Vesa Halava | Tero Harju | Tomi Kärki | Patrice Séébold

Restricted square property and infinite partial words

Turku Centre for Computer Science

TUCS Technical Report
No 930, February 2009

Restricted square property and infinite partial words

Vesa Halava
Department of Mathematics and
TUCS - Turku Centre for Computer Science
University of Turku, FI-20014 Turku, Finland
vehalava@utu.fi
Tero Harju
Department of Mathematics and
TUCS - Turku Centre for Computer Science
University of Turku, FI-20014 Turku, Finland
harju@utu.fi
Tomi Kärki
Department of Mathematics and
TUCS - Turku Centre for Computer Science
University of Turku, FI-20014 Turku, Finland
topeka@utu.fi
Patrice Séébold
LIRMM, UMR 5506 CNRS, 161 rue Ada
34392 Montpellier, France
Département Mathématiques et Informatique Appliquées,
Université Paul Valéry, Route de Mende,
34199 Montpellier Cedex 5, France
Patrice.Seebold@lirmm.fr

TUCS Technical Report
No 930, February 2009

Abstract

We prove that there exist infinitely many infinite overlap-free binary partial words containing at least one hole. Moreover, we show that these words cannot contain more than one hole and the only hole must occur either in the first or in the second position. We define that a partial word is k-overlap-free if it does not contain a factor of the form $xyxyx$ where the length of x is at least k. We prove that there exist infinitely many 2-overlap-free binary partial words containing an infinite number of holes.

Keywords: Repetition-freeness, k-free, overlap, partial words, Thue-Morse word, infinite words, restricted square property

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

Repetitions, i.e., consecutive occurrences of words within a word and especially repetition-freeness have been fundamental research subjects in combinatorics on words since the seminal papers of Thue [21, 22] in the beginning of the 20th century; see [4] to learn what Thue exactly proved. In particular, Thue showed that there exists an infinite word w over a 3-letter alphabet, which does not contain any nonempty squares xx. Moreover, he constructed an infinite binary word t which does not contain any overlaps $xyxyx$ for any words x and y with x nonempty. This celebrated word is nowadays called the Thue-Morse word, which has many surprising and remarkable properties; see [2]. As an example, we mention applying t for designing an unending play of chess [8, 15] and for solving the Burnside problem for groups [1] and semigroups [16, 17].

In [14] Manea and Mercaş considered repetition-freeness of partial words. Partial words are words with "do not know"-symbols called holes and they were first introduced by Berstel and Boasson in [5]. Motivation for the study of partial words comes from applications in word algorithms and molecular biology, in particular; see [6] for using partial words in DNA sequencing and DNA comparison. The theory of partial words has developed rapidly in the recent years and many classical topics in combinatorics on words have been revisited. Topics such as periodicity, primitivity, unbordered word, codes and equations have been considered in the first book on partial words authored by Blanchet-Sadri in 2007 [7]. See also related works by Shur and Gamzova [20], Leupold [11] and Lischke [12]. As another approach for modeling missing or uncertain information in words we want to mention word relations, a generalization of the compatibility of partial words introduced in [9].

It was shown in [14] that there exist infinitely many cube-free binary partial words containing an infinite number of holes. In this paper we give short and simple proofs that this result can be improved. The key notion is the restricted square property of infinite words over a three-letter alphabet introduced in Section 3. Using it we easily prove in Section 5 that there exist infinitely many binary partial words with an infinite number of holes which do not contain 2-overlaps, i.e., factors of the form $xyxyx$ where the length of x is at least two. We also prove that there exist infinitely many infinite overlap-free binary partial words with one hole but none with two or more holes, and that the single hole can only be either in the first or in the second position of the word.

2 Words, morphisms, and powers

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. 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 $\mathcal{A}^*$ and infinite words are denoted by $\mathcal{A}^\omega$. Let also $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. A finite word v is a *factor* of $w \in \mathcal{A}^* \cup \mathcal{A}^\omega$ if $w = xvy$, where $x \in \mathcal{A}^*$ and $y \in \mathcal{A}^* \cup \mathcal{A}^\omega$. If $x = \varepsilon$, then v is a *prefix* of w. If $v \in \mathcal{A}^* \cup \mathcal{A}^\omega$ and $w = xv$, then v is called a *suffix* of w.

A morphism on $\mathcal{A}^*$ is a mapping $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ satisfying $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in \mathcal{A}^*$. Note that φ is completely defined by the values $\varphi(a)$ for every letter a on $\mathcal{A}$. A morphism is called *prolongable on a letter* a if $\varphi(a) = aw$ for some word $w \in \mathcal{A}^+$ such that $\varphi^n(w) \neq \varepsilon$ for all integers $n \geq 1$. By the definition, $\varphi^n(a)$ is a prefix of $\varphi^{n+1}(a)$ for all integers $n \geq 0$ and the sequence $(\varphi^n(a))_{n\geq 0}$ converges to the unique infinite word generated by φ,

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

which is a fixed point of φ.

A kth *power* of a word $u \neq \varepsilon$ is the word u^k. It is the 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* or 2^+*-free* if it does not contain overlaps or, equivalently, if it does not contain kth powers for any $k > 2$. Hence, 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 an infinite square-free word over a binary alphabet but, as we will see in the next section, there exist infinite overlap-free binary words.

We generalize the notion of an overlap as follows.

Definition 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*[1]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. By the definition, it is evident that any 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.

Remark 1. Another possible definition for k-overlap-freeness is to require that a k-overlap-free word must also be cube-free. In the case of 2-overlap-free words, this new definition just means that in addition to 2-overlaps a word cannot contain short cubes aaa, where a is a letter. We want to stress that all the results in this paper are also valid for this alternative definition of 2-overlap-freeness.

[1]While it is not exactly the same, this notion of k-overlap-freeness resembles that of k-bounded overlaps introduced by Thue in [22].

3 Preliminary results

Let us consider two alphabets $\mathcal{A} = \{a, b\}$ and $\mathcal{B} = \{0, 1, 2\}$.

3.1 Overlap-free binary words

In [22] Thue introduced the following morphism $\mu\colon \mathcal{A}^* \to \mathcal{A}^*$,

$$a \mapsto ab, \quad b \mapsto ba.$$

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

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

generated by μ; see, e.g., [2] for other definitions and properties.

Proposition 1 below gives a useful property of the Thue-Morse word t. Its proof uses two already known lemmata.

The first lemma is due to Thue [22] himself; see [13] for a proof.

Lemma 1. *Let $X = \{ab, ba\}$. If $x \in X^*$, then $axa \notin X^*$ and $bxb \notin X^*$.*

The second lemma is a part of Proposition 1.7.5 in [3]; see also [18].

Lemma 2. *If x is an infinite overlap-free binary word over $\mathcal{A}$, then there exist $v \in \{\varepsilon, a, b, aa, bb\}$ and an infinite overlap-free binary word y such that $x = u\mu(y)$.*

Proposition 1. *Let t' be a suffix of the Thue-Morse word t beginning with $\mu(abaabb)$. Then the word bbt' is overlap-free.*

Proof. The word t' is overlap-free since t is overlap-free. Let us first prove that bt' is also overlap-free. Suppose that bt' contains an overlap. Since t' is overlap-free and begins with the letter a, this means that bt' begins with $bubub$ for a word $u \in \mathcal{A}^+$. By the definition of t', we must have $v = \varepsilon$ in Lemma 2 and $t' = \mu(t'')$ for some infinite word t''. This implies that t' decomposes over $\{ab, ba\}$. Thus, if $|u|$ is even, then u and bub are images of words by μ, which contradicts with Lemma 1.

Consequently, $|u|$ is odd. Thus, $ub = \mu(u'a)$ for some word $u' \in \mathcal{A}^*$ and t' begins with $\mu(u'au'a)$ which implies that t contains $u'au'a$ as a factor. From the definition of t' one has that $|u'| \geq 6$ and u' begins with $abaa$. So $u' = abaau''$ for some word $u'' \in \mathcal{A}^+$ and $u'au'a = abaau''aabaau''a$. If u'' begins or ends with a then $u'au'a$ contains aaa as a factor. Otherwise u'' begins and ends with b and $u'au'a$ contains the factor $baabaab$. In both cases this contradicts with the overlap-freeness of t.

Now we prove that bbt' is overlap-free. Suppose that bbt' contains an overlap. Since bt' is overlap-free, this means that bbt' begins with $bubub$ for a word $u \in \mathcal{A}^+$, and $ubub$ is overlap-free. Suppose that $|u|$ is odd. Since $t' = \mu(t'')$ for some infinite word t'', we have $u = b\mu(u')$ for some word $u' \in \mathcal{A}^+$. But in this case bt', which begins with ubu, has the prefix $b\mu(u')bb$. This means that t' begins with $\mu(u')bb$, which contradicts with $t' = \mu(t'')$.

Thus, $|u|$ is even. Let $u' \in \mathcal{A}^+$ be such that $u = bu'$. Now bbt' begins with $bbu'bbu'b$. Since $t' = \mu(t'')$, there exist two words u_1 and u_2 in $\mathcal{A}^+$ such that $u'b = \mu(u_1)$ and $bu' = \mu(u_2)$. Since $u'bbu'b$ is overlap-free, the word u' begins and ends with a. This implies that $u_1 = au_1'a$ and $u_2 = bu_2'b$ for some words u_1', u_2' over $\mathcal{A}$. Consequently, $bbu'bbu'b = bb\mu(u_1)\mu(u_2)b = bbab\mu(u_1')abba\mu(u_2')bab$, which implies that u_1' begins with b and u_2' ends with a. But in this case $ubub = bu'bbu'b = \mu(u_2)bb\mu(u_1)$ contains $\mu(ab)bb\mu(ab) = abbabbabba$, which contradicts with the overlap-freeness of $ubub$. □

3.2 The restricted square property

In order to prove the existence of infinite cube-free words over a two-letter alphabet from the existence of square-free words over three letters, Thue used in [21] the following morphism $\delta\colon \mathcal{B}^* \to \mathcal{A}^*$,

$$0 \mapsto a, \quad 1 \mapsto ab, \quad 2 \mapsto abb.$$

Six years later he proved the following

Proposition 2 ([22]). *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 v is square-free and does not contain 010 nor 212 as a factor.*

Here we will use the morphism δ to prove the existence of infinite 2-overlap-free binary words that are not overlap-free.[2] We need the following new notion introduced in [19].

Definition 2. An infinite word v over $\mathcal{B}$ has the *restricted square property* if, for every nonempty factor rr of v, the word r does not begin nor end with the letter 0 and the factor rr is preceded and followed by the letter 0.

Notice that if a word $v \in \mathcal{B}^\omega$ has the restricted square property then v does not begin with a square, v is overlap-free, and v does not contain 00 as a factor.

The following result is a useful analogue of Proposition 2.

Theorem 1. *Let v be an infinite word over $\mathcal{B}$ such that it does not begin with a square and the infinite word $u = \delta(v)$ over $\mathcal{A}$ does not contain the factor aaa. Then the word u is 2-overlap-free if and only if v has the restricted square property.*

[2] Thue [22] already remarked that a word $\delta(w)$, where w is square-free, may have overlaps, but if $xyxyx$ is an overlap, then x is a letter.

Note that here the word u is also cube-free.

Proof. Let u and v be as in the statement. Since u does not contain the factor aaa, v does not contain the factor 00. Let rr be a factor of v with $r \neq \varepsilon$. By the hypothesis, rr is not at the beginning of v. This means that in v the factor rr is preceded (and followed) by at least one letter.

If r begins with the letter 0, then it does not end with 0 (because 00 is not a factor of v) and it is preceded by the letter 1 or the letter 2. Consequently, $\delta(r) = asb$ and $\delta(rr)$ is necessarily preceded by b and followed by a. This means that u contains the factor $b\delta(rr)a = basbasba$, which implies that u is not 2-overlap-free. The argument is the same if r ends with the letter 0.

Now if r begins with 1 or 2, ends with 1 or 2, and rr is not followed by 0, then $\delta(r)$ begins with ab and $\delta(rr)$ is followed by ab. Hence, u is not 2-overlap-free.

To end, if r ends with 1 or 2 and is not preceded by 0, then $\delta(r)$ begins with a and ends with b, and $\delta(rr)$ is preceded by b. Since $\delta(rr)$ is followed by a, this implies that u is not 2-overlap-free.

Consequently, if u is 2-overlap-free, then v has the restricted square property.

Conversely, suppose that u is not 2-overlap-free. There are four possible cases:

1. If u contains a factor $aaxaaxaa$, then the word v contains a square beginning with 0;
2. If u contains a factor $abxabxab$, then there exists necessarily $y \in \mathcal{B}^+$ such that $abx = \delta(y)$. Thus, v contains a square yy followed by a letter 1 or 2;
3. If u contains a factor $baxbaxba$, then there exists necessarily $y \in \mathcal{B}^+$ such that $axb = \delta(y)$. Thus, v contains a square yy preceded by a letter 1 or 2;
4. If u contains a factor $bbxbbxbb$, then there exists necessarily $y \in \mathcal{B}^+$ such that $xbb = \delta(y)$. Thus, v contains a square yy preceded by a letter 2.

In the four cases v does not have the restricted square property. $\square$

3.3 2-overlap-free binary words

We consider another morphism introduced by Thue [22]: $\tau\colon \mathcal{B}^* \to \mathcal{B}^*$,

$$0 \mapsto 01201, \quad 1 \mapsto 020121, \quad 2 \mapsto 0212021.$$

Proposition 3 ([22])**.** *The infinite word $\tau^\omega(0)$ is square-free.*

Since $\tau(2) = 0212021$, the infinite word $\tau^\omega(0)$ contains the factor 212. Hence, by Proposition 2, the word $\delta(\tau^\omega(0))$ is not overlap-free. However, by Proposition 3, $\tau^\omega(0)$ has the restricted square property. Now, by the construction, the word $\tau^\omega(0)$ contains an infinite number of occurrences of $\tau(01)$:

$$\begin{array}{rcl} \tau^\omega(0) & = & u_1\tau(01)u_2\tau(01)\cdots u_k\tau(01)\cdots, u_i \in \mathcal{B}^+ \\ & = & \prod_{k=1}^{\infty} u_k\tau(01) \\ & = & \prod_{k=1}^{\infty} u_k 01201020121. \end{array}$$

Let $n \in \mathbb{N}$ and let us denote by Y_n the word obtained from $\tau^\omega(0)$ by replacing 102 by 22 in n (not necessarily consecutive) occurrences of $\tau(01)$.

Proposition 4. *For every $n \in \mathbb{N}$, the word Y_n has the restricted square property.*

Proof. We will prove that the occurrences of 22 are the only squares in the word Y_n.[3]

Suppose that Y_n contains another square $rr \neq 22$. Since the only difference between $\tau^\omega(0)$ and Y_n comes from the occurrences of 10 replaced by the letter 2 and since $\tau^\omega(0)$ is square-free, the square rr must have a common factor with at least one factor 22.

If r contains some full occurrences of 22, then replacing each of these occurrences by 102 does not change the fact that rr is a square. Hence, if there are no other occurrences of 22 intersecting with rr, then this square is a factor of $\tau^\omega(0)$, which is impossible.

Thus, we have $r = 2u2$ for some $u \in \mathcal{B}^+$. By the construction, u ends with 0120. Since 01202 is not a factor of $\tau^\omega(0)$, the only solution is that the factor rr is followed in Y_n by the letter 2. Then $u22u22$ is a factor of Y_n and the corresponding factor of $\tau^\omega(0)$ (which, by the construction, is obtained by replacing in $u22u22$ all the occurrences of 22 by 102) is also a square; a contradiction.

Consequently, Y_n contains no squares but those 22 obtained from $\tau^\omega(0)$ by replacing the factor 102 by 22 in n occurrences of $\tau(01)$. Since, by the construction of $\tau^\omega(0)$, each of these 22 is preceded and followed by the letter 0, the word Y_n has the restricted square property. □

Theorem 1 implies the following useful corollary.

Corollary 1. *The words $\delta(\tau^\omega(0))$ and $\delta(Y_n)$ for every $n \in \mathbb{N}$ are 2-overlap-free.*

Proof. We have seen that the words $\tau^\omega(0)$ and Y_n have the restricted square property. Since $\tau^\omega(0)$ is square-free, it does not begin with a square. By the proof of Proposition 4, the only squares in Y_n are occurrences of 22. Thus, Y_n does not begin with a square. To end, since $\tau^\omega(0)$ and Y_n do not contain 00 as a factor, the words $\delta(\tau^\omega(0))$ and $\delta(Y_n)$ do not contain the factor aaa. Thus, Theorem 1 implies that the words $\delta(\tau^\omega(0))$ and $\delta(Y_n)$ are 2-overlap-free. □

4 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 contain *holes*, i.e., "do not know"-letters. The holes are represented by $\diamond$, a symbol that

[3] In [10] the same technique was used to prove that there exist uncountably many almost square-free partial words over a ternary alphabet with an infinite number of holes.

does not belong to $\mathcal{A}$. Classical words (called *full* words) are only partial words without holes.

Similarly to finite words, we define that infinite partial words are 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.

5 The overlap-freeness of binary partial words

In this section we again have $\mathcal{A} = \{a, b\}$.

In [14] Manea and Mercaş proved that there exist infinitely many cube-free binary partial words containing infinitely many holes. Here we prove a stronger result about 2-overlap-free binary partial words.

Theorem 2. *There exist infinitely many* 2*-overlap-free binary partial words containing infinitely many holes.*

Proof. Let $n \in \mathbb{N}$. We have seen in Corollary 1 that $\delta(\tau^\omega(0))$ and $\delta(Y_n)$ are 2-overlap-free (but they are not overlap-free). Since Y_n is obtained from $\tau^\omega(0)$ by replacing n factors 102 by 22, the only difference between $\delta(\tau^\omega(0))$ and $\delta(Y_n)$ is that n factors $\delta(102) = ab\underline{a}abb$ in $\delta(\tau^\omega(0))$ are replaced by the factors $\delta(22) = ab\underline{b}abb$ in $\delta(Y_n)$. Let us consider the word X_n, which is obtained from $\delta(Y_n)$ by replacing $\delta(22)$ by $ab\diamond abb$. Since both $\delta(\tau^\omega(0))$ and $\delta(Y_n)$ are 2-overlap-free, also the word X_n is 2-overlap-free and contains exactly n holes.

In particular, denote by Y the word which is obtained from $\tau^\omega(0)$ by replacing 102 with 22 in every occurrence of $\tau(01)$. Let us now consider the word X where every $\delta(22)$ in $\delta(Y)$ is replaced by $ab\diamond abb$. Assume that the word X is not 2-overlap-free. Then a finite prefix of X contains a 2-overlap. This implies that, for some n, there exists a word X_n which has the same finite prefix as X. By the above, this X_n is 2-overlap-free; a contradiction.

Since $\tau^\omega(0)$ contains infinitely many occurrences of $\tau(01)$, the word X contains infinitely many holes and it is 2-overlap-free. Clearly, the word X remains 2-overlap-free if we replaced any hole by either a or b. Hence, there exists infinitely many 2-overlap-free words containing infinitely many holes. $\square$

By replacing holes with letters in 2-overlap-free binary partial words containing infinitely many holes we obtain the following corollary.

Corollary 2. *For every non-negative integer n, there exist infinitely many 2-overlap-free binary partial words containing n holes.*

In the case of (1-)overlap-free binary partial words, the situation is different, because it is not possible to construct infinite overlap-free binary partial words with more than one hole. More precisely, we prove the following theorem.

Theorem 3. *An infinite overlap-free binary partial word is either full or of the form $\diamond w$ or $x\diamond w$, where w is an infinite full word and x is a letter. There are infinitely many overlap-free words of each type.*

Proof. The case of full words follows from the existence of the Thue-Morse infinite overlap-free word t.

Now, let x, y be the two different letters of the alphabet $\mathcal{A}$ and let w be an infinite partial word over $\mathcal{A}$ containing a factor u which begins with $x\diamond$. If u begins with $x\diamond x$ or $x\diamond yy$, then u contains a cube. Thus, u begins with $x\diamond yx$. If u begins with $x\diamond yxyyy$, $x\diamond yxyyx$, $x\diamond yxyx$, or $x\diamond yxxx$, then it is not overlap-free. Hence, u begins with $x\diamond yxxy$. If u begins with $x\diamond yxxyyy$ or $x\diamond yxxyx$, it is not overlap-free. Therefore, the only remaining case is that u begins with $x\diamond yxxyyx$. But then yu and xu are not overlap-free, which implies that, if the word w is overlap-free, then the factor u can only be at the beginning of w. Moreover, this also implies that w cannot contain more than one hole. To conclude the first claim, note that removing the first letter of a word keeps the word overlap-free.

To complete the proof, it remains to show that there exist an infinite number of overlap-free binary partial words beginning with such a word u. Consider any suffix of the Thue-Morse word t beginning with $\mu(babaabb) = ba\mu(abaabb)$. By the overlap-freeness of t, this suffix is overlap-free. On the other hand, if we replace the second letter of the suffix by b, we get a word of the form bbt', where t' is a suffix of t beginning with $\mu(abaabb)$. By Proposition 1, this word is also overlap-free. Hence, we conclude that the word $b\diamond t'$ is an infinite overlap-free binary partial word.

Since the Thue-Morse word t is recurrent, i.e., each factor appears infinitely often in t, it contains an infinite number of suffixes beginning with $\mu(babaabb)$. Thus, there exist an infinite number of infinite overlap-free binary partial words beginning with $b\diamond abbaab$. We note that, by the above, the word $abbaab$ is the only possibility that may occur after $b\diamond$ in an overlap-free word. $\square$

This theorem has the following corollary, which improves Proposition 5 of Manea and Mercaş [14] that there exist infinitely many cube-free binary partial words containing exactly one hole.

Corollary 3. *There exist infinitely many infinite overlap-free binary partial words containing exactly one hole.*

6 Conclusion

In this paper we have considered k-overlap-freeness and overlap-freeness of binary partial words. In Theorem 1 we have proven a connection between 2-overlap-free words and the restricted square property. Using this result we have shown in Corollary 1 that certain binary words are 2-overlap-free. These words enable us to prove in Theorem 2 that there exist infinitely many 2-overlap-free binary partial words containing infinitely many holes. Finally, we have shown in Theorem 3 that an infinite overlap-free binary partial word is either full or of the form $\diamond w$ or $x\diamond w$, where w is an infinite full word and x is a letter. Moreover, there are infinitely many overlap-free words of each type.

References

[1] S. I. Adian, The Burnside problem and identities in groups, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas 95], Springer-Verlag, Berlin, 1979.

[2] 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.

[3] J.-P. Allouche, J. Shallit, Automatic sequences: theory, applications, generalizations, Cambridge University Press, Cambridge, UK, 2003.

[4] 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.

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

[6] F. Blanchet-Sadri, Codes, orderings, and partial words, Theoret. Comput. Sci. 329 (2004) 177–202.

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

[8] M. Euwe, Mengentheoretische Betrachtungen über das Schachspiel, Proc. Konin. Acad. Wetenschappen, Amsterdam 32 (1929) 633–642.

[9] V. Halava, T. Harju, T. Kärki, Relational codes of words, Theoret. Comput. Sci. 389 (2007) 237–249.

[10] V. Halava, T. Harju, T. Kärki, Square-free partial words, Inform. Process. Lett. 108, 290–292, 2008.

[11] P. Leupold, Partial words for DNA coding, Lecture Notes in Comput. Sci. 3384 (2005) 224–234.

[12] G. Lischke, Restorations of punctured languages and similarity of languages, MLQ Math. Log. Q. 52 (2006) 20–28.

[13] 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.

[14] F. Manea, R. Mercaş, Freeness of partial words, Theoret. Comput. Sci. 389 (2007) 265–277.

[15] M. Morse, Abstract 360: a solution of the problem of infinite play in chess, Bull. Amer. Math. Soc. 44 (1938) 632.

[16] M. Morse, G.A. Hedlund, Symbolic dynamics, Amer. J. Math 60 (1938) 815–866.

[17] M. Morse, G.A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, Duke Math. J. 11 (1944) 1–7.

[18] A. Restivo, S. Salemi, Overlap free words on two symbols, *in*: Nivat, Perrin (eds), Automata on infinite words, Lecture Notes in Comput. Sci. 192 (1984) 198–206.

[19] P. Séébold, k-overlap-free words, Preprint, JORCAD'08, Rouen, France (2008) 9 pages.

[20] A.M. Shur, Y.V. Gamzova, Partial words and the interaction property of periods, Izv. Math. 68 (2004) 405–428.

[21] A. Thue, Über unendliche Zeichenreihen, Norske Vid. Skrifter I Mat.-Nat. Kl., Christiania 7 (1906) 1–22.

[22] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norske Vid. Skrifter I Mat.-Nat. Kl., Christiania 1 (1912) 1–67.

TURKU

CENTRE *for*

COMPUTER

SCIENCE

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration

- Institute of Information Systems Sciences

ISBN 978-952-12-2251-1
ISSN 1239-1891