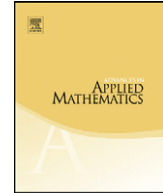




ELSEVIER

Contents lists available at ScienceDirect

## Advances in Applied Mathematics

[www.elsevier.com/locate/yaama](http://www.elsevier.com/locate/yaama)

## Balance and Abelian complexity of the Tribonacci word

Gwénaél Richomme<sup>a,1</sup>, Kalle Saari<sup>b</sup>, Luca Q. Zamboni<sup>c,d,\*</sup><sup>a</sup> *Université de Picardie Jules Verne, Laboratoire MIS (Modélisation, Information, Systèmes), 33, Rue Saint Leu, F-80039 Amiens cedex 1, France*<sup>b</sup> *Department of Mathematics, University of Turku, FI-20014, Finland*<sup>c</sup> *Université de Lyon, Université Lyon 1, CNRS UMR 5208 Institut Camille Jordan, Bâtiment du Doyen Jean Braconnier, 43, blvd du 11 novembre 1918, F-69622 Villeurbanne Cedex, France*<sup>d</sup> *Reykjavik University, School of Computer Science, Kringlan 1, 103 Reykjavik, Iceland*

## ARTICLE INFO

## Article history:

Received 8 November 2009

Accepted 18 January 2010

Available online 25 February 2010

## MSC:

68R15

## Keywords:

Abelian complexity  
Arnoux–Rauzy words  
Tribonacci  
Balanced words  
Parikh vectors

## ABSTRACT

G. Rauzy showed that the Tribonacci minimal subshift generated by the morphism  $\tau: 0 \mapsto 01, 1 \mapsto 02$  and  $2 \mapsto 0$  is measure-theoretically conjugate to an exchange of three fractal domains on a compact set in  $\mathbb{R}^2$ , each domain being translated by the same vector modulo a lattice. In this paper we study the Abelian complexity  $\rho(n)$  of the Tribonacci word  $\mathbf{t}$  which is the unique fixed point of  $\tau$ . We show that  $\rho(n) \in \{3, 4, 5, 6, 7\}$  for each  $n \geq 1$ . Our proof relies on the fact that the Tribonacci word is 2-balanced, i.e., for all factors  $U$  and  $V$  of  $\mathbf{t}$  of equal length, and for every letter  $a \in \{0, 1, 2\}$ , the number of occurrences of  $a$  in  $U$  and the number of occurrences of  $a$  in  $V$  differ by at most 2. While this result is announced in several papers, to the best of our knowledge no proof of this fact has ever been published. We offer two very different proofs: The first uses the word combinatorial properties of the generating morphism, while the second exploits the spectral properties of the incidence matrix of  $\tau$ . Although we show that  $\rho(n)$  assumes each value  $3 \leq i \leq 7$ , the sequence  $(\rho(n))_{n \geq 1}$  itself seems to be rather mysterious and may reflect some deeper properties of the Rauzy fractal.

© 2010 Elsevier Inc. All rights reserved.

\* Corresponding author at: Université de Lyon, Université Lyon 1, CNRS UMR 5208 Institut Camille Jordan, Bâtiment du Doyen Jean Braconnier, 43, blvd du 11 novembre 1918, F-69622 Villeurbanne Cedex, France.

E-mail addresses: [gwenael.richomme@univ-montp3.fr](mailto:gwenael.richomme@univ-montp3.fr) (G. Richomme), [kasaar@utu.fi](mailto:kasaar@utu.fi) (K. Saari), [zamboni@math.univ-lyon1.fr](mailto:zamboni@math.univ-lyon1.fr), [lqz@ru.is](mailto:lqz@ru.is) (L.Q. Zamboni).

<sup>1</sup> Current position: Université Paul-Valéry Montpellier 3, France.

### 1. Introduction

Given a finite non-empty set  $A$ , called the *alphabet*, we denote by  $A^*$  the free monoid generated by  $A$ . The identity element of  $A^*$ , called the *empty word*, will be denoted by  $\varepsilon$ . For any word  $u = a_1a_2 \cdots a_n \in A^*$ , the length of  $u$  is equal to  $n$  and is denoted by  $|u|$ . By convention, the length of the empty word  $\varepsilon$  is taken to be 0. For each  $a \in A$ , let  $|u|_a$  denote the number of occurrences of the letter  $a$  in  $u$ . We denote by  $A^{\mathbb{N}}$  the set of (right) infinite words on the alphabet  $A$ . Given an infinite word  $\omega = \omega_0\omega_1\omega_2 \cdots \in A^{\mathbb{N}}$ , any finite word of the form  $\omega_i\omega_{i+1} \cdots \omega_{i+n-1}$  (with  $i \geq 0$  and  $n \geq 1$ ) is called a *factor* of  $\omega$ . Let

$$\mathcal{F}_\omega(n) = \{\omega_i\omega_{i+1} \cdots \omega_{i+n-1} \mid i \geq 0\}$$

denote the set of all factors of  $\omega$  of length  $n$ , and set  $p_\omega(n) = \text{Card}(\mathcal{F}_\omega(n))$ . The function  $p_\omega : \mathbb{N} \rightarrow \mathbb{N}$  is called the *subword complexity function* of  $\omega$ . A fundamental result due to Hedlund and Morse [15] states that a word  $\omega$  is ultimately periodic if and only if for some  $n$  the subword complexity  $p_\omega(n) \leq n$ . Words of subword complexity  $p(n) = n + 1$  are called *Sturmian words*. The most well-known Sturmian word is the so-called Fibonacci word

$$\mathbf{f} = 01001010010010100101001001010010010100101001001010010 \cdots$$

fixed by the morphism  $0 \mapsto 01$  and  $1 \mapsto 0$ .

Sturmian words admit various types of characterizations of geometric and combinatorial flavours. We give two examples: In [16] Hedlund and Morse showed that each Sturmian word may be realized measure-theoretically by an irrational rotation on the circle. That is, every Sturmian word is obtained by coding the symbolic orbit of a point  $x$  on the circle (of circumference one) under a rotation by an irrational angle  $\alpha$  where the circle is partitioned into two complementary intervals, one of length  $\alpha$  and the other of length  $1 - \alpha$ . And conversely each such coding gives rise to a Sturmian word.

Sturmian words are equally characterized by a certain balance property. We begin with the following definition:

**Definition 1.1.** An infinite word  $\omega \in A^{\mathbb{N}}$  is said to be *C-balanced* ( $C$  a positive integer) if  $||U|_a - |V|_a| \leq C$  for all factors  $U$  and  $V$  of  $\omega$  of equal length, and each  $a \in A$ . We say  $\omega$  is *balanced* if it is 1-balanced.

**Theorem 1.2.** (See Theorem 2.1.5 in [13].) *An infinite word  $\omega$  is Sturmian if and only if  $\omega$  is a binary aperiodic balanced word.*

In [17,18], Rauzy showed that the regular continued fraction algorithm provides a formidable link between the arithmetic/diophantine properties of an irrational number  $\alpha$ , the ergodic/dynamical properties of a rotation by angle  $\alpha$  on the circle, and the combinatorial (balance) properties of Sturmian words.

A fundamental problem is to generalize and extend this rich interaction to higher dimensions, either by starting with a specified symbolic dynamical system, or by a class of words satisfying particular combinatorial conditions. In the case of dimension 2, there are two different dynamical systems which are natural candidates: For instance, promising results have been obtained by the third author together with Ferenczi and Holton by considering the dynamics of 3-interval exchange transformations on the unit interval [8–11], and also by Arnoux, Berthé and Ito by considering two rotations on the circle [2]. However, ever since the early work of Rauzy in [19], the most natural generalization was thought to be the one stemming from a rotation on the 2-torus. In this context, the associated symbolic counterpart is given by a class of words of subword complexity  $2n + 1$ , originally introduced by Arnoux and Rauzy in [3], and today called *Arnoux–Rauzy words*. Arnoux–Rauzy words have since been defined more generally on a  $k$ -letter alphabet for any  $k \geq 2$ . In the special case  $k = 2$ , they are precisely the Sturmian words. The *Tribonacci word*

$$\mathbf{t} = \lim_{n \rightarrow \infty} \tau^n(0) = 01020100102010 \dots,$$

defined as the fixed point of the morphism  $\tau$  given by

$$0 \mapsto 01 \quad 1 \mapsto 02 \quad 2 \mapsto 0,$$

is an Arnoux–Rauzy word on a 3-letter alphabet and constitutes the natural analogue of the Fibonacci word  $\mathbf{f}$ . In fact, in [19] Rauzy showed that the Tribonacci minimal subshift (the shift orbit closure of  $\mathbf{t}$ ) is a natural coding of a rotation on the 2-dimensional torus  $T^2$ , i.e., is measure-theoretically conjugate to an exchange of three fractal domains on a compact set in  $\mathbb{R}^2$ , each domain being translated by the same vector modulo a lattice. In [6], the third author together with Cassaigne and Ferenczi showed that there exist Arnoux–Rauzy words which are arbitrarily imbalanced, that is to say, which are not  $C$ -balanced for any positive integer  $C$ , and that such Arnoux–Rauzy words cannot be measure-theoretically conjugate to a rotation on the 2-torus.

In the first part of this paper, we prove:

**Theorem 1.3.** *The Tribonacci word  $\mathbf{t}$  is 2-balanced.*

While this result is announced in several papers (see for instance [4–6,12,23]), to the best of our knowledge no proof of this fact has ever been published. In this paper we give two different proofs of Theorem 1.3. The first is a proof by induction which uses the word combinatorial properties of the generating morphism  $\tau$ . The second proof relies on the spectral properties of the incidence matrix of  $\tau$  and is more in the spirit of the methods developed by Adamczewski in [1].

Our hope is that one of the two proofs may be extended to establish a balance property for the general  $m$ -bonacci word ( $m \geq 2$ ) defined as the fixed point of the morphism

$$0 \mapsto 01 \quad 1 \mapsto 02 \quad \dots \quad (m - 2) \mapsto 0(m - 1) \quad (m - 1) \mapsto 0.$$

Computer simulations show that the 4-bonacci word is not 2-balanced. In fact, let  $u$  be the factor of the 4-bonacci word of length 3305 occurring in position 2663 (starting with 0), and  $v$  the factor of length 3305 occurring in position 9048. Then  $|u|_1 = 891$  and  $|v|_1 = 888$ , which shows that the 4-bonacci word is not 2-balanced. It is possible that in general the  $m$ -bonacci word is  $(m - 1)$ -balanced as stated in [12].

We then apply Theorem 1.3 to study the so-called Abelian complexity of the Tribonacci word. Following [20], two words  $u$  and  $v$  in  $A^*$  are said to be *Abelian equivalent*, denoted  $u \sim_{ab} v$ , if and only if  $|u|_a = |v|_a$  for all  $a \in A$ . For instance, if we take  $A$  to be the usual alphabet  $A = \{a, b, c, \dots, z\}$ , then *elevenplustwo* is Abelian equivalent to *twelveplusone*, but neither is Abelian equivalent to *thirteen*. It is readily verified that  $\sim_{ab}$  defines an equivalence relation on  $A^*$ .

We define

$$\mathcal{F}_\omega^{ab}(n) = \mathcal{F}_\omega(n) / \sim_{ab}$$

and set  $\rho_\omega = \text{Card}(\mathcal{F}_\omega^{ab}(n))$ . The function  $\rho = \rho_\omega : \mathbb{N} \rightarrow \mathbb{N}$  which counts the number of pairwise non-Abelian equivalent factors of  $\omega$  of length  $n$  is called the *Abelian complexity* or *ab-complexity* for short. Using Theorem 1.3 we compute the ab-complexity of the Tribonacci word  $\mathbf{t}$ :

**Theorem 1.4.** *Let  $\mathbf{t}$  denote the Tribonacci word. Then  $\rho_{\mathbf{t}}(n) \in \{3, 4, 5, 6, 7\}$  for every positive integer  $n$ . Moreover, the ab-complexity of  $\mathbf{t}$  attains each of these five values.*

We also obtain several equivalent characterizations of those values  $n$  for which  $\rho_{\mathbf{t}}(n) = 3$ . As a consequence we show that  $\rho_{\mathbf{t}}(n) = 3$  for infinitely many  $n$ . Similarly we show that  $\rho_{\mathbf{t}}(n) = 7$  for infinitely many  $n$ , although the least  $n$  for which  $\rho_{\mathbf{t}}(n) = 7$  is  $n = 3914$ . We conclude with some open questions.

## 2. Two Proofs of Theorem 1.3

### 2.1. Preliminaries to Proof 1 of Theorem 1.3

Let  $\tau : \{0, 1, 2\} \rightarrow \{0, 1, 2\}^*$  denote the morphism

$$0 \mapsto 01 \quad 1 \mapsto 02 \quad 2 \mapsto 0,$$

and let

$$\mathbf{t} = \lim_{n \rightarrow \infty} \tau^n(0) = 01020100102010 \dots$$

denote the Tribonacci word.

For each  $u \in \{0, 1, 2\}^*$ , we denote by  $\Psi(u)$  the Parikh vector associated to  $u$ , that is

$$\Psi(u) = (|u|_0, |u|_1, |u|_2).$$

For each infinite word  $\omega \in \{0, 1, 2\}^{\mathbb{N}}$  let  $\Psi_\omega(n)$  denote the set of Parikh vectors of factors of length  $n$  of  $\omega$ :

$$\Psi_\omega(n) = \{ \Psi(u) \mid u \in \mathcal{F}_\omega(n) \}.$$

Thus we have

$$\rho_\omega(n) = \text{Card}(\Psi_\omega(n)).$$

We will need the following two lemmas.

**Lemma 2.1.** *Let  $U$  be a non-empty factor of  $\mathbf{t}$ . Then there exists a factor  $u$  of  $\mathbf{t}$  such that  $\Psi(U) = (|u| + \delta, |u|_0, |u|_1)$  for some  $\delta \in \{-1, 0, 1\}$ . Moreover if  $|U| \geq 3$  then  $|u| < |U|$ .*

**Proof.** It is readily seen that every non-empty factor  $U$  of  $\mathbf{t}$  may be written as either  $U = \tau(u)$ , or  $U = 0^{-1}\tau(u)$ , or  $U = \tau(u)0$ , or  $U = 0^{-1}\tau(u)0$  for some factor  $u$  of  $\mathbf{t}$ , and where  $0^{-1}\tau(u)$  indicates the deletion of the initial 0 in  $\tau(u)$ . Moreover,  $|\tau(u)|_0 = |u|$ ,  $|\tau(u)|_1 = |u|_0$ , and  $|\tau(u)|_2 = |u|_1$ . Thus we have that

$$\Psi(U) = (|u| + \delta, |u|_0, |u|_1)$$

for some  $\delta \in \{-1, 0, 1\}$ . Moreover, from above it follows that  $|U| \geq |\tau(u)| - 1$ . Also, for all  $u$  we have  $|\tau(u)| \geq |u|$ , and if  $|u| \geq 3$ , then  $|\tau(u)| \geq |u| + 2$  since either  $u$  contains at least two occurrences of 0, or  $u$  contains both an occurrence of 0 and an occurrence of 1. Finally, to see that  $|u| < |U|$  whenever  $|U| \geq 3$ , we suppose to the contrary that  $|u| \geq |U| \geq 3$ ; then

$$|U| \geq |\tau(u)| - 1 \geq |u| + 1 > |u|,$$

which is a contradiction.  $\square$

**Note.** In what follows, we will often apply the above lemma to a pair of words  $U$  and  $V$  where  $|V|_0 = |U|_0 + 2$  and  $|U|_i = |V|_i + 3$  for some  $i \in \{1, 2\}$ . Note that under these hypotheses,  $|U| \geq 7$  since  $U$  contains either at least 3 occurrences of 1, or at least 3 occurrences of 2. Since every 1 and 2 occurring in  $\mathbf{t}$  is always preceded by 0, we deduce that  $|U|_0 \geq 2$ , and hence  $|V| \geq 4$ .

**Lemma 2.2.** Let  $U$  and  $V$  be factors of  $\mathbf{t}$  satisfying the following two conditions:

$$|V|_0 = |U|_0 + 2,$$

$$|U|_i = |V|_i + 3$$

for some  $i \in \{1, 2\}$ . Then there exist factors  $u$  and  $v$  of  $\mathbf{t}$ , with  $|u| < |U|$ ,  $|v| < |V|$ ,  $|u| \leq |v|$ , and  $|u|_{i-1} = |v|_{i-1} + 3$ .

**Proof.** By the previous lemma there exist factors  $u, v$  with  $|u| < |U|$ ,  $|v| < |V|$ , such that

$$\Psi(U) = (|u| + \delta_1, |u|_0, |u|_1) \quad \text{and} \quad \Psi(V) = (|v| + \delta_2, |v|_0, |v|_1)$$

for some  $\delta_1, \delta_2 \in \{-1, 0, 1\}$ . By hypothesis we have

$$|v| + \delta_2 = |V|_0 = |U|_0 + 2 = |u| + \delta_1 + 2,$$

whence

$$|u| = |v| + \delta_2 - \delta_1 - 2 \leq |v|.$$

Finally, the condition  $|U|_i = |V|_i + 3$  for  $i \in \{1, 2\}$  implies  $|u|_{i-1} = |v|_{i-1} + 3$ .  $\square$

In what follows, we will make use the following terminology: We say two finite words  $U$  and  $V$  on the alphabet  $\{0, 1, 2\}$  are *pairwise 2-imbalanced* if  $|U| = |V|$  and  $|U|_i - |V|_i \geq 3$  for some  $i \in \{0, 1, 2\}$ .

2.2. Proof 1 of Theorem 1.3

Suppose to the contrary that  $\mathbf{t}$  is not 2-balanced. Then there exists a shortest pair of factors  $U$  and  $V$  with  $|U| = |V|$  and  $|U|_i - |V|_i \geq 3$  for some  $i \in \{0, 1, 2\}$ . If  $|U|_i - |V|_i > 3$ , then by removing the last letter from each of  $U$  and  $V$ , we would obtain a shorter pair of words of equal length which are pairwise 2-imbalanced. Thus the minimality condition on  $|U|$  implies that  $|U|_i - |V|_i = 3$ . Also, it is easily checked that all factors of length less than or equal to 3 are 2-balanced, whence  $|U| \geq 4$ . We consider three cases:  $i = 0$ , then  $i = 2$ , and finally  $i = 1$ .

**Case 1:**  $|U|_0 - |V|_0 = 3$ . By Lemma 2.1 there exist factors  $u, v$  with  $|u| < |U|$ ,  $|v| < |V|$ , such that

$$\Psi(U) = (|u| + \delta_1, |u|_0, |u|_1) \quad \text{and} \quad \Psi(V) = (|v| + \delta_2, |v|_0, |v|_1)$$

for some  $\delta_1, \delta_2 \in \{-1, 0, 1\}$ . The condition  $|U|_0 = |V|_0 + 3$  implies that

$$|u| + \delta_1 = |v| + \delta_2 + 3$$

that is

$$|u| - |v| = (3 + \delta_2 - \delta_1) > 0.$$

The condition  $|U| = |V|$  implies that

$$|u| + \delta_1 + |u|_0 + |u|_1 = |v| + \delta_2 + |v|_0 + |v|_1$$

or equivalently

$$2|u| + \delta_1 - |u|_2 = 2|v| + \delta_2 - |v|_2.$$

Thus

$$|u|_2 - |v|_2 = 2(|u| - |v|) + \delta_1 - \delta_2 = 2(3 + \delta_2 - \delta_1) + \delta_1 - \delta_2 = 6 + \delta_2 - \delta_1.$$

Let  $u'$  be the prefix of  $u$  of length  $|u| - (3 + \delta_2 - \delta_1) = |v|$ . Then,  $|u'|_2 \geq |v|_2 + 3$  contradicting the minimality of  $|U|$ .

**Case 2:**  $|U|_2 - |V|_2 = 3$ . By Lemma 2.1 there exist factors  $u, v$  with  $|u| < |U|$ ,  $|v| < |V|$ , such that

$$\Psi(U) = (|u| + \delta_1, |u|_0, |u|_1) \quad \text{and} \quad \Psi(V) = (|v| + \delta_2, |v|_0, |v|_1)$$

for some  $\delta_1, \delta_2 \in \{-1, 0, 1\}$ . The condition  $|U|_2 = |V|_2 + 3$  implies that  $|u|_1 = |v|_1 + 3$ . If  $|u| \leq |v|$ , then  $u$  and the prefix of  $v$  of length  $|u|$  are pairwise 2-imbalanced and of length less than  $|U|$ , contradicting the minimality of  $|U|$ . Thus we can suppose that

$$|u| = |v| + m \quad \text{for some } m \geq 1.$$

If  $m \geq 2$ , then we deduce that

$$|U|_0 = |u| + \delta_1 \geq |u| - 1 \geq |u| - (m - 1) = |v| + 1 \geq |v| + \delta_2 = |V|_0.$$

As  $|U|_2 = |V|_2 + 3$ ,  $|U|_0 \geq |V|_0$ , and  $|U| = |V|$ , it follows that  $|V|_1 \geq |U|_1 + 3$ , that is  $|v|_0 \geq |u|_0 + 3$ . But then  $v$  and the prefix of  $u$  of length  $|v|$  are pairwise 2-imbalanced and of length less than  $|U|$ , contradicting the minimality of  $|U|$ .

Thus we can suppose  $m = 1$ , that is  $|u| = |v| + 1$ . Again, if  $|U|_0 \geq |V|_0$ , as above we would deduce that  $|v|_0 \geq |u|_0 + 3$  which would give rise to a contradiction. So we must have that  $|U|_0 < |V|_0$ . This gives

$$|v| + 1 + \delta_1 = |u| + \delta_1 = |U|_0 < |V|_0 = |v| + \delta_2$$

that is

$$1 + \delta_1 < \delta_2$$

which in turn implies that

$$\delta_1 = -1 \quad \text{and} \quad \delta_2 = 1.$$

Thus

$$\Psi(U) = (|u| - 1, |u|_0, |u|_1) \quad \text{and} \quad \Psi(V) = (|v| + 1, |v|_0, |v|_1).$$

Finally, the condition  $|U| = |V|$  together with  $|u| = |v| + 1$ , and  $|u|_1 = |v|_1 + 3$  imply that

$$\begin{aligned} |v|_0 &= |v|_0 + (|v| + 1 + |v|_1) - (|v| + 1 + |v|_1) \\ &= |V| - (|v| + 1 + |v|_1) \\ &= |V| - (|u| + |u|_1 - 3) \\ &= |V| - (|u| + |u|_1 - 1 - 2) \\ &= |V| - (|U| - |u|_0 - 2) \\ &= |u|_0 + 2. \end{aligned}$$

In summary we have:  $|u|_1 = |v|_1 + 3$  and  $|v|_0 = |u|_0 + 2$ . Thus we can apply Lemma 2.2 which gives a contradiction to the minimality of  $|U|$ .

**Case 3:**  $|U|_1 - |V|_1 = 3$ . By Lemma 2.1 there exist factors  $u, v$  with  $|u| < |U|$ ,  $|v| < |V|$ , such that

$$\Psi(U) = (|u| + \delta_1, |u|_0, |u|_1) \quad \text{and} \quad \Psi(V) = (|v| + \delta_2, |v|_0, |v|_1)$$

for some  $\delta_1, \delta_2 \in \{-1, 0, 1\}$ . The condition  $|U|_1 = |V|_1 + 3$  implies that  $|u|_0 = |v|_0 + 3$ .

If  $|u| \leq |v|$ , then  $u$  and the prefix of  $v$  of length  $|u|$  are pairwise 2-imbalanced and of length less than  $|U|$ , contradicting the minimality of  $|U|$ . Thus we can suppose that

$$|u| = |v| + m \quad \text{for some } m \geq 1.$$

Next we proceed like in Case 2. If  $m \geq 2$ , then we deduce that

$$|U|_0 = |u| + \delta_1 \geq |u| - 1 \geq |u| - (m - 1) = |v| + 1 \geq |v| + \delta_2 = |V|_0.$$

As  $|U|_1 = |V|_1 + 3$ ,  $|U|_0 \geq |V|_0$ , and  $|U| = |V|$ , it follows that  $|V|_2 \geq |U|_2 + 3$ , that is  $|v|_1 \geq |u|_1 + 3$ . But then  $v$  and the prefix of  $u$  of length  $|v|$  are pairwise 2-imbalanced and of length less than  $|U|$ , contradicting the minimality of  $|U|$ .

Thus we can suppose  $m = 1$ , that is  $|u| = |v| + 1$ . Again, if  $|U|_0 \geq |V|_0$ , as above we would deduce that  $|v|_1 \geq |u|_1 + 3$  which would give rise to a contradiction. So we must have that  $|U|_0 < |V|_0$ . This gives

$$|v| + 1 + \delta_1 = |u| + \delta_1 = |U|_0 < |V|_0 = |v| + \delta_2$$

that is

$$1 + \delta_1 < \delta_2$$

which in turn implies that

$$\delta_1 = -1 \quad \text{and} \quad \delta_2 = 1.$$

Thus

$$\Psi(U) = (|u| - 1, |u|_0, |u|_1) \quad \text{and} \quad \Psi(V) = (|v| + 1, |v|_0, |v|_1).$$

The condition  $|U| = |V|$  together with  $|u| = |v| + 1$ , and  $|u|_0 = |v|_0 + 3$  imply that

$$\begin{aligned} |v|_1 &= |v|_1 + (|v| + 1 + |v|_0) - (|v| + 1 + |v|_0) \\ &= |V| - (|v| + 1 + |v|_0) \\ &= |V| - (|u| + |u|_0 - 3) \\ &= |V| - (|u| + |u|_0 - 1 - 2) \\ &= |V| - (|U| - |u|_1 - 2) \\ &= |u|_1 + 2. \end{aligned}$$

In summary we have

$$|u| = |v| + 1 \quad \text{and} \quad |v|_1 = |u|_1 + 2 \quad \text{and} \quad |u|_0 = |v|_0 + 3.$$

Since  $|u|, |v| \geq 3$ , Lemma 2.1 implies that there exist factors  $u', v'$  with  $|u'| < |u|$ ,  $|v'| < |v|$ , such that

$$\Psi(u) = (|u'| + \delta'_1, |u'|_0, |u'|_1) \quad \text{and} \quad \Psi(v) = (|v'| + \delta'_2, |v'|_0, |v'|_1)$$

for some  $\delta'_1, \delta'_2 \in \{-1, 0, 1\}$ .

As  $|v|_1 = |u|_1 + 2$  we deduce that  $|v'|_0 = |u'|_0 + 2$ . Similarly, the condition  $|u|_0 = |v|_0 + 3$  implies that

$$|u'| + \delta'_1 = |v'| + \delta'_2 + 3$$

so that

$$|v'| = |u'| + \delta'_1 - \delta'_2 - 3 \leq |u'| - 1.$$

Now if  $|v'| \leq |u'| - 2$ , then if  $v''$  is any factor of the Tribonacci word of length  $|u'|$  beginning with  $v'$ , we would have  $|v''|_0 \geq |v'|_0 + 1 = |u'|_0 + 3$  as every factor of length 2 or greater contains at least one occurrence of the letter 0.

So we can assume that  $|v'| = |u'| - 1$ . Then

$$|v'| + 1 + \delta'_1 = |u'| + \delta'_1 = |u|_0 = |v|_0 + 3 = |v'| + \delta'_2 + 3$$

that is

$$\delta'_1 - \delta'_2 = 2,$$

which in turn implies that

$$\delta'_1 = 1 \quad \text{and} \quad \delta'_2 = -1.$$



So

$$\begin{aligned}
 |u'|_1 &= |u'|_1 + (|v'| + |v'|_0) - (|v'| + |v'|_0) \\
 &= |u'|_1 + (|v'| + 1 + 1 + |v'|_0 - 2) - (|v'| - 1 + |v'|_0 + 1) \\
 &= |u'|_1 + (|u'| + 1 + |u'|_0) - (|v'| - 1 + |v'|_0 + 1) \\
 &= |u'|_1 + (|u| - |u'|) - (|v| - |v'|_1 + 1) \\
 &= |u| - (|v| - |v'|_1 + 1) \\
 &= |u| - (|v| + 1) + |v'|_1 \\
 &= |v'|_1.
 \end{aligned}$$

Finally,

$$\begin{aligned}
 |u'|_2 &= |u'|_2 + (|u'|_0 + |u'|_1 - 1) - (|u'|_0 + |u'|_1 - 1) \\
 &= |u'| - 1 - (|u'|_0 + |u'|_1 - 1) \\
 &= |v'| - (|v'|_0 - 2 + |v'|_1 - 1) \\
 &= |v'|_2 + 3.
 \end{aligned}$$

So in summary we have

$$|v'|_0 = |u'|_0 + 2 \quad \text{and} \quad |u'|_2 = |v'|_2 + 3$$

to which we can apply Lemma 2.2 to obtain the desired contradiction.

### 2.3. Preliminaries to proof 2 of Theorem 1.3

For the results in this section, see [14, Chapter 10].

We write  $\mathbf{t} = u_0 u_1 u_2 \dots$  with  $u_i \in \{0, 1, 2\}$ . The *Tribonacci numbers*  $T_k$  are defined by  $T_k = |\tau^k(0)|$  for  $k \geq 0$ . Hence  $T_k = T_{k-1} + T_{k-2} + T_{k-3}$  for  $k \geq 3$ , and

$$T_0 = 1, \quad T_1 = 2, \quad T_2 = 4, \quad T_3 = 7, \quad T_4 = 13, \quad \dots$$

Let

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

denote the incidence matrix of the morphism  $\tau$ , i.e.,  $M_{ij} = |\tau(j)|_i$  for all  $i, j \in \{0, 1, 2\}$ . The eigenvalues of  $M$  are the roots of the polynomial  $x^3 - x^2 - x - 1$ , which we denote by  $\beta, \alpha$ , and  $\bar{\alpha}$  where  $\beta$  is real and  $\alpha$  and  $\bar{\alpha}$  are complex conjugates. (Given a complex number  $z = a + bi$ , we denote by  $\bar{z} = a - bi$  the complex conjugate of  $z$ .) We have

$$\beta = 1.83928\dots \quad \text{and} \quad |\alpha| = |\bar{\alpha}| = 0.73735\dots$$

More information concerning the roots  $\alpha$  and  $\beta$  may be found in the comments of sequences A058265 and A000073 in “The On-Line Encyclopedia of Integer Sequences” [22]. In particular, some formulas relating  $\alpha$  and  $\bar{\alpha}$  to  $\beta$  are recalled.

It is well known that the frequencies of letters in  $\mathbf{t}$  exist (see [24]) and

$$\text{Freq}_{\mathbf{t}}(0) = \frac{1}{\beta}, \quad \text{Freq}_{\mathbf{t}}(1) = \frac{1}{\beta^2}, \quad \text{Freq}_{\mathbf{t}}(2) = \frac{1}{\beta^3}.$$

Set

$$v_{\beta} = \begin{pmatrix} \beta^{-1} \\ \beta^{-2} \\ \beta^{-3} \end{pmatrix}, \quad v_{\alpha} = \begin{pmatrix} \alpha^{-1} \\ \alpha^{-2} \\ \alpha^{-3} \end{pmatrix}, \quad v_{\bar{\alpha}} = \begin{pmatrix} \bar{\alpha}^{-1} \\ \bar{\alpha}^{-2} \\ \bar{\alpha}^{-3} \end{pmatrix}.$$

Then  $v_{\beta}, v_{\alpha}, v_{\bar{\alpha}}$  are eigenvectors of  $M$  corresponding to  $\beta, \alpha, \bar{\alpha}$ , respectively, normalized so that in each case the sum of coordinates equals 1. Let us denote by  $a_{\beta}, a_{\alpha}, a_{\bar{\alpha}}$  the complex numbers for which

$$a_{\beta}v_{\beta} + a_{\alpha}v_{\alpha} + a_{\bar{\alpha}}v_{\bar{\alpha}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The numbers  $a_{\beta}, a_{\alpha}$  and  $a_{\bar{\alpha}}$  are unique, and therefore

$$a_{\bar{\alpha}} = \bar{a}_{\alpha} \quad \text{and} \quad |a_{\alpha}| = |a_{\bar{\alpha}}| = 0.14135\dots$$

Let us define

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then we can represent letter frequencies as  $\text{Freq}_{\mathbf{t}}(i) = \langle v_{\beta}, e_i \rangle$  for each letter  $i = 0, 1, 2$ , where  $\langle \cdot, \cdot \rangle$  denotes the Hermitian scalar product.

#### 2.4. Proof 2 of Theorem 1.3

Every natural number  $N$  has a unique Zeckendorff Tribonacci representation

$$N = \sum_{k \geq 0} p_k T_k,$$

where  $p_k \in \{0, 1\}$ , all but finitely many  $p_k$  equal 0, and if  $p_k = p_{k-1} = 1$ , then  $p_{k-2} = 0$ .

It can be shown (see [14, Proposition 10.7.4, p. 510]) that for each letter  $i \in \{0, 1, 2\}$ ,

$$|u_0 \cdots u_{N-1}|_i = a_{\beta} \langle v_{\beta}, e_i \rangle \sum_{k \geq 0} p_k \beta^k + a_{\alpha} \langle v_{\alpha}, e_i \rangle \sum_{k \geq 0} p_k \alpha^k + a_{\bar{\alpha}} \langle v_{\bar{\alpha}}, e_i \rangle \sum_{k \geq 0} p_k \bar{\alpha}^k,$$

and

$$N = a_{\beta} \sum_{k \geq 0} p_k \beta^k + a_{\alpha} \sum_{k \geq 0} p_k \alpha^k + a_{\bar{\alpha}} \sum_{k \geq 0} p_k \bar{\alpha}^k.$$

These two identities imply

$$\begin{aligned}
 |u_0 \cdots u_{N-1}|_i - N \langle v_\beta, e_i \rangle &= a_\alpha \langle v_\alpha - v_\beta, e_i \rangle \sum_{k \geq 0} p_k \alpha^k + a_{\bar{\alpha}} \langle v_{\bar{\alpha}} - v_\beta, e_i \rangle \sum_{k \geq 0} p_k \bar{\alpha}^k \\
 &= 2 \operatorname{Re} \left( a_\alpha \langle v_\alpha - v_\beta, e_i \rangle \sum_{k \geq 0} p_k \alpha^k \right), \\
 &= \sum_{k \geq 0} 2 p_k \operatorname{Re} (a_\alpha (\alpha^{-(i+1)} - \beta^{-(i+1)}) \alpha^k), \tag{1}
 \end{aligned}$$

where  $\operatorname{Re}(\cdot)$  denotes the real part of a complex number. The second equality above holds because complex conjugation commutes with multiplication and addition.

The following lemma shows that in order to prove Theorem 1.3, it suffices to determine adequate upper and lower bounds for the above equation (1):

**Lemma 2.3.** *Let  $x = x_0 x_1 x_2 \cdots$  be an infinite word on a finite alphabet, and let  $b$  be a letter occurring in  $x$ . If there exist real numbers  $A, B$  and  $\gamma$  for which*

$$A < |x_0 x_1 \cdots x_{N-1}|_b - N\gamma < B$$

for all integers  $N \geq 1$ , then the word  $x$  is  $\lfloor 2(B - A) \rfloor$ -balanced with respect to the letter  $b$ , i.e.,

$$\left| |U|_b - |V|_b \right| \leq \lfloor 2(B - A) \rfloor$$

for all factors  $U, V$  of equal length, where  $\lfloor 2(B - A) \rfloor$  denotes the integer part of  $2(B - A)$ .

**Proof.** Let  $i, j \geq 0$  be integers. Since

$$|x_i x_{i+1} \cdots x_{i+N-1}|_b - N\gamma = (|x_0 x_1 \cdots x_{i+N-1}|_b - (N + i)\gamma) - (|x_0 x_1 \cdots x_{i-1}|_b - i\gamma),$$

we see that

$$A - B < |x_i x_{i+1} \cdots x_{i+N-1}|_b - N\gamma < B - A. \tag{2}$$

Consequently,

$$\left| |x_i x_{i+1} \cdots x_{i+N-1}|_b - |x_j x_{j+1} \cdots x_{j+N-1}|_b \right| < (B - A) - (A - B) = 2(B - A). \quad \square$$

We now determine good upper and lower bounds for Eq. (1). The numerical calculations in Propositions 2.4, 2.5 and 2.6 were verified independently by the first and second authors via different computer programs.

**Proposition 2.4.** *We have*

$$-0.6 < |u_0 u_1 \cdots u_{N-1}|_0 - \frac{N}{\beta} < 0.9.$$

**Proof.** By (1), we have

$$|u_0 \cdots u_{N-1}|_0 - \frac{N}{\beta} = \sum_{k \geq 0} 2p_k \operatorname{Re} \left( a_\alpha \left( \frac{1}{\alpha} - \frac{1}{\beta} \right) \alpha^k \right).$$

Set

$$S_L = \sum_{0 \leq k \leq 7} 2p_k \operatorname{Re} \left( a_\alpha \left( \frac{1}{\alpha} - \frac{1}{\beta} \right) \alpha^k \right), \quad S_R = \sum_{k \geq 8} 2p_k \operatorname{Re} \left( a_\alpha \left( \frac{1}{\alpha} - \frac{1}{\beta} \right) \alpha^k \right).$$

Then we have

$$S_L - |S_R| \leq S_L + S_R = |u_0 \cdots u_{N-1}|_0 - \frac{N}{\beta} \leq S_L + |S_R|. \tag{3}$$

Choosing the  $p_k$  so that  $p_k = 1$  if and only if  $\operatorname{Re}(a_\alpha(\frac{1}{\alpha} - \frac{1}{\beta})\alpha^k)$  is positive, and similarly for negative values, it can be seen that

$$-0.42 < S_L < 0.73. \tag{4}$$

To estimate the magnitude of  $|S_R|$ , we compute

$$\left| \frac{1}{\alpha} - \frac{1}{\beta} \right| = 1.72457 \dots \tag{5}$$

Thus

$$|S_R| \leq 2|a_\alpha| \cdot \left| \frac{1}{\alpha} - \frac{1}{\beta} \right| \sum_{k \geq 8} |\alpha|^k = 2|a_\alpha| \cdot \left| \frac{1}{\alpha} - \frac{1}{\beta} \right| \frac{|\alpha|^8}{1 - |\alpha|} < 0.17. \tag{6}$$

The claim follows from inequalities (3), (4), and (6).  $\square$

**Proposition 2.5.** We have

$$-0.775 < |u_0 u_1 \cdots u_{N-1}|_1 - \frac{N}{\beta^2} < 0.725.$$

**Proof.** Analogous to the proof of Proposition 2.4, set

$$S_L = \sum_{0 \leq k \leq 10} 2p_k \operatorname{Re} \left( a_\alpha \left( \frac{1}{\alpha^2} - \frac{1}{\beta^2} \right) \alpha^k \right), \quad S_R = \sum_{k \geq 11} 2p_k \operatorname{Re} \left( a_\alpha \left( \frac{1}{\alpha^2} - \frac{1}{\beta^2} \right) \alpha^k \right).$$

Again, by choosing the  $p_k$  appropriately, it can be verified that

$$-0.70 < S_L < 0.65.$$

To find an upper bound for  $|S_R|$ , we compute

$$\left| \frac{1}{\alpha^2} - \frac{1}{\beta^2} \right| = 1.96298\dots \tag{7}$$

Therefore,

$$|S_R| \leq 2|a_\alpha| \cdot \left| \frac{1}{\alpha^2} - \frac{1}{\beta^2} \right| \sum_{k \geq 11} |\alpha|^k = 2|a_\alpha| \cdot \left| \frac{1}{\alpha^2} - \frac{1}{\beta^2} \right| \frac{|\alpha|^{11}}{1 - |\alpha|} < 0.075.$$

These inequalities imply the claim.  $\square$

**Proposition 2.6.** *We have*

$$-0.88 < |u_0 u_1 \cdots u_{N-1}|_2 - \frac{N}{\beta^3} < 0.62.$$

**Proof.** Once again we set

$$S_L = \sum_{0 \leq k \leq 13} 2p_k \operatorname{Re} \left( a_\alpha \left( \frac{1}{\alpha^3} - \frac{1}{\beta^3} \right) \alpha^k \right), \quad S_R = \sum_{k \geq 14} 2p_k \operatorname{Re} \left( a_\alpha \left( \frac{1}{\alpha^3} - \frac{1}{\beta^3} \right) \alpha^k \right).$$

As above, it can be verified that

$$-0.8371 < S_L < 0.5764.$$

We compute

$$\left| \frac{1}{\alpha^3} - \frac{1}{\beta^3} \right| = 2.33887\dots, \tag{8}$$

and thus

$$|S_R| \leq 2|a_\alpha| \cdot \left| \frac{1}{\alpha^3} - \frac{1}{\beta^3} \right| \sum_{k \geq 14} |\alpha|^k = 2|a_\alpha| \cdot \left| \frac{1}{\alpha^3} - \frac{1}{\beta^3} \right| \frac{|\alpha|^{14}}{1 - |\alpha|} < 0.0354.$$

The claim follows from these inequalities.  $\square$

Propositions 2.4–2.6 together with Lemma 2.3 imply that  $\mathbf{t}$  is 2-balanced.

**Remark 2.7.** None of the inequalities in the claims of Propositions 2.4–2.6 are optimal.

**Corollary 2.8.** *For all integers  $n \geq 0$  and factors  $w$  of length  $T_n$ , we have*

$$\left| |\tau^n(\mathbf{0})|_i - |w|_i \right| \leq 1$$

for all letters  $i = 0, 1, 2$ .

**Proof.** Since the Tribonacci representation of  $T_n$  is  $10^{n-1}$ , we see that (1), (5), (7), and (8) imply

$$\left| \left| \tau^n(0) \right|_i - \frac{T_n}{\beta^{i+1}} \right| = \left| 2 \operatorname{Re}(a_\alpha (\alpha^{-(i+1)} - \beta^{-(i+1)}) \alpha^n) \right| \leq 2 |a_\alpha| |(\alpha^{-(i+1)} - \beta^{-(i+1)})| |\alpha| < 0.49.$$

By Propositions 2.4–2.6 and (2), we get

$$\left| \left| \tau^n(0) \right|_i - |w|_i \right| \leq \left| \left| \tau^n(0) \right|_i - \frac{T_n}{\beta^{i+1}} \right| + \left| |w|_i - \frac{T_n}{\beta^{i+1}} \right| < 1.39 < 2,$$

and the claim follows.  $\square$

### 3. Abelian complexity

In this section we study the Abelian complexity of the Tribonacci word and prove Theorem 1.4. We begin by recalling the following basic fact from [20]:

**Fact 3.1.** If an infinite word  $\omega$  has two factors  $u$  and  $v$  of the same length  $n$  for which the  $i$ th entry of the Parikh vector are  $p$  and  $p + c$  respectively, then for each  $\ell = 0, \dots, c$ , there exists a factor  $u_\ell$  of  $\omega$  of length  $n$  such that the  $i$ th entry of  $\Psi(u_\ell)$  is equal to  $p + \ell$ .

Henceforth, the Abelian complexity of the Tribonacci word  $\mathbf{t}$  is denoted by  $\rho$ .

#### 3.1. Proof of Theorem 1.4

Let us recall that a factor  $u$  of an infinite word  $\omega$  is *right* (resp. *left*) *special* if there exist distinct letters  $a$  and  $b$  such that the words  $ua$  and  $ub$  (resp.  $au$ ,  $bu$ ) are both factors of  $\omega$ . A factor which is both left and right special is called *bispecial*.

It is well known that for every  $n \geq 1$ ,  $\mathbf{t}$  has exactly one right special factor of length  $n - 1$ , which we denote by  $\mathcal{R}_{n-1}$ , and that the three words  $\mathcal{R}_{n-1}0$ ,  $\mathcal{R}_{n-1}1$ , and  $\mathcal{R}_{n-1}2$  are all factors of  $\mathbf{t}$  of length  $n$ . Writing  $\Psi(\mathcal{R}_{n-1}) = (i, j, k)$ , where  $i, j, k$  are each non-negative integers, we define

$$\text{Central}(n) = \{(i + 1, j, k), (i, j + 1, k), (i, j, k + 1)\}.$$

We have

$$\text{Central}(n) \subseteq \Psi_{\mathbf{t}}(n). \tag{9}$$

Given a vector  $\vec{v} = (\alpha, \beta, \gamma)$ , set  $\|\vec{v}\| = \max(|\alpha|, |\beta|, |\gamma|)$ . The set of vectors  $\vec{v}$  such that  $\|\vec{v} - \vec{u}\| \leq 2$  for all  $\vec{u}$  in  $\text{Central}(n)$  is given by the graph in Fig. 1 (whose vertices are vectors, and where two vertices  $\vec{u}, \vec{v}$  are joined by an undirected edge if and only if  $\|\vec{v} - \vec{u}\| = 1$ ).

Since  $\mathbf{t}$  is 2-balanced,  $\Psi_{\mathbf{t}}(n)$  is a subset of this set of twelve vectors. Moreover for the same reason, we should have  $\|\vec{v} - \vec{u}\| \leq 2$  for all  $\vec{u}, \vec{v}$  in  $\Psi_{\mathbf{t}}(n)$ . This implies that the only possibility for  $\Psi_{\mathbf{t}}(n)$  is to be a subset of one of the three sets delimited by a regular hexagon in Fig. 1, or one of the three sets delimited by an equilateral triangle of base length 2. These sets have cardinalities 7 and 6 respectively showing that  $\rho_\omega(n) \leq 7$ .

**Remark 3.2.** By computer simulation we found that

$$(\rho(n))_{n \geq 1} = 3343444344444434444444444444345544444554444 \dots$$

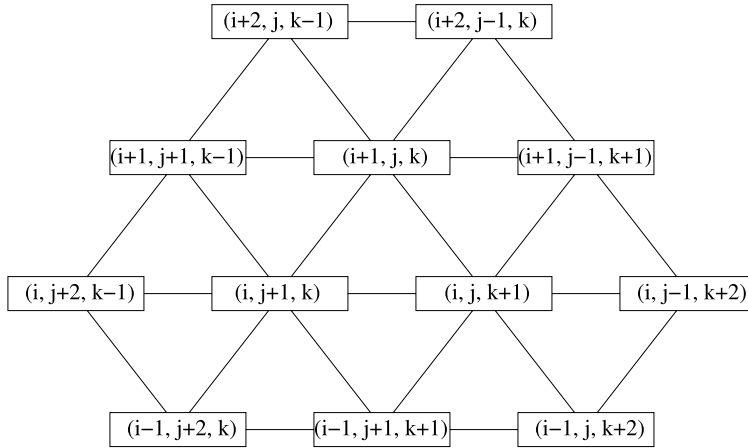


Fig. 1. Links between Parikh vectors.

In particular, the least  $n$  for which  $\rho(n) = 5$  is  $n = 30$ . We also found that the smallest  $n$  for which  $\rho(n) = 6$  is  $n = 342$ , and the smallest  $n$  for which  $\rho(n) = 7$  is  $n = 3914$ . The next four values of  $n$  for which  $\rho(n) = 7$  are  $n = 4063, 4841, 4990, 7199$ .

3.2. More on the Abelian complexity of the Tribonacci word

In this section, we characterize those  $n$  for which  $\rho_{\mathbf{t}}(n) = 3$ .

We continue to adopt the notation  $\mathcal{R}_{n-1}$ ,  $\Psi(\mathcal{R}_{n-1})$ ,  $\text{Central}(n)$ , and  $\|\vec{v}\|$  introduced in the proof of Theorem 1.4. Let  $B(n)$  be the set of vectors  $\vec{v}$  for which there is exactly one  $\vec{u}$  in  $\text{Central}(n)$  with  $\|\vec{v} - \vec{u}\| = 2$ :

$$B(n) = \{(i - 1, j + 1, k + 1), (i + 1, j - 1, k + 1), (i + 1, j + 1, k - 1)\}. \tag{10}$$

**Proposition 3.3.** *The following are equivalent:*

- (1) For all factors  $v$  and  $w$  of  $\mathbf{t}$  of length  $n$ , we have  $\| |v|_a - |w|_a \| \leq 1$  for all  $a \in \{0, 1, 2\}$ .
- (2)  $\rho_{\mathbf{t}}(n) = 3$ .
- (3)  $\Psi_{\mathbf{t}}(n) \cap B(n) = \emptyset$ .
- (4)  $\mathbf{t}$  contains a bispecial special factor of length  $n - 1$ .
- (5)  $n = 1$ , or  $n = \frac{1}{2}(T_m + T_{m+2} - 1)$  for  $m \geq 0$ , where  $(T_m)_{m \geq 0} = 1, 2, 4, 7, 13, 24, \dots$  denotes the sequence of Tribonacci numbers.

**Proof.** Equivalence (1)  $\Leftrightarrow$  (2) is an immediate consequence of Eq. (9). To see the equivalence between items (4) and (5), we note that the bispecial factors of  $\mathbf{t}$  are precisely the palindromic prefixes of  $\mathbf{t}$  (see [21]). The lengths of these words are known to be

$$\frac{T_m + (T_{m-1} + T_m) + (T_{m-2} + T_{m-1} + T_m) - 3}{2} = \frac{T_m + T_{m+2} - 3}{2}$$

(see Corollary 3.11 in [7] or Corollary 3.3 in [24]). It follows that (4)  $\Leftrightarrow$  (5). Thus it remains to prove that items (2), (3), and (4) are equivalent. We will make use of the following lemmas.

**Lemma 3.4.** Let  $n \geq 1$ , and let  $v$  be a factor of  $\mathbf{t}$ . Put  $\Psi(\mathcal{R}_{n-1}) = (i, j, k)$ , and  $\Psi(v) = (i', j', k')$ . Suppose either

- $|v| \geq n$  and either  $i' \leq i - 1$  or  $j' \leq j - 1$ ,

or

- $|v| \leq n$ , and either  $i' \geq i + 2$  or  $j' \geq j + 2$ .

Then  $\rho(n) > 3$ .

**Proof.** The first condition implies the existence of a factor  $v'$  of  $\mathbf{t}$  of length  $n$  with either  $|v'|_0 \leq i - 1$  or  $|v'|_1 \leq j - 1$ . By (9) it follows that the factors of  $\mathbf{t}$  of length  $n$  are not balanced. Similarly, the second condition implies the existence of a factor  $v'$  of  $\mathbf{t}$  of length  $n$  with either  $|v'|_0 \geq i + 2$  or  $|v'|_1 \geq j + 2$ . Again by (9) it follows that the factors of  $\mathbf{t}$  of length  $n$  are not balanced.  $\square$

For each  $n \geq 1$ , let  $\phi(n) = |\tau(\mathcal{R}_{n-1})\mathbf{0}| + 1$ .

**Lemma 3.5.** If  $\rho(n) = 3$ , then  $\rho(\phi(n)) = 3$ .

**Proof.** Assume  $\rho(n) = 3$ , and set  $N = \phi(n)$ . Let  $\Psi(\mathcal{R}_{n-1}) = (i, j, k)$ , and  $\mathcal{R}_{N-1} = (I, J, K)$ . Since  $\mathcal{R}_{N-1} = \tau(\mathcal{R}_{n-1})\mathbf{0}$ , it follows that

$$I = n; \quad J = i; \quad K = j. \tag{11}$$

Now we have

$$\Psi_{\mathbf{t}}(n) = \{(i + 1, j, k), (i, j + 1, k), (i, j, k + 1)\}$$

and

$$\{(I + 1, J, K), (I, J + 1, K), (I, J, K + 1)\} \subseteq \Psi_{\mathbf{t}}(N)$$

and we must show that in fact we have equality in the previous containment. By Theorem 1.3, the only other possible Parikh vectors of factors of  $\mathbf{t}$  of length  $N$  are one of the following six vectors

$$\begin{pmatrix} I-1 \\ J \\ K+2 \end{pmatrix}; \quad \begin{pmatrix} I-1 \\ J+2 \\ K \end{pmatrix}; \quad \begin{pmatrix} I \\ J-1 \\ K+2 \end{pmatrix}; \quad \begin{pmatrix} I \\ J+2 \\ K-1 \end{pmatrix}; \quad \begin{pmatrix} I+2 \\ J-1 \\ K \end{pmatrix}; \quad \begin{pmatrix} I+2 \\ J \\ K-1 \end{pmatrix}$$

or one of the following three vectors

$$\begin{pmatrix} I+1 \\ J+1 \\ K-1 \end{pmatrix}; \quad \begin{pmatrix} I+1 \\ J-1 \\ K+1 \end{pmatrix}; \quad \begin{pmatrix} I-1 \\ J+1 \\ K+1 \end{pmatrix}.$$

Using Lemma 3.4 we will show that if any one of these nine vectors belong to  $\Psi_{\mathbf{t}}(n)$ , then  $\rho(n) > 3$ , a contradiction. Hence  $\rho(N) = 3$ . We proceed one vector at a time.



- ◇ Suppose  $(I - 1, J, K + 2) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I - 1 &= |v| + \delta, \\ J &= i', \\ K + 2 &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n - 1 - \delta \leq n$ , and  $j' = j + 2$  which contradicts Lemma 3.4.

- ◇ Suppose  $(I - 1, J + 2, K) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I - 1 &= |v| + \delta, \\ J + 2 &= i', \\ K &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n - 1 - \delta \leq n$ , and  $i' = i + 2$  which contradicts Lemma 3.4.

- ◇ Suppose  $(I, J - 1, K + 2) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I &= |v| + \delta, \\ J - 1 &= i', \\ K + 2 &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n - \delta \leq n + 1$ ,  $i' = i - 1$ , and  $j' = j + 2$  which contradicts Lemma 3.4.

- ◇ Suppose  $(I, J + 2, K - 1) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I &= |v| + \delta, \\ J + 2 &= i', \\ K - 1 &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n - \delta \leq n + 1$ ,  $i' = i + 2$ , and  $j' = j - 1$  which contradicts Lemma 3.4.

- ◇ Suppose  $(I + 2, J - 1, K) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I + 2 &= |v| + \delta, \\ J - 1 &= i', \\ K &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n + 2 - \delta \geq n + 1$ , and  $i' = i - 1$  which contradicts Lemma 3.4.

- ◇ Suppose  $(I + 2, J, K - 1) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I + 2 &= |v| + \delta, \\ J &= i', \\ K - 1 &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n + 2 - \delta \geq n + 1$ , and  $j' = j - 1$  which contradicts Lemma 3.4. This concludes the first six cases. Now we consider the last three:

- ◇ Suppose  $(I + 1, J + 1, K - 1) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I + 1 &= |v| + \delta, \\ J + 1 &= i', \\ K - 1 &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n + 1 - \delta \geq n$ , and  $j' = j - 1$  which contradicts Lemma 3.4.

- ◇ Suppose  $(I + 1, J - 1, K + 1) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I + 1 &= |v| + \delta, \\ J - 1 &= i', \\ K + 1 &= j'. \end{aligned}$$

It follows from (11) that  $|v| = n + 1 - \delta \geq n$ , and  $i' = i - 1$  which contradicts Lemma 3.4.

- ◇ Suppose  $(I - 1, J + 1, K + 1) \in \Psi_{\mathbf{t}}(N)$ . By Lemma 2.1 there exists a factor  $v$  of  $\mathbf{t}$  with  $\Psi(v) = (i', j', k')$  and  $\delta \in \{-1, 0, 1\}$  such that

$$\begin{aligned} I - 1 &= |v| + \delta, \\ J + 1 &= i', \\ K + 1 &= j'. \end{aligned}$$

It follows from (11) that  $|v| = i' + j' + k' = n - 1 - \delta = i + j + k - \delta$ ,  $i' = i + 1$ , and  $j' = j + 1$ . Thus we have  $k' = k - 2 - \delta$ . If  $\delta = -1$ , then  $(i + 1, j + 1, k - 1) \in \Psi_{\mathbf{t}}(n)$  contradicting  $\rho(n) = 3$ . If  $\delta = 0$ , then  $|v| = n - 1$  and  $k' = k - 2$ . Thus any factor of  $\mathbf{t}$  of length  $n$  beginning with  $v$  will have at most  $k - 1$  many 2's, contradicting the balance between any two factors of length  $n$  being at most one (there is a factor of length  $n$  with Parikh vector  $(i, j, k + 1)$ ). If  $\delta = 1$ , then  $|v| = n - 2$  and  $k' = k - 3$ . Thus any factor of length  $n$  beginning with  $v$  will have at most  $k - 2$  many 2's, contradicting Theorem 1.3.

This concludes the proof of the lemma.  $\square$

**Lemma 3.6.** *If  $\Psi_{\mathbf{t}}(n) \cap B(n) \neq \emptyset$ , then  $\Psi_{\mathbf{t}}(\phi(n)) \cap B(\phi(n)) \neq \emptyset$ .*

**Proof.** Since  $\Psi(\mathcal{R}_{n-1}) = (i, j, k)$ , we have that  $\phi(n) = n + i + j + 1$ . Following (10), if  $\Psi(v) = (i - 1, j + 1, k + 1)$  for some factor  $v$  of  $\mathbf{t}$  of length  $n$ , then  $\Psi(\tau(v)0) = (n + 1, i - 1, j + 1) \in \Psi_{\mathbf{t}}(\phi(n)) \cap B(\phi(n))$ . If  $\Psi(v) = (i + 1, j - 1, k + 1)$  for some factor  $v$  of  $\mathbf{t}$  of length  $n$ , then  $\Psi(\tau(v)0) = (n + 1, i + 1, j - 1) \in \Psi_{\mathbf{t}}(\phi(n)) \cap B(\phi(n))$ . Finally if  $\Psi(v) = (i + 1, j + 1, k - 1)$  for some factor  $v$  of  $\mathbf{t}$  of length  $n$ , then  $\Psi(0^{-1}\tau(v)) = (n - 1, i + 1, j + 1) \in \Psi_{\mathbf{t}}(\phi(n)) \cap B(\phi(n))$ .  $\square$

**Lemma 3.7.** *If  $\Psi_{\mathbf{t}}(n) \cap B(n) = \emptyset$ , then  $\mathcal{R}_{n-1}$  is a bispecial factor of  $\mathbf{t}$ .*

**Proof.** We proceed by induction on  $n$ . The result is clear for  $n = 1, 2$ . Let  $N > 2$ , and suppose the result is true for all  $n < N$ . Assume  $\Psi_{\mathbf{t}}(N) \cap B(N) = \emptyset$ . Set  $\Psi(\mathcal{R}_{N-1}) = (I, J, K)$ . Since  $\mathcal{R}_{N-1}$  is right special, it ends with 0. We now show that  $\mathcal{R}_{N-1}$  also begins with 0. Suppose  $\mathcal{R}_{N-1}$  begins with 1. Then,  $1^{-1}\mathcal{R}_{N-1}20$  is a factor of  $\mathbf{t}$  of length  $N$ , and  $\Psi(1^{-1}\mathcal{R}_{N-1}20) = (I + 1, J - 1, K + 1) \in \Psi_{\mathbf{t}}(N) \cap B(N)$  contrary to our assumption. Similarly, assume  $\mathcal{R}_{N-1}$  begins with 2. Then,  $2^{-1}\mathcal{R}_{N-1}10$  is a factor of  $\mathbf{t}$  of length  $N$ , and  $\Psi(2^{-1}\mathcal{R}_{N-1}10) = (I + 1, J + 1, K - 1) \in \Psi_{\mathbf{t}}(N) \cap B(N)$  contrary to our assumption.

Having established that  $\mathcal{R}_{N-1}$  begins with 0, we can write  $\mathcal{R}_{N-1} = \tau(u)0$  for some factor  $u$  of  $\mathbf{t}$ . Also, as  $\mathcal{R}_{N-1}$  is right special, so is  $u$  and hence  $u = \mathcal{R}_{n-1}$  for some  $n$ , and so  $N = \phi(n)$ . It follows from the previous lemma that  $\Psi_{\mathbf{t}}(n) \cap B(n) = \emptyset$ . Hence by the induction hypothesis we have that  $\mathcal{R}_{n-1}$  is bispecial, and hence so is  $\tau(\mathcal{R}_{n-1})0 = \mathcal{R}_{N-1}$ .  $\square$

We are now ready to establish the remaining equivalences in Proposition 3.3: We will show that (3)  $\Rightarrow$  (4)  $\Rightarrow$  (2)  $\Rightarrow$  (3). The previous lemma states precisely that (3)  $\Rightarrow$  (4). The implication (2)  $\Rightarrow$  (3) is clear from (9) and (10). Finally to see that (4)  $\Rightarrow$  (2), we proceed by induction on  $n$ . The result is clear for  $n = 1, 2$ . Let  $N > 2$  and suppose the result is true for all  $n < N$ . Assume  $\mathcal{R}_{N-1}$  is bispecial. Then as  $\mathcal{R}_{N-1}$  begins and ends with 0, we can write  $\mathcal{R}_{N-1} = \tau(u)0$  for some factor  $u$  of  $\mathbf{t}$ . Moreover as  $\mathcal{R}_{N-1}$  is bispecial, so is  $u$ . It follows that  $u = \mathcal{R}_{n-1}$  for some  $n$ , and hence  $N = \phi(n)$ . By induction hypothesis we have that  $\rho(n) = 3$ . It now follows from Lemma 3.5 that  $\rho(N) = 3$  as required. This completes our proof of Proposition 3.3.  $\square$

As an immediate consequence of Proposition 3.3, we have that  $\rho(n) = 3$  for infinitely many values of  $n$ . We next show that the same is true for  $\rho(n) = 7$ .

**Proposition 3.8.** *The Abelian complexity of the Tribonacci word attains the value 7 infinitely often.*

**Proof.** Note that  $\mathbf{t}$  begins with the square  $(0102010)^2 = \tau^3(00)$  and so with the squares  $\tau^{n+3}(00)$  for each  $n \geq 0$ . Consider any integer  $n \geq 0$  for which all factors of length  $m = 3914$  occur in  $\tau^{n+3}(0)$  (we recall that a computer computation showed that  $\rho(3914) = 7$ ). For any factor  $y$  of length  $m$  there exists a cyclic conjugate (i.e., a circular shift)  $x$  of  $\tau^{n+3}(0)$  such that  $xy$  is a factor of  $\tau^{n+3}(00)$ . Since  $\Psi(x) = \Psi(\tau^{n+3}(0))$  and since  $\Psi(xy) = \Psi(x) + \Psi(y)$ , we deduce that  $\text{Card}(\Psi_{\mathbf{t}}(|\tau^{n+3}(0)| + m)) \geq \text{Card}(\Psi_{\mathbf{t}}(m)) = 7$ . Since the Abelian complexity of the Tribonacci word is bounded by 7, we get  $\text{Card}(\Psi_{\mathbf{t}}(|\tau^{n+3}(0)| + m)) = 7$ .  $\square$

#### 4. Conclusion

Various aspects of the Abelian complexity of the Tribonacci word remain a mystery. For instance, it is surprising to us that the value  $\rho_{\mathbf{t}}(n) = 7$  does not occur until  $n = 3914$ , but then re-occurs relatively shortly thereafter. As another example, it is verified that for all  $n \leq 184$ , if  $U$  and  $V$  are factors of  $\mathbf{t}$  of length  $n$ , with  $U$  a prefix of  $\mathbf{t}$ , then  $\|U\|_a - \|V\|_a \leq 1$ , for all  $a \in \{0, 1, 2\}$ . But then this property fails for  $n = 185$ . We list a few related open questions:

**Open problem 1.** Does the Abelian complexity of the Tribonacci word attain each value in  $\{4, 5, 6\}$  infinitely often?

**Open problem 2.** For each value  $m \in \{4, 5, 6, 7\}$ , characterize those  $n$  for which  $\rho(n) = m$ .

Another problem concerns the  $m$ -bonacci word, the generalization of the Tribonacci word (and of the celebrated Fibonacci word) to an alphabet of  $m$  letters. This word is defined as the fixed point of the morphism  $\tau_m : \{0, \dots, m-1\}^* \rightarrow \{0, \dots, m-1\}^*$  defined by  $\tau_m(i) = 0(i+1)$  for  $0 \leq i \leq m-2$  and  $\tau(m-1) = 0$ .

**Open problem 3.** Prove or disprove that the  $m$ -bonacci word is  $(m-1)$ -balanced.

## Acknowledgments

The second author is partially supported by grant no. 8121419 from the Finnish Academy. The third author is partially supported by grant no. 090038011 from the Icelandic Research Fund, and by the grant *SubTile* funded by the A.N.R. We would like to especially thank the anonymous referee for carefully reading the manuscript and for providing us with many helpful suggestions which have greatly improved the exposition of the paper.

## References

- [1] B. Adamczewski, Balances for fixed points of primitive substitutions, *Theoret. Comput. Sci.* 307 (2003) 47–75.
- [2] P. Arnoux, V. Berthé, S. Ito, Discrete planes,  $\mathbb{Z}^2$ -actions, Jacobi–Perron algorithm and substitutions, *Ann. Inst. Fourier (Grenoble)* 52 (2002) 305–349.
- [3] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité  $2n+1$ , *Bull. Soc. Math. France* 119 (1991) 199–215.
- [4] J. Berstel, Recent results on extensions of Sturmian words, in: *International Conference on Geometric and Combinatorial Methods in Group Theory and Semigroup Theory*, Lincoln, NE, 2000, *Internat. J. Algebra Comput.* 12 (2002) 371–385.
- [5] J. Berstel, Sturmian and episturmian words (a survey of some recent results), in: *CAI 2007*, in: *Lecture Notes in Comput. Sci.*, vol. 4728, 2007, pp. 23–47.
- [6] J. Cassaigne, S. Ferenczi, L.Q. Zamboni, Imbalances in Arnoux–Rauzy sequences, *Ann. Inst. Fourier (Grenoble)* 50 (2000) 1265–1276.
- [7] A. de Luca, L.Q. Zamboni, On graphs of central episturmian words, *Theoret. Comput. Sci.* 411 (2010) 70–90.
- [8] S. Ferenczi, C. Holton, L.Q. Zamboni, The structure of three-interval exchange transformations I: an arithmetic study, *Ann. Inst. Fourier (Grenoble)* 51 (2001) 861–901.
- [9] S. Ferenczi, C. Holton, L.Q. Zamboni, The structure of three-interval exchange transformations II: a combinatorial description of the trajectories, *J. Anal. Math.* 89 (2003) 239–276.
- [10] S. Ferenczi, C. Holton, L.Q. Zamboni, The structure of three-interval exchange transformations III: ergodic and spectral properties, *J. Anal. Math.* 93 (2004) 103–138.
- [11] S. Ferenczi, C. Holton, L.Q. Zamboni, Joinings of three-interval exchange transformations, *Ergodic Theory Dynam. Systems* 25 (2005) 483–502.
- [12] A. Glen, J. Justin, Episturmian words: A survey, *RAIRO-Theor. Inform. Appl.* 43 (2009) 402–433.
- [13] M. Lothaire, *Algebraic Combinatorics on Words*, *Encyclopedia Math. Appl.*, vol. 90, Cambridge University Press, 2002.
- [14] M. Lothaire, *Applied Combinatorics on Words*, *Encyclopedia Math. Appl.*, vol. 105, Cambridge University Press, 2005.
- [15] M. Morse, G.A. Hedlund, Symbolic dynamics, *Amer. J. Math.* 60 (1938) 815–866.
- [16] M. Morse, G.A. Hedlund, Symbolic dynamics II: Sturmian trajectories, *Amer. J. Math.* 62 (1940) 1–42.
- [17] G. Rauzy, Une généralisation du développement en fraction continue, *Séminaire de Théorie des Nombres Années 1975–1977*, Paris.
- [18] G. Rauzy, Échanges d’intervalles et transformations induites, *Acta Arith.* 34 (1979) 315–328.
- [19] G. Rauzy, Nombres algébriques et substitutions, *Bull. Soc. Math. France* 110 (1982) 147–178.
- [20] G. Richomme, K. Saari, L.Q. Zamboni, Abelian complexity of minimal subshifts, preprint, see <http://fr.arxiv.org/abs/0911.2914v1>, 2009.
- [21] R. Risley, L.Q. Zamboni, A generalization of Sturmian sequences: combinatorial structure and transcendence, *Acta Arith.* 95 (2000) 167–184.
- [22] N.J.A. Sloane, The on-line encyclopedia of integer sequences, published electronically at <http://www.research.att.com/~njas/sequences/>, 2009.
- [23] L. Vuillon, Balanced words, *Bull. Belg. Math. Soc. Simon Stevin* 10 (2003) 787–805.
- [24] N. Wozny, L.Q. Zamboni, Frequencies of factors in Arnoux–Rauzy sequences, *Acta Arith.* 96 (2001) 261–278.