RAIRO-Inf. Theor. Appl. **38** (2004) 163-185
DOI: 10.1051/ita:2004009

COMPLEXITY OF INFINITE WORDS ASSOCIATED WITH BETA-EXPANSIONS *

Christiane Frougny[1, 2], Zuzana Masáková[3] and Edita Pelantová[3]

Abstract. We study the complexity of the infinite word u_β associated with the Rényi expansion of 1 in an irrational base $\beta > 1$. When β is the golden ratio, this is the well known Fibonacci word, which is Sturmian, and of complexity $\mathcal{C}(n) = n + 1$. For β such that $d_\beta(1) = t_1 t_2 \cdots t_m$ is finite we provide a simple description of the structure of special factors of the word u_β. When $t_m = 1$ we show that $\mathcal{C}(n) = (m-1)n + 1$. In the cases when $t_1 = t_2 = \cdots = t_{m-1}$ or $t_1 > \max\{t_2, \ldots, t_{m-1}\}$ we show that the first difference of the complexity function $\mathcal{C}(n+1) - \mathcal{C}(n)$ takes value in $\{m-1, m\}$ for every n, and consequently we determine the complexity of u_β. We show that u_β is an Arnoux-Rauzy sequence if and only if $d_\beta(1) = t\,t \cdots t\,1$. On the example of $\beta = 1 + 2\cos(2\pi/7)$, solution of $X^3 = 2X^2 + X - 1$, we illustrate that the structure of special factors is more complicated for $d_\beta(1)$ infinite eventually periodic. The complexity for this word is equal to $2n + 1$.

Mathematics Subject Classification. 11A63, 11A67, 37B10, 68R15.

Keywords and phrases. Beta-expansions, complexity of infinite words.

* *The authors acknowledge the financial support of GA ČR 201/01/0130, the Barrande Project number 03900WJ, and the CNRS/STIC Action Spécifique 80 "Systèmes dynamiques et modélisation en algorithmique".*

[1] LIAFA, CNRS UMR 7089, 2 place Jussieu, 75251 Paris Cedex 05, France;
e-mail: `christiane.frougny@liafa.jussieu.fr`

[2] Université Paris 8.

[3] Department of Mathematics, FNSPE, Czech Technical University, Trojanova 13, 120 00 Praha 2, Czech Republic;
e-mail: `masakova@km1.fjfi.cvut.cz, pelantova@km1.fjfi.cvut.cz`

1. Introduction

In order to give a measure on the structure of an infinite sequence $v = (v_n)_{n\geq 0}$ on a finite alphabet $\mathcal{A}$, it is often useful to use the *complexity* function of v, which is defined as follows: $\mathcal{C}(n)$ is the number of factors of length n appearing in v. It is not difficult to show that an infinite word v is eventually periodic if and only if there exists some $n \in \mathbb{N}$ such that $\mathcal{C}(n) \leq n$. Thus the simplest aperiodic words have complexity $\mathcal{C}(n) = n + 1$ for all $n \in \mathbb{N}$. Such words are binary (as $\mathcal{C}(1) = 2$) and are called *Sturmian* words. The Fibonacci word is well known to be Sturmian, see for instance [10] Chapter 2. In the survey [1] there are many examples of sequences the complexity of which is known.

To study the complexity function, it is useful to know how to find all factors of length $n + 1$ starting with the factors of length n. Special role is played by those factors that have more than one extension, the so-called special factors [6]. If we describe the occurrences of special factors and determine the number of possible extensions for each of them, we can determine the complexity. For example, Sturmian words have for every n exactly one right and one left special factor of length n with two extensions, which implies $\mathcal{C}(n + 1) - \mathcal{C}(n) = 1$, thus $\mathcal{C}(n) = n + 1$. As a generalisation of Sturmian words one defines infinite words with complexity $(m-1)n+1$, which have exactly one right and one left special factor of each length with m extensions, the so-called *Arnoux-Rauzy sequences* of order m, see [2,3,7].

It turns out that for the description of special factors it is useful to distinguish two types, according to whether they can be extended to an arbitrarily long special factor, or not. The study of the complexity function is facilitated by the notions of infinite and maximal special factors, and total bispecial factors (Defs. 3.5, 4.1 and 5.1). A very useful tool for creating and verifying hypotheses about the complexity and other combinatorial properties of substitution invariant sequences is the computer program available online at [12].

In this paper we consider infinite words u_β that are fixed points of substitutions canonically associated with the Rényi expansion of 1 in base β, where $\beta > 1$ is a Parry number, that is to say a number such that the Rényi expansion of 1 is eventually periodic or finite. This substitution generates a tiling of the nonnegative real line with a finite number of tiles [8, 15]. The vertices are labelled by the set of nonnegative β-integers, which are real numbers having a polynomial beta-expansion. The most simple example of a Parry number is a quadratic Pisot unit (that is to say, β is the root > 1 of the polynomial $X^2 - aX - 1$, with $a \geq 1$, or $X^2 - aX + 1$, with $a \geq 3$). The infinite word associated with such a number β is Sturmian [9].

In our paper we provide results for infinite words associated with simple Parry numbers, *i.e.*, those for which the Rényi expansion of 1 is finite, $d_\beta(1) = t_1 \cdots t_m$, see Section 2 for precise definitions. We completely describe the structure of special factors of u_β. When $t_m = 1$, we show that $\mathcal{C}(n) = (m-1)n+1$. In the cases when $t_1 = t_2 = \cdots = t_{m-1}$ or $t_1 > \max\{t_2, \ldots, t_{m-1}\}$ we show that the first difference of the complexity function $\mathcal{C}(n+1) - \mathcal{C}(n)$ takes value in $\{m-1, m\}$ for every n, and consequently we determine the complexity of u_β. This computation uses the linear recurrent sequence G canonically associated with β, see [5].

As a consequence of our result, every word associated with a number β such that $d_\beta(1) = t_1 \cdots t_{m-1}1$ has complexity $(m-1)n+1$. We show that such u_β is a characteristic Arnoux-Rauzy sequence (of order m) if and only if $t_1 = t_2 = \cdots = t_{m-1}$ (Th. 7.2). Note that if $m = 3$ and $t_1 = t_2 = 1$, then β is the so-called Tribonacci number, and the associated sequence has been particularly studied from the point of view of coding of a rotation on the two-dimensional torus [2]. As a byproduct, we give the necessary and sufficient condition on a simple Parry number β, so that the set of factors of u_β is closed under reversal.

In Section 8 we consider an example of a number β for which $d_\beta(1)$ is eventually periodic, namely the cubic Pisot unit $\beta = 1 + 2\cos(2\pi/7)$, solution of $X^3 = 2X^2 + X - 1$. This number is well known in mathematical quasicrystal theory, because its associated cyclotomic ring presents a seven-fold symmetry [9]. We show that the complexity for this word is equal to $2n+1$, but it is not an Arnoux-Rauzy sequence. This example illustrates the fact that the structure of special factors is more complicated for numbers with infinite eventually periodic $d_\beta(1)$.

2. Definitions

In the following $\mathbb{N}$ will denote the set of nonnegative integers, and $\mathbb{N}_+$ the set of positive integers.

Words and substitutions

Let $\mathcal{A}$ be a finite alphabet. A concatenation of letters of $\mathcal{A}$ is called a *word*. The set $\mathcal{A}^*$ of all finite words (including the empty word ε) equipped with the operation of concatenation is a free monoid. The length of a word $w = w_0w_1 \cdots w_{n-1}$ is denoted by $|w| = n$. One considers also infinite words $v = v_0v_1v_2\cdots$, the set of infinite words on $\mathcal{A}$ is denoted by $\mathcal{A}^{\mathbb{N}}$. A word w is called a *factor* of $v \in \mathcal{A}^*$, resp. $\mathcal{A}^{\mathbb{N}}$, if there exist words $w^{(1)}$ in $\mathcal{A}^*$, $w^{(2)}$ in $\mathcal{A}^*$, resp. in $\mathcal{A}^{\mathbb{N}}$, such that $v = w^{(1)}ww^{(2)}$. The word w is called a *prefix* of v if $w^{(1)} = \varepsilon$. It is a *suffix* of v if $w^{(2)} = \varepsilon$. We denote by a^k the word obtained by concatenating k letters a, with the convention that if $k = 0$, $a^k = \varepsilon$. An infinite word v is said to be *eventually periodic* if it is of the form $v = wz^\omega$, where w and z are in $\mathcal{A}^*$ and $z^\omega = zzz\cdots$

A factor w of v is called a *left special* factor of v if there exist distinct letters a and b of $\mathcal{A}$ such that aw and bw are factors of v. We say that a and b are possible left *extensions* of w. Similarly, w is a *right special* factor of v, if wa and wb are factors of v. A word w is a *bispecial* factor of v if it is in the same time right special and left special. We say that a factor w of v has a unique left, resp. right, extension if there exists a unique letter $a \in \mathcal{A}$ such that aw, resp. wa, is a factor of v.

The *complexity* of an infinite word v is the function $\mathcal{C}:\mathbb{N}\to\mathbb{N}$ given by

$$\mathcal{C}(n) := \#\{v_i v_{i+1}\cdots v_{i+n-1} \mid i\in\mathbb{N}\}\cdot$$

It is not difficult to show that an infinite word v is eventually periodic if and only if there exists some $n\in\mathbb{N}$ such that $\mathcal{C}(n)\le n$. Thus the simplest aperiodic words have complexity $\mathcal{C}(n)=n+1$ for all $n\in\mathbb{N}$. Such words are binary (as $\mathcal{C}(1)=2$) and are called *Sturmian* words.

An infinite word v over a m letter alphabet is said to be an *Arnoux-Rauzy sequence* of order m if there is exactly one right special and one left special factor of each length and if moreover these factors have m right, resp. left, extensions. Its complexity is equal to $(m-1)n+1$.

In order to determine the complexity of an infinite word, we will use the following proposition.

Proposition 2.1. *Let v be in $\mathcal{A}^{\mathbb{N}}$. For every left special factor w of v we denote by $\mu(w)$ the number of possible left extensions, and denote by M_n the set of all left special factors of v of length n. Then*

$$\mathcal{C}(n+1)-\mathcal{C}(n)=\sum_{w\in M_n}\big(\mu(w)-1\big).$$

A *morphism* of the free monoid $\mathcal{A}^*$ is a map $\varphi:\mathcal{A}^*\to\mathcal{A}^*$ satisfying $\varphi(wz)=\varphi(w)\varphi(z)$ for all w and z in $\mathcal{A}^*$. Clearly, the morphism φ is determined by $\varphi(a)$ for all a in $\mathcal{A}$.

A morphism φ is called a *substitution*[1] if $\varphi(a)\neq\varepsilon$ for all a in $\mathcal{A}$ and if there exists at least one letter a in $\mathcal{A}$ such that $|\varphi(a)|>1$. An infinite word v is said to be a *fixed point* of the substitution φ, or invariant under the substitution φ, if

$$\varphi(v_0)\varphi(v_1)\varphi(v_2)\cdots = v_0v_1v_2\cdots \tag{1}$$

or $\varphi(v)=v$, after having naturally extended the action of φ to infinite words. Relation (1) implies that $\varphi(v_0)$ is of the form $\varphi(v_0)=v_0v'$ and $\varphi^n(v)=v$ for every $n\in\mathbb{N}$. The length of the word $\varphi^n(v_0)$ grows to infinity with n, therefore for every $n\in\mathbb{N}$ the word $\varphi^n(v_0)$ is a prefix of the fixed point v, formally

$$v=\lim_{n\to\infty}\varphi^n(v_0).$$

Beta-expansions

Let $\beta>1$ be a real number. The *Rényi expansion* in base β (also called the *β-expansion*) of a number x of the interval $[0,1]$ is obtained by the following greedy algorithm [13]:
denote by $\lfloor . \rfloor$ and by $\{.\}$ the integral part and the fractional part of a number.

[1]Note that there are other definitions of *substitution* in the literature.

Let $x_1 = \lfloor \beta x \rfloor$ and $r_1 = \{\beta x\}$. Then for $i \geq 2$, let $x_i = \lfloor \beta r_{i-1} \rfloor$ and $r_i = \{\beta r_{i-1}\}$. Then $x = \sum_{i \geq 1} x_i \beta^{-i}$.

The digits x_i are nonnegative integers less than β, so they are elements of the *canonical* alphabet $\mathcal{B}_\beta = \{0, \ldots, \lfloor \beta \rfloor\}$ if $\beta \notin \mathbb{N}$, which will be the case here. The β-expansion of x is denoted by $d_\beta(x) = (x_i)_{i \geq 1}$, which is an infinite word on the alphabet $\mathcal{B}_\beta$. When $d_\beta(x)$ ends with infinitely many zeroes, it is said to be *finite*, and the 0's are omitted.

Every number $\beta > 1$ is characterized by its Rényi expansion of 1, which we denote in this paper by $d_\beta(1) = (t_i)_{i \geq 1}$. Note that $t_1 = \lfloor \beta \rfloor$. Not every sequence of nonnegative integers is equal to $d_\beta(1)$ for some β. Parry in his paper [11] gives a necessary and sufficient condition: the sequence $(t_i)_{i \geq 1}$, $t_i \in \mathbb{N}$, is the Rényi expansion of 1 for some number β if and only if the sequence satisfies

$$t_j t_{j+1} t_{j+2} \cdots \quad \prec \quad t_1 t_2 t_3 \cdots \qquad \text{for every } j > 1,$$

where $\preceq$ is the lexicographical ordering.

A number β such that $d_\beta(1)$ is eventually periodic is called by Parry [11] a *beta-number*, we propose to call it a *Parry number*. When $d_\beta(1)$ is finite, β is said to be a *simple* Parry number. A strict subclass of Parry numbers is formed by *Pisot numbers* [4,14]. Recall that a Pisot number is an algebraic integer such that all the other roots of its minimal polynomial have modulus less than 1.

Substitution, infinite word and numeration system associated with a Parry number

Let β be a Parry number. One associates with β in a canonical way a substitution [8] and a linear numeration system [5]. There are two cases to consider.

Definition 2.2. Let β be a simple Parry number, *i.e.* $d_\beta(1) = t_1 t_2 \cdots t_m$, for $m \in \mathbb{N}_+$. The substitution $\varphi = \varphi_\beta$ associated with β is defined on the alphabet $\{0, 1, \ldots, m-1\}$ by

$$\begin{aligned} \varphi(0) &= 0^{t_1} 1 \\ \varphi(1) &= 0^{t_2} 2 \\ &\vdots \\ \varphi(m-2) &= 0^{t_{m-1}} (m-1) \\ \varphi(m-1) &= 0^{t_m}. \end{aligned} \tag{2}$$

The infinite word u_β associated with β is the fixed point $u_\beta = \lim_{n \to \infty} \varphi^n(0)$ of φ.

The *characteristic polynomial* of β is $P(X) = X^m - t_1 X^{m-1} - \cdots - t_m$. One associates with P a linear recurrent sequence of integers $G = (G_n)_{n \geq 1}$, defined by

$$\begin{aligned} G_0 = 1, \qquad G_i &= t_1 G_{i-1} + \cdots + t_i G_0 + 1, \qquad 1 \leq i \leq m-1, \\ G_{n+m} &= t_1 G_{n+m-1} + \cdots + t_m G_n, \qquad\quad n \geq 0. \end{aligned}$$

It is known that the set of β-expansions of numbers of $[0,1)$ and the G-numeration system for the integers define the same symbolic dynamical system [5], see also [9], and [10], Chapter 7 for a survey on numeration systems.

Note that coding given by substitution $\varphi = \varphi_\beta$ described in Definition 2.2 is uniquely decodable. Indeed, $\varphi(\{0,\dots,m-1\})$ is a suffix code.

Definition 2.3. Let $d_\beta(1)$ be infinite eventually periodic, in particular let $m \in \mathbb{N}_+$, $p \in \mathbb{N}_+$ be minimal such that $d_\beta(1) = t_1t_2\cdots t_m(t_{m+1}\cdots t_{m+p})^\omega$. The substitution $\varphi = \varphi_\beta$ associated with β is defined on the alphabet $\{0,1,\dots,m+p-1\}$ by

$$\begin{aligned} \varphi(0) &= 0^{t_1}1 \\ \varphi(1) &= 0^{t_2}2 \\ &\ \vdots \\ \varphi(m+p-2) &= 0^{t_{m+p-1}}(m+p-1) \\ \varphi(m+p-1) &= 0^{t_{m+p}}m. \end{aligned} \tag{3}$$

The infinite word u_β associated with β is the fixed point $u_\beta = \lim_{n\to\infty}\varphi^n(0)$ of φ.

Now $P(X) = X^{m+p} - t_1X^{m+p-1} - \cdots - t_{m+p} - X^m + t_1X^{m-1} + \cdots + t_m$ is the characteristic polynomial of β. One associates with P a linear recurrent sequence of integers $G = (G_n)_{n\geq 1}$, defined by

$$\begin{aligned} &G_0 = 1, \qquad G_i = t_1G_{i-1} + \cdots + t_iG_0 + 1, \qquad 1 \leq i \leq m+p-1, \\ &G_{n+m+p} = t_1G_{n+m+p-1} + \cdots + t_{m+p}G_n + G_{n+m} - t_1G_{n+m-1} - \cdots - t_mG_n. \end{aligned}$$

As in the finite case, the set of β-expansions of numbers of $[0,1)$ and the G-numeration system for the integers define the same symbolic dynamical system. Also, similarly as in the case of a simple Parry number, $\varphi(\{0,\dots,m+p-1\})$ is a suffix code and thus coding given by substitution φ of Definition 2.3 is uniquely decodable.

Example: the golden ratio

Take $\beta = \frac{1+\sqrt{5}}{2}$. Then $d_\beta(1) = 11$. The substitution φ associated with $\frac{1+\sqrt{5}}{2}$ is the Fibonacci substitution defined by

$$0 \mapsto 01\,, \qquad 1 \mapsto 0.$$

The infinite word u_β is the Fibonacci word

$$u_\beta = 0100101001\cdots$$

and the associated numeration system is the Fibonacci numeration system defined by

$$\begin{gathered} F_{n+2} = F_{n+1} + F_n \\ F_0 = 1,\ F_1 = 2. \end{gathered}$$

3. Infinite left special factors

In the next sections we consider a simple Parry number with $d_\beta(1) = t_1 \cdots t_m$. We first state some properties of the infinite word u_β that follow from the form of the substitution (2). Using this, we study the structure of left special factors of u_β, see [6] for more details on this subject.

Lemma 3.1. *The word $\varphi^n(0)$ ends with the letter $n \pmod m$ for every $n \in \mathbb{N}$.*

Proof. Follows directly from the definition of the substitution φ. ☐

Lemma 3.2.

(i) *For every $n < m$, we have*

$$\varphi^n(0) = \left(\varphi^{n-1}(0)\right)^{t_1} \left(\varphi^{n-2}(0)\right)^{t_2} \cdots \left(\varphi^{1}(0)\right)^{t_{n-1}} 0^{t_n} n. \tag{4}$$

(ii) *For every $n \geq m$, we have*

$$\varphi^n(0) = \left(\varphi^{n-1}(0)\right)^{t_1} \left(\varphi^{n-2}(0)\right)^{t_2} \cdots \left(\varphi^{n-m}(0)\right)^{t_m}. \tag{5}$$

Proof. Using the definition of φ and the fact that $\varphi^{n+1}(0) = \varphi(\varphi^n(0))$, the equation (4) follows. Equation (5) can be derived easily by induction on n. ☐

Corollary 3.3. *For every $n \in \mathbb{N}$, the word $\varphi^n(0)$ is a left special factor of u_β with m distinct left extensions.*

Proof. The statement follows using (5) and Lemma 3.1. ☐

Corollary 3.4. *The length of the word $\varphi^n(0)$ is $|\varphi^n(0)| = G_n$, $n \in \mathbb{N}$.*

Proof. According to (5), $|\varphi^n(0)|$ and the sequence $(G_n)_{n\geq 0}$ satisfy the same recurrence relation with the same initial values. ☐

Definition 3.5. An infinite word $v = v_0v_1v_2\cdots$ on the alphabet $\mathcal{A}$ is called an *infinite left special factor* of u, if for every $n \in \mathbb{N}$, the prefix $v_0 \cdots v_n$ is a left special factor of u.

Remark 3.6. Corollary 3.3 implies that u_β is an infinite left special factor of itself.

Lemma 3.7.

(i) *The image under φ of a left special factor with p left extensions is again a left special factor with p left extensions.*
(ii) *A left special factor with q left extensions ending in a letter $X \neq 0$ is the image of a uniquely determined left special factor with q left extensions.*

Proof.
(i) Let v be a left special factor of u_β with p left extensions, and let Y_1, Y_2, $\dots$, Y_p be pairwise distinct letters of the alphabet $\{0, 1, \dots, m-1\}$ such that the word u_β contains factors Y_1v, Y_2v, $\dots$, Y_pv. Then $\varphi(Y_1)\varphi(v)$, $\dots$, $\varphi(Y_p)\varphi(v)$ are

factors of u_β. Since under the substitution φ the images of distinct letters end with distinct letters, the word $\varphi(v)$ is a left special factor of u_β with $\tilde{p}$ extensions, where $\tilde{p} \geq p$. Equality $\tilde{p} = p$ follows from the second statement.

(ii) Let $v = v_0 v_1 \cdots v_k X$, $X \neq 0$ be a left special factor of u_β with q left extensions, and let $X_1 v$, $X_2 v$, $\ldots$, $X_q v$ be factors of u_β for pairwise distinct letters X_1, X_2, $\ldots$, X_q. Find $n \in \mathbb{N}$ such that all the above factors appear in the word $\varphi^n(0)$. Therefore we can find factors f_1, f_2, $\ldots$, f_q in the word $\varphi^{n-1}(0)$ such that $X_i v$ is a factor of $\varphi(f_i)$. We choose the factors f_i of $\varphi^{n-1}(0)$ so that they have minimal length. At least one of the letter X_i, say X_1, is distinct from 0. Since $\varphi(f_1)$ contains $X_1 v_0 v_1 \cdots v_k X$, using the minimality of f_1 and the form of the substitution φ, it follows that $f_1 = Y_1 w_0 w_1 \cdots w_i$ where $\varphi(w_0 w_1 \cdots w_i) = v_0 v_1 \cdots v_k X$. Hence also $f_2, \ldots, f_q$, have the form $Y_2 w_0 w_1 \cdots w_i, \ldots, Y_q w_0 w_1 \cdots w_i$, where Y_1, Y_2, $\ldots$, Y_q are pairwise distinct letters. This means that the preimage $w_0 w_1 \cdots w_i$ of the word $v_0 v_1 \cdots v_k X$ is a left special factor of u_β with $\tilde{q}$ left extensions where $\tilde{q} \geq q$. Combining the proofs of (i) and (ii) we obtain $\tilde{p} = p$ and $\tilde{q} = q$. □

Proposition 3.8. *Let $v = v_0 v_1 v_2 \cdots$ be an infinite left special factor of u_β. Then there exists an infinite left special factor w of u_β satisfying $\varphi(w) = v$.*

Proof. The form of the substitution φ ensures that v contains infinitely many letters different from 0. According to Lemma 3.7 every finite prefix $v_0 v_1 \cdots v_k X$ of v, $X \neq 0$, is the image of a uniquely determined left special factor $w_0 w_1 \cdots w_\ell$ of u_β. This proves the proposition. □

Theorem 3.9. *The infinite word u_β has a unique infinite left special factor, namely u_β itself, each prefix of which is a left special factor with m left extensions.*

Proof. Assume that there are at least two distinct infinite left special factors of u_β. Among all the pairs of infinite left special factors of u_β we choose $v^{(1)}$, $v^{(2)}$ such that $d(v^{(1)}, v^{(2)}) := \min\{k \mid v_k^{(1)} \neq v_k^{(2)}\}$ is minimal. According to the above proposition, there exist infinite left special factors $w^{(1)}$, $w^{(2)}$ of u_β satisfying $\varphi(w^{(1)}) = v^{(1)}$ and $\varphi(w^{(2)}) = v^{(2)}$. But necessarily $d(w^{(1)}, w^{(2)}) < d(v^{(1)}, v^{(2)})$ which contradicts the minimality of $d(v^{(1)}, v^{(2)})$. Moreover, Corollary 3.3 implies that every prefix of u_β is a left special factor with m left extensions. □

4. Maximal left special factors

The aim of this section is to study those left special factors that are not prefixes of any infinite left special factor.

Definition 4.1. A left special factor $v = v_0 v_1 \cdots v_k$ of the infinite word u_β is called a *maximal left special factor* of u_β if $v_0 v_1 \cdots v_k X$ is not a left special factor of u_β for any $X \in \{0, 1, \ldots, m-1\}$.

Observation 4.2. A left special factor of u_β which has a uniquely determined right extension is not a maximal left special factor.

The following notation will be used in the sequel.

Definition 4.3. Let $d_\beta(1) = t_1 t_2 \cdots t_m$. We denote

$$j_k := \min\{i \in \mathbb{N}_+ \mid i \leq k-1,\, t_{k-i} \neq 0\}\,, \qquad \text{for } 2 \leq k \leq m.$$

Such a j_k exists because $t_1 > 0$.

Lemma 4.4.

(i) *If $\varphi^n(0)$ ends with the letter $k \geq 2$, then it has the suffix $j_k 0^{t_k} k$.*

(ii) *If $\varphi^n(0)$ ends with the letter 0, then it has the suffix $j_m 0^{t_m}$.*

Proof.

(i) If $\varphi^n(0)$ ends with the letter $k \geq 2$, then Lemma 3.1 implies that $\varphi^{n-j_k-1}(0)$ ends with the letter $k - j_k - 1$ and $\varphi^{n-j_k}(0) = \varphi(\varphi^{n-j_k-1}(0))$ ends with $0^{t_{k-j_k}}(k-j_k)$, where $t_{k-j_k} \neq 0$. From the definition of j_k, the following $t_{k-j_k+1}, \ldots, t_{k-1}$ are equal to 0. Therefore
$\varphi^{n-j_k+1}(0)$ ends with $1(k-j_k+1)$,
$\varphi^{n-j_k+2}(0)$ ends with $2(k-j_k+2)$,
...
$\varphi^{n-1}(0)$ ends with $(j_k-1)(k-1)$,
$\varphi^n(0)$ ends with $j_k 0^{t_k} k$.

(ii) The proof for $\varphi^n(0)$ ending with 0 is similar. Lemma 3.1 implies that $\varphi^{n-j_m-1}(0)$ ends with the letter $m-j_m-1$ and $\varphi^{n-j_m}(0) = \varphi(\varphi^{n-j_m-1}(0))$ ends with $0^{t_{m-j_m}}(m-j_m)$, where $t_{m-j_m} \neq 0$. The following $t_{m-j_m+1}, \ldots, t_{m-1}$ are equal to 0. Therefore
$\varphi^{n-j_m+1}(0)$ ends with $1(m-j_m+1)$,
$\varphi^{n-j_m+2}(0)$ ends with $2(m-j_m+2)$,
...
$\varphi^{n-1}(0)$ ends with $(j_m-1)(m-1)$,
$\varphi^n(0)$ ends with $j_m 0^{t_m}$. ☐

Lemma 4.5. *All factors of u_β of the form $X0^rY$, where X and Y are letters $\neq 0$, and $r \in \mathbb{N}$, are the following*

$$\begin{aligned} j_k 0^{t_k} k, &\quad \text{for } k = 2, 3, \ldots, m-1,\\ k\, 0^{t_1} 1, &\quad \text{for } k = 1, 2, \ldots, m-1,\\ j_m 0^{t_1+t_m} 1. & \end{aligned}$$

Proof. It is necessary to show that the factors of the form $X0^rY$, where $X, Y \neq 0$ and $r \in \mathbb{N}$, contained in the word $\varphi^n(0)$ are of the form given above. For $n < m$ it can be verified directly from (4). For $n \geq m$, the statement can be proven by induction on n using (5) and Lemma 4.4. ☐

Corollary 4.6. *Let v be a left special factor of u_β of the form $v = \tilde{v}0^s$, $\tilde{v}$ not ending with a 0. If $s \notin \{t_1, \ldots, t_{m-1}\}$, then v has a unique right extension.*

Lemma 4.7.

(i) *The word* 0^r, *for* $1 \le r \le t_1$, *is a left special factor of* u_β *with* m *left extensions.*

(ii) *The word* 0^r, *for* $t_1 < r \le t_1 + t_m - 1$, *is a left special factor of* u_β *with* 2 *left extensions, namely* 0 *and* j_m.

(iii) *The word* $0^{t_1+t_m-1}$ *is a maximal left special factor of* u_β *if* $t_m \ge 2$.

(iv) *If* $t_m = 1$ *then* u_β *does not have a maximal left special factor of the form* 0^r, $r \in \mathbb{N}_+$.

Proof. Note that, as a consequence of Parry's relations, $t_1 \ge t_k$ for $2 \le k \le m$. The statement (i) follows from the fact that 0^r, $r \le t_1$ is a prefix of u_β, *i.e.*, the infinite left special factor with m left extensions. The statements (ii)–(iv) follow from Lemma 4.5. □

Lemma 4.8. *Every left special factor of* u_β, *which is not a prefix of* u_β, *has two left extensions.*

Proof. The proof follows from Lemmas 3.7 and 4.7. □

From now on we shall not study the number of possible left extensions of maximal left special factors.

Lemma 4.9. *Let* v *be a factor of* u_β *with the suffix* $(m-1)Y$. *Then* $Y = 0$.

Proof. If $Y \ne 0$, then $(m-1)Y$ is a factor of u_β of the form $X0^rY$, where $X = m-1$ and $r = 0$. Since $t_1 > 0$ and $r = 0$, Lemma 4.5 implies that $X = j_k$ for $2 \le k \le m-1$. However, by definition j_k satisfies $j_k \le k-1 \le m-2$. This contradiction completes the proof. □

Proposition 4.10. *For every maximal left special factor* $v = v_0 v_1 \cdots v_k$ *containing a letter* $v_j \ne 0$ *there exists a maximal left special factor* w *and an* $s \in \{t_1, t_2, \ldots, t_{m-1}\}$ *such that* $v = \varphi(w)0^s$.

Proof. Let $j = \max\{i \mid v_i \ne 0\}$. According to Lemma 3.7 there exists a left special factor $w = w_0 w_1 \cdots w_\ell$ such that $v_0 v_1 \cdots v_j = \varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)$ and thus

$$v = v_0 v_1 \cdots v_j 0^s = \varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)0^s\,, \qquad \text{where } s = k - j\,.$$

Since v is maximal, we can use Observation 4.2 and Corollary 4.6 to derive that $s \in \{t_1, t_2, \ldots, t_{m-1}\}$.

It remains to show that w is a maximal left special factor of u_β. Assume that w is not maximal, then according to Lemma 4.9 there exists a left special factor wX, where $X \ne m-1$ or a left special factor $w(m-1)0$. However, then (ii) of Lemma 3.7 implies that $\varphi(wX)$, resp. $\varphi(w(m-1)0)$, is also a left special factor. Note that v is a proper prefix of both of them, which is a contradiction with the maximality of v. □

Corollary 4.11. *If* $t_m = 1$ *then* u_β *does not have any maximal left special factor.*

Proof. This follows by combination of Proposition 4.10 and (iv) of Lemma 4.7. □

Proposition 4.10 and (iii) of Lemma 4.7 allow us to define the following sequence of factors of the infinite word u_β.

Theorem 4.12. *If $t_m > 1$, for any $n \geq 1$ there exists an integer sequence $(s_n)_{n\geq 2}$ such that*

$$\begin{aligned} U^{(1)} &= 0^{t_1+t_m-1} \\ U^{(n)} &= \varphi(U^{(n-1)})0^{s_n} \quad \text{for } n \in \mathbb{N},\ n \geq 2, \end{aligned}$$

are maximal left special factors of u_β. Conversely, for every maximal left special factor v of u_β there exists $n \in \mathbb{N}_+$ such that $v = U^{(n)}$.

To describe explicitly the sequence $\left(U^{(n)}\right)_{n\geq 1}$ means to describe explicitly the sequence $(s_n)_{n=2}^{\infty}$. This depends on the form of the Rényi expansion of 1, $d_\beta(1) = t_1t_2\cdots t_m$. Determination of s_n in general seems to be complicated. We provide the description for two classes of numbers β in the following remark and in Proposition 4.14.

Remark 4.13. If $t_1 = t_2 = \cdots = t_{m-1} =: t$, then $(s_n)_{n=2}^{\infty}$ is the constant sequence $s_n = t$, $n \geq 2$. Note that such a β is a Pisot number.

The greedy algorithm implies $t_1 \geq \max\{t_2, t_3, \ldots, t_n\}$. In case the inequality is strict, the sequence $(s_n)_{n=1}^{\infty}$ and thus also $\left(U^{(n)}\right)_{n=1}^{\infty}$ is determined by the following proposition.

Proposition 4.14. *Let $t_1 > \max\{t_2, \ldots, t_{m-1}\}$. Then for every $n \geq 2$ we have*

$$U^{(n)} = \varphi(U^{(n-1)})0^{t_i}, \qquad \textit{where } i \in \{1, 2, \ldots, m-1\},\ i = n\,(\bmod\ (m-1))\,. \quad (6)$$

Proof. We show by induction on n a stronger statement: for every $n \in \mathbb{N}_+$, $U^{(n)}$ has the form (6) and the right extensions of $U^{(n)}$ in u_β are 0 or i, more precisely, if $U^{(n)}X$ is a factor of u_β, then $X \in \{0, i\}$.

For $n = 1$ we have $U^{(1)} = 0^{t_1+t_m-1}$ and by Lemma 4.5 only $U^{(1)}0$ and $U^{(1)}1$ are factors of u_β of length $|U^{(1)}| + 1$ with prefix $U^{(1)}$. Assume that $U^{(n)} = \varphi(U^{(n-1)})0^{t_i}$, for $i \in \{1, 2, \ldots, m-1\}$, $i = n\,(\bmod\ (m-1))$, and that $U^{(n)}0$, $U^{(n)}i$ are the only factors of u_β of length $|U^{(n)}| + 1$ with the prefix $U^{(n)}$. Let us distinguish two cases:

(a) suppose that $i < m-1$. Then $\varphi(U^{(n)})0^{t_1}1$ and $\varphi(U^{(n)})0^{t_{i+1}}(i+1)$ are factors of u_β. Therefore $U^{(n+1)} = \varphi(U^{(n)})0^{t_{i+1}}$, where $i \in \{1, 2, \ldots, m-1\}$ and $i+1 = n+1\,(\bmod\ (m-1))$. Moreover, the possible right extensions of $U^{(n+1)}$ are 0 and $(i+1)$;

(b) suppose now $i = m-1$. Lemma 4.9 says that the right extension of the letter $m-1$ is always the letter 0. Thus $U^{(n)}(m-1)0$ and $U^{(n)}1$ are factors of u_β and hence also their images under φ, namely $\varphi(U^{(n)})0^{t_1+t_m}1$ and $\varphi(U^{(n)})0^{t_1}1$ are factors of u_β. Therefore $U^{(n+1)} = \varphi(U^{(n)})0^{t_1}$. Since $i = m-1 = n\,(\bmod\ (m-1))$, we have $1 = n+1\,(\bmod\ (m-1))$. Moreover, the possible right extensions of $U^{(n+1)}$ are 1 and 0. $\square$

Example 4.15. We shall now illustrate the notion of infinite left special factors and maximal left special factors on the example of u_β, where $\beta = 1 + \sqrt{3}$ is the positive solution of $X^2 = 2X + 2$, *i.e.*, $d_\beta(1) = 22$. The word

$$u_\beta = 0010010000100100001001001001001\cdots$$

is the fixed point of the substitution

$$0 \mapsto 001, \qquad 1 \mapsto 00.$$

The structure of left special factors in the word u_β can be illustrated on a tree. On the following figure every left special factor is represented as a sequence of letters along a path in the tree starting at the root ε.

```
                                                                 /1-0-0-1-0-0-0-0-1-0-0-1-0-0-···
                           /0-0-1-0-0-1-0-0-0-0-1-0-0-1-0-0<
          /1-0-0-1-0-0<                                      \0-0-1-0-0-1-0-0
ε-0-0<                 \1-0-0
          \0
```

FIGURE 1. Tree representation of the structure of left special factors in u_β for $d_\beta(1) = 22$.

The sequence of letters in the upper infinite path forms the infinite word u_β, *i.e.*, the infinite left special factor. The sequence of letters along the path from the root ε to every leaf of the tree is a maximal left special factor of u_β. For example, the figure shows the three shortest maximal left special factors

$$\begin{aligned} U^{(1)} &= 000, \\ U^{(2)} &= 00100100100 = \varphi(U^{(1)})00, \\ U^{(3)} &= 00100100001001000010010000100100 = \varphi(U^{(2)})00. \end{aligned}$$

Note that here we have $U^{(n)} = \varphi(U^{(n-1)})00$ for every $n \geq 2$ since by Remark 4.13 there is $s_n = t_1 = 2$.

From the above figure we can see that the number of left special factors of length n either is 1 or is 2. The dependence of this value on n will be studied in the next section for general β.

5. Total bispecial factors

Every left special factor of u_β is either a prefix of the infinite left special word u_β, or is a prefix of some maximal left special word $U^{(n)}$. In order to determine the complexity of u_β, we need to study the common prefixes of u_β and $U^{(n)}$. For a fixed n let $V^{(n)}$ be the maximal common prefix of u_β and $U^{(n)}$. Since $V^{(n)}$ has two right extensions and $U^{(n)}$ is a maximal left special factor, we have $V^{(n)} \neq U^{(n)}$. Therefore both $V^{(n)}X$ and $V^{(n)}Y$, $X \neq Y$ are left special factors of u_β, *i.e.*, $V^{(n)}$ is a bispecial factor of u_β.

Definition 5.1. A factor w is called a *total bispecial factor* of u_β, if there exist distinct letters $X, Y \in \{0, 1, \ldots, m-1\}$ such that both wX and wY are left special factors of u_β.

Proposition 5.2. *Let $v_0 v_1 \cdots v_k$ be a total bispecial factor of u_β such that $v_i \neq 0$ for some i, $0 \leq i \leq k$. Then there exists a total bispecial factor $w_0 w_1 \cdots w_\ell$ and an $s \in \{t_1, \ldots, t_{m-1}\}$ such that*

$$v_0 v_1 \cdots v_k = \varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)0^s.$$

Proof. Let $j = \max\{i \mid v_i \neq 0\}$. According to Lemma 3.7 there exists a left special factor $w = w_0 w_1 \cdots w_\ell$ such that $v_0 v_1 \cdots v_j = \varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)$ and thus

$$v = v_0 v_1 \cdots v_j 0^s = \varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)0^s, \qquad \text{where } s = k - j.$$

If $s \notin \{t_1, t_2, \ldots, t_{m-1}\}$, then according to Corollary 4.6, the word v has a unique right extension, which contradicts the fact that v is a bispecial factor.

It remains to show that w is a total bispecial factor. Since v is a total bispecial factor, there exist letters X, X_1, X_2 and Y, Y_1, Y_2 such that $X \neq Y$, $X_1 \neq X_2$ and $Y_1 \neq Y_2$ and that

$$\begin{array}{c} X_1\varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)0^s X \\ X_2\varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)0^s X \\ Y_1\varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)0^s Y \\ Y_2\varphi(w_0)\varphi(w_1)\cdots\varphi(w_\ell)0^s Y \end{array}$$

are factors of u_β. The properties of the substitution φ imply that u_β contains factors $\tilde{X}_1 w \tilde{X}$, $\tilde{X}_2 w \tilde{X}$, $\tilde{Y}_1 w \tilde{Y}$, and $\tilde{Y}_2 w \tilde{Y}$, for some $\tilde{X} \neq \tilde{Y}$, $\tilde{X}_1 \neq \tilde{X}_2$, and $\tilde{Y}_1 \neq \tilde{Y}_2$. Hence w is also a total bispecial factor of u_β. □

Let us recall that we have denoted by $V^{(n)}$ the maximal common prefix of u_β and $U^{(n)}$. Clearly $V^{(1)} = 0^{t_1}$. As a consequence of the above proposition we have the following statement.

Corollary 5.3. *There exists a sequence $(s_n)_{n\geq 2}$ such that*

$$\begin{aligned} V^{(1)} &= 0^{t_1} \\ V^{(n)} &= \varphi(V^{(n-1)})0^{s_n} \quad \textit{for } n \in \mathbb{N},\ n \geq 2. \end{aligned}$$

Remark 5.4. If $t_1 = t_2 = \cdots t_{m-1} = t$, then $(s_n)_{n\geq 2}$ is the constant sequence $s_n = t$ for $n \geq 2$.

Determination of the sequence $(s_n)_{n\geq 2}$ for the sequence of words $\left(V^{(n)}\right)_{n\geq 1}$ is not simple in general. However, if the assumption of Proposition 4.14 is satisfied, sequences $(s_n)_{n\geq 2}$ for generating $\left(U^{(n)}\right)_{n\geq 1}$ and $\left(V^{(n)}\right)_{n\geq 1}$ coincide.

Proposition 5.5. *Let $t_1 > \max\{t_2, \dots, t_{m-1}\}$. Then for every $n \geq 2$ we have*

$$V^{(n)} = \varphi(V^{(n-1)})0^{t_i}, \qquad \text{where } i \in \{1, 2, \dots, m-1\},\ i = n \,(\bmod\ (m-1))\,. \quad (7)$$

Proof. We show by induction on n the following statement: the word $V^{(n)}$ has the form (7) and if $V^{(n)}X$ is a factor of u_β, then $X \in \{0, i\}$, where $i \in \{1, 2, \dots, m-1\}$ and $i = n \,(\bmod\ (m-1))$.

For $n = 1$ we have $V_1 = 0^{t_1}$ and the only $V^{(1)}0$ and $V^{(1)}1$ are factors of u_β of length $|V^{(1)}| + 1$ with prefix $V^{(1)}$. The induction step is analogous to the proof of Proposition 4.14. □

Proposition 5.6. *Let $t_1 > \max\{t_2, \dots, t_{m-1}\}$ or $t_1 = t_2 = \dots = t_{m-1}$. Then for $n \geq 1$*

(1) $|U^{(n)}| = (t_m - 1)G_{n-1} + |V^{(n)}|$;

(2) $|V^{(n)}| = \sum_{i=1}^{n} c_i G_{n-i}$, *where $c_i = t_k$, where $k \in \{1, 2, \dots, m-1\}$ such that $i = k \,(\bmod\ (m-1))$;*

(3) $|U^{(n-1)}| < |V^{(n)}|$.

Proof. 1. From the construction of $|U^{(n)}|$ and $|V^{(n)}|$ and from the fact that $U^{(1)} = 0^{t_m - 1}V^{(1)}$ it follows that $U^{(n)} = \left(\varphi^{n-1}(0)\right)^{t_m - 1} V^{(n)}$. Since $|\varphi^n(0)| = G_n$ for every $n \in \mathbb{N}$, we obtain the assertion.

2. Using the relation $V^{(n)} = \varphi(V^{(n-1)})0^{t_i}$ we can easily prove by induction on n that

$$V^{(n)} = \left(\varphi^{n-1}(0)\right)^{c_1} \left(\varphi^{n-2}(0)\right)^{c_2} \cdots \left(\varphi(0)\right)^{c_{n-1}} 0^{c_n}\,,$$

where the c_i's are defined in the statement of the proposition. Therefore $|V^{(n)}| = \sum_{i=1}^{n} c_i G_{n-i}$.

3. We have to verify

$$(t_m - 1)G_{n-2} + \sum_{i=1}^{n-1} c_i G_{n-1-i} < \sum_{i=1}^{n} c_i G_{n-i}. \quad (8)$$

This is equivalent to verifying

$$(t_m - 1)G_{n-2} < \sum_{i=1}^{n-1} c_i (G_{n-i} - G_{n-1-i}) + c_n.$$

Note that $(t_m - 1)G_{n-2}$ is smaller than the first term of the sum on the right hand side of the inequality above, *i.e.*, $(t_m - 1)G_{n-2} < t_1(G_{n-1} - G_{n-2})$. This comes from the recurrence relation defining the sequence $(G_n)_{n \geq 0}$ and from the fact that $t_1 \geq t_m$. Since also $G_{n-i} > G_{n-i-1}$ the validity of the considered inequality is obvious. □

Example 5.7. Let us illustrate the notion of total bispecial factors on the case of β with $d_\beta(1) = 22$. In this case $s_n = t_1 = 2$ for $n \geq 2$ and therefore $V^{(n)} = \varphi(V^{(n-1)})00$ for $n \geq 2$. In the tree of left special factors, a total bispecial factor is represented by a path from the root ε, ending at a vertex with two sons. In Figure 1 we can observe

$$\begin{aligned} V^{(1)} &= 00\,, \\ V^{(2)} &= 00100100 = \varphi(V^{(1)})00\,, \\ V^{(3)} &= 001001000010010000100100 = \varphi(V^{(2)})00\,. \end{aligned}$$

6. Complexity for a simple Parry number

Every left special factor w of u_β is a prefix either of the infinite left special factor u_β or of a maximal left special factor $U^{(k)}$. Moreover, if the length of w satisfies $\left|U^{(k-1)}\right| < |w| \le \left|V^{(k)}\right|$ for some k, then necessarily w is a prefix of u_β. Therefore for n such that $\left|V^{(k)}\right| < n \le \left|U^{(k)}\right|$ there exist two left special factors of length n, one being a prefix of u_β and thus having m left extensions, the other being a prefix of $U^{(k)}$, and thus having 2 left extensions. It is clear that the values $\left|V^{(k)}\right|$, $\left|U^{(k)}\right|$ play an important role in determining the complexity of the infinite word u_β.

Definition 6.1. For β such that $d_\beta(1) = t_1t_2\cdots t_m$ we denote

$$\ell_0 := 0, \qquad \ell_r := |V^{(r)}| = \sum_{i=1}^{r} c_i G_{r-i}, \quad r > 0$$

where $c_i = t_k$ for $k \in \{1,2,\dots,m-1\}$ such that $i = k \,(\mathrm{mod}\ (m-1))$.

We are now in position to state the main theorem of the paper. For technical reasons we have to set $G_{-1} = 0$.

Theorem 6.2. *Let $\beta > 1$ have finite Rényi expansion of 1, $d_\beta(1) = t_1t_2\cdots t_m$.*

(1) *Suppose that $t_m = 1$. Then for every $n \in \mathbb{N}_+$ we have*

$$\mathcal{C}(n+1) - \mathcal{C}(n) = m - 1.$$

(2) *Suppose now that $t_1 > \max\{t_2,\dots,t_{m-1}\}$ or $t_1 = t_2 = \cdots = t_{m-1}$. Then for every $n \in \mathbb{N}_+$ there exists a $k \in \mathbb{N}$ such that*

$$\ell_k < n \le \ell_{k+1}$$

and

$$\mathcal{C}(n+1) - \mathcal{C}(n) = \begin{cases} m & \text{if} \quad \ell_k < n \le \ell_k + (t_m - 1)G_{k-1}, \\ m-1 & \text{if} \quad \ell_k + (t_m-1)G_{k-1} < n \le \ell_{k+1}. \end{cases}$$

Proof. Statement 1 follows directly from Theorem 3.9 and Corollary 4.11. Statement 2 is a consequence of Theorem 3.9, Propositions 4.14, 5.5 and 5.6.

For Statement 2 realize that the increase of complexity is of two types, one is due to infinite left special factors, the other to left special factors that cannot be extended to infinite left special factors. Theorem 3.9 says that there is exactly one infinite left special factor with m left extensions for every n, therefore

$$\mathcal{C}(n+1) - \mathcal{C}(n) = m - 1 + B_n, \tag{9}$$

where $B_n \geq 0$ for $n \in \mathbb{N}$. The sequence B_n determines the number of left special factors of length n that are not prefixes of any infinite left special factor. According to Theorem 4.12 and Proposition 5.6 such left special factors exist only for $|V^{(k)}| < n \leq |U^{(k)}|$ for some k. Moreover, for every n in the above interval, there is exactly one such left special factor. Therefore

$$B_n = \begin{cases} 1 & \text{if} \quad \ell_k < n \leq \ell_k + (t_m - 1)G_{k-1}, \\ 0 & \text{if} \quad \ell_k + (t_m - 1)G_{k-1} < n \leq \ell_{k+1}, \end{cases} \tag{10}$$

where we use the formulas for $|V^{(k)}|$, $|U^{(k)}|$ from Proposition 5.6. The assertion of the theorem follows easily. $\square$

We can now find an explicit formula for the complexity $\mathcal{C}(n)$ of the infinite word u_β for two classes of simple Parry numbers β.

Corollary 6.3. *If $t_m = 1$, we have*

$$\mathcal{C}(n) = (m-1)n + 1.$$

Corollary 6.4. *If $t_m > 1$ and $t_1 > \max\{t_2, \ldots, t_{m-1}\}$ or $t_1 = t_2 = \cdots = t_{m-1}$. Then the complexity of the infinite word u_β satisfies*

$$(m-1)n + 1 \leq \mathcal{C}(n) \leq mn.$$

More precisely,

$$\mathcal{C}(n) = (m-1)n + 1 + (t_m - 1)\left(\sum_{j=0}^{k-2} G_j\right) + \min\bigl\{n - \ell_k - 1, (t_m - 1)G_{k-1}\bigr\}$$

where $\ell_k < n - 1 \leq \ell_{k+1}$.

Proof. According to (9) in the proof of Theorem 6.2

$$\begin{aligned}\mathcal{C}(n) &= \mathcal{C}(1) + (m-1)(n-1) + \sum_{i=1}^{n-1} B_i \\ &= m + (m-1)(n-1) + \sum_{i=1}^{n-1} B_i = (m-1)n + 1 + \sum_{i=1}^{n-1} B_i.\end{aligned}$$

Now it is enough to calculate $\sum_{i=1}^{n-1} B_i$. Let $k \in \mathbb{N}$, such that $\ell_k < n-1 \leq \ell_{k+1}$. Then

$$\sum_{i=1}^{n-1} B_i = \sum_{\substack{1 \leq i \leq n-1 \\ B_i = 1}} 1 = \sum_{j=0}^{k-1} (t_m - 1) G_{j-1} + \min\{n-1-\ell_k, (t_m-1)G_{k-1}\}\cdot$$

For calculating the above sum we have used (10). □

Example 6.5. Again consider $d_\beta(1) = 22$. The sequence G associated with β is defined by

$$G_{n+2} = 2G_{n+1} + 2G_n$$

$$G_0 = 1, \quad G_1 = 3.$$

Since $m = 2$ and $t_1 = 2$, for $r \geq 1$ we have $\ell_r = 2\sum_{i=1}^{r} G_{r-i} = 2\sum_{i=0}^{r-1} G_i$. Corollary 6.4 gives

$$\mathcal{C}(n) = n + 1 + \sum_{j=0}^{k-2} G_j + \min\left\{ n - 1 - 2\sum_{j=0}^{k-1} G_j \, , \, G_{k-1} \right\},$$

for k such that

$$2\sum_{i=0}^{k-1} G_i < n - 1 \leq 2\sum_{i=0}^{k} G_i.$$

Let us calculate the complexity $\mathcal{C}(n)$ of u_β for several small values of n,

$$\mathcal{C}(1) = 2,\ \mathcal{C}(2) = 3,\ \mathcal{C}(3) = 4,\ \mathcal{C}(4) = 6,\ \mathcal{C}(5) = 7,\ \mathcal{C}(6) = 8,\ \mathcal{C}(7) = 9,\ \mathcal{C}(8) = 10,$$
$$\mathcal{C}(9) = 11,\ \mathcal{C}(10) = 13,\ \mathcal{C}(11) = 15,\ \mathcal{C}(12) = 17,\ \mathcal{C}(13) = 18,\ \mathcal{C}(14) = 19,\ \ldots$$

Note that the first difference $\mathcal{C}(n+1) - \mathcal{C}(n) = 2$ for $n = 3, 9, 10, 11$. These values correspond to those levels of the tree from Figure 1 which have two vertices. In general, the first difference of the complexity $\mathcal{C}(n+1) - \mathcal{C}(n)$ is equal to the number of vertices on the n-th level in the tree of left special factors.

In order to illustrate the behaviour of the complexity function let us cite the results of [16]. The author determines $\limsup \frac{\mathcal{C}(n)}{n}$ and $\liminf \frac{\mathcal{C}(n)}{n}$ for the case $d_\beta(1) = 22$,

$$\limsup_{n\to\infty} \frac{\mathcal{C}(n)}{n} = 1 + \frac{\beta+2}{9} \simeq 1.38,$$
$$\liminf_{n\to\infty} \frac{\mathcal{C}(n)}{n} = 1 + \frac{\beta-2}{4} \simeq 1.18.$$

7. Arnoux-Rauzy sequences

In this section we describe all simple Parry numbers β, for which the associated word u_β is an Arnoux-Rauzy sequence. For it we use an auxiliary result, which is however of independent interest. In the following proposition we provide necessary and sufficient conditions under which the set of factors of u_β is closed under reversal.

Proposition 7.1. *Let $d_\beta(1) = t_1 \cdots t_{m-1} t_m$. Then the set of factors of u_β is closed under reversal if and only if $t_1 = t_2 = \cdots = t_{m-1}$.*

Proof. First we show that $t_1 = t_2 = \cdots = t_{m-1} = t$ is a necessary condition. Assume that it is not true. Let k be the smallest integer in $\{2, 3, \ldots, m-1\}$, such that $t_k < t_1$. Then $j_k = 1$ (see Def. 4.3) and according to Lemma 4.5, u_β has the factor $v = 10^{t_k} k$. All factors of u_β which start with k have the prefix $k\,0^{t_1}$. Thus v does not have its reversal in u_β.

Suppose now that $t_1 = t_2 = \cdots = t_{m-1} = t$. We show that in this case u_β is a limit of palindromes, which implies that the set of factors of u_β is closed under reversal. Substitution φ is given by

$$0 \mapsto 0^t 1 \qquad 1 \mapsto 0^t 2 \qquad \ldots \qquad m-2 \mapsto 0^t(m-1) \qquad m-1 \mapsto 0^{t_m}.$$

As a simple consequence, if $w = w_0 w_1 \cdots w_\ell$ is a palindrome with $w_\ell = 0$, then $\varphi(w)0^t$ is also a palindrome. Moreover, if w is a factor of u_β, then $\varphi(w)$ is also a factor of u_β, which ends with 1. Since every non-zero letter in u_β is followed by 0^t, also $\varphi(w)0^t$ is a factor of u_β. We can thus define a sequence of palindromes in u_β by

$$W^{(0)} = 0^t, \qquad W^{(n)} = \varphi(W^{(n-1)})0^t.$$

Since $W^{(0)}$ is a prefix of u_β, also $W^{(n)}$ is a prefix of u_β for every n. Thus

$$u_\beta = \lim_{n \to \infty} W^{(n)},$$

which completes the proof. □

Let us now determine which simple Parry numbers give Arnoux-Rauzy sequences. If u_β is an Arnoux-Rauzy sequence of order m, then $d_\beta(1) = t_1 t_2 \cdots t_m$. From the definition, an Arnoux-Rauzy sequence does not have maximal special factors. Therefore as a consequence of Theorem 4.12, $t_m = 1$. In that case u_β has the desired complexity $\mathcal{C}(n) = (m-1)n + 1$.

Theorem 7.2. *Let $d_\beta(1) = t_1 \cdots t_{m-1} t_m$. The infinite word u_β is an Arnoux-Rauzy sequence (of order m) if and only if $t_1 = \cdots = t_{m-1}$ and $t_m = 1$.*

Proof. First, let i be the smallest index, $2 \le i \le m-1$, such that $t_1 > t_i$. Then the factor 0^{t_i+1} is right special, because it has two right extensions, namely $0^{t_i+1}0$ and $0^{t_i+1}1$. On the other hand the factor 10^{t_i} has two right extensions, which are $10^{t_i}i$ and $10^{t_i}0$, so it is right special. Thus there are two right special factors of length $t_i + 1$, hence the sequence is not Arnoux-Rauzy.

Suppose now that $t_1 = t_2 = \cdots = t_{m-1} = t$ and $t_m = 1$. By Theorem 3.9 and Corollary 4.11 for each $n \geq 1$ there is a unique left special factor of length n and this factor has m left extensions. Since the set of factors of u_β is closed under reversal, for every n there exists a unique right special factor of length n with m right extensions. Thus u_β is Arnoux-Rauzy. $\square$

Note that if $d_\beta(1) = t\,t \cdots t\,t_m$, then β is a Pisot number. It is also interesting to mention that if $t_m = 1$ then the unique left special factor of each length is a prefix of u_β. An Arnoux-Rauzy sequence satisfying this property is called *characteristic*, see [3].

8. Complexity for $d_\beta(1) = 2(01)^\omega$

According to (3) the infinite word u_β for β satisfying $d_\beta(1) = 2(01)^\omega$ is the fixed point of the substitution ψ defined by

$$\begin{aligned} 0 &\mapsto 001 \\ 1 &\mapsto 2 \\ 2 &\mapsto 01 \end{aligned}$$

so that $u_\beta = \lim_{n\to\infty} \psi^n(0)$. This β is the root > 1 of the polynomial $X^3 - 2X^2 - X + 1$. It is a totally real cubic Pisot number and it is equal to $1 + 2\cos(2\pi/7)$. It can be shown that its associated cyclotomic ring presents a sevenfold symmetry [9].

In this section we determine the complexity of u_β and in the same time we illustrate the obstacles that may appear if we want to do the same for general β with eventually periodic Rényi expansion of 1. For this we shall use the following notation.

Definition 8.1. Let $v = v_0 v_1 v_2 \cdots v_k$ be a word on the alphabet $\mathcal{A}$ and let $w = v_0 v_1 \cdots v_i$, for $i \leq k$. We write

$$w^{-1}v := v_{i+1} v_{i+2} \cdots v_k.$$

Lemma 8.2. *For $n \geq 3$ we have*

$$\psi^n(0) = \psi^{n-1}(0)\ \psi^{n-1}(0)\ \big(\psi^{n-3}(0)\big)^{-1}\ \psi^{n-2}(0).$$

Proof. Since for $n \geq 3$, the word $\psi^{n-3}(0)$ is always a prefix of $\psi^{n-2}(0)$, the expression $\big(\psi^{n-3}(0)\big)^{-1}\psi^{n-2}(0)$ is well defined. Since $\psi^0(0) = 0$, $\psi^1(0) = 001$, $\psi^2(0) = 0010012$ and

$$\psi^3(0) = \underbrace{0010012}_{\psi^2(0)}\quad \underbrace{0010012}_{\psi^2(0)}\ \underbrace{01}_{0^{-1}\psi^1(0)} \tag{11}$$

the statement is true for $n = 3$. For $n > 3$ the statement follows easily by induction, taking into account that $\psi(w^{-1}v) = \big(\psi(w)\big)^{-1}\psi(v)$. $\square$

Lemma 8.3. *For $n \geq 3$ we have*

$$\psi^n(0) = \psi^{n-1}(0)\ \psi^{n-2}(0)\ \cdots \psi^4(0)\ \psi^3(0)\ \psi^2(0)\ \psi^2(0)\ 0^{-1}\ \psi^{n-2}(0).$$

Proof. The relation (11) implies that the statement is true for $n = 3$. For $n > 3$ we obtain by induction

$$\psi^{n+1}(0) = \psi^n(0)\ \psi^{n-1}(0)\ \cdots \psi^4(0)\ \psi^3(0)\ \underbrace{\psi^3(0)\ \big(\psi(0)\big)^{-1}\ \psi^{n-1}(0)}_{w}.$$

The word w is according to (11) equal to

$$\begin{aligned} w &= \psi^3(0)\ \big(\psi(0)\big)^{-1}\ \psi^{n-1}(0) = \psi^2(0)\ \psi^2(0)\ 0^{-1}\ \psi(0)\ \big(\psi(0)\big)^{-1}\ \psi^{n-1}(0) \\ &= \psi^2(0)\ \psi^2(0)\ 0^{-1}\ \psi^{n-1}(0)\,, \end{aligned}$$

which completes the proof. □

Lemma 8.4. *The words $\psi^n(0)$ and $0^{-1}\psi^n(0)$ are left special factors of u_β for all $n \in \mathbb{N}$, $n \geq 3$.*

Proof. Let us denote $\psi^n(0) = A$. Since $\psi^n(0)$ is a prefix of both $\psi^{2n}(0)$ and $\psi^{2n-1}(0)$, we have $\psi^{2n}(0) = Aw_1$, $\psi^{2n-1}(0) = Aw_2$ for some words w_1, w_2. Using Lemma 8.2 we have

$$\begin{aligned} \psi^{2n}(0) &= \psi^{2n-1}(0)\, Aw_2\, \big(\psi^{2n-3}(0)\big)^{-1}\ \psi^{2n-2}(0), \\ \psi^{2n+1}(0) &= \psi^{2n}(0)\ Aw_1\, \big(\psi^{2n-2}(0)\big)^{-1}\ \psi^{2n-1}(0). \end{aligned}$$

Since $\psi^{2n-1}(0)$ ends with 1 and $\psi^{2n}(0)$ ends with 2, both $1A$ and $2A$ are factors of u_β. Thus $A = \psi^n(0)$ is a left special factor.

According to Lemma 8.3, $\psi^2(0)0^{-1}\psi^n(0)$ is a factor of $\psi^{n+2}(0)$ and thus a factor of u_β. Since $\psi^2(0)$ ends with 2, also $20^{-1}\psi^n(0)$ is a factor of u_β. But $\psi^n(0) = 00^{-1}\psi^n(0)$ is a factor of u_β. Therefore $0^{-1}\psi^n(0)$ is a left special factor of u_β with left extensions 0 and 2. □

Let us now describe the possible left extensions of the above left special factors.

Lemma 8.5. *Every left special factor of u_β of length ≥ 2 has either the prefix* 00 *or the prefix* 01*.*

(i) *The left special factor $00v$ has two left extensions, namely* 1 *or* 2*. We denote this property by $\frac{1}{2}\!>$.*
(ii) *The left special factor $01v$ has two left extensions, namely* 0 *or* 2*. We denote this property by $\frac{0}{2}\!>$.*

Proof. First realize that the definition of ψ implies for a letter X that

$$\text{if } \ X1 \ \text{ is a factor of } u_\beta \text{ then } \ X1 = 01. \tag{12}$$

We further show that

$$\text{if } \ X2 \ \text{ is a factor of } u_\beta \text{ then } \ X2 = 12. \tag{13}$$

From the definition of the substitution ψ it follows that every factor $X2$ of u_β is a suffix of the image of a factor $Y1$ under ψ. From (12) we have $Y = 0$. Since $\psi(01) = 0012$, necessarily we have $X = 1$.

As a consequence of (12) and (13), every word with the prefix 1 or 2 has a uniquely determined left extension. Thus a left special factor of u_β must start with the letter 0. Due to (13), the right extension of 0 can only be 0 or 1. Therefore every left special factor of u_β of length ≥ 2 has either the prefix 00 or the prefix 01.

The definition of the substitution ψ implies that 000 is not a factor of u_β. Therefore the left special factor with the prefix 00 must have the two left extensions $\genfrac{}{}{0pt}{}{1}{2}\!>$.

If the word 101 were a factor of u_β, then it would be a suffix of the image of the factor 02 or 22. But according to (13) these words are not factors of u_β. Thus the prefix 01 of a left special factors admits only the left extensions $\genfrac{}{}{0pt}{}{0}{2}\!>$. □

Lemma 8.6.

(i) *Every left special factor of length* ≥ 3 *with the prefix* 00 *not ending with* 0 *is the image under* ψ *of another left special factor of the same type.*

(ii) *For every left special factor* v *of length* ≥ 3 *with the prefix* 01 *not ending with* 0 *it holds that* $(01)^{-1}v$ *is the image under* ψ *of another left special factor of the same type as* v.

Proof. (i) It is obvious that a factor $v = 00v'$ with prefix 00 not ending with 0 has a uniquely determined preimage w under ψ, *i.e.*, $\psi(w) = v$. The left extensions of v, namely $\genfrac{}{}{0pt}{}{1}{2} > 00v'$ could be the images under ψ of $\genfrac{}{}{0pt}{}{0}{1} > w$ or $\genfrac{}{}{0pt}{}{2}{1} > w$ but the combination $\genfrac{}{}{0pt}{}{0}{1}\!> w$ is not admissible.

(ii) The proof is analogous. □

Proposition 8.7. *The infinite word* u_β *has a unique infinite left special factor with left extensions* $\genfrac{}{}{0pt}{}{1}{2}\!>$*, namely* u_β *itself, and a unique infinite left special factor with left extensions* $\genfrac{}{}{0pt}{}{0}{2}\!>$*, namely* $0^{-1}u_\beta$.

Proof. According to Lemma 8.4, $\psi^n(0)$ is a left special factor for every $n \in \mathbb{N}$, moreover it has the prefix 00. Therefore using Lemma 8.5 $u_\beta = \lim_{n\to\infty} \psi^n(0)$ is an infinite left special factor with left extensions $\genfrac{}{}{0pt}{}{1}{2} >$. Similarly, $0^{-1}\psi^n(0)$ is a left special factor for all $n \in \mathbb{N}$ and it has the prefix 01. Thus $0^{-1}u_\beta = \lim_{n\to\infty} 0^{-1}\psi^n(0)$ is an infinite left special factor of u_β with left extensions $\genfrac{}{}{0pt}{}{0}{2}\!>$.

We show the uniqueness of the infinite left special factors by contradiction. Among all infinite left special factors with left extensions $\genfrac{}{}{0pt}{}{1}{2}\!>$ find a pair

$$v^{(1)} = v_0^{(1)}v_1^{(1)}v_2^{(1)}\cdots, \qquad v^{(2)} = v_0^{(2)}v_1^{(2)}v_2^{(2)}\cdots$$

such that $d(v^{(1)}, v^{(2)}) := \min\{i \mid v_i^{(1)} \neq v_i^{(2)}\}$. Then according to Lemma 8.6 there exist two left special factors $w^{(1)}$, $w^{(2)}$ with left extensions $\genfrac{}{}{0pt}{}{1}{2} >$ satisfying

$\psi(w^{(1)}) = v^{(1)}$, $\psi(w^{(2)}) = v^{(2)}$, which implies $d(w^{(1)}, w^{(2)}) < d(v^{(1)}, v^{(2)})$, which is a contradiction with the choice of $v^{(1)}$, $v^{(2)}$.

The uniqueness of the infinite left special factor with left extensions $^0_2>$ is shown in the same way. □

Proposition 8.8. *The infinite word u_β has no maximal left special factor.*

Proof. From the definition of the substitution ψ it follows that a maximal left special factor v cannot end with 00. (Otherwise it has a unique right extension, namely 1, and thus cannot be maximal.) The factor v has the form $v = \tilde{v}0^s$, where $\tilde{v}$ does not end with 0 and $s \in \{0, 1\}$. If $|\tilde{v}| \geq 3$, due to Lemma 8.6, there exists a left special factor $\tilde{w}$ such that $v = \varphi(w)0^s$ or $v = 01\varphi(w)0^s$. Maximality of v implies that w is a maximal left special factor.

The only left special factor of length < 3 are 00 and 01. However, these are prefixes of the infinite left special factors u_β and $0^{-1}u_\beta$, respectively. Hence they cannot be maximal. □

By combination of the above Propositions 8.7 and 8.8 we obtain the main result of this section.

Theorem 8.9. *Let β satisfy $d_\beta(1) = 2(01)^\omega$. The complexity of the infinite word u_β is given by*

$$\mathcal{C}(n) = 2n + 1\,, \qquad n \in \mathbb{N}.$$

Remark 8.10. The infinite word u_β in this case is not an Arnoux-Rauzy sequence, because it has more than one left special factor of each length.

9. Conclusions

It is interesting to mention that the description of the structure of left special factors given here is valid for fixed points of the substitution (2) even for parameters $t_1, \dots, t_m$ which do not correspond to the Rényi expansion $d_\beta(1)$ of 1 for some number β. However, in order to provide an explicit formula for the complexity of the word u_β, it is essential to have $|V^{(n)}| < |U^{(n)}|$, which is established in our paper if the parameters satisfy the conditions $t_1 > \max\{t_2, \dots, t_{m-1}\}$ or $t_1 = t_2 = \cdots = t_{m-1}$. We have chosen such requirements in order to avoid technical complications, but we conjecture that $|V^{(n)}| < |U^{(n)}|$ is satisfied for every sequence of parameters $t_1, t_2, \dots, t_m$ corresponding to the Rényi expansion $d_\beta(1)$ for some simple Parry number β.

In the second part of our paper we focus on the cubic Pisot number $1 + 2\cos(2\pi/7)$ as an example of a Parry number with infinite Rényi expansion of 1, namely $d_\beta(1) = 2(01)^\omega$. In this case there are no maximal left special factors (which is apparently connected with the fact that β is an algebraic unit), but there are two infinite left special factors. We see that the structure of left special factors for finite and infinite $d_\beta(1)$ is essentially different. The question about complexity of u_β for general β with eventually periodic Rényi expansion $d_\beta(1)$ remains open.

Acknowledgements. We thank the referees for the accuracy of their comments, and their interesting suggestions.

References

[1] J.-P. Allouche, Sur la complexité des suites infinies. *Bull. Belg. Math. Soc. Simon Stevin* **1** (1994) 133-143.
[2] P. Arnoux et G. Rauzy, Représentation géométrique de suites de complexité $2n+1$. *Bull. Soc. Math. France* **119** (1991) 199-215.
[3] J. Berstel, Recent results on extensions of Sturmian words. *J. Algebra Comput.* **12** (2003) 371-385.
[4] A. Bertrand, Développements en base de Pisot et répartition modulo 1. *C. R. Acad. Sci. Paris* **285A** (1977) 419-421.
[5] A. Bertrand-Mathis, Comment écrire les nombres entiers dans une base qui n'est pas entière. *Acta Math. Acad. Sci. Hungar.* **54** (1989) 237-241.
[6] J. Cassaigne, Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997) 67-88.
[7] J. Cassaigne, S. Ferenczi and L. Zamboni, Imbalances in Arnoux-Rauzy sequences. *Ann. Inst. Fourier* **50** (2000) 1265-1276.
[8] S. Fabre, Substitutions et β-systèmes de numération. *Theoret. Comput. Sci.* **137** (1995) 219-236.
[9] Ch. Frougny, J.-P. Gazeau and R. Krejcar, Additive and multiplicative properties of point sets based on beta-integers. *Theoret. Comput. Sci.* **303** (2003) 491-516.
[10] M. Lothaire, *Algebraic combinatorics on words*. Cambridge University Press (2002).
[11] W. Parry, On the β-expansions of real numbers. *Acta Math. Acad. Sci. Hungar.* **11** (1960) 401-416.
[12] J. Patera, Statistics of substitution sequences. On-line computer program, available at `http://kmlinux.fjfi.cvut.cz/~patera/SubstWords.cgi`
[13] A. Rényi, Representations for real numbers and their ergodic properties. *Acta Math. Acad. Sci. Hungar.* **8** (1957) 477-493.
[14] K. Schmidt, On periodic expansions of Pisot numbers and Salem numbers. *Bull. London Math. Soc.* **12** (1980) 269-278.
[15] W.P. Thurston, *Groups, tilings, and finite state automata*. Geometry supercomputer project research report GCG1, University of Minnesota (1989).
[16] O. Turek, Complexity and balances of the infinite word of β-integers for $\beta = 1+\sqrt{3}$, in *Proc. of Words'03*, Turku. TUCS Publication **27** (2003) 138-148.

Communicated by C. Choffrut.
Received September, 2003. Accepted February 12, 2004.