Improved bounds on the length of maximal abelian square-free words

Evan M. Bullock*

Department of Mathematics
Rice University, Houston, Texas, USA
evanmb@rice.edu

Submitted: Sep 28, 2003; Accepted: Feb 9, 2004; Published: Feb 20, 2004
MR Subject Classifications: 68R15, 05D99

Abstract

A word is abelian square-free if it does not contain two adjacent subwords which are permutations of each other. Over an alphabet Σ_k on k letters, an abelian square-free word is maximal if it cannot be extended to the left or right by letters from Σ_k and remain abelian square-free. Michael Korn proved that the length $\ell(k)$ of a shortest maximal abelian square-free word satisfies $4k-7 \le \ell(k) \le 6k-10$ for $k \ge 6$. In this paper, we refine Korn's methods to show that $6k-29 \le \ell(k) \le 6k-12$ for $k \ge 8$.

1 Introduction

We consider words over an alphabet of k letters, which will be taken to be $\Sigma_k = \{0, 1, ..., k-1\}$. We say b is a *subword* of d if d is the concatenation abc, where a and c are words, possibly empty. An *abelian square* is a word of the form

$$a_1 a_2 \cdots a_n a_{\tau(1)} a_{\tau(2)} \cdots a_{\tau(n)},$$

where the a_i are letters and τ is a permutation of $\{1, \dots n\}$. A word is *abelian square-free* if it does not contain a nonempty subword which is an abelian square. For example, any word in which the same letter appears twice in a row is not abelian square-free.

A word w over Σ_k is *left-maximal* if, for every $a \in \Sigma_k$, the word aw contains an abelian square subword beginning with the initial a. In particular, an abelian square-free word w is left-maximal over Σ_k if aw contains an abelian square for every $a \in \Sigma_k$. Similarly, a word w is *right-maximal* over Σ_k if for every $a \in \Sigma_k$, wa contains an abelian square

*26 Greenough St., Newton, MA 02465

subword ending with the final a. A word is *maximal* over Σ if it is both left-maximal and right-maximal.

Abelian square-free words were first introduced by Erdős [4], who asked for a characterization of alphabet sizes for which arbitrary long abelian square-free words exist (see [5], [8], and [6]). Maximal abelian square-free words were first constructed by Zimin [9], who gave the following recursive construction: $z_1 = 0$ and $z_k = z_{k-1}(k-1)z_{k-1}$ for $k > 1$. The first four Zimin words are

$$\begin{aligned} z_1 &= 0, \\ z_2 &= 010, \\ z_3 &= 0102010, \\ z_4 &= 010201030102010. \end{aligned}$$

It is easy to check that z_k is maximal abelian square-free over Σ_k and that the length of z_k is $2^k - 1$. Thus for any positive integer k, we may define $\ell(k)$ to be the minimum length of a maximal abelian square-free word over Σ_k. Zimin's construction shows that $\ell(k) \leq 2^k - 1$. Cummings and Mays [2] gave a construction improving Zimin's bound to $\ell(k) = O(2^{k/2})$. Recently, Korn [7] constructed maximal abelian square-free words of length $6k - 10$ over Σ_k for $k \geq 4$, a dramatic improvement over previous constructions. The first four words of Korn's construction are the following:

$$\begin{aligned} y_4 &= 02120213202320, \\ y_5 &= 02312302132430234230, \\ y_6 &= 02341234021324354023452340, \\ y_7 &= 02345123450213243546502345623450. \end{aligned}$$

In Section 2, we will describe a similar construction which is shorter by two, improving the upper bound on $\ell(k)$ to $6k - 12$.

Korn also proved that any left-maximal word over Σ_k has length at least $4k - 7$. In fact, for left-maximal words this bound is optimal; taking the first $4k - 7$ letters of Korn's construction yields a left-maximal abelian square-free word. In Section 3, we will use left-maximality and right-maximality together to show that $\ell(k) \geq 6k - 29$.

For more details on the history of abelian square-free words, see [7].

2 The upper bound for $\ell(k)$

In this section, we will construct maximal abelian square-free words of length $6k - 12$ over Σ_k for $k \geq 8$, proving the following theorem.

Theorem 2.1. *For $k \geq 8$, the upper bound $\ell(k) \leq 6k - 12$ holds.*

For integers a and b with $0 \leq a \leq b < k$, let $u_{a,b}$ denote $a(a+1)\cdots(b-1)b$, the concatenation of the integers from a to b in ascending order. For integers a and b with

$1 \leq a \leq b < k$, let $v_{a,b}$ denote $a(a-1)(a+1)a\cdots(b-1)(b-2)b(b-1)$, the concatenation in ascending order of a block $i(i-1)$ for each i with $a \leq i \leq b$. We are now ready to present our construction. For $k \geq 8$, set

$$x_k = 0v_{1,k-4}(k-3)(k-1)(k-4)(k-2)u_{3,k-2}0(k-1)u_{1,k-4}1302v_{4,k-1}(k-1)$$

The first three words x_k are the following:

$$\begin{aligned} x_8 &= 010213243574634560712341302435465767, \\ x_9 &= 010213243546857345670812345130243546576878, \\ x_{10} &= 010213243546579683456780912345613024354657687989. \end{aligned}$$

We note that the words x_k are symmetric in the sense that reversing x_k produces the same result as replacing every appearance of i with $k-1-i$. Also, it is easy to check that the length of x_k is $6k-12$. Thus to prove Theorem 2.1, it will suffice to prove the following:

Lemma 2.2. *For $k \geq 8$, the word x_k is left-maximal and abelian square-free.*

Proof. First, we show that x_k is left-maximal. When the letters 0, 1, and 2 are added to the left of x_k, they are contained in abelian squares 00, 1010, and 201021, respectively.

When 3 is added to the left of x_k, the two subwords

$$30v_{1,k-4}(k-3)(k-1)(k-4)(k-2)$$

and

$$u_{3,k-2}0(k-1)u_{1,k-4}1302$$

are consecutive and each contains the letters $k-1$, $k-2$, and $k-3$ exactly one time, the letter 3 exactly three times, and every other letter exactly two times. Thus together they form an abelian square, $s_3 = 30\cdots u_{1,k-4}1302$.

Adding the first pair from the $v_{4,k-1}$ block, namely 43, to the right side of s_3 increases its length by two. This shifts the half-way point to the right by one, moving the initial 3 in the $u_{3,k-2}$ block from the right half to the left, for a net addition of a 3 to the left half and a 4 to the right half. Thus if we also change the letter initially added to the left of x_k from a 3 to a 4, we get another abelian square, $s_4 = 40\cdots u_{1,k-4}130243$. Continuing in this way, we find that when $i > 3$ is added to the left of x_k, it is included in an abelian square whose first half is

$$i0v_{1,k-4}(k-3)(k-1)(k-4)(k-2)u_{3,i-1}$$

and whose second half is

$$u_{i,k-2}0(k-1)u_{1,k-4}1302v_{4,i}.$$

Finally, we show that x_k is abelian square-free. First, we note that any non-empty subword of $0v_{1,k-4}$ contains an odd number of occurrences of some letter. In particular, it contains exactly one occurrence of the greatest letter it contains, since between two occurrences of i there is always an $i+1$. This shows that no subwords of $0v_{1,k-4}$ are abelian squares. It shows also that no subwords of x_k starting within $0v_{1,k-4}$ and ending at the end of x_k are abelian squares, since every letter occurs an even number of times in x_k. By symmetry, the same holds for $v_{4,k-1}(k-1)$.

Now, suppose w is a non-empty abelian square subword of x_k. If we delete every instance of some letter from w, then the result is an abelian square since the same number of occurrences of the given letter must have been deleted from the left half of w as from the right. The resulting abelian square is a subword, possibly empty, of the word obtained from x_k by deleting every occurrence of the same letter. Now, if we delete from x_k every letter besides the first three and the last three in the alphabet, for all k, the resulting word is

$$y = 010212\bar{2}\bar{0}\bar{1}\bar{2}\bar{1}0\bar{0}12102\bar{2}\bar{1}\bar{2}\bar{0}\bar{1}\bar{0},$$

where $\bar{i} = k-1-i$. If w contains none of the first three and last three letters, either it must lie entirely in some block $u_{i,j}$, which contains only one of any given letter and thus certainly contains no abelian squares, or it must lie entirely in some $v_{i,j}$ block, which contains no abelian squares by the work above. On the other hand, one can check that the only non-empty abelian squares in the above word y over $\{0,1,2,\bar{0},\bar{1},\bar{2}\}$ are $010212\bar{2}\bar{0}\bar{1}\bar{2}\bar{1}0\bar{0}12102$ and its reflection, $\bar{2}\bar{0}\bar{1}\bar{2}\bar{1}0\bar{0}12102\bar{2}\bar{1}\bar{2}\bar{0}\bar{1}\bar{0}$. If either of these is the result of deleting the letters $\{3,4,\ldots,k-4\}$ from w, then w must be either a word starting within the $0v_{1,k-4}$ block and ending at the end of x_k or a word starting at the beginning of x_k and ending within the $v_{4,k-1}(k-1)$ block; but we have shown above that in this case, w is not an abelian square, which is a contradiction. $\square$

3 The lower bound for $\ell(k)$

Our goal in this section is to show that $\ell(k) \geq 6k-29$ by proving the following.

Theorem 3.1. *Any maximal word over Σ_k has length at least $6k-29$. Moreover, for $k \geq 7$ any maximal word over Σ_k has length at least $5k-11$.*

The proof consists of a series of technical lemmas. We begin by recalling that if we remove every occurrence of some letter from an abelian square, we must have removed the same number of occurrences of that letter from the first half as from the second, whence the result is again an abelian square. It follows that if w is a left-maximal word over Σ_k, then the word obtained from w by deletion of every occurrence of $a \in \Sigma_k$ is left-maximal over $\Sigma_k - \{a\}$. The same holds for right-maximal words and for maximal words. (Note, however, that the word obtained from an abelian square-free word by deleting every occurrence of some letter need not be abelian square-free.) This observation on the deletion of letters provides the basic argument in our proof of the lower bound.

Lemma 3.2. *Assume that A, B, and C are positive integers with $A \geq B$ such that*

1. *no maximal word over Σ_A contains 5 or fewer occurrences of each letter,*

2. *no maximal word over* Σ_B *contains* 4 *or fewer occurrences of each letter, and*

3. *every maximal word over* Σ_{B-1} *has length at least* C.

Then for $k \geq B-1$, *every maximal word over* Σ_k *has length greater than or equal to* $\max\{5(k-B+1)+C, 6(k-A+1)+5(A-B)+C\}$.

Proof. First, we note that $6(k-A+1)+5(A-B)+C > 5(k-B+1)+C$ only when $k \geq A$ and that the result is obvious when $k = B-1$.

Let w be a maximal word over Σ_k for $k \geq B$. There must be at least $k-B+1$ letters in Σ_k that occur at least 5 times in w. If not, deleting from w every instance of $k-B$ letters including all those appearing at least 5 times would yield a maximal word over an alphabet of size $k-(k-B) = B$ in which each letter appears at most 4 times, contradicting assumption 2. Thus deleting from w every instance of $k-B+1$ letters which occur at least 5 times yields a maximal word on an alphabet of size $B-1$, whence w has length at least $5(k-B+1)+C$.

Similarly, if $k \geq A$, there must be at least $k-A+1$ letters in Σ_k which appear in w at least 6 times, and removing every instance of these $k-A+1$ letters yields a maximal word on an alphabet of size $A-1$; applying the above argument to this word shows that the length of w is at least $6(k-A+1)+5(A-1-B+1)+C$, as desired. □

The remainder of this section will be devoted to showing that the values $A = 19$, $B = 8$, $C = 24$ satisfy the hypotheses of Lemma 3.2. This will give the lower bound $6k-29$ stated in Theorem 3.1 for $k \geq 7$, and for $k < 7$, this bound follows immediately from the proof of the lower bound $\ell(k) \geq 4k-7$ in [7]. For $k \geq 7$, Lemma 3.2 will also imply the second lower bound $5k-11$ stated in Theorem 3.1, which is better when $7 \leq k < 18$.

Our arguments will be based on an analysis of left-maximal words, especially those in which each letter appears at most four times. We begin by recalling the argument used to prove the lower bound in [7].

Let w be a left-maximal word over Σ_k. For each $i \in \Sigma_k$, there is a shortest initial subword of w which yields an abelian square when i is added to it on the left. Different values of i give rise to different initial subwords, so we can permute Σ_k so that $w = w_0w_1w_2\cdots w_{k-1}e$ and the word $s_i = iw_0w_1\cdots w_i$ is an abelian square for $1 \leq i \leq k-1$. Clearly $w_0 = 0$. Also, left-maximality places no restriction on e, so in analyzing the structure of left-maximal words, we will generally be able to assume that e is empty. For $0 \leq i \leq k-1$, the word $0w_1\cdots w_{i-1}$ contains an even number of occurrences of every letter except $i-1$ and an odd number of occurrences of $i-1$. Similarly, $0w_1\cdots w_{i-1}w_i$ contains an odd number of occurrences of i and an even number of each other letter. Thus w_i contains an odd number of occurrences of $i-1$ and i, and also w_i contains an even number of occurrences of every other letter.

Now, assume $k > 1$. The halfway point of the abelian square s_{k-1} must lie in w_m for some $1 \leq m \leq k-1$. More precisely, for some m, the first half of s_{k-1} has the form $(k-1)0w_1\cdots w_{m-1}t_1$ and the second half has the form $t_2w_{m+1}\cdots w_{k-1}$, where $w_m = t_1t_2$ and t_2 may be empty.

For $m < i < k-1$, the letter i appears at least twice in the second half of s_{k-1}, namely once in w_i and once in w_{i+1}, but s_{k-1} is an abelian square, so i must also appear at least twice in the first half of s_{k-1}, and hence i must appear at least four times in w. Similarly, if $0 \leq i < m-1$, the letter i must appear twice in the first half of s_{k-1} and hence four times in w. The letter $k-1$ must appear in w_{k-1} and the letter $m-1$ must occur at least twice, as must m unless $m = k-1$. From this we can conclude that the length of w is at least $4k-3-2-2 = 4k-7$.

Lemma 3.3. *Let $w = 0w_1 \cdots w_{k-1}$ be a left-maximal word over Σ_k with $k \geq 6$ in which each letter appears at most four times. Let m be defined as above. Assume also that it is not the case that w contains a single occurrence of $k-1$ in the last six letters and no other occurrence of $k-1$. Then $m \leq 3$ and w_m contains a subword containing $m+1$, $m+2$, …, $k-1$ in some order, followed by a subword containing m, $m+1$,…, $k-1$ in increasing order. Furthermore $|w_i| = 2$ for all $1 \leq i \leq k-1$ except for $i = m$ and possibly for one other value of i where w_i contains an additional pair of $m-1$'s.*

Proof. Assume $m \geq 4$. For $0 \leq i \leq 2$, the letter i appears two times in w before the start of w_m, since $2 < m-1$. Thus 0, 1, and 2 must also appear at least two times in w_m or later, since s_{k-1} is an abelian square whose halfway point lies in w_m. Now, consider the word obtained by deleting from $0w_1 \cdots w_{m-1}$ every occurrence of a letter greater than $m-1$. The result is a left-maximal word over Σ_m. Moreover, since the letters 0, 1, and 2 each occur at most four times in w and occur at least two times in w_m or later, these three letters appear at most two times in the resulting word. However, $2 < m-1$, so none of these letters is equal to $m-1$, and the above discussion shows that in a left-maximal word over Σ_m, at most two letters besides $m-1$ can appear fewer than four times. From this contradiction we conclude that $m \leq 3$.

Now, for $1 \leq i \leq k-1$, define h_i so that $ih_1h_2 \cdots h_i$ is the first half of the abelian square $s_i = i0w_1w_2 \cdots w_i$. Then h_i must have half the length of w_i and, more specifically, h_i must contain $i-1$ and one j for each additional pair of j's contained in w_i. This is because the initial $i-1$ in s_{i-1} has been removed and an $i-1$ has been added to the right, so an $i-1$ must be moved from the right to the left for the result to be an abelian square. Similarly, if two of some letter is added to the right, one of that letter must be shifted to the left half for the result to be an abelian square.

By the definition of m, the word $h_1 \cdots h_{k-1}$ is contained in $0w_1 \cdots w_m$. Thus the letters m, $m+1$,…, $k-2$ appear in increasing order somewhere in $0w_1 \cdots w_m$, and since the m is in h_{m+1}, all are in the right half of s_m. For $m+1 \leq i \leq k-2$, the letter i must appear in the left half of s_m, since it appears in the right half. The two appearances of i in s_m, however, must both be in the same word w_j for some $j \leq m$, since the number of appearances of i in each word w_j with $j \leq m$ must be even: the only w_j in which i appears an odd number of times are those where $j = i, i+1$ and here we have assumed $i > m$. In particular, the $k-2$ in h_{k-1} must be in the same w_j as another $k-2$ in the left half of s_3, whence the m, $m+1$, …, $k-3$ in $h_{m+1}, \ldots, h_{k-2}$ must all be in the same w_j as the $k-2$. We have thus established that some w_j with $1 \leq j \leq m$ contains a subword containing $m+1$, $m+2$, …, $k-2$ in some order, followed by a subword containing m,

$m+1, \ldots, k-2$ in increasing order. Note that since $k \geq 6$, we know $m+1 \leq k-2$, so since $m+1$ occurs twice in w_j, the half-way point of s_j is between them and we may conclude that the half-way points of s_j, s_{j+1}, …, s_{k-2} all lie between letters in w_j.

Assume $j = m$. In this case, we are nearly done. For $m < l < k-1$, we have already determined the location of four occurrences of l. Moreover, if $m = 2$ or 3, there are two occurrences of 0 in $0w_1$, which must be on the left half of s_m, so there must be two more occurrences of 0 in w_m which are on the right half of s_i for all $i \geq m$. Similarly, if $m = 3$, there are two occurrences of 1 in $0w_1w_2$ so there must be two more occurrences of 1 in w_m on the right half of s_i for all $i \geq 3$. Thus $m-1$, m, and $k-1$ are the only remaining letters which might appear in a pair in w_i for some $i \neq m$. It is not possible for there to be a pair of m's, $(m-1)$'s, or $(k-1)$'s in w_i for $i < m$, since then m, $m-1$, or $k-1$ would appear three times on the left side of s_{k-1}. Similarly, since one m is already known to appear in w_{m+1} on the right side of s_{k-1}, there can not be a pair of m's in w_i for $i > m$. If a pair of $(k-1)$'s appeared in w_i for $i > m$, then another $k-1$ would appear in h_i, and this $k-1$ must therefore be part of another pair of $(k-1)$'s appearing in w_m, giving a total of at least five $(k-1)$'s. Thus $|w_i| = 2$ for all but possibly one $i \neq m$ where w_i may contain an additional pair of $(m-1)$'s, as desired. Finally, since w_{k-1} has length at most four, in order for it not to be the case that w contains exactly one occurrence of $k-1$ and that $k-1$ is one of the last six letters of w, the word w must not contain exactly one occurrence of $k-1$. Thus a pair of $(k-1)$'s appears in w_i for some i, and we have already eliminated every possibility except for $i = m$. Further, if there is a pair of $(k-1)$'s in w_m, one must be on the left half of s_m and hence before the m in h_{m+1}, and the other must be on the right half of s_{k-1} and hence after the $k-2$ in h_{k-1}.

It now remains only to show that $j = m$. If $m = 1$, then this is obvious. If $m = 3$ and $j = 1$, since the halfway point of s_{k-2} is between letters in w_j but the halfway point of s_{k-1} is after at least one letter in w_3, all of w_2 is contained in h_{k-1}, so there is a pair of 2's in w_{k-1}. The half-way point for s_3, however, is in w_1, so all four appearances of 2 are to the right of the half-way point for s_3, but there are two occurrences of 2 in s_3, one of which must be on the left half of s_3, which is a contradiction.

If $m = 3$ and $j = 2$, two 0's occur in the left half of s_2. Thus there must be two more occurrences of 0 in w_2, which must be on the right half of s_i for $i \geq 2$, but the halfway point for s_{k-1} lies in w_3 so they can't be on the right half of s_{k-1}, which is a contradiction.

Assume now that $m = 2$ and $j = 1$. We will show that in this case $k-1$ can occur at most one time and that it is one of the last six letters of w. Since 2 appears in w_1, it must appear twice in w_1. Thus four of every letter but 0, 1, and $k-1$ have already been located. Even if there are pairs of both 0's and 1's in w_{k-1}, a $k-1$ will still then be one of the last six letters of w, so it remains only to show that there can be no more than one $k-1$ in w. The number of $(k-1)$'s in w must be odd and at most four, hence either one or three. If there were three $(k-1)$'s, then one of the additional $(k-1)$'s would have to be in the right half of s_{k-1} and one would have to be in the left. Since both would have to be in the same w_i, both must be in w_2, but then the first, being in w_2 and on the left half of s_{k-1}, would be in h_{k-1}, and there would then be an additional pair of $(k-1)$'s in w_{k-1}, which is a contradiction. $\square$

The $m = 2$ and $j = 1$ case considered at the end of the proof of the lemma actually can occur, as in the word 0234512034512320430541615. We will not be interested in this case, however, because a single occurrence of $k-1$ at the end of a left-maximal word makes it impossible to extend the word to the right to make a short maximal word.

More precisely, let $w = 0w_1 \cdots w_{k-1}e$ be a maximal word over Σ_k with $k \geq 8$ in which each letter appears at most four times. Assume only one $k-1$ occurs in $0w_1 \cdots w_{k-1}$ and it is within the last 6 letters. We will show this is impossible.

For convenience, let $\tilde{w}$ be the reversal of w, and let $\tilde{w} = \tilde{w}_0\tilde{w}_1 \cdots \tilde{w}_{k-1}\tilde{e}$ be the decomposition of $\tilde{w}$ analogous to the decomposition $w = 0w_1 \cdots w_{k-1}e$, so that for some permutation σ of Σ_k, the word $\tilde{s}_i = \sigma(i)\tilde{w}_0 \cdots \tilde{w}_i$ is an abelian square for all $i \in \Sigma_k$.

Now we know $0w_1 \cdots w_{k-1}$ contains four occurrences of every letter except possibly m and $m-1$, which may occur only twice, and $k-1$, which may occur only 1 or 3 times. Thus $|e| \leq 5$, since if there were two or more $(k-1)$'s in e, then these $(k-1)$'s would surely be in the first half of $\tilde{s}_{k-1}$, which is impossible since then there would only be at most one $k-1$ in the right half $\tilde{s}_{k-1}$. Thus every occurrence of $k-1$ is in the last $5+6 = 11$ letters of w, and hence in the first 11 letters of $\tilde{w}$. However, $\tilde{w}$ is a left-maximal word over Σ_k and hence has at least $4k-7$ letters by the lower bound in [7]. Therefore, there are at least $2k-4$ letters of $\tilde{w}$ in the first half of $\tilde{s}_{k-1}$. Since $k \geq 8$, the first 12 letters of $\tilde{w}$ are all in the first half of $\tilde{s}_{k-1}$, whence every occurrence of $k-1$ in w is in the first half of $\tilde{s}_{k-1}$, which is a contradiction.

The same argument can be used to show that if $k \geq 15$, there can be no maximal word $w = 0w_1 \cdots w_{k-1}e$ over Σ_k in which each letter appears at most five times but only one $k-1$ occurs in $0w_1 \cdots w_{k-1}$ and it occurs within the last 6 letters.

The following lemma shows that $A = 19$ satisfies the first hypothesis in Lemma 3.2.

Lemma 3.4. *For $k \geq 19$, there is no maximal word over Σ_k in which each letter appears at most five times.*

Proof. Assume, for the sake of a contradiction, that there is a maximal word w' over Σ_{k+2}, for some $k \geq 17$, in which each letter appears at most five times. First, let $w' = 0w'_1 \cdots w'_{k+1}e'$ be the usual decomposition of w'. Then the number of occurrences of each letter in Σ_{k+2} other than $k+1$ in $w' = 0w'_1 \cdots w'_{k+1}$ is an even number which is at most 5, hence at most 4. On the other hand, $k+1$ may appear 5 times in $w' = 0w'_1 \cdots w'_{k+1}$. We thus delete from w every instance of the letter $k+1$ and also every instance of $\sigma(k+1)$, where $\sigma(i)$ is, as above, the letter that would have the same role as i if we reversed w. The result, after renaming letters, is a maximal word, $w = 0w_1 \cdots w_{k-1}e$, over Σ_k with reversal $\tilde{w} = \tilde{w}_0\tilde{w}_1 \cdots \tilde{w}_{k-1}\tilde{e}$ such that every letter occurs at most five times in w and at most four times in $0w_1 \cdots w_{k-1}$ and in $\sigma(0)\tilde{w}_1 \cdots \tilde{w}_{k-1}$.

We will say a letter $i \in \Sigma_k$ is *typically configured* in w if it appears in $0w_1 \cdots w_{k-1}$ exactly four times and the number of letters between the first and second occurrences and the number of letters between the second and third occurrences of i are both at least three but the number of letters between the third and fourth occurrences of i is at most two. It will suffice to show that at least $k-8$ letters are typically configured, since then the majority of the letters would be typically configured, and the same would hold for the reversal, $\tilde{w}$, so at least one letter would be typically configured for both w and $\tilde{w}$. But this is impossible, since the third and fourth occurrences of i from the left would be either the first and second or the second and third occurrences of i from the right.

By the discussion following Lemma 3.3, the hypotheses of Lemma 3.3 are satisfied by $0w_1 \cdots w_{k-1}$. Thus w_m, where $m \leq 3$, contains a subword containing $m+1$, $m+2$, …, $k-1$ in some order, followed by a subword containing m, $m+1$,…, $k-1$ in increasing order, and all but at most one of $w_6, w_7, \ldots, w_{k-1}$ are of length 2.

If all have length 2 or if w_j has length 4 for $4 \leq j \leq 6$, then $7, 8, \ldots, k-2$ are all typically configured; for if $7 \leq i \leq k-2$, then the last two occurrences of i are in w_i and w_{i+1}. Since these each have length two, the last two occurrences of i have at most two letters between them. The second and third occurrences of i certainly have at least three letters between them, since w_4 and w_5 are both between them, whereas the first and second occurrences of i have the letters 3, 4, and 5 between them.

If w_j has length four where $6 < j \leq k-1$, then $m-1$ is typically configured, as is every i satisfying $6 \leq i \leq k-2$ except for $j-1$ and j. It follows from the same argument as above that every i satisfying $6 \leq i \leq k-2$ except for $j-1$ and j is typically configured. On the other hand, $m-1$ is typically configured since the last two occurrences of $m-1$ are both in w_j, which has length four, whereas the first two occurrences, the first of which is in w_{m-1} and the second of which is in h_j which is contained in w_m, are separated from each other by h_4, h_5, h_6 and from the last two by w_4 and w_5. Thus in either case, at least $k-8$ letters are typically configured, as desired. $\square$

The next lemma shows that $B = 8$ satisfies the second hypothesis in Lemma 3.2.

Lemma 3.5. *For $k \geq 8$, there is no maximal word over Σ_k in which each letter appears at most four times.*

Proof. Deleting all instances of some letters as necessary, it suffices to consider the case where $k = 8$. Suppose that $w = 0w_1 \cdots w_7 e$ is a maximal word over Σ_8 in which each letter appears at most four times. By the discussion following Lemma 3.3, the hypotheses of Lemma 3.3 are satisfied by $0w_1 \cdots w_7$.

The length of $0w_1 \cdots w_7$ is at least $(4)(8) - 5 = 27$. The length of $w_4w_5w_6w_7$ is either 8 or 10, where in the latter case, the length of $0w_1 \cdots w_7$ is at least 29. In either case then, $0w_1w_2w_3$ has length at least 19 and the number of letters of w on the left side of s_3 is at least 9.

Now, $w_4w_5w_6w_7e$ certainly contains a 3, two 4's, two 5's, two 6's and a 7. It may also contain up to two 2's, up to two additional 3's, and an additional 7. Even if it contained all of these, however, the last 9 letters of w would include all of w_6 and w_7. This implies that the reversal of w_6w_7 is contained in the first half of $\tilde{s}_3$, and, in particular, there are two occurrences of 6 in $\tilde{w}$ which are in the first half of $\tilde{s}_3$, whence the remaining two occurrences of 6 in $\tilde{w}$ must also be in $\tilde{s}_3$.

Now, recall that by Lemma 3.3, we have $m \leq 3$ and w_m contains a subword containing $m+1$, $m+2$, …, 7 in some order, followed by a subword containing m, $m+1$,…, 7 in increasing order. In particular, the leftmost occurrence of 6 is before the subword containing m, $m+1$,..., 7 in that order. The leftmost 6 must lie on the left side of s_m, whereas the m in the second subword must lie on the right side of s_m, since it is in h_{m+1}. Thus for $i \geq m$, there are two i's to the right of the leftmost 6, at least one of which is not in s_m, so there is at most one i to the left of the leftmost 6 in $0w_1 \cdots w_m$.

Since there are at least three letters which appear twice in $\tilde{w}_4\tilde{w}_5\tilde{w}_6\tilde{w}_7$, it must be that $m = 3$ and the subword to the left of the leftmost 6 in w contains two 0's, two 1's, and two 2's. However, w_1 must consist of 0 and 1 in some order, and w_2 of 1 and 2, so the two letters in $\tilde{w}_4\tilde{w}_5\tilde{w}_6\tilde{w}_7$ which are not in a pair must all be among the first three letters of $\tilde{w}_4\tilde{w}_5\tilde{w}_6\tilde{w}_7$. But one of the two letters which is not in a pair is in $\tilde{w}_7$, which is a contradiction. $\square$

Lemma 3.5 is sharp: for $k = 7$, the word 0102132645345061230124354656 is maximal abelian square-free with exactly four appearances of each letter. Note that, together with the following lemma which shows that $C = 24$ satisfies the third hypothesis in Lemma 3.2, this example shows that $24 \leq \ell(7) \leq 28$.

Lemma 3.6. *If $k \geq 6$, and w is a maximal word over Σ_k, then $|w| \geq 4k - 4$.*

Proof. Let $w = 0w_1 \cdots w_{k-1}e$ be a maximal word over Σ_k of length at most $4k - 5$ for $k \geq 6$. Then the length of $0w_1 \cdots w_{k-1}$ is an odd number which is at least $4k - 7$, by the lower bound in [7], hence it must be either $4k - 7$ or $4k - 5$.

In the first case, there is only one occurrence of $k - 1$ in $0w_1 \cdots w_{k-1}$, and the word $0w_1 \cdots w_{k-2}$ is left-maximal over Σ_{k-1} and hence has length at least $4(k - 1) - 7$. From this we may conclude that $|w_{k-1}| \leq 4$, and every $k - 1$ in w is in $w_{k-1}e$, which has length at most 6. Now, each half of $\tilde{s}_{k-1}$ has at least $2k - 4$ letters, so at least the last $2(6) - 4 - 1 = 7$ letters of w are in the first half of $\tilde{s}_{k-1}$. However, the last 6 letters of w include every occurrence of $k - 1$ in this case, and it is impossible for every occurrence of $k - 1$ to be in the first half of $\tilde{s}_{k-1}$, which is a contradiction.

We may now assume that $w = 0w_1 \cdots w_{k-1}$, with length exactly $4k - 5$. It is not possible that w contains only a single occurrence of $k - 1$ which is in the last six letters since this would yield the same contradiction as above. Thus the hypotheses of Lemma 3.3 are satisfied, and we conclude that w must contain exactly 3 occurrences of $k - 1$, exactly 2 occurrences of m and $m - 1$, and exactly 4 occurrences of every other letter. Indeed Lemma 3.3 provides, in some sense, the location of every letter: $m \leq 3$ and w_m contains a subword containing $m + 1$, $m + 2$, ..., $k - 1$ in some order, followed by a subword containing m, $m + 1$,..., $k - 1$ in that order, whereas $|w_i| = 2$ for all $i \neq m$. In particular, $|w_{k-1}| = 2$, so there is an occurrence of $k - 1$ in the last two letters of w.

Now, we may apply the same argument to $\tilde{w}$ and conclude that there is an occurrence of $k - 1$ within the first two letters of w: since $k - 1$ is the only letter to occur exactly three times, it must also be the letter which forms an abelian square when added to the left of $\tilde{w}$, i.e. $\sigma(k - 1) = k - 1$. But the first occurrence of $k - 1$ in w is in w_m, so we may conclude that $m = 1$ and that $k - 1$ is the second letter of w. But w_{k-1} also has length two, so the last four letters of w contain one letter twice, namely $k - 2$. Applying the same argument to $\tilde{w}$, we may conclude that the first four letters of w also contain two occurrences of the same letter. But w must start with a 0, followed by a block containing $2, 3, \ldots, k-1$ in some order, and nothing else, since every occurrence of every letter has already been accounted for. Thus it is impossible that the first four letters of w include some letter twice, and we have a contradiction. $\square$

Lemma 3.6 shows in particular that every maximal word over Σ_7 has length at least 24. It follows from Lemmas 3.4, 3.5, and 3.6 that $A = 19$, $B = 8$, and $C = 24$ satisfy the hypotheses of Lemma 3.2, which completes the proof of Theorem 3.1.

Lemma 3.6 also shows that, as Korn [7] verified by a computer search, the word 01021453405120143545 is maximal abelian square-free of minimum length for $k = 6$. Since Korn's computer search was for maximal abelian square-free words rather than maximal words, it does not provide a proof of Lemma 3.6. It seems reasonable, however, to expect that Lemma 3.6 could be verified computationally. Moreover, any improvement of the bounds in Theorem 3.1 for a particular value of $k > 6$ would imply the same improvement for all greater values of k. It thus might be interesting, for example, to compute the minimum length of a maximal word over Σ_7.

4 Conclusion

Although we have determined $\ell(k)$ asymptotically, there is still a gap between the upper and lower bounds. Nor is it clear how important the original abelian square-freeness condition is: we have found no example of a k such that there are shorter maximal words than there are maximal abelian square-free words. Even within the framework of Lemma 3.2, it seems plausible that improvements might be made; although Lemma 3.5 is known to be sharp, we have no reason to believe that Lemma 3.4 is. Indeed, we know of no examples of maximal words over Σ_k, with $k > 7$, in which every letter appears at most five times. Also, one might generalize the basic problem being considered; for example, it is easy to generalize the Zimin construction to show the existence of maximal abelian lth-power-free words, and one might consider how short such words can be in the general case. The problem of characterizing alphabet sizes for which there are arbitrarily long abelian lth-power-free words was solved in the $l > 2$ case by Dekking [3]. It has also been shown by Bean, Ehrenfeucht and McNulty [1] that any abelian lth-power-free word can be extended to a maximal abelian lth-power-free word.

5 Acknowledgments

This research was done at the 2003 Research Experience for Undergraduates at the University of Minnesota Duluth. I would like to thank the program director Joseph Gallian for suggesting the problem and I would like to thank Patrick Headley, Geir Helleloid, and Phil Matchett for many useful comments on this paper. Finally, I want to thank all the other student participants and visitors for a wonderful summer. This work was supported by NSA grant MDA904-02-1-0060 and NSF grant DMS-0137611.

References

[1] D. R. Bean, A. Ehrenfeucht and G. F. McNulty, Avoidable patterns in strings of symbols, Pacific J. Math. **85** (1979), no. 2, 261–294.

[2] L. J. Cummings, M. Mays, A one-sided Zimin construction, Electron. J. Combin. 8 (2001), #R27.

[3] F. M. Dekking, Strongly nonrepetitive sequences and progression-free sets, J. Combin. Theory Ser. A **27** (1979), no. 2, 181–185.

[4] P. Erdős, Some unsolved problems, Magyar Tud. Akad. Mat. Kutató Int. Közl. **6** (1961), 221–254.

[5] A. A. Evdokimov, Strongly asymmetric sequences generated by a finite number of symbols, Dokl. Akad. Nauk SSSR **179** (1968), 1268–1271; Soviet Math. Dokl. **9** (1968), 536–539.

[6] V. Keränen, Abelian squares are avoidable on 4 letters, *Automata, languages and programming (Vienna, 1992)*, 41–52, Lecture Notes in Comput. Sci., 623, Springer, Berlin, 1992.

[7] M. Korn, Maximal abelian square-free words of short length, J. Combin. Theory Ser. A 102 (2003), 207–211.

[8] P. A. B. Pleasants, Non-repetitive sequences, Proc. Cambridge Philos. Soc. **68** (1970), 267–274.

[9] A.I. Zimin, Blocking sets of terms, Mat. Sb. (N.S.) 119 (161) (1982), 363–375, 447.