ON THE RELATION BETWEEN PERIODICITY AND UNBORDERED FACTORS OF FINITE WORDS*

ŠTĚPÁN HOLUB

Department of Algebra
Charles University in Prague
Czech Republic
holub@karlin.mff.cuni.cz

DIRK NOWOTKA

Institute for Formal Methods in Computer Science (FMI)
Universität Stuttgart
Germany
nowotka@fmi.uni-stuttgart.de

Received (Day Month Year)
Accepted (Day Month Year)
Communicated by (xxxxxxxxxx)

Finite words and their overlap properties are considered in this paper. Let w be a finite word of length n with period p and where the maximum length of its unbordered factors equals k. A word is called unbordered if it possesses no proper prefix that is also a suffix of that word. Suppose $k < p$ in w. It is known that $n \leq 2k - 2$, if w has an unbordered prefix u of length k. We show that, if $n = 2k - 2$ then u ends in ab^i, with two different letters a and b and $i \geq 1$, and b^i occurs exactly once in w. This answers a conjecture by Harju and the second author of this paper about a structural property of maximum Duval extensions. Moreover, we show here that $i < k/3$, which in turn leads us to the solution of a special case of a problem raised by Ehrenfeucht and Silberger in 1979.

Keywords: combinatorics on words; periodicity; unbordered factors; Duval extensions; Ehrenfeucht-Silberger Problem.

2000 Mathematics Subject Classification: 68R15

1. Introduction

Overlaps are one of the central combinatorial properties of words. Despite the simplicity of this concept, its nature is not very well understood and many fundamental questions are still open. For example, problems on the relation between the period of a word, measuring the self-overlap of a word, and the lengths of its unbordered factors, representing the absence of overlaps, are unsolved. The focus of this paper is on the investigation of such questions. In particular, we consider so called Duval

*The work on this article has been supported by the research project MSM 0021620839.

extensions by solving a conjecture [6, 4] about the structure of maximum Duval extensions. This result leads us to a partial answer of a problem raised by Ehrenfeucht and Silberger [5] in 1979.

When repetitions in words are considered then two notions are central: the *period*, which gives the least amount by which a word has to be shifted in order to overlap with itself, and the shortest *border*, which denotes the least (nonempty) overlap of a word with itself. Both notions are related in several ways, for example, the length of the shortest border of a word w is not larger than the period of w, and hence, the period of an unbordered word is its length, moreover, the shortest border itself is always unbordered. Deeper dependencies between the period of a word and its unbordered factors have been investigated for decades; see also the references to related work below.

Let a word w be called a *Duval extension* of u, if $w = uv$ such that u is unbordered and for every unbordered factor x of w holds $|x| \leq |u|$. Let $\pi(w)$ denote the shortest period of a word w. A Duval extension is called *nontrivial* if $|u| < \pi(w)$. It is known that $|v| \leq |u| - 2$ for any nontrivial Duval extension uv [8, 9, 10]. This bound is tight, that is, Duval extensions with $|v| = |u| - 2$ exist. Let those be called *maximum Duval extensions*. The following conjecture has been raised in [6]; see also [4].

Conjecture 1. *Let uv be a maximum Duval extension of $u = u'ab^i$ where $i \geq 1$ and a and b are different letters. Then b^i occurs only once in uv.*

This conjecture is answered positively by Theorem 18 in this paper. Moreover, we show that $i < |u|/3$ in Theorem 20, which leads us to the result that a word z with unbordered factors of length at most k and $\pi(z) > k$ that contains a maximum Duval extension uv with $|u| = k$ is of length at most $7k/3 - 2$. This solves a special case of a conjecture in [5, 1].

Previous Work. In 1979 Ehrenfeucht and Silberger [5] raised the problem about the maximum length of a word w, w.r.t. the length k of its longest unbordered factor, such that k is shorter than the period $\pi(w)$ of w. They conjectured that $|w| \geq 2k$ implies $k = \pi(w)$ where $|w|$ denotes the length of w. That conjecture was falsified shortly thereafter by Assous and Pouzet [1] by the following example:

$$w = a^n ba^{n+1} ba^n ba^{n+2} ba^n ba^{n+1} ba^n$$

where $n \geq 1$ and $k = 3n + 6$ and $\pi(w) = 4n + 7$ and $|w| = 7n + 10$, that is, $k < \pi(w)$ and $|w| = 7k/3 - 4 > 2k$. Assous and Pouzet in turn conjectured that $3k$ is the bound on the length of w for establishing $k = \pi(w)$. Duval [3] did the next step towards solving the problem. He established that $|w| \geq 4k - 6$ implies $k = \pi(w)$ and conjectures that, if w possesses an unbordered prefix of length k, then $|w| \geq 2k$ implies $k = \pi(w)$. Note that a positive answer to Duval's conjecture yields the bound $3k$ for the general question. Despite some partial results [11, 4, 7] towards a solution, Duval's conjecture was only solved in 2004 [8, 9] with a new proof given in [10]. The proof of (the extended version of) Duval's conjecture lowered the bound for Ehrenfeucht and Silberger's problem to $3k-2$ as conjectured by Assous and Pouzet [1]. However, there remains a gap of $k/3$ between that bound and the largest known example, which is given above. With this paper we take the next step towards the solution of the problem by Ehrenfeucht and Silberger by establishing the optimal bound of $7k/3$ for a special case.

2. Notation and Basic Facts

Let us fix a finite set A, called alphabet, of letters. Let A^* denote the monoid of all finite words over A including the *empty word* denoted by ε. In general, we denote variables over A by a, b, c, d and e and variables over A^* are usually denoted by f, g, h, r through z, and α, β, and γ including their subscripted and primed versions. The letters i through q are to range over the set of nonnegative integers.

Let $w = a_1a_2\cdots a_n$. The word $a_na_{n-1}\cdots a_1$ is called the *reversal* of w denoted by $\overline{w}$. We denote the length n of w by $|w|$, in particular $|\varepsilon| = 0$. If w is not empty, then let ${}^\bullet w = a_2\cdots a_{n-1}a_n$ and $w^\bullet = a_1a_2\cdots a_{n-1}$. We define ${}^\bullet\varepsilon = \varepsilon^\bullet = \varepsilon$. Let $0 \le i \le n$. Then $u = a_1a_2\cdots a_i$ is called a *prefix* of w, denoted by $u \le_{\rm p} w$, and $v = a_{i+1}a_{i+2}\cdots a_n$ is called a *suffix* of w, denoted by $v \le_{\rm s} w$. A prefix or suffix is called proper when $0 < i < n$. An integer $1 \le p \le n$ is a *period* of w if $a_i = a_{i+p}$ for all $1 \le i \le n-p$. The smallest period of w is called *the period* of w, denoted by $\pi(w)$. A nonempty word u is called a *border* of a word w, if $w = uy = zu$ for some words y and z. We call w *bordered*, if it has a border that is shorter than w, otherwise w is called *unbordered*. Note that every bordered word w has a minimum border u such that $w = uvu$, where u is unbordered.

Let $\lhd$ be a total order on A. Then $\lhd$ extends to a *lexicographic order*, also denoted by $\lhd$, on A^* with $u \lhd v$ if either $u \le_{\rm p} v$ or $xa \le_{\rm p} u$ and $xb \le_{\rm p} v$ and $a \lhd b$. Let $\overline{\lhd}$ denote a lexicographic order on the reversals, that is, $u \mathrel{\overline{\lhd}} v$ if $\overline{u} \lhd \overline{v}$. Let $\lhd^a$ and $\lhd_b$ and $\lhd_b^a$ denote lexicographic orders where the maximum letter or the minimum letter or both are fixed in the respective orders on A. We establish the following convention for the rest of this paper: in the context of a given order $\lhd$ on A, we denote the inverse order of $\lhd$ by $\blacktriangleleft$. A $\lhd$-maximal prefix (suffix) α of a word w is defined as a prefix (suffix) of w such that $v \mathrel{\overline{\lhd}} \alpha$ ($v \lhd \alpha$) for all $v \le_{\rm p} w$ ($v \le_{\rm s} w$).

The notion of maximum pre- and suffix are symmetric. It is general practice that facts involving the maximum ends of words are mostly formulated for maximum suffixes. The analogue version involving maximum prefixes is tacitly assumed.

Remark 1. *Any maximum suffix of a word w is longer than $|w| - \pi(w)$ and occurs only once in w.*

Indeed, let α be the $\lhd$-maximal suffix of u for some order $\lhd$. Then $u = x\alpha y$ and $\alpha \lhd \alpha y$ implies $y = \varepsilon$ by the maximality of α. If $w = uv\alpha$ with $|v| = \pi(w)$, then $u\alpha \le_{\rm p} w$ gives a contradiction again.

Let an integer q with $0 \le q < |w|$ be called *point* in w. A nonempty word x is called a *repetition word* at point q if $w = uv$ with $|u| = q$ and there exist words y and z such that $x \leq_s yu$ and $x \leq_p vz$. Let $\pi(w, q)$ denote the length of the shortest repetition word at point q in w. We call $\pi(w, q)$ the *local period* at point q in w. Note that the repetition word of length $\pi(w, q)$ at point q is necessarily unbordered and $\pi(w, q) \leq \pi(w)$. A factorization $w = uv$, with $u, v \neq \varepsilon$ and $|u| = q$, is called *critical*, if $\pi(w, q) = \pi(w)$, and if this holds, then q is called a *critical point*.

Let $\lhd$ be an order on A. Then the shorter of the $\lhd$-maximal suffix and the $\blacktriangleleft$-maximal suffix of some word w is called a *critical suffix* of w. Similarly, we define a *critical prefix* of w by the shorter of the two maximum prefixes resulting from some order and its inverse. This notation is justified by the following formulation of the so called critical factorization theorem (CFT) [2], which relates maximum suffixes and critical points.

Theorem 2 (CFT) *Let $w \in A^*$ be a nonempty word and γ be a critical suffix of w. Then $|w| - |\gamma|$ is a critical point.*

Let uv be a *Duval extension* of u if u is an unbordered word and every factor in uv longer than $|u|$ is bordered. A Duval extension uv of u is called trivial if $v \leq_p u$. The following fact was conjectured in [3] and proven in [8, 9, 10].

Theorem 3. *Let uv be a nontrivial Duval extension of u. Then $|v| \leq |u| - 2$.*

Following Theorem 3 let a *maximum Duval extension* of u be a nontrivial Duval extension uv with $|v| = |u| - 2$. This length constraint on v will often tacitly be used in the rest of this paper.

Let wuv be an *Ehrenfeucht-Silberger extension* of u if both uv and $\overline{wu}$ are Duval extensions of u and $\overline{u}$, respectively, moreover, uv and $\overline{wu}$ are called the Duval extensions corresponding to the Ehrenfeucht-Silberger extension of u.

Ehrenfeucht and Silberger were the first to investigate the bound on the length of a word w, w.r.t. the length k of its longest unbordered factors, such that $k < \pi(w)$. Some bounds have been conjectured. The latest such conjecture is taken from [9].

Conjecture 2. *Let wuv be a nontrivial Ehrenfeucht-Silberger extension of u. Then $|wv| < \frac{4}{3}|u|$.*

3. Periods and Maximum Suffixes

Note the following simple but noteworthy fact.

Lemma 4. *Let u be an unbordered word, and let v be a word that does not contain u. Let α be the $\lhd$-maximal suffix of u. Then any prefix w of uv such that α is a suffix of w, is unbordered.*

Proof. Certainly, $|w| \geq |u|$ by Remark 1. Suppose that w has a shortest border h. Then $|h| < |u|$ otherwise $u \leq_p h$ and u occurs in v since h is the shortest border; a contradiction. But now, h is a border of u; again a contradiction. $\square$

This implies immediately the following version of Lemma 4 for Duval extensions, which will be used frequently further below.

Lemma 5. *Let uv be a nontrivial Duval extension of u, and let α be the $\lhd$-maximal suffix of u. Then uv contains just one occurrence of α.*

The next lemma highlights an interesting fact about borders involving maximum suffixes. It will mostly be used on maximum prefixes of words, the dual to maximum suffixes, in later proofs. However, it is general practice to reason about ordered factors of words by formulating facts about suffixes rather than prefixes. Both ways are of course equivalent. We have chosen to follow general practice here despite its use on prefixes later in this paper.

Lemma 6. *Let αa be the $\lhd$-maximal suffix of a word wa where a is a letter. Let u be a word such that αa is a prefix of u and wb is a suffix of u, with $b \neq a$ and $b \lhd a$. Then u is either unbordered, or its shortest border has the length at least $|w| + 2$.*

Proof. Suppose that u has a shortest border hb. If $|h| < |\alpha|$ then $hb \leq_{\mathrm{p}} \alpha$ and $h \leq_{\mathrm{s}} \alpha$ and $hb \lhd ha$ contradict the maximality of αa. Note that $|h| \neq |\alpha|$ since $a \neq b$. If $|\alpha| < |h| \leq |w|$ then $\alpha a \leq_{\mathrm{p}} h$, and hence, αa occurs in w contradicting the maximality of αa again; see Remark 1. Hence, $|hb| \geq |w| + 2$. □

The next lemma is taken from a result in [7] about so called minimal Duval extensions. However, the shorter argument given here (including the use of Lemma 6) gives a more concise proof than the one in [7].

Lemma 7. *Let uv be a nontrivial Duval extension of u where $u = xazb$ and $xc \leq_{\mathrm{p}} v$ and $a \neq c$. Then bxc occurs in u.*

Proof. Let ya be the $\lhd^a$-maximal suffix of xa. Consider the factor $yazbxc$ of uv, which is longer than u and therefore bordered with a shortest border r. Now, Lemma 6 implies that $|r| > |xc|$, and hence, $bxc \leq_{\mathrm{s}} r$ occurs in u. □

4. Some Facts about Certain Suffixes of a Word

This section is devoted to the foundational proof technique used in the remainder of this paper. The main idea is highlighted in Lemma 8, which identifies a certain unbordered factor of a word.

Lemma 8. *Let α be the $\lhd$-maximal suffix and β be the $\blacktriangleleft$-maximal suffix of a word u, and let v be such that neither α nor β occur in uv more than once. Let a be the last letter of v and b be the first letter of x where $x \leq_{\mathrm{s}} \alpha v^{\bullet}$ and $|x| = \pi(\alpha v^{\bullet})$.*

If $\pi(\alpha v) > \pi(\alpha v^{\bullet})$, then αv is unbordered, in case $a \lhd b$, and βv is unbordered, in case $b \lhd a$.

Proof. Let γ be the longest border of $\alpha v^\bullet$. Note that $|\gamma| < |\alpha|$ since ${}^\bullet\alpha v$ does not contain the critical suffix of u, by assumption. We have $\alpha = \gamma b\alpha'$ and $\alpha v = v'\gamma a$. Note that $\pi(\alpha v^\bullet) = |v'|$, and the inequality $\pi(\alpha v) > \pi(\alpha v^\bullet)$ means $a \neq b$.

Suppose that $a \lhd b$. We claim that αv is unbordered in this case. Suppose the contrary, and let αv have a shortest border ha. Then $|h| < |\gamma|$ otherwise either $a = b$, if $|h| = |\gamma|$, or γ is not the longest border of $\alpha v^\bullet$, if $|h| > |\gamma|$; a contradiction in both cases. But now $\alpha \lhd hb\alpha'$ since $ha \leq_{\mathrm{p}} \alpha$ and $a \lhd b$ contradicting the maximality of α because $hb\alpha' \leq_{\mathrm{s}} \alpha$.

Suppose that $b \lhd a$. In this case the word βv is unbordered. To see this suppose that βv has a shortest border ha. The assumption that uv contains just one occurrence of the maximal suffixes implies that ha is a proper prefix of β. If $|h| \geq |\gamma|$ then γa occurs in u contradicting the maximality of α since $\gamma b \leq_{\mathrm{p}} \alpha \lhd \gamma a$. But now $ha \leq_{\mathrm{p}} \beta \blacktriangleleft hb\alpha'$ (since $b \lhd a$) contradicting the maximality of β. □

Proposition 9. *Let uv be a nontrivial Duval extension of u, and let α be a critical suffix w.r.t. an order $\lhd$. Then $|v| < \pi(\alpha v) \leq |u|$.*

Proof. If $|v| \geq \pi(\alpha v)$ then α occurs twice in αv contradicting Lemma 5. Suppose that $\pi(\alpha v) > |u|$, and let z be the shortest prefix of v such that already $\pi(\alpha z) > |u|$. Then $\pi(\alpha z) > \pi(\alpha z^\bullet)$, and Lemma 8 implies that either αz or βz is unbordered, where β is the $\blacktriangleleft$-maximal suffix of u. This contradicts the assumption that uv is a Duval extension, since both the candidates are longer than u, which follows from $\pi(\alpha z) > |u|$ and $|\beta| > |\alpha|$. □

5. About Maximum Duval Extensions

In this section we consider the general results of the previous section for the special case of Duval extensions, which leads is to the main results, Theorem 18 and 20. Theorem 18 confirms a conjecture in [6]. Theorem 20 constitutes a further step to answer Conjecture 2.

Definition 10. *Let uv be a Duval extension of u. The suffix s of uv is called a* trivial suffix *if $\pi(s) = |u|$ and s is of maximum length.*

Note that $s = uv$, if uv is a trivial Duval extension, and $as \leq_{\mathrm{s}} uv$ with $\pi(as) > |u|$, if uv is a nontrivial Duval extension. Moreover, Proposition 9 implies that $|s| \geq |\alpha v|$ where α is any critical suffix of u.

Let us begin with considerations about the periods of suffixes of maximum Duval extensions.

Lemma 11. *Let uv be a maximum Duval extension of u, and let $\lhd$ be an order such that the $\lhd$-maximal suffix α is critical. Then $\pi(\alpha v) = |u|$.*

Proof. It follows from Proposition 9 that $|u| - 1 \leq \pi(\alpha v) \leq |u|$ since $|v| = |u| - 2$. Suppose $\pi(\alpha v) = |u| - 1$. Let $w\alpha$ be the longest suffix of u such that $\pi(w\alpha v) = |u| - 1$.

We have $w\alpha \neq u$ since u is unbordered. We can write $w\alpha v = w\alpha v' w\alpha^\bullet$, where v' is a prefix of v such that $|w\alpha v'| = |u| - 1$. The maximality of $w\alpha$ implies that $aw\alpha$ is a suffix of u, and $bw\alpha^\bullet$ is a suffix of αv, with $a \neq b$.

Choose a letter c in $w\alpha^\bullet$ such that $c \neq a$. Such a letter exists for otherwise $aw\alpha^\bullet \in a^+$ and α is just a letter, different from a. But this implies $u \in a^+\alpha$ and $v \notin a^+$ for uv to be nontrivial, that is, $v'd \leq_{\mathrm{p}} v$ with $d \neq a$; a contradiction since $uv'd$ is unbordered in this case.

Consider the $\overline{\lhd}^c$-maximal prefix of $bw\alpha^\bullet$ denoted by bt. Note that $|t| \geq 1$. We claim that $aw\alpha v' t$ is unbordered. Suppose on the contrary that r is the shortest border of $aw\alpha v' t$. By Lemma 6 applied to the reversal of $aw\alpha v' t$, the border r is longer than $bw\alpha^\bullet$. Hence, r contains α contradicting Lemma 5. But now, since $|w\alpha v'| = |u| - 1$ and $|t| \geq 1$, the unbordered factor $aw\alpha v' t$ is longer than u; a contradiction. □

Lemma 12. *Let uv be a maximum Duval extension of u, let a be the last letter of u, and let xv be the trivial suffix of uv. Then $|\alpha| \leq |x|$ for the $\lhd^a$-maximal suffix α of any order $\lhd^a$.*

Proof. Suppose on the contrary that $|\alpha| > |x|$, which implies that the $\blacktriangleleft^a$-maximal suffix β is critical and $\beta \leq_{\mathrm{s}} x$ by Lemma 11. Since uv is nontrivial, we can write $u = u'cwba$ and $v = v'dw$ where $wba = x$.

Consider the maximum prefix t of dw with respect to any order on the reversals where d is maximal. Note that $d \leq_{\mathrm{s}} t$. The word $cwbav't$ is longer than u, therefore it is bordered. Let r be its shortest border. By Lemma 6, we have $|cw| < |r|$. Lemma 5 implies that $r = cwb$, and we have $d = b$ since $d \leq_{\mathrm{s}} t$. Note that $|t| < |bw|$ otherwise $t = bw = wb$, which implies $|u| = \pi(xv) = \pi(wbav'bw) = \pi(bwav'bw) \leq |v| + 1 < |u|$; a contradiction. Hence, $te \leq_{\mathrm{p}} bw$ for some letter $e \neq b$. Moreover, $e \neq a$ since $\beta^\bullet \leq_{\mathrm{s}} r$ and β does occur only once in βv by Lemma 5.

Consider the factor $\alpha v' te$, which is longer than u, and hence, bordered. Let s be the shortest border of $\alpha v' te$. Note that $|s| < |\beta|$ otherwise $\beta^\bullet e \leq_{\mathrm{s}} s$ contradicting the maximality of β since $\beta = \beta^\bullet a \blacktriangleleft^a \beta^\bullet e$. Let $s = \beta' e$ where $\beta' \leq_{\mathrm{s}} \beta^\bullet$. But then $\beta' e \leq_{\mathrm{p}} \alpha \lhd^a \beta' a$ and $\beta' a \leq_{\mathrm{s}} u$ contradicting the maximality of α. □

Lemma 13. *Let uv be a maximum Duval extension of $u = u'ab$ where a and b are letters. Then a occurs in u'.*

Proof. Suppose on the contrary that a does not occur in u'. Note that b occurs in u' by Lemma 7. So, we may assume that $a \neq b$. Moreover, we have that also a letter c different from a and b has to occur in u' otherwise $u = b^i ab$ and $v = b^j dv'$ for some $d \neq b$ and $j < i$, but then $ub^j d$ is unbordered; a contradiction.

Let β be the maximum suffix of u w.r.t. some order $\lhd_c^b$, and let α be a maximum suffix of u w.r.t. the order $\blacktriangleleft_c^b$. Let γ be the shorter of the two suffixes α and β, and note that $|\gamma| > 2$.

Lemma 11 implies $\pi(\gamma v) = |u|$. Let $w\gamma v$ be the trivial suffix of uv. We have that $u \neq w\gamma$ since uv is a nontrivial Duval extension of u. Therefore, we can write $u = u'dw\gamma$ and $v = v'ew\gamma^{\bullet\bullet}$ where d and e are different letters and $|w\gamma v'e| = |u|$. Note that e occurs in $u^{\bullet\bullet}$ otherwise $uv'e$ is unbordered; a contradiction. Consider an order $\lhd^e$ and let t be the $\overline{\lhd}^e$-maximal prefix of $ew\gamma^{\bullet\bullet}$.

The word $dw\gamma v't$ is longer than u, therefore it is bordered. Let r be its shortest border. By Lemma 6, we have $|dw\gamma| - 2 < |r|$. Lemma 5 implies that $|r|$ is exactly $|dw\gamma| - 1$, whence $r = dw\gamma^{\bullet}$. Clearly, the letter e is a suffix of t, and thus also of r, which implies that e is a suffix of $u^{\bullet}$; a contradiction since $e \neq a$. $\square$

The following example shows that the requirement of a maximum Duval extension is indeed necessary in Lemma 13.

Example 14. *Let a, b, and c be different letters, and consider $u = a^iba^{i+j}bcb$ and $v = a^{i+j}ba^{i-1}$ with $i, j \geq 1$. Then $u.v = a^iba^{i+j}bcb.a^{i+j}ba^{i-1}$ is a nontrivial Duval extension of length $2|u| - 4$ such that c occurs only in the second last position of u. However, a maximum Duval extension of a word $|u|$ has length $2|u| - 2$.*

The next lemma highlights a relation between the trivial suffix of a maximum Duval extension uv and the set $\mathrm{alph}(u)$ of all letters occurring in u.

Lemma 15. *Let uv be a maximum Duval extension of u and wxw be the trivial suffix of uv where $|wx| = |u|$. Then either $\mathrm{alph}(w) = \mathrm{alph}(u)$ or there exists a letter b such that $\mathrm{alph}(w) = \mathrm{alph}(u) \setminus \{b\}$ and $u = u'bb$ and bb does not occur in u'.*

Proof. Suppose contrary to the claim that $|\mathrm{alph}(w)| < |\mathrm{alph}(u)|$ and for any $b \in \mathrm{alph}(u) \setminus \mathrm{alph}(w)$ we have bb is not a suffix of u or bb occurs in $u^{\bullet\bullet}$.

Let $btwac \leq_{\mathrm{s}} u$ where $a, b, c \in \mathrm{alph}(u)$ and b does not occur in tw. Consider $btwxw$, which is longer than u and therefore has to be bordered. Let r be the shortest border of $btwxw$. Certainly, $|w| < |r|$ since $b \leq_{\mathrm{p}} r$ and $b \notin \mathrm{alph}(w)$. Moreover, $btw \leq_{\mathrm{p}} r$ implies $\pi(btwxw) \leq |u|$ contradicting the maximality of wxw. So, we note that $|w| < |r| < |btw|$.

Suppose $a \neq b$. Let $v = v'r$ and consider the factor $twacv'b$, which has to be bordered since $|twacv'b| = |twacv| - |r| + 1 > |acv| = |u|$. Let s be the shortest border of $twacv'b$. We have $|s| > |twa|$ because b is a suffix of s and does not occur in tw and $a \neq b$ by assumption. But now, $twac \leq_{\mathrm{p}} s$ contradicting Lemma 5 since wac contains a maximum suffix of u.

Suppose $a = b$. This is the only case where we need to consider that either $bb \not\leq_{\mathrm{s}} u$ or bb occurs at least twice in u. Let $d \in \mathrm{alph}(u)$ be such that $d = c$, if $c \neq b$, and d be an arbitrary letter different from b otherwise. Consider an order $\lhd_d^b$ on $\mathrm{alph}(u)$. Let α be the $\lhd_d^b$-maximal suffix of u. Note that $|\alpha| > |wbc|$ since either $c = b$ or $c = d$. If $c = b$ then $bb \leq_{\mathrm{p}} \alpha$ occurs in $u^{\bullet}$ by assumption. If $c = d$ then be occurs in $u^{\bullet}$ for some letter e by Lemma 13 where we have $be \leq_{\mathrm{p}} \alpha$ since either $d \lhd_d^b e$ or $e = d$. Since every critical suffix of u is a suffix of wbc by Lemma 11 and

$\alpha \not\leq_{\mathrm{s}} wbc$, we have that the $\blacktriangleleft_d^b$-maximal suffix β is critical and $\beta \leq_{\mathrm{s}} wbc$. Moreover, $|\beta| > 2$ since $bc \leq_{\mathrm{s}} u$ and d occurs in $u^\bullet$ by Lemma 7. We have that $\beta^{\bullet\bullet} \leq_{\mathrm{s}} w$, and hence, $\beta^{\bullet\bullet} \leq_{\mathrm{s}} r$. From $|r| < |btw|$ follows that $\beta^{\bullet\bullet}c'$ occurs in tw where c' is a letter in tw, and therefore $c' \neq b$. But this contradicts the maximality of β since $\beta^{\bullet\bullet}b \blacktriangleleft_d^b \beta^{\bullet\bullet}c'$. □

The next two results, Lemma 16 and 17, constitute a case split of the proof of Theorem 18. Namely, the cases when exactly two or more than two letters occur in a maximum Duval extension.

Lemma 16. *Let uv be a maximum Duval extension of $u = u'ab^i$ where $i \geq 1$ and $|\mathrm{alph}(u)| > 2$ and $a \neq b$. Then u' does not contain the factor b^i.*

Proof. Suppose, contrary to the claim, that b^i occurs in u'. Consider the trivial suffix $wcbv'dw$ of uv where $|cbv'dw| = |u|$ and $c \in \{a, b\}$. We can write $u = u'ewcb$ with $d \neq e$ since $|u| > |wcb|$. We have that $\mathrm{alph}(w) = \mathrm{alph}(u)$ by Lemma 15. Choose a letter f in dw such that $f \neq e$ and $f \neq c$. Let $\lhd_e^f$ be an order. Let dt be the $\overline{\lhd}_e^f$-maximal prefix of dw. The word $ewcbv't$ is longer than u, therefore it is bordered. Let r be its shortest border. By Lemma 6, we have $|dw| < |r|$. Lemma 5 implies that $|r|$ is exactly $|dwc|$, and hence, $r = ewc$. Clearly, the letter f is a suffix of t, and thus also of r, which implies that $f = c$; a contradiction. □

Lemma 17. *Let uv be a maximum Duval extension of $u = u'ab^i$ over a binary alphabet where $i \geq 1$ and $a \neq b$. Then u' does not contain the factor b^i and $awbb \leq_{\mathrm{s}} u$ and $v = v'bw$ where $wbbv$ is the trivial suffix of uv.*

Proof. Let s be the trivial suffix of uv, and let $u = u_0cwdb$ and $v = v'ew$ where $wdbv'ew = s$. Note that $c \neq e$ by the maximality of s. Let $\lhd$ be the order such that $a \lhd b$.

Suppose $c = b$ and $e = a$. Let t be the $\overline{\blacktriangleleft}$-maximal prefix of aw. Consider the factor $bwdbv't$, which is longer than $|u|$ and hence bordered. Let r be its shortest border. Lema 6 implies that $|bw| < |r|$. Lemma 5 implies that $r = bwd$, in fact, $r = bwa$ since $a \leq_{\mathrm{s}} t$. Note that $|t| \leq |w|$ otherwise $r = bwa = baw = ba^{|w|+1}$ contradicting Lemma 15. So, we have $tb \leq_{\mathrm{p}} aw$ by the maximality of t. But now wab occurs in v, and hence, the critical suffix of u occurs in v by Lemma 11 contradicting Lemma 5.

We conclude that $c = a$ and $e = b$. Consider the $\lhd$-maximal suffix β of u. Suppose contrary to the claim that b^i occurs in u'. Then $b^ja \leq_{\mathrm{p}} \beta$ for some $j \geq i$.

Let t be the $\overline{\lhd}$-maximal prefix of bw. Similarly to the reasoning above, we consider the factor $awdbv't$ and conclude that it has the border $r = awb$ and $d = b$ and $ta \leq_{\mathrm{p}} bw$. Lemma 12 implies that $\beta \leq_{\mathrm{s}} wbb$. Note that b^j is a power of b in u of maximum size and occurs in w by assumption, and hence, $b^j \leq_{\mathrm{s}} t$. But now, $b^j \leq_{\mathrm{s}} r$ and $b^{j+1} \leq_{\mathrm{s}} u$; a contradiction. □

The main result follows directly from the previous two lemmas.

Theorem 18. *Let uv be a maximum Duval extension of $u = u'ab^i$ where $i \geq 1$ and $a \neq b$. Then b^i occurs only once in uv.*

Indeed, b^i does not occur in u' by Lemma 16 and 17. If b^i occurs in $b^{i-1}v$, that is, $b^{i-1}v = wb^iv'$, then $u'abwb^i$ is unbordered; a contradiction.

Let us consider the results obtained so far for the special case of a binary alphabet in the following remark.

Remark 19. *Let $uv \in \{a, b\}^+$ be a maximum Duval extension with $b \leq_s u$, and let wv be the trivial suffix of uv.*

Theorem 18 implies that the $\lhd_a^b$-maximal suffix of u is critical and equal to b^i. Lemma 7 implies that $i \geq 2$. Lemma 15 implies that a occurs in w, and in particular, $w \in a^+bb$, if $i = 2$. Lemma 17 implies that $axb^i \leq_s u$ and $bxb^{i-2} \leq_s v$, where $w = xb^i$.

Theorem 20. *Let uv be a maximum Duval extension of $u = u'ab^i$ where $i \geq 1$ and $a \neq b$. Then $3i \leq |u|$.*

Proof. The shortest possible maximum Duval extension of a word u is of the form uv with $u = abaabb$ and $v = aaba$. This proves the claim for $i \leq 2$. Assume $i > 2$ in the following.

Let $cb^k \leq_s v$ with $c \neq b$. Lemma 11 implies that $k \geq i - 2$, and Lemma 2 yields $k \leq i - 1$. Consider the shortest border h of uv. Then $|h| < |u| - 2$ otherwise uv is trivial. Let $h = gb^k$, and let j be the maximum integer such that $gb^j \leq_p u$. Clearly, $k \leq j \leq i - 1$ since b^i occurs only as a suffix of u. Let $u = gb^jfb^i$. Note that

$$b \notin \{\text{pref}_1(g), \text{pref}_1(f), \text{suff}_1(g), \text{suff}_1(f)\} . \tag{2}$$

Next we show that b^k occurs in g or f. Suppose the contrary, that is, neither g nor f contains b^k. Consider the shortest border x of fb^iv. We have $|x| < |fb^i|$, since b^i does not occur in v. Property (2) and the assumption that b^k does not occur in f imply that $x = fb^k$. Let $v = v'fb^k$. Consider the shortest border y of $b^jfb^iv'f$. Again, we have $|y| < |b^jfb^i|$ since b^i does not occur in v, and property (2) implies that $y = b^jh$. Let $v = v''b^jfb^k$. Finally, consider the shortest border z of $uv''b^j$. Property (2) and the assumption that b^k does not occur in g or f imply that either $z = gb^j$ or $z = gb^jfb^j$. The former implies that $uv = gb^jfb^igb^jfb^k$ is a trivial Duval extension, and the latter implies that $|u| < |v|$; a contradiction in both cases.

We conclude that b^k occurs in g or f. Let $u = u_1b^mu_2b^nu_3b^i$ where u_1, u_2, and u_3 are not empty and neither begin nor end with b and $k \leq m, n \leq i - 1$. The claim is proven if $|u_1u_2u_3| > 3$ or $m = i - 1$ or $n = i - 1$. Suppose the contrary, that is, u_1, u_2, and u_3 are letters and $m = i - 2$ and $n = i - 2$ and $k = i - 2$.

Let us consider the shape of v next. Note that every factor of length 2 in v contains b otherwise there exists a prefix w of v that ends in two letters not equal to b and uw is unbordered; a contradiction. Moreover, for every power $b^{k'}$ in v holds

$i-1 \leq k'$ otherwise $w'cb^{k'}d$ is a prefix of v where c and d are letters different from b and $b^m u_2 b^n u_3 b^i w' c b^{k'} d$ is unbordered; a contradiction. Considering possible borders of words uv_1b^{i-2}, $uv_1b^{i-2}v_2b^{i-2}$, $u_2b^{i-2}u_3b^iv_1b^{i-2}v_2b^{i-2}$ and $u_3b^iv_1b^{i-2}v_2b^{i-2}v_3b^{i-2}$ we deduce that $v_1 = u_1$, $v_2 = u_2$ and $v_3 = u_3$; a contradiction since uv is assumed to be nontrivial. This proves the claim. □

Corollary 21. *Let w be a nontrivial Ehrenfeucht-Silberger extension of u such that one of its corresponding Duval extensions is of maximum length. Then $|w| < \frac{7}{3}|u| - 2$.*

Indeed, suppose on the contrary that $w = xuv$ and uv is a maximum Duval extension with $ab^i \leq_{\mathrm{s}} u$ and $|x| \geq i$ where $a \neq b$. The case where $\overline{xu}$ is a maximum Duval extension is symmetric. Now, either $b^i \leq_{\mathrm{s}} x$ or $eb^j \leq_{\mathrm{s}} x$ with $j < i$ and $e \neq b$. If $eb^j \leq_{\mathrm{s}} x$ with $j < i$ and $e \neq b$, then $eb^j u$ is unbordered; a contradiction. If $b^i \leq_{\mathrm{s}} x$ then $b^i u b^{-i}$ is unbordered by Theorem 18, and its Duval extension $b^i uv$ is trivial, since it is too long; a contradiction.

The following example is taken from [1].

Example 22. *Consider the following word xuv where we separate the factors x, u, and v for better readability*

$$x.u.v = b^{i-2}.ab^{i-1}ab^{i-2}ab^i.ab^{i-2}ab^{i-1}ab^{i-2}$$

where $i > 2$. We have that the largest unbordered factors of xuv are of length $3i$, namely the factors $u = ab^{i-1}ab^{i-2}ab^i$ and $b^iab^{i-2}ab^{i-1}a$, and $\pi(xuv) = 4i - 1$, and hence, xuv is a nontrivial Ehrenfeucht-Silberger extension of u. Note that uv is a maximum Duval extension. We have $|xuv| = 7i - 4 = \frac{7}{3}|u| - 4$.

References

[1] R. Assous and M. Pouzet. Une caractérisation des mots périodiques. *Discrete Math.*, 25(1):1–5, 1979.

[2] M. Crochemore and D. Perrin. Two-way string-matching. *J. ACM*, 38(3):651–675, 1991.

[3] J.-P. Duval. Relationship between the period of a finite word and the length of its unbordered segments. *Discrete Math.*, 40(1):31–44, 1982.

[4] J.-P. Duval, T. Harju, and D. Nowotka. Unbordered factors and Lyndon words. *Discrete Math.*, 308(11):2261–2264, 2008.

[5] A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. *Discrete Math.*, 26(2):101–109, 1979.

[6] T. Harju and D. Nowotka. Duval's conjecture and Lyndon words. TUCS Tech. Rep. 479, Turku Centre of Computer Science, Finland, 2002.

[7] T. Harju and D. Nowotka. Minimal Duval extensions. *Internat. J. Found. Comput. Sci.*, 15(2):349–354, 2004.

[8] T. Harju and D. Nowotka. Periodicity and unbordered words. In *STACS 2004 (Montpellier)*, volume 2996 of *Lecture Notes in Comput. Sci.*, pages 294–304, Berlin, 2004. Springer-Verlag.

[9] T. Harju and D. Nowotka. Periodicity and unbordered words: A proof of the extended Duval conjecture. *J. ACM*, 54(4), 2007.

[10] Š. Holub. A proof of the extended Duval's conjecture. *Theoret. Comput. Sci.*, 339(1):61–67, 2005.

[11] F. Mignosi and L. Q. Zamboni. A note on a conjecture of Duval and Sturmian words. *Theor. Inform. Appl.*, 36(1):1–3, 2002.