

# The Number of Runs in a String

Wojciech Rytter \*

Key-Words: run, string, periodicity

## Abstract

A *run* in a string is a nonextendable (with the same minimal period) periodic segment in a string. The set of runs corresponds to the structure of internal periodicities in a string. Periodicities in strings were extensively studied and are important both in theory and practice (combinatorics of words, pattern-matching, computational biology). Let  $\rho(n)$  be the maximal number of runs in a string of length  $n$ . It has been shown in [8] that  $\rho(n) = O(n)$ , the proof was very complicated and the constant coefficient in  $O(n)$  has not been given explicitly. We demystify the proof of the linear upper bound for  $\rho(n)$  and propose a new approach to the analysis of runs based on the properties of subperiods: the periods of periodic parts of the runs. We show that  $\rho(n) \leq 3.44n$  and there are at most  $O(.67n)$  runs with periods larger than 87. This supports the conjecture that the number of all runs is smaller than  $n$ . We also give a completely new proof of the linear bound and discover several new interesting “periodicity lemmas”. Our proofs are inspired by the results of [4], where the role of new periodicity lemmas has been emphasized.

---

\*Institute of Informatics, Warsaw University, Banacha 2, PL-02-097, Warsaw, Poland, e-mail: ryttermimuw.edu.pl. Research supported by the grant 4T11C04425. The preliminary version of this paper appeared in [12] with the weaker result:  $\rho(n) \leq 5n$ .

## 1 Introduction

The set of all runs in a string corresponds to the structure of its regularities. Initial interest was mostly in repetitions of the type  $xx$  (so called *squares*), [1, 10]. The number of squares, with *primitive*  $x$ , is  $\Omega(n \log n)$ , hence the number of periodicities of this type is not linear. Then, it has been discovered that the number of runs (also called maximal repetitions or repeats) is linear and consequently linear time algorithms for runs were investigated [8, 7]. The result of [8] was one of the deepest results related to combinatorics and algorithmics of strings. However the most intriguing question remained the asymptotically tight bound for the number of runs. The first bound was quite complicated and has not given any *concrete* constant coefficient in  $O(n)$  notation. This subject has been studied in [13, 14, 2]. A beautiful construction showing the lower bound of approximately  $0.927 n$  has been given in [2]. The exact number of runs has been considered for special strings: *Fibonacci words* and (more generally) *Sturmian words*, [6, 5, 11]. In this paper we make a step towards better understanding of the structure of runs. The proof of the linear upper bound is simplified and small *explicit* constant coefficient is given in  $O(n)$  notation.

Let  $\text{per}(w)$  denote the size of the smallest period of  $w$ . We say that a word  $w$  is **periodic** iff  $\text{per}(w) \leq \frac{|w|}{2}$ .

A **run** in a string  $w$  is an interval  $\alpha = [i..j]$  such that  $w[i..j]$  is a periodic word with the period  $p = \text{per}(w[i..j])$  and this period is not extendable to the left or to the right of  $[i..j]$ .

In other words,  $[i..j]$  is a run iff  $|j - i + 1| \geq 2p$ ,  $i = 1$  or  $w[i - 1] \neq w[i - 1 + p]$  and  $j = n$  or  $w[j + 1] \neq w[j + 1 - p]$ . A run  $\alpha$  can be properly included as an interval in another run  $\beta$ , but in this case  $\text{per}(\alpha) < \text{per}(\beta)$ .

The value of the run  $\alpha = [i..j]$  is  $\text{val}(\alpha) = w[i..j]$ . When it creates no ambiguity we identify sometimes runs with their values although two different runs could correspond to the identical subwords, if we disregard positions of these runs. Hence runs are also called maximal *positioned* repetitions.

Denote by  $\text{RUNS}(w)$  the set of runs of  $w$ , see Figure 1 for an example.

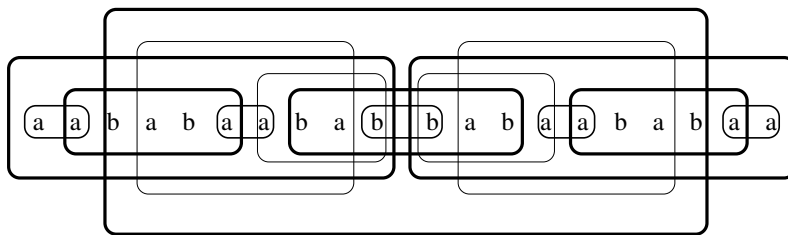


Figure 1: The structure of  $\text{RUNS}((aabab)^2(babaa)^2)$ .

Denote:  $\rho(n) = \max\{|\text{RUNS}(w)| : |w| = n\}$ .

The most interesting and open conjecture about the runs is:  $\rho(n) < n$ .

We make a small step towards proving validity of this conjecture and show that  $\rho(n) \leq 3.44 n$ . The proof of linear upper bound in [8] does not give any explicit constant coefficient at all.





**Example.**

(a) In Figure 5 the runs  $\alpha, \beta$  are in  $\mathcal{L}(11)$  (since  $\lceil \Delta^{11} \rceil = 12$  and the lengths of periodic parts of  $\alpha, \beta$  are 12, 13, respectively. At the same time they start in the same interval of length  $\lceil \frac{1}{4} \Delta^{11} \rceil = 3$ .

(b) In Figure 2 the runs  $\alpha, \beta$  are in  $\mathcal{R}(3)$  (since both have subperiod 3) and they start in the same interval of length 3.

Denote by  $\mathbf{HP}(n, p)$  the maximal number of hp-runs  $\alpha$  with  $\text{subperiod}(\alpha) \geq p$ , and by  $\mathbf{WP}(n, p)$  the maximal number of wp-runs  $\alpha$  with  $\text{period}(\alpha) \geq p$ , maximized over strings of length  $n$ .

The Key-Lemma implies in a simple way the following fact.

**Lemma 2.2**

1.  $\mathbf{HP}(n, p) \leq \frac{4}{p} \cdot n$ ;
2.  $\mathbf{WP}(n, \lceil \Delta^r \rceil) \leq 40 \cdot \Delta^{-r} \cdot n$

**Proof :**

**Point 1.** Denote by  $\mathbf{hp}(n, p)$  the maximal number of hp-runs  $\alpha$  with  $p \leq \text{subperiod}(\alpha) < 2p$ , maximized over strings of length  $n$ .

It follows directly from Lemma 2.1 that  $\mathbf{hp}(n, p) \leq \frac{2}{p} n$ . Hence the number of hp-runs with subperiod at least  $p$  is bounded from above by:

$$\begin{aligned} & \mathbf{hp}(n, p) + \mathbf{hp}(n, 2p) + \mathbf{hp}(n, 4p) + \mathbf{hp}(n, 8p) + \dots \\ & \leq 2n \cdot \left( \frac{1}{p} + \frac{1}{2p} + \frac{1}{4p} + \frac{1}{8p} + \dots \right) \leq \frac{4}{p} n \end{aligned}$$

**Point 2.** It follows directly from Lemma 2.1 that

$$|\mathcal{L}(k)| \leq 2 \cdot \left( 1 / \left( \Delta^k \cdot \frac{1}{4} \right) \right) \cdot n = 8\Delta^{-k} \cdot n$$

Consequently we have

$$\mathbf{WP}(n, \lceil \Delta^r \rceil) \leq \sum_{k=r}^{\infty} |\mathcal{L}(k)| \leq \sum_{k=r}^{\infty} 8 \cdot \Delta^{-k} \cdot n = 8\Delta^{-r} \times \frac{1}{1 - \Delta^{-1}} \cdot n = 40 \cdot \Delta^{-r} \cdot n$$

□

Observe that  $\mathbf{HP}(n, 1) = 0$  and  $\rho(n) = \mathbf{HP}(n, 2) + \mathbf{WP}(n, \lceil \Delta^0 \rceil)$ . Consequently we have directly the following corollary of Lemma 2.1.

**Corollary 2.3**  $\rho(n) \leq 42n$ .

In this way we have a very *simple* (if we disregard technicalities in the proof of Lemma 2.1) proof of a linear upper bound for  $\rho(n)$  with an explicit coefficient. In the next section we reduce the coefficient from 42 to 3.44. The reduction is done by estimating separately runs with small periodic part.

### 3 Estimating Number of Runs with Small Periodic Part

We say that a run is **large** iff  $\text{per}(\alpha) > 86$ . Otherwise the run is called a **small run**. We estimate now the number of small runs. Let  $\rho(n, k)$  be the maximal number of all runs  $\alpha$  with  $\text{per}(\alpha) \leq k$ , in a string of length  $n$ . We estimate the number of runs with small  $\text{PerPart}(\alpha)$  in a rather naive way using the following lemma.

**Lemma 3.1** For any given  $k \geq 1$  there are at most  $\frac{1}{k+1} n$  runs with  $\text{per}(\alpha) = k$  or  $\text{per}(\alpha) = 2k$ .

**Proof:** The proof of the following simple fact is illustrated in Figure 4.

**Claim 3.2** If  $u, v$  are primitive words and  $|u| = 2|v|$ , then  $vv$  is not contained in  $uu$  as a subword.

Assume that  $\alpha \prec \beta$  are two different runs with periods  $k$  or  $2k$ .

If  $\text{per}(\alpha) = \text{per}(\beta) = k$  then  $\alpha, \beta$  can have an overlap of size at most  $k - 1$ , otherwise  $\alpha, \beta$  could be merged into a single run. Hence  $\text{first}(\beta) - \text{first}(\alpha) \geq k + 1$ .

If  $\text{per}(\alpha) = k$  and  $\text{per}(\beta) = 2k$  then it is possible that  $\text{first}(\beta) - \text{first}(\alpha) = 1$ . Due to the claim the distance from  $\text{first}(\beta)$  to the occurrence of the next hp-run  $\gamma$  with period  $k$  or  $2k$  is at least  $2k+1$ . Then two consecutive distances give together  $(\text{first}(\beta) - \text{first}(\alpha)) + (\text{first}(\gamma) - \text{first}(\beta)) \geq 2k + 2$ , and ‘‘on average’’ the distance is  $k + 1$ . Therefore there are at most  $\frac{n}{k+1}$  runs with a period  $k$  or  $2k$ .  $\square$

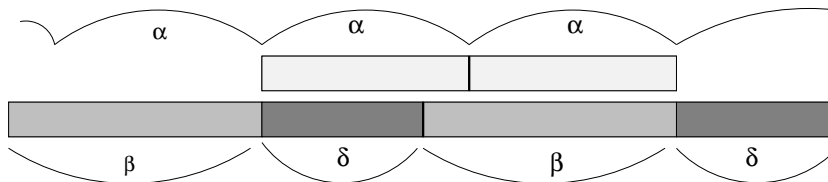


Figure 4: An occurrence of a square text  $v = \alpha\alpha$  inside a square text  $u = \beta\delta\beta\delta$ , where  $|u| = 2|v|$ , implies that  $|\alpha|$  is a period of  $u$ , consequently  $u$  is not periodic.

The last lemma motivates the introduction of the infinite set  $\Phi$ , generated by the following algorithm (which never stops).

```

 $\Phi := \emptyset; \Psi := \{1, 2, 3, \dots\};$ 
repeat forever
   $k := \min \Psi;$ 
  remove  $k$  and  $2k$  from  $\Psi;$ 
  insert  $k$  into  $\Phi;$ 

```

Define the set  $\Phi(p) = \{k \in \Phi : k \leq p\}$ . For example:

$$\Phi(34) = \{1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 16, 17, 19, 20, 21, 23, 25, 27, 28, 29, 31, 33\}$$

For  $p \geq 1$  define the numbers:

$$\mathcal{H}(p) = \sum_{k \in \Phi(p)} \frac{1}{k+1}.$$

The next lemma follows directly from Lemma 3.1 and from the structure of the set  $\Phi$ . We have, by straightforward brute-force calculations, that  $\mathcal{H}(86) \leq 2.77n$ .

**Lemma 3.3**

- (a)  $\rho(n, p) \leq \mathcal{H}(p) \times n$ .  
 (b) *There are at most  $2.77n$  small runs.*

**4 Estimating the Number  $\rho(n)$  of All Runs**

The analysis of the number of large runs is based on the *sparsity* properties of *hp-runs* and *wp-runs*. We estimate separately the number of runs in four disjoint classes:

- large hp-runs with subperiod larger than 21 (at most  $\frac{2}{11}n$ );
- large hp-runs with subperiod smaller than 22 (at most  $\frac{1}{45}n$ );
- large wp-runs; (at most  $0.4612 \cdot n$ )
- small runs (at most  $2.77n$ , due to Lemma 3.3).

Denote by  $\mathbf{hp}(n, p)$  the maximal number of hp-runs  $\alpha$  with  $p \leq \text{subperiod}(\alpha) < 2p$ , maximized over strings of length  $n$ .

**Lemma 4.1**

- (a) *There are at most  $\frac{2}{11}n$  hp-runs with subperiod larger than 21.*  
 (b) *There are at most  $\frac{1}{45}n$  large hp-runs with subperiod smaller than 22.*

**Proof :** Point (a) follows directly from Lemma 2.2. We show now Point (b). Two occurrences of two hp-runs with subperiods not exceeding 21 and with periods larger than 86 have to at least at distance  $87 - 2 \cdot 21 = 45$ , otherwise they have large overlap implying that these two runs merge into a single one (due to the periodicity lemma). Hence we have at most  $\frac{1}{45}n$  large hp-runs with subperiod smaller than 22.  $\square$

**Lemma 4.2** *There are at most  $0.4612 \cdot n$  large wp-runs.*

**Proof :** If we choose  $r = 20$  then  $\lfloor \Delta^r \rfloor = 86$  and  $\lceil \Delta^r \rceil = 87$ .

Now it follows from Lemma 2.2 that the number of large runs is bounded from above by

$$40 \cdot \left(\frac{5}{4}\right)^{-20} \cdot n \leq 0.4612 \cdot n$$

$\square$

**4.1 Main Result**

We can now combine all estimation together and proof the main result which gives a *concrete* constant coefficient in  $O(n)$  notation for  $\rho(n)$ .

**Theorem 1 [Main Result]**

- (1)  $\rho(n) \leq 3.44 n$ ;  
 (2) *There are at most  $0.67n$  large runs in a string of size  $n$ .*

**Proof :**

According to Lemma 4.2, Lemma 3.3 and Lemma 4.1 we have at most 2.77 small runs, and at most  $(\frac{2}{11} + \frac{1}{45} + 0.4612) \cdot n \leq 0.67n$  large runs.

Putting all together we get:

$$\rho(n) \leq (2.77 + 0.67n) \cdot n = 3.44n$$

□

## 5 The Proof of Point A of Lemma 2.1

We introduce a useful terminology of *neighboring runs*. We say that two different runs  $\alpha, \beta$  are **neighbors** iff there is a positive number  $\eta$  such that:

$$\text{dist}(\alpha, \beta) \leq \frac{1}{4}\eta \text{ and } \eta \leq \text{per}(\alpha), \text{per}(\beta) \leq \Delta \eta$$

Informally, two runs are neighbors iff they have similar periods and are positioned close to each other relatively to their sizes, in particular this means that

$$\text{per}(\alpha), \text{per}(\beta) \geq 4 |\text{dist}(\alpha, \beta)|.$$

**Example.** In Figure 5 we have two runs  $\alpha, \beta$  which are neighbors with  $\eta = 12$ ,  $\text{per}(\alpha) = 12$ ,  $\text{per}(\beta) = 13$  and  $\text{dist}(\alpha, \beta) = 2$ .

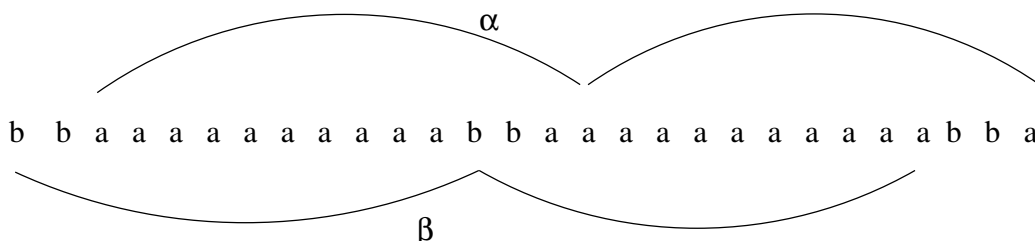


Figure 5: Two weakly-periodic runs  $\alpha, \beta$  which are neighbors with  $\eta = 12$ . We have:  $\alpha, \beta \in \mathcal{L}(11)$  and  $\text{dist}(\alpha, \beta) < \lceil \frac{1}{4}\Delta^{11} \rceil = 3$ .

### 5.1 The Three Neighbors Lemma

If  $\alpha \prec \beta$  and the *square part* of  $\beta$  is not contained in the *square part* of  $\alpha$  then we write  $\alpha \prec\prec \beta$  (see Figure 7). More formally:

$$\alpha \sqsupset \beta \text{ iff SquarePart}(\beta) \text{ is contained in SquarePart}(\alpha) \text{ as an interval}$$

$$\alpha \prec\prec \beta \text{ iff } [\alpha \prec \beta \text{ and not } (\alpha \sqsupset \beta)]$$

#### Lemma 5.1

(a) If  $\alpha \sqsupset \beta$  are distinct neighbors then  $\beta$  is an *hp-run*.

(b) If  $\alpha \prec\prec \beta$  are distinct neighbors then the prefix of  $\beta$  of size  $\text{per}(\alpha) - \delta$  has a period  $|q - p|$ , where  $\delta = \text{first}(\beta) - \text{first}(\alpha)$  and  $p = \text{per}(\alpha)$ ,  $q = \text{per}(\beta)$ .



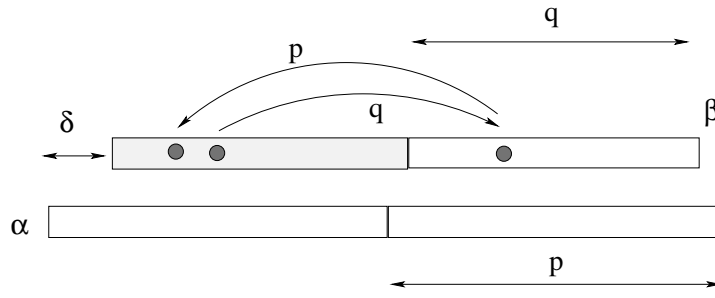


Figure 6: Two neighbors with  $\alpha \sqsubset \beta$ , a case  $\text{center}(\beta) > \text{center}(\alpha)$ . The *square part* of  $\beta$  is contained in the *square part* of  $\alpha$ . The periodic part of  $\beta$  is  $h$ -periodic, so it should have a period  $p - q$ , where  $p = \text{per}(\alpha)$ ,  $q = \text{per}(\beta)$ .

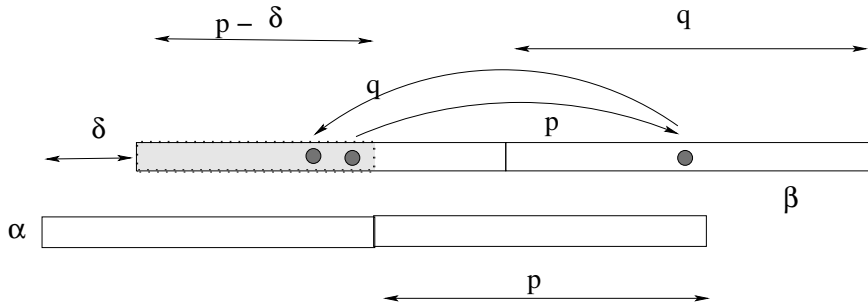


Figure 7: Two neighbors with  $\alpha \prec \prec \beta$ , the case  $p < q$ . The shaded part has the period  $|q - p|$ , where  $p = \text{per}(\alpha)$ ,  $q = \text{per}(\beta)$ .

**Proof :**

**Point (a)** We refer the reader to Figure 6, where the case  $\text{center}(\beta) > \text{center}(\alpha)$  is illustrated. Obviously  $p > q$ . It is easy to see that the whole  $\text{PerPart}(\beta)$  has a period  $\text{per}(\alpha) - \text{per}(\beta)$ .

Let  $\eta$  be the constant from the definition of neighbors, then

$$\text{per}(\alpha) - \text{per}(\beta) \leq \frac{1}{4}\eta \text{ and } |\text{PerPart}(\beta)| \geq \eta ,$$

hence  $\text{PerPart}(\beta)$  is  $h$ -periodic. The case  $\text{center}(\beta) \leq \text{center}(\alpha)$  can be considered similarly.

**Point (b)** We refer to Figure 7, when only the case  $p < q$  is shown. For each position  $i$  in the shaded area we have  $w[i] = w[i + p] = w[i + p - q]$ . The opposite case  $p > q$  can be considered similarly. This completes the proof.  $\square$

**Lemma 5.2 [The Three-Neighbors Lemma]**

*If we have three distinct runs which are pairwise neighbors with the same number  $\eta$  then at least one of them is  $h$ -periodic.*

**Proof :** Assume we have 3 runs  $\alpha_1 \prec \alpha_2 \prec \alpha_3$  which are pairwise neighbors, with periods  $p_1, p_2, p_3$ , respectively. Let  $\delta_1 = \text{first}(\alpha_2) - \text{first}(\alpha_1)$ , and  $\delta_2 = \text{first}(\alpha_3) - \text{first}(\alpha_2)$ . Then,

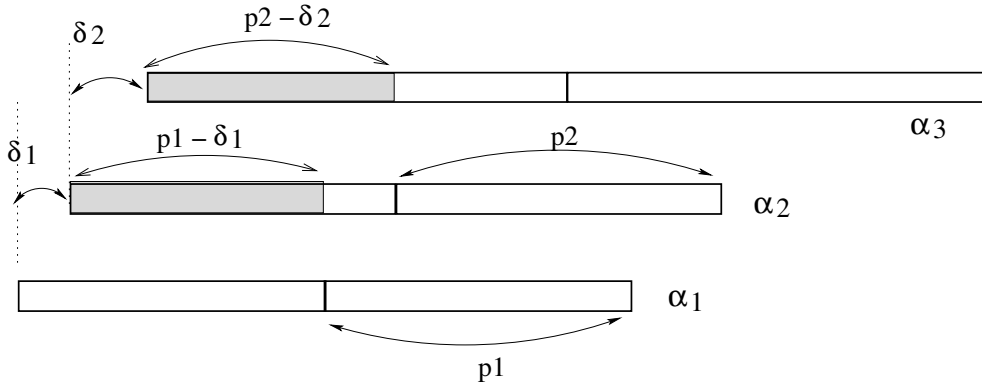


Figure 8: The Three-Neighbors Lemma, a situation when  $\alpha_1 \prec\prec \alpha_2 \prec\prec \alpha_3$ .  $\alpha_2$  should be h-periodic, since both its large suffix and large prefix have small periods.

due to Lemma 5.1 the “middle” run  $\alpha_2$  has a suffix  $\gamma_2$  of size  $p_2 - \delta_2$  with a period  $|p_3 - p_2|$  and a prefix  $\gamma_1$  of size  $p_1 - \delta_1$  with a period  $|p_2 - p_1|$ , see Figure 8.

Let  $\eta$  be the number from the definition of neighbors. We have

$$\delta_1 + \delta_2 \leq \frac{1}{4}\eta, \quad p_1 \geq \eta, \quad \text{and} \quad |\gamma_1 \cup \gamma_2| = p_2.$$

Hence:

$$|\gamma_1 \cap \gamma_2| \geq (p_2 - \delta_2) + (p_1 - \delta_1) - p_2 = p_1 - \delta_1 - \delta_2 \geq \frac{3}{4}\eta$$

We have  $|p_3 - p_2|, |p_2 - p_1| \leq \frac{1}{4}\eta$ , hence  $\text{per}(\gamma_1), \text{per}(\gamma_2) \leq \frac{1}{4}\eta$ . Due to the periodicity lemma  $\gamma_1 \cap \gamma_2$  has a period which divides periods of  $\gamma_1$  and  $\gamma_2$ , and the whole  $\alpha_2 = \gamma_1 \cup \gamma_2$  has a period of size not larger than  $\frac{1}{4}\eta$ . Consequently, the run  $\alpha_2$  is h-periodic. This completes the proof.  $\square$

## 5.2 The Proof of Point A of Lemma 2.1

If we take  $\eta = \lceil \Delta^k \rceil$  then, as a direct corollary of Lemma 5.2 we obtain Point A of Lemma 2.1.

## 6 The Proof of Point B of Lemma 2.1.

First we prove the following lemma.

### Lemma 6.1

Assume we have two distinct hp-runs  $\alpha, \beta$  with the same subperiod  $p$  and such that periodic part of one of them is a prefix of the periodic part of another. Then  $\text{dist}(\alpha, \beta) \geq p$ .

**Proof :** If  $\text{dist}(\alpha, \beta) < p$  then, due to *periodicity lemma* [9, 3, 13], the periodic part of one of the runs would have subperiod smaller than  $p$ , which contradicts the assumption that  $p$  is the smallest subperiod.  $\square$

We say that a hp-run  $\alpha = [i \dots j]$  of a string  $w$  is **left-periodic** iff  $w[i - 1] = w[i - 1 + \text{subperiod}(\alpha)]$ . The runs  $\beta, \gamma$  in Figure 2 are left-periodic. We also say that a position  $i$  in a word  $w$  *breaks* period  $p$  iff  $w[i] \neq w[i + p]$ . Hence a hp-run  $\alpha$  of a word  $w$  is *left-periodic* iff

$\text{first}(\alpha) - 1$  does not break  $\text{subperiod}(\alpha)$ . In other words the subperiod of  $\text{PerPart}(\alpha)$  continues to the left.

**Example.** In Figure 2 the runs  $\alpha, \beta, \gamma$  are shown, the first one is not left periodic and the other two are. The position  $\text{center}(\beta) - 1 = \text{center}(\gamma) - 1 = 21$  breaks subperiod 3. The periodic part of  $\beta$  is a prefix of a periodic part of  $\gamma$ .

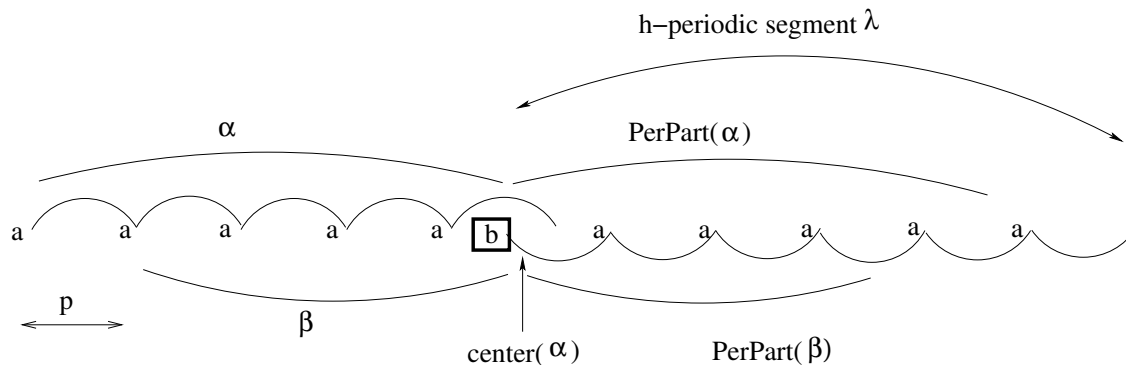


Figure 9: Two left-periodic runs. The position  $\text{center}(\alpha) - 1 = \text{center}(\beta) - 1$  breaking subperiod  $p$  is placed in a small square.  $\text{subperiod}(\alpha) = \text{subperiod}(\beta) = p$ ,  $\text{center}(\alpha) = \text{center}(\beta)$ . The second occurrences of periodic parts of  $\alpha$  and  $\beta$  start at the same position  $\text{center}(\alpha)$ , consequently  $\text{PerPart}(\beta)$  is a prefix of  $\text{PerPart}(\alpha)$ .

### Lemma 6.2

Assume two neighbors  $\alpha, \beta$  are left-periodic and  $h$ -periodic. Then  $\text{center}(\alpha) = \text{center}(\beta)$ .

#### Proof :

We first prove that positions  $\text{center}(\alpha) - 1, \text{center}(\beta) - 1$  break  $\text{subperiod}(\alpha)$ , see Figure 9. The proof is by contradiction. If it is not true then one of these runs can be extended one position to the left. This contradicts the definition of the run as a left non-extendible segment. The positions  $\text{center}(\alpha)$  and  $\text{center}(\beta)$  are positions in the same  $h$ -periodic segment  $\lambda$ , see Figure 9. They should be equal to the first position of this segment, because the next position to the left breaks the period. Hence they should be the same position, consequently  $\text{center}(\alpha) = \text{center}(\beta)$ .  $\square$

**Lemma 6.3** If  $\alpha, \beta$  are two  $hp$ -runs of a string  $w$  and satisfy for a given  $p > 1$  the inequality

$$\text{dist}(\alpha, \beta) < p \text{ and } \alpha, \beta \in \mathcal{R}(p)$$

then  $\text{subperiod}(\alpha) = \text{subperiod}(\beta)$ .

**Proof :** Assume that  $\text{first}(\alpha) \leq \text{first}(\beta)$ . If  $\text{dist}(\alpha, \beta) < p$  and  $p \leq \text{per}(\alpha), \text{per}(\beta) < 2p$  then periodic parts of  $hp$ -runs  $\alpha, \beta$  have an overlap of size at least  $\text{per}(\alpha) + \text{per}(\beta)$ . Then, due to the periodicity lemma the periodic parts  $\text{PerPart}(\alpha), \text{PerPart}(\beta)$  have the same minimal period. Consequently  $\text{subperiod}(\alpha) = \text{subperiod}(\beta)$ .  $\square$

## 6.1 The Proof of Point B of Lemma 2.1.

Due to Lemma 6.3 the HP-Runs Lemma is reduced to a slightly weaker statement:

For a given  $p > 1$  there are at most two occurrences of hp-runs with subperiod  $p$  in any interval of length  $p$ .

The proof of this fact is by contradiction. Assume we have three distinct hp-runs  $\alpha_1 \prec \alpha_2 \prec \alpha_3$  with the same subperiod  $p$  such that  $\text{dist}(\alpha_i, \alpha_j) \leq p$  for  $1 \leq i, j \leq 3$ . Then all of them are neighbors. We show that  $\alpha_2 = \alpha_3$ . Both  $\alpha_2, \alpha_3$  should be left-periodic since their subperiods extend to the left at least to  $\text{first}(\alpha_1)$ .

Therefore the runs  $\alpha_2, \alpha_3$  are h-periodic and they are neighbors. Due to Lemma 6.2  $\text{center}(\alpha_2) = \text{center}(\alpha_3)$ . Consequently periodic parts of  $\alpha_2$  and  $\alpha_3$  have occurrences starting at the same position  $\text{center}(\alpha_2)$ . If two words start at a same position then one should be a prefix of another. Consequently  $\text{PerPart}(\alpha_3)$  is a prefix of  $\text{PerPart}(\alpha_2)$ . Now, due to Lemma 6.1, if  $\alpha_2 \neq \alpha_3$  then  $\text{first}(\alpha_3) - \text{first}(\alpha_2) \geq p$ . However  $\text{first}(\alpha_3) - \text{first}(\alpha_2) < p$ . This implies that all of  $\alpha_1, \alpha_2, \alpha_3$  cannot be pairwise distinct. This contradicts the assumption and completes the proof of this fact and of Point B of Lemma 2.1.

## References

- [1] *M. Crochemore*, An optimal algorithm for computing the repetitions in a word, *Inf. Proc. Letters* 42:5(1981) 244-250
- [2] *F. Franek, R.J.Simpson, W.F.Smyth*, The maximum number of runs in a string, *Proc. 14-th Australian Workshop on Combinatorial Algorithms*, M.Miller, K. Park (editors) (2003) 26-35
- [3] *M. Crochemore, W.Rytter*, *Jewels of stringology: text algorithms*, World Scientific 2003
- [4] *Kangmin Fan, William F. Smyth, R. J. Simpson*: A New Periodicity Lemma. *CPM 2005*: 257-265
- [5] *F. Franek, A. Karaman, W.F.Smyth*, Repetitions in Sturmian strings, *TCS* 249-2 (2000) 289-303
- [6] *C. Iliopoulos, D. Moore, W.F.Smyth*, A characterization of the squares in a Fibonacci string, *TCS* 172 (1997) 281-291
- [7] *R.Kolpakov, G.Kucherov*, On maximal repetitions in words, *Journal of Discr. Algorithms* 1 (2000) 159-186
- [8] *R.Kolpakov, G.Kucherov*, Finding maximal repetitions in a word in linear time, *FOCS* (1999) 596-604
- [9] *Lothaire*, *Algebraic combinatorics on words*, Cambridge University Press
- [10] *M.G.Main, R.J.Lorentz*, An  $O(n \log n)$  algorithm for finding all repetitions in a string, *Journal of Algorithms* 5 (1984) 422-432
- [11] *W.Rytter*, The structure of subword graphs and suffix trees of Fibonacci words, in *Colloquium on Implementation and Application of Automata*, CIAA (2005)
- [12] *W.Rytter*, The number of runs in a string: improved analysis of the linear upper bound, *STACS 2006*, LNCS 3884
- [13] *W.F.Smyth*, *Computing patterns in strings*, Addison-Wesley (2003)
- [14] *W.F.Smyth*, Repetitive perhaps, but certainly not boring, *TCS* 249-2 (2000) 343-355.