Smooth Words over Arbitrary Alphabets

Valerie Berthe, S. Brlek, P. Choquette

Smooth words over arbitrary alphabets

Valérie Berthé* Srečko Brlek† Philippe Choquette‡

Abstract

Smooth infinite words over $\Sigma = \{1, 2\}$ are connected to the Kolakoski word $K = 221121\cdots$, defined as the fixpoint of the function Δ that counts the length of the runs of 1's and 2's. In this paper we extend the notion of smooth words to arbitrary alphabets and study some of their combinatorial properties. Using the run-length encoding Δ, every word is represented by a word obtained from the iterations of Δ. In particular we provide a new representation of the infinite Fibonacci word F as an eventually periodic word. On the other hand, the Thue-Morse word is represented by a finite one.

Mathematics Subject Classification: 68R15, 37B10.

Keywords: Kolakoski word, run-length encoding, Fibonacci word, smooth words, Thue-Morse word, Sturmian words.

1 Introduction

In two previous papers [3, 4] we defined the class of smooth words over the 2-letter alphabet $\Sigma = \{1, 2\}$, which is invariant under the action of the run-length encoding operator, and is related to the curious Kolakoski word

$$K = 2211212212211211221211212211211212212211212212112112 2\cdots$$

which received a noticeable attention by showing some intriguing combinatorial properties, constituting mainly a bouquet of conjectures.

The (finite) palindromes of this class are characterized by means of the left palindromic closure of the prefixes of the Kolakoski word and reveal an interesting perspective for understanding some of the conjectures [3]. In particular, recurrence, mirror invariance and permutation invariance are all direct consequences of the presence in K of these palindromes. This last assumption however remains a conjecture.

*LIRMM, 161 rue Ada 34392 Montpellier Cedex 05, France, `berthe@lirmm.fr`

†with the support of NSERC (Canada), LACIM, Département d'Informatique, UQAM, Montréal (QC) CANADA H3C 3P8, `brlek@lacim.uqam.ca`

‡York University, Département de Mathématiques et statistiques, 4700 Keele street, Toronto, Ontario CANADA M3J 1P3, `philcho@mathstat.yorku.ca`

Other regularities such as squares, overlaps and cubes can be studied in this framework. Indeed, the number of squares being finite – and consequently the number of overlaps – a simple verification shows that smooth words do not contain cubes [4], providing an extension of the work of A. Carpi [5].

In this paper we introduce the class of smooth infinite words over an arbitrary alphabet of k letters, and study some of their combinatorial properties. In particular, there is a natural map Φ between the free monoid over the k-letter alphabet and the class of smooth words. As an example we provide new characterizations of the Thue-Morse word M and the Fibonacci word F: while M is characterized by the finite word 1112113, F is represented by the infinite periodic word $112(13)^\omega$. More generally we also prove that any infinite word obtained by shifting F is represented by an ultimately periodic word ending in $(13)^\omega$, and among the Sturmian words, they appear to be the only ones with such a property.

2 Definitions and notation

Let us consider a finite *alphabet* of *letters* Σ. A *word* 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 $|w|_f$. Clearly, the length of a word is given by the number of its letters,

$$|w| = \sum_{\alpha\in\Sigma} |w|_\alpha. \tag{1}$$

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, we denote by $\mathrm{Pal}(L)$ 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

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

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 $|[u]| = |u|$.

3 Run-length encoding

The widely known *run-length encoding* 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 has been used for the enumeration of factors in the Thue-Morse sequence [2].

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

$$\Delta(w) = e_1 e_2 e_3 \cdots = \prod_{j \geq 0} e_j,$$

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

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

Example. Let $\Sigma = \{a, b, c\}$, and $w = aaabbaaaacccccaa$, then

$$\begin{aligned} w &= a^3 b^2 a^4 c^5 a^2, \\ (\Delta, \tau)(w) &= [(3, a), (2, b), (4, a), (5, c), (2, a)]. \end{aligned}$$

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) $p \in \mathrm{Pal}(\Sigma^*) \Longrightarrow \Delta(p) \in \mathrm{Pal}(\mathbb{N}^*)$.

Note that the function Δ is not bijective. Moreover this coding is easily extended to infinite words as $(\Delta, \tau) : \Sigma^\omega \longrightarrow \mathbb{N}^\omega \times \Sigma^\omega$. Note that the first component $\Delta(w)$ of a word w is a vector so that the usual operations on vectors apply.

The alphabet Σ being finite (countable), it may be identified with an integer alphabet $\Sigma \longrightarrow \mathbf{k}$ where

$$\mathbf{k} = \{1, 2, 3, \cdots, k\}, k \in \mathbb{N} .$$

Example. Let $\Sigma = \mathbf{3}$, and $w = 122231112233$, then $w = 1^1 2^3 3^1 1^3 2^2 3^2$ and

$$(\Delta, \tau)(w) = [(1, 1), (3, 2), (1, 3), (3, 1), (2, 2), (2, 3)].$$

Often, when the alphabet is small enough, the punctuation and brackets are omitted in order to manipulate the more compact notation

$$\Delta(w) = 131322.$$

In this example the coding alphabet coincides with the alphabet of w, rising the problem of the existence of fixpoints, which are considered later. Observe for the moment that in this case w does not contain blocks of the form α^{k+1}, for $\alpha \in \mathbf{k}$. The function Δ is a contraction, that is, for every word $w \in \mathbf{k}^*$ we have

$$|\Delta(w)| \leq |w| , \tag{2}$$

and equality holds when $w \in \mathbf{k}^* - \mathbf{k}^* \cdot \{1^2, 2^2, \ldots, k^2\} \cdot \mathbf{k}^* = \tau(\mathbf{k}^*)$.

The operator Δ can be iterated, provided the process is stopped when the resulting word has length 1. Note that in general the coding alphabet may change at each iteration. We restrict ourself to the set of words whose Δ-iterates are coded on some (possibly large) alphabet $\mathbf{k}$:

$$\Delta^{(m)}(\mathbf{k}) = \{w \in \mathbf{k}^* \,|\, \Delta^m(w) \in \mathbf{k} - \{1\} \text{ and } \Delta^j(w) \in \mathbf{k}^+ \text{ for } 1 \leq j \leq m-1\}$$

and denote $\Delta^{(*)}(\mathbf{k}) = \cup_{m \geq 0} \Delta^{(m)}$. Doing so we obtain the representation

$$\begin{aligned} \Phi \quad &: \quad \Delta^{(*)}(\mathbf{k}) \longrightarrow \mathbf{k}^+ , \\ \Phi(w)[j] \quad &= \quad \Delta^j(w)[0] \quad \text{for} \quad 0 \leq j \leq m. \end{aligned} \tag{3}$$

Example. Let $w = 11222333123112233111233111222$. The iteration of Δ gives:

$\Delta^0(w) = 11222333123112233111233111222,$
$\Delta^1(w) = 23311122231223,$
$\Delta^2(w) = 12331121,$
$\Delta^3(w) = 112211,$
$\Delta^4(w) = 222,$
$\Delta^5(w) = 3,$

hence $\Phi(w) = 121123$.

Let $d : \Sigma \longrightarrow \Sigma^*$ be defined by $d(\alpha) = \alpha\alpha$, which amounts to duplicate the letters. Then

$$\Delta(d(w)) = 2 * \Delta(w), \tag{4}$$

where $*$ is the scalar multiplication on the vector $\Delta(w)$. The duplication of letters in w changes only the second letter of $\Phi(d(w))$. More precisely,

$$\Phi(d(w)) = w[0] \cdot 2 * \Delta(w)[0] \cdot \Phi(\Delta^2(w)). \tag{5}$$

This property extends to r-plication of letters. Indeed, let $d : \Sigma \times \mathbb{N} \longrightarrow \Sigma^*$ be defined by $d(\alpha, r) = \alpha^r$. Then we have

Proposition 2 $\Delta \circ d(w, r) = r * \Delta(w)$.

Example. Let $w = 121121211211211211212$. Then

$\Delta^0(w) = 121121211211211211212,$
$\Delta^1(w) = 1121112121112111,$
$\Delta^2(w) = 213111313,$
$\Delta^3(w) = 1113111,$
$\Delta^4(w) = 313,$
$\Delta^5(w) = 111,$
$\Delta^6(w) = 3,$

and we have

$$\begin{array}{rcl} d(w,2) & = & 112211112211221111221111221122111122112211 \\ \Phi(d(w,2)) & = & 1221313 = 1 \cdot 2 \cdot 21313. \end{array}$$

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

$$\Delta(uv) = \Delta(u) \cdot \Delta(v) \iff \widetilde{u}[0] \neq v[0], \tag{6}$$

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 3 (Glueing Lemma) *Let $u,\, v \in \Delta^{(*)}(\mathbf{k})$. If there exists an index m such that, for all $i, 0 \le i \le m$, 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) $\Delta^i(uv) = \Delta^i(u)\Delta^i(v)$.

Proof. By definition of Φ, the last letter of $\Delta^i(u)$ is $\Phi(\widetilde{u})[0]$ and the first letter of $\Delta^i(v)$ is $\Phi(v)[0]$. If $m = 0$ then (6) applies. If $m > 0$, by iterating (6) we have for every $i \le m$,

$$\begin{aligned} \Phi(uv)[i] &= \Delta^i(uv)[0] = \Delta^i(u)[0], \\ &= \Phi(u)[i], \end{aligned}$$

which implies (i). The proof of (ii) is obvious. ◼

Example. Let $u = 11211$ and $v = 21211$. Applying Δ we obtain

$$\begin{array}{rrcl} \Delta^0(uv) = & \mathbf{1}1211 & \cdot & 21211 \\ \Delta^1(uv) = & \mathbf{2}12 & \cdot & 1112 \\ \Delta^2(uv) = & \mathbf{1}11 & \cdot & 31 \\ \Delta^3(uv) = & \mathbf{3} & \cdot & 11 \\ \Delta^4(uv) = & \mathit{1} & \cdot & 2 \\ \Delta^5(uv) = & \mathit{1} & \cdot & \mathit{1} \\ \Delta^6(uv) = & \mathit{2} & & \end{array}$$

We have $m = 3$ and $\Phi(11211 \cdot 21211) = \mathbf{1213} \cdot \Phi(12) = 1213 \cdot 112$.

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.

The infinite case

The representation Φ extends with some caution to infinite words in a natural way. Indeed, define the set of infinite *smooth* words

$$\mathcal{K}(\mathbf{k}) = \{W \in \mathbf{k}^\omega \mid \forall m \in \mathbb{N}, \Delta^m(W) \in \mathbf{k}^\omega\}.$$

The extension is $\Phi : \mathcal{K}(\mathbf{k}) \longrightarrow \mathbf{k}^\omega$, denoted and defined identically by (3). The elements of the set $L(\mathcal{K}(\mathbf{k}))$ of finite factors of $\mathcal{K}(\mathbf{k})$ are also called smooth, generalizing the definition given by F.M. Dekking [7]. The next property, easily obtained from the definition, expresses the duality between $\mathcal{K}(\mathbf{k})$ and $\mathbf{k}^\omega$.

Proposition 4 *For all $W \in \mathcal{K}(\mathbf{k})$ and all m we have $\Phi(\Delta^m(W)) = s^m \circ \Phi(W)$.*

The function Φ appears in the thesis of P. Lamas [12] and is used for a classification of infinite words. Independently, F.M. Dekking [8] used it in the case of a 2-letter alphabet in order to show the existence of words satisfying $\Delta^m(W) = W$ for every $m \in \mathbb{N}$, the Kolakoski word K corresponding to the case $m = 1$.

The glueing Lemma 3 admits an extension to infinite words: let $u \in \Delta^{(*)}(\mathbf{k})$ and $v \in \mathcal{K}(\mathbf{k})$. 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 properties in Proposition 1 imply the following closure properties:

$$U \in \Delta^m(\mathbf{k}) \iff \rho U, \widetilde{U} \in \Delta^m(\mathbf{k}),\ \forall m \geq 0; \tag{7}$$

$$U \in \mathcal{K}(\mathbf{k}) \iff \rho U \in \mathcal{K}(\mathbf{k}). \tag{8}$$

The fact that $\widetilde{u}$ does not appear in statement (8) is not surprising because closure by mirror image clearly requires to work with twosided infinite words.

4 Results

This section gathers results first on the the Thue-Morse word, then on the Fibonacci word, and ends by considering the Sturmian ones.

4.1 The Thue-Morse word

Recall that the Thue-Morse word M word may be obtained as the fixed point of the morphism $\mu : \{1, 2\} \longrightarrow \{1, 2\}^*$ defined by

$$\mu(1) = 12 \quad ; \quad \mu(2) = 21,$$

and it is easily seen that M is not smooth over a two-letter alphabet. Nevertheless, by using arbitrary intermediate coding alphabets we obtain the following representation for the iterates of $M_n = \mu^n(1)$.

Proposition 5 *The Thue-Morse word M satisfies the following conditions:*

(i) $\Phi(M_1) = 112$; $\Phi(M_2) = 1113$; $\Phi(M_3) = 111213$;

(ii) $\Phi(M_n) = 1112113$, *if* $n \geq 4$.

Proof. The first 3 conditions are easy to obtain. Let $n \geq 4$. Recall from [2] the properties that $\Delta(M_n)$ is a palindrome for all n, and also that

$$\Delta(M_n) = \begin{cases} \Delta(M_{n-2}) \cdot \Delta(M_{n-1}) \cdot \Delta(M_{n-2},) & \text{if } n \text{ is even,} \\ \\ \Delta(M_{n-1}) \cdot \Delta(M_{n-1}), & \text{if } n \text{ is odd.} \end{cases} \tag{9}$$

Moreover the only cube factor is 222 (corresponding to 112211) and its occurrences are centered on the positions $2^{2k+1}, k \geq 1$. It follows that $\Delta^2(M_n)$, is palindromic because Δ preserves palindromicity. We prove now by induction that

$$\Delta^2(M_n) = 11 \cdot x \cdot 11, \quad \text{where} \quad x \in \mathrm{Pal}(2\Sigma^*2), x_i \neq x_{i+1}. \tag{10}$$

We have $\Delta^2(M_4) = 1123211$, which establishes the base case. Assume now by induction hypothesis that (10) holds. Then from (9) we have

$$\Delta^2(M_{n+1}) = \Delta^2(M_n)1^{-1} \cdot 2 \cdot 1^{-1}\Delta^2(M_n)$$

which also satisfies (10). Therefore, for $n \geq 4$ we have

$$\begin{array}{rcl} \Delta^2(M_n) & = & 11 \cdot x \cdot 11 \\ \Delta^3(M_n) & = & 2 \cdot 1^{\alpha(n)} \cdot 2\,, \\ \Delta^4(M_n) & = & 1 \cdot \alpha(n) \cdot 1\,, \\ \Delta^5(M_n) & = & 111\,, \\ \Delta^6(M_n) & = & 3\,, \end{array}$$

where $\alpha(n) = |x|$ is an $\mathcal{O}(2^n)$ function, since $|M_n| = 2^n$. ◼

The proof above does not request $\alpha(n)$ to be explicited. However, for sake of completeness we give its values :

$$\alpha(n) = \begin{cases} \left\lfloor \dfrac{2^n - 6}{3} \right\rfloor & \text{if } n \text{ is even,} \\ \\ \left\lceil \dfrac{2^n - 6}{3} \right\rceil & \text{if } n \text{ is odd.} \end{cases}$$

4.2 The Fibonacci word

We recall the construction of the infinite Fibonacci word. 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. The infinite Fibonacci word F is obtained as the fixed point of φ, that is

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

One easily checks that F is not smooth on a 2-letter alphabet:

$$\begin{array}{rcl} \Delta^0(F) & = & 12112121121121211212\cdots \\ \Delta^1(F) & = & 1121112121112111\cdots \\ \Delta^2(F) & = & 213111313\cdots \end{array}$$

but it is smooth over the 3-letter alphabet **3** as we shall see.

The Fibonacci finite words

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

Proposition 6 *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 7 *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 6 for $2n+2$ and obtain

$$\begin{array}{rcl} \Phi(2 \cdot F_{2n+2} \cdot 1^{-1}) & = & \Phi(2 \cdot (F_{2n} F_{2n-1} F_{2n}) \cdot 1^{-1}) \\ & = & \Phi(2 \cdot (F_{2n} \cdot 1^{-1} 1 \cdot F_{2n-1} \cdot 2^{-1} 2 \cdot F_{2n}) \cdot 1^{-1}) \end{array} \quad (11)$$

Recall that Δ preserves palindromicity (Proposition 1), and that $2 \cdot F_{2n+2} \cdot 1^{-1}$ is palindromic (Proposition 6). Therefore, for every $m \leq 2n-1$ by induction hypothesis, the Δ-iterates satisfy

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

where *Last* abusively denotes the index of the last letter of a word. We may now apply the glueing Lemma 3 to equation (11) 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. ◼

Proposition 8 *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) $\Phi(F_{2n+1} \cdot 2^{-1}) = 112(13)^{n-1} \cdot 12$.

Proof. Again, we proceed by induction. It is easily verified for $n = 2$. Assume that the conditions hold until n. The recurrence relations of Proposition 6 imply that

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

By induction hypothesis we have $\Phi(F_{2n+1} \cdot 2^{-1}) = 112(13)^{n-1} \cdot 12$, and by Proposition 7 we also have $\Phi(2 \cdot F_{2n} \cdot 1^{-1}) = 2(13)^n$. This implies that

$$\Delta^m(F_{2n+1} \cdot 2^{-1}) = 31, \quad \text{and} \quad \Delta^m(2 \cdot F_{2n} \cdot 1^{-1}) = 3,$$

where $m = 2n$. Then the glueing lemma applies and yields

$$\begin{aligned}\Phi(F_{2n+2} \cdot 1^{-1}) &= \Phi(F_{2n+1} \cdot 2^{-1}) \oplus_{2n} \Phi(\Delta^{m+1}(F_{2n+2} \cdot 1^{-1})) \\ &= \Phi(F_{2n+1} \cdot 2^{-1}) \cdot \Phi(\Delta(31 \cdot 3)) \\ &= 112(13)^{n-1} \cdot 13 = 112(13)^n\end{aligned}$$

establishing condition (i). In a quite similar way, using the decomposition

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

the condition (i) and the glueing lemma one obtains the condition (ii). ◼

Proposition 9 *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 smooth 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 codes 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 **3** to conclude. ◼

The Fibonacci infinite words

The infinite Fibonacci word satisfies the following property, which is a direct consequence of Proposition 8 and 9.

Proposition 10 *The Fibonacci word F is a smooth word in $\mathcal{K}(\mathbf{3})$. Furthermore, one has $\Phi(F) = 112(13)^\omega$.*

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

Lemma 11 *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 9, 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 $2 \cdot F$

$$\begin{array}{lcllcl}
\Delta^0(2F) & = & 2 \cdot F & & = & 2 \cdot 1211212 \cdot 112112121121211211212 \cdots \\
\Delta^1(2F) & = & 1 \cdot \Delta(F) & & = & 1 \cdot 11 \cdot 2111 \cdot 2121112111212111 \cdots \\
\Delta^2(2F) & = & 3 \cdot \Delta(s^2(\Delta(F))) & & = & 3 \cdot 13 \cdot 1113131113111 \cdots \\
\Delta^3(2F) & = & \Delta^3(F) & & = & 1113111313 \cdots
\end{array}$$

that is,

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

so that $\Phi(2 \cdot F) = 213 \cdot \Phi(\Delta^3(F)) = 213 \cdot s^3 \circ \Phi(F)$, where the last equality is obtained by the duality property of Proposition 4.

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

(i) *U is a smooth word of $\mathcal{K}(\mathbf{5})$;*

(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 13 *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 $\Delta(F)$. ■

Let us come back to the proof of Theorem 12. 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 13. 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 11.

- 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 12.

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 6), 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 7 and Proposition 9, 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 $(13)^{\omega}$ for 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 $(13)^{\omega}$ 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 14 *There exist words U in $\mathcal{F}$ such that $\Phi(U)$ contains infinitely many occurrences of the letter* 2.

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 6. The set of factors of F being stable under mirror image (see for instance [13]), we have

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

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 7 and Proposition 9, 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 $(13)^\omega$.

4.3 Sturmian words

Proposition 15 *The only Sturmian words that are smooth belong to the Fibonacci shift.*

Proof. Let V be a smooth Sturmian word, defined over the alphabet $\mathbf{2}$. Since V is Sturmian, one letter, 2 say, admits isolated occurrences with consecutive occurrences always separated by factors of the form 1^x, and 1^{x+1}, for $x \geq 1$. This is a direct consequence of the fact that a Sturmian word is balanced (see for instance [6, 16]). Hence $\Delta(V)$ takes its values in $\mathbf{k+1}$.

Suppose $x \geq 2$. Then eventually no power of a letter occurs in $\Delta^2(V)$, that is, all the letters are isolated. Hence $\Delta^2(V)$ eventually ends in $11111\cdots$, which implies that $\Delta^3(V)$ is not defined. We thus get that $x = 1$ and that V ultimately belongs to $\{12, 1\}^\omega$ where the occurrences of 1 in this decomposition are isolated, with consecutive occurrences separated by either 12 or 1212.

We prove now by induction on $n \geq 1$ that if $\Delta^{n+1}(V)$ is well defined, then V belongs up to a finite prefix to $\{F_n, F_{n-1}\}^\omega$, where the occurrences of F_{n-1} are isolated with consecutive occurrences of F_{n-1} separated by F_nF_n or F_n.

The induction hypothesis holds for $n = 1$. Assume now that the induction hypothesis holds for $n \geq 1$. Hence V may be ultimately coded over $\{F_n, F_{n-1}\}$. Since F_{n-1} has isolated occurrences, then V may be ultimately coded over $\{F_n \cdot F_{n-1}, F_{n-1}\}^\omega = \{F_{n+1}, F_n\}^\omega$.

Assume F_nF_n occurs in this decomposition. Then there exists an occurrence of the form $F_nF_nF_{n+1} = F_nF_nF_n \cdot F_{n-1}$, which contradicts the fact that $F_nF_nF_n$ does not occur betwen two consecutive occurrences of F_{n-1} in the decomposition over $\{F_n, F_{n-1}\}^\omega$. Hence the occurrences of F_n are isolated.

It remains to prove that between two consecutive occurrences of F_n do only occur $F_{n+1}F_{n+1}$ and F_{n+1}. Assume that $F_{n+1} \cdot F_{n+1} \cdot F_{n+1}$ occurs in V. Then between two consecutive occurrences of F_n do only occur $F_{n+1}F_{n+1}F_{n+1}$ powers at least equal to 2 or 3 F_{n+1}. Assume n even, $n = 2k$.

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

Glueing! It yields that $\Delta^{n+4}(V)$ is not defined, hence a contradiction.

The fact that for all n, the Sturmian sequence V belongs up to a finite prefix to $\{F_n \cdot F_{n-1}, F_n\}^\omega$ implies that it has the same language as F, and hence that it belongs to $\mathcal{F}$.

5 Concluding remarks

A number of problems arise from the representation Φ, which requires to store for every infinite word $w \in \mathbf{k}^*$, the sequence $[\Delta^j(w), j = 0..\infty]$, a major drawback for efficient computations. It is therefore natural to consider the cases that avoid such constraints and this closely relies on making Φ a bijection.

Recall that Δ is not injective (see Proposition 1), so that inverting Φ requires to store the word $\tau(\Delta^j(w))$ for every $j \geq 0$. When $\mathbf{k} = \{1, 2\}$ this word is periodic so that we need to store only its first letter (see [3, 8]). This construction may be generalized to an arbitrary alphabet leading to a bijection between a subset of $\Delta^{(*)}(\mathbf{k})$ and $\mathbf{k}^+ - \mathbf{k}^* \cdot 1$. We only need to assume a cyclic order on $\mathbf{k}$ and take the subset of $\Delta^{(*)}(\mathbf{k})$ compatible with that order. The operator Δ has many fixpoints. It is well-known that in the case of a 2-letter alphabet we have

$$\Delta(K) = K, \qquad \Delta(1 \cdot K) = 1 \cdot K.$$

Clearly $K \in \mathcal{K}(\mathbf{2})$, and we have $\Phi(K) = 2^\omega$ and $\Phi(1 \cdot K) = 1^\omega$. When the alphabet has more than 2 letters ($|\mathbf{k}| > 2$), the situation is slightly different. Assume for instance the alphabet $\mathbf{k}$ to be ordered. Indeed, for any fixed $m \in \mathbf{k} - \{1\}$, $\Phi^{-1}(m^\omega)$ is a fixed point.

Example. Let $\Sigma = \mathbf{3}$ with the cyclic order $(1, 3, 2)$. Then we have two fixpoints for Δ. The first one is

$$\Phi^{-1}(2^\omega) = 221132111332132221113321113322 11\ldots$$

which is the Kolakoski analogue on 3 letters. And the second one is

$$\Phi^{-1}(3^\omega) = 33322211133221132133322211332133322 1\ldots$$

Since both of them satisfy $\Delta(1 \cdot W) = 1 \cdot W$, a canonical representative for 1^ω can be provided as in the case of the alphabet $\mathbf{2}$, by setting

$$\Phi^{-1}(1^\omega) = 1 \cdot \Phi^{-1}(3^\omega)\ .$$

More generally, we may also consider the subset of $\Delta^{(*)}(\mathbf{k})$ such that $\tau(\Delta^j(w))$ is a prefix of a fixed infinite word W without squares. Then again one obtains fixed points in which the letters appear in the order given by W.

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.