



HAL
open science

On Generating Binary Words Palindromically

Tero Harju, Mari Huova, Luca Q. Zamboni

► **To cite this version:**

Tero Harju, Mari Huova, Luca Q. Zamboni. On Generating Binary Words Palindromically. Journal of Combinatorial Theory, Series A, 2015, 129, pp.142–159. <hal-01829174>

HAL Id: hal-01829174

<https://hal.science/hal-01829174v1>

Submitted on 3 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Manuscript Number:

Title: On Generating Binary Words Palindromically

Article Type: Article

Corresponding Author: Ms. Mari Huova,

Corresponding Author's Institution: University of Turku

First Author: Tero Harju

Order of Authors: Tero Harju; Mari Huova; Luca Q Zamboni

Abstract: We regard a finite word $u = u_1 u_2 \dots u_n$ up to word isomorphism as an equivalence relation on $\{1, 2, \dots, n\}$ where i is equivalent to j if and only if $u_i = u_j$. Some finite words (in particular all binary words) are generated by palindromic relations of the form $k \sim j+i-k$ for some choice of $1 \leq i \leq j \leq n$ and $k \in \{i, i+1, \dots, j\}$. That is to say, some finite words u are uniquely determined up to word isomorphism by the position and length of some of its palindromic factors. In this paper we study the function $\mu(u)$ defined as the least number of palindromic relations required to generate u . We show that if x is an infinite word such that $\mu(u) \leq 2$ for each factor u of x , then x is ultimately periodic. On the other hand, we establish the existence of non-ultimately periodic words for which $\mu(u) \leq 3$ for each factor u of x , and obtain a complete classification of such words on a binary alphabet (which includes the well known class of Sturmian words). In contrast, for the Thue-Morse word, we show that the function μ is unbounded.

On Generating Binary Words Palindromically

Tero Harju^a, Mari Huova^{a,b,*}, Luca Q. Zamboni^{a,c,1}

^a*Department of Mathematics and Statistics & FUNDIM, University of Turku, Finland*

^b*Turku Centre for Computer Science*

^c*Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 boulevard du 11 novembre 1918, F69622 Villeurbanne Cedex, France*

Abstract

We regard a finite word $u = u_1u_2 \cdots u_n$ up to word isomorphism as an equivalence relation on $\{1, 2, \dots, n\}$ where i is equivalent to j if and only if $u_i = u_j$. Some finite words (in particular all binary words) are generated by *palindromic* relations of the form $k \sim j+i-k$ for some choice of $1 \leq i \leq j \leq n$ and $k \in \{i, i+1, \dots, j\}$. That is to say, some finite words u are uniquely determined up to word isomorphism by the position and length of some of its palindromic factors. In this paper we study the function $\mu(u)$ defined as the least number of palindromic relations required to generate u . We show that if x is an infinite word such that $\mu(u) \leq 2$ for each factor u of x , then x is ultimately periodic. On the other hand we establish the existence of non-ultimately periodic words for which $\mu(u) \leq 3$ for each factor u of x , and obtain a complete classification of such words on a binary alphabet (which includes the well known class of Sturmian words). In contrast, for the Thue-Morse word, we show that the function μ is unbounded.

Keywords: Palindromes, Sturmian words, balanced words.

2000 MSC: 68R15

1. Introduction

Palindromes play an important role in various areas of mathematics including diophantine approximation and number theory (see for instance [1, 6]), discrete math (see for instance [13, 4]), algebra (see for instance [17, 7]), biomathematics (see for instance [16]), geometric symmetry in translation surfaces associated with various dynamical systems including interval exchange transformations (see for instance [11]) and theoretical physics in the spectral properties of discrete Schrödinger operators defined on quasicrystals (see for instance [15]).

Let \mathbb{A} be a finite non-empty set of *letters*. For each positive integer n , let $\mathcal{S}(n) = \{(i, j) \mid 1 \leq i \leq j \leq n\}$. Let $u = u_1u_2 \cdots u_n \in \mathbb{A}^n$ be a word of length n with values in \mathbb{A} .

*Corresponding author.

Email addresses: harju@utu.fi (Tero Harju), mahuov@utu.fi (Mari Huova), lupastis@gmail.com (Luca Q. Zamboni)

¹Partially supported by a FiDiPro grant (137991) from the Academy of Finland and by ANR grant SUBTILE.

Let $\text{Alph}(u) = \{u_i \mid 1 \leq i \leq n\}$ denote the subset of \mathbb{A} consisting of all letters occurring in u . For each pair $(i, j) \in \mathcal{S}(n)$, we denote by $u[i, j]$ the *factor* $u_i u_{i+1} \cdots u_j$. In case $i = j$, we write $u[i]$ in lieu of $u[i, i]$. Finally, let \mathcal{P} denote the collection of all palindromes.

Definition 1. Let $u \in \mathbb{A}^n$ and $S \subseteq \mathcal{S}(n)$. We say that S *palindromically generates* u if the following three conditions are verified:

- $u[i, j] \in \mathcal{P}$ for all $(i, j) \in S$,
- for all $k \in \{1, 2, \dots, n\}$, there exists $(i, j) \in S$ with $i \leq k \leq j$,
- for each non-empty set \mathbb{B} and word $v \in \mathbb{B}^n$, if $v[i, j] \in \mathcal{P}$ for all $(i, j) \in S$ then there exists a mapping $c: \text{Alph}(u) \rightarrow \mathbb{B}$ which extends to a morphism $c: \text{Alph}(u)^* \rightarrow \mathbb{B}^*$ of words such that $c(u) = v$.

We call the elements of S *generators* (or *palindromic generators*). It follows from the definition that if a set $S \subseteq \mathcal{S}(n)$ palindromically generates two words $u \in \mathbb{A}^n$ and $v \in \mathbb{B}^n$ of length n , then u and v are word isomorphic, i.e., there is a bijection $\nu: \text{Alph}(u) \rightarrow \text{Alph}(v)$ which extends to a morphism of words such that $\nu(u) = v$. In particular, the last condition in Definition 1 states that $\text{Alph}(u)$ has the largest cardinality such that the first two conditions are satisfied.

Example 2. For each letter $a \in \mathbb{A}$, the singleton set $S = \{(1, 1)\}$ palindromically generates a . The set $S = \{(1, 2)\}$ palindromically generates the word $a^2 = aa$, while $S = \{(1, 1), (2, 2)\}$ palindromically generates the word ab . For $n \geq 3$, the set $S = \{(1, n-1), (1, n)\}$ as well as $S = \{(1, n), (2, n)\}$ palindromically generate a^n .

Given a word $u \in \mathbb{A}^n$, let $\mu(u)$ denote the infimum of the cardinality of all sets $S \subseteq \mathcal{S}(n)$ that palindromically generate u , i.e.,

$$\mu(u) = \inf\{\text{Card } S \mid S \subseteq \mathcal{S}(|u|) \text{ palindromically generates } u\}.$$

As usual, we let $\inf(\emptyset) = +\infty$. It is easily checked that, for instance, $\mu(a^n) = 1$ for $n = 1, 2$ and $\mu(a^n) = 2$ for $n \geq 3$. Also $\mu(u) < +\infty$ whenever $u \in \mathbb{A}^*$ is a nonempty word and $\text{Card } \mathbb{A} = 2$, i.e., whenever u is a binary word. Indeed, it is readily verified that for each nonempty word $u \in \{0, 1\}^*$, the set

$$S_u = \{(i, j) \mid u[i, j] = ab^k a \text{ for } \{a, b\} = \{0, 1\}, k \geq 0\}$$

palindromically generates u . Typically $\mu(u) < \text{Card } S_u$, e.g., if $u = 00101100$, then the set $S = \{(1, 2), (2, 4), (3, 5), (4, 7), (7, 8)\}$ palindromically generates u . In this example, also $(5, 6)$ is in S_u , but it is not needed in the generating set. The set S is the smallest palindromic generating set whence $\mu(00101100) = 5$. In contrast, for the ternary word $abca$ no subset S of $\mathcal{S}(4)$ palindromically generates $abca$. For this reason, we shall primarily restrict ourselves to binary words.

Given an infinite word $x \in \mathbb{A}^{\mathbb{N}}$, we are interested in the quantity

$$\psi(x) = \sup\{\mu(u) \mid u \text{ is a factor of } x\}.$$

We begin by showing that if $x \in \mathbb{A}^{\mathbb{N}}$ is aperiodic (i.e., not ultimately periodic), then $\psi(x) \geq 3$. This is stated in Proposition 10. We will also prove that $\psi(\mathbb{T}) = +\infty$ for the Thue-Morse infinite word \mathbb{T} defined as follows. Let

$$\mathbb{T} = 01101001100101101001 \dots$$

denote the Thue-Morse infinite word defined as the fixed point beginning in 0 of the morphism $\tau: \{0, 1\}^* \rightarrow \{0, 1\}^*$ given by $\tau(0) = 01$ and $\tau(1) = 10$. Indeed, in Proposition 11 we will show that for each integer $n > 0$ there exists a factor u of \mathbb{T} with $\mu(u) \geq n$.

The main result of this paper is to establish the existence of aperiodic binary words x for which $\psi(x) = 3$ and in fact to obtain a complete classification of such words. Recall that an infinite binary word $x \in \{0, 1\}^{\mathbb{N}}$ is called *Sturmian* if x contains exactly $n + 1$ factors of each given length $n \geq 1$. As is well known, each Sturmian word x is aperiodic and uniformly recurrent, i.e, each factor of x occurs in x with bounded gap. In order to state our main result we will make use of the following morphisms: For each subset $A \subseteq \{0, 1\}$, we denote by $d_A: \{0, 1\}^* \rightarrow \{0, 1\}^*$ the *doubling morphism* defined by the rule

$$d_A(a) = \begin{cases} aa & \text{if } a \in A \\ a & \text{if } a \notin A. \end{cases}$$

Definition 3. A word $y \in \{0, 1\}^{\mathbb{N}}$ is called *double Sturmian* if y is a suffix of $d_A(x)$ for some Sturmian word x and $A \subseteq \{0, 1\}$. In particular, taking $A = \emptyset$, it follows that every Sturmian word is double Sturmian.

We note that if y is double Sturmian, then there exists a Sturmian word x , a subset $A \subseteq \{0, 1\}$ and $a \in A$ such that $d_A(x) \in \{y, ay\}$.

Theorem 4. *Let $u \in \{0, 1\}^+$. Then $\mu(u) \leq 3$ if and only if u is a factor of a double Sturmian word.*

Combining Theorem 4 and Proposition 10, i.e., $\psi(x) \geq 3$ for an infinite aperiodic word x we deduce that:

Corollary 5. *Let $x \in \{0, 1\}^{\mathbb{N}}$ be aperiodic. Then $\psi(x) = 3$ if and only if x is double Sturmian. In particular, if $\psi(x) = 3$ then x is uniformly recurrent.*

2. Preliminaries

Throughout this section \mathbb{A} will denote a finite non-empty set. We denote by \mathbb{A}^+ the set of all *words* (or finite sequences) of the form $w = a_1 a_2 \cdots a_n$ with $a_i \in \mathbb{A}$. We write $|w| = n$ for the length of w . We denote the empty word by ε and set $\mathbb{A}^* = \mathbb{A}^+ \cup \{\varepsilon\}$. Given $w, u, v \in \mathbb{A}^+$ with $w = uv$, we write $v = u^{-1}w$ and $u = wv^{-1}$. For $w = a_1 a_2 \cdots a_n \in \mathbb{A}^n$ we denote by w^R the reversal (or mirror image) $a_n \cdots a_2 a_1$. A word w is a *palindrome* if $w = w^R$. Let \mathcal{P} denote the set of all palindromes (over any alphabet).

Given two non-empty words u and v , we say u is a *border* of v if u is both a proper prefix and a proper suffix of v . If v has no borders then it is said to be *unbordered*.

1
2
3
4
5 A word $u \in A^+$ is said to be *overlap-free* if it has no factors of the form $vzuzv$ where
6 v and z are nonempty words.

7 An infinite word $\omega \in \mathbb{A}^{\mathbb{N}}$ is *aperiodic* if it is not of the form $\omega = uv^{\mathbb{N}} = uvv \cdots$ for
8 some words $u, v \in \mathbb{A}^*$ with $v \neq \varepsilon$. The following result was shown by Saari in [20]. It
9 improves an earlier result of Ehrenfeucht and Silberger in [10]. We shall give a simplified
10 proof of this result.
11
12

13 **Lemma 6.** *Each aperiodic infinite word ω contains an infinite number of Lyndon words.*
14 *In particular ω has arbitrarily long unbordered factors.*
15
16

17 *Proof.* Let \preceq be a lexicographic ordering of words, and suppose to the contrary that ω
18 contains only finitely many Lyndon factors. We write $\omega = u_1 u_2 \cdots$ where for each $i \geq 2$
19 we have that u_i is the longest Lyndon word that is a prefix of the suffix $(u_1 \cdots u_{i-1})^{-1} \omega$.
20 Then, for all i , we have $u_{i+1} \preceq u_i$ since otherwise $u_i u_{i+1}$ would be a Lyndon prefix longer
21 than u_i . Thus there exists a positive integer j such that $w = u_1 \cdots u_{j-1} u_j^{\omega}$, contradicting
22 that ω is aperiodic.
23

24 The last claim now follows from the well known fact that every Lyndon word is unbor-
25 dered. Indeed, if $u = vuv$ is a Lyndon word with respect to the order \preceq , then $vvu \preceq vuv$
26 (since $vu \preceq uv$), and hence v is the empty word. \square
27
28

29 For each pair $I = (i, j) \in \mathcal{S}(n)$, we define a function
30

$$31 \quad \rho_I: \{i, i+1, \dots, j\} \rightarrow \{i, i+1, \dots, j\}$$

32 called a *reflection* by $\rho_I(k) = i + j - k$. The function ρ_I is an involution, i.e., it is a
33 permutation that satisfies $\rho_I(\rho_I(k)) = k$ for all $i \leq k \leq j$. For $u \in \mathbb{A}^n$ we have that
34 $u[i, j] \in \mathcal{P}$ if and only if $u[k] = u[\rho_I(k)]$ for each $i \leq k \leq j$. If $J = (i', j')$ with $i \leq i' \leq$
35 $j' \leq j$, we denote by $\rho_I(J)$ the reflected pair $(\rho_I(j'), \rho_I(i'))$.
36
37

38 The following lemma gives essentially a reformulation of Definition 1:
39
40

41 **Lemma 7.** *Let $u \in \mathbb{A}^n$ be such that all letters of \mathbb{A} occur in u . Let $S \subseteq \mathcal{S}(n)$. The*
42 *following conditions are equivalent:*
43

44 (i) *S palindromically generates u .*

45 (ii) *for each $k \in \{1, 2, \dots, n\}$, there exists an $(i, j) \in S$ such that $i \leq k \leq j$, and for*
46 *each $1 \leq i, j \leq n$, we have $u[i] = u[j]$ if and only if*

$$47 \quad \text{there exists a finite sequence (or path) } (I_t)_{t=1}^r \in S^r \quad (\star)$$

$$48 \quad \text{such that } j = \rho_{I_r} \rho_{I_{r-1}} \cdots \rho_{I_1}(i).$$

49
50
51
52
53
54 *Proof.* We first define relation θ as follows. Let $i\theta j$ if and only if there exists an $I \in S$
55 such that $j = \rho_I(i)$. Denote by θ^* the reflexive and transitive closure of θ . The relation θ
56 is symmetric by the definition of the mappings ρ_I and thus θ^* is an equivalence relation.
57 Here $i\theta^* j$ if and only if (\star) holds for some sequence $(I_t)_{t=1}^r$ of elements from S . Let then
58 $v \in \mathbb{B}^n$ be any word such that the factor $v[i, j]$ is a palindrome for each $(i, j) \in S$. Now,
59
60
61
62
63
64
65

$i\theta j$ implies $v[i] = v[j]$. Consequently, $i\theta^*j$ implies $u[i] = u[j]$ by transitivity. It follows that the cardinality of \mathbb{B} is at most the number of equivalence classes of θ^* .

If S palindromically generates the given word u , then, by the last condition of Definition 1, each equivalence class of θ^* corresponds to a different letter in \mathbb{A} . Therefore $u[i] = u[j]$ if and only if $i\theta^*j$. This proves the claim from (i) to (ii).

Suppose then that S satisfies (ii). First, let $I = (i, j) \in S$, and let $i \leq k \leq j$. Denote $k' = \rho_I(i)$. By (ii), we have that $u[k] = u[k']$, and therefore $u[i, j] \in \mathcal{P}$. Hence the first condition of Definition 1 holds. The second condition is part of (ii). For the third condition, by the beginning of the proof, any word $v \in \mathbb{B}^n$ for which $v[i, j]$ is a palindrome for each $(i, j) \in S$, the cardinality of \mathbb{B} is at most the cardinality of \mathbb{A} . By (ii), we have that $v[i] = v[j]$ implies $u[i] = u[j]$ which proves the claim. \square

Definition 8. Suppose $S \subseteq \mathcal{S}(n)$ palindromically generates a word $u \in \mathbb{A}^n$, and let $m \in \{1, 2, \dots, n\}$. We say m is a leaf with respect to S if there exists at most one pair $I = (i, j) \in S$ for which $i \leq m \leq j$ and $\rho_I(m) \neq m$.

Lemma 9. Let $u \in \mathbb{A}^+$. Then $\mu(v) \leq \mu(u)$ for all factors v of u .

Proof. The result is clear in case $\mu(u) = +\infty$. So suppose $S \subseteq \mathcal{S}(|u|)$ palindromically generates u and set $k = \text{Card } S$. It suffices to show that if $u = ax = yb$ for some words x and y and letters $a, b \in \mathbb{A}$, then $\max\{\mu(x), \mu(y)\} \leq k$. We prove only that $\mu(x) \leq k$ as the proof that $\mu(y) \leq k$ is completely symmetric.

Suppose $S = \{I_1, I_2, \dots, I_k\}$ palindromically generates u and let $m \in \mathbb{N}$ be the largest integer such that $I = (1, m) \in S$. Let $D = \{r \in \{1, 2, \dots, k\} \mid I_r = (1, q) \text{ with } q < m\}$. Let

$$S' = S \cup \{I'_r \mid r \in D\} \setminus \{I_r \mid r \in D\}$$

where for each $r \in D$ we set $I'_r = \rho_I(I_r) = (m - q + 1, m)$; see Fig. 1.

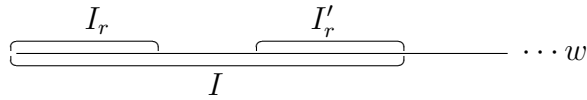


Figure 1: Reflecting the generators in S .

It follows that S' also palindromically generates u and I is the only generator in S' containing the initial position 1. Whence 1 is a leaf w.r.t. S' , and hence

$$S'' = S' \cup \{(2, m - 1)\} \setminus \{I\}$$

palindromically generates the suffix $x = a^{-1}u$. This proves the claim. \square

Proposition 10. If $x \in \mathbb{A}^{\mathbb{N}}$ is aperiodic, then $\psi(x) \geq 3$.

Proof. Let $r = \text{Card } \mathbb{A}$ and let $x \in \mathbb{A}^{\mathbb{N}}$ be aperiodic. Suppose on the contrary that $\psi(x) \leq 2$. By Lemma 6, x contains an unbordered factor w of length $|w| \geq 2r + 1$. If w is palindromically generated by a singleton set $\{I\} \subseteq \mathcal{S}(|w|)$, then w contains at

1
2
3
4
5 least $r + 1$ distinct symbols, a contradiction. If w is palindromically generated by a set
6 $\{(1, p), (q, |w|)\} \subseteq \mathcal{S}(|w|)$ of size 2, then as w is unbordered, it follows that the palindromic
7 prefix $w[1, p]$ does not overlap the palindromic suffix $w[q, |w|]$ (i.e., $p < q$). It follows again
8 that w must have at least $r + 1$ distinct symbols, a contradiction. Hence by Lemma 9,
9 $\psi(x) \geq 3$. \square

10
11 Consider the *Thue-Morse* word

$$12 \quad \mathbb{T} = 0110100110010110100101100110100110010110 \dots$$

13
14 whose origins go back to the beginning of the last century with the works of the Norwegian
15 mathematician Axel Thue [21]. The n th entry t_n of \mathbb{T} is defined as the sum modulo 2 of
16 the digits in the binary expansion of n . Alternatively, \mathbb{T} is the fixed point beginning in 0
17 of the morphism $\tau: \{0, 1\}^* \rightarrow \{0, 1\}^*$ given by $\tau(0) = 01$ and $\tau(1) = 10$.

18
19 **Proposition 11.** *For each integer $n > 0$ there exists a factor u of \mathbb{T} with $\mu(u) \geq n$, i.e.,*
20 $\psi(\mathbb{T}) = +\infty$.

21
22 *Proof.* Set $t_k = \tau^k(0)$ for $k \geq 0$. We will show that $\mu(t_{2k}) > \mu(t_{2k-2})$ for each $k > 1$ from
23 which it follows immediately that $\psi(\mathbb{T}) = +\infty$. For each $k \geq 0$ there exists $S_{2k} \subseteq \mathcal{S}(2^{2k})$
24 which palindromically generates t_{2k} and $\mu(t_{2k}) = \text{Card } S_{2k}$. We first observe that the prefix
25 t_{2k} of \mathbb{T} of length 2^{2k} is a palindrome, since $t_{2k} = t_{2k-1}^R t_{2k-1}^R$. Also, since \mathbb{T} is overlap-free
26 (see e.g. [18]), it follows that if v is a palindromic factor of t_{2k} , either v lies completely in
27 the prefix t_{2k-1} or completely in the suffix t_{2k-1}^R , or its midpoint is the midpoint of t_{2k} .
28 There necessarily exists a generator in S_{2k} that shares the middle point with t_{2k} in order
29 to relate an occurrence of 0 in the prefix t_{2k-1} with an occurrence of 0 in the suffix t_{2k-1}^R
30 of t_{2k} . Such a generator can always be replaced by the pair $F = (1, 2^{2k})$ corresponding to
31 the full palindrome t_{2k} , and thus without loss of generality we can assume that $F \in S_{2k}$.

32
33 If $I = (i, j) \in S_{2k}$ lies in the suffix t_{2k-1}^R of t_{2k} , i.e., if $i > 2^{2k-1}$, then we replace I
34 by its reflection $I' = \rho_F(I)$ which lies entirely in the first half of t_{2k} . Since $\rho_I = \rho_F \rho_{I'} \rho_F$
35 on the domain of ρ_I , the set $(S_{2k} \setminus \{I\}) \cup \{I'\}$ generates t_{2k} . In this fashion we obtain
36 a generator set S'_{2k} consisting of F and a set of pairs (i, j) where $j \leq 2^{2k-1}$. Thus $S'_{2k} \setminus \{F\}$
37 palindromically generates the prefix t_{2k-1} of t_{2k} . Since t_{2k-2} is a factor of t_{2k-1} , it follows
38 from Lemma 9 that $\mu(t_{2k}) > \mu(t_{2k-1}) \geq \mu(t_{2k-2})$ as required. \square

39
40 The following lemmas will be used in the subsequent section:

41
42 **Lemma 12.** *Suppose $u \in \mathbb{A}^n$ is palindromically generated by a set $S \subseteq \mathcal{S}(n)$. Suppose*
43 *further that there exist $p, q \in \mathbb{N}$ such that $(p, p + 2q) \in S$. Then for $A = \{u[p + q]\}$ we have*
44 $\mu(d_A(u)) \leq \text{Card } S$.

45
46 *Proof.* Let $a = u[p + q]$, and write d_a for $d_{\{a\}}$. We define first a mapping $p_{d_a}: \{1, 2, \dots, n\} \rightarrow$
47 $\mathbb{N} \cup \{(j, j + 1) \mid j \in \mathbb{N}\}$ for the positions of $u \in \mathbb{A}^n$. It maps every position of u to a
48 corresponding position of $d_a(u)$ or to a pair of positions of $d_a(u)$ depending on whether
49 the position of u has a label a that will be doubled or some other label. For convenience
50 let us denote the prefix $u[1, i]$ of u by u_i and let u_0 be the empty word.

$$51 \quad p_{d_a}(i) = \begin{cases} |d_a(u_i)| & \text{if } u[i] \neq a \\ (|d_a(u_i)| - 1, |d_a(u_i)|) & \text{if } u[i] = a. \end{cases}$$

Let $S' = \{(i', j') \mid (i, j) \in S\}$ where $i' = |d_a(u_{i-1})| + 1$ and $j' = |d_a(u_j)|$. In other words, we dilate each generator (i, j) by applying d_a to the corresponding factor $u[i, j]$. Clearly $\text{Card } S' = \text{Card } S$. We shall show that S' palindromically generates $d_a(u)$. We claim first that if a position i_1 of u is reflected to i_2 by a generator $(i_S, j_S) \in S$ then $p_{d_a}(i_1)$ is reflected to $p_{d_a}(i_2)$ by a generator $(i_{S'}, j_{S'}) \in S'$. Here $(i_{S'}, j_{S'})$ is the generator in S' corresponding to the generator (i_S, j_S) in S . Now $U = u[i_S, i_1 - 1] = u[i_2 + 1, j_S]^R$, and thus $|d_a(U)| = |d_a(U^R)| = |d_a(u[i_2 + 1, j_S])|$; see Figure 2. So if $u[i_1] \neq a$ then $p_{d_a}(i_1) - i_{S'} = j_{S'} - p_{d_a}(i_2)$ and otherwise $p_{d_a}(i_1) - (i_{S'}, i_{S'} + 1) = (j_{S'} - 1, j_{S'}) - p_{d_a}(i_2)$. Thus $p_{d_a}(i_1)$ is reflected to $p_{d_a}(i_2)$ as a single or as a pair of positions.

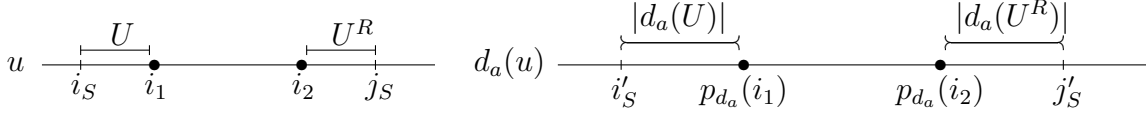


Figure 2: Reflection complies with doubling morphism d_a .

Let us denote, for any $x \in \mathbb{A}$,

$$\Omega_{u,x} = \{i \mid 1 \leq i \leq n \text{ and } u[i] = x\},$$

i.e., $\Omega_{u,x}$ is the set of occurrences of x in u . By Lemma 7, for each $i, j \in \Omega_{u,x}$ there exists a sequence $i = i_1, i_2, \dots, i_l = j$ of positions $i_m \in \Omega_{u,x}$ such that i_m is reflected to i_{m+1} by some generator in S . As we just showed there also exists a sequence $i' = p_{d_a}(i_1), p_{d_a}(i_2), \dots, p_{d_a}(i_l) = j'$ such that $p_{d_a}(i_m)$ is reflected to $p_{d_a}(i_{m+1})$ by some generator in S' . Thus if $i \in \Omega_{u,x}$ then $p_{d_a}(i) \in \hat{\Omega}_{d_a(u),x}$ where

$$\hat{\Omega}_{d_a(u),x} = \{p_{d_a}(i) \mid i \in \Omega_{u,x}\}.$$

In fact, if $x \neq a$ then $\hat{\Omega}_{d_a(u),x}$ is the same set as $\Omega_{d_a(u),x}$. The only problematic set is $\hat{\Omega}_{d_a(u),a} = \{(i', i' + 1) \mid 1 \leq i' < |d_a(u)| \text{ and for which } \exists i \in \Omega_{u,a} \text{ s.t. } p_{d_a}(i) = (i', i' + 1)\}$.

Now, consider the generator $(p', q') \in S'$ that is obtained from $(p, p + 2q) \in S$, i.e. $(p', q') = (|d_a(u_{p-1})| + 1, |d_a(u_{p+2q})|)$. The length of the palindrome determined by the generator (p', q') is even because $|d_a(u[p, p + q - 1])| = |d_a(u[p + q + 1, p + 2q])|$ and $|d_a(u[p + q])| = 2$. Hence a pair of two consecutive positions of $d_a(u)$, namely the positions of $p_{d_a}(p + q) \in \hat{\Omega}_{d_a(u),a}$, is now in the middle of this palindrome (p', q') thus these two positions reflect to each other and they have to have the same letter in word $d_a(u)$. Because of the pairwise reflections among the set $\hat{\Omega}_{d_a(u),a}$, the positions of u according to the pairs in this set contain the same letter a .

So we have that $\hat{\Omega}_{d_a(u),x} = \Omega_{d_a(u),x}$ for $x \neq a$. In the case $x = a$ the positions covered by the pairs in $\hat{\Omega}_{d_a(u),a}$ are exactly the positions in $\Omega_{d_a(u),a}$. This shows that there exists a path as described in Lemma 7 between each positions in $\Omega_{d_a(u),x}$ for any $x \in \mathbb{A}$ thus S' palindromically generates the doubled word $d_a(u)$ and this ends the proof. \square

Corollary 13. *Let $a \in \mathbb{A}$ and $u \in \mathbb{A}^n$. Then for $A = \{a\}$ we have $\mu(d_A(u)) \leq \mu(u) + 1$.*

1
2
3
4
5 *Proof.* The result is clear if the symbol a does not occur in u since in this case $d_A(u) = u$.
6 Thus we can assume that a occurs in u . Suppose $S \subseteq \mathcal{S}(n)$ palindromically generates u
7 and $\mu(u) = \text{Card } S$. If S contains a generator that determines a palindrome of odd length
8 whose center is equal to a , then by Lemma 12 we deduce that $\mu(d_A(u)) \leq \mu(u) < \mu(u) + 1$.
9 If no such generator exists, then we can add a "trivial" generator (i, i) where i is such
10 that $u[i] = a$. Now $S \cup \{(i, i)\}$ has $\mu(u) + 1$ elements and the result follows again from
11 Lemma 12. \square
12
13

14 **Example 14.** Consider a palindrome $u = aba$ which can be generated by one palin-
15 dromic generator $(1,3)$. The doubled word $d_{\{a\}}(u) = aabaa$ can be generated by the set
16 $\{(1, 5), (1, 2)\}$ and $\mu(d_{\{a\}}(u)) = \mu(u) + 1$. If we instead double the letter b then we do not
17 need any additional generators, i.e., the doubled word $d_{\{b\}}(u) = abba$ can be generated by
18 $(1, 4)$ and $\mu(d_{\{b\}}(u)) = \mu(u)$.
19
20

21 **Lemma 15.** *Let $w \in \{0, 1\}^*$ be an unbordered word of length n which is palindromically*
22 *generated by a set $S \subseteq \mathcal{S}(n)$ with $\text{Card } S = 3$. Then w has at most four leaves. Moreover,*
23 *there are at most two leaves with label 0, and at most two leaves with label 1.*
24
25

26 *Proof.* We note first that since w is unbordered, the longest palindromic prefix U and the
27 longest palindromic suffix V of w do not overlap. Indeed, if $U = ux$ and $V = xv$ then w
28 has a border x^R ; a contradiction.
29

30 Let $S = \{(1, i), (j, n), (k, m)\}$. Since the maximal palindromes U and V do not overlap,
31 we have $i < j$, and hence every path starting from a leaf has a unique continuation to a
32 new position until the path enters another leaf. Therefore, there is at most one path for
33 the letter 0 and at most one for the letter 1. These two paths necessarily consume every
34 position of w , and the endpoints of these paths are the leaves of S . \square
35
36
37

38 **3. A characterization of binary words with $\psi(x) = 3$**

39
40 We begin by recalling a few basic facts concerning Sturmian words (see [2, 18]). An
41 infinite binary word $\omega \in \{0, 1\}^{\mathbb{N}}$ is called *Sturmian* if ω contains exactly $n + 1$ factors of
42 each given length $n \geq 1$. Each Sturmian word ω is aperiodic, uniformly recurrent, closed
43 under reversal (i.e., if u is a factor of ω then so is u^R) and balanced (for any two factors
44 u and v of the same length, $||u|_a - |v|_a| \leq 1$ for each $a \in \{0, 1\}$, where $|w|_a$ denotes the
45 number of occurrences of a in w). In fact an infinite binary word $\omega \in \{0, 1\}^{\mathbb{N}}$ is Sturmian
46 if and only if it is aperiodic and balanced.
47
48

49 A factor u of a Sturmian word ω is called *right special* (resp. *left special*) if both $u0$ and
50 $u1$ (resp. $0u$ and $1u$) are factors of ω . Thus ω contains exactly one right special (resp. left
51 special) factor of each given length, and u is right special if and only if u^R is left special.
52 A factor u of ω is called bispecial if u is both right and left special. Thus, if u is a bispecial
53 factor of ω , then u is a palindrome. A binary word $x \in \{0, 1\}^*$ is called a *central* word if
54 and only if $x \in \mathcal{P}$ and $x0$ and $x1$ are both balanced.
55
56

57 We recall the following fact which is a consequence of a result in [8] (see also [5]):

58 **Lemma 16** (Proposition 22 in [2]). *A word $x \in \{0, 1\}^*$ is a central word if it is a power of*
59 *a single letter or if it satisfies the equations $x = u01v = v10u$ with $u, v \in \{0, 1\}^*$. Moreover*
60
61
62
63
64
65

1
2
3
4
5 in the latter case u and v are central words and setting $p = |u| + 2$ and $q = |v| + 2$, we
6 have that p and q are relatively prime periods of x and $\min\{p, q\}$ is the least period of x .
7

8 We will also need the following result.
9

10 **Lemma 17.** *Let u be an unbordered factor of a Sturmian word $\omega \in \{0, 1\}^{\mathbb{N}}$. Then either*
11 *$u \in \{0, 1\}$ or $u = axb$, where $\{a, b\} = \{0, 1\}$, and x is a central word.*
12

13 *Proof.* If $u \notin \{0, 1\}$, then we can write $u = axb$ with $\{a, b\} = \{0, 1\}$. We claim that x is a
14 palindrome. In fact, suppose vc (resp. dv^R) is a prefix (resp. suffix) of x with $v \in \{0, 1\}^*$
15 and $\{c, d\} = \{0, 1\}$. Since u is balanced it follows that $c = b$ and $d = a$. Since av^Rb and avb
16 are both factors of ω , it follows that v is a bispecial factor of ω , and hence a palindrome.
17 Thus avb is both a prefix and a suffix of u . Since u is unbordered, it follows that $u = avb$,
18 and hence $x = v$. Thus x is central. \square
19
20
21

22 We also recall the following extremal property of the Fine and Wilf theorem [12]
23 concerning central words due to de Luca and Mignosi [9]
24

25 **Lemma 18** (Proposition 23 in [2]). *A word x is a central word if and only if there exist*
26 *relatively prime positive integers p and q with $|x| = p + q - 2$ such that x has periods p*
27 *and q .*
28
29

30 We will also use the following property by Lothaire [19] several times to find a period
31 of a palindrome which has a palindromic prefix (and suffix).
32

33 **Lemma 19** ([19]). *If $uv = vu'$, then $|u|$ is a period of uv .*
34

35 A word $y \in \{0, 1\}^{\mathbb{N}}$ is called *double Sturmian* if y is a suffix of $d_A(x)$ for some Sturmian
36 word x and $A \subseteq \{0, 1\}$ (where d_A is the doubling morphism defined in section 1). By
37 taking $A = \emptyset$, each Sturmian word is double Sturmian.
38
39

40 **Proposition 20.** *Let y be a factor of a double Sturmian word $\omega \in \{0, 1\}^{\mathbb{N}}$. Then $\mu(y) \leq 3$.*
41

42 *Proof.* Let y be a factor of a double Sturmian word ω . Thus there exists a Sturmian
43 word $\omega' \in \{0, 1\}^{\mathbb{N}}$ and a subset $A \subseteq \{0, 1\}$ such that ω is a suffix of $d_A(\omega')$. Let y'
44 be the shortest unbordered factor of ω' such that y is a factor of $d_A(y')$. Because ω' is
45 aperiodic and uniformly recurrent, such a factor y' always exists. Now by Lemma 9 it
46 is enough to show that $d_A(y')$ is palindromically generated by a set with at most three
47 generators. By Lemma 12 it suffices to show that y' is palindromically generated by a set
48 S' with $\text{Card } S' \leq 3$ and containing two generators which determine two palindromes of
49 odd length in y' with distinct central symbols .
50
51

52 Since y' is unbordered, by Lemma 17, we may write $y' = axb$ where $x \in \{0, 1\}^*$ is a
53 central word and $\{a, b\} = \{0, 1\}$. Without loss of generality we can assume that $a = 0$ and
54 $b = 1$. We first consider the case where x is a power of a single letter. If x is empty, then
55 $y' = 01$ is palindromically generated by $S' = \{(1, 1), (2, 2)\}$. If $x = 0^n$ with $n \geq 1$, then $y' =$
56 $0^{n+1}1$ is palindromically generated by $S' = \{(1, n), (1, n+1), (n+2, n+2)\}$. If $x = 1^n$ with
57 $n \geq 1$, then $y' = 01^{n+1}$ is palindromically generated by $S' = \{(1, 1), (2, n+1), (2, n+2)\}$.
58
59
60
61
62
63
64
65

1
2
3
4
5 In all of these cases there exist two generators which determine odd length palindromes
6 with distinct central symbols.

7 Next we assume x is not a power of a single letter. By Lemma 16 there exist central
8 words u and v such that $y' = 0u01v1$. Put $U = 0u0$ and $V = 1v1$. We claim that y' is
9 palindromically generated by the set

$$10 \quad S' = \{(1, |U|), (|y'| - |V| + 1, |y'|), (2, |x| + 1)\}.$$

11 We first note that $y'[1, |U|] = U$, $y'[|y'| - |V| + 1, |y'|] = V$ and $y'[2, |x| + 1] = x$, whence
12 $y'[1, |U|]$, $y'[|y'| - |V| + 1, |y'|]$, $y'[2, |x| + 1] \in \mathcal{P}$. Next let $w \in \mathbb{B}^*$ be a word with $|w| = |y'|$.
13 Set $U' = w[1, |U|]$, $V' = w[|y'| - |V| + 1, |y'|]$, and $x' = w[(2, |x| + 1)]$ so that $w = a'x'b'$
14 with $a', b' \in \mathbb{B}$. Suppose $U', V', x' \in \mathcal{P}$. It follows, e.g., from Lemma 19 that x' has now
15 periods $|U'| = |U|$ and $|V'| = |V|$ and by Lemma 16 they are relatively prime because
16 $|U|$ and $|V|$ are. Since $|x'| = |U'| + |V'| - 2$, we deduce by Lemma 18 that x' is a central
17 word. Thus x' is either a power of a single letter or it is isomorphic to x . In the first case
18 w is also a power of a single letter. In the second case, w is word isomorphic to y' . Thus
19 in either case there exists a mapping $\nu: \mathbb{A} \rightarrow \mathbb{B}$ with $\nu(y') = w$. So y' is palindromically
20 generated by $S' = \{(1, |U|), (|y'| - |V| + 1, |y'|), (2, |x| + 1)\}$ where two of the associated
21 palindromes have odd length. Indeed, by Lemma 16, $|U|$ and $|V|$ are relatively prime and
22 $|x| = |U| + |V| - 2$. So either $|U|$ and $|V|$ are odd or only the other one is odd but then $|x|$
23 is odd, too. Because $y' \in \{0, 1\}^*$ is unbordered and palindromically generated by a set of
24 size 3, Lemma 15 gives that y' has at most two leaves with label 0 and at most two leaves
25 with label 1. Thus the central symbols of these odd palindromes are necessarily distinct
26 since the first and the last letter of y' are also leaves with different symbols. \square
27
28
29
30
31
32
33
34

35 We have an auxiliary result which shows that the generating set obtained in Propo-
36 sition 20 for $y = axb$ includes both the longest palindromic prefix and the longest palin-
37 dromic suffix of y . This follows from the following refinement of Lemma 16.

38 **Proposition 21.** *Let $x \in \{0, 1\}^*$ be a central word which is not a power of a single letter.*
39 *Let u and v be as in Lemma 16. Then $0u0$ (respectively $1v1$) is the longest palindromic*
40 *prefix (respectively suffix) of $0x1$. Moreover, two of the three palindromes $\{x, 0u0, 1v1\}$ are*
41 *of odd length and have distinct central symbols.*

42 *Proof.* Since u is a central word and hence a palindrome, we have that $0u0$ is a palindromic
43 prefix of $0x1$. It remains to show that it is the longest such prefix. Suppose to the contrary
44 that $0x1$ admits a palindromic prefix $0u'0$ with $|u'| > |u|$. Then by lemmas 16 and 19,
45 we have that $p = |u| + 2, q = |v| + 2$ and $p' = |x| - |u'|$ are each periods of x . Since
46 $p' = |x| - |u'| < |x| - |u| = q$, it follows from Lemma 16 that the $\min\{p, q\} = p$. Also, as
47 $|x| = |u| + |v| + 2 \leq |u'| + |v| + 1$, we deduce that

$$48 \quad |x| \geq |x| - |u'| + |x| - |v| - 1 = p' + p - 1 \geq p' + p - \gcd(p, p').$$

49 But since both p and p' are periods of x , it follows from the Fine and Wilf Theorem [12]
50 that x has period $\gcd(p, p')$ which by Lemma 16 is equal to p . Whence p divides p' . Let z
51 denote the suffix of x of length p' . Since $0u'0$ is a palindromic prefix of $0x1$ it follows that
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5 z begins in 0. On the other hand, since $10u$ is a suffix of x of length p and p divides p' , it
6 follows that z begins in 1. This contradiction proves that $0u0$ is the longest palindromic
7 prefix of $0x1$. Similarly one deduces that $1v1$ is the longest palindromic suffix of $0x1$.

8 By Lemma 16, p and q are relatively prime so two of the three palindromes $\{x, 0u0, 1v1\}$
9 have odd length and by Lemma 15 the central symbols of these have to be different. \square

10 We next prove the converse of Theorem 4, namely:
11

12 **Proposition 22.** *Suppose $w \in \{0, 1\}^+$ is such that $\mu(w) \leq 3$. Then w is a factor of a*
13 *double Sturmian word ω .*

14 In the following lemmas we use a and b as variables of letters such that $\{a, b\} = \{0, 1\}$.
15 Lemma 15 on the number of leaves entails some immediate restrictions on w .
16

17 **Lemma 23.** *Suppose $w \in \{a, b\}^*$ be such that $\mu(w) \leq 3$. Then*

18 (i) *The words a^3 and b^3 do not both occur in w .*

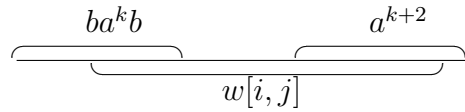
19 (ii) *For odd k , $ba^k b$ and a^{k+2} do not both occur in w .*

20 (iii) *The words $ba^k b$ and a^{k+3} do not both occur in w for any $k \geq 1$.*

21 (iv) *All three words a^{k+2} , $ba^{k+1} b$ and $ba^k b$ do not occur in w for any $k \geq 1$.*

22 *Proof.* In each of the cases we assume that w with $|w| = n$ is a minimal counter example,
23 i.e., none of its proper factors is a counter example. By the minimality assumption, w
24 begins and ends in the expressed forbidden words. Since also the reverse w^R is a counter
25 example, we may assume that the first letter of w is b .
26

27 For Case (i), let $w = b^3 u a^3$ for some u . By the minimality assumption, w is unbordered,
28 and w is palindromically generated by the set $S = \{(1, 3), (n - 2, n), (i, j)\}$ for some (i, j)
29 that determines the palindrome $w[i, j]$. Here either $i > 3$ or $j < n - 2$, since the palindrome
30 $w[i, j]$ starts and ends with the same letter. However, this means that there are three leaves
31 for a or three leaves for b . This contradicts Lemma 15.
32



35 Figure 3: The palindromic factors of w in Case 2.
36

37 For Case (ii), let $w = ba^k b u a^{k+2}$ for some u . Since w is minimal, it is unbordered, and
38 it is palindromically generated by $S = \{(1, k + 2), (n - k - 1, n), (i, j)\}$ for some (i, j) .
39 These generators determine the factors $w[1, k + 2] = ba^k b$, $w[n - k - 1, n] = a^{k+2}$ and
40 $w[i, j]$, respectively, where the palindrome $w[i, j]$ necessarily misses the first b of $w[1, k + 2]$
41 and the last two a 's of $w[n - k - 1, n]$; see Figure 3. Also, $w[1, k + 2]$ and $w[n - k - 1, n]$
42 are both of odd length, since k is assumed to be odd, and thus their middle positions are
43 leaves. This contradicts Lemma 15.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5 For Case (iii), let $w = ba^k b u a^{k+3}$. Since w is minimal, w is unbordered, and it is
6 palindromically generated by $S = \{(1, k+2), (n-k-2, n), (i, j)\}$ for some (i, j) . The
7 palindrome $w[i, j]$ misses at least the last three a 's of w . Again Lemma 15 yields a
8 contradiction.
9

10 For Case 4, consider first any factor of the form $v = ba^k b u a^{k+1}$ of w , where $ba^k b$ occurs
11 only as a prefix and a^{k+1} only as a suffix of v . Again v is unbordered, and S contains a
12 generator (i, j) , where $i > 1$ and $j < |v|$, i.e., the palindrome $v[i, j]$ misses at least the
13 suffix a and the prefix b of v . By Case (ii), $u \neq ab$ (for otherwise $k+1 \geq 3$), and hence,
14 by Lemma 15, $u \in \{\varepsilon, b\}$. Similarly, if $v = a^{k+1} u b a^k b$ then $u \in \{\varepsilon, b\}$. This proves the case
15 since if a^{k+2} and $ba^k b$ occur in w , by the above $ba^{k+1} b$ does not occur in w . \square
16
17

18 The key result for the converse is stated in Lemma 25 which gives a simplified criterion
19 for pairs witnessing that a word is unbalanced when compared to the general case of
20 Lemma 24. We have
21
22

23 **Lemma 24** (Proposition 2.1.3 in [19]). *If $w \in \{a, b\}^*$ is unbalanced then there is a palin-*
24 *drome u such that both aua and bub are factors of w .*
25

26 **Lemma 25.** *Suppose $w \in \{a, b\}^*$ is unbalanced and $\mu(w) \leq 3$. Let aua and bub be any*
27 *palindromic factors of w . Then $u = a^k$ or $u = b^k$ for some even $k \geq 0$.*
28
29

30 *Proof.* By Lemma 23 we can assume without loss of generality that b^3 is not a factor of
31 w . Also, the case for $u = \varepsilon$ is clear, and thus we can assume that $u \neq \varepsilon$.
32

33 Consider a shortest factor z of w containing both aua and bub . Then z begins with
34 aua and ends with bub , or vice versa, and z has unique occurrences of the factors aua and
35 bub . We can suppose that z starts with a , since otherwise we take the reverse of z . Now,
36 aua is a maximal palindromic prefix of z , since a longer palindromic prefix of z would
37 contain aua as a suffix. Similarly bub is a maximal palindromic suffix of z .
38

39 By Lemma 9, $\mu(z) \leq 3$. Since u is a palindrome, the factors aua and bub do not overlap.
40 Indeed, if $z = au_1 b u_2 a u_3 b$, where $u = u_1 b u_2 = u_2 a u_3$, then by taking a reverse of u , we
41 have $u_2^R b u_1^R = u_2 a u_3$, and so $b u_1^R = a u_3$; a contradiction since $a \neq b$. Since aua and bub do
42 not overlap and z is chosen to be minimal, z is unbordered, and thus Lemma 15 applies
43 to z .
44

45 Let the palindromic generators of z be $(1, p)$, (s, n) and (i, j) . Then $z[1, p]$ is a prefix
46 of the maximal palindromic prefix aua of z , $z[s, n]$ is a suffix of the maximal palindromic
47 suffix bub of z , and $z = xz[i, j]y$, where x and y are non-empty. By Lemma 15, $|x| + |y| \leq 4$,
48 and $|x|, |y| \geq 1$ since two of the leaves (one a and one b) reside at the ends of z . We have
49 $|x| \neq |y|$, since the last letter of aua and the first letter of bub cannot be reflected to
50 each other by the generator (i, j) . Therefore, either $|x| = 1$ or $|y| = 1$. We assume that
51 $|y| = 1$ (i.e., $y = b$) and $|x| > 1$, the proof for $|x| = 1$ and $|y| > 1$ being similar. We
52 now have found three leaves, two in the prefix of z and one at the end. Hence, we have
53 $z = z[1, p]vz[s, n]$ where $|v| \leq 1$ since the positions in v are also leaves.
54
55

56 Since bub is a maximal palindromic suffix of z and $y = b$, $z[i, j]$ is preceded by the letter
57 a . It follows that both leaves for the letter a occur in the prefix x , and the remaining fourth
58 leaf resides at a position for the letter b . In particular, if $v \neq \varepsilon$ then $v = b$.
59
60
61
62
63
64
65

Suppose first that $|u|$ is odd. If $|v| = 1$, and thus $v = b$, then the position of v is the fourth leaf, and $z[1, p] = aua$ or $z[s, n] = bub$. Here $z[s, n] \neq bub$ since otherwise the midpoint of bub would be a fifth leaf. Hence $z[1, p] = aua$ and the midpoint of aua is a leaf. To avoid a fifth leaf, it must be at position $i = 2$, i.e., $|u| = 1$, and consequently $z = acabcb$ where $c = a$ because b^3 does not occur in z . However, $\mu(aaabab) = 4$. Therefore, by Lemma 15, $v = \varepsilon$, in which case, $z[1, p] = aua$ and $z[s, n] = bub$, and they both have odd length. Again, necessarily $|u| = 1$ to avoid a fifth leaf, and as above $z = aaabab$ with $\mu(z) = 4$.

Consequently, $|u|$ must be even. Let $u = u_1u_2$ be such that $z[i, j] = u_2atbu_1u_2$ for some t and hence $z = au_1u_2atbu_1u_2b$. By the above, we have $|t| \leq 1$, and if $|t| = 1$, then $z[1, p] = aua$ and $z[s, n] = bub$. Here u_2 is a palindrome since $z[i, j]$ is one. Also, in all cases both u_1 and t consist of leaves only, and hence $1 \leq |u_1| + |t| \leq 2$. Since u is a palindrome, we have $u = u_1u_2 = u_2^R u_1^R = u_2 u_1^R$, and, by Lemma 19, $|u_1|$ is a period of u .

If $t \neq \varepsilon$, i.e., $t \in \{a, b\}$, then $|u_1| = 1$, and thus $u = a^k$ for some k .

Assume then that $t = \varepsilon$. Now, $|u_1| = 1$ or $|u_1| = 2$. If $|u_1| = 1$, then again $u = a^k$ for some k . Finally, if $|u_1| = 2$, then $z = z[1, p] \cdot z[s, n]$ with $z[1, p] = aua$ and $z[s, n] = bub$ by Lemma 15, since the three positions of the prefix au_1 and the last position of z are leaves. Also, $u_1 = ba$ since $z[i, j]$ and thus also its middle factor abu_1 is a palindrome. Since $z[i, j]$ is a palindrome, and $|u_1|$ is a period of u , we have $z[1, p] = aua = a(ba)^{i+1}ba$ for some i . However, now $|u|$ is odd; a contradiction. This proves the claim. \square

For $w \in \{0, 1\}^+ \cup \{0, 1\}^{\mathbb{N}}$, we define $A(w) \subseteq \{0, 1\}$ by the rule $a \in A(w)$ if and only if w has no factors of the form $ba^{2k+1}b$ for any $k \geq 0$ with $a \neq b$. If w is a finite word, we define the *lean word* of w to be the shortest word u such that $d_{A(w)}(u)$ contains w as a factor. We extend this notion to the infinite case as follows: If $w \in \{0, 1\}^{\mathbb{N}}$, we say $u \in \{0, 1\}^{\mathbb{N}}$ is the *lean word* if $d_{A(w)}(u)$ contains w as a suffix and for all $v \in \{0, 1\}^{\mathbb{N}}$, if w is a suffix of $d_{A(w)}(v)$, then u is a suffix of v . Clearly, in each case the lean word of w is uniquely determined by w .

For instance, if $w = 0010011$, then $A(w) = \{0\}$ and the lean word of w is $u = 01011$. Similarly if $w = 011001$, then $A(w) = \{0, 1\}$ and the lean word $u = 0101$. For $w = 0010110$, we have $A(w) = \emptyset$ and hence w is its own lean word.

Lemma 26. *Let $w \in \{0, 1\}^*$ be a binary word with $\mu(w) \leq 3$. Then the lean word u of w is balanced.*

Proof. Let w be a shortest counter example to the claim such that its lean word u is unbalanced. By appealing to symmetry and Lemma 23(i), we can assume that w contains no occurrences of 111.

Since u is not balanced, it has factors $0v0$ and $1v1$ for some palindrome v according to Lemma 24. Now, also $0d_{A(w)}(v)0$ and $1d_{A(w)}(v)1$ are palindromes, and they are factors of w . By Lemma 25, we have $d_{A(w)}(v) = 0^k$ for some even integer k . It follows that $v = 0^m$ for $m = k/2$ or $m = k$ depending on whether 0 is in $A(w)$ or not. Here u has factors 0^{m+2} and 10^m1 .

Suppose first that $k > 0$. Since w has factors 0^{k+2} and 10^k1 , it has no blocks of 0's of odd length by Lemma 23(iii) and (iv). Hence $0 \in A(w)$ by the definition of $A(w)$,

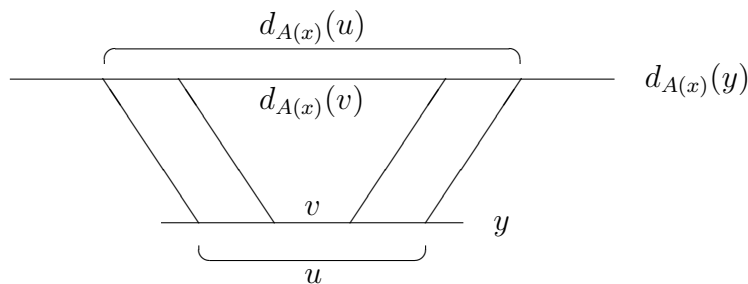


Figure 4: Proof of Proposition 5: The word v is extended to a lean factor u of $d_{A(x)}(u)$.

but then $d_{A(w)}(u)$ has factors 0^{2m+4} and $10^{2m}1$, and thus w has factors 0^{2m+3} and $10^{2m}1$ contradicting Lemma 23(iii).

Suppose then that $k = 0$. Since now 11 occurs in u but 111 does not occur in w , we have $1 \notin A(w)$. By the definition of $A(w)$, the word 010 must be a factor of u , and hence by the minimality of w , we have $u = 00(10)^r11$ for some positive $r \geq 1$. From this also $0 \notin A(w)$ follows, and hence $u = w$. However, the palindromic generators of w are now $(1, 2)$ that determines the prefix $w[1, 2] = 00$, and $(n - 1, n)$ that determines the suffix $w[n - 1, n] = 11$, and (i, j) that determines the factor $w[i, j]$. But the palindrome $w[i, j]$ cannot overlap with both $w[1, 2]$ and $w[n - 1, n]$ and thus the two 0's in $w[1, 2]$ and the latter 0 are not equivalent or the two 1's in $w[n - 1, n]$ are not equivalent to a preceding 1; a contradiction. \square

Proof of Proposition 22. By Lemma 26, if $\mu(w) \leq 3$, then w is a factor of the word $d_{A(w)}(u)$ for a balanced word u . Since the factors of Sturmian words are exactly the balanced words, see [19], the claim follows. This also proves Theorem 4. \square

Proof of Corollary 5. Let $x \in \{0, 1\}^{\mathbb{N}}$ be aperiodic. By Proposition 10, $\psi(x) \geq 3$.

If x is double Sturmian then, by Proposition 20, $\mu(w) \leq 3$ for all factors of x , and thus by definition, $\psi(x) = 3$.

For the converse, assume that $\psi(x) = 3$, and consider the set $A(x) \subseteq \{0, 1\}$ together with the lean word y of x . Then x is a suffix of $d_{A(x)}(y)$. We need to show that y is Sturmian. Let v be a factor of y , and let u be a factor of y such that u contains v and $d_{A(x)}(u)$ contains a factor of the form $ba^{2k+1}b$ with $k \geq 0$ for each $a \notin A(x)$. Hence $A(d_{A(x)}(u)) = A(x)$, and therefore u is the lean word of $d_{A(x)}(u)$; see Figure 4. By Lemma 9, $\mu(d_{A(x)}(u)) \leq 3$, and thus Lemma 26 yields that u , and hence also v , is balanced. It follows that y is Sturmian as required. \square

- [1] B. Adamczewski and Y. Bugeaud, On complexity of algebraic numbers, II. Continued fractions, Acta. Math. **195** (2005), 1–20.
- [2] J. Berstel, Sturmian and Episturmian Words (A Survey of Some Recent Results), in S. Bozpalidis and G. Rahonis, editors, CAI 2007, Lecture Notes in Computer Science **4728** Springer-Verlag 2007, 23–47.

- 1
2
3
4
5 [3] J. Berstel and A. de Luca, Sturmian words, Lyndon words and trees, *Theoret. Comput. Sci.* **178** (1997), 171–203.
6
7
8 [4] S. Brlek, S. Hamel, M. Nivat and C. Reutenauer, On the palindromic complexity of
9 infinite words, *Internat. J. Found. Comput. Sci.* **15** (2004), 293–306.
10
11 [5] A. Carpi and A. de Luca, Codes of central Sturmian words, *Theoret. Comput. Sci.*
12 **340** (2005), 220–239.
13
14 [6] S. Col, Palindromes dans les progressions arithmétiques (French), [Palindromes in
15 arithmetic progressions], *Acta Arith.* **137** (2009), 1–41.
16
17 [7] F. Deloup, Palindromes and orderings in Artin groups, *J. Knot Theory Ramifications*
18 **19** (2010), 145–162.
19
20 [8] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theo-*
21 *ret. Comput. Sci.* 183(1) (1997), 45–82.
22
23 [9] A. de Luca and F. Mignosi, Some combinatorial properties of Sturmian words, *Theo-*
24 *ret. Comput. Sci.* **136** (1994), 361–385.
25
26 [10] A. Ehrenfeucht and D. Silberger, Periodicity and unbordered segments of words,
27 *Discrete Math.* **26** (1979), 101–109.
28
29 [11] S. Ferenczi and L.Q. Zamboni, Eigenvalues and simplicity of interval exchange trans-
30 formations, *Ann. Sci. École Norm. Sup. (4)* **44** (2011), 361–392.
31
32 [12] N.J. Fine and H.S. Wilf, Uniqueness theorem for periodic functions, *Proc. Amer.*
33 *Math. Soc.* **16** (1965), 109–114.
34
35 [13] A. Glen, J. Justin, S. Widmer and L.Q. Zamboni, Palindromic richness, *European*
36 *J. Combin.* **30** (2009), 510–531.
37
38 [14] T. Harju and D. Nowotka, Minimal Duval extensions, *Internat. J. Found. Comput.*
39 *Sci.* **15** (2004), 349–354.
40
41 [15] A. Hof, O. Knill and B. Simon, Singular continuous spectrum for palindromic
42 Schrödinger operators, *Commun. Math. Phys.* **174** (1995), 149–159.
43
44 [16] L. Kari and K. Mahalingam, Watson-Crick palindromes in DNA computing, *Nat.*
45 *Comput.* **9** (2010), 297–316.
46
47 [17] C. Kassel and C. Reutenauer, A palindromization map for the free group, *Theoret.*
48 *Comput. Sci.* **409** (2008), 461–470.
49
50 [18] M. Lothaire, *Combinatorics on Words*, Cambridge University Press, 1997.
51
52 [19] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, 2002.
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

[20] K. Saari, Lyndon words and Fibonacci numbers, *J. Combin. Theory Ser. A* **121** (2013), 34–44. ArXiv:1207.4233.

[21] Thue, A.: Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I Math-Nat. Kl.* 7 (1906), 1–22.