Vesa Halava | Tero Harju | Tomi Kärki

# Square-free partial words

Turku Centre *for* Computer Science

TUCS Technical Report
No 893, May 2008

# Square-free 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

TUCS Technical Report
No 893, May 2008

**Abstract**

We say that a partial word $w$ over an alphabet $\mathcal{A}$ is square-free if every factor $xx'$ of $w$ such that $x$ and $x'$ are compatible is either of the form $\diamond a$ or $a\diamond$ where $\diamond$ is a hole and $a \in \mathcal{A}$. We prove that there exist uncountably many square-free partial words over a ternary alphabet with an infinite number of holes.

**Keywords:** Repetitions, square-freeness, partial words, Thue-Morse word, Leech word, infinite words

**TUCS Laboratory**
Discrete Mathematics for Information Technology

## 1 Introduction

Repetitions and repetition-freeness have been intensively studied in combinatorics on words during the last three decades. The seminal papers in this research are those by Thue [7, 8]. In addition to the celebrated binary Thue-Morse sequence [9], Thue showed that there exists an infinite word $w$ over a 3-letter alphabet that does not contain any squares $xx$, where $x$ is a nonempty word in $w$. In this paper we generalize this result for partial words.

Partial words are words with "do not know"-symbols $\diamond$ called holes. They were first introduced by Berstel and Boasson in [1]. The theory of partial words has developed rapidly in recent years and many classical topics in combinatorics on words have been revisited; see [2]. In [6] Manea and Mercaş considered repetition-freeness of partial words. They showed that there exist infinitely many cube-free binary partial words containing an infinite number of holes. Moreover, they constructed an infinite word over a $4$-letter alphabet such that substituting randomly any letter with a hole the word stays cube-free. Furthermore, if arbitrarily many letters with a distance at least two are replaced by holes, the word is still cube-free.

The study of repetitions in partial words was continued in [4], where the present authors proved that there exist infinitely many infinite overlap-free binary partial words with one hole. Secondly, they showed that an infinite overlap-free binary partial word cannot contain more than one hole. However, a binary partial word with an infinite number of holes can be "almost overlap-free". More precisely, it was shown in [4] that there exist infinitely many cube-free binary partial words with an infinite number of holes which do not contain a factor of the form $xyx'y'x''$ where $x, x', x''$ and, respectively $y, y'$, are pairwise compatible, the length of $x$ is at least three and $y$ is nonempty. It remained an open question, whether the length of $x$ can be reduced to two. Moreover, the question about the existence of "square-free" partial words was not considered. For square-freeness we must allow at least squares of the form $\diamond a$ and $a\diamond$ where $a$ is a letter, since repetitions of this form are unavoidable. In this paper we tackle this problem by constructing with the help of a $13$-uniform morphism an infinite square-free partial word over a ternary alphabet with an infinite number of holes.

## 2 Preliminaries

We recall some notions and notation mainly from [1]. A word $w = a_1a_2\cdots a_n$ of length $n$ over an alphabet $\mathcal{A}$ is a mapping $w\colon \{1, 2, \ldots, n\} \to \mathcal{A}$ such that $w(i) = a_i$. The elements of $\mathcal{A}$ are called letters. The length of a word $w$ is denoted by $|w|$, and the length of the empty word $\varepsilon$ is zero. An infinite word $w = a_1a_2a_3\cdots$ is 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 the set of the infinite words is denoted by $\mathcal{A}^\omega$. A finite word $v$ is a *factor* of $w$ if $w = xvy$, where $x$ is finite word and $y$ is either a finite or an infinite word. The set of factors of $w$ is denoted by $F(w)$. The word $v$ is called a *prefix* of $w$, if in the above $x = \varepsilon$. A prefix of $w$ of length $n$ is denoted by $\mathrm{pref}_n(w)$. If $w = xv$, then $v$ is called a *suffix* of $w$.

A partial word $u$ of length $n$ over the alphabet $\mathcal{A}$ is a partial function $u\colon \{1, 2, \ldots, n\} \to \mathcal{A}$. The domain $D(u)$ is the set of positions $i \in \{1, 2, \ldots, n\}$ such that $u(i)$ is defined. The set $H(u) = \{1, 2, \ldots, n\} \setminus D(u)$ is called the set of *holes*. If $H(u)$ is empty, then $u$ is a (full) word. As for full words, we denote by $|u| = n$ the length of a partial word $u$. Similarly to finite words, we define infinite partial words as partial functions from $\mathbb{N}_+$ to $\mathcal{A}$.

Let $\diamond$ be a symbol that does not belong to $\mathcal{A}$. For a partial word $u$, we define its *companion* to be the full word $u_\diamond$ over the augmented alphabet $\mathcal{A}_\diamond = \mathcal{A} \cup \{\diamond\}$ such that $u_\diamond(i) = u(i)$, if $i \in D(u)$, and $u_\diamond(i) = \diamond$, otherwise. The sets $\mathcal{A}_\diamond^*$ and $\mathcal{A}_\diamond^\omega$ correspond to the sets of finite and infinite partial words, respectively. A partial word $u$ is said to be *contained* in $v$ (denoted by $u \subset v$) if $|u| = |v|$, $D(u) \subseteq D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. Two partial words $u$ and $v$ are *compatible* (denoted by $u \uparrow v$) if there exists a (partial) word $z$ such that $u \subset z$ and $v \subset z$. Using the companions this means that we must have $u_\diamond(i) = v_\diamond(i)$ whenever neither $u_\diamond(i)$ nor $v_\diamond(i)$ is a hole $\diamond$.

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

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

which is a fixed point of $h$. A morphism $h$ is called *$k$-uniform* if $|h(a)| = k$ for all $a \in \mathcal{A}$. As an example, consider the morphism $\varphi\colon \{0,1,2\}^* \to \{0,1,2\}^*$ defined by

$$\begin{aligned} 0 &\mapsto 0121021201210, \\ 1 &\mapsto 1202102012021, \\ 2 &\mapsto 2010210120102. \end{aligned} \tag{1}$$

This morphism is 13-uniform. The word

$$\Lambda := \varphi^\omega(0) = 0121021201210120210201202120102101201021 20\cdots$$

obtained by iterating the morphism $\varphi$ turns out to be very useful when considering square-freeness of partial words. We call this word the *Leech word*; see [5].

## 3 Square-free infinite partial words

The $k$th *power* of a word $u \neq \varepsilon$ is the word $u^k = \mathrm{pref}_{k\cdot|u|}(u^\omega)$, where $u^\omega$ denotes the infinite catenation of the word $u$ with itself and $k$ is a rational number such that $k \cdot |u|$ is an integer. A partial word $u$ is called *$k$-free* if, for any nonempty factor $v$ of $u$, there does not exists a full word $x$ such that $v$ is contained in the $k$th power of $x$, *i.e.*, $v \subset x^k$. Note that, for full words, this means that $v = x^k$. If $k = 2$ or $k = 3$, then we talk about *square-free* or *cube-free* words, respectively. Moreover, a word is called *overlap-free* if it is $k$-free for any $k > 2$.

It is easy to verify that there does not exist square-free infinite words over a binary alphabet. However, the classical results by Thue state the following:

**Theorem 1** ([7, 8])**.** *There exist a binary infinite overlap-free word and an infinite square-free word over a ternary alphabet.*

The infinite overlap-free word constructed by Thue is nowadays called the *Thue-Morse word* and it is obtained as a fixed point $t = \tau^\omega(0)$ of the morphism $\tau\colon \{0,1\}^* \to \{0,1\}^*$, where $\tau(0) = 01$ and $\tau(1) = 10$. A square-free word $T$ is derived from $t$ by using the inverse of the morphism $\sigma$ for which $\sigma(a) = 011$, $\sigma(b) = 01$ and $\sigma(c) = 0$. Square-free words can also be generated by iterating uniform morphisms as was proved by Leech.

**Theorem 2** ([5])**.** *The word $\Lambda = \varphi^\omega(0)$, where $\varphi$ is defined by* (1)*, is square-free.*

We will use this result in order to prove that there exists infinitely many almost square-free ternary partial words with an infinite number of holes. As was mentioned above, we cannot avoid short squares. Namely, any word containing a hole contains also a square of the form $\diamond a$ or $a\diamond$ for some $a \in \mathcal{A}$. Hence, we modify the definition of square-freeness as follows.

**Definition 1.** A word of the form $xx'$ where $x$ and $x'$ are compatible and either $|x| > 1$ or $x = x'$ is called a *partial square*. A partial word is called *square-free* if it does not contain any partial squares.

The above definition means that a square-free partial word cannot contain any full squares or squares of the form $\diamond\diamond$. Only the unavoidable squares $\diamond a$ or $a\diamond$ are allowed.

Let us now consider the Leech word $\Lambda = \varphi^\omega(0)$. Since $\Lambda$ is a fixed point of $\varphi$, *i.e.*, $\varphi(\Lambda) = \Lambda$, the word can be decomposed into blocks $\varphi(0), \varphi(1)$ and $\varphi(2)$ of length $13$. Now define the *partial Leech word* $\hat{\Lambda}$ by replacing each block $\varphi(0)$ of $\Lambda$ by

$$\alpha = 012\diamond021201210.$$

Next we prove that $\hat{\Lambda}$ is square-free. The result means that in every block $\varphi(0)$ of $\Lambda$ the $4$th letter can be replaced by $0$ or $2$, and still the infinite word remains square-free. Hence, this construction gives an uncountable set of ternary infinite full words where the only square factors are $00$ and $22$ .

**Theorem 3.** *There exist uncountably many words over a ternary alphabet containing infinitely many holes.*

*Proof.* If the partial Leech word is not square-free, then in $\hat{\Lambda}$ there is a partial square of the form $xx'$ or $x'x$ such that, for some position $i$, we have

$$x(i) = \diamond \text{ and either } x'(i) = 0 \text{ or } x'(i) = 2. \tag{2}$$

Namely, if this is not the case, then we could replace all the holes of $x$ and $x'$ by $1$ and obtain a square in the original full word $\Lambda$, which contradicts with Theorem 2. Note also that $|x| > 1$, since by the construction there are no full squares and no factors $\diamond\diamond$ in $\hat{\Lambda}$.

Hence, let us now assume that there exists a position $i$ satisfying (2). Assume first that the position is neither the first nor the last position of the word $x$. If $x'(i) = 0$, then $x'(i+1)$ can not be a hole. Thus, we must have $x'(i)x'(i+1) = x'(i)x(i+1) = 00$, which contradicts with Theorem 2. Similarly, if $x'(i) = 2$, then $x'(i-1) \neq \diamond$ and $22$ occurs in $\hat{\Lambda}$. Again, by Theorem 2, this is not possible.

Let us then consider the case where $i = 1$, *i.e.*, the first letter of $x$ in the partial square $xx'$ or $x'x$ is a hole satisfying (2). Since $|x| > 1$ and $00$ does not occur in $\hat{\Lambda}$, the word $x'$ must begin with $20$. Moreover, it follows that a prefix of $x'$ must be contained in $z = 20212012$. Namely, for the partial square $xx'$, there is no suitable position such that $x'$ could begin inside $\varphi(0)$. On the other hand, in the case of the partial square $x'x$ we know that $x'$ ends with $012$. However, the word $z$ is not a factor of $\Lambda$, since it does not occur in any of the blocks $\varphi(0)$, $\varphi(1)$, $\varphi(2)$ and in any pairwise catenation of these block. Consequently, no factor of $\hat{\Lambda}$ is contained in $z$, which gives a contradiction.

Finally, let us assume that $i = |x|$, *i.e.*, the last position of $x$ in the partial square $xx'$ or $x'x$ is a hole satisfying (2). Using similar reasoning as above, we conclude that the suffix of $x'$ must be contained in $0120$. Now we have two possibilities. Either $i$ is a position in $\varphi(20)$ or in $\varphi(10)$. In the former case the only position where $x'$ can end is the $11$th letter of $\varphi(1)$. Hence, $x'$ ends with $21020120$ whereas $x$ ends with $01020120$, which is a contradiction. In the latter case the last letter of $x'$ is either the third letter of $\varphi(1)$ or the $10$th letter of $\varphi(2)$. Now the suffix of $x$ must be $20210120$ and the suffix of $x'$ is either $01210120$ or $10210120$. Once more we have a contradiction. Thus, we have proved that the partial word $\hat{\Lambda}$ is square-free. Finally, there are uncountably many required words, since any hole in $\hat{\Lambda}$ can be replaced by $1$ and we obtain a square-free word. □

# References


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

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

[3] V. Halava, T. Harju, T. Kärki, Relational codes of words, Theoret. Comput. Sci. 389(1-2) (2007) 237–249.

[4] V. Halava, T. Harju, T. Kärki, Overlap-freeness in infinite partial words, submitted to Theoret. Comput. Sci.

[5] J. Leech, A problem on strings of beads, Math. Gazette 41 (1957) 277–278.

[6] F. Manea, R. Mercaş, Freeness of partial words, Theoret. Comput. Sci. 389(1-2) (2007) 265–277.

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

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

[9] J.-P. Allouche, J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999, 1–16.

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 978-952-12-2097-5
ISSN 1239-1891