AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **83(3)** (2022), Pages 435–447

# On the palindromic zl-factorization and c-factorization of the generalized period-doubling sequences

Moussa Barro  Idrissa Kaboré

*Département de Mathématiques, UFR-SEA*
*Université Nazi BONI*
*Bobo-Dioulasso, Burkina Faso*
mous.barro@yahoo.com  ikaborei@yahoo.fr

K. Ernest Bognini

*Centre Universitaire de Kaya (CU-Kaya)*
*Université Joseph KI-ZERBO*
*Ouagadougou, Burkina Faso*
ernestk.bognini@yahoo.fr

**Abstract**

In this paper, we study period-doubling sequences over an ordered alphabet of size $q \geq 2$. We present properties of these words relative to the structure of their palindromic factors. The explicit formulas of the palindromic Ziv-Lempel factorization and the palindromic Crochemore factorization based on the combinatorial structure of infinite sequences are also established.

## 1 Introduction

In combinatorics on words, the study of palindromes occupies an important place. For instance, this notion is used to characterize the Sturmian words (see [11]). A palindrome is a word which is the same when read from left to right or from right to left. The factorization of a finite or infinite sequence consists of the decomposition of this word (finite or infinite) into factors with specific properties: periodicity, palindromicity, etc. It thus appears in combinatorics on words as an important tool in the understanding of structures of words. There are several types of factorization in the literature (see [5, 6, 13, 14]) including Lyndon factorization, Ziv-Lempel factorization and Crochemore factorization. The Ziv-Lempel factorization was introduced in the middle of the 20th Century (see [17]) and the Crochemore factorization at the end of the 20th Century (see [8, 9]). These two factorizations provide, respectively,

ISSN: 2202-3518 

a better comprehension of repetitive and non-repetitive factors in the infinite words. They find applications in data compression [18], word processing [8], molecular biology [7, 12], cryptography [16], etc. Since their introduction, several authors have established these factorizations for some infinite words (see [4, 5, 6, 13, 14]). It is in this context that a study of the structure of palindromes, the Ziv-Lempel factorization and the Crochemore factorization of the generalized period-doubling word in [6] was conducted. Thereafter, other variants of the Ziv-Lempel factorization and the Crochemore factorization were introduced in [14], namely the palindromic Ziv-Lempel factorization and the palindromic Crochemore factorization. In the same paper the authors established these factorizations for the $m$-bonacci words which are generalizations of the Fibonacci word.

Here, we obtain the palindromic Ziv-Lempel factorization and the palindromic Crochemore factorization of the generalized period-doubling sequences. This paper is organized as follows. After definitions and notation, we recall some useful results in Section 2. Next, we present properties on some factors in Section 3 and then we establish in Section 4 the palindromic Ziv-Lempel factorization of the generalized period-doubling sequences. Finally, in Section 5 we give the palindromic Crochemore factorization of these sequences.

## 2 Preliminaries

Let $\mathcal{A}$ be a finite alphabet. The set of finite words over $\mathcal{A}$ is denoted by $\mathcal{A}^*$ and $\varepsilon$ represents the empty word. The set of non-empty finite words (respectively, infinite words) over $\mathcal{A}$ is denoted by $\mathcal{A}^+$ (respectively, $\mathcal{A}^\omega$). The set of finite and infinite words over $\mathcal{A}$ is denoted by $\mathcal{A}^\infty$. Let $u \in \mathcal{A}^\infty$ and $v \in \mathcal{A}^*$. The word $v$ is called a factor of $u$ if there exist $u_1 \in \mathcal{A}^*$, $u_2 \in \mathcal{A}^\infty$ such that $u = u_1 v u_2$. The factor $v$ is called a prefix (respectively, suffix) if $u_1$ (respectively, $u_2$) is empty. Let $u = a_1 a_2 \cdots a_n$ be a finite word with $a_i \in \mathcal{A}$, for $i = 1, 2, \ldots, n$. The word $\overline{u} = a_n a_{n-1} \cdots a_1$ is called the reflection of $u$. The word $u$ is called a palindrome if $u = \overline{u}$. For all $u \in \mathcal{A}^*$, $|u|$ denotes the length of $u$. For $a \in \mathcal{A}$, we denote by $a^{-1}$ the inverse of $a$, that is: $aa^{-1} = a^{-1}a = \varepsilon$. If $u$ is a finite word over $\mathcal{A}$ beginning (respectively, ending) with the letter $a$ then $a^{-1}u$ (respectively, $ua^{-1}$) denotes the word obtained from $u$ by deleting its first (respectively, last) letter.

A morphism over $\mathcal{A}^*$ is a map $f : \mathcal{A}^* \longrightarrow \mathcal{A}^*$ such that $f(uv) = f(u)f(v)$ for all $u, v \in \mathcal{A}^*$. It is $k$-uniform (respectively, non-erasing), for some $k \in \mathbb{N}$, if $|f(a)| = k$ (respectively, $f(a) \neq \varepsilon$), for any $a \in \mathcal{A}$. The map $f$ is said to be a substitution if it is a non-erasing morphism over $\mathcal{A}^*$. It is said to be prolongable on $a$, if there exists $u \in \mathcal{A}^+$ such that $f(a) = au$. In this case, $f^n(a)$ is a proper prefix of $f^{n+1}(a)$, for any positive integer $n$. The sequence $(f^n(a))_{n\geq 0}$ converges to a unique infinite word denoted $f^\omega(a)$ and so is called a purely morphic word.

It is said that the set $X \subset \mathcal{A}^+$ is a code over $\mathcal{A}$ if any word $w \in \mathcal{A}^*$ admits at most one factorization into words from $X$. For more details on coding theory, we refer the reader to [1, 3, 15]. Now, we give two formal definitions of the palindromic

Ziv-Lempel factorization and the palindromic Crochemore factorization, and then we recall an important result about codes [15].

Let $\mathbf{u}$ be an infinite word. Then we have the following definitions introduced in [14].

**Definition 2.1.** The palindromic Ziv-Lempel factorization or simply **pzl**-factorization of $\mathbf{u}$ is the factorization:

$$\mathbf{pzl}(\mathbf{u}) = \prod_{k\geq 0} \tilde{z}_k = (\tilde{z}_0,\ \tilde{z}_1,\ \ldots,\ \tilde{z}_n,\ \tilde{z}_{n+1}, \ldots), \tag{1}$$

where $\tilde{z}_n$ is the shortest palindromic prefix of $\tilde{z}_n\tilde{z}_{n+1}\tilde{z}_{n+2}\cdots$ such that there is no occurrence of $\tilde{z}_n$ at any position $j < |\tilde{z}_0\tilde{z}_1\cdots\tilde{z}_{n-1}|$ in $\mathbf{u}$.

For all integers $n$, $\tilde{z}_n$ are called **pzl**-factors.

**Example 2.1.** The **pzl**-factorization of $v = abcaacbaaabc$ is:

$$\mathbf{pzl}(v) = (a, b, c, aa, cbaaabc).$$

**Definition 2.2.** The palindromic Crochemore factorization or simply **pc**-factorization of $\mathbf{u}$ is the factorization:

$$\mathbf{pc}(\mathbf{u}) = \prod_{k\geq 0} \tilde{c}_k = (\tilde{c}_0,\ \tilde{c}_1, \ldots,\ \tilde{c}_m,\ \tilde{c}_{m+1}, \ldots), \tag{2}$$

where $\tilde{c}_m$ is the longest palindromic prefix of $\tilde{c}_m\tilde{c}_{m+1}\tilde{c}_{m+2}\cdots$ occurring at least twice in $\tilde{c}_0\tilde{c}_1\cdots\tilde{c}_m$, or $\tilde{c}_m$ is just a letter if this letter does not appear in $\tilde{c}_0\tilde{c}_1\cdots\tilde{c}_{m-1}$.

**Example 2.2.** The **pc**-factorization of $w = abacabaacabaaca$ is:

$$\mathbf{pc}(w) = (a, b, a, c, aba, aca, b, aa, c, a).$$

**Proposition 2.1** (Ch. 6 of [15]). *Let $\mathcal{A}$, $\mathcal{B}$ be two finite alphabets and $f : \mathcal{A}^* \longrightarrow \mathcal{B}^*$ an injective morphism. Then we have:*

1. *If $X \subseteq \mathcal{A}^+$ is a code then $f(X)$ is a code.*
2. *If $Y \subseteq \mathcal{B}^+$ is a code then $f^{-1}(Y)$ is a code.*

The binary period-doubling sequence $\mathbf{P}_2$ is the unique fixed point of the substitution $S_2 : a \longmapsto ab,\ b \longmapsto aa$ defined over $\mathcal{A}_2 = \{a, b\}$ and beginning with $a$. Thus, we have $\mathbf{P}_2 = \lim_{n\to\infty} S_2^n(a) = S_2^\omega(a)$, whose first letters are given by:

$$\mathbf{P}_2 = abaaabababaaabaaabaaababababaaabababab\cdots .$$

This sequence has its origin in chaotic dynamics (see [2]). The name period-doubling of this sequence comes from the fact that its fundamental block is doubled in each step. It has been intensively studied in [1, 2, 4, 5, 10].

Now, let us consider the alphabet $\mathcal{A}_q = \{0, 1, 2, \ldots, q-1\}$, for a fixed integer $q \geq 3$ and the 2-uniform substitution over $\mathcal{A}_q$ defined by:

$$S_q(m) = \begin{cases} 0(m+1) & \text{if } 0 \leq m \leq q-2 \\ 00 & \text{if } m = q-1. \end{cases} \tag{3}$$

A natural generalization of binary period-doubling sequence is the unique fixed point of the substitution $S_q$ defined in (3). Let us note $w_n = S_q^n(0)$, for all $n \geq 0$. Then, $\mathbf{P}_q = \lim_{n\to\infty} w_n = S_q^\omega(0)$. For instance, if $q = 5$ then the first letters of $\mathbf{P}_5$ are given by:

$$\mathbf{P}_5 = 0102010301020104010201030102010001020103\cdots.$$

## 3 Some properties of generalized period-doubling sequences

**Proposition 3.1.** *The set $\mathbb{B}_q = S_q(\mathcal{A}_q)$ is a code over $\mathcal{A}_q$, for any integer $q \geq 2$.*

**Proof.** Since $\mathcal{A}_q$ is a code over $\mathcal{A}_q$ and $S_q$ is an injective substitution then, by Proposition 2.1, $\mathbb{B}_q$ is a code. □

**Lemma 3.1.** *Let $u$ and $v$ be two finite factors of $\mathbf{P}_q$ such that $S_q(v)$ is a factor of $S_q(u)$. Then, $v$ is a factor of $u$.*

**Proof.** If $u$ or $v$ is empty then the result can easily be checked. Suppose that $u$ and $v$ are two non-empty factors of $\mathbf{P}_q$ such that $S_q(v)$ is a factor of $S_q(u)$. Then, there exist two finite words $r$ and $t$ such that $S_q(u) = rS_q(v)t$. We continue the reasoning over the length of $r$. If $|r|$ is odd, then $S_q(v)$ can only be a power of 0 and must be preceded by 0. It follows that $v = (q-1)^k$ for some integer $k$ and that $u$ contains also $(q-1)^k$. From now on, assume that $|r|$ is even. Since $S_q$ is 2-uniform then $r = S_q(r')$ and $t = S_q(t')$ for some words $r'$ and $t'$. So, $S_q(u) = S_q(r'vt')$. Since $S_q$ is injective then, $u = r'vt'$. Thus, $v$ is a factor of $u$. □

**Theorem 3.1.** *[6] Let $v$ be a factor of $\mathbf{P}_q$ such that $|v| > 2$. Then the following assertions are equivalent.*

1. *$v$ is a palindromic factor of $\mathbf{P}_q$.*
2. *$S_q(v)0$ is a palindromic factor of $\mathbf{P}_q$.*
3. *$0^{-1}S_q(v)$ is a palindromic factor of $\mathbf{P}_q$.*

Note that 0, 00 and 000 are factors of $\mathbf{P}_q$ but not $0^k$ for any $k > 3$.

**Theorem 3.2.** *Let $v$ be a non-empty palindromic factor of $\mathbf{P}_q$ such that $v \neq 00$. Then we have:*

1. *If $v$ begins with an odd power of the letter 0, then there exists a palindromic factor $v'$ of $\mathbf{P}_q$ such that $v = S_q(v')0$.*

2. *If $v$ begins with an even power of the letter* 0 *or does not begin with the letter* 0*, then there exists a palindromic factor $v'$ of $\mathbf{P}_q$ such that $v = 0^{-1}S_q(v')$.*

**Proof.** Since $v$ is a factor of $\mathbf{P}_q$, there exist a finite word $u$ and an infinite word $\mathbf{w}$ such that $\mathbf{P}_q = uv\mathbf{w}$. As $\mathbf{P}_q$ is a fixed point of $S_q$ then $\mathbf{P}_q = S_q(uv\mathbf{w})$, with $|S_q(uv)| > |uv|$. We continue the proof by induction on $|v|$.
If $|v| = 1$ then the two properties hold. Indeed, $v$ is equal to either $S_q(\varepsilon)0$ or $0^{-1}S_q(x)$, with $x \in \mathcal{A}_q$; $\varepsilon$ and $x$ being palindromes. Suppose that $|v| > 2$.

1. Suppose that $v$ begins with an odd power of the letter 0. Then, either $v = 000$ or $v$ is of the form $v = 000v_1000$ or $v = 0yv_1y0$, with $y \in (\mathcal{A}_q - \{0\}) \cup \{\varepsilon\}$ and $v_1$ a non-empty palindromic factor of $\mathbf{P}_q$ such that $|v_1| < |v|$, with $v_1 \neq 00$.
**Case 1.** $v = 000$. So, it is sufficient to take $v' = q-1$ and we have $v = S_q(v')0$.
**Case 2.** $v = 000v_1000$. Since, $0v_10$ is a palindromic factor beginning with 0 and $|0v_10| < |v|$ then, by induction hypothesis, we have $0v_10 = S_q(v_1')0$ with $v_1'$ being a palindromic factor of $\mathbf{P}_q$. As a result, we obtain the following equalities:

$$\begin{aligned} v &= 000v_1000 \\ &= 00S_q(v_1')000 \\ &= S_q(q-1)S_q(v_1')S_q(q-1)0 \\ &= S_q((q-1)v_1'(q-1))0 \\ &= S_q(v')0, \text{ with } v' = (q-1)v_1'(q-1). \end{aligned}$$

Let us now show that $v'$ is a factor of $\mathbf{P}_q$. Since $v = S_q(v')0$ is a factor of $uv$ then, it is also a factor of $S_q(uv)$ for some word $u$. Thus, $v'$ is a factor of $uv$, by Lemma 3.1. Hence, $v'$ is a factor of $\mathbf{P}_q$.
**Case 3.** $v$ is in the form $v = 0yv_1y0$. Then by hypothesis, there exists a palindromic factor $v_1'$ of $\mathbf{P}_q$ such that $yv_1y = 0^{-1}S_q(v_1')$. Thus,

$$v = 00^{-1}S_q(v_1')0 = S_q(v_1')0.$$

2. Suppose that $v$ begins with an even power of the letter 0 or with the letter $x \neq 0$. Then, either $v = 00v_100$ or $v = xv_1x$, with $v_1$ a non-empty palindromic factor of $\mathbf{P}_q$.
**Case 1.** $v = 00v_100$. Then, by induction hypothesis there exists a palindromic factor $v_1'$ of $\mathbf{P}_q$ such that $0v_10 = S_q(v_1')0$. Thus,

$$\begin{aligned} v &= 00v_100 \\ &= 0S_q(v_1')00 \\ &= 0^{-1}S_q((q-1)v_1'(q-1)) \\ &= 0^{-1}S_q(v'), \text{ with } v' = (q-1)v_1'(q-1). \end{aligned}$$

**Case 2.** $v = xv_1x$. Note that the word $\mathbf{P}_q$ does not contain the factor $x^2$ for $x \neq 0$. Thus, $v_1$ begins with 0, since $S_q$ is 2-uniform and prolongable in 0 by all letters of $\mathcal{A}_q$. By induction hypothesis, there exists a palindromic factor $v'_1$ of $\mathbf{P}_q$ such that $v_1 = S_q(v'_1)0$. We then obtain the following equalities.

$$\begin{aligned} v &= xv_1x \\ &= xS_q(v'_1)0x \\ &= 0^{-1}0xS_q(v'_1)0x \\ &= 0^{-1}S_q(x-1)S_q(v'_1)S_q(x-1) \\ &= 0^{-1}S_q((x-1)v'_1(x-1)) \\ &= 0^{-1}S_q(v'), \text{ with } v' = (x-1)v'_1(x-1). \end{aligned}$$

We show with the same reasoning of item 1 that $v'$ is a factor of $\mathbf{P}_q$. □

Note that the only even-length palindromic factor in $\mathbf{P}_q$ is the word 00.

## 4 Palindromic zl-factorization of the sequences $\mathbf{P}_q$

In this section we present the palindromic Ziv-Lempel factorization of $\mathbf{P}_2$ (Theorem 4.1) and then of the sequences $\mathbf{P}_q$ (Theorem 4.2).

### 4.1 Palindromic zl-factorization of the sequence $\mathbf{P}_2$

We construct a sequence of finite words $(z_n)_{n\geq 0}$ over $\mathcal{A}_2$ as follows:

$$z_0 = a, \ z_1 = b, \ z_2 = aa \text{ and for all } n \geq 3,$$

$$z_n = \begin{cases} a^{-1}S_2(z_{n-1}) & \text{if } n \text{ is even} \\ S_2(z_{n-1})a & \text{otherwise.} \end{cases} \tag{4}$$

**Proposition 4.1.** *For all integers $n$, we have:*

1. *The word $z_n$ is a palindromic factor of $\mathbf{P}_2$.*
2. *The word $z_n$ is not a factor of $z_{n+1}$.*

**Proof.**

1. The proof stems from equality (4) and Theorem 3.1, in particular in the case of $q = 2$ (see [5]).

2. The second assertion is demonstrated by induction on the integer $n$. The property is checked at the initial index. Indeed, the equality (4) ensures that $z_0 = a$ is not a factor of $z_1 = b$. Suppose that $z_k$ is not a factor of $z_{k+1}$, for $k \leq n-1$, with $n \geq 1$. Let us show that $z_n$ is not a factor of $z_{n+1}$. For the sake of contradiction, let us assume that $z_n$ is a factor of $z_{n+1}$. Then there exist two non-empty finite words $u_1$ and $u_2$ such that $z_{n+1} = u_1z_nu_2$. According to the parity of the integer $n$, we have:

- If $n$ is even then, by (4), we have that $z_n$ does not begin with $a$ and $z_{n+1}$ ends with $a$. Thus $z_{n+1} = u'_1 a z_n u'_2 a$, for some non-empty words $u'_1$ and $u'_2$. Furthermore, $S_2(z_n) = z_{n+1}a^{-1} = u'_1 S_2(z_{n-1}) u'_2$, i.e, $S_2(z_{n-1})$ is a factor of $S_2(z_n)$. Hence $z_{n-1}$ is a factor of $z_n$, by Lemma 3.1. This contradicts the induction hypothesis.
- If $n$ is odd then $z_n = S_2(z_{n-1})a$ and $z_{n+1} = a^{-1}S_2(z_n)$. By using similar reasoning to the previous case, we show that $z_n$ is not a factor of $z_{n+1}$. Thus $z_n$ is not a factor of $z_{n+1}$ for all integers $n$.

□

**Lemma 4.1.** *For all integers $n \geq 1$, $z_n$ is not a factor of $\xi_n = \prod_{k=0}^{n-1} z_k$.*

**Proof.** Before proceeding to the demonstration of this lemma, we make the following remark by equality (4). For all integers $n \geq 3$ we have:

$$\xi_n = \begin{cases} S_2(\xi_{n-1})a & \text{if } n \text{ is even} \\ S_2(\xi_{n-1}) & \text{otherwise.} \end{cases} \tag{5}$$

For $n \in \{1, 2\}$, we have $z_1 = b$ (respectively, $z_2 = aa$) is not a factor of $\xi_1 = a$ (respectively, of $\xi_2 = ab$). Hence the property holds for $n = 1$ and $n = 2$.

Suppose that $z_i$ is not a factor of $\xi_i$ for $i \leq n$, with $n \geq 3$. Let us show that the property remains true for $i = n + 1$. We prove this by contradiction. Suppose that $z_{n+1}$ is a factor of $\xi_{n+1}$; then there exist two non-empty finite words $v_1$ and $v_2$ such that $\xi_{n+1} = v_1 z_{n+1} v_2$. We distinguish two cases according to the parity of the integer $n$.

- If $n$ is even then, $z_{n+1} = S_2(z_n)a$ and $\xi_{n+1} = S_2(\xi_n)$. As a result, $S_2(\xi_n) = v_1 S_2(z_n) a v_2$, i.e., $S_2(z_n)$ is a factor of $S_2(\xi_n)$. Thus $z_n$ is a factor of $\xi_n$ by Lemma 3.1. This contradicts the induction hypothesis.

- If $n$ is odd then $z_{n+1} = a^{-1}S_2(z_n)$ and $\xi_{n+1} = S_2(\xi_n)a$. Note that in this case $z_{n+1}$ does not begin with $a$ by Theorem 3.2 and $\xi_{n+1}$ ends with $a$ by (5). Thus we can write $\xi_{n+1} = S_2(\xi_n)a = v_1 z_{n+1} v_2 = v'_1 a z_{n+1} v'_2 a = v'_1 S_2(z_n) v'_2 a$, for some non-empty words $v'_1$ and $v'_2$. Therefore $S_2(\xi_n)a = v'_1 S_2(z_n) v'_2 a$. Hence $S_2(z_n)$ is a factor of $S_2(\xi_n)$. By Lemma 3.1, $z_n$ is a factor of $\xi_n$. We again obtain a contradiction with the induction hypothesis.

□

**Theorem 4.1.** *The* **pzl**-*factorization of the sequence* $\mathbf{P}_2$ *is:*

$$\mathbf{pzl}(\mathbf{P}_2) = \prod_{k \geq 0} z_k.$$

**Proof.** The proof of this theorem stems from Proposition 4.1, Lemma 4.1 and the fact that the sequence $(z_n)_{n \geq 0}$ is increasing in the sense of the length. □

**Example 4.1.** The first factors of the **pzl**-factorization of $\mathbf{P}_2$ are:

$$\mathbf{pzl}(\mathbf{P}_2) = (a, b, aa, ababa, baaabaaab, \cdots).$$

### 4.2 Palindromic zl-factorization of the sequences $\mathbf{P}_q$

Let us now consider, over $\mathcal{A}_q$ with $q \geq 3$, the sequence of finite words $(Z_n)_{n\geq 0}$ defined by:

$$Z_0 = 0, \text{ and for all } n \geq 1,$$

$$Z_n = \begin{cases} S_q(Z_{n-1})0 & \text{if } n \text{ is even} \\ 0^{-1}S_q(Z_{n-1}) & \text{otherwise.} \end{cases} \tag{6}$$

**Proposition 4.2.** *For all integers $n$, $Z_n$ is a palindromic factor of $\mathbf{P}_q$.*

**Proof.** We have $Z_0 = 0$, $Z_1 = 1$ and $Z_2 = 020$ which are palindromic factors of $\mathbf{P}_q$. Suppose that the property is true up to the index $n-1$ for $n \geq 3$ and let us show that it remains true for the index $n$. By hypothesis, $Z_k$ is a palindromic factor of $\mathbf{P}_q$ for $k \leq n-1$. We distinguish two cases according to the parity of the integer $k$.

**Case 1.** $k$ is even. Then, we have $Z_k = S_q(Z_{k-1})0$. By induction hypothesis, $Z_{k-1}$ is a palindromic factor of $\mathbf{P}_q$ and $|Z_{k-1}| > 2$. So $Z_k$ is a palindromic factor of $\mathbf{P}_q$, by Theorem 3.1.

**Case 2.** $k$ is odd. Then, we have $Z_k = 0^{-1}S_q(Z_{k-1})$. Since $|Z_{k-1}| > 2$ for $k \in \{3, 4, \ldots, n-1\}$ and $Z_{k-1}$ is a palindromic factor of $\mathbf{P}_q$ by hypothesis, it follows that the word $0^{-1}S_q(Z_{k-1}) = Z_k$ is also a palindromic factor of $\mathbf{P}_q$ by Theorem 3.1. □

**Lemma 4.2.** *For all integers $n$, $Z_n$ is not a factor of $Z_{n+1}$.*

**Proof.** For $n = 0$, $Z_0 = 0$ is not a factor of $Z_1 = 1$. Suppose that $Z_k$ is not a factor of $Z_{k+1}$, for $k < n$ and let us show that $Z_n$ is not a factor of $Z_{n+1}$. Let us assume for the sake of contradiction that $Z_n$ is a factor of $Z_{n+1}$, then there exist two finite non-empty words $u$ and $v$ such that $Z_{n+1} = uZ_nv$.

**Case 1.** $n$ is even. Then, by equality (6), we have:

$$Z_n = S_q(Z_{n-1})0 \text{ and } Z_{n+1} = 0^{-1}S_q(Z_n).$$

Moreover, $Z_{n+1}$ does not begin with the letter 0 by Theorem 3.2. By Proposition 4.2, we can write $Z_{n+1} = xwZ_n\overline{w}x$ where $x$ is different to the letter 0 and $w$ a non-empty word. It follows that, $S_q(Z_n) = 0Z_{n+1} = 0xwZ_n\overline{w}x = 0xwS_q(Z_{n-1})0\overline{w}x$, i.e, $S_q(Z_{n-1})$ is a factor of $S_q(Z_n)$. We deduce by Lemma 3.1 that $Z_{n-1}$ is a factor of $Z_n$. This contradicts the induction hypothesis.

**Case 2.** $n$ is odd. Then, by (6), we have :

$$Z_n = 0^{-1}S_q(Z_{n-1}) \text{ and } Z_{n+1} = S_q(Z_n)0.$$

By Theorem 3.2, $Z_{n+1}$ begins and ends with 0. Thus, $S_q(Z_n)0 = Z_{n+1} = 0wZ_n\overline{w}0 = 0w'0Z_n0\overline{w'}0 = 0w'S_q(Z_{n-1})0\overline{w'}0$, for some non-empty words $w$ and $w'$. Therefore, $S_q(Z_{n-1})$ is a factor of $S_q(Z_n)$. Hence, $Z_{n-1}$ is a factor of $Z_n$ by Lemma 3.1. This still contradicts the induction hypothesis.

Hence $Z_n$ is not a factor of $Z_{n+1}$ for all integers $n$. □

For all $n \geq 1$, let us put:

$$\mathrm{H}_n = \prod_{k=0}^{n-1} Z_k.$$

The remark below gives us recursive writing of $\mathrm{H}_n$.

**Remark 4.1.** *For all integers $n \geq 2$, we have:*

$$\mathrm{H}_n = \begin{cases} S_q(\mathrm{H}_{n-1}) & \text{if } n \text{ is even} \\ S_q(\mathrm{H}_{n-1})0 & \text{otherwise.} \end{cases}$$

**Lemma 4.3.** *For all integers $n$, $Z_n$ is not a factor of* $\mathrm{H}_n$.

**Proof.** The property holds for the initial index. Suppose that $Z_j$ is not a factor of $\mathrm{H}_j$ for $j \leq n-1$ and let us show that $Z_n$ is not a factor of $\mathrm{H}_n$. For the sake of contradiction, let us assume that $Z_n$ is a factor of $\mathrm{H}_n$. Then, $\mathrm{H}_n = vZ_nw$ for some non-empty words $v, w$. We separate the proof into two cases according to the parity of the integer $n$.

**Case 1.** $n$ is even. Then, $Z_n = S_q(Z_{n-1})0$. Furthermore, we have $S_q(\mathrm{H}_{n-1}) = \mathrm{H}_n = vS_q(Z_{n-1})0w$, by Remark 4.1 and assumption. Thus $Z_{n-1}$ is a factor of $\mathrm{H}_{n-1}$, by Lemma 3.1. This contradicts the induction hypothesis.

**Case 2.** $n$ is odd. Then $Z_n = 0^{-1}S_q(Z_{n-1})$. In addition, by Remark 4.1, we have $\mathrm{H}_n = S_q(\mathrm{H}_{n-1})0$. Since $Z_n$ does not begin with the letter 0 by Theorem 3.2 and $\mathrm{H}_n$ ends with 0, we have $S_q(\mathrm{H}_{n-1})0 = \mathrm{H}_n = vZ_nw = v'0Z_nw'0 = v'S_q(Z_{n-1})w'0$, for some non-empty words $v'$ and $w'$. Thus $S_q(Z_{n-1})$ is a factor of $S_q(\mathrm{H}_{n-1})$. By Lemma 3.1, we deduce that $Z_{n-1}$ is a factor of $\mathrm{H}_{n-1}$. This contradicts the induction hypothesis. □

**Theorem 4.2.** *The* **pzl***-factorization of the sequences* $\mathbf{P}_q$ *is given by:*

$$\mathbf{pzl}(\mathbf{P}_q) = \prod_{k \geq 0} Z_k.$$

**Proof.** We first show that the sequence $(\mathrm{H}_n)_{n\geq 0}$ is increasing in length, and then that it constitutes increasingly long prefixes of $\mathbf{P}_q$. Finally, we show that the $Z_n$ which compose it are the **pzl**-factors of the sequences $\mathbf{P}_q$.

- Since $Z_n \neq \varepsilon$, we have $|\mathrm{H}_{n+1}| > |\mathrm{H}_n|$ for all $n \geq 0$.
- Let us show by induction that $(\mathrm{H}_n)_{n\geq 1}$ is a sequence of prefixes of $\mathbf{P}_q$.

For $1 \leq n \leq 2$, we have respectively $\mathrm{H}_1 = 0 = S_q^0(0)$ and $\mathrm{H}_2 = Z_0Z_1 = 01 = S_q(0)$ which are prefixes of $\mathbf{P}_q$. Suppose that $\mathrm{H}_{n-1}$ is a prefix of $\mathbf{P}_q$, for $n \geq 1$. Then there exists $\mathbf{w}$, an infinite word, such that $\mathbf{P}_q = \mathrm{H}_{n-1}\mathbf{w}$. According to the parity the integer $n$, we have two cases:

**Case 1.** $n$ is even. Then $\mathrm{H}_n = S_q(\mathrm{H}_{n-1})$, by Remark 4.1. Since $\mathbf{P}_q$ is a fixed point of $S_q$, we get:

$$\begin{aligned} \mathbf{P}_q &= S_q(\mathrm{H}_{n-1}\mathbf{w}) \\ &= S_q(\mathrm{H}_{n-1})S_q(\mathbf{w}) \\ &= \mathrm{H}_nS_q(\mathbf{w}). \end{aligned}$$

**Case 2.** $n$ is odd. Then we have $\mathrm{H}_n = S_q(\mathrm{H}_{n-1})0$ by Remark 4.1. Thereafter,

$$\begin{aligned}\mathbf{P}_q &= S_q(\mathrm{H}_{n-1}\mathbf{w})\\ &= S_q(\mathrm{H}_{n-1})S_q(\mathbf{w})\\ &= S_q(\mathrm{H}_{n-1})0\mathbf{w'},\ \text{because } S_q(\mathbf{w}) = 0\mathbf{w'} \text{ for some infinite word } \mathbf{w'}\\ &= \mathrm{H}_n\mathbf{w'}.\end{aligned}$$

Thus, for all integers $n$, $\mathrm{H}_n$ is a prefix of $\mathbf{P}_q$.

• The sequence of finite words $(Z_n)_{n\geq 0}$ represents the **pzl**-factors of $\mathbf{P}_q$. Indeed, $Z_n$ is a palindromic factor of $\mathbf{P}_q$ according to Proposition 4.2. Moreover, $Z_n$ is neither a factor of $Z_{n+1}$ nor a factor of $\mathrm{H}_n$, by Lemmas 4.2 and 4.3. In addition, the sequence $(Z_n)_{n\geq 0}$ is increasing in the sense of length. Hence $Z_n$ is the shortest palindromic prefix of $Z_nZ_{n+1}\cdots$ uni-occurrent in $\mathrm{H}_{n+1} = Z_0Z_1\cdots Z_n$.

Since $(\mathrm{H}_n)_{n\geq 1}$ is a sequence of prefixes of $\mathbf{P}_q$ with increasing length,

$$\mathbf{pzl}(\mathbf{P}_q) = \lim_{n\to\infty} \mathrm{H}_n = \prod_{k\geq 0} Z_k.$$

□

**Example 4.2.** For $q = 5$, the first factors of the **pzl**-factorization of $\mathbf{P}_5$ are given by:

$$\mathbf{pzl}(\mathbf{P}_5) = (0, 1, 020, 10301, 02010401020, \cdots).$$

## 5 Palindromic c-factorization of the sequences $\mathbf{P}_q$

We begin this section by constructing the sequence $(p_n)_{n\geq 0}$ of finite words over $\mathcal{A}_q$ as follows:

$$p_0 = \varepsilon \text{ and for all } n \geq 0,\ p_{n+1} = p_n\vartheta_n p_n \text{ with } \vartheta_n = n \bmod q.$$

The sequence $(p_n)_{n\geq 1}$ constitutes the sequence of palindromic prefixes of $\mathbf{P}_q$.

**Remark 5.1.** *For all integers $n$, $p_{n+1} = S_q(p_n)0$.*

**Theorem 5.1.** *For all integers $q \geq 2$, the* **pc**-*factorization of* $\mathbf{P}_q$ *is given by:*

$$\mathbf{pc}(\mathbf{P}_q) = \prod_{k\geq 0} C_k,\ \textit{with the sequence } (C_k)_k \textit{ defined by:}$$

• *for $q = 2$, we have:*

$C_0 = a,\ C_1 = b,\ C_2 = a,\ C_3 = aa,\ C_4 = b,\ C_5 = ababa,\ C_6 = aa,$ *and for all* $k \geq 7$,

$$C_k = \begin{cases} S_2(C_{k-2})a & \textit{if } k \textit{ is even}\\ a^{-1}S_2(C_{k-2}) & \textit{otherwise;}\end{cases}$$

• *for $q \geq 3$, we have:*

$$C_0 = 0 \text{ and for all } k \in \{1, 2, 3, \ldots, q-1\},$$

$$\begin{cases} C_{2k} = p_k, \\ C_{2k-1} = k. \end{cases} \tag{7}$$

$$C_{2q-1} = 00 \text{ and for all } k \geq 2q,\ C_k = \begin{cases} 0^{-1}S_q(C_{k-2}) & \text{if } k \text{ is even} \\ S_q(C_{k-2})0 & \text{otherwise.} \end{cases} \tag{8}$$

For the proof of this theorem, we state the following lemmas.

**Lemma 5.1.** *For all integers $n$, $C_n$ is a palindromic factor of $\mathbf{P}_q$.*

**Proof.** We do the proof in case $q \geq 3$, the case $q = 2$ being similar.
**Case 1.** $n \in \{0, 1, 2, 3, \ldots, q-1\}$, $C_n$ is a palindromic factor of $\mathbf{P}_q$. Indeed, $C_n$ is either a letter or a palindromic prefix $p_n$.
**Case 2.** $n = 2q - 1$. So, $C_{2q-1} = 00$ is a palindromic factor of $\mathbf{P}_q$.
**Case 3.** $n \geq 2q$. Then, we continue the demonstration by induction on the integer $n$. The property holds for the initial index. Indeed, $C_{2q} = 0^{-1}S_q(C_{2q-2}) = 0^{-1}S_q(p_{q-1})$ which is a palindromic factor of $\mathbf{P}_q$, by Theorem 3.1. Suppose that $C_n$ is a palindromic factor of $\mathbf{P}_q$ for $n \geq 2q$ and let us show that it is the same for $C_{n+1}$. We distinguish two cases according to the parity of the integer $n$.

- If $n$ is even then, by (8), we have $C_{n+1} = S_q(C_{n-1})0$. By hypothesis, $C_{n-1}$ is a palindromic factor of $\mathbf{P}_q$ with $|C_{n-1}| > 2$. Thus $C_{n+1}$ is a palindromic factor of $\mathbf{P}_q$, according to Theorem 3.1.
- If $n$ is odd, then we have $C_{n+1} = 0^{-1}S_q(C_{n-1})$, by the equality (8). By a similar reasoning to the even case, we show that $C_{n+1}$ is a palindromic factor of $\mathbf{P}_q$.

□

For all $n \geq 1$, let us put:

$$\Omega_n = \prod_{k=0}^{n-1} C_k.$$

**Remark 5.2.** *For all integers $n \geq 2q$, we have:*

$$\Omega_n = \begin{cases} S_q(\Omega_{n-2})0 & \text{if } n \text{ is even} \\ S_q(\Omega_{n-2}) & \text{otherwise.} \end{cases}$$

**Lemma 5.2.** *For all integers $n \geq 2q$, $\Omega_n$ contains at least two occurrences of $C_{n-1}$.*

**Proof.** Let us proceed by disjunction of cases according to the parity of the integer $n$.
**Case 1.** For $n = 2q$, the property holds. Suppose that $\Omega_{2n}$ has at least two occurrences of $C_{2n-1}$, for $n \geq 2q$. Let us show that $\Omega_{2n+2}$ has at least two occurrences of $C_{2n+1}$. By the induction hypothesis, there exist some non-empty finite words $v_1$, $v_2$,

and $v$ possibily empty, such that $\Omega_{2n} = v_1 C_{2n-1} v C_{2n-1} v_2$. By Remark 5.2, we get the following equalities:

$$\begin{aligned}\Omega_{2n+2} &= S_q(\Omega_{2n})0 \\ &= S_q(v_1 C_{2n-1} v C_{2n-1} v_2)0 \\ &= S_q(v_1) S_q(C_{2n-1}) S_q(v) S_q(C_{2n-1}) S_q(v_2)0 \\ &= S_q(v_1) S_q(C_{2n-1})0v' S_q(C_{2n-1})0v_2'0, \text{ because } S_q \text{ is prolongable in } 0 \\ &= S_q(v_1) C_{2n+1} v' C_{2n+1} v_2'0, \text{ by equality (8).}\end{aligned}$$

**Case 2.** Suppose that $\Omega_{2n+1}$ has at least two occurrences of $C_{2n}$, for $n \geq 2q+1$. Let us then show that $\Omega_{2n+3}$ has at least two occurrences of $C_{2n+2}$. By Remark 5.2 and with a similar reasoning to the previous case, we obtain the result. □

**Proof of Theorem 5.1.** By Remark 5.2 and the equalities (7), (8), we deduce that the words $\Omega_n$ are increasingly long prefixes of $\mathbf{P}_q$. The Lemmas 5.1 and 5.2 assure us that for all $n \geq 0$, $C_n$ is a palindromic factor and at least bi-occurring in $\Omega_{n+1} = C_0 C_1 \cdots C_{n-1} C_n$. As a result, it represents a palindromic Crochemore factor of $\mathbf{P}_q$. Thus

$$\mathbf{pc}(\mathbf{P}_q) = \lim_{n\to\infty} \Omega_n = \prod_{k\geq 0} C_k.$$

□

**Example 5.1.** For $q = 5$, the first factors of the **pc**-factorization of $\mathbf{P}_5$ are given by:

$$\mathbf{pc}(\mathbf{P}_5) = (0, 1, 0, 2, 010, 3, 0102010, 4, 010201030102010, 00, \cdots).$$

## Acknowledgments

The authors thank the anonymous referees for careful reading and their helpful comments which improved the quality of this paper. The work reported in this paper is supported by Simons Foundation via PREMA.

## References

[1] J. P. Allouche and J. Shallit, Automatic sequences, Theory, Application, Generalization, Cambridge University Press, 2003.

[2] M. Baake and U. Grimm, Aperiodic Order, Vol. 1: A Mathematical Invitation, Cambridge University Press, 2013.

[3] J. Berstel and D. Perrin, Theory of Codes, Academic Press, 1985.

[4] J. Berstel and A. Savelli, Crochemore factorisation of Sturmian and other infinite words, In: Int. Symp. Math. Foundations of Comp. Sci. (2006), pp. 157–166; Springer, Berlin, Heidelberg.

[5] K. E. Bognini, I. Kaboré and T. Tapsoba, Some combinatorial properties and Lyndon factorization of the period-doubling word, *Far East J. Math. Sci. (FJMS)* **125**(2) (2020), 87–103.

[6] K. E. Bognini, I. Kaboré and B. Kientéga, Ziv-Lempel and Crochemore factorizations of the generalized period-doubling word, *Adv. Appl. Discrete Math.* **25**(2) (2020), 143–159.

[7] X. Chen, S. Kwong and M. Li, A compression algorithm for DNA sequences, *IEEE Eng. Med. Biol.* **20** (2001), 61–66.

[8] M. Crochemore and W. Rytter, Texts algorithms, Oxford University Press, New York, 1994.

[9] M. Crochemore, Recherche linéaire d'un carré dans un mot , *C. R. Math. Acad. Sci. Paris* **296** (1983), 781–784.

[10] D. Damanik, Local symmetries in the period-doubling sequence, *Discrete Appl. Math.* **100** (2000), 115–121.

[11] X. Droubay and G. Pirillo, Palindromes and Sturmian words, *Theoret. Comp. Sci.* **223** (1999), 73–85.

[12] M. Farach, M. O. Noordewier, S. A. Savari, L. A. Shepp, A. D. Wyner and J. Ziv, On the entropy of DNA: Algorithms and measurements based on memory and rapid convergence, *SODA* (1995), 121–124.

[13] N. Ghareghami, M. Mohammad-Noori and P. Sharifani, On z-factorization and c-factorization of standard epistrumian words, *Theoret. Comp. Sci.* **412** (2011), 5232–5238.

[14] M. Jahannia, M. Mohammad-Noori, N. Rampersad and M. Stipulante, Palindromic Ziv-Lempel and Crochemore factorizations of $m$-bonacci infinite words, *Theoret. Comp. Sci.* **790** (2019), 16–40.

[15] M. Lothaire, Algebraic combinatorics on words, (Ch. 6), Codes, Cambridge University Press, 2001.

[16] S. Mund, Ziv-Lempel complexity for periodic sequences and its cryptography application, *Lec. Notes in Comp. Sci.* **166** (1984), 230–240.

[17] J. Ziv and A. Lempel, A universal algorithm for sequential data compression, *IEEE Trans. Inf. Theor.* **23**(3) (1977), 337–343.

[18] J. Ziv and A. Lempel, Compression of individual sequences via variable-rate coding, *IEEE Trans. Inform. Theory* **24** (1978), 530–536.

(Received 30 Dec 2021; revised 23 May 2022)