# Extremal generalized smooth words * 

Geneviève Paquin ${ }^{\dagger} \ddagger \quad$ Srečko Brlek ${ }^{\ddagger}$<br>Damien Jamet ${ }^{\S}$<br>[paquin, brlek]@lacim.uqam.ca, jamet@loria.fr


#### Abstract

In this article, we consider smooth words over 2-letter alphabets $\{a, b\}$, with $a, b \in \mathbb{N}$ having same parity. We show that they all are recurrent and provide a linear algorithm computing the extremal words. Moreover, the set of factors of any infinite smooth word over an odd alphabet is closed under reversal, while it is not for even parity alphabets. The minimal word is an infinite Lyndon word if and only if either $a=1$ and $b$ odd, or $a, b$ even. We also describe a connection between generalized Kolakoski words and maximal infinite smooth words over even 2-letter alphabets. Finally, the density of letters in extremal words is $1 / 2$ for even alphabets, and $1 /(\sqrt{2 b-1}-1)$ for $a=1$ with $b$ odd.


## 1 Introduction

Smooth infinite words over $\Sigma=\{1,2\}$ form an infinite class $\mathcal{K}$ of infinite words containing the well known Kolakoski word [10]

$$
K=22112122122112112212112122112112122122112122121121122 \cdots
$$

defined as the fixed point of the run-length encoding function $\Delta$. They are characterized by the property that the orbit obtained by iterating $\Delta$ is contained in $\{1,2\}^{*}$. As a discrete dynamical system, $(\mathcal{K}, \Delta)$ is topologically conjugate of the full shift $\left(\Sigma^{*}, \sigma\right)$ where $\sigma$ is the shift operator. In the early work of Dekking [8] there are some challenging conjectures on the structure of $K$ that still remain unsolved despite the efforts devoted to the study of patterns in $K$. For instance, we know from Carpi [6] that $K$ does contain only a finite number of squares, implying by direct inspection that $K$ is cube-free, a result that was extended in [5] to the infinite class $\mathcal{K}$ of smooth words over $\Sigma=\{1,2\}$. Weakley [14] showed that the complexity function (number of factors of length $n$ ) of $\mathcal{K}$ is polynomially bounded.
In [4], a connection was established between the palindromic complexity and the recurrence of $K$. More recently, Berthé et al. [2] studied smooth words over arbitrary alphabets and obtained

[^0]a new characterization of the infinite Fibonacci word. Relevant work may also be found in [1] and in $[2,9]$, where generalized Kolakoski words are studied for arbitrary 2 -letter alphabets. Finally, in [12], the authors studied the extremal infinite smooth words, that is the minimal and the maximal ones w.r.t. the lexicographic order, over the alphabets $\{1,2\}$ and $\{1,3\}$ : a surprising link was established between the minimal infinite smooth word over $\{1,3\}$ and the Fibonacci word.
In the present work, we deal with smooth words over 2-letter alphabets $\{a, b\}$ where $a$ and $b$ are two integers having same parity, with $a<b$.
The paper is organized as follows. In section 2, we borrow from Lothaire [11] all the basic notions on combinatorics on words, while in section 3, we briefly sketch the computation of extremal infinite smooth words over a 2 -letter alphabet, and recall the main results of Paquin et al. [12]. Section 4 deals with the extremal smooth words over odd 2-letter alphabets. We generalize a result of [12] about the extremal words over $\{1,3\}$ : we show that $\Phi\left(m_{\{a, b\}}\right)=(a b)^{\omega}$ (Theorem 10 ). We deduce a linear algorithm for computing the extremal smooth words (Corollary 12). Next, we prove that the set $F(w)$ of factors of an infinite smooth word $w$ is closed under reversal, and consequently, that $w$ is recurrent (Proposition 13). Finally, we show that the minimal infinite smooth word is an infinite Lyndon word if and only if $a=1$ and that the Lyndon factorization of $\Delta\left(m_{\{a, b\}}\right)$ is an infinite sequence of finite Lyndon words (Theorem 19).
Section 5 is devoted to even 2-letter alphabets $\{a, b\}$, in which case $\Phi\left(m_{\{a, b\}}\right)=a b^{\omega}$ (Theorem 21 and Corollary 22), yielding in turn a linear algorithm to generate the extremal words. From the algorithm, we deduce that the frequencies of the letters $a$ and $b$ are $\frac{1}{2}$. Moreover, any infinite smooth word over an even 2-letter alphabet is recurrent despite the fact that the set of its factors is not closed under reversal (Proposition 25). Finally, the minimal words are infinite Lyndon words (Theorem 26), and a connection is established between generalized Kolakoski words and maximal infinite smooth words.

## 2 Preliminaries

Throughout this paper $\Sigma$ is a finite alphabet of letters equipped with an order $<$ on its letters. A finite word is a finite sequence of letters

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

of length $n$, and $w[i]$ denotes its $i$-th letter. The set of $n$-length words over $\Sigma$ is denoted $\Sigma^{n}$. By convention the empty word is denoted $\epsilon$ and its length is 0 . The free monoid generated by $\Sigma$ is defined by $\Sigma^{*}=\bigcup_{n \geq 0} \Sigma^{n}$. The set of right infinite words is denoted by $\Sigma^{\omega}$ 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 that of infinite ones. Given a word $w \in \Sigma^{*}$, a factor $f$ of $w$ is a word $f \in \Sigma^{*}$ satisfying

$$
\exists x, y \in \Sigma^{*}, w=x f y
$$

If $x=\epsilon$ (resp. $y=\epsilon$ ) then $f$ is called a prefix (resp. suffix). The set of all factors of $w$, also called the language of $w$, is denoted by $F(w)$, and those of length $n$ is $F_{n}(w)=F(w) \cap \Sigma^{n}$. Finally, $\operatorname{Pref}(w)$ denotes the set of all prefixes of $w$. The length of a word $w$ is $|w|$, and the number of occurrences of a factor $f \in \Sigma^{*}$ is $|w|_{f}$. An infinite word $w$ is said recurrent if $|w|_{f}$ is
infinite for every factor $f$. A block of length $k$ is a factor of the particular form $f=\alpha^{k}$, with $\alpha \in \Sigma$.
Over an arbitrary 2-letter alphabet $\Sigma=\{a, b\}$, there is a usual length preserving morphism, the swapping of the letters, defined by $\bar{a}=b ; \bar{b}=a$, which extends to words as follows. The complement of $u=u_{1} u_{2} \cdots u_{n} \in \Sigma^{n}$, is the word $\bar{u}=\overline{u_{1}} \overline{u_{2}} \overline{u_{3}} \cdots \overline{u_{n}}$. The reversal of $u$ is the word $\widetilde{u}=u_{n} \cdots u_{1}$.
For $u, v \in \Sigma^{*}$, we write $u \prec v$ if and only if $u$ is a proper prefix of $v$ or if there exists an integer $k$ such that $u_{i}=v_{i}$ for $i=1, \ldots, k-1$ and $u_{k}<v_{k}$. The relation $\preceq$ defined by $u \preceq v$ if and only if $u=v$ or $u \prec v$, is called the lexicographic order. That definition holds for $\Sigma^{\infty}$. Note that in general, the complementation does not preserve the lexicographic order. Indeed, when $u$ is not a proper prefix of $v$ then

$$
\begin{equation*}
u \succ v \Longleftrightarrow \bar{u} \prec \bar{v} . \tag{1}
\end{equation*}
$$

A word $u \in \Sigma^{*}$ is a Lyndon word if $u \prec v$ for all proper suffixes $v$ of $u$. For instance, the word 11212 is a Lyndon word while 12112 is not. A word of length 1 is clearly a Lyndon word. From Lothaire [11], we have the following theorem.

Theorem 1. Any non empty finite word is uniquely expressed as a non increasing product of Lyndon words.

This product is called a Lyndon factorization. Siromoney et al. [13] extended Theorem 1 to infinite words. For that purpose, they introduced infinite Lyndon words as inductive limits of sequences of finite Lyndon words:

Theorem 2. [13] Any infinite word $w$ is uniquely expressed as a non increasing product of Lyndon words, finite or infinite, in one of the two following forms:
(i) either there exists an infinite non increasing sequence of finite Lyndon words $\left(l_{k}\right)_{k \geq 0}$ such that $w=l_{0} l_{1} l_{2} \ldots$,
(ii) there exist finite Lyndon words $l_{0}, \ldots, l_{m-1}(m \geq 0)$ and an infinite Lyndon word $l_{m}$ such that $w=l_{0} l_{1} \ldots l_{m-1} l_{m}, \quad$ with $l_{0} \succeq \ldots \succeq l_{m-1} \succ l_{m}$.

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 also was used for the enumeration of factors in the Thue-Morse sequence [3]. Let $\Sigma=\{a, b\}$ be an ordered alphabet. Then every word $w \in \Sigma^{*}$ can be uniquely written as a product of factors as follows:

$$
w=a^{i_{1}} b^{i_{2}} a^{i_{3}} \ldots
$$

where $i_{k} \geq 0$. The operator giving the size of the blocks appearing in the coding is a function $\Delta: \Sigma^{*} \longrightarrow \mathbb{N}^{*}$, defined by $\Delta(w)=i_{1}, i_{2}, i_{3}, \cdots$ which is easily extended to infinite words as $\Delta: \Sigma^{\omega} \longrightarrow \mathbb{N}^{\omega}$. For instance, let $\Sigma=\{1,3\}$ and $w=13333133111$, then $w=1^{1} 3^{4} 1^{1} 3^{2} 1^{3}, \quad$ and $\quad \Delta(w)=[1,4,1,2,3]$. Often the punctuation and the parentheses are omitted in order to manipulate the more compact notation $\Delta(w)=14123$.

This example is a special case where the coding integers does not coincide with the alphabet on which is encoded $w$, so that $\Delta$ can be viewed as a partial function $\Delta:\{1,3\}^{*} \longrightarrow\{1,2,3,4\}^{*}$. Recall from [4] that $\Delta$ is not bijective since $\Delta(w)=\Delta(\bar{w})$, but commutes with the reversal ( $\sim$ ), is stable under complementation ( ${ }^{-}$) and preserves palindromicity.
The operator $\Delta$ may be iterated, provided the process is stopped when the coding alphabet changes or when the resulting word has length 1 .

Example 3. Let $w=1333111333133311133313133311133313331113331$. The successive application of $\Delta$ gives :
$\Delta^{0}(w)=1333111333133311133313133311133313331113331$;
$\Delta^{1}(w)=1333133311133313331 ;$
$\Delta^{2}(w)=131333131 ;$
$\Delta^{3}(w)=1113111 ;$
$\Delta^{4}(w)=313 ;$
$\Delta^{5}(w)=111 ;$
$\Delta^{6}(w)=3$.
The operator $\Delta$ extends to infinite words (see [4]). Define the set of infinite smooth words over $\Sigma=\{a, b\}$ by

$$
\mathcal{K}_{\Sigma}=\left\{w \in \Sigma^{\omega} \mid \forall k \in \mathbb{N}, \Delta^{k}(w) \in \Sigma^{\omega}\right\} .
$$

In $\mathcal{K}_{\Sigma}$ the operator $\Delta$ has two fixpoints, namely

$$
\Delta\left(K_{(a, b)}\right)=K_{(a, b)}, \quad \Delta\left(K_{(b, a)}\right)=K_{(b, a)},
$$

where $K_{(a, b)}$ is the generalized Kolakoski word [9] over the alphabet $\{a, b\}$ starting with the letter $a$. For instance, the Kolakoski word [10] over the alphabet $\Sigma=\{1,2\}$ is $K=K_{(2,1)}$. We recall from [12] that the right derivative is a function $D_{r}: \Sigma^{*} \rightarrow \mathbb{N}^{*}$ such that:

$$
D_{r}(w)= \begin{cases}\epsilon & \text { if } \Delta(w)=a, a<b \text { or } w=\epsilon, \\ \Delta(w) & \text { if } \Delta(w)=x b, \\ x & \text { if } \Delta(w)=x a, a<b,\end{cases}
$$

with $\Sigma=\{a, b\}$. A word $w$ is $r$-smooth (also said smooth prefix) if $\forall k \geq 0, D_{r}^{k}(w) \in \Sigma^{*}$.

A bijection $\Phi: \mathcal{K}_{\Sigma} \longrightarrow \Sigma^{\omega}$ is defined by

$$
\Phi(w)[j+1]=\Delta^{j}(w)[1], \text { for } j \geq 0,
$$

and its inverse is inductively defined as follows. Let $u \in \Sigma^{k}$, then

$$
\Phi^{-1}(u)=w_{k},
$$

where

$$
w_{n}= \begin{cases}u[k], & \text { if } n=1 ; \\ \Delta_{u[k-n+1]}^{-1}\left(w_{n-1}\right), & \text { if } 1<n \leq k .\end{cases}
$$

Example 4. For the word $w=1333111333133311133313133311133313331113331$ of Example $3, \Phi(w)=1111313$. We also find inductively, starting from the bottom, that

$$
\Phi^{-1}(\Phi(w))=w .
$$

## 3 Computation of extremal smooth words

Let $m_{\{a, b\}}\left(\right.$ resp. $\left.M_{\{a, b\}}\right)$ be the minimal (resp. maximal) infinite smooth word over the alphabet $\Sigma=\{a, b\}$ w.r.t the lexicographic order. From (1), it easily follows that $M_{\{a, b\}}=\bar{m}_{\{a, b\}}$, so that the computation of $m_{\{a, b\}}$ also yields $\bar{m}_{\{a, b\}}$, by simply exchanging the order on the alphabet. The naive algorithm for computing the minimal infinite smooth word over an alphabet $\Sigma$ consists in computing the minimal smooth prefixes of increasing length. At each step, the minimal letter of the alphabet $\Sigma$ which makes the word a smooth prefix is added. The smoothness condition is checked with the right derivative operator $D_{r}$, and ensures that the prefix computed is the prefix of at least one infinite smooth word. If we assume $a \prec b$, the corresponding algorithm is:

```
\(m_{\{a, b\}}:=a ;\)
LOOP
IF isSmooth \(\left(m_{\{a, b\}} \cdot a\right)\) THEN \(m_{\{a, b\}}:=m_{\{a, b\}} \cdot a\);
ELSE \(m_{\{a, b\}}:=m_{\{a, b\}} \cdot b\);
END IF;
EXIT WHEN length \(\left(m_{\{a, b\}}\right)=\) MaxLength;
END LOOP;
```

Observe that this algorithm is independant of the letter parities. For MaxLength $=52$, we obtain the following words:

$$
\begin{aligned}
& m_{\{1,2\}}[1 \ldots 52]=1121122121121221121121221211221221121121221121122121, \\
& M_{\{1,2\}}[1 \ldots 52]=2121122121121221121121221211221221121121221121122121 \text {, } \\
& m_{\{1,3\}}[1 \ldots 52]=1113111313111311131311131311131113131113111313111313, \\
& M_{\{1,3\}}[1 \ldots 52]=3331333131333133313133313133313331313331333131333131 \text {, } \\
& m_{\{2,4\}}[1 \ldots 52]=2222444422224444224422442222444422224444224422442222 \text {, } \\
& m_{\{3,5\}}[1 \ldots 52]=3333355555333335553335553333355555333335553335553333, \\
& m_{\{2,3\}}[1 \ldots 52]=2223332223322332223332223322333222332233322233322332 \text {, } \\
& m_{\{3,4\}}[1 \ldots 52]=3333444433334443334443333444433334443334443333444433 .
\end{aligned}
$$

With the naive algorithm, the computation of an $n$-length prefix of $m_{\{a, b\}}$ takes $\mathcal{O}\left(n^{2} \log (n)\right)$ steps: indeed, for every newly added letter to the current prefix of $m_{\{a, b\}}$, we have to check smoothness by applying the $D_{r}$ operator. To improve the amount of $D_{r}$ operations, it is convenient to add more than one letter at each step. That was already studied in [12] for $m_{\{1,2\}}$, using the De Bruijn graphs. The same idea could be applied to extremal infinite smooth words over different alphabets, but we shall prove in the next sections that more efficient algorithms exist for computing them.

## Extremal smooth words over $\{1,2\}$ and $\{1,3\}$

We recall without proof some results established in a previous paper [12].
Proposition 5. Let $m_{\{1,2\}}$ and $M_{\{1,2\}}$ be respectively the minimal and maximal infinite smooth words over $\Sigma=\{1,2\}$. Then the vertical word

$$
\begin{aligned}
& \Phi\left(m_{\{1,2\}}\right)=1212212112221121112112221111221211112222 \cdots \\
& \Phi\left(M_{\{1,2\}}\right)=2212212112221121112112221111221211112222 \cdots
\end{aligned}
$$

The next letters can be computed but no characaterization is known, so that we do not know whether $\Phi\left(m_{\{1,2\}}\right)$ or $\Phi\left(M_{\{1,2\}}\right)$ are periodic or not. Nevertheless, the minimal smooth word $m_{\{1,2\}}$ is not an infinite Lyndon word [12].
In [2] Berthé et al. showed that the infinite Fibonacci word $F$, defined as

$$
F=\lim _{n \rightarrow \infty} F_{n} \quad \text { where } \quad F_{0}=2, \quad F_{1}=1, \quad \text { and } \quad \forall n \geq 2, F_{n}=F_{n-1} F_{n-2},
$$

is not smooth over the alphabet $\Sigma=\{1,2\}$, but smooth over the alphabet $\Sigma=\{1,2,3\}$. More precisely, they proved that $\Phi(F)=112(13)^{\omega}$, the periodicity meaning that $\Delta^{k}(F)=\Delta^{k+2}(F)$ for all $k \geq 3$. In [12], the link between the Fibonacci word and the minimal infinite smooth word over $\Sigma=\{1,3\}$ is established:

Theorem 6. $m_{\{1,3\}}=\Delta^{3}(F)$.
Since $F$ and $m_{\{1,3\}}$ are in the same orbit the next properties follow immediately from properties established for the Fibonacci orbit in [2].

Corollary 7. The extremal infinite smooth words over $\Sigma=\{1,3\}$ satisfy the following conditions:
(i) $\Delta^{k}\left(m_{\{1,3\}}\right)=\Delta^{k+2}\left(m_{\{1,3\}}\right)$, for all $k \geq 0$.
(ii) $\Phi\left(m_{\{1,3\}}\right)=(13)^{\omega}$ and $\Phi\left(M_{\{1,3\}}\right)=3(31)^{\omega}$.
(iii) $m_{\{1,3\}}$ does not admit the factors 33 and 31313, and its complement, $M_{\{1,3\}}$, does not admit the complement factors 11 and 13131.
(iv) Let $m_{\{1,3\}}=11 u$, then $\Delta\left(m_{\{1,3\}}\right)=3 u$.

The close relation between the Fibonacci word and the minimal infinite smooth word also provides a recursive definition for $m_{\{1,3\}}$.
Proposition 8. Let $m_{\{1,3\}}=11 u$. Then the word $u$ is defined as

$$
u=\lim _{n \rightarrow \infty} u_{n} \quad \text { where } \quad u_{0}=11, \quad u_{1}=13, \quad \text { and } \quad \forall n \geq 2, u_{n}=u_{n-1} u_{n-2} .
$$

Finally, from property (iv) of Corollary 7, the following transducer computing the minimal infinite smooth word $m_{\{1,3\}}$ in linear time is provided.
As we shall see in the next section, for other alphabets the situation becomes simpler, a rather surprising fact.

## 4 Extremal words over odd 2-letter alphabets

In this section, we assume that the alphabet is $\Sigma=\{a, b\}$, where $a \prec b$, with $a$ and $b$ odd. The minimal (resp. the maximal) infinite smooth word is denoted $m_{\{a, b\}}$ (resp. $M_{\{a, b\}}$ ). We start by a useful lemma.

Lemma 9. Let $\Sigma=\{a, b\}$, where $a \prec b$ are odd, and let $u \in \Sigma^{k}$. Then $\Phi^{-1}(u)$ is a palindrome of odd length.

Proof. Let $w=\Phi^{-1}(u)$. We proceed by induction on the length of $u$. If $n=|u|=1$ then $w=\beta \in \Sigma$, which is a palindrome. If $n=2$ then $u=\alpha \beta$, with $\alpha, \beta \in\{a, b\}$. Then both $w=\alpha^{\beta}$ and $\Delta(w)=\beta$ are palindromic, since $a$ and $b$ are odd. It follows that they have odd lengths. Assume now that the statement is true for every $u$ such that $|u| \leq k$. By hypothesis $w=w_{1} w_{2} \ldots w_{2 j+1}=\Phi^{-1}(u)$ is a palindrome of odd length, therefore $w=w^{\prime} \cdot w_{j+1} \cdot \widetilde{w^{\prime}}$ and

$$
\left.\Delta_{\alpha}^{-1}(w)=\Delta_{\alpha}^{-1}\left(w^{\prime}\right) \cdot \Delta_{\gamma}^{-1}\left(w_{j+1}\right) \cdot \widetilde{\Delta_{\alpha}^{-1}\left(w^{\prime}\right.}\right)
$$

where each factor is a palindrome of odd length.

Theorem 10. If $a$ and $b$ are odd, with $a \prec b$, then $\Phi\left(m_{\{a, b\}}\right)=(a b)^{\omega}$.
Proof. We proceed by induction on the length of the prefixes of $u=\Phi\left(m_{\{a, b\}}\right)$. First note that $m_{\{a, b\}}$ starts with the smallest letter, namely $a$. One can easily verify that $\Phi^{-1}(a b)=a^{b}<$ $a^{a} b \cdot w=\Phi^{-1}(a a x)$, for any $x$. Assume now that $\Phi^{-1}\left((a b)^{k}\right)$ is minimal, for every $k \leq n$.


Figure a

From Figure a, we deduce that the letter $x$ is the one that makes $\Phi^{-1}\left((b a)^{n-1} b x\right)$ the smallest, and from Figure b, the letter $x$ is the one that makes $\Phi^{-1}\left((a b)^{n-1} x\right)$ the smallest. By the induction hypothesis, we get $x=a$. It follows that if $\Phi^{-1}\left((a b)^{n}\right)$ is minimal, then $\Phi^{-1}\left((a b)^{n} a\right)$ is so.


Figure c shows that the next letter $y$ is the one that makes $\Phi^{-1}\left((b a)^{n} a \bar{y}\right)$ the smallest. Figure d describes that word. The letter $\bar{y}$ is such that $\Phi^{-1}\left((a b)^{n-1} a y\right)$ is the smallest. By the induction hypothesis, we obtain $y=b$ and the conclusion follows.

Indeed, we get free the computation of $\Phi$ for the maximal word:
Corollary 11. If $a$ and $b$ are odd, with $a \prec b$, then $\Phi\left(M_{\{a, b\}}\right)=b(b a)^{\omega}$.
The periodicity of $\Phi\left(m_{\{a, b\}}\right)$ yields a linear algorithm generating the minimal (therefore the maximal) infinite smooth word for odd alphabets:

Corollary 12. Let $\alpha \in \Sigma$. The following transducer computes $m_{\{a, b\}}$.

$$
\varepsilon / a^{b} \quad a /\left(b^{b} a^{b}\right)^{\frac{a-1}{2}} \quad \alpha /\left(b^{a} a^{a}\right)^{\frac{\alpha-1}{2}} b^{a}
$$

Two long standing conjectures of Dekking [8] concern, on one hand the closure of the set $F(K)$ of factors of the Kolakoski word by reversal, and on the other hand the recurrence of $K$. These conjectures were stated for every infinite smooth word over $\{1,2\}$ in [5]. Due to the special palindromic structure of smooth words on odd alphabets (see Lemma 9) we have the following positive answer.

Proposition 13. For every infinite smooth word $w$ over an odd 2-letter alphabet, the set $F(w)$ is closed under reversal. Moreover $w$ is recurrent.

Proof. Let $f$ be a factor of $w$. Then $w=u f v$ for some $u, f \in \Sigma^{*}$ and $v \in \Sigma^{\omega}$. Since every smooth word $w$ has, by Lemma 9, arbitrarily long palindromic prefixes, there is a palindromic prefix $p$ containing $u f$ and the result follows. For the recurrence property one extra step is necessary: take a longer palindromic prefix containing $p$.

The link between the existence of arbitrarily long palindromes and the recurrence property was first observed in [4] for smooth words over $\{1,2\}$. The proof above shows that this link also exists for arbitrary alphabets.

## Lyndon factorizations

We take now a closer look to the minimal words and start with a negative result.
Lemma 14. Let $\Sigma=\{a, b\}$, with $a, b$ odd, $a \prec b$ and $a \neq 1$. Then, the minimal infinite smooth word over $\Sigma$ is not an infinite Lyndon word.

Proof. Computing $\Phi^{-1}\left((a b)^{2}\right)$ of $m_{\{a, b\}}$, we get the prefix

$$
\left(\left(a^{b} b^{b}\right)^{\frac{a-1}{2}} a^{b}\left(b^{a} a^{a}\right)^{\frac{a-1}{2}} b^{a}\right)^{\frac{b-1}{2}}\left(a^{b} b^{b}\right)^{\frac{a-1}{2}} a^{b} .
$$

We can write $m_{\{a, b\}}=a^{b} b^{b} s$, with $s \in \Sigma^{\omega}$. A suffix of $m_{\{a, b\}}$ is $a^{b} b^{a} a s^{\prime}$, with $s^{\prime} \in \Sigma^{\omega}$. Then $a^{b} b^{b} s>a^{b} b^{a} a s^{\prime}$. Thus, $m_{\{a, b\}}$ is not an infinite Lyndon word.
In Lemma 14, we assumed $a \neq 1$ to ensure that the word was starting with $a^{b} b^{b}$. In the case $a=1$,the situation is different and we will establish that $m_{\{1, b\}}$ an infinite Lyndon word. Before proving that fact, some technical results are required.
Proposition 15. Let $\Sigma=\{1, b\}$, where $b>1$ is odd, and let $w_{2 k}=\Phi^{-1}\left((1 b)^{k}\right)$. Then, the following conditions hold
(i) $w_{2 k}=\left(w_{2 k-2} w_{2 k-3}\right)^{\frac{b-1}{2}} w_{2 k-2}$, for $k \geq 2$;
(ii) $w_{2 k+1}=\left(w_{2 k-1} w_{2 k-2}\right)^{\frac{b-1}{2}} w_{2 k-1}$, for $k \geq 2$.

Proof. (by induction on $k$ ) For $k=2$, we have $w_{1}=b, w_{2}=1^{b}, w_{3}=(b 1)^{\frac{b-1}{2}} b$ and $w_{4}=$ $\left(1^{b} b\right)^{\frac{b-1}{2}} 1^{b}$, and the property is verified. Assume now that $w_{2 k}=\left(w_{2 k-2} w_{2 k-3}\right)^{\frac{b-1}{2}} w_{2 k-2}$, for $k \leq n$. Then, since the function $\Delta^{-1}$ distributes nicely because all $w_{i}$ are palindromic of odd length by Lemma 9 , we have:

$$
\begin{aligned}
w_{2(n+1)} & =\Delta_{1}^{-1}\left(w_{2 n+1}\right)=\Delta_{1}^{-1} \circ \Delta_{b}^{-1}\left(w_{2 n}\right), \\
& =\Delta_{1}^{-1} \circ \Delta_{b}^{-1}\left(\left(w_{2 n-2} w_{2 n-3}\right)^{\frac{b-1}{2}} w_{2 n-2}\right), \\
& =\Delta_{1}^{-1}\left(\left(\Delta_{b}^{-1}\left(w_{2 n-2}\right) \Delta_{b}^{-1}\left(w_{2 n-3}\right)\right)^{\frac{b-1}{2}} \Delta_{b}^{-1}\left(w_{2 n-2}\right)\right), \\
& =\Delta_{1}^{-1}\left(\left(w_{2 n-1} w_{2 n-2}\right)^{\frac{b-1}{2}} w_{2 n-1}\right), \\
& =\left(\Delta_{1}^{-1}\left(w_{2 n-1}\right) \Delta_{1}^{-1}\left(w_{2 n-2}\right)\right)^{\frac{b-1}{2}} \Delta_{1}^{-1}\left(w_{2 n}\right), \\
& =\left(w_{2 n} w_{2 n-1}\right)^{\frac{b-1}{2}} w_{2 n-1},
\end{aligned}
$$

which completes the proof of (i). The proof of (ii) is similar.
Proposition 16. Let $\Sigma=\{1, b\}$, where $b>1$ is odd, and let $w_{2 k}=\Phi^{-1}\left((1 b)^{k}\right)$. Then $w_{2 k} w_{2 k-1}$ and $w_{2 k} w_{2 k+1}$ are Lyndon words.

Proof. (by induction on $k$ ) For $k=1$, we get $w_{1}=b, w_{2}=1^{b}$ and $w_{3}=(b 1)^{\frac{b-1}{2}} b$. Then, $w_{2} w_{1}=1^{b} b$ and $w_{2} w_{3}=1^{b}(b 1)^{\frac{b-1}{2}} b$ are Lyndon words. Assume now that $w_{2 k} w_{2 k-1}$ and $w_{2 k} w_{2 k+1}$ are Lyndon words for every $k \leq n$. We first state an obvious but useful property of Lyndon words:

Lemma 17. If $u=p s$, where $p$ and $s$ are non-empty, is a finite Lyndon word, then so are $p^{j} u$ and $u s^{j}$.
(i) $w_{2 n+2} w_{2 n+1}=\left(w_{2 n} w_{2 n-1}\right)^{\frac{b-1}{2}} \cdot w_{2 n} w_{2 n+1}$, by Proposition 15. Then, using the induction hypothesis, $w_{2 n} w_{2 n-1}$ and $w_{2 n} w_{2 n+1}$ are Lyndon words, so that $w_{2 n+2} w_{2 n+1}=u^{\frac{b-1}{2}} v$, where $u$, $v$ are Lyndon words and $u$ prefix of $v$. Now Lemma 17 applies, which concludes.
(ii) $w_{2 n+2} w_{2 n+3}=w_{2 n+2} w_{2 n+1} \cdot\left(w_{2 n} w_{2 n+1}\right)^{\frac{b-1}{2}}$, by Proposition 15. Then, using (i) and the induction hypothesis, we deduce that $w_{2 n+2} w_{2 n+1}$ and $w_{2 n} w_{2 n+1}$ are Lyndon words. Then, $w_{2 n+2} w_{2 n+3}=u v^{\frac{b-1}{2}}$, where $u$ and $v$ are Lyndon words, $v$ is a suffix of $u$. Again Lemma 17 permits to conclude.

Proposition 18. Let $\Sigma=\{1, b\}$, where $b>1$ is odd, and let $w_{2 k}=\Phi^{-1}\left((1 b)^{k}\right)$. Let $L_{n}$ be the Lyndon factorization of $w_{n}$ for $n \leq k \in \mathbb{N}$. Then for $n \geq 2$,
(i) $L_{2 n}=\left(\bigodot_{i=1}^{\frac{b-1}{2}} w_{2 n-2} w_{2 n-3}\right) \cdot L_{2 n-2}$;
(ii) $L_{2 n+1}=L_{2 n-1} \cdot\left(\bigodot_{i=1}^{\frac{b-1}{2}} w_{2 n-2} w_{2 n-1}\right)$.

Proof. (by induction on $n$ ) For $n=2, w_{1}=b, w_{2}=1^{b}$, $w_{3}=(b 1)^{\frac{b-1}{2}} b, w_{4}=\left(1^{b} b\right)^{\frac{b-1}{2}} 1^{b}$ and $w_{5}=$ $\left((b 1)^{\frac{b-1}{2}} b 1^{b}\right)^{\frac{b-1}{2}} b$. Then, the corresponding Lyndon factorizations are: $L_{1}=b, L_{2}=1 \cdot 1 \cdots \cdots 1$, $L_{3}=b \cdot(1 b) \cdot(1 b) \cdots \cdot(1 b), L_{4}=\left(1^{b} b\right) \cdot\left(1^{b} b\right) \cdots \cdot\left(1^{b} b\right) \cdot 1 \cdot 1 \cdots \cdot 1$ and

$$
L_{5}=b \cdot(1 b) \cdot(1 b) \cdots \cdots(1 b) \cdot\left(\bigodot_{i=1}^{\frac{b-1}{2}} 1^{b}(b 1)^{\frac{b-1}{2}} b\right) .
$$

Both (i) and (ii) are verified. Assume now that (i) and (ii) hold for every $i \leq n$.
(i) By Proposition 15, $w_{2 n+2}=\left(w_{2 n} w_{2 n-1}\right)^{\frac{b-1}{2}} w_{2 n}$. Using the fact that $w_{2 n} w_{2 n-1}$ is a Lyndon word and that $w_{2 n}$ is a proper prefix, we deduce the Lyndon factorization $L_{2 n+2}$.
(ii) Again by Proposition 15, $w_{2 n+3}=w_{2 n+1}\left(w_{2 n} w_{2 n+1}\right)^{\frac{b-1}{2}}$. Using the fact that $w_{2 n} w_{2 n+1}$ is a Lyndon word having $w_{2 n+1}$ as a suffix, we deduce the Lyndon factorization $L_{2 n+3}$.
We are now in a position to state the main result about the Lyndon factorization of the minimal infinite smooth word $m_{\{1, b\}}$.

Theorem 19. Let $\Sigma=\{1, b\}$, where $b>1$ is odd. Then
(i) $m_{\{1, b\}}$ is an infinite Lyndon word;
(ii) the Lyndon factorization of $\Delta\left(m_{\{1, b\}}\right)$ is an infinite sequence of finite Lyndon words.

Proof. It suffices to take the limit as $n \rightarrow \infty$ of the statements in Proposition 18.

## 5 Extremal words over even 2-letter alphabets

In this section, we assume that the letters of $\Sigma=\{a, b\}$ are even, with $a<b$. We only state the results without proofs, they will appear in an extended version of this paper.

Lemma 20. Let $w \in\{a, b\}^{*}$ be a smooth word, $u=\Phi(w)$ and $|u|=n$. Then for all $i \leq n-2$, $\Delta^{i}(w)$ is a word of even length.

In the previous section, Theorem 10 states that $\Phi\left(m_{\{a, b\}}\right)=(a b)^{\omega}$ for $a, b$ odd. For an even alphabet there is an analogous result.

Theorem 21. If $a$ and $b$ are even, with $a \prec b$, then $\Phi\left(M_{\{a, b\}}\right)=b^{\omega}$.
Using the fact that $\Delta\left(m_{\{a, b\}}\right)=\Delta\left(M_{\{a, b\}}\right)$, we get:
Corollary 22. If $a$ and $b$ are even, with $a \prec b$, then $\Phi\left(m_{\{a, b\}}\right)=a b^{\omega}$.
Therefore, $M_{\{a, b\}}$ is equal to $\Delta\left(m_{\{a, b\}}\right)$ and is the generalized Kolakoski word $K_{(b, a)}$, for being a fixpoint of the function $\Delta$ over the even alphabet $\{a, b\}$. This last property yields a linear algorithm generating prefixes of the minimal (hence the maximal) infinite smooth word for an even alphabet, represented by the following transducer, where $\alpha \in\{a, b\}$.

$$
\varepsilon / a^{b-1} \quad a a / a b^{b} \quad \alpha \alpha / a^{\bar{\alpha}} b^{\bar{\alpha}}
$$

Remark 23. This transducer has two cycles (one for each letter) with same base state, and therefore any infinite path runs through these two cycles. Since an equal number of $a$ 's and $b$ 's are written in each cycle, the frequency of both letters is $\frac{1}{2}$. This again is a surprising fact: indeed, for the well-known Kolakoski word $K_{(1,2)}$ it is still a challenging conjecture. Indeed, the best known result is 0.50084 and is due to Chvátal [7], who designed an ingenious procedure for computing an approximation of the density.
In the case of an odd alphabet we uncovered in Lemma 9 the palindromic structure of the prefixes for any smooth infinite word $w$. This characterization does not hold for an even alphabet, but we can state:

Lemma 24. Let $\Sigma=\{a, b\}$, where $a \prec b$ are even, and let $u \in \Sigma^{k}$. Then for all $i \leq k-3$,

$$
\Delta^{i}\left(\Phi^{-1}(u)\right)=p^{u[k] / 2}
$$

for some $p \in \Sigma^{*}$.
It follows from Lemma 24 that every infinite smooth word over an even 2-letter alphabet is recurrent. On the other hand, we have:

Proposition 25. The set of factors of the extremal infinite smooth words over an even 2-letter alphabet is not closed under reversal.

The study of the Lyndon factorization of the minimal smooth words over an even alphabet leads to:

Theorem 26. For an even 2-letter alphabet $\{a, b\}$, with $a \prec b$, the minimal smooth word $m_{\{a, b\}}$ is an infinite Lyndon word.

An easy consequence is that the complement of the generalized Kolakoski word $K_{(b, a)}$, with $a \prec b$, is an infinite Lyndon word.

## 6 Concluding remarks

The density of letters in an infinite smooth word over $\{1,2\}$ is a still unsolved conjecture. Nevertheless for even alphabets this density is 0.5 for the extremal words. For odd alphabets of the type $\{1, b\}$, the inductive formulas in Proposition 15 enable us to compute the density for extremal words. Indeed, the density of the letter $b$ is

$$
1 /(\sqrt{2 b-1}-1)
$$

Proofs are omitted for lack of space and will appear in a full paper.
Moreover, the work presented here raises a number of questions. It is quite surprising that for alphabets of same parity, some of the Dekking conjectures are rather easy to prove: recurrence, density for extremal words on even alphabets, closure by reversal for odd alphabets. The density problem remains open for odd alphabets, as well as all the conjectures for the alphabet $\{1,2\}$, an instance of a different parities alphabet. The results presented here beg for an investigation of smooth words on different parities: study of the extremal words, combinatorial properties, Lyndon factorizations, closure properties, and so on. In another direction it would be interesting to compute the complexity function $P(n)$ in the way Weakley did for the alphabet $\{1,2\}$. The case of larger $k$-letter alphabets is also challenging.

## References

[1] A. Bergeron-Brlek, S. Brlek, A. Lacasse, X. Provençal, Patterns in Smooth Tilings, Proc. WORDS'03, 4-th International Conference on Words Turku (Finland) September 10-13 2003 TUCS General Publication No 27, August (2003), 370-381.
[2] V. Berthé, S. Brlek, P. Choquette, Smooth words over arbitrary alphabets, Theoretical Computer Science 341 (2005) 293-310.
[3] S. Brlek, Enumeration of factors in the Thue-Morse word, Discrete Applied Mathematics 24 (1989), 83-96.
[4] S. Brlek, A. Ladouceur, A note on differentiable palindromes, Theoretical Computer Science 302 (2003), 167-178.
[5] S. Brlek, S. Duluce, A. Ladouceur, L. Vuillon, Combinatorial properties of smooth infinite words, Theoretical Computer Science 352 (2006) 306-317.
[6] A. Carpi, On repeated factors in $\mathrm{C}^{\infty}$-words, Information Processing Letters, 52 (1994) 289-294.
[7] V. ChVÁtal, Notes on the Kolakoski Sequence, DIMACS Techn. Rep. 93-84, February 1994.
[8] F. M. Dekking, On the structure of self generating sequences, Séminaire de théorie des nombres de Bordeaux, exposé 31 (1980-1981).
[9] D. Jamet, G. Paquin, Discrete surfaces and infinite smooth words, FPSAC'05-17th annual International conference on Formal Power Series and Algebraic Combinatorics, Taormina (Sicile) June 20-25 2005.
[10] W. Kolakoski, Self Generating Runs, Problem 5304, American Mathematical Monthly, 72 (1965), 674. Solution: American Math. Monthly 73 (1966), 681-682.
[11] M. Lothaire, Combinatorics on words, Addison Wesley, Reading MA,1983.
[12] G.Paquin, G. Melançon, S. Brlek, Properties of the extremal infinite smooth words, Journal of Automata, Languages and Combinatorics, to appear.
[13] R. Siromoney, L. Matthew, V. R. Dare, K. G. Subramanian, Infinite lyndon words, Information Processing Letters 50 (1994), 101-104.
[14] W. D. Weakley, On the number of $C^{\infty}$-words of each length, Journal Combinatorial Theory Series A 51 no. 1 (1989), 55-62.


[^0]:    * with the support of NSERC (Canada)
    ${ }^{\dagger}$ Corresponding author
    ${ }^{\ddagger}$ LaCIM, Université du Québec à Montréal, CP 8888 Succ. Centre-Ville, Montréal (QC) CANADA H3C 3P8
    ${ }^{\S}$ LORIA - INRIA Lorraine, 615 rue du Jardin Botanique - BP 101, F-54602 Villers-lès-Nancy

