A}$ , let  $F : \mathcal{B}^{\mathbb{Z}} \rightarrow \mathcal{B}^{\mathbb{Z}}$  be a 0-quiescent CA, and let  $x \in \mathcal{B}^{\mathbb{Z}}$  be a finite configuration. If  $\forall n \in \mathbb{N}, F^n(x)_0 \in \mathcal{A}$ , then,  $(F^n(x))_{n \geq 0} \in \mathcal{S}$ .*

*Proof.* Without loss of generality, let us assume that  $\mathcal{B} = \{0, \dots, k-1\}$  for some  $k \geq 3$ . We recode the  $k$  symbols of the alphabet  $\mathcal{B}$  by binary words of length  $2k-1$ , using the map  $\tau : \mathcal{B} \rightarrow \mathcal{A}^{2k-1}$  defined as follows:

$$\tau(0) = \underbrace{00 \cdots 0}_{2k-1}, \quad \text{and for } i \in \{1, \dots, k-1\}, \tau(i) = \underbrace{10 \cdots 01}_{i-1} \underbrace{0 \cdots 0}_{2k-i-2}.$$

From a configuration  $x \in \mathcal{B}^{\mathbb{Z}}$ , the concatenation of the blocks  $\tau(x_i)$  for  $i \in \mathbb{Z}$  gives rise to a new configuration  $\tilde{x} \in \mathcal{A}^{\mathbb{Z}}$ , where we assume that the block  $\tau(x_0)$  starts at position 0 in the new configuration, meaning that  $(\tilde{x}_i)_{0 \leq i \leq 2k-2} = \tau(x_0)$ . With this recoding, it is possible to determine locally the boundaries between the blocks in the new configuration, except if we have a long range of 0’s. But in that case, since  $F$  is 0-quiescent, we know that the local function will output a 0. So, using this recoding, we can define a new CA  $G : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$  of radius  $(2k-1)r + 2k - 2 = (2k-1)(r+1) - 1$  that mimics on  $\mathcal{A}^{\mathbb{Z}}$  the action of the original CA  $F : \mathcal{B}^{\mathbb{Z}} \rightarrow \mathcal{B}^{\mathbb{Z}}$ , and  $G$  is itself a 0-quiescent CA. Since the first letter of  $\tau(0)$  and  $\tau(1)$  are respectively 0 and 1, if there is no occurrence of symbols from  $\mathcal{B} \setminus \mathcal{A}$  in the central column, then we will have:  $\forall n \in \mathbb{N}, F^n(x)_0 = G^n(x)_0$ .  $\square$

*Example 1.* In particular, if  $\mathcal{B} = \{0, 1, 2\}$ , the recoding  $\tau$  above is defined by:  $\tau(0) = 00000, \tau(1) = 11000, \tau(2) = 101000$ , and the construction answers a question asked by Rowland and Yassawi [14]. Indeed, it proves that any 3-automatic sequence  $(u_n)_{n \geq 0}$  on a binary alphabet occurs as a column of a (nonlinear) 2-state space-time diagram. On Fig. 1, we present an illustration for the 3-automatic sequence defined by  $u_n = \mathbf{1}_T$ , where  $T = \{3^n : n \in \mathbb{N}\}$ . We can use the method of Rowland and Yassawi [14] to obtain this sequence as a column of a linear CA with memory 2, with a finite initial configuration. Let us denote by  $\phi(m, n)$  the value of the cell on column  $m \in \mathbb{Z}$  and line  $n \in \mathbb{N}$ . The local rule obtained to generate the space-time diagram is given, for all  $m \in \mathbb{Z}$  and  $n \geq 2$ , by:  $\phi(m, n+2) = \phi(m+1, n) + \phi(m-2, n)$ . In particular there are only 0’s on all lines of even rank. For a better readability, we represent 0’s by empty cells. Time axis is oriented upward. The four first lines correspond to initial conditions, and the sequence  $(u_n)_{n \geq 0}$  appears in the red framed column. Fig. 2 shows the space-time diagram after recoding.

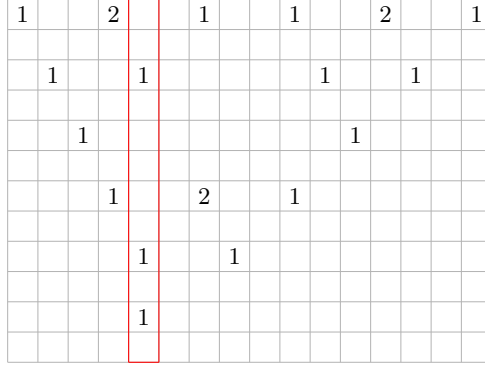


Fig. 1: Construction of the powers of three.

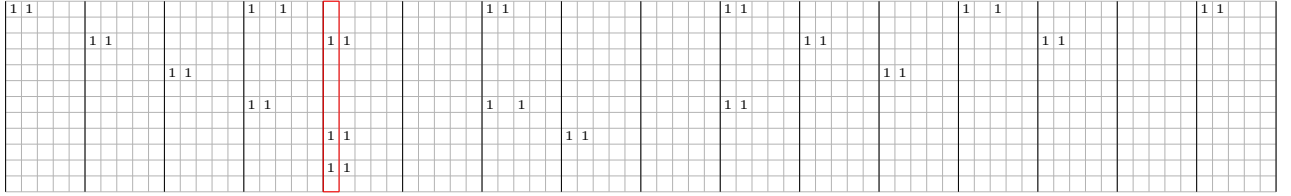


Fig. 2: Construction of the powers of three, after recoding.

*Remark 1.* For a larger alphabet  $\mathcal{A} = \{0, \dots, q - 1\}$  with  $q > 2$ , if  $\mathcal{B}$  is a finite set of symbols such that  $\mathcal{B} \supsetneq \mathcal{A}$ , it is also possible to recode the elements of  $\mathcal{B}$  by blocks made by elements of  $\mathcal{A}$  only. Let us assume that  $\mathcal{B} = \{0, \dots, k - 1\}$ , with  $k > q$ . We recode the  $k$  symbols of the alphabet  $\mathcal{B}$  by words of length  $2k - 1$ , using the map  $\omega : \mathcal{B} \rightarrow \mathcal{A}^{2k-1}$  defined as follows:

$$\begin{aligned} \omega(0) &= \underbrace{00 \cdots 0}_{2k-1}, \\ \omega(i) &= i \underbrace{0 \cdots 0}_{i-1} 1 \underbrace{0 \cdots 0}_{2k-i-2} \quad \text{for } i \in \{1, \dots, q - 1\}, \\ \text{and } \omega(i) &= 0 \underbrace{0 \cdots 0}_{i-1} 1 \underbrace{0 \cdots 0}_{2k-i-2} \quad \text{for } i \in \{q, \dots, k - 1\}. \end{aligned}$$

*Remark 2.* In order to optimize the size of the blocks in the recoding, we can use some further results on *non-overlapping codes* [2]. Two words  $u$  and  $v$  (not necessarily distinct) are said to be overlapping if a non-empty proper prefix of  $u$  is equal to a non-empty proper suffix of  $v$ , or the converse. A code  $C \subseteq \mathcal{A}^n$  is non-overlapping if for all (not necessarily distinct)  $u, v \in C$ , the words  $u$  and  $v$  are non-overlapping. For an integer  $k \in \{0, \dots, n - 1\}$ , we can build a non-overlapping code  $C_n^{(k)}$  of length  $n$  by considering the family of words  $c \in \mathcal{A}^n$  satisfying the following properties:  $c_i = 0$  for  $1 \leq i \leq k$ ,  $c_{k+1} = 1$ ,  $c_n = 1$ , and the sequence  $c_{k+2}, c_{k+3}, \dots, c_{n-1}$  does not contain  $k$  consecutive 0's (see [4, 7, 9]). Let us denote by  $F_n^{(k)}$  the number of codewords obtained, that is, the cardinality of  $C_n^{(k)}$ . One can check that the sequence  $(F_n^{(k)})_{n \geq 1}$  satisfies the recurrence relation:  $F_{n+k}^{(k)} = F_{n+k-1}^{(k)} + F_{n+k-2}^{(k)} + \cdots + F_{n+1}^{(k)} + F_n^{(k)}$ , which corresponds to the  $k$ -step Fibonacci sequence.

If  $\mathcal{B}$  is an alphabet with  $k$  letters, we can match each letter  $k \geq 1$  with a codeword of a non-overlapping code, and match 0 with the block of  $0 \cdots 0$  of the same size. The set obtained is no more exactly a non-overlapping code, but it is enough for our purpose, since we can determine locally the boundaries between blocks as soon as there is one occurrence of 1, and the CA  $F$  is assumed to be 0-quiescent. For example, for a length  $n = 9$ , if we consider the words starting with two 0's, with a 1 at the third position and a 1 at the last position, the construction above provides 13 codewords, and with the recoding of the symbol 0 by a block of 0's, this allows to handle an alphabet of 14 letters. In comparison, with the map  $\tau$  introduced earlier, as soon as the alphabet contains more than 5 letters, the length of the recoding is already larger than 9.

*Remark 3.* Prop. 1 can be strengthened in the following way. Let  $\pi : \mathcal{B} \rightarrow \mathcal{A}$  define a symbol-to-symbol projection from  $\mathcal{B}$  to  $\mathcal{A}$ . Assume that  $u \in \mathcal{B}^{\mathbb{N}}$  occurs as a column of a CA on  $\mathcal{B}$  with a finite initial configuration, and consider the sequence  $v = \pi(u)$ , defined by  $v_n = \pi(u_n)$  for all  $n \in \mathbb{N}$ . Then, the sequence  $v$  occurs as a column of a CA on  $\mathcal{A}$  with a finite initial configuration.

### 3 Polynomial sequences

In this section we present a CA that realizes the indicator function of any given polynomial  $P$  (under suitable conditions) in column 0, thus showing that  $\mathbf{1}_{P(\mathbb{N})} \in \mathcal{S}$ . As seen in Section 1, these are typical sequences that are nonautomatic. By the results of Mazoyer and Terrier [11], these sequences are constructible by CA, as the underlying set can be obtained by successive addition and multiplication operations. Our aim here is to generalize the geometric construction for squares due to Mazoyer and Terrier [11, Fig. 5], and used again by Delacourt, Poupet, Sablik and Theyssier [5, Section 3], that involves the notion of walls (*i.e.* signals of speed 0) and of directed left and right signals. While the constructions proposed by Mazoyer and Terrier for polynomials of higher degree use several signals zigzaging between the trajectories of other signals and the resulting sequence, our construction allows to realize  $\mathbf{1}_{P(\mathbb{N})}$  by hits of just one single signal at column 0.

To start with, let us shortly recall the construction concerning squares [5]. A signal (represented by an arrow, see Fig. 3) starts on level 1 to the NE-direction (right signal). Every time it meets the wall (represented by a vertical bar), the wall gets shifted by one cell to the right and the signal changes its direction from right to left. When the signal meets the 0-column, then a 1 appears, and it changes again its direction from left to right.

This construction can be generalized. We first show a way to generate from a constructed sequence the summatory sequence in column 0. Note that Mazoyer and Terrier proved a similar result for their geometric construction [11, Corollary 2].

**Proposition 2.** *Let  $f : \mathbb{N} \rightarrow \mathbb{N}$  be an ultimately strictly increasing function, and let  $g$  denote its summatory sequence defined by  $g(n) = \sum_{k=0}^n f(k)$  for all  $n \geq 0$ . Suppose that  $u = \mathbf{1}_{f(\mathbb{N})}$  belongs to  $\mathcal{S}$ . Then  $v = \mathbf{1}_{g(\mathbb{N})}$  also belongs to  $\mathcal{S}$ .*

*Proof.* Let us call  $F$  the CA that constructs  $u = \mathbf{1}_{f(\mathbb{N})}$  and denote by  $G$  the CA that we are going to construct for  $v = \mathbf{1}_{g(\mathbb{N})}$ . For the proof, it is sufficient to treat the case when  $f$  is strictly increasing for all  $n \geq 0$ , since we can treat a finite number of exceptions by introducing some appropriate initial conditions for  $G$ .

Suppose first that the following condition holds:

$$(\star) \quad 2g(i) > f(i+1) \text{ for all } i \geq 1.$$

In a second step, we will show how to transform the construction in order to handle the case when  $(\star)$  is not satisfied. But for the sake of clearness, we first assume that  $(\star)$  holds. This condition allows to split the construction into three steps.

*(i) Marking the columns:*

The first step consists in marking the columns  $f(0), f(1), \dots, f(n), \dots$ . For this, we start from the construction of  $u = \mathbf{1}_{f(\mathbb{N})}$  and for all integers  $i$ , we shift all cells of line  $i$  exactly  $i$  cells to the right, which corresponds to considering the space-time diagram of the CA  $\sigma \circ F$  (with the same initial configuration). Note that if the CA  $F$  constructing  $u$  is made of signals of different speeds, then we increase by 1 the speeds of all signals in order to get this marking of the column. In the construction of  $u = \mathbf{1}_{f(\mathbb{N})}$ , column 0 is marked by a 1 at lines  $f(0), f(1), \dots, f(n), \dots$ , therefore in the space-time diagram of  $G$ , lines  $f(0), f(1), \dots, f(n), \dots$  are marked but now in the columns  $f(0), f(1), \dots, f(n), \dots$ , that is on the diagonal of the new space-time diagram. Since  $f$  is strictly increasing, these columns are distinct.

*(ii) Constructing the walls:*

Next, we define a new signal of speed 1 that runs on the diagonal of the space-time diagram. The intersections between this signal and the (marked) columns  $f(0), f(1), \dots, f(n), \dots$  send a new signal of speed 0 which are the walls.

*(iii) Passing from one hit on the column 0 to the next one:*

Suppose we have a 1 on some line in column 0. We now show how to generate the next hit on column 0. Define a signal of speed 2 that moves to the right. When this signal meets a wall, that is a column  $f(i)$  for some integer  $i$ , the signal changes direction to the left (same absolute speed). When this signal meets the column 0 we mark a new 1. Depending on the parity, there are two ways for such a signal and a wall to meet. Either the two signals are in two columns side by side and then, the signal of speed  $-2$  starts in the same column of the last symbol of the signal of speed 2 but one line above, or there is a cell between the two signals and then, the signal of speed  $-2$  starts in the same column of the last symbol of the signal of speed 2 but two lines above. In both cases, we annihilate the wall after the signal has passed. Since  $f$  is strictly increasing, we do not have two walls in a same column, and the procedure is well-defined.

Let us now assume that we have so far constructed the first lines of the CA, in other words, we have a 1 in column 0 at lines  $f(0), f(0) + f(1), \dots, f(0) + \dots + f(i)$  for some  $i$ . We use the previous rules to continue the process. In particular, the 1 on the column 0 at line  $f(0) + \dots + f(i)$  sends a signal of speed 2 to the intersection with the wall in column  $f(i+1)$ . If  $f(i+1)$  is odd, we are in the case where the two signals are in two columns side by side. The distance crossed by the signal of speed 2 is  $\frac{f(i+1)-1}{2}$  from the initial 1 and the signal of speed  $-2$  runs the same distance. We place the new 1 in column 0 one line above. Therefore, this new 1 is on the line:

$$f(0) + \dots + f(i) + 2 \left( \frac{f(i+1) - 1}{2} \right) + 1 = f(0) + \dots + f(i) + f(i+1).$$

If  $f(i+1)$  is even, we are in the other case and the new 1 in column 0 is on the line:

$$f(0) + \dots + f(i) + \left( \frac{f(i+1)}{2} - 1 \right) + 1 + \left( \frac{f(i+1)}{2} - 1 \right) + 1 = f(0) + \dots + f(i) + f(i+1).$$

Note that the condition  $(\star)$ , which can be rewritten as:

$$f(0) + f(1) + \dots + f(i) + \frac{f(i+1)}{2} > f(i+1),$$

ensures that the wall at column  $f(i+1)$  has already been created when the signal of speed 2 arrives at column  $f(i+1)$ . Depending on the exact form of the function  $f$ , the condition  $(\star)$

is not always satisfied. In the case where  $(\star)$  is satisfied for  $n \geq n_0$ , in order to cope with this problem, one possibility consists in considering the first lines of the space-time diagram as initial conditions (which is not a loss of generality, since it amounts to introducing some auxiliary symbols in the initial configuration only, in such a way that they will fix the first few lines of the space-time diagram). Another possibility is to move the construction of the points  $f(i)$  under the main diagonal several lines lower, so that the walls appear sufficiently early. We can then use the procedure given above exactly in the same way. If there are infinitely many values of  $n$  such that  $(\star)$  is not satisfied, we use another strategy: we change the signals of speed 2 and  $-2$  at the step  $(iii)$  to signals of speed 1 and  $-1$ . In this case, the sequence that is constructed is not  $\mathbf{1}_{g(\mathbb{N})}$  but  $\mathbf{1}_{2g(\mathbb{N})}$ . Nevertheless, we can recover  $\mathbf{1}_{g(\mathbb{N})}$  with the CA  $G^2$ .  $\square$

*Example 2 (Construction of the characteristic sequence of the sum of the squares, see Fig. 3).* Let  $f(n) = n^2$  and  $g(n) = \sum_{k=0}^n k^2 = n(n+1)(2n+1)/6$ . Here, the condition  $(\star)$  is satisfied for  $n \geq 2$ , and for the construction of  $\mathbf{1}_{g(\mathbb{N})}$ , we consider the first three lines of the space-time diagram as an initial condition.

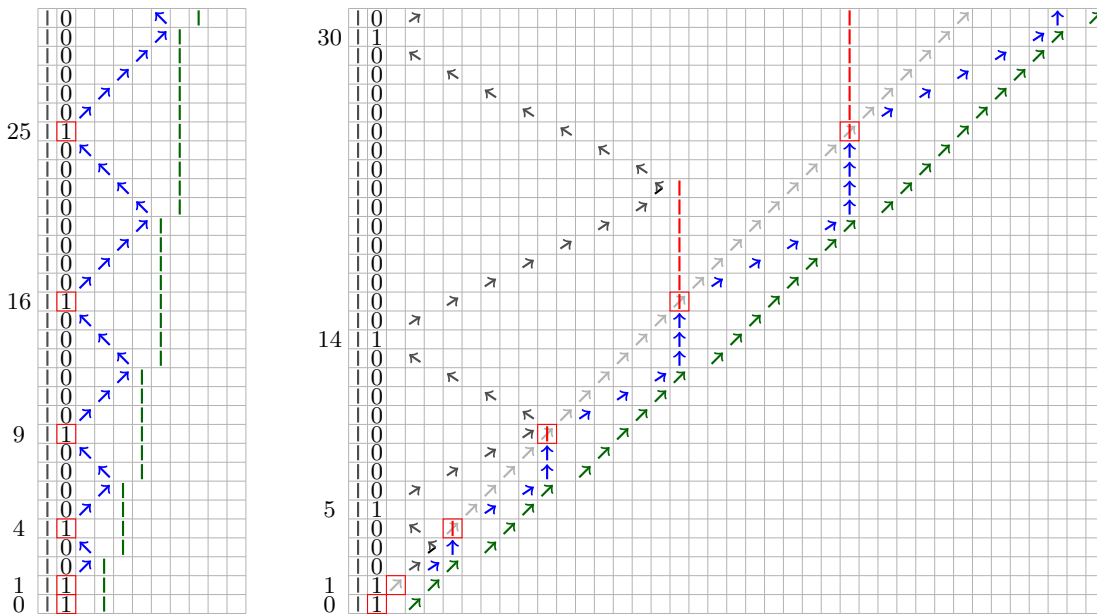


Fig. 3: CA for squares [5] (left) and for the sum of squares (right).

In Example 2, a polynomial of degree 3 is realized. This suggests that the method can be generalized to obtain the characteristic sequence of any polynomial with higher degree as a column of a CA. To begin with, we show by an explicit construction that the characteristic sequence of the cubes belongs again to  $\mathcal{S}$ .

*Example 3 (Construction of the characteristic sequence of the cubes, see Fig. 4).* For all  $n \in \mathbb{N}$  we have  $(n+1)^3 - n^3 = 3n^2 + 3n + 1$ . With the construction of the squares, it is straightforward



to obtain the characteristic sequence of  $3n^2$ : In fact, it suffices to divide the speed of signals of speed  $\pm 1$  by 3 (see Fig. 4, left). The red, blue and black cells in column 0 correspond to the term  $3n$ . Thus, the construction of the characteristic sequence of the polynomial  $3n^2 + 3n + 1$  follows. The eight first lines are initial conditions. Now, to obtain the characteristic sequence of the cubes (see Fig. 4, right), we observe that for all  $n \geq 1$ ,  $\sum_{k=0}^{n-1} ((k+1)^3 - k^3) = n^3$  and we use the construction given in the proof of Prop. 2. The first five lines are initial conditions. The construction under the diagonal is moved three lines down.

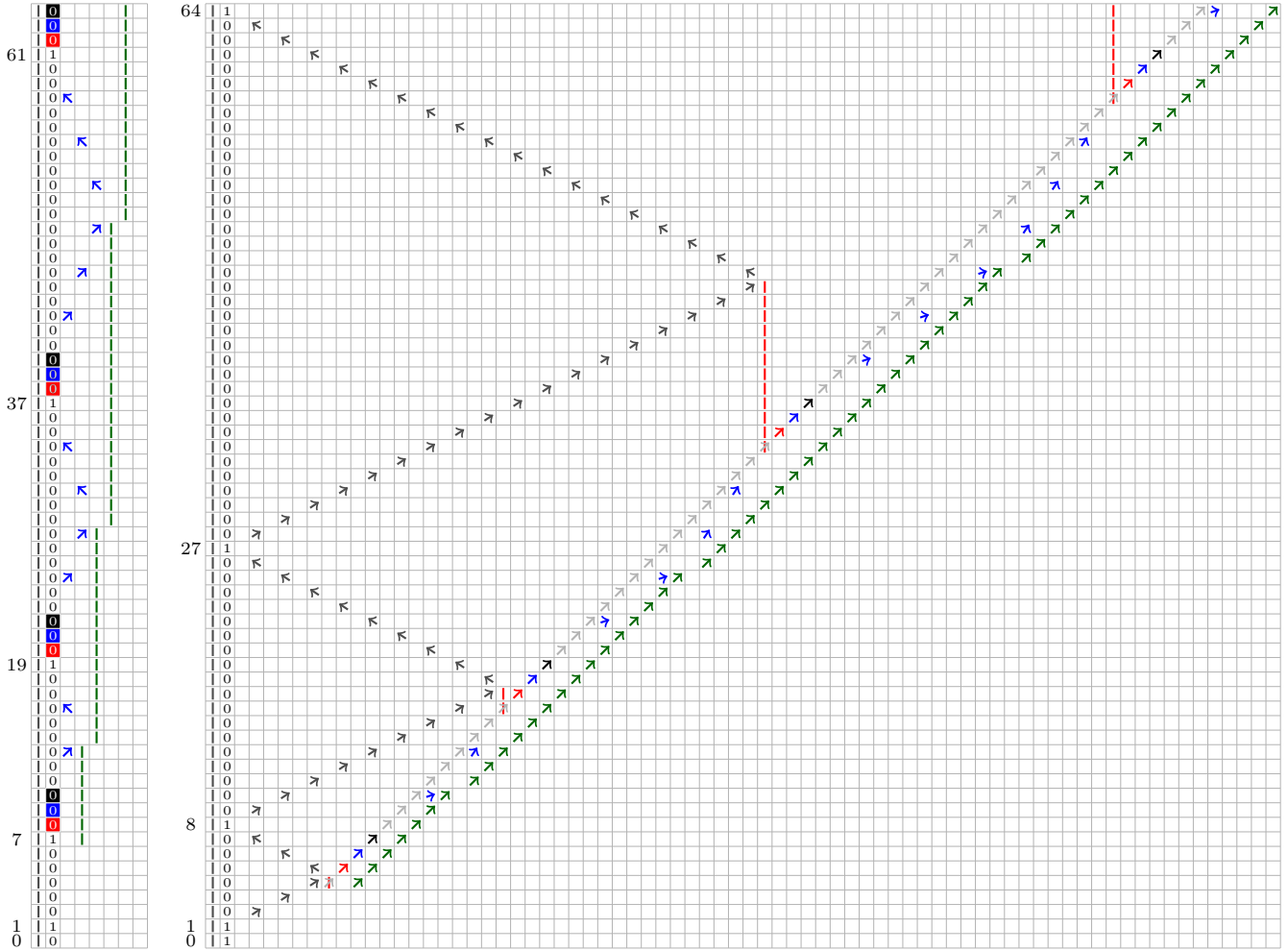


Fig. 4: CA for the difference of cubes (left) and for the cubes (right).

We finally arrive at our main result.

**Theorem 1.** *Let  $P(X) \in \mathbb{Q}[X]$  denote a polynomial of degree  $d \geq 1$  with  $P(\mathbb{N}) \subset \mathbb{N}$ . Then the sequence  $u = \mathbf{1}_{P(\mathbb{N})}$  belongs to  $\mathcal{S}$  and can be realized by the hitting of one single signal in column 0.*

*Proof.* The result of Theorem 1 is clear for all polynomials of degree 1. Indeed, if  $P$  has the form  $P(X) = aX + b$  with  $a \in \mathbb{N}, b \in \mathbb{Z}$ , one can construct the sequence by the trajectory of a signal that hits a wall in column  $\lfloor a/2 \rfloor$  (we fix the initial conditions in order to take into account the translation of  $b$ ). Note that we can deal with rational coefficients by multiplying and dividing accordingly the speed of the symbols. Suppose now that the statement is true for all polynomials of degree  $k < d$ , for some  $d \geq 2$ . We show that we can realize any polynomial  $P$  of degree  $d$  with only one hitting in column 0. For this purpose, we observe that  $Q(X) = P(X + 1) - P(X)$  is a polynomial of degree at most  $d - 1$ . Furthermore, since  $P(\mathbb{N}) \subset \mathbb{N}$ , the leading coefficient of  $P$  is positive, so that there exists  $c \in \mathbb{N}$ , depending only on  $P$ , such that the polynomial  $\tilde{Q}(X) = Q(X + c)$  satisfies  $\tilde{Q}(\mathbb{N}) \subset \mathbb{N}$ .

$$P(n) - P(0) = \sum_{k=0}^{n-1} (P(k+1) - P(k)) = \sum_{k=0}^{n-1} Q(k)$$

By induction, we know that  $\mathbf{1}_{\tilde{Q}(\mathbb{N})}$  is constructible. Taking care of the finite initial conditions due to the introduction of the integer  $c$  and to the occurrence of  $P(0)$ , Prop. 2 implies that  $\mathbf{1}_{P(\mathbb{N})}$  belongs to  $\mathcal{S}$ , and is realized by the hitting of just one signal.  $\square$

## 4 Fibonacci word

The Fibonacci sequence is the recursive sequence defined by  $F_0 = 0, F_1 = 1$ , and  $\forall n \geq 0, F_{n+2} = F_{n+1} + F_n$ . We denote by  $\mathbf{F}$  the set of numbers that appear in the Fibonacci sequence, that is,  $\mathbf{F} = \{F_n : n \in \mathbb{N}\}$ . We propose a construction of the sequence  $u = \mathbf{1}_{\mathbf{F}}$ , which by definition, satisfies:  $u_n = 1 \iff \exists k \in \mathbb{N}, n = F_k$ . Our construction presents some similarities with the one proposed by Mazoyer and Terrier for general linear recursions [11, Prop. 4].

**Proposition 3.** *The indicator sequence  $\mathbf{1}_{\mathbf{F}}$  of Fibonacci numbers belongs to  $\mathcal{S}$ .*

*Proof.* We propose an explicit construction that uses signals of speed 0 (walls),  $\pm 1$ , and 2. Let us assume that on line  $F_n$ , we have marked the positions 0,  $F_{n-2}, F_{n-1}$  and  $F_n$ . We show how to produce the same marking at time  $F_{n+1}$ . In particular, the procedure will enable us to spot the points of ordinate  $F_n$  in the column 0 of the space-time diagram. The dynamics is as follows.

- From point  $(0, F_n)$ , we send a signal of speed 1.
- From point  $(F_{n-2}, F_n)$  we send a signal of speed 1 and one of speed 2.
- From point  $(F_{n-1}, F_n)$  we send a signal of speed 0 (wall) and one of speed  $-1$ .
- From point  $(F_n, F_n)$  we send a signal of speed 0 (wall) and one of speed 1.

Using elementary arguments based only on the definition of the Fibonacci sequence, one can show that the intersections of these different signals on line  $F_{n+1}$  give points at positions 0,  $F_{n-1}, F_n$  and  $F_{n+1}$ , as illustrated on Fig. 5. In order to complete the construction, one only has to choose appropriate initialisation symbols.  $\square$

The Fibonacci word is the morphic sequence  $v \in \mathcal{A}^{\mathbb{N}}$  which is the unique fixed point of the substitution  $\sigma$  defined by  $0 \mapsto 01$  and  $1 \mapsto 0$ .



Let us define recursively a collection of finite words by  $w_1 = 0$  and for any  $n \geq 1, w_{n+1} = \sigma(w_n)$ . The first words obtained are thus:

$$\begin{aligned} w_1 &= 0 \\ w_2 &= 01 \\ w_3 &= 010 \\ w_4 &= 01001 \\ w_5 &= 01001010 \\ w_6 &= 0100101001001 \\ w_7 &= 010010100100101001010 \\ w_8 &= 0100101001001010010100100101001001 \end{aligned}$$

and the Fibonacci word  $v \in \mathcal{A}^{\mathbb{N}}$  is the limit of these words.

**Proposition 4.** *The Fibonacci word belongs to  $\mathcal{S}$ .*

*Proof (sketch).* It follows from the definition that for any  $n \geq 1$ , the length of the word  $w_n$  is equal to  $F_n$ . Furthermore, one can check that for any  $n \geq 1$ , we have:

$$w_{n+1} = w_n \cdot w_{n-1} = w_{n-1} \cdot w_{n-2} \cdot w_{n-2} \cdot w_{n-3} = \underbrace{w_{n-2} \cdot w_{n-3} \cdot w_{n-2}}_{w_n} \cdot w_{n-2} \cdot w_{n-3},$$

where the symbol  $\cdot$  represents the concatenation of words. It follows that if the word  $w_n$  has already been computed, in order to obtain the word  $w_{n+1}$ , one can first concatenate after  $w_n$  the sequence  $w_{n-2}$ , which is the one that appears at the last  $F_{n-2}$  positions of  $w_n$ , and then the sequence  $w_{n-3}$ , which appears in  $w_n$  between positions  $F_{n-2}$  and  $F_{n-2} + F_{n-3} = F_{n-1}$ . This can be done with a finite state cellular automaton: the block  $w_{n-2}$  appearing at the end of  $w_n$  is translated to the right until reaching a wall placed at position  $F_{n-2}/2$  and then translated to the left. And in the same way, the block  $w_{n-3}$  appearing between positions  $F_{n-2}$  and  $F_{n-1}$  is translated to the right until reaching a wall placed at position  $(F_{n-2} + F_{n-1})/2 = F_n/2$  and then translated to the left. Note that the walls at positions  $F_{n-2}/2$  and  $(F_{n-2} + F_{n-1})/2$  can be computed without much difficulty. The construction of the wall at position  $F_{n-2}/2$  appears on Fig. 5. For the construction of the wall at position  $(F_{n-2} + F_{n-1})/2$ , one can send signals of respective speeds 1 and  $-1$  from positions  $F_{n-2}$  and  $F_{n-1}$  at time  $F_{n-1}$ .  $\square$

## 5 Other computable sequences and open questions

Since  $\mathcal{S}$  is countable, whereas  $\mathcal{A}^{\mathbb{N}}$  is uncountable, there exist sequences of  $\mathcal{A}^{\mathbb{N}}$  that do not belong to  $\mathcal{S}$ . A sequence  $(u_n)_{n \geq 0} \in \mathcal{A}^{\mathbb{N}}$  is computable if there exists an algorithm, or more formally, a *Turing machine*, which on the input  $n$ , calculates  $u_n$ , for any  $n \geq 0$ . In other words, this means that the set  $\{n \in \mathbb{N} : u_n = 1\}$  is accepted by a Turing machine that always halts. It follows easily from the definition that all the sequences of  $\mathcal{S}$  are computable. Next proposition gives an elementary proof of the fact that  $\mathcal{S}$  is strictly contained in the set of computable sequences, using a diagonal argument. We refer to the article of Mazoyer and Terrier [11, Prop. 12 and 13] for further characterizations of  $\mathcal{S}$  in terms of computability.

**Proposition 5.** *There exist computable sequences that do not belong to  $\mathcal{S}$ .*

*Proof.* For an integer  $r \geq 0$ , let us denote by  $\mathcal{F}_r$  the set of CA  $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$  of radius  $r$ . The set  $\mathcal{F}_r$  is a finite set, of cardinality  $2^{2r+1}$ , and it can be enumerated in a canonical way, for example using Wolfram code [15]. So, the set  $\mathcal{F} = \cup_{r \geq 0} \mathcal{F}_r$  of all CA  $F : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$  is countable, and one can define a computable surjection  $\phi : \mathbb{N} \rightarrow \mathcal{F}$ .

The set  $\mathcal{C}_0(\mathcal{A})$  of finite configurations on  $\mathcal{A}^{\mathbb{Z}}$  is also countable, and can be enumerated by a computable surjection  $\psi : \mathbb{N} \rightarrow \mathcal{C}_0(\mathcal{A})$ . Let now  $\tau : \mathbb{N} \rightarrow \mathbb{N}^2$  be a computable surjection, and let us consider the following map:

$$\begin{aligned} \Pi : \mathbb{N} &\rightarrow \mathcal{S} \\ k &\mapsto (F^n(x)_0)_{n \geq 0}, \text{ where } F = \phi(\tau(k)_1) \text{ and } x = \psi(\tau(k)_2). \end{aligned}$$

The map  $\Pi$  above is surjective. Let us consider the sequence  $(u_n)_{n \geq 0}$  defined by  $u_n = 1 - \Pi(n)_n$ , for any  $n \geq 0$ . This sequence is computable, and does not belong to the image of  $\Pi$ , so it does not belong to  $\mathcal{S}$ .  $\square$

The proposition above provides an example of sequence that cannot be obtained by a CA. Nevertheless, this sequence has an abstract form, and it is an open problem to exhibit a more concrete example of (computable) sequence that does not belong to  $\mathcal{S}$ . We conclude with some open questions.

1. We have seen that the Fibonacci word belongs to  $\mathcal{S}$ . More generally, does any morphic sequence belong to  $\mathcal{S}$ ? Is there a canonical way to construct the CA using directly the form of the defining morphism of the sequence, at least for  $k$ -step recurrences over words?
2. A sequence of  $\mathcal{A}^{\mathbb{N}}$  is said to be *normal* if all finite strings of equal length occur with equal asymptotic frequency. None of the sequences that we have evocated here is normal. Is it possible to come up with a CA that gives a normal sequence such as the Champernowne sequence (concatenation of natural numbers in base 10 : 12345678910111213...), or the Copeland–Erdős sequence (concatenation of prime numbers in base 10 : 2357111317192329...)?
3. Do the binary expansions of  $\sqrt{2} = (1.011010100000100\dots)$  or  $\pi = (11.00100100001111\dots)$  belong to  $\mathcal{S}$ ? Note that it is not known whether the binary digits of these two constants give rise to normal sequences.
4. There is a close connection between one-dimensional linear CA and  $p$ -automatic sequences. Do there exist analogous statements for higher dimensional linear CA? Also, is it possible to realize any multivariate polynomial with the help of a suitable higher dimensional (nonlinear) CA?
5. Linear CA provide examples of *permutive* CA. What kind of sequences can be obtained with permutive CA?

**Acknowledgements.** The authors thank N. Fatès and E. Jeandel for fruitful discussions and the referees for valuable comments.

## References

1. Allouche, J.P., Shallit, J.: Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press, Cambridge (2003)
2. Blackburn, S.R.: Non-Overlapping Codes. IEEE Trans. Information Theory 61(9), 4890–4894 (2015)
3. Büchi, J.R.: Weak Second-Order Arithmetic and Finite Automata. In: The Collected Works of J. Richard Büchi, pp. 398–424. Springer, New York, NY (1990)
4. Chee, Y.M., Kiah, H.M., Purkayastha, P., Wang, C.: Cross-Bifix-Free Codes Within a Constant Factor of Optimality. IEEE Trans. Information Theory 59(7), 4668–4674 (2013)

5. Delacourt, M., Poupet, V., Sablik, M., Theyssier, G.: Directional dynamics along arbitrary curves in cellular automata. *Theoretical Computer Science* 412(30), 3800–3821 (Jul 2011)
6. Fischer, P.C.: Generation of Primes by a One-Dimensional Real-Time Iterative Array. *Journal of the ACM* 12(3), 388–394 (Jul 1965)
7. Gilbert, E.N.: Synchronization of binary messages. *IRE Trans. Information Theory* 6(4), 470–477 (1960)
8. Korec, I.: Real-time generation of primes by a one-dimensional cellular automaton with 11 states. In: Goos, G., Hartmanis, J., van Leeuwen, J., Prívvara, I., Ružička, P. (eds.) *Mathematical Foundations of Computer Science 1997*, vol. 1295, pp. 358–367. Springer Berlin Heidelberg, Berlin, Heidelberg (1997)
9. Levenshtein, V.I.: The maximal number of words in codes without overlap. *Problemy Peredachi Informatsii* 6(4), 88–90 (1970)
10. Litow, B., Dumas, P.: Additive cellular automata and algebraic series. *Theoretical Computer Science* 119(2), 345–354 (Oct 1993)
11. Mazoyer, J., Terrier, V.: Signals in one-dimensional cellular automata. *Theoretical Computer Science* 217(1), 53–80 (Mar 1999)
12. Minsky, M., Papert, S.: Unrecognizable Sets of Numbers. *Journal of the ACM* 13(2), 281–286 (Apr 1966)
13. Ritchie, R.W.: Finite Automata and the Set of Squares. *Journal of the ACM* 10(4), 528–531 (Oct 1963)
14. Rowland, E., Yassawi, R.: A characterization of  $p$ -automatic sequences as columns of linear cellular automata. *Advances in Applied Mathematics* 63, 68–89 (Feb 2015)
15. Wolfram, S.: Statistical mechanics of cellular automata. *Reviews of Modern Physics* 55(3), 601–644 (Jul 1983)