

A Connection Between Unbordered Partial Words and Sparse Rulers

Aleksi Saarela

Aleksi Vanhatalo

Submitted: Jan 24, 2025; Accepted: Jan 19, 2026; Published: Feb 27, 2026

© The authors. Released under the CC BY license (International 4.0).

Abstract

Partial words are words that contain, in addition to letters, special symbols \diamond called *holes*. Two partial words $a = a_0 \dots a_n$ and $b = b_0 \dots b_n$ are *compatible* if, for all i , $a_i = b_i$ or at least one of a_i, b_i is a hole. A partial word is *unbordered* if it does not have a proper nonempty prefix and a suffix that are compatible. Otherwise, the partial word is *bordered*.

A set $R \subseteq \{0, \dots, n\}$ is called a *complete sparse ruler of length n* if, for all $k \in \{0, \dots, n\}$, there exists $r, s \in R$ such that $k = r - s$. These are also known as *restricted difference bases*.

From the definitions it follows that the more holes a partial word has, the more likely it is to be bordered. By introducing a connection between unbordered partial words and sparse rulers, we improve the bounds on the maximum number of holes an unbordered partial word can have over alphabets of sizes 4 or greater. We also provide a counterexample for a previously reported theorem, depending on the reported values of the minimal number of marks in a length-135 ruler. We have not verified this value.

We then study a two-dimensional generalization of these results. We adapt methods from the one-dimensional case to solve the correct asymptotic for the number of holes that an unbordered two-dimensional binary partial word can have.

Mathematics Subject Classifications: 68R15

1 Introduction

Partial words are words with a special hole symbol, usually denoted by \diamond . They were introduced, together with a new word relation in which holes are considered to be compatible with any letter, by Berstel and Boasson in [9]. They wanted a systematic, well-defined way to study words with some letters missing and replaced by “do not know” symbols. Since then many classical word combinatorics results have been considered in the context of partial words and this new relation, see, for example, [11], [15] and [14].

Department of Mathematics and Statistics, University of Turku, Finland (amsaar@utu.fi, aevanh@utu.fi).

In [10] and [2] Blanchet-Sadri, Allen et al. considered unbordered partial words with the maximum number of holes with respect to the length of the words. They were able to find the closed form for this maximum on the binary alphabet, a general upper bound depending on the size of the alphabet, and lower bounds for alphabet sizes 3 and 4.

Sparse rulers are seemingly unrelated combinatorial objects. The intuitive idea of a sparse ruler is to have a ruler with low number of marks which is still able to measure every distance less than its length between its marks. Sparse rulers have been studied by number theorists at least from the late 1940s, for example in [13], [21] and [16] under the name of restricted difference bases.

In Section 3 we show a connection between unbordered partial words and sparse rulers. Using results from the theory of rulers we are then able to correct and improve some results on unbordered partial words. In particular, we get new best upper and lower bounds for the number of holes when considering large enough alphabets. As a necessary tool we also provide a fairly simple proof for the correctness of the currently best known ruler construction, something which was originally omitted by Wichmann in [26]. The connection and some of its corollaries were first reported in the second author's master's thesis [25] (in Finnish).

In Section 4 we consider the two-dimensional case of unbordered partial words and rulers. We solve the correct asymptotic for the number of holes in n -by- m unbordered binary partial words, and prove some other preliminary results. While 2D sparse rulers have known applications in sensor placements, see [1] and [18], we are not aware of any pure mathematical research on the subject.

2 Partial words

In this section we present the basic definitions concerning partial words. We use slightly simplified versions of some definitions, but all equivalent to more commonly used variants.

Definition 1. A partial word over an alphabet A is a word over the alphabet $A_\diamond = A \cup \{\diamond\}$ where $\diamond \notin A$. The symbol \diamond is called a hole and is thought to be a “do not know” symbol in a word over A . The set of all finite partial words is denoted by A_\diamond^* .

Partial words without holes (that is, words over the alphabet A) are full words. The length of a partial word is measured similarly to a full word. We call all elements in $A \cup \{\diamond\}$ symbols while only the elements in A are called letters.

Definition 2. Let $w = w_0w_1 \dots w_n$ be a partial word over A . Denote by $D(w)$ the set $\{i_0, i_1, \dots, i_m\}$ where $w_j \in A \iff j = i_k$, for some k . The set $D(w)$ is called the domain of w .

Now we can define the compatibility relation between partial words.

Definition 3. Partial words $w = w_0w_1 \dots w_n$ and $v = v_0v_1 \dots v_m$ are compatible if $n = m$ and, for all $i \in D(w) \cap D(v)$, we have $w_i = v_i$. This is denoted by $w \uparrow v$.

In the intuitive sense, this means that compatible partial words could be equal if all the unknown symbols could be identified.

Example 4. Let $A = \{a, b, c\}$. Then partial words abc , $\diamond cc$ and $a\diamond c$ are all elements of A_\diamond^* , all have length 3 and $abc \uparrow a\diamond c$, $\diamond cc \uparrow a\diamond c$ but $\diamond cc \not\uparrow abc$.

The main topic of this paper can now be defined.

Definition 5. A border of a partial word w is a pair (x, y) such that x is a proper prefix of w , y is a proper suffix of w and $x \uparrow y$. A partial word without any borders is called an unbordered partial word.

From the definitions it is clear that the more holes a partial word has, the more likely it is to be bordered. This leads to the interesting question of determining the maximum number of holes in length- n unbordered partial words over a given alphabet. Adding to the notation in [2] we also want a notation in the case of an arbitrary alphabet.

Definition 6. $HB_k(n)$ denotes the maximum number of holes an unbordered partial word of length n can have over an alphabet of size k . Similarly, $HB_\infty(n)$ denotes the maximal number of holes an unbordered partial word can have over an infinite alphabet, i.e., $HB_\infty(n) = HB_n(n)$.

In [2] and [10], Allen, Blanchet-Sadri and others provided a closed form of HB_2 and bounds for larger alphabet sizes. The bounds we state in Theorem 7 are a simplified version of the ones in [10]. In some particular cases of n , the lower bound for HB_4 is better in [10], but still something we will improve upon in Section 3.

Theorem 7 ([2] and [10]). *For all $n > 9$ and $k \geq 4$,*

$$n - \lceil 2\sqrt{n+3} \rceil + 2 \leq HB_3(n) \leq HB_4(n) \leq HB_k(n) \leq n - \sqrt{\frac{2k}{k-1}}(n-1).$$

3 The connection with sparse rulers

We show a new connection between unbordered partial words and an independently studied combinatorial problem of sparse rulers. This proves to be a very fruitful connection as the theory of rulers has more powerful theorems than the partial words counterparts.

Definition 8. A subset $R \subseteq \{0, 1, \dots, n\}$ containing the numbers 0 and n is a sparse ruler (or just a ruler) of length n . The elements of a ruler are called marks. The ruler R can measure the distance d if there exists $i \in R$ and $j \in R$ so that $|i - j| = d$. If the ruler can measure all distances $d \in \{0, 1, \dots, n\}$, then the ruler is complete.

Example 9. The set $\{0, 1, 4, 6\}$ can easily be verified to be an example of a complete sparse ruler of length 6.

The combinatorial optimization problem that comes with complete sparse rulers is to minimize the number of marks needed to construct a complete sparse ruler with a given length n .

Definition 10. $M_1(n)$ is the minimum number of marks a length- n complete sparse ruler can have.

The following lower bound for the function $M_1(n)$ was originally found by Rédei and Rényi [21] in 1949 and then improved marginally by Leech [16] in 1956 and then again by BernshTEYN and Tait [8] in 2019. We present the lower bound by Leech. For the upper bounds, see the section about Wichmann rulers.

Theorem 11 ([16] and [21]).

$$M_1(n) \geq \max_{0 < \delta < 2\pi} \sqrt{2n + 1 - \frac{\sin \delta}{\sin \frac{\delta}{2n+1}}} > \max_{0 < \delta < 2\pi} \sqrt{2(1 - \frac{\sin \delta}{\delta})n + 1 - \frac{\sin \delta}{\delta}}.$$

Or approximately

$$M_1(n) > \sqrt{2.43n}.$$

We now show how the rulers are linked to the unbordered partial words. A part of the proof is equivalent to an earlier result by Allen et al. ([3], Proposition 1.), but there was no mention of a connection to sparse rulers.

Theorem 12. *For all unbordered partial words w , the domain $D(w)$ is a complete sparse ruler of length $|w| - 1$. Conversely, for all complete sparse rulers R , there exist an unbordered partial word w over some alphabet such that $R = D(w)$.*

Proof. Let $w = w_0w_1 \dots w_{n-1}$ be an unbordered partial word and let $0 < l < n$. Since w has no borders, its prefix $x = w_0 \dots w_{l-1}$ and suffix $y = w_{n-l} \dots w_{n-1}$ of length l are not compatible. This means that, for every $l < n$, there exists some $i \leq l - 1$ such that we have $w_i \not\sim w_{n-l+i}$ and thus $w_i \neq \diamond$ and $w_{n-l+i} \neq \diamond$. This means that i and $n - l + i$ are elements of the set $D(w)$. Since $n - l + i - i = n - l$, the set $D(w)$ can represent the distance $n - l$ using differences. Since $l \in \{1, 2, \dots, n - 1\}$ was arbitrary, $D(w)$ can represent all numbers $n - l \in \{1, 2, \dots, n - 1\}$ by differences and thus $D(w)$ is a complete sparse ruler of length $n - 1 = |w| - 1$.

For the converse statement, assume $R = \{r_0, r_1, r_2, \dots, r_{m-1}\}$ to be a complete sparse ruler of length $n - 1$ with m marks. We construct a partial word $v = v_0v_1 \dots v_{n-1}$ over the alphabet R such that

$$v_i = \begin{cases} i, & \text{if } i \in R \\ \diamond, & \text{otherwise} \end{cases}$$

Now clearly $D(v) = R$. Now let l be a length of a proper prefix x of v and let y be the suffix of the same length. Since the ruler R can represent the length $n - l$ by difference, there are marks r_i and $r_j = r_i + n - l$ in the ruler R . Thus in the word v there are different letters placed at the index positions r_i and $r_i + n - l$. But the index $r_i + n - l$ in the word

Proof. Clear result from Theorem 11 and Corollary 13. □

Before going for the lower bound, which needs some additional work, we show an unintuitive result that is dependent on the value of $M_1(135)$. The value $M_1(135) = 21$ is reported in the Online Encyclopedia of Integer Sequences (OEIS) [19], and verified in [22]. The following theorem is a direct counterexample to a result that was thought to be already proven in [2, Theorem 4.5]. The proof in [2] indirectly assumes that the ruler function M_1 is non-decreasing, which, surprisingly, it is not.

Since we provide example words in the proof, one needs to show that $M_1(135) > 20$ to verify this counterexample independently from reported values in [19] and [22].

Corollary 15. *For some $n \in \mathbb{N}$, $HB_k(n + 1) > HB_k(n) + 1$, for all $k \geq 4$.*

Proof. From known values of the function $M_1(n)$ ([19, 22]) we see that $M_1(135) = 21$ and $M_1(138) = 20$. Then by Corollary 13 we have

$$\begin{aligned} HB_{21}(136) &= 136 - 21 = 115 \text{ and} \\ HB_{21}(139) &= 139 - 20 = 119. \end{aligned}$$

The following unbordered words prove that these values are already obtained in four letter alphabet:

$$a^4b^3\diamond^{58}a(\diamond^2c)^3(\diamond^6d)^7(\diamond^3b)^3$$

and

$$ab^3\diamond^3(a\diamond^6)^3b(\diamond^{14}c)^5(\diamond^7d)^4b^3.$$

So $HB_4(139) > HB_4(136) + 3$ from which the claim follows. □

There is currently no intuitive nor rigorous characterization of the numbers $n \in \mathbb{N}$ such that $M_1(n + 1) < M_1(n)$.

The result $HB_k(n + 1) \leq HB_k(n) + 1$ holds for $k = 2$ due to the closed form of the function HB_2 [10]. The $k = 3$ case is an open problem.

3.2 Wichmann rulers and Wichmann words

The best upper bound for the function M_1 is by Pegg [20] with a construction for large values of n and by computer search for “small” values of n ($n < 257992$). The construction for large values depends on so called Wichmann rulers, constructed by Wichmann in 1963 [26]. In this section we give a self-contained presentation of Wichmann ruler construction, including the omitted proof from Wichmann’s original paper. We then show how to turn this construction to unbordered partial words over the minimal alphabet size.

If $R = \{a_0, \dots, a_k\}$ is a sparse ruler, we say that $D = (a_1 - a_0, \dots, a_k - a_{k-1})$ is a *difference representation* of R . If $a_0 < \dots < a_k$, then the elements of D are all positive. However, this is not required, which means that a ruler has many different difference representations. On the other hand, each difference representation specifies a unique ruler since 0 is the smallest element in rulers.

Definition 16. The Wichmann ruler $w_1(r, s)$ is the ruler defined by difference representation

$$(1^r, r + 1, (2r + 1)^r, (4r + 3)^s, (2r + 2)^{r+1}, 1^r)$$

or equivalently

$$((-1)^r, (2r + 1)^{r+1}, (4r + 3)^s, (2r + 2)^{r+1}, 1^r), \tag{1}$$

where a^b means that the value a is repeated b times.

For example, the ruler $w_1(1, 1)$ has the difference representation

$$(1, 2, 3, 7, 4, 4, 1)$$

and thus $w_1(1, 1) = \{0, 1, 3, 6, 13, 17, 21, 22\}$. The difference representation (1) might better show the symmetry of the construction. For $w_1(1, 1)$ it would be $(-1, 3, 3, 7, 4, 4, 1)$.

For most lengths there is no Wichmann ruler, so we also need extended Wichmann rulers.

Definition 17. The extended Wichmann ruler $w_1(r, s, i, j)$ is the ruler defined by the difference representation

$$(1^r, r + 1, (2r + 1)^r, (4r + 3)^s, (2r + 2)^{r+1}, 1^r, (r + 1)^i, j)$$

or equivalently

$$((-1)^r, (2r + 1)^{r+1}, (4r + 3)^s, (2r + 2)^{r+1}, 1^r, (r + 1)^i, j)$$

where $j \leq r + 1$ and a^b means that the value a is repeated b times.

Extended Wichmann rulers are complete. Wichmann originally omitted the proof due to its tedious nature, so we include it to cut loose ends in our theory later.

The proof presented here is a greatly improved version of the one in the second author's master's thesis [25] (in Finnish).

Theorem 18. For all $r, s, i \in \mathbb{N}$ and $j \leq r + 1$, the ruler $w_1(r, s, i, j)$ is complete.

Proof. Firstly we note that if $w_1(r, s)$ is complete, then so is $w_1(r, s, i, j)$. This is because the ruler $w_1(r, s, i, j)$ can measure the distances that are less or equal to the length of the ruler $w_1(r, s)$ by completeness of $w_1(r, s)$. Then for the longer lengths of the type $n + l$, where n is the length of the ruler $w_1(r, s)$ and $l \leq (r + 1)i + j$, the ruler $w_1(r, s, i, j)$ has the marks $\{0, 1, \dots, r\}$ and marks $\{n + t(r + 1) \mid t \leq i\}$ and the mark $n + i(r + 1) + j$.

What is left to prove is that $w_1(r, s)$ is complete for all parameter values. So we need to prove that $w_1(r, s)$ can measure every distance from 1 to its length, which is $(4r + 3)(r + s + 1) + r$ by Definition 16. It is sufficient to prove that, for all $x \leq (4r + 3)(r + s + 1) + r - 1$, if x can be measured, then so can $x + 1$.

For the rest of the proof, we use the fact that a sparse ruler with a difference representation (b_1, \dots, b_k) can measure a distance $x \geq 1$ if and only if there exists $i \leq j$ such that $x = |b_i + \dots + b_j|$. This can be seen as follows: Let $R = \{a_0, \dots, a_k\}$ and $b_i = a_i - a_{i-1}$ for

all i . Then $a_i = a_0 + b_1 + \dots + b_i$ for all i . The distances that can be measured are exactly the numbers $|a_j - a_i| = |b_{i+1} + \dots + b_j|$ for $i \leq j$. So in our case x can be measured if and only if it is a sum of consecutive elements of (1). If the first of these elements is -1 , then we can leave it out to get a representation for $x + 1$. Similarly, if these elements are followed by a 1, then we can add it to the end to get a representation for $x + 1$. We still have to consider sums of consecutive elements of the part

$$((2r + 1)^{r+1}, (4r + 3)^s, (2r + 2)^r)$$

and sums that reach all the way to the end of (1) and end in r copies of 1. For the former sums, we have the following cases:

- If $x = i(2r + 1) + j(4r + 3)$, where $i \geq 2$ and $j < s$, then $x + 1 = (i - 2)(2r + 1) + (j + 1)(4r + 3)$.
- If $x = 2r + 1 + j(4r + 3)$, then $x + 1 = j(4r + 3) + 2r + 2$.
- If $x = i(2r + 1) + s(4r + 3) + j(2r + 2)$, where $i \geq 1$ and $j \leq r$, then $x + 1 = (i - 1)(2r + 1) + s(4r + 3) + (j + 1)(2r + 2)$.
- If $x = i(4r + 3) + j(2r + 2)$, where $i \geq 1$ and $j \leq r - 1$, then $x + 1 = (i - 1)(4r + 3) + (j + 2)(2r + 2)$.
- If $x = i(4r + 3) + r(2r + 2)$, then $x + 1 = r(-1) + (r + 1)(2r + 1) + i(4r + 3)$.

For the latter sums, we have the following cases:

- If $x = i(2r + 1) + s(4r + 3) + (r + 1)(2r + 2) + r$, where $i \leq r$, then $x + 1 = i(-1) + (r + 1)(2r + 1) + s(4r + 3) + (i + 1)(2r + 2)$.
- If $x = i(4r + 3) + (r + 1)(2r + 2) + r$, where $i \leq s - 1$, then $x + 1 = r(2r + 1) + (i + 1)(4r + 3)$.
- If $x = i(2r + 2) + r$, where $i \leq r$, then $x + 1 = (r - i)(-1) + (i + 1)(2r + 1)$.

This completes the proof. □

We can now present the proof for asymptotic growth on the number of marks that Wichmann rulers can achieve. The proof is the same as in Wichmann's original paper with a slightly larger constant. It is included so a reader interested in unbordered partial words can follow the calculations later with words in mind. The constant term in the bound could be lowered according to Pegg's result [20], but we mostly care about the asymptotic behaviour.

Theorem 19 ([26]). *For all $n \geq 213$, there exists an extended Wichmann ruler $w_1(r, s, i, j)$ of length n so that*

$$|w_1(r, s, i, j)| \leq \sqrt{3n} + 4.$$

Proof. By Definition 16, the length of a Wichmann ruler $w_1(r, s) = w_1(r, s, 0, 0)$ is $l_{r,s} = 4r(r + s + 2) + 3s + 3$ and the number of marks is $m_{r,s} = 4r + s + 3$. Let $r \geq 4$ and $s = 2r + e \in [2r - 2, 2r + 4]$. Then

$$\begin{aligned} \sqrt{3l_{r,s}} &= \sqrt{12r(r + s + 2) + 9s + 9} \\ &= \sqrt{36r^2 + 2(e + 3)6r + 6r + 9e + 9} \\ &\geq \sqrt{36r^2 + 2(e + 3)6r + (e + 3)^2} \\ &= 6r + e + 3 = 4r + s + 3 = m_{r,s}. \end{aligned}$$

An increase of parameter s by one increases the length of Wichmann ruler by $4r + 3 < 4(r + 1)$. So, for all $n \in [l_{r,2r+e}, l_{r,2r+e+1}]$, we can extend the Wichmann ruler $w_1(r, 2r + e)$ with 4 or less marks to obtain a complete ruler of length n following Definition 17 (by adding 4 marks, we can add 4 intervals of length $r + 1$ or lower). This extended Wichmann ruler has at most $\sqrt{3l_{r,2r+e}} + 4 \leq \sqrt{3n} + 4$ marks. A simple calculation shows us that $l_{r,2r+4} = l_{r+1,2(r+1)-2}$,

$$\begin{aligned} l_{r,2r+4} &= 4r(r + 2r + 4 + 2) + 3(2r + 4) + 3 \\ &= 4r(3r + 6) + 12 + 6r + 3 \\ &= 4r(3r + 3) + 12r + 12 + 6r + 3 \\ &= 4(r + 1)(3r + 3) + 6r + 3 \\ &= 4(r + 1)(r + 1 + 2(r + 1) - 2 + 2) + 3(2(r + 1) - 2) + 3 \\ &= l_{r+1,2(r+1)-2}. \end{aligned}$$

This means that, for all lengths $n \geq l_{4,6} = 213$, there exists some $r \geq 4$ such that $n \in [l_{r,2r-2}, l_{r,2r+4}]$. We already showed that for lengths in these intervals, a complete sparse ruler of length n can be constructed with $\sqrt{3n} + 4$ marks, and thus the claim is proved. \square

We now have (mostly) proven

Theorem 20 ([26]). *For all n ,*

$$M_1(n) \leq \sqrt{3n} + 4.$$

Proof. For $n \geq 213$ this follows from Theorems 18 and 19. For $n < 213$, the constant term is so generous that the claim can easily be checked by computer search, or from known values of M_1 in OEIS [19] or in [22]. \square

Now we can consider how to turn Wichmann rulers into unbordered partial words. Theorem 21 gives the minimum alphabet size where this can be done. We call the constructed words *Wichmann words*.

Theorem 21. *Partial words*

$$W(r, s) = ab^r \diamond^r (a \diamond^{2r})^r b (\diamond^{4r+2} c)^s (\diamond^{2r+1} d)^{r+1} b^r$$

and

$$W(r, s, i, j) = ab^r \diamond^r (a \diamond^{2r})^r b (\diamond^{4r+2} c)^s (\diamond^{2r+1} d)^{r+1} b^r (\diamond^r c)^i \diamond^{j-1} c$$

are unbordered for all parameter values $r, s, i \in \mathbb{N}$ and $1 \leq j \leq r + 1$.

Proof. In this proof, instead of directly showing there is no border, we equivalently show that the domain of these partial words are complete sparse rulers in such a way that the measuring can be done between marks that correspond to different letters in the word. The word is then unbordered similarly to the word in the proof of Theorem 12.

For all i and $j \leq r + 1$, we have $W(r, s, i, j) = W(r, s) (\diamond^r c)^i \diamond^{j-1} c$. Then it is quite easy to see (following the first part of Theorem 18) that if $W(r, s)$ is unbordered, then so is $W(r, s, i, j)$. So we concentrate on $W(r, s)$ for the rest of the proof.

$W(r, s)$ is constructed so that $D(W(r, s)) = w_1(r, s)$. We label the defining vector of differences in Wichmann rulers as follows:

$$(1^r, r + 1, (2r + 1)^r, (4r + 3)^s, (2r + 2)^{r+1}, 1^r) = (A, B, C, D, E, F),$$

where $A = 1^r, B = r + 1, C = (2r + 1)^r$ and so on. We consider the marks bordering two areas to belong to both. Now note that if a distance is measured using two marks within a single area, the measured distance is a multiple of that area's difference. Thus each area needs to use only at most two different letters to be able to measure within itself. Then all other measuring that the underlying Wichmann ruler does is with marks in two different areas. If the areas use disjoint alphabets, then those areas clearly are able to measure all the differences between them. Also note that area B has only a single difference, and shares all its marks with the areas A and C , so does not need a separate check. The above observations take care of all other distances except the distances from area A to areas C and F and from area C to area F . The fact that the word measures the distances between these areas has to be checked separately.

The distance between the starting a in the area A to the other a -letter in the area C is a multiple of $2r + 1$. These distances can be measured within the area C .

The distances between letters b in the area A and the letter b in the area C are in the range $[r + 1 + r(2r + 1), 2r + r(2r + 1)] = [r(2r + 2) + 1, r(2r + 2) + r]$. This range can be measured from the area E to the area F .

The measurements from the area A to the area F can always be made using the letter a from the area A or by using the d from the area F so the common letter b does not matter.

Lastly, the distances between the last letter of the area C and the ending letters of the area F are in the range $[(4r + 3)s + (2r + 2)(r + 1) + 1, (4r + 3)s + (2r + 2)(r + 1) + r]$ which is equal to $[r + 1 + (2r + 1)r + (4r + 3)s + (2r + 2), r + r + 1 + (2r + 1)r + (4r + 3)s + (2r + 2)]$. This can be measured from the area A to the second mark in the area E .

Now this word can measure all the distances similarly to a Wichmann ruler, so it is unbordered. \square

Like in the one-dimensional case, we define $M_2(n, m)$ to be the minimal number of marks in an (n, m) -ruler. Now the same game of finding good rulers and then turning them to unbordered partial words over some finite alphabet can be played in 2D.

It is easy to see that the Cartesian product of two 1D-rulers is a 2D-ruler. This gives us one way to construct 2D-rulers. However, as one might expect, this is not the optimal way. The asymptotic we get from the Cartesian product of two 1D rulers is

$$M_2(n, m) \leq 3\sqrt{mn} + \mathcal{O}(\sqrt{n} + \sqrt{m})$$

(assuming the 1D rulers we currently know are close to optimal). This growth rate is achieved by many simpler constructions, for example U-shaped walls form a better ruler than this, see [18] for that construction. The best construction we found is seen in Figure 3.

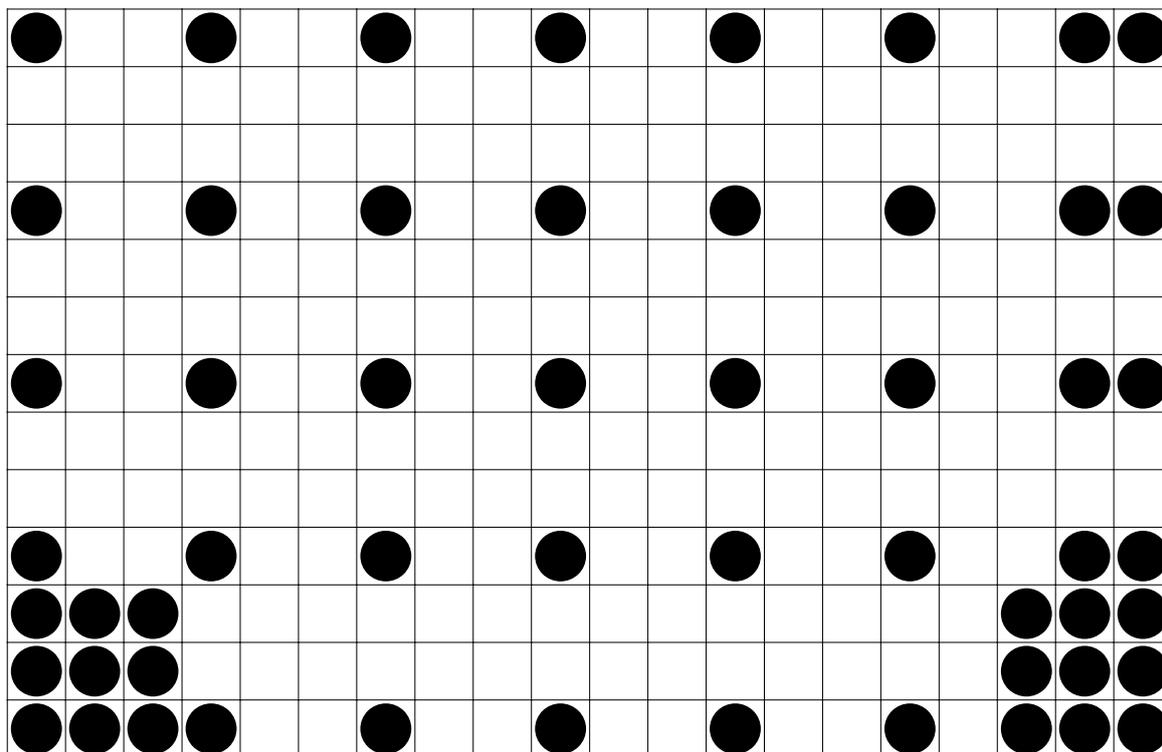


Figure 3: $(19, 12)$ -ruler constructed in Theorem 25 with parameters $l = k = 3$.

Theorem 25.

$$\sqrt{4mn + 2(m + n) + \frac{1}{4} + \frac{1}{2}} \leq M_2(n, m) \leq 2\sqrt{2}\sqrt{(n + 1)(m + 1)} + \mathcal{O}(\sqrt{n} + \sqrt{m})$$

Proof. For the upper bound, the following construction gives a working 2D-ruler. Let

$R = R_1 \cup R_2 \cup R_3 \cup R_4$ where:

$$\begin{aligned} R_1 &= \{0, 1, \dots, l-1\} \times \{0, 1, \dots, k-1\} \\ R_2 &= R_1 + (n+1-l, 0) \\ R_3 &= \{(x, y) \mid x = 0 \pmod{l} \text{ and } y = 0 \pmod{k}\} \\ R_4 &= \{(n, m), (n, y), (x, m) \mid x = 0 \pmod{l} \text{ and } y = 0 \pmod{k}\} \end{aligned}$$

for some $1 \leq l \leq n$ and $1 \leq k \leq m$. Figure 3 showcases this ruler, and it is easy to see from the figure that it can in fact measure all vectors. The idea is that we have a sparse lattice of marks covering the whole ruler (the set R_3), and then two denser clusters in the corners (the sets R_1 and R_2). Then (most) vectors can be measured as the path from marks from these clusters to marks in the sparse lattice, or vice versa. The set R_4 is a refinement of the lattice R_3 on the edges of the ruler.

Counting the number of elements in each set we have $|R_1| = |R_2| = lk$, $|R_3| = \lfloor \frac{(n+1)(m+1)}{lk} \rfloor$ and $|R_4| \leq \lceil \frac{n+1}{l} + \frac{m+1}{k} \rceil$. Now the number of elements in this ruler is

$$|R| \leq |R_1| + |R_2| + |R_3| + |R_4| \leq 2lk + \frac{(n+1)(m+1)}{lk} + \frac{n+1}{l} + \frac{m+1}{k} + 1.$$

Here if we pick $l = \lfloor \frac{\sqrt{n+1}}{\sqrt[4]{2}} \rfloor$ and $k = \lfloor \frac{\sqrt{m+1}}{\sqrt[4]{2}} \rfloor$ we obtain the claimed upper bound on the number of marks.

For the lower bound, first note that when a 2D-ruler can measure a vector (x, y) , it can also measure the vector $(-x, -y)$ with the same marks. Let $v_{n,m}$ be the number of vectors that an (n, m) -ruler needs to measure with pairwise different mark pairs. This number can be counted by first considering vectors without a zero component. We have the vectors (x, y) where $x \in \{1, 2, \dots, n\}$ and $y \in \{1, 2, \dots, m\}$. Then by changing those vectors into $(-x, y)$ we get another set of vectors that need to be measured differently from the previous vectors. Finally, we have $n + m$ vectors where one coordinate is zero. Together, this makes up $2nm + n + m$ vectors that have to be measured by different mark pairs. The number of mark pairs then has to be greater than this, so

$$\binom{M_2(n, m)}{2} \geq 2nm + n + m,$$

leading to the lower bound. □

To make the construction from the proof above into a two-dimensional unbordered word, we can fill marks in R_1 and R_2 with a single letter, say a , and R_3 and R_4 with another letter, say $b \neq a$. The problem is that the sets have common elements, and we cannot assign multiple letters to a single coordinate. To solve this problem, we can fill R_1 and R_2 with the first letter and add additional markers to measure the distances we miss by doing so. This means that we have to give in the construction new ways to measure vectors that are in the form $r_2 - (0, 0)$ and $(n, 0) - r_1$, where $r_2 \in R_2$ and $r_1 \in R_1$. This can be done by adding letters on top of the areas R_1 and R_2 . The number of added marks

is $\mathcal{O}(\sqrt{n} + \sqrt{m})$, so we keep the same growth rate of the construction with even the binary alphabet. This is shown in Figure 4. We summarize these observations in the theorem below, where $\text{HB}2_k(n, m)$ denotes the maximum number of holes a two-dimensional n -by- m unbordered partial word over a k -sized alphabet can have. Note that due the difference of indexing in rulers and words, the $2\sqrt{2}\sqrt{(n+1)(m+1)}$ term in Theorem 25 turns to $2\sqrt{2}\sqrt{nm}$.

Theorem 26. *For all $k \geq 2$ and $mn \geq 1$,*

$$nm - 2\sqrt{2}\sqrt{nm} - \mathcal{O}(\sqrt{n} + \sqrt{m}) \leq \text{HB}2_k(n, m).$$

By adapting an argument from one-dimensional case in [2], we can asymptotically solve the correct growth rate for $\text{HB}2_2$. The argument turns the unbordered partial word to a graph, and then applies Turán's theorem.

Theorem 27 (Turán's theorem [24]). *Given graph $G = (V, E)$ without $(k+1)$ -clique, then*

$$|E| \leq \frac{k-1}{2k} |V|^2$$

holds.

Theorem 28 (Essentially in [2]). *For all $k \geq 2$ and $mn \geq 1$,*

$$\text{HB}2_k(n, m) \leq mn - \sqrt{\frac{2k}{k-1}(2nm - n - m)}.$$

Proof. In the proof of Theorem 25, we saw that an (n, m) -ruler needs to have at least $2nm - n - m$ different pairs of marks. In unbordered partial words, this means that the word needs at least that many different pairs of letters such that pairs do not have the same letter. Consider the unbordered partial word over a k -sized alphabet as a graph, where the nodes are the letters and there is an edge between two letters if the letters are not compatible. Then this graph is obviously k -colorable and thus has no $(k+1)$ -clique. By Theorem 27, we then have

$$2nm - n - m \leq \frac{k-1}{2k} (mn - \text{HB}2_k(n, m))^2,$$

where by solving for $\text{HB}2_k(n, m)$ we get the proof of the theorem. □

Combining Theorems 26 and 28 we see that

$$\text{HB}2_2(n, m) = mn - 2\sqrt{2}\sqrt{mn} + \mathcal{O}(\sqrt{n} + \sqrt{m}).$$

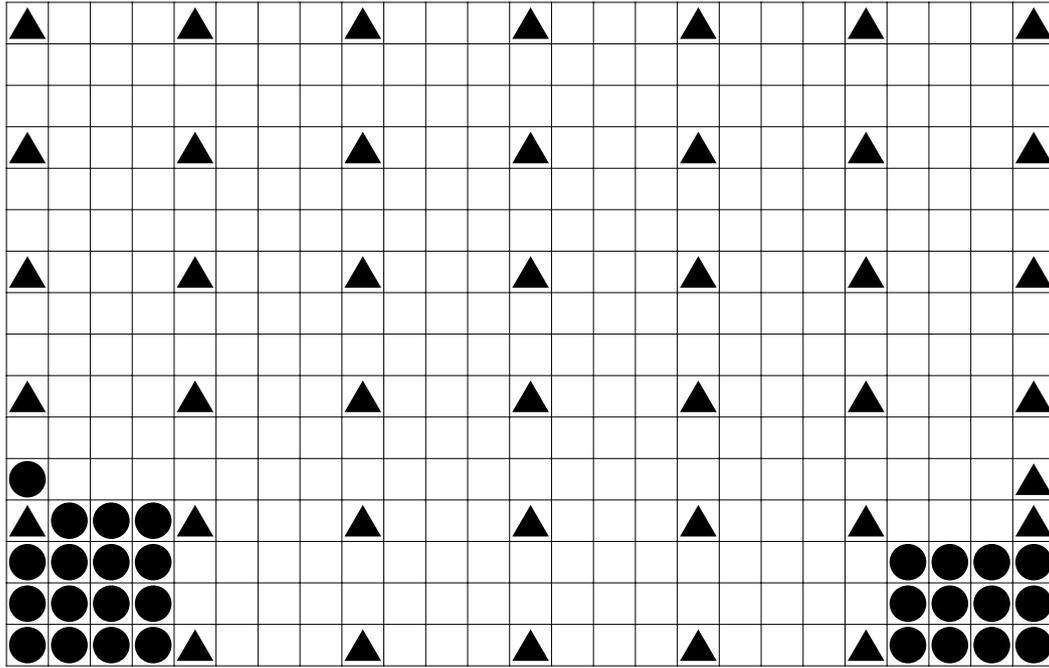


Figure 4: A 2D ruler turned into 2D unbordered binary partial word. Here the alphabet consists of disks and triangles, and an empty cell denotes a hole. Note the few extra marks placed on top of the rectangle areas compared to figure 3.

5 Quick note on cross-bifix-free codes

Lastly, we make a quick observation related to so called cross-bifix-free codes of fixed length. A code C is cross-bifix-free if, for any codewords $c_1 \in C$ and $c_2 \in C$, there is no proper prefix of c_1 which is also a suffix of c_2 . More formally $\forall c_1, c_2 \in C$:

$$c_1 = xy_1 \wedge c_2 = y_2x \implies c_1 = c_2 = x \vee x = \varepsilon.$$

These codes and, in particular, the number of codewords in maximal cross-bifix-free codes have seen recent interest; see [23] for a summary of recent results.

Related to our topic, given an alphabet A and a partial word $w \in A_\diamond^*$, we can define a language by filling all the holes in the given word by letters. These languages have been considered earlier in language theoretical settings, see [12, 17].

Definition 29.

$$L_\uparrow(w) = \{u \in A^* | w \uparrow u\}$$

If $w \in A_\diamond^*$ is an unbordered partial word with k holes, then it is easy to see that $L_\uparrow(w)$ is a cross-bifix-free code of size $|A|^k$. In the binary case, for example, this idea would give a subset of A^n of size $2^{\lfloor n-2\sqrt{n-1} \rfloor}$ by taking w to be one of the optimal unbordered binary partial words found in [2]. It is known that the maximal size of a cross-bifix-free subset of

A^n is much larger, see [23], but the partial word construction is worth mentioning because it is particularly simple, and also because it gives another connection between different research topics.

In the two-dimensional case, this idea becomes perhaps more interesting because the previously known results are not as strong as in the one-dimensional case. Again, if w is an unbordered two-dimensional partial word with k holes, then $L_{\uparrow}(w)$ is a cross-bifix-free code of size $|A|^k$. Therefore, we can use Theorem 26 to get a cross-bifix-free set of binary n -by- m words of size $2^{nm-2\sqrt{2}\sqrt{nm}-O(\sqrt{n}+\sqrt{m})}$, where the exact value of the exponent is easy to calculate once the seed word is fixed. This could be compared with existing constructions, such as the ones in [5, 6, 7], and we suspect it could be used as a basis for a more complicated construction. A detailed comparison and a study of improved constructions is a potential direction for future research.

6 Conclusions

In Sections 2 and 3 we showed a new connection between unbordered partial words and complete sparse rulers. Using this connection and earlier ruler results, we get new bounds on the function HB_k when $k \geq 4$. The new bounds are

$$n - \sqrt{3n - 3} - 4 \leq \text{HB}_4(n) \leq \text{HB}_k(n) \leq \text{HB}_{\infty}(n) \leq n - \sqrt{2.43n},$$

where the lower bound asymptotically beats previous lower bounds and the upper bound does so when $k \geq 6$.

We also provided a direct counterexample to the claim that was made in [2, Theorem 4.5]: $\text{HB}_k(n+1) \leq \text{HB}_k(n) + 1$, for all n and k . The counterexample shows that this is not the case for $k \geq 4$. Interestingly, for $k = 2$, the claim holds, but the case $k = 3$ is unclear.

Conjecture 30. For all $n \geq 1$,

$$\text{HB}_3(n+1) \leq \text{HB}_3(n) + 1.$$

Given that the best known ruler construction can be converted to a word over an alphabet of size 4 and the ruler construction is thought to be optimal, it really seems that the four-letter alphabet gives the optimal unbordered partial words. This was conjectured already by Blanchet-Sadri et al. [10], and now the conjecture has new interest in the study of sparse rulers.

Conjecture 31. For all $n \geq 1$,

$$\text{HB}_4(n) = \text{HB}_{\infty}(n).$$

We then in Section 4 considered two-dimensional generalizations of these ideas. Unsurprisingly, in two dimensions it is easy to come up with asymptotically better rulers than just a Cartesian product of two one-dimensional rulers. We showed one example, which could also be turned into binary unbordered two-dimensional words with just a few extra marks. This reaches the optimal asymptotic for n -by- m unbordered binary partial words when both n and m are taken to infinity. Without explicit conjecture, we state as open the problem of finding better two-dimensional rulers and unbordered partial words.

Acknowledgements

Authors were supported by the Academy of Finland under grants 339311 and 346566.

References

- [1] Aboumahmoud, I., Muqaibel, A., Alhassoun, M., Alawsh, S.: A review of sparse sensor arrays for two-dimensional direction-of-arrival estimation. *IEEE Access* **9**, 92999–93017 (2021). doi:[10.1109/ACCESS.2021.3092529](https://doi.org/10.1109/ACCESS.2021.3092529)
- [2] Allen, E., Blanchet-Sadri, F., Bodnar, M., Bowers, B., Hidakatsu, J., Lensmire, J.: Combinatorics on partial word borders. *Theoretical Computer Science* **609**, 469–493 (2016). doi:[10.1016/j.tcs.2015.11.006](https://doi.org/10.1016/j.tcs.2015.11.006)
- [3] Allen, E., Blanchet-Sadri, F., Byrum, C., Cucuringu, M., Mercaş, R.: Counting bordered partial words by critical positions. *The Electronic Journal of Combinatorics* **18**(1), #P138 (2011). doi:[10.37236/625](https://doi.org/10.37236/625)
- [4] Anselmo, M., Giammarresi, D., Madonia, M.: Avoiding overlaps in pictures. In: 19th International Conference on Descriptive Complexity of Formal Systems (DCFS). pp. 16–32. Springer International Publishing (2017). doi:[10.1007/978-3-319-60252-3_2](https://doi.org/10.1007/978-3-319-60252-3_2)
- [5] Anselmo, M., Giammarresi, D., Madonia, M.: Non-expandable non-overlapping sets of pictures. *Theoretical Computer Science* **657**, 127–136 (2017). doi:[10.1016/j.tcs.2016.09.025](https://doi.org/10.1016/j.tcs.2016.09.025)
- [6] Barucci, E., Bernini, A., Bilotta, S., Pinzani, R.: Non-overlapping matrices. *Theoretical Computer Science* **658**, 36–45 (2017). doi:[10.1016/j.tcs.2016.05.009](https://doi.org/10.1016/j.tcs.2016.05.009)
- [7] Barucci, E., Bernini, A., Bilotta, S., Pinzani, R., et al.: Non-overlapping matrices via Dyck words. *Enumerative Combinatorics and Applications*. **1** (2021). doi:[10.54550/ECA2021V1S2R9](https://doi.org/10.54550/ECA2021V1S2R9)
- [8] Bernshteyn, A., Tait, M.: Improved lower bound for difference bases. *Journal of Number Theory* **205**, 50–58 (2019). doi:[10.1016/j.jnt.2019.05.002](https://doi.org/10.1016/j.jnt.2019.05.002)
- [9] Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. *Theoretical Computer Science* **218**(1), 135–141 (1999). doi:[10.1016/S0304-3975\(98\)00255-2](https://doi.org/10.1016/S0304-3975(98)00255-2)
- [10] Blanchet-Sadri, F., Allen, E., Byrum, C., Mercaş, R.: How many holes can an unbordered partial word contain? In: International Conference on Language and Automata Theory and Applications. pp. 176–187. Springer (2009). doi:[10.1007/978-3-642-00982-2_15](https://doi.org/10.1007/978-3-642-00982-2_15)
- [11] Blanchet-Sadri, F., Mercaş, R., Scott, G.: A generalization of Thue freeness for partial words. *Theoretical Computer Science* **410**(8-10), 793–800 (2009). doi:[10.1016/j.tcs.2008.11.006](https://doi.org/10.1016/j.tcs.2008.11.006)
- [12] Dassow, J., Manea, F., Mercaş, R.: Regular languages of partial words. *Information Sciences* **268**, 290–304 (2014). doi:[10.1016/j.ins.2013.12.032](https://doi.org/10.1016/j.ins.2013.12.032)

- [13] Erdős, P., Gál, I.: On the representation of $1, 2, \dots, n$ by differences. *Indagationes Math* **10**(9), 379–382 (1948)
- [14] Halava, V., Harju, T., Kärki, T.: Square-free partial words. *Information Processing Letters* **108**(5), 290–292 (2008). doi:[10.1016/j.ipl.2008.06.001](https://doi.org/10.1016/j.ipl.2008.06.001)
- [15] Halava, V., Harju, T., Kärki, T., Séébold, P.: Overlap-freeness in infinite partial words. *Theoretical Computer Science* **410**(8-10), 943–948 (2009). doi:[0.1016/j.tcs.2008.12.041](https://doi.org/0.1016/j.tcs.2008.12.041)
- [16] Leech, J.: On the representation of $1, 2, \dots, n$ by differences. *Journal of the London Mathematical Society* **s1-31**(2), 160–169 (1956). doi:[10.1112/jlms/s1-31.2.160](https://doi.org/10.1112/jlms/s1-31.2.160)
- [17] Lischke, G.: Restorations of punctured languages and similarity of languages. *Mathematical Logic Quarterly* **52**(1), 20–28 (2006). doi:[10.1002/malq.200510017](https://doi.org/10.1002/malq.200510017)
- [18] Liu, C., Vaidyanathan, P.: Two-dimensional sparse arrays with hole-free coarray and reduced mutual coupling. In: 2016 50th Asilomar Conference on Signals, Systems and Computers. pp. 1508–1512 (2016). doi:[10.1109/ACSSC.2016.7869629](https://doi.org/10.1109/ACSSC.2016.7869629)
- [19] OEIS Foundation Inc.: Entry A046693 in The On-Line Encyclopedia of Integer Sequences (2024), <https://oeis.org/A046693>
- [20] Pegg, E.: Hitting all the marks: Exploring new bounds for sparse rulers and a wolfram language proof., <https://blog.wolfram.com/2020/02/12/hitting-all-the-marks-exploring-new-bounds-for-sparse-rulers-and-a-wolfram-language-proof/>
- [21] Rédei, L., Rényi, A.: On the representation of the numbers $1, 2, \dots, n$ by means of differences. *Matematicheskii Sbornik* **24**(66), 385–389 (1949), <https://www.mathnet.ru/eng/sm5985>
- [22] Schwartz, F., Schröder, Y., Wolf, L., Schoebel, J.: Large minimum redundancy linear arrays: Systematic search of perfect and optimal rulers exploiting parallel processing. *IEEE Open Journal of Antennas and Propagation* **2**, 79–85 (2020). doi:[10.1109/OJAP.2020.3043541](https://doi.org/10.1109/OJAP.2020.3043541)
- [23] Stanovnik, L., Moškon, M., Mraz, M.: In search of maximum non-overlapping codes. *Designs, Codes and Cryptography* pp. 1–28 (2024). doi:[10.1007/s10623-023-01344-z](https://doi.org/10.1007/s10623-023-01344-z)
- [24] Turán, P.: On an external problem in graph theory. *Matematikai és Fizikai Lapok* **48**, 436–452 (1941)
- [25] Vanhatalo, A.: Reunattomien osittaissanojen ja harvojen viivoittimien yhteydestä (2024), <https://urn.fi/URN:NBN:fi-fe2024032713407>, master’s thesis, University of Turku
- [26] Wichmann, B.: A note on restricted difference bases. *Journal of the London Mathematical Society* **s1-38**(1), 465–466 (1963). doi:[10.1112/jlms/s1-38.1.465](https://doi.org/10.1112/jlms/s1-38.1.465)