# Column representation of episturmian words in cellular automata

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

[1] FIT, Czech Technical University in Prague, Czech Republic
[2] Loria, Université de Lorraine, France
dolcefra@fit.cvut.cz, pierre-adrien.tahay@univ-lorraine.fr

**Abstract.** We prove that for every strict standard episturmian word that is fixed by a morphism it is possible to construct a one-dimensional cellular automaton such that the word is represented in a chosen column in its space-time diagram.

**Keywords:** episturmian words · Sturmian words · cellular automata

## 1 Introduction

In this contribution we prove the following result.

**Theorem 1.** *A stric epistandard word that is a fixed point of a morphism can be represented as a column in the space-time diagram of a one-dimensional cellular automaton.*

This can be seen as a generalization of a similar result concerning Sturmian words with a quadratic slope [3]. For this purpouse, instead of the notion of continued fraction expansion used in [3] – that is well defined in the binary case, but more complicated to deal with in case of larger alphabets – we use a slightly modified notion of directive word, as well as an infinite family of prefixes associated to it.

## 2 Episturmian words

We refer to [7] for all undefined terms. 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., $\{\mathtt{0}, \mathtt{1}\}$. *Episturmian words* are a generalization of these words to larger alphabets (we refer to [5] for a survery on episturmian words).

If both sequences $\mathtt{0}\mathbf{w}$ and $\mathtt{1}\mathbf{w}$ are Sturmian, we call $\mathbf{w}$ a *standard Sturmian word*. It is known that for every Sturmian word there exists a standard one having the same set of factors. In [4] , Droubay, Justin and Pirillo generalize the work of de Luca [2] and define a *standard episturmian word*, also called an *epistandard word*, as a word obtained by iterative palindromic closure: An infinite word $\mathbf{w}$ is epistandard iff there exists an infinite word $\Delta = a_1 a_2 \cdots$ over $\mathcal{A}$, and an infinite sequence $(u_n)_{n\geq 1}$ in $\mathcal{A}^*$ such that $u_1 = \varepsilon$, and $u_{n+1} = (u_n a_n)^{(+)}$ for every positive $n$, where $u^{(+)}$ is the shortest palindrome having $u$ as a prefix. The word $\Delta$ is called the *directive word* of $\mathbf{w}$. In analogy with the binary case, for every *episturmian word* there exists an epistandard word having the same set of factors (such word is unique, except for the periodic case).

An episturmian (resp., epistandard) word is called *strict* if every letter occurs infinitely often in its directive word. In this case, the factor complexity is $\mathcal{C}_{\mathbf{w}}(n) = (d-1)n + 1$, with $d$ the size of the alphabet.

*Example 1.* The *d-bonacci word* is the strict epistandard word over the alphabet $\{\mathtt{0}, \mathtt{1}, \ldots, (\mathtt{d}-\mathtt{1})\}$, defined as fixed point of the morphism $\mathtt{i} \mapsto \mathtt{0}(\mathtt{i}+\mathtt{1})$ for $0 \leq i < d-1$ and $(\mathtt{d}-\mathtt{1}) \mapsto \mathtt{0}$. Its directive sequence is the periodic word $\Delta = (\mathtt{01}\cdots(\mathtt{d}-\mathtt{1}))^{\omega}$. For $d = 2$ and $d = 3$ we obtain respectively the well-known *Fibonacci word* $\mathbf{f} = \mathtt{0100101001001010010100100}\cdots$ and the *Tribonacci word* $\mathbf{t} = \mathtt{0102010010201010201001020}\cdots$ .

In the following, we consider the directive sequences in their *multiplicative form*, i.e., in the form $\Delta = a_1^{e_1} a_2^{e_2} \cdots$ where $a_i \neq a_{i+1}$ and $e_i > 0$ for every positive integer $i$. In other words, we cluster the runs of each letter.

For every letter $a \in \mathcal{A}$ let us consider the morphism $\psi_a : a \mapsto a;\ b \mapsto ab$ for $b \neq a$. Given an epistandard word $\mathbf{w}$ with directive sequence $\Delta = a_1^{e_1} a_2^{e_2} \cdots$, we define the morphisms $\sigma_n = \psi_{a_1}^{e_1} \psi_{a_2}^{e_2} \cdots \psi_{a_n}^{e_n}$ and the prefixes $w_n = \sigma_n(a_{n+1})$, where $\sigma_0 = \mathrm{id}$. Note that these prefixes are different from the palindromic ones seen above. Indeed, we have $u_n = w_{n-2} w_{n-3} \cdots w_0$ (see [6]).

Let $\ell(k)$ be the last occurrence in $\Delta$ of the run of $a_{k+1}$ before the $k-$th run. Notice that $\ell(k)$ is defined only starting from the second run of the letter $a_k$ in $\Delta$.

*Example 2.* Let us consider the infinite aperiodic sequence $\Delta = \mathtt{0}^1\mathtt{1}^2\mathtt{0}^3\mathtt{1}^4\mathtt{0}^5\mathtt{1}^6\cdots$ over the alphabet $\{\mathtt{0}, \mathtt{1}\}$. Then we have $a_i = \mathtt{0}$ when $i$ is odd, and $a_i = \mathtt{1}$ when $i$ is even. It is easy to see that $\ell(0)$ and $\ell(1)$ are not defined, and that $\ell(n) = n-2$ for every $n \geq 2$.

*Example 3.* Let us consider the periodic sequence $\Delta = (\mathtt{012}\cdots(\mathtt{d}-\mathtt{1}))^{\omega}$ over the alphabet $\{\mathtt{0}, \mathtt{1}, \ldots(\mathtt{d}-\mathtt{1})\}$. Then $a_k = (\mathtt{k}-\mathtt{1}) \pmod d$ for every positive integer $k$ and $\ell(k)$ is defined only for $k \geq d$, in which case it is equal to $\ell(k) = k - d$.

**Proposition 1 ([10]).** *Let $\mathbf{w}$ be an epistandard word with directive sequence $\Delta = a_1^{e_1} a_2^{e_2} \cdots$. The prefix $w_k$ is given by $w_k = \left(\prod_{i=1}^{k} w_{k-i}^{e_{k-i+1}}\right) a_{k+1}$ if the letter $a_{k+1}$ is not a factor of $w_k$, and by $w_k = \left(\prod_{i=1}^{k-\ell(k)-1} w_{k-i}^{e_{k-i+1}}\right) w_{\ell(k)}$ otherwise.*

*Example 4.* Let $\mathbf{w}$ be the $d$-bonacci word defined in Example 1. According to Proposition 1 the prefixes $w_k$ are obtained as $w_k = \left(\prod_{i=0}^{k} w_{k-i}^{1}\right)(\mathtt{k}-\mathtt{1})$

for $0 \leq k < d$ (e.g., $w_0 = \mathtt{0}$, $w_1 = \mathtt{0} \cdot \mathtt{1}$, $w_2 = \mathtt{01} \cdot \mathtt{0} \cdot \mathtt{2}$, etc.) and $w_n = w_{n-1}^1 \cdots w_{n-d+1}^1 w_{n-d}$ for $k > d$.

An epistandard word $\mathbf{w}$ is called *regular* if its directive word $\Delta$ is regular, i.e., if the runs of the letters in $\Delta$ appear in lexicographic (circular) order (see, e.g., [10]). Note that not every regular word is periodic (see Example 2), and not every periodic word is regular (e.g., $\Delta = (\mathtt{abac})^\omega$).

*Example 5.* Let $\mathbf{w}$ be a regular epistandard word with periodic directive word $\Delta = a_1^{e_1} a_2^{e_2} \cdots$ of period $p = md$, where $d$ the cardinality of the alphabet. Then, $w_k = w_{k-1}^{e_k} \cdots w_0^{e_1} a_{k+1}$ for $1 \leq k < d$ and $w_k = w_{k-1}^{e_k} \cdots w_{k-d+1}^{e_{k-d+2}} w_{k-d}$ for $k \geq d$.

The following result is a main ingredient to prove Theorem 1.

**Theorem 2 ([4]).** *A strict epistandard word is the fixed point of a morphism if and only if its directive sequence $\Delta$ is periodic.*

## 3 Cellular automata

In the following we use the terminology developed by Mazoyer and Terrier in [9] and Marcovici, Stoll and Tahay in [8].

**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 acting uniformly and synchronously on the configuration space. More precisely, there exists an integer $r \in \mathbb{N}$ called the* radius *of the CA, and a* local rule *$\tau : \mathcal{A}^{2r+1} \to \mathcal{A}$ such that for every $\mathbf{x} = (a_k)_{k\in\mathbb{Z}}$ and for every $k \in \mathbb{Z}$, we have $T(\mathbf{x})_k = \tau((a_{k+i})_{-r\leq i\leq r})$.*

When the alphabet $\mathcal{A}$ is understood, we call *cellular automaton* just the map $T$. The elements of $\mathcal{A}^\mathbb{Z}$ are called *configurations*. 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.

In a space-time diagram it is also possible to "transmit information" through *signals*, by connecting 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 [9] for a formal definition). When $m = m'$, we call such a signal a *vertical signal* or a signal of *infinite* slope. We suppose that a signal stops whenever a new signal starts. Signals are usually "porous", i.e., they do not interact between each other. In some case, however, we also need to consider "concrete" signals, called *walls*. Intuitively, a *wall* is a vertical signal such that whenever a given signal hit it, this signal "bounces" back. When two signals meet, we can *mark* the cell at the intersection, i.e., assign to it a letter from the alphabet $\mathcal{A}$, and define new signals starting from it.

Given a letter $z \in \mathcal{A}$ we define the class of cellular automata

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

where $T$ is *$z$-quiescent* if $T(z^\mathbb{Z}) = z^\mathbb{Z} = \cdots zzz \cdots$, and a configuration $\mathbf{x} = (a_k)_{k\in\mathbb{Z}}$ is *finite* if the set $\{k \in \mathbb{Z} : a_k \neq z\}$ is finite.

## 4 Construction of the automaton

We say that a CA *recognizes* an infinite word $\mathbf{w} = a_1 a_2 \cdots$ if this word appears vertically in the zero column, i.e., if $(T^n(\mathbf{w})_0)_{n>0} = \mathbf{w}$. To prove our main result we proceed in two steps.

1. First, we construct a CA recognizing the infinite word $\mathbb{1}_{\{|w_n|\}_{n\geq 0}}$ coding the lengths of the prefixes $w_n$ of our epistandard word $\mathbf{w}$.
2. Then, we construct a CA that recognizes $\mathbf{w}$, by constructing at each step the prefix $w_k$ using the previous prefixes.

The automaton in the first step above can be constructed using a modified version of the following result.

**Proposition 2 ([9]).** *Let $(S_n)_{n\geq 0}$ be an integer sequence defined by $S_{n+p} = \sum_{i=0}^{p-1} \alpha_i S_{n+i}$, where $p, \alpha_i \in \mathbb{N}$. Then $\mathbb{1}_{\{S_n\}_{n\geq 0}} \in \mathcal{S}$.*

To construct the automaton in second step we first show that it is possible to "copy-paste" a letter in a cell of a CA in the same column and in a chosen row above. We then use this construction to copy at each step the right number of copies of each of the $k - \ell(k)$ prefixes of $w_k$, according to the formula in Proposition 1.

Note that in both steps, the hypothesis of periodicity of the sequence $\Delta$ is essential to guarantee that the cellular automaton is defined over a finite set $\mathcal{A}$.

## 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. A. de Luca. Sturmian words: structure, combinatorics and their arithmetics. *Theoret. Comput. Sci.*, 183:123–138, 1997.
3. F. Dolce and P.-A. Tahay. Column representation of Sturmian words in cellular automata. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13257:127–138, 2022.
4. X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255:539–553, 2001.
5. A. Glen and J. Justin. Episturmian words: a survey. *RAIRO - Theor. Inform. Appl.*, 43:403–442, 2009.
6. J. Justin and G. Pirillo. Episturmian words and episturmian morphisms,. *Theoret. Comput. Sci.*, 276:281–313, 2002.
7. M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
8. 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.
9. J. Mazoyer and V. Terrier. Signals in one-dimensional cellular automata. *Theoret. Comput. Sci.*, 217:53–80, 1999.
10. J. Peltomäki. Initial nonrepetitive complexity of regular episturmian words and their diophantine exponents. *European Journal of Combinatorics*, 118, 2024.