k-overlap-free binary words

Patrice Séébold

k-overlap-free binary words

Patrice Séébold *
LIRMM, UMR 5506 CNRS
161 rue Ada
34392 Montpellier, FRANCE
e-mail: Patrice.Seebold@lirmm.fr

August 29, 2010

Abstract

We study k-overlap-free binary infinite words that are binary infinite words which can contain only overlaps $xyxyx$ with $|x| \leq k-1$. We prove that no such word can be generated by morphism, except if $k=1$. On the other hand, for every $k \geq 2$, there exist k-overlap-free binary infinite words which are not $(k-1)$-overlap-free. As an application, we prove that, for every integer n, there exists infinitely many k-overlap-free binary infinite partial words with n holes.

1 Introduction

Repetitions, i.e., consecutive occurrences of a given factor within a word, and especially repetition-freeness have been fundamental research subjects in combinatorics on words since the seminal papers of Thue [11, 12] in the beginning of the 20th century (see also [2]). In particular, Thue and Morse independently showed the existence of an overlap-free binary infinite word (the Thue-Morse word [8, 12]), i.e., an infinite word using only two different letters and which does not contain any factor $xyxyx$ with x a non-empty word.

In the present paper we study the case where x must be of length at least k, that is, k-overlap-free binary infinite words which do not contain any factor $xyxyx$ with $|x| \geq k$.

The paper is organized as follows. After general definitions and notations given in Section 2, the notion of k-overlap-freeness is introduced in Section 3 where it is proved that no k-overlap-free binary infinite word can be generated by morphism, except if $k=1$. In Section 4 we introduce the concept of 0-limited square property (a word has this property if the squares it contains have a particular form) to prove that, for every integer k, there exist k-overlap-free binary infinite words that are not $(k-1)$-overlap-free. In Section 5 we consider the particular case of k-overlap-free words which do not contain cubes of some letters. Section 6 is then dedicated to an application to partial words[1].

2 Preliminaries

Generalities on combinatorics on words can be found, e.g., in [7].

Let $\mathcal{A}$ be a finite alphabet. The elements of $\mathcal{A}$ are called *letters*. A word $w = a_1a_2\cdots a_n$ of length n over the alphabet $\mathcal{A}$ is a mapping $w\colon \{1,2,\ldots,n\} \to \mathcal{A}$ such that $w(i) = a_i$. The length of a word w is denoted by $|w|$, and ε is the empty word of length zero. For a word w and a letter a, $|w|_a$ denotes the number of occurrences of the letter a in the word w. By a (right) infinite word $w = a_1a_2a_3\cdots$ we mean a mapping w from the positive integers $\mathbb{N}_+$ to the alphabet $\mathcal{A}$ such that $w(i) = a_i$. The set of all finite words is denoted by $\mathcal{A}^*$, infinite words are denoted by $\mathcal{A}^\omega$ and $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. A finite word v is a *factor*

*Département Mathématiques Informatique et Applications, Université Paul Valéry, Route de Mende, 34199 Montpellier Cedex 5, France

[1]A preliminary version of this paper was presented at JORCAD'08 [10]. Some results about the case $k=2$ appeared in [6]. But in these two papers, the 0-limited square property was replaced by the restricted square property, a much more restrictive condition.

of $w \in \mathcal{A}^* \cup \mathcal{A}^\omega$ if $w = xvy$, where $x \in \mathcal{A}^*$ and $y \in \mathcal{A}^* \cup \mathcal{A}^\omega$. Words xv and vy are respectively called *a prefix* and *a suffix* of w.

A morphism on $\mathcal{A}^*$ is a mapping $f\colon \mathcal{A}^* \to \mathcal{A}^*$ satisfying $f(xy) = f(x)f(y)$ for all $x, y \in \mathcal{A}^*$. The morphism f is *erasing* if there exists $a \in \mathcal{A}$ such that $f(a) = \varepsilon$. Note that f is completely defined by the values $f(a)$ for every letter a on $\mathcal{A}$. A morphism is called *prolongable on a letter* a if $f(a) = aw$ for some word $w \in \mathcal{A}^+$ such that $f^n(w) \neq \varepsilon$ for all integers $n \geq 1$. This implies that $f^n(a)$ is a prefix of $f^{n+1}(a)$ for all integers $n \geq 0$ and a is a *growing letter for* f, that is, $|f^n(a)| < |f^{n+1}(a)|$ for every $n \in \mathbb{N}$. Consequently, the sequence $(f^n(a))_{n\geq 0}$ converges to the unique infinite word generated by f from the letter a,

$$f^\omega(a) := \lim_{n\to\infty} f^n(a) = awf(w)f^2(w)\cdots,$$

which is a fixed point of f.

A morphism $f\colon \mathcal{A}^* \to \mathcal{A}^*$ *generates an infinite word* w *from a letter* $a \in \mathcal{A}$ if there exists $p \in \mathbb{N}$ such that the morphism f^p is prolongable on a. We say that the morphism f *generates an infinite word* if it generates an infinite word from at least one letter.

A kth *power* of a word $u \neq \varepsilon$ is the word u^k prefix of length $k \cdot |u|$ of u^ω, where u^ω denotes the infinite catenation of the word u, and k is a rational number such that $k \cdot |u|$ is an integer. A word w is called k*-free* if there does not exist a word x such that x^k is a factor of w. If $k = 2$ or $k = 3$, then we talk about *square-free* or *cube-free* words, respectively. An *overlap* is a word of the form $xyxyx$ where $x \in \mathcal{A}^+$ and $y \in \mathcal{A}^*$. A word is called *overlap-free* if it does not contain overlaps. Therefore, it can contain squares but it cannot contain any longer repetitions such as overlaps or cubes. For example, over the alphabet $\{a, b\}$ the word *abbabaa* is overlap-free but it contains squares *bb*, *aa*, and *baba*. It is easy to verify that there does not exist a square-free infinite word over a binary alphabet, but as we recall in the next section there exist overlap-free binary infinite words.

In all the paper we will use the two alphabets $\mathcal{A} = \{a, b\}$, $\mathcal{B} = \{0, 1, 2\}$.

3 k-overlap-free binary words

In [12], Thue introduced the morphism

$$\begin{array}{cccc} \mu: & \mathcal{A}^* & \to & \mathcal{A}^* \\ & a & \mapsto & ab \\ & b & \mapsto & ba \end{array}$$

The Thue-Morse word is the overlap-free binary infinite word

$$t := \lim_{n\to\infty} \mu^n(a) = abbabaabbaababbaba\cdots$$

generated by μ from the letter a (see, e.g., [1] for other definitions and properties, see also [2] for a translation of the contribution of Thue to the combinatorics on words). Another overlap-free binary infinite word is t', the word generated by the morphism μ from the letter b. Note that the word t' can be obtained from the word t by exchanging all the a's and b's.

We generalize the notion of overlap with the following definition.

Definition 3.1 *A* k-overlap *is a word of the form* $xyxyx$ *where* x *and* y *are two words with* $|x| = k$. *A word is* k-overlap-free[2] *if it does not contain* k*-overlaps.*

For example, the word *baabaab* is not overlap-free but it is 2-overlap-free while the word *baabaaba* is not.

It is important to note that a k-overlap-free word can contain a^{3k-1} for each letter a. More generally, a k-overlap-free word can contain ℓ-powers u^ℓ where the value of ℓ, which can be greater than k, depends on the word u. For example, the word $u = ababababab$, which is a 5-power, is 3-overlap-free. On the contrary, the word $v = abaabaaba$, which is only a 3-power and which is also such that $|v| < |u|$, is a

[2] While it is not exactly the same, this notion of k-overlap-freeness ressembles that of k-bounded overlaps introduced by Thue in [12].

3-overlap thus not being 3-overlap-free! This peculiarity is one reason for restraining the definition, for example to the case of cube-free words. However, such a restriction seems to be very drastic and it is generally enough to avoid the powers of letters. In the present paper, in Section 5, we study the change in our results when restraining to the case of words without a^3 ... and we will see that the results are not very different.

By definition, it is evident that every k-overlap-free word is also k'-overlap-free for $k' \geq k$. Note that a word is 1-overlap-free if and only if it is overlap-free. So, an overlap-free infinite word is a k-overlap-free infinite word for every positive integer k.

It is a well-known important property that t and t' are the only overlap-free binary infinite words which are generated by morphism (see, e.g., [5], [9]). Since k-overlap-freeness does not imply ℓ-overlap-freeness for $\ell < k$, k-overlap-freeness is weaker than overlap-freeness when $k \geq 2$. Therefore, we might suppose that there exist binary infinite words, generated by morphism, that are k-overlap-free for some $k \geq 2$ but that are not overlap-free (therefore different from t and t'). In fact, rather surprisingly, for that property, k-overlap-freeness does not give more than only (1)-overlap-freeness.

Theorem 3.2 *Let $k \in \mathbb{N}_+$ and let w be a k-overlap-free binary infinite word. Then w is generated by a morphism if and only if $w = t$ or $w = t'$.*

Proof. The only if part is obvious since $t = \mu^\omega(a)$ and $t' = \mu^\omega(b)$ are k-overlap-free for every positive integer k.

Conversely, as we have already seen, if an overlap-free ($k = 1$) binary infinite word is generated by a morphism then $w = t$ or $w = t'$. Thus it remains to prove that an infinite word which contains an overlap but is k-overlap-free for some integer $k \geq 2$ cannot be generated by a morphism.

Assume, contrary to what we want to prove, that an infinite word w over $\mathcal{A}$ which contains an overlap but is k-overlap-free for some integer $k \geq 2$, is generated by a morphism f. Then there exists a positive integer r such that f^r is prolongable on the first letter of w. Without loss of generality we may assume that this first letter is the letter a. In particular, w begins with $(f^r)^n(a)$ for every $n \in \mathbb{N}$. Moreover, by definition a is a growing letter for f^r, which implies that there exists a positive integer N_0 such that $|f^{rN_0}(a)| \geq k$.

If $f(b) = \varepsilon$ or if $|f(a)|_b = 0$ (which means $f(a) = a^p$, $p \geq 2$, because a is a growing letter for f), then w is the periodic word $(f(a))^\omega$ which contains arbitrarily large powers of $f(a)$.

If $f(b) = b^p$, $p \geq 2$, or if $f^r(a)$ ends with b and $f(b) = b$ (which implies that $f^{rn}(a)$ ends with b^n for every integer n), then w contains arbitrarily large powers of b.

If $f(a)$ ends with a and begins with ab^pa, $p \geq 1$, and $f(b) = b$, then w, which begins with $(f^r)^2(a)$, contains a factor $auaua$ with $u = b^p$.

If w contains an overlap $auaua$ as a factor, then $f^{rN_0}(auaua)$ is also a factor of w.

The only remaining case is w contains an overlap $bubub$ as a factor and $f(b) = xay$ for some words $x, y \in \mathcal{A}^*$. Then $f^{rN_0}(bubub)$ contains the factor $f^{rN_0}(a)f^{rN_0}(yux)f^{rN_0}(a)f^{rN_0}(yux)f^{rN_0}(a)$.

Consequently, since $f^{rN_0}(w) = w$ (because $w = (f^r)^\omega(a)$), in all cases w contains a k-overlap, which contradicts with the hypothesis. ■

Remark. Although the case of alphabets with more than two letters is out of the scope of the present paper, one can notice that Theorem 3.2 is no more true if we consider larger alphabets. Indeed, over a 3-letter alphabet it is possible, for every integer $k \geq 2$, to find k-overlap-free words that are not $(k-1)$-overlap-free and that are generated by morphism.

For example, let us consider the morphism

$$\begin{array}{llll} \mu_c: & (\mathcal{A} \cup \{c\})^* & \to & (\mathcal{A} \cup \{c\})^* \\ & a & \mapsto & ac^{3(k-1)}b \\ & b & \mapsto & ba \\ & c & \mapsto & c \end{array}$$

This morphism is obtained from the morphism μ by adding $c^{3(k-1)}$ at the middle of $\mu(a)$. Since μ is an overlap-free morphism, the only overlaps in the word $\mu_c^\omega(a)$ are powers of the letter c. Therefore the word $\mu_c^\omega(a)$ is k-overlap-free but not $(k-1)$-overlap-free.

4 The 0-limited square property

We have seen with Theorem 3.2 that the Thue-Morse words t and t' are the only k-overlap-free binary infinite words generated by morphism, whatever be the value of k. So it is natural to ask about the existence of k-overlap-free binary infinite words with $k \geq 2$ (then, of course, not generated by morphism), which are not ℓ-overlap-free for $\ell < k$. The answer is given in the present section where a family of such words is characterized.

Before this, we have to recall some works of Thue.

In order to prove the existence of infinite cube-free words over a two-letter alphabet from square-free words over three letters, Thue used in [11] the application

$$\begin{array}{rccl} \delta: & \mathcal{B}^* & \rightarrow & \mathcal{A}^* \\ & 0 & \mapsto & a \\ & 1 & \mapsto & ab \\ & 2 & \mapsto & abb \end{array}$$

Six years later he proved the following result.

Proposition 4.1 *[12] [2] Let $u \in \mathcal{A}^\omega$ and $v \in \mathcal{B}^\omega$ be such that $\delta(v) = u$. The word u is overlap-free if and only if the word v is square-free and does not contain* 010 *nor* 212 *as a factor.* ■

Thue also remarked that if the word $\delta(w)$ is not overlap-free for a square-free word w (thus containing 010 or 212) then every overlap $xyxyx$ in $\delta(w)$ is such that x is a single letter. Therefore, it suffices to prove the existence of a square-free ternary infinite word containing either 010 or 212 over $\mathcal{B}$ to obtain a 2-overlap-free binary infinite word that is not overlap-free. Here again such a word is found in [12].

Let τ be the morphism

$$\begin{array}{rccl} \tau: & \mathcal{B}^* & \rightarrow & \mathcal{B}^* \\ & 0 & \mapsto & 01201 \\ & 1 & \mapsto & 020121 \\ & 2 & \mapsto & 0212021 \end{array}$$

Proposition 4.2 *[12] The word $\tau^\omega(0)$ is square-free and it contains* 212 *as a factor.* ■

Now, to prove the existence, for every integer $k \geq 2$, of k-overlap-free binary infinite words that are not $(k-1)$-overlap-free, we generalize Thue's idea with the following notion.

Definition 4.3 *An infinite word v over $\mathcal{B}$ has the* 0-limited square property *if*

- *the word v does not contain* 00 *as a factor,*
- *whenever v contains a non-empty square rr as a factor, then, in v, the factor rr is preceded (if it is not a prefix of v) and followed by the letter 0.*

Note that if a word $v \in \mathcal{B}^\omega$ has the 0-limited square property then v is overlap-free and if v contains a non-empty square rr as a factor, the word r does not begin nor end with the letter 0.

The following corollary is straightforward from Proposition 4.2 because each square-free word obviously has the 0-limited square property.

Corollary 4.4 *The word $\tau^\omega(0)$ has the* 0*-limited square property.* ■

Now, let k, p be two integers with $k \geq 2$ and $1 \leq p \leq k-1$. We associate to (k, p) the application

$$\begin{array}{rccl} \delta_{k,p}: & \mathcal{B}^* & \rightarrow & \mathcal{A}^* \\ & 0 & \mapsto & a^{k-p} \\ & 1 & \mapsto & a^{k-p}b^p \\ & 2 & \mapsto & a^{k-p}b^{p+1} \end{array}$$

Of course, $\delta_{2,1} = \delta$ thus our affirming that this is a generalization of Thue's idea.

Theorem 4.5 *Let $u \in \mathcal{A}^\omega$ and $v \in \mathcal{B}^\omega$ be such that $\delta_{k,p}(v) = u$. If the word v has the 0-limited square property then the word u is k-overlap-free.*

Proof. Suppose that u is not k-overlap-free. Since $u = \delta_{k,p}(v)$, the following cases are possible:

- u contains a factor $a^k x a^k x a^k$

  If $|x|_b = 0$, or if $x = zx'$ with $|z| \geq k - 2p + 1$ and $|z|_b = 0$ then u contains $a^{k-p}a^{k-p}a$ which means that v contains 00.

  Henceforth, u contains a factor $a^n a^k a^m x' a^k a^m x' a^k$ with $n+m+k = 2(k-p)$, i.e., $k-p = n+m+p$, and x' begins with the letter b. Therefore, u contains a factor $a^{n+m+p}a^{k-p}x'a^{p+m}a^{k-p}x'a^{p+m}a^{p+n}$, which implies that v contains a factor $0yy$ where y is such that $\delta_{k,p}(y) = a^{k-p}x'a^{p+m}$ (in particular, $y \neq \varepsilon$). But in this case, y necessarily ends with 0 because $p + m \geq p \geq 1$. Therefore, either yy is followed by the letter 0 implying that v contains 00 as a factor, or yy is not followed by the letter 0.

- u contains a factor $a^n b^{p+1} a^m x a^n b^{p+1} a^m x a^n b^{p+1} a^m$ with $n + m = k - p - 1$ (this includes the case where u contains $b^k x b^k x b^k$ when $p = k - 1$)

  In this case, v contains a factor $2y2y2$ with $\delta_{k,p}(y2) = a^m x a^n b^{p+1}$.

- u contains a factor $a^n b^p a^m x a^n b^p a^m x a^n b^p a^m$ with $n + m = k - p$

  Here, two cases are possible.

  1. $m \neq 0$
     Then u contains a factor $b^p a^m x a^n b^p a^m x a^n b^p a^{k-p}$, which implies that v contains a square yy, preceded by 1 or 2, with $\delta_{k,p}(y) = a^m x a^n b^p$.
  2. $m = 0$ (then $n = k - p$)
     Then u contains a factor $a^{k-p}b^p x a^{k-p} b^p x a^{k-p} b^p$, which implies that v contains a square yy, followed by 1 or 2, with $\delta_{k,p}(y) = a^{k-p} b^p x$.

- u contains a factor $b^n a^{k-p} b^m x b^n a^{k-p} b^m x b^n a^{k-p} b^m$ with $n + m = p$

  Here again, two cases are possible.

  1. $m \neq 0$
     Then u contains a factor $a^{k-p} b^m x b^n a^{k-p} b^m x b^n a^{k-p} b^p$, which implies that v contains a square yy, followed by 1 or 2, with $\delta_{k,p}(y) = a^{k-p} b^m x b^n$.
  2. $m = 0$ (then $n = p$)
     Then u contains a factor $b^p a^{k-p} x b^p a^{k-p} x b^p a^{k-p}$, which implies that v contains a square yy, preceded by 1 or 2, with $\delta_{k,p}(y) = a^{k-p} x b^p$.

In all the cases v has not the 0-limited square property. ■

Conditions given in Definition 4.3 are not sufficient to guarantee that the word v has the 0-limited square property when $u = \delta_{k,p}(v)$ is k-overlap-free. For example, the word $v = 0\tau^\omega(0)$ contains only one square, the factor 00 which v begins with. But, since the word $\tau^\omega(0)$ has the 0-limited square property, the word $\delta_{k,p}(\tau^\omega(0))$ is k-overlap-free which implies that $\delta_{k,p}(v)$ is also k-overlap-free (otherwise, $\delta_{k,p}(v)$ begins with a k-overlap whose prefix is $a^{k-p}a^{k-p}a^{k-p}$, implying that $\delta_{k,p}(\tau^\omega(0))$ contains an occurrence of this factor $a^{k-p}a^{k-p}a^{k-p}$ from which $\tau^\omega(0)$ contains 00, a contradiction). However, it is possible to obtain an equivalence by giving conditions on the words u and v.

Corollary 4.6 *Let u, an infinite word over $\mathcal{A}$, which does not contain the factor $a^{2(k-p)+1}$, and v, an infinite word over $\mathcal{B}$ which does not begin with a square, be such that $\delta_{k,p}(v) = u$. The word u is k-overlap-free if and only if the word v has the 0-limited square property.*

Remark that here the word u is also $[2(k-p)+1]$-free.

Proof. Let u and v be as in the statement. It is of course equivalent that u does not contain the factor $a^{2(k-p)+1}$ and v does not contain the factor 00, thus our assuming that 00 is not a factor of v.

From Theorem 4.5, it suffices to prove the necessary condition.

Let rr be a factor of v with $r \neq \varepsilon$. According to the hypothesis, rr is not at the beginning of v which means that in v, rr is preceded (and followed) by at least one letter.

- If r begins with the letter 0 then, since 00 is not a factor of v, r does not end with 0. Thus $\delta_{k,p}(r) = a^{k-p}sb^p$. For the same reason, rr is preceded by the letter 1 or by the letter 2, so $\delta_{k,p}(rr)$ is preceded by b^p. Whatever be the letter following rr, $\delta_{k,p}(rr)$ is followed by a^{k-p}. Consequently, u contains the factor $b^p\delta_{k,p}(rr)a^{k-p} = b^pa^{k-p}sb^pa^{k-p}sb^pa^{k-p}$, a k-overlap. This implies that u is not k-overlap-free.

- If r ends with the letter 0 then, since v does not contain 00 as a factor, r does not begin with 0, which implies that $\delta_{k,p}(r)$ begins with $a^{k-p}b^p$. Moreover, in v, the factor rr is followed either by 1 or by 2. Then u contains the factor $\delta_{k,p}(rr)a^{k-p}b^p = a^{k-p}b^psa^{k-p}b^psa^{k-p}b^p$, a k-overlap. This implies that u is not k-overlap-free.

- Now if r begins with 1 or 2, and rr is not followed by 0 then $\delta_{k,p}(r)$ begins with $a^{k-p}b^p$ and $\delta_{k,p}(rr)$ is followed by $a^{k-p}b^p$, which means that u is not k-overlap-free.

- Finally, if r ends with 1 or 2, and rr is not preceded by 0 then $\delta_{k,p}(r)$ begins with a^{k-p} and ends with b^p, and $\delta_{k,p}(rr)$ is preceded by b^p. This implies that, since $\delta_{k,p}(rr)$ is followed by a^{k-p}, u is not k-overlap-free.

Consequently, if u is k-overlap-free then v has the 0-limited square property. ■

Theorem 4.5 gives the first part of the answer to the question we asked at the beginning of this section by showing the existence of k-overlap-free binary infinite words for every integer $k \geq 2$. It remains to prove that some words u satisfying Theorem 4.5 can effectively be constructed containing $(k-1)$-overlaps. This is done by using again Thue's morphism τ.

Proposition 4.7 *For every integer $k \geq 2$, the word $\delta_{k,k-1}(\tau^\omega(0))$ is k-overlap-free but it contains $(k-1)$-overlaps.*

Proof. Since from Corollary 4.4 the word $\tau^\omega(0)$ has the 0-limited square property, the word $\delta_{k,k-1}(\tau^\omega(0))$ is k-overlap-free from Theorem 4.5.

Now, we know from Proposition 4.2 that $\tau^\omega(0)$ contains 212 as a factor. Therefore, $\delta_{k,k-1}(\tau^\omega(0))$ contains the factor $\delta_{k,k-1}(212) = ab^kab^{k-1}ab^k$, which implies that the $(k-1)$-overlap $b^{k-1}ab^{k-1}ab^{k-1}$ is a factor of $\delta_{k,k-1}(\tau^\omega(0))$. ■

5 Strongly k-overlap-free binary words

A k-overlap-free binary infinite word must contain occurrences of a^2 or b^2 (or both). For if it were not the case the word would be $(ab)^\omega$ or $(ba)^\omega$ which obviously contains k-overlaps for every $k \in \mathbb{N}$.

As mentioned after Definition 3.1, the particular case of k-overlap-free binary infinite words without x^3 for some letter $x \in \mathcal{A}$ is of interest. We define such words as follows.

Definition 5.1 *A word over $\mathcal{A}$ is* x-strongly k-overlap-free *if it is k-overlap-free and if it does not contain x^3, where x is a letter. A word is* strongly k-overlap-free *if it is x-strongly k-overlap-free for every letter $x \in \mathcal{A}$.*

For example, the word a^5 is 2-overlap-free; it is b-strongly 2-overlap-free, but it is not strongly 2-overlap-free because it is not a-strongly 2-overlap-free.

Notice that there exists effectively strongly k-overlap-free words that are not $(k-1)$-overlap-free: for example, from Proposition 4.2, the word $\delta(\tau^\omega(0))$ is strongly 2-overlap-free without being overlap-free.

Since every strongly k-overlap-free binary infinite word is k-overlap-free and since the Thue-Morse words t and t' are cube-free, Theorem 3.2 remains true in the present case.

Theorem 5.2 *Let $k \in \mathbb{N}_+$ and let w be a strongly k-overlap-free binary infinite word. Then w is generated by a morphism if and only if $w = t$ or $w = t'$.*

It is obvious that $\mu(u)$ does not contain neither a^3 nor b^3, whatever be the value of u. This implies that if $\mu(u)$ is a k-overlap-free binary infinite word then it is indeed strongly k-overlap-free. Let us recall the two lemmas used by Thue to prove that the Thue-Morse word t is overlap-free.

Lemma 5.3 *Let $\Sigma = \{ab, ba\}$. If $u \in \Sigma^*$ then $aua \notin \Sigma^*$ and $bub \notin \Sigma^*$.*

Lemma 5.4 *A word $u \in \mathcal{A}^* \cup \mathcal{A}^\omega$ is overlap-free if and only if the word $\mu(u)$ is overlap-free.*

The following result is an extension of Lemma 5.4.

Proposition 5.5 *Let $w \in \mathcal{A}^* \cup \mathcal{A}^\omega$ and let $k \in \mathbb{N}_+$. The word w is k-overlap-free if and only if the word $\mu(w)$ is strongly $(2k-1)$-overlap-free.*

Proof. If $k = 1$, the equivalence is true from Lemma 5.4, thus our assuming that $k \geq 2$.

If the word w is not k-overlap-free then it contains a factor $XYXYX$ with $|X| = k$. This implies that $\mu(w)$, which contains the factor $\mu(X)\mu(Y)\mu(X)\mu(Y)\mu(X)$ with $|\mu(X)| = 2k$, is not $(2k-1)$-overlap-free.

Conversely, if the word $\mu(w)$ is not $(2k-1)$-overlap-free then it contains a factor $XxYXxYXx$ where $X, Y \in A^*$, $|X| = 2k-2$, and $x \in A$.

If $|Y|$ is even then $Y \neq \varepsilon$. For if not $\mu(w)$ would contain $XxXxXx$ which implies that both X and xXx are in Σ^*, a contradiction with Lemma 5.3. So, let $Z \in A^*$ and $y, z \in A$ be such that $Y = Zyz$. Then $XxYXxYXx = XxZyzXxZyzXx$ which implies that both X, xZy, yz, zXx, and Z are elements of Σ^*.

From Lemma 5.3, $X \in \Sigma^*$ and $zXx \in \Sigma^*$ imply $x \neq z$, and $Z \in \Sigma^*$ and $xZy \in \Sigma^*$ imply $x \neq y$. Therefore, $y = z$ which contradicts with $yz \in \Sigma^*$.

Consequently, $|Y|$ is odd so $|XxYXxYXx|$ is odd, and two cases are possible depending on whether, in $\mu(w)$, the factor $XxYXxYXx$ appears at an even index or at an odd index.

- $\mu(w) = \mu(w_1)XxYXxYXxy\mu(w_2)$ for a letter y.

  In this case, by definition of μ, the letter y is also the first letter of Y. This implies that $XxYXxYXxy = \mu(ZY'ZY'Z)$ with $\mu(Z) = Xxy$. Since $|Xxy| = 2k$, $|Z| = k$ and the word w is not k-overlap-free.

- $\mu(w) = \mu(w_1)yXxYXxYXx\mu(w_2)$ for a letter y.

  In this case, since $k \geq 2$ one has $X \neq \varepsilon$, so let $X = zX'$, $z \in A$, $X' \in A^+$. Then $yXxYXxYXx = yzX'xYzX'xYzX'x$, and by definition of μ, the letter y is also the last letter of Y. This implies that $yzX'xYzX'xYzX'x = \mu(ZY'ZY'Z)$ with $\mu(Z) = yzX'x$. Since $|yzX'x| = 2k$, $|Z| = k$ and the word w is not k-overlap-free. ■

Now, we consider the application $\delta_{k,k-1}$ $(k \geq 2)$ already used above. Since $\delta_{k,k-1}$ is defined by $\delta_{k,k-1}(0) = a$, $\delta_{k,k-1}(1) = ab^{k-1}$, $\delta_{k,k-1}(2) = ab^k$, it is straightforward that if $u \in \mathcal{A}^* \cup \mathcal{A}^\omega$ is such that $u = \delta_{k,k-1}(v)$ for some $v \in \mathcal{B}^* \cup \mathcal{B}^\omega$ then u contains a^3 if and only if v contains 00. Consequently, from Theorem 4.5, if $u \in \mathcal{A}^\omega$ and $v \in \mathcal{B}^\omega$ are such that $u = \delta_{k,k-1}(v)$ then u is a-strongly k-overlap-free whenever v has the 0-limited square property. We have see above that if $k = 2$, i.e., in the case of Thue's original application δ, the word u is strongly 2-overlap-free.

Now we notice that, in the case of $\delta_{k,k-1}$, the statement of Corollary 4.6 can be simplified because $2(k-p)+1 = 3$ when $p = k-1$.

Corollary 5.6 *Let $u \in \mathcal{A}^\omega$, and v, an infinite word over $\mathcal{B}$ which does not begin with a square, be such that $\delta_{k,k-1}(v) = u$. The word u is a-strongly k-overlap-free if and only if the word v has the 0-limited square property.*

In this section we have seen that the results given in Section 4 remain the same when adding the condition that words over $\mathcal{A}$ do not contain cubes of some single letter, in particular in using the application $\delta_{k,k-1}$. In the next section we give another interesting use of this application.

6 k-overlap-free binary partial words

A *partial word* u of length n over an alphabet $\mathcal{A}$ is a partial function $u\colon \{1, 2, \ldots, n\} \to \mathcal{A}$. This means that in some positions the word u contains *holes*, i.e., "do not know"-letters. The holes are represented by $\diamond$, a symbol that does not belong to $\mathcal{A}$. Classical words (called *full* words) are only partial words without holes. Partial words were first introduced by Berstel and Boasson [3] (see also [4]).

Similarly to finite words, we define infinite partial words to be partial functions from $\mathbb{N}_+$ to $\mathcal{A}$. We denote by $\mathcal{A}_\diamond^*$ and $\mathcal{A}_\diamond^\omega$ the sets of finite and infinite partial words, respectively.

A partial word $u \in \mathcal{A}_\diamond^*$ is a *factor* of a partial word $v \in \mathcal{A}_\diamond^* \cup \mathcal{A}_\diamond^\omega$ if there exist words $x, u' \in \mathcal{A}_\diamond^*$ and $y \in \mathcal{A}_\diamond^* \cup \mathcal{A}_\diamond^\omega$ such that $v = xu'y$ with $u'(i) = u(i)$ whenever neither $u(i)$ nor $u'(i)$ is a hole $\diamond$. Prefixes and suffixes are defined in the same way.

For example, let $u = ab\diamond bba\diamond a$. The length of u is $|u| = 8$, and u contains two holes in positions 3 and 7. Let $v = aa\diamond bb\diamond ba\diamond abbaa\diamond$. The word v contains the word u as a factor in positions 3 and 8. The word u is a suffix of the word v.

Note that a partial word is a factor of all the (full) words of the same length in which each $\diamond$ is replaced by any letter of $\mathcal{A}$. We call these (full) words the *completions* of the partial word. In the previous example, if $\mathcal{A} = \{a, b\}$, the partial word u has four completions: *ababbaaa*, *ababbaba*, *abbbbaaa*, and *abbbbaba*.

Let k be a rational number. A partial word u is *k-free* if all its completions are k-free. Overlaps, k-overlaps, overlap-freeness, and k-overlap-freeness of partial words are defined in the same manner.

In [6] it is proved that overlap-free binary infinite partial words cannot contain more than one hole, when 2-overlap-free binary infinite partial words can contain infinitely many holes. Here we complete this last result by the following theorem.

Theorem 6.1 *For every integer $k \geq 2$ and for every non-negative integer n, there exist infinitely many k-overlap-free binary infinite partial words containing n holes, and being not $(k-1)$-overlap-free.*

Proof of Theorem 6.1 is constructive and needs some preliminaries.

The word $\tau^\omega(0)$ contains an infinite number of occurrences of $\tau(01)$:

$$\begin{aligned}
\tau^\omega(0) &= \tau(01)u_1\tau(01)u_2\cdots u_\ell\tau(01)u_{\ell+1}\cdots, u_i \in \mathcal{B}^+ \\
&= \textstyle\prod_{\ell=1}^{\infty} \tau(01)u_\ell \\
&= \textstyle\prod_{\ell=1}^{\infty} 01201020121u_\ell.
\end{aligned}$$

For every integer $n \geq 0$, let Y_n be the word obtained from $\tau^\omega(0)$ by replacing 102 by 22 in n (not necessarily consecutive) occurrences of $\tau(01)$. Of course $Y_0 = \tau^\omega(0)$.

Proposition 6.2 *For every $n \in \mathbb{N}$, the word Y_n has the 0-limited square property.*

Proof. In [6], it is proved that the occurrences of 22 are the only squares in the word Y_n. Consequently, Y_n does not contain 00 as a factor. Moreover, Y_n contains no squares but those 22 obtained from $\tau^\omega(0)$ by replacing the factor 102 by 22 in n occurrences of $\tau(01)$, that is in n factors 01020. This implies that each of these 22 is preceded and followed by the letter 0. Therefore, since the word Y_n fulfills the conditions of Definition 4.3 it has the 0-limited square property. ■

Corollary 6.3 *For every integers $k \geq 2$ and p, $1 \leq p \leq k-1$, and for every integer $n \geq 0$, the word $\delta_{k,p}(Y_n)$ is k-overlap-free.*

Proof. By Proposition 6.2, the word Y_n has the 0-limited square property which implies, by Theorem 4.5, that $\delta_{k,p}(Y_n)$ is k-overlap-free. ■

In particular, for every integer $n \geq 0$, the words $\delta_{k,k-1}(\tau^\omega(0))$ and $\delta_{k,k-1}(Y_n)$ are k-overlap-free.

Proof of Theorem 6.1.

$$\begin{aligned}
\delta_{k,k-1}(\tau(01)) &= \delta_{k,k-1}(0120)\delta_{k,k-1}(102)\delta_{k,k-1}(0121) \\
&= \delta_{k,k-1}(0120)ab^{k-1}\underline{a}ab^k\delta_{k,k-1}(0121)
\end{aligned} \tag{1}$$

and

$$\begin{aligned}
\delta_{k,k-1}(0120\,22\,0121) &= \delta_{k,k-1}(0120)\delta_{k,k-1}(22)\delta_{k,k-1}(0121) \\
&= \delta_{k,k-1}(0120)ab^{k-1}\underline{b}ab^k\delta_{k,k-1}(0121)
\end{aligned} \tag{2}$$

Let us define Z_n to be the word obtained from $\delta_{k,k-1}(\tau^\omega(0))$ by replacing n (not necessarily consecutive) occurrences of $\delta_{k,k-1}(\tau(01))$ by $\delta_{k,k-1}(0120)ab^{k-1}\diamond ab^k\delta_{k,k-1}(0121)$.

From Corollary 6.3, and equations (1) and (2) above, for every integer $n \geq 0$, the word Z_n is k-overlap-free. Moreover, from Proposition 4.7, Z_n is not $(k-1)$-overlap-free. ■

Corollary 6.4 *For every integer $k \geq 2$, there exist infinitely many k-overlap-free binary infinite partial words containing infinitely many holes, and being not $(k-1)$-overlap-free.*

Proof. Considering the words Z_n defined in the proof of Theorem 6.1, and making n tend to infinity, we deduce that the word $\prod_{\ell=1}^{\infty} \delta_{k,k-1}(0120)ab^{k-1} \diamond ab^k \delta_{k,k-1}(0121)u_\ell$ has the required property.

Now, we can choose to leave out a finite number of substitutions of the factor 102 by 22. Since the number of such choices is infinite, the result follows. ■

Acknowledgments

I am greatly indebted to Professor Gwénaël Richomme whose comments and suggestions on a preliminary version of this paper were very useful.

References

[1] J.-P. Allouche, J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in: C. Ding. T. Helleseth, H. Niederreiter (Eds.), *Sequences and Their Applications*, Proceedings of SETA'98, Springer-Verlag (1999), 1–16.

[2] J. Berstel, Axel Thue's work on repetitions in words, *in*: Leroux, Reutenauer (eds), Séries formelles et combinatoire algébrique, Publications du LaCIM, Université du Québec à Montréal, Montréal (1992) 65–80. *See also* Axel Thue's papers on repetitions in words: a translation, *Publications du LaCIM*, Département de mathématiques et d'informatique, Université du Québec à Montréal **20** (1995), 85 pages.

[3] J. Berstel, L. Boasson, Partial words and a theorem of Fine and Wilf, *Theoret. Comput. Sci.* **218** (1999), 135–141.

[4] F. Blanchet-Sadri, *Algorithmic Combinatorics on Partial Words*, Chapman & Hall/CRC Press, Boca Raton, FL, 2007.

[5] J. Berstel, P. Séébold, A characterization of overlap-free morphisms, *Discrete Appl. Math.* **46** (1993), 275–281.

[6] V. Halava, T. Harju, T. Kärki, P. Séébold, Overlap-freeness in infinite partial words, *Theoret. Comput. Sci.* **410** (2009), 943–948.

[7] M. Lothaire, *Combinatorics on Words*, vol. 17 of *Encyclopedia of Mathematics and Applications*, Addison-Wesley, Reading, Mass., 1983.
Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, Cambridge, UK, 1997.

[8] M. Morse, Recurrent geodesics on a surface of negative curvature, *Trans. Amer. Math. Soc.* **22** (1921), 84–100.

[9] P. Séébold, Sequences generated by infinitely iterated morphisms, *Discrete Appl. Math.* **11** (1985), 255–264.

[10] P. Séébold, k-overlap-free words, Preprint, JORCAD'08, Rouen, France (2008), 47–49.

[11] A. Thue, Über unendliche Zeichenreihen, *Christiania Vidensk.-Selsk. Skrifter. I. Mat. Nat. Kl.* **7** (1906), 1–22.

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