Counting Bordered Partial Words by Critical Positions*

Emily Allen[1]   F. Blanchet-Sadri[2]   Cameron Byrum[3]

Mihai Cucuringu[4]   Robert Mercaş[5]

June 8, 2011

This paper is dedicated to Professor Pál Dömösi
on the occasion of his 65th birthday.

Abstract

A *partial word*, sequence over a finite alphabet that may have some
undefined positions or holes, is *bordered* if one of its proper prefixes is
compatible with one of its suffixes. The number theoretical problem
of enumerating all bordered full words (the ones without holes) of a
fixed length n over an alphabet of a fixed size k is well known. In this
paper, we investigate the number of bordered partial words having h
holes with the parameters k, n. It turns out that all borders of a full
word are *simple*, and so every bordered full word has a unique minimal

[1]Department of Mathematical Sciences, Carnegie Mellon University, 5032 Forbes Ave.,
Pittsburgh, PA 15289, USA

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

[3]Department of Mathematics, University of Mississippi, P.O. Box 1848, University, MS
38677, USA

[4]Department of Mathematics, Princeton University, Washington Road, Princeton, NJ
08544–1000, USA

[5]GRLMC, Universitat Rovira i Virgili, Campus Catalunya, Departament de Filologies
Romàniques, Av. Catalunya, 35, Tarragona, 43002, Spain

1

border no longer than half its length. Counting bordered partial words is made extremely more difficult by the failure of that combinatorial property since there is now the possibility of a minimal border that is *nonsimple*. Here, we give recursive formulas based on our approach of the so-called *simple and nonsimple critical positions*.

Keywords: Theory of formal languages; Number theory; Combinatorics on words; Partial words; Bordered partial words; Simple border; Simply bordered partial words; Critical positions.

AMS Subject Classification: 68R15; 68Q45; 05A05

1   Introduction

The two fundamental concepts of *primitive words* and *bordered words* are highly connected in areas including coding theory, combinatorics on words, formal language theory, and text algorithms [8, 9, 12, 16]. A primitive word is a sequence that cannot be written as a power of another sequence, while a bordered word is a sequence such that at least one of its nonempty proper prefixes is one of its suffixes. For example, $\underline{ab}a\underline{ab}$ is bordered with border ab while $abaabb$ is unbordered. The numbers of primitive and bordered words of a fixed length over an alphabet of a fixed size are well known, the number of primitive words being related to the Möbius function [12]. In 1999, being motivated by a practical problem on gene comparison, Berstel and Boasson used the terminoloy of *partial words* for sequences over a finite alphabet that may have some "do not know" symbols or "holes" denoted by $\diamond$'s [2]. For instance, $a\diamond bca\diamond b$ is a partial word with two holes over the three-letter alphabet $\{a, b, c\}$. Actually in 1974, partial words were introduced as strings with don't-cares by Fischer and Paterson in [10]. Partial words are a special case of what are variously called "generalized" or "indeterminate" or "degenerate" strings, which were first discussed in 1987 by Abrahamson in [1] and which have been studied by several authors since 2003. Combinatorial properties of partial words have been investigated, and connections have been made, in particular, with problems concerning primitive sets of integers, lattices, vertex connectivity in graphs, etc [4].

Primitive partial words were introduced by Blanchet-Sadri in 2005 [3]. Testing whether or not a partial word is primitive can be done in a way similar to that of words [6]. The problem of counting primitive partial words with h holes of length n over a k-size alphabet was initiated in [4]. There, formulas for $h = 1$ and $h = 2$ were given through a constructive approach, and some bounds were also provided for $h > 2$. *Bordered partial words* were also introduced in [3], two types of borders being identified:

2

simple and *nonsimple*. A partial word is called *unbordered* if it does not have any border. For the finite alphabet $\{a, b, c\}$, the partial word $a\diamond\diamond b$ has both a simple border ab and a nonsimple border aab, the first one being minimal, while the partial word $a\diamond bc$ is unbordered (the $\diamond$ symbol represents an undefined position or a "hole," and matches every letter of the alphabet).

In this paper, we investigate the problem of enumerating all bordered partial words with h holes of length n over a k-letter alphabet, a problem that yields some recursive formulas. It turns out that every bordered full word (one without holes) of length n has a unique minimum-length border no longer than $\lfloor \frac{n}{2} \rfloor$. When we allow words to have holes (even when we allow only one hole), counting bordered partial words is made extremely more difficult by the failure of that combinatorial property since there is now the possibility of a minimal border that is nonsimple as in $a\diamond bb$. Thus, we will restrict our attention almost exclusively to partial words with one hole. Note that several counting problems for partial words have been proved to be "hard" by Manea and Tiseanu in [13].

The contents of our paper are as follows: In Section 2, we define the notion of bordered partial words and discuss some of their properties. The *simply bordered* partial words are introduced (partial words that have a minimal border that is simple), and there we also count the number of bordered full words (every bordered full word turns out to be simply bordered). In Section 3, we give a formula for the number of simply bordered partial words of length n with h holes over a k-letter alphabet. Our approach is a recursive one, dependent only on the number of "perfect squares" (bordered partial words of even length, that have a minimal border length equal to exactly half their length). In Section 4, we introduce the notion of *critical positions* that once their letters are changed into holes create borders, and investigate the number of bordered partial words with the parameters h, k, n and with respect to these positions. Using the two distinct border notions, depending on the type of border created, the critical positions are divided into "simple critical positions" and "nonsimple critical positions." Under these conditions, the previously defined concept of perfect squares can be expressed in terms of the critical positions. Using independent recursive formulas, we compute the exact number of simple and nonsimple critical positions, and in Section 5, we achieve our main goal of calculating the number of bordered partial words of length n with one hole over an alphabet of size k answering an open problem of [5].

We first review basic concepts on words and partial words. Let A be a nonempty finite set called an *alphabet*. Elements of A are called *letters* and finite sequences of letters from A are called *(full) words* over A. A *partial*

word over A is a sequence of symbols from the alphabet A enlarged with the *hole* symbol, denoted by $\diamond$, that is a sequence of symbols from $A_\diamond = A \cup \{\diamond\}$. Note that every full word is also a partial word. The set of all full words over A is denoted by A^*, the set of all partial words over A by $A_\diamond^*$. The empty word is denoted by ε. We denote by $|u|$ the length of a full or a partial word u (the length of the empty word is 0). We say that position i in u is part of the *domain* of u, denoted by $i \in D(u)$, if the symbol at position i, denoted by $u(i)$, is from A, and i belongs to the *set of holes* of u, denoted by $i \in H(u)$, otherwise. A word over A is a partial word over A with an empty set of holes. The labelling of the positions of a partial word start at 0.

If u and v are two partial words of equal length, then u is said to be contained in v, denoted by $u \subset v$, if $u(i) = v(i)$ for all $i \in D(u)$. The partial words u and v are called *compatible*, denoted by $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$, in which case we denote by $u \vee v$ the least upper bound of u and v. For example, $u = aba\diamond a$ and $v = a\diamond\diamond ba$ are compatible, and $(u \vee v) = ababa$.

A *(strong) period* of a partial word u over A is a positive integer p such that $u(i) = u(j)$ whenever $i, j \in D(u)$ and $i \equiv j \bmod p$. In such a case, we call u *(strongly) p-periodic*. Similarly, a *weak period* of u is a positive integer p such that $u(i) = u(i+p)$ whenever $i, i+p \in D(u)$. In such a case, we call u *weakly p-periodic*. The partial word $abb\diamond bbcbb$ is weakly 3-periodic but is not strongly 3-periodic. The latter shows a difference between partial words and full words since every weakly p-periodic full word is strongly p-periodic. Another difference worth noting is the fact that even if the length of a partial word u is a multiple of a weak period of u, then u is not necessarily a power of a shorter partial word. A partial word u is *nonperiodic* if it is not p-periodic for any positive integer p, $p < |u|$.

2   Bordered partial words

A nonempty partial word u is *unbordered* if no nonempty partial words x_1, x_2, v, w exist such that $u = x_1 v = w x_2$ and $x_1 \uparrow x_2$. If such nonempty words exist, then x exists such that $x_1 \subset x$ and $x_2 \subset x$ and we call u *bordered* and call x a *border* of u. It is easy to see that if u is unbordered and $u \subset u'$, then u' is unbordered as well. A border x of u is called *minimum* if $|x| > |y|$ implies that y is not a border of u.

Note that there are two types of borders. Writing u as $x_1 v = w x_2$ where $x_1 \subset x$ and $x_2 \subset x$, we say that x is an *overlapping (nonsimple)* border if

$|x| > |v|$, and a *nonoverlapping (simple)* border otherwise. The partial word $u = a \diamond \diamond ab$ is bordered with the simple border ab and nonsimple border aab, the first one being minimal, while the partial word $ab \diamond c$ is unbordered. We have that 2 is a simple border length of $u = a \diamond \diamond ab$ and 3 is a nonsimple border length of u. Here the minimal border length, which is 2, is simple.

Proposition 1. *Let u be a partial word.*

1. *If $0 < l < |u|$, then u has a border of length l if and only if u has weak period $|u| - l$.*

2. *If $0 < l \leq \lfloor \frac{|u|}{2} \rfloor$, then u has a border of length l if and only if u has strong period $|u| - l$.*

Proof. Set $|u| = n$. For the backward implication of Statement 1, assume that u has weak period $n - l$. Write u as $u = u_1 u_2 \ldots u_d x$ where for $1 \leq i \leq d$, $|u_i| = n - l$, $|x| = j = n \bmod (n - l)$ and $d = \lfloor \frac{n}{n-l} \rfloor$. Note that $u_i \uparrow u_{i+1}$ for all $1 \leq i < d$ and x must be compatible with the prefix y of u_d of length j. We may also assume that x is not empty since otherwise u would have the border $(u_1 u_2 \ldots u_{d-1}) \vee (u_2 u_3 \ldots u_d)$ of length l.

We begin with the case when $d = 1$, that is, $u = u_1 x$. Obviously, $x \vee y$ is a border of length l because x is compatible with y. Now suppose $d \geq 2$. In this case, a border of length l is a nonsimple border. Let v, w be the prefix and suffix of u of length l, that is, $v = u_1 u_2 \ldots u_{d-1} y$ and $w = u_2 u_3 \ldots u_d x$:

$$
\begin{array}{ccccccc}
u & = & u_1 & u_2 & u_3 & \ldots & u_d & x \\
& & u_1 & u_2 & \ldots & u_{d-1} & y
\end{array}
$$

Since $u_i \uparrow u_{i+1}$ for $1 \leq i < d$ and $x \uparrow y$, it follows that $v \uparrow w$.

For the forward implication, it is easy to see that if u has a border of length $l \leq \lfloor \frac{n}{2} \rfloor$, then u has strong period $n - l$, and thus weak period $n - l$. If $\lfloor \frac{n}{2} \rfloor < l < n$, then the result follows similarly as above.

Note that for Statement 2, since any strong period is a weak period, we have that if u has strong period $n - l$, then u has a border of length l. $\square$

Note that the partial word $u = aaa \diamond aba$ has a border of length 5 but is not strongly 2-periodic. Hence the bound on l in Statement 2 of Proposition 1.

We call a bordered partial word u *simply bordered* if a minimal border x exists satisfying $|u| \geq 2|x|$.

Proposition 2 ([7]). *Let u be a nonempty bordered partial word. Let x be a minimal border of u, and set $u = x_1 v = w x_2$ where $x_1 \subset x$ and $x_2 \subset x$. Then the following hold:*

5

1. *The partial word x is unbordered.*

2. *If x_1 is unbordered, then $u = x_1 u' x_2 \subset x u' x$ for some u'.*

Note that Proposition 2 implies that if u is a full bordered word, then $x_1 = x$ is unbordered. In this case, $u = x u' x$ where x is the minimal border of u. Hence a bordered full word is always simply bordered.

Note that because borderedness in partial words is defined via containment, it does not make sense to talk about *the* minimal border of a partial word, there could be many possible borders of a certain length.

We will denote by $U_{h,k}(n)$ (respectively, $B_{h,k}(n)$) the number of unbordered (respectively, bordered) partial words of length n with h holes over a k-letter alphabet. Clearly,

$$B_{h,k}(n) = \binom{n}{h} k^{n-h} - U_{h,k}(n), n \geq h$$

The problem of enumerating all unbordered full words of length n over a k-letter alphabet yields to a conceptually simple recursive formula [11, 17, 15]: $U_{0,k}(0) = 1$, $U_{0,k}(1) = k$, and for $n > 0$,

$$U_{0,k}(2n) = kU_{0,k}(2n-1) - U_{0,k}(n)$$

$$U_{0,k}(2n+1) = kU_{0,k}(2n)$$

These equalities can be seen from the fact that if a word has odd length $2n + 1$, then it is unbordered if and only if it is unbordered after removing the middle letter. If a word has even length $2n$, then it is unbordered if and only if it is obtained from an unbordered word of length $2n - 1$ by adding a letter next to the middle position unless doing so creates a word that is a perfect square.

Using these formulas and Proposition 2, we can easily obtain a formula for counting bordered full words. The number of full words of length n over a k-letter alphabet that have a minimal border of length l is

$$U_{0,k}(l)k^{n-2l}$$

Then we have that

$$B_{0,k}(n) = \sum_{l=1}^{\lfloor \frac{n}{2} \rfloor} U_{0,k}(l)k^{n-2l}$$

When we allow words to have holes, counting bordered partial words is made more difficult since they are not necessarily simply bordered.

We end this section with two propositions that give properties for bordered partial words with one hole that will be useful in the sequel.

Proposition 3. *Let u be a partial word with one hole that has a minimal border that is nonsimple, and let x, y denote a prefix and a suffix of u such that $x \uparrow y$. Then both $\|H(x)\| = \|H(y)\| = 1$.*

Proof. Let x, y denote a prefix and a suffix of u such that $x \uparrow y$ with the length of x being minimal. Assume now that y is a full word and that x contains $\diamond$. We let $x = x_1 x_2$ and $y = y_1 y_2 = x_2 y'$ such that $x_1 \uparrow y_1$ and $x_2 \uparrow y_2$ and $\|H(x_1)\| = 1$. Because x_2, y_2 are full words it follows that $x_2 = y_2$. We can now write x as $x = x_2' y''$ where $x_2 \uparrow x_2'$ and $y' \uparrow y''$. This implies that u has a prefix x_2' and a suffix y_2 that are compatible, that is, u has a border shorter than $|x|$, which implies a contradiction. $\square$

Proposition 4. *Let u be a nonperiodic bordered partial word with one hole. There exists a unique integer l with $\lceil \frac{n}{2} \rceil \le l < n$ such that u has a border of length l.*

Proof. First, a result of [4] states that if x, y and z are partial words such that $|x| = |y| > 0$, then $xz \uparrow zy$ if and only if xzy is weakly $|x|$-periodic. Second, a result of [2] states that if a partial word x with one hole is weakly p-periodic and weakly q-periodic and $|x| \ge p+q$, then x is $\gcd(p,q)$-periodic. Now, let us assume that there exist more than one border of u of length at least $\lceil \frac{n}{2} \rceil$. Hence, we can write $u = x_1 y_1 z y_2 x_2$ with $x_1 y_1 z \uparrow z y_2 x_2$ and $x_1 y_1 z y_2 \uparrow y_1 z y_2 x_2$, where $|x_1| = |x_2|$ and $|y_1| = |y_2|$. Here $l = |x_1 y_1 z|$ and $l + |y_2|$ are lengths of the borders. From $x_1 y_1 z \uparrow z y_2 x_2$, we get that u is weakly $|x_1 y_1|$-periodic. Also, from $x_1 y_1 z y_2 \uparrow y_1 z y_2 x_2$ we get that u is weakly $|x_1|$-periodic. Since u is weakly $|x_1|$-periodic and weakly $|x_1 y_1|$-periodic, and $|u| \ge |x_1| + |x_1 y_1|$, we have that u is $\gcd(|x_1|, |x_1 y_1|)$-periodic, a contradiction with the fact that u is nonperiodic. $\square$

3 Counting simply bordered partial words

When counting bordered partial words, we cannot assume that the length of a minimal border x satisfies $|x| \le \lfloor \frac{n}{2} \rfloor$, as there is now the possibility that a partial word has a minimal nonsimple border. The bordered partial words where this inequality is satisfied are the simply bordered ones.

Let $S_{h,k}(n)$ be the number of simply bordered partial words of length n with h holes over a k-letter alphabet, and let $\mathcal{S}_{h,k}(n)$ be the set of such partial words. Clearly if $h > n$, then $S_{h,k}(n) = 0$. Note that $S_{0,k}(n) = B_{0,k}(n)$. In this section we give a formula for $S_{h,k}(n)$.

The case when n is odd is easy to deal with. We can obtain all simply bordered partial words of odd length just by inserting a letter or a $\diamond$

in the middle position of a simply bordered word of even length. If the inserted symbol is a letter, then $kS_{h,k}(n-1)$ distinct words in $\mathcal{S}_{h,k}(n)$ can be generated. The case when we insert a $\diamond$ produces $S_{h-1,k}(n-1)$ words. Thus,

$$S_{h,k}(n) = kS_{h,k}(n-1) + S_{h-1,k}(n-1), \; n \text{ odd}$$

A similar argument gives a recurrence relation for the number of partial words that are not simply bordered. Let $N_{h,k}(n)$ be the number of partial words with h holes, of length n, over a k-letter alphabet that are not simply bordered. Obviously we can find the value of this function by subtracting the value of $S_{h,k}(n)$ from the total number of partial words with those parameters (that is, $N_{h,k}(n) = \binom{n}{h}k^{n-h} - S_{h,k}(n)$). We have

$$N_{0,k}(n) = U_{0,k}(n)$$

since a full word that is not simply bordered is an unbordered full word. It is easy to see that $N_{1,k}(0) = 0$, $N_{1,k}(1) = 1$, $N_{1,k}(2) = 0$, and for $h > 1$ that $N_{h,k}(1) = 0$ and $N_{h,k}(2) = 0$. Now, for $h > 0$, the following formula holds:

$$N_{h,k}(n) = kN_{h,k}(n-1) + N_{h-1,k}(n-1), \; n \text{ odd}$$

Although the approach for the case when $n = 2m$ is similar, it yields a more complicated formula. We construct the words in $\mathcal{S}_{h,k}(2m)$ by inserting two symbols of $A \cup \{\diamond\}$ into simply bordered partial words of length $2m-2$ with h, $h-1$ or $h-2$ holes.

We write $w \rightarrow w'$, if $w' = a_0 a_1 \ldots a_{2m-1}$, $w = b_0 b_1 \ldots b_{2m-3}$, $a_i = b_i$ for $i \in [0..m-2]$, and $a_{i+2} = b_i$ for $i \in [m-1..2m-3]$:

$$
\begin{array}{ccccccccccc}
w' & = & a_0 & a_1 & \ldots & a_{m-2} & a_{m-1} & a_m & a_{m+1} & a_{m+2} & \ldots & a_{2m-1} \\
w & = & b_0 & b_1 & \ldots & b_{m-2} & & & b_{m-1} & b_m & \ldots & b_{2m-3}
\end{array}
$$

We denote by $\mathcal{W}_{h,k}(2m)$ the set of partial words w' of length $2m$ with h holes over a k-letter alphabet A such that, for some $w \in \mathcal{S}_{h',k}(2m-2)$ we have $w \rightarrow w'$, where $h' \in \{h, h-1, h-2\}$, and by $W_{h,k}(2m)$ the cardinality of $\mathcal{W}_{h,k}(2m)$.

We analyze three cases, depending on whether a_{m-1} and a_m are letters of A or $\diamond$'s.

Case 1. $a_{m-1} \in A$ and $a_m \in A$

Since a_{m-1} and a_m can be any letters, this case creates $k^2 S_{h,k}(2m-2)$ new words in $\mathcal{S}_{h,k}(2m)$.

Case 2. $a_{m-1} = \diamond$ and $a_m = \diamond$

Since w has $h-2$ holes, this case yields $S_{h-2,k}(2m-2)$ words.

Case 3. $a_{m-1} \in A$ and $a_m = \diamond$, or $a_{m-1} = \diamond$ and $a_m \in A$

The two cases are identical and both create $kS_{h-1,k}(2m-2)$ words in $\mathcal{S}_{h,k}(2m)$ since we can pick any letter in the alphabet and w has $h-1$ holes. This gives us a total of

$$W_{h,k}(2m) = k^2 S_{h,k}(2m-2) + 2kS_{h-1,k}(2m-2) + S_{h-2,k}(2m-2)$$

Now, let $S_{h,k}(n,l)$ (respectively, $S'_{h,k}(n,l)$) represent the number of partial words with h holes of length n over a k-letter alphabet that have a border of length l (respectively, a minimal border of length l), and by $\mathcal{S}_{h,k}(n,l)$ and $\mathcal{S}'_{h,k}(n,l)$ respectively the sets of such words.

Proposition 5. *The following equality holds for $h \geq 2$, $k \geq 2$ and $m \geq 1$:*

$$W_{h,k}(2m) = \sum_{l=1}^{m-1} S'_{h,k}(2m,l)$$

In other words, $W_{h,k}(2m)$ represents the total number of simply bordered partial words with h holes of length $2m$ over a k-letter alphabet, which have a minimal border x where $1 \leq |x| < m$.

Proof. Pick $w \in \mathcal{W}_{h,k}(2m)$. From the definition of $\mathcal{W}_{h,k}(2m)$, it follows that there exists $u \in \mathcal{S}_{h',k}(2m-2)$ for some $h' \in \{h, h-1, h-2\}$ such that $u \to w$. Since $u \in \mathcal{S}_{h',k}(2m-2)$, there must exist $l < m$ such that u has a minimal border of length equal to l. Thus, $w \in \mathcal{S}'_{h,k}(2m,l)$ and obviously $w \in \bigcup_{l=1}^{m-1} \mathcal{S}'_{h,k}(2m,l)$.

Now pick $w \in \bigcup_{l=1}^{m-1} \mathcal{S}'_{h,k}(2m,l)$. Say w has a minimal border of length $l < m$. If we take out the two middle positions in w and denote the resulting word by u, we have that $u \to w$ and $u \in \mathcal{S}_{h',k}(2m-2)$ for some $h' \in \{h, h-1, h-2\}$. Furthermore, since $l < m$, the two middle positions we eliminated from w do not affect the length of a minimal border of u, which will still have length l. Since $u \to w$ it follows that $w \in \mathcal{W}_{h,k}(2m)$.

It is obvious that for $l \neq l'$ it holds that $\mathcal{S}'_{h,k}(2m,l) \cap \mathcal{S}'_{h,k}(2m,l') = \emptyset$, since all words in the former set have a minimal border of length l while the latter has only words with minimal border of length l'. Note that unbordered partial words of length $2m-2$ cannot create any bordered partial words with a border of length less than or equal to $m-1$.

We may now say that

$$W_{h,k}(2m) = \| \bigcup_{l=1}^{m-1} \mathcal{S}'_{h,k}(2m,l)\| = \sum_{l=1}^{m-1} S'_{h,k}(2m,l)$$

which concludes our proof. $\qquad\square$

Corollary 1. *The following equality holds for $h \geq 2$, $k \geq 2$ and $m \geq 1$:*

$$S_{h,k}(2m) = k^2 S_{h,k}(2m-2) + 2k S_{h-1,k}(2m-2) + S_{h-2,k}(2m-2) + S'_{h,k}(2m, m)$$

Corollary 2. *The following equality holds for $h \geq 2$, $k \geq 2$ and $m \geq 1$:*

$$N_{h,k}(2m) = k^2 N_{h,k}(2m-2) + 2k N_{h-1,k}(2m-2) + N_{h-2,k}(2m-2) - S'_{h,k}(2m, m)$$

Proof. Note that $N_{h,k}(2m) = \binom{2m}{h} k^{2m-h} - S_{h,k}(2m)$. The result easily follows from Corollary 1 and the fact that $\binom{2m}{h} = \binom{2m-2}{h-2} + 2\binom{2m-2}{h-1} + \binom{2m-2}{h}$. $\qquad\square$

In the next section, we will express $S'_{h,k}(2m, m)$ in terms of "critical pairs."

4 Counting bordered partial words by critical pairs

In this section, we count the simple critical pairs determined by a word of length n, as well as the nonsimple ones. Most of the recurrences obtained are for full words, since our goal, see Section 5, is to calculate the number of bordered partial words with one hole by critical pairs.

We start with a definition.

Definition 1. *A partial word u is said to* generate *a partial word v if $v \subset u$ and $\|H(u)\| + 1 = \|H(v)\|$. For the unique i with $u(i) \in A$ and $v(i) = \diamond$, we say v is generated by redefining Position i or letter $u(i)$ in u.*

For example, the word *ababb* generates $\diamond babb$, $a\diamond abb$, $ab\diamond bb$, $aba\diamond b$, $abab\diamond$ by replacing a letter with a hole. Only $a\diamond abb$ is unbordered. Furthermore, $a\diamond abb$ can be obtained as well from *aaabb*.

For an alphabet of size k and a partial word v with h holes, there are hk words that generate v. It is easy to see that if v is unbordered and is generated by a partial word u, then u is unbordered as well. Hence, every unbordered word over a k-letter alphabet with h holes can be generated by hk unbordered words with $h-1$ holes each.

Any given partial word of length n with $h-1$ holes can generate $n-h+1$ partial words each with h holes. Given an unbordered word of length n with $h-1$ holes, we wish to determine how many of the $n-h+1$ words that it generates will also be unbordered. Note that the words generated by an unbordered word may be bordered. Thus, to find $U_{h,k}(n)$, it suffices to count the total number of unbordered words with h holes generated by each of the unbordered words with $h-1$ holes, and divide by hk.

Definition 2. *Given a partial word u and Position i, $0 \leq i < |u|$, we say that the pair (u,i) is a* critical pair *for the border length l if u does not have a border of length l, but the word generated by redefining Position i in u has a border of length l. We say (u,i) is a* simple critical pair *if it is a critical pair for a simple border length, and a* nonsimple critical pair *if it is a critical pair for a nonsimple border length and is not a critical pair for any simple border length.*

For example, consider the word $u = abaababb$. We have that $(u,0)$, $(u,2)$, $(u,6)$, $(u,7)$ are simple critical pairs and $(u,3)$ is a nonsimple critical pair. For the rest of the paper, whenever we fix a word u, we will refer to $u(i)$ as the *critical letter* and i as the *critical position* of the critical pair (u,i).

Proposition 6. *Let u be a partial word with prefix x of length l and suffix y of length l, where $l < \left\lceil \frac{|u|}{2} \right\rceil$.*

- *A pair (u,i), with i in the prefix x, is a simple critical pair for the border length l if and only if $x = x_1 u(i) x_2$ and $y = y_1 u(n-i-1) y_2$ where $x_1 \uparrow y_1$, $x_2 \uparrow y_2$, and $u(i) \not\uparrow u(n-i-1)$.*

- *A pair (u,i), with i in the suffix y, is a simple critical pair for the border length l if and only if $x = x_1 u(n-i-1) x_2$ and $y = y_1 u(i) y_2$ where $x_1 \uparrow y_1$, $x_2 \uparrow y_2$, and $u(i) \not\uparrow u(n-i-1)$.*

Note that we allow x_1, x_2, y_1 and y_2 to be empty.

Definition 3. *Let $\mathcal{C}_{h,k}(n,l)$ be the set of pairs (u,i) such that u is an unbordered partial word of length n over a k-letter alphabet with h holes, where (u,i) is a critical pair for the border length l but (u,i) is not a critical pair for any border length less than l. Denote by $C_{h,k}(n,l)$ the cardinality of $\mathcal{C}_{h,k}(n,l)$. Let*

$$\mathcal{C}_{h,k}(n) = \bigcup_{l=1}^{n-1} \mathcal{C}_{h,k}(n,l) \ \text{ and } \ C_{h,k}(n) = \sum_{l=1}^{n-1} C_{h,k}(n,l)$$

Note that, if u is an unbordered word of length n over an alphabet of size k with $h-1$ holes, and v is the word generated by redefining Position i in u, then v is bordered if and only if $(u,i) \in \mathcal{C}_{h-1,k}(n)$.

For example, when $n = 5$, we have $U_{0,2}(5) = 12$ and $U_{1,2}(5) = 4$. The following are the unbordered words which begin with an a:

Unbordered word u with no hole	Pairs which are not critical	Generated word v with one hole
$u_1 = aaaab$	none	none
$u_2 = aaabb$	$(u_2, 1)$	$a \diamond abb$
$u_3 = aabbb$	$(u_3, 3)$	$aab \diamond b$
$u_4 = aabab$	$(u_4, 3)$	$aab \diamond b$
$u_5 = abbbb$	none	none
$u_6 = ababb$	$(u_6, 1)$	$a \diamond abb$

Here,

$$
\begin{aligned}
\mathcal{C}_{0,2}(5,1) &= \{(u_1, 0), (u_2, 0), (u_3, 0), (u_4, 0), (u_5, 0), (u_6, 0), \\
&\quad (u_1, 4), (u_2, 4), (u_3, 4), (u_4, 4), (u_5, 4), (u_6, 4), \\
&\quad (\overline{u_1}, 0), (\overline{u_2}, 0), (\overline{u_3}, 0), (\overline{u_4}, 0), (\overline{u_5}, 0), (\overline{u_6}, 0), \\
&\quad (\overline{u_1}, 4), (\overline{u_2}, 4), (\overline{u_3}, 4), (\overline{u_4}, 4), (\overline{u_5}, 4), (\overline{u_6}, 4)\} \\
\mathcal{C}_{0,2}(5,2) &= \{(u_1, 1), (u_4, 1), (u_5, 3), (u_6, 3), \\
&\quad (\overline{u_1}, 1), (\overline{u_4}, 1), (\overline{u_5}, 3), (\overline{u_6}, 3)\} \\
\mathcal{C}_{0,2}(5,3) &= \{(u_1, 2), (u_4, 2), (u_5, 2), (u_6, 2), \\
&\quad (\overline{u_1}, 2), (\overline{u_4}, 2), (\overline{u_5}, 2), (\overline{u_6}, 2)\} \\
\mathcal{C}_{0,2}(5,4) &= \{(u_1, 3), (u_2, 2), (u_2, 3), (u_3, 1), (u_3, 2), (u_5, 1), \\
&\quad (\overline{u_1}, 3), (\overline{u_2}, 2), (\overline{u_2}, 3), (\overline{u_3}, 1), (\overline{u_3}, 2), (\overline{u_5}, 1)\}
\end{aligned}
$$

where $\overline{u}$ denotes the complement of u, and $C_{0,2}(5) = 52$ (here $\overline{u}(i) = a$ if $u(i) = b$, and $\overline{u}(i) = b$ if $u(i) = a$).

Proposition 7. *The following equality holds for $h \geq 1$, $k \geq 2$ and $n \geq h$:*

$$
U_{h,k}(n) = \frac{(n - h + 1)U_{h-1,k}(n) - C_{h-1,k}(n)}{hk}
$$

Proof. Redefining Position i in an unbordered word u with $h - 1$ holes of length n over a k-letter alphabet will generate an unbordered word if and only if (u, i) is not a critical pair. Since each unbordered word of length n with $h-1$ holes has $n-h+1$ letters, there are $(n-h+1)U_{h-1,k}(n)-C_{h-1,k}(n)$ pairs which are not critical. However, each unbordered word with h holes will be generated hk times. $\square$

Furthermore it is easy to get a formula for $S'_{1,k}(2m, m)$.

Proposition 8. *The following equality holds for $k \geq 2$ and $m \geq 1$:*

$$
S'_{1,k}(2m, m) = \frac{C_{0,k}(2m, m)}{k - 1}
$$

12

Proof. A partial word, u, of length $2m$, with one hole and a minimal border of length m can be generated by exactly k full words: one perfect square and $k-1$ unbordered partial words. We have that u will be generated by an unbordered partial word if and only if the word which generates it is in a critical pair in $\mathcal{C}_{0,k}(2m, m)$ with the position in the critical pair taking any value from the remaining $k-1$ letters of the alphabet, that do not determine a border of length m. $\qquad\square$

To determine $C_{h,k}(n)$, it will be useful to distinguish those pairs which are simple critical and those which are nonsimple critical.

Definition 4. *Let $\mathcal{C}_{h,k}(n, S)$ denote the set of simple critical pairs in $\mathcal{C}_{h,k}(n)$ and $\mathcal{C}_{h,k}(n, N)$ the set of nonsimple critical pairs in $\mathcal{C}_{h,k}(n)$.*

The following equations, which are consequences of the definition, hold:

$$
\begin{array}{rcl}
\mathcal{C}_{h,k}(2m, S) & = & \bigcup_{l=1}^{m} \mathcal{C}_{h,k}(2m, l) \\
\mathcal{C}_{h,k}(2m+1, S) & = & \bigcup_{l=1}^{m} \mathcal{C}_{h,k}(2m+1, l) \\
\mathcal{C}_{h,k}(2m, N) & = & \bigcup_{l=m+1}^{2m-1} \mathcal{C}_{h,k}(2m, l) \\
\mathcal{C}_{h,k}(2m+1, N) & = & \bigcup_{l=m+1}^{2m} \mathcal{C}_{h,k}(2m+1, l)
\end{array}
$$

Observe that $C_{h,k}(n) = C_{h,k}(n, S) + C_{h,k}(n, N)$ where $C_{h,k}(n, S)$ (respectively, $C_{h,k}(n, N)$) denotes the cardinality of $\mathcal{C}_{h,k}(n, S)$ (respectively, $\mathcal{C}_{h,k}(n, N)$). Also, we have that $C_{h,k}(1, S) = C_{h,k}(1, N) = 0$, as well as $C_{h,k}(2, S) = 2U_{h,k}(2)$ and $C_{h,k}(2, N) = 0$.

Note that for a simple critical pair, the position appears in either the prefix or the suffix determined by the simple border length, but not in both the prefix and the suffix. Since, for any given length, we will consider both a word and its reversal, we make the following remark.

Remark 1. *Exactly half of the simple critical pairs are critical in a prefix, and half in a suffix.*

For the unbordered word $u = ababb$, $(u, 0)$ is critical in a prefix while $(u, 3)$ and $(u, 4)$ are critical in a suffix, and for the unbordered word $u = bbaba$, $(u, 0)$ and $(u, 1)$ are critical in a prefix while $(u, 4)$ is critical in a suffix.

For a nonsimple critical pair (u, i) where u is a full word, i appears in exactly one nonsimple border because the word generated by redefining Position i will not have a simple border according to Definition 2, and thus, will be a nonperiodic bordered partial word with one hole. In this case, by Proposition 4, the nonsimple border length l is unique. In addition, for that

nonsimple border length l, Position i will appear in both the prefix and the suffix of length l because a word with one hole will have the hole in both the prefix and the suffix of that length, according to Proposition 3.

Nonsimple critical pairs may be critical in only the prefix, only the suffix, or both. To illustrate this, if $u_1 = abbbb$, then $(u_1, 1)$ is a nonsimple critical pair which is critical only in the suffix

$$
\begin{array}{cccc}
a_0 & a_1 & a_2 & a_3 \\
a & b & b & b \\
\underline{b} & b & b & b \\
a_1 & a_2 & a_3 & a_4
\end{array}
$$

But if $u_2 = abccc$, then $(u_2, 1)$ is a nonsimple critical pair which is critical only in both the prefix and the suffix

$$
\begin{array}{cccc}
a_0 & a_1 & a_2 & a_3 \\
a & \underline{b} & c & c \\
\underline{b} & c & c & c \\
a_1 & a_2 & a_3 & a_4
\end{array}
$$

Assume an unbordered full word $u = a_0 a_1 \ldots a_{n-1}$ of length n has a nonsimple critical pair (u, i) for the border length l. Consider the letters a_{l-n+i}, which is in the prefix and corresponds to a_i in the suffix, and a_{n-l+i}, which is in the suffix and corresponds to a_i in the prefix. Figure 1 denotes this situation (if $a_i = a_{l-n+i}$ we say that (u, i) is only critical in the prefix, and if $a_i = a_{n-l+i}$ we say that (u, i) is critical only in the suffix). In the figure, the prefix of length l of u, $a_0 \ldots a_{l-1}$, has been aligned with the suffix of length l of u, $a_{n-l} \ldots a_{n-1}$.

Figure 1: Nonsimple critical pair

Proposition 9. *Considering all full words of a given length over a k-letter alphabet, $\frac{2}{k}$ of the nonsimple critical pairs are critical in only the prefix or the suffix.*

Proof. Referring to Figure 1, the factor of the prefix preceding a_{l-n+i} agrees with the factor of the suffix preceding a_i, so we have $a_0 \ldots a_{l-n+i-1} = $

14

$a_{n-l} \ldots a_{i-1}$. The factor of length $l - n + i$ of the prefix preceding a_i, which is equal to $a_{n-l} \ldots a_{i-1}$, is equal to the factor of length $l - n + i$ of the suffix preceding a_{n-l+i}, and thus, $a_{n-l} \ldots a_{i-1} = a_{2n-2l} \ldots a_{n-l+i-1}$. This gives us $a_0 \ldots a_{l-n+i-1} = a_{2n-2l} \ldots a_{n-l+i-1}$. Similarly, $a_{l-n+i+1} \ldots a_{2l-n-1} = a_{n-l+i+1} \ldots a_{n-1}$. It must be that $a_{l-n+i} \neq a_{n-l+i}$, or u would have a border of length $2l - n$ with prefix $a_0 \ldots a_{2l-n-1}$ and suffix $a_{2n-2l} \ldots a_{n-1}$. Hence, given distinct letters a_{l-n+i} and a_{n-l+i}, we have (u, i) critical in only the prefix or the suffix for exactly two letters of the k-letter alphabet (that is when $a_i = a_{l-n+i}$ or $a_i = a_{n-l+i}$). □

4.1 Counting simple critical pairs

To count the simple critical pairs determined by a word of length n, we can count the critical pairs for the border lengths 1 through $\lfloor \frac{n}{2} \rfloor$.

Proposition 10. *The equality $C_{0,k}(2m + 1, S) = kC_{0,k}(2m, S)$ holds for $m \geq 1$ and $k \geq 2$.*

In Proposition 11, we will determine simple critical pairs in perfect squares. To illustrate our ideas, if $v_1 = abccc$ then $(v_1, 0)$ is critical in the prefix of length 1, $(v_1, 4)$ is critical in the suffix of length 1, and $(v_1, 1)$ is critical in both the prefix and suffix of length 4. For the word $w_1 = v_1 v_1 = abcccabccc$, only $(w_1, 0)$ and $(w_1, 9)$ are simple critical pairs. Now if $v_2 = abbbb$ then $(v_2, 0)$ is critical in the prefix of length 1, $(v_2, 4)$ is critical in the suffix of length 1, $(v_2, 3)$ is critical in the suffix of length 2, $(v_2, 2)$ is critical in the suffix of length 3, and $(v_2, 1)$ is critical in the suffix of length 4. For $w_2 = v_2 v_2 = abbbbabbbb$, $(w_2, 0)$, $(w_2, 6)$, $(w_2, 7)$, $(w_2, 8)$ and $(w_2, 9)$ are simple critical pairs.

Proposition 11. *The following equality holds for $k \geq 2$ and $m \geq 1$:*

$$C_{0,k}(2m, S) = kC_{0,k}(2m - 1, S) - C_{0,k}(m, S) - \frac{2}{k}C_{0,k}(m, N) + C_{0,k}(2m, m)$$

Proof. Consider the unbordered full word $u = a_0 \ldots a_{2m-1}$, of length $2m$, and $u' = a_0 \ldots a_{m-1} a_{m+1} \ldots a_{2m-1}$ of length $2m - 1$. The pair (u, i) is simple critical if and only if by redefining Position i, u generates a word with a minimal simple border of length m, or (u', i) is a simple critical pair when $0 \leq i < m$ or $(u', i - 1)$ is a simple critical pair when $m < i < 2m$. There are $C_{0,k}(2m, m)$ pairs for which a redefined position generates words with minimal borders of length m. For every unbordered word u' of length $2m - 1$, we can construct k words of length $2m$ which will all be unbordered

15

except for the perfect squares. Thus, $\sum_{l=1}^{m-1} C_{0,k}(2m,l)$ equals $kC_{0,k}(2m-1,S)$ plus $C_{0,k}(2m,m)$ minus the number of simple critical pairs in perfect squares of length $2m$.

Let v be an unbordered full word of length m and $w = vv$. Position i, $0 \le i < m$, where (w,i) is a simple critical pair, will create a critical pair (v,i) only in the prefix of the same length of v. Similarly, Position i, $m < i < 2m$, where (w,i) is a simple critical pair, will create a critical pair $(v, i-m)$ only in the suffix of v of that length. In addition, a position which is critical only in a prefix of v will determine a simple critical pair in the prefix of w of the same length, and a position which is critical only in a suffix of v will create a simple critical pair in the suffix of w of that length.

Thus, the number of simple critical pairs in perfect squares of length $2m$ equals the number of pairs that are critical only in a prefix or a suffix of an unbordered word of length m. $\qquad \square$

To illustrate ideas of the proof of Proposition 12, first $u = u_1 u_2 = (a\underline{b}aab)(a\underline{b}aab)$ is bordered with a border length 2, while $u' = u'_1 u_2 = (a\underline{a}aab)(a\underline{b}aab)$ is unbordered. In the latter example, $(u',1)$ is simple critical for the border length 2. Here, u_2 is bordered with a border length $2 < 5$. Now $u = u_1 u_2 = (aaa\underline{b}b)(aaa\underline{a}b)$ is bordered with a border length 4, while $u' = u'_1 u_2 = (aaa\underline{c}b)(aaa\underline{a}b)$ is unbordered. Here, $(u_2,3)$ is critical in the prefix and not critical in the suffix for the border length 4, and $(u',3)$ is simple critical for that border length. Last $u = u_1 u_2 = (aaa\underline{a}b)(aaa\underline{c}b)$ and $u' = u'_1 u_2 = (aaa\underline{b}b)(aaa\underline{c}b)$ are unbordered. Here, $(u_2,3)$ is critical in both the prefix and suffix for the border length 4, and $(u,3)$ and $(u',3)$ are critical for the border length 5.

Proposition 12. *The following equality holds for $k \ge 2$ and $m \ge 2$:*

$$C_{0,k}(2m,m) = 2(k-1)\left(mU_{0,k}(m) - \frac{1}{2}C_{0,k}(m,S) - \frac{1}{k}C_{0,k}(m,N)\right)$$

Proof. Consider a full word of length $2m$ over a k-letter alphabet, $u = a_0 \dots a_{2m-1}$. We wish to determine when $(u,i) \in \mathcal{C}_{0,k}(2m,m)$, for all $i \in [0..2m-1]$. To simplify this, we will count only the pairs $(u,i) \in \mathcal{C}_{0,k}(2m,m)$ with $i \in [0..m-1]$ and multiply by 2 (refer to Remark 1). We have $(u,i) \in \mathcal{C}_{0,k}(2m,m)$ if and only if u is unbordered, (u,i) is a critical pair for the border length m, but (u,i) is not a critical pair for any border length less than m. This is equivalent to having all three of the following conditions hold:

1. (u, i) is not critical for any border length less than m

2. u does not have a border of length less than or equal to m

3. $u = va_i w v a_{m+i} w$ for some full words v, w with $|v| = i$, $|w| = m - i - 1$, and $a_i \neq a_{m+i}$

Therefore, to count $C_{0,k}(2m, m)$ it suffices to count the number of words u for which the above three conditions hold.

Define $u_1 = va_i w$ and $u_2 = va_{m+i}w$. Note that, if u_2 is bordered with a border length $l < m$, then it must be the case that either u has also a border of length l or (u, i) is a critical pair for the border length l. So, for Conditions 1 and 2 to hold, u_2 must be unbordered. Thus, to count $C_{0,k}(2m, m)$, we may count the pairs of unbordered words u_2 and letters $a_i \neq a_{m+i}$, and then subtract the number of those pairs which still cause u to violate either Condition 1 or Condition 2. Given an index i, the number of pairs of unbordered words u_2 and letters $a_i \neq a_{m+i}$ is equal to $(k-1)U_{0,k}(m)$.

Now we have the following additional condition for u to be constructed by an unbordered word u_2 and letter $a_i \neq a_{m+i}$, but for u to still violate Condition 1 or 2:

4. $(u_2, i) \in \mathcal{C}_{0,k}(m, l)$ is critical in the prefix and not critical in the suffix (note that Position i in u_2 is Position $m + i$ in u)

First, we prove that if $(u_2, i) \in \mathcal{C}_{0,k}(m, l)$ is critical in the prefix but not the suffix, then u violates either Condition 1 or 2. Considering the prefix and suffix of u of length l, it must be that either u has a border of length l or (u, i) is critical for l. Now, assume that u is bordered with a border of length $l < m$. For this not to be a border of u_2, it must be that $(u_2, i) \in \mathcal{C}_{0,k}(m, l)$ is a critical pair only in the prefix, because the suffixes of u and u_2 of length l are equivalent. Assuming that (u, i) is critical for a border length $l < m$, by the same reasoning, $(u_2, i) \in \mathcal{C}_{0,k}(m, l)$ is a critical pair only in the prefix of length l.

By Proposition 9, the number of pairs (u_2, i) for which Condition 4 holds is $\frac{1}{2}C_{0,k}(m, S) + \frac{1}{k}C_{0,k}(m, N)$. Thus,

$$C_{0,k}(2m, m) = 2(k - 1)(mU_{0,k}(m) - \frac{1}{2}C_{0,k}(m, S) - \frac{1}{k}C_{0,k}(m, N))$$

$\square$

Corollary 3. *The following equality holds for $k \geq 2$ and $m \geq 2$:*

$$C_{0,k}(2m, S) = kC_{0,k}(2m-1, S) - kC_{0,k}(m, S) - 2C_{0,k}(m, N) + 2(k-1)mU_{0,k}(m)$$

Proof. Follows immediately from Proposition 11 and Proposition 12. $\square$

4.2 Counting nonsimple critical pairs

To count the nonsimple critical pairs determined by a word of length n, we can count the critical pairs for the border lengths $\lfloor \frac{n}{2} \rfloor + 1$ through $n - 1$. Recall that if a position in a full word is critical for a minimal nonsimple border length, then it is critical for exactly one nonsimple border length. This is due to the fact that the word generated by redefining that position will have a unique nonsimple border.

If v is the word generated by redefining a position from a nonsimple critical pair of a full word u of length n, for a border length $n-l$, $0 < l \le \lceil \frac{n}{2} \rceil - 1$, then, by Proposition 1, v is weakly l-periodic with one hole. For $q = \lceil \frac{n}{l} \rceil - 1$ and $r = n - ql$, note that $q \ge 2$ and $0 < r \le l$. We claim that for some full words v_1, v_2, w, u must be of the form $(v_1 a v_2 w)^{q_1} (v_1 b v_2 w)^{q_2} (v_1 c v_2 w)^{q_3} v_1 c v_2$, where a, b, c are distinct letters, $q_1 > 0$, $q_2 \in \{0, 1\}$, $q_3 \ge 0$, $q = q_1 + q_2 + q_3$, $l = |v_1 a v_2 w|$ and $r = |v_1 c v_2|$. In this form, only the positions determined by the letters a, b, c can be part of nonsimple critical pairs, depending on the values of q_1, q_2 and q_3.

To see this, consider first the case where $v = u_1 w u_2 w \dots u_q w u_{q+1}$ with $|u_1 w| = |u_2 w| = \cdots = |u_q w| = l$, $|u_1| = |u_2| = \cdots = |u_{q+1}| = r$, and $u_i = v_1 \diamond v_2$ for some $1 \le i \le q + 1$. Then there exist letters of the alphabet a, c such that $u_j = v_1 a v_2$ for $1 \le j < i$, and $u_j = v_1 c v_2$ for $i < j \le q + 1$. Note that $a \ne c$ (otherwise, u would have a simple border of length $r < n-l$). There are three possibilities to consider:

- If $i = 1$, then $(u, |v_1|)$ is a simple critical pair, a contradiction.

- If $i = q + 1$, then $(u, ql + |v_1|)$ is a simple critical pair, a contradiction.

- If $1 < i < q + 1$, then there are three possibilities:

  - If the hole corresponds to an a in u, then $q_1 = i > 1, q_2 = 0, q_3 = q - i \ge 0$.

  - If the hole corresponds to a c in u, then $q_1 = i - 1 > 0, q_2 = 0, q_3 = q - i + 1 > 0$.

  - If the hole corresponds to a b in u, distinct from both a and c, then $q_1 = i - 1 > 0, q_2 = 1, q_3 = q - i \ge 0$.

Consider now the case where $v = w u_1 w u_2 \dots w u_q w$ where $|w u_1| = |w u_2| = \cdots = |w u_q| = l$, $|w| = r$, and $u_i = v_1 \diamond v_2$ for some $1 \le i \le q$. In this case, u would have a simple border of length $r < n - l$.

Proposition 13. *Let v_1, v_2, w be full words, $q_1 > 0$, $q_2 \in \{0, 1\}$, $q_3 \geq 0$, and a, b, c distinct letters. Then the word $u' = v_1 a v_2 w v_1 c v_2$ is unbordered if and only if the word $u = (v_1 a v_2 w)^{q_1} (v_1 b v_2 w)^{q_2} (v_1 c v_2 w)^{q_3} v_1 c v_2$ is unbordered.*

Proof. Let $l = |v_1 a v_2 w|$, $r = |v_1 c v_2|$, and $i = |v_1|$. A simple border of u' will be a simple border of u, so if u' is bordered, then u is bordered. First, let us assume that $q_2 = 0$. If u has a border of length smaller than $|v_1 a v_2|$, then u' has a border of the same length. And if u has a border of length at least $|v_1 a v_2|$, then for the last a, $u((q_1 - 1)l + i)$, appearing in the prefix, the factor of the prefix of length strictly less than $r + l$ and beginning with $v_1 a v_2$ corresponds to a factor of the suffix of the same length, ending with $v_1 c v_2$. These factors represent a border of $v_1 a v_2 w v_1 c v_2$.

Now, let us assume that $q_2 = 1$. Given a simple border length of u, at least one of the prefix or suffix must be strongly l-periodic. If u has a simple border, where $u(q_1 l + i) = b$ is in the prefix, then the corresponding position in the suffix is a b. But then, the last a, $u((q_1 - 1)l + i)$, in the prefix, must also correspond to a b in the suffix, since the suffix is strongly l-periodic. Similarly, if $u(q_1 l + i) = b$ were to appear in the suffix, then the first c, $u((q_1 + 1)l + i)$, in the suffix, must also correspond to a b. Now, if u has a simple border in which $u(q_1 l + i) = b$ does not appear, then we proceed as in the case where $q_2 = 0$. $\square$

In addition, we have that, if $v_1 a v_2 w v_1 c v_2$ is unbordered, then $v_1 a v_2 v_1 c v_2$ is unbordered. This is because any simple border of $v_1 a v_2 v_1 c v_2$ must be a border of $v_1 a v_2 w v_1 c v_2$. Thus, if we can count the unbordered full words of the form $v_1 a v_2 w v_1 c v_2$ for a certain $r = |v_1 c v_2|$ and $l = |v_1 a v_2 w|$, then we can count the number of ways to construct u so it has critical pairs for the border length $n - l$.

Definition 5. *Let $k \geq 2$, $n > 0$ and $r > 0$ be integers.*

- *If $n \leq r$, then $F_{0,k}(n, r) = 0$.*

- *If $n > r$, then set $n = l + r$ for some $l > 0$. Then $F_{0,k}(n, r)$ is the number of unbordered full words of length n that have a prefix $v_1 a v_2$ and a suffix $v_1 c v_2$ of length r, where a, c are distinct letters of the alphabet of size k, and v_1, v_2 are full words.*

Proposition 14. *The following equality holds for $k \geq 2$, $n \geq 2$, $0 < l \leq \lceil \frac{n}{2} \rceil - 1$, where $q = \lceil \frac{n}{l} \rceil - 1$ and $r = n - ql$:*

$$C_{0,k}(n, n - l) = k(q - 1)F_{0,k}(l + r, r)$$

Therefore,

$$C_{0,k}(n, N) = \sum_{l=1}^{\lceil \frac{n}{2} \rceil - 1} k(q-1)F_{0,k}(l+r, r)$$

Proof. Consider the unbordered full word

$$u = (v_1 a v_2 w)^{q_1} (v_1 b v_2 w)^{q_2} (v_1 c v_2 w)^{q_3} v_1 c v_2$$

of length n, where $q_1 > 0$, $q_2 \in \{0, 1\}$, $q_3 \geq 0$, $q = q_1 + q_2 + q_3$, $l = |v_1 a v_2 w|$ and $r = |v_1 c v_2|$. Set $|v_1| = i$. When $q_1 = 1$ and $q_2 = 0$, we have

$$u = v_1 u(i) v_2 w (v_1 c v_2 w)^{q_3} v_1 u(ql + i) v_2$$

Since the first a, $u(i)$, and the last c, $u(ql + i)$, determine the simple critical pairs (u, i) and $(u, ql + i)$ for the border length r (here $r < n - l$), only the first c, $u(l + i)$, will create a nonsimple critical pair for the border length $n - l$ in the suffix. The prefix and suffix of u of length $n - l$ are respectively:

$$v_1 a v_2 (w v_1 c v_2)^{q_3}$$
$$v_1 \underline{c} v_2 (w v_1 c v_2)^{q_3}$$

When $q_2 = 0$ and $q_3 = 0$, we have $u = (v_1 a v_2 w)^{q_1} v_1 c v_2$ and only the last a, $u((q-1)l + i)$, will determine a critical pair for the border length $n - l$ in the prefix:

$$(v_1 a v_2 w)^{q_1 - 1} v_1 \underline{a} v_2$$
$$(v_1 a v_2 w)^{q_1 - 1} v_1 c v_2$$

When $q_1 > 1$, $q_2 = 0$ and $q_3 > 0$, we have $u = (v_1 a v_2 w)^{q_1} (v_1 c v_2 w)^{q_3} v_1 c v_2$ and both the last a, $u((q_1 - 1)l + i)$, in the prefix, and the first c, $u(q_1 l + i)$, in the suffix, will create critical pairs for the border length $n - l$ (in this case, q_1 may assume any of the values 2 through $q - 1$):

$$(v_1 a v_2 w)^{q_1 - 1} v_1 \underline{a} v_2 w v_1 c v_2 (w v_1 c v_2)^{q_3 - 1}$$
$$(v_1 a v_2 w)^{q_1 - 1} v_1 \underline{c} v_2 w v_1 c v_2 (w v_1 c v_2)^{q_3 - 1}$$

When $q_2 = 1$, we have $u = (v_1 a v_2 w)^{q_1} (v_1 b v_2 w)(v_1 c v_2 w)^{q_3} v_1 c v_2$, only the b, $u(q_1 l + i)$, creates a critical pair for the border length $n - l$, both in the prefix and the suffix, and q_1 may assume any of the values 1 through $q - 1$. Given distinct letters a and c, there are $k - 2$ choices for the letter b such that it is distinct from both a and c. The prefix and suffix of u of length $n - l$ are respectively:

$$(v_1 a v_2 w)^{q_1 - 1} v_1 a v_2 w v_1 \underline{b} v_2 (w v_1 c v_2)^{q_3}$$
$$(v_1 a v_2 w)^{q_1 - 1} v_1 \underline{b} v_2 w v_1 c v_2 (w v_1 c v_2)^{q_3}$$

Thus, we have

$$C_{0,k}(n, n - l) = (2 + 2(q - 2) + (q - 1)(k - 2))F_{0,k}(l + r, r)$$

Note that if a letter other than the first a, $u(i)$, or the last c, $u(ql + i)$, determines a critical pair for the nonsimple border length $n - l$, then it does not create a simple critical pair for any border length. For a simple border of u, either the prefix, the suffix, or both, will be strongly l-periodic. Thus, for a pair (u, j), critical for the nonsimple border length $n - l$, to be a simple critical pair with $j \neq i, j \neq ql + i$, either the prefix of the simple border is $(v_1 a v_2 w)^{q_1} v_1 u(j) v_2'$, where $u(j) \neq a$, and the suffix is strongly l-periodic, or the suffix of the simple border is $v_1' u(j) v_2 (w v_1 c v_2)^{q_3}$, where $u(j) \neq c$, and the prefix is strongly l-periodic, for some prefix v_2' of v_2 or some suffix v_1' of v_1. Without loss of generality, assume (u, j) is critical in the prefix. Then the letter corresponding to j in the suffix is an a, because it must agree with the other a's in the prefix. However, then $v_1 a v_2'$ is a border of u. $\square$

Proposition 15. *The following equality holds for $k \geq 2$ and $r \geq 1$:*

$$F_{0,k}(2r, r) = (k - 1)r U_{0,k}(r) - \frac{1}{k} C_{0,k}(r, N)$$

Proof. Here $F_{0,k}(2r, r)$ counts the unbordered words of length $2r$ of the form $u = v_1 a v_2 v_1 c v_2 = v_1 u(i) v_2 v_1 u(r + i) v_2$ with $a \neq c$ (where $|v_1| = i$). Note that, it must be the case that either $u_1 = v_1 u(i) v_2$, $u_2 = v_1 u(r + i) v_2$, or both, are unbordered. This is because, if both the full words u_1 and u_2 are bordered, then they both have only simple borders. If $u(i)$ does not appear in the minimal border of u_1, then u has the same border. If $u(r + i)$ does not appear in the minimal border of u_2, then u has the same border. Thus, it must be that both minimal borders are simple with $u(i)$ and $u(r + i)$ in the same position, and either both appear in the prefixes or suffixes. If $u(i)$, $u(r + i)$ appear in the prefixes, then let x_1 be the prefix of u_1 and x_2 be the suffix of u_1 that create a border of u_1 such that $|x_1| \geq |v_1 u(i)|$. In this case, x_2 being a suffix of v_2, we see that x_1 and x_2', the suffix of length $|x_2|$ of u_2, create a border of u. A similar argument can be made if $u(i)$, $u(r + i)$ appear in the suffixes. Thus, it suffices to count the number of pairs u_1, u_2 which are both unbordered, and add the number of pairs u_1, u_2 where exactly one is unbordered. For every word u_1, we will have $k - 1$ choices for u_2, such that $u(r + i) \neq u(r)$.

Given an unbordered word u_1 and position i, the following hold for u and u_2.

- If (u_1, i) is not a critical pair, then u and u_2 will be unbordered.

- If (u_1, i) is a simple critical pair in the prefix, then u will be unbordered and u_2 will be bordered if there exists a critical border length in which the position corresponding to $u(i)$ is a c. Otherwise u_2 will be unbordered.

- If (u_1, i) is a simple critical pair in the suffix, then u will be bordered if there exists a critical border length in which the position corresponding to $u(i)$ is a c. Otherwise u will be unbordered. The word u_2 is unbordered if and only if u is unbordered.

- If (u_1, i) is a nonsimple critical pair in only the prefix, then u and u_2 will be unbordered.

- If (u_1, i) is a nonsimple critical pair only in the suffix, then u will be unbordered for the $k - 2$ of $k - 1$ letters which are neither a nor c and correspond to $u(i)$ appearing critically in the suffix. The word u_2 will be unbordered when u is unbordered.

- If (u_1, i) is a nonsimple critical pair in both the prefix and suffix, then both u and u_2 are unbordered.

Let α be the proportion of words u_1 in which (u_1, i) is a simple critical pair in the prefix and there exists a critical border length in which the position corresponding to $u(i)$ is a c. This is equal to the proportion of words u_1 in which (u_1, i) is a simple critical pair in the suffix and there exists a critical border length in which the position corresponding to $u(i)$ is a c. For the case $k = 2$, we have $\alpha = 1$.

Since we can choose either u_1 or u_2 to be unbordered, we add twice the number of words where u_2 is bordered and add once the number of words where both u_1 and u_2 are unbordered. Then we have

$$F_{0,k}(2r, r) = (k-1)\left(rU_{0,k}(r) - C_{0,k}(r, S) - C_{0,k}(r, N) + 2\alpha\left(\frac{1}{2}C_{0,k}(r, S)\right)\right.$$

$$+(1-\alpha)\frac{1}{2}C_{0,k}(r, S) + (1-\alpha)\frac{1}{2}C_{0,k}(r, S) + \frac{1}{2}\left(\frac{2}{k}\right)C_{0,k}(r, N) + \frac{k-2}{k}C_{0,k}(r, N)$$

$$\left. + \frac{1}{2}\left(\frac{k-2}{k-1}\right)\left(\frac{2}{k}\right)C_{0,k}(r, N)\right) = (k-1)rU_{0,k}(r) - \frac{1}{k}C_{0,k}(r, N)$$

$\square$

Proposition 16. *The following equalities hold for $k \geq 2$, $n \geq 2$, $r \geq 1$, $q = \lceil \frac{n}{n-r} \rceil - 1$, and $r' = n - q(n-r)$:*

$$
F_{0,k}(n,r) = \begin{cases}
0 & \text{if } n \leq r \\
U_{0,k}(n) & \text{if } 1 = r < n \\
qF_{0,k}(n - r + r', r') & \text{if } 1 < r < n < 2r \\
(k-1)rU_{0,k}(r) - \frac{1}{k}C_{0,k}(r, N) & \text{if } 1 < r < n = 2r \\
kF_{0,k}(n - 1, r) & \text{if } n > 2r, n \text{ odd} \\
kF_{0,k}(n - 1, r) - F_{0,k}(\frac{n}{2}, r) & \text{if } n > 2r, n \text{ even}
\end{cases}
$$

Proof. Note that for $n > 2r$, in order to construct an unbordered full word of length $n = r + l$ of the form $v_1 a v_2 w v_1 c v_2$ where $l = |v_1 a v_2 w|$, $r = |v_1 c v_2|$, and a, c are distinct letters, we may take a word of the same form $v_1 a v_2 w' v_1 c v_2$, of length $n - 1 = r + (l - 1)$, and add a letter to the center. However, we need to consider those cases when a perfect square is created. Thus, for $n > 2r$, n odd, we have

$$F_{0,k}(n,r) = kF_{0,k}(n - 1, r)$$

And for $n > 2r$, n even, we have

$$F_{0,k}(n,r) = kF_{0,k}(n - 1, r) - F_{0,k}(\frac{n}{2}, r)$$

Next, we need to compute $F_{0,k}(n,r)$ for values of n from 2 to $2r$. We have that $F_{0,k}(n,r) = 0$, for $n \in [2..r]$. For $r = 1$, $n > r$, $F_{0,k}(n,1)$ is the number of unbordered words of length n. Thus, $F_{0,k}(n,1) = U_{0,k}(n)$. For $2r = n > r > 1$, $F_{0,k}(n,r)$ is computed according to Proposition 15.

For $2r > n > r > 1$, $F_{0,k}(n,r)$ represents the number of unbordered words u of length n with a prefix $v_1 a v_2$ of length r, and a suffix $v_1 c v_2$ of length r. Since $(u, |v_1|)$ and $(u, n - 1 - |v_2|)$ are both critical pairs for the nonsimple border length r, we can count the words of length n with critical pairs for the border length r. These words can be constructed from words of the form $u' = v_1' a v_2' w v_1' c v_2'$, where $|v_1' a v_2' w| = n - r$ and $|v_1' c v_2'| = r'$. However, we have the additional requirement that $(u, |v_1|)$ and $(u, n-1-|v_2|)$ are only critical in either the prefix or the suffix, and not in both. From Proposition 14, for $q_1 = 1, q_2 = 0$, u has one nonsimple critical pair which is critical only in the suffix. For $q_2 = 0, q_3 = 0$, u has one nonsimple critical pair which is critical only in the prefix. For $q_1 > 1$, $q_2 = 0$ and $q_3 > 0$, u has two nonsimple critical pairs, one of which is critical only in the prefix and the other critical only in the suffix (q_1 may assume any of the values 2

through $q - 1$). For $q_2 = 1$, all nonsimple critical pairs are critical in both the prefix and the suffix, and thus, words of this form are not considered. Therefore, we can use the following recursive formula for n from $r + 1$ to $2r - 1$:

$$F_{0,k}(n, r) = (2 + q - 2)F_{0,k}(n - r + r', r') = qF_{0,k}(n - r + r', r')$$

with $q = \lceil \frac{n}{n-r} \rceil - 1$ and $r' = n - q(n - r)$, such that $r' < n - r < r$. $\qquad \square$

5 Counting bordered partial words with one hole

We can use the previously obtained results to calculate $U_{1,k}(n)$, for any integers $k \geq 2$, $n \geq 2$.

Theorem 1. *The following equalities hold for $k \geq 2$:*

- *If $m \geq 1$, then*

$$U_{1,k}(2m) = \frac{2mU_{0,k}(2m) - C_{0,k}(2m, S) - C_{0,k}(2m, N)}{k}$$

- *If $m \geq 0$, then*

$$U_{1,k}(2m+1) = \frac{(2m + 1)kU_{0,k}(2m) - kC_{0,k}(2m, S) - C_{0,k}(2m + 1, N)}{k}$$

where $C_{0,k}(2m, S)$ is computed according to Proposition 10 and Corollary 3, and $C_{0,k}(2m, N), C_{0,k}(2m+1, N)$ are computed according to Propositions 14, 15 and 16.

As an example, consider $n = 4$. By Proposition 10 and Corollary 3, we have

$$
\begin{aligned}
C_{0,k}(4, S) &= kC_{0,k}(3, S) - kC_{0,k}(2, S) - 2C_{0,k}(2, N) + 4(k - 1)U_{0,k}(2) \\
&= k^2C_{0,k}(2, S) - 2k^2(k - 1) + 4(k - 1)(k^2 - k) \\
&= (k - 1)(2k^3 - 2k^2 + 4k^2 - 4k) \\
&= 2k^4 - 6k^2 + 4k
\end{aligned}
$$

By Propositions 14 and 16, we have

$$
\begin{aligned}
C_{0,k}(4, N) &= k(3 - 1)F_{0,k}(1 + 1, 1) \\
&= 2k((k - 1)U_{0,k}(1) - \tfrac{1}{k}C_{0,k}(1, N)) \\
&= 2kU_{0,k}(2) \\
&= 2k^3 - 2k^2
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
C_{0,k}(4) &= C_{0,k}(4, S) + C_{0,k}(4, N) \\
&= 2k^4 + 2k^3 - 8k^2 + 4k
\end{aligned}
$$

By Proposition 7, we have

$$
\begin{aligned}
U_{1,k}(4) &= \frac{4U_{0,k}(4) - C_{0,k}(4)}{k} \\
&= \frac{4(k^4 - k^3 - k^2 + k) - (2k^4 + 2k^3 - 8k^2 + 4k)}{k} \\
&= 2k^3 - 6k^2 + 4k
\end{aligned}
$$

Table 1 gives formulas for $U_{h,k}(n)$ when $0 \le n \le 7$.

We can also use the previously obtained results to calculate $S_{1,k}(n)$, for any integers $k \ge 2$, $n \ge 1$.

Theorem 2. *The following equalities hold for $k \ge 2, m \ge 1$:*

$$
\begin{aligned}
S_{1,k}(2m) &= k^2 S_{1,k}(2m-2) + 2k \sum_{l=1}^{m-1} U_{0,k}(l) k^{2m-2-2l} \\
&\quad + 2(m U_{0,k}(m) - \tfrac{1}{2} C_{0,k}(m, S) - \tfrac{1}{k} C_{0,k}(m, N)) \\
S_{1,k}(2m+1) &= k S_{1,k}(2m) + \sum_{l=1}^{m} U_{0,k}(l) k^{2m-2l}
\end{aligned}
$$

where $S_{1,k}(0) = 0$, $S_{1,k}(1) = 0$, and where $C_{0,k}(m, S)$ is computed according to Proposition 10 and Corollary 3, and $C_{0,k}(m, N)$ is computed according to Propositions 14, 15 and 16.

As an example, consider $n = 4$. We have

$$
\begin{aligned}
S_{1,k}(2) &= k^2 S_{1,k}(0) + 2k(0) + 2(U_{0,k}(1) - \tfrac{1}{2} C_{0,k}(1, S) - \tfrac{1}{k} C_{0,k}(1, N)) \\
&= k^2(0) + 2(k - \tfrac{1}{2}(0) - \tfrac{1}{k}(0)) \\
&= 2k
\end{aligned}
$$

and consequently,

$$
\begin{aligned}
S_{1,k}(4) &= k^2 S_{1,k}(2) + 2k U_{0,k}(1) k^0 \\
&\quad + 2(2 U_{0,k}(2) - \tfrac{1}{2} C_{0,k}(2, S) - \tfrac{1}{k} C_{0,k}(2, N)) \\
&= 2k^3 + 2k^2 + 2(2(k^2 - k) - \tfrac{1}{2} 2(k^2 - k) - \tfrac{1}{k}(0)) \\
&= 2k^3 + 4k^2 - 2k
\end{aligned}
$$

Tables 2 and 3 give formulas for $S_{h,k}(n)$ when $2 \le n \le 7$. Note that $S_{h,k}(0) = S_{h,k}(1) = 0$ for $h \ge 0$.

			$k=2$	$k=3$	$k=4$
$U_{h,k}(0) =$	1	if $h=0$	1	1	1
	0	if $h\geq 1$	0	0	0
		total:	1	1	1
$U_{h,k}(1) =$	k	if $h=0$	2	3	4
	1	if $h=1$	1	1	1
	0	if $h\geq 2$	0	0	0
		total:	3	4	5
$U_{h,k}(2) =$	$k^2 - k$	if $h=0$	2	6	12
	0	if $h\geq 1$	0	0	0
		total:	2	6	12
$U_{h,k}(3) =$	$k^3 - k^2$	if $h=0$	4	18	48
	0	if $h\geq 1$	0	0	0
		total:	4	18	48
$U_{h,k}(4) =$	$k^4 - k^3 - k^2 + k$	if $h=0$	6	48	180
	$2k^3 - 6k^2 + 4k$	if $h=1$	0	12	48
	0	if $h\geq 2$	0	0	0
		total:	6	60	228
$U_{h,k}(5) =$	$k^5 - k^4 - k^3 + k^2$	if $h=0$	12	144	720
	$3k^4 - 6k^3 - k^2 + 4k$	if $h=1$	4	84	384
	0	if $h\geq 2$	0	0	0
		total:	16	228	1104
$U_{h,k}(6) =$	$k^6 - k^5 - k^4 + k^2$	if $h=0$	20	414	2832
	$4k^5 - 6k^4 - 6k^3 + 8k^2$	if $h=1$	16	396	2304
	$4k^4 - 16k^3 + 20k^2 - 8k$	if $h=2$	0	48	288
	0	if $h\geq 3$	0	0	0
		total:	36	858	5424
$U_{h,k}(7) =$	$k^7 - k^6 - k^5 + k^3$	if $h=0$	40	1242	11328
	$5k^6 - 7k^5 - 6k^4 + 7k^3 - 3k^2 + 4k$	if $h=1$	52	1632	12192
	$9k^5 - 26k^4 + 9k^3 + 24k^2 - 16k$	if $h=2$	8	492	3456
	$2k^4 - 12k^3 + 22k^2 - 12k$	if $h=3$	0	0	48
	0	if $h\geq 4$	0	0	0
		total:	100	3366	27024

Table 1: Formulas for $U_{h,k}(n)$ when $0 \leq n \leq 7$

			$k=2$	$k=3$	$k=4$
$S_{h,k}(2) =$	k	if $h=0$	2	3	4
	$2k$	if $h=1$	4	6	8
	1	if $h=2$	1	1	1
	0	if $h\geq 3$	0	0	0
		total:	7	10	13
$S_{h,k}(3) =$	k^2	if $h=0$	4	9	16
	$2k^2+k$	if $h=1$	10	21	36
	$3k$	if $h=2$	6	9	12
	1	if $h=3$	1	1	1
	0	if $h\geq 4$	0	0	0
		total:	21	40	65
$S_{h,k}(4) =$	k^3+k^2-k	if $h=0$	10	33	76
	$2k^3+4k^2-2k$	if $h=1$	28	84	184
	$6k^2$	if $h=2$	24	54	96
	$4k$	if $h=3$	8	12	16
	1	if $h=4$	1	1	1
	0	if $h\geq 5$	0	0	0
		total:	71	184	373
$S_{h,k}(5) =$	$k^4+k^3-k^2$	if $h=0$	20	99	304
	$2k^4+5k^3-k^2-k$	if $h=1$	66	285	812
	$8k^3+4k^2-2k$	if $h=2$	76	246	568
	$10k^2$	if $h=3$	40	90	160
	$5k$	if $h=4$	10	15	20
	1	if $h=5$	1	1	1
	0	if $h\geq 6$	0	0	0
		total:	213	736	1865
$S_{h,k}(6) =$	$k^5+k^4-k^2$	if $h=0$	44	315	1264
	$2k^5+6k^4+4k^3-8k^2+2k$	if $h=1$	164	1014	3720
	$10k^4+14k^3-12k^2+3k$	if $h=2$	230	1089	3276
	$20k^3$	if $h=3$	160	540	1280
	$15k^2$	if $h=4$	60	135	240
	$6k$	if $h=5$	12	18	24
	1	if $h=6$	1	1	1
	0	if $h\geq 7$	0	0	0
		total:	671	3112	9805

Table 2: Formulas for $S_{h,k}(n)$ when $2 \leq n \leq 6$

$S_{h,k}(7) =$			$k=2$	$k=3$	$k=4$
	$k^6 + k^5 - k^3$	if $h=0$	88	945	5056
	$2k^6 + 7k^5 + 5k^4 - 8k^3 + k^2$	if $h=1$	372	3357	16144
	$12k^5 + 20k^4 - 8k^3 - 5k^2 + 2k$	if $h=2$	624	4281	16824
	$30k^4 + 14k^3 - 12k^2 + 3k$	if $h=3$	550	2709	8396
	$35k^3$	if $h=4$	280	945	2240
	$21k^2$	if $h=5$	84	189	336
	$7k$	if $h=6$	14	21	28
	1	if $h=7$	1	1	1
	0	if $h \geq 8$	0	0	0
		total:	2013	12448	49025

Table 3: Formulas for $S_{h,k}(7)$

6   Conclusion

In [14], Moore et al. discuss how to count and generate all words of length n constructed using exactly k letters that give rise to distinct border arrays, for all positive integers n and k, providing an algorithm that computes all such words in constant time per word. Recall that the border array β corresponding to a word u of length n over a k-letter alphabet is a word of same length over the alphabet $\{0, 1, \ldots, n-1\}$ such that for $0 \leq j < n$, $\beta(j)$ is the length of the longest border of the prefix of length $j+1$ of u. For example, the binary words *abaab* and *ababb* of length 5 give rise to the distinct border arrays 00112 and 00120, respectively.

A possible topic to investigate in the future is the generalization to partial words of the enumeration of strings with distinct border arrays. Unfortunately, a translation of the results from [14] to partial words is not trivial. This is due, in particular, to the fact that the authors use a tree structure over the alphabet in order to illustrate the smallest canonical strings that give each of the border arrays. Their tree starts with the letter a at the first level, giving the border array 0, and continues to the second level with b and a, giving all possible border arrays of length two, that is, 00 and 01. Continuing in this manner, the authors are able to generate all possible border arrays, by giving the smallest canonical string in each class.

Let us now consider this approach in the context of partial words. Note that the root of the tree would be in this case $\diamond$, which is the smallest lexicographical symbol. It follows that all canonical strings of length two

start with $\diamond$ and so do all canonical strings of length three. Note however that Moore et al.'s tree construction is not suitable for partial words since the border array 001, associated with the canonical string $ab\diamond$, cannot be obtained in such a way. Thus, not only using such a tree structure we cannot get all canonical strings associated with the border arrays, but moreover, some border arrays are missing.

We observe that, except for partial words of length 1, all border arrays associated with partial words with at least one hole have at least one position greater than 0. Note that it may also be possible to change some positions of a word with holes while preserving the border array. Consider for example the word $u = abaab$ with border array 00112. The partial words $ab\diamond ab$, $abaa\diamond$ and $ab\diamond a\diamond$ all have the same border array as u (they belong to the same equivalence class); however, changing position 1 in u with a hole, to obtain $v = a\diamond aab$, does "increase" the border array from 00112 to 01232. Note that the 2 (respectively, 3) in position 2 (respectively, 3) is due to the presence of a nonsimple border. Determining which positions in a word can be changed to holes while preserving the border array becomes an interesting problem.

Acknowledgement

References

[1] K. Abrahamson. Generalized string matching. *SIAM Journal on Computing*, 16(6):1039–1051, 1987.

[2] J. Berstel and L. Boasson. Partial words and a theorem of Fine and Wilf. *Theoretical Computer Science*, 218:135–141, 1999.

[3] F. Blanchet-Sadri. Primitive partial words. *Discrete Applied Mathematics*, 148(3):195–213, 2005.

[4] F. Blanchet-Sadri. *Algorithmic Combinatorics on Partial Words*. Chapman & Hall/CRC Press, Boca Raton, FL, 2008.

[5] F. Blanchet-Sadri. Open problems on partial words. In G. Bel-Enguix, M. D. Jiménez-López, and C. Martín-Vide, editors, *New Developments in Formal Languages and Applications*, volume 3, chapter 2, pages 11–58. Springer-Verlag, Berlin, 2008.

[6] F. Blanchet-Sadri and A. R. Anavekar. Testing primitivity on partial words. *Discrete Applied Mathematics*, 155(3):279–287, 2007.

[7] F. Blanchet-Sadri, C. D. Davis, J. Dodge, R. Mercaş, and M. Moorefield. Unbordered partial words. *Discrete Applied Mathematics*, 157(5-6):890 – 900, 2009.

[8] C. Choffrut and J. Karhumäki. Combinatorics of Words. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 329–438. Springer-Verlag, Berlin, 1997.

[9] M. Crochemore and W. Rytter. *Jewels of Stringology*. World Scientific, NJ, 2003.

[10] M. J. Fischer and M. S. Paterson. String-matching and other products. MAC Technical Memorandum 41, 1974.

[11] H. Harborth. Endliche 0–1-folgen mit gleichen teilblöcken. *Journal für die reine und angewandte Mathematik*, 271:139–154, 1974.

[12] M. Lothaire. *Combinatorics on Words*. Cambridge University Press, Cambridge, 1997.

[13] F. Manea and C. Tiseanu. Hard counting problems for partial words. In A.-H. Dediu, H. Fernau, and C. Martín-Vide, editors, *LATA 2010, 4th International Conference on Language and Automata Theory and Applications, Trier, Germany*, volume 6031 of *Lecture Notes in Computer Science*, pages 426–438, Berlin, Heidelberg, 2010. Springer-Verlag.

[14] D. Moore, W. F. Smyth, and D. Miller. Counting distinct strings. *Algorithmica*, 23(1):1–13, 1999.

[15] M. Régnier. Enumeration of bordered words the language of the laughing cow. *RAIRO-Theoretical Informatics and Applications*, 26:303–317, 1992.

[16] W. F. Smyth. *Computing Patterns in Strings*. Pearson Addison-Wesley, 2003.

[17] P. Tolstrup Nielsen. A note on bifix-free sequences. *IEEE Transactions on Information Theory*, 19:704–706, 1973.