

# Dyck Words, Lattice Paths, and Abelian Borders

F. Blanchet-Sadri

Department of Computer Science  
University of North Carolina  
P.O. Box 26170  
Greensboro, NC 27402-6170, USA

blanchet@uncg.edu

Kun Chen

k\_chen2@uncg.edu

Kenneth Hawes

Department of Mathematics  
University of Virginia  
P.O. Box 400137  
Charlottesville, VA 22904-4137, USA

kgh5cc@virginia.edu

We use results on Dyck words and lattice paths to derive a formula for the exact number of binary words of a given length with a given minimal abelian border length, tightening a bound on that number from Christodoulakis et al. (Discrete Applied Mathematics, 2014). We also extend to any number of distinct abelian borders a result of Rampersad et al. (Developments in Language Theory, 2013) on the exact number of binary words of a given length with no abelian borders. Furthermore, we generalize these results to partial words.

## 1 Introduction

*Abelian borders* and *abelian periods* are defined similarly to their classical counterparts based on abelian equivalence. Two words are *abelian equivalent* if they have the same *Parikh vector* (a vector that records the frequency of each letter occurring in the word); equivalently, if the two words are the same up to a permutation of characters. For example, the word 010000010010100 has, in particular, an abelian border of length 3 since the prefix of length 3, 010, is abelian equivalent to the suffix of length 3, 100. It has an abelian period of length 4 as the following factorization suggests: 0100 | 0001 | 0010 | 100. Connections between these concepts exist (see, e.g., [7]). Abelian versions of some well-known results such as Fine and Wilf's periodicity theorem [11, 2] and the critical factorization theorem [1] were proposed. Several algorithms that efficiently compute abelian periods in a given word were developed [12, 15, 18]. Abelian concepts were reviewed in [9]; applications include sequencing from compomers [3], permutation pattern discovery in biosequences [14], gapped permutation patterns for comparative genomics [21], and jumbled pattern matching [5, 4, 6].

Christodoulakis et al. [10] studied abelian borders in words over an alphabet of size two. They investigated the number of such words with fixed minimal border length by establishing connections with Dyck words. A *Dyck word* is a word over the alphabet  $\{0, 1\}$  that has the same number of zeros and ones and that has no prefix with more ones than zeros, e.g., 0100110101 is a Dyck word but 0110 is not. The language of Dyck words is equivalent to the language of balanced parentheses: 0101000111 corresponds to  $()()((( )))$ . The number of Dyck words of length  $2n$  is given by the Catalan number  $C_n = \frac{1}{n+1} \binom{2n}{n}$ . Christodoulakis et al. [10] gave bounds on the number of binary words of a given length with no abelian borders and on the number of binary words of a given length with at least one abelian border. They also described an algorithm to find the minimal abelian border of a binary word of length  $n$  in  $\Theta(\sqrt{n})$  time on average (when the word has an abelian border). They left as open problems the derivation of tighter bounds and the generalization to larger alphabet sizes.

Rampersad et al. [22] investigated the number of words with or without abelian borders by establishing connections with *lattice paths*; in particular, they used lattice paths to find the exact number of binary words of a given length with no abelian borders. Similarly to a construction given in [22], we can

visualize a word  $w$  over  $\{0, 1\}$  as an ordered pair of increasing lattice paths, where one path corresponds to the prefixes of the word, and the other corresponds to the suffixes. A step to the right represents the letter 0, and a step upward represents the letter 1. An intersection between the two paths after  $k$  steps indicates an abelian border of length  $k$ . If we graph the distance in steps between the prefix path and the suffix path, we get a *Motzkin path* which consists of steps diagonally upward corresponding to 0, steps diagonally downward corresponding to 1, and steps straight forward corresponding to 2. If we remove the straight forward steps, we get a *Dyck path*. A binary word has an abelian border of a given length if and only if the corresponding Dyck path is encoded by a Dyck word or the bitwise negation of a Dyck word [10].

Since lattice paths have been studied independently of the (relatively new) concept of abelian borders (see, e.g., [13, 16]), in this paper we use known results about lattice paths to discover properties of words with abelian borders. Among other results, by counting pairs of lattice paths by intersections, we derive a formula, that involves the Catalan number, for the exact number of binary words of a given length with a given minimal abelian border length, tightening the abovementioned bound on that number from Christodoulakis et al. [10]. We also extend the abovementioned result of Rampersad et al. [22] to any number of distinct abelian borders. Furthermore, we generalize these results to partial words. Other enumerative applications of non-intersecting lattice paths can be found in [17].

Our paper is the first attempt to study abelian borders of partial words (abelian periods were studied in [2]). Our paper's contents are as follows: In Section 2, we review some basic concepts on partial words such as abelian borders. In Section 3, we describe an  $O(n)$  time algorithm to compute the minimal abelian border of a given partial word of length  $n$  over an alphabet of any size. It is an adaptation of an algorithm from [10]. In Section 4, we extend to partial words a relation from [10] between minimal abelian borders and Dyck words. In Section 5, we apply results from intersecting lattice paths to the enumeration of abelian borders for total words. We also count abelian bordered partial words using lattice paths. Finally in Section 6, we conclude with some remarks.

## 2 Preliminaries

Let  $\Sigma$  be a finite and non-empty set of characters, called an *alphabet*. A sequence of characters from  $\Sigma$  is referred to as a *total word* over  $\Sigma$ . A *partial word* over  $\Sigma$  is a sequence of characters from  $\Sigma_\diamond = \Sigma \cup \{\diamond\}$ , where  $\diamond$ , a new character which is not in  $\Sigma$ , is the “don't care” or “hole” character (it represents an undefined position). A partial word may have zero or more holes, while a total word has zero holes. A total word is also a partial word. In the rest of the paper, when we refer to a “letter” we mean a character from the alphabet  $\Sigma$  (we will never call the  $\diamond$  character a letter since it does not belong to  $\Sigma$ ). We will use “word” and “total word” as equivalent terms. The *length* of a partial word  $w$ , denoted by  $|w|$ , is the number of  $\Sigma_\diamond$  characters in  $w$ , e.g., if  $w = abba\diamond\diamond c$  then  $|w| = 8$ . The *empty word* is the word of length zero and we let  $\varepsilon$  denote it. The set of all words over  $\Sigma$  is denoted by  $\Sigma^*$ . Similarly, the set of all non-empty words over  $\Sigma$  is denoted by  $\Sigma^+$ . We let  $\Sigma^n$  denote the set of all words of length  $n$  over  $\Sigma$ . We can similarly define  $\Sigma_\diamond^*$ ,  $\Sigma_\diamond^+$ , and  $\Sigma_\diamond^n$  for partial words over  $\Sigma$ .

A *completion* over  $\Sigma$  of a partial word  $w$ , denoted by  $\hat{w}$ , is a total word constructed by replacing each  $\diamond$  in  $w$  with some letter in  $\Sigma$  (i.e., we fill all the holes in  $w$ ). If  $w$  has one or more holes and  $\Sigma$  has more than one letter, then there is more than one distinct completion of  $w$  and in this case  $\hat{w}$  denotes an arbitrary completion of  $w$  unless otherwise specified.

A partial word  $u$  is a *factor* of a partial word  $w$  if there exist (possibly empty) partial words  $x, y$  such that  $w = xuy$ . We say that  $u$  is a *prefix* of  $w$  if  $x = \varepsilon$ . Similarly,  $u$  is a *suffix* of  $w$  if  $y = \varepsilon$ . Starting

numbering positions from 0, we let  $w[i]$  denote the character in position  $i$  of  $w$  and let  $w[i..j]$  denote the factor of  $w$  from position  $i$  to position  $j$  (inclusive). We let  $w^m$  denote  $w$  concatenated with itself  $m$  times.

If  $u$  and  $v$  are partial words of equal length over  $\Sigma$ , then  $u$  is *contained* in  $v$ , denoted by  $u \subset v$ , if  $u[i] = v[i]$  for all  $i$  such that  $u[i] \in \Sigma$ ; in other words, the set of non-hole positions in  $u$  is a subset of the set of non-hole positions in  $v$ . Partial words  $u$  and  $v$  are *compatible*, denoted by  $u \uparrow v$ , if there exists a partial word  $w$  such that  $u \subset w$  and  $v \subset w$ , i.e., there are completions  $\hat{u}$  of  $u$  and  $\hat{v}$  of  $v$  such that  $\hat{u} = \hat{v}$ .

The number of occurrences of a letter  $a$  in a partial word  $w$  is denoted by  $|w|_a$ . The *Parikh vector* of  $w$  over  $\Sigma = \{a_0, \dots, a_{k-1}\}$  is defined as  $\psi(w) = \langle |w|_{a_0}, \dots, |w|_{a_{k-1}} \rangle$ . Note that we do not count occurrences of the  $\diamond$  character in Parikh vectors. A partial word  $u$  is *abelian compatible* with a partial word  $v$ , denoted by  $u \hat{\phi} v$ , if a permutation of  $u$  is compatible with  $v$  (this implies that  $u$  and  $v$  are of equal length). For example, the partial words  $ab \diamond bb \diamond$  and  $bb \diamond ab$  are abelian compatible. Note that for total words  $u$  and  $v$  of same length, we have  $u \hat{\phi} v$  if and only if  $\psi(u) = \psi(v)$ , i.e., if  $u$  and  $v$  are permutations of one another (we also say that  $u$  and  $v$  are *abelian equivalent*).

For a non-empty partial word  $w$ , if non-empty partial words  $x_1, x_2, u, v$  exist such that  $w = x_1 u = v x_2$  and  $x_1 \hat{\phi} x_2$ , we call  $w$  *abelian bordered*. In this case, a total word  $x$  exists such that  $x_1 \hat{\phi} x$  and  $x_2 \hat{\phi} x$ , and  $x$  is called an *abelian border* of  $w$  (we also call  $x_1, x_2$  abelian borders of  $w$ ). Two abelian borders  $x, y$  of  $w$  are *equivalent* if and only if  $|x| = |y|$ . We refer to non-equivalent abelian borders as *distinct*. An abelian border  $x$  of  $w$  is *minimal* if  $|x| > |y|$  implies that  $y$  is not an abelian border of  $w$  and  $x$  is *maximal* if  $|x| < |y|$  implies that  $y$  is not an abelian border of  $w$ .

Equivalence of abelian borders being an equivalence relation, we identify abelian borders by their equivalence classes to avoid counting equivalent borders multiple times, but we will not refer to the equivalence classes explicitly. Since we are only interested in identifying and counting distinct abelian borders, we hereafter use the phrase *the abelian border of length  $k$*  to mean the equivalence class of abelian borders of length  $k$ , and we refer to this equivalence class by one of its representatives.

If a total word of length  $n$  has an abelian border of length  $k$ , then it must also have an abelian border of length  $n - k$ . For example, the word  $babbbba$  of length 8 has abelian borders of lengths 2, 3, 4, 5, and 6. Note that equivalent abelian borders of a total word always come in pairs of the same length. Note also that for a total word of length  $n$ , abelian borders come in complementary pairs whose lengths sum to  $n$  (if  $n$  is even, then any abelian border of length  $\frac{n}{2}$  is equivalent to its complementary partner) [10]. However for a partial word of length  $n$ , we may have an abelian border of length greater than  $\lfloor \frac{n}{2} \rfloor$  with no complementary abelian border of shorter length. For example,  $bb \diamond a$  has no abelian borders of length two or less, but it does have the non-abelian border  $bba$  ( $bb \diamond \hat{\phi} bba$  and  $b \diamond a \hat{\phi} bba$ ).

The following proposition shows that all partial words of length at least two without abelian borders are total words.

**Proposition 1.** *If  $w$  is a partial word of length greater than one, with at least one hole, then  $w$  has an abelian border.*

### 3 Computing the minimal abelian border

We now present an algorithm, based on algorithm SHORTEST-ABELIAN-BORDER in [10]. Our algorithm computes the length of the minimal abelian border of a given non-empty partial word  $w$  (over an alphabet of any size) and runs in  $O(n)$  time, where  $n$  is the length of the input  $w$ . We provide the pseudo-code.

Algorithm MINIMAL-ABELIAN-BORDER-PARTIAL works as follows. We can assume without loss of generality that  $w$  is a partial word over the alphabet  $\{0, \dots, \sigma - 1\}$ . We define  $V$  as a vector which gives the difference between the Parikh vectors of a prefix and a suffix of equal length, as defined in [10].

**Algorithm 1** MINIMAL-ABELIAN-BORDER-PARTIAL( $w, n, \sigma$ )

---

**Require:**  $w$ , a partial word with letters in  $\Sigma = \{0, 1, \dots, \sigma - 1\}$ , and  $n \geq 1$ , the length of  $w$

---

```

1:  $V \leftarrow \langle 0, 0, \dots, 0 \rangle$ ,  $|V| = 0$ , and  $holes = 0$ ;
2: if  $w[0] = \diamond$  or  $w[n-1] = \diamond$  then
3:   return 1;
4: else
5:    $V[w[0]] = V[w[0]] + 1$  and  $V[w[n-1]] = V[w[n-1]] - 1$ ;
6:   if  $V[w[0]] = 0$  then
7:     return 1;
8:   else
9:      $|V| = 2$ ;
10:  for  $i \leftarrow 1$  to  $n-1$  do
11:    if  $w[i] = \diamond$  then
12:       $holes = holes + 1$ ;
13:    else
14:       $V[w[i]] = V[w[i]] + 1$ ;
15:      if  $V[w[i]] > 0$  then
16:         $|V| = |V| + 1$ ;
17:      else if  $V[w[i]] \leq 0$  then
18:         $|V| = |V| - 1$ ;
19:      if  $w[n-1-i] = \diamond$  then
20:         $holes = holes + 1$ ;
21:      else
22:         $V[w[n-1-i]] = V[w[n-1-i]] - 1$ ;
23:        if  $V[w[n-1-i]] \geq 0$  then
24:           $|V| = |V| - 1$ ;
25:        else if  $V[w[n-1-i]] < 0$  then
26:           $|V| = |V| + 1$ ;
27:        if  $|V| \leq holes$  then
28:          return  $i + 1$ ;
29:  return  $n$ ;
```

---

We define the magnitude of  $V$ , denoted by  $|V|$ , as the sum of the absolute values of the entries in  $V$ . This variable  $|V|$  represents the “absolute total difference” between two Parikh vectors. The variable  $holes$  keeps track of the number of holes we have encountered so far in  $w$ .

Every time we read a letter from the left side of the partial word, we increment the entry of  $V$  corresponding to that letter, e.g., if we read the letter 3 then we increment  $V[3]$ . Similarly, every time we read a letter on the right side, we decrement the corresponding entry of  $V$ . This process of incrementing and decrementing corresponds to calculating Parikh vectors for the prefix and suffix and then subtracting them. Each time we increment or decrement an entry in  $V$ , we check the new value of that entry and adjust  $|V|$  accordingly. For example, if we increment  $V[w[i]]$  by 1 and  $V[w[i]] \leq 0$  after this action, then  $|V|$  must be decreased by 1. Every time we read a  $\diamond$  instead of a letter, we count it by incrementing  $holes$  but do nothing to  $V$  since holes are not counted in Parikh vectors.

Each time we finish moving “inward” by one position on each side, we check to see if we have found an abelian border. This is equivalent to checking whether the prefix up to the current position is abelian

compatible with the suffix of the same length; this is equivalent to checking whether  $|V|$  is less than or equal to the total number of holes in both the prefix and the suffix.

Unlike the algorithm in [10] for total words, we must run through the entire length of the partial word instead of stopping halfway. If the input partial word has no abelian borders, then we return the length of the partial word. Note that by Proposition 1, this can only occur when the input is a total word.

**Proposition 2.** *Algorithm MINIMAL-ABELIAN-BORDER-PARTIAL computes the minimal abelian border for any partial word of length  $n$  over any finite alphabet and runs in  $O(n)$  time.*

## 4 Abelian borders and Dyck words

We extend a result from [10], which relates minimal abelian borders to Dyck words (see Proposition 3 below). We start with a lemma and a definition.

**Lemma 1.** *A partial word  $w$  has an abelian border of length  $k$  if and only if there exists a completion of  $w^2$  with an abelian border of length  $k$ .*

**Definition 1.** *Let  $w$  be a total word over  $\{0, 1\}$  with  $|w| = n \geq 2$ . Let  $1 \leq k \leq n$ .*

- The  $\mathcal{Y}_k$ -form of  $w$ , denoted by  $\mathcal{Y}_k(w)$ , is a total word over  $\{0, 1, 2\}$  of length  $k$  defined by

$$\mathcal{Y}_k(w)[i] = \begin{cases} 0, & \text{if } w[i] = 0 \text{ and } w[n-1-i] = 1; \\ 1, & \text{if } w[i] = 1 \text{ and } w[n-1-i] = 0; \\ 2, & \text{if } w[i] = w[n-1-i]. \end{cases}$$

- The  $\mathcal{Z}_k$ -form of  $w$ , denoted by  $\mathcal{Z}_k(w)$ , is a total word over  $\{0, 1\}$  defined by removing the twos from  $\mathcal{Y}_k(w)$ .

For example, if  $w = 1010011001$ , then  $\mathcal{Y}_5(w) = 22100$  and  $\mathcal{Z}_5(w) = 100$ . If  $w = 010$ , then  $\mathcal{Y}_3(w) = 222$  and  $\mathcal{Z}_3(w) = \varepsilon$ . Note that  $\mathcal{Z}_k(w)$  is a subsequence of (not necessarily consecutive) letters of  $w$ .

A Dyck word of length  $2n \geq 0$  is a total word over  $\{0, 1\}$  which consists of  $n$  zeros and  $n$  ones arranged so that no prefix of the word has more ones than zeros. We call a  $\mathcal{Z}_k$ -form Dyckian if it is a Dyck word or the bitwise negation of a Dyck word. For a given partial word  $w$  of length  $n$  with  $h$  holes, we let  $D(w)$  denote the set of all  $\mathcal{Z}_k((\hat{w}^2)_i)$  such that  $1 \leq k \leq n$ ,  $0 \leq i \leq 2^h - 1$ ,  $(\hat{w}^2)_i$  is a completion of  $w^2$ , and  $\mathcal{Z}_k((\hat{w}^2)_i)$  is Dyckian. Note that there are  $2^{2h}$  distinct completions of  $w^2$  (since  $w^2$  has  $2h$  holes) and they can be ordered lexicographically. For example, let  $w = 11\circ 0$ . Then the completions of  $w^2 = 11\circ 011\circ 0$  are  $(\hat{w}^2)_0 = 11001100$ ,  $(\hat{w}^2)_1 = 11001110$ ,  $(\hat{w}^2)_2 = 11101100$ ,  $(\hat{w}^2)_3 = 11101110$ . So  $D(w) = \{\mathcal{Z}_4((\hat{w}^2)_0), \mathcal{Z}_3((\hat{w}^2)_1), \mathcal{Z}_4((\hat{w}^2)_3)\}$ . Looking at  $D(w)$ , the following proposition implies that the minimal abelian border of  $w$  has length 3. The proof's main idea comes from [10, Lemma 1].

**Proposition 3.** *A partial word  $w$  over  $\{0, 1\}$  of length at least two has a minimal abelian border of length  $\min\{k \mid \text{there exists } i \text{ such that } \mathcal{Z}_k((\hat{w}^2)_i) \in D(w)\}$ .*

## 5 Abelian borders and lattice paths

We can represent a word over  $\{0, 1\}$  as an ordered pair of paths of equal length on the lattice  $\mathbb{Z}^2$ . The first path corresponds to the prefixes, and the second path corresponds to the suffixes. Both paths begin at the origin (the southwest corner of the lattice). For the prefix, we start with the leftmost letter and step right every time we see a zero and up every time we see a one. For the suffix, we start with the rightmost letter and step in the same manner. Note that both paths are monotonically increasing, also called *minimal*.

**Lemma 2** ([22]). *A total word  $w$  over  $\{0, 1\}$  has an abelian border of length  $k$  if and only if the lattice paths for  $w$  meet after  $k$  steps.*

Figure 1 gives an example of a lattice representation of a word that has abelian borders. If we draw the paths diagonally (taking a step diagonally upward for a zero and diagonally downward for a one) and treat them as discrete functions of the prefix (suffix) length, then we can easily take the absolute difference between these functions, and this difference is precisely  $|V|$  as in Algorithm MINIMAL-ABELIAN-BORDER-PARTIAL. Note that since we are dealing with total words, we have an abelian border of length  $k$  if and only if  $|V| = 0$  after  $k$  steps, and this corresponds to a meeting of the two paths.

We can also think of the graph of  $|V|$  as corresponding to the  $\mathcal{Y}_k$ -form of  $w$ , where a zero gets a step diagonally upward, a one gets a step diagonally downward, and a two gets a step straight forward. Note that this is a *Motzkin path*. Similarly, we can extend this correspondence to  $\mathcal{L}_k$ -forms by removing all of the straight steps from the graph. Just as we saw a relationship between  $\mathcal{L}_k$ -forms, Dyck words, and abelian borders in Proposition 3, here we have that a word has an abelian border of length  $k$  if and only if the  $\mathcal{L}_k$  graph is a *Dyck path*. A Dyck path of length  $2n$  is a sequence of  $n$  upward steps and  $n$  downward steps that starts at the origin and never passes below the horizontal axis. As the name suggests, Dyck paths are coded by Dyck words [13].

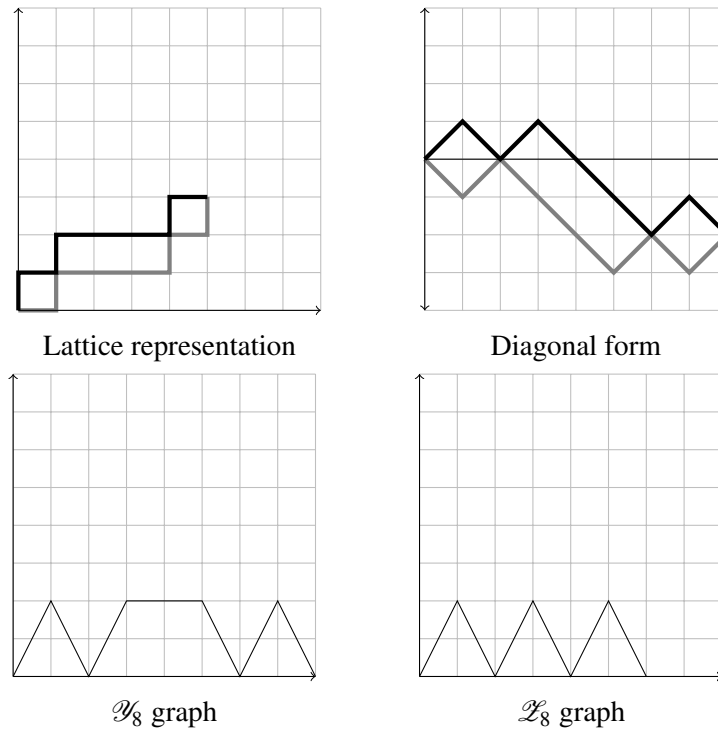


Figure 1: The word  $w = 01000101$  has abelian borders of length 2 and 6;  $\mathcal{Y}_8(w) = 10122010$  and  $\mathcal{L}_8(w) = 101010$ .

### 5.1 The total word case

First, we count binary words of a given length with a given minimal abelian border length. Theorem 4 in [10] gives a bound on that number, but by counting pairs of lattice paths we can find the exact number.

Our formula involves the *Catalan number*,  $C_n = \frac{1}{n+1} \binom{2n}{n}$ , which enumerates Dyck words of length  $2n$ .

**Theorem 1.** *The number of binary total words of length  $n$  with minimal abelian border of length  $k$ ,  $1 \leq k \leq \lfloor \frac{n}{2} \rfloor$ , is  $2^{n-2k+1} C_{k-1}$ .*

*Proof.* The result can easily be checked for  $k = 1$ , so suppose that  $k \geq 2$ . Clearly, we want to count the number of ordered pairs of increasing lattice paths of equal length which meet for the first time after  $k$  steps. For a given rectangle with dimensions  $r \times (n - r)$  on the lattice, the number of ordered pairs of increasing lattice paths which begin at the southwest corner of the rectangle and meet at the northeast corner (without meeting in between) is given by

$$2N_{n-1,r} = \frac{2}{n-1} \binom{n-1}{r} \binom{n-1}{r-1}$$

where  $N_{n,r} = \frac{1}{n} \binom{n}{r} \binom{n}{r-1}$  is the *Narayana number* [16].

For an abelian border of length  $k$ , we want to find all such paths for all rectangles such that the southwest corner is at the origin and the perimeter of the rectangle is  $2k$ . It is easy to see that there are  $k - 1$  such rectangles. For each rectangle, we calculate the number of permissible pairs of paths, giving a total of

$$\sum_{r=1}^{k-1} \frac{2}{k-1} \binom{k-1}{r} \binom{k-1}{r-1}.$$

There is a relationship between the Narayana numbers and the Catalan numbers, namely

$$\sum_{r=1}^n N_{n,r} = \sum_{r=1}^n \frac{1}{n} \binom{n}{r} \binom{n}{r-1} = \frac{1}{n+1} \binom{2n}{n} = C_n, \quad (1)$$

which can be seen by the fact that  $N_{n,r}$  counts the number of Dyck paths of length  $2n$  that have  $r$  peaks [19]. Using Equation (1), we get

$$\sum_{r=1}^{k-1} \frac{2}{k-1} \binom{k-1}{r} \binom{k-1}{r-1} = 2C_{k-1}.$$

Since we do not care what happens after the minimal abelian border, every position numbered from  $k$  to  $n - k - 1$  can either be one of two distinct letters, so we also need to count each of these possibilities. There are  $2^{n-2k}$  such configurations possible, so for the total number of binary words of length  $n$  with minimal abelian border of length  $k$  we arrive at  $2^{n-2k+1} C_{k-1}$ .  $\square$

Second, we can also use lattice paths to find the exact number of binary total words of a given length that have no abelian borders. This result was proved by Rampersad et al. in [22] using a similar method with Motzkin paths.

**Theorem 2.** *The number of binary partial words of length  $n$  with no abelian borders is equal to the number of binary total words of length  $n$  with no abelian borders, which is  $\binom{n}{\frac{n}{2}}$  when  $n$  is even, and  $2 \binom{n-1}{\frac{n-1}{2}}$  when  $n$  is odd.*

Third, we use lattice path methods to arrive at a more general formula for counting binary total words of a given length with a given number of distinct abelian borders. As was the case with Theorem 1, the following theorem uses formulas for counting the number of ordered pairs of lattice paths that intersect in particular ways [16]. Note that when  $m = 0$ , this reduces to Theorem 2.

**Theorem 3.** *The number of binary total words of length  $n$  with  $m$  distinct abelian borders is*

$$\begin{cases} 2^{\frac{m}{2}+1} \frac{\binom{n-\frac{m}{2}-1}{\frac{n}{2}}!}{\binom{n-m-1}{\frac{n-m}{2}}!}, & \text{if } n \text{ and } m \text{ are both even;} \\ 2^{\frac{m-1}{2}} (m+1) \frac{\binom{n-\frac{m+3}{2}}{\frac{n}{2}}!}{\binom{n-m-1}{\frac{n-m}{2}}!}, & \text{if } n \text{ is even and } m \text{ is odd;} \\ 2^{\frac{m}{2}+1} \binom{n-\frac{m}{2}-1}{\frac{n-1}{2}}, & \text{if } n \text{ is odd and } m \text{ is even;} \\ 0, & \text{if } n \text{ and } m \text{ are both odd.} \end{cases}$$

*Proof.* When  $n$  and  $m$  are even, the number of binary words of length  $n$  with  $m$  distinct abelian borders is precisely the number of ordered pairs of minimal lattice paths of length  $\frac{n}{2}$  which start at the origin, intersect exactly  $\frac{m}{2}$  times, and end at different points. By [16], this is equal to

$$\begin{aligned} & \left( \# \text{ of such paths that intersect } \frac{m}{2} \text{ times and end anywhere} \right) \\ & - \left( \# \text{ of such paths that intersect } \frac{m}{2} - 1 \text{ times and end at the same point} \right). \end{aligned}$$

If  $p = \frac{n}{2}$  and  $q = \frac{m}{2}$ , then this quantity is

$$2^q \binom{2p-q}{p} - q2^q \frac{(2p-q-1)!}{p!(p-q)!}. \tag{2}$$

By substituting our fractions into Equation (2) and simplifying, we get our result.

When  $n$  is even and  $m$  is odd, first recall that for total words, abelian borders come in complementary pairs whose lengths sum to  $n$ . The only abelian border length that does not have a complementary partner which is distinct is the abelian border of length  $\frac{n}{2}$ , so the only way a word can have an odd number  $m$  of abelian borders is if it has an abelian border of this length. This corresponds to pairs of minimal lattice paths of length  $\frac{n}{2}$  which begin at the origin, end at the same point, and intersect  $\frac{m-1}{2}$  times (not including the origin or the end point). This is given by

$$2^{\frac{m-1}{2}} (m+1) \frac{\binom{n-\frac{m+3}{2}}{\frac{n}{2}}!}{\binom{n-m-1}{\frac{n-m}{2}}!}.$$

When  $n$  is odd and  $m$  is even, note that if a given word of length  $n$  has prefix and suffix lattice paths which after  $\lfloor \frac{n}{2} \rfloor$  steps have intersected  $r$  times, then the word has a total of  $2r$  distinct abelian borders. Therefore, for a word with  $m$  distinct abelian borders, the number of options for the first  $\lfloor \frac{n}{2} \rfloor$  steps is

$$2^{\frac{m}{2}} \binom{n-\frac{m}{2}-1}{\frac{n-1}{2}}.$$

Since the first  $\lfloor \frac{n}{2} \rfloor$  steps determine all but one of the letters in the word, and the middle letter does not affect the number of abelian borders, the total number of binary words of length  $n$  with  $m$  abelian borders is twice the above quantity. Finally, when  $n$  and  $m$  are odd, it follows immediately from the above that an odd number of abelian borders can only occur when the length is even. Therefore, there are no words of odd length having an odd number of distinct abelian borders.  $\square$



## 5.2 The partial word case

We generalize Theorem 1 and Theorem 3 to partial words. First, we count the number of partial words of a given length with a given minimal abelian border length.

**Theorem 4.** *The number of partial words of length  $n$  with  $h$  holes over an alphabet of size  $\sigma$  with minimal abelian border of length  $k$ ,  $n \geq 2k$ , is*

$$\sum_{j=0}^h g_{\sigma}(k, j) \binom{n-2k}{h-j} \sigma^{n-2k-h+j},$$

where  $g_{\sigma}(k, j)$  is the number of words of length  $2k$  over an alphabet of size  $\sigma$  with  $j$  holes and a minimal abelian border of length  $k$ .

*Proof.* For  $0 \leq j \leq h$  there are  $g_{\sigma}(k, j)$  choices for the prefix and suffix of length  $k$  containing a total of  $j$  holes, and there are  $n - 2k$  remaining positions in the partial word, exactly  $h - j$  of which must be holes. For each position that is a letter, there are  $\sigma$  possibilities, and there are  $n - 2k - h + j$  such positions.  $\square$

In the  $k = 1$  case, we can compute the  $g_{\sigma}(k, j)$ 's. For  $h = 1$ , the number of partial words of length  $n$  with  $h$  holes over an alphabet of size  $\sigma$  with minimal abelian border of length one, where  $n \geq 2$ , is

$$\sum_{j=0}^1 g_{\sigma}(1, j) \binom{n-2}{1-j} \sigma^{n+j-3},$$

where  $g_{\sigma}(1, 0) = \sigma$  and  $g_{\sigma}(1, 1) = 2\sigma$ . For  $h \geq 2$ , the number of partial words of length  $n$  with  $h$  holes over an alphabet of size  $\sigma$  with minimal abelian border of length one, where  $n \geq 2$ , is

$$\sum_{j=0}^h g_{\sigma}(1, j) \binom{n-2}{h-j} \sigma^{n-h+j-2},$$

where  $g_{\sigma}(1, 0) = \sigma$ ,  $g_{\sigma}(1, 1) = 2\sigma$ ,  $g_{\sigma}(1, 2) = 1$ , and  $g_{\sigma}(1, j) = 0$  for  $j \in [3..h]$ . In the  $\sigma = 2$  case, we can compute the  $g_{\sigma}(k, j)$ 's using the next corollary. To illustrate the proof, the total words over  $\{0, 1\}$ , up to renaming of letters and reversal, of length  $2k = 10$  with minimal abelian border length  $k = 5$  are

$$\begin{array}{c} 0000\underline{10000}1, 0001\underline{1000}11, 0001\underline{10010}1, 0001\underline{10100}1, \\ 0010\underline{1000}11, 0010\underline{10010}1, 0011\underline{10110}1, \end{array}$$

the underlined positions are the  $k$  ways to insert a hole while preserving the minimal abelian border.

**Corollary 1.** *For  $k \geq 2$  and  $0 \leq j \leq k$ , the equality  $g_2(k, j) = 2 \binom{k}{j} C_{k-1}$  holds. Thus for  $k \geq 2$ , the number of binary partial words with  $h$  holes of length  $n$ ,  $n \geq 2k$ , with minimal abelian border of length  $k$  is*

$$f(n, k, h) = 2^{n-2k+1} C_{k-1} \sum_{j=0}^h \binom{k}{j} \binom{n-2k}{h-j} 2^{j-h}.$$

Second, we count the number of partial words of a given length with a given number of distinct abelian borders. In the proof of Proposition 1, we saw that a partial word of length  $n$  with one hole must have at least an abelian border of length  $n - 1$ . By induction on  $h$ , if we insert a hole into a partial word with  $h - 1$  holes, we are guaranteed to get an additional abelian border of length  $n - h$ .

**Theorem 5.** *A partial word  $w$  of length  $n$  with  $h$  holes,  $0 \leq h < n$ , has at least  $h$  distinct abelian borders. In particular,  $w$  has at least abelian borders of lengths  $n - 1, n - 2, \dots, n - h$ .*

We next focus on the one-hole case. Experimental data suggest some elegant formulas in this case.

**Theorem 6.** *If  $n$  is even, then the number of binary partial words of length  $n$  with exactly one hole and  $n - 1$  distinct abelian borders is  $4(3^{n/2} - 2^{n/2})$ .*

*Proof.* It is clear that for  $n$  even and  $m = n - 1$ , each binary partial word of length  $n$  with exactly one hole and  $m$  distinct abelian borders generates, through renaming of letters and reversal, three other partial words with the same properties. So to get the number of partial words satisfying the desired properties, we need to find the number of “unique” partial words satisfying them and multiply by 4; “unique” is defined in terms of the hole being in the first half of the partial word and the first non-hole letter being 0. The first few values are given in Table 1. For example, for  $n = 6$ , there are 9 unique partial words with  $\diamond$  in position 0, 6 with  $\diamond$  in position 1, and 4 with  $\diamond$  in position 2, for a total of 19.

length \ $i$	0	1	2	...	$\frac{n-2}{2}$
2	1				
4	3	2			
6	9	6	4		
8	27	18	12		
10	81	54	36		
$\vdots$	$\vdots$	$\vdots$	$\vdots$		
$n$	$2^0 \times 3^{(\frac{n}{2}-1)}$	$2^1 \times 3^{(\frac{n}{2}-2)}$	$2^2 \times 3^{(\frac{n}{2}-3)}$	...	$2^{(\frac{n}{2}-1)} \times 3^0$

Table 1: Number of unique binary partial words of even length  $n$  with one hole in position  $i$  and with  $n - 1$  distinct abelian borders.

Thus, the number of unique partial words  $N$  for a given  $n$  is

$$N = 2^0 \times 3^{(\frac{n}{2}-1)} + 2^1 \times 3^{(\frac{n}{2}-2)} + 2^2 \times 3^{(\frac{n}{2}-3)} + \dots + 2^{(\frac{n}{2}-1)} \times 3^0.$$

This is a geometric series with ratio  $\frac{2}{3}$ , so  $N = 3^{n/2} - 2^{n/2}$ .

For  $n = 2$ , there is one unique partial word. For  $n = 4$ , there are three unique partial words with  $\diamond$  in position 0 and two with  $\diamond$  in position 1. Then to build the partial words of length  $n + 2$  from those of length  $n$ , when  $n \geq 4$ , we proceed as follows. Each partial word  $a_0 \cdots a_{\frac{n-2}{2}} a_{\frac{n}{2}} \cdots a_{n-1}$  of length  $n$  gives rise to a partial word of length  $n + 2$  by inserting a pair of suitable letters between  $a_{\frac{n-2}{2}}$  and  $a_{\frac{n}{2}}$ , to obtain three partial words of the form  $a_0 \cdots a_{\frac{n-2}{2}} aba_{\frac{n}{2}} \cdots a_{n-1}$ . Then there are  $2^{\frac{n}{2}}$  additional unique partial words of length  $n + 2$  with  $\diamond$  in position  $\frac{n}{2}$ .  $\square$

It is clear that each binary partial word of even length  $n$  with exactly one hole having a minimal abelian border of length  $n - 1$  generates, through renaming of letters and reversal, three other partial words with the same properties. So to get the number of partial words satisfying the desired properties, we need to find the number of “unique” partial words satisfying them and multiply by 4; “unique” is defined in terms of the hole being in the first half of the partial word and the first non-hole letter being 0.

**Lemma 3.** *If a binary partial word of even length  $n$ ,  $n \geq 4$ , with exactly one hole in the first half having a minimal abelian border of length  $n - 1$  starts with 0, then it ends with 11. Therefore if such  $w = uv$ , with  $|u| = |v|$ , is unique, then  $|u|_0 > |v|_0$  and  $|u|_1 < |v|_1 - 1$ .*

**Lemma 4.** *The number of unique binary partial words of even length  $n$ ,  $n \geq 4$ , with exactly one hole in the first half and no 1 in the first half having a minimal abelian border of length  $n - 1$  is  $(\frac{n}{2} - 1)2^{(\frac{n}{2}-2)}$ .*

**Lemma 5.** *The number of unique binary partial words of even length  $n$ ,  $n \geq 4$ , with exactly one hole in the first half and exactly one 0 in the first half having a minimal abelian border of length  $n - 1$  is 1; such partial word is of the form  $0 \diamond 1^{n-2}$ .*

**Proposition 4.** *The number of unique binary partial words of even length  $n$ ,  $n \geq 6$ , with exactly one hole in the first half and exactly two 0's in the first half having a minimal abelian border of length  $n - 1$  is  $\frac{n^2}{4} - \frac{n}{2} - 2$ .*

*Proof.* We proceed by induction. For the basis  $n = 6$ , the unique partial words are  $0 \diamond 0011$ ,  $0 \diamond 0111$ ,  $00 \diamond 011$ ,  $00 \diamond 111$  and the result holds. For the inductive step, we show how to build the unique partial words  $w' = a_0 \cdots a_{\frac{n}{2}-1} ab a_{\frac{n}{2}} \cdots a_{n-1}$  of length  $n + 2$  with minimal abelian border of length  $(n + 2) - 1$  and exactly two 0's in the first half, where  $ab \in \{00, 01, 10, 11\}$ , from the unique partial words  $w = a_0 \cdots a_{\frac{n}{2}-1} a_{\frac{n}{2}} \cdots a_{n-1}$  of length  $n$  with minimal abelian border of length  $n - 1$  but not necessarily with exactly two 0's in the first half (there could be only one 0 in the first half).

By Lemma 3, the first half of  $w$ , say  $u$ , has more 0's than the second half, say  $v$ . If we start with a unique partial word  $w$  with  $|u|_0 = 2$ , then we obtain a unique partial word  $w'$  that satisfies our desired properties by inserting the pair  $ab = 11$  of letters between  $a_{\frac{n}{2}-1}$  and  $a_{\frac{n}{2}}$ . There are  $\frac{n^2}{4} - \frac{n}{2} - 2$  such  $w$ 's by the inductive hypothesis. For some of these  $w$ 's, it is also possible to insert the pair  $ab = 10$ , i.e., when  $|v|_0 = 0$ . There are  $n - 4$  such  $w$ 's resulting in  $n - 4$  unique  $w'$  with 10 in the middle. Indeed when  $n = 6$ , the  $n - 4 = 2$  unique  $w'$  are  $0 \diamond 010111$  and  $00 \diamond 10111$ . When  $n > 6$ , we first claim that  $w'[\frac{n}{2} + 2..n + 1] = 1^{\frac{n}{2}}$ . Indeed by Lemma 3, the number of 0's in  $w'[\frac{n}{2} + 1..n + 1]$  is zero or one, so it is one in this case. We now claim that  $w'[1] \neq 1$ ; otherwise,  $w'$  has an abelian border of length  $(n + 2) - 2 < (n + 2) - 1$ . If  $w'[0..1] = 0 \diamond$ , then there are  $\frac{n}{2} - 2$  positions left in the first half to position the other 0, while if  $w'[0..1] = 00$ , then there are  $\frac{n}{2} - 2$  positions left in the first half to position the  $\diamond$ .

If we start with a unique partial word  $w$  with  $|u|_0 = 1$ , then  $w = 0 \diamond 1^{n-2}$  by Lemma 5. We obtain a unique partial word  $w'$  that satisfies our desired properties by inserting  $ab = 00$  or  $ab = 01$ .

There are also two unique  $w'$  with a  $\diamond$  in position  $\frac{n}{2}$ . When  $n = 6$ , the two unique  $w'$  are  $001 \diamond 0111$  and  $001 \diamond 1111$ . When  $n > 6$ , we first claim that  $w'[\frac{n}{2} + 2..n + 1] = 1^{\frac{n}{2}}$ . Indeed by Lemma 3, the number of 0's in  $w'[\frac{n}{2} + 1..n + 1]$  is zero or one. If it is one and that 0 appears in  $w'[\frac{n}{2} + 2..n + 1]$ , then  $w'$  has an abelian border of length  $\frac{n}{2} + 2 < (n + 2) - 1$ , implying a contradiction. We now claim that  $w'[1] = 0$ ; otherwise,  $w'$  has an abelian border of length  $(n + 2) - 2 < (n + 2) - 1$ . So the two unique  $w'$  are  $001^{\frac{n}{2}-2} \diamond 01^{\frac{n}{2}}$  and  $001^{\frac{n}{2}-2} \diamond 11^{\frac{n}{2}}$ . Altogether, there are  $\frac{n^2}{4} - \frac{n}{2} - 2 + n - 4 + 2 + 2 = \frac{(n+2)^2}{4} - \frac{n+2}{2} - 2$  unique  $w'$ .  $\square$

**Proposition 5.** *The number of unique binary partial words of even length  $n$ ,  $n \geq 6$ , with exactly one hole in the first half and exactly one 1 in the first half having a minimal abelian border of length  $n - 1$  is  $(\frac{n}{2} - 2)(2^{\frac{n}{2}-2})(\frac{n}{2} - 3) + 1$ .*

*Proof.* We proceed by induction. For the basis  $n = 6$ , the unique partial word is  $0 \diamond 1111$  by Lemma 5 and the result holds. For the inductive step, we show how to build the unique partial words  $w' = a_0 \cdots a_{\frac{n}{2}-1} ab a_{\frac{n}{2}} \cdots a_{n-1}$  of length  $n + 2$  with minimal abelian border of length  $(n + 2) - 1$  and exactly one 1 in the first half, where  $ab \in \{00, 01, 10, 11\}$ , from the unique partial words  $w = a_0 \cdots a_{\frac{n}{2}-1} a_{\frac{n}{2}} \cdots a_{n-1}$  of length  $n$  with minimal abelian border of length  $n - 1$  but not necessarily with exactly one 1 in the first half (there could be no 1 in the first half).

Let  $u$  be the first half of  $w$  and  $v$  be the second half. If we start with a unique partial word  $w$  with  $|u|_1 = 1$ , then we obtain a unique partial word  $w'$  that satisfies our desired properties by inserting the pair  $ab = 00$  or  $ab = 01$  of letters between  $a_{\frac{n}{2}-1}$  and  $a_{\frac{n}{2}}$ . There are  $(\frac{n}{2} - 2)(2^{\frac{n}{2}-2})(\frac{n}{2} - 3) + 1$  such  $w$ 's by the inductive hypothesis. If we start with a unique partial word  $w$  with  $|u|_1 = 0$ , then there are  $(\frac{n}{2} - 1)2^{\frac{n}{2}-2}$

such  $w$ 's by Lemma 4. We obtain a unique partial word  $w'$  that satisfies our desired properties by inserting  $ab = 11$ . For some of those  $w$ 's, we can also insert  $ab = 10$ . To see how many such  $w'$ 's there are, the  $\diamond$  can be in any position in  $\{1, \dots, \frac{n}{2} - 1\}$ . The last two positions are 11. There are  $2^{\binom{n}{2}-2} - 1$  choices for the  $\frac{n}{2} - 2$  positions after the middle 10. The  $-1$  comes from the fact that these positions cannot be all 0's (otherwise, there would be an abelian border of length  $\frac{n}{2} + 1 < (n+2) - 1$ ).

There are also  $(\frac{n}{2} - 3)2^{\binom{n}{2}-1} + 2$  unique  $w'$  with a  $\diamond$  in position  $\frac{n}{2}$ . We first claim that  $w'[1] = 0$ ; otherwise,  $w'$  has an abelian border of length  $(n+2) - 2 < (n+2) - 1$ . We next claim that for  $2 \leq i < \frac{n}{2}$ , if  $w'[0..i] = 0^i 1$  then  $w'[n-i+1..n-1] \neq 0^{i-1}$ ; otherwise,  $w'$  has an abelian border of length  $(n+2) - (i+1) < (n+2) - 1$ . So if the 1 in the first half is in position 2, there are  $(2^1 - 1)2^{\binom{n}{2}-2}$  unique corresponding  $w'$ , if the 1 in the first half is in position 3, there are  $(2^2 - 1)2^{\binom{n}{2}-3}$  unique corresponding  $w'$ ,  $\dots$ , and if the 1 in the first half is in position  $\frac{n}{2} - 1$ , there are  $(2^{\binom{n}{2}-2} - 1)2^1$  unique corresponding  $w'$ . So, we get a total of  $(\frac{n}{2} - 3)2^{\binom{n}{2}-1} + 2$  unique  $w'$  with a  $\diamond$  in position  $\frac{n}{2}$ .  $\square$

**Theorem 7.** *The number of binary partial words of even length  $n$  with exactly one hole having a minimal abelian border of length  $n - 1$  is  $\Theta(2^{n-1})$ .*

*Proof.* Let  $f(n, n-1, 1)$  denote the number of binary partial words of even length  $n$  with exactly one hole having a minimal abelian border of length  $n - 1$ . For  $n = 2$ , there is one unique partial word with  $\diamond$  in position 0. For  $n = 4$ , there is one unique partial word with  $\diamond$  in position 1 by Lemma 4 or Lemma 5. Then, for  $n \geq 4$ , to build the unique partial words of length  $n + 2$  with minimal abelian border of length  $(n+2) - 1$  from those of length  $n$  with minimal abelian border of length  $n - 1$ , we proceed as follows. Each unique binary partial word  $w = a_0 \cdots a_{\frac{n}{2}-1} a_{\frac{n}{2}} \cdots a_{n-1}$  of even length  $n$  with exactly one hole having a minimal abelian border of length  $n - 1$  gives rise to a partial word of length  $n + 2$  by inserting a pair  $ab$  of letters,  $ab \in \{00, 01, 11\}$ , between  $a_{\frac{n}{2}-1}$  and  $a_{\frac{n}{2}}$ , to obtain three unique partial words of the form  $w' = a_0 \cdots a_{\frac{n}{2}-1} ab a_{\frac{n}{2}} \cdots a_{n-1}$ . For some  $w$ 's, it is sometimes also possible to insert the pair  $ab = 10$ .

To see this, let  $w$  be as above. By our convention,  $w$  starts with 0, and by Lemma 3,  $w$  ends with 11. Since  $w$  has no abelian border of any length in  $\{1, \dots, \frac{n}{2}\}$ ,  $w'$  has no such abelian border. Clearly,  $w'$  has also no abelian border of any length in  $\{\frac{n}{2} + 2, \dots, n\}$ . If  $ab \in \{00, 11\}$ , then  $a_0 \cdots a_{\frac{n}{2}-1} a$  is not abelian compatible with  $ba_{\frac{n}{2}} \cdots a_{n-1}$  and such  $w'$  has no abelian border of length  $\frac{n}{2} + 1$ . If  $ab = 01$ , then  $a_0 \cdots a_{\frac{n}{2}-1} a$  is not abelian compatible with  $ba_{\frac{n}{2}} \cdots a_{n-1}$  since the number of 0's in  $a_0 \cdots a_{\frac{n}{2}-1}$  is greater than the number of 0's in  $a_{\frac{n}{2}} \cdots a_{n-1}$  by Lemma 3, and so such  $w'$  has no abelian border of length  $\frac{n}{2} + 1$ .

Then there are  $2^{\binom{n}{2}-2} + 2^{n-4}$  additional unique partial words  $w'$  of length  $n + 2$  with  $\diamond$  in position  $\frac{n}{2}$ . To see this, we first claim that  $w'[1] = 0$ ; otherwise,  $w'$  has an abelian border of length  $(n+2) - 2 < (n+2) - 1$ . So we can write  $w' = 00u \diamond v 11$ , where  $u$  and  $v$  are total words of length  $\frac{n}{2} - 2$  and  $\frac{n}{2} - 1$ , respectively. There are  $2^{\frac{n}{2}-2}$  possibilities for  $u$ , ranging from  $u = 1^{\frac{n}{2}-2}$  to  $u = 0^{\frac{n}{2}-2}$ . On the one hand, by the proof of Proposition 4, if  $u = 1^{\frac{n}{2}-2}$  then there are 2 unique partial words of length  $n + 2$  with  $\diamond$  in position  $\frac{n}{2}$ , i.e.,  $001^{\frac{n}{2}-2} \diamond 01^{\frac{n}{2}}$  and  $001^{\frac{n}{2}-2} \diamond 11^{\frac{n}{2}}$ . On the other hand if  $u = 0^{\frac{n}{2}-2}$ , there are  $2^{\frac{n}{2}-1}$  unique such partial words. Altogether, there are  $2 + 4 + 6 + \dots + 2^{\binom{n}{2}-1}$  unique partial words of length  $n + 2$  with  $\diamond$  in position  $\frac{n}{2}$ , for a total of  $2(1 + 2 + 3 + \dots + 2^{\binom{n}{2}-2}) = 2^{\binom{n}{2}-2} + 2^{n-4}$ .

Thus, the number of unique partial words with exactly one hole of length  $n + 2$  having a minimal abelian border of length  $(n+2) - 1$  satisfies

$$\frac{3f(n, n-1, 1)}{4} + 2^{\binom{n}{2}-2} + 2^{n-4} < \frac{f(n+2, (n+2)-1, 1)}{4} < \frac{4f(n, n-1, 1)}{4} + 2^{\binom{n}{2}-2} + 2^{n-4}.$$

From this we deduce that, for  $n$  sufficiently large, if  $f(n, n-1, 1)$  is  $O(2^{n-1})$  then  $f(n+2, (n+2)-1, 1)$  is  $O(2^{(n+2)-1})$ , and if  $f(n, n-1, 1)$  is  $\Omega(2^{n-1})$  then  $f(n+2, (n+2)-1, 1)$  is  $\Omega(2^{(n+2)-1})$ .  $\square$

Theorem 7 equivalently states that the number of binary partial words of even length  $n$  with exactly one hole and one abelian border is  $\Theta(2^{n-1})$ . This is implied by Theorem 5 that guarantees at least a minimal border of length  $n - 1$  when a partial word has at least one hole. We can prove that for  $n$  sufficiently large, the number of binary partial words of even length  $n$  with exactly one hole and  $m$ ,  $m > 1$  odd, distinct abelian borders is bounded from above by the number of binary partial words of length  $n$  with exactly one hole and one abelian border. The proof is based on the following lemmas.

First, we consider unique binary partial words of even length  $n$  with exactly one hole in position 0 and  $m$  distinct abelian borders.

**Lemma 6.** *Let  $m > 1$  be odd. If a binary partial word of even length  $n$ ,  $n \geq 4$ , has exactly one hole in position 0 and  $m$  distinct abelian borders, then three of these abelian borders have lengths  $1, \frac{n}{2}, n - 1$ . Furthermore, for  $k \in \{1, \dots, \frac{n}{2} - 1\}$ , we have that  $k$  is an abelian border length if and only if  $n - k$  is an abelian border length.*

Similarly, we can show that for odd  $m$ , if a binary partial word of odd length  $n$ ,  $n \geq 3$ , has exactly one hole and  $m$  distinct abelian borders, then the hole is not in position 0.

**Lemma 7.** *Let  $m > 1$  be odd. If a partial word  $w = a_0 \cdots a_{\frac{n}{2}-1} a_{\frac{n}{2}} \cdots a_{n-1}$  of even length  $n$ ,  $n \geq 6$ , over the alphabet  $\{0, 1\}$  has exactly one hole in position 0 and  $m$  distinct abelian borders, then for  $j \in \{2, \dots, \frac{n}{2} - 1\}$  and  $ab \in \{00, 11\}$ , the partial word  $w' = a_0 \cdots a_{j-1} aa_j \cdots a_{\frac{n}{2}-1} a_{\frac{n}{2}} \cdots a_{n-j-1} ba_{n-j} \cdots a_{n-1}$  has exactly  $m$  distinct abelian borders.*

Now, we consider unique binary partial words of even length  $n$  with exactly one hole in position 1 and  $m$  distinct abelian borders.

**Lemma 8.** *Let  $m$  be odd. If a binary partial word of even length  $n$ ,  $n \geq 6$ , has exactly one hole in position 1 and  $m$  distinct abelian borders, then  $n - 1$  is an abelian border length and for  $k \in \{2, \dots, \frac{n}{2} - 1\}$  we have that  $k$  is an abelian border length if and only if  $n - k$  is an abelian border length. Consequently, either both  $1, \frac{n}{2}$  are abelian border lengths or none is an abelian border length.*

If  $m = 3$  in Lemma 8, then an abelian border of length 1 would imply an abelian border of length 2 (and also one of length  $n - 2$ ), so there would be at least four distinct border lengths, a contradiction. So neither 1 nor  $\frac{n}{2}$  is an abelian border length in the  $m = 3$  case when the position of the hole is 1.

**Lemma 9.** *Let  $m$  be odd. If a partial word  $w = a_0 \cdots a_{\frac{n}{2}-1} a_{\frac{n}{2}} \cdots a_{n-1}$  of even length  $n$ ,  $n \geq 6$ , over  $\{0, 1\}$  has exactly one hole in position 1 and  $m$  distinct abelian borders, then for at least three pairs  $ab$  in  $\{00, 01, 10, 11\}$ , the partial word  $w' = a_0 \cdots a_{\frac{n}{2}-1} aba_{\frac{n}{2}} \cdots a_{n-1}$  has exactly  $m$  distinct abelian borders.*

Next, we consider unique binary partial words of even length  $n$  with exactly one hole in position  $i$ , where  $2 \leq i \leq \frac{n}{2} - 1$ , and  $m$  distinct abelian borders.

**Lemma 10.** *Let  $m > 1$  be odd. If a binary partial word of even length  $n$ ,  $n \geq 8$ , has exactly one hole in position  $i$ ,  $2 \leq i \leq \frac{n}{2} - 1$ , and  $m$  distinct abelian borders, one whose length is 1, then two of these abelian borders have lengths  $n - 2, n - 1$ . Furthermore, for  $k \in \{i + 1, \dots, \frac{n}{2} - 1\}$ , we have that  $k$  is an abelian border length if and only if  $n - k$  is an abelian border length.*

**Lemma 11.** *Let  $m > 1$  be odd. If a partial word  $w = a_0 \cdots a_{\frac{n}{2}-1} a_{\frac{n}{2}} \cdots a_{n-1}$  of even length  $n$ ,  $n \geq 8$ , over the alphabet  $\{0, 1\}$  has exactly one hole in position  $i$ ,  $2 \leq i \leq \frac{n}{2} - 1$ , and  $m$  distinct abelian borders, one whose length is 1, then for at least three pairs  $ab$  in  $\{00, 01, 10, 11\}$ , the partial word  $w' = a_0 \cdots a_{\frac{n}{2}-1} aba_{\frac{n}{2}} \cdots a_{n-1}$  has exactly  $m$  distinct abelian borders.*

The  $m = 1$  case of the next theorem follows from Theorem 7. The general  $m > 1$  odd case is similar to the  $m = 1$  case, the only difference is that we can have some more trivial values. We can prove the result based on the location of the hole, inserting three or four pairs in  $\{00, 01, 10, 11\}$  in suitable positions of partial words of even length  $n$  to get partial words of length  $n + 2$ , and finding the dominant term.

**Theorem 8.** *If  $m$  is odd, then the number of binary partial words of length  $n$  with exactly one hole and  $m$  distinct abelian borders is  $\Theta(2^{n-1})$ .*

## 6 Conclusion

Computer experiments strongly suggest that the following conjectures are true. The integer  $A191386(n)$  in Sloane's Online Encyclopedia of Integer Sequences (OEIS) gives the number of ascents of length 1 in all dispersed Dyck paths of length  $n$ . A *dispersed Dyck path* is one or more Dyck paths connected by one or more horizontal steps [8], or equivalently a Motzkin path of length  $n$  with no horizontal steps at positive heights. An *ascent* is a maximal sequence of upward steps. The first few values of this sequence (starting at  $n = 0$ ) are 0, 0, 1, 2, 5, 10, 23, 46, 102, 204, 443, 886, 1898, 3796, 8054.

**Conjecture 1.** *The number of binary partial words of length  $n$  with exactly one hole having a minimal abelian border of length  $k < n$  is*

$$f(n, k, 1) = \begin{cases} A191386(n), & \text{if } n \geq 3 \text{ is odd and } k = n - 1; \\ 2A191386(n - 1), & \text{if } n \geq 4 \text{ is even and } k = n - 1; \\ 2A191386(n - 2), & \text{if } n \geq 5 \text{ is odd and } k = n - 2; \\ 4A191386(n - 3), & \text{if } n \geq 6 \text{ is even and } k = n - 2. \end{cases}$$

Note that Theorem 5 suggests that Conjecture 1 is true if and only if the number of binary partial words of length  $n$  with exactly one hole and exactly one abelian border is  $A191386(n)$  when  $n$  is odd and  $2A191386(n - 1)$  when  $n$  is even.

**Conjecture 2.** *The set of binary total words without abelian borders is non-context-free.*

## Acknowledgements

Project sponsored by the National Security Agency under Grant Number H98230-15-1-0232. The United States Government is authorized to reproduce and distribute reprints notwithstanding any copyright notation herein. This manuscript is submitted for publication with the understanding that the United States Government is authorized to reproduce and distribute reprints. This material is based upon work supported by the National Science Foundation under Grant No. DMS-1060775. We thank Benjamin De Winkle for his very useful comments and suggestions. We also thank the referees of preliminary versions of this paper for their very valuable comments and suggestions.

## References

- [1] S. V. Avgustinovich, J. Karhumäki & S. Puzynina (2012): *On Abelian versions of the critical factorization theorem. RAIRO-Theoretical Informatics and Applications*, 46(1), pp. 3–15, doi:10.1051/ita/2011121.
- [2] F. Blanchet-Sadri, S. Simmons, A. Tebbe & A. Veprauskas (2013): *Abelian periods, partial words, and an extension of a theorem of Fine and Wilf. RAIRO-Theoretical Informatics and Applications*, 47, pp. 215–234, doi:10.1051/ita/2013034.
- [3] S. Böcker (2004): *Sequencing from compomers: Using mass spectrometry for DNA de novo sequencing of 200+ nt. Journal of Computational Biology*, 11(6), pp. 1110–1134, doi:10.1089/cmb.2004.11.1110.
- [4] P. Burcsi, F. Cicalese, G. Fici & Z. Lipták (2012): *Algorithms for jumbled pattern matching in strings. International Journal of Foundations of Computer Science*, 23(2), pp. 357–374, doi:10.1142/S0129054112400175.

- [5] P. Burcsi, F. Cicalese, G. Fici & Z. Lipták (2012): *On approximate jumbled pattern matching in strings*. *Theory of Computing Systems*, 50(1), pp. 35–51, doi:10.1007/s00224-011-9344-5.
- [6] A. Butman, R. Eres & G.M. Landau (2004): *Scaled and permuted string matching*. *Information Processing Letters*, 92(6), pp. 293–297, doi:10.1016/j.ipl.2004.09.002.
- [7] E. Charlier, T. Harju, S. Puzynina & L. Q. Zamboni (2015): *Abelian bordered factors and periodicity*. arXiv:1501.07464v1 [cs.FL] 29 Jan 2015.
- [8] W. Y. C. Chen & P. L. Guo (2012): *Oscillating rim hook tableaux and colored matchings*. *Advances in Applied Mathematics*, 48, pp. 393–406, doi:10.1016/j.aam.2011.06.007.
- [9] M. Christodoulakis & M. Christou (2012): *Abelian concepts in strings: a review*. In J. Holub, B. W. Watson, and J. Zdarek, editors: *Festschrift for Borivoj Melichar*, Czech Technical University in Prague, Czech Republic, pp. 19–45.
- [10] M. Christodoulakis, M. Christou, M. Crochemore & C. S. Iliopoulos (2014): *Abelian borders in binary words*. *Discrete Applied Mathematics*, 171, pp. 141–146, doi:10.1016/j.dam.2014.02.012.
- [11] S. Constantinescu & L. Ilie (2006): *Fine and Wilf’s theorem for Abelian periods*. *Bulletin of the European Association for Theoretical Computer Science*, 89, pp. 167–170.
- [12] M. Crochemore, C. S. Iliopoulos, T. Kociumaka, M. Kubica, J. Pachocki, J. Radoszewski, W. Rytter, W. Ty-  
czyński & T. Waleń (2013): *A note on efficient computation of all Abelian periods in a string*. *Information  
Processing Letters*, 113, pp. 74–77, doi:10.1016/j.ipl.2012.11.001.
- [13] E. Deutsch (1999): *Dyck path enumeration*. *Discrete Mathematics*, 204, pp. 167–202, doi:10.1016/S0012-  
365X(98)00371-9.
- [14] R. Eres, G. M. Landau & L. Parida (2004): *Permutation pattern discovery in biosequences*. *Journal of  
Computational Biology*, 11(6), pp. 1050–1060, doi:10.1089/cmb.2004.11.1050.
- [15] G. Fici, T. Lecroq, A. Lefebvre & E. Prieur-Gaston (2014): *Algorithms for computing Abelian periods of  
words*. *Discrete Applied Mathematics*, 163, pp. 287–297, doi:10.1016/j.dam.2013.08.021.
- [16] I. Gessel, W. Goddard, W. Shur, H. Wilf & L. Yen (1996): *Counting pairs of lattice paths by intersections*.  
*Journal of Combinatorial Theory, Series A*, 74, pp. 173–187, doi:10.1006/jcta.1996.0047.
- [17] I. Gessel & G. Viennot (1985): *Binomial determinants, paths, and hook length formulae*. *Advances in  
Mathematics*, 58, pp. 300–321, doi:10.1016/0001-8708(85)90121-5.
- [18] T. Kociumaka, J. Radoszewski & W. Rytter (2013): *Fast algorithms for Abelian periods in words and greatest  
common divisor queries*. In N. Portier and T. Wilke, editors: *STACS 2013, 30th International Symposium on  
Theoretical Aspects of Computer Science, Kiel, Germany, LIPIcs 20*, Schloss Dagstuhl - Leibniz-Zentrum  
für Informatik, Germany, pp. 245–256, doi:10.4230/LIPIcs.STACS.2013.245.
- [19] V. P. Kostov & B. Z. Shapiro (2009): *Narayana numbers and Schur-Szegő composition*. *Journal of Approx-  
imation Theory*, 161, pp. 464–476, doi:10.1016/j.jat.2008.10.013.
- [20] B. Lindstrom (1973): *On the vector representation of induced matroids*. *Bulletin of the London Mathematical  
Society*, 5, pp. 85–90, doi:10.1112/blms/5.1.85.
- [21] L. Parida (2006): *Gapped permutation patterns for comparative genomics*. In P. Bücher and B. M. E. Moret,  
editors: *WABI 2006, International Workshop on Algorithms in Bioinformatics, LNCS 4715*, Springer, Berlin,  
Heidelberg, pp. 376–387, doi:10.1007/11851561\_35.
- [22] N. Rampersad, M. Rigo & P. Salimov (2013): *On the number of Abelian bordered words*. In M.-P. Béal and  
O. Carton, editors: *DLT 2013, 17th International Conference on Developments in Language Theory, Paris-  
Est, France, LNCS 7907*, Springer, Berlin, Heidelberg, pp. 420–432, doi:10.1007/978-3-642-38771-5\_37.