#A24 INTEGERS 9 (2009), 281-310

FACTOR COMPLEXITY OF INFINITE WORDS ASSOCIATED WITH NON-SIMPLE PARRY NUMBERS

Karel Klouda
FNSPE, Czech Technical University in Prague and *LIAFA, Université Denis-Diderot (Paris VII)*
karel@kloudak.eu

Edita Pelantová
Czech Technical University in Prague and *Doppler Institute for Mathematical Physics and Applied Mathematics, Prague*
edita.pelantova@fjfi.cvut.cz

Received: 10/31/08, Accepted: 2/13/09

Abstract

The factor complexity of the infinite word $\mathbf{u}_\beta$ canonically associated with a non-simple Parry number β is studied. Our approach is based on the notion of special factors introduced by Berstel and Cassaigne. At first, we give a handy method for determining infinite left special branches; this method is applicable to a broad class of infinite words which are fixed points of a primitive substitution. In the second part of the article, we focus on infinite words $\mathbf{u}_\beta$ only. To complete the description of their special factors, we define and study (a, b)-maximal left special factors. This enables us to characterize non-simple Parry numbers β for which the word $\mathbf{u}_\beta$ has affine complexity.

1. Introduction

The aim of this work is to compute the *factor complexity function* $\mathcal{C}(n)$ of the infinite word $\mathbf{u}_\beta$ associated with β-expansions [26], where β is a *non-simple Parry number*. The definition of Parry numbers is connected with the Rényi expansion of unity $d_\beta(1)$. Parry numbers are those β for which $d_\beta(1)$ is eventually periodic. Positional numerical systems with a Parry number as a base have a nice behavior. For example, if we consider β-integers, i.e., real numbers with vanishing β-fractional part in their β-expansion, then the distances between two consecutive β-integers take only finitely many values. In fact, this property can be used as an equivalent definition of Parry numbers. In this sense, positional numeration systems based on Parry numbers are a natural generalization of the classical decimal or binary systems. Let us mention that even the innocent looking rational base $\beta = \frac{3}{2}$ brings into numeration systems phenomena never observed before [1].

The most prominent Parry number is the golden mean $\tau = \frac{1+\sqrt{5}}{2}$ with $d_\tau(1) = 11$. The infinite word associated with τ is the famous Fibonacci chain, i.e., the word generated by the substitution $0 \mapsto 01$ and $1 \mapsto 0$. The Fibonacci chain codes the distances between τ-integers. Fabre in [14] showed that for any Parry number there exists a canonical substitution over a finite alphabet such that its unique fixed point u_β represents the distribution of β-integers on the real line.

We acknowledge financial support by grants MSM6840770039 & LC06002 of Ministry of Education, Youth, and Sports of the Czech Republic and by grant 201/09/0584 of the Czech Science Foundation.

β-integers attracted attention of physicists after the discovery of quasicrystals in 1982 [27]. The τ-integers were shown to be a suitable tool for describing atomic positions in solid materials with long range order and non-crystalographical five-fold symmetry [22], [4]. The knowledge of the factor complexity of the Fibonacci chain is the first step towards the description of variability of local configurations in quasicrystals [21].

Parry numbers are split into two groups: a Parry number β is called *simple* if the Rényi expansion of unity $d_\beta(1)$ has only a finite number of nonzero elements, otherwise β is non-simple. The questions concerning the factor complexity of words $\mathbf{u}_\beta$ associated with simple Parry numbers were discussed in [17] and [5]. Of course, since among the $\mathbf{u}_\beta$ one can find some Sturmian sequences and Arnoux-Rauzy words, the complexity of $\mathbf{u}_\beta$ for some specific values of β was known earlier.

The first non-simple Parry number β for which the factor complexity of $\mathbf{u}_\beta$ was precisely determined is such that $d_\beta(1) = 2(01)^\omega$, i.e., β is a root of the polynomial $x^3 - 2x^2 - x + 1$. This non-simple Parry number appears naturally when describing the model of quasicrystals with seven-fold symmetry [16]. The first attempt to study the factor complexity of $\mathbf{u}_\beta$ for a broader class of non-simple Parry numbers can be found in [18].

Since any infinite word $\mathbf{u}_\beta$ is the fixed point of a primitive substitution, the factor complexity of $\mathbf{u}_\beta$ can be estimated from above by a linear function, see [25]. Moreover, we know that the first difference of complexity is bounded by a constant [10, 23]. Nevertheless, in general, it is hard to find an explicit formula for the complexity function of an infinite word $\mathbf{u}$ and it seems it holds also for the case of $\mathbf{u}_\beta$. However, we are able to find all *left special factors* that, in a certain sense, completely determine the factor complexity. The notion of (right) special factor was introduced by Berstel [6] in 1980 and considerably enhanced by Cassaigne in his paper [11] in 1997. We introduce another slight enhancement, a tool that will help us to identify all *infinite left special branches* of fixed point of substitutions satisfying some natural assumption. Further, the knowledge of the structure of left special factors will allow us to identify all non-simple Parry numbers β for which the complexity of $\mathbf{u}_\beta$ is affine: The complexity of $\mathbf{u}_\beta$ is affine if and only if $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$ (Theorem 60).

2. Parry Numbers and Associated Infinite Words

For each $x \in [0,1)$ and for each $\beta > 1$, using a greedy algorithm, one can obtain the unique *β-expansion* $(x_i)_{i\geq 1}$, $x_i \in \mathbb{N}$, $0 \leq x_i < \beta$, of the number x such that

$$x = \sum_{i\geq 1} x_i \beta^{-i} \quad \text{and} \quad \sum_{i\geq k} x_i \beta^{-i} < \beta^{-k+1}.$$

By shifting, each non-negative number has a β-expansion. For $x \in [0,1)$, the β-expansion can also be computed by using the piecewise linear map $T_\beta : [0,1] \to [0,1)$

defined as

$$T_\beta(x) = \{\beta x\},$$

where $\{\beta x\}$ is the fractional part of the real number βx. The sequence $d_\beta(x) = x_1x_2x_3\cdots$ is obtained by iterating T_β with

$$x_i = \lfloor \beta T_\beta^{i-1}(x) \rfloor.$$

If we put $x = 1$ in the previous formula and denote $\lfloor \beta T_\beta^{i-1}(1) \rfloor$ by t_i, then we obtain the sequence $d_\beta(1) = t_1t_2\cdots$ which is called the *Rényi expansion of unity*. Parry [24] showed that $d_\beta(1)$ plays a very important role in the theory of β-numeration. Among other things, it allows us to define Parry numbers.

Definition 1. A real number $\beta > 1$ is said to be a Parry number if $d_\beta(1)$ is eventually periodic. In particular,

(a) if $d_\beta(1) = t_1\cdots t_m$ is finite, i.e., it ends in infinitely many zeros, then β is a *simple Parry number*,

(b) if it is not finite, i.e., $d_\beta(1) = t_1\cdots t_m(t_{m+1}\cdots t_{m+p})^\omega$, then β is called a *non-simple Parry number*.

Note, that the parameters $m, p > 0$ are taken the least possible. It implies that $t_m \neq t_{m+p}$ which will be a very important fact. Another crucial property of $d_\beta(1)$ is the following *Parry condition* [24] valid for all $\beta > 1$:

$$t_jt_{j+1}t_{j+2}\cdots \quad \prec \quad t_1t_2t_3\cdots \qquad \text{for every } j > 1\,, \tag{1}$$

where $\prec$ is the (strict) lexicographical ordering. In particular, notice the important fact that $t_1 > 0$.

As the infinite word $\mathbf{u}_\beta$ is tightly connected with a geometrical interpretation of β-integers, we first introduce β-integers along with some of their properties.

Definition 2. *The real number x is a β-*integer *if the β-expansion of $|x|$ is of the form $\sum_{i=0}^k a_i\beta^i$, where $a_i \in \mathbb{N}$. The set of all β-integers is denoted by $\mathbb{Z}_\beta$.*

The definition of β-integers coincides with the definition of classical integers in the case of β in $\mathbb{Z}$. But there are several new phenomena linked with the notion of β-integers when β is not an integer. For our purposes, the most interesting difference between classical integers and β-integers is the difference in their distribution on the real line. While the classical integers are distributed equidistantly, i.e., gaps between two consequent integers are always of the same length 1, the lengths of gaps between β-integers can take their values even in an infinite set. More precisely, Thurston [28] proved the following theorem.

Theorem 3. *Let $\beta > 1$ be a real number and $d_\beta(1) = (t_i)_{i\geq 1}$. Then the length of gaps between neighbors in $\mathbb{Z}_\beta$ takes values in the set $\{\triangle_0, \triangle_1, \ldots\}$, where*

$$\triangle_i = \sum_{k\geq 1} \frac{t_{k+i}}{\beta^k}, \quad \textit{for } i \in \mathbb{N}.$$

Corollary 4. *The set of lengths of gaps between two consecutive β-integers is finite if and only if β is a Parry number. Moreover, if β is a simple Parry number, i.e., $d_\beta(1) = t_1 \cdots t_m$, the set reads $\{\triangle_0, \triangle_1, \ldots \triangle_{m-1}\}$, if β is a non-simple Parry number, i.e., $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$, we obtain $\{\triangle_0, \triangle_1, \ldots \triangle_{m+p-1}\}$.*

Now, let us suppose that we have drawn the non-negative β-integers on the real line and assume that β is a Parry number. If we read the length of gaps from zero to the right, we obtain an infinite sequence, say $\{\triangle_{i_k}\}_{k\geq 0}$. Further, if we read only indices, we obtain an infinite word over the alphabet $\{0, \ldots, m-1\}$ in the case of simple Parry numbers, and over the alphabet $\{0, \ldots, m+p-1\}$ in the non-simple case. The obtained infinite word is just the word $\mathbf{u}_\beta$ we are interested in. However, there exists another way to define it. Fabre [14] proved that $\mathbf{u}_\beta$ can be defined as the unique fixed point of a substitution φ_β canonically associated with a Parry number β and defined as follows.

Definition 5. For a simple Parry number β the canonical substitution φ_β over the alphabet $\mathcal{A} = \{0, 1, \ldots, m-1\}$ is defined by

$$\begin{array}{rcl}
\varphi_\beta(0) & = & 0^{t_1}1, \\
\varphi_\beta(1) & = & 0^{t_2}2, \\
 & \vdots & \\
\varphi_\beta(m-2) & = & 0^{t_{m-1}}(m-1), \\
\varphi_\beta(m-1) & = & 0^{t_m}.
\end{array}$$

Definition 6. For a non-simple Parry number β the canonical substitution φ_β over the alphabet $\mathcal{A} = \{0, 1, \ldots, m+p-1\}$ is defined by

$$\begin{array}{rcl}
\varphi_\beta(0) & = & 0^{t_1}1, \\
\varphi_\beta(1) & = & 0^{t_2}2, \\
 & \vdots & \\
\varphi_\beta(m-1) & = & 0^{t_m}m, \\
\varphi_\beta(m) & = & 0^{t_{m+1}}(m+1), \\
 & \vdots & \\
\varphi_\beta(m+p-2) & = & 0^{t_{m+p-1}}(m+p-1), \\
\varphi_\beta(m+p-1) & = & 0^{t_{m+p}}m.
\end{array}$$

We see that the definition of φ_β is given by $d_\beta(1)$ and that the only difference between simple and non-simple cases lies in the images of the last letters $m-1$ and $m+p-1$ respectively. While in the simple case the last letters of images $\varphi_\beta(k), k = 0, 1, \ldots, m-1$, are all distinct and so the images form a suffix-free code, in the non-simple case either $\varphi_\beta(m-1) = 0^{t_m}m$ is a suffix of $\varphi_\beta(m+p-1) = 0^{t_{m+p}}m$ or vice versa. As we will see later on, this property is crucial from the point of view of computing the complexity of the infinite word $\mathbf{u}_\beta$.

Definition 7. *Let $\beta > 1$ be a Parry number. The unique fixed point of the canonical substitution φ_β is denoted by*

$$\mathbf{u}_\beta = \lim_{n\to\infty} \varphi_\beta^n(0) = \varphi_\beta^\infty(0).$$

The uniqueness of $\mathbf{u}_\beta$ follows from the definitions of φ_β, the letter 0 is the only admissible starting letter of a fixed point because $t_1 > 0$.

3. Special Factors and Factor Complexity

In this section, we will recall the notion of special factors of an arbitrary infinite word and we will explain how the structure of special factors of an infinite word determines its factor complexity. To be able to do it, we need some usual basic notation, see [11] for more.

Definition 8. *Let $\mathcal{A} = \{0, 1, \ldots, q-1\}, q \geq 1$, be a finite alphabet. An* infinite word *over the alphabet $\mathcal{A}$ is a sequence $\mathbf{u} = (u_i)_{i\geq 1}$ where $u_i \in \mathcal{A}$ for all $i \geq 1$. If $v = u_j u_{j+1} \cdots u_{j+n-1}$, $j, n \geq 1$, then v is said to be a* factor *of $\mathbf{u}$ of length n and the index j is an* occurrence *of v, the empty word ϵ is the factor of length 0.*

By $\mathcal{L}_n(\mathbf{u})$ we denote the set of all factors of $\mathbf{u}$ of length $n \in \mathbb{N}$, the language *of $\mathbf{u}$ is then the set $\mathcal{L}(\mathbf{u}) = \bigcup_{n\in\mathbb{N}} \mathcal{L}_n(\mathbf{u})$.*

Definition 9. *Let $\mathbf{u}$ be an infinite word over an alphabet $\mathcal{A}$. The function $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$ is the* factor complexity function *of $\mathbf{u}$. We further define the first difference of the complexity by $\triangle\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n)$.*

In what follows, we shall restrict ourself to those infinite words which are fixed point of some substitution (morphism) φ defined over a finite alphabet $\mathcal{A}$. We shall further assume that φ is injective and primitive.

Definition 10. *A substitution φ is* primitive *if there exists $k \in \mathbb{N}$ such that for all $a, b \in \mathcal{A}$ the word $\varphi^k(a)$ contains b.*

Equivalently, φ is primitive if the incidence matrix M_φ is primitive.

There are several well-known properties of the complexity function $\mathcal{C}$.

Proposition 11. *The following hold:*

(i) *For each infinite word* $\mathbf{u}$, $0 \leq \mathcal{C}(n) \leq (\#\mathcal{A})^n$,

(ii) *if* $\mathbf{u}$ *is eventually periodic, then* $\mathcal{C}(n)$ *is eventually constant,*

(iii) $\mathbf{u}$ *is aperiodic if and only if* $\mathcal{C}(n)$ *is unbounded and* $\mathcal{C}(n)$ *is unbounded if and only if* $\triangle\mathcal{C}(n) \geq 1$, *for all* $n \in \mathbb{N}$,

(iv) *if* $\mathbf{u}$ *is a fixed point of a primitive substitution, then* $\mathcal{C}(n)$ *is a sublinear function, i.e.,* $\mathcal{C}(n) \leq an + b$, *for some* $a, b \in \mathbb{N}$,

(v) *if* $\mathbf{u}$ *is a fixed point of a primitive substitution, then* $\triangle\mathcal{C}(n)$ *is bounded.*

Items $(i) - (iii)$ are obvious, (iv) is due to [25], (v) was proved in [23] and in a more general context in [10].

It is also well known that any fixed point of a primitive substitution is uniformly recurrent, i.e., each factor occurs infinitely many times and the gaps between its two consecutive occurrences are bounded in length. This implies that each factor is extendable both to the right and to the left.

Definition 12. *Let* v *be a factor of* $\mathbf{u}$, *the set of* left extensions *of* v *is defined as*

$$\mathrm{Lext}(v) = \{a \in \mathcal{A} \mid av \in \mathcal{L}(\mathbf{u})\}.$$

If $\#\mathrm{Lext}(v) \geq 2$, *then* v *is said to be a* left special (LS) factor *of* $\mathbf{u}$.

In the analogous way we define the set of right extensions $\mathrm{Rext}(\mathbf{u})$ *and a* right special (RS) factor. *If* v *is both left and right special, then it is called* bispecial.

The connection between (left) special factors and the complexity follows from the following reasoning. Let us suppose that $\mathcal{L}_n(\mathbf{u}) = \{v_1, \ldots, v_k\}, k \geq 1$, and let $\mathrm{Lext}(v_i) = \{a_1^{(i)}, \ldots, a_{\ell_i}^{(i)}\}$, $\ell_i \geq 1, i = 1, \ldots, k$. Now, it is not difficult to realize that

$$\mathcal{L}_{n+1}(\mathbf{u}) = \{a_1^{(1)}v_1, \ldots, a_{\ell_1}^{(1)}v_1, a_1^{(2)}v_2, \ldots, a_{\ell_{k-1}}^{(k-1)}v_{k-1}, a_1^{(k)}v_k, \ldots, a_{\ell_k}^{(k)}v_k\},$$

i.e., by concatenating all factors of length n and all their left extensions we obtain all factors of length $n+1$. It implies that

$$\#\mathcal{L}_{n+1}(\mathbf{u}) - \#\mathcal{L}_n(\mathbf{u}) = \triangle\mathcal{C}(n) = \sum_{\substack{v \in \mathcal{L}_n(\mathbf{u}) \\ v \text{ is LS}}} (\#\mathrm{Lext}(v) - 1). \tag{2}$$

Hence, if we know all LS factors along with the number of their left extensions, we are able to evaluate the complexity $\mathcal{C}(n)$ using this formula.

3.1. Classification of LS factors

Let $a, b \in \mathrm{Lext}(v)$ be left extensions of a factor v of $\mathbf{u}$, which means both av and bv are factors of $\mathbf{u}$. If there exists a letter $c \in \mathrm{Rext}(av) \cap \mathrm{Rext}(bv)$, we say that v can be extended to the right such that it remains LS with left extensions a, b; indeed $a, b \in \mathrm{Lext}(vc)$.

Definition 13. *Let $a, b \in \mathrm{Lext}(v)$ be distinct left extensions of an LS factor v of $\mathbf{u}$. We say that v is an (a, b)-*maximal LS factor *if $\mathrm{Rext}(av) \cap \mathrm{Rext}(bv) = \emptyset$; in words, v cannot be extended to the right such that it remains LS with left extensions a, b.*

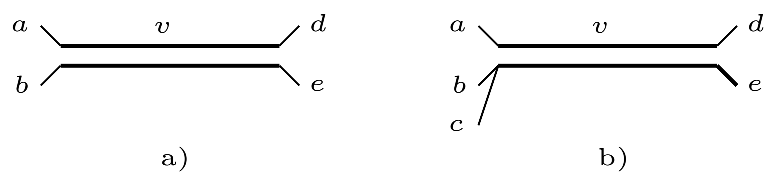

Figure 1: Two types of (a, b)-maximal LS factor v.

In general, there are two types of (a, b)-maximal LS factors both depicted in Figure 1. In Case a), a and b are the only left extensions of v and so v cannot be extended to the right and remain LS. In Case b), v can be prolonged by letter e such that ve is still an LS factor but it looses its left extension a.

It can also happen that a factor v with left extensions a and b is extendable to the right infinitely many times. In this way we obtain an infinite LS branch.

Definition 14. *An infinite word $\mathbf{w}$ is an* infinite LS branch *of $\mathbf{u}$ if each prefix of $\mathbf{w}$ is a LS factor of $\mathbf{u}$. We put*

$$\mathrm{Lext}(\mathbf{w}) = \bigcap_{v \text{ prefix of } \mathbf{w}} \mathrm{Lext}(v).$$

Clearly, we have that $\#\mathrm{Lext}(\mathbf{w}) \geq 2$ since each prefix of infinite LS branch $\mathbf{w}$ is an LS factor having at least two left extensions.

Proposition 15. *The following hold:*

(i) *If $\mathbf{u}$ is eventually periodic, then there is no infinite LS branch of $\mathbf{u}$,*

(ii) *if $\mathbf{u}$ is aperiodic, then there exists at least one infinite LS branch of $\mathbf{u}$,*

(iii) *if $\mathbf{u}$ is a fixed point of a primitive substitution, then the number of infinite LS branches is bounded.*

Item (i) is obvious, and (iii) is a direct consequence of (2) and Proposition 11 (v). Item (ii) is a direct consequence of the famous König's infinity lemma [20] applied on sets $V_1, V_2, \ldots$, where the set V_k comprises all LS factors of length k and where $v_1 \in V_i$ is connected by an edge with $v_2 \in V_{i+1}$ if v_1 is a prefix of v_2.

Taking all this together, our aim is to find all (a, b)-maximal LS factors and also all infinite LS branches of $\mathbf{u}$.

Remark 16. The term "special factor" (for us it was RS factor) was introduced in 1980 [6] and it has been used for computing the factor complexity since then (eg. [7], [13]). The notations introduced above are based on Cassaigne's article [11]. An (a, b)-maximal factor is a new term; actually it is a special case of a *weak bispecial factor* proposed there. It is also shown in the article that bispecial factors determine the second difference of the complexity similarly to the way that LS factors determine the first difference of the complexity.

Remark 17. *Everything that has been (and will be) defined or shown for LS factors can be defined or shown similarly for RS factors.*

3.2. How To Find Infinite LS Branches

Before introducing a new notion, let us consider the example substitution

$$\varphi : 1 \mapsto 1211, 2 \mapsto 311, 3 \mapsto 2412, 4 \mapsto 435, 5 \mapsto 534 \tag{3}$$

with $\mathbf{u} = \varphi^\infty(1)$. Further, let w be an LS factor (or infinite LS branch) of $\mathbf{u}$ with left extensions 1 and 2. Is $\varphi(w)$ again an LS factor? From Figure 2 (first line) we see that it is not since the letter 1 is its only left extension. In order to obtain an LS factor, we have to append as a prefix the factor 11 which is the longest common suffix of $\varphi(1) = 1211$ and $\varphi(2) = 311$, and then $11\varphi(w)$ is an LS factor with left extensions 2 and 3. In the case when $\mathrm{Lext}(w) = \{2, 3\}$ (second line in Figure 2), $\varphi(w)$ is an LS factor since the longest common suffix of $\varphi(2) = 311$ and $\varphi(3) = 2412$ is the empty word ϵ.

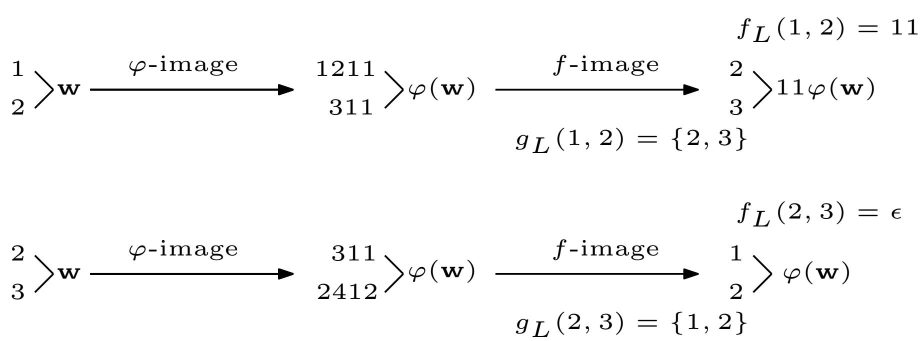

Figure 2: Images of LS factors.

Definition 18. Let φ be a substitution defined over an alphabet $\mathcal{A}$. For each pair of distinct letters $a, b \in \mathcal{A}$ we define $f_L(a,b)$ as the longest common suffix of words $\varphi(a)$ and $\varphi(b)$.

Definition 19. If v be a prefix of a word w, then $v^{-1}w$ is the word w without the prefix v. If v is a suffix of w, we define wv^{-1} analogously.

Definition 20. Let φ be an injective substitution defined over an alphabet $\mathcal{A}$ having a fixed point $\mathbf{u}$. For each unordered pair of distinct letters $a, b \in \mathcal{A}$ such that $\mathrm{Rext}(a) \cap \mathrm{Rext}(b) \neq \emptyset$ we define the set $g_L(a,b)$ as follows.

(i) If $f_L(a,b)$ is a proper suffix of both $\varphi(a)$ and $\varphi(b)$, then $g_L(a,b)$ contains just the last letters of factors $\varphi(a)(f_L(a,b))^{-1}$ and $\varphi(b)(f_L(a,b))^{-1}$.

(ii) If $f_L(a,b) = \varphi(a)$ (i.e., W.L.O.G., $|\varphi(a)| < |\varphi(b)|$), then $g_L(a,b)$ contains the last letter of the factor $\varphi(b)(f_L(a,b))^{-1}$ and all the last letters of factors $\varphi(c)$, where $c \in \mathrm{Lext}(a)$ such that $\mathrm{Rext}(ca) \cap \mathrm{Rext}(b) \neq \emptyset$.

Assumption 21. *Let* $\mathbf{u}$ *be a fixed point of an injective substitution* φ. *We suppose that for all* $a, b \in \mathcal{A}$ *the number of elements of* $g_L(a,b)$ *is two, whenever* $g_L(a,b)$ *is defined.*

Moreover, if $f_L(a,b) = \varphi(a)$, *then the longest common suffix of* $\varphi(ca)$ *and* $\varphi(b)$ *is* $\varphi(a)$ *for all* $c \in \mathcal{A}$ *such that* $\mathrm{Rext}(ca) \cap \mathrm{Rext}(b) \neq \emptyset$.

Assumption 21 is valid for all suffix-free substitutions since $g_L(a,b)$ from point (i) of Definition 20 always contains just two elements, and the case when $f_L(a,b) = \varphi(a)$ never happens. If $f_L(a,b) = \varphi(a)$, then Assumption 21 says that if v is an LS factor with $\mathrm{Lext}(v) = \{a, b\}$, then the last letter e of $\varphi(c)$ is the same for all $c \in \mathrm{Lext}(av)$ and, moreover, $e\varphi(a)$ is not a suffix of $\varphi(b)$ – in other words, for each LS factor v the factor $f_L(a,b)\varphi(v)$ is again LS. We will see that this complicated assumption is satisfied for the (not suffix-free) substitution φ_β, where β is a non-simple Parry number.

Definition 22. Let φ be a substitution satisfying Assumption 21. Then for each LS factor (or infinite LS branch) w having distinct left extensions a and b we define *f-image* of w as the factor $f_L(a,b)\varphi(w)$.

With respect to the preceding discussion, Assumption 21 says that f-image is always an LS factor and it has just two left extensions, namely two elements of $g_L(a,b)$, corresponding to the two original left extensions a and b.

Assumption 21 along with the notation introduced above allows us to define the following graph.

Definition 23. Let φ be a substitution defined over an alphabet $\mathcal{A}$ satisfying Assumption 21. We define a directed labelled graph GL_φ as follows:

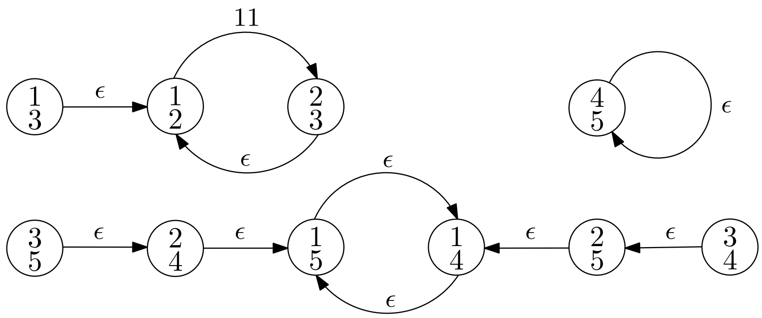

Figure 3: The graph GL_φ for the Substitution (3).

(i) vertices of GL_φ are unordered pairs of distinct letters a, b such that $\mathrm{Rext}(a) \cap \mathrm{Rext}(b) \neq \emptyset$,

(ii) if $g_L(a,b) = \{c,d\}$, then there is an edge from vertex (a,b) to vertex (c,d) labelled by $f_L(a,b)$.

In fact, the crucial result of Assumption 21 is that out-degree of each vertex is exactly one. The graph GL_φ for our example substitution is drawn in Figure 3, this substitution satisfies Assumption 21 for it is suffix-free.

Now, let us consider the case when $\mathbf{w}$ is an infinite LS branch with $a, b \in \mathrm{Lext}(\mathbf{w}), a \neq b$. Obviously, f-image of $\mathbf{w}$ is uniquely given for this (a,b). For primitive substitutions even an "f-preimage" of each infinite LS branch exists, it is a direct consequence of the fact proved by Mossé [23, Theorem 2] that a primitive substitution with an aperiodic fixed point is recognizable.

Definition 24. Let φ be a primitive substitution on a finite alphabet A and let $\mathbf{u} = \mathbf{u}_0\mathbf{u}_1\cdots$ be its aperiodic fixed point. Define $f(n) = |\varphi(\mathbf{u}_0\cdots\mathbf{u}_n)|$. Then φ is *recognizable* if there exists a *context length* $L > 0$ such that for any factor w of $\mathbf{u}$ of length at least $2L$ there exist i, j with $0 \leq i \leq L, |w| - L \leq j \leq |w|$ and unique factor v such that $\varphi(v) = \mathbf{u}_i \cdots \mathbf{u}_j$ and whenever $\mathbf{u}_m \cdots \mathbf{u}_{m+|w|-1} = w$, then there exist i', j' such that $f(i') = m + i, f(j') = m + j$ and $\mathbf{u}_{i'} \cdots \mathbf{u}_{j'} = v$.

In words, φ is recognizable if each sufficiently long factor w of its fixed point (possibly without a prefix and suffix of bounded length) has a unique decomposition into words $(\varphi(a))_{a\in\mathcal{A}}$, i.e., there exists a "central part" of w which has a unique φ-preimage.

In a certain context, the notion of recognizable substitutions coincides with the notion of *circular DOL-languages*; for details see e.g. [9].

Lemma 25. *Let an infinite word* $\mathbf{u}$ *be a fixed point of a primitive substitution* φ *satisfying Assumption 21. Then for each infinite LS branch* $\mathbf{w}$ *of* $\mathbf{u}$ *with* $a, b \in \mathrm{Lext}(\mathbf{w})$,

$a \neq b$, there exists at least one infinite LS branch $\overline{\mathbf{w}}$ with $c, d \in \mathrm{Lext}(\overline{\mathbf{w}}), c \neq d$, such that f-image $f_L(c,d)\varphi(\overline{\mathbf{w}})$ of $\overline{\mathbf{w}}$ equals $\mathbf{w}$ and $g_L(c,d) = \{a,b\}$.

Proof. Let L be the context length of φ. Since each prefix of $\mathbf{w}$ is a factor of $\mathbf{u}$, the recognizability of φ implies that there exists $0 \leq i \leq L$ such that $\varphi(\overline{\mathbf{w}}) = \mathbf{w}_i\mathbf{w}_{i+1}\cdots$, where $\overline{\mathbf{w}}$ is a unique infinite word whose each prefix is a factor of $\mathbf{u}$. If i is taken the least possible, then $\overline{\mathbf{w}}$ is an infinite LS branch as well as $\mathbf{w}$. Moreover, due to Assumption 21, there exist distinct $c, d \in \mathrm{Lext}(\overline{\mathbf{w}})$ such that $\mathbf{w} = f_L(c,d)\varphi(\overline{\mathbf{w}})$ and $\{a,b\} = g_L(c,d)$. □

Theorem 26. *Let $\mathbf{u}$ be a fixed point of a primitive injective substitution φ satisfying Assumption 21 and let $\mathbf{w}$ be an infinite LS branch with $a, b \in \mathrm{Lext}(\mathbf{w}), a \neq b$. Then either $\mathbf{w}$ is a* periodic point *of φ, i.e.,*

$$\mathbf{w} = \varphi^{\ell}(\mathbf{w}) \quad \textit{for some } \ell \geq 1, \tag{4}$$

and (a,b) is a vertex of a cycle in GL_φ labelled by ϵ only or $\mathbf{w} = s\varphi^{\ell}(s)\varphi^{2\ell}(s)\cdots$ is the unique solution of the equation

$$\mathbf{w} = s\varphi^{\ell}(\mathbf{w}), \tag{5}$$

where (a,b) is a vertex of a cycle in GL_φ containing at least one edge with a non-empty label, ℓ is the length of this cycle and

$$s = f_L(g_L^{\ell-1}(a,b))\cdots\varphi^{\ell-2}(f_L(g_L(a,b))\varphi^{\ell-1}(f_L(a,b)). \tag{6}$$

Proof. Due to Assumption 21 and Lemma 25, both the f-image and the f-preimage of $\mathbf{w}$ exist and they are unique. The uniqueness of f-preimage follows from the fact that the number of infinite LS branches is finite (Proposition 15 (*iii*)). Thus, f-image is a one-to-one mapping on the finite set of all ordered pairs

$$\{((c,d), \overline{\mathbf{w}})\},$$

where $\overline{\mathbf{w}}$ is an infinite LS branch of $\mathbf{u}$ and (c,d) is an unordered pair of letters such that $c, d \in \mathrm{Lext}(\overline{\mathbf{w}})), c \neq d$. The f-image can be viewed as a permutation on this finite set and so it decomposes the set to independent cycles as depicted in Figure 4.

Let us consider separately two cases:

(a) The vertex (a,b) is a vertex of a cycle in GL_φ of length k labelled by ϵ only. In this case, the f-image coincides with φ and thus all the infinite words $\mathbf{w}, \varphi^k(\mathbf{w}), \varphi^{2k}(\mathbf{w}), \ldots$ are infinite LS branches with left extensions a and b. Since the number of infinite LS branches is finite and each of them has a unique f-image and f-preimage, this sequence of infinite LS branches is periodic and there must exist $m \geq 1$ such that $\mathbf{w} = \varphi^{mk}(\mathbf{w})$ and hence $\mathbf{w}$ is a periodic point of order ℓ, where ℓ divides mk (see also Remark 27).

(b) The vertex (a, b) is a vertex of a cycle of length k with at least one edge labelled by a non-empty word. We prove that $\ell = k$. Putting $\mathbf{w}^{(1)} = \mathbf{w}$, after applying f-image on $\mathbf{w}^{(1)}$ k times, we obtain an infinite LS branch $\mathbf{w}^{(2)} = s\varphi^k(\mathbf{w}^{(1)})$ again having left extensions a and b ($s \neq \epsilon$ is given by (6) for $\ell = k$). Continuing this way, we obtain a sequence of infinite LS branches defined by the equations $\mathbf{w}^{(m+1)} = s\varphi^k(\mathbf{w}^{(m)}), m = 1, 2, \ldots$. This sequence is periodic for the same reason as the sequence in case (a) and hence these equations have a unique solution, namely the constant sequence $\mathbf{w}^{(m)} = s\varphi^k(s)\varphi^{2k}(s)\cdots$, $m = 1, 2, \ldots$.

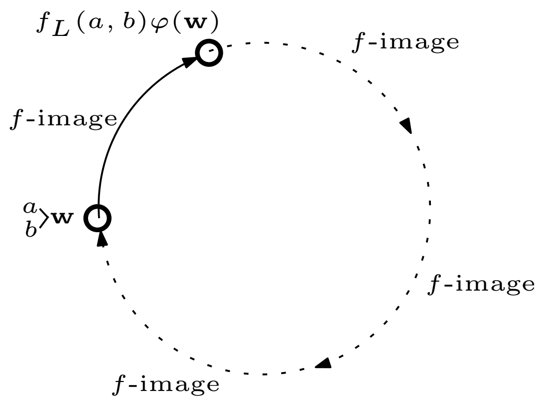

Figure 4: Circular structure of infinite LS branches.

□

Our example substitution φ (see (3)) has five periodic points

$$\varphi^\infty(1), \varphi^\infty(4), \varphi^\infty(5), (\varphi^2)^\infty(2), (\varphi^2)^\infty(3).$$

It is an easy exercise to show that $\mathrm{Lext}(1) = \{1, 2, 3, 4, 5\}, \mathrm{Lext}(2) = \{1, 4, 5\}$, $\mathrm{Lext}(3) = \{1, 4, 5\}, \mathrm{Lext}(4) = \{1, 2, 3\}$, and $\mathrm{Lext}(5) = \{1, 2, 3\}$. Looking at the graph GL_φ depicted in Figure 3, we see that $\varphi^\infty(4), \varphi^\infty(5)$ are not infinite LS branches, as none of the vertices, $(1, 2), (2, 3)$ and $(1, 3)$, is a vertex of a cycle labelled by ϵ only. Hence, only $\varphi^\infty(1), (\varphi^2)^\infty(2), (\varphi^2)^\infty(3)$ are infinite LS branches with left extensions $1, 4, 5$.

As for infinite LS branches corresponding to Equation (5), in the case of our example, there is only one cycle not labelled by the empty word: between vertices $(1, 2)$ and $(2, 3)$. There are two (= the length of the cycle) equations corresponding to this cycle:

$$\mathbf{w} = \varphi(11)\varphi^2(\mathbf{w}) \quad \text{and} \quad \mathbf{w} = 11\varphi^2(\mathbf{w}).$$

They give us two infinite LS branches

$$\varphi(11)\varphi^3(11)\varphi^5(11)\cdots,$$
$$11\varphi^2(11)\varphi^4(11)\cdots,$$

the former having left extensions 1 and 2 and the latter 2 and 3.

Remark 27. Consider an infinite LS branch of a fixed point of a substitution ψ having two distinct left extensions a and b, and let (a, b) be a vertex of a cycle of length k in GL_ψ whose edges are all labelled by ϵ. Regarding the relation between the length k of the cycle and the integer ℓ from (4), ℓ can be equal to, greater or less than k.

In the case of our example substitution φ, $\ell = 1$ and $k = 2$ for the vertex $(1, 4)$ and the infinite LS branch $\varphi^\infty(1)$, $\ell = 2$ and $k = 2$ for the vertex $(1, 4)$ and the infinite LS branch $(\varphi^2)^\infty(2)$ and, finally, $\ell = 2$ and $k = 1$ for the vertex $(4, 5)$ and the infinite LS branch $(\varphi^2)^\infty(2)$.

It can even happen that ℓ is not a multiple of k: Consider the substitution

$$\psi : 1 \mapsto 243, 2 \mapsto 32, 3 \mapsto 121, 4 \mapsto 41.$$

The graph GL_ψ contains two cycles all of whose edges have label ϵ, namely, the loop at vertex $(1, 3)$ and the cycle between vertices $(1, 2)$ and $(2, 3)$. There are no other cycles in GL_ψ, i.e., the only candidates for infinite LS branches are the periodic points $(\psi^3)^\infty(1), (\psi^3)^\infty(2)$ and $(\psi^3)^\infty(3)$. Using the same steps as in the case of the example substitution above, we can show that $(\psi^3)^\infty(1)$ is an infinite LS branch such that 2 and 3 are its left extensions. Hence, for this pair, we have $\ell = 3$ and $k = 2$.

Remark 28. Assumption 21 could be reformulated into a weaker form but to do so, it would require the introduction of rather complicated notation. The important fact here is that the canonical substitution φ_β satisfies Assumption 21.

4. Infinite LS Branches of $\mathbf{u}_\beta$

At first, let us recall known results for simple Parry numbers. The substitution φ_β from Definition 5 is suffix-free and it implies that it satisfies Assumption 21. As mentioned earlier, the last letters of images of letters are all distinct and so $f_L(a, b) = \epsilon$ for all pairs $a, b \in \mathcal{A}$. The graph GL_{φ_β} then looks as in Figure 5. It contains $m - 1$ cycles labelled by ϵ only and hence the only candidate for being an infinite LS branch is the unique fixed (and periodic) point of φ_β, namely $\mathbf{u}_\beta$ with $\mathrm{Lext}(\mathbf{u}_\beta) = \mathcal{A}$. The same result is proved in [17] using different techniques.

Theorem 29. ([5],[17]). *Let $\beta > 1$ be a simple Parry number with $d_\beta(1) = t_1 \cdots t_m$ and let $\mathbf{u}_\beta$ be the fixed point of the canonical substitution φ_β given by Definition 5. Then*

(i) *if $t_1 = t_2 = \cdots = t_{m-1}$ or $t_1 > \max\{t_2, \ldots, t_{m-1}\}$, the exact value of $\mathcal{C}(n)$ is known [17],*

(ii) *in particular, $(m - 1)n + 1 \le \mathcal{C}(n) \le mn$, for all $n \ge 1$,*

(iii) *$\mathcal{C}(n)$ is affine if and only if the following two conditions are fulfilled:*

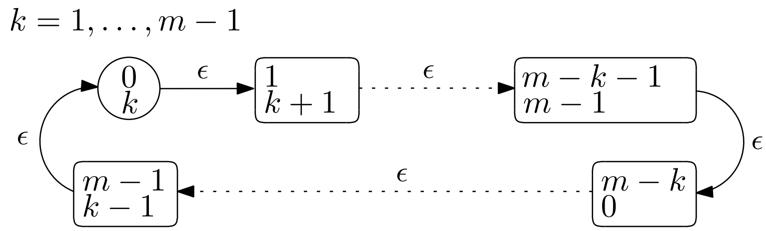

Figure 5: GL_{φ_β} for Simple Parry β.

(1) $t_m = 1$,
(2) *for all* $i = 2, 3, \ldots, m-1$ *we have*

$$t_i t_{i+1} \cdots t_{m-1} t_1 \cdots t_{i-1} \quad \preceq \quad t_1 t_2 \cdots t_{m-1}.$$

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

In this paper, we will find the necessary and sufficient condition for the complexity being affine in the case of non-simple Parry numbers. We will see that it is more restrictive than the one from point (iii).

4.1. Infinite LS Branches in Case of Non-Simple Parry Numbers

In this section, we will apply the hitherto introduced theory on the fixed point $\mathbf{u}_\beta$ of the substitution φ_β, where β is a non-simple Parry number. To be able to do so, we need some more notation and simple but useful technical lemmas.

Definition 30. For all $k, \ell \in \mathbb{N}$, we define an addition $\oplus : \mathbb{N} \times \mathbb{N} \to \mathcal{A}$ as follows:

$$k \oplus \ell := \begin{cases} k + \ell & \text{if } k + \ell < m + p, \\ m + (k + \ell - m \bmod p) & \text{otherwise.} \end{cases}$$

Similarly, if used with parameters t_i, we define for all $k, \ell \in \mathbb{N}, k + \ell > 0$,

$$t_{k \oplus \ell} := \begin{cases} t_{k+\ell} & \text{if } 0 < k + \ell < m + p + 1, \\ t_{m+1+(k+\ell-m-1 \bmod p)} & \text{otherwise.} \end{cases}$$

In fact, the addition $\oplus$ tracks the last letters of the words $\varphi_\beta^n(0), n = 0, 1, \ldots$, or, if used with parameters t_i, the indices of letters in the infinite word $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$. Note that these two cases are not the same, e.g., $(m+p-1) \oplus 1 = m$ but $t_{(m+p-1)\oplus 1} = t_{m+p}$. We can rewrite the definition of the substitution φ in a simpler form

$$\varphi_\beta(k) = 0^{t_{k+1}}(k \oplus 1), \qquad \forall k \in \mathcal{A}.$$

Further, employing the new notation and the definition of the substitution φ_β, one can easily prove the following simple observations.

Lemma 31. *For the substitution φ_β the following hold:*

(i) for all $n \in \mathbb{N}$ and all $k \in \mathcal{A}$

$$\varphi_\beta^n(k) = (\varphi_\beta^{n-1}(0))^{t_{k\oplus 1}}(\varphi_\beta^{n-2}(0))^{t_{k\oplus 2}} \cdots (\varphi_\beta(0))^{t_{k\oplus(n-1)}} 0^{t_{k\oplus n}}(k \oplus n),$$

(ii) if avb is a factor of $\mathbf{u}_\beta$, $v \in \mathcal{A}^*$ and $a, b \neq 0$, then there exists a unique factor v' such that $\varphi_\beta(v') = vb$.

Our aim is to obtain the graph GL_{φ_β}; thus, we need to know left extensions of letters and also all $g_L(a,b)$.

Definition 32. Let us define for all $k \in \mathcal{A}$, $k \neq 0$, a function $z : \{1, \ldots, m+p-1\} \to \{0, 1, \ldots, m+p-2\}$ by

$$z(k) = \max\{j \in \mathbb{N} \mid 0^j \text{ is a suffix of } t_1 t_2 \cdots t_k\}.$$

For $k \in \{m, \ldots, m+p-1\}$ we also define a function $y : \{m, \ldots, m+p-1\} \to \{0, 1, \ldots, p-1\}$ by

$$y(k) = \begin{cases} \max\{j \in \mathbb{N} \mid 0^j \text{ is a suffix of } t_{m+1}t_{m+2}\cdots t_{m+p}t_{m+1}\cdots t_k\} & \text{if } k > m, \\ \max\{j \in \mathbb{N} \mid 0^j \text{ is a suffix of } t_{m+1}t_{m+2}\cdots t_{m+p}\} & \text{if } k = m. \end{cases}$$

Further, we define

$$\ell_0 = \begin{cases} 0 & \text{if } t_1 > 1, \\ 1 + \max\{j \in \mathbb{N} \mid 0^j \text{ is a prefix of } t_2 t_3 \cdots t_m\} & \text{otherwise} \end{cases}$$

and finally we put $t = \min\{t_m, t_{m+p}\}$.

Note that $z(k)$ and $y(k)$ can return different values for $k \geq m$, and a necessary condition for $z(k) \neq y(k)$ is that $t = 0$ and $z(\ell) \neq y(\ell)$ for all $m \leq \ell < k$. Due to the Parry condition (1) we must have $1 \leq \ell_0 \leq m-1$, as the case $d_\beta(1) = 10\cdots 0(t_{m+1}\cdots t_{m+p-1}1)^\omega$ is not admissible.

Lemma 33. *For $\mathbf{u}_\beta$, the fixed point of φ_β, the following hold:*

(i) $\mathrm{Lext}(0) = \{\ell_0, \ldots, m+p-1\} \supset \{m, \ldots, m+p-1\}$,

(ii) $\mathrm{Lext}(k) = \{z(k)\}$, *for $k \in \{1, 2, \ldots, m-1\}$,*

(iii) $\mathrm{Lext}(k) = \{z(k), y(k)\}$, *for $k \in \{m, m+1 \ldots, m+p-1\}$.*

Proof. (ii) Each letter $k > 0$ can appear in $\mathbf{u}_\beta$ in the image of $k-1$, namely $\varphi_\beta(k-1) = 0^{t_k}k$. If $t_k > 0$, then $z(k) = 0 \in \mathrm{Lext}(k)$, if $t_k = 0$, we consider $\varphi_\beta^2(k-2) = \varphi_\beta(0^{t_{k-1}})k = (0^{t_1}1)^{t_{k-1}}k$. Again, if $t_{k-1} > 0$, then $z(k) = 1 \in \mathrm{Lext}(k)$, otherwise we continue in the same way. Since $t_1 > 0$, this process is finite.

(iii) The letter m can appear in $\mathbf{u}_\beta$ not only in the image $\varphi_\beta(m-1)$ (i.e., case (ii)) but also in $\varphi_\beta(m+p-1) = 0^{t_{m+p}}m$. If we realize this other possible origin of occurrences of the letters $m, m+1, \ldots, m+p-1$, then the proof is the same as for (ii).

(i) If $t_1 > 1$, then 00 is a factor of $\mathbf{u}$ and $\ell_0 = 0$. Hence, for all $n \in \mathbb{N}$, the word $\varphi_\beta^n(0)0 = \cdots(0 \oplus n)0$ is a factor of u as well and thus $\mathrm{Lext}(0) = \mathcal{A}$.

If $t_1 = 1$, this implies that $t_i \in \{0,1\}$ for $i = 1, \ldots, m+p$. Since either $\ell_0 = 1$ and $t_2 = 1$ or $\ell_0 > 1$, $t_2 = \cdots = t_{\ell_0} = 0$ and $t_{\ell_0+1} = 1$, we have $\varphi_\beta^{\ell_0}(01) = \varphi_\beta((\ell_0-1)\ell_0) = \ell_0 0(\ell_0+1)$, and hence, $\ell_0, \ell_0+1, \ldots, m+p-1 \in \mathrm{Lext}(0)$. But $d_\beta(1)$ cannot contain a sequence of consecutive 0's shorter than ℓ_0 due to Parry condition (1) and so ℓ_0 is the least letter in $\mathrm{Lext}(0)$. □

The previous lemma allows us to get auxiliary results about prefixes of all LS factors of $\mathbf{u}_\beta$.

Corollary 34. *If v is an LS factor of $\mathbf{u}_\beta$ containing at least one nonzero letter, then one of the following factors is a prefix of v:*

(i) $0^{t_1}1$,

(ii) $0^t m$,

(iii) $0^{t_k}k$, *if $k > m$ and $t = t_{m+1} = t_{m+2} = \cdots = t_{k-1} = 0$.*

Proof. Taking into account the definition of φ_β, each LS factor of $\mathbf{u}_\beta$ containing at least one nonzero letter must begin in $0^{t_k}k, k \in \mathcal{A} \setminus \{0\}$ or $0^{t_{m+p}}m$. Of course, $0^{t_k}k$ is then LS as well. Consider $k \in \mathcal{A}$ different from 1 and m. In order for the factor $0^{t_k}k = \varphi(k-1)$ to be LS, the letter $k-1$ must have at least two distinct left extensions. It means, according to Lemma 33, that $k-1 \geq m$ and $z(k-1) \neq y(k-1)$. Item (iii) then follows from the definition of the functions z and y. Note that the factors from item (iii) are successive images of the factor $0^t m$ in the case when $t = 0$. □

Lemma 35. *For the fixed point $\mathbf{u}_\beta$ of φ_β, the following hold:*

(i) if (k,ℓ) is an unordered pair of distinct letters of $\mathcal{A}$ such that $\mathrm{Rext}(k) \cap \mathrm{Rext}(\ell) \neq \emptyset$, *and $(k,\ell) \neq (m-1, m+p-1)$, then $f_L(k,\ell) = \epsilon$ and $g_L(k,\ell) = \{k \oplus 1, \ell \oplus 1\}$,*

(ii) $f_L(m-1, m+p-1) = 0^t m$ and $g_L(m-1, m+p-1) = \{0, z\}$, where

$$z = \begin{cases} 1 + z(m-1) & \text{if } t_m < t_{m+p}, \\ 1 + z(m+p-1) & \text{if } t_{m+p} < t_m. \end{cases} \tag{7}$$

Proof. (i) follows directly from the definitions of g_L, f_L and φ_β. (ii) is a simple consequence of Lemma 33. Note that if $t_m > t_{m+p} \geq 0$, then $z(m+p-1) = y(m+p-1)$. ☐

This lemma along with Lemma 33 implies the following.

Corollary 36. *The substitution φ_β from Definition 6 satisfies Assumption 21.*

Now, we know all we need to be able to construct the graph GL_{φ_β}. For the case when $t_1 > 1$, the graph is depicted in Figure 6. Since $\mathrm{Lext}(0) = \mathcal{A}$, all possible unordered pairs of letters are vertices of the graph. If z is not a multiple of p (i.e., the decision condition $z = sp$ in Figure 6 returns *no*), then the graph contains only cycles with edges labelled by ϵ. If $z = sp$ for a certain positive integer s, then there is a cycle on the vertices $(0, z), (1, z \oplus 1), \ldots, (m-1, z \oplus m-1)$, where the edge from the vertex $(m-1, z \oplus m-1)$ to the vertex $g_L(m-1, z \oplus m-1) = (0, z)$ is labelled by $f_L(m-1, z \oplus m-1) = 0^t m$.

If $t_1 = 1$, the graph GL_{φ_β} is the same as in Figure 6, but we have to remove vertices (k, ℓ), where $k < \ell_0$ or $\ell < \ell_0$ and $(k, \ell) \neq (0 \oplus n, z \oplus n)$ for any $n \in \mathbb{N}$, because such pairs of letters are not left extensions of any LS factor, i.e., $\mathrm{Rext}(k) \cap \mathrm{Rext}(l) = \emptyset$ (see Definition 23). For our purpose, it is important that the structure of cycles is the same for arbitrary value of t_1.

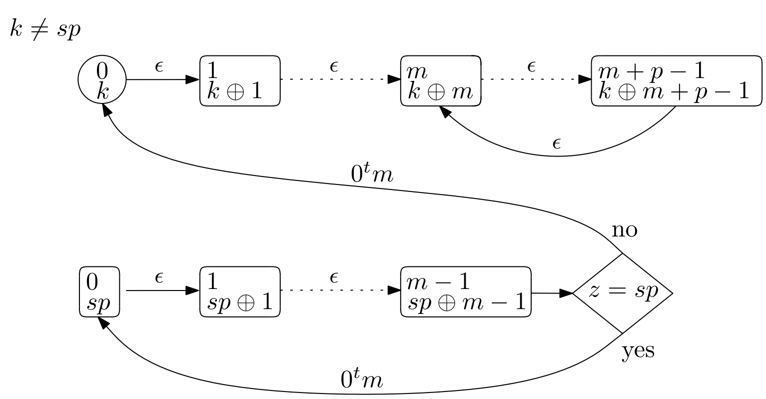

Figure 6: GL_{φ_β} for non-simple Parry β, s is a positive integer.

Since the fact whether z is or is not a multiple of p is crucial for the structure of cycles in GL_{φ_β}, we introduce the following set.

Definition 37. A non-simple Parry number $\beta > 1$ is an element of the set $\mathcal{S}$ if and only if there exists a positive integer s such that $z = sp$, where z is defined in (7).

Note that if $p = 1$, then $\beta \in \mathcal{S}$.

Employing Lemmas 33 and 35 and Definition 32, one can easily prove the following.

Lemma 38. *A non-simple Parry number $\beta > 1$ belongs to $\mathcal{S}$ if and only if one of the following conditions is satisfied:*

a) $d_\beta(1) = t_1 \cdots t_m (0 \cdots 0 t_{m+p})^\omega$ *and* $t_m > t_{m+p}$,

b) $d_\beta(1) = t_1 \cdots \underbrace{t_{m-qp}}_{\neq 0} \underbrace{0 \cdots 0}_{qp-1} t_m (t_m + 1 \cdots t_{m+p})^\omega$, $q \geq 1$, *and* $t_m < t_{m+p}$.

Putting this all together, we obtain a proof of the following proposition which gives us the complete list of infinite LS branches of $\mathbf{u}_\beta$ for all non-simple Parry numbers.

Proposition 39. *Let $\beta > 1$ be a non-simple Parry number and let $\mathbf{u}_\beta$ be the fixed point of the canonical substitution φ_β. Then:*

(i) if $p > 1$, then $\mathbf{u}_\beta$ is an infinite LS branch with left extensions $\{m, m+1, \ldots, m+p-1\}$,

(ii) if $\beta \notin \mathcal{S}$, then $\mathbf{u}_\beta$ is the unique infinite LS branch,

(iii) if $\beta \in \mathcal{S}$, then there are m infinite LS branches

$$0^t m \varphi^m(0^t m) \varphi^{2m}(0^t m) \cdots,$$
$$\vdots$$
$$\varphi^{m-1}(0^t m) \varphi^{2m-1}(0^t m) \varphi^{3m-1}(0^t m) \cdots.$$

There are no other infinite LS branches of $\mathbf{u}_\beta$.

5. Maximal LS Factors

As explained earlier, in order to determine the complexity of an infinite word, we need to find all infinite LS branches as well as all (a, b)-maximal LS factors. The structure of (a, b)-maximal LS factors is not so simple as the one of infinite LS branches, but still it can be described using the notion of f-image. To define an f-image for (a, b)-maximal LS factors, we need Assumption 21 to be satisfied also for g_R – we will say that the *right version of* Assumption 21 is satisfied.

Lemma 40. *For the substitution φ_β and for all distinct $a, b \in \mathcal{A}$ we have $f_R(a, b) = 0^{t_{a,b}}$, where*

$$t_{a,b} = \min\{t_a, t_b\}. \tag{8}$$

Thus, the right version of Assumption 21 is satisfied for φ_β is prefix-free.

Definition 41. Let a, b, c, d be letters of a finite alphabet $\mathcal{A}$ such that $a \neq c$ and $b \neq d$. A factor $v \in \mathcal{A}^+$ is an $(a-c, b-d)$-*bispecial* factor of an infinite word $\mathbf{u}$ defined over $\mathcal{A}$ if both avc and bvd are factors of $\mathbf{u}$.

Definition 42. Let a substitution φ defined over a finite alphabet $\mathcal{A}$ satisfy the left and right versions of Assumption 21 and let v be an $(a-c, b-d)$-bispecial factor of a fixed point of φ. Then $f_L(a,b)\varphi(v)f_R(c,d)$ is said to be the *f-image* of v.

Obviously, the f-image of v is $(\tilde{a}-\tilde{c}, \tilde{b}-\tilde{d})$-bispecial, where $g_L(a,b) = \{\tilde{a}, \tilde{b}\}$ and $g_R(c,d) = \{\tilde{c}, \tilde{d}\}$. An LS factor v having $a, b \in \mathrm{Lext}(v)$ is (a,b)-maximal if $\mathrm{Rext}(av) \cap \mathrm{Rext}(bv) = \emptyset$. Thus, it is as well an $(a-c, b-d)$-bispecial for all $c \in \mathrm{Rext}(av)$ and $d \in \mathrm{Rext}(bv)$. Are f-images of v again $(g_L(a,b))$-maximal? This question will be discussed only for our particular case of φ_β.

Lemma 43. *Let v be a bispecial factor of $\mathbf{u}_\beta$ having left extensions a and b. If its f-image*

$$f_L(a,b)\varphi_\beta(v)f_R(c,d) = f_L(a,b)\varphi_\beta(v)0^{t_{c\oplus 1, d\oplus 1}},$$

is $(g_L(a,b))$-maximal, then $c \in \mathrm{Rext}(av)$ and $d \in \mathrm{Rext}(bv)$ satisfy

$$\begin{aligned} t_{c\oplus 1} &\geq \max\{t_{e\oplus 1, f\oplus 1} \mid e \in \mathrm{Rext}(av), f \in \mathrm{Rext}(bv)\}, \\ t_{d\oplus 1} &\geq \max\{t_{e\oplus 1, f\oplus 1} \mid e \in \mathrm{Rext}(av), f \in \mathrm{Rext}(bv)\}. \end{aligned} \tag{9}$$

Proof. As we have already mentioned, for any $e \in \mathrm{Rext}(av)$ and $f \in \mathrm{Rext}(bv)$, the factor $f_L(a,b)\varphi_\beta(v)f_R(e,f) = f_L(a,b)\varphi_\beta(v)0^{t_{e\oplus 1, f\oplus 1}}$ is bispecial and therefore LS as well. These LS factors differ only in the length of the strings of zeros $0^{t_{e\oplus 1, f\oplus 1}}$ being their suffixes. Clearly, the $(g_L(a,b))$-maximal LS factor among these LS factors must be the longest one, i.e., the length of its corresponding string of zeros is greater than or equal to $t_{e\oplus 1, f\oplus 1}$ for all $e \in \mathrm{Rext}(av)$ and $f \in \mathrm{Rext}(bv)$. □

Definition 44. An f-image of a bispecial factor v having left extensions a and b

$$f_L(a,b)\varphi_\beta(v)f_R(c,d),$$

where $c \in \mathrm{Rext}(av)$ and $d \in \mathrm{Rext}(bv)$ satisfy (9), is said to be the *max-f-image* of v.

The following lemma is crucial for understanding the structure of the max-f-images of (a,b)-maximal factors.

Lemma 45. *If $\ell, k \in \mathcal{A}$, $\ell \neq k$, and $t_{\ell\oplus 1}t_{\ell\oplus 2}\cdots \succeq t_{k\oplus 1}t_{k\oplus 2}\cdots$, then for all $n \in \mathbb{N}$ the longest common prefix of the factors $\varphi_\beta^n(k)$ and $\varphi_\beta^n(\ell)$, denoted by $\mathrm{lcp}(\varphi_\beta^n(k), \varphi_\beta^n(\ell))$, satisfies*

$$\mathrm{lcp}(\varphi_\beta^n(k), \varphi_\beta^n(\ell)) = \varphi_\beta^n(k)(k \oplus n)^{-1},$$

i.e., $\varphi_\beta^n(k)$ *without the last letter* $k \oplus n$.

Moreover, denote by c *the letter such that* $(\mathrm{lcp}(\varphi_\beta^n(k), \varphi_\beta^n(\ell)))c$ *is a prefix of* $\varphi_\beta^n(\ell)$. *Then,* $t_{c\oplus 1}t_{c\oplus 2}\cdots \succeq t_{k\oplus(n+1)}t_{k\oplus(n+2)}\cdots$ *for all* $n \in \mathbb{N}$.

Proof. The case $n = 0$ is trivial. The rest of the proof is carried on by induction on n. We have

$$\begin{aligned} \varphi_\beta^{n+1}(k) &= (\varphi_\beta^n(0))^{t_{k\oplus 1}}\varphi_\beta^n(k\oplus 1), \\ \varphi_\beta^{n+1}(\ell) &= (\varphi_\beta^n(0))^{t_{k\oplus 1}}(\varphi_\beta^n(0))^{t_{\ell\oplus 1}-t_{k\oplus 1}}\varphi_\beta^n(\ell\oplus 1). \end{aligned} \tag{10}$$

If $t_{\ell\oplus 1} = t_{k\oplus 1}$, we apply the induction hypothesis on $\mathrm{lcp}(\varphi_\beta^n(k\oplus 1), \varphi_\beta^n(\ell\oplus 1))$ and if $t_{\ell\oplus 1} > t_{k\oplus 1}$, then on $\mathrm{lcp}(\varphi_\beta^n(k\oplus 1), \varphi_\beta^n(0))$ (see the Parry condition (1)).

As for the second part of the statement, the letter c is given by (10) and this along with the Parry condition concludes the proof. □

Lemma 46. *Let* $n \in \mathbb{N}$. *The* n*-th max-*f*-image of a bispecial factor* v *with left extensions* a *and* b*, i.e., the factor we obtain if we apply* n *times the mapping max-*f*-image on* v*, equals*

$$\overline{v} = s\varphi_\beta^n(v)\mathrm{lcp}(\varphi_\beta^n(c), \varphi_\beta^n(d)),$$

where $c \in \mathrm{Rext}(av)$, $d \in \mathrm{Rext}(bv)$, s *is given by (cf.* (6)*)*

$$s = f_L(g_L^{n-1}(a,b))\cdots\varphi^{n-2}(f_L(g_L(a,b))\varphi^{n-1}(f_L(a,b)), \tag{11}$$

and

$$\begin{aligned} t_{c\oplus 1}t_{c\oplus 2}\cdots &\succeq t_{c'\oplus 1}t_{c'\oplus 2}\cdots, \\ t_{d\oplus 1}t_{d\oplus 2}\cdots &\succeq t_{d'\oplus 1}t_{d'\oplus 2}\cdots \end{aligned}$$

for all $c' \in \mathrm{Rext}(av)$ *and* $d' \in \mathrm{Rext}(bv)$.

Proof. The case $n = 0$ is obvious, we carry on by induction on n. Let us assume, without loss of generality, that

$$t_{c\oplus 1}t_{c\oplus 2}\cdots \succeq t_{d\oplus 1}t_{d\oplus 2}\cdots$$

and that $g_L^n(a,b) = \{\tilde{a}, \tilde{b}\}$. Hence, we have

$$\overline{v} = s\varphi_\beta^n(v)\varphi_\beta^n(d)(d\oplus n)^{-1}$$

and

$$\mathrm{Rext}(\tilde{b}\overline{v}) = \{d' \oplus n \mid t_{d'\oplus 1}\cdots t_{d'\oplus n} = t_{d\oplus 1}\cdots t_{d\oplus n}\}.$$

Further, if $c' \in \mathrm{Rext}(\tilde{a}\overline{v})$, then, due to Lemma 45,

$$t_{c'\oplus 1}t_{c'\oplus 2}\cdots \succeq t_{d'\oplus(n+1)}t_{d'\oplus(n+2)}\cdots$$

for all $d' \oplus n \in \mathrm{Rext}(\tilde{b}\overline{v})$. But $t_{d\oplus(n+1)} \geq t_{d'\oplus(n+1)}$ for all $d' \oplus n \in \mathrm{Rext}(\tilde{b}\overline{v})$, and so the max-$f$-image of $\overline{v}$ equals

$$f_L(g_L^n(a,b))\varphi_\beta(\overline{v})0^{t_{d\oplus(n+1)}} = f_L(g_L^n(a,b))\varphi_\beta(s)\varphi_\beta^{n+1}(v)\mathrm{lcp}(\varphi_\beta^{n+1}(c),\varphi_\beta^{n+1}(d)).$$

□

Every bispecial factor v having left extensions a and b has a unique max-f-image. Since the substitution φ_β is injective, the structure of max-f-images cannot be circular as it is for f-images of infinite LS branches; v cannot be the k-th max-f-image of itself for any k. However, the notion of a max-f-image allows us to describe all (a,b)-maximal factors of $\mathbf{u}_\beta$ for all $a,b \in \mathcal{A}$. We will prove that each (a,b)-maximal factor is the k-th max-f-image either of 0^{t_1-1} if $t_1 > 1$ or of 0 if $t_1 = 1$, for some $k \in \mathbb{N}$. A sketch of the proof is as follows. Let v be an (a,b)-maximal factor containing at least two nonzero letters. Employing item (ii) of Lemma 31, one can find a bispecial factor $\overline{v}$ such that its max-f-image is v. Again, if $\overline{v}$ contains at least two nonzero letters, we find a bispecial factor $\overline{\overline{v}}$ such that its max-f-image is $\overline{v}$. In this way, we obtain a bispecial factor containing at most one nonzero letter such that its k-th max-f-image equals v. According to Corollary 34, the only candidates for such bispecial factors are of the form 0^s or $0^t m 0^q$, where $1 \leq s \leq t_1$ and $0 \leq q \leq t_1$. Note that 0^{t_1+1} cannot be a factor of $\mathbf{u}_\beta$ and that is why we consider $s, q \leq t_1$. In the case when $t = 0$, words $0^{t_k}k0^q$, with $k > m, t_{m+1} = \cdots = t_{k-1} = 0$, could also be taken as candidates but we do not consider them as they are just prefixes of $\varphi_\beta^{k-m}(m0^q)$. The following two lemmas tell us that 0^{t_1-1} (resp. 0 if $t_1 = 1$) is the only candidate.

Lemma 47. *Let $t_1 > 1$ and $k \in \mathbb{N}$. Then the k-th max-f-image of factors 0^{t_1}, 0^s and $0^t m 0^q$, where $1 \leq s < t_1 - 1$ and $0 \leq q \leq t_1$, is not (a,b)-maximal for any distinct letters a and b.*

Proof. First, consider 0^{t_1} with distinct left extensions a and b. We have that $\mathrm{Lext}(0^{t_1}) = \mathrm{Lext}(0^{t_1}1)$ and $\mathrm{Rext}(0^{t_1}) = \{k \in \mathcal{A}\backslash\{0\} \mid t_k = t_1 \text{ or } k = m, t_{m+p} = t_1\}$. For each $k \in \mathrm{Rext}(0^{t_1})$, we must have $t_{k\oplus1}t_{k\oplus2}\cdots \prec t_2t_3\cdots$ (see Parry condition (1)) and, due to Lemma 46, the k-th max-f-image of 0^{t_1} is a prefix of a k-th f-image of the LS factor $0^{t_1}1$, both having the same left extensions.

Similar arguments can be used in order to prove that the k-th max-f-image of 0^s is always a prefix of the k-th f-image of the LS factor 0^{t_1-1}. Again, $\mathrm{Lext}(0^s) = \mathrm{Lext}(0^{t_1-1})$ and the rest is implied directly by the Parry condition.

Finally, consider the LS factor $0^t m 0^q$ having just two left extensions 0 and z (see (7)). According to Lemma 46, the m-th max-f-image of 0^{t_1-1} with left extensions 0 and p equals

$$0^t m \varphi_\beta^m(0^{t_1-1})\varphi_\beta^m(1)(m+1)^{-1} = 0^t m 0^{t_1} 1 \cdots . \tag{12}$$

Indeed, $\mathrm{Rext}(00^{t_1-1}) = \{k \in \mathcal{A} \setminus \{0\} | t_k = t_1 \text{ or } k = m, t_{m+p} = t_1\}$, and $0 \in \mathrm{Rext}(p0^{t_1-1})$. Therefore the fact that $t_{k\oplus1}t_{k\oplus2}\cdots \prec t_2t_3\cdots$ and the Parry condition imply that the m-th max-f-image is $0^t m\varphi_\beta^m(0^{t_1-1})\mathrm{lcp}(\varphi_\beta^m(0), \varphi_\beta^m(1))$. Thus, $0^t m 0^q$, as a prefix of (12), is not $(0, z)$-maximal. ☐

Lemma 48. *Let $t_1 = 1$ and $k \in \mathbb{N}$. Then $t = 0$ and the k-th max-f-image of the factor $m0^q$, where $0 \leq q \leq 1$, is not (a, b)-maximal for any distinct letters a and b.*

Proof. As in the proof of the previous lemma, we can prove that the $(m - \ell_0)$-th max-f-image of 0 with left extensions ℓ_0 and $\ell_0 + p$ is the factor

$$m\varphi_\beta^m(1)(m+1)^{-1}, \tag{13}$$

where, according to item (i) of Lemma 31,

$$\varphi_\beta^m(1) = (\varphi_\beta^{m-1}(0))^{t_2}(\varphi_\beta^{m-2}(0))^{t_3} \cdots (\varphi_\beta(0))^{t_m} 0^{t_{m+1}}(m+1).$$

In order that $m0$ may be $(0, z)$-maximal, we must have $\varphi_\beta^m(1) = m + 1$, and thus $t_2 = \cdots = t_{m+1} = 0$. Then we have $t_{m+p} = 1$ and $t_1 t_2 \cdots \prec t_{m+p} t_m \cdots t_{m+p} t_m \cdots$, a contradiction with the Parry condition. ☐

Proposition 49. *Let v be an (a, b)-maximal factor of $\mathbf{u}_\beta$. Then there exists $k \in \mathbb{N}$ such that v is the k-th max-f-image of*

(i) $0^{t_1 - 1}$ *if* $t_1 > 1$,

(ii) 0 *if* $t_1 = 1$.

Proof. We will prove that if v contains at least two nonzero letters, then it is the k-th max-f-image of a bispecial factor of the form 0^s or $0^t m 0^q$, where $1 \leq s \leq t_1$ and $0 \leq q \leq t_1$. The rest of the proof then follows from the previous two lemmas.

Let us assume that v contains at least two nonzero letters. Then, due to item (ii) of Lemma 31, $v = f_L(a', b')\varphi_\beta(\overline{v})f_R(c', d')$, where $\overline{v}$ is an $(a' - c', b' - d')$-bispecial factor such that v is the max-f-image of $\overline{v}$ and $g_L(a', b') = \{a, b\}$. Analogously, if $\overline{v}$ contains at least two nonzero letters, there exists an $(a'' - c'', b'' - d'')$-bispecial factor $\overline{\overline{v}}$ which is an f-preimage of $\overline{v}$. But it must also be a max-f-preimage. Indeed, if it is not, then $\overline{v}0^{q'}$ is also an f-image of $\overline{\overline{v}}$ having the left extensions a' and b' for some $q' > 0$ and so v cannot be (a, b)-maximal as it is a proper prefix of the max-f-image of the LS factor $\overline{v}0^{q'}$ with the left extensions a and b. Using this argument iteratively, we will obtain a bispecial factor of the form 0^s or $0^t m 0^q$ such that v is its k-th max-f-image. ☐

In fact, the previous proposition along with Lemma 46 provides us with the complete list of (a, b)-maximal factors. However, in the last section of this paper we will need to know some details to be able to determine under which conditions the complexity of $\mathbf{u}_\beta$ is affine.

Corollary 50. *If $d_\beta(1) \neq t_1(0\cdots 0(t_1-1))^\omega$, then the k-th max-f-image of the factor* (12) *is $(g_L^k(0,z))$-maximal for all $k \in \mathbb{N}$.*

If $\beta \notin \mathcal{S}$, then the k-th max-f-image reads

$$\varphi_\beta^k(0^t m)\varphi_\beta^{m+1}(0^{t_1-1})\varphi_\beta^{m+1}(1)(m \oplus k)^{-1}.$$

Proof. The factor (12) is always LS with just two left extensions 0 and z. Therefore, it is $(0,z)$-maximal if it is neither a prefix of any infinite LS branch nor a proper prefix of the k-th max-f-image of itself for any $k > 0$.

In the case when $\beta \notin \mathcal{S}$, the longest common prefix of the k-th max-f-image of the factor (12) and of the unique infinite LS branch $\mathbf{u}_\beta$ equals

$$\varphi_\beta^k(0^t m)(m \oplus k)^{-1}.$$

Hence, either it is non-empty and shorter than the longest common prefix of the $(k+1)$-th max-f-image of (12) and $\mathbf{u}_\beta$, or it is empty, $k < p$, and $t = t_{m+1} = \cdots = t_{m+k} = 0$ (or only $t = 0$ for $k = 0$). In the latter case, the k-th max-f-image of (12) begins in letter $m + k$ which is different from the first letters of $\mathbf{u}_\beta$ and of all other max-f-images of (12). Putting all this together, the k-th max-f-image of (12) is neither a prefix of $\mathbf{u}_\beta$ nor of the ℓ-th max-f-image of (12) for any $\ell \neq k$.

If $\beta \in \mathcal{S}$, then $\mathbf{u}_\beta$ is not the only infinite LS branch; there are m other branches:

$$\mathbf{u}_1 = 0^t m \varphi_\beta^m(0^t m)\varphi_\beta^{2m}(0^t m)\cdots \tag{14}$$

and $\mathbf{u}_\ell = \varphi_\beta^{\ell-1}(\mathbf{u}_1), l = 2, \ldots, m$. To finish the proof, we have to foreclose the possibility that the factor (12) is a prefix of $\mathbf{u}_1$. Looking at (14) and (12), we see that it happens only if $t = t_1 - 1$ and $m = 1$, in other words, if $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$. The proof that the factor (12) is not a prefix of any max-f-image of itself is analogous to the one above. □

Let us state the immediate consequence of the previous corollary and its proof.

Corollary 51. *Each LS factor in $\mathbf{u}_\beta$ is a prefix of an infinite LS branch if and only if $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$.*

Corollary 52. *If $t_1 > 1$, then the k-th max-f-image of 0^{t_1-1} with left extensions 0 and a is a $(g_L^k(0,a))$-maximal factor for all $a \in \mathcal{A} \setminus \{0,z\}$ and for all $0 \leq k < m$.*

Moreover, put

$$k_0 = \begin{cases} -1 & \text{if } t \neq t_1 - 1, \\ 0 & \text{if } t = t_1 - 1 \text{ and } t_2 \neq t_{m+1}, \\ \max\{\ell \in \mathbb{N} \mid t_{\ell+1} = t_{m\oplus\ell}\} & \text{otherwise,} \end{cases} \tag{15}$$

then the k-th max-f-image of 0^{t_1-1} is also a $(g_L^k(0,z))$-maximal factor for all $k_0 < k < m$.

Proof. We have

$$\mathrm{Rext}(00^{t_1-1}) = \{k \in \mathcal{A} \setminus \{0\} \mid t_k = t_1 \text{ or } k = m, t_{m+p} = t_1\}$$

and for all $a \in \mathcal{A} \setminus \{0\}$ we have $k \in \mathrm{Rext}(a0^{t_1-1})$ if and only if $k = 0$ or both the following conditions are satisfied:

(i) $z(k) = a - 1$ or $y(k) = a - 1$,

(ii) $t_k = t_1 - 1$, or $k = m$ and $t_{m+p} = t_1 - 1$.

The intersection of $\mathrm{Rext}(00^{t_1-1})$ and $\mathrm{Rext}(a0^{t_1-1})$ is non-empty if and only if $a = z$ and $t = t_1 - 1$. In other words, if and only if 0^{t_1-1} is a prefix of $0^t m$, which is an LS factor having just two left extensions 0 and z.

Similarly, we can prove that the k-th max-f-image of 0^{t_1-1} is a $(g_L^k(0,a))$-maximal factor for all $a \in \mathcal{A} \setminus \{0, z\}$. In the same way, the k-th max-f-image of 0^{t_1-1}, namely

$$\varphi_\beta^k(0^{t_1-1})\varphi_\beta^k(1)(k+1)^{-1},$$

is $(g_L^k(0,z))$-maximal if it is not a prefix of the LS factor

$$\varphi_\beta^k(0^t m) = \varphi_\beta^k(0^t)\varphi_\beta^k(m)$$

having the left extensions $g_L^k(0,z)$. The proof then follows from item (i) of Lemma 31 and Lemma 45 applied on $\varphi_\beta^k(1)$ and $\varphi_\beta^k(m)$. □

Taking into account Lemmas 33 and 49, one can prove the following corollary using analogous techniques as in the proof of the Corollary 52. Note that $\mathrm{Rext}(\ell_0 0) = \{k \in \mathcal{A} \mid z(k-1) = \ell_0 - 1 \text{ or } y(k-1) = \ell_0 - 1\}$ and $\mathrm{Rext}(a0) = \{1\}$ for all $a > \ell_0$, i.e., 0 is $(\ell_0, \ell_0 + z)$-maximal if it is not a prefix of the ℓ_0-th max-f-image of the factor (12) which reads

$$\varphi_\beta^{\ell_0}(m)\varphi_\beta^{m+\ell_0}(1)(1 \oplus (m+\ell_0))^{-1} = \varphi_\beta^{\ell_0}(m)\varphi_\beta^{m}(\ell_0+1)(1 \oplus (m+\ell_0))^{-1}.$$

Corollary 53. *If $t_1 = 1$, then the k-th max-f-image of 0 is a $(g_L^k(\ell_0, a+\ell_0))$-maximal factor for all letters $a > \ell_0$, $a \neq z$ and for all $0 \leq k < m - \ell_0$.*

Moreover, the k-th max-f-image of 0 is $(g_L^k(\ell_0, z+\ell_0))$-maximal if $k_0 \geq \ell_0$ and $k = k_0 - \ell_0, k_0 - \ell_0 + 1, \ldots, m - \ell_0$, where k_0 is defined by (15).

6. Affine Complexity

The aim of the present section is to find a necessary and sufficient condition for the factor complexity of $\mathbf{u}_\beta$ being affine. In order for the complexity to be affine, the first difference of complexity $\triangle\,\mathcal{C}(n)$ must be constant. The following lemma says when $\triangle\mathcal{C}(n)$ can change its value. The proof is an immediate consequence of (2).

Lemma 54. *Let* $\mathbf{u}$ *be an infinite word over a finite alphabet.*

(i) *If* $\triangle\,\mathcal{C}(n+1) >\triangle\,\mathcal{C}(n)$*, then the number of LS factors of length* $n+1$ *is greater than the number of LS factors of length* n*.*

(ii) *If* $\triangle\,\mathcal{C}(n+1) <\triangle\,\mathcal{C}(n)$*, then* $\mathbf{u}$ *contains an* (a,b)*-maximal factor of length* n *for some letters* a *and* b*.*

That is, the complexity is affine if either $\mathbf{u}$ does not contain any (a,b)-maximal factor and all infinite LS branches have empty common prefix, or if each (a,b)-maximal factor of length n is "compensated" by appearance of a "new" LS factor of length $n+1$. Examples of the first case are Arnoux-Rauzy words, all of whose LS factors are prefixes of a unique infinite LS branch. As for the latter case, the appearance of a "new" LS factor of length $n+1$ means there is an LS factor v of length n and its right extensions c and d such that vc and vd are both LS, i.e., v is the longest common prefix of two different LS factors – Cassaigne [11] calls such LS factors *strong bispecial*.

Since $\mathbf{u}_\beta$ always comprises at least one (a,b)-maximal factor, each such (a,b)-maximal factor must be as long as the longest common prefix of two different LS factors in order that the complexity may be affine. We will prove that it is only possible if the number of (a,b)-maximal factors is finite, thus, in the case of $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$.

Lemma 55. *If* $k_0 < m-1$*, where* k_0 *is defined by* (15)*, then the factor complexity of* $\mathbf{u}_\beta$ *is not affine.*

Proof. First assume that $t_1 > 1$. If $k_0 < m-1$, then the (k_0+1)-th max-f-image of 0^{t_1-1} is $g_L^{k_0}(0,z)$-maximal. Consider the longest common prefix of the LS factor $\varphi_\beta^{k_0}(0^t m)$ having the left extensions $g_L^{k_0}(0,z)$ and of the infinite LS branch $\mathbf{u}_\beta$ if $p > 1$, or of the LS factor $\varphi_\beta^{m-1}(0^{t_1-1})$ with the left extensions $m-1$ and m if $p=1$ (in such a case $\mathbf{u}_\beta$ is not an infinite LS branch). This factor equals $\varphi_\beta^{k_0}(0^t m)(m\oplus k_0)^{-1}$ which is a prefix of the (k_0+1)-th max-f-image of 0^{t_1-1} and hence it is not (a,b)-maximal for any distinct $a,b\in\mathcal{A}$. Overall, $\triangle\,\mathcal{C}(n_0) <\triangle\,\mathcal{C}(n_0+1)$, where n_0 is the length of the factor $\varphi_\beta^{k_0}(0^t m)(m\oplus k_0)^{-1}$.

Similar arguments work for the case $t_1 = 1$, where we replace k_0+1 by $k_0-\ell_0+1$ and the factor 0^{t_1-1} by 0. $\square$

Lemma 56. *If* $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$*, then the factor complexity of* $\mathbf{u}_\beta$ *is affine, namely* $\mathcal{C}(n) = pn+1, n\in\mathbb{N}$*.*

Proof. In this case, $t = t_1 - 1$ and so $k_0 = 0 = m-1$. Hence, the $(0,a)$-maximal factor 0^{t_1-1} is at the same time the longest common prefix of the only infinite LS branches $\mathbf{u}_\beta$ and $0^t m\varphi_\beta(0^t m)\varphi_\beta^2(0^t m)\cdots$. But 0^{t_1-1} is the only (a,b)-maximal factor and prefixes of these two infinite LS branches are the only LS factors of $\mathbf{u}_\beta$; thus, the proof is complete. □

Lemma 57. *If $\beta \in \mathcal{S}$ and $d_\beta(1) \neq t_1(0\cdots 0(t_1-1))^\omega$, then the factor complexity of $\mathbf{u}_\beta$ is not affine.*

Proof. In the case when $p > 1$, there are $m+1$ infinite LS branches given by Proposition 39. Let us denote them by $\mathbf{u}_0 = \mathbf{u}_\beta, \mathbf{u}_1, \ldots, \mathbf{u}_m$ and put

$$n_0 = \max\{|v| \mid v = \mathrm{lcp}(\mathbf{u}_i, \mathbf{u}_j), i \neq j, i, j = 0, 1, \ldots, m\}.$$

We have $\triangle\mathcal{C}(n) \geq \#\mathrm{Lext}(\mathbf{u}_0) - 1 + \sum_{k=1}^{m} \#\mathrm{Lext}(\mathbf{u}_k) - 1 \geq p - 1 + m$ for all $n > n_0$. Due to Corollary 50, we know that there exist infinitely many $(g_L^k(0,z))$-maximal factors, $k = 0, 1, \ldots$, and hence there must exist an LS factor of length $n_1 > n_0$ which is not a prefix of any LS branch. Therefore $\triangle\mathcal{C}(n_1) > m + p - 1 = \triangle\mathcal{C}(1)$.

In the case of $p = 1$, the proof is analogous. The only difference is that there are only m infinite LS branches since $\mathbf{u}_\beta$ is not one. □

Remark 58. *For the word $\mathbf{u}_\beta$ with $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$ we may easily describe all left special factors. If the length of the period p is greater than 1, each LS factor is a prefix of one of two infinite LS branches $\mathbf{u}_\beta$ and $0^{-1}\mathbf{u}_\beta$. If $p = 1$, then $\mathbf{u}_\beta$ is not an infinite LS branch and thus every LS factor is a prefix of the unique infinite LS branch $0^{-1}\mathbf{u}_\beta$. Hence, we obtain the known result that $\mathbf{u}_\beta$ is Sturmian if and only if $d_\beta(1) = t_1(t_1-1)^\omega$. We were pointed out by Christiane Frougny that numbers β satisfying $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$ are Pisot units. Such Parry number β is a root of the polynomial $x^{p+1} - t_1 x^p - x + 1$.*

Lemma 59. *Let $\beta \notin \mathcal{S}$ and let $k_0 \geq m-1$. Then the factor complexity of $\mathbf{u}_\beta$ is not affine.*

Proof. As shown in the proof of Lemma 55, the k-th max-f-image of 0^{t_1-1} (resp. 0 if $t_1 = 1$) is equal to the longest prefix of some two LS factors for $k = 0, 1, \ldots, m-1$. In order for the complexity to be affine, also all consecutive max-f-images of the factor (12) must be as long as the longest common prefix of some two LS factors.

Let $t_1 > 1$. Then the factor (12) must be of the same length as the longest common prefix of $\mathbf{u}_\beta$ and the m-th max-f-image of itself. Remember that the longest common prefix of $\mathbf{u}_\beta$ and the k-th max-f-image of (12) is the k-th max-f-image of 0^{t_1-1} for $k = 0, 1, \ldots, m-1$. Formally,

$$\begin{aligned} |0^t m\varphi_\beta^m(0^{t_1-1}1)(1+m)^{-1}| &= |\mathrm{lcp}(\mathbf{u}_\beta, \varphi_\beta^m(0^t m)\varphi_\beta^{2m}(0^{t_1-1}1)(1\oplus(2m))^{-1})| \\ &= |\varphi_\beta^m(0^t m)(m\oplus m)^{-1}| \end{aligned}$$

which is never satisfied for $|\varphi_\beta^m(0^t m)| \le |\varphi_\beta^m(0^{t_1-1}1)|$.

Let $t_1 = 1$. Following the same reasoning as for the case $t_1 > 1$, a necessary condition for the complexity to be affine is that the factor (12),

$$m\varphi_\beta^m(1)(1+m)^{-1},$$

must be of the same length as the longest common prefix of the $(m-\ell_0)$-th max-f-image of itself and $\mathbf{u}_\beta$, namely

$$|\mathrm{lcp}(\mathbf{u}_\beta, \varphi_\beta^{m-\ell_0}(m)\varphi_\beta^{2m-\ell_0}(1)(1 \oplus (2m-\ell_0))^{-1})| = |\varphi_\beta^{m-\ell_0}(m)(m \oplus (m-\ell_0))^{-1}|,$$

which is never satisfied for $|\varphi_\beta^{m-\ell_0}(m)| \le |\varphi_\beta^m(1)|$. $\square$

Putting all lemmas of this section together, we obtain the main theorem of this paper.

Theorem 60. *Let β be a non-simple Parry number. The factor complexity of $\mathbf{u}_\beta$ is affine if and only if $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$.*

7. Conclusion

Among infinite words $\mathbf{u}_\beta$ associated with Parry numbers we may identify Arnoux-Rauzy words. An infinite word is said to be Arnoux-Rauzy of order ℓ, if for any length $n \in \mathbb{N}$ there exists exactly one left special factor and one right special factor both of length n and, moreover, these special factors have just ℓ left and ℓ right extensions, respectively. Arnoux-Rauzy words can be considered as a natural generalization of Sturmian words to more letter alphabets.

It is easy to see that $\mathbf{u}_\beta$ is a Sturmian word if and only if $d_\beta(1) = t_1 1$ or $d_\beta(1) = t_1(t_1-1)^\omega$. The word $\mathbf{u}_\beta$ is an Arnoux-Rauzy word of order $m \ge 3$ if and only if $d_\beta(1) = t_1^{m-1}1$, see [17] and [2]. It means that there is no Arnoux-Rauzy word over three or more letter alphabet associated with non-simple Parry number. A direct consequence of the definition of Arnoux-Rauzy words is that the complexity of Arnoux-Rauzy word is affine and that any left (right) special factor is a prefix (suffix) of an infinite left (right) special branch.

In the previous section, we have proved that the infinite word $\mathbf{u}_\beta$ associated with a non-simple Parry number β has affine complexity if and only if $d_\beta(1) = t_1(0\cdots 0(t_1-1))^\omega$. In fact, we have proved that the complexity is affine if and only if any left special factor of $\mathbf{u}_\beta$ is a prefix of an infinite left special branch (Corollary 51). The validity of the same statement for infinite words associated with simple Parry numbers is proven in [5]. However, this equivalence is not a general rule for the factor complexity of fixed points of primitive morphisms. For a counter example see [12] and [15].

It is known that Sturmian words have many equivalent definitions, see [8] for more. In 2001 Vuillon [30] showed that a binary infinite word is Sturmian if and only if each its factor has exactly two return words. In article [29] Vuillon introduced the property R_ℓ: an infinite word satisfies the property R_ℓ if each its factor has exactly ℓ return words. Therefore, words with R_ℓ can be considered as another generalization of Sturmian words. In [19] Justin and Vuillon proved that Arnoux-Rauzy words of order ℓ have the property R_ℓ. Applying Theorem 4.5 of [3], we see that all u_β over an ℓ-letter alphabet with affine complexity have also the property R_ℓ.

Acknowledgement

We thank Christiane Frougny for many fruitful discussions, Petr Kůrka for his text on Symbolic dynamics, available at his homepage, from where we took the definition of recognizable substitutions and the anonymous referee for his suggestions to improve the paper.

References

[1] S. Akiyama, C. Frougny, and J. Sakarovitch. Powers of rationals modulo 1 and rational base number systems. *Israel J. Math* **168** (2008), 53–91.

[2] L. Balková and Z. Masáková. Palindromic complexity of infinite words associated with non-simple Parry numbers. *Theor. Inform. Appl.* **43** (2009), 145–163.

[3] L. Balková, E. Pelantová, and W. Steiner. Sequences with constant number of return words. *Monatsh. Math.* **155** (2008), no. (3-4), 251–263.

[4] D. Barache, B. Champagne, and J.-P. Gazeau. *Quasicrystals and discrete geometry*, chapter Pisot-cyclotomic quasilattices and their symmetry semigroups, pages 15–66. Amer. Math. Soc., 1998.

[5] J. Bernat, Z. Masáková, and E. Pelantová. On a class of infinite words with affine factor complexity. *Theoret. Comput. Sci.* **389** (2007), no. (1-2), 12–25.

[6] J. Berstel. Mots de Fibonacci. Séminaire d'informatique théorique, LITP, Paris, Année 1980/81, 57-78.

[7] J. Berstel. Properties of infinite words: recent results. In *Proceedings of the 6th Annual Symposium on Theoretical Aspects of Computer Science on STACS 89*, pages 36–46, New York, NY, USA, 1989. Springer-Verlag New York, Inc.

[8] J. Berstel. Sturmian and episturmian words: A survey of some recent results. *Lecture notes in computer science* **4728** (2007), 23–47.

[9] J. Cassaigne. An algorithm to test if a given circular HD0L-language avoids a pattern. In *IFIP World Computer Congress'94*,459–464. North-Holland, 1994.

[10] J. Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in Language Theory*, pages 25–34. World Scientific, 1996.

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

[12] R. V. Chacon. Weakly mixing transformations which are not strongly mixing. *Proc. Amer. Math. Soc.* **22** (1969), no. 3, 559–562.

[13] A. de Luca and S. Varricchio. On the factors of the Thue-Morse word on three symbols. *Inform. Process. Lett.* **27** (1988), no. 6, 281–285.

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

[15] S. Ferenczi. Les transformations de Chacon : combinatoire, structure géométrique, lien avec les systèmes de complexité $2n + 1$. *Bull. Soc. Math. France* **123** (1995), no. 2, 271–292.

[16] C. Frougny, J.-P. Gazeau, and R. Krejcar. Additive and multiplicative properties of point sets based on beta-integers. *Theor. Comput. Sci.* **303** (2003) no. 2-3, 491–516.

[17] C. Frougny, Z. Masáková, and E. Pelantová. Complexity of infinite words associated with beta-expansions. *Theor. Inform. Appl.***38** (2004), 163–185.

[18] C. Frougny, Z. Masáková, and E. Pelantová. Infinite special branches in words associated with beta-expansions. *Discrete Math. Theor. Comput. Sci.* **9** (2007), 125–144.

[19] J. Justin and L. Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.* **34** (2000), 343–356.

[20] D. König. *Theorie der endlichen und unendlichen Graphen.* Akademische Verlagsgesellschaft, Leipzig, 1936.

[21] Z. Masáková, J. Patera, and J. Zich. Classification of Voronoi and Delone tiles in quasicrystals: I. General method. *J. Phys. A: Math. Gen.* **16** (2003), 1869–1894.

[22] R. V. Moody and J. Patera. Quasicrystals and icosians. *J. Phys. A: Math. Gen.* **26** (1993), 2829–2853.

[23] B. Mossé. Notions de reconnaissabilité pour les substitutions et complexité des suites automatiques. *Bull. Soc. Math. France***124** (1996),101–108.

[24] W. Parry. On the β-expansions of real numbers. *Acta Math. Acad. Sci. Hunger.* **11** (1960), 401–416.

[25] M. Queffélec. *Substitution dynamical systems–spectral analysis*, *Lecture Notes in Mathematics* **1284**, Springer-Verlag, Berlin, 1987.

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

[27] D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn. Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.***53** (1984), no. 20, 1951–1953.

[28] W. Thurston. Groups, tilings and finite state automata. AMS Colloquium Lecture Notes, 1989.

[29] L. Vuillon. On the number of return words in infinite words with complexity $2n+1$. Technical Report **15** (2000), LIAFA.

[30] L. Vuillon. A characterization of Sturmian words by return words. *Eur. J. Comb.* **22** (2001), 263–275.