ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus Mathématique

Jason Bell, Chris Schulz and Jeffrey Shallit

Consecutive Power Occurrences in Sturmian Words

Volume 362 (2024), p. 1273-1278

Online since: 5 November 2024

https://doi.org/10.5802/crmath.644

The Comptes Rendus. Mathématique *are a member of the Mersenne Center for open scientific publishing*
www.centre-mersenne.org — e-ISSN : 1778-3569

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mathématique
2024, Vol. 362, p. 1273-1278
https://doi.org/10.5802/crmath.644

Research article / *Article de recherche*
Theoretical computer science / *Informatique théorique*

Consecutive Power Occurrences in Sturmian Words

Occurrences consécutives de puissance dans les mots sturmiens

Jason Bell [a], **Chris Schulz** [a] **and Jeffrey Shallit** [*,b]

[a] University of Waterloo, Department of Pure Mathematics, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

E-mails: jpbell@uwaterloo.ca, chris.schulz@uwaterloo.ca, shallit@uwaterloo.ca

Abstract. We show that every Sturmian word has the property that the distance between consecutive ending positions of cubes occurring in the word is always bounded by 10 and this bound is optimal, extending a result of Rampersad, who proved that the bound 9 holds for the Fibonacci word. We then give a general result showing that for every $e \in [1,(5+\sqrt{5})/2)$ there is a natural number N, depending only on e, such that every Sturmian word has the property that the distance between consecutive ending positions of e-powers occurring in the word is uniformly bounded by N.

Résumé. Nous montrons que la distance entre deux positions finales consécutives de cubes apparaissant dans un mot sturmien est toujours inférieure ou égale à 10 et que cette valeur est optimale, étendant ainsi un résultat de Rampersad, qui a démontré que cette distance est majorée par 9 pour le mot de Fibonacci. Nous donnons ensuite un résultat général montrant que pour tout $e \in [1,(5+\sqrt{5})/2)$ il existe un entier naturel N, dépendant uniquement de e, tel que la distance entre deux positions finales consécutives de puissances e apparaissant dans un mot sturmien est uniformément majorée par N.

Keywords. Sturmian word, cube, periodicity, balanced word.

Mots-clés. Mot Sturmien, cube, périodicité, mot équilibré.

2020 Mathematics Subject Classification. 68R15.

Funding. The authors were supported by Discovery Grants from the National Sciences and Engineering Research Council of Canada.

Manuscript received 19 February 2024, revised 26 March 2024, accepted 21 April 2024.

1. Introduction

In this note we are concerned with the occurrences of powers in words. We say that a length-n word $w = w[0..n-1]$ has *period* p if $w[i] = w[i+p]$ for $i \in \{0,\ldots,n-p-1\}$. The smallest such period is called *the* period and is written $\operatorname{per}(w)$. The exponent of a length-n word w is defined to be $\exp(w) = n/\operatorname{per}(w)$. For example, $\exp(\texttt{entente}) = 7/3$. For a real number $e \geq 1$, we say a word w is an *e-power* if $\lceil e \operatorname{per}(w) \rceil$ is equal to $|w|$, the length of w. In particular, a finite word

---

[*] Corresponding author

being an e-power and having exponent e are not in general the same: the exponent is necessarily rational, while there exist e-powers for every real number $e \geq 1$. A 2-power is called a *square*, like the English word `murmur`. A 3-power is called a *cube*, like the English word `shshsh`.

Let $\alpha = (1+\sqrt{5})/2$ be the golden ratio and let $\mathbf{x}$ be any infinite word. In 1997 the third author of this paper conjectured, and Mignosi, Restivo, and Salemi later proved [10], that every sufficiently long prefix of $\mathbf{x}$ has a suffix of exponent $\geq \alpha^2$ if and only if $\mathbf{x}$ is ultimately periodic. This is an example of the "local periodicity implies global periodicity" phenomenon.

Furthermore, the constant α^2 is best possible. Let $\mathbf{f} = \mathbf{f}[0]\mathbf{f}[1]\mathbf{f}[2]\cdots = 01001010\cdots$ denote the infinite Fibonacci word [4], the fixed point of the morphism sending $0 \to 01$ and $1 \to 0$. Mignosi, Restivo, and Salemi [10] also showed that for all $\epsilon > 0$ every sufficiently long prefix of $\mathbf{f}$ has a suffix that is an $\alpha^2 - \epsilon$ power. For example, there is a square suffix of every prefix of length $n \geq 6$ in $\mathbf{f}$, and there is a $\frac{5}{2}$-power suffix of every prefix of length $n \geq 220$ in $\mathbf{f}$. An explicit version of this result was recently proved by the third author [13].

More generally, given an infinite word $\mathbf{x}$ and an exponent e, one can consider the list of all positive integers $p_1 < p_2 < \cdots$ marking the position where some e-power in $\mathbf{x}$ ends. For example, for the Fibonacci word we have

- $(010)^3$ ends at position 13
- $(01001)^3$ ends at position 22
- $(10010)^3$ ends at position 23
- $(010)^3$ ends at position 26
- $(010)^3$ ends at position 34

and so forth.

Recently Rampersad [12] obtained an explicit description of the positions where cubes end in the Fibonacci word $\mathbf{f}$.

In particular, he observed that if $p_1 < p_2$ are two consecutive positions where cubes end in $\mathbf{f}$, then $p_2 - p_1 \leq 9$, and more precisely $p_2 - p_1 \in \{1,2,3,4,8,9\}$. (Note that he phrased his discussion in terms of the lengths of *runs* of consecutive positions where there are *no cubes ending*, so the numbers he presented differ by 1 from ours.)

Since the Fibonacci word is the simplest of a much larger class of binary infinite words called Sturmian words, this naturally raises the question of whether a similar result holds for this much larger class. Sturmian words have several equivalent descriptions, as follows:

(a) Infinite words of the form $(\lfloor (n+1)\gamma + \beta \rfloor - \lfloor n\gamma + \beta \rfloor)_{n \geq 1}$ for real $0 < \gamma < 1$, $0 \leq \beta < 1$, where γ is irrational. Here γ is called the *slope* of the word and β is called the *intercept*.
(b) Infinite words having exactly $n+1$ distinct factors of length n for all $n \geq 0$. Here by a *factor* we mean a consecutive block lying inside the word.
(c) Infinite aperiodic binary balanced words, that is, words such that for all factors x, y of the same length, and all $a \in \{0,1\}$, the inequality $\big||x|_a - |y|_a\big| \leq 1$ holds, where $|z|_a$ denotes the number of occurrences of the letter a in z.

For more information about this class of words, see, for example, [9, Chapter 2].

In this note we generalize Rampersad's result to all Sturmian words.

Theorem 1. *Let* $\mathbf{x}$ *be a Sturmian word. Then the maximum gap between positions where cubes end in* $\mathbf{x}$ *is at most* 10*, and this bound is optimal.*

In fact, our proof shows that the gap between consecutive occurrences of cubes having period at most 5 is at most 10 for every Sturmian word; furthermore, the proof shows that the optimal bound 10 is achieved by the Sturmian characteristic word with slope $\sqrt{2}-1$ and intercept 0.

It is natural to ask whether a similar result to Theorem 1 holds for powers other than cubes. It is known that the Fibonacci word does not have β-powers for $\beta = (5+\sqrt{5})/2$, and so it is necessary to restrict to exponents that are less than β when considering the gap question over all Sturmian words. Once this condition is imposed, however, we are able to again prove a general result about the existence of uniform bounds on gaps of e-powers in Sturmian words for every $e < \beta$.

Theorem 2. *Let* $\mathbf{x}$ *be a Sturmian word and let* $e \in [1,(5+\sqrt{5})/2)$*. Then there are natural numbers* N *and* k*, depending only on* e*, such that the maximum gap between the ending positions of* e*-powers of period at most* k *in* $\mathbf{x}$ *is at most* N*.*

2. Proof of Theorem 1

We begin with a lemma. As it turns out, it suffices to consider words of bounded period.

Lemma 3. *Every balanced binary word of length* 17 *contains a cube of period at most* 5*, and the bound* 17 *is best possible.*

Proof. We enumerate all 594 binary balanced words of length 17 and check. For length 16, the word 0010100101001001 is balanced but contains no cube. □

Proof of Theorem 1. We claim that every balanced binary word of length 32 contains at least two consecutive cube occurrences, each of period at most 5. From Lemma 3, we know the prefix of length 17 contains a cube of period at most 5, and so does the suffix of length 17. Since in a word of length 32 these two cubes can overlap in at most two symbols, they must be distinct cube occurrences. Thus the maximum possible gap must appear in some balanced word of length 32. It then suffices to examine all balanced words of this length (there are 3650) and compute all gaps between consecutive ending positions of cubes of period 5 in all of them. The longest is 10.

To prove that 10 is optimal, we use the theorem-prover `Walnut`; see [11, 14] for more details about it. We now consider the Sturmian characteristic word with slope $\sqrt{2}-1$ and intercept 0. This word has an associated Ostrowski numeration system based on the Pell numbers, $1,2,5,12,29,70,169,\ldots$; this system is built-in to `Walnut` and is invoked by saying `msd_pell` (see [1–3]). We now use the following `Walnut` code:

```
reg odd0 msd_pell "(0|1|2)*(1|2)0(00)*":
def root2 "?msd_pell (n>=0) & $odd0(n+1)":
combine RT2 root2:
# make a DFAO RT2 for Sturmian sequence w/slope sqrt(2)-1

def end_in_cube "?msd_pell Ei,j (j>=1) & (i+3*j=n) &
   At (t<2*j) => RT2[i+t]=RT2[i+j+t]":
# is there a cube ending at position n

def two_consec "?msd_pell (i<j) & $end_in_cube(i) &
   $end_in_cube(j) & At (i<t & t<j) => ~$end_in_cube(t)":
# are i and j two consecutive cube-ending positions?

def two_consec_dist "?msd_pell (n>0) & Ei,j i+n=j & $two_consec(i,j)":
# it accepts only 1, 110 and 200, i.e., 1, 7 and 10 in Pell
```

The code first creates a deterministic finite-state automaton with output, named RT2, for the Sturmian sequence $\mathbf{x}$ with slope $\sqrt{2}-1$ and intercept 0. It next determines the set of natural numbers n for which there is a cube occurring in $\mathbf{x}$ that ends at position n. Finally, the code finds the set of gaps between consecutive ending positions of cubes and determines that the only gaps that occur are the natural numbers whose expansions in the Pell numeration system are $1, 110$, and 200. These are the numbers $1,7,10$ and so the bound of 10 is best possible. □

Remark 4. We note that the proof in fact shows the slightly stronger claim that the gap between consecutive occurrences of cubes of period at most 5 is at most 10 for every Sturmian word and that the bound 10 is achieved by the Sturmian characteristic word with slope $\sqrt{2}-1$ and intercept 0.

3. General exponents

We now give the proof of Theorem 2. We recall that a *partial function* f from a set X to a set Y is simply a map defined on a non-empty subset U of X that maps U into the set Y. In the case that the map is defined on all of X, the map is called a *total function*. We write $f : X \rightharpoonup Y$ for a partial function from X to Y.

Given two partial functions $f, g : \mathbb{N} \rightharpoonup \mathbb{R}_{\geq 0}$, we declare that $f \preceq g$ if $f(i) \leq g(i)$ for all i in the domain of both f and g. We note that $\preceq$ is not a partial order on partial functions, but it is, however, a partial order on total functions. We recall that for a right-infinite word $\mathbf{x}$ over a finite alphabet, we have a subword complexity function $p_{\mathbf{x}}$ whose value at n is the number of distinct length-n factors of $\mathbf{x}$. If w is a finite word, we have a partial function $p_w : \mathbb{N} \rightharpoonup \mathbb{N}$ that is defined on the nonnegative integers that are less than or equal to the length of w and whose value at i is the number of factors of w of length i for all i less than or equal to the length of w.

Recall that an infinite word $\mathbf{x}$ is *uniformly recurrent* if for each of its finite factors v, there is an integer $n(v)$ such that every block of length $n(v)$ in $\mathbf{x}$ contains a copy of v.

Lemma 5. *Let Σ be a finite alphabet, let $h : \mathbb{N} \to \mathbb{N}$ be a total function, and let S be a non-empty set of words over Σ. Suppose that every uniformly recurrent word $\mathbf{x}$ with $p_{\mathbf{x}} \preceq h$ has the property that $\mathbf{x}$ contains some word from S as a factor. Then there exists a finite subset S_0 of S and a natural number N, depending only on h and S_0, such that, whenever $\mathbf{x}$ is a right-infinite word with $p_{\mathbf{x}} \preceq h$, every factor of $\mathbf{x}$ of length N contains a factor from S_0.*

Proof. Let U_n denote the set of words u over Σ of length n such that $p_u \preceq h$ and such that u does not have an element of S as a factor. Then $\bigcup_{n\geq 0} U_n$ is closed under the process of taking factors (where we take U_0 to be the empty word). If there is some $N \geq 1$ such that U_N is empty, then letting S_0 be the set of words in S of length at most N, we see that whenever $\mathbf{x}$ is a right-infinite word with $p_{\mathbf{x}} \preceq h$, every factor of $\mathbf{x}$ of length N contains a factor from S_0.

Thus we may assume that each U_n is non-empty. By König's infinity lemma (see [8] or [7, Section 2.3.4.3]), there is a right-infinite word $\mathbf{x}$ with the property that all of its finite factors are in the union of the U_i. In particular, $p_{\mathbf{x}} \preceq h$ and $\mathbf{x}$ has no words from S as a factor. Then by Furstenberg's theorem [6], there is a uniformly recurrent right-infinite word $\mathbf{y}$ whose finite factors are all factors of $\mathbf{x}$, so $p_{\mathbf{y}} \preceq p_{\mathbf{x}} \preceq h$ and $\mathbf{y}$ has no occurrences of factors from S, a contradiction. □

Proof of Theorem 2. Let $h(n) = n+1$ for $n \geq 0$ and let P_e denote the set of words w over Σ that are e-powers. By work of Damanik and Lenz [5], every Sturmian word has an element from P_e as a factor. Since every uniformly recurrent word whose subword complexity function is $\preceq h$ is either Sturmian or periodic and since periodic words always contain elements from P_e as factors, Lemma 5 gives that there is a finite subset Q_e of P_e and some number $n = n(e)$ such that every length-n factor of a Sturmian word $\mathbf{x}$ has an occurrence of a word from Q_e. In particular, if we take k to be the supremum of periods in Q_e and $N = 2n$, we see that gaps between consecutive e-powers of period at most k are bounded by N. □

4. Other exponents

When one considers Theorem 2, a natural question that arises is whether one can bound the associated quantities k and N explicitly in terms of the exponent $e \in [1, (5+\sqrt{5})/2)$. This appears to be somewhat subtle in general, but using the same computational approach as we employed in establishing Theorem 1, we can obtain results for several other exponents.

For each exponent e we report the smallest n such that such that every balanced word of length n contains a factor that is an e-power, the largest period p of an e-power in those words, the largest gap g between ending positions of e-powers, and a quadratic irrational γ realizing this gap g.

Table 1. Data for some exponents

e	n	p	g	γ
$5/2$	9	3	6	$(5+\sqrt{5})/10$
$8/3$	15	5	9	$\sqrt{2}-1$
3	17	5	10	$\sqrt{2}-1$
$16/5$	30	8	17	$(25-\sqrt{5})/62$
$23/7$	50	13	27	$(59+\sqrt{5})/158$
$10/3$	69	18	37	$(217-\sqrt{5})/298$

It would be interesting to better understand the relationship between the largest gap g and the exponent e as e approaches $(5+\sqrt{5})/2$.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

References

[1] A. R. Baranwal, *Decision Algorithms for Ostrowski-automatic Sequences*, Memoir, University of Waterloo, School of Computer Science, 2020.

[2] A. R. Baranwal, L. Schaeffer and J. Shallit, "Ostrowski-automatic sequences: theory and applications", *Theor. Comput. Sci.* **858** (2021), pp. 122–142.

[3] A. R. Baranwal and J. Shallit, "Critical exponent of infinite balanced words via the Pell number system", in *Combinatorics on words*, Springer, 2019, pp. 80–92.

[4] J. Berstel, "Fibonacci words—a survey", in *The Book of L* (G. Rozenberg and A. Salomaa, eds.), Springer, 1986, pp. 13–27.

[5] D. Damanik and D. Lenz, "The index of Sturmian sequences", *Eur. J. Comb.* **23** (2002), no. 1, pp. 23–29.

[6] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, 1981, pp. xi+203.

[7] D. E. Knuth, *The art of computer programming. Vol. 1. Fundamental algorithms*, Third edition edition, Addison-Wesley Publishing Group, 1997, pp. xx+650.

[8] D. König, "Über eine Schlußweise aus dem Endlichen ins Unendliche", *Acta Litt. Sci. Szeged* **3** (1927), pp. 121–130.

[9] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, 2002, pp. xiv+504.

[10] F. Mignosi, A. Restivo and S. Salemi, "Periodicity and the golden ratio", *Theor. Comput. Sci.* **204** (1998), no. 1-2, pp. 153–167.

[11] H. Mousavi, "Automatic theorem proving in `Walnut`", 2016. https://arxiv.org/abs/1603.06017.

[12] N. Rampersad, "Prefixes of the Fibonacci word that end with a cube", *C. R. Math. Acad. Sci. Paris* **361** (2023), pp. 323–332.

[13] J. Shallit, "Prefixes of the Fibonacci word", 2023. https://arxiv.org/abs/2302.04640.

[14] J. Shallit, *The logical approach to automatic sequences – exploring combinatorics on words with `Walnut`*, Cambridge University Press, 2023, pp. xv+358.