*Discrete Mathematics and Theoretical Computer Science* DMTCS vol. **9:2**, 2007, 125–144

# *Infinite special branches in words associated with beta-expansions*

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

[1] *LIAFA, CNRS UMR 7089, 2 place Jussieu, 75251 Paris Cedex 05, France, & Université Paris 8*
[2] *Department of Mathematics, FNSPE, Czech Technical University, Trojanova 13, 120 00 Praha 2, Czech Republic*

*received 13 Jan 2005, revised 20 Jun 2005, accepted 30 Jun 2005.*

A Parry number is a real number $\beta > 1$ such that the Rényi $\beta$-expansion of 1 is finite or infinite eventually periodic. If this expansion is finite, $\beta$ is said to be a *simple* Parry number. Remind that any Pisot number is a Parry number. In a previous work we have determined the complexity of the fixed point $u_\beta$ of the canonical substitution associated with $\beta$-expansions, when $\beta$ is a simple Parry number. In this paper we consider the case where $\beta$ is a non-simple Parry number. We determine the structure of infinite left special branches, which are an important tool for the computation of the complexity of $u_\beta$. These results allow in particular to obtain the following characterization: the infinite word $u_\beta$ is Sturmian if and only if $\beta$ is a quadratic Pisot unit.

**Keywords:** factor complexity function, Parry numbers

## 1 Introduction

A natural way to understand the combinatorial structure of an infinite word consists in studying its (factor) complexity function. The complexity function of an infinite word $u$ over a finite alphabet is defined as follows: $\mathcal{C}(n)$ is the number of finite factors of length $n$ appearing in $u$. Clearly an infinite word $u$ is eventually periodic if and only if there exists some $n$ in $\mathbb{N}$ such that $\mathcal{C}(n) \leqslant n$. The simplest aperiodic words have complexity $\mathcal{C}(n) = n + 1$ for all $n$ in $\mathbb{N}$. Such words are called *Sturmian* words. The Fibonacci word is well known to be Sturmian, see for instance [13, Chapter 2]. The Fibonacci word is associated in a canonical way with the golden ratio.

Numeration systems where the base $\beta$ is not an integer have been wisely studied, and one can find definitions and results in [13, Chapter 7]. Let us recall the particular role played by some numbers. A *Pisot number* $\beta$ is an algebraic integer such that all its Galois conjugates are less than 1 in modulus. The golden ratio and the natural integers are Pisot numbers. Numeration systems in base a Pisot number play an important role in the modelization of quasicrystals. The first quasicrystal was discovered in 1984: it is a solid structure presenting a local symmetry of

1365–8050 

order 5, *i.e.*, a local invariance under rotation of $\pi/5$, and it is linked to the golden ratio and to the Fibonacci substitution. The Fibonacci word is a historical model of a one-dimensional mathematical quasicrystal.

More generally, let $\beta > 1$ be a real number such that the Rényi expansion of 1 in base $\beta$ is eventually periodic or finite; these numbers are called *Parry numbers* (simple if the Rényi expansion of 1 is finite). It is known that a Pisot number is a Parry number, [4, 18].

With a Parry number $\beta$ is associated a substitution $\varphi_\beta$ having a fixed point $u_\beta$ (see next section for definitions). This substitution generates a tiling of the non-negative real line with a finite number of tiles [19, 8]. The vertices are labelled by the elements of the set of non-negative $\beta$-integers, which are real numbers having a $\beta$-expansion with no fractional part. The notion of *$\beta$-integer* has been introduced in the domain of quasicrystallography, see for instance [10].

Let us recall known results on complexity of infinite words $u_\beta$ associated with $\beta$-expansions. The infinite word $u_\beta$ is aperiodic and therefore has complexity $\mathcal{C}(n) \geqslant n+1$. Another general result about the complexity of $u_\beta$ is derived from the fact that the complexity of a fixed point of a primitive substitution is a sublinear function, see for instance [1]. A substitution $\xi$ over an alphabet $\mathcal{A}$ is called *primitive*, if there exists $k$ in $\mathbb{N}$ such that every letter $b$ of $\mathcal{A}$ appears in the $k$-th iteration $\xi^k(a)$ of every letter $a$ in $\mathcal{A}$. It can be easily seen that the canonical substitution $\varphi_\beta$ satisfies this condition. Therefore the complexity of the infinite word $u_\beta$ is a sublinear function.

For the determination of the complexity $\mathcal{C}(n)$ of an infinite word $u$ it is enough to know the first difference $\Delta\mathcal{C}(n) := \mathcal{C}(n+1) - \mathcal{C}(n)$ for all $n$ in $\mathbb{N}$. One can compute $\Delta\mathcal{C}(n)$ with the knowledge of left special factors (or right special factors) of $u$, introduced by Cassaigne [5]. Left special factors are factors of $u$ which have at least two left extensions in $u$. The description of all left special factors and the cardinality of the corresponding left extensions is crucial for the determination of the complexity.

In a previous work [11] we have studied the complexity of infinite words associated with $\beta$-expansions for a simple Parry number $\beta$. We recall some of the results. An infinite left special branch is an infinite word whose every prefix is a left special factor. Letters which belong to the extension of every prefix of an infinite left special branch are called its extensions.

**Theorem 1** [11] *Let $\beta$ be a simple Parry number with $d_\beta(1) = t_1 t_2 \cdots t_m$. Then $u_\beta$ has a unique infinite left special branch, namely $u_\beta$ itself. Moreover, the left extensions of $u_\beta$ are $\{0, 1, \ldots, m-1\}$.*

This result implies that $\Delta\mathcal{C}(n) \geqslant m-1$. On the other hand, in the same paper we show using the structure of maximal left special factors that under some additional weak conditions one has $\Delta\mathcal{C}(n) \leqslant m$. The complexity of infinite words $u_\beta$ for simple Parry numbers $\beta$ is therefore well understood.

The present work is devoted to the case of Parry numbers $\beta$ which are not simple. We describe the structure of infinite left special branches in the case that, in $d_\beta(1) = t_1 t_2 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$, all the coefficients $t_i$'s are positive (the integers $m$ and $p$ are chosen minimal). The situation is rather different according to whether the length $p$ of the periodic part of $d_\beta(1)$ is greater than or equal to 2, or is equal to 1.

If $p \geqslant 2$ then $u_\beta$ has a unique infinite left special branch, namely $u_\beta$ itself. Unlike the case of simple Parry numbers, here the set of left extensions of $u_\beta$ is equal to $\{m, m+1, \ldots, m+p-1\}$, Theorem 10.

If the length of the period of $d_\beta(1)$ is $p = 1$, then $u_\beta$ is not an infinite left special branch of itself, but there are $m$ different infinite left special branches, each of them having extension with two letters only, Theorem 13.

We then give an exact formula for the value of the complexity function in the case that $d_\beta(1) = ab^\omega$, Theorem 18; then $\beta$ is a quadratic Pisot number. From this and previous results from [10] and [11] follows that $u_\beta$ is a Sturmian word if and only if $\beta$ is a quadratic Pisot unit, *i.e.*, $\beta$ is the largest of the roots of $x^2 = qx + 1$, $q \geqslant 1$, or $x^2 = qx - 1$, $q \geqslant 3$.

Sturmian words can be characterized by the property that they are binary words with exactly one left and exactly one right special factor of each length. A generalization of this notion are the Arnoux-Rauzy words, defined as words over an alphabet with $k$ letters, having exactly one left and exactly one right special factor of each length with $k$ extensions. In [11] we have shown that, for a simple Parry number $\beta$, $u_\beta$ is an Arnoux-Rauzy word if and only if $\beta$ is a *confluent* Pisot number, that is to say if $d_\beta(1) = tt \cdots t1$, see [9] for definitions and properties of confluent numeration systems. Note that in [2] we completely describe the palindromic complexity of the word $u_\beta$ when $\beta$ is a confluent Pisot number. Here we are only able to prove that for non-simple Parry numbers $\beta$ with all the coefficients $t_i$'s positive, the infinite word $u_\beta$ is not an Arnoux-Rauzy word, Corollary 5.

We end this paper with some open problems. The situation is quite different if we allow some of the coefficients $t_i$'s to vanish. In [11] we have considered the Pisot number $\beta$ with $d_\beta(1) = 2(01)^\omega$, which appears in quasicrystallography as it presents a 7-fold symmetry. We have shown that in this case $u_\beta$ has two distinct infinite left special branches, and the complexity is equal to $\mathcal{C}(n) = 2n + 1$.

From a result of Cassaigne [6] we know that the first difference $\Delta\mathcal{C}(n)$ is bounded from above by a constant, for any word $u_\beta$ when $\beta$ is a Parry number. Computer experiments suggest the following conjecture: For all $n$ in $\mathbb{N}$

$$\#\mathcal{A} - 1 \leqslant \Delta\mathcal{C}(n) \leqslant \#\mathcal{A}$$

where $\#\mathcal{A}$ denotes the cardinality of the alphabet of $u_\beta$.

## 2 Preliminaries

### 2.1 Words and complexity

A word (finite or infinite) on a finite alphabet $\mathcal{A}$ is an arbitrary (finite or infinite) concatenation of letters in the alphabet. The length of a finite word $w$, *i.e.*, the number of its letters, is denoted by $|w|$. The set of finite words over the alphabet $\mathcal{A}$ is denoted by $\mathcal{A}^*$. Equipped with the operation of concatenation it is a free monoid. We will use the following notation $a^k = \underbrace{aa \cdots a}_{k \text{ times}}$.

A morphism on $\mathcal{A}^*$ is a map $\xi : \mathcal{A}^* \to \mathcal{A}^*$ for which $\xi(uw) = \xi(u)\xi(w)$ for all $u$ and $w$ in $\mathcal{A}^*$. A morphism is uniquely determined by the words $\xi(a)$ for all $a$ in $\mathcal{A}$. If $\xi(a)$ is a non-empty word for all $a \in \mathcal{A}$ and if there exists a letter $a_0$ in $\mathcal{A}$ such that $\xi(a_0) = a_0 w$ for some non-empty word $w$, then the morphism $\xi$ is called a *substitution*. The reader will find related results on substitutions in [16].

The action of a substitution can be naturally extended to infinite words $u = u_0u_1u_2\cdots$ by the prescription $\xi(u) = \xi(u_0)\xi(u_1)\xi(u_2)\cdots$. An infinite word $u$ is *invariant* under a substitution $\xi$ (or is its fixed point), if $\xi(u) = u$. The infinite word $\lim_{n\to\infty}\xi^n(a_0)$ is a fixed point of $\xi$.

A finite word $w = w_0w_1\cdots w_{n-1}$ is called a *factor* of $u$ of length $n$ if $w_0w_1\cdots w_{n-1} = u_iu_{i+1}\cdots u_{i+n-1}$ for some $i$ in $\mathbb{N}_0$ (we denote by $\mathbb{N}$ the set of positive integers and by $\mathbb{N}_0$ the set of non-negative integers). The set of factors of $u$ of all lengths is denoted by $F(u)$. The *factor complexity* (or simply *complexity*) of the infinite word $u$ is the mapping $\mathcal{C} : \mathbb{N} \to \mathbb{N}$ defined by

$$\mathcal{C}(n) := \#\{w \mid w \in F(u),\, |w| = n\}\,.$$

One can compute $\Delta\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n)$ with the knowledge of left special factors or right special factors of $u$. We shall focus on left special factors defined as follows. Let $w$ be a factor of an infinite word $u$. The set of all letters $a$ such that $aw$ is a factor of $u$ is called *left extension* of $w$ and denoted by $\mathrm{ext}(w)$,

$$\mathrm{ext}(w) := \{a \in \mathcal{A} \mid aw \in F(u)\}\,.$$

If $\#\mathrm{ext}(w) \geqslant 2$, then $w$ is called a *left special factor* of $u$.

The description of all left special factors and the cardinality of the corresponding left extensions is crucial for determination of the complexity, since the following simple relation holds

$$\mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{w\in F(u),|w|=n} \Big(\#\mathrm{ext}(w) - 1\Big)\,. \tag{1}$$

An infinite word $v = v_0v_1v_2\cdots$ is called an *infinite left special branch* of $u$ if every prefix of $v$ is a left special factor of $u$. Clearly, since $\mathrm{ext}(w') \supseteq \mathrm{ext}(w)$ for $w$ and $w'$ in $F(u)$ such that $w'$ is a prefix of $w$, the left extension of all sufficiently large prefixes of $v$ is constant. Thus we can define the left extension of the infinite left special branch $v$ by

$$\mathrm{ext}(v) := \bigcap_{w \text{ prefix of } v} \mathrm{ext}(w)\,.$$

A left special factor $w$ is called a *maximal left special factor*, if $\#\mathrm{ext}(wa) = 1$ for all letters $a$ in $\mathcal{A}$. It is clear that every left special factor is a prefix of a maximal left special factor or of an infinite left special branch.

## 2.2 $\beta$-expansions and substitutions

For a real base $\beta > 1$, every non-negative real number $x$ can be represented in the form

$$x = x_k\beta^k + x_{k-1}\beta^{k-1} + x_{k-2}\beta^{k-2} + \cdots = \sum_{i=-\infty}^{k} x_i\beta^i\,,$$

where $x_i \in \{0, 1, \ldots, \lceil\beta\rceil - 1\}$. Moreover, if we require that for every $N \in \mathbb{Z}$, $N \leqslant k$, the condition $0 \leqslant x - \sum_{i=N}^{k} x_i\beta^i < \beta^N$ is fulfilled, then such a representation is unique, it is called the *$\beta$-expansion* of $x$ and it is denoted by

$$(x)_\beta = x_kx_{k-1}\cdots x_1x_0\bullet x_{-1}x_{-2}\cdots\,.$$

The coefficients $x_i$ can be obtained by a greedy algorithm [17]. A representation having only finitely many non-zero coefficients is said to be *finite*, and the trailing zeroes are omitted.

Numbers $x$ such that the $\beta$-expansion of $|x|$ is of the form $(|x|)_\beta = x_k x_{k-1} \cdots x_0 \bullet 000 \cdots$ are called *$\beta$-integers*, see [10], and the set of $\beta$-integers is denoted by $\mathbb{Z}_\beta$. If $\beta$ is an integer $> 1$, we have $\mathbb{Z}_\beta = \mathbb{Z}$. If $\beta$ is not an integer, the structure of $\mathbb{Z}_\beta$ is determined by the so-called *Rényi expansion* $d_\beta(1)$ *of 1*, which is defined using the map $T_\beta(x) := \beta x - \lfloor \beta x \rfloor$ as the sequence

$$d_\beta(1) = t_1 t_2 t_3 \cdots , \qquad \text{where} \qquad t_i = \lfloor \beta T_\beta^{i-1}(1) \rfloor \quad \text{for } i \in \mathbb{N}$$

Parry in [14] has shown that a sequence of non-negative integers $t_1 t_2 t_3 \cdots$ is a Rényi expansion $d_\beta(1)$ of 1 for some $\beta > 1$ if and only if $t_i t_{i+1} t_{i+2} \cdots$ is lexicographically strictly smaller than $t_1 t_2 t_3 \cdots$ for every $i \in \mathbb{N}$, $i > 1$. This in particular implies that $t_i \leqslant t_1$ for all $i \in \mathbb{N}$.

From the construction of $d_\beta(1)$ it is obvious that $1 = \sum_{i=1}^{\infty} t_i \beta^{-i}$. It can be shown that consecutive $\beta$-integers have distances of the form $D_k = \sum_{i=1}^{\infty} t_{k+i} \beta^{-i}$ for $k \in \mathbb{N}_0$. The set of distances $\{D_k \mid k \in \mathbb{N}_0\}$ is thus finite if and only if the Rényi expansion $d_\beta(1)$ is eventually periodic [19]. Numbers with eventually periodic $d_\beta(1)$ are called *Parry numbers*. Obviously, a Parry number is an algebraic integer. A Parry number is called *simple* if $d_\beta(1)$ is finite, *i.e.*,

$$d_\beta(1) = t_1 t_2 \cdots t_m \tag{2}$$

where $m$ in $\mathbb{N}$ is such that $t_m \neq 0$. Otherwise

$$d_\beta(1) = t_1 t_2 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega \tag{3}$$

where in the period at least one of the coefficients is non-zero. Preperiod and period are not given uniquely, we shall assume that they have minimal length. In this case $t_m \neq t_{m+p}$.

When $\beta$ is a simple Parry number satisfying (2), the set $\mathbb{Z}_\beta$ has $m$ different distances $\{D_i \mid i = 0, 1, \ldots, m-1\}$ between neighbours. If $\beta$ is a Parry number with $d_\beta(1)$ of the form (3), then there are $m + p$ distances $\{D_i \mid i = 0, 1, \ldots, m+p-1\}$. Identifying the distance $D_i$ with $i$, the sequence of distances between non-negative $\beta$-integers forms an infinite word $u_\beta$ in the alphabet $\{0, 1, \ldots, m-1\}$, resp. $\{0, 1, \ldots, m+p-1\}$.

The infinite word $u_\beta$ associated with the sequence of distances between consecutive $\beta$-integers for a Parry number $\beta$ is the fixed point of a substitution $\varphi = \varphi_\beta$ canonically associated with $\beta$ [8]. For a simple Parry number $\beta$ satisfying (2) this substitution is defined on the alphabet $\mathcal{A} = \{0, 1, \ldots, m-1\}$ by

$$\begin{array}{rcl} \varphi(0) & = & 0^{t_1} 1 \\ \varphi(1) & = & 0^{t_2} 2 \\ & \vdots & \\ \varphi(m-2) & = & 0^{t_{m-1}} (m-1) \\ \varphi(m-1) & = & 0^{t_m} \end{array} \tag{4}$$

Thus $u_\beta = \lim_{n \to \infty} \varphi^n(0)$.

For instance take $\beta = \frac{1+\sqrt{5}}{2}$ the golden ratio. Then $d_\beta(1) = 11$, and the golden ratio is a simple Parry number. The substitution $\varphi$ associated with $\frac{1+\sqrt{5}}{2}$ is the Fibonacci substitution defined by

$$\varphi(0) = 01 \, , \qquad \varphi(1) = 0 \, .$$

The infinite word $u_\beta$ is the Fibonacci word

$$u_\beta = 0100101001\cdots .$$

In case the Rényi expansion is infinite eventually periodic (3) the substitution is defined on the alphabet $\mathcal{A} = \{0, 1, \ldots, m+p-1\}$ by

$$\begin{array}{rcl} \varphi(0) & = & 0^{t_1}1 \\ \varphi(1) & = & 0^{t_2}2 \\ & \vdots & \\ \varphi(m+p-2) & = & 0^{t_{m+p-1}}(m+p-1) \\ \varphi(m+p-1) & = & 0^{t_{m+p}}m \end{array} \tag{5}$$

and $u_\beta = \lim_{n\to\infty} \varphi^n(0)$.

## 3 Factors of $u_\beta$ of special type

From now on we shall consider Parry numbers $\beta$, which are not simple, *i.e.*, such that their Rényi expansion is of the form $d_\beta(1) = t_1 t_2 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$.

With the exception of Lemma 2 and Lemma 3 we focus on those $\beta$ which have all their coefficients in $d_\beta(1)$ positive, *i.e.*, $t_i \geqslant 1$ for all $i$. Note that this assumption together with the Parry's lexicographic condition on $d_\beta(1)$ implies that $t_1 \geqslant 2$. Such conditions help us to avoid technicalities, the ideas however could be generalized.

Lemma 2 can be proved easily by induction with the following notation: Let $v = w^k v'$, then $w^{-k}v = v'$. The symbol $\circ$ stands for concatenation.

**Lemma 2** *For every $n$ in $\mathbb{N}$,*

$$\begin{aligned} \varphi^{n+m+p}(0) = {} & \big(\varphi^{n+m+p-1}(0)\big)^{t_1} \big(\varphi^{n+m+p-2}(0)\big)^{t_2} \cdots \big(\varphi^{n}(0)\big)^{t_{m+p}} \\ & \circ \big(\varphi^{n}(0)\big)^{-t_m} \cdots \big(\varphi^{n+m-2}(0)\big)^{-t_2} \big(\varphi^{n+m-1}(0)\big)^{-t_1} \varphi^{n+m}(0) . \end{aligned}$$

The following lemma can be simply derived from the definition of the morphism $\varphi$.

**Lemma 3** *Let $n$ be in $\mathbb{N}$. The word $\varphi^{m+n}(0)$ ends with the letter $m+k$, where $k \in \{0, 1, \ldots, p-1\}$ and $k \equiv n \pmod p$.*

From now on we assume that in $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$, for each $i \geqslant 1$, $t_i \geqslant 1$, because in that case non-zero letters in the infinite word $u_\beta$ are separated by blocks of 0's. The following result describes the form of these blocks.

**Proposition 4** *All factors of the infinite word $u_\beta$ of the form $X0^sY$, where $X$ and $Y$ are non-zero letters and $s$ is in $\mathbb{N}_0$, are precisely the following ones:*

$$\begin{array}{ll} k0^{t_1}1 & \text{for } k = 1, 2, \ldots, m+p-1, \\ 10^{t_k}k & \text{for } k = 1, 2, \ldots, m+p-1, \\ 10^{t_{m+p}}m . & \end{array} \tag{6}$$

**Proof:** Since the substitution $\varphi$ is primitive, every factor appears infinitely many times in its fixed point $u_\beta$. From the definition of $\varphi$ we see that for $k \neq m$, the desired blocks are of the form $X0^{t_k}k$, where $X \neq 0$. For $k = m$, the blocks are of the form $X0^{t_m}m$ and $X0^{t_{m+p}}m$.

Consider a block $X0^{t_k}k$ for $k \geqslant 2$. Such a block is a suffix of $\varphi\big(0^{t_{k-1}}(k-1)\big)$. Since $t_{k-1} > 0$, we have $\varphi\big(0^{t_{k-1}}(k-1)\big) = w0^{t_1}10^{t_k}k$ for some word $w$, which implies that $X = 1$.

If $k = m$, the block $X0^{t_{m+p}}m$ is a suffix of $\varphi\big(0^{t_{m+p-1}}(m+p-1)\big)$. Since $t_{m+p-1} > 0$, we have $\varphi\big(0^{t_{m+p-1}}(m+p-1)\big) = w0^{t_1}10^{t_{m+p-1}}m$ for some word $w$, which implies that $X = 1$.

Let us describe the blocks of the form $X0^{t_1}1$. Since $u_\beta = \lim_{n\to\infty} \varphi^n(0)$, $\varphi^2(0) = \big(0^{t_1}1\big)^{t_1}0^{t_2}2$ and $t_1 \geqslant 2$, the word $u_\beta$ contains the factor 10. Then it has also the factor $\varphi(10) = 0^{t_2}20^{t_1}1$, and thus also 20. Similarly $u_\beta$ has a factor $k0$ for every $k \in \{1, 2, \dots, m+p-2\}$. This implies that $u_\beta$ has factors $\varphi(k0) = 0^{t_k}(k+1)0^{t_1}1$. Thus the desired blocks are $k0^{t_1}1$ for $k = 1, 2, \dots, m+p-1$.

Discussing all possibilities for the letter $Y$, we have described all the blocks of the form $X0^sY$, $X, Y \neq 0$. $\square$

The above proposition implies an interesting corollary. For all considered non-simple Parry numbers $\beta$, the set of factors of $u_\beta$ is not closed under mirror image. For, taking a $k$ such that $t_k < t_1$, the set of factors contains the word $10^{t_k}k$ but not the word $k0^{t_k}1$. However, invariance of the set of factors under mirror image is a necessary condition for a word to be an Arnoux-Rauzy word [7].

**Corollary 5** *Let $d_\beta(1) = t_1 \cdots t_m(t_{m+1} \cdots t_{m+p})^\omega$ with $t_i > 0$ for all $1 \leqslant i \leqslant m+p$. Then $u_\beta$ is not an Arnoux-Rauzy word.*

For further considerations we will use the following notation.

**Remark 1** *Assume that for some non-zero letter $Y$ there exist distinct integers $s, r$ such that $X_10^rY$ and $X_20^sY$ are in $F(u)$ for some letters $X_1 \neq 0$ and $X_2 \neq 0$. From Proposition 4 it is obvious that necessarily $Y = m$. The corresponding blocks are $10^{t_m}m$ and $10^{t_{m+p}}m$. In the sequel we use the following notation*

$$t := \min\{t_m, t_{m+p}\}\,. \tag{7}$$

# 4 Pre-images of left special factors

The description of the structure of left special factors in the infinite word $u_\beta$ is facilitated by the fact that every left special factor is given by an image of a shorter left special factor under the substitution. In this section we study the action of the substitution $\varphi$ on the left special factors, depending on their left extensions.

**Lemma 6** *Let $w$ be a left special factor of $u_\beta$ containing a non-zero letter and such that $0 \in \mathrm{ext}(w)$. Then $\mathrm{ext}(w) = \{0, 1\}$ and there exists a factor $\hat{w}$ of $u_\beta$ such that $w = 0^tm\varphi(\hat{w})0^s$ where $s \in \mathbb{N}_0$, $s \leqslant t_1$ and $\mathrm{ext}(\hat{w}) \supseteq \{m-1, m+p-1\}$ and $t$ is given by (7).*

**Proof:** Let $w$ be a left special factor of $u_\beta$ containing a non-zero letter and such that $0 \in \mathrm{ext}(w)$. Then there exists a prefix of $w$ of the form $0^\ell Y$ for some non-zero letter $Y$ and some $\ell \in \mathbb{N}_0$. Let $X \in \mathrm{ext}(w)$, $X \neq 0$. Since $X0^\ell Y$ and $0^{\ell+1}Y$ are factors of $u_\beta$, Remark 1 implies that $Y = m$, $\ell = t$, and $X = 1$. The factor $w$ is therefore of the form $w = 0^tmw'0^s$ for some $s \in \mathbb{N}_0$ and some word $w'$, which is either empty or ends in a non-zero letter. If $w'$ is empty (which corresponds to the case that $w$ contained exactly one non-zero letter), then the proof is finished taking $\hat{w}$ as the empty word. Otherwise $w'$ is a factor of $u_\beta$ ending in a non-zero letter. Form (6) it follows that $s \leqslant t_1$.

It remains to show that $w' = \varphi(\hat{w})$ for some word $\hat{w}$ such that $\mathrm{ext}(\hat{w}) \supseteq \{m-1, m+p-1\}$. We show it for $t_{m+p} > t_m$, the opposite case is analogous. If $t_{m+p} > t_m$, then words $10^{t_m}mw'$ and $0^{t_{m+p}}mw'$ are factors of $u_\beta$. From the definition of $\varphi$ it follows that there exists a word $\hat{w}$ such that $0^{t_m}mw' = \varphi\big((m-1)\hat{w}\big)$ and $0^{t_{m+p}}mw' = \varphi\big((m+p-1)\hat{w}\big)$, and both words $(m-1)\hat{w}$ and $(m+p-1)\hat{w}$ are factors of $u_\beta$. Thus $\hat{w}$ is a left special factor of $u_\beta$ and its left extension contains letters $m-1$ and $m+p-1$, what was to be shown. □

**Lemma 7** *Let $w$ be a left special factor of $u_\beta$ containing a non-zero letter and suppose there exist two distinct non-zero letters $X$ and $Y$ such that $X$ and $Y$ are in* $\mathrm{ext}(w)$. *Then there exists a left special factor $\hat{w}$ and an integer $s \in \mathbb{N}_0$, $s \leqslant t_1$, such that $w = \varphi(\hat{w})0^s$. Moreover,* $\mathrm{ext}(w)$ *is formed by the suffixes of length 1 of the words $\varphi(Z)$ for all $Z \in \mathrm{ext}(\hat{w})$.*

**Proof:** Lemma 6 implies that $0 \notin \mathrm{ext}(w)$. Let us write $w$ in the form $w = w'0^s$, where $w'$ ends in a non-zero letter. Obviously $s \leqslant t_1$. If $\mathrm{ext}(w) = \{X_1, \dots, X_k\}$, then $X_1w', \dots, X_kw'$ are factors of $u_\beta$. Since $X_i \neq 0$, for all $i = 1, 2, \dots, k$, the definition of $\varphi$ implies that there exist letters $Z_1, \dots, Z_k$ and a word $\hat{w}$ such that $Z_1\hat{w}, \dots, Z_k\hat{w}$ are factors of $u_\beta$ and $X_1w', \dots, X_kw'$ are suffixes of words $\varphi\big(Z_1\hat{w}\big), \dots, \big(Z_k\hat{w}\big)$. □

## 5 Infinite left special branches for $d_\beta(1)$ with period of length $p \geqslant 2$

In this section we focus on the case that $p \geqslant 2$. We shall see further on that words $u_\beta$ such that $p = 1$ differ substantially from this case.

Combination of Lemma 2 and 3 leads to the fact that the word $\varphi^n(0)$ has at least $p$ left extensions, namely that $\mathrm{ext}\big(\varphi^n(0)\big) \supseteq \{m, m+1, \dots, m+p-1\}$.

**Corollary 8** *Let $p \geqslant 2$. Then $u_\beta$ is an infinite left special branch of itself and*

$$\mathrm{ext}(u_\beta) \supseteq \{m, m+1, \dots, m+p-1\}\,.$$

The following lemma shows that the letters $0, 1, \dots, m-1$ do not belong to the extension of any infinite left special branch.

**Lemma 9** *Let $p \geqslant 2$. There exists a constant $K$ such that every left special factor $w$, whose left extension contains a letter $i \in \{0, 1, \dots, m-1\}$, has length $|w|$ bounded by $K$.*

**Proof:** Obviously, the largest factor of $u_\beta$ not containing a non-zero letter is $0^{t_1}$. Suppose that $w$ is a left special factor containing a non-zero letter.

First assume that $\mathrm{ext}(w)$ contains the letter 0. According to Lemma 6 we have $\mathrm{ext}(w) = \{0, 1\}$ and there exist a word $w_1$ and a non-negative integer $s_1 \leqslant t_1$ such that

$$w = 0^t m\varphi(w_1)0^{s_1}\,, \qquad \mathrm{ext}(w_1) \supseteq \{m-1, m+p-1\}\,.$$

Let us discuss the case $p \geqslant m$. Lemma 7 implies that there exist a sequence of words $w_2, w_3, \ldots, w_m$ and a sequence of non-negative integers $s_2, s_3, \ldots, s_m \leqslant t_1$ such that

$$\begin{array}{rclll} w_1 & = & \varphi(w_2)0^{s_2} & & \mathrm{ext}(w_2) \supseteq \{m-2, m+p-2\}\,, \\ w_2 & = & \varphi(w_3)0^{s_3} & & \mathrm{ext}(w_3) \supseteq \{m-3, m+p-3\}\,, \\ & \vdots & & & \\ w_{m-1} & = & \varphi(w_m)0^{s_m} & & \mathrm{ext}(w_m) \supseteq \{0, p\}\,. \end{array}$$

Since $p \neq 1$, the latter implies using Lemma 6 that $w_m$ does not contain a non-zero letter, *i.e.*, $w_m = 0^{s_{m+1}}$, where $s_{m+1} \leqslant t_1$. The left special factor $w$ is therefore of the form

$$w = 0^t m \varphi^m\big(0^{s_{m+1}}\big)\varphi^{m-1}\big(0^{s_m}\big)\cdots\varphi^2\big(0^{s_3}\big)\varphi\big(0^{s_2}\big)0^{s_1} \tag{8}$$

Since $s_i \leqslant t_1$ and $t \leqslant t_1$, the length of the word $w$ is bounded,

$$|w| \leqslant 1 + t_1 + \big|\varphi^m(0^{t_1})\varphi^{m-1}(0^{t_1})\cdots\varphi(0^{t_1})0^{t_1}\big|\,, \tag{9}$$

where the bound on the right hand side does not depend on $w$. Let us now discuss the case $m > p$. Again, using Lemma 7 we construct the sequence $w_2, w_3, \ldots$. For $w_p$ we obtain that $\mathrm{ext}(w_p) \supseteq \{m-p, m\}$. In the next step, two cases may occur, namely $\mathrm{ext}(w_{p+1}) \supseteq \{m-p-1, m-1\}$ and $\mathrm{ext}(w_{p+1}) \supseteq \{m-p-1, m+p-1\}$. For both of these cases we obtain that $\mathrm{ext}(w_m) \supseteq \{0, X\}$, with $X \geqslant 2$, and thus using Lemma 6 the word $w_m$ is of the form $w_m = 0^{s_m+1}$ for some $s_{m+1} \leqslant t_1$. Thus the original left special factor $w$ can be written in the form (8) and its length is bounded as in (9).

It remains to prove that $|w|$ is bounded, if $\mathrm{ext}(w)$ does not contain 0. Let us denote $k = \min\{i \mid i \in \mathrm{ext}(w)\}$. From the assumptions of the lemma we have $1 \leqslant k \leqslant m-1$. Pick $j$ in $\mathrm{ext}(w)$ different from $k$. Repeated application of Lemma 7 ($k$ times) leads to

$$w = \varphi^k(w_k)\varphi^{k-1}\big(0^{s_k}\big)\varphi^{k-2}\big(0^{s_{k-1}}\big)\cdots\varphi\big(0^{s_2}\big)0^{s_1}\,,$$

for some sequence of non-negative integers $s_1, \ldots, s_k \leqslant t_1$ and a left special factor $w_k$ of $u_\beta$, satisfying $\mathrm{ext}(w_k) \supseteq \{0, X\}$, $X \geqslant 1$. For such left special factor $w_k$ the statement of this lemma has been already proved in the first part of the proof. Thus the length $|w_k|$ is bounded by (9). Since $k \leqslant m-1$, also the length of the word $|w|$ is bounded. □

For the proof of the main theorem of this section we need to measure the diversity of two distinct infinite words $u = u_1u_2u_3\cdots$ and $v = v_1v_2v_3\cdots$. For that we introduce

$$d(u,v) := \min\{i \in \mathbb{N} \mid v_i \neq u_i\}\,. \tag{10}$$

The reciprocal value of $d(u,v)$ is used for defining a distance between infinite words.

**Theorem 10** *Let $p \geqslant 2$. Then $u_\beta$ has a unique infinite left special branch, namely $u_\beta$ itself. Moreover,* $\mathrm{ext}(u_\beta) = \{m, m+1, \ldots, m+p-1\}$.

**Proof:** Let $u$ be an infinite left special branch of $u_\beta$. Using Lemma 9 for every sufficiently long prefix $w$ of $u_\beta$ we have $\mathrm{ext}(w) \subseteq \{m, m+1, \ldots, m+p-1\}$. Thus using Lemma 7 every infinite left special branch of $u_\beta$ is the image $\varphi(\hat{u})$ of some infinite left special branch $\hat{u}$. The statement of the theorem will be proved by contradiction. Assume that $u = u_1u_2u_3\cdots$ and $v = v_1v_2v_3\cdots$ are distinct infinite left special branches of $u_\beta$. Assume that among all infinite left special branches we have chosen the pair $u, v$ so that $d(u, v)$ is minimal. We find infinite special branches $\hat{u}, \hat{v}$ in such a way that $\varphi(\hat{u}) = u$ and $\varphi(\hat{v}) = v$. From the definition of $\varphi$ it is obvious that $d(\hat{u}, \hat{v}) < d(u, v)$, which contradicts the choice of $u, v$. Thus we have proved that there exists at most one infinite left special branch. According to Corollary 8 we derive that the unique left special branch is $u_\beta$ and that $\mathrm{ext}(u_\beta) = \{m, m+1, \ldots, m+p-1\}$. $\square$

## 6 Infinite left special branches for $d_\beta(1)$ with period of length $p = 1$

Let us study the structure of infinite left special branches in case that the length of the period in the Rényi expansion $d_\beta(1)$ is $p = 1$. Here we work with the alphabet $\{0, 1, \ldots, m-1, m\}$, and the substitution $\varphi$ is given as

$$0 \mapsto 0^{t_1}1, \quad 1 \mapsto 0^{t_2}2, \quad \ldots, \quad (m-1) \mapsto 0^{t_m}m, \quad m \mapsto 0^{t_{m+1}}m.$$

**Lemma 11** *Let $p = 1$. Then there exists a constant $K$ such that every left special factor $w$ of $u_\beta$ such that $\{i, j\} \subset \mathrm{ext}(w)$ and $|i - j| \geqslant 2$ has length smaller than $K$.*

**Proof:** Let $w$ be a left special factor of $u_\beta$ such that $\{i, j\} \subset \mathrm{ext}(w)$. Without loss of generality assume that $i + 2 \leqslant j$. Either $w$ is of the form $w = 0^s$, then clearly $|w| \leqslant t_1$. Or $w$ contains a non-zero letter, and then using Lemma 6 we have $i \neq 0$. Repeated application of Lemma 7 ($i$ times) leads to

$$w = \varphi^i(w_i)\varphi^{i-1}(0^{s_i})\cdots\varphi^2(0^{s_3})\varphi(0^{s_2})0^{s_1}$$

for some non-negative integers $s_1, s_2, \ldots, s_i \leqslant t_1$, where $w_i$ is a left special factor of $u_\beta$ satisfying $\mathrm{ext}(w_i) \supset \{0, X\}$ for $X \geqslant 2$. Lemma 6 then implies that $w_i = 0^{s_{i+1}}$ for some $s_{i+1} \leqslant t_1$. Since $i \leqslant m - 2$, the length of the word $w$ is bounded by a constant independent of $w$. $\square$

**Corollary 12** *Let $p = 1$ and let $u$ be an infinite left special branch of $u_\beta$. Then $\mathrm{ext}(u) = \{i, i+1\}$ for some $i \in \{0, 1, 2, \ldots, m-1\}$.*

**Theorem 13** *Let $p = 1$. Define infinite words*

$$\begin{aligned}
v^{(0)} &:= 0^t m \varphi^m(0^t m)\varphi^{2m}(0^t m)\varphi^{3m}(0^t m)\cdots, \\
v^{(1)} &:= \varphi\big(v^{(0)}\big), \\
v^{(2)} &:= \varphi\big(v^{(1)}\big), \\
&\;\;\vdots \\
v^{(m-1)} &:= \varphi\big(v^{(m-2)}\big).
\end{aligned}$$

*For every $i \in \{0, 1, \ldots, m-1\}$, the word $v^{(i)}$ is an infinite left special branch with left extension $\mathrm{ext}(v^{(i)}) = \{i, i+1\}$. The infinite word $u_\beta$ has no other infinite left special branches.*

**Proof:** Let $u$ be an infinite left special branch of $u_\beta$ such that $0 \in \mathrm{ext}(u)$. Corollary 12 implies that $\mathrm{ext}(u) = \{0,1\}$. Due to Lemma 6 there exists an infinite left special branch of $u^{(1)}$ such that $u = 0^t\varphi\big(u^{(1)}\big)$ and $\mathrm{ext}(u^{(1)}) = \{m-1, m\}$. Lemma 7 and Lemma 11 implies that there exists an infinite left special branch $u^{(2)}$ so that $u^{(1)} = \varphi\big(u^{(2)}\big)$ and $\mathrm{ext}(u^{(2)}) = \{m-2, m-1\}$. In this way we obtain a sequence of infinite left special branches $u^{(0)}, \ldots, u^{(m)}$, where $u^{(m-1)} = \varphi\big(u^{(m)}\big)$ and $\mathrm{ext}(u^{(m)}) = \{0,1\}$. Together we have

$$u = 0^t m\varphi^m\big(u^{(m)}\big)\,. \tag{11}$$

We have shown that to every infinite left special branch $u$ with $\mathrm{ext}(u) = \{0,1\}$ one can find another infinite left special branch $u^{(m)}$ with the same extension. Moreover (11) holds true.

We show by contradiction that there cannot exist more than one infinite left special branch with $\mathrm{ext}(u) = \{0,1\}$. Assume the opposite, *i.e.*, that both $u, v$ are infinite left special branches with extension $\mathrm{ext}(u) = \mathrm{ext}(v) = \{0,1\}$, moreover assume that the pair $u, v$ is such that $d(u,v)$ is minimal, where $d(uv, v)$ is defined in (10). Now we find $u^{(m)}$ and $v^{(m)}$ so that (11) is satisfied. From the definition of $\varphi$ we obviously have $d\big(u^{(m)}, v^{(m)}\big) < d(u,v)$, but $u^{(m)}$, $v^{(m)}$ are infinite left special branches with extension $\{0,1\}$, which contradicts the choice of $u, v$.

If there exists an infinite left special branch $u$ with extension $\mathrm{ext}(u) = \{0,1\}$, it must satisfy the equation of words

$$u = 0^t m\varphi^m(u)\,. \tag{12}$$

Thus $u = 0^t m\hat{u}$ for some infinite word $\hat{u}$. Substituting into (12) we obtain

$$0^t m\hat{u} = 0^t m\varphi^m\big(0^t m\hat{u}\big) = 0^t m\varphi^m(0^t m)\varphi^m(\hat{u})\,.$$

Thus $\hat{u} = \varphi^m(0^t m)\hat{\hat{u}}$ for some infinite word $\hat{\hat{u}}$, *i.e.*, $u = 0^t m\varphi^m(0^t m)\hat{\hat{u}}$. Repeated substitution into (12) shows that the infinite word $v^{(0)}$ defined in the assertion of the theorem is a unique candidate to be the infinite left special branch with extension $\{0,1\}$. In order to prove that $v^{(0)}$ is indeed the infinite left special branch, it suffices to realize that $0^t m$ is a left special factor and use an auxiliary statement:

If $0^t mw$ for some finite word $w$ is a left special factor with $\mathrm{ext}(0^t mw) \supseteq \{0,1\}$, then $0^t m\varphi^m\big(0^t m\big)\varphi^m(w)$ is also a left special factor and its extension contains $\{0,1\}$.

Lemma 7 implies that every left special branch with left extension $\{i, i+1\}$ is the $i$-th iteration under $\varphi_\beta$ of an infinite left special branch with extension $\{0,1\}$. Since such a branch is unique, namely $v^{(0)}$, the only infinite left special branch with extension $\{i, i+1\}$ is $v^{(i)} = \varphi^i\big(v^{(0)}\big)$. This completes the proof. □

## 7 Complexity of words $u_\beta$ for quadratic Parry numbers

Quadratic Parry numbers have a simple characterization [12], namely that a quadratic irrational $\beta > 1$ is a Parry number if and only if it is a Pisot number, i.e., an algebraic integer with all conjugates in modulus strictly smaller than 1. In the quadratic case it results to all solutions of equations

$$\begin{array}{lll} x^2 = qx + r, & q, r \in \mathbb{N}, & q \geqslant r, \quad \text{or} \\ x^2 = qx - r, & q, r \in \mathbb{N}, & q \geqslant r + 2\,. \end{array}$$

Thus the Rényi expansion $d_\beta(1)$ has for quadratic Parry numbers the form

$$d_\beta(1) = qr \qquad \text{or} \qquad d_\beta(1) = (q-1)(q-r-1)^\omega\,.$$

Here we shall study the case of non-simple Parry numbers. Note that $q-1 > 0$ and $q-r-1 > 0$ and thus we can use the results derived in previous sections under the condition that the coefficients of the Rényi expansion of 1 are all positive. For simplicity of notation we denote $a = q-1$ and $b = q-r-1$, thus the Rényi expansion of 1 to be considered is $d_\beta(1) = ab^\omega$. The corresponding infinite word $u_\beta$ is a word over the binary alphabet $\{0,1\}$ invariant under the substitution

$$\varphi(0) = 0^a1\,, \quad \varphi(1) = 0^b1\,, \qquad \text{where } \ a > b \geqslant 1 \tag{13}$$

As a result of Section 6, we know that the word $u_\beta$ has a unique infinite left special branch, say $v$, which satisfies the word equation

$$v = 0^b1\varphi(v) = 0^b1\varphi(0^b1)\varphi^2(0^b1)\cdots\,. \tag{14}$$

In order to determine the complexity of the word $u_\beta$, we need to determine also the maximal left special factors.

**Remark 2**

*(i)* *According to Proposition 4 there exist only two factors of $u_\beta$ of the form $X0^sY$, $X \neq 0$ and $Y \neq 0$, namely $10^a1$ and $10^b1$.*

*(ii)* *Since $a > b$, every maximal left special factor, which contains at least one letter 1 has the suffix $10^b$.*

**Proposition 14**

*(i)* *Every maximal left special factor $w$, which contains at least one letter 1 is of the form $w = 0^b1\varphi(\hat{w})0^b$, where $\hat{w}$ is a maximal left special factor.*

*(ii)* *If $a = b+1$, then there is no maximal left special factor of the form $0^s$, $s$ in $\mathbb{N}$; if $a \geqslant b+2$, then the maximal left special factor not containing a letter 1 is $0^{a-1}$.*

**Proof:** Suppose that the maximal left special factor $w$ contains a letter 1. Since $0w$ and $1w$ are in $F(u_\beta)$, statement (i) of Remark 2 implies that $w = 0^b1w'$ for some word $w'$. If $w'$ does not contain a letter 1, then (ii) of Remark 2 implies $w = 0^b10^b$, which is a prefix of the infinite left special branch $v$ of $u_\beta$, (cf. (14)). This is a contradiction with the maximality of $w$. Therefore using the statement ii) of Remark 2 the word $w$ must be of the form $w = 0^b1w''10^b$ for some non-empty word $w''$. Since both $w = 00^b1w''10^b$ and $w = 10^b1w''10^b$ are factors of $u_\beta$, there exists a word $\hat{w}$ such that $\varphi(\hat{w}) = w''1$, and $1\hat{w}$ and $0\hat{w}$ belong to $F(u_\beta)$, *i.e.*, $\hat{w}$ is a left special factor of $u_\beta$. Assume that $\hat{w}$ is not a maximal special factor. Then there exists a letter $Y \in \{0,1\}$, such that $0\hat{w}Y$ and $01\hat{w}Y$ are factors of $u_\beta$. The images of these two factors, namely $0^a1\varphi(\hat{w})\varphi(Y)$ and $0^a10^b1\varphi(\hat{w})\varphi(Y)$ are also elements of $F(u_\beta)$. Therefore $0^b1\varphi(\hat{w})\varphi(Y)$ is a left special factor and $w = 0^b1\varphi(\hat{w})0^b$ is its proper prefix which contradicts the maximality of $w$.

The statement (ii) of the proposition is obvious from (i) of Remark 2. $\square$

**Proposition 15** *Let $w$ be a maximal left special factor of $u_\beta$. Then $0^b1\varphi(w)0^b$ is also a maximal left special factor of $u_\beta$.*

**Proof:** Since $0w$ and $1w$ are factors of $u_\beta$, then also the words $\varphi(0w) = 0^a1\varphi(w)$ and $\varphi(01w) = 0^a10^b1\varphi(w)$ belong to $F(u_\beta)$. Since $a > b$, the word $0^b1\varphi(w)$ is a left special factor of $u_\beta$. The word $0^b1\varphi(w)$ has the suffix 1, and the letter 1 is always followed by $0^b$. Thus $0^b1\varphi(w)0^b$ is also a left special factor. If it is not maximal, then either $0^b1\varphi(w)0^b1$ or $0^b1\varphi(w)0^a1$ is also a left special factor and thus $w1$ or $w0$ is a left special factor, which contradicts the maximality of $w$. $\Box$

Combining Propositions 14 and 15 we obtain the following corollary.

**Corollary 16** *Let $u_\beta$ be the infinite word invariant under the substitution* (13).

1. *The word $u_\beta$ contains maximal left special factors if and only if $a \geqslant b+2$.*
2. *Let $a \geqslant b+2$. Let $(U^{(k)})_{k\in\mathbb{N}}$ be the sequence of words satisfying the recurrent relation*

$$\begin{aligned} U^{(1)} &= 0^{a-1}\,, \\ U^{(k)} &= 0^b\varphi(U^{(k-1)})0^b\,, \quad k \geqslant 2\,. \end{aligned}$$

   *The set of words $\{U^{(k)} \mid k \in \mathbb{N}\}$ coincides with the set of all maximal left special factors of $u_\beta$.*

For a binary alphabet, the formula (1) for the first difference of the complexity function has the form

$$\Delta\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n) = \text{number of left special factors of length } n.$$

If $a = b+1$, then the number $\beta$ is a solution of $x^2 = (a+1)x - 1$ and thus it is an algebraic unit. (Recall that an algebraic unit is a root of a polynomial with integer coefficients such that the leading and absolute coefficients are $\pm 1$.) According to the statement 1 of Corollary 16, the corresponding word $u_\beta$ has no maximal left special factors, and therefore every left special factor of $u_\beta$ is a prefix of the unique infinite left special branch of $u_\beta$. This implies that $\Delta\mathcal{C}(n) = 1$ for every $n \in \mathbb{N}$. We deduce the following corollary, which says that $u_\beta$ is a Sturmian word for $r = a - b = 1$.

**Corollary 17** *Let $\beta > 1$ and let $\beta^2 = q\beta - 1$. Then the complexity $\mathcal{C}(n)$ of the infinite word $u_\beta$ satisfies $\mathcal{C}(n) = n+1$ for every $n \in \mathbb{N}$.*

In the following we therefore focus on the case $r \geqslant 2$, *i.e.*, $a \geqslant b+2$. Then every left special factor is a prefix of the infinite left special branch $v$ or of some maximal left special factor $U^{(k)}$, $k \in \mathbb{N}$. Note that it is not excluded that some of the left special factors are common prefixes of both. Therefore we define a sequence of words $(V^{(k)})_{k\in\mathbb{N}}$ as

$$V^{(k)} = \text{the maximal common prefix of } v \text{ and } U^{(k)}.$$

This means that for every $k \in \mathbb{N}$ there exists an infinite word $R_1^{(k)}$ and a non-empty finite word $R_2^{(k)}$ such that $v = V^{(k)}R_1^{(k)}$ and $U^{(k)} = V^{(k)}R_2^{(k)}$, and the words $R_1^{(k)}$, $R_2^{(k)}$ do not have a common prefix. Clearly $V^{(1)} = 0^b$.

The recurrent relation for $U^{(k)}$ and the word equation (14) for the infinite left special branch $v$ imply that

$$\begin{aligned} U^{(k)} &= 0^b 1\varphi(U^{(k-1)})0^b = 0^b 1\varphi(V^{(k-1)})\varphi(R_2^{(k-1)})0^b \\ v &= 0^b 1\varphi(v) = 0^b 1\varphi(V^{(k-1)})\varphi(R_1^{(k-1)}) \end{aligned} \tag{15}$$

Since one of the words $R_1^{(k-1)}$, $R_2^{(k-1)}$ begins with 1 and the other with 0, then one of the words $\varphi(R_1^{(k-1)})$, $\varphi(R_2^{(k-1)})$ has the prefix $0^b1$ and the other the prefix $0^a1$. Comparing the words $U^{(k)}$ and $v$ in the relations (15) we obtain

$$V^{(k)} = 0^b 1\varphi(V^{(k-1)})0^b\,.$$

Thus the word sequences $(U^{(k)})_{k\in\mathbb{N}}$ and $(V^{(k)})_{k\in\mathbb{N}}$ are given by the same recurrence relation; they differ by the initial word $U^{(1)} = 0^{a-1}$, $V^{(1)} = 0^b$. For the lengths of words of these sequences we therefore have

$$|U^{(k)}| = |V^{(k)}| + (a-b-1)|\varphi^{(k-1)}(0)|\,, \qquad \text{for } k\in\mathbb{N}\,. \tag{16}$$

From the definition of $U^{(k)}$ and $V^{(k)}$ it is obvious that the sequences of lengths $(|U^{(k)}|)_{k\in\mathbb{N}}$, $(|V^{(k)}|)_{k\in\mathbb{N}}$ are strictly increasing. In order to describe these sequences, we define the integer sequences $(G_k)_{k=0}^\infty$, $(H_k)_{k=0}^\infty$ by

$$\begin{aligned} G_0 &= H_0 = 1\,, \\ G_k &= |\varphi^k(0)|\,, \quad H_k = |\varphi^k(1)|\,, \quad \text{for } k\in\mathbb{N}\,. \end{aligned}$$

Since

$$\begin{aligned} \varphi^{(k)}(0) = \varphi^{(k-1)}(0^a1) &= (\varphi^{(k-1)}(0))^a\varphi^{(k-1)}(1)\,, \\ \varphi^{(k)}(1) = \varphi^{(k-1)}(0^b1) &= (\varphi^{(k-1)}(0))^b\varphi^{(k-1)}(1)\,, \end{aligned}$$

we obtain for the sequences $(G_k)_{k=0}^\infty$, $(H_k)_{k=0}^\infty$ the recurrence relations $G_0 = H_0 = 1$,

$$\begin{aligned} G_k &= aG_{k-1} + H_{k-1}\,, \\ H_k &= bG_{k-1} + H_{k-1}\,, \end{aligned} \qquad \text{for } k \geqslant 1\,. \tag{17}$$

One can easily show by induction on $k$ that for $k\geqslant 2$

$$V^{(k)} = \varphi(1)\varphi^2(1)\cdots\varphi^{(k-1)}(1)\varphi^{(k-1)}(0^b)\varphi^{(k-2)}(0^b)\cdots\varphi(0^b)0^b\,.$$

For the length $|V^{(k)}|$ we thus have

$$|V^{(k)}| = b\sum_{i=0}^{k-1} G_i + \sum_{i=1}^{k-1} H_i = \sum_{i=1}^{k} H_i - 1 \tag{18}$$

and using the relation (16)

$$|U^{(k)}| = \sum_{i=1}^{k} H_i - 1 + (a-b-1)G_{k-1}\,. \tag{19}$$

In order to show that for every $n \in \mathbb{N}$ there exists at most one left special factor of length $n$ which is not prefix of the infinite left special branch, it suffices to verify that

$$|V^{(k)}| > |U^{(k-1)}| \quad \text{for every} \quad k \geqslant 2\,.$$

We use (18) and (19) to see that it suffices to verify

$$H_k > (a-b-1)G_{k-2}\,. \tag{20}$$

Since $G_k$, $H_k$ are positive integers, from the relations (17) we easily obtain

$$G_k \quad > \quad aG_{k-1}\,, \tag{21}$$

$$G_{k+1} \quad = \quad (a+1)G_k - (a-b)G_{k-1}\,. \tag{22}$$

In order to verify the inequality (20) we rewrite its left hand side using the relations (17), (22) and estimate it using (21), to obtain

$$\begin{aligned} H_k = \;& G_{k+1} - aG_k = (b+1)G_{k-1} - (a-b)G_{k-2} > \\ > \;& a(b+1)G_{k-2} - (a-b)G_{k-2} = (ab+b)G_{k-2}\,. \end{aligned}$$

Since for positive integers $a, b$, $a > b \geqslant 1$ it holds that $ab + b > a - b - 1$, the inequality (20) is satisfied.

The elements of the sequences $(|U^{(k)}|)_{k\in\mathbb{N}}$, $(|V^{(k)}|)_{k\in\mathbb{N}}$ are thus ordered in the following way:

$$|V^{(1)}| < |U^{(1)}| < |V^{(2)}| < |U^{(2)}| < |V^{(3)}| < |U^{(3)}| < \cdots \tag{23}$$

Using the relations (18) and (22) we can easily derive that $|V^{(k)}|$ satisfies

$$|V^{(k+2)}| \;=\; (a+1)|V^{(k+1)}| \;-\; (a-b)|V^{(k)}| \;+\; 3b{-}a{+}1\,.$$

Taking (19), we see that $|U^{(k)}|$ satisfies the same recurrence relations but with different initial values.

The following theorem summarizes the results about the complexity function for $u_\beta$ when $\beta$ is a quadratic Parry number. The statement concerning non-simple Parry numbers follows from the above considerations; the part for simple Parry numbers can be derived from the general theorem proved in [11].

**Theorem 18** *Let $\beta$ be a quadratic Parry number. Then the complexity $\mathcal{C}(n)$ of the infinite word $u_\beta$ satisfies*

$$\mathcal{C}(n+1) - \mathcal{C}(n) = \begin{cases} 1, & \text{if} \quad \mathcal{U}_{k-1} \;<\; n \;\leqslant\; \mathcal{V}_k \quad \text{for some} \;\; k \in \mathbb{N}\,, \\ 2, & \text{if} \quad \mathcal{V}_k \;<\; n \;\leqslant\; \mathcal{U}_k \quad \text{for some} \;\; k \in \mathbb{N}\,, \end{cases}$$

*where $(\mathcal{U}_k)_{k\in\mathbb{N}_0}$ and $(\mathcal{V}_k)_{k\in\mathbb{N}_0}$ are integer sequences defined by the recurrences*

$$\begin{aligned} \mathcal{U}_{k+2} \quad &= \quad q\,\mathcal{U}_{k+1} - \; r\,\mathcal{U}_k \;+\; 2q{-}3r{-}1\,, \qquad \text{with} \quad \mathcal{U}_0 = -1, \quad \mathcal{U}_1 = q-2\,, \\ \mathcal{V}_{k+2} \quad &= \quad q\,\mathcal{V}_{k+1} - \; r\,\mathcal{V}_k \;+\; 2q{-}3r{-}1\,, \qquad \text{with} \quad \mathcal{V}_0 = -1, \quad \mathcal{V}_1 = q-r-1\,, \end{aligned}$$

*if* $\beta^2 = q\beta - r$, *for* $q, r \in \mathbb{N}$, $q \geqslant r + 2$, *and*

$$\begin{array}{rclll} \mathcal{U}_{k+2} & = & q\,\mathcal{U}_{k+1} + r\,\mathcal{U}_k + 2q\,, & \textit{with} \quad \mathcal{U}_0 = 1 - r^{-1}, & \mathcal{U}_1 = q + r - 1\,, \\ \mathcal{V}_{k+2} & = & q\,\mathcal{V}_{k+1} + r\,\mathcal{V}_k + 2q\,, & \textit{with} \quad \mathcal{V}_0 = 0, & \mathcal{V}_1 = q\,, \end{array}$$

*if* $\beta^2 = q\beta + r$, *for* $q, r \in \mathbb{N}$, $q \geqslant r$.

Note that the sequences $(\mathcal{U}_k)_{k\in\mathbb{N}_0}$ and $(\mathcal{V}_k)_{k\in\mathbb{N}_0}$ are formally defined even in the unitary case $r = 1$, but then they coincide and the first difference of complexity is constantly equal to 1, whence $\mathcal{C}(n) = n + 1$ for $n \in \mathbb{N}$. On the other hand, if $\beta$ is not a unit, then the above theorem implies that $\mathcal{C}(n)$ is not a linear function.

In order to characterize Parry numbers $\beta$ for which $u_\beta$ is a Sturmian word, it suffices to realize that $\mathcal{C}(1) = 2$ implies that $\beta$ is a quadratic integer. Thus we have the following corollary.

**Corollary 19** *The infinite word $u_\beta$ associated with a Parry number $\beta$ is Sturmian if and only if $\beta$ is a quadratic Pisot unit.*

# 8 Open problems and one more example

**1)** We have described the infinite left special branches only for those Parry numbers $\beta$ such that in $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$ all coefficients are positive. Under such assumptions and if $p \geqslant 2$, the infinite word $u_\beta$ has a unique infinite left special branch. The situation is quite different if we allow that some of the coefficients $t_i$ vanish. In [11] we consider the Pisot number $\beta$ with $d_\beta(1) = 2(01)^\omega$, which has two distinct infinite left special branches. The complexity in this case is equal to $\mathcal{C}(n) = 2n + 1$.

**2)** The studied infinite words $u_\beta$ are invariant under a primitive substitution and therefore have sublinear complexity $\mathcal{C}(n)$. From the result of Cassaigne [6] it follows that the first difference $\Delta\mathcal{C}(n)$ is bounded from above by a constant. All our computer experiments show that the first difference of complexity of $u_\beta$ for a Parry number $\beta$ takes at most two values. More precisely, denote $\#\mathcal{A}$ the cardinality of the alphabet of $u_\beta$.

**Conjecture 20** *For all $n$ in $\mathbb{N}$*

$$\#\mathcal{A} - 1 \leqslant \Delta\mathcal{C}(n) \leqslant \#\mathcal{A}.$$

Let us mention the results and examples which support this conjecture.

- In [11] it is shown that for $d_\beta(1) = t_1 t_2 \cdots t_m$ satisfying $t_1 > \max\{t_2, \ldots, t_{m-1}\}$ or $t_1 = t_2 = \cdots = t_{m-1}$ the conjecture holds true.
- The relation $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$ implies that $\beta$ is a root of the polynomial

$$\begin{aligned} &(x^p - 1)(x^m - t_1 x^{m-1} - t_2 x^{m-2} - \cdots - t_m) - \\ &\qquad - t_{m+1} x^{p-1} - t_{m+2} x^{p-2} - \cdots - t_{m+p-1} x - t_{m+p}\,. \end{aligned} \tag{24}$$

If this polynomial is irreducible, *i.e.*, $\beta$ is an algebraic integer of degree $m+p$, we can use the result of Tijdeman [20], which implies that

$$\Delta\mathcal{C}(n) \geqslant m+p-1 = \#\mathcal{A}-1\,.$$

Let us mention that for many non-simple Parry numbers $\beta$ the polynomial (24) is indeed irreducible.

- When $d_\beta(1) = t_1\cdots t_m(t_{m+1})^\omega$ and $t_i > 0$ for all $i$, Theorem 13 implies that $\Delta\mathcal{C}(n) \geqslant m = \#\mathcal{A}-1$. If moreover $m=1$, then according to Theorem 18 we have $\Delta\mathcal{C}(n) \leqslant \#\mathcal{A}$.
- Most interesting is the case when $d_\beta(1) = t_1\cdots t_m(t_{m+1}\cdots t_{m+p})^\omega$ with $p \geqslant 2$. The alphabet of the infinite word $u_\beta$ is $\mathcal{A} = \{0,1,\ldots,m+p-1\}$, while Theorem 10 implies only $\Delta\mathcal{C}(n) \geqslant p-1$ instead of expected $m+p-1$. The remaining contribution to the first difference of complexity can thus be obtained uniquely by maximal left special factors.

**Example 1** For illustration, let us study the structure of left special factors of the infinite word $u_\beta$ with $d_\beta(1) = 543(12)^\omega$. Such $\beta$ is a root of the irreducible polynomial $x^5-5x^4-5x^3-3x^2+3x+1$. For the study of the complexity of the infinite word $u_\beta$ we have used the computer program [15]. The infinite word $u_\beta$ over the alphabet $\mathcal{A} = \{0,1,2,3,4\}$ is the fixed point of the substitution

$$\varphi(0) = 0^51,\ \varphi(1) = 0^42,\ \varphi(2) = 0^33,\ \varphi(0) = 04,\ \varphi(0) = 0^23\,.$$

Although according to Theorem 10 the infinite word $u_\beta$ is an infinite left special branch with extension $\{3,4\}$, some prefixes of $u_\beta$ have larger extension, namely

the prefix of $u_\beta$ of length $n \leqslant 4$ has the extension $\{0,1,2,3,4\}$,
the prefix of $u_\beta$ of length $5 \leqslant n \leqslant 28$ has the extension $\{1,2,3,4\}$,
the prefix of $u_\beta$ of length $29 \leqslant n \leqslant 167$ has the extension $\{2,3,4\}$,
all prefixes of $u_\beta$ longer than 167 are left special factors with extension only $\{3,4\}$.

Figure 1 shows schematically the tree of left special factors of $u_\beta$. The cardinality of the extensions of the factors of given length is marked by the thickness of the lines.

Let us denote by $U^{(1)}$ the prefix of $u_\beta$ of length 4, by $U^{(2)}$ the prefix of $u_\beta$ of length 28, and by $U^{(3)}$ the prefix of $u_\beta$ of length 167. It is easy to see that

$$\begin{array}{rclrcl} U^{(1)} &=& 0^4\,, & \mathrm{ext}(U^{(1)}) &=& \{0,1,2,3,4\}\,,\\ U^{(2)} &=& \varphi(U^{(1)})0^4\,, & \mathrm{ext}(U^{(2)}) &=& \{1,2,3,4\}\,,\\ U^{(3)} &=& \varphi(U^{(2)})0^3\,, & \mathrm{ext}(U^{(3)}) &=& \{2,3,4\}\,. \end{array}$$

The first maximal left special factor is

$$U^{(4)} = 0^23\varphi(U^{(3)})0\,, \quad \text{with}\ \ \mathrm{ext}(U^{(4)}) = \{0,1\}\,.$$

This factor is of length 974 and the longest common prefix of $U^{(4)}$ and $u_\beta$ is $0^2$. Therefore in Figure 1 the broken line corresponding to $u^{(4)}$ and the half-line corresponding to $u_\beta$ have a common segment of length 2.

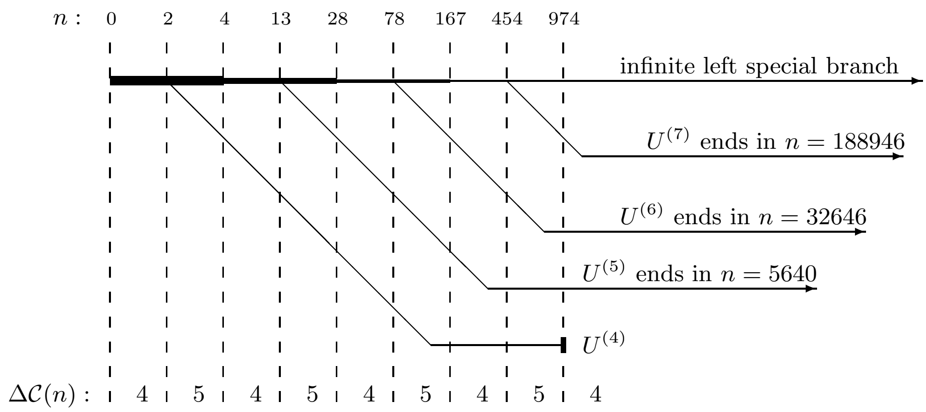


**Fig. 1:** Tree of left special factors of the infinite word $u_\beta$ with $d_\beta(1) = 543(12)^\omega$

It can be shown that every other maximal left special factor can be obtained by the recurrence

$$U^{(n)} = \varphi(U^{(n-1)})0^{s_n}\,, \qquad \text{where} \qquad s_n = \begin{cases} 1, & \text{if } n \text{ is even,} \\ 2, & \text{if } n \text{ is odd.} \end{cases}$$

Moreover, for all $n$ the extension satisfies $\#\mathrm{ext}(U^{(n)}) = 2$. The lengths of $U^{(n)}$ for $n = 5, 6, 7$ and of the common prefixes of $U^{(n)}$ and $u_\beta$ are given in Figure 1.

Let us mention that even the factors $U^{(1)}$, $U^{(2)}$, $U^{(3)}$ can be in some sense understood as maximal left special factors. For although they can be extended to the right by a letter $a \in \mathcal{A}$ in such a way that $U^{(i)}a$ is a left special factor, however for all such $a$, we have $\#\mathrm{ext}(U^{(i)}) > \#\mathrm{ext}(U^{(i)}a) \geqslant 2$.

Every left special factor $w$ of length $n > 167$ has in its extension only 2 letters, *i.e.*, every such left special factor contributes to the first difference of complexity by 1. $\Delta\mathcal{C}(n)$ is thus equal to the number of left special factors of length $n$. The bottom line in Figure 1 shows the increment of complexity which is equal to 4 or 5 for every $n$, which again supports Conjecture 20.

Notice that in this example the structure of maximal left special factors is essentially different from that of $u_\beta$ for quadratic non-simple Parry numbers $\beta$, see Section 7. Here the inequalities (23) say that the common prefix of $u_\beta$ and $U^{(n)}$ is longer than the previous maximal left special factor $U^{(n-1)}$, *i.e.*, the maximal left special factors do not overlap, unlike to our example.

**3)** As the last open problem let us mention the question for which $\beta$ the infinite word $u_\beta$ has other left special factors than prefixes of an infinite left special branch. This question is discussed in [11] for simple Parry numbers $\beta$, *i.e.*, for $d_\beta(1) = t_1 \cdots t_m$. It is shown that if $t_m \geqslant 2$, then $u_\beta$ has infinitely many maximal left special factors.

The condition $t_m \geqslant 2$ is however not necessary for existence of maximal left special factors. For example, the infinite word $u_\beta$ corresponding to the simple Parry number $\beta$ such that $d_\beta(1) = 101000101$ has a maximal left special factor [3]. In [11] it is shown that if $t_1 > \max\{t_2, \ldots, t_{m-1}\}$ or $t_1 = t_2 = \cdots = t_{m-1}$, then $u_\beta$ has a maximal left special factor if and only if $t_m \geqslant 2$.

For non-simple Parry numbers $\beta$ the question of existence of maximal left special factors has been treated only in the quadratic case.

## Acknowledgements

The authors acknowledge the financial support of the Czech Science Foundation GA ČR 201/05/0169.

## References

[1] J.-P. Allouche, Sur la complexité des suites infinies, *Bull. Belg. Math. Soc. Simon Stevin* **1** (1994), 133–143.

[2] P. Ambrož, Ch. Frougny, Z. Masáková and E. Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers, *Ann. Inst. Fourier* **56** (2006), 2131–2160.

[3] J. Bernat, V. Berthé, *oral communication*, (2004).

[4] A. Bertrand, Développements en base de Pisot et répartition modulo 1, *C. R. Acad. Sc. Paris*, Série A, t.285, (1977) 419–421.

[5] J. Cassaigne, Complexité et facteurs spéciaux, *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997), 67–88.

[6] J. Cassaigne, Special factors of sequences with linear subword complexity, *Proceedings of DLT'95*, Word Sci. Publishing, River Edge, NY, (1996), 25–34.

[7] X. Droubay, J. Justin, G. Pirillo, Epi-Sturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001), 539–553.

[8] S. Fabre, Substitutions et $\beta$-systèmes de numération, *Theoret. Comput. Sci.* **137** (1995), 219–236.

[9] Ch. Frougny, Confluent linear numeration systems, *Theoret. Comput. Sci.* **106** (1992), 183–219.

[10] Ch. Frougny, J.-P. Gazeau, R. Krejcar, Additive and multiplicative properties of point sets based on beta-integers, *Theoret. Comput. Sci.* **303** (2003), 491–516.

[11] Ch. Frougny, Z. Masáková, E. Pelantová, Complexity of infinite words associated with $\beta$-expansions, *RAIRO Theor. Inf. Appl.* **38** (2004), 162–184. Corrigendum, *RAIRO Theor. Inf. Appl.* **38** (2004), 269–271.

[12] Ch. Frougny and B. Solomyak, Finite beta-expansions, *Ergo. Th. Dyn. Syst.* **12** (1992), 713–723.

[13] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, Cambridge (2002).

[14] W. Parry, On the $\beta$-expansions of real numbers, *Acta Math. Acad. Sci. Hungar.* **11** (1960), 401–416.

[15] J. Patera, *Statistics of substitution sequences*, `http://kmlinux.fjfi.cvut.cz/~patera/SubstWords.cgi`

[16] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics **1794**, Springer (2002).

[17] A. Rényi, Representations for real numbers and their ergodic properties, *Acta Math. Acad. Sci. Hung.* **8** (1957), 477–493.

[18] K. Schmidt, On periodic expansions of Pisot numbers and Salem numbers, *Bull. London Math. Soc.* **12** (1980) 269–278.

[19] W.P. Thurston, *Groups, tilings, and finite state automata*, Lecture notes, AMS colloquium lectures, AMS, Providence, RI (1990).

[20] R. Tijdeman, On the minimal complexity of infinite words, *Indag. Math. N.S.* **10** (1999), 123–129.