

On Periodic Properties of Circular Words

László Hegedüs^{a,*}, Benedek Nagy^{a,b}

^a*Department of Computer Science, Faculty of Informatics,
University of Debrecen, Debrecen, Hungary*

^b*Department of Mathematics, Faculty of Arts and Sciences,
Eastern Mediterranean University, Famagusta, North Cyprus, Mersin-10, Turkey*

Abstract

The conjugacy relation defines a partition of words into equivalence classes. We call these classes circular words. Periodic properties of circular words are investigated in this article. The Periodicity Theorem of Fine and Wilf does not hold for weak periods of circular words; instead we give a strict upper bound on the length of a non-unary circular word that has two given relatively prime weak periods. Weak periods also lead to a way of representing circular words in a more compact form. We investigate in which cases are these representations unique or minimal. We will also analyze weak periods of circular Thue-Morse, Fibonacci and Christoffel words.

Keywords: circular words, weak period, combinatorics on words

1. Introduction

Combinatorics on words is a relatively young subfield of computer science which mostly deals with the structure of finite and infinite sequences of symbols or words. It has strong connections to various areas of mathematics, namely algebra, number theory, game theory, etc. The first book on the subject was written by several authors under the pen name M. Lothaire (see [8]), but most researchers consider the papers of A. Thue to be one of the first publications in this field [17, 18]. The wide range of applications in computer science (e.g., in string matching, data compression, bioinformatics, etc.) show the importance of investigating this area.

Most publications are about linear (finite or infinite) words and only a small fraction of them present properties of circular words (sometimes called cyclic words [4] or necklaces [16]). Recent research shows that circular sequences can also appear in nature. For example, circular DNA sequences can be formed if the right enzymes are introduced to a collection (or so called soup) of DNA strands.

*Corresponding author

Email addresses: `hegedus.laszlo@inf.unideb.hu` (László Hegedüs),
`nbenedek.inf@gmail.com` (Benedek Nagy)

It can be done in a formal way, e.g., in the theory of H systems [11] introduced by T. Head. Periodic discrete functions can be considered to be circular words too. One of the most famous results about periods of linear words belongs to N. J. Fine and H. S. Wilf [5]. We will show later that this theorem does not necessarily apply to weak periods of circular words.

Though most of the research is done on linear words, there are some interesting articles on circular words too. In relation to Weinbaum factorizations the reader can consult [4]. Chapter 4 of the dissertation of D. Nowotka [10] deals with unbordered conjugates of words. Results about pattern avoidance of circular words can be found in [3, 6, 15]. Furthermore, [12, 13] contains some applications to integer sequences. The definition of circular words is related to cyclic (or circular) codes [19] that are of great interest among mathematicians. This confirms that the role of circular words in mathematics and computer science is not negligible.

This article extends the results presented in [7] where the authors gave the definition of weak and strong periods of circular words and presented results about the relation between different weak periods of a circular word. We also give improved proofs of two main theorems. We begin our paper by defining the basic notations and notions of combinatorics on words (both linear and circular) in Section 2. Then we summarize the properties of strong and weak periods in Sections 3 and 4, respectively. In Section 5 we present some results about representations of circular words. At the end, in Section 6 we discuss some properties of circular words that are obtained from well known sequences.

2. Preliminaries

We will use the following notions and notation throughout the text. An *alphabet* is a non-empty set of symbols and it is denoted by Σ . Finite sequences of the elements of Σ are called (linear) *words* over Σ . The set of all words over Σ is denoted by Σ^* . The *length* of the word $w \in \Sigma^*$ (denoted by $|w|$) is the number of all the symbols in w . There exists a unique word of length zero which is called the *empty word* (λ). The set of non-empty words over an alphabet Σ is $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$. A word $v \in \Sigma^*$ is a *factor* of $w \in \Sigma^*$ if there exist $x, y \in \Sigma^*$ such that $w = xvy$. Furthermore, if $x = \lambda$ (resp. $y = \lambda$), then v is a *prefix* (resp. *suffix*) of w . We say that a word w is *square free* if it does not have any factor xx , where x is a non-empty word. For a word w and integer $0 \leq k \leq |w|$, we denote the set of *length k factors* of w by $\mathcal{F}_k(w)$.

For arbitrary positive integers p and q , we use $(p \bmod q)$ to denote the remainder of $\frac{p}{q}$. Let $w = w_1 \dots w_n$ (for some $w_1, \dots, w_n \in \Sigma$) and $p \in \mathbb{N}$ (where \mathbb{N} denotes the set of non-negative integers). Then $w^{\frac{p}{n}} = w^{\lfloor \frac{p}{n} \rfloor} w'$, where $w' = w_1 \dots w_{(p \bmod n)}$. From now on we will always refer to the i th position of a word $w \in \Sigma^*$ as w_i .

A word $w \in \Sigma^+$ is *primitive* if there is no word $v \in \Sigma^*$ such that $w = v^p$ where p is a positive integer greater than one. Words x and y are *conjugates* if there exist words $u, v \in \Sigma^*$ such that $x = uv$ and $y = vu$. A positive integer p

with $p \leq n$ is a *period* of $w = w_1 \dots w_n$ if $w_i = w_{i+p}$ for all $i = 1, \dots, n - p$. A word $v \in \Sigma^*$ is a *border* of $w \in \Sigma^*$ if v is a prefix and also a suffix of w .

Remark 1. Note that a border of a word can overlap with itself. For example, the word *abababa* has borders *a*, *aba* and *ababa*.

There is a connection between borders and periods. A word w has a border b if and only if w has period $|w| - |b|$.

One of the central results in combinatorics on words is the theorem of Fine and Wilf (see [5]) generalized to words. Let $\gcd(p, q)$ denote the greatest common divisor of p and q .

Lemma 1 (Periodicity Theorem of Fine and Wilf). *Let w be an arbitrary word over some alphabet Σ . If w has two distinct periods p and q such that $p + q - \gcd(p, q) \leq |w|$, then $\gcd(p, q)$ is also a period of w .*

The lemma implies that, for long enough words, we can obtain new periods from the ones that we already have. We can also conclude that, if a non-unary word w has periods p, q such that $\gcd(p, q) = 1$, then $|w| < p + q - 1$.

Lyndon and Schützenberger stated the following, which characterizes the relation between a word and its non-trivial borders [9].

Lemma 2 (Lyndon and Schützenberger). *Let $x \in \Sigma^+$, $y, s \in \Sigma^*$ be arbitrary words. Then $xs = sy$ if and only if there exist $u \in \Sigma^+$, $v \in \Sigma^*$ and $k \geq 0$ such that $x = uv$, $y = vu$ and $s = (uv)^k u = u(vu)^k$.*

We continue with the definition of Christoffel words [1]. Let us have an ordered alphabet $\Sigma = \{a, b\}$ with $a < b$. Take two relatively prime positive integers, which we denote by p and q ; and let $n = p + q$. The *Christoffel word* w of slope $\frac{p}{q}$ over Σ is the word $w_1 \dots w_n$, where

$$w_i = \begin{cases} a & \text{if } i \cdot p \bmod n > (i-1) \cdot p \bmod n \\ b & \text{if } i \cdot p \bmod n < (i-1) \cdot p \bmod n \end{cases}$$

for all $i = 1, \dots, n$.

The geometrical interpretation of a Christoffel word of slope $\frac{p}{q}$ is as follows. Construct a path from $(0, 0)$ to (q, p) in the (horizontal and vertical) grid lines of the integer lattice $\mathbb{Z} \times \mathbb{Z}$ that satisfies the following two conditions:

1. the path lies below the line segment that begins at $(0, 0)$ and ends at (q, p) ,
2. there are no points of $\mathbb{Z} \times \mathbb{Z}$ between the path and the line segment.

From the 4-connected path $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ we construct the Christoffel word $w_1 w_2 \dots w_n$ as follows

$$w_i = \begin{cases} a & \text{if } x_i = x_{i-1} + 1 \text{ or } i = 1, \\ b & \text{if } y_i = y_{i-1} + 1, \end{cases}$$

for all $i = 1, 2, \dots, n$.

where $w^{(\ell)} = w_{1+\ell}, \dots, w_n w_1, \dots, w_{n-1}$. In the rest of the paper w_o will always denote the set of all conjugates of w , but we will usually call it a (circular) word.

Note that w_o consists exactly of the length $|w|$ factors of ww . That is, $w_o = \mathcal{F}_{|w|}(ww)$.

Definition 1. The positive integer p is a *weak* (resp. *strong*) *period* of a circular word w_o if p is a period of at least one (resp. all) $v \in w_o$.

3. Strong periods of circular words

Let us consider an arbitrary primitive word $w \in \Sigma^+$. Since w has at least one conjugate which is unbordered (see e.g., page 56 of [10]), the only strong period of w_o is $|w|$.

The following statement describes the relation of different strong periods of a circular word. It is a direct consequence of Lemma 1. Note that, if an integer p is a period of all conjugates of a word w , then it is also a period of ww .

Proposition 1. *Let $w \in \Sigma^+$ be an arbitrary non-empty word. The only strong periods of w_o are*

$$p, 2p, 3p, \dots, np,$$

where $n = \frac{|w|}{p}$ for some p that divides $|w|$.

Example 3. Consider the word $(ababab)_o$, which has strong periods 2, 4 and 6.

Remark 2. Proposition 1 is also related to periodic properties of infinite words. More precisely, it is a consequence of the fact that all periods of an infinite word are multiples of its smallest period.

From the previous statements we see that only some special types of circular words have non-trivial strong periods. Sequences like these (i.e., periodic infinite words) have been investigated in the literature. Some generalizations are mentioned in Section 7 that may still lead to interesting results about strong periods.

Now we continue with analyzing weak periods that turn out to be more interesting in our case, as we will see in the next section.

4. Weak periods of circular words

Recall that a positive integer p is a weak period of w_o if p is a period of at least one conjugate of w . First of all, we present a result that shows that the case of circular words is entirely different from the case of (linear) words.

Proposition 2. *The Periodicity Theorem of Fine and Wilf, i.e., Lemma 1 does not hold for weak periods of circular words.*

PROOF. The proof goes by example. Consider the word $aabab$ over the binary alphabet $\{a, b\}$. The following table shows the conjugates of this word with their periods.

$aabab$	$- 5$
$ababa$	$- 2, 4, 5$
$babaa$	$- 5$
$abaab$	$- 3, 5$
$baaba$	$- 3, 5$

We see that both 2 and 3 are weak periods of $(aabab)_o$ and $2 + 3 - 1 \leq 5$, but 1 is not a weak period of $(aabab)_o$. \square

In fact, for any $p > 2$, we can find infinitely many circular words w_o of length $2p - 1$ such that p and $p - 1$ are weak periods of w_o , but 1 is not a weak period of w_o . For example, let $w = ab^{p-1}ab^{p-2}$. Clearly, w has period p . Notice that $w^{(2)} = b^{p-2}ab^{p-2}ab$ has period $p - 1$. Thus w_o must have p and $p - 1$ as weak periods and it is not a unary word.

The following Lemma states a necessary and sufficient condition for a circular word $w_o \in \{a, b\}^*$ to have weak periods p and $p - 1$ for some $0 \leq p \leq |w|$.

Lemma 3. *Let $w \in \{a, b\}^*$. If $|w| \geq 2$, then there is an integer $p \in \mathbb{N}$ such that w_o has weak periods p and $p - 1$ if and only if w_o has a factor aa or bb .*

PROOF. \Rightarrow If aa is a factor of w_o , then there is a conjugate aua of w which has period $|w| - 1$. Moreover $|w|$ is a period of all conjugates of w , therefore both $|w| - 1$ and $|w|$ are weak periods of w_o .

\Leftarrow Assume that w_o has weak periods $p, p - 1$, but does not have aa or bb as factor. But in this case w_o must be equal to $((ab)^n)_o$. The only weak periods of this circular word are of the form $2m$ where $m \leq n$. In this case $|p - q| \geq 2$ (here $|p - q|$ denotes the absolute value of $p - q$) for any two distinct weak periods p, q . This is a contradiction. \square

Lemma 3 is not necessarily true if we consider alphabets of more than two symbols.

Example 4. Consider the word $w = abacbabc$ which does not have factors aa , bb , or cc . The following table represents the conjugates of w with their periods.

$abacbabc$	$- 8$	$bacbabac$	$- 5, 8$
$bacbacba$	$- 3, 6, 8$	$acbabcab$	$- 5, 8$
$acbabcab$	$- 8$	$cbabacba$	$- 5, 8$
$cbacabab$	$- 8$	$bababcac$	$- 8$

We see that both 5 and 6 are weak periods of w_o .

For weak periods that are small enough we can state the following Lemma, which is based on the connection of borders and periods.

Lemma 4. *Let $w \in \Sigma^*$ be an arbitrary word. If $p \leq \frac{|w|}{2}$ is a weak period of w_o , then $|w| - p$ is also a weak period of w_o .*

PROOF. Assume that $p \leq \frac{|w|}{2}$ is a period of $w^{(k)}$. Then $(w^{(k)})^{(p)}$ has a border of length p , from which we obtain the period $|w| - p$. Thus, $|w| - p$ is a weak period of w_o . \square

What if we would like to find all weak periods of a circular word w_o ? One way to do this is by taking every element $v \in w_o$ one by one and calculating each period of v . Since there is an equivalence relation on w_o , the periodic properties of an element in w_o have some effect on the periodic properties of the others. The following theorem specifies necessary and sufficient conditions that must be satisfied by a circular word to have some specific weak period.

Theorem 1. *Let w be a non-empty word over an arbitrary alphabet Σ . If p is a weak period of w_o , then there exists a word $x \in \Sigma^*$ with $|x| = |w| - p$ such that $x^{(\ell)}x$ is a factor of ww for some $\ell \geq 0$.*

Conversely, the following statements hold:

1. *If ww contains a factor xx for some $x \in \Sigma^+$ with $|x| < |w|$, then $p = |w| - |x|$ is a weak period of w_o .*
2. *If there exists an integer $p > 0$ such that the followings are satisfied:*
 - *$|w| = k \cdot p + r$ for some $k > 0$, $0 \leq r < p$ and*
 - *there exists a word $v \in \Sigma^*$ of length p , such that ww has factor*

$$(v^{k-1}v')^{(s \cdot |v|)}(v^{k-1}v')^{((s-1) \cdot |v|)}$$

for some $0 < s < k$, with v' a length r prefix of v ,

then p is a weak period of w_o .

PROOF. Clearly, if w has period p , then it has a border of length $|w| - p$, thus ww has a square factor xx of length $2 \cdot (|w| - p)$.

We can assume that p is a period of some $u \in w_o$. If $p > \lfloor \frac{|w|}{2} \rfloor$, then every cyclic shift of uu (including ww) must have a square factor xx of length $2 \cdot (|w| - p)$. Otherwise, if $p \leq \lfloor \frac{|w|}{2} \rfloor$, then $u = v^k v'$ for some $v, v' \in \Sigma^*$ and $k > 0$ integer, where $|v| = p$ and v' is a prefix of v .

We have to consider various cases depending on the amount of cyclic shift operations required to obtain ww from uu . Assume that $(uu)^{(j)} = ww$.

1. If $j \leq p$, then $(uu)^{(j)}$ has a square $xx = (v^{k-1}v')^2 = x^{(0)}x$.
2. If $p < j \leq kp$, then let $s = \lfloor \frac{j}{p} \rfloor$. Then there is a factor $v^{k-s-1}v'v^s v^{k-s}v'v^{s-1}$ in $(uu)^{(j)}$. Thus, x can be chosen to be $v^{k-s}v'v^{s-1}$. In this case $\ell = |v| = p$.
3. In the case of $kp \leq j \leq |u|$, $(uu)^{(j)} = zv^k v'v^k z'$, where $z'z = v'$. Here, $(v^{k-1}v')^2$ is clearly a factor of the word $(uu)^{(j)}$, thus $\ell = 0$ and $x = v^{k-1}v'$.

We continue with the proof of the converse statements. (1) Suppose that w is a nonempty word and ww has factor xx . If x is a border of w or xx is a factor of w , then the statement is trivial. Now assume that $ww = yxyy'$ such that $|y| < |y'|$. Then $w = yxx' = x''y'$ where $x'x'' = x$ and $|x'| = |y'| - |y|$. Notice that $w^{(-|x'|)} = x'yx = x'x''y''$ (with y'' a prefix of y') has x as a border. That is, $w^{(-|x'|)}$ has period $p = |w| - |x|$ which is therefore a weak period of w_o . The proof of the case $|y| > |y'|$ goes similarly.

(2) By the second case listed above in the proof, ww must have a factor $v^{k-s-1}v'v^sv^{k-s}v'v^{s-1}$, where v, k, s are as specified above. Notice that if there exists an integer $p > 0$ such that the conditions given in the theorem are satisfied, then

$$\begin{aligned} ww &= y(v^{k-1}v')^{(s \cdot |v|)}(v^{k-1}v')^{((s-1) \cdot |v|)}y' = \\ &= yv^{k-s-1}v'v^sv^{k-s}v'v^{s-1}y' = \\ &= yv^{k-s-1}v'v^kv'v^{s-1}y' \end{aligned}$$

for some $y, y' \in \Sigma^*$. Thus

$$(ww)^{(|y| + (k-s-1) \cdot p + r)} = v^kv'v^{s-1}y'yv^{k-s-1}v'$$

which has a length $|w|$ prefix $u \in w_o$ that has period p . Therefore, w_o has weak period p . \square

The converse statement requires further explanation. We can see that the factors $x^{(\ell)}x$ of ww depend on both p and ℓ , thus the existence of an arbitrary $x^{(\ell)}x$ factor of ww does not necessarily imply w_o having the weak period $|w| - |x|$. For example, take the word $w = abccb$. Then $ww = abccbabccb$ has a factor $bccb = cb^{(1)}cb$, but 3 is not a weak period of w_o . Part (1) of the converse statement covers the case when the circular word w_o has a weak period p with $p > \lfloor \frac{|w|}{2} \rfloor$. Surely, this does not cover all the converse cases. For example, take the circular word $(abbbaabaabbba)_o$ which has weak period 5, but there is no square of length 16 in $abbbaabaabbbaabbbaabaabbba$. Another example is the circular word $(bcababca)_o$ over the three letter alphabet which has weak period 3. Circular words like these satisfy the conditions described in (2).

Let us investigate now the relation of various weak periods of a circular word. In the case of linear words Lemma 1 gives a rather strong restriction on the length of words that can have two given, relatively prime periods. As we mentioned before, if a non-unary linear word w has periods p and q , such that $\gcd(p, q) = 1$, then $|w| < p + q - 1$.

The next theorem gives an upper bound on the length of a non-unary circular word with periods that are relatively prime.

Theorem 2. *Let $w_o \in \Sigma^*$ be a circular word with weak periods p and q (with $p > q \geq 2$). Furthermore, let $m = (p \bmod q)$. If $\gcd(p, q) = 1$ and w_o is not a unary word, then the length of w satisfies*

$$|w| \leq \begin{cases} p + q \cdot \lfloor \frac{p}{q} \rfloor & \text{if } m = 1, \\ p + q \cdot \lceil \frac{p}{q} \rceil & \text{if } m > 1. \end{cases}$$

PROOF. Consider, first, the case of $m = 1$. We construct a circular word of length $p + q \cdot \lfloor \frac{p}{q} \rfloor$ that has weak periods p and q . Let $v \in \Sigma^*$ be some non-unary word of length q . Now let $u = v^{\frac{p}{q}} = v^{\lfloor \frac{p}{q} \rfloor} v_1$. Clearly, $w = uv^{\lfloor \frac{p}{q} \rfloor}$ has period p and $w^{(p)} = v^{\lfloor \frac{p}{q} \rfloor} u = v^{\lfloor \frac{p}{q} \rfloor} v^{\lfloor \frac{p}{q} \rfloor} v_1$ has period q . Thus w_o has weak periods p and q .

Now we show that there are no non-unary circular words longer than w having weak periods both p and q . Let $z \in \Sigma^*$ be some word of length $|w| + r$ for some $r > 0$. We have to show that z is a unary word for any $r > 0$.

- In fact, if $r = 1$, then z is periodic with period p (since $|z| = 2p$) and so is $z^{(\ell)}$ for all $\ell = 1, \dots, |z|$. Thus there is a conjugate $z^{(k)}$ of z that has periods p and q . Then, by Lemma 1, $\gcd(p, q) = 1$ is also a period of $z^{(k)}$ thus $z^{(k)}$ is a unary word and so is z .
- Also, if $r \geq p$, then every conjugate of z has a factor $(z_1 \dots z_p z_1 \dots z_p)^{(k)}$ for some $k \in \mathbb{N}$. Since q is a weak period of z , it must be a period of some conjugate of z . Then, following the previous argument, z is unary by Lemma 1.
- Now, we can assume that $1 < r < p$. The word z can be written in the form

$$z = (z_1 \dots z_p)^{2 + \frac{r-1}{p}} = (z_1 \dots z_p)^2 z_1 \dots z_{r-1}.$$

We can see that $z^{(\ell)}$ has factor $(z_1 \dots z_p)^2$ for all $\ell < r$, thus cannot have period q unless z is unary. This leaves the shifts of the form

$$z_{p-(\ell-r)} \dots z_p z_1 \dots z_{r-1} z_1 \dots z_p z_1 \dots z_{p-\ell+r-1},$$

where $\ell \geq r$. Assume that one of these shifts has period q . Now, if $p - \ell + r \geq q$, then $z_1 \dots z_p z_1 \dots z_{p-\ell+r-1}$ and also z must have period 1 by Lemma 1. Thus $p - \ell + r < q$, that is, $p - q + r < \ell$. In this case, z_{p-q} must appear in $z_{p-(\ell-r)} \dots z_p$. Then either

$$z_{p-q+1} \dots z_p = z_1 \dots z_q,$$

or

$$z_{p-q+1} \dots z_p = z_1 \dots z_{r-1} z_1 \dots z_{q-r+1}.$$

Lemma 1 applies in the first case, because the word $z_1 \dots z_p z_1 \dots z_q$ has periods p and q . Thus it has period $\gcd(p, q) = 1$. In the second case $r - 1$ is also a period of $z_1 \dots z_q$. Thus, we can conclude by Lemma 1 that $z_1 \dots z_p z_1 \dots z_{r-1} z_1 \dots z_{q-r+1} = z_1 \dots z_p z_1 \dots z_q$ must be unary, because it has periods p and q .

In the case of $m > 1$, consider the non-unary word v of length q with a border of length $q - m$. We will denote this border with s . Now let $u = v^k v'$, such that $|u| = p$, and $v' = v_1 \dots v_m$. From u and v we can obtain $w = uv^{\lceil \frac{p}{q} \rceil}$. Notice that $v^{\lceil \frac{p}{q} \rceil} = v^{k+1}$, thus $w = v^k v' v^{k+1} = u u s$. Since s is a border of v , it is a prefix of u . From this, it follows that w has period $|u| = p$. If we shift w by

p positions to the left, then we get $w^{(p)} = v^{k+1}v^k v'$, which clearly has period q . Thus we have constructed a non-unary word w of length $p + q \cdot \lceil \frac{p}{q} \rceil$, such that w_o has weak periods p and q . Showing that all z_o of length $|z| = |w| + r$ (for some $r > 0$) with weak periods p and q can only be unary, can be done in a similar way to the case of $m = 1$. \square

Theorem 3. *For given $n \geq 3$ the maximal number of weak periods of non-unary circular words of length n over $\{a, b\}$ is*

$$\begin{aligned} \lfloor \frac{n}{2} \rfloor + \lfloor \frac{n}{4} \rfloor + 1 & \quad \text{if } n \text{ is odd,} \\ \frac{n}{2} + 1 & \quad \text{if } n \text{ is even and } (n \bmod 3) = 2, \\ \frac{n}{2} & \quad \text{otherwise.} \end{aligned}$$

PROOF. The set of weak periods of $w_o \in \Sigma^*$ ($|w| = n$) is a subset P_w of $\{1, \dots, n\}$. Since we are only considering non-unary words, we can make use of Theorem 2 to give some restrictions on the maximal number of distinct elements from the set. Also, it is convenient to investigate the elements of P_w in an increasing order.

Consider that the smallest element of P_w is k for some $k \in \{2, \dots, n\}$. Then $t \cdot k \in P_w$ for every $t = 1, \dots, \lfloor \frac{n}{k} \rfloor$. That makes exactly $\lfloor \frac{n}{k} \rfloor$ elements that are definitely in P_w . Note, that every element ℓ of P_w with $\ell \leq \lfloor \frac{n}{2} \rfloor$ must be a multiple of k , otherwise, by Theorem 2, w_o would be unary. Thus, the number of distinct elements of P_w is at most $\lfloor \frac{n}{2k} \rfloor + \lfloor \frac{n}{2} \rfloor + 1$. Choosing $k = 2$ maximizes the value. Thus, $|P_w| \leq \lfloor \frac{n}{4} \rfloor + \lfloor \frac{n}{2} \rfloor + 1$.

Note that if k divides n , then the weak periods given by Lemma 4 have already been counted among the multiples of k . So is n . Thus, the upper limit for $|P_w|$ in this case is $|P_w| \leq \lfloor \frac{n}{4k} \rfloor + \frac{n}{2} + 1 - \lfloor \frac{n}{4k} \rfloor - 1 = \lfloor \frac{n}{2} \rfloor$.

To further sharpen the limit, we consider the three cases separately:

1. If n is odd, then the limit $|P_w| \leq \lfloor \frac{n}{4} \rfloor + \lfloor \frac{n}{2} \rfloor + 1$ is sharp, since circular words of the form $((ab)^{\lfloor \frac{n}{2} \rfloor} a)_o$ have exactly $\lfloor \frac{n}{4} \rfloor + \lfloor \frac{n}{2} \rfloor + 1$ weak periods. That is, n , the multiples of two and the ones that can be derived from them by using Lemma 4.
2. The case when n is even and $(n \bmod 3) = 2$. If n is even, we notice, that $P_w = \{2, 4, \dots, n\}$ implies that w_o does not have any odd weak periods. Clearly, it is of no use to choose another even number for the smallest weak period. Thus, we have to investigate whether an odd k would lead to a higher number of weak periods. As we mentioned before, if k divides n , then $|P_w| \leq \lfloor \frac{n}{2} \rfloor$, therefore we only have to consider any odd k that does not divide n . Then, for every weak period $p > \lfloor \frac{n}{2} \rfloor$, either $p \in \{n-1, \dots, n - (n \bmod k) + 1\}$ or $p = n - t \cdot k$ for some $t \in \{1, \dots, \lfloor \frac{n}{k} \rfloor\}$. Thus, $|P_w| \leq \lfloor \frac{n}{2k} \rfloor + \lfloor \frac{n}{k} \rfloor + (n \bmod k) - 1 \leq \lfloor \frac{3n}{2k} \rfloor + (n \bmod k) - 1$. Choosing $k = 3$ maximizes this value, that is $|P_w| \leq \lfloor \frac{3n}{6} \rfloor + 1 = \frac{n}{2} + 1$. This limit is sharp, since circular words of the form $((aba)^{\lfloor \frac{n}{3} \rfloor} ab)_o$ have weak periods $n, n-1, t \cdot 3$ for $t = 1, \dots, \lfloor \frac{n}{3} \rfloor$, and $n - t \cdot 3$ for $t = 1, \dots, \lfloor \frac{n}{6} \rfloor$.

3. Similarly to case (2), we can either choose $k = 2$ or some other integer that does not divide n . Choosing $k = 2$, we get, that $P_w = \{2, 4, \dots, n\}$, while $|P_w| \leq \lfloor \frac{3n}{2k} \rfloor + (n \bmod k) - 1$ for all other values of k . Observe, that $\lfloor \frac{3n}{2k} \rfloor + (n \bmod k) - 1 \leq \frac{n}{2}$ for any $k > 2$. The smallest weak period of w_o in this case is 2 (or possibly, 3 if $(n \bmod 3) = 1$). Thus, $|P_w| \leq \frac{n}{2}$. The limit is clearly sharp, since the circular word $((ab)^{\frac{n}{2}})_o$ has just as many weak periods. \square

5. Representations of circular words

Every word w can be represented by its length and its generator u , i.e., the length p prefix of w , where p is the smallest period of w (see e.g., pages 10–11 of [16]). That is, $w = u^{\frac{|w|}{|u|}}$, where $|u| = p$, which can also be written as a pair $(u, |w|) \in \Sigma^* \times \mathbb{N}$. In this section we discuss analogous representations of circular words.

Definition 2. A pair $(u, n) \in \Sigma^* \times \mathbb{N}$ is a *representation* of the circular word w_o over Σ if $|u| \leq n$, $n = |w|$ and $(u^{\frac{n}{|u|}})_o = w_o$.

Definition 3. A *minimal representation* of a circular word w_o over Σ is a representation (u, n) of w_o , such that $|u| \leq |u'|$ for any other representation (u', n) of w_o .

We remark that in a minimal representation (u, n) the word u is primitive. Every circular word has a minimal representation, since all of them have a smallest weak period, but not all representations of circular words are minimal. For example, consider the representation $r_1 = (aaaba, 9)$ of the circular word $w_o = (aaabaaaab)_o$. This circular word has a representation $r_2 = (abaa, 9)$ too and $|abaa| < |aaaba|$, thus r_1 is not the minimal representation of w_o (in fact, it is r_2).

It is also true that a circular word can have more than one minimal representation. For example, $(ababa, 12)$, $(babaa, 12)$, $(abaab, 12)$ and $(baaba, 12)$ are all minimal representations of the circular word $(ababaababab)_o$. Note that $(aabab, 12)$ is not a minimal representation of this circular word, since it represents $(aababaababaa)_o$.

Clearly, if $n = k \cdot |u|$ for some $k \in \mathbb{N}$ in a minimal representation (u, n) , then (u^ℓ, n) is also a minimal representation of the same circular word for all $\ell = 0, \dots, |u| - 1$.

Suppose that $w = u^m u'$ for some $u \in \Sigma^*$ where u' is a non empty prefix of u and $m > 0$. Then the word $w' = wu^k$ has a cyclic shift $w'^{(k \cdot |u|)} = u^{m+k} u'$. Thus the circular word w'_o has a representation $(u, |w| + k \cdot |u|)$.

Theorem 4. Let (u, n) be a representation of w_o . Suppose that u has border b , that is, $u = bx = yb$, and $n = 2 \cdot |u| - |b|$. Then (y, n) is also a representation of w_o . Moreover, if b is the longest non-trivial border of u , then (y, n) is a minimal representation of w_o .

PROOF. Let us have a representation (u, n) of w_o that satisfies the assumption, that is, u has border b and $n = 2 \cdot |u| - |b|$. Then u is of the form $u = bx = yb$ for some $x, y \in \Sigma^*$ and $w_o = (uy)_o = (yby)_o$. By Lemma 2, yby has period $|y|$, thus w_o has weak period $|y|$ and a representation (y, n) .

If b is the longest non-trivial border of u , then y is the shortest root of u , thus (y, n) is a minimal representation of w_o . \square

6. Circular words formed from well known sequences

6.1. Circular Thue-Morse words

The morphism $\tau : \{a, b\}^* \rightarrow \{a, b\}^*$ defined as

$$\tau(a) = ab \quad \text{and} \quad \tau(b) = ba.$$

is called the *Thue-Morse morphism* and the word $T_n = \tau^n(a)$ (where $n \geq 0$) is the n th *Thue-Morse word*. [17, 18] Let us consider the complement homomorphism $\iota : \{a, b\}^* \rightarrow \{a, b\}^*$, defined by $\iota(a) = b$ and $\iota(b) = a$. For a shorter notation we denote $\iota(w)$ by \overline{w} .

The following two lemmas are well known results about Thue-Morse words. They can be found, for example, in [16].

Lemma 5. *For all $n \geq 0$, $T_{n+1} = T_n \overline{T_n}$.*

Lemma 6. *For all $n \geq 2$, the borders of T_n are*

$$T_{n-2}, T_{n-4}, \dots, T_{(n \bmod 2)}.$$

The previous properties of Thue-Morse words, lead us to the next result about minimal representations of circular Thue-Morse words.

Theorem 5. *Let $n \geq 3$ and $(u, 2^n)$ be a minimal representation of the circular word $(T_n)_o$. Then*

$$|u| = 2^n - (2^{n-2} + 2^{n-3}) = 5 \cdot 2^{n-3}.$$

PROOF. Recall from Theorem 1 that there is a strong relationship between subwords of ww and the weak periods of w_o . We will use this property in our proof. First we show that the two longest square factors of the word $T_n T_n$ are in fact the squares $(T_{n-2} T_{n-3})^2$ and $(\overline{T_{n-2} T_{n-3}})^2$. From the definition of Thue-Morse words we know that

$$\begin{aligned} (T_n T_n) &= (T_{n-1} \overline{T_{n-1}} = T_{n-2} \overline{T_{n-2} T_{n-2}} T_{n-2})^2 = \\ &= (T_{n-3} \overline{T_{n-3} T_{n-3}} \underbrace{T_{n-3} \overline{T_{n-3}}}_{T_{n-2}} T_{n-3} T_{n-3} \overline{T_{n-3}})^2. \end{aligned}$$

Notice that $T_n T_n$ contains the squares $(T_{n-2} T_{n-3})^2$ and $(\overline{T_{n-2} T_{n-3}})^2$ starting at positions $3 \cdot |T_{n-3}|$ and $7 \cdot |T_{n-3}|$, respectively. No other parts of $T_n T_n$ contain a longer square because of the complement homomorphism.

Now, by Theorem 1 we get that $(T_n)_o$ has a weak period of length $5 \cdot 2^{n-3}$. This weak period is minimal, since it is calculated from the longest square factors. Thus the length of u in a minimal representation $(u, 2^n)$ of $(T_n)_o$ is $2^n - (2^{n-2} + 2^{n-3})$. \square

For $n \geq 1$, the square-free Thue-Morse word (e.g., in [14]) S_n is the word $S_n = s_1 s_2 \dots s_{2^n-1}$, where for all $i = 1 \dots 2^n - 1$,

$$s_i = \begin{cases} a & \text{if } (T_n)_i (T_n)_{i+1} = ab \\ b & \text{if } (T_n)_i (T_n)_{i+1} = ba \\ c & \text{if } (T_n)_i = (T_n)_{i+1}, \end{cases}$$

where $(T_n)_i$ is the letter at position i in T_n .

Similarly to the case of the minimal representation of Thue-Morse circular words over the two letter alphabet we have the following theorem for square-free Thue-Morse words.

Corollary 1. *Let $n \geq 3$ and $(u, 2^n - 1)$ a minimal representation of the circular word $(S_n)_o$. Then*

$$|u| = 5 \cdot 2^{n-3}.$$

PROOF. This fact is a direct consequence of Theorem 5. \square

6.2. Circular Fibonacci words

The morphism $\varphi : \{a, b\}^* \rightarrow \{a, b\}^*$ defined as $\varphi(a) = ab$ and $\varphi(b) = a$ is called the *Fibonacci morphism* and the word $F_n = \varphi^n(b)$ (where $n \geq 0$) is the *nth Fibonacci word* [16].

The following three lemmas can be found, for example, in [16].

Lemma 7. *The Fibonacci words satisfy the recurrence*

$$F_0 = b, \quad F_1 = a, \quad F_n = F_{n-1} F_{n-2}.$$

Lemma 8. *For every $n \geq 3$, F_n has borders*

$$F_{n-2}, F_{n-4}, \dots, F_{2-(n \bmod 2)}.$$

Lemma 9. *A square u^2 appears in the infinite Fibonacci word (i.e., in $\varphi^\infty(b)$) if and only if u is a cyclic shift of some F_n ($n \geq 0$).*

Theorem 6. *Let $n \geq 4$ and $(u, |F_n|)$ a minimal representation of the circular word $(F_n)_o$. Then*

$$|u| = |F_{n-2}|.$$

Moreover, there is a minimal representation $(u, |F_n|)$ of F_n , such that $u = F_{n-2}$.

PROOF. Lemma 7 states that $F_n = F_{n-1}F_{n-2} = F_{n-2}F_{n-3}F_{n-2}$. It is also true that F_{n-3} is a prefix of F_{n-2} . Thus the shifted word $F_n^{(|F_{n-1}|)} = F_{n-2}F_{n-2}F_{n-3}$ has period $|F_{n-2}|$, which is then a weak period of $(F_n)_o$. Clearly, $(F_{n-2}, |F_n|)$ is a representation of $(F_n)_o$. Moreover, if we use Lemma 7 again, we get that $F_{n-2}F_{n-3}F_{n-2} = F_{n-3}F_{n-4}F_{n-3}F_{n-3}F_{n-4}$. It is clear that by iterating this method of rewriting, we can only find squares that are shorter than or equal to F_{n-2}^2 . \square

6.3. Circular Christoffel words

Example 5. Consider the Christoffel word w of slope $\frac{5}{4}$. The following table contains its conjugates with their periods.

$ababababb$	$- 9$	$ababbabab$	$- 5, 9$
$babababba$	$- 7, 9$	$babbababa$	$- 7, 9$
$abababbab$	$- 7, 9$	$abbababab$	$- 7, 9$
$bababbaba$	$- 5, 9$	$bbabababa$	$- 9$
$babababab - 2, 4, 6, 8, 9$			

Thus the circular Christoffel word w_o of slope $\frac{5}{4}$ has weak periods 2, 4, 5, 6, 7, 8, 9.

In the previous example one may notice that all non-trivial weak periods of w_o are either multiples of 2, or of the form $9 - k \cdot 2$ for some $1 \leq k \leq \lfloor \frac{9}{2 \cdot 2} \rfloor = 2$. For $k = 1$ we have $9 - 2 = 7$ and the slope of the dual of w is exactly $\frac{2}{7}$.

The following lemma is about the local periods of Christoffel words. [2]

Lemma 10. Let $w = aub$ be a Christoffel word of slope $\frac{p}{q}$. Then u has periods p^* and q^* , where $p^*p, q^*q \equiv 1 \pmod{p+q}$.

We can make a similar observation in the case of all circular Christoffel words as stated by the following theorem.

Theorem 7. Let w be the Christoffel word of slope $\frac{p}{q}$, $n = p + q$ and $m = \min\{p^*, q^*\}$, where $\frac{p^*}{q^*}$ is the slope of the dual of w . Then the set of all non-trivial weak periods of w_o is

$$\left\{ k \cdot m \mid 1 \leq k \leq \left\lfloor \frac{n}{m} \right\rfloor \right\} \cup \left\{ n - k \cdot m \mid 1 \leq k \leq \left\lfloor \frac{n}{2m} \right\rfloor \right\}.$$

PROOF. According to Theorem 1, we have to show that each square in ww has length $2k \cdot m$ for some $1 \leq k \leq \left\lfloor \frac{m}{n} \right\rfloor$. This is a direct consequence of Lemma 10. The existence of weak periods $n - k \cdot m$ follows from Lemma 4. \square

Thus each circular Christoffel word of slope $\frac{p}{q}$ has a representation $(u, p+q)$, with $|u| = \min\{p^*, q^*\}$.

7. Conclusions

We have given general definitions of weak and strong periods of circular words that lead to some interesting questions. The following problems arise:

1. Is there a generalization of the theorem of Fine and Wilf to weak periods of circular words? Theorem 2 covers the case of relatively prime weak periods.
2. Give a necessary and sufficient condition for a representation to be minimal.
3. It is clear that all weak periods can be calculated in $O(n^2)$ time by calculating the border array for all n shifts of a word. Is there a more efficient algorithm for finding all weak periods of a circular word?

8. Acknowledgments

Some parts of these results were presented at the WORDS 2013 conference in Turku, Finland [7]. The authors would like to thank the members of the audience for their valuable and useful comments. The work is supported by the TÁMOP 4.2.2/C-11/1/KONV-2012-0001 and 4.2.2/B-10/1-2010-0024 projects. The projects are implemented through the New Hungary Development Plan, co-financed by the European Social Fund and the European Regional Development Fund.

References

- [1] Berstel, J., Lauve, A., Reutenauer, C., Saliola, F. V.: *Combinatorics on Words: Christoffel Words and Repetitions in Words*. CRM Monograph Series, vol. 27., 2008
- [2] Berthé, V., de Luca, A. and Reutenauer, C.: *On an involution of Christoffel words and Sturmian morphisms*. European Journal of Combinatorics, vol. 29(2), 2008, pp. 535–553.
- [3] Currie, J. D. and Fitzpatrick, D. S.: *Circular words avoiding patterns*. In Proceedings of the 6th International Conference on Developments in Language Theory. LNCS 2450. Springer-Verlag, 2002, pp. 319–325.
- [4] Diekert, V., Harju, T. and Nowotka, D.: *Factorizations of cyclic words*. In Workshop on Words and Automata at CSR, 7. 2006.
- [5] Fine, N. J. and Wilf, H. S.: *Uniqueness theorems for periodic functions*. In Proceedings of the American Mathematical Society, vol. 16., 1965, pp. 109–114.
- [6] Fitzpatrick, D. S.: *There are binary cube-free circular words of length n contained within the Thue-Morse word for all positive integers n* . In Ars Combinatorica, vol. 74., 2005.

- [7] Hegedüs, L. and Nagy, B.: *Periodicity of circular words*. In Local Proceedings of WORDS 2013, TUCS Lecture Notes No. 20, 2013, pp. 45–56.
- [8] Lothaire, M.: *Combinatorics on words*. Addison-Wesley, 1983.
- [9] Lyndon, R. C. and Schützenberger, M. P.: *The equation $a^M = b^N c^P$ in a free group*. In Michigan Math. J. vol 9, issue 4 (1962), pp. 289–298.
- [10] Nowotka, D.: *Periodicity and unbordered factors of words*. TUCS Dissertations No. 50, 2004.
- [11] Păun, G., Rozenberg, G. and Salomaa, A.: *DNA computing: New computing paradigms*. Springer-Verlag, 1998.
- [12] Rittaud, B. and Vivier, L.: *Circular words and applications*. In Proceedings of Words 2011, Electronic Proceedings in Theoretical Computer Science, vol. 63., 2011, pp. 31–36.
- [13] Rittaud, B. and Vivier, L.: *Circular words and three applications: factors of the Fibonacci word, \mathcal{F} -adic numbers, and the sequence 1, 5, 16, 45, 121, 320, ...* In Funct. Approx. Comment. Math. vol. 47(2)., 2012, pp. 207–231.
- [14] Shallit, J.: *A Second Course in Formal Languages and Automata Theory* Cambridge University Press, 2008
- [15] Shur, A. M.: *On ternary square-free circular words*. In The Electronic Journal of Combinatorics, **17**, 2010.
- [16] Smyth, W.: *Computing patterns in strings*. Addison-Wesley, 2003.
- [17] Thue, A.: *Über unendliche Zeichenreihen*. In Kra. Vidensk. Selsk. Skrifter, I. Mat. Nat. Kl. vol. 7., 1906, pp. 1–22.
- [18] Thue, A.: *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*. In Kra. Vidensk. Selsk. Skrifter, I. Mat. Nat. Kl. vol. 46., 1912, pp. 1–67.
- [19] van Lint, J. H.: *Introduction to coding theory*. Graduate Texts in Mathematics 86, Springer Verlag, 1998.