A NEW CHARACTERIZATION OF THE FIBONACCI WORD

V. BERTHÉ, S. BRLEK, AND P. CHOQUETTE

Abstract. Using the run-length encoding Δ, the Fibonacci infinite word F possesses a natural representation obtained from the iterations of Δ. This representation is the eventually periodic word $112(13)^\omega$ which can be extended to the shifted Fibonacci words.

1. Introduction

Let $\varphi : \{1,2\}^* \longrightarrow \{1,2\}^*$ be the morphism defined by

$$\varphi(1) = 12 \quad ; \quad \varphi(2) = 1,$$

and let $F_n = \varphi^n(1)$ be the n-th iterate, also called the n-th Fibonacci word. We then have

$$\begin{aligned} F_0 &= 1, \\ F_1 &= 12, \\ F_2 &= 121, \\ F_3 &= 12112, \\ F_4 &= 12112121, \end{aligned}$$

and the infinite Fibonacci word F is obtained as the fixed point of φ, that is

$$F = \lim_{n\to\infty} F_n = \varphi^\omega(1) = 1211212112112121121211211212112112\cdots.$$

The combinatorial and arithmetic properties of F have been widely studied and it has a dominant role in the theory of Sturmian words.

The *run-length encoding* Δ is used in many applications as a method for compressing data. For instance, the first step in the algorithm used for compressing the data transmitted by Fax machines consists of a run-length encoding of each line of pixels. It also has been used for the enumeration of factors in the Thue-Morse sequence [2]. It is defined as follows. Let $\Sigma = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be a finite alphabet. Then every word $w \in \Sigma^*$ can be uniquely written as a product of factors as follows

$$w = \alpha_{m_1}^{e_1}\alpha_{m_2}^{e_2}\alpha_{m_3}^{e_3}\cdots$$

where $\alpha_{m_j} \in \Sigma$ and the exponents $e_j \geq 0$. Hence the coding is realized by a function

$$(\Delta, \tau) : \Sigma^* \longrightarrow \mathbb{N}^* \times \Sigma^*$$

where the first component is the function $\Delta : \Sigma^* \longrightarrow \mathbb{N}^*$, defined by

$$\Delta(w) = e_1e_2e_3\cdots = \prod_{j\geq 0} e_j,$$

with the support of NSERC (Canada).

and the second component is the function $\tau : \Sigma^* \longrightarrow \Sigma^*$ induced by the congruence $\equiv$ defined by

$$\alpha^2 \equiv \alpha,\ \forall \alpha \in \Sigma.$$

Note that the alphabet Σ may be identified with a subset of $\mathbb{N}$ and we shall denote $\mathbf{k} = \{1, 2, \ldots, k\} \subset \mathbb{N}$ for a fixed integer k. The operator Δ can be iterated, provided the process is stopped when the resulting word has length 1, and can also be extended to infinite words. For instance we have

$$\begin{array}{rcl} \Delta^0(F) & = & 12112121121121211212\cdots, \\ \Delta^1(F) & = & 1121112121112111\cdots, \\ \Delta^2(F) & = & 213111313\cdots, \\ \Delta^3(F) & = & 1113111\cdots, \\ \Delta^4(F) & = & 313\cdots. \end{array}$$

It is not difficult to see that the process can be reversed: $\Delta^i(F)$ may be retrieved from $\Delta^{i+1}(F)$ with the knowledge of $\tau(\Delta^i(F))$. It turns out that in the case of the Fibonacci word F, not only the alphabet is bounded but also $\tau(\Delta^i(F))$ is eventually periodic. Therefore F is completely determined by the characteristic sequence

$$\Phi(F) = (\Delta^i(F)[0])_{i=0..\infty} = 112(13)^\omega.$$

We also show that the shifted sequences of F also share this property.

2. Definitions and notation

A *word* over a finite *alphabet* of *letters* Σ is a finite sequence of letters

$$w : [0, n-1] \longrightarrow \Sigma\ , \quad n \in \mathbb{N},$$

of length n, and $w[i]$ or w_i denotes its letter of index i. The set of n-length words over Σ is denoted Σ^n. By convention the *empty* word is denoted ε and its length is 0. The free monoid generated by Σ is defined by $\Sigma^* = \bigcup_{n\geq 0} \Sigma^n$. The set of right infinite words is denoted by Σ^ω and $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$. Adopting a consistent notation for sequences of integers, $\mathbb{N}^* = \bigcup_{n\geq 0} \mathbb{N}^n$ is the set of finite sequences and $\mathbb{N}^\omega$ is those of infinite ones. Given a word $w \in \Sigma^*$, a *factor* f of w is a word $f \in \Sigma^*$ satisfying

$$\exists x, y \in \Sigma^*, w = xfy.$$

If $x = \varepsilon$ (resp. $y = \varepsilon$) then f is called *prefix* (resp. *suffix*). The set of all factors of w, called the *language* of w, is denoted by $L(w)$, and those of length n is $L_n(w) = L(w) \cap \Sigma^n$. Finally $\mathrm{Pref}(w), \mathrm{Suff}(w)$ denote respectively the set of all prefixes and suffixes of w. The length of a word w is $|w|$, and the number of occurrences of a factor $f \in \Sigma^*$ is $|w|_f$. A *block* of length k is a factor of the particular form $f = \alpha^k$, with $\alpha \in \Sigma$. If $w = pu$, and $|w| = n, |p| = k$, then $p^{-1}w = w[k]\cdots w[n-1] = u$ is the word obtained by erasing p. As a special case, when $|p| = 1$ we obtain the *shift* function defined by $s(w) = w_1 \cdots w_{n-1}$. Clearly the shift extends to right infinite words.

The *reversal* (or mirror image) $\widetilde{u}$ of $u = u_0u_1\cdots u_{n-1} \in \Sigma^n$ is the unique word satisfying

$$\widetilde{u_i} = u_{n-1-i},\ \forall\, i, 0 \leq i \leq n-1.$$

A *palindrome* is a word p such that $p = \widetilde{p}$, and for a language L, $\mathrm{Pal}(L)$ denotes the set of its palindromic finite factors. Over any finite alphabet $\Sigma = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$, there is a usual length preserving morphism, defined for every permutation $\rho : \Sigma \longrightarrow \Sigma$ of the letters, which extends to words by composition

$$[0, n-1] \xrightarrow{u} \Sigma \xrightarrow{\rho} \Sigma \ ,$$

defined by $\rho u = \rho u_0 \ \rho u_1 \ \rho u_2 \cdots \rho u_{n-1}$.

This definition extends as usual to infinite words $\mathbb{N} \longrightarrow \Sigma$. The occurrences of factors play an important role and an infinite word w is recurrent if it satisfies the condition

$$u \in L(w) \implies |w|_u = \infty \ .$$

Clearly, every periodic word is recurrent, and there exist recurrent but non-periodic words, the Thue-Morse word M being one of these [15]. Finally, two words u and v are *conjugate* when there are words x, y such that $u = xy$ and $v = yx$. The conjugacy class of a word u is denoted by $[u]$, and the length is invariant under conjugacy so that it makes sense to define $|[u]| = |u|$.

Checking that Δ commutes with the mirror image, is stable under permutation and preserves palindromicity is straightforward:

Proposition 1. *The operator Δ satisfies the conditions*

(a) $\Delta(\widetilde{u}) = \widetilde{\Delta(u)}$, *for all* $u \in \Sigma^*$;
(b) $\Delta(\rho u) = \Delta(u)$, *for all* $u \in \Sigma^*$ *and every permutation* $\rho : \Sigma \longrightarrow \Sigma$;
(c) $p \in \mathrm{Pal}(\Sigma^*) \implies \Delta(p) \in \mathrm{Pal}(\mathbb{N}^*)$.

Note that Δ is not distributive on concatenation in general. Nevertheless

(1) $$\Delta(uv) = \Delta(u) \cdot \Delta(v) \iff \widetilde{u}[0] \neq v[0],$$

that is to say if and only if the last letter of u differs from the first letter of v. This property can be extended to iterations and yields the following useful lemma.

Lemma 2 (Glueing Lemma)**.** *Let $u \cdot v \in \mathrm{Pref}(F_n)$ for some n. If there exists an index m such that, for all $i, 0 \leq i \leq m$, the last letter of $\Delta^i(u)$ differs from the first letter of $\Delta^i(v)$, and $\Delta^i(u) \neq 1$, $\Delta^i(v) \neq 1$, then*

(i) $\Phi(uv) = \Phi(u)[0..m] \cdot \Phi \circ \Delta^{m+1}(uv)$;
(ii) $\Delta^i(uv) = \Delta^i(u)\Delta^i(v)$.

The glueing operation is denoted by $\oplus$:

$$\Phi(u) \oplus_m \Phi(v) = \Phi(u)[0..m] \cdot \Phi \circ \Delta^{m+1}(uv),$$

and observe that the glueing lemma may also be generalized (by associativity) to the concatenation of more than two words.

Example. Let $u = 1211$ and $v = 21211$. Iterating Δ on uv yields

$$\begin{array}{rrcl} \Delta^0(uv) = & \mathbf{1}211 & \cdot & 21211 \\ \Delta^1(uv) = & \mathbf{1}12 & \cdot & 1112 \\ \Delta^2(uv) = & \mathbf{2}1 & \cdot & 31 \\ \Delta^3(uv) = & \mathit{1}1 & \cdot & 11 \\ \Delta^4(uv) = & \mathit{4} & & \end{array}$$

In this case we have $m = 2$ and $\Phi(1211 \cdot 21211) = \mathbf{112} \cdot \Phi(1111) = 112 \cdot 14$.

The glueing Lemma 2 admits an extension to infinite words: let $u \in \Delta^{(*)}(\mathrm{Pref}(F))$ and $v \in \Delta^{(*)}(\mathrm{Suff}(F))$. If there exists an index m such that the last letter of $\Delta^m(u)$ differs from the first letter of $\Delta^m(v)$, then

$$\Phi(u) \oplus_m \Phi(v) = \Phi(u)[0..m] \cdot \Phi \circ \Delta^{m+1}(uv).$$

The glueing lemma is fundamental for establishing the claim results and most of the proofs are based on induction, use canonical factorizations of the Fibonacci finite words F_n, where the glueing lemma applies.

3. Results

The Fibonacci words F_n satisfy many characteristic properties and we state without proof the ones that will be used hereafter:

Proposition 3. *For all $n \geq 0$ the following properties hold:*

(a) $F_{n+3} = F_{n+2} \cdot F_{n+1}$, *and* $F_{n+4} = F_{n+2} \cdot F_{n+1} \cdot F_{n+2}$;
(b) $2 \cdot F_{2n+2} \cdot 1^{-1}$ *and* $1 \cdot F_{2n+1} \cdot 2^{-1}$ *are palindromic factors.*
(c) *The set $\{F_{n+1}, F_n\}$ is an ω-code, that is, every word in $\{1,2\}^\omega$ admits at most one $\{F_n, F_{n-1}\}$-factorization.*

In the finite case we have the following property.

Proposition 4. *The sequence of finite Fibonacci words satisfies for all $n \geq 0$ the conditions*

(i) $\Phi(2 \cdot F_{2n+2} \cdot 1^{-1}) = 2(13)^{n+1}$;
(ii) $\Phi(1 \cdot F_{2n+1} \cdot 2^{-1}) = 12(13)^n$.

Proof. We proceed by induction. A straightforward verification establishes the base of the induction for $n = 0, 1, 2, 3$. Assume now the conditions hold until $n-1$. In order to establish (i) we use the recurrence relations of Proposition 3 for $2n+2$ and obtain

$$\begin{aligned} \Phi(2 \cdot F_{2n+2} \cdot 1^{-1}) &= \Phi(2 \cdot (F_{2n} F_{2n-1} F_{2n}) \cdot 1^{-1}) \\ &= \Phi(2 \cdot (F_{2n} \cdot 1^{-1} 1 \cdot F_{2n-1} \cdot 2^{-1} 2 \cdot F_{2n}) \cdot 1^{-1}) \end{aligned} \tag{2}$$

Recall that Δ preserves palindromicity (Proposition 1), and that $2 \cdot F_{2n+2} \cdot 1^{-1}$ is palindromic (Proposition 3). Therefore, for every $m \leq 2n-1$ by induction hypothesis, the Δ-iterates satisfy

$$\begin{aligned} \Delta^m(2 \cdot F_{2n} \cdot 1^{-1})[0] &= \Delta^m(2 \cdot F_{2n} \cdot 1^{-1})[\mathit{Last}] \\ &\neq \Delta^m(1 \cdot F_{2n-1} \cdot 2^{-1})[0] = \Delta^m(1 \cdot F_{2n-1} \cdot 2^{-1})[\mathit{Last}], \end{aligned}$$

where *Last* abusively denotes the index of the last letter of a word. We may now apply the glueing Lemma 2 to equation (2) in order to obtain

$$\Delta^{2n-1}(2 \cdot F_{2n+2} \cdot 1^{-1}) = 313\,,$$

from which one concludes that

$$\begin{aligned} \Phi(2 \cdot F_{2n+2} \cdot 1^{-1}) &= 2(13)^{n-1}1 \oplus_{2n-1} \Phi \circ \Delta^{2n}(2 \cdot F_{2n+2} \cdot 1^{-1}) \\ &= 2(13)^{n-1}1 \cdot \Phi(313) \\ &= 2(13)^{n+1}. \end{aligned}$$

The proof of (ii) is similar and is left to the reader. □

In a similar way one can establish the following result.

Proposition 5. *The sequence of Fibonacci words satisfies for all* $n \geq 2$ *the conditions*

(i) $\Phi(F_{2n} \cdot 1^{-1}) = 112(13)^{n-1}$;
(ii) $\Phi(F_{2n+1} \cdot 2^{-1}) = 112(13)^{n-1} \cdot 12$.

We proceed now with showing that the alphabet used in the iterations of Δ is bounded.

Proposition 6. *The words*

$$F_{2n} \cdot 1^{-1},\ F_{2n+2} \cdot 2^{-1},\ 2 \cdot F_{2n+2} \cdot 1^{-1},\ 1 \cdot F_{2n+1} \cdot 2^{-1},\ n \in \mathbb{N},$$

are words in $\Delta^{(*)}(\mathbf{3})$.

Proof. First, we prove by induction on $n \geq 1$ that there exist two uniquely and well defined words V_n and W_n such that

$$\begin{array}{rclrcl} \Phi(V_n) & = & (13)^n, & \Phi(W_n) & = & 3(13)^n, \\ \Delta(V_n) & = & W_{n-1}, & \Delta(W_n) & = & V_n, \\ V_n & \in & \{1,3\}^*, & W_n & \in & \{1,3\}^*, \end{array}$$

and two consecutive occurrences in V_n or in W_n of the letter 3 are separated by 111 or 1.

One has $V_1 = 111$, $W_1 = 313$, $V_2 = 1113111$, $W_2 = 313111313$. Assume that the induction hypothesis holds for $n \geq 2$. The word V_{n+1} is uniquely determined by its first letter 1 and the fact that $\Delta(V_{n+1}) = W_n$. Similarly, W_{n+1} is uniquely determined. Since the $3's$ are separated by either 1 or 111, then 313 always code 1113111 in V_{n+1}, whereas the word 31113 always codes 111313111, which implies the desired property on V_{n+1}. The proof is similar for W_{n+1}.

Observe that we have proved that $V_n \in \{13, 11\}^*$, that is, V_n can be encoded over the alphabet $\{A, B\}$, where $A = 13$, $B = 11$, and that $V_{n+1} = \phi(V_n)$, where ϕ is defined by $\phi : A \mapsto ABA$, $B \mapsto AB$ (ϕ is the square of the Fibonacci morphism up to the alphabet).

Now, we have $\Delta^3(F_{2n} \cdot 1^{-1}) = V_{n-1}$, $\Delta^3(F_{2n+2} \cdot 2^{-1})$ is a prefix of V_n, $\Delta(2 \cdot F_{2n+2} \cdot 1^{-1}) = V_{n+1}$, and $\Delta^2(1 \cdot F_{2n+1} \cdot 2^{-1}) = V_n$, so that it only remains to check that the first interations of Δ produce words over the alphabet $\mathbf{3}$ to conclude. □

The infinite Fibonacci word satisfies the following property, which is a direct consequence of Proposition 5 and 6.

Proposition 7. *The word* F *satisfies* $\Phi(F) = 112(13)^\omega$ *and* $\Delta^*(F) \subset \mathbf{3}^\omega$.

It is well known that the Fibonacci word F does not contain cubes, and for the Δ-iterates the following patterns are avoided.

Lemma 8. *The factors* 33 *and* 31313 *never occur in* $\Delta^k(F)$*, for every* $k \geq 2$*. The factors* 22 *and* 21212 *never occur in* $\Delta(F)$.

Proof. One checks that 33 and 22 never occur in $\Delta^k(F)$, for $k \leq 2$. According to the proof of Proposition 6, 33 never occurs in V_n, for all n and hence in F. Assume now that the factor 31313 occurs in $\Delta^k(F)$, for some $k \geq 2$. Since 33 does not occur in $\Delta^{k-1}(F)$ (if $k = 2$, consider 22), then $\Delta(31313) = 11111 \in \Delta^k(F)$, which implies that the letter 5 occurs in $\Delta^{k+1}(F)$, a contradiction. The same argument applies for 21212. □

Let $\mathcal{F}$ denote the *Fibonacci shift*, that is, the set of infinite words having exactly the same factors as the Fibonacci word F; let us recall that $\mathcal{F}$ is the closure in $\{1,2\}^\omega$ of the orbit $\{s^k(F);\ k \in \mathbb{N}\}$ of F.

Example. $\Phi(2 \cdot F) = 213 \cdot (s^3 \circ \Phi)(F) = 2(13)^\omega$. Indeed by applying the glueing lemma, we have the following iterations of Δ on $2 \cdot F$

$$\begin{array}{lclcl}
\Delta^0(2F) & = & 2 \cdot F & = & 2 \cdot 1211212 \cdot 11211212112121121 1211212 \cdots \\
\Delta^1(2F) & = & 1 \cdot \Delta(F) & = & 1 \cdot 11 \cdot 2111 \cdot 2121112111212111 \cdots \\
\Delta^2(2F) & = & 3 \cdot \Delta(s^2(\Delta(F))) & = & 3 \cdot 13 \cdot 1113131113111 \cdots \\
\Delta^3(2F) & = & \Delta^3(F) & = & 1113111313 \cdots
\end{array}$$

that is,

$$\Phi(2F) = 2 \oplus_0 \Phi(1 \cdot \Delta(F)) = 213 \oplus_2 \Phi(\Delta^3(2F)),$$

so that $\Phi(2 \cdot F) = 213 \cdot \Phi(\Delta^3(F)) = 213 \cdot s^3 \circ \Phi(F)$.

We know that $\Phi(F)$ is eventually periodic so that the following question is natural: does such a behaviour extend to other words in the Fibonacci shift $\mathcal{F}$? More precisely is this property characteristic of the Fibonacci language or does it hold only for particular sequences of the Fibonacci shift? The next theorem answers this question:

Theorem 9. *Every word $U \in \mathcal{F}$ satisfies the following properties:*

(i) *U is a word of $\mathbf{5}^\omega$);*
(ii) *for every $k \geq 2$, $s(\Delta^k(U)) \in \{1,3\}^*$;*
(iii) *every factor of $\Delta^k(U)$ having 3 or 111 for prefix occurs in $\Delta^k(F)$;*
(iv) *if U belongs to the two-sided orbit under the shift s of F, that is, if there exists $n \in \mathbb{N}$ such that either $U = s^n(F)$ or $F = s^n(U)$, then $\Phi(U)$ eventually ends with $(13)^\omega$.*

Proof. The remaining of this section will be devoted to the proof of this theorem which requires several steps. We need first a preliminary lemma to state the base case of an induction property that we prove below.

Lemma 10. *Let $U \in \mathcal{F}$. Then $\Delta(U) \in \{1,2\}^\omega$ and we have:*

(i) *two consecutive occurrences of the letter 2 in $\Delta(U)$ are separated by 1 or 111; 2 occurs infinitely often;*
(ii) *every factor having 2 or 111 for prefix occurs in $\Delta(F)$.*

Proof. Since $F = \varphi(F)$ it follows that $22, 111 \notin L(F) = L(U)$. Therefore two consecutive occurrences of 2 are separated by 1 or 11 in U, which implies that $\Delta(U) \in \{1,2\}^\omega$.

(i) Since $22 \notin U$, every occurrence of 2 in $\Delta(U)$ codes an occurrence of 11 in U. Let us prove that $11111 \notin L(\Delta(U))$. By contradiction, assume that 11111 is a factor, then 11111 would code an occurrence of either 121212 or 212121 in U, but neither word is a factor of F. Furthermore, two consecutive occurrences of 2 in $\Delta(U)$ cannot be separated by an even number of 1's: indeed, either the first or the last 2 would code 22 in U, which ends the proof of this statement.

(ii) Let w be a factor of $\Delta(U)$ whose prefix is either the letter 2 or the factor 111. It codes uniquely a factor in U and in F, implying that it belongs to $\Delta(F)$. $\square$

Let us come back to the proof of Theorem 9. We prove by induction the following assertions, where $x_k = 2$ if $k = 1$ and 3 otherwise;

(1) $\Delta^k(U)$ is well defined;
(2) $\Delta^k(U) \in \mathbf{5}^\omega$; $(s \circ \Delta^k)(U)) \in \{1, x_k\}^\omega$;
(3) two successive occurrences of x_k are separated either by 1 or 111; the letter x_k occurs infinitely often;
(4) every factor of $\Delta^k(U)$ having x_k or 111 for prefix occurs in $\Delta^k(F)$.

The induction property holds for $k = 1$ by Lemma 10. Fix now an integer $k \geq 1$ and assume that the induction property holds for both k and $k-1$. For the sake of simplicity, we assume that $k \geq 2$ and replace x_k by its value 3. The proof proceeds exactly in the same way when $k = 1$, $x_k = 2$. We only need to use the fact that 22 does not occur in $\Delta^0(U) = U$.

Observe first that the factors 33 and 31313 do not occur in $\Delta^k(U)$, and 33 does not occur in $\Delta^{k-1}(U)$, according to Assertion 4 and Lemma 8.

- From Assertions 1, 2 and 3 above, $\Delta^{k+1}(U)$ is easily seen to be well defined.
- We have three cases to consider.
  - If $\Delta^k(U)[0] = 3$, then $\Delta^{k+1}(U) \in \{1,3\}^\omega$, by Assertion 3.
  - If $\Delta^k(U)$ has 1^y3 ($y \geq 1$) for prefix, then $\Delta^{k+1}(U) = y\Delta(s^y \circ \Delta^k(U))$, and $s \circ \Delta^{k+1}(U) \in \{1,3\}^\omega$.
  - If $\Delta^k(U)[0] = y \neq 1, 3$, then $\Delta^k(U)$ has $y1^z$ ($z \geq 1$) for prefix, since the factor 33 cannot occur in $\Delta^{k-1}(U)$. If z is even, then Assertion 2 implies that $y1^z3$ would code a factor of the form $r^y(3131)^{z/2}333$ in $\Delta^k(U)$ ($r \in \mathbf{5}$), a contradiction with the fact that $33 \notin L(\Delta^k(U))$. If $z \geq 5$, then $y1^z3$ would code a factor of the form r^y31313, a contradiction with the fact that $31313 \notin L(\Delta^k(U))$. We have thus proved that $y \in \{1,3\}$, which implies that $(s \circ \Delta^{k+1}(U)) \in \{1,3\}^\omega$.

  Note that the first letter of $\Delta^{k+1}(U)$ is smaller than or equal to 5, since 31313 does not occur in $\Delta^{k-1}(U)$. Hence, $\Delta^{k+1}(U) \in \mathbf{5}^\omega$.
- The factor $33 \notin L(\Delta^{k+1}(U))$, otherwise 333 would occur in $\Delta^k(U)$. Hence every occurrence of the letter 3 in $\Delta^{k+1}(U)$ codes 111 in $\Delta^k(U)$. The factor $311113 \notin L(\Delta^{k+1}(U))$, otherwise it would code 1113131333 in $\Delta^k(U)$, contradicting the fact that 33 does not occur in $\Delta^k(U)$. Similarly, the factor $311111 \notin L(\Delta^{k+1}(U))$, otherwise it would code 11131313 in $\Delta^k(U)$, but 31313 does not occur in $\Delta^k(U)$. At last, the factor $3113 \notin L(\Delta^{k+1}(U))$, since otherwise it would code 11131333 in $\Delta^k(U)$, again a contradiction.

  Hence two consecutive occurrences in $\Delta^{k+1}(U)$ of 3 are separated either by 1 or 111, and the letter 3 occurs infinitely often.
- Let w be a factor of $\Delta^{k+1}(U)$ whose prefix is either 3 or the factor 111. It codes uniquely a factor in $\Delta^k(U)$ also starting with either 3 or 111, and belonging thus by Assertion 4 to $\Delta^k(F)$; therefore w belongs to $\Delta^{k+1}(F)$.

It remains now to prove that $\Phi(U)$ ultimately ends in $(13)^\omega$ if U is an image or a preimage of F under the action of the shift s to complete the proof of Theorem 9.

Assume first that U is a shifted image of the Fibonacci word F, that is, there exists $k \in \mathbb{N}$ such that $U = s^k(F)$. Let us now introduce a suitable factorization of $2F$. For that purpose, let us first observe that $F = \varphi^{2n+1}(F)$ can be uniquely decomposed over the ω-code $\{F_{2n}, F_{2n+1}\}$ (see Proposition 3), and even over the

ω-code $\{F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}, F_{2n+2}\cdot F_{2n+1}\}$. Hence we may factorize $2F$ over

$$\{2\cdot F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}, 2\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}\}.$$

Furthermore, the first term of this factorization is easily seen by induction to be $2\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}$, whereas its second term is $2\cdot F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}$. One has $U = s^{k+1}(2F)$. Let $n \geq 2$ be large enough such that $|F_{2n+3}| > k+1$. Let us write $2F_{2n+2}\cdot F_{2n+1}2^{-1}$ as

$$2\cdot F_{2n+2}\cdot F_{2n+1} = P_k\cdot Q_k,$$

where P_k is the prefix of $2F$ of length $k+1$; hence $2F = P_k\cdot U$, and

$$U = Q_k\cdot s^{|F_{2n+3}|}(2\cdot F),$$

i.e.,

$$U \in Q_k\cdot\{2\cdot F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}, 2\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}\}^{\omega},$$

the first term of this factorization being $2\cdot F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}$.

Let us observe that

$$2\cdot F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1} = (2\cdot F_{2n+2}\cdot 1^{-1})\cdot(1\cdot F_{2n+1}\cdot 2^{-1})\cdot(2\cdot F_{2n}\cdot 1^{-1})\cdot(1\cdot F_{2n+1}\cdot 2^{-1}),$$

and

$$2\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1} = (2\cdot F_{2n+2}\cdot 1^{-1})\cdot(1\cdot F_{2n+1}\cdot 2^{-1}).$$

Let us first prove that $\Phi(s^{|F_{2n+3}|}(2F)) = 2(13)^{n+1}112(13)^{\omega}$. Following Proposition 4 and Proposition 6, the glueing lemma applies, and implies that the first terms of $\Phi(s^{|F_{2n+3}|}(2F))$ are $2(13)^n$; let us note that $\Delta^{2n+1}(2\cdot F_{2n+2}\cdot 1^{-1}) = 111$, $\Delta^{2n+1}(1\cdot F_{2n+1}\cdot 2^{-1}) = 3$, $\Delta^{2n+1}(2\cdot F_{2n}\cdot 1^{-1}) = 1$. Hence

$$\Delta^{2n+1}(2\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}) = 111\cdot 3.$$

$$\Delta^{2n+1}(2\cdot F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}) = 111\cdot 3\cdot 1\cdot 3.$$

One concludes by considering the next values of Δ^k, $2n+2 \leq k \leq 2n+6$ and using the fact that $\Phi(2F) = \Phi(2\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}\cdot s^{|F_{2n+3}|}(2F)) = 2(13)^{\omega}$.

Let us prove that $\Phi(Q_k\cdot s^{|F_{2n+3}|}(2F))$ and $\Phi(s^{|F_{2n+3}|}(2F))$ ulimately coincide. Let m be the smallest integer such that $\Delta^m(Q_k) = 1$. One checks that $m \leq 2n+5$. Let us distinguish two cases according to the parity of m, and apply the glueing lemma, by noticing that the first term of the decomposition of $s^{|F_{2n+3}|}(2F)$ is $2\cdot F_{2n+2}\cdot F_{2n+2}\cdot F_{2n+1}\cdot 2^{-1}$.

- Assume that m is even. Assume furthermore $m \leq 2n$. Then the factor $\Delta^m(s^{|F_{2n+3}|}(2F))$ admits 313111313 as a prefix since $\Phi(s^{|F_{2n+3}|}(2F)) = 2(13)^{n+1}112(13)^{\omega}$. Hence $\Delta^{m+1}(Q_k\cdot s^{|F_{2n+3}|}(2F))$ admits 11113111 as a prefix, which implies that $\Delta^{m+2}(Q_k\cdot s^{|F_{2n+3}|}(2F))$ admits 413 as a prefix; one deduces that $\Phi(Q_k\cdot s^{|F_{2n+3}|}(2F))$ and $\Phi(s^{|F_{2n+3}|}(2F))$ coincide for indices larger than $m+3$. If $m = 2n+2$, then $\Delta^{2n}(s^{|F_{2n+3}|}(2F))$ admits 3111313 as a prefix, and similarly one checks that $\Phi(Q_k\cdot s^{|F_{2n+3}|}(2F))$ ends in $(13)^{\omega}$ from indices larger than or equal to $2n+5$. If $m = 2n+4$, then one checks that $\Phi(Q_k\cdot s^{|F_{2n+3}|}(2F))$ and $\Phi(s^{|F_{2n+3}|}(2F))$ coincide for indices larger than $2n+6$.
- Assume that m is odd. This implies that $\Delta^{m-1}(Q_k) = 2$. Assume that $m \leq 2n+1$. One checks that $\Delta^m(Q_k\cdot s^{|F_{2n+3}|}(2F))$ admits 11113 as a prefix, and thus $\Phi(Q_k\cdot s^{|F_{2n+3}|}(2F))$ and $\Phi(s^{|F_{2n+3}|}(2F))$ coincide for indices larger than $m+2$. If $m = 2n+3$, $\Phi(Q_k\cdot s^{|F_{2n+3}|}(2F))$ ends in $(13)^{\omega}$ from for indices larger than $2n+6$. If $m = 2n+5$, one checks that

$\Phi(Q_k \cdot s^{|F_{2n+3}|}(2F))$ and $\Phi(s^{|F_{2n+3}|}(2F))$ coincide for indices larger than $2n+8$.

One thus deduces that $\Phi(U)$ ultimately terminates in $(13)^\omega$.

Assume now that U is a preimage of F under an iterate of s, that is, there exists k such that $s^k(U) = F$. Since both $2F$ and $1F$ belong to $\mathcal{F}$, then U is either a preimage or $2F$ or of $1F$, that is, there exists a finite word P_U such that either $U = P_U \cdot 2F$ or $U = P_U \cdot 1F$. Using the factorizations, respectively, of $2F$ over $\{2 \cdot F_{2n+2} \cdot F_{2n+2} \cdot F_{2n+1} \cdot 2^{-1}, 2 \cdot F_{2n+2} \cdot F_{2n+1} \cdot 2^{-1}\}$ or $1F$ over $\{1 \cdot F_{2n+1} \cdot F_{2n+1} \cdot F_{2n} \cdot 2^{-1}, 1 \cdot F_{2n+1} \cdot F_{2n} \cdot 2^{-1}\}$ we may apply the same reasoning as above. Let us recall that $\Phi(2F) = 2(13)^\omega$, whereas one checks that $\Phi(1F) = 12(13)^\omega$. One thus obtains that $\Phi(P_U \cdot 2F)$ and $\Phi(P_U \cdot 1F)$ ultimately coincide with respectively $\Phi(2F)$ or $\Phi(1F)$, which ends the proof. □

We have thus proved that words that are images or preimages of F under the shift s eventually end with $(13)^\omega$. The next proposition states that this property does not hold for all words in $\mathcal{F}$, that is, there exist words U with the same set of factors as F for which $\Phi(U)$ presents a different behaviour.

Proposition 11. *There exist words U in $\mathcal{F}$ such that $\Phi(U)$ contains infinitely many occurrences of the letter* 2.

Proof. Let us exhibit an example of a Sturmian word U in $\mathcal{F}$ such that $\Phi(U)$ does not ultimately end in $(13)^\omega$. Let U be the limit word in $\{1,2\}^\omega$ of the sequence of finite words

$$U_n = (1 \cdot (F_7 \cdot F_{10}) \cdots (F_{2^k-1} \cdot F_{2^k+2}) \cdots (F_{2^n-1} \cdot F_{2^n+2}) \cdot 1^{-1}), n \geq 3.$$

This sequence of words converges for the usal topology on $\{1,2\}^\omega$ and for every n, U_n is a factor of the Fibonacci word F as we shall see now. Indeed, following [9], every finite concatenation of F_n's with decreasing order of indices and where no two consecutive indices occur, is a prefix of the Fibonacci word F. Hence

$$F_{2^n+2} \cdot F_{2^n-1} \cdots F_{10} \cdot F_7$$

is a prefix of F. Since $2F$ is also a Sturmian word in $\mathcal{F}$, $2 \cdot F_{2^n+2} \cdot F_{2^n-1} \cdots F_{10} \cdot F_7$ is also a factor of F. But

$$\begin{aligned} &2 \cdot F_{2^n+2} \cdot F_{2^n-1} \cdots F_{10} \cdot F_7 \cdot 2^{-1} = \\ &\quad (2 \cdot F_{2^n+2} \cdot 1^{-1}) \cdot (1 \cdot F_{2^n-1} \cdot 2^{-1}) \cdots (2 \cdot F_{10} \cdot 1^{-1}) \cdot (1 \cdot F_7 \cdot 2^{-1}) \end{aligned}$$

is a concatenation of palindromes by Proposition 3. The set of factors of F being stable under mirror image (see for instance [13]), we have

$$\begin{aligned} (1 \cdot F_7 \cdot 2^{-1}) \cdot (2 \cdot F_{10} \cdot 1^{-1}) \quad &\cdots \quad (1 \cdot F_{2^n-1} \cdot 2^{-1}) \cdot (2 \cdot F_{2^n+2} \cdot 1^{-1}) \\ &= \quad 1 \cdot (F_7 \cdot F_{10}) \cdots (F_{2^n-1} \cdot F_{2^n+2}) \cdot 1^{-1} \end{aligned}$$

is a factor of F. Hence the word U belongs to $\mathcal{F}$ since it is a limit of factors of the Fibonacci word, and admits for every n, U_n as a prefix. Consider now the following factorization

$$\begin{aligned} &(1 \cdot F_{2^n-1} \cdot 2^{-1}) \cdot (2 \cdot F_{2^n+2} \cdot 1^{-1}) = \\ &\quad (1 \cdot F_{2^n-1} \cdot 2^{-1}) \cdot (2 \cdot F_{2^n} \cdot 1^{-1}) \cdot (1 \cdot F_{2^n-1} \cdot 2^{-1})(2 \cdot F_{2^n} \cdot 1^{-1}). \end{aligned}$$

Following Proposition 4 and Proposition 6, the glueing lemma applies. One has $\Delta^{2^n}(1\cdot F_{2^n-1}\cdot 2^{-1}) = 1$, $\Delta^{2^n}(1\cdot F_{2^n+1}\cdot 2^{-1}) = 111$, and $\Delta^{2^n}(2\cdot F_{2^n}\cdot 1^{-1}) = 3$. Hence

$$\begin{aligned}
\Delta^{2^n}(1\cdot F_{2^n-1}\cdot F_{2^n+2}\cdot 2^{-1}) &= 1\cdot 3\cdot 111\cdot 3,\\
\Delta^{2^n+1}(1\cdot F_{2^n-1}\cdot F_{2^n+2}\cdot 2^{-1}) &= 1\cdot 1\cdot 3\cdot 1,\\
\Delta^{2^n+2}(1\cdot F_{2^n-1}\cdot F_{2^n+2}\cdot 2^{-1}) &= 2\cdot 1\cdot 1\\
\Delta^{2^n+3}(1\cdot F_{2^n-1}\cdot F_{2^n+2}\cdot 2^{-1}) &= 1\cdot 2\\
\Delta^{2^n+4}(1\cdot F_{2^n-1}\cdot F_{2^n+2}\cdot 2^{-1}) &= 1\cdot 1\\
\Delta^{2^n+5}(1\cdot F_{2^n-1}\cdot F_{2^n+2}\cdot 2^{-1}) &= 2.
\end{aligned}$$

By applying the glueing lemma, one proves by induction that

$$\Delta^{2^{n-1}+8}(U_n) = \Delta^{2^{n-1}+8}((1\cdot F_{2^n-1}\cdot 2^{-1})\cdot(2\cdot F_{2^n+2}\cdot 1^{-1})),$$

which implies $\Phi(U)[2^n+2] = 2$, for all $n \geq 3$. □

Remark One can in fact prove that there exist uncountably many words U in $\mathcal{F}$ such that $\Phi(U)$ does not ultimately end in $(13)^\omega$.

References

[1] J. Berstel, Axel Thue's papers on repetition words: a translation, *Publications du LaCIM*, (1995).

[2] S. Brlek, Enumeration of factors in the Thue-Morse word, *Discrete Appl. Math.* **24** (1989) 83–96.

[3] S. Brlek, A. Ladouceur, A note on differentiable palindromes, *Theoret. Comput. Sci.*, **302** (2003) 167–178

[4] S. Brlek, S. Dulucq, A. Ladouceur, L. Vuillon, Combinatorial properties of smooth infinite words, MFCS'04, Prague, Czech Republic (2004) submitted.

[5] A. Carpi, Repetitions in the Kolakovski sequence, *Bull. of the EATCS,* **50** (1993) 194–196.

[6] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory*, **7** (1973 138–153.

[7] F. M. Dekking, On the structure of self generating sequences, *Séminaire de théorie des nombres de Bordeaux*, exposé 31, 1980–1981.

[8] F. M. Dekking, What is the long range order in the Kolakoski sequence, *in R. V. Moody (ed.), The mathematics of Long-Range Aperiodic Order*, Kluwer Academic Publishers (1997) 115–125.

[9] J.-M. Dumont, A. Thomas, Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theoret. Comp. Sci.* **65** (1989), 153–169.

[10] W. Kolakoski, Self Generating Runs, Problem 5304, *American Math. Monthly*, **72** (1965) 674. Solution: *American Math. Monthly*, **73** (1966) 681–682.

[11] A. Ladouceur, Outil logiciel pour la combinatoire des mots, *Mém. Maitrise en Math., UQAM*, AC20U5511 M6258, 1999.

[12] P. Lamas, Contribution à l' étude de quelques mots infinis, *Mém. Maitrise en Math., UQAM*, AC20U5511 M4444, 1995.

[13] M. Lothaire, Algebraic Combinatorics on words, Cambridge University Press, 2002.

[14] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theoret. Comput. Sci.*, **183** (1997) 45–82.

[15] M. Morse, 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] A. Thue, Über unendliche Zeichenreihen, *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.*, Christiana, Nr. **7** (1906) 1–22.

[18] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.*, Christiana, Nr. **1** (1912) 1–67.

LIRMM, 161 rue Ada, 34392 Montpellier Cedex 05, FRANCE
E-mail address: berthe@lirmm.fr

LaCIM, Université du Québec à Montréal, Montréal (QC) CANADA H3C 3P8
E-mail address: brlek@lacim.uqam.ca

Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto (ON) CANADA M3J1P3
E-mail address: philcho@mathstat.yorku.ca