A generalization of Thue freeness for partial words

By: Francine Blanchet-Sadri, Robert Mercaş, and Geoffrey Scott

Made available courtesy of Elsevier: http://www.elsevier.com

Abstract:
This paper approaches the combinatorial problem of Thue freeness for partial words. Partial words are sequences over a finite alphabet that may contain a number of "holes". First, we give an infinite word over a three-letter alphabet which avoids squares of length greater than two even after we replace an infinite number of positions with holes. Then, we give an infinite word over an eight-letter alphabet that avoids longer squares even after an arbitrary selection of its positions are replaced with holes, and show that the alphabet size is optimal. We find similar results for overlap-free partial words.
Keywords: Combinatorics on words, Partial words, Thue–Morse word, Freeness, Square-freeness, Overlap-freeness

Article:
1. Introduction
A well-known result from Thue [10,11] states that over a three-letter alphabet there exist infinitely many square-free words, and that over a binary alphabet there exist infinitely many overlap-free words. Because his results were published in obscure Norwegian journals, they remained unknown for a long time and were independently rediscovered by Arshon in 1937 and by Morse and Hedlund between 1938 and 1944. For more information see [2].

A partial word is a sequence of symbols over a finite alphabet that may contain a number of "holes". The first studies of partial words in 1999 by Berstel and Boasson were motivated by gene alignment [1,3]. In this paper, we explore basic freeness properties of partial words, generalizing the well-known freeness properties of full words.

The contents of our paper are as follows: In Section 2, we review concepts to be used throughout the paper. In Section 3, we introduce the concept of a non-trivial square and construct an infinite partial word over a ternary alphabet containing an arbitrary number of holes and avoiding all squares except ones of the form $a\lozenge$ or $\lozenge a$. There, we also construct an infinite word over an eight-letter alphabet that remains non-trivial-square-free after an arbitrary selection of its positions are replaced by holes. Furthermore, we show that there is no infinite word over a smaller alphabet satisfying this property. In Section 4, we study overlap-free partial words. We show that the word constructed in Section 3 is overlap-free, and we give bounds for the alphabet size necessary to construct an infinite word which remains overlap-free after an arbitrary selection of its positions are replaced by holes. We end the paper with a series of conclusions and suggestions for future work.

2. Preliminaries
Let A be a nonempty finite set, called an *alphabet*. An element $a \in A$ is called a letter or symbol. A finite word $w = a_0 ... a_{n-1}$, of length n is a finite concatenation of symbols $a_i \in A$, for $0 \le i < n$. The length of w is denoted by $|w|$. The *empty word*, denoted by ε, is the unique word of length zero. The set of all finite words over an alphabet A is denoted by A^*. It is a monoid under the associative binary operation defined by concatenation of words, with ε serving as the identity element, and it is referred to as the *free monoid* over A. Similarly, the set of

all nonempty words over A is denoted by A^+. It is a semigroup under the operation of concatenation, and is called the *free semigroup* over A. We denote by A^n the set of all words of length n over A.

The ith-power of a word $\underline{w}$ is defined recursively as
$$w^i = \begin{cases} \varepsilon & \text{if } i = 0 \\ ww^{i-1} & \text{if } i > 0. \end{cases} \tag{1}$$
A *partial word* of length n over A is defined using a partial function $w : \{0, \dots, n-1\} \to A$. For each $i < n$, we say that i is in the *domain* of w, denoted by $D(w)$, if $w(i)$ is defined. Otherwise, we say that i is in the *hole set* of w, denoted by $H(w)$. A word, or *full word*, is a partial word with an empty hole set.

Given a partial word w, the *companion* of w, denoted $w_\Diamond$, is the total function $w_\Diamond : \{0, \dots, n-1\} \to A \cup \{\Diamond\}$ defined by
$$w_\Diamond(i) = \begin{cases} w(i) & \text{if } i \in D(w) \\ \Diamond & \text{otherwise} \end{cases} \tag{2}$$
where $\Diamond$ is a new symbol which is not in the alphabet A, and which acts as a "do not know" symbol. If $A_\Diamond = A \cup \{\Diamond\}$, then the set of all finite partial words over A is denoted by $A_\Diamond^*$, and the set of all partial words of length n over A is denoted by $A_\Diamond^n$. Because the map $w \mapsto w_\Diamond$ is bijective, we can extend our definitions of concatenation and powers to partial words intuitively.

A partial word u is a *factor* of a partial word w if there exist, possibly empty, partial words x, y such that $w = xuy$. We say that u is a prefix of w, denoted by $u \le w$, if $x = \varepsilon$. Similarly, we say that u is a *suffix* of w if $y = \varepsilon$. A *factorization* of a partial word w is a sequence of partial words $w_0, w_1, \dots, w_i$ such that $w = w_0 w_1 \dots w_i$.

Partial words u and v are equal if $|u| = |v|$, $D(u) = D(v)$, and $u(i) = v(i)$ for all $i \in D(u)$. If $|u| = |v|$, $D(u) \subset D(v)$, and $u(i) = v(i)$ for all $i \in D(u)$, then u is said to be *contained* in v, denoted by $u \subset v$. We say u and v are *compatible*, denoted by $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$. We note that $u \uparrow v$ implies $v \uparrow u$. A nonempty partial word w is called *primitive* if there does not exist a partial word u such that $w \subset u^k$, for $k \ge 2$.

A *morphism* is a mapping $\phi : A^* \to B^*$ that satisfies $\phi(xy) = \phi(x)\phi(y)$, for all $x, y \in A^*$, where A and B denote alphabets. Since A^* is a free monoid, ϕ is completely defined by $\phi(a)$, for all $a \in A$, and $\phi(\varepsilon) = \varepsilon$. We say that a morphism ϕ is *prolongable* on $a \in A$, if $\phi(a) = aw$, where $w \in A^*$.

The most frequently used method to define infinite words is that of iterating a morphism. More precisely, we assume that $\phi : A^* \to A^*$ is a morphism prolongable on $a \in A$. Consequently, $\phi^i(a)$ is a prefix of $\phi^{i+1}(a) = \phi(\phi^i(a))$ with $i \ge 1$. Thus, the limit (the infinite word) $w = \lim_{i \to \infty} \phi^i(a)$ exists. This infinite word is said to be defined by iterating ϕ, and has the property of being a fixed point of the morphism ϕ.

Example 1. (*The Thue–Morse word*). Let $\phi : \{a, b\}^* \to \{a, b\}*$ be the morphism defined by $\phi(a) = ab$ and $\phi(b) = ba$. We define $\phi^0(a) = a$ and $\phi^{i+1}(a) = \phi(\phi^i(a))$ with $i \ge 0$, and note that $\phi^{i+1}(a) = \phi^i(a)\overline{\phi^i(a)}$, where $\bar{x}$ is the word obtained from x by replacing each occurrence of a with b and each occurrence of b with a. We define the Thue–Morse word by $\tau = \lim_{i \to \infty} \phi^i(a)$. The Thue–Morse word is a fixed point for the morphism ϕ.

We say that an infinite word w is *k-free* if there exists no word x such that x^k is a factor of w and $x \ne \varepsilon$. A word is called *overlap-free* if it does not contain any factor of the form $ayaya$ with $a \in A$. It is clear that any word w which is overlap-free is also k-free, for $k \ge 3$. For simplicity, a word that is 2-free is said to be *square-free*, and a word that is 3-free is said to be *cube-free*.

We call the partial word w *k-free*, if for any factor $x_0 x_1 \dots x_{k-1}$ of w there does not exist a partial word u such that $x_i \subset u$ for all $0 \le i < k$. A partial word is called *overlap-free* if it does not contain any factor of the form $a_0 w_0 a_1 w_1 a_2$ where a_0, a_1, a_2 are pairwise compatible symbols of $A_\Diamond$ and w_0, w_1 are compatible partial words in

$A_\Diamond^*$.

Throughout the paper, we will use the following well-known result regarding the Thue–Morse infinite word defined in Example 1.

Theorem 1 *(Thue Theorem [10,11]). The Thue–Morse word is overlap -free.*
The infinite partial words we describe in this paper are obtained from infinite full words by applying morphisms.

3. Square-freeness

A well-known result from Thue [10,11] states that over a three-letter alphabet there exist infinite words that are square-free. To generalize Thue's result, we wish to find a square-free partial word with infinitely many holes, and an infinite full word that remains square-free even after replacing an arbitrary selection of positions with holes. Unfortunately, every partial word containing at least one hole and having length at least two contains a square (either $a\Diamond$ or $\Diamond a$ cannot be avoided, where a denotes a letter from our alphabet).

Motivated by these observations, a partial word u such that $u \subset w^2$ for some word w is called a *square*. A *trivial square* is one of the form $a\Diamond$ or $\Diamond a$ or $\Diamond ab\Diamond$ for any distinct letters *a*, *b*. Any other square is called a *non-trivial square*. We call a word *non-trivial square*-free if it contains no non-trivial squares. Concepts similar to non-trivial squares have been investigated in the context of full words. In [9], several iterating morphisms are given for infinite words avoiding large squares. In particular, the authors give an infinite binary word avoiding squares yy with $|y| \geq 4$ and an infinite binary word avoiding all squares except 0^2, 1^2, and $(01)^2$ using a construction that is somewhat simpler than the original one from Fraenkel and Simpson [6].

Definition 1. Inserting a hole is defined as replacing a letter with a hole in a fixed position of a word (the length of the word remains the same). We impose the restriction that holes should be sparse in the sense that every two holes must have at least two non-hole symbols between them.

Without imposing this restriction, it would always be possible to obtain non-trivial squares of the form $\Diamond^2$ and $\Diamond a \Diamond b \Diamond c$, where *a*, *b*, *c* are letters of the alphabet. If the restriction on squares of length 4 would not be imposed, then we would have to change the restriction regarding how close a hole can be to another hole.

With these restrictions, the study of square-free partial words becomes much more subtle and interesting. In Section 3.1 we find a partial word with infinitely many holes avoiding all squares except ones of the form $a\Diamond$ or $\Diamond a$, and in Section 3.2 we find an infinite full word that remains non-trivial-square-free even after replacing an arbitrary selection of positions with holes. As a visual aid, throughout this section, we will underline the first and $(n + 1)$th symbol in a factor that is a square of length $2n$.

3.1. Square -free partial words
The following theorem gives an infinite word over a three-letter alphabet that avoids all squares except ones of the form $a\Diamond$ or $\Diamond a$ after we replace an infinite number of its positions with holes.

Theorem 2. *There exist infinitely many infinite partial words with infinitely many holes over a three-letter alphabet that do not contain any squares other than squares of the form $\Diamond a$ or $a\Diamond$.*

Proof. Let σ be the fixed point of the morphism $\phi : \{a, b, c\}^* \rightarrow \{a, b, c\}^*$ with $\phi(a) = abc$, $\phi(b) = ac$ and $\phi(c) = b$. In [7], it is shown that σ is square-free. We can define the word σ' by applying a morphism δ on the word σ that replaces a with $\phi^4(a)'$, b with $\phi^4(b)$, and *c* with $\phi^4(c)$ where

$$\phi 4(a) = abcacbabcbac\underline{a}bcacbacabcb$$
and
$$\phi 4(a)' = abcacbabcbac\underline{\Diamond}bcacbacabcb.$$

Here the a representing the 13th symbol of $\phi^4(a)$ is changed into a $\lozenge$. Set $\sigma = a_0 a_1 \ldots$, and let $\sigma' = b_0 b_1 \ldots$ be the partial word $\delta(\sigma)$. We claim that σ' satisfies the desired property.

First, σ' contains no squares of length less or equal to 4 other than $c\lozenge$ and $\lozenge b$. To see this, it is enough to check the word $bac\lozenge bca$. Now, assume that σ' contains a non-trivial square. Then there exist integers $i \geq 0$, $k > 0$ such that $b_i b_{i+1} \ldots b_{i+k-1} \uparrow b_{i+k} b_{i+k+1} \ldots b_{i+2k-1}$. Since σ itself is square-free, the square in σ' must contain a hole. If $k < 7$, then the square factor is also a factor of $\phi^4(a)'$. It can be checked explicitly that $\phi^4(a)'$ is non-trivial-square-free. Therefore, $k > 7$. We proceed by showing that if $b_{i+j} = \lozenge$, then $b_{i+k+j} \in \{\lozenge, a\}$ and that if $b_{i+k+j} = \lozenge$, then $b_{i+j} \in \{\lozenge, a\}$. This will show that every hole in σ' can be filled with the letter a while preserving the square factor in the word. However, the result of filling all holes in σ' with the letter a is the square-free word σ, so we will arrive at a contradiction. Since both implications are proved using the same logic, we will only show that if $b_{i+j} = \lozenge$, then $b_{i+k+j} \in \{\lozenge, a\}$. Let us consider the possibilities where the hole can appear. Suppose $b_{i+j} = \lozenge$.

- If $0 \leq j < k - 2$, then $b_{i+j} \ldots b_{i+j+2} = \lozenge bc$. It is easy to check, by looking at the description of ϕ, that the only factors of σ' compatible with $\lozenge bc$ are $\lozenge bc$ and abc. Since $b_{i+k+j} \ldots b_{i+k+j+2}$ must be compatible with $\lozenge bc$, it follows that $b_{i+k+j} \in \{\lozenge, a\}$.

- If $5 \leq j < k$, then $b_{i+j-5} \ldots b_{i+j} = bcbac\lozenge$. It is easy to check that the only factors of σ' compatible with $bcbac\lozenge$ are $bcbac\lozenge$ and $bcbaca$. Therefore, $b_{i+k+j} \in \{\lozenge, a\}$.

- If $k - 2 \leq j < 5$, then $b_{i+j-1} \ldots b_{i+j+1} = c\lozenge b$. Since the only factors of σ' compatible with $c\lozenge b$ are $c\lozenge b$ and cab, it follows that $b_{i+k+j} \in \{\lozenge, a\}$.

Corollary 1. *There exist infinitely many infinite partial words with an arbitrary number of holes over a three-letter alphabet that do not contain any squares other than squares of the form $\lozenge a$ or $a\lozenge$.*

Proof. If not all a's are replaced by $\phi^4(a)'$ (some could be replaced by $\phi^4(a)$ instead), then we get the result with an arbitrary number of holes.

3.2. Generalization of square -freeness

We now turn our attention to words that remain non-trivial-square-free after replacing an arbitrary collection of their positions with holes. Here, we give an infinite word over an eight-letter alphabet that remains non-trivial-square-free even after an arbitrary selection of its positions are replaced with holes, and show that the alphabet size of eight is optimal.

We begin by stating an obvious remark that will be used several times throughout this section.

Remark 1. Let $t_0 = a_0 a_1 a_2$ and $t_1 = b_0 b_1 b_2$ be full words. It is possible to insert holes into t_0 and t_1 such that the resulting partial words are compatible if and only if there exists i such that $a_i = b_i$ (or the letters in position i of t_0 and t_1 are equal). This is due to Definition 1 that states that every two holes must have at least two non-hole symbols between them.

To insert holes in $t_0 = abb$ and $t_1 = acc$ in order to make them compatible (with the convention in Definition 1), one can create $t_0' = a\lozenge b$ and $t_1' = ac\lozenge$ respectively. However, this is impossible when $t_0 = abb$ and $t_1 = bcc$.

Proposition 1. *Let t be a full word over an alphabet A. If every factor of length n of t contains n distinct elements of A, then it is impossible to insert holes into t such that the resulting partial word contains a non-trivial square $w_0 w_1$ with $w_0 \uparrow w_1$ and $|w_0| = |w_1| < n$.*

Proof. All positions i and $i + k$ have a different letter for $3 \leq k < n$ and thus the position i or $i + k$ must gain a hole. So there must be two holes at distance 1 or 2.

Theorem 3. *There exists an infinite word over an eight-letter alphabet that remains non-trivial-square-free after an arbitrary insertion of holes.*

Proof. Let σ be the fixed point of the morphism $\phi : \{a, b, c\}^* \to \{a, b, c\}^*$ with $\phi(a) = abc$, $\phi(b) = ac$ and $\phi(c) = b$. Recall that σ is square-free.

We construct the desired word t by applying a uniform morphism δ on the word σ that replaces
- *a* with *defghijk*,
- *b* with *deghfkij*, and
- *c* with *dehfgjki*.

We claim that *t* satisfies our desired properties.

Assume that it is possible to change a selection of positions in *t* to holes such that the resulting partial word *t'* contains a non-trivial square. It is clear that *t'* has no factors compatible with *aa* or *abab* other than ones of the form $a\Diamond$ or $\Diamond a$ or $\Diamond ba\Diamond$, for any letters $a, b \in \{d, e, f, g, h, i, j, k\}$. Therefore, we can restrict our attention to factors of the form $w_0 w_1$ with $w_0 \uparrow w_1$ and $|w_0| = |w_1| \geq 3$. That is, if $t = a_0 a_1 a_2 \ldots$ and $t' = b_0 b_1 b_2 \ldots$, then there exist $i \geq 0$ and $k \geq 3$ such that
$$b_i b_{i+1} b_{i+2} \ldots b_{i+k-1} \uparrow b_{i+k} b_{i+k+1} b_{i+k+2} \ldots b_{i+2k-1}.$$

There are two cases to be analyzed:

Case 1. $k \equiv 0 \bmod 8$
Setting $k = 8(m + 1)$, note that $a_i a_{i+1} a_{i+2} \ldots a_{i+k-1}$ is of the form
$$w_{00} \delta(c_0) \delta(c_1) \ldots \delta(c_{m-1}) w_{01}$$
and $a_{i+k} a_{i+k+1} a_{i+k+2} \ldots a_{i+2k-1}$ is of the form
$$w_{10} \delta(c_{m+1}) \delta(c_{m+2}) \ldots \delta(c_{2m}) w_{11}$$
with $w_{01} w_{10} = \delta(c_m)$, $|w_{pr}| = |w_{qr}|$ with $w_{pr}, w_{qr} \in \{d, e, f, g, h, i, j, k\}^*$ and $c_l \in \{a, b, c\}$ for all $p, q, r \in \{0, 1\}$ and $l \in \{0, 1, \ldots, 2m\}$.

Also note that if $c_p \neq c_{m+p+1}$ for any $0 \leq p < m$, then it is impossible to insert holes into $\delta(c_p)$ and $\delta(c_{m+p+1})$ such that the resulting partial words are compatible. Therefore, $c_p = c_{m+p+1}$ for all $0 \leq p < m$.

If $|w_{01}| \geq 5$, then by Remark 1, w_{11} must be a prefix of $\delta(c_m)$. Hence,
$$\underline{c_0}\, c_1 \ldots c_{m-1} c_m \underline{c_{m+1}} c_{m+2} \ldots c_{2m} c_m$$
is a factor of σ. Since σ is square-free and $c_p = c_{m+p+1}$ for $0 \leq p < m$, this is a contradiction. If $|w_{01}| < 5$, then $|w_{00}| \geq 4$ and it follows that w_{00} and w_{10} are suffixes of $\delta(c_m)$. Then
$$\underline{c_m} c_0 c_1 \ldots c_{m-1} \underline{c_m} c_{m+1} c_{m+2} \ldots c_{2m}$$
is a factor of σ. Since σ is square-free, this is a contradiction.

Case 2. $k \not\equiv 0 \bmod 8$
Suppose that $a_{i+l} = a_{i+k+l} = d$ for some $0 \leq l < k - 4$. Then the words
$$a_{i+l+1} \ldots a_{i+l+4} \text{ and } a_{i+k+l+1} \ldots a_{i+k+l+4}$$
can only be *efgh*, *eghf*, or *ehfg*. Since $k \not\equiv 0 \bmod 8$, it follows that $a_{i+l+1} \ldots a_{i+l+4}$ is different from $a_{i+k+l+1} \ldots a_{i+k+l+4}$. However, if we select any two different strings from *efgh*, *eghf* and *ehfg*, it is easy to see that they cannot be made compatible through the introduction of holes. Therefore, it is clear that $b_{i+l+1} \ldots b_{i+l+4}$ is not compatible with $b_{i+k+l+1} \ldots b_{i+k+l+4}$. This contradicts with the assumption that
$$b_i b_{i+1} b_{i+2} \ldots b_{i+k-1} \uparrow b_{i+k} b_{i+k+1} b_{i+k+2} \ldots b_{i+2k-1}.$$

Therefore, there is no *l* satisfying $0 \leq l < k - 4$ such that $a_{i+l} = a_{i+k+l} = d$. In fact, this argument remains true if we replace the letter *d* with any letter in the set $\{d, e, f, g, h, i, j, k\}$. Thus, there exists no *l* satisfying $0 \leq l < k - 4$ such that $a_{i+l} = a_{i+k+l}$. By Remark 1, it follows that $a_{i+l} = a_{i+k+l}$ for some $0 \leq l < 3$. If $k \geq 7$, this same l would

satisfy $0 \leq l < k - 4$. Therefore, $k < 7$.

We observe that every factor of length six of t contains no repeated letters. By Proposition 1, it follows that $k = 6$. Every factor of length 12 in t is contained in $\delta(c_1)\delta(c_2)\delta(c_3)$ for some $c_i \in \{a, b, c\}$. We used a computer program to check that it is impossible to insert holes into any of the above factors to create a square.

Since all cases lead to contradiction we conclude that t satisfies the desired properties.

Remark 2. Note that the word t' constructed in Theorem 3 is also cube-free.

Of course, it is natural to ask whether such a word can be constructed over a smaller alphabet. This question is intimately related to the study of full words of the form $v_0 a w a v_1$, where $a \in A$ and $v_0, v_1, w \in A^*$.

Proposition 2. *Let $t = v_0 a w a v_1$ be a full word over the alphabet A, where $a \in A$ and $v_i, w \in A^*$. If any of the following hold, then it is possible to insert holes into t so that the resulting partial word contains a non-trivial square:*

1. $|w| = 2$ *and* $|t| \geq 6$,
2. $|w| = 3$, $|t| \geq 8$ *and* $|v_i| \geq 1$,
3. $|w| = 4$ *and* $|v_i| \geq 2$,
4. $|w| = 5$, $|t| \geq 15$, $|v_i| \geq 4$ *and* $|A| \leq 7$.

Proof. Let $b_i \in A$. For Statement 1, if t has factors of the form $ab_0b_1ab_2b_3$, $b_0ab_1b_2ab_3$, or $b_0b_1ab_2b_3a$, then by replacing b_0 and b_3 with holes into t we get partial words containing factors that are squares of the form $\underline{a}\Diamond \underline{b_1 a b_2}\Diamond$, $\Diamond \underline{ab_1 b_2}\underline{a}\Diamond$, or $\Diamond \underline{b_1 a b_2}\Diamond a$ respectively.

For Statement 2, if t has a factor $b_0ab_1b_2b_3ab_4b_5$ or $b_0b_1ab_2b_3b_4ab_5$, we can insert holes into t such that the resulting partial word has square factors $\Diamond \underline{ab_1}\Diamond \underline{b_3}a\Diamond b_5$ or $\underline{b_0}\Diamond ab_2 \Diamond \underline{b_4}a\Diamond$ respectively.

For Statement 3, if t has a factor of the form $b_0b_1ab_2b_3b_4b_5ab_6b_7$, we can insert holes into t such that the resulting partial word has the square factor $\Diamond \underline{b_1}a\Diamond \underline{b_3}\underline{b_4}\Diamond ab_6\Diamond$.

For Statement 4, if t has a factor of the form
$$b_0b_1b_2b_3ab_4b_5b_6b_7b_8ab_9b_{10}b_{11}b_{12}$$
then we argue as follows. If $b_i = b_j$ for any $4 \leq i < j < 9$, then by the previous three statements we can insert holes into the factor such that the resulting partial word contains a non-trivial square (note that if $j = i + 1$ or $j = i + 2$, we could create the non-trivial squares $\underline{b_i b_i}$ or $\underline{b_i}\underline{b_i}b_k$ for some k). For the same reason, $b_i \neq a$ for $4 \leq i < 9$. Therefore, we assume that the letters b_i for $4 \leq i < 9$ are pairwise nonequal and distinct from a. Similarly, we can assume that $b_9 \neq b_i$ for $5 \leq i < 9$ and $b_9 \neq a$. If $b_9 = b_4$, then we can insert holes into the factor such that the resulting partial word contains the square $\Diamond \underline{b_3}ab_4\Diamond b_6\underline{b_7}\Diamond ab_9b_{10}\Diamond$. Thus, the letters b_i for $4 \leq i < 10$ are pairwise nonequal and distinct from a. Using the same logic, the letters b_i for $3 \leq i < 9$ are pairwise nonequal and distinct from a. Since $\|A\| \leq 7$, we must have $b_3 = b_9$.

Next, b_{10} must be distinct from a and b_i for $6 \leq i < 10$, so either $b_{10} = b_5$ or $b_{10} = b_4$. If $b_{10} = b_5$, we can insert holes into the factor such that the resulting partial word contains the non-trivial square $\Diamond \underline{b_3}a\Diamond b_5 b_6\underline{b_7}\Diamond ab_9b_{10}\Diamond$. Therefore, $b_{10} = b_4$. Using the same logic, we find that $b_2 = b_8$.

Finally, we can insert holes into our factor such that the obtained partial word contains the non-trivial square $\Diamond \underline{b_2}b_3\Diamond \underline{b_4}b_5\Diamond \underline{b_7}b_8\Diamond b_9b_{10}\Diamond b_{12}$.

Corollary 2. *Let t be an infinite word over an alphabet A such that any partial word obtained by inserting holes in t is non-trivial-square-free. Then $||A|| \geq 8$.*

Proof. Let t be an infinite word over the alphabet $A = \{a_0, a_1, \ldots, a_6\}$, where $a_i \neq a_j$ for all $0 \leq i < j \leq 6$. If t has a factor of the form $v_0 a w a v_1$, where $a \in A$, $v_i, w \in A^*$, $2 \leq |w| \leq 5$ and $|v_i| \geq 4$, then according to the previous proposition it is possible to introduce holes into t to create square factors (note that if $|w| = 1$, then we can replace w with $\lozenge$ to create a non-trivial square of the form $a\lozenge ab$). To avoid this, t must have a factor of the form
$$\underline{a_0} a_1 a_2 a_3 a_4 a_5 a_6 \underline{a_0} a_1 a_2 a_3 a_4 a_5 a_6$$
up to an isomorphism between the letters. This implies that t contains squares that will certainly be preserved when holes are added. Therefore, at least eight letters are needed to create an infinite word satisfying our conditions.

4. Overlap-freeness

In this section we will extend the concept of overlap-freeness to partial words. We use the standard definition of overlap-freeness given in the preliminaries, but we still adhere to the restriction described in Definition 1 when replacing an arbitrary selection of positions in a word with holes. In Section 4.1 we find an overlap-free partial word with infinitely many holes, and in Section 4.2 we find an infinite full word that remains overlap-free even after replacing an arbitrary selection of positions with holes. As a visual aid, we will underline the a_i's of the overlapping factor $\underline{a_0} w_0 \underline{a_1} w_1 \underline{a_2}$ to distinguish an overlap present in a sequence of letters.

4.1. Overlap -free partial words

In [8], the question was raised as to whether there exist overlap-free infinite partial words, and to construct them over a binary alphabet if such exist. In this section, we construct overlap-free infinite partial words with one hole over a two-letter alphabet and show that none exists with more than one hole. In addition, we show that there exist infinitely many overlap-free infinite partial words with an arbitrary number of holes over a three-letter alphabet.

Proposition 3. *There exist overlap-free infinite binary partial words containing one hole.*

Proof. Recall that the Thue–Morse word is overlap-free. We claim that the Thue–Morse word preceded by a hole, $\lozenge\tau$, is also overlap-free. Let ϕ be the Thue–Morse morphism. Because r is overlap-free, any overlap occurring in or must contain the hole. It suffices, therefore, to show that $\lozenge\phi^i(a)$ is overlap-free for any positive i. Note that
$$\phi^{i+3}(a) = \phi^i(a) \overline{\phi^i(a)} \overline{\phi^i(a)} \phi^i(a) \overline{\phi^i(a)} \phi i(a) \phi^i(a) \overline{\phi^i(a)}$$
contains a copy of both a $\phi^i(a)$ and $b\phi^i(a)$ (due to the factors $\overline{\phi^i(a)} \phi^i(a)$ and $\phi^i(a) \phi^i(a)$). Since the Thue–Morse word is overlap-free, $\phi^{i+3}(a)$ is as well. Therefore, neither $a\phi^i(a)$ nor $b\phi^i(a)$ contain an overlap. This implies that $\lozenge\phi^i(a)$ is overlap-free. It is clear that $\lozenge\overline{\tau}$ is overlap-free as well.

Remark 3. Over a binary alphabet all words of length greater than six with a hole in the third position contain an overlap. To see this, note that if the partial word has a factor of the form $a\lozenge a$, $aa\lozenge$ or $\lozenge aa$, then it clearly contains an overlap. Therefore, we can assume that any overlap-free binary word with a hole in the third position has a prefix of the form $ab\lozenge ab$. If this factor is followed by aa, then the word contains the overlap $\underline{a}b\lozenge\underline{a}ba\underline{a}$. Similarly, if the factor is followed by ab, ba, or bb, it will contain $\underline{\lozenge}a\underline{b}ab$, $\underline{a}b\lozenge\underline{a}bb\underline{a}$, or bbb respectively.

Proposition 4. *There is no infinite overlap-free binary partial word with more than one hole.*

Proof. To see this, note that by Remark 3, an infinite overlap-free binary partial word cannot contain a hole after the second position. However, it cannot contain holes in both the first and second positions, as an overlap of the form $\lozenge\lozenge a$ would clearly appear. Thus, only one hole is allowed.

We now prove that the word given in Theorem 2 is also overlap-free.

Proposition 5. *There exist infinitely many overlap-free infinite partial words with an arbitrary number of holes over a three-letter alphabet.*

Proof. In Theorem 2 we showed that the word σ' constructed there does not contain any squares other than squares of the form $\lozenge b$ or $c\lozenge$. Because σ is square-free and hence overlap-free, any overlap in σ' must contain a hole. So it remains only to show that σ' contains no overlaps of the form $a_0a_1a_2$ with $a_0, a_1, a_2 \subset b$ or $a_0, a_1, a_2 \subset c$. However, any such overlapping factor is so small that it would be contained in $\phi^4(a)'$. It is easy to check that $\phi^4(a)'$ does not contain any such overlapping factor.

4.2. Generalization of overlap -freeness

In the previous section, we gave infinite words that are overlap-free even after selected symbols in the words were changed to holes. In this section, we give infinite overlap-free words over a six-letter alphabet that remain overlap-free even after an arbitrary selection of their positions are changed to holes, and show that none exists over a four-letter alphabet.

Proposition 6. *There is no infinite word over a four-letter alphabet that remains overlap-free after an arbitrary selection of its positions are changed to holes.*

Proof. Assume that such a word t exists over the four-letter alphabet A. Clearly, it contains no factors of the form bba or bab where $a, b \in A$, since holes could be introduced to form the overlap factors $bb\lozenge$ and $b\lozenge b$ respectively. If the word contains no factor of the form ba_0a_1b where $a_i \in A$ for all i, then every factor of t of length four contains no repeated letters. This would imply that t is of the form

$$... \, a_0a_1a_2a_3a_0a_1a_2a_3a_0a_1a_2a_3....$$

Therefore, we can assume that t has a factor of the form $a_0a_1a_2ba_3a_4ba_5a_6$, where $b \neq a_i$, for all $i > 0$. If $a_5 = a_3$, then $\underline{\lozenge}ba_3\underline{a_4}ba_5\underline{\lozenge}$ is an overlap (symmetrically, $a_4 \neq a_2$ to avoid the overlap $\underline{\lozenge}a_2b\underline{\lozenge}a_4b\underline{a_5}$). Therefore, $a_5 \neq a_3$. Similarly, we must have $a_5 \neq a_4$, $a_5 \neq a_6$, $a_4 \neq a_6$ and $a_4 \neq a_3$ to avoid the overlaps $a_4\lozenge a_5$, $\lozenge a_5a_6$, $\underline{\lozenge}ba_3\underline{a_4}b\lozenge\underline{a_6}$, and $\lozenge a_3a_4$ respectively. Since b, a_4, a_5 and a_6 are pairwise nonequal, they must be four different letters. Since A is a four-letter alphabet and a_3 is distinct from b, a_4 and a_5, it follows that $a_3 = a_6$.

We use similar logic to determine that $a_1 = a_4$. We arrive at our desired contradiction by introducing holes to get the overlap $\underline{\lozenge}a_1a_2\lozenge\underline{a_3}a_4\lozenge a_5\underline{a_6}$

This proposition gives us a lower bound of five for a minimum alphabet size necessary to construct a word that is overlap-free after an arbitrary selection of its positions are changed to holes.

Theorem 4. *There exists an infinite word over a six-letter alphabet that remains overlap -free after an arbitrary insertion of holes.*

Proof. Let σ be the fixed point of the morphism $\phi : \{a, b, c\}^* \to \{a, b, c\}^*$ with $\phi(a) = abc$, $\phi(b) = ac$ and $\phi(c) = b$. Recall that σ is square-free.

We construct the desired word t by applying a uniform morphism δ on the word σ that replaces

- a with *defghi*,
- b with *degifh*, and
- c with *dehfig*.

We claim that t satisfies our desired properties.

Assume that it is possible to change a selection of positions in t to holes such that the resulting partial word t'

contains an overlap. That is, if $t = a_0a_1a_2 \ldots$ and $t' = b_0b_1b_2 \ldots$ then there exist integers $i \geq 0$, $k > 0$ such that
$$b_ib_{i+1}b_{i+2} \ldots b_{i+k-1} \uparrow b_{i+k}b_{i+k+1}b_{i+k+2} \ldots b_{i+2k-1}$$
and b_i, b_{i+k} and b_{i+2k} are pairwise compatible.

There are two cases to be analyzed:

Case 1. $\mathrm{k} \equiv 0 \bmod 6$

Setting $k = 6(m + 1)$, note that $a_ia_{i+1}a_{i+2} \ldots a_{i+k-1}$ is of the form
$$w_{00}\delta(c_0)\delta(c_1) \ldots \delta(c_{m-1})w_{01}$$
and $a_{i+k}a_{i+k+1}a_{i+k+2} \ldots a_{i+2k-1}$ is of the form
$$w_{10}\delta(c_{m+1})\delta(c_{m+2}) \ldots \delta(c_{2m})w_{11}$$
with $w_{01}w_{10} = \delta(c_m)$, $|w_{pr}| = |w_{qr}|$ with w_{pr}, $w_{qr} \in \{d, e, f, g, h, i\}^*$ and $c_l \in \{a, b, c\}$ for all $p, q, r \in \{0, 1\}$ and $l \in \{0, 1, \ldots, 2m\}$.

If $c_p \neq c_{m+p+1}$ for any $0 \leq p < m$, then it is impossible to insert holes into $\delta(c_p)$ and $\delta(c_{m+p+1})$ such that the resulting partial words are compatible. Therefore, $c_p = c_{m+p+1}$ for all $0 \leq p < m$.

If $|w_{01}| \geq 5$, then by Remark 1, w_{11} must be a prefix of $\delta(c_m)$ (otherwise, the third, fourth, and fifth letters of w_{01} are not equal to the third, fourth, and fifth letters of w_{11} respectively, and it would be impossible to introduce holes into w_{01} and w_{11} to make them compatible). Hence,
$$\underline{c_0}c_1 \ldots c_{m-1}c_m\underline{c_{m+1}}c_{m+2} \ldots c_{2m}c_m$$
is a factor of σ. Since σ is square-free and $c_p = c_{m+p+1}$ for $0 \leq p < m$, this is a contradiction.

If $|w_{01}| \leq 3$, then $|w_{00}| \geq 3$ and it follows that w_{00} is a suffix of $\delta(c_m)$. Then
$$\underline{c_m}c_0c_1 \ldots c_{m-1}\underline{c_m}c_{m+1}c_{m+2} \ldots c_{2m}$$
is a factor of σ. Since σ is square-free, this is a contradiction.

The only remaining case is $|w_{01}| = 4$. Let a' be the first letter of w_{10}. Then $w_{01}a'$ and $w_{11}a_{i+2k}$ can be made compatible by the insertion of holes. Again by Remark 1, $w_{11}a_{i+2k}$ is a prefix of $\delta(c_m)$ and we see that
$$\underline{c_0}c_1 \ldots c_{m-1}c_m\underline{c_{m+1}}c_{m+2} \ldots c_{2m}c_m$$
is again a factor of σ, a contradiction.

Case 2. $\mathrm{k} \not\equiv 0 \bmod 6$

Throughout this case, we assume that $k \geq 5$. Because every overlap with $k < 5$ is contained in a factor of the form $\delta(c_0)\delta(c_1)\delta(c_2)$ for $c_i \in \{a, b, c\}$ where $c_0c_1c_2$ is a factor of σ, it can be checked exhaustively that t' contains no overlaps with $k < 5$.

Suppose that $a_{i+l} = a_{i+k+l} = h$ for some $0 \leq l < k - 2$. Then the words
$$a_{i+l+1} \ldots a_{i+l+3} \text{ and } a_{i+k+l+1} \ldots a_{i+k+l+3}$$
can only be def, deh, ide, or fig. Since $\mathrm{k} \not\equiv 0 \bmod 6$, it follows that $a_{i+l+1} \ldots a_{i+l+3}$ and $a_{i+k+l+1} \ldots a_{i+k+l+3}$ are nonequal, and we do not have the case where one of $a_{i+l+1} \ldots a_{i+l+3}$ and $a_{i+k+l+1} \ldots a_{i+k+l+3}$ is def while the other is deh. Because of Remark 1, it is clear that $b_{i+l+1} \ldots b_{i+l+3}$ is not compatible with $b_{i+k+l+1} \ldots b_{i+k+l+3}$. This either contradicts with the assumption that
$$b_ib_{i+1}b_{i+2} \ldots b_{i+k-1} \uparrow b_{i+k}b_{i+k+1}b_{i+k+2} \ldots b_{i+2k-1}$$
or with the assumption that b_{i+k} and b_{i+2k} are compatible.

Therefore, there is no l satisfying $0 < l < k - 2$ such that $a_{i+l} = a_{i+k+l} = h$. In fact, it is easy to check that the argument presented above remains true if we replace the letter h with the letter f, g, or i. Also note that because $k \not\equiv 0 \bmod 6$, there is no l satisfying $a_{i+l} = a_{i+k+l} = d$ or $a_{i+l} = a_{i+k+l} = e$ (since this is possible only if $k \equiv 0 \bmod 6$

for both d and e). However, note that by Remark 1, there must be some $0 < l < 2$ satisfying $a_{i+l} = a_{i+k+l}$. Since $k \geq 5$, it follows that this l satisfies $0 < l < k - 2$, a contradiction.

Since all cases lead to contradiction we conclude that t remains overlap-free after an arbitrary selection of its positions are replaced with holes.

5. Conclusion
The paper extends in a natural way the concepts of square- and overlap-freeness of words to partial words. Some of the problems left open in [8] are solved here: (1) An overlap-free infinite partial word over a binary alphabet is proved to be easily constructible using the algorithms from [8]. It is also shown that this kind of words will contain at most one hole. (2) The existence of overlap-free infinite partial words over a three-letter alphabet and containing an infinity of holes is proven.

In addition, we have shown that there exists an infinite overlap-free word over a six-letter alphabet in which we can randomly replace positions by holes and obtain in this way an infinite partial word that is overlap-free, and have proved that such a word does not exist over a four-letter alphabet. The case of a five-letter alphabet remains open.

Conjecture 1. *There exists an infinite word over a five-letter alphabet that remains overlap-free after an arbitrary insertion of holes.*

As a direction for future work, we propose the extension of the concept of square-free (respectively, overlap-free or cube-free) morphism to partial words. From [8] and this paper, some of the properties of this kind of morphisms have already started to be obvious. An even further analysis might bring us more properties that such a morphism should fulfill. Following the approach of Dejean [5], another interesting problem to analyze would be the identification of the exact value of k (related to k-freeness) for a given alphabet size. This value would represent the repetition threshold in an n-letter alphabet. If for full words this value has been investigated [4], for partial words this value has not yet been looked into.

References
[1] J. Berstel, L. Boasson, Partial words and a theorem of Fine and Wilf, Theoret. Comput. Sci. 218 (1999)135–141.
[2] D.R. Bean, A. Ehrenfeucht, G.F. McNulty, Avoidable patterns in strings of symbols, Pacific J. Math. 85 (1979) 261–294.
[3] F. Blanchet-Sadri, Algorithmic Combinatorics on Partial Words, Chapman & Hall/CRC Press, 2007.
[4] A. Carpi, On Dejean's conjecture over larger alphabets, Theoret. Comput. Sci. 385 (2007)137–151.
[5] F. Dejean, Sur un théorème de Thue, J. Combin. Theory Ser. A 13 (1972) 90–99.
[6] A.S. Fraenkel, J. Simpson, How many squares must a binary sequence contain?, Electron. J. Combin. 2 339 #R2 (1995).
[7] M. Lothaire, Combinatorics on Words, Cambridge University Press, 1997.
[8] F. Manea, R. Mercaş, Freeness of partial words, Theoret. Comput. Sci. 389 (2007) 265–277.
[9] N. Rampersad, J. Shallit, M.-W. Wang, Avoiding large squares in infinite binary words, Theoret. Comput. Sci. 339 (2005)19–34.
[10] A. Thue, Über unendliche zeichenreihen. Norske Vid. Selsk. Skr.I, Mat. Nat. Kl. Christiana 7 (1906)1–22; Reprinted in T. Nagell, A. Selberg, S. Selberg, K. Thalberg (Eds.), Selected Mathematical Papers of Axel Thue, Universitetsforlaget, Oslo, Norway, 1977, pp. 139–158.
[11] A. Thue, Über die gegenseitige lage gleicher teile gewisser zeichenreihen. Norske Vid. Selsk. Skr.I, Mat. Nat. Kl. Christiana 1(1912)1–67; Reprinted in T. Nagell, A. Selberg, S. Selberg, K. Thalberg (Eds.), Selected Mathematical Papers of Axel Thue, Universitetsforlaget, Oslo, Norway, 1977, pp. 413–478.