

# Improved Upper Bounds on all Maximal $\alpha$ -gapped Repeats and Palindromes

Tomohiro I<sup>1</sup> and Dominik Köppl<sup>\*2</sup>

<sup>1</sup>Kyushu Institute of Technology, Japan

<sup>2</sup>Department of Computer Science, TU Dortmund, Germany

August 19, 2018

## Abstract

We show that the number of all maximal  $\alpha$ -gapped repeats and palindromes of a word of length  $n$  is at most  $3(\pi^2/6 + 5/2)\alpha n$  and  $7(\pi^2/6 + 1/2)\alpha n - 3n - 1$ , respectively.

**Keywords:** Combinatorics on words; Repetitions; Gapped repeats; Gapped palindromes

## 1 Introduction

Gapped repeats and gapped palindromes are a generalization of squares and palindromes. They are motivated by DNA and RNA structures [12, Sect. 7.11.2], modelling different types of tandem repeats (i.e. repetitions), interspersed repeats (e.g., [20, 13]), as well as inverted repeats (e.g., [24, 23]) or hairpin structures (e.g., [21]). Such structures are important for the analysis of structural and functional information of genetic sequences, as can be seen by the number of software products devoted to finding those (see references in [6, 18, 22]).

Given a word  $w$ , a *gapped repeat* is a triple of integers  $(i_\lambda, i_\rho, u)$  with the properties (a)  $0 < i_\rho - i_\lambda$ , and (b)  $w[i_\lambda..i_\lambda + u - 1] = w[i_\rho..i_\rho + u - 1]$ . A variant are *gapped palindromes* with the properties (a)  $0 \leq i_\rho - i_\lambda$ , and (b)  $w[i_\lambda..i_\lambda + u - 1]$  is equal to the reverse of  $w[i_\rho..i_\rho + u - 1]$ . In both cases (repeats or palindromes),  $w[i_\lambda..i_\lambda + u - 1]$  and  $w[i_\rho..i_\rho + u - 1]$  are called *left* and *right arm*, respectively. Given a real number  $\alpha \geq 1$ ,  $(i_\lambda, i_\rho, u)$  is called  *$\alpha$ -gapped* if  $i_\rho - i_\lambda \leq \alpha u$ . A gapped repeat is *maximal* if its arms can be extended neither to their left nor to their right sides (to form a larger gapped repeat). Similarly, a gapped palindrome is *maximal* if it can be extended neither inwards nor outwards. Maximal  $\alpha$ -gapped repeats and palindromes starred in several recent papers [16, 17, 5, 10]. The most intriguing questions are:

1. How to compute all maximal  $\alpha$ -gapped repeats/palindromes efficiently, and:
2. What is the maximum number of maximal  $\alpha$ -gapped repeats/palindromes in a word?

---

\*Corresponding author.

E-mail addresses: [tomohiro@ai.kyutech.ac.jp](mailto:tomohiro@ai.kyutech.ac.jp) (Tomohiro I) and [dominik.koeppl@tu-dortmund.de](mailto:dominik.koeppl@tu-dortmund.de) (Dominik Köppl).

Previously, the second question was answered with  $\mathcal{O}(\alpha^2 n)$  [16, 17], subsequently with  $\mathcal{O}(\alpha n)$  [5], and finally with  $18\alpha n$  and  $28\alpha n + 7n$  for maximal  $\alpha$ -gapped repeats and maximal  $\alpha$ -gapped palindromes, respectively [10]. Following this line of achievements, this article gives yet another improvement to those answers:

- The number of all maximal  $\alpha$ -gapped repeats in a word of length  $n$  is at most  $3(\pi^2/6 + 5/2)\alpha n$  (Theorem 4.3).
- The number of all maximal  $\alpha$ -gapped palindromes in a word of length  $n$  is at most  $7(\pi^2/6 + 1/2)\alpha n - 3n - 1$  (Theorem 4.9).

The improvement of the upper bound on the number of all maximal  $\alpha$ -gapped repeats is a small refinement step (in Lemmas 3.4 to 3.6), whereas our new upper bound on the number of all maximal  $\alpha$ -gapped palindromes involves a more thorough analysis (in Lemma 4.7). Here, the main difference to [10] is that we define a periodic gapped palindrome to have a left arm with a sufficiently long periodic suffix (instead of prefix).

We present these improvements in a more generalized setting that supports overlaps (previous results assumed that  $i_\lambda + u \leq i_\rho$ ). Supporting overlaps is a generalization, since the set of all maximal  $\alpha$ -gapped repeats *supporting* overlaps ( $i_\lambda + 1 \leq i_\rho$ ) consists of all maximal  $\alpha$ -gapped repeats *prohibiting* overlaps ( $i_\lambda + u \leq i_\rho$ ) and those maximal gapped repeats *whose arms overlap* ( $i_\lambda + 1 \leq i_\rho \leq i_\lambda + u - 1$ ). This generalization makes the maximality property more natural, since a left/right extension of a gapped repeat (resp. an inward extension of a gapped palindrome) is always a gapped repeat (resp. gapped palindrome). For instance, an equivalent definition for an  $\alpha$ -gapped repeat  $(i_\lambda, i_\rho, u)$  to be maximal is that it holds the two conditions

- (a)  $i_\lambda = 1$  or  $w[i_\lambda - 1] \neq w[i_\rho - 1]$  and
- (b)  $i_\rho + u - 1 = |w|$  or  $w[i_\rho + u] \neq w[i_\lambda + u]$ .

If we prohibit overlaps, then a gapped repeat  $(i_\lambda, i_\rho, u)$  with  $i_\lambda + u = i_\rho$  and  $w[i_\lambda - 1] = w[i_\rho - 1]$  (i.e., condition (a) is violated) cannot be extended to the left (since its arms would overlap), which seems odd according to this definition.

**Example 1.1.** The first two characters of  $w = \mathbf{aaa}$  form a gapped repeat  $(1, 2, 1)$ . The right extensions  $(1, 2, 2)$  of both arms is only a gapped repeat if overlaps are supported. Similarly,  $(1, 3, 1)$  is a gapped palindrome, but the inward extension  $(1, 2, 2)$  is a gapped palindrome only if overlaps are supported. When prohibiting overlaps, only the gapped repeat  $(1, 3, 1)$  is maximal. When supporting overlaps, the gapped repeat  $(1, 2, 2)$  and the gapped palindromes  $(1, 1, 2)$ ,  $(2, 2, 2)$ , and  $(1, 1, 3)$  are also maximal.

Another interesting property of supporting overlaps is that the maximal  $\alpha$ -gapped repeats in  $w$  are a representation of the  $\alpha$ -gapped repeats in  $w$ , since an  $\alpha$ -gapped repeat is contained in a maximal  $\alpha$ -gapped repeat. This is not true when prohibiting overlaps: In this setting, an  $\alpha$ -gapped repeat that is not contained in a maximal  $\alpha$ -gapped repeat is contained in a run, as already noted by Kolpakov et al. [17, Fig. 3 and 4].

A natural question arising from this generalization is whether we can still compute the set of all maximal  $\alpha$ -gapped repeats and palindromes within the same bounds when supporting overlaps. We can answer this question affirmatively in the penultimate section of this article. Throughout this article, we heavily borrow the notations and ideas evolved by Gawrychowski et al. [10] and Kolpakov et al. [17].

The question on the number of all maximal  $\alpha$ -gapped repeats and palindromes has also found an answer for the average case, which was studied by Duchon et al. [7] who analyzed the *expected* number of maximal  $\alpha$ -gapped repeats/palindromes in a word.

## 2 Preliminaries

A (real) **interval**  $\mathcal{I} = [b, e] \subset \mathbb{R}$  for  $b, e \in \mathbb{R}$  is the set of all real numbers  $i \in \mathbb{R}$  with  $b \leq i \leq e$ . We write  $[b, e)$ ,  $(b, e]$  or  $(b, e)$  if  $e$ ,  $b$ , or both values are not included in the interval. For an interval  $\mathcal{I}$ ,  $\mathbf{b}(\mathcal{I})$  and  $\mathbf{e}(\mathcal{I})$  denote the beginning and end of  $\mathcal{I}$ , respectively.

A special kind of intervals are integer intervals  $\mathcal{I} = [b..e]$ , where  $\mathcal{I}$  is the set of consecutive integers from  $b = \mathbf{b}(\mathcal{I}) \in \mathbb{Z}$  to  $e = \mathbf{e}(\mathcal{I}) \in \mathbb{Z}$ , for  $b \leq e$ . We write  $|\mathcal{I}|$  to denote the length of  $\mathcal{I}$ ; i.e.,  $|\mathcal{I}| = \mathbf{e}(\mathcal{I}) - \mathbf{b}(\mathcal{I}) + 1$ .

Let  $\Sigma$  be a finite alphabet; an element of  $\Sigma$  is called **character**.  $\Sigma^*$  denotes the set of all finite **words** over  $\Sigma$ . The **length** of a word  $w \in \Sigma^*$  is denoted by  $|w|$ . For  $v = xuy$  with  $x, u, y \in \Sigma^*$ , we call  $x$ ,  $u$  and  $y$  a **prefix**, **factor**, and **suffix** of  $v$ , respectively. We denote by  $w[i]$  the character occurring at position  $i$  in  $w$ , and by  $w[i..j]$  the factor of  $w$  starting at position  $i$  and ending at position  $j$ , consisting of the catenation of the characters  $w[i], \dots, w[j]$ , where  $1 \leq i \leq j \leq n$ ;  $w[i..j]$  is the empty word if  $i > j$ . By  $w^\top$  we denote the **reverse** of  $w$ .

The notation  $w[b..e]$  can be ambivalent: it can denote both a factor and the occurrence of this factor starting at position  $b$  in  $w$ . The second entity is called the segment<sup>1</sup>  $w[b..e]$ : A **segment**  $w[b..e]$  of a word  $w$  is the occurrence of a factor  $f$  equal to  $w[b..e]$  in  $w$ ; we say that  $f$  **occurs** at position  $b$  in  $w$ . While a factor is identified only by a sequence of characters, a segment is also identified by its position in the word. A conclusion is that segments are always unique, while a word may contain multiple occurrences of the same factor. We use the same notation for defining factors and segments of a word. For two segments  $u$  and  $\bar{u}$  of a word  $w$ , we write  $u \equiv \bar{u}$  if they start at the same position in  $w$  and have the same length. We write  $u = \bar{u}$  if the factors identifying these segments are the same (hence  $u \equiv \bar{u} \Rightarrow u = \bar{u}$ ). We implicitly use segments both like factors of  $w$  and as intervals contained in  $[1..|w|]$ , e.g., we write  $u \subseteq \bar{u}$  if two segments  $u := w[b..e], \bar{u} := w[\bar{b}..\bar{e}]$  of  $w$  satisfy  $[b..e] \subseteq [\bar{b}..\bar{e}]$ , i.e.,  $\mathbf{b}(\bar{u}) \leq \mathbf{b}(u) \leq \mathbf{e}(u) \leq \mathbf{e}(\bar{u})$ .

A **period** of a word  $w$  over  $\Sigma$  is a positive integer  $p < |w|$  such that  $w[i] = w[j]$  for all  $i$  and  $j$  with  $1 \leq i, j \leq |w|$  and  $i \equiv j \pmod{p}$ . A word  $w$  whose smallest period is at most  $\lfloor |w|/2 \rfloor$  is called **periodic**; otherwise,  $w$  is called **aperiodic**. A **repetition** in a word  $w$  is a periodic factor; a **run** is a maximal repetition; the **exponent** of a run is the (rational) number of times the smallest period fits in that run. The exponent of a run  $r$  is denoted by  $\exp(r)$ . The sum of the exponents of runs in the word  $w$  is denoted by  $\mathcal{E}(w)$ . We use the following results from the literature:

**Lemma 2.1** ([8]). Given a word  $w$  with two periods  $p$  and  $p'$  such that  $p + p' \leq |w|$ , the greatest common divisor  $\gcd(p, p')$  of  $p$  and  $p'$  is also a period of  $w$ .

**Corollary 2.2.** A periodic factor  $u$  in a word  $w$  with the smallest period  $p$  cannot have two distinct occurrences  $u_\lambda$  and  $u_\rho$  in  $w$  with  $|\mathbf{b}(u_\lambda) - \mathbf{b}(u_\rho)| < p$ .

*Proof.* Since the smallest period of  $u$  is  $p$ ,  $|u| > 2p$  holds. Assume for a contradiction that two distinct occurrences  $u_\lambda$  and  $u_\rho$  of  $u$  exist in  $w$  with a distance  $\delta := \mathbf{b}(u_\rho) - \mathbf{b}(u_\lambda)$  such that  $0 < \delta < p$  (see also Fig. 1).

<sup>1</sup>This notion was coined in [5].

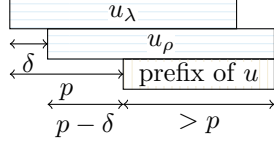


Figure 1: Setting of the proof of Corollary 2.2 with  $\delta < p$ . There are two occurrences  $u_\lambda$  and  $u_\rho$  of  $u$  with an overlap of  $2p - \delta$  characters. Both occurrences induce a run of period  $\delta$ . There are at least three occurrences of  $u$ 's prefix of length  $p + 1$  (starting at  $\mathbf{b}(u_\lambda)$ ,  $\mathbf{b}(u_\rho)$ , and  $\mathbf{b}(u_\lambda) + p$ ).

Since  $|u_\lambda \cap u_\rho| \geq 2p - \delta \geq p$ ,  $\delta$  is a period of  $u$ . Additionally, since  $u$  has the smallest period  $p$ , there is another occurrence of a prefix of  $u$  starting at  $\mathbf{b}(u_\lambda) + p - \delta$  with a length of at least  $p + \delta > p$ . Hence,  $p - \delta$  is also a period of  $u$ . Because the sum of both periods  $\delta$  and  $p - \delta$  is less than  $|u|$ , Lemma 2.1 states that  $\gcd(\delta, p - \delta) < p$  is a period of  $u$ . This contradicts the fact that  $p$  is the smallest period of  $u$ .  $\square$

**Lemma 2.3** ([1]). For a word  $w$ ,  $\mathcal{E}(w) < 3|w|$ .

Instead of working with triples of integers  $(i_\lambda, i_\rho, u)$  as in Section 1 when representing gapped repeats and palindromes, we stick to pairs of segments  $(w[i_\lambda..i_\lambda + u - 1], w[i_\rho..i_\rho + u - 1])$  for convenience: For a word  $w$ , we call a pair of segments  $(u_\lambda, u_\rho)$  a ***gapped repeat*** (resp. ***gapped palindrome***) with ***period***  $q = \mathbf{b}(u_\rho) - \mathbf{b}(u_\lambda)$  iff

- $\mathbf{b}(u_\lambda) + 1 \leq \mathbf{b}(u_\rho)$  and  $u_\rho = u_\lambda$  in the case of a gapped repeat, or
- $\mathbf{b}(u_\lambda) \leq \mathbf{b}(u_\rho)$  and  $u_\rho = u_\lambda^\top$  in the case of a gapped palindrome (it is possible that  $u_\lambda \equiv u_\rho$ ).

The segments  $u_\lambda$  and  $u_\rho$  are called left and right ***arm***, respectively. The value  $\mathbf{b}(u_\rho) - \mathbf{e}(u_\lambda) - 1$  is called the ***gap***, and is the distance between both arms in case that it is positive. For  $\alpha \geq 1$ , the gapped repeat or gapped palindrome  $(u_\lambda, u_\rho)$  is called  ***$\alpha$ -gapped*** iff its period  $q$  is at most  $\alpha|u_\lambda|$ .

Given a gapped repeat  $(u_\lambda, u_\rho)$ , it is called ***maximal*** iff the characters to the immediate left and to the immediate right of its arms differ (as far as they exist), i.e.,

- $w[\mathbf{b}(u_\lambda) - 1] \neq w[\mathbf{b}(u_\rho) - 1]$  (or  $\mathbf{b}(u_\lambda) = 1$ ) and
- $w[\mathbf{e}(u_\lambda) + 1] \neq w[\mathbf{e}(u_\rho) + 1]$  (or  $\mathbf{e}(u_\rho) = |w|$ ).

Similarly, a gapped palindrome  $(u_\lambda, u_\rho)$  is called ***maximal*** iff it can be extended neither inwards nor outwards, i.e.,

- $w[\mathbf{b}(u_\lambda) - 1] \neq w[\mathbf{e}(u_\rho) + 1]$  (or  $\mathbf{b}(u_\lambda) = 1$  or  $\mathbf{e}(u_\rho) = n$ ) and
- $w[\mathbf{e}(u_\lambda) + 1] \neq w[\mathbf{b}(u_\rho) - 1]$ .

Let  $\mathcal{G}_\alpha(w)$  (resp.  $\mathcal{G}_\alpha^\top(w)$ ) denote the set of all maximal  $\alpha$ -gapped repeats (resp. palindromes) in  $w$ .

Gapped palindromes generalize the definition of ordinary palindromes: A gapped palindrome  $(u_\lambda, u_\rho)$  is an ***ordinary palindrome*** if  $u_\lambda \equiv u_\rho$ . For a maximal gapped palindrome with a gap  $\mathbf{b}(u_\rho) - \mathbf{e}(u_\lambda) - 1 \leq 1$  it follows that  $u_\lambda \equiv u_\rho$  (otherwise it could be extended inwards).

### 3 Improved Point Analysis

A pair of integers is called a ***point***. In [10], a certain subset  $C$  of maximal  $\alpha$ -gapped repeats and maximal  $\alpha$ -gapped palindromes are mapped to points injectively. The cardinality of  $C$  is estimated with the property

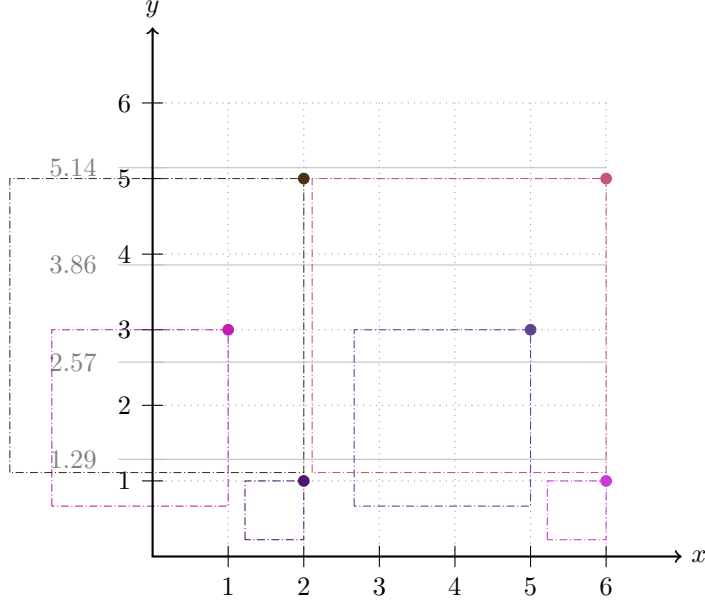


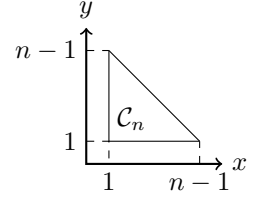
Figure 2:  $7/9$ -cover of the points  $\{(4x-2-(y+1 \bmod 2), 2y-1) \mid 1 \leq x, y \leq 3\} \subset \mathbb{N}^2$ . The dash-dotted square of a point  $\vec{p}$  comprises all points that are  $7/9$ -covered by  $\vec{p}$  (the square of  $\vec{p}$  is the square that has  $\vec{p}$  as its top right vertex). A point  $(x, y)$  with  $y = 1$  only  $7/9$ -covers itself. The light-gray dotted lines create the grid  $\mathbb{N}^2$ . For each integer  $i$  with  $1 \leq i \leq 4$  the  $y$ -axis is label with the value  $i/\gamma$  with  $\gamma := 7/9$ . Additionally, there is a gray horizontal line having the height  $i/\gamma$ .

that every point of  $C$  has a large vicinity that does not contain another point of  $C$ . This vicinity is given formally by the following definition:

**Definition 3.1.** For a real number  $\gamma$  with  $\gamma \in (0, 1]$ , we say that a point  $(\hat{x}, \hat{y}) \in \mathbb{Z}^2$   $\gamma$ -covers a point  $(x, y) \in \mathbb{Z}^2$  iff  $\hat{x} - \gamma\hat{y} \leq x \leq \hat{x}$  and  $\hat{y}(1 - \gamma) \leq y \leq \hat{y}$ .

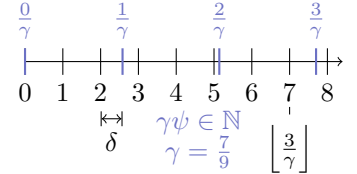
Figure 2 gives an example for  $\gamma := 7/9$ . In Lemma 7 of [10], it is shown that  $|C| < 3n/\gamma$  holds for every set of points  $C \subseteq [1..n]^2$  with the property that no two distinct points in  $C$   $\gamma$ -cover the same point. In the following, we devise an improved version of this lemma to upper bound the number of the  $\beta$ -aperiodic repeats/palindromes.

For our purpose, it is sufficient to focus on the set  $\mathcal{C}_n := \{(x, y) \in \mathbb{Z}^2 \mid 1 \leq y \leq n-1 \text{ and } 1 \leq x \leq n-y\}$ , since we will later show that we can map all maximal  $\alpha$ -gapped repeats/palindromes to the set injectively. Before that, we introduce two small helper lemmas that improve an inequality needed in Lemma 3.4:



**Lemma 3.2.** Given a real interval  $\mathcal{I} := [\psi - 1/\gamma, \psi)$  with  $\gamma, \psi \in \mathbb{R}$  and  $\gamma \in (0, 1]$ ,

$$|\mathcal{I} \cap \mathbb{Z}| = \begin{cases} \lfloor 1/\gamma \rfloor + 1 & \text{if } 0 < \psi - \lfloor \psi \rfloor \leq \delta, \\ \lfloor 1/\gamma \rfloor & \text{otherwise,} \end{cases}$$



where  $\mathcal{I} \cap \mathbb{Z} = \{i \in \mathbb{Z} \mid i \in \mathcal{I}\}$  and  $\delta := 1/\gamma - \lfloor 1/\gamma \rfloor$ .

*Proof.* In the case that  $\psi = \lfloor \psi \rfloor$  (i.e.,  $\psi \in \mathbb{Z}$ ),  $\mathbf{b}(\mathcal{I}) = \psi - 1/\gamma \leq \psi - \lfloor 1/\gamma \rfloor \in \mathcal{I} \cap \mathbb{Z}$ . Hence,  $\{\psi - \lfloor 1/\gamma \rfloor, \dots, \psi - 1\} = \mathcal{I} \cap \mathbb{Z}$ , and  $|\mathcal{I} \cap \mathbb{Z}| = \lfloor 1/\gamma \rfloor$ .

In the case that  $0 < \psi - \lfloor \psi \rfloor \leq \delta$ , we have  $\psi - \delta \leq \lfloor \psi \rfloor$ , and therefore  $\mathbf{b}(\mathcal{I}) = \psi - 1/\gamma = \psi - \lfloor 1/\gamma \rfloor - \delta \leq \lfloor \psi \rfloor - \lfloor 1/\gamma \rfloor \in \mathcal{I} \cap \mathbb{Z}$ . Hence,  $\{\lfloor \psi \rfloor - \lfloor 1/\gamma \rfloor, \dots, \lfloor \psi \rfloor\} = \mathcal{I} \cap \mathbb{Z}$ , and  $|\mathcal{I} \cap \mathbb{Z}| = \lfloor 1/\gamma \rfloor + 1$  (because  $\lfloor \psi \rfloor < \psi$ ).

The remaining case is that  $\psi - \lfloor \psi \rfloor > \delta$ . With  $\lfloor \psi \rfloor < \psi - \delta = \psi - 1/\gamma + \lfloor 1/\gamma \rfloor$ , we obtain that  $\mathfrak{b}(\mathcal{I}) = \psi - 1/\gamma > \lfloor \psi \rfloor - \lfloor 1/\gamma \rfloor \notin \mathcal{I} \cap \mathbb{Z}$ . Hence,  $\{\lfloor \psi \rfloor - \lfloor 1/\gamma \rfloor + 1, \dots, \lfloor \psi \rfloor\} = \mathcal{I} \cap \mathbb{Z}$ , and  $|\mathcal{I} \cap \mathbb{Z}| = \lfloor 1/\gamma \rfloor$ .  $\square$

**Lemma 3.3.** Given the function  $g_\gamma : \mathbb{N} \rightarrow \mathbb{N}$  with  $g_\gamma(i) := |\{y \in \mathbb{N} \mid (i-1)/\gamma \leq y < i/\gamma\}|$  for  $1 \leq i \leq \lceil n\gamma \rceil$ , and a nonincreasing function  $f : \mathbb{N} \rightarrow \mathbb{R}$ , the inequality

$$\sum_{i=1}^{\lceil n\gamma \rceil} (f(i)g_\gamma(i)) \leq \sum_{i=1}^{\lceil n\gamma \rceil} f(i)/\gamma \quad (1)$$

holds for every natural number  $n$  and every real number  $\gamma \in (0, 1]$ .

*Proof.* We set  $Y_i := \{y \in \mathbb{N} \mid (i-1)/\gamma \leq y < i/\gamma\}$ . Our task is to upper bound the sizes of  $Y_i$ , since  $g_\gamma(i) = |Y_i|$ . It is clear that  $|Y_i| \leq \lfloor 1/\gamma \rfloor + 1$ . Since  $Y_1$  cannot contain zero, it holds that  $|Y_1| \leq \lfloor 1/\gamma \rfloor$  (if  $1/\gamma \in \mathbb{N}$  then  $|Y_1| = 1/\gamma - 1$ , otherwise  $|Y_1| = \lfloor 1/\gamma \rfloor$ ). For  $i \geq 2$ , Lemma 3.2 provides that

$$|Y_i| = \lfloor 1/\gamma \rfloor + 1 \text{ iff } 0 < i/\gamma - \lfloor i/\gamma \rfloor \leq \delta, \text{ where } \delta := 1/\gamma - \lfloor 1/\gamma \rfloor < 1. \quad (2)$$

Having Eq. (2), Eq. (1) is a conclusion of the following game estimating the cumulative sum of  $f(i)/\gamma - f(i)g_\gamma(i)$ : The game is divided in  $\lceil n\gamma \rceil$  rounds. In the  $i$ -th round ( $1 \leq i \leq \lceil n\gamma \rceil$ ), we receive a credit of  $(1/\gamma - \lfloor 1/\gamma \rfloor)f(i) = \delta f(i)$ , but we additionally pay  $f(i)$  from the credit when  $g_\gamma(i) = \lfloor 1/\gamma \rfloor + 1$ . If the credit does not become negative, it holds that  $\sum_{i=1}^{\lceil n\gamma \rceil} (f(i)g_\gamma(i)) \leq \sum_{i=1}^{\lceil n\gamma \rceil} f(i)/\gamma$  (which is what we want to show in this proof).

Let  $i_1, i_2, \dots$  be the sequence of integers such that  $g_\gamma(i_j) = \lfloor 1/\gamma \rfloor + 1$  for each  $j$ . After sorting this sequence ascendingly, it holds that  $\delta i_j > j$  for every  $j$ . To see this, we write  $i/\gamma - \lfloor i/\gamma \rfloor = i/\gamma - i \lfloor 1/\gamma \rfloor - \lfloor i/\gamma - i \lfloor 1/\gamma \rfloor \rfloor = \delta i - \lfloor \delta i \rfloor$ , and apply Eq. (2): First,  $\delta i_1 \geq 1$ , since otherwise ( $\delta i_1 < 1$ ) we obtain a contradiction to Eq. (2) with  $\delta i_1 - \lfloor \delta i_1 \rfloor = \delta i_1 > 2\delta$  (remember that  $i_1 \geq 2$  because  $|Y_1| \leq \lfloor 1/\gamma \rfloor$ ). Next, assume that there exists a  $j \geq 2$  such that  $j \leq \delta i_j < \delta i_{j+1} < j+1$ . Then  $\delta i_{j+1} - \lfloor \delta i_{j+1} \rfloor \geq \delta(i_j+1) - \lfloor \delta i_j \rfloor > \delta$  (since  $\delta i_j - \lfloor \delta i_j \rfloor > 0$ ), a contradiction that Eq. (2) holds for  $i_{j+1}$ . We conclude that  $\delta i_j > j$  for every  $j$ .

Back to our game, we claim that there is at least  $(\delta i_j - j)f(i_j)$  credit remaining after the  $i_j$ -th round. When reaching the  $i_1$ -th round, we have already gathered a credit of  $\sum_{i=1}^{i_1} \delta f(i)$ . Remember that we have to pay the amount  $f(i_1)$ . From our gathered credit we can pay  $f(i_1)$  with  $s := \delta f(1) + \delta f(2) + \dots + \delta f(i_1 - 1) + (1 - \delta(i_1 - 1))f(i_1)$ : First,  $s$  is smaller than our gathered credit, since  $f(i_1) < \delta i_1 f(i_1)$ , and hence  $(1 - \delta(i_1 - 1))f(i_1) < \delta f(i_1)$ . Second,  $s \geq f(i_1)$ , because  $\delta(i_1 - 1)f(i_1) \leq \sum_{i=1}^{i_1-1} \delta f(i)$  (remember that  $f$  is nonincreasing). By paying the amount  $s$ , a credit of at least  $f(i_1)(\delta i_1 - 1)$  remains.

Under the assumption that our claim holds after the  $i_j$ -th round for an integer  $j \in \mathbb{N}$ , we show that the claim holds after the  $i_{j+1}$ -th round, too: According to our assumption, we have gathered a credit of at least  $(\delta i_j - j)f(i_j) + \sum_{i=i_j+1}^{i_{j+1}} \delta f(i)$  at the beginning of the  $i_{j+1}$ -th round. We pay the amount  $f(i_{j+1})$  with  $s := (\delta i_j - j)f(i_j) + \delta f(i_j + 1) + \dots + \delta f(i_{j+1} - 1) + (j + 1 - \delta(i_{j+1} - 1))f(i_{j+1})$ . First,  $s$  is smaller than our gathered credit, since  $\delta i_{j+1} > j + 1$ , and hence  $(j + 1 - \delta(i_{j+1} - 1))f(i_{j+1}) < \delta f(i_{j+1})$ . Second,  $s \geq f(i_{j+1})$ , because  $\delta(i_{j+1} - 1)f(i_{j+1}) \leq (\delta i_j - j)f(i_j) + jf(i_{j+1}) + \sum_{i=i_j+1}^{i_{j+1}-1} \delta f(i)$ . Similar to the  $i_1$ -th round, a credit of at least  $(\delta i_{j+1} - j - 1)f(i_{j+1})$  remains.  $\square$

**Lemma 3.4.** Let  $\gamma$  be a real number with  $\gamma \in (0, 1]$ , and  $C \subseteq \mathcal{C}_n$  be a set of points such that no two distinct points in  $C$   $\gamma$ -cover the same point. Then  $|C| < n\pi^2/(6\gamma)$ . In particular,  $|C| \leq n(\pi^2/6 - 3/4)$  for  $\gamma = 1$ .

*Proof.* Given that a point  $\vec{p}$  in  $\mathbb{Z}^2$  is  $\gamma$ -covered by a point  $(\hat{x}, \hat{y})$  of  $C$  with  $(i-1)/\gamma \leq \hat{y} < i/\gamma$  for a positive integer  $i$ , we assign  $\vec{p}$  the weight  $1/i^2$ . Otherwise ( $\vec{p}$  is not  $\gamma$ -covered by any point of  $C$ ), we assign  $\vec{p}$  the weight zero. Let us fix a point  $(\hat{x}, \hat{y}) \in C$  with  $(i-1)/\gamma \leq \hat{y} < i/\gamma$  for an integer  $i$ . We have  $\hat{x} - i < \hat{x} - \gamma\hat{y} \leq \hat{x} - (i-1)$ , and these inequalities also hold when substituting  $\hat{x}$  with  $\hat{y}$ , i.e.,  $\hat{y} - i < \hat{y} - \gamma\hat{y} \leq \hat{y} - (i-1)$ . There are exactly  $i^2$  points  $(x, y) \in \mathbb{Z}^2$  that are  $\gamma$ -covered by  $(\hat{x}, \hat{y})$ , since for each of them it holds that  $\hat{x} - i < \hat{x} - \gamma\hat{y} \leq \hat{x} - (i-1) \leq x \leq \hat{x}$  and  $\hat{y} - i < \hat{y} - \gamma\hat{y} \leq \hat{y} - (i-1) \leq y \leq \hat{y}$ . Therefore, the sum of the weights of the points that are  $\gamma$ -covered by  $(\hat{x}, \hat{y})$  is one. As a consequence, the size of  $C$  is equal to the sum of the weights of all points in  $\mathbb{Z}^2$ . In the following, let  $w(\vec{p})$  denote the weight of a point  $\vec{p}$ . In what follows, we upper bound the sum of all weights.

First, we fix an integer  $y$  with  $1 \leq y \leq n$ , and show that the sum of the weights of all points  $(\cdot, y)$  is at most  $n/i^2$ , where  $i$  is the integer with  $(i-1)/\gamma \leq y < i/\gamma$ . Given an integer  $x \in \mathbb{Z}$ , we conclude by the definition of  $\mathcal{C}_n$  that

$$w(x, y) \begin{cases} \leq 1/i^2 & \text{for } 1 \leq x \leq n - y, \text{ and} \\ = 0 & \text{for } x \geq n - y + 1. \end{cases}$$

The sum  $\sum_{x=-\infty}^1 w(x, y)$  is maximized to  $1/i$  when each point in  $E := \{(x, y) \in \mathbb{Z}^2 \mid 2 - i \leq x \leq 1\}$  with  $|E| = i$  has weight  $1/i^2$ , and the other points  $\{(x, y) \in \mathbb{Z}^2 \mid x \leq 1 - i\}$  are not  $\gamma$ -covered. This can be seen by the following fact: A point  $(x, y)$  with  $x \leq 1 - i$  can only be  $\gamma$ -covered by a point  $(\hat{x}, \hat{y}) \in \mathcal{C}_n$  when  $\hat{x} - \gamma\hat{y} \leq x \leq 1 - i$ , or equivalently  $i \leq \gamma\hat{y}$  (the smallest value for  $\hat{x}$  is one). Assume that such a point  $(\hat{x}, \hat{y})$  exists. Then there is an integer  $j$  with  $i < j$  such that  $\gamma\hat{y} < j$  and  $(j-1)/\gamma \leq \hat{y} < j/\gamma$ . Since  $1 - j \leq \hat{x} - \gamma\hat{y} \leq x \leq 1$ , there are at most  $|\{(x, y) \mid 2 - j \leq x \leq 1\}| = j$  many different values for  $x$ . Furthermore, since  $(\hat{x}, \hat{y}) \in \mathcal{C}_n$   $\gamma$ -covers  $(x, y)$ , it is not possible that another element of  $\mathcal{C}_n$   $\gamma$ -covers  $(x', y)$  with  $x' < x$  (otherwise it would also cover  $(x, y)$ ). In total, the sum under consideration  $\sum_{x \leq 1} w(x, y)$  can be at most  $1/j$ , which is less than  $1/i$ . With  $\sum_{x \leq 1} w(x, y) \leq 1/i$  we obtain  $\sum_{x \in \mathbb{Z}} w(x, y) \leq (n - y - 1 + i)/i^2 \leq (n - y + \gamma y)/i^2 \leq n/i^2$ .

Having computed  $\sum_{x \in \mathbb{Z}} w(x, y)$  for a fixed  $y$ , we compute the sum over all  $y$  with  $y \in \mathbb{Z}$ . First, we deal with the special case that  $\gamma = 1$ . That is because it is the only case where  $w(\cdot, 0)$  might not be zero (given  $(\hat{x}, \hat{y}) \in \mathcal{C}_n$  and  $\gamma < 1$ , it holds that  $\hat{y} \geq 1$  and therefore  $0 < \hat{y} - \gamma\hat{y}$ ). A point  $(x, y)$  is 1-covered by  $(\hat{x}, \hat{y}) \in \mathcal{C}_n$  iff  $0 \leq y \leq \hat{y}$  and  $\hat{x} - \hat{y} \leq x \leq \hat{y}$  hold. The weight of a point  $(x, 0)$  with  $0 \leq x \leq n - 1$  is maximized to  $1/2^2$  if it is  $\gamma$ -covered by a point  $(\hat{x}, \hat{y}) \in \mathcal{C}_n$  with the lowest possible value of  $\hat{y}$ , which is one. We conclude that  $\sum_{x \in \mathbb{Z}} w(x, 0) \leq n/2^2$ . With the same argument we conclude that  $\sum_{x \in \mathbb{Z}} w(x, y) \leq n/(y+1)^2$  for every positive integer  $y$ . Summing up everything yields  $\sum_{(x, y) \in \mathbb{Z}^2} w(x, y) \leq n/2^2 + n \sum_{y=1}^n (1/(y+1)^2) = n/4 + n \sum_{i=2}^{\infty} (1/i^2) = n/4 + n\pi^2/6 - n = n\pi^2/6 - 3n/4$  due to the Basel problem.

Finally we consider the case that  $\gamma < 1$ . The idea is to cover the interval  $[1..n-1]$  with the sets  $Y_i := \{y \in \mathbb{N} \mid (i-1)/\gamma \leq y < i/\gamma\}$  for  $1 \leq i \leq \lceil n\gamma \rceil$ . Since a point  $(x, y_i)$  with  $y_i \in Y_i$  has a weight of at most  $1/i^2$ , summing up all weights gives  $\sum_{(x, y) \in \mathbb{Z}^2} w(x, y) \leq \sum_{i=1}^{\lceil n\gamma \rceil} n|Y_i|/i^2$ . To compute  $|Y_i|$ , we use the function  $g_\gamma(i) := |Y_i|$  as defined in Lemma 3.3. With  $g_\gamma$  the upper bound of  $\sum_{(x, y) \in \mathbb{Z}^2} w(x, y)$  can be stated as  $\sum_{i=1}^{\lceil n\gamma \rceil} (g_\gamma(i)n/i^2)$ . Since  $g_\gamma(i) \leq \lfloor 1/\gamma \rfloor + 1$ , it is easy to see that  $\sum_{i=1}^{\lceil n\gamma \rceil} (g_\gamma(i)n/i^2) < n(\lfloor 1/\gamma \rfloor + 1) \sum_{i=1}^{\lceil n\gamma \rceil} (1/i^2) < n(\lfloor 1/\gamma \rfloor + 1)\pi^2/6$ . By defining the non-increasing function  $f$  with  $f(i) := n/i^2$ , Lemma 3.3 yields  $\sum_{i=1}^{\lceil n\gamma \rceil} (g_\gamma(i)n/i^2) = \sum_{i=1}^{\lceil n\gamma \rceil} g_\gamma(i)f(i) \leq (n/\gamma) \sum_{i=1}^{\lceil n\gamma \rceil} (1/i^2) < \sum_{i=1}^{\infty} n/(\gamma i^2) = n\pi^2/(6\gamma)$ , which is also an upper bound of  $|C|$ .  $\square$

By restricting the subset  $C \subseteq \mathcal{C}_n$  in Lemma 3.4 to be additionally bijective to the set of all maximal  $\alpha$ -gapped repeats or palindromes, we can refine the upper bound attained in Lemma 3.4. For the maximal  $\alpha$ -gapped repeats, we follow the approach of Gawrychowski et al. [10] who map a maximal  $\alpha$ -gapped repeat  $(u_\lambda, u_\rho)$  with period  $q := b(u_\rho) - b(u_\lambda)$  to  $(e(u_\lambda), q)$ . It holds that  $(e(u_\lambda), q) \in \mathcal{C}_n$ , because  $e(u_\rho)$  and  $q$  are positive, and  $e(u_\lambda) + q = e(u_\rho) \leq n$ . In particular we have  $e(u_\lambda) \leq n - 1$ , since otherwise ( $e(u_\lambda) = n$ ) both endings  $e(u_\rho)$  and  $e(u_\lambda)$  would be equal, and therefore  $u_\lambda \equiv u_\rho$  (a contradiction to the definition of gapped repeats). Let  $\varphi$  denote the mapping from  $(u_\lambda, u_\rho)$  to  $(e(u_\lambda), q)$ , and let  $\varphi(\mathcal{G}_\alpha(w)) := \{\varphi(u_\lambda, u_\rho) \mid (u_\lambda, u_\rho) \text{ is a maximal } \alpha\text{-gapped repeat}\} \subset \mathcal{C}_n$  denote the image of  $\varphi$ . The following lemma bounds the size of  $\varphi(\mathcal{G}_\alpha(w))$  to be roughly at half of the size of  $\mathcal{C}_n$ , a fact that will be used in Lemma 3.6.

**Lemma 3.5.** If  $(x, y) \in \varphi(\mathcal{G}_\alpha(w))$ , then  $(x + 1, y) \notin \varphi(\mathcal{G}_\alpha(w))$ .

*Proof.* Let  $(u_\lambda, u_\rho)$  be a maximal  $\alpha$ -gapped  $\beta$ -aperiodic repeat with period  $q = b(u_\rho) - b(u_\lambda)$ , and  $(x, y) := \varphi(u_\lambda, u_\rho) = (e(u_\lambda), q)$ . If  $(x + 1, y) \in \varphi(\mathcal{G}_\alpha(w))$ , then  $w[x + 1] = w[e(u_\lambda) + 1] = w[x + y + 1] = w[e(u_\rho) + 1]$ , which contradicts the maximality of  $(u_\lambda, u_\rho)$ .  $\square$

With Lemma 3.5 we attain a version of Lemma 3.4 tailored to subsets of  $\varphi(\mathcal{G}_\alpha(w))$ :

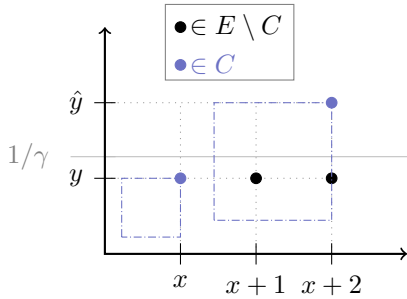


Figure 3: Setting of the proof of Lemma 3.6, where the point  $(x, y) \in C$ , but  $(x + 1, y) \notin C$  with  $w(x + 1, y) > 0$ . Thus  $(x + 1, y)$  is  $\gamma$ -covered by a point  $(\hat{x}, \hat{y}) \in C$  ( $\hat{x} = x + 2$  in this figure). Like in Fig. 2, the dash-dotted rectangle of a point  $\vec{p} \in C$  comprises all points that are  $\gamma$ -covered by  $\vec{p}$ . The points that are  $\gamma$ -covered by  $(\hat{x}, \hat{y})$  are contained in the top right dashed rectangle. It can be seen that  $(x + 2, y)$  is also  $\gamma$ -covered by  $(\hat{x}, \hat{y})$ , and therefore cannot be in  $C$ .

**Lemma 3.6.** Let  $\gamma$  be a real number with  $\gamma \in (0, 1]$ . A set of points  $C \subseteq \varphi(\mathcal{G}_\alpha(w))$  such that no two distinct points in  $C$   $\gamma$ -cover the same point obeys the inequality  $|C| < n(\pi^2/6 - 1/2)/\gamma$ .

*Proof.* If  $\gamma = 1$ , Lemma 3.4 already gives  $|C| < n\pi^2/6 - 3n/4 < n\pi^2/6 - n/2$ . For the case  $\gamma < 1$ , we focus on the points  $E := \{(x, y) \mid 1 \leq x \leq n \text{ and } y < 1/\gamma\}$ . In the proof of Lemma 3.4, we used the weights  $w(\cdot)$  of all points in  $\mathbb{Z}^2$  as an upper of  $|C|$ . There, we bounded the sum of the weights of all points in  $E$  by  $n/\gamma$  (assign each point the weight 1). We can refine this upper bound by halving the weights of the points in  $E$ . We justify this with the following analysis.

First, each point  $(n, y) \in E$  has weight zero, since there is no point  $(\hat{x}, \hat{y}) \in C$  (and even in  $\mathcal{C}_n$ ) with  $n \leq \hat{x}$ . Thus the sum of the weights of the points  $(n - 1, y)$  and  $(n, y)$  is at most one, for every  $y$  with  $1 \leq y < 1/\gamma$ .

Second, a point  $(x, y) \in E \cap C$  can only cover itself, since  $y < 1/\gamma$ . Consequently, a point  $(x, y) \in E \setminus C$  can have a weight of at most  $1/2^2 = 1/4$ , since all points  $(\hat{x}, \hat{y}) \in C \setminus E$  have  $\hat{y} \geq 1/\gamma$ . Given that  $E \cap C = \emptyset$ , the total weight of all points in  $E$  is at most  $(1/4)|E|$ .

Finally, suppose there is a point  $(x, y) \in E \cap C$ . Then  $w(x, y) = 1$ . Given  $x \leq n - 2$ ,  $(x + 1, y) \in \mathcal{C}_n$ , but  $(x + 1, y) \notin C$  according to Lemma 3.5. We consider two cases:

- $w(x + 1, y) = 0$ . Then both points  $(x, y)$  and  $(x + 1, y)$  together have a weight of one.



- $w(x+1, y) > 0$ , see also Fig. 3. Since  $(x+1, y) \notin C$ ,  $w(x+1, y) \leq 1/4$ , i.e., it is  $\gamma$ -covered by a point  $(\hat{x}, \hat{y}) \in C \setminus E$ . Since  $\hat{y} \geq 1/\gamma$ , the point  $(\hat{x}, \hat{y})$   $\gamma$ -covers at least four points (including itself). Since  $w(x, y) = 1$ ,  $(\hat{x}, \hat{y})$  cannot  $\gamma$ -cover  $(x, y)$ . Instead, it  $\gamma$ -covers the point  $(x+2, y)$ . We conclude that  $(x+2, y) \notin C$ . All three points  $(x, y)$ ,  $(x+1, y)$ , and  $(x+2, y)$  have a total weight of at most  $1 + 1/4 + 1/4 = 3/2$ .

In both cases, a node has the average weight of at most  $1/2$ . Summing up all average weights yields the total weight of all points in  $E$ , which is at most  $(1/2)|E| = n/(2\gamma)$ .

Following the proof of Lemma 3.4, our modification of the weights modifies the nonincreasing function  $f$ , which is now defined by  $f(1) := n/2$  and  $f(i) := n/i^2$  for  $i \geq 2$ . Modifying  $f$  yields the upper bound  $\sum_{i=1}^{\lceil n\gamma \rceil} (f(i)g_\gamma(i)) \leq n(1/2 + \sum_{i=2}^{\lceil n\gamma \rceil} (1/i^2))/\gamma < n(\pi^2/6 - 1/2)/\gamma$  on the size of  $C$ .  $\square$

## 4 On the Number of all Maximal $\alpha$ -gapped Repeats and Palindromes

Like previous approaches [10, 17], we split the set of all maximal  $\alpha$ -gapped repeats  $\mathcal{G}_\alpha(w)$  and the set of all maximal  $\alpha$ -gapped palindromes  $\mathcal{G}_\alpha^\top(w)$  into subsets, and analysis of these subsets individually.

Given a positive number  $\beta \in \mathbb{R}$ , a maximal  $\alpha$ -gapped repeat  $(u_\lambda, u_\rho)$  (resp. a maximal  $\alpha$ -gapped palindrome  $(u_\lambda, u_\rho)$  with  $u_\lambda \neq u_\rho$ ) belongs to the set of all maximal  $\alpha$ -gapped  $\beta$ -periodic repeats  $\beta\mathcal{P}_\alpha(w)$  (resp. palindromes  $\beta\mathcal{P}_\alpha^\top(w)$ ) if  $u_\rho$  contains a periodic prefix of length at least  $\beta|u_\lambda|$ . We call the elements of  $\beta\mathcal{P}_\alpha(w)$  and  $\beta\mathcal{P}_\alpha^\top(w)$   $\beta$ -**periodic**. A maximal  $\alpha$ -gapped repeat is called  $\beta$ -**aperiodic** if it is not  $\beta$ -periodic. A maximal  $\alpha$ -gapped palindrome  $(u_\lambda, u_\rho)$  is called  $\beta$ -**aperiodic** if it is neither  $\beta$ -periodic nor a maximal ordinary palindrome. The set of all maximal  $\alpha$ -gapped  $\beta$ -aperiodic repeats and the set of all maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindromes are denoted by  $\overline{\beta\mathcal{P}_\alpha}(w)$  and  $\overline{\beta\mathcal{P}_\alpha^\top}(w)$ , respectively.

Let us summarize for the gapped palindromes that we partition the set of all maximal  $\alpha$ -gapped palindromes  $\mathcal{G}_\alpha^\top(w)$  in

- the set of all maximal  $\alpha$ -gapped  $\beta$ -periodic palindromes  $\beta\mathcal{P}_\alpha^\top(w)$ ,
- the set of all maximal ordinary palindromes, and
- the set of all maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindromes  $\overline{\beta\mathcal{P}_\alpha^\top}(w)$ .

The size of the second set is known to be at most  $2|w| - 1$ . In the following, we give an upper bound on the number of maximal  $\alpha$ -gapped palindromes that are  $\beta$ -periodic or  $\beta$ -aperiodic with Lemma 4.1 or Corollary 4.6, respectively.

**Lemma 4.1.** Let  $w$  be a string. Further let  $\alpha$  and  $\beta$  be two real numbers with  $\alpha > 1$  and  $0 < \beta < 1$ . Then

( $\mathcal{P}$ )  $|\beta\mathcal{P}_\alpha(w)|$  is at most  $2\alpha\mathcal{E}(w)/\beta$ , and

( $\mathcal{P}^\top$ )  $|\beta\mathcal{P}_\alpha^\top(w)|$  is at most  $2(\alpha - 1)\mathcal{E}(w)/\beta + 2n$ .

*Proof.* Let  $(u_\lambda, u_\rho) \in \beta\mathcal{P}_\alpha(w)$  (resp.  $\in \beta\mathcal{P}_\alpha^\top(w)$ ) be a maximal  $\alpha$ -gapped  $\beta$ -periodic repeat (resp. palindrome). By definition, the right arm  $u_\rho$  has a periodic prefix  $s_\rho$  of length at least  $\beta|u_\rho|$ . Let  $r_\rho$  denote the run that

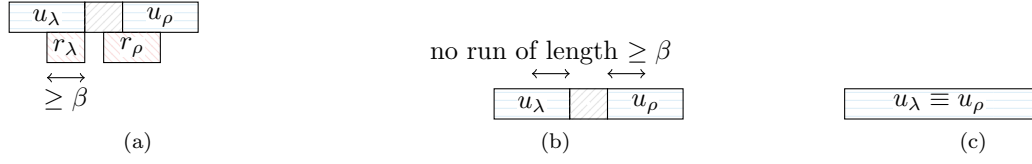


Figure 4: Types of maximal  $\alpha$ -gapped palindromes  $(u_\lambda, u_\rho)$  under consideration. The segments  $r_\lambda$  and  $r_\rho$  are runs. The gapped palindromes in Fig. (a) and Fig. (b) are  $\beta$ -periodic and  $\beta$ -aperiodic, respectively. The gapped palindrome in Fig. (c) is an ordinary palindrome.

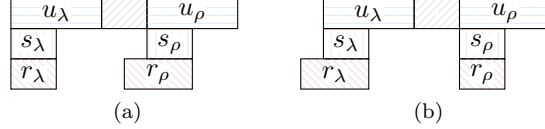


Figure 5: Setting of the proof of Lemma 4.1( $\mathcal{P}$ ). Each figure shows a maximal  $\alpha$ -gapped  $\beta$ -periodic repeat  $(u_\lambda, u_\rho)$  and the periodic prefixes  $s_\lambda$  and  $s_\rho$  of its respective arms  $u_\lambda$  and  $u_\rho$ . The periodic prefixes are contained in the runs  $r_\lambda$  and  $r_\rho$ , respectively. The equation (a)  $\mathbf{b}(u_\lambda) = \mathbf{b}(r_\lambda)$  or (b)  $\mathbf{b}(u_\rho) = \mathbf{b}(r_\rho)$  must hold. By the maximality property of runs,  $\mathbf{e}(r_\lambda) = \mathbf{e}(s_\lambda)$  and  $\mathbf{e}(r_\rho) = \mathbf{e}(s_\rho)$ , i.e.,  $s_\lambda \equiv r_\lambda \cap u_\lambda$  and  $s_\rho \equiv r_\rho \cap u_\rho$ .

generates  $s_\rho$ , i.e.,  $s_\rho \subseteq r_\rho$ . The two segments  $s_\rho$  and  $r_\rho$  have the smallest period  $p$  in common. By the definition of the gapped repeats (resp. palindromes), there is a prefix  $s_\lambda$  of  $u_\lambda$  (resp. suffix  $s_\lambda$  of  $u_\lambda$ ) with  $s_\lambda = s_\rho$  (resp.  $s_\lambda = s_\rho^\top$ ). Let  $r_\lambda$  be the run generating  $s_\lambda$ . By definition,  $r_\lambda$  has the same smallest period  $p$  as  $r_\rho$ .

**( $\mathcal{P}$ ) Gapped Repeats.** Since  $(u_\lambda, u_\rho)$  is maximal,  $\mathbf{b}(u_\lambda) = \mathbf{b}(r_\lambda)$  or  $\mathbf{b}(u_\rho) = \mathbf{b}(r_\rho)$  must hold (see Fig. 5); otherwise we could extend  $(u_\lambda, u_\rho)$  to the left.

The periodic  $\alpha$ -gapped repeat  $(u_\lambda, u_\rho)$  is uniquely determined by its arm-period  $q$  and

- (a)  $r_\lambda$  in case  $\mathbf{b}(u_\lambda) = \mathbf{b}(r_\lambda)$ , or
- (b)  $r_\rho$  in case  $\mathbf{b}(u_\rho) = \mathbf{b}(r_\rho)$ .

Since  $(u_\lambda, u_\rho)$  is  $\alpha$ -gapped, it holds that  $q \leq \alpha u$  with  $u := |u_\lambda|$ . We analyze Case (a), where  $\mathbf{b}(u_\lambda) = \mathbf{b}(s_\lambda) = \mathbf{b}(r_\lambda)$  holds. Case (b) is treated exactly in the same way by symmetry. The gapped repeat  $(u_\lambda, u_\rho)$  is identified by its arm-period  $q$  and  $r_\lambda$ . We fix  $r_\lambda$  and pose the question of how many maximal periodic gapped repeats can be generated by  $r_\lambda$ . We answer this question by counting the number of possible values for the arm-period  $q$ . Since the starting position  $\mathbf{b}(s_\rho) = \mathbf{b}(u_\rho) = \mathbf{b}(u_\lambda) + q = \mathbf{b}(r_\lambda) + q$  of the periodic segment  $s_\rho$  is determined by  $q$ , two possible values of  $q$  must have a distance of at least  $p$  due to Corollary 2.2.

It is left to find a lower and an upper bound for the value of  $q$ . For the lower bound,  $q = \mathbf{b}(u_\rho) - \mathbf{b}(u_\lambda) = \mathbf{b}(s_\rho) - \mathbf{b}(s_\lambda)$  is at least  $u \geq 2p$  in case that the arms  $u_\lambda$  and  $u_\rho$  are not overlapping. Otherwise (in case

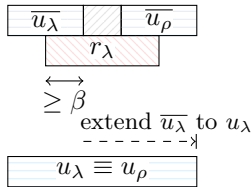


Figure 6: A  $\beta$ -periodic gapped palindrome  $(\overline{u}_\lambda, \overline{u}_\rho)$  with a run  $r_\lambda$  covering the suffix (resp. prefix) of length  $\beta$  of the left arm  $\overline{u}_\lambda$  (resp. right arm  $\overline{u}_\rho$ ). According to the proof of Lemma 4.1( $\mathcal{P}^\top$ ), we can make  $(\overline{u}_\lambda, \overline{u}_\rho)$  maximal by extending it inwards to form an ordinary palindrome  $(u_\lambda, u_\rho)$  with  $u_\lambda \equiv u_\rho$ , which is of the same type as Fig. 4c.

the arms overlap),  $q$  is at least  $p$  because the starting positions of the periodic segments  $s_\lambda$  and  $s_\rho$  have a distance of at least  $p$  (again due to Corollary 2.2). For the upper bound, with  $u \leq |s_\lambda|/\beta$  and  $q \leq \alpha u$ , we obtain  $q \leq |s_\lambda| \alpha/\beta \leq |r_\lambda| \alpha/\beta$ .

Finally, we have that  $p \leq q \leq |r_\lambda| \alpha/\beta$ . Then the number of possible arm-periods  $q$  is at most  $|r_\lambda| \alpha/(\beta p) = \exp(r_\lambda) \alpha/\beta$ . Overall, the number of all maximal  $\alpha$ -gapped repeats is at most  $\alpha \mathcal{E}(w)/\beta$  for the case  $\mathbf{b}(u_\lambda) = \mathbf{b}(r_\lambda)$ . Since Case (b) with  $\mathbf{b}(u_\rho) = \mathbf{b}(r_\rho)$  is symmetric, we get the total upper bound  $2\alpha \mathcal{E}(w)/\beta$ .

**( $\mathcal{P}^\top$ ) Gapped Palindromes.** If  $r_\lambda \equiv r_\rho$  (see Fig. 6), then either  $\mathbf{b}(u_\rho) - \mathbf{e}(u_\lambda) \leq 2$  (i.e.,  $(u_\lambda, u_\rho)$  is an ordinary palindrome), or  $(u_\lambda, u_\rho)$  is not maximal. That is because of the following: Assume that  $r_\lambda$  contains  $s_\lambda$  and  $s_\rho$ . Then we have  $w[\mathbf{e}(s_\lambda) + 1] = w[\mathbf{e}(s_\lambda) - p + 1] = w[\mathbf{b}(s_\rho) + p - 1] = w[\mathbf{b}(s_\rho) - 1]$ , where the first and third equality follows from  $|s_\rho| = |s_\lambda| \geq 2p$ , and the second equality follows from  $s_\rho = s_\lambda^\top$ .

From now on, we assume that  $r_\lambda \not\equiv r_\rho$ . Since  $(u_\lambda, u_\rho)$  is maximal,  $\mathbf{e}(u_\lambda) = \mathbf{e}(r_\lambda)$  or  $\mathbf{b}(u_\rho) = \mathbf{b}(r_\rho)$  must hold; otherwise we could extend  $(u_\lambda, u_\rho)$  inwards. This means that  $(u_\lambda, u_\rho)$  is uniquely determined by the gap  $v := \mathbf{b}(u_\rho) - \mathbf{e}(u_\lambda) - 1$  and

- (a)  $r_\lambda$  in case  $\mathbf{e}(u_\lambda) = \mathbf{e}(r_\lambda)$ , or
- (b)  $r_\rho$  in case  $\mathbf{b}(u_\rho) = \mathbf{b}(r_\rho)$ .

Since ordinary palindromes are excluded from the set of all maximal  $\alpha$ -gapped  $\beta$ -periodic palindromes, the gap  $v$  is at least two. Cases (a) and (b) are depicted in Fig. 7.



Figure 7: Setting of the proof of Lemma 4.1( $\mathcal{P}^\top$ ). Each figure depicts a maximal  $\alpha$ -gapped  $\beta$ -periodic palindrome  $(u_\lambda, u_\rho)$  with the periodic suffix  $s_\lambda$ . The periodic suffix  $s_\lambda \equiv r_\lambda \cap u_\lambda$  of  $u_\lambda$  and the periodic prefix  $s_\rho \equiv r_\rho \cap u_\rho$  of  $u_\rho$  are the intersections of the runs  $r_\lambda$  and  $r_\rho$  with the respective arms. By the maximality property of runs, the equation (a)  $\mathbf{e}(u_\lambda) = \mathbf{e}(r_\lambda)$  or (b)  $\mathbf{b}(u_\rho) = \mathbf{b}(r_\rho)$  must hold.

We analyze Case (a) with  $\mathbf{e}(s_\lambda) = \mathbf{e}(r_\lambda)$ , Case (b) is treated exactly in the same way by symmetry. The gapped palindrome  $(u_\lambda, u_\rho)$  is identified by its gap  $v \geq 2$  and  $r_\lambda$ . We fix  $r_\lambda$  and count the number of possible values of  $v$ . Since the starting position  $\mathbf{b}(s_\rho) = \mathbf{e}(r_\lambda) + v + 1$  of the periodic segment  $s_\rho$  is determined by  $v$ , two possible values of  $v$  must have a distance of at least  $p$  due to Corollary 2.2. Since  $|u_\lambda| \leq |s_\lambda|/\beta$  and  $(u_\lambda, u_\rho)$  is  $\alpha$ -gapped,  $v \leq (\alpha - 1)|u_\lambda| \leq (\alpha - 1)|s_\lambda|/\beta$ . Then the number of possible values for  $v$  is bounded by  $1 + |s_\lambda|(\alpha - 1)/(\beta p) = 1 + |r_\lambda|(\alpha - 1)/(\beta p) = 1 + \exp(r_\lambda)(\alpha - 1)/\beta$ . In total, the number of maximal  $\alpha$ -gapped palindromes in this case is bounded by  $n + (\alpha - 1)\mathcal{E}(w)/\beta$  for the case  $\mathbf{e}(u_\lambda) = \mathbf{e}(r_\lambda)$ . Case (b) is symmetric, leading to the bound of  $2(\alpha - 1)\mathcal{E}(w)/\beta + 2n$  in total.  $\square$

We are already done with the analysis of the gapped repeats, since we can use a result of Gawrychowski et al. [10] that does not assume that  $\mathbf{e}(u_\lambda) < \mathbf{b}(u_\rho)$ , and therefore supports gapped repeats with overlap:

**Lemma 4.2** ([10, Lemma 9]). Given a string  $w$ , and two real numbers  $\alpha, \beta$  with  $\alpha > 1$  and  $2/3 \leq \beta < 1$ , the points mapped by two different maximal  $\beta$ -aperiodic  $\alpha$ -gapped repeats in  $\overline{\beta\mathcal{P}_\alpha}(w)$  cannot  $\frac{1-\beta}{\alpha}$ -cover the same point.

Combining Lemmas 4.2 and 3.6 with Lemma 4.1( $\mathcal{P}$ ) gives the final result:

**Theorem 4.3.** Given a real number  $\alpha$  with  $\alpha > 1$  and a word  $w$  of length  $n$ , the number of all  $\alpha$ -gapped repeats  $|\mathcal{G}_\alpha(w)|$  is less than  $3(\pi^2/6 + 5/2)\alpha n$ .

*Proof.* Combining the results of Lemmas 4.2 and 3.6 yields that the set  $\overline{\beta\mathcal{P}_\alpha}(w)$  has at most  $(\pi^2/6 - 1/2)\alpha n / (1 - \beta)$  elements. Summing up the sizes of both sets yields  $|\mathcal{G}_\alpha(w)| < 2\alpha \mathcal{E}(w) / \beta + (\pi^2/6 - 1/2)\alpha n / (1 - \beta)$ . This number becomes minimal with  $|\mathcal{G}_\alpha(w)| < 9\alpha n + 3(\pi^2/6 - 1/2)\alpha n = 3(\pi^2/6 + 5/2)\alpha n$  when setting  $\beta$  to  $2/3$ .  $\square$

It is left to analyze all maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindromes. To apply the results of Section 3, we map maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindromes to points. Gawrychowski et al. [10] map a maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindrome  $(u_\lambda, u_\rho)$  to the point  $(\mathbf{e}(u_\lambda), v)$ , where  $v := \mathbf{b}(u_\rho) - \mathbf{e}(u_\lambda) - 1$  is the gap of  $(u_\lambda, u_\rho)$ . Since  $(u_\lambda, u_\rho)$  is  $\beta$ -aperiodic, the gap  $v$  is at least two (otherwise it could be extended inwards to a maximal ordinary palindrome). With  $\mathbf{e}(u_\lambda) + v = \mathbf{b}(u_\rho) - 1 \leq n - 1$ , we conclude that  $(\mathbf{e}(u_\lambda), v) \in \mathcal{C}_n$ . However, this mapping seems not useful in combination with *our* definition of the  $\beta$ -periodic gapped palindromes. Defining periodic gapped palindromes to have a left arm with a periodic *suffix* (instead of prefix as in [10]) invalidates the proof of Lemma 12 in [10]. There, we fail to transfer the contradiction in the Sub-Case 2b with  $2z - \delta < 0$  to our new definition: We want to derive a contradiction by showing that  $u_\lambda$  has a sufficiently large periodic suffix  $s_\lambda$  (in [10, Lemma 12], it was shown that  $u_\lambda$  has a sufficiently large periodic prefix). However, we have not found a way to upper bound the length of  $u_\lambda$ , and thus, we are not able to show that the periodic suffix  $s_\lambda$  is sufficiently large in relation to  $|u_\lambda|$ .

To solve this problem, we define an alternative mapping  $\varphi_\top$  that maps a maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindrome  $(u_\lambda, u_\rho)$  of a word of length  $n$  to the point

$$(m, d) := \varphi_\top(u_\lambda, u_\rho) := (\lceil (\mathbf{b}(u_\lambda) + \mathbf{e}(u_\lambda))/2 \rceil, \lfloor (\mathbf{b}(u_\rho) + \mathbf{e}(u_\rho))/2 \rfloor - \lceil (\mathbf{b}(u_\lambda) + \mathbf{e}(u_\lambda))/2 \rceil).$$

Let  $\varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w)) = \{\varphi_\top(u_\lambda, u_\rho) \mid (u_\lambda, u_\rho) \text{ is a maximal } \alpha\text{-gapped } \beta\text{-aperiodic palindrome}\} \subset \mathcal{C}_n$  be the image of  $\varphi_\top$ . The first coordinate  $m$  is the (integer) position nearest to the mid-point  $(\mathbf{b}(u_\lambda) + \mathbf{e}(u_\lambda))/2$  (tie-breaking to the *right*) of the left arm, and  $m + d$  is the position nearest to the mid-point  $(\mathbf{b}(u_\rho) + \mathbf{e}(u_\rho))/2$  (tie-breaking to the *left*) of the right arm (in particular,  $w[m] = w[m + d]$ ). The mapping  $\varphi_\top$  is injective because we can retrieve the pair of segments  $(u_\lambda, u_\rho)$  by computing the maximal inward and outward matches at the positions  $m$  and  $m + d$ . Since  $m$  and  $d$  are positive integers with  $m + d \leq n$ , we conclude that  $(m, d) \in \mathcal{C}_n$ . For convenience, we give an alternative definition of  $(m, d)$  using the function<sup>2</sup>  $c(i) := (i + 1 \bmod 2)/2$  such that  $c(i) = 0$  if  $i$  is odd, and  $c(i) = 1/2$  if  $i$  is even. With  $c(\mathbf{b}(u_\lambda) + \mathbf{e}(u_\lambda) + 1) = c(2\mathbf{b}(u_\lambda) + |u_\lambda|) = c(|u_\lambda|)$  we get  $(m, d) = ((\mathbf{b}(u_\lambda) + \mathbf{e}(u_\lambda))/2 + c(|u_\lambda|), q - 2c(|u_\lambda|))$ , where  $q := \mathbf{b}(u_\rho) - \mathbf{b}(u_\lambda)$  is the period of  $(u_\lambda, u_\rho)$  (see also Fig. 8).

**Fact 4.4.** Given a maximal gapped palindrome  $(u_\lambda, u_\rho)$  with  $u := |u_\lambda|$  and  $(m, d) := \varphi_\top(u_\lambda, u_\rho)$ , it holds that

$$(a) \quad |u/2 - c(u)| \in \{1/2, 3/2, 5/2, \dots\},$$

<sup>2</sup>The function is called  $c$  since it calculates a *correction* term that adds with  $(\mathbf{b}(u_\lambda) + \mathbf{e}(u_\lambda))/2$  up to the mid-point  $m$  of the left arm.

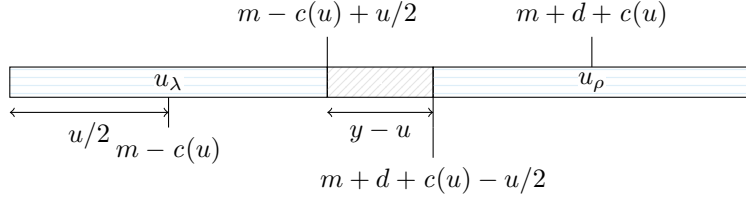


Figure 8: A gapped palindrome  $(u_\lambda, u_\rho)$  with  $u = |u_\lambda|$  mapped to the point  $(m, d)$ .

(b)  $\mathbf{b}(u_\lambda) = m - c(u) - u/2 + 1/2$ ,

(c)  $\mathbf{e}(u_\lambda) = m - c(u) + u/2 - 1/2$ , and

(d)  $\mathbf{b}(u_\rho) = m + d + c(u) - u/2 + 1/2$ .

(e) If  $d \leq 2$ , then  $(u_\lambda, u_\rho)$  is a maximal *ordinary* palindrome.

**Lemma 4.5.** Given a maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindrome  $(u_\lambda, u_\rho)$  with  $u := |u_\lambda|$  and  $(m, d) = \varphi_\top(u_\lambda, u_\rho)$ ,  $(m + i, d - 2i) \notin \varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w))$  for every integer  $i$  with  $-\lfloor u/2 \rfloor - 1 \leq i \leq -1$  or  $1 \leq i \leq \lceil u/2 \rceil$ .

*Proof.* For every integer  $i$  with  $-\lfloor u/2 \rfloor \leq i \leq -1$  or  $1 \leq i \leq \lceil u/2 \rceil - 1$  (excluding  $-\lfloor u/2 \rfloor - 1$  and  $\lceil u/2 \rceil$  as stated in the claim), the maximal inward and outward matches at the positions  $m + i$  and  $m + d - i$  yields  $(u_\lambda, u_\rho)$ , and thus,  $(m + i, d - 2i)$  cannot be in  $\varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w))$  due to the injectivity of  $\varphi_\top$  (cf. Fig. 9). If  $i$  is  $-\lfloor u/2 \rfloor - 1$  or  $\lceil u/2 \rceil$ , the point  $(m + i, d - 2i) \in \mathbb{Z}^2$  is not in  $\varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w))$  because the pair of positions  $m + i$  and  $m + d - i$  is where the inward or outward match from the positions  $m$  and  $m + i$  fails.  $\square$

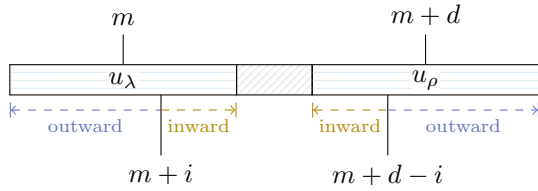


Figure 9: Setting of Lemma 4.5. A gapped palindrome  $(u_\lambda, u_\rho)$  is mapped to the point  $(m, d)$ . It can be restored by longest common prefix and suffix queries at the positions  $(m + i, d - 2i)$  for every integer  $i$  with  $-\lfloor u/2 \rfloor - 1 \leq i \leq -1$  or  $1 \leq i \leq \lceil u/2 \rceil$ , where  $u := |u_\lambda| = |u_\rho|$ .

Due to Lemma 4.5, each point  $(m, d) \in \varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w))$  has at least one distinct point that is not in the image of  $\varphi_\top$ . For instance, we count the point  $(m + 1, d - 2) \in \mathcal{C}_n \setminus \varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w))$  ( $d \geq 3$  according to Fact 4.4(e)) for each  $(m, d) \in \varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w))$ , and each counted point is counted only once. With this insight, we can prove the next corollary in exactly the same way as Lemma 3.6.

**Corollary 4.6.** Let  $\gamma$  be a real number with  $\gamma \in (0, 1]$ . A set of points  $C \subseteq \varphi_\top(\overline{\beta\mathcal{P}_\alpha}^\top(w))$  such that no two distinct points in  $C$   $\gamma$ -cover the same point obeys the inequality  $|C| < n(\pi^2/6 - 1/2)/\gamma$ .

*Proof.* With the same definition of  $E$  as in the proof of Lemma 3.6, it is left to show that the sum of the weights of all points in  $E$  is at most  $n/(2\gamma)$ .

Unlike the proof of Lemma 3.6, we can take a shortcut with the following observation: Only the highest points in  $E$  can be  $\gamma$ -covered by a point from  $C \setminus E$ .<sup>3</sup> To see this, let  $(x, y) \in E$  be a point with  $y < 1/\gamma - 1$ . Assume that  $(\hat{x}, \hat{y})$   $\gamma$ -covers  $(x, y)$ , then  $\hat{y} - \gamma\hat{y} \leq y < 1/\gamma - 1$ , or equivalently  $\hat{y} < 1/\gamma$ . This means that  $(x, y) = (\hat{x}, \hat{y})$ .

<sup>3</sup>This holds also in the case of maximal  $\alpha$ -gapped repeats in the proof of Lemma 3.6. However, this trick does not lead to anything useful there.

We conclude that every point  $(x, y) \in E$  with  $w(x, y) = 1$  belongs to  $C$ , and therefore (a)  $(x+1, y-2) \notin C$  according to Lemma 4.5, and (b)  $w(x+1, y-2) = 0$  according to the above observation. Hence, both points  $(x, y)$  and  $(x+1, y-2)$  have a total weight of 1 (remember that  $w(n, y-2) = 0$  in any case, cf. proof of Lemma 3.6). For  $y \leq 2$ , there is no aperiodic  $\alpha$ -gapped palindrome that can be mapped to a point  $(x, y)$  according to Fact 4.4(e).

Although a highest point  $(x, y)$  (with  $1/\gamma - 1 \leq y$ ) can be  $\gamma$ -covered by a point in  $C \setminus E$ , one of its neighbors  $(x-1, y)$  or  $(x+1, y)$  has to be  $\gamma$ -covered by the same point, such that the sum of the weights of both points is at most  $1/2$ . The total weight of all points in  $E$  is therefore at most  $(1/2)|E| \leq n/(2\gamma)$ .  $\square$

Corollary 4.6 finally leads us to the connection between the  $\gamma$ -cover property and the maximal  $\alpha$ -gapped palindromes:

**Lemma 4.7.** Let  $w$  be a word, and  $\alpha$  and  $\beta$  two real numbers with  $\alpha > 1$  and  $6/7 \leq \beta < 1$ . The points mapped by two different maximal gapped palindromes in  $\overline{\beta\mathcal{P}}_\alpha^\top(w)$  cannot  $\frac{1-\beta}{\alpha}$ -cover the same point.

*Proof.* Let  $(u_\lambda, u_\rho)$  and  $(\overline{u}_\lambda, \overline{u}_\rho)$  be two different maximal  $\alpha$ -gapped palindromes in  $\overline{\beta\mathcal{P}}_\alpha^\top(w)$ . Set  $u := |u_\lambda| = |u_\rho|$  and  $\overline{u} := |\overline{u}_\lambda| = |\overline{u}_\rho|$ . Let  $(m, d)$  and  $(\overline{m}, \overline{d})$  be the points mapped from  $(u_\lambda, u_\rho)$  and  $(\overline{u}_\lambda, \overline{u}_\rho)$ , respectively. Assume, for the sake of contradiction, that both points  $\frac{1-\beta}{\alpha}$ -cover the same point  $(x, y)$ .

Let  $z := |m - \overline{m}|$ , and let  $s_\lambda := u_\lambda \cap \overline{u}_\lambda$  be the overlap of  $u_\lambda$  and  $\overline{u}_\lambda$ . Let  $s := |s_\lambda|$ , and let  $s_\rho$  (resp.  $\overline{s}_\rho$ ) be the reverse copy of  $s_\lambda$  based on  $(u_\lambda, u_\rho)$  (resp.  $(\overline{u}_\lambda, \overline{u}_\rho)$ ), i.e.,  $s_\lambda = s_\rho^\top = \overline{s}_\rho^\top$  with  $\mathbf{b}(s_\rho) = \mathbf{b}(u_\rho) + \mathbf{e}(u_\lambda) - \mathbf{e}(s_\lambda)$  and  $\mathbf{b}(\overline{s}_\rho) = \mathbf{b}(\overline{u}_\rho) + \mathbf{e}(\overline{u}_\lambda) - \mathbf{e}(s_\lambda)$ .

**Sub-Claim.** The overlap  $s_\lambda$  is not empty, and  $\mathbf{b}(s_\rho) \neq \mathbf{b}(\overline{s}_\rho)$ .

**Sub-Proof.** First we show that  $s_\lambda$  is not empty. If  $m = \overline{m}$ , it is clear that  $s_\lambda$  contains  $w[m]$ . Without loss of generality, assume that  $m < \overline{m}$  for this sub-proof (otherwise exchange  $(u_\lambda, u_\rho)$  with  $(\overline{u}_\lambda, \overline{u}_\rho)$ ). By combining (a) the  $(1-\beta)/\alpha$ -cover property with (b) the fact that  $(\overline{u}_\lambda, \overline{u}_\rho)$  is  $\alpha$ -gapped and (c) the constraint  $6/7 \leq \beta < 1$ , we obtain  $\overline{m} - \overline{u}/2 \stackrel{(c)}{\leq} \overline{m} - (1-\beta)\overline{u} \stackrel{(b)}{\leq} \overline{m} - \overline{d}(1-\beta)/\alpha \stackrel{(a)}{\leq} x \stackrel{(a)}{\leq} m < \overline{m}$ . This long inequality says that the text position  $m$  is contained in  $\overline{u}_\lambda$ , which implies that  $s_\lambda$  is not empty. If  $s_\rho$  and  $\overline{s}_\rho$  start at the same position, then expanding the arms  $s_\lambda$  and  $s_\rho$  ( $\equiv \overline{s}_\rho$ ) to the left and right yields the arms  $u_\lambda \equiv \overline{u}_\lambda$  and  $u_\rho \equiv \overline{u}_\rho$ , which implies that  $(u_\lambda, u_\rho)$  and  $(\overline{u}_\lambda, \overline{u}_\rho)$  are the same gapped repeat, a contradiction.  $\blacksquare$

Without loss of generality let  $d \leq \overline{d}$ . With the  $(1-\beta)/\alpha$ -cover property we obtain

$$\overline{d} - \frac{\overline{d}(1-\beta)}{\alpha} \leq y \leq d \leq \overline{d}. \quad (3)$$

The difference  $\delta := \overline{d} - d \geq 0$  can be estimated by

$$\delta \leq \overline{d}(1-\beta)/\alpha \leq \overline{u}(1-\beta). \quad (4)$$

Equation (3) can also be used to lower bound  $u$  in terms of  $\overline{d}$  due to the fact that  $(u_\lambda, u_\rho)$  is  $\alpha$ -gapped:

$$u \geq d/\alpha \geq \frac{\overline{d}}{\alpha} \left(1 - \frac{1-\beta}{\alpha}\right) \geq \overline{d}\beta/\alpha. \quad (5)$$

**Outline.** In the following we conduct a thorough case analysis. In each case we show the contradiction that

$(u_\lambda, u_\rho)$  or  $(\bar{u}_\lambda, \bar{u}_\rho)$  is  $\beta$ -periodic. We prove each case in a similar way: We first show that the intersection of  $s_\rho$  and  $\bar{s}_\rho$  is large enough such that it induces a repetition on  $s_\rho \cup \bar{s}_\rho$ . Subsequently, we find a run covering  $s_\rho \cup \bar{s}_\rho$ , and another run covering  $s_\lambda$ . However, since  $s_\lambda$  is the suffix of  $u_\lambda$  (resp.  $\bar{u}_\lambda$ ), we can conclude that  $(u_\lambda, u_\rho)$  (resp.  $(\bar{u}_\lambda, \bar{u}_\rho)$ ) is  $\beta$ -periodic.

Before starting with the case analysis, we introduce a general property of the starting positions  $\mathbf{b}(s_\rho)$  and  $\mathbf{b}(\bar{s}_\rho)$  needed for the analysis. Adding up the equalities of Fact 4.4(c) and (d) gives  $\mathbf{b}(u_\rho) + \mathbf{e}(u_\lambda) = 2m + d$ . With that we obtain  $\mathbf{b}(s_\rho) = \mathbf{b}(u_\rho) + \mathbf{e}(u_\lambda) - \mathbf{e}(s_\lambda) = 2m + d - \mathbf{e}(s_\lambda)$ . Hence, the distance between the starting positions of  $s_\rho$  and  $\bar{s}_\rho$  is given by

$$|\mathbf{b}(s_\rho) - \mathbf{b}(\bar{s}_\rho)| = \begin{cases} 2z + \delta & \text{if } m \leq \bar{m}, \\ 2z - \delta & \text{if } m > \bar{m} \text{ and } \mathbf{b}(s_\rho) > \mathbf{b}(\bar{s}_\rho), \text{ or} \\ \delta - 2z & \text{if } m > \bar{m} \text{ and } \mathbf{b}(s_\rho) < \mathbf{b}(\bar{s}_\rho). \end{cases} \quad (6)$$

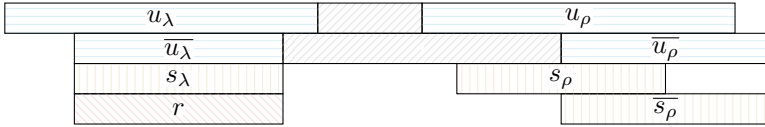


Figure 10: Sub-Case 1a in the proof of Lemma 4.7 with  $m \leq \bar{m}$  and  $\mathbf{b}(u_\lambda) \leq \mathbf{b}(\bar{u}_\lambda) \leq \mathbf{e}(\bar{u}_\lambda) \leq \mathbf{e}(u_\lambda)$ .

**Case 1:**  $m \leq \bar{m}$ . Since  $\bar{m} - \bar{d}(1 - \beta)/\alpha \leq x \leq m \leq \bar{m}$  (due to the  $(1 - \beta)/\alpha$ -cover property),

$$z = \bar{m} - m \leq \bar{d}(1 - \beta)/\alpha \leq \bar{u}(1 - \beta), \quad (7)$$

because  $(\bar{u}_\lambda, \bar{u}_\rho)$  is  $\alpha$ -gapped. Due to Eq. (6), the starting positions of both right copies  $\bar{s}_\rho$  and  $s_\rho$  differ by  $\mathbf{b}(\bar{s}_\rho) - \mathbf{b}(s_\rho) = 2z + \delta > 0$ . By Eqs. (4) and (7), we get

$$2z + \delta \leq 3\bar{d}(1 - \beta)/\alpha \leq 3\bar{u}(1 - \beta). \quad (8)$$

Depending on the relations  $\mathbf{b}(u_\lambda) \leq \mathbf{b}(\bar{u}_\lambda)$  and  $\mathbf{e}(u_\lambda) \leq \mathbf{e}(\bar{u}_\lambda)$ , we split the case in four sub-cases. However, one of the four sub-cases with  $\mathbf{b}(\bar{u}_\lambda) < \mathbf{b}(u_\lambda)$  and  $\mathbf{e}(\bar{u}_\lambda) < \mathbf{e}(u_\lambda)$  already leads to a contradiction (without proving that one left arm has a periodic suffix): Assume that both inequalities  $\mathbf{b}(\bar{u}_\lambda) < \mathbf{b}(u_\lambda)$  and  $\mathbf{e}(\bar{u}_\rho) < \mathbf{e}(u_\rho)$  hold for the sake of contradiction. Under these assumptions, with Fact 4.4(b) it must hold that  $\mathbf{b}(\bar{u}_\lambda) + 1/2 = \bar{m} - c(\bar{u}) - \bar{u}/2 + 1 \leq m - c(u) - u/2 = \mathbf{b}(u_\lambda) - 1/2$  and  $\mathbf{e}(\bar{u}_\rho) + 1/2 = \bar{m} - c(\bar{u}) + \bar{u}/2 + 1 \leq m - c(u) + u/2 = \mathbf{e}(u_\rho) - 1/2$ . Adding the left sides and the right sides of both inequalities gives  $\bar{m} - m \leq c(\bar{u}) - c(u) - 1 < 0$ , which contradicts that  $\bar{m} - m \geq 0$ .

Thus, it is enough to consider the following three sub-cases 1a, 1b, and 1c.

**Sub-Case 1a:**  $s_\lambda \equiv \bar{u}_\lambda$ , see Figure 10. Since  $\bar{u}/(2z + \delta) \geq \bar{u}/(3\bar{u}(1 - \beta)) \geq 7/3 > 2$  holds (due to Eq. (8)) for  $6/7 \leq \beta < 1$ , we conclude that  $s_\rho = \bar{u}_\rho$  is periodic, which means that  $(\bar{u}_\lambda, \bar{u}_\rho) \in \beta\mathcal{P}_\alpha^I(w)$ , a contradiction.

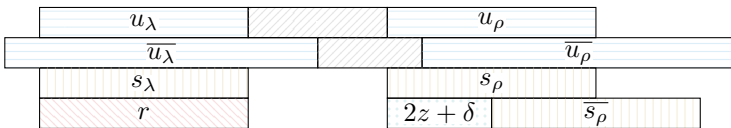


Figure 11: Sub-Case 1b in the proof of Lemma 4.7 with  $m \leq \bar{m}$  and  $\mathbf{b}(\bar{u}_\lambda) \leq \mathbf{b}(u_\lambda) \leq \mathbf{e}(u_\lambda) \leq \mathbf{e}(\bar{u}_\lambda)$ .

**Sub-Case 1b:**  $s_\lambda \equiv u_\lambda$ , see Figure 11. Recall that  $u = s \geq \bar{d}\beta/\alpha$  by Eq. (5). It follows from Eq. (8) and  $6/7 \leq \beta < 1$  that  $s/(2z + \delta) \geq \bar{d}\alpha\beta/(3\bar{d}\alpha(1 - \beta)) = \beta/(3(1 - \beta)) \geq 2$ . Hence  $s_\rho \equiv u_\rho$  is periodic, which means that  $(u_\lambda, u_\rho) \in \beta\mathcal{P}_\alpha^\Gamma(w)$ , a contradiction.

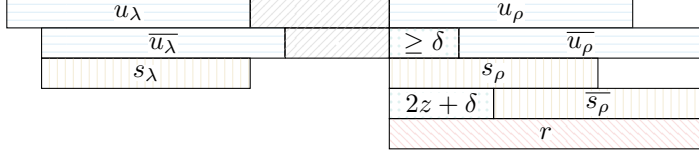


Figure 12: Sub-Case 1c in the proof of Lemma 4.7 with  $m \leq \bar{m}$  and  $\mathbf{b}(u_\lambda) < \mathbf{b}(\bar{u}_\lambda) \leq \mathbf{e}(u_\lambda) < \mathbf{e}(\bar{u}_\lambda)$ . The second inequality holds because the overlap  $s_\lambda$  cannot be empty due to the sub-claim.

**Sub-Case 1c:**  $\mathbf{b}(u_\lambda) < \mathbf{b}(\bar{u}_\lambda)$  and  $\mathbf{e}(u_\lambda) < \mathbf{e}(\bar{u}_\lambda)$ , see Figure 12. Since  $s_\lambda$  is a suffix of  $u_\lambda$  and a prefix of  $\bar{u}_\lambda$ , the reverse copies  $s_\rho$  and  $\bar{s}_\rho$  are a prefix of  $u_\rho$  and a suffix of  $\bar{u}_\rho$ , respectively. We have  $1 \leq \mathbf{b}(\bar{u}_\lambda) - \mathbf{b}(u_\lambda) = \bar{m} - \bar{u}/2 - c(\bar{u}) - (m - u/2 - c(u))$ . A simple reshaping leads to  $\bar{m} - \bar{u}/2 - (m - u/2) \geq 1 + c(\bar{u}) - c(u)$ . This inequality yields  $\mathbf{b}(\bar{u}_\rho) - \mathbf{b}(u_\rho) = \bar{m} + \bar{d} - \bar{u}/2 + c(\bar{u}) - (m + d - u/2 + c(u)) \geq \delta + 1 + 2(c(\bar{u}) - c(u)) \geq \delta \geq 0$ . This means that  $\mathbf{b}(s_\rho) = \mathbf{b}(u_\rho) \leq \mathbf{b}(\bar{u}_\rho) \leq \mathbf{b}(\bar{s}_\rho) \leq \mathbf{e}(\bar{s}_\rho) = \mathbf{e}(\bar{u}_\rho)$ . With  $\mathbf{b}(s_\rho) \leq \mathbf{b}(\bar{u}_\rho) = \mathbf{e}(\bar{s}_\rho) - \bar{u} + 1 = \mathbf{b}(\bar{s}_\rho) + s - \bar{u}$ , it follows that  $s \geq \bar{u} - (2z + \delta) > 2z + \delta$  because  $\bar{u}/(2z + \delta) \geq \bar{u}/(3\bar{u}(1 - \beta)) \geq 7/3 > 2$  holds for  $6/7 \leq \beta < 1$ . Since  $\mathbf{b}(\bar{s}_\rho) - \mathbf{b}(s_\rho) = \mathbf{e}(\bar{s}_\rho) - \mathbf{e}(s_\rho) = 2z + \delta$ ,  $s_\rho \cap \bar{s}_\rho \neq \emptyset$ . This means that  $\bar{u}_\rho \subset s_\rho \cup \bar{s}_\rho$ , and that  $\bar{u}_\rho$  is periodic with a period of at most  $2z + \delta$ , a contradiction.

**Case 2:**  $m > \bar{m}$ . Since  $m - d(1 - \beta)/\alpha \leq x \leq \bar{m} < m$ ,

$$z = m - \bar{m} \leq d(1 - \beta)/\alpha \leq \bar{d}(1 - \beta)/\alpha \leq \bar{u}(1 - \beta). \quad (9)$$

Due to Eq. (6), the starting positions of both right copies differ by  $|\mathbf{b}(s_\rho) - \mathbf{b}(\bar{s}_\rho)| = |2z - \delta|$ . Equation (6) with Eqs. (4) and (9) yields

$$|2z - \delta| \leq \begin{cases} 2z \leq 2\bar{d}(1 - \beta)/\alpha \leq 2\bar{u}(1 - \beta) & \text{if } \mathbf{b}(s_\rho) > \mathbf{b}(\bar{s}_\rho), \text{ or} \\ \delta \leq \bar{d}(1 - \beta)/\alpha \leq \bar{u}(1 - \beta) & \text{if } \mathbf{b}(s_\rho) < \mathbf{b}(\bar{s}_\rho). \end{cases} \quad (10)$$

Like Case 1, we split the case into sub-cases depending on the relation of the starting and of the ending positions of the left arms. The sub-case with  $\mathbf{b}(u_\lambda) < \mathbf{b}(\bar{u}_\lambda)$  and  $\mathbf{e}(u_\lambda) < \mathbf{e}(\bar{u}_\lambda)$  already leads to a contradiction, which can be seen by an argument that is similar to the one used in Case 1 due to symmetry.

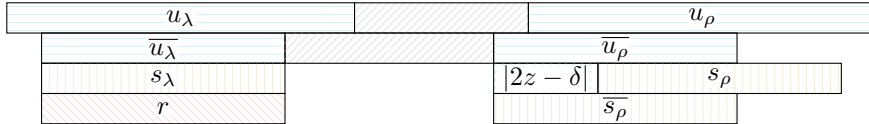


Figure 13: Sub-Case 2a in the proof of Lemma 4.7 with  $m > \bar{m}$  and  $\mathbf{b}(u_\lambda) \leq \mathbf{b}(\bar{u}_\lambda) \leq \mathbf{e}(\bar{u}_\lambda) \leq \mathbf{e}(u_\lambda)$ .

**Sub-Case 2a:**  $s_\lambda \equiv \bar{u}_\lambda$ , see Figure 13. Since  $s/|2z - \delta| \geq \bar{u}/(2\bar{u}(1 - \beta)) = 1/(2(1 - \beta)) \geq 7/2 > 2$  holds (due to Eq. (10)) for  $6/7 \leq \beta < 1$ , the distance between  $\mathbf{b}(s_\rho)$  and  $\mathbf{b}(\bar{s}_\rho)$  is small enough such that  $s_\rho = \bar{u}_\rho$  is periodic, which means that  $(\bar{u}_\lambda, \bar{u}_\rho) \in \beta\mathcal{P}_\alpha^\Gamma(w)$ , a contradiction.

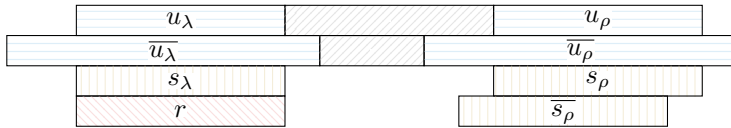


Figure 14: Sub-Case 2b in the proof of Lemma 4.7 with  $m > \bar{m}$  and  $\mathbf{b}(\bar{u}_\lambda) \leq \mathbf{b}(u_\lambda) \leq \mathbf{e}(u_\lambda) \leq \mathbf{e}(\bar{u}_\lambda)$ .

**Sub-Case 2b:**  $s_\lambda \equiv u_\lambda$ , see Figure 14. Recall that  $u = s \geq \bar{d}\beta/\alpha$  by Eq. (5). It follows from  $6/7 \leq \beta < 1$



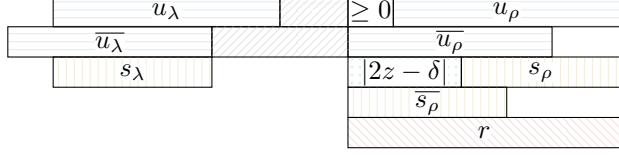


Figure 15: Subsub-Case 2c-i in the proof of Lemma 4.7 with  $m > \bar{m}$ ,  $\mathbf{b}(\bar{u}_\lambda) < \mathbf{b}(u_\lambda) \leq \mathbf{e}(\bar{u}_\lambda) < \mathbf{e}(u_\lambda)$  and  $\mathbf{b}(\bar{u}_\rho) \leq \mathbf{b}(u_\rho)$ .

and Eq. (10) that  $s/|2z - \delta| \geq \bar{d}\alpha\beta/(2\bar{d}\alpha(1 - \beta)) = \beta/(2(1 - \beta)) \geq 3 > 2$ . Hence  $s_\rho \equiv u_\rho$  is periodic, which means that  $(u_\lambda, u_\rho) \in \beta\mathcal{P}_\alpha^\Gamma(w)$ , a contradiction.

**Sub-Case 2c:**  $\mathbf{b}(u_\lambda) > \mathbf{b}(\bar{u}_\lambda)$  and  $\mathbf{e}(u_\lambda) > \mathbf{e}(\bar{u}_\lambda)$ . Since  $s_\lambda$  is a prefix of  $u_\lambda$  and a suffix of  $\bar{u}_\lambda$ , the reverse copies  $s_\rho$  and  $\bar{s}_\rho$  are a suffix of  $u_\rho$  and a prefix of  $\bar{u}_\rho$ , respectively.

**Subsub-Case 2c-i:**  $\mathbf{b}(u_\rho) \geq \mathbf{b}(\bar{u}_\rho)$ , see Figure 15. Recall that  $u \geq \bar{d}\beta/\alpha$  by Eq. (5). It follows from  $6/7 \leq \beta < 1$  and Eq. (10) that  $u/|2z - \delta| \geq \bar{d}\alpha\beta/(2\bar{d}\alpha(1 - \beta)) = \beta/(2(1 - \beta)) \geq 3 > 2$ . With  $\mathbf{b}(u_\rho) \geq \mathbf{b}(\bar{u}_\rho)$ , this case is symmetric to Sub-Case 1c, leading to the result that  $u_\rho \subset s_\rho \cup \bar{s}_\rho$ , and that  $u_\rho$  is periodic with a period of at most  $|2z - \delta|$ , a contradiction.

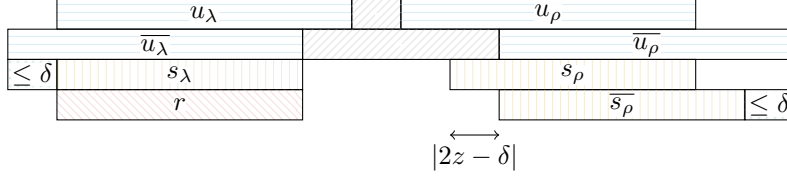


Figure 16: Subsub-Case 2c-ii in the proof of Lemma 4.7 with  $m > \bar{m}$ ,  $\mathbf{b}(\bar{u}_\lambda) < \mathbf{b}(u_\lambda) \leq \mathbf{e}(\bar{u}_\lambda) < \mathbf{e}(u_\lambda)$  and  $\mathbf{b}(u_\rho) < \mathbf{b}(\bar{u}_\rho)$ .

**Subsub-Case 2c-ii:**  $\mathbf{b}(u_\rho) < \mathbf{b}(\bar{u}_\rho)$ , see Figure 16. It follows from  $\mathbf{b}(u_\rho) < \mathbf{b}(\bar{u}_\rho)$  that  $\mathbf{b}(u_\rho) - \mathbf{b}(\bar{u}_\rho) = m + d - u/2 + c(u) - (\bar{m} + \bar{d} - \bar{u}/2 + c(\bar{u})) = \mathbf{b}(u_\lambda) - \mathbf{b}(\bar{u}_\lambda) - \delta + 2(c(u) - c(\bar{u})) \leq -1$ , which leads to  $\mathbf{b}(u_\lambda) - \mathbf{b}(\bar{u}_\lambda) \leq \delta + 2(c(\bar{u}) - c(u)) - 1 \leq \delta$ . Combining this inequality with Eq. (4) gives  $s = \bar{u} - (\mathbf{b}(u_\lambda) - \mathbf{b}(\bar{u}_\lambda)) \geq \bar{u} - \delta \geq \beta\bar{u}$ . With Eq. (10) this yields  $s/|2z - \delta| \geq \beta\bar{u}/(2\bar{u}(1 - \beta)) = \beta/(2(1 - \beta)) \geq 3 > 2$  under the presumption that  $6/7 \leq \beta < 1$ . This means that  $\bar{u}_\lambda$  has a periodic suffix of length  $\beta\bar{u}$ , and that  $(\bar{u}_\lambda, \bar{u}_\rho) \in \beta\mathcal{P}_\alpha^\Gamma(w)$ , a contradiction.  $\square$

Combining the results of Corollary 4.6 and Lemma 4.7 immediately gives the following corollary:

**Corollary 4.8.** Given two real numbers  $\alpha$  and  $\beta$  with  $\alpha > 1$  and  $6/7 \leq \beta < 1$ , and a word  $w$  of length  $n$ , the number of all maximal  $\alpha$ -gapped  $\beta$ -aperiodic palindromes is bounded by the inequality  $|\beta\overline{\mathcal{P}}_\alpha^\Gamma(w)| < \alpha n(\pi^2/6 - 1/2)/(1 - \beta)$ .

**Theorem 4.9.** Given a real number  $\alpha$  with  $\alpha > 1$ , and a word  $w$  of length  $n$ , the number of all maximal  $\alpha$ -gapped palindromes  $|\mathcal{G}_\alpha^\Gamma(w)|$  less than  $7(\pi^2/6 + 1/2)\alpha n - 3n - 1$ .

*Proof.* Combining the results of Lemma 4.1 and Corollary 4.8 yields

$$\begin{aligned}
|\mathcal{G}_\alpha^\Gamma(w)| &= \underbrace{2n - 1}_{\text{max. palindromes}} + \underbrace{|\beta\mathcal{P}_\alpha^\Gamma(w)|}_{\beta\text{-periodic}} + \underbrace{|\beta\overline{\mathcal{P}}_\alpha^\Gamma(w)|}_{\beta\text{-aperiodic}} \\
&< 2n - 1 + \underbrace{2(\alpha - 1)\frac{\mathcal{E}(w)}{\beta}}_{\text{Lemma 4.1}} + \underbrace{2n + \left(\frac{\pi^2}{6} - \frac{1}{2}\right)\frac{\alpha n}{1 - \beta}}_{\text{Corollary 4.8}}
\end{aligned}$$

for every  $6/7 \leq \beta < 1$ . Applying Lemma 2.3, the term on the right side is upper bounded by  $4n - 1 + 2(\alpha - 1)(3n/\beta) + (\pi^2/6 - 1/2)\alpha n/(1 - \beta)$ . This number is minimal when  $\beta = 6/7$ , yielding the bound  $4n - 1 + 7n(\alpha - 1) + 7(\pi^2/6 - 1/2)\alpha n = 7(\pi^2/6 + 1/2)\alpha n - 3n - 1$ .  $\square$

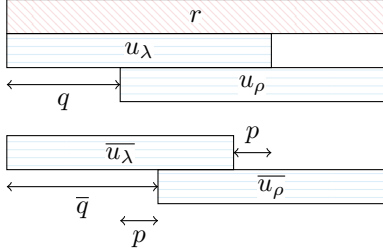


Figure 17: Two gapped repeats  $(u_\lambda, u_\rho)$  and  $(\bar{u}_\lambda, \bar{u}_\rho)$  with overlapping arms. Both gapped repeats are within a run  $r$ . They are maximal if their arms border the run  $r$ . Each such maximal gapped repeat with overlapping arms has a period ( $q$  or  $\bar{q}$  in the figure) that is a multiple of  $r$ 's period  $p$ .

## 5 A Linear Time Algorithm on Integer Alphabets

In this algorithmic section, we are given a word  $w$  of length  $n$  on an *integer* alphabet  $\Sigma$  as input such that  $|\Sigma| = n^{\mathcal{O}(1)}$ . We focus on the maximal  $\alpha$ -gapped repeats/palindromes  $(u_\lambda, u_\rho)$  with  $\mathbf{e}(u_\lambda) \geq \mathbf{b}(u_\rho)$ , which we call  $\alpha$ -gapped repeats (resp. palindromes) **with overlap**. In the following, we provide an  $\mathcal{O}(n)$  time algorithm that finds all  $\alpha$ -gapped repeats/palindromes with overlap.

When studying  $\alpha$ -gapped repeats/palindromes with overlap, we can neglect the parameter  $\alpha$ , because a gapped repeat/palindrome  $(u_\lambda, u_\rho)$  whose arms overlap obeys the inequality  $\mathbf{b}(u_\rho) - \mathbf{b}(u_\lambda) < |u_\lambda| \leq \alpha |u_\lambda|$  for every  $\alpha \geq 1$ . For a gapped palindrome  $(u_\lambda, u_\rho)$  with  $\mathbf{e}(u_\lambda) \geq \mathbf{b}(u_\rho)$ , we already know that either  $(u_\lambda, u_\rho)$  is not maximal, or  $u_\lambda \equiv u_\rho$ . Hence, a maximal gapped palindrome with an overlap is equal to a maximal ordinary palindrome. It is well known that maximal ordinary palindromes can be found in  $\mathcal{O}(n)$  time [19].

In what follows, we focus on the maximal gapped repeats with overlap. Given a maximal gapped repeat  $(u_\lambda, u_\rho)$  with period  $q := \mathbf{b}(u_\rho) - \mathbf{b}(u_\lambda) < |u_\lambda|$ , it induces a square with  $w[\mathbf{b}(u_\lambda).. \mathbf{b}(u_\lambda) + q - 1] = w[\mathbf{b}(u_\rho).. \mathbf{b}(u_\rho) + q - 1]$ . The square induces a run  $r$  whose minimal period  $p$  divides  $q$  (also observed in [5, Conclusions]). Both arms  $u_\lambda$  and  $u_\rho$  are contained in  $r$ . Because  $(u_\lambda, u_\rho)$  is maximal,  $\mathbf{b}(u_\lambda) = \mathbf{b}(r)$  and  $\mathbf{e}(u_\rho) = \mathbf{e}(r)$  hold; otherwise we could extend the arms to the left or to the right, respectively. This means that the left arm  $u_\lambda$  covers at least the segment  $w[\mathbf{b}(r).. \mathbf{b}(r) + \exp(r)p/2]$  (otherwise the arms would not overlap). Since  $q$  is a multiple of  $p$ , the number of different lengths of  $u_\lambda$  is bounded by  $\exp(r)/2$ . Figure 17 illustrates two maximal gapped repeats with overlap within the same run.

Our idea is to probe at the borders of each run  $r$  for all possible values of  $q$  to find all gapped repeats whose arms (a) overlap and (b) are contained in  $r$ . The positions of the runs and their periods can be computed in linear time [15, 1]. Since a run  $r$  contains  $\mathcal{O}(\exp(r))$  maximal  $\alpha$ -gapped repeats with overlap, we need  $\mathcal{O}(n)$  time to process all runs due to Lemma 2.3. We report each such gapped repeat exactly once because a gapped repeat  $(u_\lambda, u_\rho)$  with overlap is uniquely defined by its period and the borders of the run containing  $u_\lambda$  and  $u_\rho$ .

This algorithm is sufficient for saying that we can compute all maximal  $\alpha$ -gapped repeats (resp. palindromes) supporting overlaps, since this set consists of the set of all maximal  $\alpha$ -gapped repeats (resp. palindromes) without overlap and all maximal gapped repeats (resp. palindromes) with overlap:

**Theorem 5.1.** Given a word  $w$  of length  $n$  on an integer alphabet, we can compute all maximal  $\alpha$ -gapped repeats  $\mathcal{G}_\alpha(w)$  and all maximal  $\alpha$ -gapped palindromes  $\mathcal{G}_\alpha^\tau(w)$  in  $\mathcal{O}(\alpha n)$  time.

*Proof.* We compute the set of all maximal  $\alpha$ -gapped repeats (resp. palindromes) without overlap with the algorithm of Gawrychowski et al. [10] in  $\mathcal{O}(\alpha n)$  time. The above described algorithm computes the missing maximal gapped repeats with overlap in  $\mathcal{O}(n)$  time. The set of all maximal gapped palindromes is represented by the set of all maximal (ordinary) palindromes, which can be computed in  $\mathcal{O}(n)$  time as described above.  $\square$

## 6 Conclusion

We provided a thorough analysis on the maximum number of all maximal  $\alpha$ -gapped repeats and palindromes, for which we achieved the bounds of  $3(\pi^2/6 + 5/2)\alpha n$  and  $7(\pi^2/6 + 1/2)\alpha n - 3n - 1$ , respectively, for a word of length  $n$ . Our proofs work for both supporting overlaps and prohibiting overlaps, and thus generalize the analysis of former studies. Our study does not lead to a dead end, as can be seen by the following open problems:

**Generalizing Gaps.** A generalization of  $\alpha$ -gapped repeats are  $(f, g)$ -gapped repeats, i.e., gapped repeats  $(u_\lambda, u_\rho)$  with the additional property that  $g(|u_\lambda|) \leq \mathbf{b}(u_\rho) - \mathbf{e}(u_\lambda) - 1 \leq f(|u_\lambda|)$  for two functions  $f, g : \mathbb{N} \rightarrow \mathbb{R}$ . The  $(f, g)$ -gapped repeats with  $f(j) := 1, g(j) = \alpha j$  are exactly the  $\alpha$ -gapped repeats without overlap. Kolpakov [14] showed that the number of all maximal  $(f, g)$ -gapped repeats is bounded by

$$\mathcal{O}\left(n \left(1 + \max\left(\sup_{j \in \mathbb{N}} (1/j)(f(j) - g(j)), \sup_{j \in \mathbb{N}} |f(j+1) - f(j)|, \sup_{j \in \mathbb{N}} |g(j+1) - g(j)|\right)\right)\right).$$

Shaping the upper bound, or devising a lower bound for certain  $f$  and  $g$  is left for future work.

Regarding the algorithmic part, Brodal et al. [3] presented an algorithm computing all maximal  $(f, g)$ -gapped repeats in  $\mathcal{O}(n \lg n + \text{occ})$  time, where  $\text{occ}$  is the number of occurrences. In the light that we achieved  $\mathcal{O}(\alpha n)$  running time for finding all maximal  $\alpha$ -gapped repeats, it looks feasible to devise an algorithm whose running time depends linearly on  $n$  and on the values of  $f$  and  $g$ . Needless to say,  $(f, g)$ -gapped palindromes are also an unexplored topic.

**Online Algorithm.** To the best of our knowledge, there has not yet been an algorithm devised for computing all maximal  $\alpha$ -gapped repeats/palindromes of a given word *online*. We are aware of the algorithm of Fujishige et al. [9] finding all gapped palindromes with a fixed gap  $\mathbf{b}(u_\rho) - \mathbf{e}(u_\lambda) - 1 = c$  for a constant  $c$  in  $\mathcal{O}(n \lg \sigma)$  time online while taking  $\mathcal{O}(n)$  words of working space.

**Distinct Sets.** From literature it is already known that searching all distinct squares [2, 4] or all distinct ordinary palindromes [11] of a word of length  $n$  can be done in  $\mathcal{O}(n)$  time. A natural extension is computing all distinct  $\alpha$ -gapped repeats/palindromes, for which we are unaware of any results, both on the combinatorial (like giving an upper bound on the number of all distinct  $\alpha$ -gapped repeats/palindromes) and on the algorithmic aspects.

## Acknowledgements

We thank the anonymous reviewer who spotted a mistake in counting the number of possible values for the gap studied in the proof of Lemma 4.1( $\mathcal{P}^\tau$ ) (we previously bounded this number by  $|s_\lambda|(\alpha - 1)$ ). The mistake

is also present for counting the number of different gaps and distances in [10, Lemma 8]. However, we could come up with a fix in the case for the gapped repeats by restricting the distance to be at least  $p$  (or  $2p$  in case of the gapped repeats without overlaps considered in [10, Lemma 8a]).

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- [1] H. Bannai, T. I. S. Inenaga, Y. Nakashima, M. Takeda, and K. Tsuruta. The "runs" theorem. *SICOMP*, 46(5):1501–1514, 2017.
- [2] H. Bannai, S. Inenaga, and D. Köppl. Computing all distinct squares in linear time for integer alphabets. In *Proc. CPM*, volume 78 of *LIPICs*, pages 22:1–22:18, 2017.
- [3] G. S. Brodal, R. B. Lyngsø, C. N. S. Pedersen, and J. Stoye. Finding maximal pairs with bounded gap. In *Proc. CPM*, volume 1645 of *LNCS*, pages 134–149, 1999.
- [4] M. Crochemore, C. S. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, and T. Walen. Extracting powers and periods in a word from its runs structure. *TCS*, 521:29–41, 2014.
- [5] M. Crochemore, R. Kolpakov, and G. Kucherov. Optimal bounds for computing  $\alpha$ -gapped repeats. In *Proc. LATA*, volume 9618 of *LNCS*, pages 245–255, 2016.
- [6] A. L. Delcher, S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg. Alignment of whole genomes. *Nucleic Acids Research*, 27(11):2369–2376, 1999.
- [7] P. Duchon, C. Nicaud, and C. Pivoteau. Gapped pattern statistics. In *Proc. CPM*, volume 78 of *LIPICs*, pages 21:1–21:12, 2017.
- [8] N. J. Fine and H. S. Wilf. Uniqueness theorem for periodic functions. *Proc. AMS*, 16:109–114, 1965.
- [9] Y. Fujishige, M. Nakamura, S. Inenaga, H. Bannai, and M. Takeda. Finding gapped palindromes online. In *Proc. IWOCA*, volume 9843 of *LNCS*, pages 191–202, 2016.
- [10] P. Gawrychowski, T. I. S. Inenaga, D. Köppl, and F. Manea. Tighter bounds and optimal algorithms for all maximal  $\alpha$ -gapped repeats and palindromes. *TOCS*, 62(1):162–191, 2018.
- [11] R. Groult, É. Prieur, and G. Richomme. Counting distinct palindromes in a word in linear time. *IPL*, 110(20):908–912, 2010.
- [12] D. Gusfield. *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press, 1997.
- [13] J. Jurka. Repeats in genomic DNA: mining and meaning. *Current Opinion in Structural Biology*, 8(3): 333–337, 1998.
- [14] R. Kolpakov. On the number of gapped repeats with arbitrary gap. *Theor. Comput. Sci.*, 723:11–22, 2018.

- [15] R. Kolpakov and G. Kucherov. Finding maximal repetitions in a word in linear time. In *Proc. FOCS*, pages 596–604, 1999.
- [16] R. Kolpakov and G. Kucherov. Searching for gapped palindromes. *TCS*, 410(51):5365–5373, 2009.
- [17] R. Kolpakov, M. Podolskiy, M. Posypkin, and N. Khrapov. Searching of gapped repeats and subrepetitions in a word. *JDA*, 46-47:1–15, 2017.
- [18] S. Kurtz, J. V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, and R. Giegerich. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, 29(22):4633–4642, 2001.
- [19] G. Manacher. A new linear-time “on-line” algorithm for finding the smallest initial palindrome of a string. *JACM*, 22(3):346–351, 1975.
- [20] R. V. Samonte and E. E. Eichler. Segmental duplications and the evolution of the primate genome. *Nature Reviews Genetics*, 3(1):65–72, 2002.
- [21] P. Svoboda and A. D. Cara. Hairpin RNA: a secondary structure of primary importance. *Cellular and Molecular Life Sciences*, 63(7):901–908, 2006.
- [22] M. Tarailo-Graovac and N. Chen. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, 25:4.10.1–4.10.14, 2009.
- [23] B. Trombetta and F. Cruciani. Y chromosome palindromes and gene conversion. *Human Genetics*, 136(5):605–619, May 2017.
- [24] P. Warburton, J. Giordano, F. Cheung, Y. Gelfand, and G. Benson. Inverted repeat structure of the human genome: The X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Research*, 14:1861–1869, 2004.