How Many Holes Can an Unbordered Partial Word Contain?*

F. Blanchet-Sadri[1], Emily Allen[2], Cameron Byrum[3], and Robert Mercaş[4]

[1] University of North Carolina, Department of Computer Science, P.O. Box 26170, Greensboro, NC 27402–6170, USA blanchet@uncg.edu
[2] Carnegie Mellon University, Department of Mathematical Sciences, 5032 Forbes Ave., Pittsburgh, PA 15289, USA
[3] University of Mississippi, Department of Mathematics, P.O. Box 1848, University, MS 38677, USA
[4] GRLMC, Universitat Rovira i Virgili, Plaça Imperial Tárraco, 1, Tarragona, 43005, Spain robertmercas@gmail.com

Abstract. Partial words are sequences over a finite alphabet that may have some undefined positions, or "holes," that are denoted by $\diamond$'s. A nonempty partial word is called *bordered* if one of its proper prefixes is compatible with one of its suffixes (here $\diamond$ is compatible with every letter in the alphabet); it is called *unbordered* otherwise. In this paper, we investigate the problem of computing the maximum number of holes a partial word of a fixed length can have and still fail to be bordered.

1 Introduction

Motivated by a practical problem in gene comparison, Berstel and Boasson introduced the notion of *partial words*, or sequences over a finite alphabet that may contain some "holes" denoted by $\diamond$'s [1]. 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, partitions of integers and their generalizations, vertex connectivity in graphs, etc [3].

An unbordered word is a word such that none of its proper prefixes is one of its suffixes. *Unbordered partial words* were defined in [2], and two types of borders were identified: *simple* and *nonsimple*. In this paper, we investigate the maximum number of holes an unbordered partial word of length n over a k-letter alphabet can have.

The contents of our paper are as follows: In Section 2, we compute the maximum number of holes a nonsimply bordered partial word of length n over a

* This material is based upon work supported by the National Science Foundation under Grant No. DMS–0754154. We thank the referees of a preliminary version of this paper for their very valuable comments and suggestions. A World Wide Web server interface has been established at www.uncg.edu/cmp/research/countingpwords for automated use of the program. This work was done during the fourth author's stay at the University of North Carolina at Greensboro.

k-letter alphabet may have, and show that this maximum number is constant over all integers $k \geq 2$. In Section 3, we investigate the maximum number of holes an unbordered partial word of length n over a k-letter alphabet may have, obtaining an upper bound. An exact formula for $k = 2$ and lower bounds for $k \geq 3$ are also derived. Experimental evidence suggests that for $k \geq 4$, the numbers are constant. In Section 4, we conclude with some remarks. We end this section with an overview of basic concepts of combinatorics on partial words.

Let A be a nonempty finite set of symbols called an *alphabet*. Each element $a \in A$ is called a *letter*. A *full word* u over A is a finite sequence of letters from A. A *partial word* u over A is a finite sequence of symbols from $A_\diamond = A \cup \{\diamond\}$, the alphabet A being extended with the "hole" symbol $\diamond$ (a *full word* is a partial word that does not contain the $\diamond$ symbol). We will denote by u_i the symbol at position i of the partial word u.

The *length* of a partial word u is denoted by $|u|$ and represents the total number of symbols in u. The *empty word* is the sequence of length zero and is denoted by ε. For a partial word u, the powers of u are defined recursively by $u^0 = \varepsilon$ and for $n \geq 1$, $u^n = uu^{n-1}$. The set of all words over the alphabet A is denoted by A^*, while the set of all partial words over A is denoted by $A^*_\diamond$.

If u and v are two partial words of equal length, then u is said to be *contained in* v, denoted $u \subset v$, if $u_i = v_i$, whenever $u_i \in A$. Partial words u and v are *compatible* if there exists a partial word w such that $u \subset w$ and $v \subset w$. This is denoted by $u \uparrow v$. A partial word u is a *factor* of a partial word v if there exist partial words x, y such that $v = xuy$. The factor u is *proper* if $u \neq \varepsilon$ and $u \neq v$. We say that u is a *prefix* of v if $x = \varepsilon$ and a *suffix* of v if $y = \varepsilon$.

A partial word u is called *unbordered* if no nonempty partial words x_1, x_2, v, w exist such that $u = x_1v = wx_2$ and $x_1 \uparrow x_2$. If such nonempty words exist, then there exists a partial word x such that $x_1 \subset x$ and $x_2 \subset x$. In this case, we call u *bordered* and call x a *border* of u. A border x of u is called *minimal* if $|y| < |x|$ implies that y is not a border of u. We have two distinct types of borders: For a partial word $u = x_1v = wx_2$ where $x_1 \subset x$ and $x_2 \subset x$, we say that x is a *simple* border if $|x| \leq |v|$, and a *nonsimple* border otherwise. A bordered partial word u is called *simply bordered* if a minimal border x exists such that $|u| \geq |2x|$. For example, the word $a\diamond b\diamond bb$ has the simple and minimal border abb and the nonsimple border $abbbb$, and thus it is simply bordered.

2 Simply Bordered Partial Words

Once the number of holes in a partial word of a fixed length reaches a certain bound, the word will have a simple border. In this section, we give a closed formula for that bound and show that it is constant over all alphabets of size at least two.

Let us first recall a result regarding bordered partial words.

Proposition 1 ([5]). *Suppose that u is a nonempty partial word that is bordered. Let x be a minimal border of u. Say $u = x_1v = wx_2$, where $x_1 \subset x$ and*

$x_2 \subset x$. Then (1) the partial word x is unbordered, and (2) in the case where x_1 is unbordered, $u = x_1 u' x_2 \subset x u' x$ for some u'.

It follows from Proposition 1 that if u is a full bordered word, then $x_1 = x$ is unbordered. In this case, x is the minimal border of u and $u = xu'x$. Thus, a bordered full word is always simply bordered and has a unique minimal border. Since borders for partial words are defined using containment, it is possible to have numerous borders having the same length. Thus, a partial word does not necessarily have a unique minimal border.

In [4], an open problem related to borderedness in the context of partial words was suggested by the fact that every partial word of length five that has more than two holes is simply bordered. The partial word $aa\diamond\diamond b$ shows that this bound on the number of holes for length five is tight. For length six, every partial word with more than two holes is simply bordered as well. What is the maximum number of holes $m_k(n)$ a partial word of length n over an alphabet of size k can have and still fail to be simply bordered? Some values for small n follow: $m_2(5) = 2$, $m_2(6) = 2$, $m_2(7) = 3$, $m_2(8) = 4$, $m_2(9) = 5$, $m_2(10) = 5$, $m_2(11) = 6$, $m_2(12) = 7$, $m_2(13) = 8$, $m_2(14) = 8$, and $m_2(15) = 9$. We end this section by answering this open problem.

Theorem 1. *For $k \geq 2$ and $l \geq 1$, the following equalities hold: $m_k(1) = 1$, $m_k(2l) = 2l - \left(\left\lfloor\sqrt{l}\right\rfloor + \left\lceil\frac{l}{\lfloor\sqrt{l}\rfloor}\right\rceil\right)$ and $m_k(2l+1) = m_k(2l) + 1$.*

Proof. Let A be a k-letter alphabet where $k \geq 2$, and let a, b be two distinct letters of A. We prove the lower bound by constructing a partial word $w(n)$ of length n over A with $m_k(n)$ holes, that is not simply bordered. Take $w(1) = \diamond$, and for $l \geq 1$,

$$w(2l) = (a\diamond^{\lfloor\sqrt{l}\rfloor-1})^{\frac{l}{\lfloor\sqrt{l}\rfloor}}\diamond^{l-\lfloor\sqrt{l}\rfloor}b^{\lfloor\sqrt{l}\rfloor}$$
$$w(2l+1) = (a\diamond^{\lfloor\sqrt{l}\rfloor-1})^{\frac{l}{\lfloor\sqrt{l}\rfloor}}\diamond^{l+1-\lfloor\sqrt{l}\rfloor}b^{\lfloor\sqrt{l}\rfloor}$$

where a fractional power of the form $(a_0 \dots a_{i-1})^{\frac{mi+j}{i}}$ with $0 \leq j < i$ is equal to $(a_0 \dots a_{i-1})^m a_0 \dots a_{j-1}$. It is easy to check that for this construction we never have a prefix of $w(n)$ of length at most $\lfloor\frac{n}{2}\rfloor = l$ compatible with a suffix. For $n \geq 4$, this is due to the fact that no factor of length $\lfloor\sqrt{l}\rfloor + 1$ of the prefix is compatible with the suffix $\diamond b^{\lfloor\sqrt{l}\rfloor}$, since each such factor has an a among its last $\lfloor\sqrt{l}\rfloor$ positions.

Now, we prove the upper bound. Let us observe that the simply bordered words of odd length are not influenced by the middle character. Hence, this character can always be replaced by a hole so that the number of holes is maximal. Because of that we can only look at the even length case. Let us consider a partial word $w = a_0 \dots a_{2l-1}$ of length $n = 2l \geq 4$ that is not simply bordered. Obviously both a_0 and a_{2l-1} are distinct letters of the alphabet A in order to avoid a trivial one-letter border. Note that any two factors $a_0 \dots a_{i-1}$ and $a_{2l-i} \dots a_{2l-1}$ differ in at least one position for any $0 < i \leq l$. In order to avoid having the second half of w formed only of letters, we need in the first half at least two occurrences of letters. Let us suppose that a_i is the second occurrence of a letter in the first half of w (the first occurrence is a_0, that is, a_0 and a_i are letters and between them there are only holes). This implies that $a_{2l-i} \dots a_{2l-1} \in A^*$. In other words, the suffix of length l of the word ends with a word of length i over A, since otherwise we again would get compatibility for a shorter factor. Now if we look at the prefix of length $2i$, we observe that we need a second incompatibility relation with the suffix of the same length. This implies that there exists another occurrence of a letter either in the prefix at a position j, with $j \leq 2i$, or in the suffix at position j, with $j > n - 2i$. Continuing the reasoning, and looking at the problem for the following occurrences of letters in each half, we will finally get an expression of the form $i + \frac{l}{i}$ for which we have to find the minimum value, for $0 < i \leq l$. Calculating the first derivative of $i + \frac{l}{i}$ and equating to zero, we get that $i = \sqrt{l}$. Hence the minimum number of letters (it is the number of holes that we wish to be maximized) is $\left\lfloor \sqrt{l} \right\rfloor + \left\lceil \frac{l}{\lfloor \sqrt{l} \rfloor} \right\rceil$, i.e., the number of consecutive occurrences of letters from the end of the word plus the number of occurrences of letters in the first half of the word. Furthermore, we observe that the upper bound coincides with the lower bound and the obtained computer values. □

Note that Theorem 1 implies that the equality $m_k(n) = m_2(n)$ holds for all $k \geq 2, n \geq 1$.

3 Bordered Partial Words

The previously defined concept of the maximum number of holes a "nonsimply bordered" partial word may have can be extended to an "unbordered" partial word. Let $\hat{m}_k(n)$ be the maximum number of holes a partial word of length n over a k-letter alphabet can have and still fail to be bordered. For all integers $k \geq 2$ and $n \geq 1$, the inequality

$$\hat{m}_k(n) \leq m_2(n) \tag{1}$$

holds. To see this, consider a partial word u of length n over a k-letter alphabet with more than $m_k(n)$ holes. The word u necessarily has a simple border, so u is bordered and $\hat{m}_k(n)$ cannot be greater than $m_k(n)$. The inequality then follows by Theorem 1 which implies that $m_k(n) = m_2(n)$.

We now give an interval of values for $\hat{m}_k(n)$ for $k \geq 3$ and $n \geq 1$. We start with two lemmas.

Lemma 1. *The inequality $\hat{m}_k(st) \geq (s-1)(t-1)$ holds for all integers $k \geq 3$ and $s, t \geq 1$.*

Proof. Let a, b, c be distinct letters of an alphabet of size at least three. For all integers $i, j \geq 0$, $(a\diamond^i)^j bc^i$ is an unbordered word of length $(i+1)(j+1)$. Indeed, assume that $i, j \geq 1$ (the case where $i = 0$ or $j = 0$ is similar). For a border of length l with $1 \leq l \leq i$, the prefix $a\diamond^{l-1}$ will correspond to the suffix c^l. If $i+1 \leq l \leq j(i+1)$, an a will appear within the last $(i+1)$ positions of the prefix, and the corresponding position in the suffix will be b or c. If $(i+1)j+1 \leq l \leq (i+1)j+i$, the prefix will end with $bc^{i'}$ for some $0 \leq i' < i$. Since the suffix will end with c^i, the b in the prefix will not agree with the corresponding letter of the suffix. An unbordered word of length st can be constructed as described above with $i = s-1$ and $j = t-1$ with $(s-1)(t-1)$ holes. □

Lemma 2. *The equality* $\hat{m}_k(s^2) = (s-1)^2$ *holds for all integers* $k \geq 3$ *and* $s \geq 1$.

Proof. To demonstrate the lower bound $\hat{m}_k(s^2) \geq (s-1)^2$, construct a word as in the proof of Lemma 1 with $i = j = s-1$. This unbordered word will have length s^2 and $(s-1)^2$ holes. To demonstrate the upper bound $\hat{m}_k(s^2) \leq (s-1)^2$, assume u is an unbordered word with $h = (s-1)^2$ holes. It suffices to show that the length of u is at least s^2, which is equivalent to showing that u has at least $2\sqrt{h}+1 = 2(s-1)+1 = 2s-1$ letters. Without loss of generality, assume that u begins with an a. In order for u to be constructed such that the number of holes is maximized, we can assume that the a is followed by a string of $\diamond$'s of length l and then another letter. Thus, if we look at the last $l+1$ symbols in the word u, they must all be letters different from a. Otherwise, we would be able to construct a suffix of length at most l which would be compatible with the corresponding prefix. Repeating this procedure, the next letter is going to appear at most every $l+1$ symbols. In order to maximize the number of holes in u with respect to its length, u must be constructed so each repetition of holes has the same length. Thus, if there are r repetitions of the l holes, then there are at most rl total holes, and at least $r+l+1$ letters, since there is one letter for every repetition of holes and $l+1$ letters at the end. If we minimize the function $r+l+1$ with respect to $rl = h$ we get $r+l+1 \geq \sqrt{h}+\sqrt{h}+1 = 2\sqrt{h}+1$. Hence, there are at least $2s-1$ letters, and there are $(s-1)^2$ holes. So the length of u is at least s^2, and $\hat{m}_k(s^2) \leq (s-1)^2$. □

Theorem 2. *The inequalities* $(\lfloor\sqrt{n}\rfloor - 1)^2 \leq \hat{m}_k(n) \leq (\lceil\sqrt{n}\rceil - 1)^2$ *hold for all integers* $k \geq 3$ *and* $n \geq 1$.

Proof. The fact that $\hat{m}_k(n)$ is increasing is straightforward, since creating a word of length $n+1$ that is unbordered and has as many holes as the word of length n is easily done just by adding at the end of the word of length n a letter different from the one at the first position. Obviously, the newly created word will be bordered only if the word of length n is bordered. The bounds are an immediate consequence of Lemma 2. □

We end this section by refining the upper bound (1).

Proposition 2. *For all integers* $k \geq 2$ *and* $n \geq 2$, *we have the upper bound*

$$\hat{m}_k(n) \leq \left\lfloor n - \sqrt{\frac{2k}{k-1}(n-1)} \right\rfloor$$

Proof. Consider a partial word u of length n over a k-letter alphabet. Say $u = x_1v = wx_2$, for some partial words x_1, x_2, v and w with x_1, x_2 of length i. For u not to have a border of length i, there must exist a pair of corresponding positions from x_1, x_2 whose letters are noncompatible. Since there exist $n-1$ possible border lengths for u, there must exist at least $n-1$ such pairs of noncompatible letters for u to be unbordered.

For a given number of letters $n-h$, the maximum number of noncompatible pairs will occur when each symbol of the alphabet appears equally, which would be $\frac{n-h}{k}$ times. Thus, the maximum number of noncompatible pairs is bounded above by

$$\left(\frac{n-h}{k}\right)^2 (k-1+k-2+\cdots+1) = \left(\frac{n-h}{k}\right)^2 \left(\frac{k(k-1)}{2}\right)$$

If there are strictly less than $n-1$ noncompatible pairs of letters in u, then u is necessarily bordered. So when $n-1 > \frac{(n-h)^2(k-1)}{2k}$ holds, u will be bordered. Thus, a word with $h > n - \sqrt{\frac{2k}{k-1}(n-1)}$ holes is necessarily bordered. So we have $\hat{m}_k(n) \leq \lfloor n - \sqrt{\frac{2k}{k-1}(n-1)} \rfloor$, for all integers $k \geq 2$ and $n \geq 2$. □

3.1 A Formula for $\hat{m}_2(n)$

First, we consider the 2-letter alphabet $\{a, b\}$. For $n \geq 1$, the upper bound

$$\hat{m}_2(n) \leq \lfloor n - 2\sqrt{n-1} \rfloor \tag{2}$$

follows from Proposition 2 by letting $k = 2$ (note that the case when $n = 1$ is trivial since $\diamond$ is an unbordered word of length one with one hole). We will show that this upper bound is also a lower bound.

Proposition 3. *For all integers $i, j, k \geq 0$ where $k \leq i$, the partial word given by $(a\diamond^i)^j a\diamond^k ab^{i+1}$ is an unbordered word of length $(i+1)(j+1)+k+2$.*

Proof. Assume that $i, j \geq 1$ (the other cases are similar). Consider a prefix of length l with $1 \leq l < (i+1)$. This gives us the prefix $a\diamond^{l-1}$ and the corresponding suffix b^l. Thus, there is no border of this length. Next, consider a prefix of length l with $(i+1) \leq l < (i+1)j+k+2$. Since an a will appear within at least the last $i+1$ letters of the prefix and the corresponding position in the suffix will be b, there cannot be a border of this length. Now, consider a prefix of length l with $(i+1)j+k+2 \leq l \leq (i+1)(j+1)+k+1$. The prefix ends with $ab^{i'}$ with $i' \leq i$. Since the suffix ends with b^{i+1}, the last a in the prefix does not agree with the corresponding b in the suffix. □

Proposition 4. *For all integers $n \geq 5$, we have the lower bound*

$$\hat{m}_2(n) \geq \lfloor n - 2\sqrt{n-1} \rfloor$$

Proof. First, assume there exists an integer $l \geq 2$ such that $n = l^2 + 1$. We construct the binary word $(a\diamond^{l-1})^{l-1}ab^l$ of length l^2+1 which is unbordered for all integers $l \geq 2$ by Proposition 3. This word has $(l-1)^2$ holes. Making the substitution $n = l^2 + 1$ we have

$$\lfloor n - 2\sqrt{n-1} \rfloor = \lfloor l^2 + 1 - 2\sqrt{l^2} \rfloor = l^2 + 1 - 2l = (l-1)^2$$

Thus, there exists an unbordered word of length n with $\lfloor n - 2\sqrt{n-1} \rfloor$ holes.

Now, assume n cannot be written in the form $l^2 + 1$ for any integer l. Let

$$i = \left\lfloor \frac{-1 + \sqrt{1 + 4(n-2)}}{2} \right\rfloor + 1, j = \lceil \sqrt{n} \rceil - 3, k = n - (i+1)(j+1) - 2$$

Let $u = (a\diamond^i)^j a \diamond^k ab^{i+1}$ whose length is given by $(i+1)(j+1) + k + 2$ which is equivalent to $(i+1)(j+1) + n - (i+1)(j+1) - 2 + 2 = n$. The number of holes in u is given by $ij + k = ij + n - (i+1)(j+1) - 2 = n - i - j - 3 = n - \left\lfloor \frac{-1+\sqrt{1+4(n-2)}}{2} \right\rfloor - \lceil \sqrt{n} \rceil - 1$. In order to show that u has $\lfloor n - 2\sqrt{n-1} \rfloor$ holes, it suffices to show $\left\lfloor \frac{-1+\sqrt{1+4(n-2)}}{2} \right\rfloor + \lceil \sqrt{n} \rceil + 1 = \lceil 2\sqrt{n-1} \rceil$. First we note that for any integer $n \geq 5$, there exists a unique integer $m \geq 2$ such that $(m-1)^2 < n \leq m^2$. The next four bounds will be useful:

First, for $(m-1)(m-2) + 2 \leq n < m(m-1) + 2$, we have that $0 \leq 4(m-1)(m-2) \leq 4(n-2) < 4m(m-1)$. Thus, after adding 1 to, taking the square root of, subtracting 1 from, and dividing by 2 each part of the inequality yields $m - 2 \leq \frac{-1+\sqrt{1+4(n-2)}}{2} < m - 1$, and we get $\left\lfloor \frac{-1+\sqrt{1+4(n-2)}}{2} \right\rfloor = m - 2$.

Second, for $(m-1)^2 < n \leq m^2$, we have that $(m-1) < \sqrt{n} \leq m$ and so $\lceil \sqrt{n} \rceil = m$.

Third, for $(m-1)^2 + 1 < n \leq m(m-1) + 1$, we have $0 \leq (m-1)^2 + 1 < n \leq m(m-1) + 1.25$. Thus, after subtracting 1 from, taking the square root of, and multiplying by 2 each part of the inequality, this yields $2m - 2 < \lceil 2\sqrt{n-1} \rceil \leq 2m - 1$. Hence, we have $\lceil 2\sqrt{n-1} \rceil = 2m - 1$.

Fourth, for $m(m-1)+2 \leq n \leq m^2+1$, we have $m(m-1)+1.25 < n \leq m^2+1$. Thus, after subtracting 1 from, taking the square root of, and multiplying by 2 each part of the inequality, this yields $2m - 1 < 2\sqrt{n-1} \leq 2m$. Thus, we have $\lceil 2\sqrt{n-1} \rceil = 2m$.

Now, if $(m-1)^2 + 1 < n \leq m(m-1) + 1$, then $\left\lfloor \frac{-1+\sqrt{1+4(n-2)}}{2} \right\rfloor = m - 2$, $\lceil \sqrt{n} \rceil = m$, and $\lceil 2\sqrt{n-1} \rceil = 2m - 1$. If $m(m-1) + 2 \leq n \leq m^2$, then $\left\lfloor \frac{-1+\sqrt{1+4(n-2)}}{2} \right\rfloor = m - 1$, $\lceil \sqrt{n} \rceil = m$, and $\lceil 2\sqrt{n-1} \rceil = 2m$.

Finally we claim that u is unbordered. By Proposition 3, it suffices to show that $k \leq i$. This is equivalent to demonstrating

$$n \leq 2\lceil \sqrt{n} \rceil - \left\lfloor \frac{-1 + \sqrt{1 + 4(n-2)}}{2} \right\rfloor + \lceil \sqrt{n} \rceil \left\lfloor \frac{-1 + \sqrt{1 + 4(n-2)}}{2} \right\rfloor - 1$$

Again, let m be the unique integer such that $(m-1)^2 < n \leq m^2$. If $(m-1)^2+1 < n \leq m(m-1)+1$, then $\left\lfloor \frac{-1+\sqrt{1+4(n-2)}}{2} \right\rfloor = m-2$ and $\lceil \sqrt{n} \rceil = m$. So we have $n \leq m(m-1)+1 = 2m-(m-2)+m(m-2)-1$. If $m(m-1)+1 < n \leq m^2$, then $\left\lfloor \frac{-1+\sqrt{1+4(n-2)}}{2} \right\rfloor = m-1$ and $\lceil \sqrt{n} \rceil = m$. We get $n \leq m^2 = 2m-(m-1)+m(m-1)-1$. □

Theorem 3. *For any integer $n \geq 1$, $\hat{m}_2(n) = \lfloor n - 2\sqrt{n-1} \rfloor$.*

Proof. For $n = 1$, the result is trivial as mentioned earlier. For $n = 2$, note that a word with at least one hole necessarily has a border of length one. An unbordered word of length two with no hole is ab, and $\hat{m}_2(2) = 0 = \lfloor 2 - 2\sqrt{1} \rfloor$. For $n = 3$, $\hat{m}_2(3) = 0 = \lfloor 3 - 2\sqrt{2} \rfloor$, and an example of an unbordered word of length three with no hole is abb. As in the case of words of length two, a word that has one hole will be bordered. For $n = 4$, we can argue similarly. Thus, we have as example $abbb$, and $\hat{m}_2(4) = 0 = \lfloor 4 - 2\sqrt{3} \rfloor$. For $n \geq 5$, the result follows from (2) and Proposition 4. □

3.2 A Lower Bound for $\hat{m}_3(n)$

Now, we consider the 3-letter alphabet $\{a, b, c\}$. For $n \geq 2$, the upper bound

$$\hat{m}_3(n) \leq \left\lfloor n - \sqrt{3(n-1)} \right\rfloor \tag{3}$$

follows from Proposition 2 by letting $k = 3$. We will give a lower bound for $\hat{m}_3(n)$.

Proposition 5. *For all integers $i, j, k \geq 0$ with $k \leq i$, the partial word given by $(a\diamond^i)^j a\diamond^k c^i b$ is an unbordered word of length $(i+1)(j+1)+k+1$.*

Proof. Assume that $i, j, k > 0$ (the other cases are similar). Consider a possible border length l with $1 \leq l \leq i+1$. This yields a prefix that begins with a and a suffix which begins in b or c, so there is no border of length l. If $i+2 \leq l \leq j(i+1)+1$, we have the letter a within the last $i+1$ positions of the prefix which will correspond with c or b in the last $i+1$ positions of the suffix, so there is no border of this length. If $j(i+1)+2 \leq l \leq j(i+1)+1+k$, we have a appearing within the last $k+1$ positions of the prefix, and since $k \leq i$, the last $k+1$ positions of the suffix are c's and b's. Finally, if $j(i+1)+k+2 \leq l \leq (j+1)(i+1)+k$, we have a prefix that ends in c and a suffix which ends in b. □

Proposition 6. *For all integers $i, j \geq 2$ and $k \geq 0$, the partial word given by $(a\diamond^i)^j (b\diamond^{i+1})^k c^i b$ is an unbordered word of length $(i+1)(j+k+1)+k$.*

Proof. Assume that $i \geq 2, j \geq 2, k \geq 1$ (the case where $k = 0$ is similar). Consider a possible border length l with $1 \leq l \leq i+1$. Our prefix will begin with a which is not equal to the corresponding b or c in the suffix. For $i+2 \leq$

$l \leq j(i+1)$, we have a within the last $i+1$ positions of the prefix, and the last $i+1$ positions of the suffix are $c^i b$. So there is no border of length l. For $j(i+1)+1 \leq l \leq j(i+1)+k(i+2)$, we have one of the following three cases:

If our prefix ends in b, then we have $l = j(i+1)+m(i+2)+1$ for some integer m with $0 \leq m < k$. In this case, the a at position $l-1-m(i+2)-2(i+1)$ of the prefix will correspond with b at this position of the suffix. So there is no border of length l. If our prefix ends with $b\diamond^{i'}$ such that $1 \leq i' \leq i$, then our prefix contains the letter b within the last $i+1$ positions, but not at the last position. However, the suffix will have c's in all of these positions. If our prefix ends with $b\diamond^{i+1}$ so that we have $l = j(i+1)+m(i+2)$ where $1 \leq m \leq k$, then the a at position $l-m(i+2)-(i+1)$ of the prefix will correspond with b at this position of the suffix. So there is no border of length l.

Finally, consider the case where $j(i+1)+k(i+2)+1 \leq l \leq j(i+1)+k(i+2)+i$. We will have a prefix which ends with c and a suffix which ends with b, so we have no border for this length. □

Proposition 7. *For any integer $n > 9$, we have the lower bound $\hat{m}_3(n) \geq n - \lceil 2\sqrt{n+3} \rceil + 2$.*

Proof. Let $l = \lceil \sqrt{n} \rceil$, so that $l \geq 4$ and $(l-1)^2 < n \leq l^2$. To show that $\hat{m}_3(n) \geq n - \lceil 2\sqrt{n+3} \rceil + 2$ is equivalent to showing that

$$\hat{m}_3(n) \geq \begin{cases} n - 2\lceil \sqrt{n} \rceil + 1 & \text{if } l^2 - 2 \leq n \leq l^2 \\ n - 2\lceil \sqrt{n} \rceil + 2 & \text{if } l(l-1) - 2 \leq n \leq l^2 - 3 \\ n - 2\lceil \sqrt{n} \rceil + 3 & \text{if } (l-1)^2 + 1 \leq n \leq l(l-1) - 3 \end{cases}$$

We consider the following five cases, and in each case demonstrate that there exists an unbordered word of length n with the required number of holes. Note that in each of the cases we have $l = \lceil \sqrt{n} \rceil$.

First, if $(l-1)^2 + 1 \leq n \leq l(l-1) - 3$, then let $u = (a\diamond^i)^j (b\diamond^{i+1})^k c^i b$ where $i = l-2$, $j = (l-1)l - (n+1)$, and $k = n - (l-1)^2$. The length of u is $(i+1)(j+k+1)+k = (l-1)((l-1)l-(n+1)+n-(l-1)^2+1)+n-(l-1)^2 = n$. The number of holes in u is $ij+(i+1)k = (l-2)(l(l-1)-(n+1))+(l-1)(n-(l-1)^2) = n-2l+3$. By Proposition 6, to show u is unbordered it suffices to show $i, j \geq 2$. This case only holds for $l \geq 4$, so $i \geq 2$. Since $n \leq l(l-1) - 3$, we have $2 \leq l(l-1) - n - 1 = j$. Thus, there exists an unbordered word of length n with $n - 2\lceil \sqrt{n} \rceil + 3$ holes.

Second, if $l(l-1) - 2 \leq n \leq l(l-1)$, then let $u = (a\diamond^i)^j a\diamond^k c^i b$ where $i = l-1$, $j = l-3$, and $k = n-(l-1)^2$. The length of u is $(i+1)(j+1) + k + 1 = l(l-2) + n - l^2 + 2l - 1 + 1 = n$, and the number of holes in u is $ij + k = (l-1)(l-3) + n - l^2 + 2l - 1 = n - 2l + 2$. By Proposition 5, u is unbordered if $k \leq i$. Since $n \leq l(l-1)$, we have $n - l^2 + 2l - 1 = k \leq l - 1 = i$. Thus, there exists an unbordered word of length n with $n - 2\lceil \sqrt{n} \rceil + 2$ holes.

Third, if $l(l-1)+1 \leq n \leq l^2-4$, then let $u = (a\diamond^i)^j (b\diamond^{i+1})^k c^i b$ where $i = l-1$, $j = l^2 - 2 - n$, and $k = n - l(l-1)$. The length of u is $(i+1)(j+k+1)+k = l(l^2 - 2 - n + n - l^2 + l + 1) + n - l(l-1) = n$, and the number of holes in u is

$ij+(i+1)k=(l-1)(l^2-2-n)+l(n-l(l-1))=n-2l+2$. By Proposition 6, u is unbordered if $i,j\geq 2$. Since $n\geq 7$, it must be that $l\geq 3$, and so $i\geq 2$. Since $n\leq l^2-4$, we have $2\leq l^2-2-n=j$. Thus, there exists an unbordered word of length n with $n-2\lceil\sqrt{n}\rceil+2$ holes.

Fourth, if $n=l^2-3$, then let $u=(a\diamond^i)^2(b\diamond^{i+1})^k c^i b$ where $i=l-2$, $j=2$, and $k=l-3$. The length of u is $(i+1)(j+k+1)+k=(l-1)(2+l-3+1)+l-3=l^2-3=n$, and the number of holes in u is $ij+(i+1)k=(l-2)(2)+(l-1)(l-3)=l^2-2l-1=l^2-3-2l+2=n-2l+2$. By Proposition 6, u is unbordered since $j=2$ and $i\geq 2$, because $n\geq 7$ and $l\geq 4$. Thus, there exists an unbordered word of length n with $n-2\lceil\sqrt{n}\rceil+2$ holes.

Fifth, if $l^2-2\leq n\leq l^2$, then let $u=(a\diamond^i)^j a\diamond^k c^i b$ where $i=l-1$, $j=l-2$, and $k=n-l(l-1)-1$. The length of u is $(i+1)(j+1)+k+1=l(l-1)+n-l(l-1)-1+1=n$. The number of holes in u is $ij+k=(l-1)(l-2)+n-l(l-1)-1=n-2l+1$. By Proposition 5, u is unbordered if $k\leq i$. Since $n\leq l^2$, we have $n-l^2+l-1=k\leq l-1=i$. Thus, there exists an unbordered word of length n with $n-2\lceil\sqrt{n}\rceil+1$ holes. □

Note that our upper bound and lower bound for $\hat{m}_3(n)$ are equal for $n\leq 27$. We believe that our lower bound is tight and have the following conjecture.

Conjecture 1. The equality $\hat{m}_3(n)=n-\lceil 2\sqrt{n+3}\rceil+2$ holds for all $n\geq 6$.

3.3 A Lower Bound for $\hat{m}_4(n)$

Finally, we consider the 4-letter alphabet $\{a,b,c,d\}$. By letting $k=4$ in Proposition 2, we have the upper bound

$$\hat{m}_4(n)\leq\left\lfloor n-\sqrt{\frac{8}{3}(n-1)}\right\rfloor \tag{4}$$

for $n\geq 2$. We will give a lower bound for $\hat{m}_4(n)$.

Proposition 8. *The partial word $a\diamond^i(b\diamond^{i+1})^j c^i d$ is an unbordered word of length $(i+2)(j+1)+i$, for all $i,j\geq 0$ and distinct letters a,b,c,d.*

Proof. We assume that $i,j\geq 1$ (the other cases are similar). Consider a border length l. If $1\leq l\leq i+1$, then we have a prefix which begins with a and a suffix which begins with c or d. If $i+2\leq l<(j+1)(i+2)$, then the prefix ends in either $b\diamond^{i'}$ or $b\diamond^{i+1}$, where $0\leq i'\leq i$. If the prefix ends with $b\diamond^{i'}$, then we have the letter b appearing within the last $i+1$ positions of the prefix which will correspond with either c or d in the suffix. If the prefix ends with $b\diamond^{i+1}$, then our prefix will begin with a, and our suffix will begin with b. If $(j+1)(i+2)\leq l\leq 2i+1+j(i+2)$, then our prefix ends with the letter c while our suffix ends with the letter d. In each case, there is no border of length l. □

Proposition 9. *For integers $n \geq 7$, we have the lower bound*

$$\hat{m}_4(n) \geq \begin{cases} l(l-2) & \text{if } n = l^2 - 2 \text{ for some integer } l \\ l^2 - l - 1 & \text{if } n = l^2 + l - 2 \text{ for some integer } l \\ \hat{m}_3(n) & \text{otherwise} \end{cases}$$

Proof. First, suppose that $n = l^2 - 2$ for some integer l. Let $i = j = l - 2$. The word $a\diamond^i(b\diamond^{i+1})^j c^i d$ is unbordered by Proposition 8. The length of this word is $2i + 2 + j(i + 2) = 2(l - 2) + 2 + (l - 2)l = l^2 - 2 = n$. The number of holes in the word is $i + j(i + 1) = l - 2 + (l - 2)(l - 1) = l(l - 2)$.

Next, suppose that $n = l^2 + l - 2$ for some integer l. Now let $i = l - 2, j = l - 1$. We have the word $a\diamond^i(b\diamond^{i+1})^j c^i d$, which is unbordered by Proposition 8. The length of this word is $2i + 2 + j(i + 2) = 2(l - 2) + 2 + (l - 1)l = l^2 + l - 2 = n$. The number of holes in this word is $i + j(i + 1) = l - 2 + (l - 1)(l - 1) = l^2 - l - 1$.

For all other n, consider an unbordered word with $\hat{m}_3(n)$ holes, which is still unbordered over an alphabet of size 4. $\square$

Note that our lower bound can be improved when $n = 24, 35, 48, 63, 80$ and 99. For instance, $\hat{m}_4(24) = 16 > \hat{m}_3(24) = 15$.

$\hat{m}_4(24) = 16$	$a\diamond^1 b\diamond^2 c\diamond^2 c\diamond^3 b\diamond^8 add$
$\hat{m}_4(35) = 25$	$a\diamond^2 a\diamond^2 b\diamond^3 c\diamond^3 c\diamond^4 b\diamond^{11} addd$
	$a\diamond^2 d\diamond^2 d\diamond^3 b\diamond^3 b\diamond^4 a\diamond^{11} cccd$
$\hat{m}_4(48) = 36$	$a\diamond^1 b\diamond^2 b\diamond^2 c\diamond^3 c\diamond^3 c\diamond^3 a\diamond^4 a\diamond^{18} bddd$
	$a\diamond^1 d\diamond^2 a\diamond^2 a\diamond^3 b\diamond^3 b\diamond^3 b\diamond^4 c\diamond^{18} dcdd$
	$a\diamond^1 d\diamond^2 d\diamond^2 b\diamond^3 b\diamond^3 b\diamond^3 a\diamond^4 a\diamond^{18} cccd$
$\hat{m}_4(63) \geq 49$	$a\diamond^2 a\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^4 c\diamond^4 a\diamond^5 c\diamond^{22} dcddd$
	$a\diamond^2 a\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 c\diamond^4 c\diamond^4 a\diamond^5 c\diamond^{22} bdddd$
	$a\diamond^2 a\diamond^3 d\diamond^3 a\diamond^3 d\diamond^3 d\diamond^4 b\diamond^4 b\diamond^5 a\diamond^{22} ccccd$
$\hat{m}_4(80) \geq 64$	$a\diamond^1 b\diamond^2 b\diamond^2 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 a\diamond^4 a\diamond^{34} bddd$
	$a\diamond^1 d\diamond^2 a\diamond^2 a\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^4 c\diamond^{34} dcdd$
	$a\diamond^1 d\diamond^2 d\diamond^2 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 a\diamond^4 a\diamond^{34} cccd$
$\hat{m}_4(99) \geq 81$	$a\diamond^2 a\diamond^2 b\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^3 c\diamond^4 b\diamond^{43} addd$
	$a\diamond^2 d\diamond^2 d\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^3 b\diamond^4 a\diamond^{43} cccd$

This leads us to the following conjecture.

Conjecture 2. The equality $\hat{m}_4(l^2 - 1) = (l - 1)^2$ holds for all $l > 2$.

4 Conclusion

The following conjecture is somehow natural, since increasing the length of a partial word by one is possible through the addition of at most one hole.

Conjecture 3. The inequalities $\hat{m}_k(n) \leq \hat{m}_k(n + 1) \leq \hat{m}_k(n) + 1$ hold for all $k \geq 2, n \geq 1$.

Proposition 10. *If Conjecture 3 holds, then for all $k \geq 3$ and $n \geq 2$:*

$$\hat{m}_k(n) \leq \min(\left\lfloor n - \sqrt{\frac{2k}{k-1}(n-1)} \right\rfloor, n + 1 - 2\lfloor\sqrt{n}\rfloor) \quad (5)$$

Proof. Using Proposition 2 we have the first bound. Now, from Lemma 2 and Theorem 2 we have that $\hat{m}_k(\lfloor\sqrt{n}\rfloor^2) = (\lfloor\sqrt{n}\rfloor - 1)^2$. Since Conjecture 3 holds, it follows that $\hat{m}_k(n) \leq \hat{m}_k(\lfloor\sqrt{n}\rfloor^2) + n - \lfloor\sqrt{n}\rfloor^2 = (\lfloor\sqrt{n}\rfloor - 1)^2 + n - \lfloor\sqrt{n}\rfloor^2 = n + 1 - 2\lfloor\sqrt{n}\rfloor$. □

Furthermore, we notice that for $k = 4$ and $n \geq 64$, the first bound in (5) is always greater than the second one. This implies that for $n \geq 64$, $\hat{m}_4(n)$ is bounded by $n + 1 - 2\lfloor\sqrt{n}\rfloor)$. Moreover, if Conjecture 3 is true, then this bound should be increasing. It follows that the upper bound for $\hat{m}_4(l^2)$ equals the upper bound for $\hat{m}_4(l^2 - 1)$, for any $l > 2$, and, if Conjecture 2 holds, the bound is sharp in these cases.

Let us denote by $UB(n)$ the upper bound (4) if Conjecture 3 does not hold, and the upper bound $\min(\left\lfloor n - \sqrt{\frac{8}{3}(n-1)} \right\rfloor, n + 1 - 2\lfloor\sqrt{n}\rfloor)$ otherwise, and by $LB(n)$ the lower bound of Proposition 9. Computer obtained data on all unbordered words of length $n \leq 1,000,000$ shows that $LB(n)$ differs from $UB(n)$ by at most 1 when Conjecture 3 holds. For instance, if $n \leq 10,000$, $LB(n)$ and $UB(n)$ differ for $5,131 + 93$ lengths (the 93 disappearing when Conjecture 2 holds). Note that the percentage of mismatches between $LB(n)$ and $UB(n)$ is actually decreasing, the more words we consider.

	$N = 10,000$	$N = 100,000$	$N = 1,000,000$
$\max_{n=1}^{N}(UB(n) - LB(n)) =$	38	114	366
If Conjecture 3 holds, then $\sum_{n=1}^{N}(UB(n) - LB(n)) =$	5,224	50,692	502,474
If Conjectures 2 and 3 hold, then $\sum_{n=1}^{N}(UB(n) - LB(n)) =$	5,131	50,383	501,481

Conjecture 4. The equality $\hat{m}_k(n) = \hat{m}_4(n)$ holds for all $k \geq 4, n \geq 1$.

References

1. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. Theoret. Comput. Sci. **218** (1999) 135–141
2. Blanchet-Sadri, F.: Primitive partial words. Discrete Appl. Math. **148** (2005) 195–213
3. Blanchet-Sadri, F.: Algorithmic Combinatorics on Partial Words. Chapman & Hall/CRC Press (2007)
4. Blanchet-Sadri, F.: Open Problems on Partial Words. In Bel-Enguix, G., Jiménez-López, M.D., Martín-Vide, C. (eds.): New Developments in Formal Languages and Applications. Ch. 2, Vol. 3. Springer-Verlag, Berlin (2007) 11–58
5. Blanchet-Sadri, F., Davis, C., Dodge, J., Mercaş, R., Moorefield, M.: Unbordered partial words. Discrete Appl. Math. (to appear doi:10.1016/j.dam.2008.04.004)