Avoidable Binary Patterns in Partial Words*

F. Blanchet-Sadri[1], Robert Mercaş**[2], Sean Simmons[3], and Eric Weissenstein[4]

[1] Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402–6170, USA, blanchet@uncg.edu

[2] GRLMC, Departament de Filologies Romàniques, Universitat Rovira i Virgili, Av. Catalunya 35, Tarragona, 43002, Spain, robertmercas@gmail.com

[3] Department of Mathematics, The University of Texas at Austin, 2515 Speedway Rm 8, Austin, TX 78712–0233, USA

[4] Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Amos Eaton 301, 110 8th Street, Troy, NY 12180, USA

Abstract. The problem of classifying all the avoidable binary patterns in words has been completely solved (see Chapter 3 of M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, 2002). Partial words represent sequences that may have some undefined positions called holes. In this paper, we show that, if we do not substitute any variable of the pattern by a trivial partial word consisting of only one hole, the avoidability index of the pattern remains the same as in the full word case.

1 Introduction

A *pattern* p is a word over an alphabet E of *variables*, denoted by $\alpha, \beta, \gamma, \ldots$, and the associated set, over a finite alphabet A, is built by replacing p's variables with non-empty words over A so that the occurrences of the same variable be replaced with the same word.

The concept of *unavoidable pattern*, see Section 2, was introduced, in the context of full words, by Bean, Ehrenfeucht and McNulty [1] (and by Zimin who used the terminology "blocking sets of terms" [2]). Although they characterized such patterns (in fact, avoidability can be decided using the Zimin algorithm by reduction of patterns), there is no known characterization of the patterns unavoidable over a k-letter alphabet (also called k-unavoidable). An alternative is to find all unavoidable patterns for a fixed alphabet size. The unary patterns, or powers of a single variable α, were investigated by Thue [3,4]: α is unavoidable, $\alpha\alpha$ is 2-unavoidable but 3-avoidable, and α^m with $m \geq 3$ is 2-avoidable. Schmidt proved that there are only finitely many binary patterns, or patterns over $E =$

* This material is based upon work supported by the National Science Foundation under Grant No. DMS–0754154. The Department of Defense is also gratefully acknowledged. A World Wide Web server interface has been established at www.uncg.edu/cmp/research/unavoidablesets4 for automated use of the program.

** Author's work was partially supported by the Research Grant No. 1323 U07 E30 N-2008/InvAct/Bel, G./BJ01 of the University Rovira i Virgili.

$\{\alpha, \beta\}$, that are 2-unavoidable [5, 6]. Later on, Roth showed that there are no binary patterns of length six or more that are 2-unavoidable [7]. The classification of unavoidable binary patterns was completed by Cassaigne [8], who showed that $\alpha\alpha\beta\beta\alpha$ is 2-avoidable.

In this paper, our goal is to classify binary patterns with respect to partial word *non-trivial avoidability.* A partial word is a sequence of symbols from a finite alphabet that may have some undefined positions, called holes, and denoted by $\diamond$'s, and a pattern is called non-trivial if none of its variables is substituted by only one hole. Here $\diamond$ is *compatible* with, or matches, every letter of the alphabet. In this context, in order for a pattern p to occur in a partial word, it must be the case that for each variable α of p, all its substituted partial words be pairwise compatible.

The contents of our paper is as follows: In Section 2, we start our investigation of avoidability of binary patterns in partial words. There, we explain that in order to classify all binary patterns with respect to our concept of non-trivial avoidability, we are left with studying five patterns. In Section 4 using iterated morphisms, we construct infinite binary partial words with infinitely many holes that avoid the patterns $\alpha\beta\alpha\beta\alpha$, $\alpha\beta\alpha\beta\beta\alpha$ and $\alpha\alpha\beta\alpha\beta\beta$. In Section 5 using non-iterated morphisms, we construct such words that avoid the patterns $\alpha\alpha\beta\beta\alpha$ and $\alpha\beta\alpha\alpha\beta$. This concludes the fact that all binary patterns 2-avoidable in full words are also non-trivially 2-avoidable in partial words.

We end this section with some preliminaries. For more information regarding concepts on partial words, the reader is referred to [9].

Let A be a non-empty finite set of symbols called an *alphabet.* Each element $a \in A$ is called a *letter*. A *(full) word* over A is a sequence of letters from A. A *partial word* over A is a sequence of symbols from $A_\diamond = A \cup \{\diamond\}$, the alphabet A being augmented with the "hole" symbol $\diamond$ (a full word is a partial word without holes). We denote by $u(i)$ the symbol at position i of a partial word u. The *length* of u is denoted by $|u|$ and represents the number of symbols in u. The *empty word* is the sequence of length zero and is denoted by ε. The set containing all full words (respectively, non-empty full words) over A is denoted by A^* (respectively, A^+), while the set of all partial words (respectively, non-empty partial words) over A is denoted by $A_\diamond^*$ (respectively, $A_\diamond^+$).

If u and v are two partial words of equal length, then u is said to be *contained in* v, denoted $u \subset v$, if $u(i) = v(i)$ for all i such that $u(i) \in A$. Partial words u and v are *compatible*, denoted $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$. If u and v are non-empty compatible partial words, then uv is called a *square*.

A partial word u is a *factor* of a partial word v if there exist x, y such that $v = xuy$ (the factor u is *proper* if $u \neq \varepsilon$ and $u \neq v$). We say that u is a *prefix* of v if $x = \varepsilon$ and a *suffix* of v if $y = \varepsilon$.

2 Avoidability on Partial Words

Let E be a non-empty finite set of symbols, distinct from A, whose elements are denoted by α, β, γ, etc. Symbols in E are called *variables*, and words in E^* are called *patterns*. For the remaining of this paper, we only consider binary alphabets and patterns, hence we can fix $A = \{a, b\}$ and $E = \{\alpha, \beta\}$. Moreover, we define $\overline{a} = b$ and $\overline{b} = a$, and similarly $\overline{\alpha} = \beta$ and $\overline{\beta} = \alpha$.

The *pattern language*, over A, associated with a pattern $p \in E^*$, denoted by $p(A_\diamond^+)$, is the subset of $A_\diamond^*$ containing all partial words compatible with $\varphi(p)$, where φ is any non-erasing morphism from E^* to A^*. A partial word $w \in A_\diamond^*$ *meets* the pattern p (or p *occurs* in w) if for some factorization $w = xuy$, we have $u \in p(A_\diamond^+)$. Otherwise, w *avoids* p.

To be more precise, let $p = \alpha_0 \cdots \alpha_m$, where $\alpha_i \in E$ for $i = 0, \ldots, m$. Define an *occurrence* of p in a partial word w as a factor $u_0 \cdots u_m$ of w, where for all $i, j \in \{0, \ldots, m\}$, if $\alpha_i = \alpha_j$, then $u_i \uparrow u_j$. We call such an occurrence *non-trivial* if $u_i \neq \diamond$, for all $i \in \{0, \ldots, m\}$. We call a word *non-trivially p-free* if it contains no non-trivial occurrences of p. Note that these definitions also apply to (one-sided) infinite partial words w over A which are functions from $\mathbb{N}$ to $A_\diamond$.

Considering the pattern $p = \alpha\beta\beta\alpha$, the language associated with p over the alphabet $\{a, b\}$ is $p(\{a, b, \diamond\}^+) = \{u_1 v_1 v_2 u_2 \mid u_1, u_2, v_1, v_2 \in \{a, b, \diamond\}^+$ such that $u_1 \uparrow u_2$ and $v_1 \uparrow v_2\}$. The partial word $ab\diamond ba\diamond bba$ meets p (take $\varphi(\alpha) = bb$ and $\varphi(\beta) = a$), while the word $\diamond babbbaaab\diamond$ avoids p.

Let p and p' be two patterns. If p' meets p, then p *divides* p', which we denote by $p \mid p'$. For example, $\alpha\alpha \nmid \alpha\beta\alpha$ but $\alpha\alpha \mid \alpha\beta\alpha\beta$.

A pattern $p \in E^*$ is *k-avoidable* if there are infinitely many partial words in $A_\diamond^*$ that avoid p, where A is any alphabet of size k. On the other hand, if every long enough partial word in $A_\diamond^*$ meets p, then p is *k-unavoidable* (it is also called unavoidable over A). The *avoidability index* of p is the smallest integer k such that p is k-avoidable, or is ∞ if p is unavoidable.

Remark 1. Let $p, p' \in E$ be such that p divides p'. If an infinite partial word avoids p, then it also avoids p'.

In the context of full words all binary patterns' avoidability index have been characterized [10]. Since a full word is a partial word without holes, the avoidability index of a binary pattern in full words is not greater than the avoidability index of that pattern in partial words. Thus, all unavoidable binary patterns in full words have avoidability index ∞ in partial words as well.

In [11], it was shown that there exist infinitely many partial words with infinitely many holes over a 3-letter alphabet that non-trivially avoid $\alpha\alpha$, and so the avoidability index of $\alpha\alpha$ in partial words is 3. Since in full words all binary patterns with avoidability index 3 are divisible by $\alpha\alpha$, using Remark 1 we conclude that all 3-avoidable binary patterns in full words also have avoidability index 3 in the context of partial words.

Thus, according to Theorem 3.3.3 of [10], if we can find the avoidability index of $\alpha\alpha\alpha$, $\alpha\beta\alpha\beta\alpha$, $\alpha\beta\alpha\beta\beta\alpha$, $\alpha\alpha\beta\alpha\beta\beta$, $\alpha\beta\alpha\alpha\beta$ and $\alpha\alpha\beta\beta\alpha$, then we will have completed the classification of the binary patterns in terms of avoidability in partial words.

First let us recall that in [12], the case of patterns of the form α^m, $m \geq 3$, was considered, the avoidability index in partial words being 2. Furthermore, in [11, 13] it was shown that the pattern $\alpha\beta\alpha\beta\alpha$ is trivially 2-unavoidable, but it is 3-avoidable in partial words.

In this paper, our main result is the following.

Theorem 1. *With respect to non-trivial avoidability in partial words, the avoidability index of a binary pattern is the same as in the full word case.*

3 Binary Patterns 2-Avoidable by Iterated Morphisms

Let us recall the iterative Thue-Morse morphism ϕ such that $\phi(a) = ab$ and $\phi(b) = ba$. It is well known that $\phi^\omega(a)$ avoids $\alpha\beta\alpha\beta\alpha$ [14].

Proposition 1. *Over a binary alphabet there exist infinitely many infinite partial words, containing exactly one hole, that non-trivially avoid $\alpha\beta\alpha\beta\alpha$.*

Proof. Let $p = \alpha\beta\alpha\beta\alpha$, and t be the fixed point of the Thue-Morse morphism. We show that there exist infinitely many positions in t in which one can replace the letter at that position with a hole and obtain a new word t' that is still non-trivially p-free. Also, since all factors of the infinite Thue-Morse word t (powers of ϕ) avoid p, it follows that any occurrence of p in t' must contain the hole.

Let $x_1, x_2, x_3 \subset x$ and $y_1, y_2 \subset y$, for some partial words $x_1, x_2, x_3, x, y_1, y_2, y$ such that $|x|, |y| \geq 1$. We start by proving that there does not exist a non-trivial occurrence of p, $x_1y_1x_2y_2x_3$, in t' such that $|x| \geq 8$ or $|y| \geq 8$. We proceed by contradiction. We analyze several cases based on the possible positions of the hole.

Assume that the hole is in x_1. Note that this case is symmetrical to when the hole is in x_3 (we are implicitly using the fact that if w is a factor of the Thue-Morse word t, then so is $\text{rev}(w)$). Since t is overlap-free, it follows that the only possibility is to have in t a factor of the form $x'cx''yx'\bar{c}x''yx'\bar{c}x''$, with $c \in \{a, b\}$, and $x_1 = x'\diamond x''$, $x_2 = x_3 = x'\bar{c}x'' = x$ and $y_1, y_2 = y$, for some words $x', x'' \in \{a, b\}^*$ with $|x'x''| \geq 7$ or $|y| > 7$ (moreover, x' is non-empty since, otherwise, t would contain the factor $x''y\bar{c}x''y\bar{c}x''$ which is impossible since t is p-free). Looking at the symbols that precede and follow c in x_1 and $\bar{c}$ in x_2, we get that if $|x'x''| \geq 7$ either $\bar{c}c\bar{c}c\bar{c}$ is a factor of $x_2 = x\bar{c}y$ when c is preceded by c in xcy, or $c\bar{c}c\bar{c}c$ is a factor of $x_1 = xcy$ when c is preceded by $\bar{c}$ in xcy, and if $|y| > 7$ either $c\bar{c}c\bar{c}c$ is a factor of $x_1 = xcy$ when c is preceded by c in x_1, or $\bar{c}c\bar{c}c\bar{c}$ is a factor of $x_2 = x\bar{c}y$ when c is preceded by $\bar{c}$ in x_1. All cases lead to contradiction with the fact that t is overlap-free.

Let us illustrate by an example how this works. Let us consider the case when c is preceded by a c, $|x'| = 1$ and $|y| > 7$. We look at the factors x_1y_1 and x_2y_2 that differ at only one position. We have that y_1 starts with $\bar{c}$, such that ccc is not a prefix of our factor. It follows that $x''y_2$ starts with $\bar{c}c$ such that we do not get the cube $\overline{ccc}$ in t. But, in x_1y_1 we have the factor $cc\overline{c}c$, which must be followed by $\overline{c}$. Again looking at the prefix of $x''y_2$, we get $\overline{c}c\overline{c}c$. It follows that $ccx''y$ has as prefix $cc\overline{c}c\overline{c}c$ which contains an overlap, a contradiction with the fact that t is overlap-free.

Assume that the hole is in y_1. This case is symmetrical to when the hole is in y_2. In this case, since t is overlap-free, the only possibility is to have in t a factor of the form $xy'cy''xy'\overline{c}y''x$, with $c \in \{a,b\}$, and $x_1 = x_2 = x_2 = x$, $y_1 = y'\diamond y''$, $y_2 = y'\overline{c}y''$, for some words $y', y'' \in \{a,b\}^*$ with $|y'| + |y''| \geq 1$. Since at least one of y' and y'' is non-empty, and either $|x| > 7$ or $|y'y''| \geq 7$, the proof follows from the previous case.

Finally, assume that the hole is in x_2. Since t is overlap-free, the only possibility is to have in t a factor of the form $x'cx''yx'\overline{c}x''yx'cx''$, with $c \in \{a,b\}$, and $x_1 = x_3 = x'cx''$, $x_2 = x'\overline{c}x''$ and $y_1 = y_2 = y$, for some words $x', x'' \in \{a,b\}^*$ with $|x'| + |x''| \geq 1$. Again, since at least one of x' and x'' is non-empty, and either $|y| > 7$ or $|x'x''| \geq 7$, the proof follows from the first case.

We have thus shown that if $|x| \geq 8$ or $|y| \geq 8$, there are no non-trivial occurrences $x_1y_1x_2y_2x_3$, where $x_1, x_2, x_3 \subset x$ and $y_1, y_2 \subset y$, of p in t', therefore any such occurrence in t' must satisfy $|x_1y_1x_2y_2x_3| < 40$. We now put a hole at Position 47 of $t_7 = \phi^7(a)$:

$$t'_7 = abbabaabbaababbabaababbaabbabaabbaababbaabbabaa\underline{b}abbabaabbaababbab$$
$$aababbaabbabaababbabaabbaababbaabbabaabbaababbabaababbaabbabaab$$

It is not hard to verify (using a computer program) that there are no occurrences of p in t'_7. Moreover we have placed the hole more than 40 positions from either end of t_7. Note that t_7 occurs as a factor of t infinitely often. We can choose any arbitrary occurrence of the factor t_7 in t, and replace it with t'_7 in order to obtain an infinite word with one hole that non-trivially avoids p. □

Theorem 2. *Over a binary alphabet there exist infinitely many partial words, containing infinitely many holes, that non-trivially avoid the pattern* $\alpha\beta\alpha\beta\alpha$.

Proof. Let us give a sketch of the proof. We can obtain such words t', by replacing arbitrarily many occurrences of t_7 in t with the factor t'_7 obtained in the manner discussed above.

In order to prove the theorem we assume there exists some occurrence of p in t' having the form $x_1y_1x_2y_2x_3$, such that $x_1, x_2, x_3 \subset x$ and $y_1, y_2 \subset y$ for some words $x, y \in \{a,b\}^*$, and having the minimum number of holes possible. Because of the way we insert holes, there are at least 127 letters in between any two holes, hence, $|xy| \geq 43$. Moreover, since the number of holes is minimum, one of the inequalities $x_1 = x_2 = x_3$ or $y_1 = y_2$ must fail. This implies that the position corresponding to the hole in one of the other factors must be an a. By looking at the factors preceding and following the hole, respectively the ones preceding and following the a in the corresponding factor, we get into contradictions either regarding valid factors of w or the minimality of the number of holes. □

Let us move on and take ν to be the morphism that maps a to aab and b to bba. Define the sequence produced by ν as $t_0 = a$, and $t_n = \nu(t_{n-1})$. Recall that ν avoids $\alpha\beta\alpha\beta\beta\alpha$ and $\alpha\alpha\beta\alpha\beta\beta$ [10].

Proposition 2. *For any $n \geq 0$, $t_{n+1} = t_n t_n \overline{t_n}$, where t_i is the ith iteration of the sequence produced by ν.*

Proof. We proceed by induction. Note that $t_1 = aab = t_0 t_0 \overline{t_0}$. Assume $t_n = t_{n-1} t_{n-1} \overline{t_{n-1}}$, for some integer $n > 0$. Thus, $t_{n+1} = \nu(t_n) = \nu(t_{n-1} t_{n-1} \overline{t_{n-1}})$, and so $t_{n+1} = \nu(t_{n-1})\nu(t_{n-1})\nu(\overline{t_{n-1}}) = t_n t_n \overline{t_n}$. □

Proposition 3. *Over a binary alphabet there exist infinitely many infinite partial words, containing exactly one hole, that avoid the pattern $\alpha\beta\alpha\beta\beta\alpha$.*

Proof. Let $p = \alpha\beta\alpha\beta\beta\alpha$, and let t be the fixed point of the morphism ν. We show that there exist infinitely many positions in t in which one can replace the letter at that position with a hole and obtain a new word t' that is still non-trivially p-free. Note that since t avoids p, it follows that for p to occur in t' it must contain the hole.

First let us replace Position 58 of t_5 by a hole:

$$
\begin{aligned}
t_5 = \ & aabaabbbaaabaabbbabbabbaaabaabaabbbaaabaabbbabbabbaaabbbab\underline{b}aaabbbab\\
& baaabaabaabbbaaabaabbbaaabaabbbabbabbaaabaabaabbbaaabaabbbabbabbaa\\
& abbbabbaaabbbabbaaabaabaabbbabbabbaaabbbabbaaabaabaabbbabbabbaaabbb\\
& abbaaabaabaabbbaaabaabbbaaabaabbbabbabbaaab
\end{aligned}
$$

It is easy to verify with a computer program that the resulting word has no occurrences of p with $|\alpha|, |\beta| < 9$, since the hole is more than 58 positions from either end of t_5.

Let $x_1, x_2, x_3 \subset u$ and $y_1, y_2, y_3 \subset z$ for some partial words x, y such that $|x|, |y| \geq 1$. We prove that there does not exist an occurrence of the factor $x_1 y_1 x_2 y_2 y_3 x_3$ in t' such that $|x| \geq 9$ or $|y| \geq 9$.

Assume that the hole is in x_1 (the cases when the hole is in y_1, x_2 or y_2 are similar). Since $|x_1 y_1| > 9$, it follows that the hole is either preceded by bab or followed by aaa. Because of this, it must be the case that the a from $x_2 y_2$ corresponding to the hole (if there were an occurrence of b corresponding to the hole then t would contain p which is a contradiction), it is either preceded by bab or followed by aaa. We get that t either contains the factor $baba$ or the factor $aaaa$, a contradiction with the construction of t.

If the hole is in x_3 (the case when the hole is in y_3 is similar), since $|y_3 x_3| > 9$, it follows that the hole is either preceded by bab or followed by aaa. Comparing it to the a from $y_1 x_2$ corresponding to the hole, we have that the a is either preceded by bab or followed by aaa. We get once more that t either contains the factor $baba$ or the factor $aaaa$, a contradiction with the construction of t.

Note that t_5 occurs as a factor of t infinitely often. We can choose any arbitrary occurrence of the factor t_5 in t, and place a hole at Position 58 to obtain an infinite word with one hole that avoids p. □

Remark 2. If uu is a factor of t, the fixed point of the morphism ν, for some word u of length $|u| > 3$, it must be the case that $|u| = 0 \bmod 3$. Moreover, for all different occurrences of the same factor v of length $|v| > 3$, there exist unique words x, y, z such that $v = x\nu(y)z$, with $|x|, |z| < 3$. In other words, all occurrences of the same factor start at the same position of an iteration of ν.

Theorem 3. *Over a binary alphabet there exist infinitely many partial words, containing infinitely many holes, that avoid the pattern $\alpha\beta\alpha\beta\beta\alpha$.*

Proof. Let us denote by t_5' the word obtained by replacing the letter a at Position 58 by a $\diamond$ in t_5, and by t' the word where infinitely many occurrences of the factor t_5, that start at an even position in t, have been replaced by t_5'.

Assume, to get a contradiction, that the pattern occurs somewhere in t'. It must be the case that there exists a factor $x_1y_1x_2y_2y_3x_3$ that contains h holes, (the case $h = 1$ is proved in Proposition 3), with all x_i's and all y_i's pairwise compatible for all $i \in \{1, 2, 3\}$, and no occurrence of the pattern with less than h holes exists.

It is obvious that $h > 1$ according to the previous proposition. If there exists a hole in x_i and $|x_i| > 4$, for $0 < i \leq 3$, then there exists x_j, with $j \neq i$, that has a factor that is compatible with a word from $\{\underline{a}aaab, b\underline{a}aaa, ab\underline{a}aa, bab\underline{a}a, bbab\underline{a}\}$ (note that the underlined letter is the one that corresponds to the hole in x_i), and if x_j has a hole, then the hole does not correspond to the hole in x_i. Note that it is impossible to have a hole at another position than the underlined one, in any of the previously mentioned factors. We conclude that x_j has no holes. But, in this case we would have that t contains one of the factors $aaaa$, $abaaa$ or $baba$, which is a contradiction. The same proof works for y_i, where $0 < i \leq 3$.

Thus, either $|\alpha| \leq 4$ and y_i contains no holes for $0 < i \leq 3$, or $|\beta| \leq 4$ and x_i contains no holes for $0 < i \leq 3$ (otherwise we would have that $|x_1y_1x_2y_2y_3x_3| \leq 24$ contains more than two holes, which is a contradiction since between each two holes there are at least 72 symbols according to our construction).

Let us first assume that the hole is in α. If x_1 contains the hole then, since $|x_1| \leq 4$ and y_1 contains no hole, looking at the factor following $\diamond$ we conclude that the corresponding position in x_2 must also contain a hole. Now, if x_2 contains the hole then, it follows from the previous observation that x_1 has to contain a hole, and moreover, since y_1 and y_3 contain no holes, looking at the factor preceding the hole, we get that x_3 has a hole at the corresponding position. In the case x_3 has a hole, according to the previous observation, it must be the case that x_2 has a hole. We conclude that if x_i has a hole, then $x_i = x_j$, for all $i, j \in \{1, 2, 3\}$. Hence, there exists an occurrence of the pattern having no holes, a contradiction.

Since the case when the hole is in β is similar, we conclude that t' does not contain any occurrence of the pattern $\alpha\beta\alpha\beta\beta\alpha$. □

Proposition 4. *Over a binary alphabet there exist infinitely many infinite partial words, containing exactly one hole, that avoid the pattern $\alpha\alpha\beta\alpha\beta\beta$.*

The word is obtained by placing a hole at Position 57 of t_5:

$$t_5 = aabaabbbaaabaabbbabbabbaaabaabaabbbaaabaabbbabbabbaaabbba\underline{b}baaabbbab$$

baaabaabaabbbaaabaabbbaaabaabbbabbabbaaabaabaabbbaaabaabbbabbabbaa
abbbabbaaabbbabbaaabaabaabbbabbabbaaabbbabbaaabaabaabbbabbabbaaabbb
abbaaabaabaabbbaaabaabbbaaabaabbbabbabbaaab

The proof is similar to the one for the pattern $\alpha\beta\alpha\beta\beta\alpha$.

Theorem 4. *Over a binary alphabet there exist infinitely many partial words, containing infinitely many holes, that avoid the pattern* $\alpha\alpha\beta\alpha\beta\beta$.

The proof of this result is similar to the one for the pattern $\alpha\beta\alpha\beta\beta\alpha$.

4 Binary Patterns 2-Avoidable by Non-Iterated Morphisms

Let us now look at the pattern $\alpha\alpha\beta\beta\alpha$. Let $A = \{a, b\}$ and $A' = A \cup \{c\}$, and let $\psi : A'^* \to A'^*$ be the morphism defined by $\psi(a) = abc$, $\psi(b) = ac$ and $\psi(c) = b$. We know that $\psi^\omega(a)$ avoids $\alpha\alpha$, in other words it is square-free [10].

Furthermore, define the morphism $\chi : A'^* \to A^*$ such that $\chi(a) = aa$, $\chi(b) = aba$, and $\chi(c) = abbb$. If $w = \chi(\psi^\omega(a))$, then we know from [8] that w does not contain any occurrence of $\alpha\alpha\beta\beta\alpha$. Moreover, denote by χ_4 the application of χ to the fourth iteration of ψ. Since $\psi^4(a)$ occurs infinitely often as a factor of $\psi^\omega(a)$, it follows that $\chi_4(a) = \chi(\psi^4(a))$ occurs infinitely often as a factor of w. Hence, we can write $w = w_0\chi_4 w_1\chi_4 w_2\chi_4 \cdots$, for some words w_i with $|w_i| > 1$, for all i. Now, let us replace Position 23 of χ_4 by a $\diamond$ and denote the new partial word by χ_4':

$aaabaabbbaaabbbabaaaaba\underline{a}bbbabaaaabbbaaabaabbbaaabbbabaaaabbbaaabaabbbaba$

Lemma 1. *Let* u_1u_2 *denote a factor of* $w' = w_0\chi_4' w_1\chi_4' w_2\chi_4' \cdots$ *that was obtained by inserting holes in* v_1v_2*, a factor of* w *with* $u_1 \subset v_1$ *and* $u_2 \subset v_2$*. If* $u_1 \uparrow u_2$*, but* $v_1 \neq v_2$*, then* $|u_1| \leq 4$*, more specifically, either* u_1 *or* u_2 *is in* $\{\diamond, \diamond b, \diamond bb, a\diamond, a\diamond b, ba\diamond, a\diamond bb\}$.

Proof. Obviously, if $u_1 \uparrow u_2$ but $v_1 \neq v_2$, then a hole appears in u_1 and there is no hole at the corresponding position in u_2, or vice versa. Without loss of generality we can assume that the hole appears in u_1. Assume that $u_1 \notin \{\diamond, \diamond b, \diamond bb, a\diamond, a\diamond b, a\diamond bb, ba\diamond\}$. It follows that u_1 has as a factor $\diamond bbb$, $aba\diamond$ or $ba\diamond b$. Moreover, note that the only time bbb appears in w' is in the factors $abbba$ and $\diamond bbba$. Similarly, the only time aba appears as a factor of w' is as a factor of $abaa$ and $aba\diamond$, and the only time a word x compatible with $ba\diamond b$ appears in w' is when $x = ba\diamond b$ or $x = baab$. We see that in all of these cases the corresponding factor in u_2 must be $abbba$, $abaa$ or $baab$, a contradiction since we always have $v_1 = v_2$. □

Lemma 2. *There exists no factor* uu *of* $w = \chi(\psi^\omega(a))$*, such that either* $aaab$ *or* aba *is a prefix of* u.

Proof. Assume there exists an u with prefix $aaab$ so that $uu = w(i) \cdots w(i + 2l - 1)$, for some integers i, l with $l > 3$. Note that $aaab$ only appears as a prefix of $\chi(x)$, for some word $x \in \{a, b, c\}^+$. Moreover, since the second u also starts with $aaab$, we have that $u = \chi(x)$. Hence, uu is actually $\chi(xx)$, for some word $x \in \{a, b, c\}^+$. It follows that $\phi^\omega(a)$ contains the square xx, which is a contradiction with the nature of the ψ morphism. Similarly, aba only appears as an image of $\chi(b)$. □

Theorem 5. *Over a binary alphabet there exist infinitely many partial words, containing infinitely many holes, that avoid the pattern $\alpha\alpha\beta\beta\alpha$.*

Proof. Let $p = \alpha\alpha\beta\beta\alpha$. To prove this claim, we assume that w' is not p-free and get a contradiction. Let $x_1'x_2'y_1'y_2'x_3'$ be an occurrence of p in w', and denote by $x_1x_2y_1y_2x_3$ the factor of w in which holes were inserted to get p. Note that if $x_1 = x_2 = x_3$ and $y_1 = y_2$, then we have an occurrence of p in w, which would be a contradiction. Therefore one of the inequalities fails. Also, note that if $x_i \neq x_j$ then either x_i' or x_j' contains a hole, where $i, j \in \{1, 2, 3\}$, while if $y_1 \neq y_2$ then either y_1' or y_2' contains a hole. Moreover, if $x_1 \neq x_2$ or $y_1 \neq y_2$, according to Lemma 1, it must be the case that x_1 or x_2 or, y_1 or y_2 are in $\{\diamond, \diamond b, \diamond bb, a\diamond, a\diamond b, a\diamond bb, ba\diamond\}$. By looking at the factor χ_4', it is easy to check that the only possibilities are for x_1' and y_1' to be in $\{\diamond, \diamond b\}$ and for x_2' and y_2' to be in $\{\diamond, a\diamond, a\diamond b\}$.

If $y_1' = \diamond b$, it is easy to check that x_3' must start with aab. According to Lemma 2, we cannot have that $x_1 = x_2$. It follows, according to Lemma 1, that $|x_1'| \leq 4$, a contradiction with the factor preceding the hole. The proof is identical for the case when $x_1' = \diamond b$. If $y_1' = \diamond$, then x_3' starts with bba. Thus, x_2' starts with bba and x_1' ends in ab or $\diamond b$. It follows that x_2' ends in ab or $\diamond b$, thus, y_1' is preceded by ab, which is a contradiction. If $x_1' = \diamond$, then y_1' and y_2' start with bba and x_3' is a b. From Lemma 1 we get that $y_1 = y_2$. Thus, y_1y_2bb is a factor of w. It follows that for some word $x \in \{a, b\}^+$, $y_1y_2bb = bb\chi(xc)\chi(xc)$ is a factor of w. We get a contradiction with the fact that $\phi^\omega(a)$ is square-free.

If x_2' or y_2' are in $\{\diamond, a\diamond, a\diamond b\}$ then we get that either y_1' or x_3' are in $\{bbba, bba\}$. A contradiction is reached again with the help of Lemma 1 and the fact that $\phi^\omega(a)$ is square-free.

The final case that needs to be analyzed is when $x_1 = x_2$ and $y_1 = y_2$, and $x_3' \uparrow x_1$ and $x_3 \neq x_1$. Let us denote $x = x_1 = x_2$ and $y = y_1 = y_2$. We get that w' has $xxyyx_3'$ as a factor and there exists at least one hole in x_3' that corresponds to b's in x_1' and x_2'.

If $|x| > 4$ it follows that x has $aba\underline{b}$, $ba\underline{b}b$, $\underline{b}bbb$ as a factor (the underlined b represents the letter corresponding to the hole in x_3'). Since none of these are possible factors of w, we conclude that it is impossible. Hence, it must be the case that $x_3' \in \{\diamond, \diamond b, \diamond bb, a\diamond, a\diamond b, a\diamond bb, ba\diamond\}$. It follows that $xx \in \{b^2, (bb)^2, (bbb)^2, (ab)^2, (abb)2, (abbb)^2, (bab)^2\}$. But, only bb is a possible factor of w. It is easy to check that in this case $|y| > 6$ and we conclude that y has either aaa, aba, $baaa$ or $baba$ as a prefix. In all of these cases, using Lemma 2 we reach a contradiction.

Since all cases lead to contradictions we conclude that $\alpha\alpha\beta\beta\alpha$ is trivially-avoidable over a binary alphabet. □

Finally, let us look at the pattern $\alpha\beta\alpha\alpha\beta$. According to [10, Lemma 3.3.2], $\gamma(\psi^\omega(a))$ avoids $\alpha\beta\alpha\alpha\beta$, where $\gamma : \{a, b, c\}^* \to \{a, b\}^*$ with $\chi(a) = aaa$, $\chi(b) = bbb$, and $\chi(c) = ababab$. Moreover, the only squares that occur in $w = \gamma(\psi^\omega(a))$ are $a^2, b^2, (aa)^2, (ab)^2, (ba)^2, (bb)^2$ and $(baba)^2$. As a last thing, note that $\psi^\omega(a)$ does not contain any of the factors aba or cbc.

Now let us replace Position 84 of $\gamma_5 = \gamma(\psi^5(a))$ by a $\diamond$ and denote the new partial word by γ_5':

$$\begin{aligned}\gamma_5' =\, & aaabbbabababaaaabababbbbaaabbbabababbbbaaaabababaaabbbabababaaaabab\\ & abbbbaaaabababaa\underline{a}bbbabababbbbaaabbbabababaaaabababbbbaaabbbabababbbb\\ & aaaabababaaabbbabababbbbaaabbbabababaaaabababbbbaaaababab\end{aligned}$$

Moreover, let us denote by w' the word obtained from w after the insertion of a hole at Position 84 of an occurrence of γ_5.

Proposition 5. *Over a binary alphabet there exist infinitely many infinite partial words, containing exactly one hole, that non-trivially avoid $\alpha\beta\alpha\alpha\beta$.*

Proof. Let us assume, to get a contradiction, that there exists an occurrence of $p = \alpha\beta\alpha\alpha\beta$ in w', and denote this occurrence by $x_1y_1x_2x_3y_2$, for $x_i, y_j \in \{a, b\}^+$, $0 < i \leq 3$ and $0 < j < 3$. It must be the case that either $x_1 = x_2 = x_3 = x$ or $y_1 = y_2 = y$, for some words x and y, but not both. Note that if the variable containing the hole has length greater than 5 then the corresponding variable has as a factor $baa\underline{b}$, $aa\underline{b}bbb$ or $a\underline{b}bbba$ (the underlined b stands for the hole position in the first variable). Only the last of these words is a valid factor of w. Moreover, note that actually this represents the prefix of the variable since, having an extra symbol in front would give us the factor $aabbbb$ which is not a valid one for w.

If the hole is in one of the α's, note that, since $x_2 \uparrow x_3$ and w contains no squares of length greater than four, it must be the case that the hole is either in x_2 or x_3. That implies that w' has either the factor $abbbbax'ya\diamond bbbax'abbbbax'y$, or $abbbbax'yabbbbax'a\diamond bbbax'y$, where $x = abbbbax'$, for some non-empty word x'. In the first case, this implies that y starts with ab, giving us in w the square $bbbbax'ab\ bbbax'ab$, a contradiction. In the second case, since x' is non-empty and it is preceded by $\diamond bbba$, we conclude that x' starts with $abab$. Looking at the prefix of x_2, we get the factor $abbbbabab$, which is a contradiction with the fact that $\psi^\omega(a)$ does not contain cbc as a factor. If the hole is in β, then $x \in \{a, b, aa, ab, ba, bb, baba\}$. Thus, we get one of the factors $xa\diamond bbbay'xxabbbbay'$ or $xabbbbay'xxa\diamond bbbay'$, where $y = abbbbay'$ for some word y'. Note that, replacing x with any of the possible values gives us a contradiction with either the factor preceding the hole or the construction of w. We conclude that none of the cases are possible.

Hence, it must be that the variable containing the hole has length at most five. If we denote by z the variable containing the hole and by z' one of the variables corresponding to z, $z \uparrow z'$, we get that $z \in \{a\diamond, \diamond b, aa\diamond, a\diamond b, \diamond bb, aa\diamond b, a\diamond bb, \diamond bbb, aa\diamond bb, a\diamond bbb, \diamond bbba\}$ and $z' \in \{ab, bb, aab, abb, bbb, aabb, abbb, bbbb, aabbb, abbbb, bbbba\}$ (z' cannot contain the factor $baab$). By looking at the possible factors preceding and following the holes, and the squares that can be found in w, we conclude that if the hole is in x_1, then $x_1 \in \{a\diamond, \diamond b\}$, if the hole is in x_2, then $x_2 = \diamond b$, and it is impossible to have the hole in x_3.

If $x_1 = \diamond b$, then it follows that y has bb as a prefix, and a contradiction is reached with the fact that $bbbbb$ is a factor of w. If $x_1 = a\diamond$, it follows that y_1 starts with $bbbab$, and so, we get that $ababy_2$ determines in $\psi^\omega(a)$ the factor cbc, which is a contradiction.

If $x_2 = \diamond b$, then w contains the factor $bbabababy'\underline{a}bbbabababy'$ (the underlined letter is the one that we changed into a hole), where $y = abababy'$ for some word y'. It can be checked that in this case y' ends in $aabababaa$, and in order to avoid having the square $acac$ in $\psi^\omega(a)$, it must be the case that y' is always followed by ab. Hence, w has the factor $xyxxy$, where $x = b$ and $y = abababy'ab$, a contradiction.

If the hole is in β, since p has $\alpha\alpha$ as a factor, and the only possible squares in w are $a^2, b^2, (aa)^2, (ab)^2, (ba)^2, (bb)^2$ and $(baba)^2$, by looking at the possible factor preceding and following the hole, we conclude that y_1 cannot contain the hole and the only possibility for y_2 to contain the hole is when $y_2 = \diamond$. But, in this case the occurrence of the pattern is a trivial one, hence, we get a contradiction.

Since all cases lead to contradiction, the conclusion follows. □

Theorem 6. *Over a binary alphabet there exist infinitely many partial words, containing infinitely many holes, that non-trivially avoid the pattern $\alpha\beta\alpha\alpha\beta$.*

Proof. Let us denote by w' the word obtained from w after an infinite number of non-overlapping occurrences of γ_5 starting at an even position have been replaced by γ_5'. Furthermore, let us assume, to get a contradiction, that the pattern $p = \alpha\beta\alpha\alpha\beta$ is unavoidable and denote by $x_1y_1x_2x_3y_2$ an occurrence of p containing $h > 1$ holes, such that no occurrence of the pattern p having less than h holes appears in w'.

Since, according to Proposition 5, $h > 1$ and the distance between every two holes is at least 170, it follows that $|\alpha\beta| > 85$. Thus, there exist $z \in \{x_1, x_2, x_3, y_1, y_2\}$ and a variable $z' \in \{x_1, x_2, x_3, y_1, y_2\}$ distinct from z, with $z \uparrow z'$ and, $z = z_1\diamond z_2$ and $z' = z_1'bz_2'$, for some words z_i, z_i' with $z_i \uparrow z_i'$, for $0 < i < 3$. If $|z_1| > 2$, it follows that z_1' has a suffix compatible with baa. Since the only factor in w' compatible with baa is baa we conclude that $z_1'b$ has $baab$ as a suffix, which is a contradiction with the fact that $baab$ is a valid factor of w. It follows that $|z_1| < 3$. Moreover, if $z \in \{y_1, y_2\}$, since y_1 is preceded by x_1 and y_2 is preceded by x_3, we get $x_1y_1 \uparrow x_3y_3$. If $|\alpha| > 2$ a conclusion similar to the previous one is reached. It follows that $0 < |x_1z_1| < 3$. In this case $|z_2| > 82$ and we get that the hole is followed by $bbbab$. So the prefix of length five of bz_2', $bbbbab$, represents a factor of the image of $\gamma(cbc)$. This is a contradiction since, cbc is not a factor of $\psi^\omega(a)$. Thus, $z \in \{x_1, x_2, x_3\}$.

Note that, if for all $\diamond$'s in x_2 the corresponding position in x_3 is an a or a $\diamond$, or vice versa, then x_2, x_3 are compatible with an element of $\{aa, ab, ba, bb, baba\}$, since these are the only possible squares of length greater than one in w. Since none of these creates a valid factor, we conclude that there exists a $\diamond$ in x_2 such that the corresponding position in x_3 is a b, or vice versa.

Let us assume that $x_2 = z$ and $x_3 = z'$. The other case is similar. It follows that w' has $z_1 \diamond z_2 z_1' b z_2'$ as a factor, where $|z_1| < 3$. It can be checked that unless $x_2 = \diamond b$ and $x_3 = bb$, then $|z_2| > 5$. If z_2 has $bbbab$ as a factor, since the only factor of w' compatible with it is $bbbab$, we conclude that z_2' has $bbbab$ as a factor. This implies that the prefix of length six of bz_2', a factor of w', was determined by $\gamma(cbc)$, with cbc factor of $\psi^\omega(a)$, which is a contradiction. It must be the case that $x_2 = \diamond b$, $x_3 = bb$ and $x_1 \in \{\diamond b, bb\}$. If $x_1 = \diamond b$, it follows that y_1 has bb as a prefix. but, since $x_2 x_3 = \diamond bbb$, it follows that y_2 has ab as a prefix, a contradiction. Hence, it must be that $x_1 y_1 x_2 x_3 y_2 = bby_1 \diamond bbby_2$. But since all the holes in y_1 correspond to a's or $\diamond$'s in y_2, and vice versa, it follows that, replacing all holes but the one in x_2 we get an occurrence of the pattern having only one hole. This is a contradiction with Proposition 5. The conclusion follows. □

Since all these patterns prove to have a non-trivial avoidability index 2, the result of Theorem 1 follows.

References

1. Bean, D.R., Ehrenfeucht, A., McNulty, G.: Avoidable patterns in strings of symbols. Pacific J. Math. **85** (1979) 261–294
2. Zimin, A.I.: Blocking sets of terms. Math. USSR Sbornik **47** (1984) 353–364
3. Thue, A.: Über unendliche Zeichenreihen. Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana **7** (1906) 1–22 (Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, Norway (1977), pp. 139–158).
4. Thue, A.: Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana **1** (1912) 1–67 (Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, Norway (1977), pp. 413–478).
5. Schmidt, U.: Motifs inévitables dans les mots. Rapport LITP 86–63, Paris VI (1986)
6. Schmidt, U.: Avoidable patterns on two letters. Theoret. Comput. Sci. **63**(1) (1989) 1–17
7. Roth, P.: Every binary pattern of length six is avoidable on the two-letter alphabet. Acta Inform. **29**(1) (1992) 95–107
8. Cassaigne, J.: Unavoidable binary patterns. Acta Inform. **30** (1993) 385–395
9. Blanchet-Sadri, F.: Algorithmic Combinatorics on Partial Words. Chapman & Hall/CRC Press (2008)
10. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press (2002)
11. Blanchet-Sadri, F., Mercaş, R., Scott, G.: A generalization of Thue freeness for partial words. Theoret. Comput. Sci. **410**(8-10) (2009) 793–800
12. Manea, F., Mercaş, R.: Freeness of partial words. Theoret. Comput. Sci. **389**(1-2) (2007) 265–277
13. Halava, V., Harju, T., Kärki, T., Séébold, P.: Overlap-freeness in infinite partial words. Theoret. Comput. Sci. **410**(8-10) (2009) 943–948
14. Lothaire, M.: Combinatorics on Words. Cambridge University Press (1997)