Abelian Square-Free Partial Words*

F. Blanchet-Sadri[1], Jane I. Kim[2], Robert Mercaş[3],
William Severa[4], and Sean Simmons[5]

[1] Department of Computer Science, University of North Carolina,
P.O. Box 26170, Greensboro, NC 27402–6170, USA, blanchet@uncg.edu
[2] Department of Mathematics, Columbia University,
2960 Broadway, New York, NY 10027-6902, USA
[3] Harriet L. Wilkes Honors College, Florida Atlantic University,
5353 Parkside Dr., Jupiter, FL 33458, USA
[4] GRLMC, Universitat Rovira i Virgili, Departament de Filologies Romàniques,
Av. Catalunya 35, Tarragona, 43002, Spain, robertmercas@gmail.com
[5] Department of Mathematics, University of Texas at Austin,
1 University Station C1200, Austin, TX 78712-0233, USA

Abstract. Erdös raised the question whether there exist infinite abelian square-free words over a given alphabet (words in which no two adjacent subwords are permutations of each other). 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 (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. 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.

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 a Norwegian mathematician Axel Thue [1–3] at the beginning of the last century. One of the remarkable discoveries made by Thue is that the 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*.

Erdös [4] 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

* 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/abelianrepetitions for automated use of programs.

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 Erdös' problem was taken in [5], where it was shown that the 25th abelian powers were avoidable in the binary case. Later on, Pleasants [6] showed that there exists an infinite abelian square-free word over five letters, using a uniform iterated morphism of size fifteen. This result was improved in [7] using uniform morphisms of size five.

Dekking [8] proved that over a binary alphabet there exists a word that is abelian 4-free. Moreover, using $\mathbb{Z}_7$ instead of $\mathbb{Z}_5$, in the proof of this 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 [9], 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 are obtained by cyclic permutation of letters in the preceding words.

Most of the currently known methods, [10], for constructing arbitrarily long abelian square-free words over a four-letter alphabet are based on the structure of this endomorphism g. Moreover, it is shown that no smaller uniform morphism works here! In [11] 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*, 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) [12]. 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 with problems concerning primitive sets of integers, lattices, vertex connectivity, etc [13].

In [14], 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 [15], the authors settled the question of overlap-freeness by showing that over a two-letter alphabet there exist overlap-free infinite partial words with at most 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. The problem of square-freeness in partial words is settled in [15] and [16] 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 [14, 15, 17], 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 degree of the word. Furthermore in [18], 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 size of alphabet 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 without expanding beyond the minimal four-letter alphabet. For more than one hole, the minimal number of letters is at least five, when such words exist. 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 discuss some constructions for the finite case. Finally in Section 6, we conclude with some directions for future work.

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 ith symbol of w, where $0 \leq i < |w|$. The *empty word* is the sequence of length zero and is denoted by ε. The set of all words over A is denoted by A^*, while the set of all partial words over A is denoted by $A_\diamond^*$. A (right) (resp., two-sided) infinite partial word is a function $w : \mathbb{N} \to A_\diamond$ (resp., $w : \mathbb{Z} \to A_\diamond$).

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 $y = \varepsilon$. 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 $\mathrm{rev}(w)$. 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 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 $a \in A$.*

A morphism $\phi : A^* \to B^*$ is called *abelian square-free* if $\phi(w)$ is abelian square-free whenever w is abelian square-free.

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. Using a computer it can be checked that the maximum length of an abelian square-free partial word, over such an alphabet, is six. So to construct infinite partial words with a finite number of holes, we need at least four letters. Let us first state some remarks.

Remark 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 between positions 0 and 1, 3 and 4, and 5 and 6, yields *adbacdbda* which is abelian square-free.

Remark 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 Erdös in [4]. 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 [9,11,19], five-letter [6], and larger alphabets [20]. 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. Let u, v be partial words of equal length. If uv is an abelian square, then so is any permutation of u concatenated with any permutation of v.

Theorem 1. *There exists an infinite abelian square-free partial word with one hole over a four-letter alphabet.*

Proof. We use an abelian square-free morphism $\phi : A^* \rightarrow A^*$, where $A = \{a, b, c, d\}$, provided by Keränen [19] that is defined by

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

The length of each image is 102 and the Parikh vector of each is a permutation of $P(\phi(a)) = \langle 21, 31, 27, 23 \rangle$. We show that the word $\diamond\phi^n(a)$ is abelian square-free for all integers $n \geq 0$. Since ϕ is abelian square-free, it is sufficient to check if we have abelian squares uv that start with the hole, for $|u| = |v|$.

We refer to the factors created by the images of ϕ as blocks. 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 $\diamond\phi(w_0 e)$, and $|x| < 102$. If we delete the same number of occurrences of any given block present in both $\phi(w_0)$ and $\phi(w_1)$, we claim that we only need to consider the case where $0 \leq |w_1| \leq |w_0| < 2$ (the case $|w_1| > |w_0|$ obviously leads to $|u| < |v|$, which is a contradiction). If this were not the case, then the reduced w_0 and w_1 would have no letter in common. 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.

For example, the system for letter a is determined by

$$|u'|_{\phi(a)} + |u'|_{\phi(b)} + |u'|_{\phi(c)} + |u'|_{\phi(d)} = |v|_{\phi(b)} + |v|_{\phi(c)} + |v|_{\phi(d)} + \Lambda$$

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

$$\begin{aligned}
21|u'|_{\phi(a)} + 23|u'|_{\phi(b)} + 27|u'|_{\phi(c)} + 31|u'|_{\phi(d)} &= 23|v|_{\phi(b)} + 27|v|_{\phi(c)} + 31|v|_{\phi(d)} + \lambda_a\\
31|u'|_{\phi(a)} + 21|u'|_{\phi(b)} + 23|u'|_{\phi(c)} + 27|u'|_{\phi(d)} &= 21|v|_{\phi(b)} + 23|v|_{\phi(c)} + 27|v|_{\phi(d)} + \lambda_b\\
27|u'|_{\phi(a)} + 31|u'|_{\phi(b)} + 21|u'|_{\phi(c)} + 23|u'|_{\phi(d)} &= 31|v|_{\phi(b)} + 21|v|_{\phi(c)} + 23|v|_{\phi(d)} + \lambda_c\\
23|u'|_{\phi(a)} + 27|u'|_{\phi(b)} + 31|u'|_{\phi(c)} + 21|u'|_{\phi(d)} &= 27|v|_{\phi(b)} + 31|v|_{\phi(c)} + 21|v|_{\phi(d)} + \lambda_d
\end{aligned}$$

The parameter Λ is an error term taking values in $\{-1, 0, 1\}$, but can only be non-zero for one system of equations (this is because it obviously replaces only one of the images). Each λ_i represents the error caused by $\diamond$, $\phi(e)$ or x. Using Gaussian elimination, it is easy to see that this system is inconsistent provided that some λ_i is distinct from 0. However, the hole at the beginning ensures at least one non-zero λ_i. Thus, $w_0 = w_1 = \varepsilon$ yielding $uv = \diamond\phi(e)x$, or $w_0 = f \in A$ and $w_1 = \varepsilon$ yielding $uv = \diamond\phi(f)\phi(e)x$. It is easy to verify that all such partial words are abelian square-free. □

Corollary 1. *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 $\phi'(w) = \text{rev}(\phi(w))$ with $\phi : A^* \to A^*$ being the morphism from the proof of Theorem 1. 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 $n \geq 0$. 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 integers $n \geq 0$.

We show that $\phi'^n(a)\diamond\chi^n(e) \in \{a, b, c, d, e\}^*$, is abelian square-free for all integers $n \geq 0$. 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 integer $n \geq 0$. 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 2. *The word $\phi'^n(a)\diamond efg\diamond\phi^n(a) \in \{a, b, c, d.e, f, g\}^*_\diamond$ is a two-sided infinite abelian square-free partial word with two holes over a seven-letter alphabet.*

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. 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 $|u| \leq 10, |v| \leq 10$ or with $|u| = 11, |v| = 5$ contain abelian squares (and consequently so do the words with $|u| = 11$ and $|v| \geq 5$).

Proposition 1. *Over a four-letter alphabet, there exists no two-sided infinite abelian square-free partial word with one hole, and all right infinite partial words contain at most one hole.*

In addition, over a four-letter alphabet, for all words u and v, $|u|, |v| \leq 12$, the partial word $\diamond u\diamond v\diamond$ contains an abelian square.

Proposition 2. *If a finite partial word over a four-letter alphabet contains at least three holes, then it has an abelian square.*

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.

Theorem 2. *There exists an abelian square-free partial word with infinitely many holes over a seven-letter alphabet.*

Proof. According to [9], there exists an infinite abelian square-free word w over a four-letter alphabet $A = \{a, b, c, d\}$. Furthermore, there exist some distinct letters $x, y, z \in A$ so that for infinitely many j's we have $w(j-1) = z$, $w(j) = x$, and $w(j+1) = y$. Let k_0 be the smallest integer such that $w(k_0 - 1) = z$, $w(k_0) = x$ and $w(k_0+1) = y$. Then, define k_j recursively, where k_j is the smallest integer such that $k_j > 5k_{j-1}$, $w(k_j - 1) = z$, $w(k_j) = x$ and $w(k_j + 1) = y$. Moreover, define $A' = A \cup \{e, f, g\}$, where $e, f, g \notin A$.

Note that in order to avoid abelian squares, the holes must be somehow sparse. We now define an infinite partial word w'. For any integer $i \geq 0$, there exists some integer $j \geq 0$ so that $k_j - 1 \leq i < k_{j+1} - 1$. If $j \equiv 0 \bmod 5$, then for $i = k_j - 1$, let $w'(i) = e$; for $i = k_j$, let $w'(i) = \diamond$; for $i = k_j + 1$, let $w'(i) = f$. If $j \not\equiv 0 \bmod 5$ and $i = k_j$, then let $w'(i) = g$. For all other i's, let $w'(i) = w(i)$. Clearly, w' has infinitely many holes. The modulo 5 here helps prevent the creation of squares, by assuring that the occurrences of a letter grow faster than the ones of the hole.

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 $J_1 = \{j \mid i \leq k_j \leq i+l\}$ and $J_2 = \{j \mid i+l+1 \leq k_j \leq i+2l+1\}$. Then $|J_1| < 3$ and $|J_2| < 2$, which implies $|J_1 \cup J_2| < 4$. To see this, first assume that $|J_2| > 1$. Note that there exists $j \in J_2$ so that $j+1 \in J_2$. However, this implies that $l = i+2l+1-(i+l+1) \geq k_{j+1} - k_j > k_j > i+l \geq i+l-i = l$, a contradiction. Now assume that $|J_1| > 2$. Then there are at least two occurrences of the letter g in u, and for each occurrence of g there must also be a g or a hole in v. However, g's and holes only occur when $i = k_j$ for some j, so this implies $|J_2| \geq 2$, which violates the claim that $|J_2| < 2$.

Next, we want to show that no holes occur in the abelian square uv. We prove none occurs in u, the case when the hole is in v being similar. The occurrence of a hole in u implies that there exists j so that $i \leq k_j \leq i+l$. Note that this implies $l > 0$, since otherwise $uv = \diamond w(i+1)$ would be a trivial square. Therefore either e or f occurs in u, since u must contain either $w'(k_j - 1)$ or $w'(k_j + 1)$. Assume that e occurs, the f case being similar. Then v must contain either e or a hole, but that implies $i + l \leq k_{j+5} - 1 \leq i + 2l + 1$, since $w'(k_{j+5} - 1)$ is the next occurrence in w' of either e or a hole. Thus $j, j+1, \ldots, j+4 \in J_1 \cup J_2$ which implies that $|J_1 \cup J_2| \geq 5 > 3$, a contradiction, so no such hole can exist. Therefore, all symbols in uv are letters in A'. By Remark 2, since w' contains an abelian square, w must also contain an abelian square. □

Using a similar construction we can reduce the alphabet size to six.

Theorem 3. *There exists an abelian square-free partial word with infinitely many holes over a six-letter alphabet.*

The next question is whether or not it is possible to construct such partial words over a five-letter alphabet. Although somehow superfluous, the previous two theorems give both the method and history that were used to prove our main result. First let us state two lemmas that help us achieve our goal.

Lemma 1. *Let z be a word which is not an abelian square, x (resp., y) be a prefix (resp., suffix) of $\phi(e)$, where ϕ is defined as in Theorem 1 and $e \in \{b, c, d\}$. No word of the form $\phi(z)y$, $a\phi(z)y$ or $x\phi(z)$, preceded or followed by a hole, is an abelian square, unless either ez or ze is an abelian square.*

Lemma 2. *Let z be a word which is not an abelian square, x (resp., y) be a prefix (resp., suffix) of $\phi(a)$, where ϕ is defined as in Theorem 1. Then, no word of the form $\diamond x\phi(z)y$ is an abelian square, unless az or za is an abelian square.*

Theorem 4. *There exists an abelian square-free partial word with infinitely many holes over a five-letter alphabet.*

Proof. Let us denote by w the infinite abelian square-free full word over $A = \{a, b, c, d\}$ from the proof of Theorem 1. There exist infinitely many j's such that $w[j-101..j] = \phi(a)$. Let k_0 be the smallest integer so that $w[k_0-101..k_0] = \phi(a)$. Then define k_j recursively, where k_j is the smallest integer such that $k_j > 5k_{j-1}$ and $w[k_j - 101..k_j] = \phi(a)$.

Construct an infinite partial word w' over $A \cup \{e\}$ by introducing factors in w as follows. Let $j \geq 0$. 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. Moreover, if the holes are not taken into consideration, since no two e's are next to each other, by Remark 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 $J_1 = \{j \mid i \leq k_j \leq i + l\}$ and $J_2 = \{j \mid i + l + 1 \leq k_j \leq i + 2l + 1\}$. Then $|J_1| < 4$ and $|J_2| < 2$, which implies $|J_1 \cup J_2| < 5$. As in Theorem 2, it is trivial to show that $|J_2| < 2$. Now assume that $|J_1| > 3$. Then there are at least seven occurrences of the letter e in u, and for each occurrence of e there must also be an e or a hole in v. However, this implies $|J_2| \geq 2$, which violates the fact that $|J_2| < 2$.

Next, we want to show that no holes occur in the abelian square uv. First observe that v cannot contain more than four e's, since otherwise, $|J_1| > 6$, a contradiction. If the last position of u is a hole, then v contains an e. If no e occurs in u, then the hole in u and the e in v are cancelling each other, giving us a factor of the original word, which is abelian square-free. So there must exist an e in u. But, this implies that there exists an abelian square of one of the forms $e\phi(z_0)\diamond e\phi(z_1)y$ or $ae\phi(z_0)\diamond e\phi(z_1)y$, for some words $z_0, z_1, y \in A^*$, with $|z_0| = |z_1|$ and $|y| \in \{1, 2\}$. After cancelling the e's and any common letters from z_0 and z_1, we get that $P(a\phi(z_0))$ and $P(\phi(z_1)y)$ differ in only one component (by only one), and z_0, z_1 have different letters (otherwise the letters would cancel each other). It is easy to see that this is impossible.

Let us now assume that the last position of v is a hole. If v would contain any e's, then we would get a contradiction with the fact that $|J_2| < 2$. Moreover, u does not contain any e's, since otherwise the hole and the e would cancel each other and we would get that the original word is not abelian square-free. Hence, there exist words $x, z \in A^*$ with $|x| < 102$, such that $x\phi(z)\diamond$ is an abelian square. By Lemma 1, this is impossible. If v has a hole in any other position, then v also contains an e. Again we get that u contains an e, and so, if an abelian square exists, it would be of one of the forms $e\phi(z)y\diamond e$ or $ea\phi(z)y\diamond e$, for some words y, z with $|y| < 102$. After cancelling the e's, this is also impossible by Lemma 1.

Now let us consider the case when a position in u, other than the last one, is a hole. If $|J_1| = 1$, since u contains the $\diamond$ and an e, then v also contains an e. Hence, we have that either $\diamond ex\phi(z)e$ or $\diamond ex\phi(z)ea$, for some words x, z with $|x| < 102$, are abelian squares, which is a contradiction by Lemma 1. If $|J_1| = 3$, then the only possibilities are that either $ae\phi(z_0)\diamond e\phi(z_1)eaebXeceae\phi(z_2)y$ or $e\phi(z_0)\diamond e\phi(z_1)eaebXeceae\phi(z_2)y$ are abelian squares. Since, the hole can be taken to one end and the e's and the common images of ϕ cancel, we get that either $\diamond a\phi(z)y$ or $\diamond\phi(z)y$ are abelian squares, for some y with $|y| < 102$. According to Lemma 1 no such factors preceded by the hole would create an abelian square. If $|J_1| = 2$, then the case when X' comes after the hole in u is impossible, since then the length of v would be greater than that of u. If X' comes before the hole in u, then u contains one more e than the suffix of X' from u.

We reach a contradiction with Lemma 2, hence, all symbols in uv correspond to letters in $A \cup \{e\}$. The conclusion follows as in the proof of Theorem 2. □

5 The distinct 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 [21] 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 2. **[21]** *Let* $\{a_0, \ldots, a_{k-1}\}$ *be a* k*-letter alphabet. The Zimin words* z_i *are defined by* $z_0 = a_0$ *for* $i = 0$*, and* $z_i = z_{i-1}a_iz_{i-1}$ *for* $1 \leq i < k$*.*

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 3. *Let* $\{a_0, \ldots, a_{k-1}\}$ *be a* k*-letter alphabet. For* $1 < i < k$*, 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 $\diamond bacab, ab\diamond cab, bac\diamond ba, bacab\diamond$. □

In [22], 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 3. **[22]** *Let* $\{a_0, \ldots, a_{k-1}\}$ *be a* k*-letter alphabet. Left Zimin words* y_i *are defined recursively as follows: For* $i = 0$, $y_0 = a_0$. *For* $i = 1, \ldots, k-1$, $y_i = y_{i-1}a_i z_{\lfloor \frac{i-1}{2} \rfloor}$, *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 $i = 0, \ldots, k-1$.

Proposition 4. *Let* $\{a_0, \ldots, a_{k-1}\}$ *be a* k*-letter alphabet. For each* $5 \le i < k$, *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 $k = 6$, we find by exhaustive search that no hole can replace any letter of y_5 without creating an abelian square. Assuming that the result is true for $y_5, \ldots, y_{k-1}$, consider $y_k = y_{k-1}a_k z_{\lfloor \frac{k-1}{2} \rfloor}$, where $z_{\lfloor \frac{k-1}{2} \rfloor}$ is a Zimin word. By Proposition 3, 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. □

In [23], 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 4. **[23]** *Let* $\{a_0, \ldots, a_{k-1}\}$ *be a* k*-letter alphabet, where* $k \ge 4$. *The words* v_i *are defined recursively by* $v_0 = a_2a_1$ *for* $i = 0$, *and* $v_i = v_{i-1}a_{i+2}a_{i+1}$ *for* $1 \le i \le k-3$. *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, $w_4 = acdbcdacbdcedacdecda$.

Proposition 5. *Let* $A = \{a_0, \ldots, a_{k-1}\}$ *be a* k*-letter alphabet, where* $k \ge 4$, *and* $w_{k-1} \in A^*$ *be constructed according to Definition 4. The replacement of any letter in* w_{k-1} *with a hole results in a word containing an abelian square.*

Proof. After replacing the first or last letter with a hole, v_{k-3} remains abelian square-free. 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 would 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.

It is not possible to replace letters with holes in the subword $v_{k-3}[1..|v_{k-3}|-1)$ 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 $a_0ua_1ua_0v_{k-3}[0..2j+4)$. 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 second factor has an extra occurrence of a_{j+2} compared to the first factor. Since an a_{j+1} from the first factor is replaced by a $\diamond$, this yields an abelian square, with the $\diamond$ corresponding to a_{j+2} in the prefix of v_{k-3}. $\square$

6 Conclusion

As a possible topic for future work, we propose the study of avoidance of abelian powers greater than two. From [8], we know that over a binary alphabet, we can construct an infinite word that avoids 4-powers. In the context of partial words, which abelian powers can be avoided over a binary alphabet? In this case, certain repetitions are created, since, if w is an abelian square-free partial word of length n and a is the most common letter of w, then it is easy to see that $|w|_a + |w|_\diamond \leq \lceil \frac{n}{2} \rceil$.

Another interesting topic is to replace letters with holes in arbitrary positions of a word, without changing the word's properties of abelian repetition-freeness. Note that introducing holes two positions apart would create an abelian square, $\diamond ab\diamond$. Moreover, note that even if the word is defined in such a way that each two occurrences of a letter are four symbols apart, the resulting word could contain an abelian square. For example, if we consider

$$a_1a_2a_3a_4\underline{a_0}a_5a_6a_7a_8\underline{a_0}a_9a_{10}a_{11}a_{12}$$

in order to avoid abelian squares, no two a_i's are a_j's, no two a_j's are a_k's, and all a_i's, a_j's, and a_k's are different, for $0 < i \leq 4 < j \leq 8 < k \leq 12$. Also, if we allow for holes to be three positions apart, only one of the a_k's equals an a_j.

Also, investigating the number of abelian square-free partial words over an alphabet of size five would be interesting. This number has been studied in [10] and [19] for full words over an alphabet of size four.

References

1. 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).
2. 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).

3. Berstel, J.: Axel Thue's work on repetitions in words. In Leroux, P., Reutenauer, C., eds.: Invited Lecture at the 4th Conference on Formal Power Series and Algebraic Combinatorics. (1992) 65–80
4. Erdös, P.: Some unsolved problems. Magyar Tudományos Akadémia Matematikai Kutató Intézete **6** (1961) 221–254
5. Evdokimov, A.: Strongly asymmetric sequences generated by a finite number of symbols. Doklady Akademii Nauk SSSR **179** (1968) 1268–1271 in Russian. English translation in *Soviet mathematics - Doklady* **9** (1968) 536–539.
6. Pleasants, P.: Non repetitive sequences. Proceedings of the Cambridge Philosophical Society **68** (1970) 267–274
7. Justin, J.: Characterization of the repetitive commutative semigroups. Journal of Algebra **21** (1972) 87–90
8. Dekking, F.: Strongly non-repetitive sequences and progression-free sets. Journal of Combinatorial Theory **27**(2) (1979) 181–185
9. Keränen, V.: Abelian squares are avoidable on 4 letters. In Kuich, W., ed.: ICALP '92: Proceedings of the 19th International Colloquium on Automata, Languages and Programming, London, UK, Springer-Verlag (1992) 41–52
10. Carpi, A.: On the number of abelian square-free words on four letters. Discrete Applied Mathematics **81**(1-3) (1998) 155–167
11. Keränen, V.: New abelian square-free DT0L-languages over 4 letters. In: Proceedings of the Fifth International Arctic Seminar, Murmansk, Russia, Murmansk State Pedagogical Institute (2002)
12. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. Theoretical Computer Science **218** (1999) 135–141
13. Blanchet-Sadri, F.: Algorithmic Combinatorics on Partial Words. Chapman & Hall/CRC Press (2008)
14. Manea, F., Mercaş, R.: Freeness of partial words. Theoretical Computer Science **389**(1-2) (2007) 265–277
15. Blanchet-Sadri, F., Mercaş, R., Scott, G.: A generalization of Thue freeness for partial words. Theoretical Computer Science **410**(8-10) (2009) 793–800
16. Halava, V., Harju, T., Kärki, T.: Square-free partial words. Information Processing Letters **108**(5) (2008) 290–292
17. Blanchet-Sadri, F., Mercaş, R., Rashin, A., Willett, E.: An answer to a conjecture on overlaps in partial words using periodicity algorithms. In Dediu, A.H., Ionescu, A.M., Martín-Vide, C., eds.: Language and Automata Theory and Applications 2009. Volume 5457 of Lecture Notes in Computer Science., Berlin, Germany, Springer-Verlag (2009) 188–199
18. Halava, V., Harju, T., Kärki, T., Séébold, P.: Overlap-freeness in infinite partial words. Theoretical Computer Science **410**(8-10) (2009) 943–948
19. Keränen, V.: A powerful abelian square-free substitution over 4 letters. Theoretical Computer Science **410**(38-40) (2009) 3893–3900
20. Evdokimov, A.: The existence of a basis that generates 7-valued iteration-free sequences. Diskretnyĭ Analiz **18** (1971) 25–30
21. Zimin, A.I.: Blocking sets of terms. Mathematics of the USSR Sbornik **47** (1984) 353–364
22. Cummings, L.J., Mays, M.: A one-sided Zimin construction. The Electronic Journal of Combinatorics **8** (2001)
23. Korn, M.: Maximal abelian square-free words of short length. Journal of Combinatorial Theory, Series A **102** (2003) 207–211