Column representation of Sturmian words in cellular automata*

Francesco Dolce[1] and Pierre-Adrien Tahay[2]

[1] FIT, Czech Technical University in Prague, Czech Republic
[2] FNSPE, Czech Technical University in Prague, Czech Republic
dolcefra@fit.cvut.cz, pierre.adrien.tahay@cvut.cz

Abstract. We prove that, given a Sturmian word $\mathbf{w}$ with quadratic slope, it is possible to construct a one-dimensional cellular automaton such that $\mathbf{w}$ is represented in a chosen column in its space-time diagram. Our proof is constructive and use the continued fraction expansion of the slope of the word.

Keywords: Sturmian words · cellular automata · quadratic numbers · continued fraction expansion

1 Introduction

Sturmian words are infinite words over a binary alphabet that have exactly $n+1$ factors of length n for each non-negative n. Their origin can be traced back to the astronomer J. Bernoulli III. Their first in-depth study is by Morse and Hedlund [17]. Many combinatorial properties were described in the paper by Coven and Hedlund [5]. Sturmian words are one of the most studied topics in combinatorics on words. They can be defined in different ways and have various interpretations in several domains, including combinatorics and discrete geometry. A possible way to describe them is by using mechanical words.

In this paper we consider characteristic Sturmian words with quadratic slope (see Section 2 for precise definitions). In particular, given such a Sturmian word $\mathbf{w}$ with continued fraction of its slope $\alpha = [0, 1+b_1, b_2, \ldots, b_m, \overline{a_1, a_2, \ldots, a_k}]$ and corresponding directive sequence $\Delta = (b_1, b_2, \cdots, b_m, (a_1, a_2, \cdots, a_k)^\omega)$, we have $\mathbf{w} = \lim_{n\to\infty} w_n$, where $w_0 = \mathtt{a}, w_1 = \mathtt{b}, w_n = w_{n-1}^{b_n} w_{n-2}$ for $1 \leq n \leq m$ and $w_{m+n} = w_{m+n-1}^{a_{(n \bmod k)}} w_{m+n-2}$ for $n > 0$ (where $a_0 := a_k$). We use such a characterisation to construct a machine, called cellular automaton, that will "print" us the infinite word $\mathbf{w}$. A cellular automaton is a dynamical system defined by an infinite string of symbols over an alphabet and a map, called local rule, that transforms every symbol of the string according to its neighbourhood (see Section 3 for a more precise definition). A classical example of a 2-dimensional

* The research received funding from the Ministry of Education, Youth and Sports of the Czech Republic through the projects CZ.02.1.01/0.0/0.0/16_019/0000765 and CZ.02.1.01/0.0/0.0/16_019/0000778.

cellular automaton is given by 1970 Conway's Game of Life. In spite of its very simple definition, the Game of Life has some quite remarkable properties. Indeed, Rendell proved that starting from it, it is possible to simulate any Turing machine [18]. In this paper we consider one-dimensional cellular automata. The initial configuration will be given by an infinite string over a (finite) alphabet.

Our main result is the following one.

Theorem 1. *A Sturmian word with quadratic slope can be represented as a column in the space-time diagram of a one-dimensional cellular automaton.*

The task of representing a sequence over a finite alphabet in the space-time diagram of a cellular automaton is a non-trivial and still not entirely explored topic. One of the first results on the subject is the construction of the characteristic sequence of primes numbers, done by Fischer in 1965, using a cellular automaton with more than 30,000 states [8]. In 1997, Korec gives another construction with only 11 states [13]. In 1999, Mazoyer and Terrier establish several geometric constructions of increasing functions, which they call Fischer constructible [16]. A very interesting result is given by Rowland and Yassawi in 2015 [19]: they give a complete characterisation of the construction of q-automatic sequence, with q a power of a prime number, in the columns of the space-time diagram of linear cellular automata. Finally, in 2018, Marcovici, Stoll and Tahay construct different non-automatic sequences, such as the characteristic sequences of polynomials (squares, cubes, etc.) and the Fibonacci word [15]. Our main result in this paper can be seen as an extension of the construction obtained by Marcovici, Stoll and Tahay for this last infinite word. While in [15] the authors use *ad hoc* properties of Fibonacci numbers and Fibonacci finite words, in this article we consider the development of the continued fraction associated a Sturmian word with quadratic slope to define a new algorithm. Such an algorithm could even be generalised to larger families of infinite words (see Section 6).

2 Preliminaries

In this section we recall some basic definitions on finite and infinite words. For all undefined terms we refer to [14]. We denote the set of integers, of positive integers and of non-negative integers respectively by $\mathbb{Z}$, $\mathbb{Z}^+$ and $\mathbb{N}$, while $\mathbb{Q}$ and $\mathbb{R}$ denote the set of rational and of real numbers.

2.1 Words

An *alphabet* Σ is a (finite) set of symbols called *letters*. The set of *finite words* Σ^* over Σ is the free monoid having neutral element the *empty word* ε. We also denote by Σ^+ the free semigroup over Σ, e.g., $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$. The *length* of a word $w = a_0 a_1 \cdots a_{n-1}$, with $a_i \in \Sigma$, is the non-negative integer $|w| = n$. The length of ε is considered to be 0. When it is possible to write $w = pus$, with $p, u, s \in \Sigma^*$, we call p (resp., s, u) a *prefix* (resp., *suffix*, *factor*) of w.

An *infinite word* over Σ is a sequence $\mathbf{w} = a_0a_1a_2\cdots$, with $a_i \in \Sigma$ for all i. Similarly to finite words we can define the set of infinite words $\Sigma^{\mathbb{N}}$ and extend in a natural way to $\Sigma^* \cup \Sigma^{\mathbb{N}}$ the notions of prefix, suffix and factor. An infinite word $\mathbf{w}$ is *eventually periodic* if $\mathbf{w} = uv^\omega = uvvv\cdots$. When $u = \varepsilon$ we say that $\mathbf{w}$ is *purely periodic*. An infinite word that is not eventually periodic is called *aperiodic*. The *factor complexity* of an infinite word $\mathbf{w}$ is the mapping $\mathcal{C}_\mathbf{w} : \mathbb{N} \to \mathbb{N}$ defined by $\mathcal{C}_\mathbf{w}(n) = \#\{u \mid u \text{ is a factor of } \mathbf{w} \text{ and } |u| = n\}$.

Aperiodic infinite words with the lowest possible factor complexity, i.e., such that $\mathcal{C}_\mathbf{w}(n) = n+1$ for all $n \in \mathbb{N}$, are called *Sturmian words* (for other equivalent definitions see [1]). It follows from the definition that all Sturmian words are defined over a binary alphabet, e.g., $\{\mathtt{a}, \mathtt{b}\}$. If both sequences $\mathtt{a}\mathbf{w}$ and $\mathtt{b}\mathbf{w}$ are Sturmian, we call $\mathbf{w}$ a *characteristic Sturmian word*. The family of Sturmian words coincide with the family of irrational mechanical words, as well as the family of binary balanced aperiodic words (for more on balanced words see, for instance, [1,6,10]). In particular, characteristic Sturmian words correspond to balanced irrational mechanical words with intercept equal to the slope[14]. We denote by $\mathbf{c}_\alpha$ the unique characteristic Sturmian word with slope (and intercept) α. Since two Sturmian words with the same slope (but different intercept) share the same set of factors, we will focus only on the study of characteristic Sturmian words. Note that every mechanical word with rational slope is a purely periodic word.

Example 1. Let us consider the well-known *Fibonacci word* $\mathbf{f} = \mathtt{abaababaab}\cdots$. It is defined as the fixed point of the morphism sending $\mathtt{a} \mapsto \mathtt{ab}$ and $\mathtt{b} \mapsto \mathtt{a}$. The word $\mathbf{f}$ is the characteristic Sturmian word $\mathbf{c}_{1/\varphi^2}$, where $\varphi = \frac{1+\sqrt{5}}{2}$.

2.2 Continued fraction expansion

Let θ be a real number. A continued fraction expansion of θ is defined as $[c_0, c_1, c_2, \ldots]$ whenever

$$\theta = c_0 + \cfrac{1}{c_1 + \cfrac{1}{c_2 + \ddots}}$$

with $c_0 \in \mathbb{Z}$ and $c_i \in \mathbb{Z}^+$ for every positive i. It is known that if $\theta \in \mathbb{Q}$ then there exist exactly two continued fraction expansion of θ and they are both finite. On the other hand every positive irrational number corresponds to a unique infinite continued fraction expansion with $c_0 \in \mathbb{N}$ and $c_i \in \mathbb{Z}^+$ for all $i \geq 1$. Note that, if $c_0 = 0$ then $0 \leq \theta \leq 1$. If θ is a quadratic irrational, then its continued fraction expansion is eventually periodic, that is it will be of the form

$$\theta = [b_0, \ldots, b_m, \overline{a_1, \ldots, a_k}] = [b_0, \ldots, b_m, a_1, \ldots, a_k, a_1, \ldots, a_k, \ldots]\,.$$

Example 2. The golden ratio $\varphi = \frac{1+\sqrt{5}}{2}$ is a quadratic irrational number. Its continued fraction expansion is $\varphi = [\overline{1}] = [1, 1, 1, \ldots]$ The continued fraction expansions of e and π are $[2, 1, 2, 1, 1, 4, 1, 1, 6, \ldots]$ and $[3, 7, 15, 1, 292, 1, 1, 1, 2, \ldots]$ respectively (sequences A003417 and A001203 in the OEIS [20]).

2.3 Standard sequences and directed sequences

Let $\Delta = (d_n)_{n\geq 1}$ be an integer sequence with $d_1 \in \mathbb{N}$ and $d_n \in \mathbb{Z}^+$ for every positive integer n. The *standard sequence* associated to Δ is the sequence of finite words $(w_n)_{n\geq -1}$ defined by

$$w_{-1} = \mathtt{b}, \quad w_0 = \mathtt{a}, \quad w_n = w_{n-1}^{d_n} w_{n-2} \quad \text{for every } n \geq 1.$$

The sequence Δ is also called the *directive sequence* of $(w_n)_{n\geq -1}$. Note that if $d_1 > 0$, every w_n starts with $\mathtt{a}$. Otherwise, w_n starts with $\mathtt{b}$ for every $n \neq 0$. Let us consider the infinite word $\mathbf{w} = \lim_{n\to\infty} w_n$. Such an infinite word is well defined and w_n is a prefix of $\mathbf{w}$ for every positive n. We say that Δ (resp. $(w_n)_{n\geq -1}$) is the *directive sequence* (resp. the *standard sequence*) of $\mathbf{w}$.

Example 3. Let us consider the directive sequence $\Delta = (1^\omega) = (1, 1, 1, \ldots)$. The associated standard sequence is $f_{-1} = \mathtt{b}, f_0 = \mathtt{a}$, and $f_n = f_{n-1} f_{n-2}$ for every $n \geq 1$. It is known that the Fibonacci word $\mathbf{f}$ defined in Example 1 can be obtained as the limit $\mathbf{f} = \lim_{n\to\infty} f_n$.

Proposition 1 ([14]). *Let α be an irrational number with $0 < \alpha < 1$ having continued fraction expansion $\alpha = [0, d_1 + 1, d_2, d_3, \ldots]$, and let $(w_n)_{n\geq -1}$ be the standard sequence associated to $(d_1, d_2, d_3, \ldots)$. Then $\mathbf{c}_\alpha = \lim_{n\to\infty} w_n$.*

Note that mechanical words are defined for $0 \leq \alpha \leq 1$. It is possible to generalise such definition to every $\alpha \in \mathbb{R}$ [14, Remark 2.1.12]: the fraction expansion in the statement of Proposition 1 would be $\alpha = [c_0, d_1 + 1, d_2, d_3, \ldots]$ with $c_0 \in \mathbb{Z}$, but the standard associated sequence would not change.

Example 4. As seen in Example 3, the directive sequence of the Fibonacci word $\mathbf{f}$ is (1^ω). The continued fraction expansion of the corresponding irrational slope (see also Example 1) is $\frac{1}{\varphi^2} = \left[0, 2, \overline{1}\right]$.

Example 5. Let us consider the characteristic Sturmian word $\mathbf{v}$ having associated directive sequence $\Delta = ((1, 2)^\omega)$. We have $\mathbf{v} = \lim_{n\to\infty} v_n$, where the standard sequence $(v_n)_{n\geq -1}$ is defined by $v_{-1} = \mathtt{b}, v_0 = \mathtt{a}$, $v_{2n+1} = v_{2n} v_{2n-1}$ and $v_{2n} = v_{2n-1}^2 v_{2n-2}$ for every $n \in \mathbb{N}$. Since $\frac{3-\sqrt{3}}{3} = \left[0, 2, \overline{2, 1}\right]$, we have, according to Proposition 1, $\mathbf{v} = \mathbf{c}_{(3-\sqrt{3})/3}$.

In the next sections we will also need the lengths of the prefixes of a characteristic Sturmian sequence. Given a standard sequence $(w_n)_{n\geq -1}$ we define for every integer $n \geq -1$ the number $W_n = |w_n|$.

Example 6. Let $(f_n)_{n\geq -1}$, and $(v_n)_{n\geq -1}$ be the standard sequences defined in Examples 3, and 5. Let $F_n = |f_n|$ and $V_n = |v_n|$. For every $n \in \mathbb{N}$ we have the relations $F_n = F_{n-1} + F_{n-2}$, $V_{2n+1} = V_{2n} + V_{2n-1}$ and $V_{2n} = 2V_{2n-1} + V_{2n-2}$.

3 Cellular automata

In the following we use the terminology developed by Mazoyer and Terrier in [16] and Marcovici, Stoll and Tahay in [15].

Definition 1. *A one-dimensional* cellular automaton *(CA) is a dynamical system $(\mathcal{A}^{\mathbb{Z}}, T)$, where $\mathcal{A}$ is a finite set, and where the map $T : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ is defined by a local rule which acts uniformly and synchronously on the configuration space. More precisely, there exists an integer $r \in N$ called the* radius *of the CA, and a* local rule *$\tau : \mathcal{A}^{2r+1} \to \mathcal{A}$ such that for every $\mathbf{x} = (x_k)_{k\in\mathbb{Z}}$ and for every $k \in \mathbb{Z}$, we have $T(\mathbf{x})_k = \tau((x_{k+i})_{-r\le i\le r})$.*

When the set $\mathcal{A}$ is understood, we will call *cellular automaton* just the map T. The elements of $\mathcal{A}^{\mathbb{Z}}$ are called *configurations*. By the Curtis-Hedlund-Lyndon Theorem [9], a map $T : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ is a CA if and only if it is continuous with respect to the product topology and it commutes with the shift map σ defined by $\sigma(\mathbf{x})_k = x_{k-1}$, for every configuration $\mathbf{x} = (x_k)_{k\in\mathbb{Z}}$ and every $k \in \mathbb{Z}$. Let $\mathtt{0} \in \mathcal{A}$ and $T : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ a cellular automaton. We say that T is $\mathtt{0}$-*quiescent* if $T(\mathtt{0}^{\mathbb{Z}}) = \mathtt{0}^{\mathbb{Z}} = \cdots\mathtt{000}\cdots$. A configuration $\mathbf{x} = (x_k)_{k\in\mathbb{Z}}$ is called *finite* if the set $\{k \in \mathbb{Z} : x_k \neq \mathtt{0}\}$ is finite. A cellular automaton can be visualized by using a space-time diagram which is a 2-dimensional grid where each cell contains an element of the set $\mathcal{A}$ and is represented by a space coordinate and a time coordinate.

Let us consider the set

$$\mathcal{S} = \Big\{(T^n(\mathbf{x})_0)_{n\ge 0} \in \mathcal{A}^{\mathbb{N}} : T \text{ is a } \mathtt{0}\text{-quiescent CA on } \mathcal{A}^{\mathbb{Z}} \text{ and } \mathbf{x} \text{ is finite}\Big\}.$$

In other words, $\mathcal{S}$ is the set of sequences of $\mathcal{A}^{\mathbb{N}}$ that can occur as the first column (and thus as any column) in the space-time diagram of some one-dimensional $\mathtt{0}$-quiescent CA, starting from a finite initial configuration. This set corresponds to the set of *Fischer's produced* sequences in [16].

In a space-time diagram it is also possible to "transmit information" through *signals*, that is to connect two cells (m, n) and $(m', n+t)$ through a monotonous path; we call *slope* of the signal the number $\frac{t}{m'-m}$ (see [16, Definitions 3 and 4] for a formal definition). When $m = m'$, we call such a signal a *vertical signal* or a signal of *infinite* slope. For the sake of simplicity, we usually represent a signal as a straight line between the cells (m, n) and $(m', n + t)$. Signals are usually "porous", i.e., they do not interact between each other. In some case, however, we also need to consider "concrete" signals. In particular, let us define two distinct kinds of walls. We say that a wall is of type (i) whenever a given signal hitting the wall *bounces* from the cell just above, i.e., when a given signal of slope d arrives in a cell (ℓ, t), then such a signal dies and a new signal of slope $-d$ starts from the cell $(\ell, t + 1)$. We say that a wall is of type (ii) whenever a given signal hitting the wall bounces from the same cell, i.e., when a given signal of slope d arrives in a cell (ℓ, t), then such a signal dies and a new signal of slope $-d$ starts from the same cell (ℓ, t). We usually represent a wall of type (i) as a line inside the column $\ell + 1$ when the signals comes from the left (resp. $\ell - 1$ when the signal comes from the right), and a wall of type (ii) as a rectangle containing the cells in the column ℓ (see Figure 1).

Fig. 1. Walls of type (i) (on the left) and of type (ii) (on the right).

When two signals meet, we can *mark* the cell at the intersection, i.e., assign to it a value from the set $\mathcal{A}$, and define new signals starting from it.

4 Construction of numbers

To prove our main result we proceed in two steps. First let us construct a CA recognising the lengths W_n of the prefixes w_n of our Sturmian word $\mathbf{w}$.

Let $X \subset \mathbb{Z}$. Let us denote by $\mathbb{1}_X$ the *characteristic function* of X, that is the map $\mathbb{1}_X : \mathbb{Z} \to \{0, 1\}$ defined by $\mathbb{1}_X(x) = 1$ iff $x \in X$.

Proposition 2 ([16]). *Let $(S_n)_{n\geq 0}$ be an integer sequence defined by $S_{n+p} = \sum_{i=0}^{p-1} a_i S_{n+i}$, where $p, a_i \in \mathbb{N}$. Then $\mathbb{1}_{\{S_n\}_{n\geq 0}} \in \mathcal{S}$.*

Mazoyer and Terrier give an explicit method to build this sequences in a column of a CA. We propose here a different construction for a particular case that will be necessary for representing a Sturmian word of quadratic slope.

Proposition 3. *Let $(d_n)_{n\geq 1}$ be an eventually periodic integer sequence with $d_1 \in \mathbb{N}$ and $d_i \in \mathbb{Z}^+$ for every $i \geq 2$. Let $(S_n)_{n\geq 0}$ be the integer sequence defined by $S_n = d_n S_{n-1} + S_{n-2}$ for every $n \geq 0$, with $S_{-1}, S_0 \in \mathbb{Z}^+$. Then $\mathbb{1}_{\{S_n\}_{n\geq 0}} \in \mathcal{S}$.*

Proof (Sketch). Since the sequence $(d_n)_{n\geq 1}$ is eventually periodic, then there exist $m \in \mathbb{N}$, $k \in \mathbb{Z}^+$ and $b_1, \ldots, b_m, a_1, \ldots a_k \in \mathbb{N}$ such that $(d_n)_{n\geq 1} = (b_1, \ldots, b_m, (a_1, \ldots, a_k)^\omega)$. Note that we can consider the first S_m rows of the cellular automaton as initial conditions, i.e., we can start the construction from the row of rank S_m. Let $n \geq 1$ be an integer. We are going to consider two distinct cases according to the value of d_{n+1}.

Let us first suppose that $d_{n+1} \neq 1$. Assume that we have already marked the cells $(0, S_n)$, (S_{n-2}, S_n), (S_{n-1}, S_n), $(S_{n-1} + S_{n-2}, S_n)$ and (S_n, S_n). We claim that we can mark the cells $(0, S_{n+1})$, (S_{n-1}, S_{n+1}), (S_n, S_{n+1}), (S_n+S_{n-1}, S_{n+1}) and (S_{n+1}, S_{n+1}) In order to do that, we use the relation

$$S_{n+1} = S_n \ + \ d_n(d_{n+1}-1)S_{n-1} \ + \ (d_{n+1}-1)S_{n-2} \ + \ S_{n-1}.$$

The idea is to consider intermediate rows, such that their distance is given by the addends in the previous sum. When two signals meet, they died and we can use the cell on the intersection to define other signals. The slope of each signal is determined by the ratio between the difference between the time coordinates and the difference between the space coordinates. For example, the slope of the signal between the cell $(0, S_n)$ and the cell $(S_{n-1}, S_n + d_n(d_{n+1}-1)S_{n-1})$ is

$$\frac{(S_n + d_n(d_{n+1}-1)S_{n-1}) \ - \ S_n}{S_{n-1} \ - \ 0} = d_n(d_{n+1}-1)$$

When $d_{n+1} = 1$ the construction is different. In this case the three rows S_n, $S_n+d_n(d_{n+1}-1)S_{n-1}$ and $d_{n+1}S_n$ coincide, as well as the two rows S_n+S_{n-1} and S_{n+1} (resp. the two columns S_n+S_{n-1} and S_{n+1}). We start with the cells $(0, S_n)$, (S_{n-2}, S_n), (S_{n-1}, S_n) and (S_n, S_n) and we claim that we can mark the cells $(0, S_{n+1})$, (S_{n-1}, S_{n+1}), (S_n, S_{n+1}) and (S_{n+1}, S_{n+1}).

The construction of the sequence $(S_n)_{n\geq 0}$ is illustrated in Figures 2 and 3. Using these figures it is not hard to recover exact definitions of the signals. For sake of simplicity we use the same colour when two signals have the same slope and we represent only the cells on the intersections between two signals.

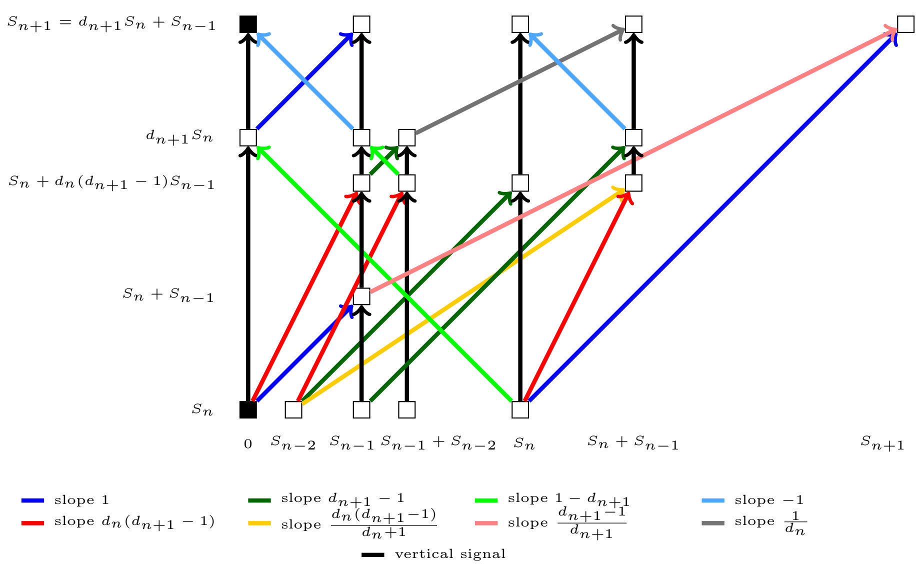

Fig. 2. Construction of the number sequence $(S_n)_{n\geq 0}$ when $d_{n+1} \neq 1$.

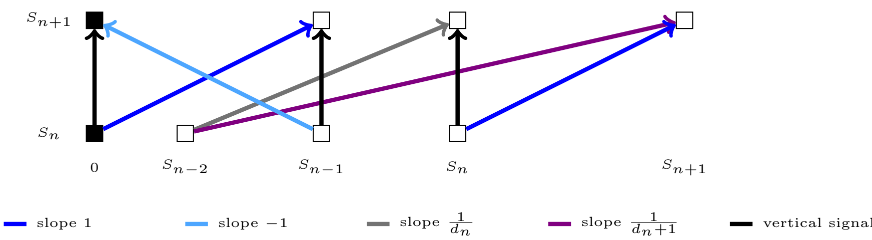

Fig. 3. Construction of the number sequence $(S_n)_{n\geq 0}$ when $d_{n+1} = 1$.

In particular, in both cases we are able to mark the cell $(0, S_n)$. Hence we can mark the sequence $\{S_n\}_{n\geq 0}$ on the column 0. To complete the proof it is enough to put the letter `1` in the cells $(0, S_n)$, for every $n \geq 0$ and the letter `0` in all other cells in the column 0.

Note that the hypothesis of eventual periodicity of the sequence $(d_n)_{n\geq 1}$ in the previous proof is essential to guarantee that the cellular automaton is defined over a finite set $\mathcal{A}$. Indeed, since the signals (and their slope) are periodically repeated, we have $\mathrm{Card}(\mathcal{A}) = \mathcal{O}(k)$, where k is the length of the maximum between the pre-period and the eventual period of $(d_n)_{n\geq 1}$.

Example 7. Let us consider the word $\mathbf{f}$ defined in Example 1 and the associated numerical sequence $(F_n)_{n\geq 0}$. According to Proposition 3 we have $\mathbb{1}_{\{F_n\}_{n\geq 0}} \in \mathcal{S}$.

5 Construction of prefixes

In this section we prove our main theorem. In order to do that, we need some preliminary results.

Proposition 4. *Let $\Delta = (d_n)_{n\geq 1}$ be an eventually periodic integer sequence with $d_1 \geq 0$ and $d_n \geq 1$ for every $n > 1$; let $(w_n)_{n\geq -1}$ be the standard sequence associated to Δ defined by $w_{-1} = \mathtt{b}, w_0 = \mathtt{a}$, and $w_n = w_{n-1}^{d_n} w_{n-2}$ for every $n \geq 1$. Then $\mathbf{w} = \lim_{n\to\infty} w_n \in \mathcal{S}$*

In order to prove Proposition 4, we need the following result stating that a letter in a cell of a CA can be recopied in the same column and in any row above. Moreover, we can do it by using only walls and signals of slope 1 and -1.

Lemma 1. *Let $T : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ be a CA, $a \in \mathcal{A}$, and $n, m, t, t' \in \mathbb{Z}$ with $m > 0$, and $t > t' \geq 0$. Suppose that the cells (n, t'), $(n + m, t')$ and (n, t) are marked, the last one with a. Then it is possible to mark the cell $(n, t + m)$ with a.*

Proof. Without loss of generality, suppose that $n = 0$ and $t' = 0$. and so a is in the cell $(0, t)$. Let us prove that we can recopy a into the cell $(0, t + m)$. To mark the wall we are going to use the cell $(m, 0)$. We consider two distinct cases, according to the parity of m.

Let m be odd. We consider a wall of type (i) to the right of the column $\lfloor \frac{m}{2} \rfloor$. Such a wall can be defined, for instance, using two signals of slope 1 and -1 starting respectively at $(0,0)$ and at $(m,0)$ (see left of Figure 4).

We send a signal of slope 1 from the cell $(0,t)$; when this signal touches the wall we define a new signal of slope -1 (starting from the cell above); when this new signal meets the column 0, we write the letter a. Since $\lfloor \frac{m}{2} \rfloor = \frac{m-1}{2}$, we have that the new a is exactly on the row $t + \frac{m-1}{2} + \frac{m-1}{2} + 1 = t + m$. as wanted.

Suppose now that m is even. We consider this time a wall of type (ii) in the column $\frac{m}{2}$. Such a wall can be defined, for instance, using two signals of slope 1 and -1 starting respectively at $(0,0)$ and at $(m,0)$ (see right of Figure 4). We send a signal of slope 1 from the cell $(0,t)$; when this signal hit the wall, we define a new signal of slope -1 (starting from the same cell); when this new signal meets the column 0, we write the letter a. The new a is exactly on the row $t + \frac{m}{2} + \frac{m}{2} - 1 + 1 = t + m$, as wanted.

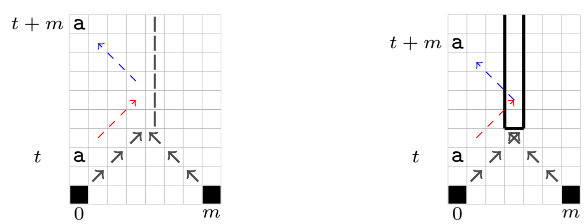

Fig. 4. Recopying of letters.

We can now prove Proposition 4.

Proof (of Proposition 4). Let us denote $S_n = |w_n|$ for all $n \geq -1$. For every $n \geq 1$ we have $S_n = d_n S_{n-1} + S_{n-2}$. Since, for all $n \geq 2$, w_{n-3} is a suffix of w_{n-1}, the word $w_{n-3}w_{n-2}$ is a suffix of w_n. Suppose the word w_n constructed for a given $n \geq 2$ and suppose that the last letter of w_n is in the cell $(0, S_n)$.

Let us first suppose that $d_n \neq 1$. We will show that it is possible to construct the word w_{n+1} with its last letter in the cell $(0, S_{n+1})$. In order to do that, we will use the relation $w_{n+1} = w_n((w_{n-2}^{d_{n-1}} w_{n-3})^{d_n} w_{n-2})^{d_{n+1}-1} w_{n-2}^{d_{n-1}} w_{n-3}$. Let us take up the construction of Proposition 3 until the number S_n. Moreover, in the column 0 we mark the cell $(0, S_{n-1} + d_{n-1}(d_n - 1)S_{n-2})$. For this, we define a signal of slope $-(d_{n-1}(d_n - 1))$ from the cell (S_{n-2}, S_{n-1}). This signal meets the vertical signal defined from $(0, S_{n-1})$ in the required cell $(0, S_{n-1} + d_{n-1}(d_n - 1)S_{n-2})$. From this cell we define a signal $\mathbf{P_1^{(n-1)}}$ of slope 1 and from the cells $(S_{n-2}, S_{n-1} + d_{n-1}(d_n - 1)S_{n-2})$, $(S_{n-1}, S_{n-1} + d_{n-1}(d_n - 1)S_{n-2})$ and $(S_{n-1} + S_{n-2}, S_{n-1} + d_{n-1}(d_n - 1)S_{n-2})$ we define three signals $\mathbf{N_1^{(n-1)}}$, $\mathbf{N_2^{(n-1)}}$ and $\mathbf{N_3^{(n-1)}}$ of slope -1. At the intersection of these with the signal $\mathbf{P_1^{(n-1)}}$ we define three walls $\mathbf{M_1^{(n-1)}}$, $\mathbf{M_2^{(n-1)}}$ and $\mathbf{M_3^{(n-1)}}$ in the columns S_{n-2}, S_{n-1} and

$S_{n-1}+S_{n-2}$ respectively. These walls will be of type (i) or of type (ii) according to the parity of the columns where $\mathbf{N_i^{(n-1)}}$ originate. Suppose we have already recopied the words w_{n-3} and w_{n-2} in the suffix of w_n, i.e., we have found the two corresponding heights in the column 0. Now, from these two words, we are going to define signals of slope 1 that will stop against one of the three walls previously defined and send back signals of slope -1 that will recopy the same word in the column 0. Using Lemma 1 we can recopy one by one the letters of w_{n-2} and w_{n-3}. First, we send a signal of slope 1 from each letter of w_{n-2}. When this signal meets the wall $\mathbf{M_1^{(n-1)}}$ we send a signal of slope -1 until the column 0. Since $\mathbf{N_1^{(n-1)}}$ is generated from the column S_{n-2} the letters of w_{n-2} are recopied in the same column but above at distance S_{n-2}. This means that we have recopied a second word w_{n-2} above the first. We repeat this procedure d_{n-1} times and we get the word $w_{n-2}^{d_{n-1}}$ above the w_n already constructed. Next, we send a signal of slope 1 from each letter of w_{n-3} to the wall $\mathbf{M_3^{(n-1)}}$ and, in a similar way, from there we send a signal of slope -1 to the column 0. The letters of w_{n-3} are recopied in the same column but above at distance $S_{n-1}+S_{n-2}=d_{n-1}S_{n-2}+S_{n-3}+S_{n-2}$. Thus, we have recopied a word w_{n-3} above the word $w_{n-3}w_{n-2}w_{n-2}^{d_{n-1}}$, where the subword $w_{n-3}w_{n-2}$ corresponds to a suffix of w_n. So far we have constructed the word $w_n(w_{n-2}^{d_{n-1}}w_{n-3})$ on the column 0. For the next step we need to use also the word w_{n-2} in the suffix of w_n. This time, we send signals of slope 1 from cells in column 0 to the wall $\mathbf{M_3^{(n-1)}}$ and from there, signals of slope -1 to the column 0. Therefore, the letters of w_{n-2} are recopied in the same column but above at distance $S_{n-1}+S_{n-2}=(d_{n-1}+1)S_{n-2}+S_{n-3}$. Hence we obtain the word $w_n(w_{n-2}^{d_{n-1}}w_{n-3})w_{n-2}$ in the column 0. By using the wall $\mathbf{M_1^{(n-1)}}$ the word $w_n((w_{n-2}^{d_{n-1}}w_{n-3})w_{n-2}^{d_{n-1}}$ can be obtained. From every letter of the word w_{n-3} we send signals of slope 1 to the wall $\mathbf{M_2^{(n-1)}}$ and, from there, signals of slope -1 back to column 0. The letters of w_{n-3} are recopied in the same column but above at distance $S_{n-1}=d_{n-1}S_{n-2}+S_{n-3}$. Hence, we obtain the word $w_n(w_{n-2}^{d_{n-1}}w_{n-3})w_{n-2}^{d_{n-1}}w_{n-3}$. Similarly, it is easy to obtain the word $w_n(w_{n-2}^{d_{n-1}}w_{n-3})^{d_n}=w_n(w_{n-2}^{d_{n-1}}w_{n-3})^{d_n-1}(w_{n-2}w_{n-2}^{d_{n-1}-1}w_{n-3})$. Following the same idea, we use the wall $\mathbf{M_2^{(n-1)}}$ to recopy w_{n-2} in the column 0 but $S_{n-1}=(d_{n-1}-1)S_{n-2}+S_{n-3}+S_{n-2}$ cells above and so, we obtain the word $w_n(w_{n-2}^{d_{n-1}}w_{n-3})^{d_n}w_{n-2}$. Since the suffix of this word is $w_{n-3}w_{n-2}$ the previous steps can be applied again to obtain the word $w_n((w_{n-2}^{d_{n-1}}w_{n-3})^{d_n}w_{n-2})^{d_{n+1}-1}$ which also has the suffix $w_{n-3}w_{n-2}$. Thus we easily obtain the word
$w_n((w_{n-2}^{d_{n-1}}w_{n-3})^{d_n}w_{n-2})^{d_{n+1}-1}w_{n-2}^{d_{n-1}}w_{n-3}=w_{n+1}$.

Let us now consider the case $d_n=1$. Here we have $S_n=S_{n-1}+S_{n-2}$ and the cell $(S_{n-1}+S_{n-2},S_{n-1}+d_{n-1}(d_n-1)S_{n-2})=(S_n,S_{n-1})$ is not marked. Therefore the wall $\mathbf{M_3^{(n-1)}}$ can no longer be defined as before. This time, we define a wall $\mathbf{M_4^{(n)}}$ at the intersection of a signal $\mathbf{P_2^{(n)}}$ a slope 1 starting from $(0,S_n)$ and a signal $\mathbf{N_4^{(n)}}$ of slope -1 starting from (S_n,S_n). The suffix $w_{n-2}w_{n-3}$ of w_n is below the signals used to define the wall $\mathbf{M_4^{(n)}}$, therefore, we have to construct the word $w_{n-3}w_{n-2}$ in $w_n(w_{n-2}^{d_{n-1}}w_{n-3}w_{n-2})^{d_{n+1}-1}w_{n-2}^{d_{n-1}}w_{n-3} = w_{n+1}$ in another way. We have to copy the letters of the word $w_{n-3}w_{n-2}$ in the column 0 above at distance $S_{n-3}+S_{n-2}+d_{n-1}S_{n-2} = S_{n-1}+S_{n-2} = S_n$. In order to do that, we define a wall $\mathbf{M_5^{(n-1)}}$ of type (i) or (ii) (according to the parity) by intersecting a signal $\mathbf{P_3^{(n-1)}}$ of slope 1 from (S_{n-2}, S_{n-1}) and a signal $\mathbf{N_2^{(n-1)}}$ of slope -1 from (S_{n-1}, S_{n-1}) defined as in the previous case. Such a wall is in the column $S_{n-2}+\left\lfloor \frac{S_{n-1}-S_{n-2}}{2} \right\rfloor = S_n - S_{n-1} + \left\lfloor S_{n-1} - \frac{S_n}{2} \right\rfloor = \left\lfloor \frac{S_n}{2} \right\rfloor$. We send a signal of slope 1 from each letter of $w_{n-3}w_{n-2}$ to the wall $\mathbf{M_5^{(n-1)}}$ and from there we send a signal of slope -1 until the column 0. Therefore, we have constructed the word $w_n w_{n-2}^{d_{n-1}} w_{n-3}w_{n-2}$. The rest of the construction is the same as before, with $\mathbf{M_4^{(n)}}$ playing the role of the wall $\mathbf{M_3^{(n-1)}}$ for the rows above the word $w_{n-3}w_{n-2}$.

Note that the signals $\mathbf{P_i^{(n)}}$ can also be used to destroy the walls $\mathbf{M_i^{(n-1)}}$ previously constructed. Formally, when a signal $\mathbf{P_i^{(n)}}$ meets a wall $\mathbf{M_i^{(n-1)}}$, the last one is destroyed and the signal $\mathbf{P_i^{(n)}}$ continues its move.

Example 8. The Fibonacci word $\mathbf{f}$ defined in Example 1 is in $\mathcal{S}$.

We can now prove our main result.

Proof (of Theorem 1). Let $\mathbf{w}$ be a Sturmian word of quadratic slope, $\alpha = [0, 1+b_1, b_2, \ldots, b_m, \overline{a_1, a_2, \ldots, a_k}]$ the continued fraction expansion of its slope, and $(w_n)_{n\geq -1}$ the standard sequence associated to the eventually periodic integer sequence $\Delta = (b_1, \cdots, b_m, (a_1, \cdots, a_k)^\omega)$ so that Δ is an eventually periodic integer sequence. Following Proposition 1, we have $\mathbf{w} = \lim_{n\to\infty} w_n$. Using Propositions 3 and 4, it is clear that $\mathbf{w} \in \mathcal{S}$.

6 Conclusions

To prove our results we used the continued fraction expansion associated with $\mathbf{w}$. However, to use Proposition 4 it is enough to know how to decompose each prefix w_n in compositions of powers of smaller prefixes. A different approach could be to use the morphisms

$$G = \begin{cases} \mathtt{a} \mapsto \mathtt{a} \\ \mathtt{b} \mapsto \mathtt{ab} \end{cases} \qquad \text{and} \qquad D = \begin{cases} \mathtt{a} \mapsto \mathtt{ba} \\ \mathtt{b} \mapsto \mathtt{b} \end{cases}.$$

Indeed, it is known that for every standard Sturmian sequence $\mathbf{w}$ there exist a unique sequence of words $\left(\mathbf{w}^{(i)}\right)_i$ and an infinite sequence $(\psi_i)_i \in \{G, D\}^{\mathbb{N}}$ of morphisms such that $\mathbf{w} = \lim_{n\to\infty} \psi_0\psi_1 \ldots \psi_n(w^{(n)})$ (see, for instance, [12]). Using these notions, and the strictly related notion of $\mathcal{S}$-adicity, we think it is possible to generalise our results to larger classes of words and languages, namely Arnoux-Rauzy words and dendric words (see, e.g., [2,3,7]).

Blanchard and Kůrka [4] consider a larger family of languages that can be recognised by a non-deterministic Turing machine. This family contains the languages corresponding to quadratic numbers but also the ones corresponding to Hurwitz numbers, such as e (a Hurwitz number is an irrational such that its continued fraction expansion is a polynomial mixture [11]). An interesting question is whether it is possible to generalise our construction to such a family as well.

References

1. Ľ. Balková, E. Pelantová, and Š. Starosta. Sturmian jungle (or garden?) on multiliteral alphabets. *RAIRO - Theor. Inf. Appl.*, 44:443–470, 2010.
2. V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176:521–550, 2015.
3. V. Berthé, F. Dolce, F. Durand, J. Leroy, and D. Perrin. Rigidity and substitutive tree words. *Int. J. of Found. Comp. S.*, 29:705–720, 2018.
4. F. Blanchard and P. Kůrka. Language complexity of rotations and Sturmian sequences. *Theoret. Comput. Sci.*, 209:179–193, 1998.
5. E. M. Coven and G. A. Hedlund. Sequences with minimal block growth. *Math. Sys. Theory*, 7:138–153, 1973.
6. F Dolce, Ľ. Dvořáková, and E. Pelantová. On balanced sequences and their asymptotic critical exponent. *Proceedings LATA 2021, Lecture Notes in Comput. Sci.*, 12638:293–304, 2021.
7. F. Dolce and D. Perrin. Eventually dendric shift spaces. *Ergod. Theor. Dyn. Syst.*, 41:2023–2048, 2021.
8. P. C. Fischer. Generation of primes by a one-dimensional real-time iterative array. *J. Assoc. Comput. Mach.*, 12:388–394, 1965.
9. G. A. Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Math. Sys. Theory*, 3:320–375, 1969.
10. P. Hubert. Suites équilibrées. *Theoret. Comput. Sci.*, 242:91–108, 2000.
11. A. Hurwitz. Über die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt. *Math. Ann.*, 46:273–3284, 1895.
12. J. Justin and G. Pirillo. Episturmian words and episturmian morphisms,. *Theoret. Comput. Sci.*, 276:281–313, 2002.
13. I. Korec. Real-time generation of primes by a one-dimensional cellular automaton with 11 states. In *Mathematical Foundations of Computer Science 1997*, volume 1295, pages 358–367. Berlin, Heidelberg, 1997.
14. M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
15. I. Marcovici, T. Stoll, and P.-A. Tahay. Construction of some nonautomatic sequences by cellular automata. *Lecture Notes in Comput. Sci.*, 10875:113–126, 2018.
16. J. Mazoyer and V. Terrier. Signals in one-dimensional cellular automata. *Theoret. Comput. Sci.*, 217:53–80, 1999.
17. M. Morse and G.A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Am. J. Math.*, 62:1–42, 1940.
18. Paul Rendell. *Turing machine universality of the game of life*, volume 18 of *Emergence, Complexity and Computation*. Springer, Cham, 2016.
19. E. Rowland and R. Yassawi. A characterization of p-automatic sequences as columns of linear cellular automata. *Adv. in Appl. Math.*, 63:68–89, 2015.
20. N.J.A. Sloane. On-line Encyclopedia of Integer Sequences. URL: `https://oeis.org/`.