

On two conjectures of Shallit about Thue-Morse-like sequences

Lubomíra Dvořáková^{a,*}, Savinien Kreczman^{b,1}, Edita Pelantová^a

^a*Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering,
Czech Technical University in Prague, Trojanova 13, Praha 2, 120 00, Czech Republic*

^b*Department of Mathematics, University of Liège,
Allée de la Découverte 12, Liège, 4000, Belgium*

Abstract

We study a class of infinite words \mathbf{x}_k , $k \in \mathbb{N}$, $k \geq 1$, recently introduced by J. Shallit. This class includes the Thue-Morse sequence \mathbf{x}_1 , the Fibonacci-Thue-Morse sequence \mathbf{x}_2 , and the Allouche-Johnson sequence \mathbf{x}_3 . Shallit stated and for $k = 3$ proved two conjectures on properties of \mathbf{x}_k . The first conjecture concerns the factor complexity, the second one the critical exponent of these words. We confirm the validity of both conjectures for every k .

Keywords: morphism, Thue-Morse sequence, factor complexity, critical exponent, asymptotic critical exponent, bispecial factors

1. Introduction

In a recent paper [1], Shallit leverages **Walnut** to prove a number of properties related to the Narayana morphism $\mu: 0 \mapsto 01, 1 \mapsto 2, 2 \mapsto 0$, its fixed point the Narayana word, and the Narayana numeration system. The latter is the positional numeration system with weights (U_n) satisfying the recurrence relation $U_{n+3} = U_{n+2} + U_n$ with initial conditions 1, 2, 3.

*Corresponding author.

Email addresses: `lubomira.dvorakova@fjfi.cvut.cz` (Lubomíra Dvořáková),
`savinien.kreczman@uliege.be` (Savinien Kreczman), `edita.pelantova@fjfi.cvut.cz`
(Edita Pelantová)

¹Supported by the FNRS Research Fellow grant 1.A.789.23F, by a Fédération Wallonie-Bruxelles Travel Grant and by a WBI Excellence WORLD grant. Avec le soutien de Wallonie-Bruxelles International.

This numeration system is one instance of a family defined by the recurrence relation $U_{n+k} = U_{n+k-1} + U_n$ with the initial conditions $1, \dots, k$. The case $k = 1$ corresponds to the usual binary numeration system, and $k = 2$ to the Zeckendorf numeration system [2]. This family of numeration systems is linked to the family of recursively defined Hofstadter functions, see [3, 4].

Shallit [1] established a connection to a sequence introduced by Allouche and Johnson [5]: the n -th term of this sequence is the parity of the number of 1's in the Narayana representation of n . Again, this is the $k = 3$ representative of a family which contains the Thue-Morse sequence itself ($k = 1$) and the Fibonacci-Thue-Morse sequence ([6], $k = 2$). We call other representatives of this family *Thue-Morse-like sequences*, although Shallit simply calls them \mathbf{x}_k .

Shallit shows that for $k \geq 2$, the sequence \mathbf{x}_k is the image under the projection π of the fixed point $\mathbf{u}_k = \xi_k^\omega(0)$ of the morphism $\xi = \xi_k$ defined over the alphabet $\mathcal{A} = \{0, 1, \dots, k-1, 0', 1', \dots, (k-1)'\}$ as follows

$$\begin{array}{ccc} \xi_k : & \begin{array}{l} 0 \rightarrow 01 \\ 1 \rightarrow 2 \\ 2 \rightarrow 3 \\ \vdots \\ (k-2) \rightarrow (k-1) \\ (k-1) \rightarrow 0' \end{array} & \text{and} \quad \begin{array}{l} 0' \rightarrow 0'1' \\ 1' \rightarrow 2' \\ 2' \rightarrow 3' \\ \vdots \\ (k-2)' \rightarrow (k-1)' \\ (k-1)' \rightarrow 0. \end{array} \end{array} \quad (1.1)$$

The projection $\pi = \pi_k$ we apply to \mathbf{u}_k is given by

$$\pi(a) = \begin{cases} 0, & \text{if } a = 0 \text{ or } a \in \{1', 2', \dots, (k-1)'\}; \\ 1, & \text{if } a = 0' \text{ or } a \in \{1, 2, \dots, k-1\}. \end{cases} \quad (1.2)$$

Having proven results for the Allouche-Johnson sequence \mathbf{x}_3 and comparing to the known results for \mathbf{x}_1 and \mathbf{x}_2 , Shallit formulates two conjectures regarding the family \mathbf{x}_k . The first conjecture concerns factor complexity, the second one critical exponent. See Section 2 for definitions of these notions.

Conjecture 1 ([1], Conjecture 45). Let $k \geq 1$. The first difference of factor complexity of \mathbf{x}_k , for n large enough, takes the values $4k - 2$ and $4k$ only.

Conjecture 2 ([1], Conjecture 44). Let $k \geq 1$. The sequence \mathbf{x}_k has critical exponent $k + 1$, which is attained by the words 0^{k+1} and 1^{k+1} . It contains no factor of length $2n + k$ and period n , and therefore has asymptotic critical exponent equal to 2.

The famous Thue-Morse sequence \mathbf{x}_1 is known to satisfy both conjectures. Already Thue showed that \mathbf{x}_1 is overlap-free and contains squares of unbounded size, see [7]. Its critical exponent equals 2, which is minimal among binary sequences. Its asymptotic critical exponent also equals 2 and the first difference of factor complexity takes only the values 2 and 4 for positive arguments, see [8] and [9]. The cases of $k = 2$ (the Fibonacci-Thue-Morse sequence \mathbf{x}_2) and $k = 3$ (the Allouche-Johnson sequence \mathbf{x}_3) were studied already ([6], resp. [1]). Both conjectures hold for these particular cases. However, the methods of the latter paper cannot be used to prove these conjectures for all k .

The main goal of this article is to prove both conjectures for all k . For this purpose, we first study bispecial factors in the fixed point \mathbf{u}_k of ξ_k in Section 3 and we show that the language of \mathbf{u}_k is overlap-free. Section 4 examines properties of the Thue-Morse-like sequence $\mathbf{x}_k = \pi(\mathbf{u}_k)$, again with a focus on bispecial factors, and prepares tools to prove both conjectures. The proofs are presented in Sections 5 (Conjecture 1) and 6 (Conjecture 2).

2. Preliminaries

An *alphabet* \mathcal{A} is a finite set, its elements are *letters*. A *word* u over \mathcal{A} of *length* n is a finite sequence $u = u_0u_1 \cdots u_{n-1}$ of letters $u_j \in \mathcal{A}$ for all $j \in \{0, 1, \dots, n-1\}$. The length of u is denoted $|u|$. The set of all finite words over \mathcal{A} is denoted \mathcal{A}^* . The set \mathcal{A}^* equipped with concatenation as the operation forms a monoid with the *empty word* ε as the neutral element. Consider $u, p, s, v \in \mathcal{A}^*$ such that $u = pvs$, then the word p is called a *prefix*, the word s a *suffix* and the word v a *factor* of u . The *longest common prefix* of two words u and z is denoted $\text{lcp}\{u, z\}$, it is the longest factor of u and z that is a prefix of both u and z . The *longest common suffix* (lcs) is defined analogously. A *sequence* (or an *infinite word*) \mathbf{v} over \mathcal{A} is an infinite sequence $\mathbf{v} = v_0v_1v_2 \cdots$ of letters $v_j \in \mathcal{A}$ for all $j \in \mathbb{N}$. A *word* w over \mathcal{A} is called a *factor* of the sequence $\mathbf{v} = v_0v_1v_2 \cdots$ if there exists $j \in \mathbb{N}$ such that $w = v_jv_{j+1}v_{j+2} \cdots v_{j+|w|-1}$. The index j is called an *occurrence* of w . If $j = 0$, then w is a *prefix* of \mathbf{v} . Consider a factor w of an infinite word $\mathbf{v} = v_0v_1v_2 \cdots$. Let $j < \ell$ be two consecutive occurrences of w in \mathbf{u} . Then the word $v = v_jv_{j+1} \cdots v_{\ell-1}$ is a *return word* to w in \mathbf{v} and vw is a *complete return word* to w in \mathbf{v} .

The *language* $\mathcal{L}(\mathbf{v})$ of a sequence \mathbf{v} is the set of factors occurring in \mathbf{v} .

The *factor complexity* of a sequence \mathbf{v} is the map $\mathcal{C} : \mathbb{N} \rightarrow \mathbb{N}$, where

$$\mathcal{C}(n) = \#\{w \in \mathcal{L}(\mathbf{v}) : |w| = n\}.$$

To evaluate the factor complexity, Cassaigne [10] introduced the concept of special factors. A factor w of a sequence \mathbf{v} is *left special* if $aw, bw \in \mathcal{L}(\mathbf{v})$ for at least two distinct letters $a, b \in \mathcal{A}$. Such letters a, b are called *left extensions* of w . The set of all left extensions of w in \mathbf{v} is denoted $\text{Lext}(w)$. A *right special* factor, *right extensions* and $\text{Rext}(w)$ are defined analogously. The factor w is *bispecial* if it is both left special and right special.

The *first difference of factor complexity* is a mapping $\Delta\mathcal{C} : \mathbb{N} \rightarrow \mathbb{N}$ defined by $\Delta\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n)$. The first difference of factor complexity of a binary sequence \mathbf{v} satisfies

$$\Delta\mathcal{C}(n) = \#\{w \in \mathcal{L}(\mathbf{v}) : w \text{ is left special in } \mathbf{v} \text{ and } |w| = n\}. \quad (2.1)$$

A sequence \mathbf{v} is *recurrent* if each factor of \mathbf{v} has infinitely many occurrences in \mathbf{v} . Moreover, a recurrent sequence \mathbf{v} is *uniformly recurrent* if for every $n \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that each factor of \mathbf{v} of length N contains all factors of \mathbf{v} of length n .

A sequence \mathbf{v} is *eventually periodic* if there exist words u (possibly empty) and v such that $\mathbf{v} = uv^\omega$, where ω denotes an infinite repetition. The sequence \mathbf{v} is called *aperiodic* otherwise. Let \mathbf{v} be a uniformly recurrent and aperiodic sequence. Then every left special factor v in \mathbf{v} is a prefix of a bispecial factor in \mathbf{v} . There exists a unique shortest bispecial factor having a prefix v .

A *morphism* is a map $\psi : \mathcal{A}^* \rightarrow \mathcal{B}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for all words $u, v \in \mathcal{A}^*$. The morphism ψ can be naturally extended to a sequence $\mathbf{v} = v_0v_1v_2\cdots$ over \mathcal{A} by setting $\psi(\mathbf{v}) = \psi(v_0)\psi(v_1)\psi(v_2)\cdots$. If a morphism $\psi : \mathcal{A}^* \rightarrow \mathcal{A}^*$ satisfies $\psi(a) \neq \varepsilon$ for every letter $a \in \mathcal{A}$ and $\psi(b) = bw$ for some $b \in \mathcal{A}$ and $w \in \mathcal{A}^*, w \neq \varepsilon$, then there exists a sequence \mathbf{v} having the prefix $\psi^n(b)$ for every $n \in \mathbb{N}$, thus $\psi(\mathbf{v}) = \mathbf{v}$ and \mathbf{v} is a *fixed point* of ψ . In the sequel, we use the notation $\psi^\omega(b)$. A morphism $\psi : \mathcal{A}^* \rightarrow \mathcal{A}^*$ is *primitive* if there exists $k \in \mathbb{N}$ such that $\psi^k(a)$ contains all letters of \mathcal{A} for each letter $a \in \mathcal{A}$. It is known that a fixed point of a primitive morphism is uniformly recurrent [11].

Let \mathbf{u} be a sequence over an alphabet \mathcal{A} and let $\psi : \mathcal{A}^* \rightarrow \mathcal{B}^*$ be a morphism. Consider a factor w of $\psi(\mathbf{u})$. We say that (w_1, w_2) is a *synchronizing point* of w if $w = w_1w_2$ and for all $p, s \in \mathcal{L}(\psi(\mathbf{u}))$ and $v \in \mathcal{L}(\mathbf{u})$ such that

$\psi(v) = pws$ there exists a factorization $v = v_1v_2$ of v with $\psi(v_1) = pw_1$ and $\psi(v_2) = w_2s$. We denote the synchronizing point by $w_1 \bullet w_2$.

Let u be a word of length $q = |u| \geq 1$. If z of length p is a prefix of u^ω , we say that q is a *period* of z and write $z = u^r$, where $r = \frac{p}{q}$. For instance, the Czech word **kačka**² has period 3 and $\text{kačka} = (\text{kač})^{\frac{5}{3}}$. The *critical exponent* of a sequence \mathbf{v} is defined as

$$E(\mathbf{v}) = \sup\{r \in \mathbb{Q} : u^r \text{ is a non-empty factor of } \mathbf{v}\}.$$

Its asymptotic version is called the *asymptotic critical exponent*, denoted $E^*(\mathbf{v})$. If $E(\mathbf{v}) = \infty$, then $E^*(\mathbf{v}) = \infty$. Otherwise, it is defined as

$$E^*(\mathbf{v}) = \lim_{n \rightarrow \infty} \sup\{r \in \mathbb{Q} : u^r \text{ is a non-empty factor of period } \geq n \text{ of } \mathbf{v}\}.$$

To compute the (asymptotic) critical exponent of a uniformly recurrent aperiodic sequence, it suffices to know bispecial factors and their shortest return words.

Theorem 3 ([12], Theorem 3). *Let \mathbf{v} be a uniformly recurrent aperiodic sequence. Let $(w_n)_{n \in \mathbb{N}}$ be the sequence of all bispecial factors in \mathbf{v} ordered by length. For every $n \in \mathbb{N}$, let r_n be the shortest return word to the bispecial factor w_n in \mathbf{v} . Then*

$$E(\mathbf{v}) = 1 + \sup\left\{\frac{|w_n|}{|r_n|} : n \in \mathbb{N}\right\} \quad \text{and} \quad E^*(\mathbf{v}) = 1 + \limsup_{n \rightarrow \infty} \frac{|w_n|}{|r_n|}.$$

3. Bispecial factors of the fixed point $\mathbf{u} = \xi^\omega(0)$

Let us fix $k \geq 4$ as Conjectures 1 and 2 are proven for $k < 4$. We denote $\mathbf{u} = \mathbf{u}_k$ the fixed point of $\xi = \xi_k$, defined in (1.1). Our first goal is to study bispecial factors and their shortest return words. This knowledge will enable us in the sequel to determine the first difference of complexity of $\mathbf{x} = \pi(\mathbf{u})$ (Conjecture 1) and using Theorem 3 we will show that there are no overlaps in \mathbf{u} , which is essential for the proof of Conjecture 2.

²kačka: Czech crown (informally), duck (in a Slovak dialect)

3.1. Properties of the fixed point \mathbf{u}

We denote \mathcal{A} the alphabet, i.e., $\mathcal{A} = \{0, 1, 2, \dots, (k-1), 0', 1', 2', \dots, (k-1)'\}$. In our proofs we will often use symmetry of the language of \mathbf{u} .

Observation 4. The language of \mathbf{u} is closed under the letter permutation $\ell \leftrightarrow \ell'$ for each $\ell \in \{0, 1, \dots, k-1\}$. Applying this exchange to every letter occurring in $w \in \mathcal{L}(\mathbf{u})$, we obtain the factor $w' \in \mathcal{L}(\mathbf{u})$ and call it *twin* of w .

Let us list some useful basic observations that follow from the form of the morphism ξ , defined in (1.1).

Observation 5. The following statements apply also for twins.

1. The synchronizing points of letters are $\bullet 0, 1\bullet, \bullet 2\bullet, \dots, \bullet(k-1)\bullet$.
2. $\xi^j(0) = 01 \cdots j$ for $j \in \{1, 2, \dots, k-1\}$.
3. $\xi^k(0) = 01 \cdots (k-1)0' = \xi^{k-1}(0)0'$.
4. If $\xi^n(0)$ ends in 0 or $0'$, then $n = mk$ for some $m \in \mathbb{N}$.
5. $\xi^{k-1}(1) = 0'$, hence $\xi^n(1)$ starts in $0'$ for each $n \geq k-1$.
6. The letter j occurs in \mathbf{u} only as a suffix of $01 \cdots j$ for $j \in \{1, 2, \dots, k-1\}$.
7. Except for its first occurrence, the letter 0 occurs in \mathbf{u} only as a suffix of the following factors:
 - $012 \cdots \ell 0$, where $\ell \in \{1, 2, \dots, (k-1)\}$;
 - $0'1'2' \cdots \ell' 0$, where $\ell' \in \{1', 2', \dots, (k-1)'\}$;
 - $012 \cdots (k-1)0'0$;
 - $0'1'2' \cdots (k-1)'00$.
8. The word 00 occurs in \mathbf{u} only as a prefix of 001.
9. The word $(k-1)0$ occurs in \mathbf{u} only as a prefix of $(k-1)01$.
10. Let $w \in \mathcal{L}(\mathbf{u})$, $w \neq \varepsilon$, and $\#\text{Rext}(w) \geq 2$. Then $\#\text{Rext}(w) \leq 3$ and 0 or $0'$ belong to $\text{Rext}(w)$. If $\#\text{Rext}(w) = 3$, then $\{0, 0'\} \subset \text{Rext}(w)$.

3.2. Klouda's method for determining bispecial factors

To describe the structure of all bispecial factors, we use a method by Klouda [13]. This method does not work directly with bispecial factors, but with bispecial triplets.

The first part of a triplet T is an element of the so-called left forky set F_L , the last part of the triplet is an element of the so-called right forky set F_R and the middle part is a bispecial factor itself.

Forky sets The left forky set F_L is formed by a finite number of unordered pairs of words $\{s_L, t_L\}$ from \mathcal{A}^* . Analogously, the elements of the right forky set F_R are denoted $\{s_R, t_R\}$. Consult Definition 20 in [13] to see that for our substitution ξ we can take the forky set F_R and F_L as follows.

- The right forky set F_R is formed by pairs $\{s_R, t_R\}$ of two kinds:
The first kind of pairs together with their twins are as follows (we consider all $\ell \in \{1, 2, \dots, k-2\}$):

$$\begin{array}{llll} \{0, (k-1)'0\}, & \{01, (k-1)'0'\}, & \{00', (k-1)'0'\}, & \{00, (k-1)'0'\}, \\ \{0, \ell\}, & \{0, \ell'0'\}, & \{0, \ell'0\}, & \{0, \ell'(\ell+1)'\}, \\ \{0, 0'\}, & \{0, k-1\}; & & \end{array}$$

The second kind of pairs in F_R together with their twins are as follows (we consider all $\ell, m \in \{1, 2, \dots, k-1\}$):

$$\begin{array}{l} \{\ell, m'\}, \\ \{\ell 0, m0\}, \quad \{\ell(\ell+1), m\}, \quad \{\ell 0, m0'\}, \quad \{\ell 0', m\} \quad \text{for } \ell < m, \\ \{\ell 0, m(m+1)\} \quad \text{for } \ell < m \leq k-2. \end{array}$$

- $F_L = \{\{a, b\} : a, b \in \mathcal{A}, a \neq b\}$.

Let us stress two important properties of F_R : the first letter of s_R is distinct from the first letter of t_R and for every pair $\{s_R, t_R\} \in F_R$ there exists a unique pair $\{\tilde{s}_R, \tilde{t}_R\} \in F_R$ such that $p\tilde{s}_R$ is a prefix of $\xi(s_R)$ and $p\tilde{t}_R$ is a prefix of $\xi(t_R)$ with $p = \text{lcp}\{\xi(s_R), \xi(t_R)\}$. The unique pair $\{\tilde{s}_R, \tilde{t}_R\}$ is denoted $f_R(\{s_R, t_R\})$. Note that for every pair $\{s_R, t_R\} \in F_R$ of the second kind, $\text{lcp}\{\xi(s_R), \xi(t_R)\} = \varepsilon$.

The set F_L satisfies the analogous properties which allow us to define $f_L(\{s_L, t_L\})$. Here, $\text{lcs}\{\xi(s_L), \xi(t_L)\}$ is empty for every pair from F_L .

Bispecial triplets Given the sets F_L and F_R , we call *bispecial triplet* a triplet $T = (\{s_L, t_L\}, w, \{s_R, t_R\})$ such that

- $\{s_L, t_L\} \in F_L$ and $\{s_R, t_R\} \in F_R$;
- both $s_L w s_R$ and $t_L w t_R$ are factors of \mathbf{u} . Obviously, w is a bispecial factor of \mathbf{u} .

Transformation of triplets We define a mapping Φ on the set of bispecial triplets

$$\Phi(\{s_L, t_L\}, w, \{s_R, t_R\}) = (\{\tilde{s}_L, \tilde{t}_L\}, \tilde{w}, \{\tilde{s}_R, \tilde{t}_R\}),$$

where $\{\tilde{s}_L, \tilde{t}_L\} = f_L(\{s_L, t_L\})$, $\{\tilde{s}_R, \tilde{t}_R\} = f_R(\{s_R, t_R\})$ and

$$\tilde{w} = \text{lcs}\{\xi(s_L), \xi(t_L)\} \varphi(w) \text{lcp}\{\xi(s_R), \xi(t_R)\}. \quad (3.1)$$

Clearly, $\Phi(\{s_L, t_L\}, w, \{s_R, t_R\})$ is a bispecial triplet, too.

Theorem 6. ([13], Theorem 36) *Let \mathbf{u} be a fixed point of a primitive morphism φ over \mathcal{A} and let F_L be a left forky set and F_R a right forky set. Then there exists a finite set \mathcal{I} of bispecial triplets such that every bispecial factor of \mathbf{u} occurs as the middle element of a triplet $\Phi^n(T)$ for some $n \in \mathbb{N}$ and $T \in \mathcal{I}$.*

As follows from the proof of the previous theorem, to find the set of initial triplets \mathcal{I} , it suffices to consider only triplets where the bispecial factor w does not have a synchronizing point. For our substitution ξ , the only bispecial factor without the synchronizing point is the empty word. Hence, the set \mathcal{I} contains only triplets of the form $T_I = (\{s_L, t_L\}, \varepsilon, \{s_R, t_R\})$. In our specific case of substitution ξ , the bispecial factor associated with the triplet $\Phi(T_I)$ by (3.1) equals $\tilde{w} = \text{lcs}\{\xi(s_L), \xi(t_L)\} \xi(\varepsilon) \text{lcp}\{\xi(s_R), \xi(t_R)\} = \text{lcp}\{\xi(s_R), \xi(t_R)\}$. Thus, only elements of F_R for which $\text{lcp}\{\xi(s_R), \xi(t_R)\} \neq \varepsilon$ give rise to non-empty bispecial factors. All such pairs or their twins are listed below.

$$\{0, (k-1)'0\}, \quad \{01, (k-1)'0'\}, \quad \{00', (k-1)'0'\} \quad \text{and} \quad \{00, (k-1)'0'\}. \quad (3.2)$$

By successively applying the transformation Φ to the above initial triplets, we will see in the next subsection that only elements $\{s_R, t_R\} \in F_R$ of the first kind appear in $\Phi^n(T_I)$.

3.3. List of bispecial triplets in \mathbf{u}

Now we can find all bispecial factors of \mathbf{u} . As we explained, in the pair $\{s_R, t_R\}$ of any initial triplet (up to the twin relation) the first digit of s_R must be 0 and the first digit of t_R must be $(k-1)'$. Hence the last digit of t_L must be $(k-2)'$. We will discuss four types of initial triplets T with the middle element $w = \varepsilon$ according with the pairs $\{s_R, t_R\}$ listed in (3.2).

We can observe for any initial bispecial triplet T that after a finite number of steps the bispecial triplet $\Phi^n(T)$ has the element from the right forky set $\{s_R, t_R\} = \{0, 0'\}$. In other words, every sufficiently long bispecial factor of \mathbf{u} has right extensions 0 and $0'$. Consequently, if w is the bispecial factor associated with the bispecial triplet $\Phi^n(T)$, then $\xi(w)$ is the bispecial factor associated with the bispecial triplet $\Phi^{n+1}(T)$.

In the role of $\{s_L, t_L\}$ we sometimes write $\{*, a\}$. By this notation we mean a pair where one element is $a \in \mathcal{A}$ and the other element is any letter of \mathcal{A} different from a .

Type 1 starting with $T = (\{*, (k-2)'\}, \varepsilon, \{0, (k-1)'0\})$:

$$\begin{aligned}\Phi(T) &= (\{*, (k-1)'\}, 0, \{1, 0\}) \\ \Phi^2(T) &= (\{*, 0\}, \xi(0), \{2, 0\}) \\ &\vdots \\ \Phi^{k-1}(T) &= (\{*, k-3\}, \xi^{k-2}(0), \{k-1, 0\}) \\ \Phi^k(T) &= (\{*, k-2\}, \xi^{k-1}(0), \{0', 0\}).\end{aligned}$$

Type 2 starting with $T = (\{*, (k-2)'\}, \varepsilon, \{01, (k-1)'0'\})$:

$$\begin{aligned}\Phi(T) &= (\{*, (k-1)'\}, 0, \{12, 0'\}) \\ \Phi^2(T) &= (\{*, 0\}, \xi(0), \{23, 0'\}) \\ &\vdots \\ \Phi^{k-1}(T) &= (\{*, k-3\}, \xi^{k-2}(0), \{(k-1)0', 0'\}) \\ \Phi^k(T) &= (\{*, k-2\}, \xi^{k-1}(0)0' = \xi^k(0), \{0', 1'\}) \\ &\vdots \\ \Phi^{2k-2}(T) &= (\{*, (k-4)'\}, \xi^{2k-2}(0), \{0', (k-1)'\}) \\ \Phi^{2k-1}(T) &= (\{*, (k-3)'\}, \xi^{2k-1}(0), \{0', 0\}).\end{aligned}$$

Type 3 starting with $T = (\{(k-1)', (k-2)'\}, \varepsilon, \{00', (k-1)'0'\})$:

$$\begin{aligned}\Phi(T) &= (\{0, (k-1)'\}, 0, \{10', 0'\}) \\ \Phi^2(T) &= (\{1, 0\}, \xi(0), \{20', 0'\}) \\ &\vdots \\ \Phi^{k-1}(T) &= (\{k-2, k-3\}, \xi^{k-2}(0), \{(k-1)0', 0'\}) \\ &\text{coincides with } \Phi^{k-1}(T) \text{ in Type 2.}\end{aligned}$$

Type 4 starting with $T = (\{(k-1)', (k-2)'\}, \varepsilon, \{00, (k-1)'0'\})$:

$$\begin{aligned}
\Phi(T) &= (\{0, (k-1)'\}, 0, \{10, 0'\}) \\
\Phi^2(T) &= (\{1, 0\}, \xi(0), \{20, 0'\}) \\
&\vdots \\
\Phi^{k-2}(T) &= (\{k-3, k-4\}, \xi^{k-3}(0), \{(k-2)0, 0'\}) \\
\Phi^{k-1}(T) &= (\{k-2, k-3\}, \xi^{k-2}(0), \{(k-1)0, 0'1'\}) \\
\Phi^k(T) &= (\{k-1, k-2\}, \xi^{k-1}(0)0' = \xi^k(0), \{0, 1'2'\}) \\
&\vdots \\
\Phi^{2k-2}(T) &= (\{(k-3)', (k-4)'\}, \xi^{2k-2}(0), \{0, (k-1)'0\}) \\
\Phi^{2k-1}(T) &= (\{(k-2)', (k-3)'\}, \xi^{2k-1}(0)0, \{1, 0\}) \\
&\vdots \\
\Phi^{3k-2}(T) &= (\{k-3, k-4\}, \xi^{3k-2}(0)\xi^{k-1}(0), \{0', 0\}).
\end{aligned}$$

Proposition 7. *Every bispecial factor of \mathbf{u} or its twin is equal to*

$$u = \xi^n(0) \quad \text{or} \quad v = \xi^{n+2k-1}(0)\xi^n(0) \quad \text{for some } n \in \mathbb{N}.$$

Moreover, $\text{Lext}(u) = \mathcal{A}$, while $\#\text{Lext}(v) = 2$.

Proof. The statement follows from the fact that all bispecial factors are associated with bispecial triplets. According to the description of bispecial triplets, see Types 1 to 4 (do not forget to consider twins), we know that $\xi^n(0)$ has all letters as left extensions and $\xi^{n+2k-1}(0)\xi^n(0)$, coming only from Type 4, has two left extensions. \square

3.4. The sequence \mathbf{u} is overlap-free

Let us first state a useful lemma on properties of the images of 0.

Lemma 8. *Consider the sequence $\mathbf{u} = \xi^\omega(0)$.*

1. $\xi^n(0)$ has synchronizing points $\bullet \xi^n(0) \bullet$ for each $n \in \mathbb{N}, n \geq 1, n \neq k$.
2. The shortest return word to $\xi^n(0)$ is $\xi^n(0)$ for each $n \leq k$.
3. $\xi^k(0)\xi^k(0)$ is always followed by $0'$.

Proof. 1. If $\xi^n(0)$ ends in $a \notin \{0, 0'\}$, then the statement is clear by Item 1 of Observation 5. If $\xi^n(0)$ ends in 0 or $0'$, then $n = mk$ for some $m \in \mathbb{N}$ by Item 4 of Observation 5. Observing the list of bispecial triplets, we can see that $\text{Rext}(\xi^n(0)) = \{0, 0'\}$ for $n \geq 2k-1$. This implies that for $m \geq 2$, the only right extensions of $\xi^{mk}(0)$ are 0 and $0'$, but then there are synchronizing points $\bullet \xi^{mk}(0) \bullet 0$ or $\bullet \xi^{mk}(0) \bullet 0'$.

2. $\xi^n(00)$ occurs in \mathbf{u} , so $\xi^n(0)$ is a return word to itself. By Item 2 and 3 of Observation 5 the factor $\xi^n(0)$ has all distinct letters. If there is a return word to $\xi^n(0)$ shorter than $\xi^n(0)$, then $\xi^n(0)$ is a (possibly fractional) power of this word, but then $\xi^n(0)$ cannot consist of mutually distinct letters. Consequently, $\xi^n(0)$ is the shortest return word to $\xi^n(0)$.
3. By Observation 5 we have $\xi^k(0)\xi^k(0) = \xi^{k-1}(0)0'\xi^{k-1}(0)0'$. By Item 1 $\xi^{k-1}(0)$ occurs only as the $(k-1)$ -st image of 0. Therefore $\xi^k(0)\xi^k(0) = \xi^{k-1}(010)0'$ occurs only as a prefix of the $(k-1)$ -st image of 0101 or 0100'. However, 0100' is not a factor of \mathbf{u} and 0101 has a unique right extension 2 (by Item 8 of Observation 5), hence $\xi^k(0)\xi^k(0)$ occurs only as a prefix of $\xi^{k-1}(01012) = \xi^k(0)\xi^k(0)0'1'$.

□

To show that \mathbf{u} does not contain overlaps, we will apply Theorem 3. For this purpose, we need some more knowledge on return words to bispecial factors.

Lemma 9. *The shortest return word r to a bispecial factor w in \mathbf{u} satisfies:*

- $r = w$ if $w = \xi^n(0)$ for $n \in \mathbb{N}$;
- $|r| \geq |w|$ if $w = \xi^{n+2k-1}(0)\xi^n(0)$ for $n \in \mathbb{N}$.

Proof. For $w = \xi^n(0)$, where $0 \leq n \leq k$, Item 2 of Lemma 8 implies that the shortest return word to w equals w (moreover, each return word to w contains w as its prefix). If $n > k$, then w has synchronizing points $\bullet w \bullet$ by Item 1 of Lemma 8. If rw is a complete return word to such bispecial factor w , then $rw = \xi(r'\xi^{n-1}(0))$, where r' is a return word to $\xi^{n-1}(0)$. Hence $r = \xi(r')$. This proves the first claim.

To get the second claim, let us explain that the factor $\xi^{2k-1}(0)$ is not a return word to $\xi^{2k-1}(0)0$ because $\xi^{2k-1}(0)\xi^{2k-1}(0)$ is always followed by $0'$. By the same arguments as above, we get that $\xi^{2k-1}(0)\xi^{2k-1}(0)$ occurs only as a prefix of the $(k-1)$ -st image of $\xi^k(0)\xi^k(0)$. By Item 3 of Lemma 8, $\xi^k(0)\xi^k(0)$ is always followed by $0'$, which implies that $\xi^{2k-1}(0)\xi^{2k-1}(0)$ is always followed by $0'$. Consequently, $\xi^{2k-1}(0)0$ is a prefix of any return word to $\xi^{2k-1}(0)0$. Thus, using Item 1 of Lemma 8, every return word to $w = \xi^{n+2k-1}(0)\xi^n(0)$, where $n \neq k$, is the n -th image of $\xi^{2k-1}(0)0$ and has therefore w as prefix. Moreover, for $n = k$, the factor $w = \xi^{3k-1}(0)\xi^k(0) = \xi^{3k-1}(0)\xi^{k-1}(0)0'$ occurs only as the prefix of the $(k-1)$ -st image of $\xi^{2k}(0)00'$

or $\xi^{2k}(0)01$. Since the last letter of $\xi^{2k}(0)$ is 0, the word $\xi^{2k}(0)00'$ is not a factor of \mathbf{u} by Item 8 of Observation 5. Consequently, $w = \xi^{3k-1}(0)\xi^k(0)$ also occurs only as the k -th image of $\xi^{2k-1}(0)0$ and each return word to w has w as its prefix. The second claim then follows. \square

Proposition 10. *The sequence $\mathbf{u} = \xi^\omega(0)$ does not contain overlaps and contains squares of unbounded size. Hence, the critical and asymptotic critical exponent of \mathbf{u} equal 2.*

Proof. There exist squares of unbounded size since $(\xi^n(0))^2$ is a factor of \mathbf{u} for any $n \in \mathbb{N}$.

To show that there are no overlaps, using Theorem 3, it is sufficient to show that $\frac{|w|}{|r|} \leq 1$ for any bispecial factor w of \mathbf{u} and its shortest return word r . This follows from Lemma 9. \square

4. Properties of the projection π

Let us recall that we fix $k \geq 4$. We denote $\xi = \xi_k$ and $\pi = \pi_k$ the morphisms defined in (1.1), and $\mathbf{u} = \mathbf{u}_k$ the fixed point of ξ and $\mathbf{x} = \mathbf{x}_k = \pi(\mathbf{u})$. The language of \mathbf{x} inherits some symmetry from the language of $\mathbf{u} = \xi^\omega(0)$. Once again, we use the notion of twin.

Observation 11. Since the language of \mathbf{u} is closed under the letter permutation $\ell \leftrightarrow \ell'$ for each $\ell \in \{0, 1, \dots, k-1\}$ and since $\pi(\ell') = 1 - \pi(\ell)$ for each $\ell \in \{0, 1, \dots, k-1\}$, the language of \mathbf{x} is closed under the letter exchange $0 \leftrightarrow 1$. If w' is obtained by this letter exchange from $w \in \mathcal{L}(\mathbf{x})$, we call w' a *twin* of w .

First, we will examine for how long two words can have the same projection without matching letters in the same positions. An example of such words $u, v \in \mathcal{L}(\mathbf{u})$ with $\pi(u) = \pi(v)$ are $u = u_1u_2u_3 = 201$ and $v = v_1v_2v_3 = 0'1'0'$.

We will use the following observation.

Observation 12. Let $w \in \mathcal{L}(\mathbf{u})$. If $\pi(w) = 0101$, then $w = 0101$.

Proof. Denote $w = w_1w_2w_3w_4$. Using Observation 5 and the definition of π , we have $w_3w_4 = 01$ or $w_4 = 0'$. Similarly, we find that either $w_2w_3 = 0'1'$ or $w_3 = 0$, and that either $w_1w_2 = 01$ or $w_2 = 0'$. Now, we use Observation 5 to rule out some combinations of those cases:

- If $w_3w_4 = 00'$, then Item 7 of Observation 5 implies $w_2 = (k-1)'$, a contradiction.
- If $w_2w_3w_4 = 0'1'0'$, then it must be the image of $0'0'$ and must thus be preceded by the image of $(k-1)$ by the same Item, so $w_1 = 0'$, a contradiction.
- If $w_2w_3w_4 = 0'01$, then the same Item forces again $w_1 = (k-1)$, a contradiction.

The only remaining possibility is $w = 0101$, which concludes the proof. \square

Lemma 13. *Let $u = u_1u_2 \cdots u_n$ and $v = v_1v_2 \cdots v_n$ belong to $\mathcal{L}(\mathbf{u})$. Assume that $\pi(u) = \pi(v)$ and the letters $u_\ell, v_\ell \in \mathcal{A}$ satisfy $u_\ell \neq v_\ell$ for every $\ell = 1, 2, \dots, n$. Then $n \leq k+1$.*

If, moreover, $u0$ and $v0$ belong to $\mathcal{L}(\mathbf{u})$, or $u0'$ and $v0'$ belong to $\mathcal{L}(\mathbf{u})$, then $n \leq k-1$ and $\pi(u) = \pi(v)$ equals 1^n or 0^n .

Proof. First assume that the last letters u_n and v_n belong to $\{1, 2, \dots, k-1\}$. Say $u_n = i$ and $v_n = j$, where $1 \leq i < j \leq k-1$. Item 6 of Observation 5 gives that $\pi(u) = \pi(v)$ is a suffix of 1^i . Clearly, $n = |\pi(u)| < k-1$. Analogously, if u_n and v_n belong to $\{1', 2', \dots, (k-1)'\}$, we get that $\pi(u) = \pi(v)$ is a suffix of 0^i for some $i < k-1$.

Let $u_n = i \in \{1, 2, \dots, k-1\}$ and $v_n = 0'$.

Case $\pi(u_n) = \pi(u_{n-1}) = 1$.

We consider $n \geq 3$ (if $n \leq 2$, there is no need to prove anything). By Observation 5, necessarily $i \geq 2$ and $v_{n-1} \in \{0', \ell\}$ for some $\ell \in \{1, 2, \dots, k-1\}$, $\ell \neq i-1$. The same observation implies that

- $u_{n-2}u_{n-1}u_n = (i-2)(i-1)i$ and $\pi(u_{n-2}u_{n-1}u_n)$ is a suffix of 01^i ;
- $v_{n-2}v_{n-1}v_n = (k-1)0'0'$, resp. $(\ell-1)\ell 0'$ and $\pi(v_{n-2}v_{n-1}v_n)$ is a suffix of 01^{k+1} , resp. $01^{\ell+1}$.

Since $\pi(u) = \pi(v)$, we deduce that $\pi(u) = \pi(v) = 1^j$ for some $j \leq i$ and thus $n = |\pi(u)| = j \leq i \leq k-1$.

Case $\pi(u_n) \neq \pi(u_{n-1}) = 0$.

Item 7 of Observation 5 implies that $v_{n-1}v_n$ equals $00'$ or $\ell'0'$ for some $\ell \in \{1, 2, \dots, k-1\}$. Note that this situation cannot happen if $v0$ or

$v0'$ is from $\mathcal{L}(\mathbf{u})$, since by the same observation, none of the factors $\ell'0'0$, $00'0'$, $\ell'0'0'$, $00'0$ belong to the language of \mathbf{u} . Consequently, the second part of the lemma is already shown.

To complete the proof, i.e., to show that $n \leq k + 1$, we assume that $n \geq 4$ (otherwise there is nothing to prove). By Observation 12, $\pi(u)$ cannot have a suffix 0101. Hence either $\pi(u_{n-1}) = \pi(u_{n-2})$ or $\pi(u_{n-2}) = \pi(u_{n-3})$.

We apply the already proved statement to $\tilde{n} = n - 1$ or to $\tilde{n} = n - 2$ and we deduce that $\tilde{n} \leq k - 1$. In particular, $n \leq k + 1$, as desired.

□

Lemma 14. *Let $w \neq \varepsilon$ be a bispecial factor of \mathbf{u} . Assume that there exist $x, y \in \text{Rext}(w)$ such that $x \neq y$ and $\pi(x) = \pi(y)$. Then w or its twin equals $\xi^\ell(0)$, where $0 \leq \ell \leq 2k - 2$ and $\ell \neq k - 1$.*

Moreover, if $u, v \in \mathcal{A}^$ are such that $wxu, wyv \in \mathcal{L}(\mathbf{u})$ and $\pi(wxu) = \pi(wyv)$, then*

- $|u| = |v| \leq 1$, if $0 \leq \ell \leq k - 2$;
- $|u| = |v| = 0$, if $k \leq \ell \leq 2k - 2$.

Proof. Recall that in Section 3.3 we have obtained a complete description of the bispecial triplets in \mathbf{u} with a nonempty bispecial factor. Using this description, we may also find all possible right extensions of these factors. The following observations rule out many possibilities for w .

- Considering Type 4, we note that the bispecial factor $w = \xi^{n+2k-1}(0)\xi^n(0)$ has two right extensions $x = 0$ and $y = n + 1$ for $n \in \{0, \dots, k - 2\}$ and it has two right extensions $x = 0$ and $y = 0'$ for $n \geq k - 1$, therefore $\pi(x) \neq \pi(y)$.
- Considering Types 1 to 3, we can see that the bispecial factor $w = \xi^n(0)$ for $n \geq 2k - 1$ has two right extensions $x = 0$ and $y = 0'$, hence again $\pi(x) \neq \pi(y)$.
- Considering Type 1, the right extensions of $w = \xi^{k-1}(0)$ are only $x = 0$ and $y = 0'$, thus once more $\pi(x) \neq \pi(y)$.

The first part of the statement is thus already proven, as the x, y from the statement exist only when w or its twin is $\xi^\ell(0)$ for $0 \leq \ell \leq k-2$ or $k \leq \ell \leq 2k-2$.

To prove the second part of the statement, we examine the right extensions of these words. These extensions are listed in Table 1. Letters after a '-' character can be uniquely deduced from what stands before. We used the fact that the factors $(k-1)'0'$ and 00 in \mathbf{u} are always followed by $1'$ and 1 respectively, compare with Items 9 and 8 Observation 5. Note also that s_R from Type 1 can be discarded as it is a prefix of the s_R of Type 2. Similarly, t_R from Type 1 can be discarded for $1 \leq \ell \leq k-2$ as the other right extensions are all projected to the other letter, and for $k \leq \ell \leq 2k-2$ as it is equal to s_R from Type 4. Examining Table 1, we see that the right extensions of these w cannot exceed the stated lengths while keeping equal projections, as expected. \square

w	$w t_R - \tilde{u}$ (Type 2)	$w s_R$ (Type 2)	$w s_R$ (Type 3)	$w s_R - \tilde{v}$ (Type 4)
ε	$w (k-1)' 0' - 1'$	$w 01$	$w 00'$	$w 00-1$
0	$w 0' - 1' 2'$	$w 12$	$w 10'$	$w 10-1$
$\xi(0)$	$w 0' - 1' 2'$	$w 23$	$w 20'$	$w 20-1$
\vdots	\vdots	\vdots	\vdots	\vdots
$\xi^{k-2}(0)$	$w 0' - 1' 2'$	$w (k-1)0'$	$w (k-1)0'$	$w (k-1)0-1$
$\xi^k(0)$	$w 1' 2'$			$w 0-1$
$\xi^{k+1}(0)$	$w 2' 3'$			$w 0-1$
\vdots	\vdots			\vdots
$\xi^{2k-2}(0)$	$w (k-1)' 0$			$w 0-1$

Table 1: Right extensions of bispecial factors in the case when the first letters of s_R and t_R have the same projection by π .

Lemma 15. *Let $\ell \in \mathbb{N}$ satisfy $2 \leq \ell \leq k-2$ or $k+2 \leq \ell \leq 2k-2$ and $w = \xi^\ell(0)$. Assume that*

1. $x, y \in \text{Rext}(w)$ are such that $x \neq y$ and $\pi(x) = \pi(y)$ and
2. $u, v \in \mathcal{A}^*$ are such that $uw x, vw y \in \mathcal{L}(\mathbf{u})$ and $\pi(u) = \pi(v)$ is a prefix of 1^{k-1} or 0^{k-1} .

Then $|u| = |v| \leq k - 3$.

Proof. For bispecial factors $w = \xi^\ell(0)$ with ℓ specified in the assumptions of the lemma, Table 2 shows all possible right extensions x, y which satisfy Item 1.

If $2 \leq \ell \leq k - 2$, the table shows that the letter $(\ell - 1)$ is the left extension of the factor $w0'$. By Observation 5, the factor $w0'$ is preceded by $01 \cdots (\ell - 2)(\ell - 1)$. As $\pi(01 \cdots (\ell - 2)(\ell - 1)) = 01^{\ell-1}$, the factor u satisfying Item 2 of the lemma is of length at most $\ell - 1 \leq k - 3$.

If $k + 2 \leq \ell \leq 2k - 2$, Table 2 says that the factor $w0$ has the left extension $j' := (\ell - k - 1)'$, and therefore $w0$ is preceded by the factor $0'1' \cdots (j - 1)'j'$. As $\pi(0'1' \cdots (j - 1)'j') = 10^j$, the factor v satisfying Item 2 is of length at most $j = \ell - k - 1 \leq k - 3$. \square

w	$t_L w t_R - \tilde{u}$	$s_L w y$
ε	$(k-2)' w (k-1)'0'$	$* w 0$
0	$(k-1)' w 0'-1'$	$* w 1$
$\xi(0)$	$0 w 0'-1'$	$* w 2$
$\xi^2(0)$	$1 w 0'-1'$	$* w 3$
\vdots	\vdots	\vdots
$\xi^{k-2}(0)$	$(k-3) w 0'1'$	$* w (k-1)$
$\xi^k(0)$	$(k-2) w 1'2'$	$(k-1) w 0$
$\xi^{k+1}(0)$	$(k-1) w 2'3'$	$0' w 0$
$\xi^{k+2}(0)$	$0' w 3'4'$	$1' w 0$
\vdots	\vdots	\vdots
$\xi^{2k-2}(0)$	$(k-4)' w (k-1)'0$	$(k-3)' w 0$

Table 2: Left and right extensions of bispecial factors $\xi^\ell(0)$ for $0 \leq \ell \leq 2k - 2$, $\ell \neq k - 1$, where y stands for the first letter of s_R , $*$ is any letter from \mathcal{A} different from t_L . Notation $t_L w t_R - \tilde{u}$ means that the factor $t_L w t_R$ is in \mathbf{u} always followed by \tilde{u} .

5. Proof of Conjecture 1

We have prepared everything to reveal a close relation between left special factors in \mathbf{u} and \mathbf{x} , which is crucial for the proof of Conjecture 1.

Proposition 16. *Let w be a left special factor of \mathbf{x} such that $|w|$ is at least $|\xi^{2k-2}(0)| + k + 2$. Then there exist $i \in \mathbb{N}, 1 \leq i \leq k$, and a left special factor f of \mathbf{u} such that w or its twin equals $w = 0^{i-1}\pi(f)$.*

Proof. Denote $n = |w|$. As $0w$ and $1w$ belong to $\mathcal{L}(\mathbf{x})$, there exist factors $u = u_1u_2 \cdots u_n$ and $v = v_1v_2 \cdots v_n$ and letters a, b such that $au, bv \in \mathcal{L}(\mathbf{u})$ and $0w = \pi(au)$ and $1w = \pi(bv)$. We denote

$$i = \min\{\ell : u_\ell = v_\ell\} \quad \text{and} \quad j = \max\{\ell : u_iu_{i+1} \cdots u_\ell = v_iv_{i+1} \cdots v_\ell\}.$$

By Lemma 13, $i \leq k + 2 \leq n$. Thus the letter $u_i = v_i$ has two left extensions. In particular, by Observation 5, $u_i \in \{0, 0'\}$. Again, by Lemma 13, second part, $i \leq k$ and $\pi(u_1u_2 \cdots u_{i-1}) = 1^{i-1}$ or 0^{i-1} . The definition of i implies that the factor $f = u_iu_{i+1} \cdots u_j = v_iv_{i+1} \cdots v_j$ is a left special factor in $\mathcal{L}(\mathbf{u})$. To prove the proposition we just need to prove that $j = n$.

Assume the contrary, i.e., that $j < n$. Then f is also a right special factor with two right extensions $u_{j+1} \neq v_{j+1}$ and $\pi(u_{j+1}) = \pi(v_{j+1})$. By Lemma 14, the length $j - i + 1$ of f is at most $|\xi^{2k-2}(0)|$. This implies

$$n \geq |\xi^{2k-2}(0)| + k + 2 \geq j - i + 1 + k + 2 \geq j + 3.$$

By the same lemma, $\pi(u_{j+2}) \neq \pi(v_{j+2})$ or $\pi(u_{j+3}) \neq \pi(v_{j+3})$, which contradicts the fact that $\pi(u) = \pi(v)$. \square

Proposition 17. *Let $f \in \mathcal{L}(\mathbf{u})$ be a left special factor of \mathbf{u} such that 0 is the first letter of f .*

1. *If f is a prefix of \mathbf{u} , then $0^{i-1}\pi(f)$ and $1^{i-1}\pi(f)$ are left special factors of \mathbf{x} for every $i \in \{1, 2, \dots, k\}$.*
2. *If f is not a prefix of \mathbf{u} , then there exist a unique $c \in \{0, 1\}$ and a unique $i \in \{1, 2, \dots, k\}$ such that $w = c^{i-1}\pi(f)$ is a left special factor in \mathbf{x} .*

Proof. Let us recall that by Observation 5 the following hold for every $j \in \{1, 2, \dots, k - 1\}$.

- The factor $j0$ occurs in \mathbf{u} only as a suffix of the factor $01 \cdots (j - 1)j0$.
- The factor $j'0$ occurs in \mathbf{u} only as a suffix of the factor $0'1' \cdots (j - 1)'j'0$.
- The factor 00 occurs in \mathbf{u} only as a suffix of the factor $0'1' \cdots (k - 1)'00$.

- The factor $0'0$ occurs in \mathbf{u} only as a suffix of the factor $01 \cdots (k-1)0'0$.

Let $a \in \mathcal{A}$ be a left extension of f . Since f starts with 0, the factor $\pi(af)$ occurs in \mathbf{x} only as a suffix of

- $01^j\pi(f)$, if $a = j \in \{1, \dots, k-1\}$,
- $10^j\pi(f)$, if $a = j' \in \{1', \dots, (k-1)'\}$,
- $10^k\pi(f)$, if $a = 0$,
- $01^k\pi(f)$, if $a = 0'$.

Let us stress that to distinct letters that could be left extensions of f , we assigned distinct factors in the previous list.

Let $w \in \mathcal{L}(\mathbf{u})$ be the shortest bispecial factor having f as its prefix.

1. If f is a prefix of \mathbf{u} , the bispecial factor w equals $\xi^n(0)$ for some $n \in \mathbb{N}$ and thus $\text{Lext}(w) = \text{Lext}(f) = \mathcal{A}$ due to Proposition 7. Using the previous list, we see that for every $\ell \in \{1, 2, \dots, k\}$,

$$01^{\ell-1}\pi(f), 11^{\ell-1}\pi(f), 10^{\ell-1}\pi(f), \text{ and } 00^{\ell-1}\pi(f)$$

all belong to $\mathcal{L}(\mathbf{x})$. Therefore, $1^{\ell-1}\pi(f)$ and $0^{\ell-1}\pi(f)$ are left special factors in \mathbf{x} .

2. If f is not a prefix of \mathbf{u} , the bispecial factor w equals $\xi^{n+2k-1}(0)\xi^n(0)$ for some $n \in \mathbb{N}$ and thus $\#\text{Lext}(w) = \#\text{Lext}(f) = 2$. Then $\pi(f)$ preceded by the longest common suffix of the factors corresponding to the two letters from $\text{Lext}(f)$ in the above list is a left special factor in \mathbf{x} .

□

The following theorem confirms the validity of Conjecture 1.

Theorem 18. *For sufficiently large N , the first difference of factor complexity of the sequence \mathbf{x} satisfies*

$$\Delta\mathcal{C}(N) \in \{4k-2, 4k\}.$$

Proof. By Formula (2.1), we have to determine the number of left special factors in \mathbf{x} of length N .

Since $\mathbf{u} = \xi^\omega(0)$, the bispecial factor $\xi^n(0)$ is a prefix of \mathbf{u} for every $n \in \mathbb{N}$ and thus any prefix of \mathbf{u} is a left special factor in \mathbf{u} . By Proposition 17, for every $i \in \{1, 2, \dots, k\}$ and every prefix f of \mathbf{u} of length $N - i + 1$, the words $0^{i-1}\pi(f)$ and $1^{i-1}\pi(f)$ are left special factors of length N in \mathbf{x} . As the language of \mathbf{x} is closed under exchange $0 \leftrightarrow 1$, the same fact is true for the twins of the above mentioned left special factors. It is easy to see that for $N > k$, all these left special factors differ. Thus, for every N we have at least $4k - 2$ left special factors in \mathbf{x} .

Let us consider a left special factor f of \mathbf{u} which is not a prefix of \mathbf{u} and such that 0 is the first letter of f . By Proposition 7, the shortest bispecial factor of \mathbf{u} with the prefix f has the form $v = \xi^{n+2k-1}(0)\xi^n(0)$. We will use the following properties of v .

- The longest common prefix of the bispecial factor $v = \xi^{n+2k-1}(0)\xi^n(0)$ and \mathbf{u} is $w = \xi^{n+2k-1}(0)$.

Indeed, on the one hand $w0$ is a prefix of v as 0 is a prefix of $\xi^n(0)$, and on the other hand we show that $w0'$ is a prefix of $\xi^{n+2k}(0)$ which is a prefix of \mathbf{u} . We have

$$\xi^{n+2k}(0) = \xi^{n+2k-1}(01) = w\xi^{n+2k-1}(1),$$

where by Item 5 of Observation 5, $0'$ is a prefix of $\xi^{n+2k-1}(1)$.

- $|v| + k - 1 < |\xi^{n+2k}(0)|$ for every $n \in \mathbb{N}$.

Indeed,

$$\begin{aligned} |v| + k - 1 &= |\xi^{n+2k-1}(0)\xi^n(0)| + k - 1 \\ &\leq |\xi^{n+2k-1}(0)\xi^n(01 \dots (k-1))| \\ &< |\xi^{n+2k-1}(0)\xi^{n+k}(0)| = |\xi^{n+2k-1}(0)\xi^{n+k}(0')| \\ &= |\xi^{n+2k-1}(0)\xi^{n+2k-1}(1)| = |\xi^{n+2k}(0)|. \end{aligned}$$

We use the first property of $v = \xi^{n+2k-1}(0)\xi^n(0)$ mentioned above and the constant c from Item 2 of Proposition 17. They imply that for every N such that $|\xi^{n+2k-1}(0)| + c \leq N \leq |v| + c - 1$, we have a new left special factor of length N . Since its twin is left special, too, we have to add 2 to the number of left special factors of length N .

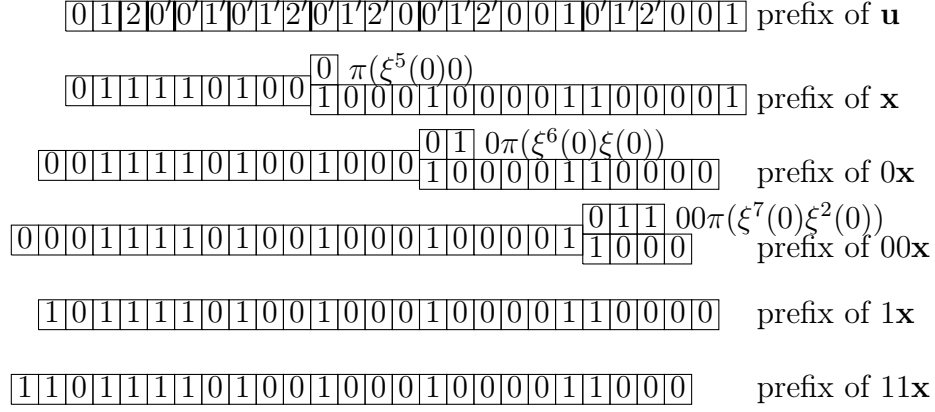


Figure 1: All left special factors of \mathbf{x}_k , where $k = 3$, of length N with $10 \leq N \leq 25$, are prefixes of the depicted factors or their twins.

For every $n \in \mathbb{N}$, we have to consider left special factors in \mathbf{x} arising from the bispecial factor $\xi^{n+2k-1}(0)\xi^n(0)$. The second property of v implies that the lengths of left special factors arising from $\xi^{n+2k-1}(0)\xi^n(0)$ are strictly smaller than the length of those arising from $\xi^{n+2k}(0)\xi^{n+1}(0)$.

Proposition 16 confirms that we have already accounted for all possible left special factors of length N .

□

The structure of left special factors is illustrated in Figure 1. In the proofs, we limit our consideration to $k \geq 4$ because Conjecture 1 has been already proven for $k \leq 3$. Nevertheless, due to space constraints, Figure 1 exhibits the situation for $k = 3$.

6. Proof of Conjecture 2

Using the fact that \mathbf{u} contains no overlaps, see Proposition 10, and our knowledge of the left and right extensions of bispecial factors in \mathbf{u} , see Lemma 14, we are ready to prove Conjecture 2.

Theorem 19. *The sequence \mathbf{x} has critical exponent $k + 1$, which is attained by the words 0^{k+1} and 1^{k+1} . It contains no factor of length $2n + k$ and period n , and therefore has asymptotic critical exponent equal to 2.*

Proof. By Observation 5, the words $1^{k+1}, 0^{k+1}$ are in $\mathcal{L}(\mathbf{x})$, but $1^{k+2}, 0^{k+2}$ are not. Therefore, we consider $n \geq 2$.

Assume the contrary to the statement of the theorem, i.e., that there exists in \mathbf{x} an overlap of size k . Equivalently, denoting $m = n + k$, there exists a factor of \mathbf{u} of length $2m - k$, say,

$$w = w_1 w_2 \cdots w_{m-k} w_{m-k+1} \cdots w_m w_{m+1} \cdots w_{2m-k} \in \mathcal{L}(\mathbf{u})$$

such that $m \geq k + 2$ and the projection by π of the prefix of w of length m coincides with the projection by π of the suffix of w of length m , formally,

$$\pi(w_1 w_2 \cdots w_{m-k} w_{m-k+1} \cdots w_m) = \pi(w_{m-k+1} \cdots w_m w_{m+1} \cdots w_{2m-k}). \quad (6.1)$$

Let us denote

$$i = \min\{\ell : w_\ell = w_{m-k+\ell}\}.$$

Lemma 13 implies $i \leq k + 2$. If i were at least 2, then the letter $w_i = w_{m-k+i}$ would have two distinct letters in its left extension, namely w_{i-1} and $w_{m-k+i-1}$. Hence, w_i is either 0 or 0'. Put $u = w_1 w_2 \cdots w_{i-1}$ and $v = w_{m-k+1} w_{m-k+2} \cdots w_{m-k+i-1}$. Clearly, $\pi(u) = \pi(v)$. Since $u0', v0' \in \mathcal{L}(\mathbf{u})$ or $u0, v0 \in \mathcal{L}(\mathbf{u})$, Lemma 13 implies that $i - 1 \leq k - 1$. We have shown

$$\text{either } i = 1 \quad \text{or} \quad i \leq k. \quad (6.2)$$

Denote

$$j = \max\{\ell : w_i w_{i+1} \cdots w_\ell = w_{m-k+i} w_{m-k+i+1} \cdots w_{m-k+\ell}\}.$$

Obviously, $i \leq j \leq m$. If j were at least $m - k + i$, then $w_i w_{i+1} \cdots w_{m-k+j}$ would be an overlap in \mathbf{u} , which contradicts Proposition 10. From now on, we assume that

$$j \leq m - k + i - 1. \quad (6.3)$$

If j were equal to m , then the previous inequality would force $i \geq k + 1$ – this contradicts (6.2).

Necessarily, $j \leq m - 1$. The definition of j forces $w_{j+1} \neq w_{m-k+j+1}$. In particular,

$$w_i w_{i+1} \cdots w_j = w_{m-k+i} w_{m-k+i+1} \cdots w_{m-k+j} \quad \text{is a right special factor in } \mathbf{u},$$

with two extensions $w_{j+1} \neq w_{m-k+j+1}$ having the same projection by π . This means that $f := w_i w_{i+1} \cdots w_j$ is a suffix of a bispecial factor $\xi^\ell(0)$ or its twin considered in Lemma 14.

Due to (6.1), we also have

$$\pi(w_i w_{i+1} \cdots w_m) = \pi(w_{m-k+i} w_{m-k+i+1} \cdots w_{2m-k}).$$

Using Lemma 14 with $w = w_i w_{i+1} \cdots w_j = w_{m-k+i} w_{m-k+i+1} \cdots w_{m-k+j}$ and xu prefix of $w_{j+1} \cdots w_m$ and yv prefix of $w_{m-k+j+1} \cdots w_{2m-k}$, since $|u| = |v| \leq 1$, we have $j+2 \geq m$, which leads to $j \geq m-2$.

If i were at most $k-2$, then $j \geq m-2 = m-k+k-2 \geq m-k+i$, which again contradicts (6.3). From now on we may restrict our consideration to the values i, j such that

$$k-1 \leq i \leq k \quad \text{and} \quad m-2 \leq j \leq m-1. \quad (6.4)$$

Since $k \geq 4$, we have $i \geq 3$. The definition of i implies $w_{i-1} \neq w_{m-k+i-1}$. Therefore, the factor

$$f = w_i \cdots w_j = w_{m-k+i} \cdots w_{m-k+j} \quad \text{is bispecial.} \quad (6.5)$$

More precisely, f or its twin equals $\xi^\ell(0)$ for some $\ell \in \{0, 1, \dots, 2k-2\} \setminus \{k-1\}$.

As $\pi(w_1 \cdots w_j w_{j+1}) = \pi(w_{m-k+1} \cdots w_{m-k+j} w_{m-k+j+1})$, one can derive $\pi(ufx) = \pi(vfy)$ denoting

$$u = w_1 \cdots w_{i-1}, \quad x = w_{j+1} \quad \text{and} \quad v = w_{m-k+1} \cdots w_{m-k+i-1}, \quad y = w_{m-k+j+1}.$$

Note that $\pi(u)$ is 0^{i-1} or 1^{i-1} from Lemma 13. Under the assumption $\ell \notin \{0, 1, k, k+1\}$, Lemma 15 forces $|u| = |v| = i-1 \leq k-3$. This contradicts (6.4). Consequently, it remains to treat $\ell \in \{0, 1, k, k+1\}$.

Let us recall that $f = w_i \cdots w_j$ and the index $m-k+i$ is also an occurrence of f in w . By (6.3) and (6.4)

$$j+1 \leq m-k+i = \underbrace{m-2-j}_{\leq 0} + j + \underbrace{i-k}_{\leq 0} + 2 \leq j+2,$$

i.e., the first occurrence of f ends at position j and simultaneously f occurs again at position $j+1$ or $j+2$. Note that $z_1 := w_{j+1}$ and $z_2 := w_{m-k+j+1}$ belong to the right extensions of f , $z_1 \neq z_2$ and $\pi(z_1) = \pi(z_2)$. Hence,

either (A): the index i is an occurrence of ffz_2 and z_1 is a prefix of f ;

or (B): the index i is an occurrence of fz_1fz_2 .

We look at Table 2 to find $\{z_1, z_2\}$, to determine what precedes fz if $z \in \{z_1, z_2\}$ and to determine what follows fz .

Case $f = 0$: $\{z_1, z_2\} = \{0', 1\}$. As none of z_1 and z_2 is a prefix of f , the situation (A) is excluded. By the table, the factor $f0' = 00'$ is preceded by $(k-1)'$ and followed by $1'$. Both possibilities for (B), namely $fz_1fz_2 = f0'f1$ and $fz_1fz_2 = f1f0'$ lead to a contradiction.

Case $f = \xi(0) = 01$: $\{z_1, z_2\} = \{0', 2\}$ and again (A) is excluded as 0 is a prefix of f . By Table 2, the factor $f0' = \xi(0)0'$ is preceded by 0 and followed by $1'$. Both possibilities $fz_1fz_2 = f2f0'$ and $fz_1fz_2 = f0'f2$ lead to a contradiction.

Case $f = \xi^k(0)$: $\{z_1, z_2\} = \{0, 1'\}$. By Table 2, the factor $f1' = \xi^k(0)1'$ is followed by $2'$. Hence, $z_1 \neq 1'$, i.e., $z_2 = 1'$. By the table, the factor $fz_2 = \xi^k(0)1'$ is preceded by $(k-2)$. This yields a contradiction, as the factor $f = \xi^k(0)$ has a suffix $0'$ and $z_1 = 0$.

Case $f = \xi^{k+1}(0)$: $\{z_1, z_2\} = \{0, 2'\}$. By Table 2, the factor $f2' = \xi^{k+1}(0)2'$ is followed by $3'$. Hence, $z_1 \neq 2'$, i.e., $z_2 = 2'$. But the factor $fz_2 = \xi^{k+1}(0)2'$ is preceded by $(k-1)$. This yields a contradiction, as the factor $f = \xi^{k+1}(0)$ has a suffix $1'$ and $z_1 = 0$.

We have proved that the assumption that there exists an overlap of size k in \mathbf{x} leads to a contradiction. Therefore, Conjecture 2 is proven. \square

7. Comments and open problem

- It is well known that the factor complexity \mathcal{C} of a coding of a fixed point of a primitive morphism is sublinear, i.e., there exists a constant c such that $\mathcal{C}(n) \leq cn$ for every $n \in \mathbb{N}$. The validity of Conjecture 1 shows that the constant c cannot be bounded even for binary words.
- Binary infinite words can have an arbitrarily large difference between the critical exponent and the asymptotic critical exponent. Sturmian sequences are examples of such sequences with language closed under reversal. Their (asymptotic) critical exponent may be computed using

the continued fraction expansion coefficients of their slope [14]: while the critical exponent is larger than any coefficient, sequences sharing the same suffix for the continued fraction of their slope have the same asymptotic critical exponent. The validity of Conjecture 2 implies that the Thue-Morse-like sequences serve as another example, this time of sequences with language closed under letter exchange.

- Let us fix $k \in \mathbb{N}$ and denote $x_0x_1x_2 \dots$ the Thue-Morse-like sequence $\mathbf{x}_k = \pi(\mathbf{u}_k)$ studied in this paper. In the case $k = 3$, Theorem 43 in [1] states that every $n \in \mathbb{N}$ can be written as a sum of two integers from the set $J_0 = \{i : x_i = 0\}$ and also as a sum of two integers from the set $J_1 = \{i : x_i = 1\}$. Our computer experiments suggest that this statement is valid for every sequence \mathbf{x}_k with $k \geq 2$, although it is known to be false for the Thue-Morse sequence \mathbf{x}_1 [15].

References

- [1] J. Shallit, The Narayana morphism and related words (2025). [arXiv:2503.01026](#).
- [2] E. Zeckendorf, Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas, Bulletin de la société royale des sciences de Liège 41 (1972) 179–182.
- [3] D. S. Meek, G. H. J. V. Rees, The solution of an iterated recurrence, The Fibonacci Quarterly 22 (2) (1984) 101–104. doi:10.1080/00150517.1984.12429900.
- [4] P. Letouzey, Generalized Hofstadter functions g, h and beyond: numeration systems and discrepancy (2025). [arXiv:2502.12615](#).
- [5] J.-P. Allouche, T. Johnson, Narayana’s Cows and Delayed Morphisms, in: Journées d’Informatique Musicale, île de Tatihou, France, 1996. URL <https://hal.science/hal-02986050>
- [6] J. Shallit, Subword complexity of the Fibonacci–Thue–Morse sequence: The proof of Dekking’s conjecture, Indagationes Mathematicae 32 (3) (2021) 729–735. doi:<https://doi.org/10.1016/j.indag.2021.03.004>.

- [7] J. Berstel, D. Perrin, The origins of combinatorics on words, *European Journal of Combinatorics* 28 (3) (2007) 996–1022. doi:<https://doi.org/10.1016/j.ejc.2005.07.019>.
- [8] S. Brlek, Enumeration of factors in the Thue-Morse word, *Discrete Applied Mathematics* 24 (1) (1989) 83–96. doi:[https://doi.org/10.1016/0166-218X\(92\)90274-E](https://doi.org/10.1016/0166-218X(92)90274-E).
- [9] A. de Luca, S. Varicchio, Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups, *Theoretical Computer Science* 63 (1989) 333–348. doi:[https://doi.org/10.1016/0304-3975\(89\)90013-3](https://doi.org/10.1016/0304-3975(89)90013-3).
- [10] J. Cassaigne, Complexité et facteurs spéciaux (Complexity and special factors), *Bulletin of the Belgian Mathematical Society - Simon Stevin* 4 (1) (1997) 67–88.
URL <http://eudml.org/doc/119937>
- [11] M. Queffélec, Substitution dynamical systems - spectral analysis, *Lecture Notes in Mathematics* 1294 (1987). doi:<https://doi.org/10.1007/BFb0081893>.
- [12] F. Dolce, L. Dvořáková, E. Pelantová, On balanced sequences and their critical exponent, *Theoretical Computer Science* 939 (2023) 18–47. doi:<https://doi.org/10.1016/j.tcs.2022.10.014>.
- [13] K. Klouda, Bispecial factors in circular non-pushy D0L languages, *Theoretical Computer Science* 445 (2012) 63–74. doi:<https://doi.org/10.1016/j.tcs.2012.05.007>.
- [14] D. Damanik, D. Lenz, The index of Sturmian sequences, *European Journal of Combinatorics* 23 (1) (2002) 23–29. doi:<https://doi.org/10.1006/eujc.2000.0496>.
- [15] A. Rajasekaran, J. Shallit, T. Smith, Additive number theory via automata theory, *Theory of Computing Systems* 64 (3) (2020) 542–567. doi:[10.1007/s00224-019-09929-9](https://doi.org/10.1007/s00224-019-09929-9).