


A word of low complexity without uniform frequencies

JULIEN CASSAIGNE[†] and IDRISSA KABORÉ[‡] 

[†] *Institut de Mathématiques de Marseille, 163 avenue de Luminy, case 907, F-13288
Marseille Cedex 9, France*

(e-mail: julien.cassaigne@math.cnrs.fr)

[‡] *UFR-Sciences Exactes et Appliquées, Université Nazi Boni, 01 BP 1091
Bobo-Dioulasso 01, Burkina Faso*

(e-mail: ikaborei@yahoo.fr)

(Received 06 October 2022 and accepted in revised form 17 April 2023)

Abstract. In this paper, we construct a uniformly recurrent infinite word of low complexity without uniform frequencies of letters. This shows the optimality of a bound of Boshernitzan, which gives a sufficient condition for a uniformly recurrent infinite word to admit uniform frequencies.

Key words: infinite word, substitution, complexity, uniform frequencies

2020 Mathematics Subject Classification: 37B10, 37A25, 68R15 (Primary)

1. Introduction

Let us consider an infinite word u over a finite alphabet. We can naturally associate a subshift with it. The goal of this paper is to describe some ergodic properties of this subshift. By Oxtoby's theorem, we know that the subshift is uniquely ergodic if and only if, in u , each finite word has uniform frequency. Moreover, the subshift is minimal if u is uniformly recurrent.

For a long time, people have tried to find some conditions on infinite words which imply one of these properties. Keane gave in [7] a uniformly recurrent infinite word with complexity $3n + 1$ (from a 4-interval exchange map) which does not have uniform frequencies. Later, Boshernitzan, in [2], proved that a uniformly recurrent infinite word admits uniform frequencies if either of the following sufficient conditions is satisfied:

$$\liminf \frac{\mathbf{p}(n)}{n} < 2 \quad \text{or} \quad \limsup \frac{\mathbf{p}(n)}{n} < 3,$$

where \mathbf{p} denotes the complexity function of u (see [1, Ch. 4, by Cassaigne and Nicolas], for more details on the subject).



The bound of the second sufficient condition of Boshernitzan being already optimal by Keane's result, our goal in this work is to establish the optimality of the bound of the first sufficient condition. Additionally, we succeed to construct a uniformly recurrent infinite word without uniform frequencies, the complexity function of which satisfies $\liminf \mathbf{p}(n)/n = 2$.

This result can be viewed as part of the more general question of relating some properties of the complexity function and of the ergodic measures of a subshift. The goal is then to bound the number of ergodic measures of the subshift in terms of the complexity function.

Boshernitzan was the first to look at this question, see [2]. During his PhD thesis, see [8, Ch. 5] and [1, Ch. 7, by Ferenczi and Monteil], Monteil has proved the same result with different techniques: if $\limsup \mathbf{p}(n)/n < K$ for some integer $K \geq 3$, then the subshift has at most $K - 2$ ergodic measures. Since that time, more results have appeared in the same vein. Cyr and Kra have also obtained similar results, see [4, 5]. In the first paper, they prove that the bound of Boshernitzan is sharp. In the second one, they construct minimal subshifts with complexity function arbitrarily close to linear but having uncountably many ergodic measures. We can also cite Damron and Fickenscher [6] who obtained the bound $K + 1/2$ under a condition on the bispecial words.

The construction presented in this paper was first announced at a conference in 2007, and published without proof in [1] (Proposition 7.5.12). For several reasons, including the fact that the more general results discussed above had appeared in the meantime, its complete proof was not published then. Nevertheless, it was pointed out to us that our proof is of a different nature, with an explicit construction of the infinite word, and that it was thus interesting by itself.

After the preliminaries (§2), we construct an infinite word in §3, then we show in §4 that this word is without uniform frequencies of letters, we study the complexity of this word in §5, and finally we prove that it is uniformly recurrent in §6.

2. Preliminaries

In all that follows, we consider the alphabet $\mathcal{A} = \{0, 1\}$. Let \mathcal{A}^* denote the set of the finite words on alphabet \mathcal{A} , and ε the empty word. For all u in \mathcal{A}^* , $|u|$ denotes the length of the word u (the number of letters it contains, with $|\varepsilon| = 0$) and for any letter x of \mathcal{A} , $|u|_x$ denotes the number of occurrences of the letter x in u . We call the Parikh vector of a finite word u the vector denoted by U and defined by $U = \begin{pmatrix} |u|_0 \\ |u|_1 \end{pmatrix}$.

A finite word u of length n formed by repeating a single letter x is typically denoted by x^n . We define the n th power of a finite word w as being the concatenation of n copies of w ; it is denoted by w^n . An infinite word is an infinite sequence of letters of \mathcal{A} . Let \mathcal{A}^ω denote the set of infinite words on \mathcal{A} .

We say that a finite word v is a factor of u if there exist two words u_1 and u_2 on the alphabet \mathcal{A} such that $u = u_1 v u_2$; we also say that u contains v . The number of different pairs (u_1, u_2) such that $u = u_1 v u_2$ is called the number of occurrences of v in u and is denoted by $|u|_v$, generalizing the notation $|u|_x$ for $x \in \mathcal{A}$ defined above. The factor v is said to be a prefix (respectively suffix) of u if u_1 (respectively u_2) is the empty word.

For any word u , the set of factors of length n is denoted by $\mathcal{L}_n(u)$. The set of all factors of u is simply denoted by $\mathcal{L}(u)$.

Definition 2.1. Let u be an infinite word on the alphabet $\mathcal{A} = \{0, 1\}$. A factor v of u is said to be:

- a right special factor if $v0$ and $v1$ are both factors of u , and a left special factor if $0v$ and $1v$ are both factors of u ;
- a bispecial factor of u if v is simultaneously a right special factor and a left special factor of u ;
- a strong bispecial factor of u if $0v0, 0v1, 1v0, 1v1$ are factors of u and a weak bispecial factor if uniquely $0v0$ and $1v1$, or $0v1$ and $1v0$, are factors of u ;
- a neutral bispecial factor of u if v is a bispecial factor of u which is neither strong nor weak.

An infinite word u is said to be recurrent if any factor of u appears infinitely often. An infinite word u is uniformly recurrent if for all $n \in \mathbb{N}$, there exists N such that any factor of u of length N contains all the factors of u of length n .

Definition 2.2. Let u be an infinite word on an alphabet \mathcal{A} . The complexity function of u is a function counting the number of distinct factors of u of length n for any given n . It is denoted by \mathbf{p} and satisfies

$$\mathbf{p}(n) = \#\mathcal{L}_n(u).$$

We also define the functions \mathbf{s} and \mathbf{b} , respectively called first difference and second difference of the complexity of u , as follows: $\mathbf{s}(n) = \mathbf{p}(n + 1) - \mathbf{p}(n)$ and $\mathbf{b}(n) = \mathbf{s}(n + 1) - \mathbf{s}(n)$.

On a binary alphabet, the function \mathbf{s} counts the number of special factors of a given length in u . Let \mathbf{m} denote the map from $\mathcal{L}(u)$ to $\{-1, 0, +1\}$ defined by

$$\text{for all } v \in \mathcal{L}(u), \mathbf{m}(v) = \begin{cases} -1 & \text{if } v \text{ is weak bispecial,} \\ +1 & \text{if } v \text{ is strong bispecial,} \\ 0 & \text{otherwise.} \end{cases}$$

The following formula was given by the first author in [3].

PROPOSITION 2.1. *If u is a recurrent binary infinite word, then*

$$\text{for all } n \geq 0, \quad \mathbf{s}(n) = 1 + \sum_{\substack{w \in \mathcal{L}(u) \\ |w| < n}} \mathbf{m}(w) = 1 + \sum_{\substack{w \text{ bispecial} \\ |w| < n}} \mathbf{m}(w).$$

This relation allows to compute the complexity $\mathbf{p}(n)$ when we are able to describe the set of strong and weak bispecial factors of the binary infinite word u .

Definition 2.3. Two bispecial factors v and w of an infinite word u on the alphabet $\{0, 1\}$ are said to have the same type if they are both strong, weak, or neutral. In other words, the bispecial factors v and w have the same type if $\mathbf{m}(v) = \mathbf{m}(w)$.

Definition 2.4. [1, Ch. 7, by Ferenczi and Monteil] Let u be an infinite word on an alphabet \mathcal{A} .

- We say that u admits frequencies if for any factor w , and any sequence (u_n) of prefixes of u such that $\lim_{n \rightarrow \infty} |u_n| = \infty$, then $\lim_{n \rightarrow \infty} |u_n|_w / |u_n|$ exists.
- We say that u admits uniform frequencies if for any factor w , and any sequence (u_n) of factors of u such that $\lim_{n \rightarrow \infty} |u_n| = \infty$, then $\lim_{n \rightarrow \infty} |u_n|_w / |u_n|$ exists.

In [7], M. Keane gave an example of a uniformly recurrent infinite word with complexity $3n + 1$ which does not admit uniform frequencies. Later, Boshernitzan [2] obtained the following results.

THEOREM 2.1. (Boshernitzan [2]) *Let u be an infinite word on an alphabet \mathcal{A} . Then, u admits uniform frequencies if its complexity function satisfies at least one of the following conditions:*

- $\liminf \mathbf{p}(n)/n < 2$;
- $\limsup \mathbf{p}(n)/n < 3$.

The example of Keane ensures that constant 3 is optimal in the second condition, that is, it cannot be replaced with a larger constant.

3. Construction of a class of uniformly recurrent words

Let $(l_i), (m_i), (n_i)$ be three integer sequences which are strictly increasing and satisfy the following conditions:

- $0 < l_i < m_i < n_i$;
- m_i/l_i increases exponentially to $+\infty$;
- n_i/m_i increases exponentially to $+\infty$.

Let us define in \mathcal{A}^* two sequences (u_i) and (v_i) in the following way: $u_0 = 0, v_0 = 1$ and for all $i \in \mathbb{N}, u_{i+1} = u_i^{m_i} v_i^{l_i}$ and $v_{i+1} = u_i^{n_i} v_i^{m_i}$. The sequence (u_i) converges toward an infinite word u .

For $i \geq 1$, consider the substitution σ_i defined by $\sigma_i(0) = 0^{m_i} 1^{l_i}, \sigma_i(1) = 0^{m_i} 1^{n_i}$. Then, we have

$$u_i = \sigma_0 \sigma_1 \sigma_2 \dots \sigma_{i-1}(0) \quad \text{and} \quad v_i = \sigma_0 \sigma_1 \sigma_2 \dots \sigma_{i-1}(1).$$

THEOREM 3.1. *Any infinite word u so defined is uniformly recurrent.*

The proof of Theorem 3.1 is given in §6 at the end of this paper.

4. The word u is without uniform frequencies

LEMMA 4.1. *For all $i \geq 1$, we have:*

- (1) $|u_i|_0 / |u_i| \geq (1 + (l_0/m_0))^{-1} \prod_{j=1}^{i-1} (1 + (l_j n_{j-1} / m_j l_{j-1}))^{-1}$;
- (2) $|v_i|_1 / |v_i| \geq \prod_{j=0}^{i-1} (1 + (m_j/n_j))^{-1}$.

Proof. (1) Lower bound on $|u_{i+1}|_0 / |u_{i+1}|$. First, for all $i \geq 0$, we have $|u_i| \leq |v_i|$ since $|u_0| = |v_0| = 1$ and u_i is a strict prefix of v_i for $i \geq 1$. Then, $|v_i| / |u_i| = m_{i-1} |u_{i-1}| + n_{i-1} |v_{i-1}| / m_{i-1} |u_{i-1}| + l_{i-1} |v_{i-1}| \leq n_{i-1} / l_{i-1}$ since $l_{i-1} < n_{i-1}$ for $i \geq 1$.

As

$$|u_{i+1}|_0 = m_i|u_i|_0 + l_i|v_i|_0 \geq m_i|u_i|_0$$

and

$$|u_{i+1}| = m_i|u_i| + l_i|v_i| = |u_i| \left(m_i + l_i \frac{|v_i|}{|u_i|} \right),$$

we deduce the following inequalities:

$$|u_{i+1}| \leq |u_i| \left(m_i + l_i \frac{n_{i-1}}{l_{i-1}} \right)$$

and

$$\frac{|u_{i+1}|_0}{|u_{i+1}|} \geq \left(1 + \frac{l_i}{m_i} \frac{n_{i-1}}{l_{i-1}} \right)^{-1} \cdot |u_i|_0/|u_i|.$$

Thus,

$$\frac{|u_i|_0}{|u_i|} \geq \frac{|u_1|_0}{|u_1|} \prod_{j=1}^{i-1} \left(1 + \frac{l_j n_{j-1}}{m_j l_{j-1}} \right)^{-1}.$$

(2) Lower bound on $|v_{i+1}|_1/|v_{i+1}|$. We have

$$|v_{i+1}|_1 = m_i|u_i|_1 + n_i|v_i|_1 \geq n_i|v_i|_1 \text{ and } |v_{i+1}| = m_i|u_i| + n_i|v_i| \leq |v_i|(m_i + n_i),$$

since $|u_i| \leq |v_i|$. So,

$$\frac{|v_{i+1}|_1}{|v_{i+1}|} \geq \frac{n_i}{m_i + n_i} \cdot \frac{|v_i|_1}{|v_i|}.$$

Hence,

$$\frac{|v_i|_1}{|v_i|} \geq \prod_{j=0}^{i-1} \left(1 + \frac{m_j}{n_j} \right)^{-1} \quad \square$$

In the rest of the paper, we shall fix

$$l_i = 2^{2 \cdot 2^i + 4}, \quad m_i = 2^{8 \cdot 2^i}, \quad \text{and} \quad n_i = 2^{10 \cdot 2^i} \quad \text{for } i \geq 0. \quad (*)$$

Remark 4.1. The particular choice of sequences in (*) was designed to get simple bounds for frequencies (in Lemma 4.3) and to ensure the desired ordering of bispecial factors (in Lemma 5.4). It is not the only possible choice, and we could have stated results for a larger class of words. However, this would have made computations more complicated and less explicit, while our purpose here is to construct one example as explicit as possible.

With (*), the inequalities of Lemma 4.1 become:

- (1) $|u_i|_0/|u_i| \geq \prod_{j=1}^i 1/(1 + 2^{-2^j})$;
- (2) $|v_i|_1/|v_i| \geq \prod_{j=1}^i 1/(1 + 2^{-2^j})$.

So we get the following lemma.

LEMMA 4.2. For all $i \geq 1$,

$$\min \left(\frac{|u_i|_0}{|u_i|}, \frac{|v_i|_1}{|v_i|} \right) \geq \prod_{j=1}^i \frac{1}{1 + 2^{-2^j}}.$$

Then we have the following lemma.

LEMMA 4.3. For all $i \in \mathbb{N}$,

$$\frac{|u_i|_0}{|u_i|} + \frac{|v_i|_1}{|v_i|} \geq \frac{3}{2}.$$

Proof. For $i = 0$, the inequality is evident.

For $i \geq 1$, write $P_i = \prod_{j=1}^i 1/(1 + 2^{-2^j})$. The sequence (P_i) is decreasing. Let us show, by induction, that $\frac{4}{3}P_i = 1/(1 - 2^{-2^{i+1}})$.

We have $\frac{4}{3}P_0 = 1/(1 - 2^{-2})$.

Assuming that for some $i \geq 0$, $\frac{4}{3}P_i = 1/(1 - 2^{-2^{i+1}})$, it follows that

$$\frac{4}{3}P_{i+1} = \frac{4}{3}P_i \times \frac{1}{1 + 2^{-2^{i+1}}} = \frac{1}{1 - 2^{-2^{i+1}}} \times \frac{1}{1 + 2^{-2^{i+1}}} = \frac{1}{1 - 2^{-2^{i+2}}}.$$

So, for all $i \in \mathbb{N}$,

$$P_i = \frac{3}{4} \times \frac{1}{1 - 2^{-2^{i+1}}}.$$

Hence, with Lemma 4.2, we get

$$\frac{|u_i|_0}{|u_i|} + \frac{|v_i|_1}{|v_i|} \geq 2 \times \frac{3}{4} \times \frac{1}{1 - 2^{-2^{i+1}}} \geq \frac{3}{2}. \quad \square$$

THEOREM 4.1. The letters of the word u do not admit uniform frequencies.

Proof. If the letters of u had uniform frequencies, then the frequencies of 0 and 1, respectively denoted by $\mathbf{f}_u(0)$ and $\mathbf{f}_u(1)$, would satisfy $\mathbf{f}_u(0) = \lim_{i \rightarrow \infty} |u_i|_0/|u_i|$, $\mathbf{f}_u(1) = \lim_{i \rightarrow \infty} |v_i|_1/|v_i|$, and $\mathbf{f}_u(0) + \mathbf{f}_u(1) = 1$, which is contradictory with Lemma 4.3. \square

5. Complexity of u

To estimate the complexity of u , we are going to observe its bispecial factors.

Notation 5.1. Let $h, i \in \mathbb{N}$. We let $u_i^{(h)}$ denote the finite word

$$u_i^{(h)} = \sigma_h \sigma_{h+1} \sigma_{h+2} \dots \sigma_{h+i-1}(0)$$

and $u^{(h)}$ denotes the infinite word $\lim_{i \rightarrow \infty} u_i^{(h)}$. Note that $u_i^{(0)} = u_i$ and $u^{(0)} = u$.

Definition 5.1. A factor of $u^{(h)}$ is said to be short if it does not contain 10 as a factor. A factor of $u^{(h)}$ which is not short is said to be long.

LEMMA 5.1. (Synchronization lemma) Let w be a long factor of $u^{(h)}$. Then there exist $x, y \in \mathcal{A}$ and $v \in \mathcal{A}^*$ such that xvy is a factor of $u^{(h+1)}$ and $w = s\sigma_h(v)p$, where s is

a non-empty suffix of $\sigma_h(x)$ and p is a non-empty prefix of $\sigma_h(y)$. Moreover, the triple (s, v, p) is unique.

Proof. Since w is long, it cannot occur inside the image of one letter. Any occurrence of w in u is therefore of the form $s\sigma_h(v)p$, so existence follows.

Uniqueness is a consequence of the fact that 10 occurs in $u^{(h)}$ only at the border between two images of letters under σ_h . □

LEMMA 5.2

- (1) The short and strong bispecial factors of $u^{(h)}$ are ε and 1^{l_h} .
- (2) The short and weak bispecial factors of $u^{(h)}$ are 0^{m_h-1} and 1^{n_h-1} .

Proof. Let us first observe that in $u^{(h)}$, the factor 01 is always preceded by 10^{m_h-1} . Therefore, a bispecial factor containing 01 must also contain 10 and is long.

Then the short bispecial factors are all of the form 0^k or 1^k , $k \geq 0$. We see that ε is strong bispecial (extensions $00, 01, 10, 11$); 0^k ($0 \leq k < m_h - 1$) is neutral bispecial (extensions $00^k0, 00^k1, 10^k0$), as well as 1^k ($1 \leq k < n_h - 1, k \neq l_h$); 1^{l_h} is strong bispecial (extensions $01^{l_h}0, 01^{l_h}1, 11^{l_h}0, 11^{l_h}1$); 0^{m_h-1} is weak bispecial (extensions $00^{m_h-1}1$ and $10^{m_h-1}0$), as well as 1^{n_h-1} ; 0^{m_h} and 1^{n_h} are not special, and 0^k ($k > m_h$) and 1^k ($k > n_h$) are not factors. □

LEMMA 5.3. Let w be a factor of $u^{(h)}$. Then the following assertions are equivalent:

- (1) w is a long bispecial factor of $u^{(h)}$;
- (2) there exists a bispecial factor v of $u^{(h+1)}$ such that $w = \widehat{\sigma}_h(v)$, where $\widehat{\sigma}_h(v) = 1^{l_h}\sigma_h(v)0^{m_h}1^{l_h}$.

Moreover, v and w have the same type and $|v| < |w|$.

Proof. First, let us observe this fact: if a finite word v is a factor of $u^{(h+1)}$, then $\widehat{\sigma}_h(v) = 1^{l_h}\sigma_h(v)0^{m_h}1^{l_h}$ is a factor of $u^{(h)}$. Now, let us consider a bispecial factor v of $u^{(h+1)}$. Therefore, the words $\widehat{\sigma}_h(0v)$, $\widehat{\sigma}_h(1v)$, $\widehat{\sigma}_h(v0)$, and $\widehat{\sigma}_h(v1)$ are factors of $u^{(h)}$; moreover, $0\widehat{\sigma}_h(v)$ and $1\widehat{\sigma}_h(v)$ are respectively the suffix of the first two words, whereas $\widehat{\sigma}_h(v)0$ and $\widehat{\sigma}_h(v)1$ are respectively the prefix of the last two words. Hence, the word $w = \widehat{\sigma}_h(v)$ is bispecial in $u^{(h)}$, and $\mathbf{m}(w) \geq \mathbf{m}(v)$.

Conversely, let w be a long bispecial factor of $u^{(h)}$. Then, according to the synchronization lemma, we can write w uniquely in the form $s\sigma_h(v)p$, where s and p are respectively the non-empty suffix and prefix of images of letters.

As $0w$ and $1w$ are factors of $u^{(h)}$, and $\sigma_h(v)p$ starts with 0 , it follows that $0s0$ and $1s0$ are factors of $u^{(h)}$. This is only possible if $s = 1^{l_h}$ ($s = 1^k$ with $1 \leq k < l_h$ or $l_h < k < n_h$ are excluded since $0s0 \notin L(u^{(h)})$); $s = 0^k1^{l_h}$ with $1 \leq k < m_h$ and $s = 0^k1^{n_h}$ with $0 \leq k < m_h$ are excluded since $1s0 \notin \mathcal{L}(u^{(h)})$; and $s = 0^{m_h}1^{l_h}$ and $s = 0^{m_h}1^{n_h}$ are excluded since $0s0 \notin \mathcal{L}(u^{(h)})$.

Similarly, $1p0$ and $1p1$ are factors of $u^{(h)}$, and this is only possible if $p = 0^{m_h}1^{l_h}$. Therefore, $w = \widehat{\sigma}_h(v)$.

If w extends as awb with $a, b \in \mathcal{A}$, then v also extends as avb . Therefore, $\mathbf{m}(v) \geq \mathbf{m}(w)$. It follows that $\mathbf{m}(v) = \mathbf{m}(w)$: v and w have the same type. Moreover, it is clear that $|v| < |w|$. □

In fact, long bispecial factors of $u^{(h)}$ are the images under $\widehat{\sigma}_h$ of the ‘less long’ bispecial factors of $u^{(h+1)}$. Thus, step by step, any non-neutral bispecial factor w of $u^{(h)}$ of given type will be written in the form $\widehat{\sigma}_h \widehat{\sigma}_{h+1} \dots \widehat{\sigma}_{h+i-1}(v)$, where v is a short bispecial factor of $u^{(h+i)}$ with the same type.

We will call bispecial factors of rank i , ($i \geq 0$) of $u^{(h)}$, and write $a_i^{(h)}$, $b_i^{(h)}$, $c_i^{(h)}$, $d_i^{(h)}$ as the following words:

$$\begin{aligned} a_i^{(h)} &= \widehat{\sigma}_h \widehat{\sigma}_{h+1} \dots \widehat{\sigma}_{h+i-1}(\varepsilon); \\ b_i^{(h)} &= \widehat{\sigma}_h \widehat{\sigma}_{h+1} \dots \widehat{\sigma}_{h+i-1}(1^{l_{h+i}}); \\ c_i^{(h)} &= \widehat{\sigma}_h \widehat{\sigma}_{h+1} \dots \widehat{\sigma}_{h+i-1}(0^{m_{h+i-1}}); \\ d_i^{(h)} &= \widehat{\sigma}_h \widehat{\sigma}_{h+1} \dots \widehat{\sigma}_{h+i-1}(1^{n_{h+i-1}}). \end{aligned}$$

The short bispecial ε , 1^l , $0^{m_{h-1}}$, and $1^{n_{h-1}}$ of $u^{(h)}$ are the bispecial factors of rank 0: $a_0^{(h)}$, $b_0^{(h)}$, $c_0^{(h)}$, and $d_0^{(h)}$.

The non-neutral bispecial factors of u are therefore $a_i = a_i^{(0)}$, $b_i = b_i^{(0)}$, $c_i = c_i^{(0)}$, and $d_i = d_i^{(0)}$, for $i \geq 0$.

Definition 5.2. Let $V, W \in \mathbb{R}^2$. Let us write $V < W$ when $W - V$ has non-negative entries and $V \neq W$.

PROPOSITION 5.1. *Let v, w, v', w' be four words such that $v' = \widehat{\sigma}_i(v)$ and $w' = \widehat{\sigma}_i(w)$. Then their Parikh vectors satisfy*

$$V < W \implies V' < W'.$$

Proof. Assume that $V < W$. Then, $|v|_0 \leq |w|_0$, $|v|_1 \leq |w|_1$, and $|v| < |w|$. On the one hand, we have $|v'|_0 = m_i(|v| + 1)$ and $|w'|_0 = m_i(|w| + 1)$; hence, $|v'|_0 < |w'|_0$. On the other hand, we have $|v'|_1 = l_i|v|_0 + n_i|v|_1 + 2l_i$ and $|w'|_1 = l_i|w|_0 + n_i|w|_1 + 2l_i$; so $|v'|_1 \leq |w'|_1$. Finally, $|v'| = |v'|_0 + |v'|_1 < |w'|_0 + |w'|_1 = |w'|$. \square

LEMMA 5.4. *For all $i \geq 0$, let A_i, B_i, C_i, D_i be the Parikh vectors corresponding to the non-neutral bispecial factors of u , a_i, b_i, c_i, d_i . Then, we have*

$$\text{for all } i \geq 1, \quad D_{i-1} < B_i < C_i < A_{i+1} < D_i.$$

Proof. Applying $\widehat{\sigma}_{i-1}$ on the words $b_0^{(i)}$, $c_0^{(i)}$, $\widehat{\sigma}_i(a_0^{(i+1)}) = 1^{l_i} 0^{m_i} 1^{l_i}$, and $d_0^{(i)}$, we get the following words:

$$\begin{aligned} d_0^{(i-1)} &= 1^{n_{i-1}-1}; \\ b_1^{(i-1)} &= 1^{l_{i-1}} (0^{m_{i-1}} 1^{n_{i-1}})^{l_i} 0^{m_{i-1}} 1^{l_{i-1}}; \\ c_1^{(i-1)} &= 1^{l_{i-1}} (0^{m_{i-1}} 1^{l_{i-1}})^{m_i-1} 0^{m_{i-1}} 1^{l_{i-1}}; \\ a_2^{(i-1)} &= 1^{l_{i-1}} (0^{m_{i-1}} 1^{n_{i-1}})^{l_i} (0^{m_{i-1}} 1^{l_{i-1}})^{m_i} (0^{m_{i-1}} 1^{n_{i-1}})^{l_i} 0^{m_{i-1}} 1^{l_{i-1}}; \\ d_1^{(i-1)} &= 1^{l_{i-1}} (0^{m_{i-1}} 1^{n_{i-1}})^{n_i-1} 0^{m_{i-1}} 1^{l_{i-1}}. \end{aligned}$$

The Parikh vectors corresponding to these words are:

$$\begin{aligned}
 D_0^{(i-1)} &= \begin{pmatrix} 0 \\ n_{i-1} - 1 \end{pmatrix}; \\
 B_1^{(i-1)} &= \begin{pmatrix} m_{i-1}(l_i + 1) \\ n_{i-1}l_i + 2l_{i-1} \end{pmatrix}; \\
 C_1^{(i-1)} &= \begin{pmatrix} m_i m_{i-1} \\ l_{i-1}(m_i + 1) \end{pmatrix}; \\
 A_2^{(i-1)} &= \begin{pmatrix} m_{i-1}(m_i + 2l_i + 1) \\ l_{i-1}(m_i + 2) + 2l_i n_{i-1} \end{pmatrix}; \\
 D_1^{(i-1)} &= \begin{pmatrix} m_{i-1}n_i \\ n_{i-1}(n_i - 1) + 2l_{i-1} \end{pmatrix}.
 \end{aligned}$$

From (*), we have

$$n_{i-1}l_i + l_{i-1} < l_{i-1}m_i, m_i + 2l_i + 1 < n_i, l_{i-1}m_i + 2n_{i-1}l_i < n_{i-1}(n_i - 1).$$

It follows the inequalities:

$$D_0^{(i-1)} < B_1^{(i-1)} < C_1^{(i-1)} < A_2^{(i-1)} < D_1^{(i-1)}.$$

Applying $\widehat{\sigma}_{i-2}$ on the words $d_0^{(i-1)}$, $b_1^{(i-1)}$, $c_1^{(i-1)}$, $a_2^{(i-1)}$, and $d_1^{(i-1)}$, we get the words $d_1^{(i-2)}$, $b_2^{(i-2)}$, $c_2^{(i-2)}$, $a_3^{(i-2)}$, and $d_2^{(i-2)}$. By Proposition 5.1, it results in the following inequalities:

$$D_1^{(i-2)} < B_2^{(i-2)} < C_2^{(i-2)} < A_3^{(i-2)} < D_2^{(i-2)}.$$

And so, after the i th iteration, we get

$$D_{i-1}^{(0)} < B_i^{(0)} < C_i^{(0)} < A_{i+1}^{(0)} < D_i^{(0)}. \quad \square$$

LEMMA 5.5. For all $i \geq 0$,

$$|b_i| < |c_i| < |a_{i+1}| < |d_i| < |b_{i+1}|.$$

Proof.

- For $i \geq 1$, the inequalities $|b_i| < |c_i| < |a_{i+1}| < |d_i| < |b_{i+1}|$ follow from Lemma 5.4.
- For $i = 0$, recall that $|b_0| = l_0$, $|c_0| = m_0 - 1$, $|a_1| = 2l_0 + m_0$, $|d_0| = n_0 - 1$, and $|b_1| = l_1(m_0 + n_0) + m_0 + 2l_0$. So,

$$|b_0| < |c_0| < |a_1| < |d_0| < |b_1|. \quad \square$$

LEMMA 5.6. The function s associated with the word u satisfies:

$$\text{for all } n \in \mathbb{N}, \quad s(n) = \begin{cases} 1 & \text{if } n = 0, \\ 2 & \text{if } n \in \bigcup_{i \geq 0} ([|c_i| + 1, |a_{i+1}|] \cup [|d_i| + 1, |b_{i+1}|]) \cup]0, |b_0|], \\ 3 & \text{if } n \in \bigcup_{i \geq 0} ([|b_i| + 1, |c_i|] \cup [|a_{i+1}| + 1, |d_i|]). \end{cases}$$

Proof. Let $n \in \mathbb{N}$. We know that $a_i, b_i, c_i,$ and $d_i,$ for $i \geq 0,$ are the only bispecial factors of u which are strong or weak. Hence, using Proposition 2.1, we have

$$\begin{aligned} s(n) &= 1 + \sum_{\substack{w \text{ bispecial} \\ |w| < n}} \mathbf{m}(w) \\ &= 1 + \#\{i \geq 0 : |a_i| < n\} \\ &\quad + \#\{i \geq 0 : |b_i| < n\} \\ &\quad - \#\{i \geq 0 : |c_i| < n\} \\ &\quad - \#\{i \geq 0 : |d_i| < n\}. \end{aligned}$$

Since for $m \in [1, |b_0| - 1],$ there is not a strong or weak bispecial factor of u with length $m,$ we have, for $0 < n \leq |b_0|:$

$$s(n) = 1 + \sum_{\substack{w \text{ bispecial} \\ |w| \leq n-1}} \mathbf{m}(w) = 1 + \mathbf{m}(\varepsilon) = 2.$$

Suppose $n > |b_0|.$ Then, there exists $i \in \mathbb{N}$ such that $n \in [|b_i| + 1, |b_{i+1}|].$ Since the sequences $|a_i|, |b_i|, |c_i|,$ and $|d_i|$ are increasing, we are in one of the following cases:

- $n \in [|b_i| + 1, |c_i|],$ then $s(n) = 1 + (i + 1) + (i + 1) - (i) - (i) = 3;$
- $n \in [|c_i| + 1, |a_{i+1}|],$ then $s(n) = 1 + (i + 1) + (i + 1) - (i + 1) - (i) = 2;$
- $n \in [|a_{i+1}| + 1, |d_i|],$ then $s(n) = 1 + (i + 2) + (i + 1) - (i + 1) - (i) = 3;$
- $n \in [|d_i| + 1, |b_{i+1}|],$ then $s(n) = 1 + (i + 2) + (i + 1) - (i + 1) - (i + 1) = 2.$

□

THEOREM 5.1. *The complexity function \mathbf{p} of u satisfies:*

$$\text{for all } n \geq 1, \quad \mathbf{p}(n) \leq 3n + 1.$$

Proof. By Lemma 5.6, $s(n) \leq 3$ for all $n \geq 0.$ As $\mathbf{p}(n) = \mathbf{p}(0) + \sum_{m=0}^{n-1} s(m),$ it follows that $\mathbf{p}(n) \leq \mathbf{p}(0) + 3(n) = 3n + 1.$ □

PROPOSITION 5.2. *Let v, w, v', w' be four finite words such that $v' = \widehat{\sigma}_i(v)$ and $w' = \widehat{\sigma}_i(w).$ Then for all $\lambda > 0,$ we have*

$$W > \lambda \left[V + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right] \implies W' > \lambda \left[V' + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right].$$

Proof. Assume that $W > \lambda[V + \begin{pmatrix} 1 \\ 1 \end{pmatrix}].$ Since $|v'|_0 = m_i(|v| + 1)$ and $|v'|_1 = l_i|v|_0 + n_i|v|_1 + 2l_i,$ then

$$v' = \begin{pmatrix} m_i & m_i \\ l_i & n_i \end{pmatrix} v + \begin{pmatrix} m_i \\ 2l_i \end{pmatrix}.$$

Here, W' is written in the same way. It follows

$$W' - \lambda \left[V' + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right] = \begin{pmatrix} m_i & m_i \\ l_i & n_i \end{pmatrix} (W - \lambda V) + (1 - \lambda) \begin{pmatrix} m_i \\ 2l_i \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Since

$$W > \lambda V + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad (1 - \lambda) \begin{pmatrix} m_i \\ 2l_i \end{pmatrix} > -\lambda \begin{pmatrix} m_i \\ 2l_i \end{pmatrix},$$

it follows that

$$\begin{aligned} W' - \lambda \left[V' + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right] &> \lambda \left[\begin{pmatrix} m_i & m_i \\ l_i & n_i \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} m_i + 1 \\ 2l_i + 1 \end{pmatrix} \right] \\ &= \lambda \begin{pmatrix} m_i - 1 \\ n_i - l_i - 1 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned} \quad \square$$

This proposition allows us to prove the following lemma.

LEMMA 5.7. For all $i \geq 0$,

$$B_{i+1} > l_{i+1} \left[D_i + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right].$$

Proof. Let us choose an integer $i \geq 1$. Then, we have $b_1^{(i)} = \widehat{\sigma}_i(b_0^{(i+1)}) = 1^{l_i} (0^{m_i} 1^{n_i})^{l_{i+1}} 0^{m_i} 1^{l_i}$ and $d_0^{(i)} = 1^{n_i-1}$; the corresponding Parikh vectors are $B_1^{(i)} = \begin{pmatrix} l_{i+1}m_i+m_i \\ l_{i+1}n_i+2l_i \end{pmatrix}$ and $D_0^{(i)} = \begin{pmatrix} 0 \\ n_i-1 \end{pmatrix}$. It follows the inequality:

$$B_1^{(i)} > l_{i+1} \left[D_0^{(i)} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right].$$

By regressive induction on $j \leq i$, suppose that

$$B_{i+1-j}^{(j)} > l_{i+1} \left[D_{i-j}^{(j)} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right],$$

where $B_{i+1-j}^{(j)}$ and $D_{i-j}^{(j)}$ are respectively Parikh vectors of the words $b_{i+1-j}^{(j)}$ and $d_{i-j}^{(j)}$. Thus, by Proposition 5.2,

$$B_{i+2-j}^{(j-1)} > l_{i+1} \left[D_{i-j+1}^{(j-1)} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]$$

since $B_{i+2-j}^{(j-1)}$ is the Parikh vector of $b_{i+2-j}^{(j-1)} = \widehat{\sigma}_{j-1}(b_{i+1-j}^{(j)})$ and $D_{i-j+1}^{(j-1)}$ is the Parikh vector of $d_{i-j+1}^{(j-1)} = \widehat{\sigma}_{j-1}(d_{i-j}^{(j)})$. So,

$$B_{i+1-j}^{(j)} > l_{i+1} \left[D_{i-j}^{(j)} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right] \quad \text{for } 0 \leq j \leq i.$$

In the inequality above, we find the lemma by making $j = 0$. □

THEOREM 5.2. The complexity function \mathbf{p} of u satisfies $\liminf \mathbf{p}(n)/n = 2$.

Proof. We have $\mathbf{s}(n) = 2$ for $|d_i| < n \leq |b_{i+1}|$. So,

$$\mathbf{p}(|b_{i+1}|) = \mathbf{p}(|d_i| + 1) + 2(|b_{i+1}| - |d_i| - 1).$$

By Lemma 5.6, we have $\mathbf{p}(n) \leq 3n + 1$ and we deduce that

$$\begin{aligned} \mathbf{p}(|b_{i+1}|) &\leq (3|d_i| + 4) + 2(|b_{i+1}| - |d_i| - 1) \\ &\leq 2|b_{i+1}| + \frac{1}{l_{i+1}}|b_{i+1}| \end{aligned}$$

since $|b_{i+1}| > l_{i+1}(|d_i| + 2)$ as $B_{i+1} > l_{i+1}[D_i + (\frac{1}{l_{i+1}})]$. So,

$$\frac{\mathbf{p}(|b_{i+1}|)}{|b_{i+1}|} \leq 2 + \frac{1}{l_{i+1}} \quad \text{and} \quad \lim_{i \rightarrow \infty} \frac{\mathbf{p}(|b_{i+1}|)}{|b_{i+1}|} = 2.$$

As $\mathbf{p}(n) \geq 2n$ since $\mathbf{s}(n) \geq 2$ for $n \geq 1$, we get $\liminf \mathbf{p}(n)/n = 2$. □

6. Proof of Theorem 3.1

Now, with Notation 5.1, we are able to explain the proof of Theorem 3.1.

Proof. Let us show that for $i \geq 0$, there exists N_i such that any factor of u of length N_i contains the prefix u_i . Note that u does not contain 1^{n_0+1} .

- (i) For $i = 0$, any factor of u of length $N_0 = n_0 + 1$ contains the prefix $0 = u_0$.
- (ii) For $i \geq 1$, any factor of $u^{(i)}$ of length $N_0^{(i)} = n_i + 1$ contains the prefix $0 = u_0^{(i)}$ of $u^{(i)}$. Thus, any factor of $u^{(i-1)}$ of length

$$N_1^{(i-1)} = (m_{i-1} + n_{i-1})(N_0^{(i)} + 1)$$

contains $\sigma_{i-1}(0) = u_1^{(i-1)}$.

By regressive induction on j , suppose that for some $j \leq i - 1$, there exists $N_{i-j}^{(j)}$ such that any factor of $u^{(j)}$ of length $N_{i-j}^{(j)}$ contains the word $u_{i-j}^{(j)}$. Then, any factor of $u^{(j-1)}$ of length

$$N_{i-j+1}^{(j-1)} = (m_{j-1} + n_{j-1})(N_{i-j}^{(j)} + 1)$$

contains $\sigma_{j-1}(u_{i-j}^{(j)}) = u_{i-j+1}^{(j-1)}$.

So, for $0 \leq j \leq i - 1$, there exists $N_{i-j}^{(j)}$ such that any factor of $u^{(j)}$ of length $N_{i-j}^{(j)}$ contains the word $u_{i-j}^{(j)}$.

Consequently, letting $N_i = N_i^{(0)}$, it follows that any factor of $u = u^{(0)}$ of length N_i contains the word u_i . This completes the proof. □

Acknowledgement. The authors would like to thank CNRS DERCI DSCA for its support during this work.

REFERENCES

[1] V. Berthé and M. Rigo (Eds). *Combinatorics, Automata and Number Theory (Encyclopedia of Mathematics and its Applications, 135)*. Cambridge University Press, Cambridge, 2010.
 [2] M. Boshernitzan. A unique ergodicity of minimal symbolic flows with minimal block growth. *J. Anal. Math.* **44** (1985), 77–96.
 [3] J. Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc.* **4** (1997), 67–88.

- [4] V. Cyr and B. Kra. Counting generic measures for a subshift of linear growth. *J. Eur. Math. Soc. (JEMS)* **21** (2019), 355–380.
- [5] V. Cyr and B. Kra. Realizing ergodic properties in zero entropy subshifts. *Israel J. Math.* **240** (2020), 119–148.
- [6] M. Damron and J. Fickenscher. The number of ergodic measures for transitive subshifts under the regular bispecial condition. *Ergod. Th. & Dynam. Sys.* **42** (2022), 86–140.
- [7] M. Keane. Non-ergodic interval exchange transformations. *Israel J. Math.* **26** (1977), 188–196.
- [8] T. Monteil. Illumination dans les billards polygonaux et dynamique symbolique. *PhD Thesis*, Institut de Mathématiques de Luminy, Université de la Méditerranée, 2005.