

Accepted Manuscript

Critical exponents of infinite balanced words

Narad Rampersad, Jeffrey Shallit, Élise Vandomme

PII: S0304-3975(18)30638-8
DOI: <https://doi.org/10.1016/j.tcs.2018.10.017>
Reference: TCS 11775

To appear in: *Theoretical Computer Science*

Received date: 19 June 2018
Revised date: 4 October 2018
Accepted date: 15 October 2018



Please cite this article in press as: N. Rampersad et al., Critical exponents of infinite balanced words, *Theoret. Comput. Sci.* (2018), <https://doi.org/10.1016/j.tcs.2018.10.017>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The final publication is available at Elsevier via <https://doi.org/10.1016/j.tcs.2018.10.017>. © 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Critical exponents of infinite balanced words

Narad Rampersad^{1,*}

University of Winnipeg (Math/Stats), Winnipeg, MB, R3B 2E9, CANADA

Jeffrey Shallit¹

School of Computer Science, University of Waterloo, Waterloo, ON, N2L 3G1, CANADA

Élise Vandomme

*Laboratoire de Combinatoire et d'informatique Mathématique (LaCIM)
Université du Québec à Montréal, Montréal, QC, H3C 3P8, CANADA*

Abstract

Over an alphabet of size 3 we construct an infinite balanced word with critical exponent $2 + \sqrt{2}/2$. Over an alphabet of size 4 we construct an infinite balanced word with critical exponent $(5 + \sqrt{5})/4$. Over larger alphabets, we give some candidates for balanced words (found computationally) having small critical exponents. We also explore a method for proving these results using the automated theorem prover Walnut.

Keywords: infinite word, balanced word, Sturmian word, critical exponent

This paper is dedicated to the memory of Maurice Nivat.

1. Introduction

A word w (finite or infinite) is *balanced* if, for any two factors u and v of w of the same length, the number of occurrences of each alphabet symbol in u and v differ by at most 1. Vuillon [25] gives a survey of some of the work done previously on balanced words. It is well-known that over a binary alphabet

*Corresponding author

Email addresses: `n.rampersad@uwinnipeg.ca` (Narad Rampersad),
`shallit@uwaterloo.ca` (Jeffrey Shallit), `elise.vandomme@lacim.ca` (Élise Vandomme)

¹The author is supported by an NSERC Discovery Grant.

the class of infinite aperiodic balanced words is exactly the class of Sturmian words. Sturmian words have been studied extensively (see the survey by Berstel and Séébold [2]), and in particular, much is known about the repetitions that occur in Sturmian words. The *critical exponent* of an infinite word w is the supremum of the set of exponents of fractional powers appearing in w . Damanik and Lenz [4] and Justin and Pirillo [12] gave an exact formula for the critical exponent of a Sturmian word. The Fibonacci word has critical exponent $(5 + \sqrt{5})/2$ [15], and furthermore, by the formula previously mentioned [4, 12], this is minimal over all Sturmian words. In other words, over a binary alphabet, the least critical exponent among all infinite balanced words is $(5 + \sqrt{5})/2$. However, little is known about the critical exponents of infinite balanced words over larger alphabets. In this paper we aim to construct infinite balanced words over a given alphabet having the smallest critical exponent possible. Over an alphabet of size 3 we construct a balanced word with critical exponent $2 + \sqrt{2}/2$. Over an alphabet of size 4 we construct a balanced word with critical exponent $(5 + \sqrt{5})/4$. Over larger alphabets, we give some candidates for balanced words (found computationally) having small critical exponents. We also explore a method for proving these results using the automated theorem prover Walnut [17].

2. Preliminaries

For basic definitions of standard terms used in combinatorics on words, the reader may consult the book by Lothaire [14]. We let $|w|$ denote the length of a finite word w , and if a is a letter of the alphabet, we let $|w|_a$ denote the number of occurrences of a in w .

Definition 1. A word w (finite or infinite) over an alphabet A is *balanced* if for every $a \in A$ and every pair u, v of factors of w with $|u| = |v|$ we have

$$||u|_a - |v|_a| \leq 1.$$

Let u be a finite word and write $u = u_0u_1 \cdots u_{n-1}$, where the u_i are letters. A positive integer p is a *period* of u if $u_i = u_{i+p}$ for all i . Let $e = |u|/p$ and let

z be the prefix of u of length p . We say that u has *exponent* e and write $u = z^e$. The word z is called a *fractional root* of u . Note that a word may have multiple periods, and consequently, multiple exponents and fractional roots. The word u is *primitive* if the only integer exponent of u is 1. Let w be a finite or infinite word. The largest $r \in \mathbb{N}$ (if it exists) such that u^r is a factor of w is the (integral) *index* of u in w .

Definition 2. The *critical exponent* of an infinite word w is

$$\begin{aligned} E(w) &= \sup\{r \in \mathbb{Q} : \text{there is a finite, non-empty factor of } w \text{ with exponent } r\} \\ &= \inf\{r \in \mathbb{Q} : \text{there is no finite, non-empty factor of } w \text{ with exponent } r\}. \end{aligned}$$

The infinite words studied in this paper are constructed by modifying Sturmian words. The structure of such words are determined by a parameter α , which is an irrational real number between 0 and 1, called the *slope*, and more specifically, by the continued fraction expansion $\alpha = [0, d_1, d_2, d_3, \dots]$, where $d_i \in \mathbb{Z}$ for $i \geq 0$.

Definition 3. The *characteristic Sturmian word with slope* α (see [1, Chapter 9]) is the infinite word c_α obtained as the limit of the sequence of *standard words* s_n defined by

$$s_0 = 0, \quad s_1 = 0^{d_1-1}1, \quad s_n = s_{n-1}^{d_n} s_{n-2}, \quad n \geq 2.$$

We write $c_\alpha[i]$ to denote the i -th letter of c_α , where we index starting from 1. For $n \geq 2$, we also define the *semi-standard words*

$$s_{n,t} = s_{n-1}^t s_{n-2},$$

where $t \in \mathbb{Z}$ and $1 \leq t < d_n$.

We also make use of the *convergents* of α , namely

$$\frac{p_n}{q_n} = [0, d_1, d_2, d_3, \dots, d_n],$$

where

$$\begin{aligned} p_{-2} &= 0, & p_{-1} &= 1, & p_n &= d_n p_{n-1} + p_{n-2} \text{ for } n \geq 0; \\ q_{-2} &= 1, & q_{-1} &= 0, & q_n &= d_n q_{n-1} + q_{n-2} \text{ for } n \geq 0. \end{aligned}$$

The convergents have the following approximation property:

$$\left| \alpha - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}. \quad (1)$$

The following fact is classical:

$$\frac{q_{n+1}}{q_n} = [d_{n+1}, d_n, \dots, d_1]. \quad (2)$$

Another classical result is that the continued fraction expansion for α is ultimately periodic if and only if α is a quadratic irrational. We will need the reverse direction of the following result as well.

Theorem 1 ([13]). *The sequences $(q_n)_{n \geq 0}$ and $(p_n)_{n \geq 0}$ satisfy a linear recurrence with constant coefficients if and only if α is a quadratic irrational. For both sequences the linear recurrence is the same and has the form*

$$q_{n+2s} - tq_{n+s} + (-1)^s q_n = 0, \quad n \geq r, \quad (3)$$

for some r, s, t .

It is easy to see that $|s_n| = q_n$ for $n \geq 1$. We need a more precise result:

Lemma 2 ([1, Lemma 9.1.9]). *For $n \geq 0$ we have $|s_n|_0 = q_n - p_n$ and $|s_n|_1 = p_n$.*

Next we introduce the *Ostrowski α -numeration system* [19] (see also [1, Section 3.9]). Each non-negative integer N can be represented uniquely as $b_j b_{j-1} \cdots b_0$, where

$$N = \sum_{0 \leq i \leq j} b_i q_i,$$

where the b_i are integers satisfying:

1. $0 \leq b_0 < d_1$,
2. $0 \leq b_i \leq d_{i+1}$, for $i \geq 1$, and
3. for $i \geq 1$, if $b_i = d_{i+1}$, then $b_{i-1} = 0$.

We call the string $b_j b_{j-1} \cdots b_0$ the *Ostrowski α -representation* of N . The next two results give the connection between characteristic Sturmian words and the Ostrowski numeration system.

Theorem 3 ([1, Theorem 9.1.13]). *Let $N \geq 1$ be an integer with Ostrowski α -representation $b_j b_{j-1} \cdots b_0$. Then the length- N prefix of c_α is equal to $s_j^{b_j} s_{j-1}^{b_{j-1}} \cdots s_0^{b_0}$.*

Theorem 4 ([1, Theorem 9.1.15]). *Let $N \geq 1$ be an integer with Ostrowski α -representation $b_j b_{j-1} \cdots b_0$. Then $c_\alpha[N] = 1$ if and only if $b_j b_{j-1} \cdots b_0$ ends with an odd number of 0's.*

One characteristic Sturmian word is of particular significance. Let $\phi = (1 + \sqrt{5})/2$. The *Fibonacci word* is the characteristic Sturmian word

$$c_\theta = 010010100100101001010010010100 \cdots$$

with slope $\theta := 1/\phi^2 = [0, 2, \bar{1}]$. We call the corresponding standard words the *finite Fibonacci words*:

$$f_0 = 0, \quad f_1 = 01, \quad f_2 = 010, \quad \dots$$

It is easy to verify that for each $i \geq 2$ the finite Fibonacci word f_i has length F_{i+2} (the $(i+2)$ -th Fibonacci number) and has F_{i+1} 0's and F_i 1's. It is also well-known that the infinite Fibonacci word is fixed by the morphism that maps $0 \rightarrow 01$ and $1 \rightarrow 0$. Mignosi and Pirillo [15] showed that $E(c_\theta) = 2 + \phi$. The more general results of Damanik and Lenz [4] and Justin and Pirillo [12] show that this is minimal over all Sturmian words.

3. Constructing balanced words

We need a characterization of recurrent aperiodic balanced words due to Hubert [11] (Graham [9] gave an equivalent characterization), which is based on the following notion: a word y has the *constant gap* property if, for each letter a , there is some number d such that the distance between successive occurrences of a in y is always d . Constant gap words are periodic, but not all periodic words are constant gap words. For example, the periodic word $(0120)^\omega$ is *not* a constant gap word, since occurrences of 0's are sometimes at distance 3 from the previous occurrence and sometimes at distance 1. On the other hand, the

word $(0102)^\omega$ is a constant gap word, since the distance between successive 0's is always 2, and the distance between successive 1's (and similarly 2's) is always 4.

Theorem 5 ([11]). *A recurrent aperiodic word x is balanced if and only if x is obtained from a Sturmian word u over $\{0, 1\}$ by the following procedure: Let $y = y_0y_1 \cdots$ and $y' = y'_0y'_1 \cdots$ be constant gap sequences over disjoint alphabets A and B respectively. For each $i \geq 0$, replace the i -th occurrence of 0 (resp. 1) in u by the letter y_i (resp. y'_i).*

For example, if u is a Sturmian word that begins with 01010010100101010010100, $y = (0102)^\omega$, and $y' = (34)^\omega$, then we have $x = 03140230410324031042301 \cdots$.

Note that we shall only consider recurrent infinite words in this paper, since a well-known result of Furstenberg [7] states that for every infinite word x there exists a uniformly recurrent infinite word x' such that every factor of x' is a factor of x .

For $3 \leq k \leq 10$ we define an infinite word x_k constructed “à la Hubert” from a Sturmian word c_α , where we set α , y and y' according to Table 1.

k	α	c.f.	y	y'
3	$\sqrt{2} - 1$	$[0, \bar{2}]$	$(01)^\omega$	2^ω
4	$1/\phi^2$	$[0, 2, \bar{1}]$	$(01)^\omega$	$(23)^\omega$
5	$\sqrt{2} - 1$	$[0, \bar{2}]$	$(0102)^\omega$	$(34)^\omega$
6	$(78 - 2\sqrt{6})/101$	$[0, 1, 2, 1, 1, \overline{1, 1, 1, 2}]$	0^ω	$(123415321435)^\omega$
7	$(63 - \sqrt{10})/107$	$[0, 1, 1, 3, \overline{1, 2, 1}]$	$(01)^\omega$	$(234526432546)^\omega$
8	$(23 + \sqrt{2})/31$	$[0, 1, 3, 1, \bar{2}]$	$(01)^\omega$	$(234526732546237526432576)^\omega$
9	$(23 - \sqrt{2})/31$	$[0, 1, 2, 3, \bar{2}]$	$(01)^\omega$	$(234567284365274863254768)^\omega$
10	$(109 + \sqrt{13})/138$	$[0, 1, 4, 2, \bar{3}]$	$(01)^\omega$	$(234567284963254768294365274869)^\omega$

Table 1: c_α and constant gap words y and y' for the construction of x_k .

We will show that

$$E(x_3) = 2 + \frac{\sqrt{2}}{2} \text{ and } E(x_4) = 1 + \frac{\phi}{2}.$$

For $k \geq 5$, computer calculations suggest that

$$E(x_k) = \frac{k-2}{k-3}.$$

For $5 \leq k \leq 9$, we have established via backtracking searches that there are no infinite balanced words with critical exponent less than this value; consequently, if we could prove that $E(x_k)$ has the claimed value for these alphabet sizes, this would show that this is indeed the least possible critical exponent for balanced words over a k -letter alphabet. However, the backtracking algorithm used to establish this is not the usual one, but instead involves backtracking over the tree of *standard pairs*. This is an (infinite) binary tree with root $(0, 1)$; each vertex (u, v) has children (u, uv) and (vu, v) (see [5, p. 254] for more details). Every finite balanced binary word appears as a factor of either u or v for some node (u, v) appearing in this tree.

Pseudocode for the backtracking search is given in Algorithm 1. We assume the existence of two subroutines: $\text{CriticalExponent}(x)$, which computes the largest exponent over all factors of x , and $\text{GenerateConstantGapWords}(A)$, which returns the set² of all primitive roots of words with the constant gap property over the alphabet A . In Step 6 we do not consider $i > k/2$ because the standard pairs as enumerated contain both the word x and its complement. If the algorithm with input η and k terminates, we have proven that no balanced word over a k -letter alphabet has critical exponent less than η .

4. Establishing the critical exponent

We first observe that if x is an infinite ternary word obtained from a Sturmian word u as described by Theorem 5, then $E(u) \geq E(x) \geq E(u)/2$. The first

²For more information on how to enumerate this set, see the paper by Goulden et al. [8]. Constant gap words over a k -letter alphabet are equivalent to *exact covering systems* of size k . In the paper by Goulden et al., they count a subset of these consisting of the *natural exact covering systems*; however, the numbers of exact covering systems and natural exact covering systems are equal up to size 12.

Algorithm 1 Find balanced words with critical exponent $< \eta$

```

1: procedure STANDARDPAIRSBACKTRACK( $\eta, k$ )     $\triangleright k$  is the alphabet size
2:    $Q \leftarrow \text{Queue}((0, 1))$ 
3:   while  $Q$  is non-empty do
4:      $(u, v) \leftarrow Q.\text{dequeue}()$ 
5:      $x \leftarrow$  the longer of  $u$  and  $v$ 
6:     for  $i \leftarrow 1, \dots, \lfloor k/2 \rfloor$  do
7:        $CG1 \leftarrow \text{GenerateConstantGapWords}(\{0, 1, \dots, i-1\})$ 
8:        $CG2 \leftarrow \text{GenerateConstantGapWords}(\{i, i+1, \dots, k-1\})$ 
9:        $x' \leftarrow x$ 
10:      for all  $w \in CG1$  do
11:        for all  $w' \in CG2$  do
12:          for  $j \leftarrow 0, \dots, |x| - 1$  do
13:             $count0 \leftarrow 0; count1 \leftarrow 0$ 
14:            if  $x[j] = 0$  then
15:               $count0 \leftarrow count0 + 1$ 
16:               $x'[j] \leftarrow w[(count0 - 1) \bmod |w|]$ 
17:            else
18:               $count1 \leftarrow count1 + 1$ 
19:               $x'[j] \leftarrow w'[(count1 - 1) \bmod |w'|]$ 
20:            end if
21:          end for
22:          if  $\text{CriticalExponent}(x') < \eta$  then
23:             $Q.\text{enqueue}((u, uv)); Q.\text{enqueue}((vu, v))$ 
24:            goto 3
25:          end if
26:        end for
27:      end for
28:    end for
29:  end while
30: end procedure

```

inequality is clear, since the Hubert construction cannot create repetitions in x of higher exponent than the ones in u . For the second, note that up to permutation of the alphabet, the only choices for y and y' are $(01)^\omega$ and 2^ω (or vice-versa), and thus, any repetition in u of exponent $e \geq 2$ will produce a repetition in x of exponent $e/2$.

Proposition 6. *The word x_3 has critical exponent*

$$E(x_3) = 2 + \frac{\sqrt{2}}{2} \approx 2.7071.$$

Proof. As stated in Table 1, let $\alpha = \sqrt{2} - 1 = [0, \bar{2}]$ and let c_α be the characteristic Sturmian word with slope α . That is, the word c_α is the infinite word obtained as the limit of the sequence of standard words s_k defined by

$$s_0 = 0, \quad s_1 = s_01, \quad s_k = s_{k-1}^2 s_{k-2}, \quad k \geq 2$$

Let x_3 be the word over $\{0, 1, 2\}$ obtained from c_α by replacing the 0's in c_α by the periodic sequence $(01)^\omega$ and by replacing the 1's with 2's. We have $s_1 = 01$, $s_2 = 01010$, $s_3 = (01010)^2 01$, etc., and

$$c_\alpha = 01010010100101010010100101010010100101001010010 \dots$$

and

$$x_3 = 021201202102120210212012021201202102120 \dots$$

Let $(z')^e$ be a repetition of exponent $e \geq 2$ in x_3 ($e \in \mathbb{Q}$). Then, by applying the morphism that sends $\{0, 1\} \rightarrow 0$ and $2 \rightarrow 1$ to x_3 , we see that there is a corresponding repetition z^e of the same length in c_α . Suppose that z is primitive. By [20, Corollary 4.6] (originally due to Damanik and Lenz [4]) z is either a conjugate of one of the standard words s_k defined above or a conjugate of one of the semi-standard words

$$s_{k,1} = s_{k-1} s_{k-2}, \quad k \geq 2.$$

Suppose that z is a conjugate of a standard word. Note that $|s_k|_0$ is odd for every $k \geq 1$. Hence $|z|_0$ is odd. It follows that $z'z'$ cannot occur in x_3 (the

alternations of 0 and 1 will not “match up” in the two copies of z'), and so there is no repetition $(z')^e$ in x_3 .

Now suppose that z is a conjugate of a semi-standard word. Then $|z| = q_{k-2} + q_{k-1}$ for some $k \geq 2$. From [12, Theorem 4(i)], one finds that the longest factor of c_α with this period has length $2(q_{k-2} + q_{k-1}) + q_{k-1} - 2$. It follows that

$$\begin{aligned} e &\leq \frac{2(q_{k-2} + q_{k-1}) + q_{k-1} - 2}{q_{k-2} + q_{k-1}} \\ &= 2 + \frac{q_{k-1} - 2}{q_{k-2} + q_{k-1}} \\ &= 2 + \frac{1 - 2/q_{k-1}}{1 + q_{k-2}/q_{k-1}}. \end{aligned}$$

It is easy to show that $q_{k-2} = p_{k-1}$; using this fact together with (1) we find that $q_{k-2}/q_{k-1} > \sqrt{2} - 1 - 1/q_{k-1}^2$. Thus, we have

$$e < 2 + \frac{1 - 2/q_{k-1}}{\sqrt{2} - 1/q_{k-1}^2}.$$

The fraction on the right clearly tends to $\sqrt{2}/2$ as $k \rightarrow \infty$, and is increasing for $k \geq 1$, so the convergence is from below. Thus $e < 2 + \sqrt{2}/2$.

Indeed, for every $k \geq 2$, there are such factors with exponent

$$2 + \frac{1 - 2/q_{k-1}}{1 + q_{k-2}/q_{k-1}} \xrightarrow{k \rightarrow \infty} 2 + \frac{\sqrt{2}}{2},$$

where the convergence is from below. Note that $|s_{k,1}|_0$ is even for every $k \geq 2$ and so every such repetition z^e in c_α gives rise to a repetition $(z')^e$ in x_3 , since $|z|_0$ in this case is even.

Finally, suppose that z is not primitive. By [21, Proposition 4.6.12], the critical exponent of c_α is $3 + \sqrt{2}$, so clearly z cannot have exponent ≥ 3 . If z is a square we have

$$e < \frac{3 + \sqrt{2}}{2} < 2 + \frac{\sqrt{2}}{2}.$$

Thus $E(x_3) = 2 + \frac{\sqrt{2}}{2}$. □

Next we show that the exponent in the previous result is the least possible over 3 letters.

Proposition 7. *Every balanced word x over a 3-letter alphabet has critical exponent*

$$E(x) \geq 2 + \frac{\sqrt{2}}{2} \approx 2.7071.$$

Proof. Let $\alpha = [0, d_1, d_2, d_3, \dots]$ and let c_α be the Sturmian word with slope α . We substitute the 0's in c_α with periodic word y and the 1's with periodic word y' . We don't have much choice on 3 letters, so let $y = (01)^\omega$ and $y' = 2^\omega$. Let x be the resulting word and suppose that x is cubefree.

Consider the standard words $s_0 = 0$, $s_1 = 0^{d_1-1}1$ and $s_k = s_{k-1}^{d_k} s_{k-2}$ for $k \geq 2$. Let $k \geq 2$. We apply [20, Theorem 4.5(iii)] (which is a restatement of a result of [4]), which states that the index of the reversal of s_k (denoted s_k^R) in c_α is $d_{k+1} + 2$.

The first observation is that the number of 0's in s_k must be odd. If $|s_k|_0$ is even then $(s_k^R)^3$ occurs in c_α and produces a cube in x , even after replacement of the 0's by $(01)^\omega$.

Next we observe that d_{k+2} must be even. The number of 0's in s_{k+2} is $d_{k+2}|s_{k+1}|_0 + |s_k|_0$, and since $|s_k|_0$ and $|s_{k+1}|_0$ are odd, we must have d_{k+2} even. Now if $d_{k+2} \geq 4$, then the index of s_{k+1}^R is at least 6, and $(s_{k+1}^R)^6$ will produce a cube in x after replacement of the 0's with $(01)^\omega$.

So $d_k = 2$ for $k \geq 4$; i.e., $\alpha = [0, d_1, d_2, d_3, \bar{2}]$. As in Proposition 6, we apply [12, Theorem 4(i)], which states that c_α contains repetitions whose fractional roots are conjugates of the semi-standard words $s_{k-1}s_{k-2}$ (which have an even number of 0's) and whose exponents are of the form $2 + (q_k - 2)/(q_{k-1} + q_k)$. Because the fractional roots have an even number of 0's, after replacement of the 0's by $(01)^\omega$, the corresponding factor of the word x still has exponent $2 + (q_k - 2)/(q_{k-1} + q_k)$. A calculation similar to the one in the proof of Proposition 6 shows that as $k \rightarrow \infty$, this quantity again converges to $2 + \sqrt{2}/2$ (regardless of the first few d_i 's). The same argument (counting parities of 1's this time) applies if $y = 2^\omega$ and $y' = (01)^\omega$. \square

In order to establish the critical exponent for x_4 , we first need a technical

lemma concerning repetitions in the Fibonacci word³.

Lemma 8. *Let w be a factor of the Fibonacci word. Write $w = x^f$, where $f \in \mathbb{Q}$ and $|x|$ is the least period of w . Suppose that w has another representation $w = z^e$, where $e \in \mathbb{Q}$, z is primitive, and $|x| < |z|$. Then $e < 1 + \phi/2$.*

Proof. First, recall that the critical exponent of the Fibonacci word is $2 + \phi$ [15]. Let $g = \lceil f - 1 \rceil$ and write $x^f = x^g x'$ (so x' may possibly equal x). We must have $e < 2$, since otherwise we would have $|w| \geq |z| + |x|$, and so by the Fine–Wilf Theorem (see [1, Theorem 1.5.6]), w would have period $\gcd(|x|, |z|)$, contradicting the minimality of x ($|z|$ cannot be a multiple of $|x|$, since z is primitive). We thus write $w = z z'$, where z' is a non-empty prefix of z . For the same reason we have $e < 2$ we must also have $|z'| < |x|$. We now consider several cases.

Case 1: $|z'| < |x'|$. Then x' has z' as both a prefix and a suffix. It follows that x' has period $|x'| - |z'|$ and hence exponent

$$\frac{|x'|}{|x'| - |z'|} < 2 + \phi.$$

This implies that

$$|z'| < \left(\frac{1 + \phi}{2 + \phi}\right) |x'| < \left(\frac{1 + \phi}{2 + \phi}\right) \cdot \frac{1}{2} |w|.$$

Thus

$$\frac{|z'|}{|w|} < \left(\frac{1 + \phi}{2 + \phi}\right) \cdot \frac{1}{2},$$

and so

$$e = \frac{|w|}{|z|} < \frac{1}{1 - \left(\frac{1 + \phi}{2 + \phi}\right) \cdot \frac{1}{2}} \approx 1.5669 < 1 + \frac{\phi}{2}.$$

Case 2: $|z'| > |x'|$. Then we have $g > 1$.

³Arseny Shur has pointed out to us that a result of Pirillo [22] can be used in place of Lemma 8 in the argument that follows to obtain our desired result. Nevertheless, we give the proof of this lemma here, since it motivated us to prove the last results of this section.

Subcase 2a: $g = 2$ and $|x'| \geq |x|/2$. Then since $|z'| < |x|$, we have

$$\frac{|z'|}{|w|} < \frac{|x|}{|xx'|} \leq \frac{2}{5},$$

and so

$$e = \frac{|w|}{|z|} \leq \frac{1}{1 - \frac{2}{5}} = \frac{5}{3} < 1 + \frac{\phi}{2}.$$

Subcase 2b: $g = 2$ and $|x'| < |x|/2$. Then xx' has z' as both a prefix and a suffix. It follows that xx' has period $|xx'| - |z'|$ and hence exponent

$$\frac{|xx'|}{|xx'| - |z'|} < 2 + \phi.$$

This implies that

$$|z'| < \left(\frac{1 + \phi}{2 + \phi}\right) |xx'|.$$

Thus

$$\frac{|z'|}{|w|} < \frac{\left(\frac{1 + \phi}{2 + \phi}\right) |xx'|}{|xx'|} < \left(\frac{1 + \phi}{2 + \phi}\right) \cdot \frac{3}{5},$$

where we have used the fact that $|x'| < |x|/2$ to obtain the last inequality.

Therefore,

$$e = \frac{|w|}{|z|} < \frac{1}{1 - \left(\frac{1 + \phi}{2 + \phi}\right) \cdot \frac{3}{5}} \approx 1.7673 < 1 + \frac{\phi}{2}.$$

Subcase 2c: $g = 3$. Then $|xx'| < |w|/2$ and $|z'| < |xx'|$, so we can apply the argument of Case 1 with xx' in place of x' . \square

Proposition 9. *The word x_4 has critical exponent $E(x_4) = 1 + \frac{\phi}{2} \approx 1.8090$.*

Proof. Let c_θ be the Fibonacci word and let

$$x_4 = 021031201301203102130120130210310213012 \dots$$

be constructed from c_θ as described above with $y = (01)^\omega$ and $y' = (23)^\omega$.

Consider a repetition of the form $(z')^e$ in x_4 , where $e \in \mathbb{Q}$ and $e > 1$. If this repetition occurs at some position in x_4 , then there is a repetition z^e of the same length at the same position in c_θ . We will assume that $|(z')^{e-1}| \geq 3$:

for the case $|(z')^{e-1}| < 3$, one can verify with a little thought (or by computer) that $e < 1 + \phi/2$. We therefore have that z^{e-1} contains at least one 0 and one 1, and so $(z')^{e-1}$ contains at least one letter from $\{0, 1\}$ and at least one letter from $\{2, 3\}$. We claim that

$$|z'|_0 + |z'|_1 \equiv 0 \pmod{2} \text{ and } |z'|_2 + |z'|_3 \equiv 0 \pmod{2}. \quad (4)$$

To see this, note that the first occurrence of a letter from $\{0, 1\}$ in z' must match the first occurrence of such a letter in z'^{e-1} . Since occurrences of 0 and 1 alternate in z' , the total number of such letters in z' must therefore be even. The same argument applies to the letters from $\{2, 3\}$.

Currie and Saari [3, Theorem 3] characterized the minimal fractional roots of factors of Sturmian words. In the case of the Fibonacci word, this characterization states that the minimal fractional root of any factor of the Fibonacci word is a conjugate of a finite Fibonacci word.

Thus, if z is the minimal fractional root of the repetition z^e , then $|z|$ equals F_i (the i -th Fibonacci number) for some i , and furthermore, we have $|z|_0 = F_{i-1}$ and $|z|_1 = F_{i-2}$. Equation (4) thus implies that

$$F_{i-1} = |z|_0 = |z'|_0 + |z'|_1 \equiv 0 \pmod{2} \text{ and } F_{i-2} = |z|_1 = |z'|_2 + |z'|_3 \equiv 0 \pmod{2},$$

which is impossible.

Now suppose that z is primitive but is not the minimal fractional root of the repetition z^e . Then by Lemma 8, we have $e < 1 + \phi/2$, as required.

Finally, if z is not primitive, then there exists a word v such that $z = v^2$ or $z = v^3$. Thus either $z^e = v^{2e}$ or $z^e = v^{3e}$. Since $E(c_\theta) = 2 + \phi$, it follows that either $2e < 2 + \phi$ or $3e < 2 + \phi$. Both cases lead to the inequality $e < 1 + \phi/2$, as required.

Finally, note that since $E(c_\theta) = 2 + \phi$, there is a sequence of repetitions in c_θ with exponents $2 + r_j/s_j$, with r_j/s_j converging to ϕ from below. Consider such a repetition and write it as $V^{2+r_j/s_j} = Z^{1+r_j/(2s_j)}$. Then V is a conjugate of a finite Fibonacci word, and so for some i we have

$$2F_{i-1} = 2|V|_0 = |Z|_0 \equiv 0 \pmod{2} \text{ and } 2F_{i-2} = 2|V|_1 = |Z|_1 \equiv 0 \pmod{2},$$

so there is a corresponding repetition $(Z')^{1+r_j/(2s_j)}$ in x_4 . Thus $E(x_4) = 1 + \phi/2$. \square

A backtracking computer search (see Algorithm 1.) shows that there is no infinite balanced word over a 4-letter alphabet with critical exponent < 1.8088 . The minimal possible critical exponent among all infinite balanced words over 4 letters is thus between 1.8088 and $1 + \frac{\phi}{2} \approx 1.8090$.

Note that an important element of the previous proof was the result of Lemma 8 concerning non-minimal periods of the Fibonacci word. Little seems to be known on this topic, so we conclude this section with the following result.

Proposition 10. *Let w be a factor of a Sturmian word s and let $\{p_1 < p_2 < \dots < p_r\}$ be the set of periods of w . For $i \in \{1, \dots, r\}$ let P_i be the prefix of w of length p_i and for $i \geq 2$, write $P_i = P_{i-1}E_i$. Then for $i \in \{2, \dots, r\}$ the word E_i is a conjugate of either a standard word s_k or a semi-standard word $s_{k,t}$ for some k, t . Furthermore, the lengths of the E_i are non-increasing.*

Proof. We need the following fact (see [16]): If a word w has periods p and q with $q < p$, then the prefix of w of length $|w| - q$ has period $p - q$. We also use the result of [3, Theorem 3]: If w is a non-empty factor of s then the minimal fractional root of w is a conjugate of a standard or semi-standard word.

For $i \in \{2, \dots, r\}$, consider the prefix of w of length $|w| - p_{i-1}$. This prefix has period $p_i - p_{i-1}$ and corresponding fractional root E_i . We claim that this is its minimal fractional root. Suppose to the contrary that it has a smaller fractional root of length q . Then w has period $p_{i-1} + q$, which is less than p_i , which is a contradiction. By [3, Theorem 3], we have that E_i is a conjugate of a standard word s_k or a semi-standard word $s_{k,t}$.

To show that the lengths of the E_i are non-increasing, suppose to the contrary that $|E_{i+1}| > |E_i|$. Then the suffix of w of length $|w| - p_{i-1}$ has period $|E_i|$ and so does the suffix of w of length $|w| - p_i$. As w has period p_i , it therefore has period $p_i + |E_i| < p_{i+1}$, which is a contradiction. This completes the proof. \square

Corollary 11. *Let w be a factor of the Fibonacci word c_θ and let $\{p_1 < p_2 <$*

$\dots < p_r\}$ be the set of periods of w . For $i \in \{1, \dots, r\}$ let P_i be the prefix of w of length p_i and for $i \geq 2$, write $P_i = P_{i-1}E_i$. Then for $i \in \{2, \dots, r\}$ the word E_i is a conjugate of a finite Fibonacci word.

For more on words with multiple periods, the reader might consult [23, 24].

5. A computational approach

We can also attempt to establish the critical exponents of each x_k using the Walnut theorem-proving software [17] and the methods of Du, Mousavi, Schaeffer, and Shallit [6, 18]. The method is based on two things. The first is the fact that the Fibonacci word c_θ is a *Fibonacci-automatic sequence*: that is, the terms of c_θ can be computed by a finite automaton that takes a number n written in the Fibonacci numeration system as input and outputs the n -th term of c_θ . The second important element to the method of Mousavi et al. is that the addition relation $\{(x, y, z) \in \mathbb{N}^3 : x + y = z\}$ can be recognized by a finite automaton that reads its input in the Fibonacci numeration system.

To extend this method to an arbitrary characteristic word c_α , we first need to show that the terms of c_α can be computed by a finite automaton that takes the Ostrowski α -representation of n as input. This is immediate from Theorem 4 above. The second thing we need is the recognizability of the addition relation for the Ostrowski α -numeration system. Hieronymi and Terry [10] showed that addition is indeed recognizable in the case when α is a quadratic irrational. We now show that when α is a quadratic irrational, the Hubert construction applied to c_α results in a word that is automatic for the Ostrowski α -numeration system.

Theorem 12. *Let α be a quadratic irrational and let c_α be the characteristic Sturmian word with slope α . Let x be any word obtained by replacing the 0's in c_α with a periodic sequence y and replacing the 1's with a periodic sequence y' . Then x is Ostrowski α -automatic.*

Proof. Let p and p' be the periods of y and y' respectively. We need to show that there is a deterministic finite automaton with output that takes the Ostrowski

α -representation of n as input and outputs $x[n]$. To compute $x[n]$ it suffices to be able to compute $c_\alpha[n]$ as well as the number of 0's modulo p and 1's modulo p' in the length- n prefix of c_α . Let $b_j b_{j-1} \cdots b_0$ be the Ostrowski α -representation of n . By Theorem 4, there is an automaton that computes $c_\alpha[n]$ from $b_j b_{j-1} \cdots b_0$. By Lemma 2 and Theorem 3 the number of 0's in the length n prefix of c_α is

$$b_s(q_s - p_s) + b_{s-1}(q_{s-1} - p_{s-1}) + \cdots + b_0(q_0 - p_0).$$

Since α is a quadratic irrational, its continued fraction expansion is ultimately periodic. By Theorem 1, the sequences $(q_i)_{i \geq 0}$ and $(p_i)_{i \geq 0}$ both satisfy the same linear recurrence relation, and hence so does $(q_i - p_i)_{i \geq 0}$. It follows that $((q_i - p_i) \bmod p)_{i \geq 0}$ is ultimately periodic. Based on this ultimately periodic sequence, it is easy to construct an automaton to compute

$$(b_s(q_s - p_s) + b_{s-1}(q_{s-1} - p_{s-1}) + \cdots + b_0(q_0 - p_0)) \bmod p$$

given $b_s b_{s-1} \cdots b_0$ as input. A similar argument applies for computing the number of 1's in the length- n prefix of c_α . This completes the proof. \square

As an example, Figure 1 shows the Fibonacci-base automaton that generates the word x_4 (the labels on the states indicate the output symbol for that state).

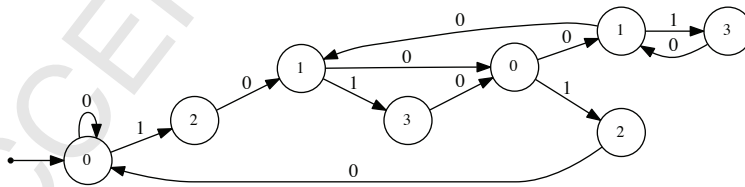


Figure 1: Fibonacci-base automaton for x_4

Let X denote the automaton in Figure 1 in the subsequent Walnut code. First, we compute the periods p such that a repetition with exponent at least $5/3$ and period p occurs in x_4 :

```
eval periods_of_high_powers "?msd_fib Ei (p>=1) &
  (Aj (3*j <= 2*p) => X[i+j]=X[i+j+p])";
```

The language accepted by the resulting automaton is $0^*1001000^*$; i.e., representations of numbers of the form $F_n + F_{n-3} = 2F_{n-1}$.

```
reg pows msd_fib "0*1001000*";
```

Next we compute pairs (n, p) such that x_4 has a factor of length $n + p$ with period p , and furthermore that factor cannot be extended to a longer factor of length $n + p + 1$ with the same period.

```
def maximal_reps "?msd_fib Ei (Aj (j<n) =>
  X[i+j]=X[i+j+p]) & (X[i+n]!=X[i+n+p])";
```

We now compute pairs (n, p) where p has to be of the form $0^*1001000^*$ and $n + p$ is the longest length of any factor having that period.

```
eval highest_powers "?msd_fib (p >= 1) & $pows(p) & $maximal_reps(n,p) &
  (Am $maximal_reps(m,p) => m <= n)";
```

The output of this last command is an automaton accepting pairs (n, p) having the form

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}^* \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}^* \begin{pmatrix} 0 \\ 0 \end{pmatrix} \left\{ \epsilon, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}.$$

So when $p = 2F_{i-1}$ we see that $n = F_i - 2$. Thus, the maximal repetitions of “large exponent” in x_4 have exponent of the form $1 + (F_i - 2)/(2F_{i-1})$. Using (1) and (2) we find that

$$\frac{F_i - 2}{2F_{i-1}} = \frac{F_i}{2F_{i-1}} - \frac{1}{F_{i-1}} < \frac{1}{2} \left(\phi + \frac{1}{F_{i-1}^2} \right) - \frac{1}{F_{i-1}} < \frac{\phi}{2}.$$

Hence these exponents converge to $1 + \phi/2$ from below, and we conclude that x_4 has critical exponent $1 + \phi/2$.

In principle, if the addition automaton of [10] were implemented in Walnut, it would be possible to carry out a similar proof for x_5 , x_6 , etc., provided the resulting computation was feasible.

6. Future work

The obvious open problem is to establish the claimed critical exponents for x_k for $5 \leq k \leq 10$, and more generally, show that for $k \geq 5$, the least critical exponent for an infinite balanced word over a k -letter alphabet is $(k-2)/(k-3)$. A first, possibly difficult, step would be to get a better understanding of the non-minimal periods of factors of Sturmian words.

- [1] J.-P. Allouche and J. Shallit, *Automatic Sequences*, Cambridge 2003.
- [2] J. Berstel and P. Séébold, “Sturmian words”, Chapter 2 of M. Lothaire, *Algebraic Combinatorics on Words*, Encyc. Math. Appl. Vol. 30, Cambridge, 2002, pp. 45–110.
- [3] J. Currie and K. Saari, Least periods of factors of infinite words, *Theor. Inform. Appl.* 43 (2009), 165–178.
- [4] D. Damanik and D. Lenz, The index of Sturmian sequences, *Europ. J. Combinatorics* 23 (2002), 23–29.
- [5] A. de Luca, Combinatorics of standard Sturmian words, in *Structures in Logic and Computer Science*, pp. 249–267, LNCS 1261, Springer, 1997.
- [6] C. F. Du, H. Mousavi, L. Schaeffer, and J. Shallit, Decision Algorithms for Fibonacci-Automatic Words, III: Enumeration and Abelian Properties, *Int. J. Found. Comput. Sci.* 27 (2016), 943–963.
- [7] H. Furstenberg, Poincaré recurrence and number theory, *Bull. Amer. Math. Soc.* 5 (1981), 211–234.
- [8] I. Goulden, A. Granville, B. Richmond, and J. Shallit, Natural exact covering systems and the reversion of the Möbius series. Preprint: <https://arxiv.org/abs/1711.04109> (2017).
- [9] R. L. Graham, Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta] : n = 1, 2, \dots\}$, *J. Combin. Theory (A)* 15 (1973), 354–358.

- [10] P. Hieronymi and A. Terry, *Ostrowski numeration systems, addition and finite automata*. To appear in Notre Dame J. Formal Logic. Preprint: <https://arxiv.org/abs/1407.7000> (2015).
- [11] P. Hubert, Suites équilibrées, Theoret. Comput. Sci. 242 (2000), 91–108.
- [12] J. Justin and G. Pirillo, Fractional powers in Sturmian words, Theoret. Comput. Sci. 255 (2001), 363–376.
- [13] H. W. Lenstra, Jr. and J. Shallit, Continued fractions and linear recurrences, Math. Comp. 61 (1993), 351–354.
- [14] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press (2002).
- [15] F. Mignosi and G. Pirillo, Repetitions in the Fibonacci infinite word, RAIRO Theoret. Inform. Appl. 26 (1992), 199–204.
- [16] F. Mignosi, A. Restivo, and S. Salemi, Periodicity and the golden ratio, Theoret. Comput. Sci. 204 (1998), 153–167.
- [17] H. Mousavi, Automatic theorem proving in Walnut. Preprint: <https://arxiv.org/abs/1603.06017> (2016).
- [18] H. Mousavi, L. Schaeffer, and J. Shallit, Decision Algorithms for Fibonacci-Automatic Words, I: Basic Results, RAIRO Theor. Inform. Appl. 50 (2016), 39–66.
- [19] A. Ostrowski, Bemerkungen zur Theorie der diophantischen Approximationen, Hamb. Abh. (in German) 1 (1922), 77–98.
- [20] J. Peltomäki, Characterization of repetitions in Sturmian words: a new proof. Inform. Process. Lett. 115 (2015), 886–891.
- [21] J. Peltomäki, *Privileged Words and Sturmian Words*, PhD thesis, TUCS Dissertations No. 214, August 2016. Available at <http://www.utupub.fi/handle/10024/124473>

- [22] G. Pirillo, *Fibonacci numbers and words*, Discrete Math. 173 (1997), 197–207.
- [23] R. Tijdeman and L. Zamboni, Fine and Wilf words for any periods. Indag. Math. 14 (2003), 135–147.
- [24] R. Tijdeman and L. Zamboni, Fine and Wilf words for any periods II. Theoret. Comput. Sci. 410 (2009) 3027–3034.
- [25] L. Vuillon, Balanced words, Bull. Belg. Math. Soc. Simon Stevin 10 (2003), 787–805.