Cross-bifix-free sets via Motzkin paths generation

Elena Barucci*   Stefano Bilotta*   Elisa Pergola*
Renzo Pinzani*   Jonathan Succi*

Abstract

Cross-bifix-free sets are sets of words such that no prefix of any word is a suffix of any other word. In this paper, we introduce a general constructive method for the sets of cross-bifix-free $q$-ary words of fixed length. It enables us to determine a cross-bifix-free words subset which has the property to be non-expandable.

1 Introduction

A cross-bifix-free set of words (also called cross-bifix-free code) is a set where, given any two words over an alphabet, possibly the same, any prefix of the first one is not a suffix of the second one and vice-versa. Cross-bifix-free sets are involved in the study of frame synchronization which is an essential requirement in a digital communication systems to establish and maintain a connection between a transmitter and a receiver.

Analytical approaches to the synchronization acquisition process and methods for the construction of sequences with the best aperiodic autocorrelation properties [1, 2, 3, 4] have been the subject of numerous analyses in the digital transmission.

The historical engineering approach started with the introduction of bifix, a name proposed by J. L. Massey as acknowledged in [5]. It denotes a subsequence that is both a prefix and suffix of a longer observed sequence.

In [4] the notion of a distributed sequences is introduced, where the synchronization word is not a contiguous sequence of symbols but is instead interleaved into the data stream. In [6] is showed that the distributed sequence entails a simultaneous search for a set of synchronization words. Each word in the set of sequences is required to be bifix-free. In addition, they arises a new requirement that no prefix of any length of any word in the set is a suffix of any other word in the set. This property of the set of synchronization words was termed as cross-bifix-free.

The problem of determining such sets is also related to several other scientific applications, for instance in pattern matching [7] and automata theory [8].

---

*Dipartimento di Matematica e Informatica ‘U.Dini’, Università degli Studi di Firenze, Viale G.B. Morgagni 65, 50134 Firenze, Italy (e-mail: elena.barucci@unifi.it; stefano.bilotta@unifi.it; elisa.pergola@unifi.it; renzo.pinzani@unifi.it; jons10@hotmail.it).
Several methods for constructing cross-bifix-free sets have been recently proposed as in [9, 10, 11]. In particular, once the cardinality $q$ of the alphabet and the length $n$ of the words are fixed, a matter is the construction of a cross-bifix-free set with the cardinality as large as possible. An interesting method has been proposed in [9] for words on a binary alphabet. This specific construction reveals interesting connections to the Fibonacci sequence of numbers. In a recent paper [11] the authors revisit the construction in [9] and generalize it obtaining cross-bifix-free sets having greater cardinality over an alphabet of any size $q$. They also show that their cross-bifix-free sets have a cardinality close to the maximum possible. To our knowledge this is the best result in the literature about the greatest size of cross-bifix-free sets.

For the sake of completeness we note that an intermediate step between the original method [9] and its generalization [11] has been proposed in [10] and it is constituted by a different construction of binary cross-bifix-free sets based on lattice paths which allows to obtain greater values of cardinality if compared to the ones in [9].

In this study, we revisit the construction in [10]. We give a new construction of cross-bifix-free sets that generalizes the construction of [10] in order to extend the construction to $q$-ary alphabets for any $q$, $q > 2$. This approach enables us to obtain cross-bifix-free sets having greater cardinality than the ones presented in [11], for the initial values of $n$. This new result extends the theory of cross-bifix-free sets and it could be used to improve some technical applications.

This paper is organized as follows. In Section 2 we give some preliminaries and describe the adopted notation. In Section 3 we present a new construction of cross-bifix-free sets in the $q$-ary alphabet and in Section 4 we analyze the sizes of the sets of our construction in comparison to the ones in the literature.

## 2 Basic definitions and notations

Let $\mathbb{Z}_q = \{0, 1, \cdots, q - 1\}$ be an alphabet of $q$ elements. A (finite) sequence of elements in $\mathbb{Z}_q$ is called (finite) word. The set of all words over $\mathbb{Z}_q$ having length $n$ is denoted by $\mathbb{Z}_q^n$. A consecutive sequence of $m$ element $a \in \mathbb{Z}_q$ is denoted by the short form $a^m$. Let $w \in \mathbb{Z}_q^n$, then $|w|_a$ denotes the number of occurrences of $a$ in $w$, being $a \in \mathbb{Z}_q$. Let $w = uzv$, then $u$ is called a prefix of $w$ and $v$ is called a suffix of $w$. A bifix of $w$ is a subsequence of $w$ that is both its prefix and suffix.

A word $w \in \mathbb{Z}_q^n$ is said to be bifix-free or unbordered [12] if and only if no prefix of $w$ is also a suffix of $w$. Therefore, $w$ is bifix-free if and only if $w \neq uzu$, being $u$ any necessarily non-empty word and $z$ any word. Obviously, a necessary condition for $w$ to be bifix-free is that the first and the last letters of $w$ must be different.

**Example 2.1** In $\mathbb{Z}_2 = \{0, 1\}$, the word $11010100$ of length $n = 9$ is bifix-free, while the word $101001010$ contains two bifixes, 10 and 1010.
Let $BF_q(n)$ denote the set of all bifix-free words of length $n$ over an alphabet of fixed size $q$ (for more details about this topic see [12]).

Given $q > 1$ and $n > 1$, two distinct words $w, w' \in BF_q(n)$ are said to be cross-bifix-free if and only if no strict prefix of $w$ is also a suffix of $w'$ and vice-versa.

**Example 2.2** The binary words 111010100 and 110101010 in $BF_2(9)$ are cross-bifix-free, while the binary words 111001100 and 110011010 in $BF_2(9)$ have the cross-bifix 1100.

A subset of $BF_q(n)$ is said to be a cross-bifix-free set if and only if for each $w, w'$, with $w \neq w'$, in this set, $w$ and $w'$ are cross-bifix-free. This set is said to be non-expandable on $BF_q(n)$ if and only if the set obtained by adding any other word in $BF_q(n)$ is not a cross-bifix-free set. A non-expandable cross-bifix-free set on $BF_q(n)$ having maximal cardinality is called a maximal cross-bifix-free set on $BF_q(n)$.

In a recent paper [11] the authors provide a general construction of cross-bifix-free sets over a $q$-ary alphabet. Below, we recall such generation for the family of cross-bifix-free sets in $\mathbb{Z}_n^q$.

For any $2 \leq k \leq n - 2$, the cross-bifix-free set $S_{k,q}(n)$ in [11] is the set of all words $s = s_1s_2\cdots s_n$ in $\mathbb{Z}_n^q$ that satisfy the following two properties:

1) $s_1 = \cdots = s_k = 0$, $s_{k+1} \neq 0$ and $s_n \neq 0$,
2) the subsequence $s_{k+2}\ldots s_{n-1}$ does not contain $k$ consecutive 0’s.

Let $$F_{k,q}(n) = \begin{cases} q^n & \text{if } 0 \leq n < k, \\ (q - 1) \sum_{l=1}^{k} F_{k,q}(n-l) & \text{if } n \geq k, \end{cases}$$
be the sequence enumerating the words in $\mathbb{Z}_n^q$ avoiding $k$ consecutive zero’s [13]. Then, from the above definition of $S_{k,q}(n)$, we have

$$|S_{n,q}^{(k)}| = (q - 1)^2 F_{k,q}(n - k - 2).$$

For any fixed $n$ and $q$, the largest size of $|S_{n,q}^{(k)}|$ is denoted by $S(n, q)$ and it is given by the following expression as in [11]

$$S(n, q) = \max\{(q - 1)^2 F_{k,q}(n - k - 2) : 2 \leq k \leq n - 2\}.$$

This result allows to obtain non-expandable cross-bifix-free sets in the $q$-ary alphabet having cardinality close to the maximum.

In the present paper we introduce an alternative constructive method for the generation of cross-bifix-free set in $\mathbb{Z}_q$. Our approach is based on the study of lattice path in the discrete plane and it moves from the construction in [10].

Each word $w \in \mathbb{Z}_n^q$ can be represented as a lattice path of $\mathbb{N}^2$ running from $(0, 0)$ to $(n, 0)$ having the following properties:
- the element 0 corresponds to a fall step which is defined by $(1, -1)$,
- the element 1 corresponds to a rise step which is defined by $(1, 1)$,
- the elements $2, \ldots, q - 1$ correspond respectively to a colored level step which is defined by $(1, 0)$ and it is labeled by one of the $q - 2$ fixed colors.

For example, in Table 1 and Table 2 showed an equivalence between elements and steps of lattice paths in the alphabets $\mathbb{Z}_3$ and $\mathbb{Z}_4$, respectively.

**Table 1:** Equivalence between symbols and steps for $\mathbb{Z}_3 = \{0, 1, 2\}$.

| Symbol | Step  | Color | Representation |
|--------|-------|-------|----------------|
| 0      | $(1, -1)$ | -     |               |
| 1      | $(1, 1)$  | -     |               |
| 2      | $(1, 0)$  | Black |               |

**Table 2:** Equivalence between symbols and steps for $\mathbb{Z}_4 = \{0, 1, 2, 3\}$.

| Symbol | Step  | Color | Representation |
|--------|-------|-------|----------------|
| 0      | $(1, -1)$ | -     |               |
| 1      | $(1, 1)$  | -     |               |
| 2      | $(1, 0)$  | Red   |               |
| 3      | $(1, 0)$  | Green |               |

From now on, we will refer interchangeably to words or their graphical representations on the discrete plane, that is paths. The definition of bifix-free and cross-bifix-free can be easily extended to paths.

A $k$-colored Motzkin path of length $n$ is a lattice path of $\mathbb{N}^2$ running from $(0, 0)$ to $(n, 0)$ that never goes below the $x$-axis and whose admitted steps are rise steps, fall steps and $k$-colored level steps (for more details about this copy see [14]).

For example, the left side of Fig. 1 shows a Motzkin path in $\mathbb{Z}_3$ having length 6, while the path in its right side is not a Motzkin path since it crosses the $x$-axis.

We denote by $\mathcal{M}_k(n)$ the set of all $k$-colored Motzkin paths of length $n$, and let $M_k(n)$ be the size of $\mathcal{M}_k(n)$.
Figure 1: Words 121002, 100212 and the equivalent paths. The first one is a Motzkin word.

\begin{center}
\begin{tikzpicture}
\draw (0,0) -- (0,1) -- (1,0) -- (2,0) -- (3,0);
\draw (0,0) -- (0,1) -- (1,0) -- (2,0) -- (1,1);
\end{tikzpicture}
\end{center}

Proposition 2.1 For any \( n \geq 0 \) and \( k \geq 3 \), \( M_k(n) \) is given by the following expression

\[ M_k(n+1) = kM_k(n) + \sum_{i=0}^{n-1} M_k(i)M_k(n-1-i) \]

with \( M_k(0) = 1 \) and \( M_k(1) = k \).

Proof. If \( n = 0 \), \( M_k(n) \) contains the empty path only, then \( M_k(0) = 1 \). If \( n = 1 \), \( M_k(n) \) only contains those paths obtained by a level step, thus \( M_k(1) = k \).

Let \( n \geq 1 \) and \( w \in M_k(n+1) \). There are two cases: \( w \) begins with a level step or \( w \) begins with a rise step. In the first case we have that \( w = h\alpha \) where \( h \) is a level step and \( \alpha \in M_k(n) \), then the number of this first kind of paths is equal to \( kM_k(n) \).

Otherwise, we have that \( w = uad\beta \) where \( u \) is a rise step, \( d \) is a fall step, \( \alpha \in M_k(n) \) and \( \beta \in M(n-1-i) \) with \( 0 \leq i \leq n-1 \). Then the number of this latter kind of paths is equal to \( \sum_{i=0}^{n-1} M_k(i)M_k(n-1-i) \).

Thus,

\[ M_k(n+1) = kM_k(n) + \sum_{i=0}^{n-1} M_k(i)M_k(n-1-i) \].

\[ \blacksquare \]

A word \( w \in \mathbb{Z}_q^n \) is called \((q-2)\)-colored Motzkin word if the equivalent lattice path is a \((q-2)\)-colored Motzkin path.

For our purposes, it is useful to denote by \( \hat{M}_{q-2}(n) \) the set of all elevated \((q-2)\)-colored Motzkin words of length \( n \), defined as

\[ \hat{M}_{q-2}(n) = \{ 1\alpha 0 : \alpha \in M_{q-2}(n-2) \} \].

For example, in Fig. 2 two words in \( \hat{M}_1(6) \) are depicted.

In the next section of the present paper we are interested in determining one among all the possible non-expandable cross-bifix-free sets of words of fixed length \( n > 1 \) on \( \mathbb{Z}_q^n \). We denote this set by \( \text{CBFS}_q(n) \).
Figure 2: An example of elevated Motzkin words

\[1 \ 2 \ 1 \ 2 \ 0 \ 0\]
\[1 \ 2 \ 2 \ 2 \ 2 \ 0\]

Figure 3: Graphical representation of the set \(A_q(n), \ n \geq 3\)

\[\alpha \in \mathcal{M}_{q-2}(i) \quad \beta \in \hat{\mathcal{M}}_{q-2}(n-i)\]

3 On the non-expandability of \(\text{CBFS}_q(n)\)

In this section we define the set \(\text{CBFS}_q(n)\) which is formed by the union of three sets of \((q-2)\)-colored Motzkin paths denoted by \(A_q(n), B_q(n)\) and \(C_q(n)\), with \(q \geq 3\) and \(n \geq 3\), respectively.

Let

\[A_q(n) = \left\{ \alpha \beta : \alpha \in \mathcal{M}_{q-2}(i), \beta \in \hat{\mathcal{M}}_{q-2}(n-i) \right\} \setminus \left\{ \alpha \beta : \alpha, \beta \in \hat{\mathcal{M}}_{q-2} \left( \frac{n}{2} \right) \right\}\]

with \(0 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor\), be the set of words composed by a \((q-2)\)-colored Motzkin word \(\alpha\) of length \(i\), and a elevated \((q-2)\)-colored Motzkin word \(\beta\) of length \(n-i\) (see Fig. 3). If \(n\) is even, we need to remove the words composed by two elevated subwords of the same length. On the other side, if \(n\) is odd, we assume the set \(\left\{ \alpha \beta : \alpha, \beta \in \hat{\mathcal{M}}_{q-2} \left( \frac{n}{2} \right) \right\}\) empty, since it does not exists any path of non-integer length.

Then, the enumeration of the set \(A_q(n)\) is given by the following expression

\[|A_q(n)| = \sum_{i=0}^{\left\lfloor n/2 \right\rfloor} M_{q-2}(i) M_{q-2}(n-i-2) - \left[ M_{q-2} \left( \frac{n}{2} - 2 \right) \right]^2.\]

Let

\[B_q(n) = \left\{ 1\alpha \beta : \alpha \in \mathcal{M}_{q-2}(i), \beta \in \hat{\mathcal{M}}_{q-2}(n-i-1) \right\}\]

with \(0 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor - 1\), be the set of words composed by a rise step, a \((q-2)\)-colored Motzkin word \(\alpha\) of length \(i\), and a elevated \((q-2)\)-colored Motzkin word \(\beta\) of length \(n-i-1\) (see Fig. 4).
Figure 4: Graphical representation of the set $B_q(n)$, $n \geq 3$

\[ \alpha \in M_{q-2}(i) \quad \beta \in \hat{M}_{q-2}(n-i-1) \]

Figure 5: Graphical representation of the set $C_q(n)$, $n \geq 3$

\[ C_q(n) = \left\{ \gamma : \gamma \in M_{q-2}(n-1), \gamma \neq \alpha \beta v, \beta \in M_{q-2}(j) \right\} \]

Then, the enumeration of the set $B_q(n)$ is given by the following expression

\[ |B_q(n)| = \sum_{i=0}^{[n/2]-1} M_{q-2}(i) M_{q-2}(n-i-3). \]

Let

\[ C_q(n) = \left\{ \gamma : \gamma \in M_{q-2}(n-1), \gamma \neq \alpha \beta v, \beta \in M_{q-2}(j) \right\} \]

with $j \geq \lceil \frac{n}{2} \rceil$, be the set of words composed by a $(q-2)$-colored Motzkin word $\gamma$ of length $n-1$ that avoids elevated $(q-2)$-colored Motzkin words of length $j$, and a fall step (see Fig. 5).

Then, the enumeration of the set $C_q(n)$ is given by the following expression

\[ |C_q(n)| = M_{q-2}(n-1) - \sum_{k=\lfloor n/2 \rfloor}^{n-1} \sum_{i=0}^{n-1-k} M_{q-2}(i) M_{q-2}(k-2) M_{q-2}(n-1-i-k). \]

Note that, in order to obtain the size $|C_q(n)|$ we need to subtract from all words $\gamma$ of length $n-1$ those containing a elevated Motzkin subword $\beta$ of length greater than or equal to $\lceil n/2 \rceil$, and $\gamma$ can contain one of those subwords at most. Then, for $k = \lfloor n/2 \rfloor, \ldots, n-1$ we need to remove the words $u \beta v$, with $u \in M_{q-2}(i), \beta \in M_{q-2}(k), v \in M_{q-2}(n-1-i-k)$ and $0 \leq i \leq n-1-k$.

At this point, we define the set $CBFS_q(n)$ as follows

\[ CBFS_q(n) = A_q(n) \cup B_q(n) \cup C_q(n) \]
that is the union of the above described sets. For instance, in Fig. 6 the set \( CBFS_3(4) \) is depicted.

**Figure 6:** Graphical representation of the set \( CBFS_3(4) \)

```
  1 2 2 0  1 1 0 0  2 1 2 0  2 2 1 0
  1 1 2 0  1 2 1 0  2 2 2 0
```

**Proposition 3.1** The set \( CBFS_q(n) \) is a cross-bifix-free set on \( BF_q(n) \), for any \( q \geq 3 \) and \( n \geq 3 \).

**Proof.** Let \( w, w' \in CBFS_q(n) \). Let \( u \) be a prefix of \( w \) and \( v \) be a suffix of \( w' \) such that \( |u| = |v| \). We need to check that in each case the prefix \( u \) does not match with the suffix \( v \).

1. Let \( w \in A_q(n) \) and \( w' \in A_q(n) \cup B_q(n) \).
   - For each prefix \( u \) of \( w \) we have \( |u|_0 \leq |u|_1 \) and if \( |u| > \lceil \frac{n}{2} \rceil \), then \( |u|_0 < |u|_1 \).
   - For each suffix \( v \) of \( w' \) we have \( |v|_0 \geq |v|_1 \) and if \( |v| < \lceil \frac{n+1}{2} \rceil \), then \( |v|_0 > |v|_1 \).
   - Let \( |u| = |v| = l \), if either \( l < \lceil \frac{n+1}{2} \rceil \) or \( l > \lceil \frac{n}{2} \rceil \), then \( u \) does not match with \( v \). So we have to check the case \( \lceil \frac{n+1}{2} \rceil \leq l \leq \lceil \frac{n}{2} \rceil \).
   - If \( n \) is odd, it does not exist an integer \( l \) satisfying \( \lceil \frac{n+1}{2} \rceil \leq l \leq \lceil \frac{n}{2} \rceil \), otherwise if \( n \) is even, the case \( \lceil \frac{n+1}{2} \rceil \leq l \leq \lceil \frac{n}{2} \rceil \) is verified only for \( l = \frac{n}{2} \).
   - Therefore let \( n \) be even and \( l = \frac{n}{2} \). In this case \( |u|_0 \leq |u|_1 \) and \( |v|_0 \geq |v|_1 \).
   - At this point \( u \) can match with \( v \) only if \( |v|_0 = |v|_1 \), and this can happen only if \( v \) is an elevated Motzkin word. Suppose now that \( u = v \), so \( u \) should be a elevated Motzkin word too, and they have both length \( \frac{n}{2} \). In this case, \( w \) should be a word composed of two elevated Motzkin subwords of the same length, but such a word does not exists in \( CBFS_q(n) \) since the set \( \{ \alpha \beta : \alpha, \beta \in \tilde{M}_{q-2}(\frac{n}{2}) \} \) is not included in it, thus \( u \) does not match with \( v \).

2. Let \( w \in B_q(n) \) and \( w' \in A_q(n) \cup B_q(n) \).
   - For each prefix \( u \) of \( w \) we have \( |u|_0 < |u|_1 \), and for each suffix \( v \) of \( w' \) we have \( |v|_0 \geq |v|_1 \), thus \( u \) does not match with \( v \).

3. Let \( w \in C_q(n) \) and \( w' \in A_q(n) \cup B_q(n) \).
   - For each prefix \( u \) of \( w \) we have \( |u|_0 \leq |u|_1 \). For each suffix \( v \) of \( w' \) we have
\( |v|_0 \geq |v|_1 \) and if \( |u| < \lfloor \frac{n+1}{2} \rfloor \), then \( |v|_0 > |v|_1 \).

Let \( |u| = |v| = l \). If \( l < \lfloor \frac{n+1}{2} \rfloor \), then \( u \) does not match with \( v \). So we have to check the case \( l \geq \lfloor \frac{n+1}{2} \rfloor \). In this case \( v \) contains a elevated Motzkin subword of length \( \lfloor \frac{n+1}{2} \rfloor = \lfloor \frac{n}{2} \rfloor \) at least, and \( u \) does not match with \( v \), since \( u \) avoids such subwords.

4. Let \( w \in CBFS_q(n) \) and \( w' \in C_q(n) \).

For each prefix \( u \) of \( w \) we have \( |u|_0 \leq |u|_1 \), and for each suffix \( v \) of \( w' \) we have \( |v|_0 > |v|_1 \), thus \( u \) cannot match with \( v \).

We proved that \( CBFS_q(n) \) is a cross-bifix-free set on \( BF_q(n) \), for any \( q \geq 3 \) and \( n \geq 3 \).

\[ \blacksquare \]

**Proposition 3.2** The set \( CBFS_q(n) \) is a non-expandable cross-bifix-free set on \( BF_q(n) \), for any \( q \geq 3 \) and \( n \geq 3 \).

**Proof.** Let \( w \in BF_q(n) \setminus CBFS_q(n) \) and \( W = CBFS_q(n) \cup \{w\} \). If \( w \) begins with 0 then \( W \) is not cross-bifix-free since any word in \( CBFS_q(n) \) ends with 0. If \( w \) ends with 1 then \( W \) is not cross-bifix-free since any word in \( A_q(n) \) begins with 1. If \( w \) ends with a letter \( k \neq 0,1 \) then \( W \) is not cross-bifix-free since the suffix \( k \) of \( w \) matches, for instance, with the prefix \( k \) of the word \( k^{n-1}0 \in C_q(n) \). Consequently we have to consider \( w \) as a word beginning with a non-zero letter and ending with 0.

Let \( h = |w|_1 - |w|_0 \) be the ordinate of the last point of the path corresponding to \( w \). We now need to distinguish three different cases: \( h > 0 \), \( h < 0 \) and \( h = 0 \).

If \( h > 0 \), \( w \) can be written as (see Fig. 7)

\[ w = \phi 1 \mu_1 1 \mu_2 \cdots 1 \mu_h, \]

where \( \phi \) is a word satisfying \(|\phi|_1 = |\phi|_0 \) and not beginning with 0, and \( \mu_1, \ldots, \mu_h \) are \((q-2)\)-colored Motzkin words with \( \mu_h \) non-empty as \( w \) ends with 0.

In this case, if \(|\mu_h| = l \leq n - 2 \), considering for instance the word \( u = 1\mu_h 2^{n-l-2}0 \in A_q(n) \) we can clearly see that \( 1\mu_h \) is a cross-bifix between \( w \) and \( u \), and then \( W \) is not cross-bifix-free. On the other hand, if \(|\mu_h| = n-1 \), then necessarily \( h = 1 \) and \( w = 1\mu_1 \). So, \( w \) can be written as \( w = 1\alpha\beta \), where \( \alpha \in M_{q-2}(i), \beta \in M_{q-2}(n-i-1) \) with \( i > \lfloor \frac{n}{2} \rfloor \) (otherwise \( w \in B_q(n) \)). In this case, for instance, the word \( \beta 12^{i-1}0 \in A_q(n) \) has a cross-bifix with \( w \), thus \( W \) is not a cross-bifix-free-set.

If \( h < 0 \), \( w \) can be written as (see Fig. 8)

\[ w = \mu_{-h} 0 \cdots \mu_2 0 \mu_1 0 \phi \]
Figure 7: Graphical representation of $w$, in the case $h > 0$

\[
\begin{array}{cccccccc}
\phi & 1 & \mu_1 & 1 & \mu_2 & \cdots & 1 & \mu_h \\
\end{array}
\]

where $\phi$ is a word satisfying $|\phi|_1 = |\phi|_0$ and ending with 0, and $\mu_1, \ldots, \mu_h$ are $(q-2)$-colored Motzkin words with $\mu_h$ non-empty as $w$ begins with a non-zero letter.

Figure 8: Graphical representation of $w$, in the case $h < 0$

\[
\begin{array}{cccccccc}
\mu_{-h} & 0 & \cdots & \mu_2 & 0 & \mu_1 & 0 & \phi \\
\end{array}
\]

In this case, if $|\mu_{-h}| = l \leq n - 2$, considering for instance the word $u = 12^{n-l-2}\mu_{-h}0 \in A_q(n)$ we can clearly see that $\mu_{-h}0$ is a cross-bifix between $w$ and $u$, and then $W$ is not cross-bifix-free. On the other hand, if $|\mu_{-h}| = n - 1$, then necessarily $h = -1$ and $w = \mu_10$. So, $w$ can be written as $w = \alpha\beta\delta0$, where $\beta \in \mathcal{M}_{q-2}(j)$ with $j \geq \lfloor \frac{n}{2} \rfloor$ (otherwise $w \in C_q(n)$), and $\alpha, \delta$ any two $(q-2)$-colored Motzkin words of the appropriate length. In this case, for instance, the word $2^{n-j-1}\alpha\beta \in A_q(n)$ has a cross-bifix with $w$, thus $W$ is not a cross-bifix-free-set.

Finally, if $h = 0$, the path associated to $w$ can either remain above $x$-axis or fall below it.

In the first case let $i$, with $\lfloor \frac{n}{2} \rfloor \leq i < n$, be the last $x$-coordinate of the path intercepting the $x$-axis. Notice that $i$ cannot be less than $\lfloor \frac{n}{2} \rfloor$, otherwise $w \in A_q(n)$. We can write $w = \alpha\beta$, where $\alpha$ is a non-empty word in $\mathcal{M}_{q-2}(i)$ and $\beta \in \mathcal{M}_{q-2}(n-i)$. We now need to take into consideration two different cases: $i = \lfloor \frac{n}{2} \rfloor$ and $i > \lfloor \frac{n}{2} \rfloor$. In the first case $\alpha \in \mathcal{M}_{q-2}(\frac{n}{2})$, otherwise $w \in A_q(n)$, then, for instance, the word $2^{n/2}\alpha \in A_q(n)$ has a cross-bifix with $w$. In the latter case, for instance, the word $\beta 2^{i-1}0 \in C_q(n)$ has a cross-bifix with $w$, so that $W$ is not a cross-bifix-free-set.

In the other case the path associated to $w$ crosses the $x$-axis. Let $i$, with
0 < i < n, be the first x-coordinate of the path crossing x-axis. We can write $w = \alpha 0\phi$, where $\alpha$ is a non-empty word in $M_{q-2}(i)$. In this case, for instance, the word $12^{n-i-2}0 \in A_q(n)$ has a cross-bifix with $w$, then $W$ is not a cross-bifix-free-set.

We proved that $CBFS_q(n)$ is a non-expandable cross-bifix-free set on $BF_q(n)$, for any $q \geq 3$ and $n \geq 3$. ■

4 Sizes of Cross-Bifix-Free sets for Small Lengths

In this section we present some interesting results concerning the size of $CBFS_q(n)$ compared to the ones in [11].

For fixed $n$ and $q$, we recall that the size of $q$-ary cross-bifix-free sets given in [11] is obtained by

$$S(n, q) = \max\{(q - 1)^2F_{k,q}(n - k - 2) : 2 \leq k \leq n - 2\}$$

which is proved to be nearly optimal.

In Table III is shown the values of $S(n, q)$ and $|CBFS_q(n)|$ for $3 \leq q \leq 6$ and $n \leq 16$. For the initial values of $n$, we can observe that the sizes obtained by our construction are greater than the size $S(n, q)$. In particular, the number of the initial values of $n$ for which $|CBFS_q(n)|$ is greater grows with $q$ and this trend can be easily verified by experimental results.

In order to improve the values of the size $S(n, q)$ for the initial size of $n$, we can consider the following expression

$$S^*(n, q) = \max\{(q - 1)^2F_{k,q}(n - k - 2) : 1 \leq k \leq n - 2\},$$

where $k$ can assume also the value 1. When $k = 1$, in the case of small $n$ and large $q$, we obtain cross-bifix-free sets having cardinality greater than the one proposed in [11].

In Table IV is shown the values of $S^*(n, q)$ and $|CBFS_q(n)|$ for $3 \leq q \leq 6$ and $n \leq 16$. Also in this situation, we can observe that the sizes obtained by our construction are greater than the size $S(n, q)$ in a range of values of $n$. In particular, the range of values of $n$ for which $|CBFS_q(n)|$ is greater grows with $q$ and this trend can be easily verified by experimental results.
Table 3: Comparing the values from [11] with $CBFS_q(n)$, for $3 \leq q \leq 6$

| n | $|CBFS_3(n)|$ | $S(n, 3)$ | $|CBFS_4(n)|$ | $S(n, 4)$ |
|---|---|---|---|---|
| 3 | 4 | 4 | 9 | 9 |
| 4 | 7 | 4 | 25 | 9 |
| 5 | 16 | 12 | 72 | 36 |
| 6 | 36 | 32 | 225 | 135 |
| 7 | 87 | 88 | 712 | 513 |
| 8 | 210 | 240 | 2 334 | 1 944 |
| 9 | 535 | 656 | 7 868 | 7 371 |
| 10 | 1 350 | 1 792 | 26 731 | 27 945 |
| 11 | 3 545 | 4 896 | 93 175 | 105 948 |
| 12 | 9 205 | 13 376 | 324 520 | 401 679 |
| 13 | 24 698 | 36 544 | 1 157 031 | 1 522 881 |
| 14 | 65 467 | 99 840 | 4 104 449 | 5 773 680 |
| 15 | 178 375 | 272 768 | 14 874 100 | 21 889 683 |
| 16 | 480 197 | 745 216 | 53 514 974 | 82 990 689 |

| n | $|CBFS_5(n)|$ | $S(n, 5)$ | $|CBFS_6(n)|$ | $S(n, 6)$ |
|---|---|---|---|---|
| 3 | 16 | 16 | 25 | 25 |
| 4 | 61 | 16 | 121 | 25 |
| 5 | 224 | 80 | 550 | 150 |
| 6 | 900 | 384 | 2 739 | 875 |
| 7 | 3 595 | 1 856 | 13 260 | 5 125 |
| 8 | 15 014 | 8 960 | 67 740 | 30 000 |
| 9 | 63 135 | 43 264 | 342 676 | 175 625 |
| 10 | 271 136 | 208 896 | 1 787 415 | 1 028 125 |
| 11 | 1 178 677 | 1 008 640 | 9 324 647 | 6 018 750 |
| 12 | 5 167 953 | 4 870 144 | 49 456 240 | 35 234 375 |
| 13 | 22 986 100 | 23 515 136 | 263 776 127 | 206 265 625 |
| 14 | 102 403 229 | 113 541 120 | 1 417 981 855 | 1 207 500 000 |
| 15 | 463 098 075 | 548 225 024 | 7 688 015 908 | 7 068 828 125 |
| 16 | 2 089 302 415 | 2 647 064 576 | 41 785 951 916 | 41 381 640 625 |
Table 4: Comparing the values from $S^*(n,q)$ with $CBFS_q(n)$, for $3 \leq q \leq 6$

| $n$ | $|CBFS_3(n)|$ | $S^*(n,3)$ | $|CBFS_4(n)|$ | $S^*(n,4)$ |
|-----|--------------|------------|--------------|------------|
| 3   | 4            | 4          | 9            | 9          |
| 4   | 7            | 8          | 25           | 27         |
| 5   | 16           | 16         | 72           | 81         |
| 6   | 36           | 32         | 233          | 243        |
| 7   | 87           | 88         | 712          | 729        |
| 8   | 210          | 240        | 2 334        | 2 187      |
| 9   | 535          | 656        | 7 868        | 7 371      |
| 10  | 1 350        | 1 792      | 26 731       | 27 945     |
| 11  | 3 545        | 4 896      | 93 175       | 105 948    |
| 12  | 9 205        | 13 376     | 324 520      | 401 679    |
| 13  | 24 698       | 36 544     | 1 157 031    | 1 522 881  |
| 14  | 65 467       | 99 840     | 4 104 449    | 5 773 680  |
| 15  | 178 375      | 272 768    | 14 874 100   | 21 889 683 |
| 16  | 480 197      | 745 216    | 53 514 974   | 82 990 089 |

| $n$ | $|CBFS_5(n)|$ | $S^*(n,5)$ | $|CBFS_6(n)|$ | $S^*(n,6)$ |
|-----|--------------|------------|--------------|------------|
| 3   | 16           | 16         | 25           | 25         |
| 4   | 61           | 64         | 121          | 125        |
| 5   | 224          | 256        | 550          | 625        |
| 6   | 900          | 1 024      | 2 739        | 3 125      |
| 7   | 3 595        | 4 096      | 13 260       | 15 625     |
| 8   | 15 014       | 16 384     | 67 740       | 78 125     |
| 9   | 63 135       | 65 536     | 342 676      | 390 625    |
| 10  | 271 136      | 262 144    | 1 787 415    | 1 953 125  |
| 11  | 1 178 677    | 1 048 576  | 9 324 647    | 9 765 625  |
| 12  | 5 167 953    | 4 870 144  | 49 456 240   | 48 828 125 |
| 13  | 22 986 100   | 23 515 136 | 263 776 127  | 244 140 625|
| 14  | 102 403 229  | 113 541 120| 1 417 981 855| 1 220 703 125|
| 15  | 463 098 075  | 548 225 024| 7 688 015 908| 7 068 828 125|
| 16  | 2 089 302 415| 2 647 064 576| 41 785 951 916| 41 381 640 625|
5 Conclusions and further developments

In this paper, we introduce a general constructive method for cross-bifix-free sets in the $q$-ary alphabet based upon the study of lattice paths on the discrete plane. This approach enables us to obtain the cross-bifix-free set $\mathcal{CBFS}_q(n)$ having greater cardinality than the ones proposed in [11], for the initial values of $n$.

Moreover, we prove that $\mathcal{CBFS}_q(n)$ is a non-expandable cross-bifix-free set on $BF_q(n)$, i.e. $\mathcal{CBFS}_q(n) \cup \{w\}$ is not a cross-bifix-free set on $BF_q(n)$, for any $w \in BF_q(n) \setminus \mathcal{CBFS}_q(n)$.

The non-expandable property is obviously a necessary condition to obtain a maximal cross-bifix-free set on $BF_q(n)$, anyway the problem of determine maximal cross-bifix-free sets is still open and no general solution has been found yet.

References

[1] R. H. Barker. Group synchronizing of binary digital systems. *Communication theory*, W. Jackson, Ed. London, U.K.: Butterworth, 273-287, 1953.

[2] J. Lindner. Binary sequences up to length 40 with best possible autocorrelation function. *Electronics letters*, vol. 11, 507, 1975.

[3] R. A. Scholtz. Frame synchronization techniques. *IEEE Transactions on Communications*, vol. COM-20, 1204-1213, August 1980.

[4] A. J. de Lind van Wijngaarden, T. J. Willink. Frame Synchronization Using Distributed Sequences. *IEEE Transactions on Communications*, vol. 48, No.12, 2127-2138, 2000.

[5] P. T. Nielsen. On the Expected Duration of a Search for a Fixed Pattern in Random Data. *IEEE Transactions on Information Theory*, vol. IT-29, 702-704, September 1973.

[6] D. Bajic, J. Stojanovic. Distributed Sequences and Search Process. *IEEE International Conference on Communications ICC2004*, Paris, 514-518, June 2004.

[7] M. Crochemore, C. Hancart, T. Lecroq. Algorithms on strings. Cambridge University Press, Cambridge 2007.

[8] J. Berstel, D. Perrin, C. Reutenauer. Codes and Automata (Encyclopedia of Mathematics and its Applications). Cambridge University Press, Cambridge 2009.

[9] D. Bajic. On Construction of Cross-Bifix-Free Kernel Sets. 2nd MCM COST 2100, TD(07)237, Lisbon, Portugal, February 2007.
[10] S. Bilotta, E. Pergola, R. Pinzani. A new approach to cross-bifix-free sets. *IEEE Transactions on Information Theory* **58** 4058–4063, 2012.

[11] Chee, Y. M., Kiah, H. M., Purkayastha, P. and Wang, C. Cross-bifix-free codes within a constant factor of optimality. *IEEE Transactions on Information Theory* **59** 4668–4674, 2013.

[12] P. T. Nielsen. A Note on Bifix-Free Sequences. *IEEE Transactions on Information Theory*, vol. IT-29, 704-706, September 1973.

[13] Levesque, C.: On $m$-th order linear recurrences. Fibonacci Quart. **23** (1985) 290–293

[14] E. Barcucci, A. Del Lungo, E. Pergola, R. Pinzani. A construction for enumerating $k$-coloured Motzkin paths. Lecture Notes in Computer Science 959, 254–263, 1995.