Journal of Combinatorial Theory, Series A 119 (2012) 257–270

Contents lists available at SciVerse ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

Avoiding abelian squares in partial words

F. Blanchet-Sadri [a], Jane I. Kim [b], Robert Mercaş [c], William Severa [d], Sean Simmons [e], Dimin Xu [f]

[a] Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402-6170, USA
[b] Department of Mathematics, Columbia University, 2960 Broadway, New York, NY 10027-6902, USA
[c] GRLMC, Departament de Filologies Romàniques, Universitat Rovira i Virgili, Av. Catalunya 35, Tarragona, 43002, Spain
[d] Department of Mathematics, University of Florida, 358 Little Hall, P.O. Box 118105, Gainesville, FL 32611-8105, USA
[e] Department of Mathematics, The University of Texas at Austin, 1 University Station C1200, Austin, TX 78712-0233, USA
[f] Department of Mathematics, Bard College, 30 Campus Road, Annandale-on-Hudson, NY 12504, USA

ARTICLE INFO

Article history:
Received 13 January 2011
Available online 9 September 2011

Keywords:
Combinatorics on words
Partial words
Abelian squares
Morphisms
Parikh vectors

ABSTRACT

Erdős raised the question whether there exist infinite abelian square-free words over a given alphabet, that is, words in which no two adjacent subwords are permutations of each other. It can easily be checked that no such word exists over a three-letter alphabet. However, infinite abelian square-free words have been constructed over alphabets of sizes as small as four. In this paper, we investigate the problem of avoiding abelian squares in partial words, or sequences that may contain some holes. In particular, we give lower and upper bounds for the number of letters needed to construct infinite abelian square-free partial words with finitely or infinitely many holes. Several of our constructions are based on iterating morphisms. In the case of one hole, we prove that the minimal alphabet size is four, while in the case of more than one hole, we prove that it is five. We also investigate the number of partial words of length n with a fixed number of holes over a five-letter alphabet that avoid abelian squares and show that this number grows exponentially with n.

1. Introduction

Words or strings belong to the very basic objects in theoretical computer science. The systematic study of word structures (combinatorics on words) was started by Norwegian mathematician Axel Thue [31,32,2] at the beginning of the last century. One of the discoveries made by Thue is that the

E-mail address: blanchet@uncg.edu (F. Blanchet-Sadri).

0097-3165/$ – see front matter
doi:10.1016/j.jcta.2011.08.008

consecutive repetitions of non-empty factors (squares) can be avoided in infinite words over a three-letter alphabet. Recall that an infinite word w over an alphabet is said to be *k-free* if there exists no word x such that x^k is a factor of w. For simplicity, a word that is 2-free is said to be *square-free*. After Thue's time, repetition-free words have found applications in various research areas like commutative semigroups [9,12,17], formal languages [23,24], unending games [28], symbolic dynamics [27, 28], computer-assisted music analysis [22], cryptography [1,30], and bioinformatics [26].

Erdős [13] raised the question whether *abelian squares* can be avoided in infinitely long words, i.e., whether there exist infinite abelian square-free words over a given alphabet. An abelian square is a non-empty word uv, where u and v are permutations of each other. For example, *abcacb* is an abelian square. A word is called *abelian square-free*, if it does not contain any abelian square as a factor. For example, the word *abacaba* is abelian square-free, while *abcdadcada* is not (it contains the subword *cdadca*). It is easily seen that abelian squares cannot be avoided over a three-letter alphabet. Indeed, each word of length eight over three letters contains an abelian square. A first step in solving special cases of Erdős' problem was taken in [14], where it was shown that the 25th abelian powers were avoidable in the binary case. Later on, Pleasants [29] showed that there exists an infinite abelian square-free word over five letters, using a uniform iterated morphism of size fifteen. Justin [17] proved that over a binary alphabet there exists a word that is abelian 5-free, using a uniform morphism of size five. This result was improved by Dekking [11] to abelian 4-free, using a non-uniform morphism. Moreover, using $\mathbb{Z}_7$ instead of $\mathbb{Z}_5$, in the proof of this latter result, we get that over a ternary alphabet an abelian 3-free infinite word is constructible. The problem of whether abelian squares can be avoided over a four-letter alphabet was open for a long time. In [18], using an interesting combination of computer checking and mathematical reasoning, Keränen proves that abelian squares are avoidable on four letters. To do this, he presents an abelian square-free morphism $g:\{a,b,c,d\}^* \to \{a,b,c,d\}^*$ whose size is $|g(abcd)| = 4 \times 85$:

$$g(a) = \texttt{abcacdcbcdcadcdbdabacabadbabcbdbcbacbcdcacb}$$
$$\texttt{abdabacadcbcdcacdbcbacbcdcacdcbdcdadbdcbca}$$

and the image of the letters b, c, d, that is, the words $g(b)$, $g(c)$, $g(d)$, are obtained by cyclic permutation of letters in the preceding words. Moreover in [8], it is shown that no smaller uniform morphism works here! In [19] a completely new morphism of length 4×98, possessing similar properties for iterations, is given.

Now let us move to partial words. Being motivated by a practical problem on gene comparison, Berstel and Boasson introduced the notion of *partial words*, which are sequences over a finite alphabet that may have some undefined positions or holes (the $\diamond$ symbol represents a hole and matches every letter of the alphabet) [3]. For instance, $a\diamond bca\diamond b$ is a partial word with two holes over the three-letter alphabet $\{a, b, c\}$. Several interesting combinatorial properties of partial words have been investigated, and connections have been made, in particular, with problems concerning primitive sets of integers, lattices, vertex connectivity in graphs, etc., [4].

In [25], the question was raised as to whether there exist cube-free infinite partial words, and an optimal construction over a binary alphabet was given (a partial word w is called *k-free*, if for every factor $x_0x_1\cdots x_{k-1}$ of w there does not exist a word u, such that for each i, the defined positions of x_i match the corresponding positions of u). In [7], the authors settled the question of overlap-freeness by showing that over a two-letter alphabet there exist overlap-free infinite partial words with one hole and none exists with more than one hole, and that a three-letter alphabet is enough for an infinity of holes. An *overlap* represents a word consisting of two overlapping occurrences of the same factor. More precisely, an overlap is a factor of the form $a_0w_0a_1w_1a_2$, where w_0 and w_1 are compatible, $w_0, w_1 \in A_\diamond^*$, and a_i and a_j are compatible, $a_i, a_j \in A_\diamond$ for $0 \leqslant i, j \leqslant 2$ (for the definition of compatibility, see the section on preliminaries), and a partial word is called *overlap-free* if it does not contain any such factor. The problem of square-freeness in partial words is settled in [7] and [15] where it is shown that a three-letter alphabet is enough for constructing such words. Quite naturally, all the constructions of these words are done by iterating morphisms, most of them uniform, similarly or directly implied by the original result of Thue. Moreover, in [25,7,6], the concept of repetitions is also solved in more general terms. The authors show that, for given alphabets, replacing arbitrary positions of some infinite words by holes, does not change the repetition freeness (that is, if the original word is k-free for some k, then the constructed partial word is also k-free). Furthermore in [16], the authors show that there exist binary words that are 2-overlap-free.

This paper focuses on the problem of avoiding abelian squares in partial words. In Section 2, we give some preliminaries on partial words. In Section 3, we explore the minimal alphabet size needed for the construction of (two-sided) infinite abelian square-free partial words with a given finite number of holes. In particular, we construct an abelian square-free infinite partial word with one hole over the minimal four-letter alphabet. For more than one hole, the minimal number of letters is at least five, when such words exist. We also construct an infinite word over a 12-letter alphabet that remains abelian square-free even after an arbitrary position is replaced with a hole. In Section 4, we prove by explicit construction the existence of abelian square-free partial words with infinitely many holes. The minimal alphabet size turns out to be five for such words. In Section 5, we investigate in particular the number of partial words of length n over a five-letter alphabet that avoid abelian squares and show that this number grows exponentially with n. Finally in Section 6, we discuss some constructions for the finite case.

Several times in the paper, we mention properties to be checked with a computer. A World Wide Web server interface at

`www.uncg.edu/cmp/research/abelianrepetitions`

has been established for automated use of an abelian square-free checker, as well as an abelian square-free word generator. Our applets will help readers verify these properties.

2. Preliminaries

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 that does not contain the $\diamond$ symbol).

The *length* of a partial word w is denoted by $|w|$ and represents the number of symbols in w, while $w(i)$ represents the symbol at position i of w, where $0 \leqslant i < |w|$. The *empty word* is the sequence of length zero and is denoted by ε. The set of distinct letters in w, or the *alphabet* of w, is denoted by $\alpha(w)$. For instance, the partial word $w = ab\diamond bba\diamond$, where a, b are distinct letters of the alphabet A, satisfies $\alpha(w) = \{a, b\}$. The set of all words over A is denoted by A^*, while the set of all partial words over A is denoted by $A_\diamond^*$. Similarly, the set of all non-empty words over A is denoted by A^+, while the set of all non-empty partial words over A is denoted by $A_\diamond^+$. A (right) (respectively, two-sided) infinite partial word is a function $w : \mathbb{N} \to A_\diamond$ (respectively, $w : \mathbb{Z} \to A_\diamond$).

Let u and v be 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, then uv is called a *square*. Whenever we refer to a square uv, it implies that $u \uparrow v$.

A partial word u is a *factor* or *subword* of a partial word v if there exist x, y such that $v = xuy$. We say that u is a *prefix* of v if $x = \varepsilon$ and a *suffix* of v if $y = \varepsilon$. The prefix u of v is called *proper* if $u \neq v$. If $w = a_0a_1 \cdots a_{n-1}$, then $w[i..j) = a_i \cdots a_{j-1}$ and $w[i..j] = a_i \cdots a_j$. The *reversal* of a partial word $w = a_0a_1 \cdots a_{n-1}$, where each $a_i \in A_\diamond$, is simply the word written backwards $a_{n-1} \cdots a_1a_0$, and is denoted $\text{rev}(w)$. For partial words u and v, $|u|_v$ denotes the number of occurrences of v found in u. The *Parikh vector* of a word $w \in A^*$, denoted by $P(w)$, is defined as $P(w) = \langle |w|_{a_0}, |w|_{a_1}, \ldots, |w|_{a_{\|A\|-1}} \rangle$, where $A = \{a_0, a_1, \ldots, a_{\|A\|-1}\}$ (here $\|A\|$ denotes the cardinality of A).

A word $uv \in A^+$ is called an *abelian square* if $P(u) = P(v)$. A word w is *abelian square-free* if no factor of w is an abelian square.

Definition 2.1. A partial word $w \in A_\diamond^+$ is an *abelian square* if it is possible to substitute letters from A for each hole in such a way that w becomes an abelian square full word. The partial word w is *abelian square-free* if it does not have any full or partial abelian square, except those of the form $\diamond a$ or $a\diamond$, where $a \in A$ (which we call *trivial* abelian squares).

A *morphism* $\phi : A^* \to B^*$ is a mapping from the free monoid A^* to the free monoid B^* that satisfies $\phi(ww') = \phi(w)\phi(w')$ for all $w, w' \in A^*$; in particular $\phi(\varepsilon) = \varepsilon$. The *reversal* of ϕ is the morphism $\mathrm{rev}(\phi) : A^* \to B^*$ that maps each $a \in A$ to $\mathrm{rev}(\phi(a))$. A *substitution* $\sigma : A^* \to B^*$ is a morphism from A^* into a monoid of B^*. A morphism ϕ is called *abelian square-free* if $\phi(w)$ is abelian square-free whenever w is abelian square-free (a similar definition holds for a substitution σ).

Throughout the paper, we will reserve letters like x, y, z to denote factors and letters like a, b, c to denote elements of the alphabet.

3. The infinite case with a finite number of holes

It is not hard to check that every abelian square-free full word over a three-letter alphabet has length less than eight. It is quite straightforward to check that the maximum length of an abelian square-free partial word with at least one hole, over such an alphabet, is six. So to construct infinite abelian square-free partial words with a finite number of holes, we need at least four letters. Let us first state some remarks.

Remark 3.1. Let $w \in A^*$ be an abelian square-free word. Inserting a new letter a, $a \notin A$, between arbitrary positions of w (so that aa does not occur) yields a word $w' \in (A \cup \{a\})^*$ that is abelian square-free.

Consider *abacba* which is abelian square-free. Inserting letter d after positions 0, 3 and 4, yields *adbacdbda* which is abelian square-free.

Remark 3.2. Let $uv \in A^*$ with $|u| = |v|$, $a \in A$ and $b \notin A$. Replace a number of a's in u and the same number of a's in v with b's, yielding a new word $u'v'$. If uv is an abelian square, then $u'v'$ is an abelian square. Similarly, if uv is abelian square-free, then $u'v'$ is abelian square-free.

The question whether there exist infinite abelian square-free full words over a given alphabet was originally raised by Erdős in [13]. As mentioned above, no such word exists over a three-letter alphabet. However, infinite abelian square-free full words are readily available over a four-letter [18–20], five-letter [29], and larger alphabets [14]. These infinite words are created using repeated application of morphisms, where most of these morphisms are abelian square-free. We now investigate the minimum alphabet size needed to construct infinite abelian square-free partial words with a given finite number of holes.

Remark 3.3. Let u, v be partial words of equal length. If uv is an abelian square, then so is any concatenation of permutations of u and v.

In the paper, we will be using an abelian square-free morphism $\phi : A^* \to A^*$, where $A = \{a, b, c, d\}$, that is provided by Keränen [20] and that is defined by

$$\begin{aligned}
\phi(a) = {} & \mathtt{abcacdcbcdcadbdcadabacadcdbcabcbdbadbdcbabcbdcdacd} \\
& \mathtt{cbcacbcdbcbabdbabcabadcbcdcbadbabcbabdbcdbdadbdcbca} \\
\phi(b) = {} & \mathtt{bcdbdadcdadbacadbabcbdbadacdcbcdcacbacadcbcdcadabda} \\
& \mathtt{dcdbdcdacdcbcacbcdbcbadcdadcbacbcdcbcacdacabacadcdb} \\
\phi(c) = {} & \mathtt{cdacabadabacbdbacbcdcacbabdadcdadbdcbdbadcdadbabcab} \\
& \mathtt{adacadabdadcdbdcdacdcbadabadcbdcdadcdbdabdbcbdbadac} \\
\phi(d) = {} & \mathtt{dabdbcbabcbdcacbdcdadbdcbcabadabacadcacbadabacbcdbc} \\
& \mathtt{babdbabcabadacadabdadcbabcbadcadabadacabcacdcacbabd}
\end{aligned}$$

The length of each image is 102 and the Parikh vector of each is a permutation of $P(\phi(a)) = \langle 21, 31, 27, 23\rangle$ ($P(\phi(b)) = \langle 23, 21, 31, 27\rangle$, $P(\phi(c)) = \langle 27, 23, 21, 31\rangle$, and $P(\phi(d)) = \langle 31, 27, 23, 21\rangle$). We refer to the factors created by the images of ϕ as *blocks*.

Theorem 3.4. *There exists an infinite abelian square-free partial word with one hole over a four-letter alphabet.*

Proof. We show that the word $\diamond\phi^n(a)$ is abelian square-free, for all integers $n \geqslant 0$. Since ϕ is abelian square-free, it is sufficient to check if we have abelian squares uv that start with the hole, for $|u| = |v|$.

Now, assume that some prefix uv of $w = \diamond\phi^n(a)$ is an abelian square. We can write $uv = \diamond\phi(w_0)\phi(e)\phi(w_1)x$, where $e \in A$, $w_0, w_1, x \in A^*$ are such that $\diamond\phi(w_0)$ is a prefix of u, u is a proper prefix of $\diamond\phi(w_0e)$, and x is a proper prefix of the image by ϕ of some letter. More specifically, we can write $\phi(e) = y_0y_1$, where $y_0 \in A^*$, $y_1 \in A^+$, $u = \diamond\phi(w_0)y_0$ and $v = y_1\phi(w_1)x$. Then $|u| = |v|$, $|\diamond y_0| \equiv |y_1x| \bmod 102$, $|\diamond y_0| \leqslant 102$, and $|y_1x| < 204$. This implies $0 \leqslant |w_0| - |w_1| < 2$.

Denoting by u' the word obtained from u after replacing the hole by a letter in A so that $P(u') = P(v)$, we can build a system of equations for each letter in A. If we denote $N_a = |w_0|_a - |w_1|_a$, $N_b = |w_0|_b - |w_1|_b$, $N_c = |w_0|_c - |w_1|_c$, and $N_d = |w_0|_d - |w_1|_d$, then the system for letter a is determined by

$$21N_a + 23N_b + 27N_c + 31N_d = \lambda_a$$

Note that the image $\phi(a)$ does not overlap any $\phi(ff')$ for f, f' in A.

The number of occurrences of a (respectively, b, c, d) in u' and v must be equal, so we construct the system of equations:

$$\begin{aligned}
21N_a + 23N_b + 27N_c + 31N_d &= \lambda_a\\
31N_a + 21N_b + 23N_c + 27N_d &= \lambda_b\\
27N_a + 31N_b + 21N_c + 23N_d &= \lambda_c\\
23N_a + 27N_b + 31N_c + 21N_d &= \lambda_d
\end{aligned}$$

Each λ_i is related to the number of occurrences of letter i created by $\diamond$, $\phi(e)$ and x. Then $\lambda_i = |y_1|_i + |x|_i - |y_0|_i - \diamond_i$, where $\diamond_i = 1$ if $\diamond$ is replaced by i, and $\diamond_i = 0$ otherwise.

We can, without loss of generality, delete the same number of occurrences of any given block present in both $\phi(w_0)$ and $\phi(w_1)$. It can be checked by computer that the only scenarios that lead to non-negative integer solutions for $\diamond_i, |y_0|_i, |y_1|_i, |x|_i$, $i \in \{a, b, c, d\}$, are when $w_1 = \varepsilon$, and $w_0 = \varepsilon$ or $w_0 \in A$ (note that because of the hole at the beginning, there is one $\diamond_i$ that must be 1 while all others must be 0). Thus, either $w_0 = w_1 = \varepsilon$, or $w_0 = f \in A$ and $w_1 = \varepsilon$. For the first case, $uv = \diamond\phi(e)x$, while for the second case, $uv = \diamond\phi(f)\phi(e)x$. It is easy to verify by computer that no such partial word is an abelian square. □

Corollary 3.5. *There exists a two-sided infinite abelian square-free partial word with one hole over a five-letter alphabet.*

Proof. For a word w, let $\phi'(w) = \mathrm{rev}(\phi(w))$ with $\phi : A^* \to A^*$, where $A = \{a, b, c, d\}$. Hence, $\phi'(w)$ is abelian square-free for all abelian square-free words w and $\phi'^n(a)\diamond$ is abelian square-free, for all integers $n \geqslant 0$. Also, let $\chi : B^* \to B^*$, where $B = \{b, c, d, e\}$, be the morphism that is constructed by replacing each a in the definition of ϕ with a new letter e. By construction, χ is an abelian square-free morphism and $\diamond\chi^n(e)$ is abelian square-free, for all $n \geqslant 0$.

We show that $\phi'^n(a) \diamond \chi^n(e) \in \{a, b, c, d, e\}^*$ is abelian square-free, for all $n \in \mathbb{N}$. Suppose to the contrary that there exists an abelian square w, which is a subword of $\phi'^n(a) \diamond \chi^n(e)$, for some n. Then, the word w must contain parts of both $\phi'(a)$ and $\chi(e)$. Therefore, at least one half, called u, is a subword of either $\phi'^n(a)$ or $\chi^n(e)$ meaning it contains either a or e but not both and it does not contain the hole. Whereas the other half of w, called v, necessarily contains the other letter and the hole. Since v contains a letter that u does not, and u has no holes, w is not an abelian square. □

Corollary 3.6. *For all integers $n \geqslant 0$,*

$$\big(\mathrm{rev}(\phi)\big)^n(a) \diamond efg \diamond \phi^n(a) \in \{a, b, c, d.e, f, g\}^*_\diamond$$

is an abelian square-free partial word with two holes over a seven-letter alphabet.

Proof. The word $(\mathrm{rev}(\phi))^n(a) \diamond efg \diamond \phi^n(a)$ is abelian square-free, for all integers $n \geqslant 0$. If not, then there exists a factor w that is an abelian square and it must contain e or g. By the same logic as in the proof of Corollary 3.5, w must be centered between the holes. Obviously, no such word is an abelian square. □

Using a computer program, we have checked that over a four-letter alphabet all words of the form $u \diamond v$, where $|u| = |v| = 12$, contain an abelian square. Clearly, this implies that all partial words with a factor of this form also contain an abelian square. It follows that, over a four-letter alphabet, an infinite abelian square-free partial word containing more than one hole, must have all holes within the first 12 positions. We have also checked that all partial words $\diamond u \diamond v$ with $|u| \leqslant 10$ and $|v| \leqslant 10$, or with $|u| = 11$ and $|v| = 5$ contain abelian squares (and consequently so do the words with $|u| = 11$ and $|v| \geqslant 5$).

Proposition 3.7. *Over a four-letter alphabet, there exists no two-sided infinite abelian square-free partial word with one hole, and all right infinite abelian square-free partial words contain at most one hole.*

In addition, over a four-letter alphabet, for all words u and v, $|u|, |v| \leqslant 12$, the partial word $\diamond u \diamond v \diamond$ contains an abelian square. So if a finite partial word over a four-letter alphabet contains at least three holes, then it has an abelian square.

To end this section, we prove the following proposition that constructs an infinite word that remains abelian square-free even after replacing an arbitrary position with a hole.

Proposition 3.8. *There exists an infinite word over a* 12*-letter alphabet so that, if we replace any position in the word with a hole, the resulting partial word contains no abelian squares.*

Proof. We know that there exist infinite words w_0 over $A_0 = \{a_0, b_0, c_0, d_0\}$, w_1 over $A_1 = \{a_1, b_1, c_1, d_1\}$, and w_2 over $A_2 = \{a_2, b_2, c_2, d_2\}$ that avoid abelian squares (see [18]). Then we can let

$$v = w_0(0)w_1(0)w_2(0)w_0(1)w_1(1)w_2(1)w_0(2)w_1(2)w_2(2)\cdots$$

where v is an infinite word over $A_0 \cup A_1 \cup A_2$. We claim that, no matter where you put the hole into v, the resulting word contains no abelian squares.

To see this, replace a position in v with a hole to produce v'. For the sake of contradiction, assume that $u'_0 u'_1 = v'[i+1..i+l]v'[i+l+1..i+2l]$ is a non-trivial abelian square in v'. Let $u_0 = v[i+1..i+l]$ and $u_1 = v[i+l+1..i+2l]$ be the corresponding words in v. Without loss of generality, we can assume that the $\diamond$ in v' replaced a letter from the alphabet A_0. Since $u'_0 u'_1$ is an abelian square, by definition we can replace the $\diamond$ with some letter to get a full word $v_0 v_1$, where $v_0 v_1$ is an abelian square. Assume we replace it with a letter in A_1, the other cases being similar. Note that the construction of $v_0 v_1$ from $u_0 u_1$ did not affect those letters in $u_0 u_1$ that had been taken from the alphabet A_2. Let $\psi : (A_0 \cup A_1 \cup A_2)^* \to {A_2}^*$ be the morphism defined as follows: if $a \in A_0 \cup A_1 \cup A_2$ then

$$\psi(a) = \begin{cases} a & \text{if } a \in A_2 \\ \varepsilon & \text{otherwise} \end{cases}$$

Then it is clear that $\psi(v_0 v_1)$ is either an abelian square, or else it is empty. Since every subword of v of length 3 or greater contains a letter in A_2, if $\psi(v_0 v_1)$ is empty then we must have that

$|v_0| = |v_1| = 1$, which implies that $u_0'u_1'$ is trivial. Therefore assume it is not empty. By construction we find that $\psi(v_0v_1) = \psi(u_0u_1)$ is a subword of w_2. This contradicts the fact w_2 is abelian square-free. Since $\|A_0 \cup A_1 \cup A_2\| = 12$ the theorem follows. □

Whether or not the alphabet size of twelve in Proposition 3.8 is optimal remains an open problem.

4. The case with infinitely many holes

The next question is how large should the alphabet be so that an abelian square-free partial word with infinitely many holes can be constructed. In this section, we construct such words over a minimal alphabet size of five.

First let us state three lemmas that help us achieve our goal.

Lemma 4.1. *Let z be a word which is not an abelian square, x (respectively, y) be a suffix (respectively, prefix) of $\phi(f)$, where $f \in \{b, c, d\}$. Then the following hold:*

1. *No word of the form $\diamond\phi(z)y$, $\phi(z)y\diamond$, $\diamond a\phi(z)y$ or $a\phi(z)y\diamond$ is an abelian square, unless zf is an abelian square.*
2. *No word of the form $\diamond x\phi(z)$, $x\phi(z)\diamond$ or $\diamond x\phi(z)a$ is an abelian square, unless fz is an abelian square.*

Proof. Let us assume that we can get an abelian square uv of one of the mentioned forms in statement 1 (the proof for statement 2 is similar). Then there exists a factorization $\phi(z_0)x_0x_1\phi(z_1)$ of $\phi(z)$, for some words x_0, x_1, z_0, z_1 such that $z = z_0\phi^{-1}(x_0x_1)z_1$ and $|x_0x_1| = 102$, so that $\phi(z_0)x_0$ is in the u-part and $x_1\phi(z_1)$ is in the v-part. After cancelling letters from both z_0 and z_1, we see that it is enough to consider the cases where $\alpha(z_0) \cap \alpha(z_1) = \emptyset$, and $|z_0| = |z_1|$ or $|z_0| = |z_1| + 1$ or $|z_1| = |z_0| + 1$. Since, having several occurrences of the same letter in either z_0 or z_1 increases the number of occurrences of one letter too fast after the application of ϕ, it is enough to consider the cases where $|z_0| + |z_1| \leqslant 4$, that is, $|z| \leqslant 5$. All words of the form $\diamond\phi(z)y$, $\phi(z)y\diamond$, $\diamond a\phi(z)y$ or $a\phi(z)y\diamond$, with z of length at most 5 and z not an abelian square, where y is a prefix of $\phi(f)$ for some $f \in \{b, c, d\}$, can be checked by computer. None of these is an abelian square unless zf is an abelian square. □

Lemma 4.2. *Let z be a word which is not an abelian square, x (respectively, y) be a suffix (respectively, prefix) of $\phi(a)$. Then, no word of the form $\diamond x\phi(z)y$ or $x\phi(z)y\diamond$ is an abelian square, unless az or za is an abelian square.*

Proof. Following the proof of Lemma 4.1, here we need to check by computer all words $\diamond x\phi(z)y$ or $x\phi(z)y\diamond$, with z of length at most 5 and z not an abelian square, where x is a suffix and y is a prefix of $\phi(a)$. None of these is an abelian square unless az or za is an abelian square. □

Lemma 4.3.

1. *Let z be a word which is not an abelian square, x be the suffix of length 101 of $\phi(a)$, y be a prefix of $\phi(f)$, where $f \in \{a, b, c, d\}$. Then, no word of the form $\diamond x\phi(z)y$ is an abelian square, unless az, zf or azf is an abelian square.*
2. *Let z be a word which is not an abelian square, y be the prefix of length 101 of $\phi(a)$, x be a suffix of $\phi(f)$, where $f \in \{a, b, c, d\}$. Then, no word of the form $\diamond x\phi(z)y$ is an abelian square, unless za, fz or fza is an abelian square.*

Proof. The result follows similarly as for Lemmas 4.1 and 4.2 and an exhaustive check using a computer. □

Theorem 4.4. *There exists an abelian square-free partial word with infinitely many holes over a five-letter alphabet.*

Proof. Let $w = \phi^{\omega}(a)$, which is an infinite abelian square-free full word over the four-letter alphabet $A = \{a, b, c, d\}$. There exist infinitely many j's such that $w[j-101..j] = \phi(a)$. Let k_0 be the smallest integer so that $w[k_0 - 101..k_0] = \phi(a)$ (note that $k_0 = 101$). Then define k_j recursively, where k_j is the smallest integer satisfying $k_j > 5k_{j-1}$ and $w[k_j - 101..k_j] = \phi(a)$.

Note that in order to avoid abelian squares, the holes must be somehow sparse. Construct an infinite partial word w' over $A \cup \{e\}$, where $e \notin A$, by introducing factors in w as follows. For all integers $j \geqslant 0$, do the following. If $i = k_j$ and $j \equiv 0 \bmod 5$, then introduce $\diamond e$ between positions i and $i+1$ of w. If $i = k_j$ and $j \not\equiv 0 \bmod 5$, then introduce four e's in the image of $\phi(a)$ that ends at position i, in the following way: setting $w[k_j - 101..k_j] = \phi(a) = abXca$, where $X \in A^*$, the word $abXca$ is replaced with $X' = eaebXceae$. Clearly, w' has infinitely many holes. The modulo 5 assures that in every finite prefix of w', there are much more e's than $\diamond$'s. Moreover, if the holes are not taken into consideration, since no two e's are next to each other, by Remark 3.1, the word is still abelian square-free.

In order to prove that w' has no abelian squares, we assume that it has one and get a contradiction. Let uv be an occurrence of an abelian square, where $u = w'[i..i+l]$ and $v = w'[i+l+1..i+2l+1]$ for some i, l. Let k'_j, $j \geqslant 0$, be the sequence in w' corresponding to the sequence k_j, $j \geqslant 0$, in w, that is, if k_j is the last position of an occurrence of $abXca$, then k'_j is the last position of the corresponding occurrence of $abXca\diamond e$ or $eaebXceae$. Let $J_1 = \{j \mid i \leqslant k'_j \leqslant i+l\}$ and $J_2 = \{j \mid i+l+1 \leqslant k'_j \leqslant i+2l+1\}$. Then $\|J_1\| \leqslant 3$ and $\|J_2\| \leqslant 1$, which implies that $\|J_1 \cup J_2\| \leqslant 4$. To see that $\|J_2\| \leqslant 1$, assume $\|J_2\| > 1$, then note that there exist some $j, j+1 \in J_2$. However, this implies that $l = (i+2l+1) - (i+l+1) \geqslant k'_{j+1} - k'_j > k'_j > i + l \geqslant l$, a contradiction. Now assume that $\|J_1\| > 3$. Then there are at least ten occurrences of the letter e in u, and for each occurrence of e there must also be an e or a $\diamond$ in v. However, this implies $\|J_2\| > 1$, which violates the fact that $\|J_2\| \leqslant 1$.

Next, we want to show that no holes occur in the abelian square uv. Observe that v cannot contain more than four e's, since otherwise, there exists $j > 0$ such that $k'_{j-1} < i + l + 1 \leqslant k'_j \leqslant i + 2l + 1 < k'_{j+1} \leqslant i + 2l + 106$. Thus, as $k'_{j+1} > 4k'_j$ and $k'_j \geqslant 104$, $l + 1 \leqslant i + l + 1 < 2k'_j < k'_{j+1} - k'_j - 104 \leqslant (i + 2l + 106) - (i + l + 105) = l + 1$, a contradiction. Observe also since $\|J_1 \cup J_2\| \leqslant 4$, by construction of w', word uv contains at most one hole.

Firstly, we prove that v contains no hole. Let us assume that the last position of v is a hole. If v contains any e's, then there exists j such that $k'_{j-1} < i + l + 1 \leqslant k'_j < i + 2l + 1 = k'_{j+1} - 1$, and as $k'_{j+1} > 3k'_j$, we get $l + 1 \leqslant i + l + 1 < 2k'_j < (k'_{j+1} - 1) - (k'_j - 1) \leqslant (i + 2l + 1) - (i + l) = l + 1$, a contradiction; note that here, the $\diamond$ is from the block corresponding to k'_{j+1}, while the e's are from the block corresponding to k'_j. Moreover, u does not contain any e's, since otherwise the $\diamond$ and the e in uv cancel each other and we get that the original word w is not abelian square-free. Hence, there exist $x, z \in A^*$, where x, $|x| < 102$, is a suffix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$, such that $uv = x\phi(z)\diamond$ is an abelian square and z ends with the letter a by construction (and by definition of k'_{j+1}). Note that z is abelian square-free as $\phi(z)$ is a factor of w. By Lemma 4.1 or Lemma 4.2, we get that the original word w contains an abelian square, which is a contradiction.

Now, let us assume that v has a hole at any position other than the last one. Then v contains at least one e. Note that v cannot contain more than one e, since otherwise, as $\|J_2\| \leqslant 1$, there exists j such that $k'_{j-1} < i + l + 1 < k'_j \leqslant i + 2l + 1 < k'_{j+1} \leqslant i + 2l + 106$, and a contradiction is reached as above. If the $\diamond$ in v corresponds to an e in u, then cancelling them as well as all other e's in uv gives us a factor of the original w, which is abelian square-free. So the $\diamond$ in v corresponds to a letter in u that is not e, and we have $|u|_e = |v|_e = 1$. Here $u = eu'$ or $u = aeu'$ for some $u' \in A^*$, and $v = v'\diamond ev''$ for some $v', v'' \in A^*$. So there exist $z_0, z_1, y \in A^*$, with y a prefix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$, such that $uv = e\phi(z_0)\diamond e\phi(z_1)y$ or $uv = ae\phi(z_0)\diamond e\phi(z_1)y$. Using Remark 3.3, we obtain an abelian square of the form $e\phi(z)y\diamond e$ or $ae\phi(z)y\diamond e$, where $z = z_0z_1$. After cancelling the two e's, this is also impossible by Lemmas 4.1 and 4.2 and our hypothesis that w is abelian square-free. Note that z_0z_1f is a factor of w with f the letter such that y is a prefix of $\phi(f)$ and to get a contradiction with Lemma 4.1 or 4.2, z_0z_1f must be abelian square-free which can be deduced from the fact that w is abelian square-free.

Secondly, we prove that the last position of u cannot be a hole. Assume the last position of u is a hole. Then the first position of v is an e. The partial word v cannot contain more than one e, since otherwise, there exists j such that $i+l+1 = k'_j \leqslant i+2l+1 < k'_{j+1} \leqslant i+2l+106$ and again we reach a contradiction. If no e occurs in u, then the $\diamond$ in u and the e in v cancel each other, giving us a factor of the original word w, which is abelian square-free. So $|u|_e = |v|_e = 1$. This implies that $uv = e\phi(z_0)\diamond e\phi(z_1)y$ or $uv = ae\phi(z_0)\diamond e\phi(z_1)y$ for some $z_0, z_1, y \in A^*$, with y a prefix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$. After cancelling the e's, setting $z = z_0z_1$, and using Remark 3.3, we get abelian squares of the form $\diamond\phi(z)y$ or $\diamond a\phi(z)y$. Thus, it is easy to connect previously presented arguments with Lemmas 4.1 and 4.2, and again a contradiction is reached with the fact that the original word w is abelian square-free.

Thirdly, we prove that u cannot have a hole. We only need to consider the case when a position in u, other than the last one, is a hole. Let k'_j denote the position of the e in u immediately following the $\diamond$. Since u contains a $\diamond$ and an e, v contains at least one e. If $|v|_e = |u|_e + 1$, then the $\diamond$ and the e's in uv can be cancelled giving us a factor of w, which is abelian square-free. Thus $|v|_e = |u|_e$. If $k'_{j+1} \leqslant i+l$, then $|u|_e \geqslant 5$ and we get a contradiction since both $|uv|_\diamond \leqslant 1$ and $|v|_e \leqslant 4$ hold. Thus, $k'_{j+1} > i+l$.

First, assume that $\|J_1\| = 3$. In this case, $|u|_e \geqslant 6$ and we reach a contradiction as before.

Now, assume that $\|J_1\| = 2$. Here, $i \leqslant k'_{j-1} < k'_j \leqslant i+l < k'_{j+1}$. Since $|v|_e \leqslant 4$, u contains one, two, or three of the e's from a suffix of the factor X' ending at position k'_{j-1}. Let us denote this suffix by X'_u.

Suppose that $k'_{j+1} > i+2l+1$. Then v contains two or three of the e's from a prefix of the factor X' ending at position k'_{j+1}. Let us denote this prefix by X'_v. So $uv = X'_u\phi(z_0)\diamond e\phi(z_1)X'_v$, where $z_0, z_1 \in A^*$. After using Remark 3.3 and cancelling the e's in uv, we are left with an abelian square of the form $\diamond x\phi(z)y$, where $z = z_0z_1$ is abelian square-free, x is a suffix of $\phi(a)$, and y is a prefix of $\phi(a)$. We reach a contradiction with Lemma 4.2. Next, suppose that $k'_{j+1} \leqslant i+2l+1$.

First, assume that X'_u contains one e, that is, $X'_u = ae$ or $X'_u = e$. There are three possibilities: $|u|_e = |v|_e = 2$ or $|u|_e = |v|_e = 3$ or $|u|_e = |v|_e = 4$. The first possibility contradicts the fact that k'_{j+1}, being in v, gives rise to four e's. For the second possibility, u has one of the e's from the factor X' ending at k'_{j+1} while v contains the other three e's. Here, $i+l < k'_{j+1} < i+l+106$. After using Remark 3.3 and cancelling the e's from u and v, we obtain an abelian square of the form $\diamond a\phi(z)y$ or $\diamond\phi(z)y$, where y is a prefix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$, and z is abelian square-free. Using Lemmas 4.1 and 4.2, we reach a contradiction. For the third possibility, u contains the first two e's of the factor X' ending at k'_{j+1} while v contains the other two e's. In addition, v contains the first two e's of the factor corresponding to k'_{j+2} (note that $k'_{j+2} > i+2l+1$). Here the abelian squares we get are of the form $\diamond a\phi(z)y$ or $\diamond\phi(z)y$, where y is a prefix of $\phi(a)$. Lemma 4.2 yields a contradiction.

Now, assume that X'_u contains two e's, that is, $3 \leqslant |X'_u| \leqslant 103$. So $|u|_e = |v|_e = 3$ or $|u|_e = |v|_e = 4$. The first possibility is easily eliminated since the factor X' ending at k'_{j+1} has four e's. As to the second possibility, u contains the first e of the factor X' ending at k'_{j+1} while v contains the other three e's. In addition, v contains the first e of the factor X' ending at k'_{j+2} (note that $k'_{j+2} > i+2l+1$). Again using Remark 3.3 and cancelling the e's from uv, we obtain an abelian square of the form $\diamond x\phi(z)$ or $\diamond x\phi(z)a$, where x is a suffix of $\phi(a)$ with $1 \leqslant |x| \leqslant 101$. A contradiction follows with the help of Lemma 4.2.

Next, assume that X'_u contains three e's, that is, $X'_u = aebXceae$ or $X'_u = ebXceae$. It must be the case that v contains four e's. Thus, after the cancellation of the e's in uv and suitable permutations, we obtain an abelian square of the form $\diamond\phi(z)y$ or $\diamond x\phi(z)y$, where x is the suffix of length 101 of $\phi(a)$, y is a proper prefix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$, and z is not an abelian square. For the first form, a contradiction follows using Lemmas 4.1 and 4.2, while for the second form, a contradiction is reached using Lemma 4.3.

Finally, assume that $\|J_1\| = 1$. Here, $i < k'_j \leqslant i+l < k'_{j+1}$. Let us first consider the case where $k'_{j+1} > i+2l+1$. The situations where $|u|_e = |v|_e = 3$ or $|u|_e = |v|_e = 4$ are clearly impossible. The situation where $|u|_e = |v|_e = 2$ is also impossible as this would imply that $v = aebXcea$ or $v = aebXce$

or $v = ebXcea$ or $v = ebXce$, while there must be at least two blocks of length 102 between k'_j, which lies in u, and $k'_{j+1} \in \{i+2l+2, i+2l+3\}$, which lies outside uv. The situation where $|u|_e = |v|_e = 1$ is possible in case the e in v is from the factor X' ending at k'_{j+1}. Here, $uv = x\phi(z_0)\diamond e\phi(z_1)e$ or $uv = x\phi(z_0)\diamond e\phi(z_1)ea$ for some $x, z_0, z_1 \in A^*$ such that x is a suffix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$. Cancelling the e's in uv, using Remark 3.3, and setting $z = z_0z_1$ lead to $\diamond x\phi(z)$ or $\diamond x\phi(z)a$ being an abelian square. In either case we get a contradiction with Lemma 4.1 or Lemma 4.2.

Let us now consider the case where $k'_{j+1} \leqslant i+2l+1 < k'_{j+2}$. The situations where $|u|_e = |v|_e = 1$ or $|u|_e = |v|_e = 2$ are impossible since the factor X' ending at k'_{j+1}, call it X'_{uv}, gives rise to four e's. On the other hand, the situation where $|u|_e = |v|_e = 3$ is possible in case the first two e's from X'_{uv} are in u while the other two are in v, and an additional e comes from the factor X' ending at k'_{j+2}. In this case, after cancelling the four e's in X'_{uv}, we obtain an abelian square of the form $x\phi(z_0)\diamond e\phi(z_1)e$ or $x\phi(z_0)\diamond e\phi(z_1)ea$ for some $x, z_0, z_1 \in A^*$ such that x is a suffix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$. Using Remark 3.3 and setting $z = z_0z_1$, either $\diamond ex\phi(z)e$ or $\diamond ex\phi(z)ea$ is an abelian square. After cancelling the e's, in either case we get a contradiction with Lemma 4.1 or Lemma 4.2. The situation where $|u|_e = |v|_e = 4$ is possible in case the first three e's from X'_{uv} are in u while the other one is in v, and three additional e's come from the factor X' ending at k'_{j+2}. In this case, after cancelling the e's in uv and using Remark 3.3, we get an abelian square of the form $\diamond x\phi(z)$ or $\diamond x\phi(z)y$, where x is a suffix of $\phi(a)$ or $\phi(b)$ or $\phi(c)$ or $\phi(d)$, y is a prefix of $\phi(a)$ of length 101, and $z \in A^*$ is abelian square-free. This yields a contradiction with Lemmas 4.1, 4.2, or Lemma 4.3.

Hence all symbols in uv correspond to letters in $A \cup \{e\}$. By Remark 3.1, since w' contains an abelian square, w must also contain an abelian square, a contradiction. □

5. Number of avoiding partial words

In this section, we investigate the number of abelian square-free partial words. This number has been studied in [8] and [20] for full words over an alphabet of size four.

Theorem 5.1. *Let c_n denote the number of partial words over a four-letter alphabet of length n with one hole, of the form $\diamond w$, that avoid abelian squares. Then there exist $N > 0$ and $\beta > 1$ such that for all $n > N$, $c_n \geqslant \beta^{-51}\beta^n$.*

Proof. Keränen in [20] provides an abelian square-free substitution σ_{109}: $A^* \to A^*$, where $A = \{a, b, c, d\}$ as follows (for convenience, we just denote σ_{109} by σ). The 12 image words of $\sigma(a)$ have the form $v_{16}v_4v_{27}v_3v_{59}$ or

$$\texttt{abcacdcbcdcadcdb}v_4\texttt{badacdadbdcdbdabdbcbabcbdcb}v_3$$

$$\texttt{bdcdadcdbcbabcbdcbcacdcacbadabcbdcbcadbabcbabdbcdbdadbdcbca}$$

with 12 distinct pairs (v_4, v_3) from $\{abcd, abdc, adbc, dabc\} \times \{acd, adc, cad\}$. The subscripts of the factors v_{16}, v_4, v_{27}, v_3 and v_{59} indicate their lengths. Here, $\sigma(b) = p(\sigma(a))$, $\sigma(c) = p(\sigma(b))$, and $\sigma(d) = p(\sigma(c))$, where $p(a) = b$, $p(b) = c$, $p(c) = d$, and $p(d) = a$. The Parikh vectors are $P(x_1) = \langle 21, 31, 29, 28\rangle$, $P(x_2) = \langle 28, 21, 31, 29\rangle$, $P(x_3) = \langle 29, 28, 21, 31\rangle$, and $P(x_4) = \langle 31, 29, 28, 21\rangle$, where $x_1 \in \sigma(a), x_2 \in \sigma(b), x_3 \in \sigma(c)$, and $x_4 \in \sigma(d)$.

Let w be a prefix of length $n-1$ of some word in $\sigma^\omega(a)$. We show that $\diamond w$ is an abelian square-free partial word. Suppose $\diamond w$ contains an abelian square, so the square has to contain the hole. Using the same method as in Theorem 3.4, let it be $uv = \diamond\sigma(w_0)\sigma(e)\sigma(w_1)x$, where $e \in A$, $w_0, w_1, x \in A^*$, $\diamond\sigma(w_0)$ is a prefix of u, u is a proper prefix of $\diamond\sigma(w_0e)$, and $|x| < 109$. Recall that $|u| = |v|$. Reduce w_0 and w_1 so that they do not share any common letter. Since $|w_1| \leqslant |w_0|$ for the equality $|u| = |v|$ to hold, we only need to check the case when $0 \leqslant |w_0| - |w_1| < 2$. Building the system of equations for the number of occurrences of each letter as we did before, we only need to check partial words of the form $\diamond\sigma(e)x$ and $\diamond\sigma(f)\sigma(e)x$, where $f \neq e \in A$. Using a computer program, we have checked both types and they are not abelian squares.

Keränen in [20] establishes the number of abelian square-free words of sufficiently long length n over a four-letter alphabet as being greater than $\beta^{-50}\beta^n$ with $\beta = (12)^{1/m} \approx 1.02306$, where $m = 109$, so the growth rate of the number of abelian square-free partial words with one hole of the form $\diamond w$ of length n over a four-letter alphabet can be derived. Let $n - 51 = qm + r$, $0 \leqslant r < m$. Note that $|v_{16}v_4v_{27}v_3| = 50$, and we do not count $\diamond$. There are at least

$$(12)^{q+1} = \left(\beta^m\right)^{q+1} = \beta^{qm+m} = \beta^{n-51-r+m} = \beta^{n-51}\beta^{-r+m} > \beta^{-51}\beta^n$$

abelian square-free partial words of the desired form. □

Theorem 5.2. *Consider $h > 0$. Let $c_{n,h}$ denote the number of partial words over a five-letter alphabet of length n with h holes that avoid abelian squares. Then there exist $N > 0$, $r > 1$ and $\beta > 0$ such that for all $n > N$, $c_{n,h} \geqslant \beta r^n$.*

Proof. Let $w = \phi^\omega(a)$, which is an infinite abelian square-free full word over $A = \{a, b, c, d\}$. Referring to the construction in the proof of Theorem 4.4, we can introduce $\diamond$'s and e's into w in such a way that the resulting partial word is abelian square-free. Here, we use the construction of Theorem 4.4, except that we limit the number of holes. More specifically, let k_0 be the smallest integer so that $w[k_0 - 101..k_0] = \phi(a)$, and in general k_j the smallest integer so that $k_j > 5k_{j-1}$ and $w[k_j - 101..k_j] = \phi(a)$. Then we construct w' as follows. If $0 \leqslant j < 5h$, $j \equiv 0 \bmod 5$, then introduce $\diamond e$ between positions k_j and $k_j + 1$ of w. If $0 \leqslant j < 5h$, $j \not\equiv 0 \bmod 5$, then replace the block $w[k_j - 101..k_j] = \phi(a) = abXca$ with $eaebXceae$. Repeating the proof of Theorem 4.4, it follows that w' avoids abelian squares, and moreover, w' has h holes (note that referring to the proof of Theorem 4.4, the assumed abelian square uv will still have only one hole in this new slightly modified construction, and so the arguments will not be affected).

Let δ be the largest integer so that $w'(\delta) = \diamond$, and let M be the largest integer so that $w'(M) = e$. Then consider $N = 2M$, and $n > N$. It is easy to see that if $f(m) = \lfloor \frac{m}{N} \rfloor + m$, then $f(m+1) - f(m) \in \{1, 2\}$ for all m. Consequently, there exists an integer m so that either $\lfloor \frac{m}{N} \rfloor + m = n$ or $\lfloor \frac{m}{N} \rfloor + m - 1 = n$. Note that in either case $\frac{nN}{N+1} \leqslant m \leqslant \frac{(n+2)N}{N+1}$.

Consider the case $\lfloor \frac{m}{N} \rfloor + m = n$, the other case being similar. Then let v be the prefix of w' of length m (in the other case, v would be of length $m - 1$). We can construct a partial word v' from v by introducing an e right before $v(iN)$ or right after $v(iN)$, for each i where $1 \leqslant i \leqslant \lfloor \frac{m}{N} \rfloor$. Note that since we have two choices for each i, there are $2^{\lfloor \frac{m}{N} \rfloor} = 2^{n-m} \geqslant 2^{n - \frac{(n+2)N}{N+1}}$ possible v' that can be constructed. Furthermore, $|v'| = \lfloor \frac{m}{N} \rfloor + m = n$.

We want to show that v' avoids abelian squares. In order to see this, begin by assuming that $u_0u_1 = v'[l..l+l')v'[l+l'..l+2l')$ is an abelian square in v'. Then note that $l + 2l' \geqslant N$, since otherwise u_0u_1 is a subword of v, which is impossible since v avoids abelian squares. This implies that $l + l' \geqslant \frac{l+2l'}{2} \geqslant \frac{N}{2} = M$. However, we know that if $v'(i) = \diamond$ then $i \leqslant \delta < M$, so u_1 does not contain any holes. On the other hand, since $\phi^\omega(a)$ is abelian square-free, it follows that u_0u_1 contains at least one hole, and, thus, u_0 contains at least one hole.

Since $l + l' \geqslant M$ and M is the largest integer so that $w'(M) = e$, $u_1 = v'[l+l'..l+2l')$ does not contain any e's other than those inserted before or after $v(iN)$ for each i, $1 \leqslant i \leqslant \lfloor \frac{m}{N} \rfloor$, to construct v' from $v = w'[0..m)$, except possibly the e at position M in case $l + l' = M$ (when $l = 0$). As mentioned before, every hole that occurs in v' has index less than or equal to δ. Since u_0 contains a hole, this implies that $l \leqslant \delta < M \leqslant l + l'$. It follows that u_0 contains the partial word $v'[\delta..M)$ as a factor, so u_0 contains the e's of that factor (at least 16 e's). Moreover, u_0 contains at most one e less than u_1 coming from the e's inserted before or after the $v(iN)$'s (half of these are in the first part and half are in the second part, or there is a difference of one e when $l = 0$). Hence, $|u_0|_e \geqslant |u_1|_e + 14$. On the other hand, $|u_0|_e \leqslant |u_1|_e$ since u_1 does not contain any holes and u_0u_1 is an abelian square. This is a contradiction, so v' is abelian square-free.

Therefore, each v' has length n, avoids abelian squares, and contains h holes. Moreover there are at least $2^{n - \frac{(n+2)N}{N+1}}$ such v'. So setting $\beta = 2^{-\frac{2N}{N+1}}$ and $r = 2^{\frac{1}{N+1}}$, it follows that $c_{n,h} \geqslant \beta r^n$, which concludes the proof. □

6. The finite case

Finite abelian square-free words are difficult to characterize and to build without the aid of a computer. This is due to the fact that they have very little structure. However, there are a few special constructions, such as Zimin words, that have been investigated. In this section, we show that the replacement of letters with holes in these words result in partial words that are not abelian square-free.

Zimin words were introduced in [33] in the context of blocking sets. Due to their construction, Zimin words are not only abelian square-free, but also maximal abelian square-free in the sense that any addition of letters, from the alphabet they are defined on, to their left or right introduces an abelian square.

Definition 6.1. (See [33].) Let $\{a_0, \ldots, a_{k-1}\}$ be a k-letter alphabet. The Zimin words z_i are defined by $z_0 = a_0$, and $z_i = z_{i-1}a_iz_{i-1}$ for $1 \leqslant i < k$.

Note that $|z_i| = 2^{i+1} - 1$ and $P(z_i) = \langle 2^i, 2^{i-1}, \ldots, 2, 1\rangle$ for all $i = 0, \ldots, k-1$.

Proposition 6.2. *Let $\{a_0, \ldots, a_{k-1}\}$ be a k-letter alphabet. For i with $1 < i < k$, the replacement of any letter in z_i with a hole yields a word with an abelian square.*

Proof. The replacement of any letter in an odd position yields an abelian square factor compatible with $abab$ for some letters a, b. For an even position, the factor is of one of the forms $\diamond bacab, ab\diamond cab, bac\diamond ba, bacab\diamond$. $\square$

In [10], Cummings and Mays introduced a modified construction, which they named a one-sided Zimin construction. The resulting words are much shorter than Zimin words.

Definition 6.3. (See [10].) Let $\{a_0, \ldots, a_{k-1}\}$ be a k-letter alphabet. Left Zimin words y_i are defined recursively as follows: $y_0 = a_0$ and, for $1 \leqslant i \leqslant k-1$, $y_i = y_{i-1}a_iz_{\lfloor \frac{i-1}{2} \rfloor}$, where $z_{\lfloor \frac{i-1}{2} \rfloor}$ is a Zimin word over $\{a_0, a_2, \ldots, a_{i-1}\}$ whenever i is odd and $\{a_1, a_3, \ldots, a_{i-1}\}$ whenever i is even. Right Zimin words can be defined similarly.

For example, $y_4 = abacbdacaebdb$ and $y_5 = abacbdacaebdbfacaeaca$.

Note that left and right Zimin words are symmetric, and both one-sided constructions have Parikh vector $P(y_i) = \langle 2^{\lfloor \frac{i+1}{2} \rfloor}, 2^{\lfloor \frac{i}{2} \rfloor}, \ldots, 4, 2, 2, 1\rangle$. Furthermore, y_i is a left maximal abelian square-free word over the alphabet $\{a_0, a_1, \ldots, a_i\}$, for each $i = 0, \ldots, k-1$ [10].

Proposition 6.4. *Let $\{a_0, \ldots, a_{k-1}\}$ be a k-letter alphabet. For each i with $6 \leqslant i < k$, the replacement of any letter in y_i with a hole results in a word containing an abelian square.*

Proof. We prove the result by induction on k. For $k = 7$, we find by exhaustive search that no hole can replace any letter of y_6 without creating an abelian square. Assuming that the result is true for $y_6, \ldots, y_{k-1}$, consider $y_k = y_{k-1}a_kz_{\lfloor \frac{k-1}{2} \rfloor}$, where $z_{\lfloor \frac{k-1}{2} \rfloor}$ is a Zimin word. By Proposition 6.2, it is not possible to place holes in $z_{\lfloor \frac{k-1}{2} \rfloor}$ while remaining abelian square-free. Replacing a_k with a hole yields $\diamond z_{\lfloor \frac{k-1}{2} \rfloor}$, which is an abelian square since $z_{\lfloor \frac{k-1}{2} \rfloor}$ is a maximal abelian square-free word. And by the inductive hypothesis, no hole can replace a letter in y_{k-1} without the resulting word having abelian square factors. $\square$

In [21], Korn gives a construction that provides shorter maximal abelian square-free words. The words' construction is very different from the variations on Zimin words.

Definition 6.5. (See [21].) Let $\{a_0,\ldots,a_{k-1}\}$ be a k-letter alphabet, where $k \geqslant 4$. The words v_i are defined recursively by $v_0 = a_2a_1$, and $v_i = v_{i-1}a_{i+2}a_{i+1}$ for $1 \leqslant i \leqslant k-3$. Then $w_{k-1} = a_0ua_1ua_0v_{k-3}a_0ua_{k-1}ua_0$, where $u = a_2 \cdots a_{k-2}$.

For example, $w_4 = acdbcdacbdcedacdecda$, where $u = cd$ and $v_{k-3} = v_2 = cbdced$. Consider the prefix $aubua = acdbcda$ of w_4 as well as the factor $aubuav_2[0..4] = acdbcdacbdce$. We have that $aubu(0) = acdbc$ has an extra occurrence of c compared to $u(1)av_2[0..2] = dacbd$, while the latter has an extra occurrence of d compared to the former. If for instance the d from the first u is replaced by $\diamond$, this yields an abelian square, with the $\diamond$ corresponding to e in the suffix of $v_2[0..4] = cbdce$. The same statement holds when considering the second u. These observations lead to the following proposition.

Proposition 6.6. *Let $A = \{a_0,\ldots,a_{k-1}\}$ be a k-letter alphabet, where $k \geqslant 4$, and $w_{k-1} \in A^*$ be constructed according to Definition 6.5. The replacement of any letter in w_{k-1} with a hole results in a word containing an abelian square.*

Proof. Note that every letter in v_{k-3}, with the exception of a_1 and a_{k-1}, occurs exactly twice. Moreover, if a hole replaces any letter in v_{k-3}, at a position other than the first or the last one, then we get a factor of either the form $a_la_{l-1}\diamond a_l$ or $a_l\diamond a_{l+1}a_l$ for some l. Note that both these partial words represent abelian squares. Thus it is impossible to replace letters with holes in $v_{k-3}[1..|v_{k-3}|-1)$ of w_{k-1} while keeping abelian square-freeness. Replacing the first (last) letter of v_{k-3} with a hole yields the abelian square $a_0ua_1ua_0\diamond$ ($\diamond a_0ua_{k-1}ua_0$).

Consider now the subword $a_0ua_1ua_0$ of w_{k-1} (the proof is similar for the subword $a_0ua_{k-1}ua_0$). Clearly, replacing a_0 or a_1 with a hole yields an abelian square. Note that the equality $2|u|+2 = |v_{k-3}|$ holds. When a hole replaces the letter at position j in any of the u's, consider the factor $a_0ua_1ua_0v_{k-3}[0..2j+2] = a_0ua_1ua_0v_{k-3}[0..2j]a_{j+1}a_{j+3}$. Since $u[0..j) = a_2\cdots a_{j+1}$ and $v_{k-3}[0..2j] = a_2a_1a_3a_2\cdots a_{j-2}a_ja_{j-1}a_{j+1}a_ja_{j+2}$, we have that $a_0u[0..j)u[j..|u|)a_1u[0..j)$ has an extra occurrence of a_{j+1} compared to $u[j..|u|)a_0v_{k-3}[0..2j]$, while the latter has an extra occurrence of a_{j+2} compared to the former. If $u(j) = a_{j+2}$ from the first u is replaced by a $\diamond$, this yields an abelian square, with the $\diamond$ corresponding to a_{j+3} in the suffix of $v_{k-3}[0..2j+2]$ (same statement holds when considering the second u). $\square$

Acknowledgments

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. Robert Mercaş' work was partially supported by Research Grant No. 1323 U07 E30 N-2008/InvAct/Bel, G./BJ01 and by Research Grant No. 1355 U07 E30 N-2010PFR-URV-B2-02 of the University Rovira i Virgili. Part of this paper was presented at LATA 2010 [5]. A research assignment from the University of North Carolina at Greensboro for F. Blanchet-Sadri is gratefully acknowledged. Some of this assignment was spent at the LIAFA: Laboratoire d'Informatique Algorithmique: Fondements et Applications of Université Paris Denis Diderot, Paris, France. We thank the referees of a preliminary version of this paper for their very valuable comments and suggestions.

References

[1] E. Andreeva, C. Bouillaguet, P.-A. Fouque, J. Hoch, J. Kelsey, A. Shamir, S. Zimmer, Second preimage attacks on dithered hash functions, in: N. Smart (Ed.), Advances in Cryptology – EUROCRYPT 2008, in: Lecture Notes in Comput. Sci., vol. 4965, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 270–288.
[2] J. Berstel, Axel Thue's work on repetitions in words, in: P. Leroux, C. Reutenauer (Eds.), Invited Lecture at the 4th Conference on Formal Power Series and Algebraic Combinatorics, 1992, pp. 65–80.
[3] J. Berstel, L. Boasson, Partial words and a theorem of Fine and Wilf, Theoret. Comput. Sci. 218 (1999) 135–141.
[4] F. Blanchet-Sadri, Algorithmic Combinatorics on Partial Words, Chapman & Hall/CRC Press, Boca Raton, FL, 2008.
[5] F. Blanchet-Sadri, J.I. Kim, R. Mercaş, W. Severa, S. Simmons, Abelian square-free partial words, in: A.-H. Dediu, H. Fernau, C. Martín-Vide (Eds.), LATA 2010, 4th International Conference on Language and Automata Theory and Applications, Trier, Germany, in: Lecture Notes in Comput. Sci., vol. 6031, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 94–105.

[6] F. Blanchet-Sadri, R. Mercaş, A. Rashin, E. Willett, An answer to a conjecture on overlaps in partial words using periodicity algorithms, in: A.-H. Dediu, A.M. Ionescu, C. Martín-Vide (Eds.), LATA 2009, 3rd International Conference on Language and Automata Theory and Applications, Tarragona, Spain, in: Lecture Notes in Comput. Sci., vol. 5457, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 188–199.
[7] F. Blanchet-Sadri, R. Mercaş, G. Scott, A generalization of Thue freeness for partial words, Theoret. Comput. Sci. 410 (8–10) (2009) 793–800.
[8] A. Carpi, On the number of abelian square-free words on four letters, Discrete Appl. Math. 81 (1–3) (1998) 155–167.
[9] R. Cori, M. Formisano, Partially abelian square-free words, RAIRO Theor. Inform. Appl. 24 (1990) 509–520.
[10] L.J. Cummings, M. Mays, A one-sided Zimin construction, Electron. J. Combin. 8 (2001).
[11] F.M. Dekking, Strongly non-repetitive sequences and progression-free sets, J. Combin. Theory Ser. A 27 (2) (1979) 181–185.
[12] V. Dickert, Research topics in the theory of free partially commutative monoids, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS 40 (1990) 479–491.
[13] P. Erdős, Some unsolved problems, Magyar Tudományos Akadémia Matematikai Kutató Intézete Közl. 6 (1961) 221–254.
[14] A.A. Evdokimov, Strongly asymmetric sequences generated by a finite number of symbols, Dokl. Akad. Nauk SSSR 179 (1968) 1268–1271 (in Russian). English translation: Soviet Math. Dokl. 9 (1968) 536–539.
[15] V. Halava, T. Harju, T. Kärki, Square-free partial words, Inform. Process. Lett. 108 (5) (2008) 290–292.
[16] V. Halava, T. Harju, T. Kärki, P. Séébold, Overlap-freeness in infinite partial words, Theoret. Comput. Sci. 410 (8–10) (2009) 943–948.
[17] J. Justin, Characterization of the repetitive commutative semigroups, J. Algebra 21 (1972) 87–90.
[18] V. Keränen, Abelian squares are avoidable on 4 letters, in: W. Kuich (Ed.), ICALP 1992, 19th International Colloquium on Automata, Languages and Programming, in: Lecture Notes in Comput. Sci., vol. 623, Springer-Verlag, Berlin, 1992, pp. 41–52.
[19] V. Keränen, New abelian square-free DT0L-languages over 4 letters, in: 5th International Arctic Seminar, Murmansk State Pedagogical Institute, Murmansk, Russia, 2002.
[20] V. Keränen, A powerful abelian square-free substitution over 4 letters, Theoret. Comput. Sci. 410 (38–40) (2009) 3893–3900.
[21] M. Korn, Maximal abelian square-free words of short length, J. Combin. Theory Ser. A 102 (2003) 207–211.
[22] V. Laakso, Musical rendering of an infinite repetition-free string, in: C. Gefwert, P. Orponen, J. Seppänen (Eds.), Logic, Mathematics and the Computer, vol. 14, Finnish Artificial Intelligence Society, Symposiosarja, Hakapaino, Helsinki, 1996, pp. 292–297.
[23] M. Lothaire, Combinatorics on Words, Cambridge University Press, Cambridge, 1997.
[24] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.
[25] F. Manea, R. Mercaş, Freeness of partial words, Theoret. Comput. Sci. 389 (1–2) (2007) 265–277.
[26] S.M. Mirkin, Expandable DNA repeats and human disease, Nature 447 (2007) 932–940.
[27] M. Morse, G.A. Hedlund, Symbolic dynamics, II. Sturmian trajectories, Amer. J. Math. 62 (1) (1940) 1–42.
[28] M. Morse, G.A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, Duke Math. J. 11 (1) (1944) 1–7.
[29] P.A.B. Pleasants, Non repetitive sequences, Math. Proc. Cambridge Philos. Soc. 68 (1970) 267–274.
[30] R.L. Rivest, Abelian Square-Free Dithering for Iterated Hash Functions, MIT, 2005, http://people.csail.mit.edu/rivest/publications.html.
[31] A. Thue, Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I Mat. Nat. Kl. Christiana 7 (1906) 1–22, reprinted in: T. Nagell (Ed.), Selected Mathematical Papers of Axel Thue, Universitetsforlaget, Oslo, Norway, 1977, pp. 139–158.
[32] 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 (Ed.), Selected Mathematical Papers of Axel Thue, Universitetsforlaget, Oslo, Norway, 1977, pp. 413–478.
[33] A.I. Zimin, Blocking sets of terms, Math. USSR Sbornik 47 (1984) 353–364.