PERIODS OF FACTORS OF THE FIBONACCI WORD

KALLE SAARI

Abstract. We show that if w is a factor of the infinite Fibonacci word, then the least period of w is a Fibonacci number.

1. Introduction

The Fibonacci word is arguably one of the most studied infinite sequence in combinatorics on words, see e.g., [17, 9, 6, 12, 18, 21, 16, 22, 19, 7, 15, 11, 13]. It is one of the simplest non-periodic infinite words, and certainly the simplest Sturmian word providing insight to the properties of all Sturmian words [14]. The Fibonacci word has been used to prove optimality of various results ranging from text algorithms to the periodicity of infinite words [4, 5, 15, 20], see also [2]. In the future, the finite Fibonacci words might have an equally important role in theoretical computer science as what Fibonacci numbers have in mathematics today.

The aim of this paper is to present, to our knowledge, a new property of the Fibonacci word. We show that if a word is a nonempty factor of the Fibonacci word, then its least period is a Fibonacci number. This is a tightening of a folklore property of the Fibonacci word stating that if a square is a factor of the Fibonacci word, then "root" of the square is a conjugate of a finite Fibonacci word [21], see also [19].

An outline of this paper follows. In the next section we present some definitions and a number of well-known properties relevant to our discussion. In Section 3 we present and prove the aforementioned property of the Fibonacci word. Section 4 concludes this paper with some comments on how the new property relates to some old ones and an idea for a generalization of this property to Sturmian words.

2. Definitions and Preliminary Results

In this section we present necessary definitions and properties of the Fibonacci word, following the terminology of [14]. For unexplained notions, we refer to [3].

Let us define a sequence of words $(f_n)_{n\geq -1}$ as follows:

$$f_{-1} = 1, \qquad f_0 = 0, \qquad \text{and} \qquad f_n = f_{n-1}f_{n-2},$$

for $n \geq 1$. The words f_n are referred to as the *finite Fibonacci words*. The limit $\mathbf{f} = \lim_{n\to\infty} f_n$ is called the (*infinite*) *Fibonacci word*.

Lemma 1 (de Luca [6])**.** *The finite Fibonacci words are primitive.*

We denote $F_n = |f_n|$, so that the numbers $F_{-1}, F_0, F_1, F_2, \ldots$ correspond to the Fibonacci numbers $1, 1, 2, 3 \ldots$. For notational purposes, we refer to a number F_{-2} in some of the lemmas that follow, and therefore we define $F_{-2} = 0$.

The author was supported by the Finnish Academy under grant 8206039.

Let $w = w_1 w_2 \cdots w_n$, with w_i letters. An integer $p \geq 1$ is *a period* of w if $i \equiv j \pmod{p}$ implies $w_i = w_j$ for all $i, j = 1, 2, \ldots n$. The least period of a word w is called *the period* of w, and denoted by $\mathrm{p}(w)$. The rational number $|w|/\mathrm{p}(w)$ is called the *order* of w, and denoted by $\mathrm{ord}(w)$. If $\mathrm{p}(w) = |w|$, then w is called *unbordered*; otherwise w is called *bordered*.

Let $\mathrm{C}\colon \{0,1\}^* \to \{0,1\}^*$ denote a cyclic permutation on words defined by

$$\mathrm{C}(\varepsilon) = \varepsilon, \qquad \mathrm{C}(ax) = xa,$$

where ε is the empty word and $a \in \{0,1\}$.

For $n \geq 1$, let g_n denote an auxiliary word

$$g_n = \mathrm{C}^{F_{n-1}}(f_n) = f_{n-2} f_{n-1}.$$

The following lemma is proved, e.g., in [1].

Lemma 2. *For $n \geq 2$, we have*

$$f_n f_{n-1} = f_{n-1} g_n, \qquad \textit{and} \qquad f_{n-1} f_n = f_n g_{n-1}.$$

Furthermore, for all $n \geq 1$, the words f_n and g_n differ only by the last two letters.

The following lemma is also proved in [14, Chapter 2].

Lemma 3 (Morse and Hedlund [17])**.** *For all $n \geq 0$, the Fibonacci word has exactly $n+1$ distinct factors of length n.*

It is well-known, and easy to see, that f_n^2 occurs in $\mathbf{f}$ for all $n \geq 0$. Since f_n is primitive and since there are $F_n + 1$ factors of length F_n, there exists precisely one factor of length F_n of $\mathbf{f}$ that is not a conjugate of f_n. This word is termed *singular*, and we denote it by c_n.

Lemma 4 (Wen and Wen [22])**.** *Singular words satisfy the recursive formula*

$$c_0 = 1, \qquad c_1 = 00, \qquad c_2 = 101, \qquad c_n = c_{n-2} c_{n-3} c_{n-2} \tag{1}$$

for $n \geq 3$.

Lemma 5 (Séébold [21])**.** *If a word u^2 is a factor of $\mathbf{f}$, then u is a conjugate of some finite Fibonacci word.*

We say that an infinite word is *kth-power-free* for some real number $k \geq 1$ if it does not have a factor of order greater than or equal to k. The previous lemma with Lemma 1 imply the following.

Lemma 6 (Karhumäki [12])**.** *The Fibonacci word is 4th-power-free.*

Actually much more can be said about the repetitions in $\mathbf{f}$, see [16], but the previous result is more suitable for our purposes.

The following three lemmas are proved in [14, Chapter 8].

Lemma 7 (Mignosi et al. [15])**.** *If a word w has two periods p and q with $q < p \leq |w|$, then the prefix of w of length $|w| - q$ has a period $p - q$.*

Lemma 8 (Mignosi et al. [15])**.** *If words uv and vw have a period p and $|v| \geq p$, then the word uvw has a period p.*

Lemma 9 (Fine and Wilf [10])**.** *Suppose that a word w has periods p and q. If*

$$|w| \geq p + q - \gcd(p,q),$$

then w has a period $\gcd(p,q)$.

3. Main Result

In this section we show that the least period of any factor of the Fibonacci word is a Fibonacci number. We start by proving this for factors whose length is a Fibonacci number, see Theorem 1.

Lemma 10. *If a word w is a bordered factor of* $\mathbf{f}$ *with $|w| = F_n$ for some $n \geq 0$, then either F_{n-2} or F_{n-1} is period of w.*

Proof. The claim is readily verified for $n = 0, 1, 2$, so we may suppose that $n \geq 3$. The word w is either a conjugate of f_n or it equals the singular word c_n. In the latter case, Equation (1) shows that w has a period F_{n-1}. Therefore we may assume that w is a conjugate of f_n, say

$$w = \mathrm{C}^i(f_n),$$

where $0 \leq i < F_n$. We have three cases to consider.

Suppose first that $0 \leq i \leq F_{n-1} - 2$. Since $n \geq 3$, Lemma 2 implies that

$$f_n^2 = f_{n-1}f_{n-2}f_{n-1}f_{n-2} = f_{n-1}f_{n-1}g_{n-2}f_{n-2}.$$

Notice that the word w occurs in f_n^2 at position $i \leq F_{n-1} - 2$. Since g_{n-2} and f_{n-2} differ only by the last two letters, and since f_{n-2} is a prefix of f_{n-1}, we see that w is a factor of f_{n-1}^3. Therefore w has a period F_{n-1}.

Suppose then that $F_{n-1} \leq i \leq F_n - 2$. Since $n \geq 3$, Lemma 2 implies that

$$\mathrm{C}^{F_{n-1}}(f_n^2) = f_{n-2}f_{n-1}f_{n-2}f_{n-1} = f_{n-2}f_{n-2}g_{n-1}f_{n-1}.$$

Notice that the word w occurs in $\mathrm{C}^{F_{n-1}}(f_n^2)$ at position $i - F_{n-1} \leq F_{n-2} - 2$. Since g_{n-1} and f_{n-1} differ only by the last two letters, and clearly f_{n-1} is a prefix of f_{n-2}^2, we see that w is a factor of f_{n-2}^3. Therefore w has a period F_{n-2}.

Suppose finally that $i = F_{n-1} - 1$ or $i = F_n - 1$. For all other values of i we just showed that $\mathrm{C}^i(f_n)$ is bordered. Since f_n is primitive, and every primitive word over two letters has at least two unbordered conjugates (the lexicographically smallest and largest conjugates), it follows that the words $\mathrm{C}^i(f_n)$ for $i = F_{n-1} - 1$ and $i = F_n - 1$ are unbordered. This completes the proof. ∎

The following lemma can be proved by a simple induction.

Lemma 11. *For all $n \geq 3$, we have*

$$F_n \geq 4F_{n-3} \qquad \text{and} \qquad F_{n-2} < \tfrac{1}{2}F_n < F_{n-1}$$

Now we are ready to prove a theorem that is a part of our main result, Theorem 2.

Theorem 1. *If a word w is a factor of* $\mathbf{f}$ *with $|w| = F_n$ for some $n \geq 0$, then the period of w equals F_{n-2}, F_{n-1}, or F_n.*

Proof. The claim is readily verified for $n = 0, 1, 2$, so we may suppose that $n \geq 3$. We divide the proof into several cases.

Suppose first that $\mathrm{p}(w) < F_{n-2}$. Then $2\,\mathrm{p}(w) < |w|$, and so the word w has a prefix of the form u^2 with $|u| = \mathrm{p}(w)$. Lemma 5 implies that the word u is a conjugate of f_k for some $k \leq n-3$, and therefore $p(w) \leq F_{n-3}$. But now Lemma 11 implies that

$$\mathrm{ord}(w) = \frac{|w|}{\mathrm{p}(w)} \geq \frac{F_n}{F_{n-3}} \geq 4,$$

contradicting Lemma 6.

If $\mathrm{p}(w) = F_{n-2}$, the claim holds.

Suppose then that $F_{n-2} < \mathrm{p}(w) < \frac{1}{2}F_n$. Since $|w| = F_n$, it follows that w has a prefix of the form u^2 with $|u| = \mathrm{p}(w)$. By Lemma 5, we have that u is a conjugate of some finite Fibonacci word. But since $F_{n-2} < \mathrm{p}(w) < F_{n-1}$ by Lemma 11, we see that $|u| = \mathrm{p}(w)$ is not a Fibonacci number, a contradiction.

Suppose now that $\frac{1}{2}F_n \le \mathrm{p}(w) < F_{n-1}$. It follows from Lemma 10 that F_{n-1} is a period of w. By applying Lemma 7, we see that the prefix of w of length $|w| - \mathrm{p}(w)$, denote it by u, has a period $F_{n-1} - \mathrm{p}(w)$. Consequently,

$$\mathrm{ord}(u) = \frac{|w| - \mathrm{p}(w)}{F_{n-1} - \mathrm{p}(w)} = \frac{F_n - \mathrm{p}(w)}{F_{n-1} - \mathrm{p}(w)}. \tag{2}$$

Notice that the function $(F_n - x)/(F_{n-1} - x)$ is increasing for $x < F_{n-1}$. Denoting $x = \frac{1}{2}F_n$, we have

$$\frac{F_n - x}{F_{n-1} - x} = \frac{F_n}{2F_{n-1} - F_n} = \frac{F_n}{F_{n-3}} \ge 4,$$

and therefore this inequality holds for all $\frac{1}{2}F_n \le x < F_{n-1}$. In particular it holds for $x = \mathrm{p}(w)$, and hence the Equation (2) gives $\mathrm{ord}(u) \ge 4$. Since u is a factor of $\mathbf{f}$, this contradicts Lemma 6.

If $\mathrm{p}(w) = F_{n-1}$, the claim holds.

Suppose finally that $\mathrm{p}(w) > F_{n-1}$. Then Lemma 10 implies that w is unbordered, and therefore $\mathrm{p}(w) = F_n$. The proof is complete. □

Lemma 12. *If a word w is a factor of $\mathbf{f}$ with $F_{n-1} \le |w| < F_n$ and $n \ge 0$, then w has a period that is a Fibonacci number.*

Proof. The claim is readily verified for $n = 0, 1, 2$, so we may suppose that $n \ge 3$. There exists a factor of $\mathbf{f}$, say z, such that w is a proper prefix of z and $|z| = F_n$.

If z is bordered, then by Theorem 1 either $\mathrm{p}(z) = F_{n-2}$ or $\mathrm{p}(z) = F_{n-1}$, and consequently w has a period F_{n-2} or F_{n-1}.

Suppose then that z is unbordered. By the proof of Lemma 10, we know that either

$$z = \mathrm{C}^{F_{n-1}-1}(f_n), \qquad \text{or} \qquad z = \mathrm{C}^{F_n-1}(f_n). \tag{3}$$

Write $z = ua$, where $u \in \{0,1\}^*$ and $a \in \{0,1\}$. Then we have either

$$au = \mathrm{C}^{F_{n-1}-2}(f_n), \qquad \text{or} \qquad au = \mathrm{C}^{F_n-2}(f_n).$$

Since $n \ge 3$, these words are distinct from the words in (3). Hence the word ua is a bordered factor of $\mathbf{f}$ of length F_n, and therefore either $\mathrm{p}(au) = F_{n-2}$ or $\mathrm{p}(au) = F_{n-1}$ by Theorem 1. Hence u has a period F_{n-2} or F_{n-1}, and so does the word w because w is a prefix of u. This completes the proof. □

As an immediate corollary, we can determine the unbounded factors of the Fibonacci word.

Corollary 1. *If a word w is an unbounded factor of $\mathbf{f}$, then w is a conjugate of f_n for some $n \ge -1$.*

We now need only one more lemma before proving the main theorem.

Lemma 13. *Suppose that k and m are integers with $2 \le k < m$. Then*

$$4\gcd(F_k, F_m) \le F_m.$$

Proof. The integers F_k, F_{k+1}, F_{k+2} are pairwise coprime. Therefore the claim holds if $m \leq k+2$. On the other hand, if $m > k+2$ the claim follows from Lemma 11 because $F_m \geq F_{k+3} \geq 4F_k$. □

The following result is the main theorem of this paper.

Theorem 2. *If a word w is a nonempty factor of the Fibonacci word, then the period of w is a Fibonacci number.*

Proof. We prove the claim by induction on $|w|$. The claim is readily verified for $|w| = 1, 2, \ldots, 8$. Suppose now that $|w| \geq 9$.

Lemma 12 implies that there exists an integer $n \geq 0$ such that the word w has a period F_n. If $|w| = F_n$, the claim follows from Theorem 1. Therefore we may suppose that $|w| > F_n$, and thus we have $\mathrm{p}(w) < |w|$. Let us write $w = axb$, where a and b are letters.

Claim A. We have either

$$\mathrm{p}(w) = \mathrm{p}(ax) \qquad \text{or} \qquad \mathrm{p}(w) = \mathrm{p}(xb). \tag{4}$$

To derive a contradiction, let us suppose that the claim does not hold. Then $\mathrm{p}(ax) < \mathrm{p}(w)$ and $\mathrm{p}(xb) < \mathrm{p}(w)$. Now we have $|x| \geq \mathrm{p}(ax)$. Indeed, if $|x| < \mathrm{p}(ax)$, then $\mathrm{p}(ax) = |ax|$, and so $\mathrm{p}(w) = |w|$, a contradiction. Consequently, we have $\mathrm{p}(ax) \neq \mathrm{p}(xb)$. Namely, if $\mathrm{p}(ax) = \mathrm{p}(xb)$, then Lemma 8 implies that $\mathrm{p}(w) = \mathrm{p}(ax)$, a contradiction. Therefore the induction assumption implies that there exist two distinct integers k and m such that $\mathrm{p}(ax) = F_k$ and $\mathrm{p}(xb) = F_m$. Since $|ax| = |xb| \geq 8$ and $\mathbf{f}$ is 4th-power-free, it follows that $m, k \geq 2$. Without loss of generality, we may assume that $k < m$. Then, since $F_k < F_m < F_n$, we get

$$|x| = |w| - 2 \geq F_n - 1 \geq F_m + F_k - 1 \geq F_m + F_k - \gcd(F_k, F_m).$$

Observe that both F_k and F_m are periods of x. Hence Lemma 9 implies that x has a period $\gcd(F_k, F_m)$. Therefore Lemma 13 implies

$$\mathrm{ord}(x) = \frac{|x|}{\mathrm{p}(x)} \geq \frac{F_m + F_k - 1}{\gcd(F_k, F_m)} \geq \frac{F_m}{\gcd(F_k, F_m)} \geq 4.$$

Hence the word x is factor of $\mathbf{f}$ with $\mathrm{ord}(x) \geq 4$, contradicting Lemma 6. Claim A is thus proved.

Now we can finish the proof of Theorem 2. By the induction assumption, both $\mathrm{p}(ax)$ and $\mathrm{p}(xb)$ are Fibonacci numbers. Thus the period of w is also a Fibonacci number by Claim A. The proof is now complete. □

4. Discussion

Theorem 2 is a fundamental property of the Fibonacci word: It establishes a nontrivial connection between the Fibonacci word and Fibonacci numbers. Hence the author was expecting to find it, or some other property directly implying it, in the literature on the Fibonacci word. The properties most similar to Theorem 2 the author was able to find are the following: Pirillo [19] shows that if a word of the form u^r with $r > (2+\varphi)/2$, where φ is the golden ratio, is a factor of $\mathbf{f}$, then u is a conjugate of a finite Fibonacci word. In [6], de Luca and De Luca show that a word w is a factor of a Sturmian word if and only if its fractional root is a conjugate of a standard word. However, though these results come close to ours, they do not seem to directly imply Theorem 2.

A conjecture of a generalization to all Sturmian words follows. Let t denote a Sturmian word with slope α, and let the continued fraction expansion of α be

$$\alpha = [\,0, 1 + d_1, d_2, d_3, \dots\,].$$

Denote

$$q_{-1} = q_0 = 1 \qquad \text{and} \qquad q_n = d_n q_{n-1} + q_n$$

for $n \geq 1$. Finally, define

$$\Pi(\alpha) := \bigcup_{n\geq 0} \{\, iq_n + q_{n-1} : \; i = 0, 1, \dots, d_n \,\}.$$

We conjecture that if a word w is a nonempty factor of t, then the least period of w is in the set $\Pi(\alpha)$. A proof of this conjecture might follow the lines of the proof of Theorem 2, the main problem would be to show the following: If a word $w = axb$ is an unbordered finite Sturmian word, then either $\mathrm{p}(w) = \mathrm{p}(ax)$ or $\mathrm{p}(w) = \mathrm{p}(xb)$.

References

[1] Berstel, J. On the index of Sturmian words. *Jewels are forever*, 287–294, *Springer, Berlin,* 1999.

[2] Cassaigne, J. Is the Fibonacci word always optimal? Talk given in the *Workshop on Fibonacci Words*, Turku, September 27 - 30, 2006.

[3] Choffrut, C.; Karhumäki, J. Combinatorics of words. *Handbook of formal languages, Vol. 1*, 329–438, *Springer, Berlin,* 1997.

[4] Crochemore, M. An Optimal Algorithm for Computing the Repetitions in a Word. *Inf. Process. Lett.* 12 (1981), no. 5, 244–250.

[5] Crochemore, M.; Rytter, W. Squares, cubes, and time-space efficient string searching. *Algorithmica* 13 (1995), no. 5, 405–425.

[6] de Luca, A. A combinatorial property of the Fibonacci words. *Inform. Process. Lett.* 12 (1981), no. 4, 193–195.

[7] de Luca, A. A Division Property of the Fibonacci Word. *Inf. Process. Lett.* 54 (1995), no. 6, 307–312.

[8] de Luca, A.; De Luca, A. Some characterizations of finite Sturmian words. *Theoret. Comput. Sci.* 356 (2006), no. 1-2, 118–125.

[9] Duval, J.-P. Contribution à la combinatoire du monoï de libre, Thèse de Doctorat d'Etat, Université de Rouen. 1980.

[10] Fine, N. J.; Wilf, H. S. Uniqueness theorems for periodic functions, *Proc. Amer. Math. Soc.* **16** 1965, 109–114.

[11] Fraenkel, Aviezri S.; Simpson, Jamie The exact number of squares in Fibonacci words. WORDS (Rouen, 1997). *Theoret. Comput. Sci.* 218 (1999), no. 1, 95–106.

[12] Karhumäki, J. On cube-free ω-words generated by binary morphisms. *Discrete Applied Mathematics* 5, 1983, 279–297.

[13] Kolpakov, Roman; Kucherov, Gregory On maximal repetitions in words. *J. Discrete Algorithms (Oxf.)* 1 (2000), no. 1, 159–186.

[14] Lothaire, M. Algebraic Combinatorics on Words, *Cambridge University Press*, 2002.

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

[16] Mignosi, F.; Pirillo, G. Repetitions in the Fibonacci infinite word. *RAIRO Inform. Théor. Appl.* 26 (1992), no. 3, 199–204

[17] Morse, M.; Hedlund, G. A. Symbolic Dynamics II. Sturmian Trajectories, *Amer. J. Math.* 60 (1940), 1–42.

[18] Pansiot, J.-J. Mots Infinis de Fibonacci et Morphismes Itérés. *RAIRO Inform. Théor. Appl.* 17 (1983), 131–135.

[19] Pirillo, G. Fibonacci numbers and words. *Discrete Math.* 173 (1997), no. 1-3, 197–207.

[20] Saari, K. Everywhere α-Repetitive Sequences and Sturmian Words. To appear in the Proceedings of CSR 2007.

[21] Séébold, P. Propriétés combinatoires des mots infinis engendrés par certains morphismes, *Thèse de doctorat*, Rapp. Tec. L.I.T.P., 85-14, 1985.

[22] Wen, Z.-X.; Wen, Z.-Y. Some properties of the singular words of the Fibonacci word, *European J. Combin.* 15, 1994, 587–598.

Department of Mathematics and Turku Centre for Computer Science, University of Turku, 2001 4 Turku, Finland

E-mail address: kasaar@utu.fi