Comment.Math.Univ.Carolin. 52,1 (2011) 1–20

A length bound for binary equality words

Jana Hadravová

Abstract. Let w be an equality word of two binary non-periodic morphisms $g, h : \{a, b\}^* \to \Delta^*$ with unique overflows. It is known that if w contains at least 25 occurrences of each of the letters a and b, then it has to have one of the following special forms: up to the exchange of the letters a and b either $w = (ab)^i a$, or $w = a^i b^j$ with $\gcd(i, j) = 1$.

We will generalize the result, justify this bound and prove that it can be lowered to nine occurrences of each of the letters a and b.

Keywords: combinatorics on words, binary equality languages

Classification: 68R15

1. Introduction

Equality language $\mathrm{Eq}(g, h)$ of morphisms $g, h : \Sigma^* \to \Delta^*$ consists of all their solutions, that is, of all words satisfying equality $g(w) = h(w)$. The concept of equality language was first introduced in [18] and since then has been widely studied. Equality languages achieved particular importance in the representation theory of formal languages since every recursively enumerable language can be effectively found as a morphic image of an equality language, see [1].

It is also well known, due to [16], that it is undecidable whether an equality language contains a nonempty word (an algorithmic problem known as the *Post Correspondence Problem*, or the PCP). Nevertheless, the problem turned out to be significantly different in the binary case. The decidability of the binary variant of PCP was announced by Ehrenfeucht, Karhumäki and Rozenberg in [3]. However, their proof contains a gap (see [7]); a full proof based on a similar approach is given by Halava, Harju and Hirvensalo in [6].

It should be also mentioned that the binary case of the PCP is decidable in polynomial time (see [8, 9]). For $|\Sigma| = 3$ it is already a long-standing open problem whether the equality set has to be regular, see [13] and [14].

The structure of binary equality languages has been first studied in [2] and [4] and later in series of papers [10], [11], [12]. It has been shown that binary equality languages are always generated by at most two words, provided that both morphisms are non-periodic (the periodic case being rather easy). It is also known that if the set $\mathrm{Eq}(g, h)$ is generated by two distinct generators, then these generators are of the form ba^i and $a^i b$.

A first step in the characterization of single generators of binary equality language was made in [5]. It was claimed there that a *simple solution*, that is, a solution with unique overflows, which is long enough in both letters a and b has to be of the form $w = (ab)^i a$ or $w = a^i b^j$ with $\gcd(i, j) = 1$ (up to the exchange of letters). The minimal requirement for the number of the letters a and b was fixed to nine, however rigorous proof was not given. The aim of this paper is to fill in this "white space" and provide the generalization of the result.

2. Basic concepts and definitions

The standard terminology and basic facts of combinatorics on words (see for example [15] and [17]) will be used across the text. Particularly, the reader should recall that a binary morphism $g : \{a, b\}^* \to \Delta^*$ is called *non-periodic* if $g(a)$ and $g(b)$ do not commute. If the image words $g(a)$ and $g(b)$ start with different letters, then we shall say that g is a *marked morphism*. We will use $u \leq_{\mathrm{p}} v$ when u is a prefix of v and $u <_{\mathrm{p}} v$ when u is a nonempty proper prefix of v. Similarly, $u \leq_{\mathrm{s}} v$ expresses the fact that u is a suffix of v and $u <_{\mathrm{s}} v$ means that u is a nonempty proper suffix of u. The greatest common prefix of two words u and v will be denoted by $u \wedge v$. (One-way) infinite word composed from infinite number of copies of a word u will be denoted u^ω. It should be also mentioned that the primitive root of a word u is the shortest word p such that $u = p^k$ for some positive k.

Binary morphims have the following very important property: For each non-periodic binary morphism g there is a uniquely given marked (non-periodic) binary morphism g_m and a word z_g such that for all words $w \in \{a, b\}^*$ we have $g(w) = z_g g_m(w) z_g^{-1}$. It is not so difficult to see that z_g is in fact equal to $g(ab) \wedge g(ba)$.

Let $g, h : \{a, b\}^* \to \Delta^*$ be two binary non-periodic morphisms. A word w is a *solution* of g, h if $g(w) = h(w)$. A solution w is called *simple* if all overflows are unique. That is, if w_1, $w_1 u$, w_2 and $w_2 u'$ are prefixes of w^ω such that

$$g(w_1)z = h(w_2) \quad \text{and} \quad g(w_1 u)z = h(w_2 u')$$

for some word z, then $|u| = |u'| = k|w|$ for some $k \in \mathbb{N}_+$.

A generalization of the concept of simple solution leads to the definition of *block* as a pair of two words (e, f) such that $g(e) = h(f)$ and which is simple in the aforementioned sense; that is, if w_1, $w_1 u$ and w_2, $w_2 u'$ are prefixes of e^ω, f^ω resp. such that

$$g(w_1)z = h(w_2) \quad \text{and} \quad g(w_1 u)z = h(w_2 u')$$

for some word z, then $|u| = k|e|$ and $|u'| = k|f|$ for some $k \in \mathbb{N}_+$.

In what follows we will be interested only in simple solutions and blocks.

Now, we are going to generalize the definitions given above. We will define a *cyclic solution* and a *cyclic block*. First though, let us fix the notation of (one-way) infinite words and intervals in words:

Let $u = u_0 \dots u_{n-1}$ be a finite word with its letters denoted by u_i, $0 \le i < n-1$. We define an infinite word starting at the i-th position of u by:

$$u[i,\infty] = u_i u_{i+1} \dots u_{n-1} u_0 u_1 \dots .$$

For two integers $0 \le i < j \le n-1$ we define the *interval* $u[i,j]$ by:

$$u[i,j] = u_i u_{i+1} \dots u_{j-1}.$$

In what follows we will use the definition of interval in a broader sense; if $i \ge j$, then we will use $u[i,j]$ instead of $u[i,\infty][0, j-i+n]$. Note that each letter u_i can be seen as $u[i,i+1]$; and a word $u[i,i]$ is a conjugate of u.

Notice that the definition of the interval $u[i,j]$ for $i \ge j$ is very natural when the word u is seen as a cyclic word. This motivates the following crucial definition:

Definition. Let $g, h : \{a,b\}^* \to \Delta^*$ be morphisms. A *cyclic solution* of g, h is an ordered quadruple $(w, \mathbf{c}, G, H)$ where $w = w_0 w_1 \cdots w_{|w|-1} \in \{a,b\}^+$, $\mathbf{c} \in \Delta^+$, $|\mathbf{c}| = |g(w)| = |h(w)|$ and $G, H : \mathbb{Z}_{|w|} \to \mathbb{Z}_{|\mathbf{c}|}$ are injective mappings such that

$$\mathbf{c}[G(i), G(i+1)] = g(w_i) \quad \text{and} \quad \mathbf{c}[H(i), H(i+1)] = h(w_i),$$

for all $i \in \mathbb{Z}_{|w|}$.

Note that in the previous definition $\mathbf{c}$ is a conjugate of $g(w)$ (and $h(w)$) and the injective mappings G, H define the ending and starting positions of image words inside the solution. Therefore, the overflows are words $\mathbf{c}[G(r), H(t)]$ and their position in the solution is uniquely given by the pair (r,t).

We will see later in Example 1 that the definition of cyclic solution indeed non-trivially generalizes the definition of (ordinary) solution. Moreover, the cyclicity of the solution simplifies the notion of overflows and allows us to avoid completely the use of the infinite words in the definition of *simple cyclic solution*:

Definition. Let $(w, \mathbf{c}, G, H)$ be a cyclic solution of g, h. We say that $(w, \mathbf{c}, G, H)$ is *simple* if

$$\mathbf{c}[G(r_1), H(t_1)] = \mathbf{c}[G(r_2), H(t_2)]$$

implies $(r_1, t_1) = (r_2, t_2)$.

A generalization of the simple cyclic solution leads to the definition of *cyclic block*:

Definition. Let $g, h : \{a,b\}^* \to \Delta^*$ be morphisms. A *cyclic block* of g, h is an ordered pentuple $(e, f, \mathbf{c}, G, H)$ where $e = e_0 e_1 \cdots e_{|e|-1} \in \{a,b\}^+$, $f = f_0 f_1 \cdots f_{|f|-1} \in \{a,b\}^+$, $\mathbf{c} \in \Delta^+$, $|\mathbf{c}| = |g(e)| = |h(f)|$ and $G : \mathbb{Z}_{|e|} \to \mathbb{Z}_{|\mathbf{c}|}$, $H : \mathbb{Z}_{|f|} \to \mathbb{Z}_{|\mathbf{c}|}$ are injective mappings such that

$$\mathbf{c}[G(i), G(i+1)] = g(e_i) \quad \text{and} \quad \mathbf{c}[H(j), H(j+1)] = h(f_j),$$

for all $i \in \mathbb{Z}_{|e|}$, $j \in \mathbb{Z}_{|f|}$. Moreover, we require that it is simple, that is, whenever

$$\mathbf{c}[G(r_1), H(t_1)] = \mathbf{c}[G(r_2), H(t_2)],$$

necessarily $(r_1, t_1) = (r_2, t_2)$.

Note that a simple cyclic solution $(w, \mathbf{c}, G, H)$ can be expressed as a cyclic block $(w, w, \mathbf{c}, G, H)$.

In order to further clarify the relation between the definitions of a solution and a cyclic solution, suppose that we are given a pair of (not necessarily marked) morphisms g and h, with a simple solution w. Now, w can be seen as a simple cyclic solution $(w, g(w), G, H)$ of g and h satisfying in addition that $G(0) = H(0)$. Consider marked versions $g_{\mathbf{m}}$ and $h_{\mathbf{m}}$ of g and h. Morphisms $g_{\mathbf{m}}, h_{\mathbf{m}}$ have a cyclic solution $(w, g(w), G_{\mathbf{m}}, H_{\mathbf{m}})$ given by

$$G_{\mathbf{m}}(j) = (G(j) + |z_g|) \bmod |g(w)|,$$
$$H_{\mathbf{m}}(j) = (H(j) + |z_h|) \bmod |g(w)|.$$

Notice that $(w, g(w), G_{\mathbf{m}}, H_{\mathbf{m}})$ is also a simple cyclic solution.

The following slightly technical definition of *p-synchronized overflows* will play an important role in our proof.

Definition. We say that a cyclic block $(e, f, \mathbf{c}, G, H)$ of morphisms g, h has k *p-synchronized overflows* if p is a primitive word and there is a k-tuple

$$((r_1, t_1), \ldots, (r_k, t_k)) \in (\mathbb{Z}_{|e|} \times \mathbb{Z}_{|f|})^k$$

of overflows which has the following properties:

1. for all $i \in \{1, \ldots, k-1\}$ there is $l_i \in \mathbb{N}_+$ such that

$$\mathbf{c}[G(r_i), H(t_i)] = p^{l_i} \mathbf{c}[G(r_{i+1}), H(t_{i+1})],$$

   and $\mathbf{c}[G(r_k), H(t_k)]$ is a nonempty prefix of p^ω;
2. r_i are pairwise distinct and t_i are pairwise distinct;
3. for each $i \in \{1, \ldots, k\}$ there is some $0 \leq m < |h(b)|$ such that

$$G(r_i) = H(t_i - 1) + m,$$

   and $f_{t_i - 1} = b$.

Informally, these are just different suffixes of $h(b)$ (by condition 3) which are overflows and at the same time are prefixes of p^ω for some primitive word p (by condition 1). Moreover, by condition 2 these overflows can neither start nor end at the same position of the word $\mathbf{c}$.

The following example illustrates previous definitions.

Example 1. Let g, h be morphisms given by:

$$g(a) = (aab)^2 a, \qquad g(b) = ab,$$
$$h(a) = a, \qquad h(b) = (baa)^3 ba.$$

They have a simple cyclic solution $((ab)^2 a, \mathbf{c}, G, H)$ where $\mathbf{c} = (aab)^8 a$, and the mappings $G, H : \mathbb{Z}_5 \to \mathbb{Z}_{25}$ are given by:

$$G(0) = 0, \quad G(1) = 7, \quad G(2) = 9, \quad G(3) = 16, \quad G(4) = 18,$$
$$H(0) = 1, \quad H(1) = 2, \quad H(2) = 13, \quad H(3) = 14, \quad H(4) = 0.$$

The cyclic solution is depicted in Figure 1.

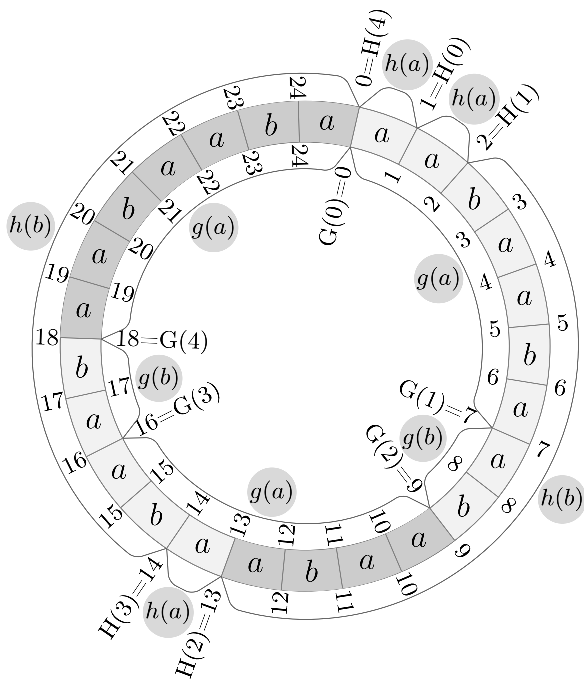

FIGURE 1: Simple cyclic solution $((ab)^2 a, (aab)^8 a, G, H)$.

It is possible to verify that g and h have no equality word. On the other hand, every equality word can be trivially viewed as a cyclic solution. This example therefore shows that the concept of cyclic solution generalizes nontrivially the concept of equality word.

Marked versions of the morphisms g and h are the following:

$$g(a) = (aba)^2 a, \qquad g(b) = ba,$$
$$h(a) = a, \qquad h(b) = (baa)^3 ba.$$

Their simple cyclic solution is $((ab)^2 a, \mathbf{c}, G_m, H_m)$ where $H_m = H$, and G_m is given by:

$$G_{\mathbf{m}}(j) = (G(j) + 1) \bmod 25.$$

Notice that in our graphical representation in Figure 1 it only means shifting the inner circle by one clockwise.

The example also features two aab-synchronized overflows, which are emphasized in Figure 1. They are given by pairs $(2, 2)$ and $(4, 4)$ since

$$\mathbf{c}[G(2), H(2)] = \mathbf{c}[9, 13] = (aab)a \quad \text{and} \quad \mathbf{c}[G(4), H(4)] = \mathbf{c}[18, 0] = (aab)(aab)a.$$

Notice that the cyclicity of the solution allows to speak easily for example about the overflow $(aab)^2 a(aab)^4 a$, which is given by $\mathbf{c}[G(4), H(2)]$. One of the main advantage of simple cyclic solutions in comparison with (ordinary) simple solutions is that the definition of simple cyclic solution does not need to employ infinite words.

The concept of p-synchronized overflows was introduced in [5] and it has been proved there that the existence of five p-synchronized overflows for some primitive word p inside the solution guarantees the special form of the solution: up to the exchange of the letters a and b either $w = (ab)^i a$ or $w = a^i b^j$ where $\gcd(i, j) = 1$. The result means a first significant step in the classification of single generated binary equality languages:

Lemma 1. *Let $g, h : \{a, b\}^* \to \Delta^*$ be non-periodic morphisms and let w be their simple solution. If $|w|_b \geq 9$ and $|w|_a \geq 9$, then, up to the exchange of the letters a and b, either*

$$w = (ab)^i a$$

or

$$w = a^i b^j$$

with $\gcd(i, j) = 1$.

Although the lemma itself does not speak about p-synchronized overflows, they are used in the proof as a key ingredient to connect assumption that $|w|_b \geq 9$ and $|w|_a \geq 9$ with the resulting structure of the solution. The key part of the proof uses the fact that the p-synchronized overflows are generated as a consequence of the assumption that $|w|_b \geq 9$ and $|w|_a \geq 9$. This is formulated in the following claim:

Lemma 2. *Let $(w, \mathbf{c}, G, H)$ be a simple cyclic solution of marked morphisms $g, h : \{a, b\}^* \to \Delta^*$. Let $h(b)$ be the longest of the image words $g(a), g(b), h(a)$ and $h(b)$. If $|w|_b \geq 9$, then there is a primitive word p such that*

- *$(w, \mathbf{c}, G, H)$ has five p-synchronized overflows;*
- *$h(b)$ is a factor of p^ω; and*
- *at least one of the words $g(a)$ or $g(b)$ is longer than p.*

Notice that in the foregoing lemma the condition $|w|_a \geq 9$ of Lemma 1 is missing. This is due to the fact that $h(b)$ is supposed to have the maximal length among the words $g(a)$, $g(b)$, $h(a)$ and $h(b)$. This distinguishes letters a and b and allows to drop the assumption on $|w|_a$.

Although the proof of Lemma 2 was hinted in [5] for much more generous bound of 25 bs inside the solution, the rigorous proof was omitted due to its complicity and length. The aim of this paper is to fill in this missing part in the proof of Lemma 1. Moreover, we will generalize Lemma 2 in order to be able to use it with cyclic blocks as well:

Main Lemma. *Let $(e, f, \mathbf{c}, G, H)$ be a cyclic block of marked morphisms $g, h : \{a, b\}^* \to \Delta^*$ and suppose that both e and f are factors of a word w such that $(w, \mathbf{c}', G', H')$ is a cyclic solution of g, h. Let $h(b)$ be the longest of the image words $g(a), g(b), h(a)$ and $h(b)$. If $|f|_b \geq 9$, then there is a primitive word p such that*

- *$(e, f, \mathbf{c}, G, H)$ has five p-synchronized overflows;*
- *$h(b)$ is a factor of p^ω; and*
- *at least one of the words $g(a)$ or $g(b)$ is longer than p.*

Notice that the foregoing lemma indeed generalizes Lemma 2 since every simple cyclic solution $(w, \mathbf{c}, G, H)$ of morphisms g, h is in fact a cyclic block $(w, w, \mathbf{c}, G, H)$ of the same pair of morphisms.

The proof of the Main Lemma will be given by combinatorial analysis in the last section.

We will finish this part by two definitions. First, let us define the *g-cover*:

Definition. Let $(e, f, \mathbf{c}, G, H)$ be a cyclic solution of morphisms $g, h : \{a, b\}^* \to \Delta^*$ and let $k, \ell \in \mathbb{Z}_{|\mathbf{c}|}$. An ordered pair $(m, u) \in \mathbb{Z}_{|e|} \times \{a, b\}^+$ is said to be the *g-cover of an ordered pair* (k, ℓ) if u and m are such that the word $g(u)$ defined as

$$g(u) = \mathbf{c}[G(m), k]\mathbf{c}[k, \ell]\mathbf{c}[\ell, G(m + |u|)]$$

is the shortest possible.

The last definition introduces a *true h-occurrence* of a word in a cyclic block:

Definition. Given a word v we say that $(k,l) \in \mathbb{Z}_{|\mathbf{c}|} \times \mathbb{Z}_{|\mathbf{c}|}$ is a *true h-occurrence of v* in $(e,f,\mathbf{c},G,H)$ if there are $i,j \in \mathbb{Z}_{|f|}$ satisfying $f[i,j] = v$ and $H(i) = k$, $H(j) = l$.

Example 2. In Figure 1 we have two true h-occurrences of b, namely $(2,13)$ and $(14,0)$. Moreover, g-cover of the first true h-occurrence of b is $(0,aba)$; the second one has the g-cover $(2,aba)$.

3. Auxiliary lemmas

In this section we will present combinatorial lemmas which will be needed in the proof of the Main Lemma. First, let us state without a proof the well known Periodicity Lemma:

Lemma 3 (Periodicity Lemma)**.** *Let p, q be primitive words. If p^ω and q^ω have a common factor of the length at least $|p|+|q|-1$, then p and q are conjugate.*

The Periodicity Lemma can be equally formulated in the following way: a word w with periods both n and m and longer than $n+m-1$ has also a period $\gcd(m,n)$.

The reader should be also familiar with the fact that two words u and v commute iff they have the same primitive root.

Just to recall the properties of primitive words, we have the next easy lemma:

Lemma 4. *Let p be a primitive word. If there are words u and v such that upv is a factor of p^ω, then u is a suffix of the word p^k and v is a prefix of p^k for sufficiently large $k \in \mathbb{N}_+$.*

The following combinatorial lemmas are mainly based on the Periodicity Lemma and explore the various periodicity properties of words.

Lemma 5. *Let u_1, u_2, v_1, v_2 be words such that*

$$\begin{aligned} u_1v_1 &= u_2v_2, \\ v_1u_1 &= v_2u_2, \end{aligned}$$

and $u_1 <_{\mathrm{p}} u_2$. Then the words $u_2u_1^{-1}$ and u_1v_2 have the same primitive root.

PROOF: Since

$$(u_1v_2)(u_2u_1^{-1}) = u_1v_1u_1u_1^{-1} = u_1v_1 = (u_2u_1^{-1})(u_1v_2),$$

words u_1v_2 and $u_2u_1^{-1}$ commute; therefore, they have the same primitive root. □

Lemma 6. *Let w be a word such that $w = s_1p_1 = s_2p_2 = s_3p_3$ for*

$$\begin{aligned} s_3 <_{\mathrm{s}} s_2 <_{\mathrm{s}} s_1 \\ p_1 <_{\mathrm{p}} p_2 <_{\mathrm{p}} p_3. \end{aligned}$$

Then w has a period

$$p = \gcd(|s_2| - |s_3|, |s_1| - |s_2|)$$

and $|w| \geq 2p$.

Proof: Obviously, the word w has periods both $|s_2| - |s_3|$ and $|s_1| - |s_2|$. Since

$$|w| = |s_3| - |s_3| + |s_2| - |s_2| + |s_1| + |p_1| \geq |s_2| - |s_3| + |s_1| - |s_2|,$$

we deduce from the Periodicity Lemma that w has a period $\gcd(|s_2| - |s_3|, |s_1| - |s_2|)$, which concludes the proof. □

Lemma 7. *Let w be a word such that*

$$w = vt_1 = r_2 v^j t_2 = r_3 v,$$

$j \geq 1$, and $r_2 <_{\mathrm{s}} r_3$, $t_2 <_{\mathrm{p}} t_1$. Then w has a period

$$\gcd(|t_1| - |t_2|, |r_3| - |r_2|).$$

Proof: It is easy to see that w has a periods both $|t_1| - |t_2|$ and $|r_3| - |r_2|$. Notice that

$$|t_1| - |t_2| + |r_3| - |r_2| = 2|w| - 2|v| - |t_2| - |r_2| = |w| + (j-2)|v|.$$

Therefore by the Periodicity Lemma, in the case $j \leq 2$, a period of w is $\gcd(|t_1| - |t_2|, |r_3| - |r_2|)$. It is easy to see that lemma is satisfied if $|r_2| = |t_2|$. Indeed, if $|r_2| = |t_2|$, then $|t_1| - |t_2| = |r_3| - |r_2|$ and the claim holds trivially. Let us discuss the remaining cases. The claim obviously holds for all w such that $|w| \leq 2|v|$ since in this case $j \leq 2$. Now, we proceed by induction on the length of the word w. By symmetry, we can suppose that $|r_2| > |t_2|$. Let t_1' be a prefix of t_1 such that $vt_1' = r_2 v$. We first show that t_2 is both a prefix and a suffix of t_1'. Since $|t_1'| = |r_2| > |t_2|$ and $t_2 <_{\mathrm{p}} t_1$, we can see easily that t_2 is a prefix of t_1'. On the other hand, since $r_2 v <_{\mathrm{s}} r_3 v$ both t_1' and t_2 are suffixes of w, and therefore suffix comparable.

We will split the proof into the following two cases.

Case $|vt_1'| \leq |v^j t_2|$. We will show that $w \in p^+$ and

$$\gcd(|t_1| - |t_2|, |r_3| - |r_2|) = k|p|,$$

$k \geq 1$, where p is the primitive root of v. Since t_2 is a suffix t_1' and $vt_1' \leq_{\mathrm{s}} v^j t_2$, we obtain $vt_1' t_2^{-1} \leq_{\mathrm{s}} v^j$. Let p be the primitive root of v. Then $pt_1' t_2^{-1}$ is a suffix of p^k for some sufficiently large $k \in N_+$. By the primitivity of p, we obtain that $t_1' t_2^{-1} \in p^*$. Moreover, since t_2 is a proper suffix of t_1', we get $t_1' t_2^{-1} \in p^+$. Since t_1' is bordered by t_2, we obtain that vt_1' is a prefix of p^ω longer than p. Recalling that p is the primitive root of v and v is a suffix of vt_1', we get that $t_1' \in p^+$ and

$t_2 \in p^+$. Since $r_2 v = v t_1'$ and $r_3 = r_2 v^j t_2 v^{-1}$, we have also $r_3 \in p^+$ and $r_2 \in p^+$. Finally, $w = r_2 v^j t_2$ leads to $w \in p^+$ and

$$\gcd(|t_1| - |t_2|, |r_3| - |r_2|) = k\,|p|,$$

$k \geq 1$, which completes the proof of this case.

Case $|vt_1'| > |v^j t_2|$. Let $w' = vt_1'$. Then

$$w' = vt_1' = r_2' v^j t_2 = r_3' v,$$

where $r_2' = vt_1'(v^j t_2)^{-1}$ and $r_3' = vt_1' v^{-1}$. Since $r_2' \leq_{\mathrm{s}} r_2$, $r_3' \leq_{\mathrm{s}} r_3$ and $|r_2'| \leq |r_3'|$, we obtain $r_2' \leq_{\mathrm{s}} r_3'$. By assumption, w' has a period

$$\pi = \gcd(|t_1'| - |t_2|, |r_3'| - |r_2'|).$$

We will prove that π divides both $|t_1| - |t_2|$ and $|r_3| - |r_2|$ and π is a period of w.

We have

$$|r_3'| - |r_2'| = |r_2| - |vt_1'| + |v^j t_2| = |r_3 v| - |vt_1'| = |r_3| - |r_2|.$$

Then π divides $|r_3| - |r_2|$ and since $|r_3| - |r_2|$ is a period of w, π is also a period of w. The last step is to prove that π divides $|t_1| - |t_2|$, which is true due to equation

$$|t_1| - |t_2| = (|t_1| - |t_1'|) + (|t_1'| - |t_2|) = |r_3| - |r_2| + |t_1'| - |t_2|.$$

The proof is now complete. □

Lemma 8. *Let w be a word such that*

$$w = r_1 v t_1 = r_2 v^j t_2 = r_3 v t_3,$$

$j \geq 1$, and $r_2 <_{\mathrm{s}} r_3$, $t_2 <_{\mathrm{p}} t_1$. Then w has a period

$$\gcd(|t_1| - |t_2|, |r_3| - |r_2|)$$

and $t_3 \leq_{\mathrm{p}} v^{j-1} t_2$.

Proof: From Lemma 7 it follows that the word $w[|r_1|, |r_3 v|]$ has a period

$$q = \gcd(|t_1| - |t_2|, |r_3| - |r_2|).$$

The word $w[0, |r_3 v|]$ has a period $|r_3| - |r_2|$; therefore, has a period q. Similarly, the word $w[|r_1|, |w|]$ has periods both $|t_1| - |t_2|$ and q. Since common factor of words $w[0, |r_3 v|]$ and $w[|r_1|, |w|]$ is the word $w[|r_1|, |r_3 v|]$, which is longer than q, w has a period q. Word t_3 is a prefix of $v^{j-1} t_2$ because $|r_3| - |r_2|$ is a period of w. □

We will finish this section by lemmas which are not strictly combinatorial but will be needed in the proof.

Lemma 9. *Let g be a marked morphism and u, v, w be words satisfying*

$$g(u) \wedge w <_{\mathrm{p}} g(v) \wedge w.$$

Then $g(u) \wedge w = g(u \wedge v)$.

Proof: It is easy to check that if $u \wedge w <_{\mathrm{p}} v \wedge w$ for arbitrary three words, then $u \wedge v = u \wedge w$. This and the fact that marked morphisms satisfy

$$g(u) \wedge g(v) = g(u \wedge v)$$

completes the proof. □

Lemma 10. *Let $(e, f, \mathbf{c}, G, H)$ be a cyclic block for marked morphisms $g, h : \{a, b\}^* \to \Delta^*$ and let $(k_1, l_1), (k_2, l_2)$ be two p-synchronized overflows where p is the primitive root of $g(a)$. Then either $e \in a^+$ or there exists a pair of indices (r_1, q_1) such that $G(r_1) = H(q_1)$ and $e[r_1 - 1, r_1] = a$.*

Proof: Suppose that $e \notin a^+$. Then

$$g(e[k_1, \infty]) \wedge p^\omega = \mathbf{c}[G(k_1), \infty] \wedge p^\omega <_{\mathrm{p}} g(a)^\omega \wedge p^\omega$$
$$g(e[k_2, \infty]) \wedge p^\omega = \mathbf{c}[G(k_2), \infty] \wedge p^\omega <_{\mathrm{p}} g(a)^\omega \wedge p^\omega.$$

By Lemma 9 we obtain the existence of r_1, r_2 such that $e[k_1, r_1] \in a^+$ and $e[k_2, r_2] \in a^+$. Notice that from the definition of p-synchronized overflows we know that $\mathbf{c}[G(k_i), H(l_i)]$ is a prefix of p^ω for both $i \in \{1, 2\}$. Moreover, by the same definition

$$\mathbf{c}[G(k_1), H(l_1)] = p^l \mathbf{c}[G(k_2), H(l_2)].$$

Therefore,

$$\mathbf{c}[G(k_2), H(l_2)]^{-1} p^\omega = \mathbf{c}[G(k_1), H(l_1)]^{-1} p^\omega.$$

Now, we know that:

$$\mathbf{c}[H(l_1), G(r_1)] = \mathbf{c}[H(l_1), \infty] \wedge \mathbf{c}[G(k_1), H(l_1)]^{-1} p^\omega,$$
$$\mathbf{c}[H(l_2), G(r_2)] = \mathbf{c}[H(l_2), \infty] \wedge \mathbf{c}[G(k_2), H(l_2)]^{-1} p^\omega.$$

Since every cyclic block is simple, we have up to the order of indices:

$$\mathbf{c}[H(l_1), \infty] \wedge \mathbf{c}[G(k_1), H(l_1)]^{-1} p^\omega <_{\mathrm{p}} \mathbf{c}[H(l_2), \infty] \wedge \mathbf{c}[G(k_2), H(l_2)]^{-1} p^\omega.$$

Since

$$\mathbf{c}[H(l_1), \infty] = h(f[l_1, \infty]) \text{ and } \mathbf{c}[H(l_2), \infty] = h(f[l_2, \infty]),$$

we finally obtain an inequality

$$(*) \qquad h(f[l_1, \infty]) \wedge \mathbf{c}[G(k_1), H(l_1)]^{-1} p^\omega <_{\mathrm{p}} h(f[l_2, \infty]) \wedge \mathbf{c}[G(k_1), H(l_1)]^{-1} p^\omega.$$

Now, we can again apply Lemma 9 on (*) and obtain an index q_1 such that

$$\begin{aligned}\mathbf{c}[H(l_1), H(q_1)] &= h(f[l_1, q_1]) \\ &= h(f[l_1, \infty]) \wedge \mathbf{c}[G(k_1), H(l_1)]^{-1} p^\omega = \mathbf{c}[H(l_1), G(r_1)].\end{aligned}$$

Then $H(q_1) = G(r_1)$, which is what we wanted to prove. From $e[k_1, r_1] \in a^+$ follows the rest. $\square$

4. Proof of Main Lemma

We shall assume that $h(b)$ is the longest of all four image words, that is,

$$|g(a)| \leq |h(b)|, \quad |g(b)| \leq |h(b)| \quad \text{and} \quad |h(a)| \leq |h(b)|.$$

The aim of this section is to show that five p-synchronized overflows are created by cumulating bs in f. We will provide an upper bound c_b for number of bs in f such that unless $|f|_b < c_b$, a cyclic block $(e, f, \mathbf{c}, G, H)$ necessarily has to have five p-synchronized overflows. This part will show that this bound can be lowered to 9 occurrences of b.

We will have a look at possible forms of g-covers of true h-occurrences of b in f. It has been shown in [5] that possible forms of g-covers of true h-occurrences of b inside a solution are quite restricted:

Lemma 11. *Let $(w, \mathbf{c}, G, H)$ be a cyclic solution for morphisms g, h such that morphism g is marked. Let (k, ℓ) be a true h-occurrence of b and let (m, u) be its g-cover. Then u belongs to the one of the following sets:*

$$a^+ \qquad b^+ \qquad a^+b^+ \qquad b^+a^+ \qquad a^+b^+a^+ \qquad b^+a^+b^+.$$

Since e and f are factors of some cyclic solution $(w, \mathbf{c}', G', H')$, we can use this fact and restrict ourselves only to these six types of g-covers.

We now proceed to prove that either a cyclic block $(e, f, \mathbf{c}, G, H)$ has five p-synchronized overflows for some primitive word p or $|f|_b < 9$.

Before the proof is given we will fix the notation and prove few auxiliary claims.

First, notice that since $h(b)$ is of the maximal length, every g-cover (m, u) of a true h-occurrence of b can be rewritten as $u = u_1 u_2$, for some nonempty words u_1, u_2. Because of this property we can apply combinatorial lemmas mentioned in the previous section.

As we have already seen in Lemma 11, every true h-occurrence of b has to possess exactly one of the six variants of cover. Thus, we will divide true h-occurrences of b into six classes depending on the variant of their cover. To simplify things slightly, we shall use the notation according to the following table, which features variables representing the number of true h-occurrences of b with the specific cover variant:

# of true h-occurrences of b	cover variant
x_1	a^+
x_2	$a^+b^+a^+$
x_3	b^+a^+
y_1	b^+
y_2	$b^+a^+b^+$
y_3	a^+b^+

The following lemma presents the way how the g-cover of a true h-occurrence can be combined with n p-synchronized overflows in order to create $n+1$ p-synchronized overflows. In this way we can increase the number of p-synchronized overflows.

Lemma 12. *Let $(e, f, \mathbf{c}, G, H)$ be a cyclic block and let (m, uv) be the g-cover of (k, l), where (k, l) is a true h-occurrence of b. Suppose that u, v are nonempty words and there is a primitive word p such that*

(1) *$h(b)$ is a factor of p^ω,*
(2) *there are p-synchronized overflows $(r_1, t_1), \ldots, (r_n, t_n)$, $n \geq 2$, such that $r_i \neq m + |u|$ and $t_i \neq H^{-1}(l)$ for all $i \in \{1, \ldots, n\}$,*
(3) *$p \leq_{\mathrm{s}} \mathbf{c}[k, G(m + |u|)]$ or $p \leq_{\mathrm{p}} \mathbf{c}[G(m + |u|), l]$.*

Then $(r_1, t_1), \ldots, (r_n, t_n), (m + |u|, H^{-1}(l))$ are up to the order p-synchronized overflows.

Proof: First recall that according to the definition of true h-occurrence of b we have that $\mathbf{c}[k, l] = h(b)$ and moreover, $k = H^{-1}(l) - 1$. Since (m, uv) is the g-cover of (k, l), we obtain from its definition that

$$G(m + |u|) = k + j,$$

where $0 < j < |h(b)|$. Therefore, $\mathbf{c}[G(m + |u|, l]$ is a suffix of $h(b)$ and the third condition in the definition of p-synchronized overflows is satisfied.

Notice also that by the second assumption of this lemma, starting and ending positions of overflows $(r_1, t_1), \ldots, (r_n, t_n)$ are different than those of $(m + |u|, H^{-1}(l))$.

It remains to show that the overflows $(r_1, t_1), \ldots, (r_n, t_n), (m + |u|, H^{-1}(l))$ are indeed p-synchronized and satisfy (up to their order) the first condition of the definition of p-synchronized overflows. But, since $h(b)$ is a factor of p^ω, this fact easily follows from the third assumption of this lemma together with Lemma 4. $\square$

Notice that the third condition of the previous lemma has a particular importance, since it "fixes" the overflow $\mathbf{c}[G(m + |u|), l]$ in accordance with distribution of ps over $h(b)$.

The next claim presents key ideas involved in combining g-covers of the same kind into p-synchronized overflows.

Claim 1. Let $(e, f, \mathbf{c}, G, H)$ be a cyclic block of morphisms g, h. Then the following combinatorial properties hold.

(I.) If $x_1 \geq 2$, then $(e, f, \mathbf{c}, G, H)$ has x_1 p-synchronized overflows, $g(a) = p^i$, $i \geq 2$ and $h(b)$ is a factor of p^ω longer than $2\,|p|$.

(II.) If $x_2 \geq 3$, then $(e, f, \mathbf{c}, G, H)$ has x_2 p-synchronized overflows, $p \leq_{\mathrm{p}} g(a)$ and $h(b)$ is a factor of p^ω longer than $2\,|p|$.

(III.) If $x_3 \geq 3$, then $(e, f, \mathbf{c}, G, H)$ has x_3 p-synchronized overflows, $p \leq_{\mathrm{p}} g(a)$, $p \leq_{\mathrm{s}} g(b)$ and $h(b)$ is a factor of p^ω longer than $2\,|p|$.

PROOF: (I.) We will use Lemma 5. Let (k_1, l_1), (k_2, l_2) be true h-occurrences of b and let (m_1, aa^{j_1}), (m_2, aa^{j_2}) be their respective g-covers. Let

$$u_1 = \mathbf{c}[G(m_1), k_1] \qquad v_1 = \mathbf{c}[k_1, G(m_1 + 1)]$$
$$u_2 = \mathbf{c}[G(m_2), k_2] \qquad v_2 = \mathbf{c}[k_2, G(m_2 + 1)].$$

Since we are in a cyclic block, which is simple, we can suppose that $u_2 <_{\mathrm{p}} u_1$. Notice that from the definition of g-cover we have

$$u_1 v_1 = u_2 v_2 = g(a),$$
$$v_1 u_1 = v_2 u_2 \leq_{\mathrm{p}} h(b).$$

It follows from Lemma 5 that the words $u_1 u_2^{-1}$ and $u_2 v_1$ have the same primitive root p. Then

$$g(a) = u_1 v_1 = u_1 u_2^{-1} u_2 v_1 = p^i,$$

for some $i \geq 2$. Since $h(b)$ is a factor of $g(a)^\omega$, it is obviously factor of p^ω as well. We shall prove that $(m_1 + 1, H^{-1}(l_1))$ and $(m_2 + 1, H^{-1}(l_2))$ are p-synchronized overflows. Obviously, $m_1 \neq m_2$ and $l_1 \neq l_2$. Since

$$\mathbf{c}[m_1 + 1, H^{-1}(l_1)] = v_1^{-1} h(b),$$
$$\mathbf{c}[m_2 + 1, H^{-1}(l_2)] = v_2^{-1} h(b),$$

we obtain that

$$\mathbf{c}[m_1 + 1, H^{-1}(l_1)]\mathbf{c}[m_2 + 1, H^{-1}(l_2)]^{-1} = v_1^{-1} v_2 = u_1 u_2^{-1} \in p^+,$$

and $(m_1 + 1, H^{-1}(l_1))$ and $(m_2 + 1, H^{-1}(l_2))$ are p-synchronized overflows. It is easy to see that each g-cover of the type a^+ of another true h-occurrence of b satisfies condition of Lemma 12 and therefore we can add these g-covers one by one to the already found p-synchronized overflows.

(II.) and (III.) Remaining two claims can be proved in a similar way using Lemma 8 and Lemma 6 respectively. □

Notice that because of symmetry of xs and ys (the covers are the same up to the exchange of letters a and b) the foregoing lemma can be reformulated with ys instead of xs. In what follows we will call this "reformulated" version of Claim 1 its dual form.

In particular, Claim 1 means that if $x_i \geq 5$ or $y_i \geq 5$ for any $i \in \{1,2,3\}$, then the conclusion of the Main Lemma holds. Notice that under the previous observation if now $|f|_b \geq 25$, then the conclusion of the Main Lemma holds simply by the pigeonhole principle. The rest of the section is dedicated to lowering this bound.

Claim 2. Let $(e, f, \mathbf{c}, G, H)$ be a cyclic block of marked morphisms g, h. The following properties hold.

(I.) If $x_1 \neq 0$, then $y_2 = 0$.
(II.) If $x_1 \geq 2$, then $y_3 \leq 1$.
(III.) If $x_1 \geq 2$, then $y_1 \leq 1$.

Proof: (I.) Suppose that $y_2 \geq 1$. Then

$$h(b) = s_2 g(a)^i p_2,$$

where $i \geq 1$ and $p_2 \leq_{\mathrm{p}} g(b)^\omega$. Since $x_1 \neq 0$, the word $h(b)$ is a factor of p^ω, where p is the primitive root of $g(a)$. By Lemma 4, we obtain that $p_2 \leq_{\mathrm{p}} p^\omega$. We have a contradiction with g being marked.

(II.) Since $x_1 \geq 2$, by Claim 1(I.) there is a primitive word p such that $g(a) = p^i$, $i \geq 2$ and $h(b)$ is longer than $2|p|$ with a period $|p|$. If $y_3 \geq 2$, then

$$h(b) = s_1 p_1 = s_2 p_2,$$

where $s_1 <_{\mathrm{s}} s_2 \in \mathrm{Suf}(g(a)^+) = \mathrm{Suf}(p^+)$ and $p_2 <_{\mathrm{p}} p_1 \in \mathrm{Pref}(g(b)^+)$. First notice that $|s_2| - |s_1|$ is a period of $h(b)$. Since p is a primitive word and $|h(b)| > 2|p|$, by the Periodicity Lemma this period has to be longer or equal to $|p|$. Then $p \leq_{\mathrm{s}} s_2$ and $h(b) = upp_2$, where $up = s_2$. By Lemma 4 we have that $p_2 \leq_{\mathrm{p}} p^\omega$, which is a contradiction with g being marked.

(III.) Suppose that $x_1 \geq 2$ and $y_1 \geq 2$. Then by Claim 1(I.) (and its dual form for ys) there are two p-synchronized overflows $(k_1, l_1), (k_2, l_2)$ and two s-synchronized overflows $(m_1, n_1), (m_2, n_2)$, where p is the primitive root of $g(a)$ and s is the primitive root of $g(b)$. From Lemma 10 it follows that there are $r_1, r_2, q_1, q_2 \in \mathbb{Z}_{|w|}$ such that $G(r_1) = H(q_1)$, $G(r_2) = H(q_2)$ and $b = e[r_1 - 1, r_1]$, $a = e[r_2 - 1, r_2]$. Since $(e, f, \mathbf{c}, G, H)$ is a cyclic block, which is simple, we have $r_1 = r_2$ and $a = e[r_2 - 1, r_2] = e[r_1 - 1, r_1] = b$ is a contradiction. $\square$

In a view of previous claim we can now without difficulty see that if $|f|_b \geq 17$, then we get the desired conclusion of the Main Lemma.

Last part of this section investigates possible combination of g-covers of different kind. Again, as the previous claim, this claim has its dual version obtained by exchanging xs by ys (and vice versa).

Claim 3. Let $(e, f, \mathbf{c}, G, H)$ be a cyclic block of marked morphisms g, h. If one of the following properties is satisfied, then the conclusion of the Main Lemma holds.

(I.) $x_1 + x_3 \geq 6$.
(II.) $x_1 \geq 2$ and $(x_2 \geq 3$ or $x_3 \geq 3)$.
(III.) $x_2 + x_3 \geq 5$.
(IV.) $x_3 \geq 3$ and $y_3 \geq 3$ and $x_1 + x_3 \geq 5$.
(V.) $x_3 \geq 3$ and $x_3 + y_2 + x_2 \geq 5$.

Proof: (I.) Suppose that $x_1 + x_3 \geq 6$. If $x_3 \geq 5$ or $x_1 \geq 5$, then we shall use Claim 1(III.) or Claim 1(I.). Therefore, $4 \geq x_1 \geq 2$ and $4 \geq x_3 \geq 2$. It follows from Claim 1(I.) that there are x_1 p-synchronized overflows such that p is the primitive root of $g(a)$, both $g(a)$ and $h(b)$ are longer than $2\,|p|$ and $h(b)$ is factor p^ω. Let $(m_1, b^{i_1}a^{j_1}), \ldots, (m_{x_3}, b^{i_{x_3}}a^{j_{x_3}})$ be the g-covers of $(r_1, t_1), \ldots, (r_{x_3}, t_{x_3})$, pairwise different true h-occurrences of b in $(e, f, \mathbf{c}, G, H)$. We will show that the inequality

$$|\mathbf{c}[G(m_n + i_n), t_n]| \geq |p|$$

holds for at least $x_3 - 1$ different g-covers of true h-occurrences of b. Then from the primitivity of p we obtain $p \leq_{\mathrm{p}} \mathbf{c}[G(m_n + i_n), t_n]$ for $x_3 - 1$ g-covers and we can gradually apply Lemma 12 and increase the number p-synchronized overflows up to the number $x_1 + x_3 - 1 \geq 5$, which completes the proof.

For contradiction, suppose that for two different indices n_1, n_2, it holds:

$$|\mathbf{c}[G(m_{n_1} + i_{n_1}), t_{n_1}]| < |p|$$
$$|\mathbf{c}[G(m_{n_2} + i_{n_2}), t_{n_2}]| < |p|\,.$$

Then $h(b)$ has a period q such that $q < |p|$. Since $|h(b)| > |p| + q$ we obtain from the Periodicity Lemma a contradiction with the primitivity of p.

(II.) Suppose that $x_1 \geq 2$ and $x_2 \geq 3$. Then by Claim 1(I.) and Claim 1(II.), there are x_1 p-synchronized overflows and x_2 s-synchronized overflows such that $g(a) = p^i$, $i \geq 2$, and s is a prefix of $g(a)$. Moreover, $h(b)$ is a factor of both p^ω and s^ω which is longer than $\max(2\,|p|\,, 2|s|)$. From the Periodicity Lemma and the primitivity of both words p and s follows that p and s are conjugates. Therefore, $p = s$. Since p is primitive and $h(b)$ is a factor of p^ω, we get from Lemma 12 that $(e, f, \mathbf{c}, G, H)$ posses $x_1 + x_2$ p-synchronized overflows and the conclusion of the Main Lemma holds. Case $x_1 \geq 2$ and $x_3 \geq 3$ is similar.

(III.) Obviously, either $x_2 \geq 3$ or $x_3 \geq 3$. Suppose that both $x_2 \neq 0$ and $x_3 \neq 0$ otherwise conclusion holds according to Claim 1. Let $x_2 \geq 3$. From Claim 1(II.) we obtain the existence of x_2 p-synchronized overflows such that p is a prefix of $g(a)$ and $h(b)$ is a factor of p^ω longer than $2\,|p|$. Let $(n, a^ib^ja^k)$ be the g-cover of (r_1, t_1), a true h-occurrence of b in $(e, f, \mathbf{c}, G, H)$, such that $|\mathbf{c}[G(n{+}i{+}j), t_1]| > |p|$. Let (m, b^sa^l) be the g-cover of (r_2, t_2), a true h-occurrence of b in $(e, f, \mathbf{c}, G, H)$. Notice, that such g-covers indeed exist since $x_2 \geq 3$ and $x_3 \neq 0$.

If $|\mathbf{c}[G(m+s),t]| < |p|$, then by looking at a prefix of $h(b)$, we obtain an equality:

$$\mathbf{c}[r_2, G(m+s)] = \mathbf{c}[r_1, G(n+i+j)]u$$

for some nonempty word $u \leq_{\mathrm{p}} g(a)^\omega$. Since $g(b) \leq_{\mathrm{s}} \mathbf{c}[r_1, G(n+i+j)]$ and $\mathbf{c}[r_2, G(m+s)] \leq_{\mathrm{s}} g(b)^s$, we can apply Lemma 4 and obtain that u is prefix comparable with the primitive root of $g(b)$. This gives us a contradiction with g being marked. Therefore, $p \leq_{\mathrm{p}} \mathbf{c}[G(m+s),t]$, and we can again gradually apply Lemma 12 ending with $x_2 + x_3 \geq 5$ p-synchronized overflows.

Now, let $x_3 \geq 3$. Using again Claim 1(III.) leads to the existence of x_3 p-synchronized overflows such that $p \leq_{\mathrm{p}} g(a)$, $p \leq_{\mathrm{s}} g(b)$ and $h(b)$ is a factor of p^ω longer than $2\,|p|$. Let $(m, a^i b^j a^k)$ be the g-cover of (r,t), a true h-occurrence of b in $(e, f, \mathbf{c}, G, H)$. Since $p \leq_{\mathrm{s}} g(b)$ we obtain from the primitivity of p that $p \leq_{\mathrm{s}} \mathbf{c}[r, G(m+i+j)]$. Proceeding by the application of Lemma 12 eventually leads to $x_2 + x_3 \geq 5$ p-synchronized overflows.

(IV.) Suppose that $x_3 \geq 3$, $y_3 \geq 3$ and $x_1 + x_3 \geq 5$. Applying Claim 1(III.) on both $x_3 \geq 3$ and $y_3 \geq 3$ leads to x_3 p-synchronized overflows and y_3 s-synchronized overflows where p and s are primitive words given by Claim 1(III.); that is, $p \leq_{\mathrm{p}} g(a)$, $p \leq_{\mathrm{s}} g(b)$ and $s \leq_{\mathrm{p}} g(b)$, $s \leq_{\mathrm{s}} g(a)$ with $h(b)$ being a factor of both p^ω and s^ω longer than $\max(2|p|, 2|s|)$. Notice that we have used the duality of xs and ys and have applied dual form of Claim 1 as well.

From the Periodicity Lemma it follows that p and s are conjugates; therefore of the same length. Let $(m, a^i a^j)$ be the g-cover of (r,t), a true h-occurrence of b in $(e, f, \mathbf{c}, G, H)$. Since $|h(b)| > 2\,|p|$ we can see that either $|\mathbf{c}[r, G(m+i)]| \geq |p|$ or $|\mathbf{c}[G(m+i),t]| \geq |p|$. If $|\mathbf{c}[r, G(m+i)]| \geq |p|$, then from $s \leq_{\mathrm{s}} g(a)$ we have as well $s \leq_{\mathrm{s}} \mathbf{c}[r, G(m+i)]$. From Lemma 4 it follows that $\mathbf{c}[G(m+i),t]$ is prefix comparable with s; therefore, $g(a)$ is prefix comparable with s. Since also $s \leq_{\mathrm{p}} g(b)$, we have obtained a contradiction with g being marked. Therefore, necessarily $p \leq_{\mathrm{p}} \mathbf{c}[G(m+i),t]$ and we can apply Lemma 12 and finally get $x_3 + x_1 \geq 5$ p-synchronized overflows inside $(e, f, \mathbf{c}, G, H)$.

(V.) Suppose that $x_3 \geq 3$. Then from Claim 1(III.) we get the existence of x_3 p-synchronized overflows such that p is a suffix of $g(b)$, a prefix of $g(a)$ and $h(b)$ is a factor of p^ω longer than $2\,|p|$. We have already seen in the third part of this proof that in this case we can gradually add x_2 p-synchronized overflows resulting from g-covers of the type $a^+b^+a^+$. Suppose now that $y_2 \geq 1$ and let $(m, b^i a^j b^k)$ be the g-cover of (r,t), a true h-occurrence of b in $(e, f, \mathbf{c}, G, H)$. From the primitivity of p it follows that $p \leq_{\mathrm{p}} \mathbf{c}[G(m+i),t]$ and we can apply Lemma 12. We have proved that $(e, f, \mathbf{c}, G, H)$ has $x_3 + y_2 + x_2 \geq 5$ p-synchronized overflows and the proof is complete. □

Now, we have everything prepared to finally present the proof of the Main Lemma.

Proof of Main Lemma: From the assumption $|f|_b \geq 9$ we obtain

$$\sum_{i=1}^{3} x_i + \sum_{i=1}^{3} y_i \geq 9.$$

We will proceed by case analysis based on Claim 1, Claim 2 and Claim 3.

Case $x_1 \geq 2$ or $y_1 \geq 2$. Suppose that $x_1 \geq 2$, the other case can be dealt with in a similar way. From Claim 2 we necessarily have $y_2 = 0$, $y_3 \leq 1$ and $y_1 \leq 1$. If $x_1 + x_3 \geq 6$, then the Main Lemma holds according to Claim 3(I.). On the other hand if $x_1 + x_3 \leq 5$ we obtain an inequality

$$\underbrace{x_1 + x_3}_{\leq 5} + x_2 + \underbrace{y_1}_{\leq 1} + \underbrace{y_2 + y_3}_{\leq 1} \geq 9.$$

Then $x_2 \geq 2$ and it follows from dual form of Claim 2(I.) that $y_1 = 0$. Therefore, $x_2 \geq 3$. Since $x_1 \geq 2$, the Main Lemma holds according to Claim 3(II.).

Case $x_1 \leq 1$ and $y_1 \leq 1$. If $x_1 = y_1 = 1$, then by Claim 2(I.) we have $x_2 = y_2 = 0$. Notice that for y_1 we have used dual form of Claim 2. Then

$$\underbrace{x_1 + y_1}_{=2} + \underbrace{x_2 + y_2}_{=0} + x_3 + y_3 \geq 9$$

and $x_3 + y_3 \geq 7$. Therefore, $x_3 \geq 4$ or $y_3 \geq 4$. Suppose that $x_3 \geq 4$. In case that $x_3 \geq 5$, we can apply Claim 1(III.) and obtain desired conclusion. If $x_3 = 4$, then $y_3 \geq 3$ and the Main Lemma holds by Claim 3(IV.). Similarly, we can deal with the possibility $y_3 \geq 4$; notice that in this case we would use dual forms of Claim 1 and Claim 3.

Suppose now that $x_1 \leq 1$ and $y_1 = 0$. If $x_2 + x_3 \geq 5$ or $y_2 + y_3 \geq 5$, then the proof is complete due to Claim 3(III.) (and its dual form). In case that $x_2 + x_3 \leq 4$ and $y_2 + y_3 \leq 4$, we have

$$\underbrace{x_1}_{\leq 1} + \underbrace{y_1}_{=0} + \underbrace{x_2 + x_3}_{\leq 4} + \underbrace{y_2 + y_3}_{\leq 4} \geq 9.$$

Therefore, $x_1 = 1$, $x_2 + x_3 = 4$ and $y_2 + y_3 = 4$. From $x_1 = 1$ it follows by Claim 2(I.) that $y_2 = 0$; therefore $y_3 = 4$. Now, if $x_2 \geq 1$, then the Main Lemma holds due to dual form of Claim 3(V.). On the other hand, if $x_2 = 0$, then we can apply the fourth part of the same claim, which completes the proof. □

5. Towards the classification of non-simple solutions

We have seen that many occurrences of b inside a cyclic block lead to the existence of five p-synchronized overflows for some primitive word p and moreover, partially reveal the structure of image words $h(b)$ and $g(b)$ or $g(a)$.

We know that when dealing with a simple cyclic solution instead of a cyclic block, these assumptions give us already very strong knowledge about the solution itself; it is either $(ab)^i a$ or $a^j b^i$, with $\gcd(i, j) = 1$.

However, since the Main Lemma is formulated more generally, we have the possibility to go even a little bit further and look at the structure of non-simple solutions as well. Taking an advantage of the fact that non-simple solutions of marked morphisms are composed from blocks, we can also apply the Main Lemma to a sufficiently "long" block inside a solution, and get the existence of five p-synchronized overflows for some primitive word p inside the block. An impact of the existence of five p-synchronized overflows inside one of the blocks on the structure of the whole solution is the question for further research.

References

[1] Culik K. II, *A purely homomorphic characterization of recursively enumerable sets*, J. Assoc. Comput. Mach. **26** (1979), no. 2, 345–350.

[2] Culik K. II, Karhumäki J., *On the equality sets for homomorphisms on free monoids with two generators*, RAIRO Inform. Théor. **14** (1980), no. 4, 349–369.

[3] Ehrenfeucht A., Karhumäki J., Rozenberg G., *The (generalized) post correspondence problem with lists consisting of two words is decidable*, Theor. Comput. Sci. **21** (1982), 119–144.

[4] Ehrenfeucht A., Karhumäki J., Rozenberg G., *On binary equality sets and a solution to the test set conjecture in the binary case*, J. Algebra **85** (1983), 76–85.

[5] Hadravová J., Holub Š., *Large simple binary equality words*, Developments in Language Theory, Lecture Notes in Comput. Sci, 5257, Springer, Berlin, 2008, pp. 396–407.

[6] Halava V., Harju T., Hirvensalo M., *Binary (generalized) post correspondence problem*, Theor. Comput. Sci. **276** (2002), no. 1–2, 183–204.

[7] Halava V., Holub Š., *Reduction tree of the binary generalized post correspondence problem*, Internat. J. Found. Comput. Sci., to appear.

[8] Halava V., Holub Š., *Binary (generalized) post correspondence problem is in P*, Turku Centre for Computer Science, 785, 2006.

[9] Holub Š., *Binary morphisms with stable suffix complexity*, Internat. J. Found. Comput. Sci., to appear.

[10] Holub Š., *Binary equality sets are generated by two words*, J. Algebra **259** (2003), no. 1, 1–42.

[11] Holub Š., *A unique structure of two-generated binary equality sets*, Developments in Language Theory, Lecture Notes in Comput. Sci., 2450, Springer, Berlin, 2003, pp. 245–257.

[12] Holub Š., *Binary equality languages for periodic morphisms*, Algebraic Systems, Formal Languages and Conventional and Unconventional Computation Theory, RIMS Kokyuroku, vol. 1366, Kyoto University, Kyoto, 2004.

[13] Karhumäki J., *On recent trends in formal language theory*, 14th International Colloquium on Automata, languages and programming (Karlsruhe, 1987), Springer, Berlin, 1987, pp. 136–162.

[14] Karhumäki J., *Open problems and exercises on words and languages (invited talk)*, in Proceedings of Conference on Algebraic Information, Aristotle University of Thessaloniki, 2005, pp. 295–305.

[15] Lothaire M., *Combinatorics on Words*, Addison-Wesley, Reading, Mass., 1983.

[16] Post E.L., *A variant of a recursively unsolvable problem*, Bull. Amer. Math. Soc. **52** (1946) 264–268.

[17] Rozenberg G., Salomaa A., *Handbook of Formal Languages, Vol. 1: Word, Language, Grammar*, Springer, New York, 1997.

[18] Salomaa A., *Equality sets for homomorphisms of free monoids*, Acta Cybernetica **4** (1978), no. 1, 127–139.

Charles University, Faculty of Mathematics and Physics, Department of Algebra, Sokolovská 83, 186 75 Praha 8, Czech Republic

E-mail: hadravova@ff.cuni.cz

(*Received* October 6, 2010)