Non-Overlapping Matrices via Dyck Words

Elena Barcucci*, Antonio Bernini*, Stefano Bilotta† and Renzo Pinzani*

Dipartimento di Matematica e Informatica "U. Dini".
Email: elena.barcucci@unifi.it, antonio.bernini@unifi.it, renzo.pinzani@unifi.it

†*Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Firenze.*
Email: stefano.bilotta@unifi.it

ABSTRACT: We define a set of binary matrices where any two of these can not be placed one on another such that the corresponding entries coincide. The rows of the matrices are obtained by means of Dyck words. The cardinality of the set of such matrices involves Catalan numbers.

1. Introduction

In the last few years several researchers focused their attention to the topic of non-overlapping matrices which can be seen as a generalization of the more famous subject of non-overlapping words. These words are a part of the wide area of Theory of Codes, where, among many interesting questions and applications, the definition of sets of words with particular characteristics is considered. Much care has been paid to matters involving overlaps among the words of a set (or, more precisely, among their prefixes or suffixes). For example, since their introduction, various kinds of codes have been considered: prefix-free, suffix-free and fix-free codes [15,18], comma-free codes [16], bifix-free codes [17], non-overlapping codes [12], cross-bifix-free codes [2,3,11] (often called non-overlapping codes), variable length non-overlapping codes [10]. More recently, cross-bifix-free codes have been deeply studied thanks to their involvement in the development of DNA-based storage systems for big amounts of data [19,20]. In literature several approaches can be found providing different definitions of cross-bifix-free words [7,11,13] and concerning their listing in a Gray code sense [8,9].

The generalization of the notion of non-overlapping strings to two dimensions appears, probably for the first time in a rigorous manner, in the article [5] where only square matrices are considered. Here, the notion of prefix (and suffix) of a string is simply translated into a square matrix anchored in the top-left or bottom-right entry. Nevertheless, in other papers [4,6] more general matrices are considered and the the notion of prefix and suffix is also improved. In particular, a set of non-overlapping matrices is defined over a finite alphabet where the rows of the matrices (which are considered as words) are subjected to some constraints. More precisely, some entries of the matrices are fixed and the rows must avoid certain consecutive patterns of symbols of the alphabet.

In the present paper we give a new set of non-overlapping matrices over the binary alphabet. The rows are not restricted words, in the sense that there are no patterns to be avoided. Here, we use the well known Dyck words which allow us to define the matrices of the set in a simpler way with respect to [4,6]. Moreover, being Dyck words enumerated by the famous Catalan numbers, we are able to find a closed formula for the cardinality of the set, depending only on the dimension of the matrices. A further remark lies in the fact that in each matrix the numbers of 1's and 0's are substantially balanced, thanks to the use of Dyck words, differently from

the matrices defined in [4,6]. We conclude by observing that the proofs of the properties of the matrices involve only the examination of prefixes or suffixes of the strings constituting their rows, by using standard techniques typical of non-overlapping words.

2.   Definitions and background

The matrices we are dealing with in the present paper are binary matrices and their rows are generated using particular binary strings (Dyck words). Before describing the construction of the main set (next section), we give a formal definition of non-overlapping matrices, together with some tools useful in the sequel. We also recall some concepts about strings.

Definition 2.1. Let u and v be two strings over an alphabet $\Sigma = \{\sigma_1, \sigma_2 \ldots, \sigma_q\}$. They are said *non-overlapping strings* if any non-empty proper prefix of u is different from any non-empty proper suffix of v, and vice versa. If $u = v$, then the string u is said *self non-overlapping*.

If some prefix (suffix) of u matches with some suffix (prefix) of v, then u and v are *overlapping strings*. When a factor of length k of consecutive symbols $\sigma_j \in \Sigma$ occurs in a string u, we denote it by σ_j^k. For example, the binary string $u = 1000110$ can be indicated as $u = 10^3 1^2 0$.

We extend the notion of non-overlapping strings to the case of matrices:

Definition 2.2. Let A and B be two distinct matrices in $M_{m \times n}$, the set of all matrices with m rows and n columns. A and B are said *non-overlapping* if all the following conditions are satisfied by A and B:

- there do not exist two block partitions

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

  such that $A_{11}, B_{22} \in M_{r \times s}$, with $1 \le r \le m-1$, $1 \le s \le n-1$, and either $A_{11} = B_{22}$ nor $A_{12} = B_{21}$, or $A_{21} = B_{12}$, or $A_{22} = B_{11}$.

- there do not exist two block partitions

$$A = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} \quad B = \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$$

  such that $A_{11}, B_{21} \in M_{r \times n}$, with $1 \le r \le m-1$, and either $A_{11} = B_{21}$, or $A_{21} = B_{11}$.

- there do not exist two block partitions

$$A = \begin{bmatrix} A_{11} & A_{12} \end{bmatrix} \quad B = \begin{bmatrix} B_{11} & B_{12} \end{bmatrix}$$

  such that $A_{11}, B_{12} \in M_{m \times s}$, with $1 \le s \le n-1$, and either $A_{11} = B_{12}$, or $A_{12} = B_{11}$.

In Figure 1 three pairs of overlapping matrices contradicting the three above points are depicted. Informally, Definition 2.2 states that two matrices A and B of the same dimension do not overlap if there is no a rigid movement of B on A (note that rotations are not allowed) such that at the end the corresponding entries of B and A coincide. More precisely, after the rigid slipping, a common area (*control window*) is formed. This common area is a rectangular array (the intersection between A and B) consisting of a finite number of 1×1 cells of the discrete plane, each containing an entry of A and an entry of B. If all the entries of A coincide with the corresponding entries of B, then the two matrices are said overlapping matrices, otherwise they are two non-overlapping matrices. Note that the roles of A and B can be interchanged so that we can assume, without loss of generality, that B is the matrix that moves on A.

Finally, Definition 2.2 can be naturally extended to the case $A = B$, and the matrix is said *self non-overlapping*.

Definition 2.3. A set $S_{m \times n} \subset M_{m \times n}$ is called *non-overlapping* if each matrix of $S_{m \times n}$ is self non-overlapping and for any two matrices $A, B \in S_{m \times n}$ they are non-overlapping matrices.
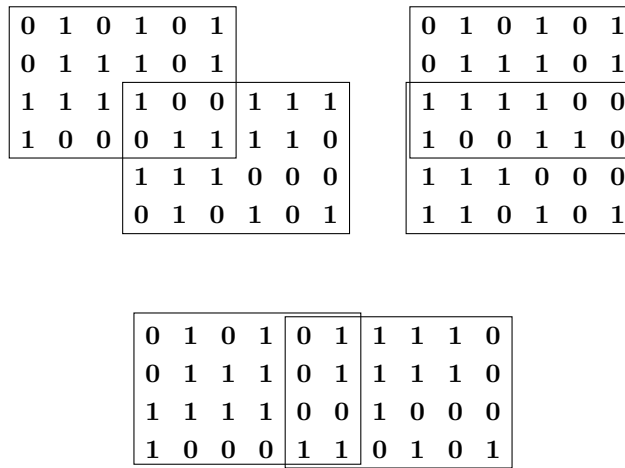
$$
\begin{array}{cccccc}
0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\
 & & & 1 & 1 & 1 & 0 & 0 & 0 \\
 & & & 0 & 1 & 0 & 1 & 0 & 1
\end{array}
\qquad
\begin{array}{cccccc}
0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1
\end{array}
$$

$$
\begin{array}{cccccccc}
0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1
\end{array}
$$

Figure 1: Three pairs of overlapping matrices of 4×6.

If $A \in M_{m \times n}$, we indicate, as usual, by $a_{i,j}$ the entry in the i-th row and j-th column, so that $A = (a_{i,j})$, with $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. A generic row of A is denoted by A_i. Moreover, in the following, if u is a binary string, we denote its length by $|u|$. The number of 1's and the number of 0's will be denoted by $|u|_1$ and $|u|_0$, respectively.

Definition 2.4. Let $\ell \geqslant 1$. A *Dyck word* v is a binary string of length 2ℓ such that $|v|_1 = |v|_0$ and $|\gamma|_1 \geq |\gamma|_0$ for each prefix γ of v. The set of all the Dyck words of length 2ℓ is denoted by $D_{2\ell}$.

Note that in a Dyck word it is also $|\delta|_1 \leqslant |\delta|_0$ for any suffix δ of v. For example if $\ell = 4$, some Dyck words of length 8 are $11110000, 10101010, 11001100$.

Definition 2.5. A Dyck word v is a *type α* Dyck word if for each proper prefix γ, we have $|\gamma|_1 > |\gamma|_0$.

Clearly, in a `type` α Dyck word it is also $|\delta|_1 < |\delta|_0$ for any proper suffix δ of v. Then, a `type` α Dyck word is a self non-overlapping string [11]. Moreover, for each `type` α Dyck word v of length 2ℓ, it is $v = 1u0$ for some Dyck word $u \in D_{2\ell-2}$. For example if $\ell = 4$, two `type` α Dyck words of length 8 are $v = 11110000$ and $v = 11100100$ where $u = 111000$ and $u = 110010$, respectively, while the string 10101010 is not a `type` α Dyck word.

We conclude this section by recalling that a Dyck word $u \in D_{2\ell}$ can be read as Dyck path which is a lattice path starting from $(0,0)$, ending at $(2\ell, 0)$, never crossing the x-axis and using only north-east and south-east steps. A `type` α Dyck word $v = 1u0$, with $|v| = 2\ell$ and $u \in D_{2\ell-2}$, is a Dyck path which touches the x-axis only at the beginning and at the end.

3. A set of non-overlapping binary matrices

In this section we are going to define a particular set of non-overlapping binary matrices, the set $L_{m \times n}$. We distinguish two cases depending on the number n of columns: n even and n odd.

3.1 The set $L_{m \times n}$ with n even

Definition 3.1. Let $m \geqslant 2$ and $n \geqslant 4$ with n even. We denote by $L_{m \times n} \subset M_{m \times n}$ the set of matrices satisfying the following conditions. Let $A = (a_{i,j})$ an element of $L_{m \times n}$.

- The first row, A_1, is a *type α* Dyck word of length n: $A_1 = 1u0$ where $u \in D_{n-2}$. Moreover, we require that the first row A_1 is the same for all the matrices of $L_{m \times n}$ and it can not appear in any other row.

- For $i = 2, \ldots, m-1$, the row A_i is a string of length n chosen among the following different five types of strings:

- *type 1*: a Dyck word of length n ($A_i \in D_n$).
- *type 2*: a Dyck word of length $n-2$ different from u preceded by 11 ($A_i = 11w$ where $w \in D_{n-2}$ and $w \neq u$).
- *type 3*: a Dyck word of length $n-2$ different from u followed by 00 ($A_i = w00$ where $w \in D_{n-2}$ and $w \neq u$).
- *type 4*: a Dyck word of length $n-2$ different from u preceded by 01 ($A_i = 01w$ where $w \in D_{n-2}$ and $w \neq u$).
- *type 5*: a Dyck word of length $n-2$ followed and preceded by 0 ($A_i = 0w0$ where $w \in D_{n-2}$).

  We refer to a row A_i ($i = 2, 3, \ldots, m-1$) as a *type k* row, with $k = 1, 2, 3, 4, 5$. Note that for *type 2, type 3, type 4* rows it is required $w \neq u$, where u is the Dyck word used to generate A_1.

- The last row A_m, is a *type 1, type 2* or *type 3* row (*type 4* and *type 5* are not allowed in the last row).

In Figure 2 an element of the set $L_{7 \times 8}$ is given. The *italic* entries correspond to the ones needed to obtain a type 2, type 3, type 4 or type 5 row.

1	1	0	1	0	1	0	0	type α row
1	*1*	1	1	0	0	1	0	type 2 row
1	0	1	1	0	0	*0*	*0*	type 3 row
0	*1*	1	1	0	0	1	0	type 4 row
0	1	0	1	0	1	0	*0*	type 5 row
1	0	1	0	1	0	1	0	type 1 row
1	0	1	0	1	0	1	0	type 1 row

Figure 2: An element of $L_{7 \times 8}$.

Proposition 3.2. *$L_{m \times n}$ with n even is a non-overlapping set.*

Proof. Let A and B two matrices in $L_{m \times n}$ (possibly the same). Since we have to check if A and B overlap, we immediately notice that the first row of A or B is always involved in the control window. We consider two different kinds of overlaps: *vertical overlap* when the i-th column of A is superimposed on the i-th column of B and *skew overlap* in all the other cases.

If a vertical overlap occurs, then A and B can not overlap since, looking at the control window (Figure 3), the row A_1 ($B_1 = A_1$) is over a row B_j (A_j) and, by Definition 3.1, $A_1 \neq B_j$ ($B_1 \neq A_j$), for $j = 2, 3, \ldots, m$.

1	1	1	0	0	0
1	1	1	0	0	0

Figure 3: Vertical overlap

If A and B skew overlap, then two cases must be considered: the control window contains a prefix of A_1 (or $B_1 = A_1$) or the control windows contains a suffix of A_1 or B_1.

- Let γ be a proper prefix of A_1 which is involved in the control window: it is superimposed with a proper suffix δ of B_j, $j = 1, 2, \ldots, m$. Depending on the type of the row B_j, we have different cases:
  - B_j is a type 1 row: since $|\gamma|_1 > |\gamma|_0$ and $|\delta|_0 \geqslant |\delta|_1$, the prefix γ and the suffix δ do not match, then rows A_1 and B_j, and consequently the matrices A and B, do not overlap.

- B_j is a `type 2` or a `type 4` row: if $|\delta| \leqslant n - 2$, then the same argument of the above case can be used to conclude that A and B do not overlap, while, if $|\delta| = n - 1$, then we observe that $\delta = 1 b_{j,3} b_{j,4} \ldots b_{j,n}$ and $\gamma = 1u$ where $u \in D_{n-2}$, so that γ and δ do not match since, by Definition 3.1, it is $u \neq b_{j,3} b_{j,4} \ldots b_{j,n}$. Therefore, A and B do not overlap.

  - B_j is a `type 3` or a `type 5` row: since $|\gamma|_1 > |\gamma|_0$ and $|\delta|_0 > |\delta|_1$, matrices A and B can not overlap.

  Notice that in the above bullet, the roles of A and B can be interchanged, leading to the same conclusion.

- Let δ be a proper suffix of A_1 which is involved in the control window: it is superimposed with a proper prefix γ of B_j, for some $j = 1, 2, \ldots, m$. Depending on the type of the row B_j, we have the following different cases:

  - B_j is a `type 1` row: since $|\delta|_0 > |\delta|_1$ and $|\gamma|_1 \geqslant |\gamma|_0$, the prefix γ and the suffix δ do not match, then rows A_1 and B_j, and consequently the matrices A and B, do not overlap.

  - B_j is a `type 2` row: since $|\delta|_0 > |\delta|_1$ and $|\gamma|_1 > |\gamma|_0$, similarly to the above case, matrices A and B can not overlap.

  - B_j is a `type 3` row: if $|\gamma| \leqslant n - 2$, no match is possible (see before) between γ and δ (so that A and B do not overlap), while, if $|\gamma| = n - 1$, then we observe that $\gamma = b_{j,1} b_{j,2} \ldots b_{j,n-2} 0$ and $\delta = u0$ where $u \in D_{n-2}$, so that γ and δ do not match since, by Definition 3.1, it is $u \neq b_{j,1} b_{j,2} \ldots b_{j,n-2}$. Therefore A and B can not overlap.

  - B_j is a `type 4` row: in this case, by Definition 3.1, it is $j \neq m$. If $|\gamma| = 1$, then $\gamma = b_{1,j} = 0$ and $\delta = a_{1,n} = 0$ (each row of a matrix in the set ends with 0) form a match. Nevertheless, there exists an index i (with $2 \leqslant i \leqslant m - 1$) such that $b_{1,m} = 1$ is over the entry $a_{i,n}$ which is 0. Therefore, no overlapping between A and B is possible (Figure 4). If $|\gamma| > 1$, then $|\gamma|_1 \geqslant |\gamma|_0$ and $|\delta|_0 > |\delta|_1$ which does not allow A and B to overlap.

  - B_j is a `type 5` row: also in this case $j \neq m$. If $|\gamma| = 1$, the matrices A and B are arranged as in the above case and they do not overlap (Figure 4). If $|\gamma| > 1$, then $\gamma = 0w$ where w is a proper prefix of a Dyck word, so that $|\gamma|_1 \geqslant |\gamma|_0$ (if $|w|_1 > |w|_0$) or $|\gamma|_1 + 1 = |\gamma|_0$ (if $|w|_1 = |w|_0$). In the first case γ and δ do not match since it is $|\delta|_0 > |\delta|_1$. In the latter, the string w would be a suffix of A_1. This is not possible since A_1 is a `type` α row. We conclude that also in this case A and B do not overlap.
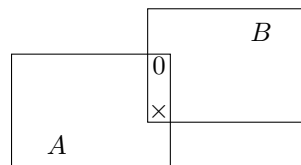

Figure 4: The prefix of a type 4 or type 5 row in B composed by the only first 0 is superimposed on the last 0 of the first row in A. The $\times$ represents a mismatch.

Notice that from the proof above we can deduce that A_1 and any other row of `type 1, 2` or `3` are non-overlapping strings. □

3.2 The set $L_{m \times n}$ with n odd

Definition 3.3. Let $m \geqslant 2$ and $n \geqslant 5$ with n odd. We denote by $L_{m \times n} \subset M_{m \times n}$ the set of matrices satisfying the following conditions. Let $A = (a_{i,j})$ an element of $L_{m \times n}$.

- The first row, A_1, is a *type* α Dyck word of length $n - 1$ preceded by 1: $A_1 = 11u0$ where $u \in D_{n-3}$. Moreover, we require that the first row is the same for all the matrices of $L_{m \times n}$ and it can not appear in any other row.

- For $i = 2, \ldots, m - 1$, the row A_i is a string of length n chosen among the following different three types of strings:

- **type 6**: a Dyck word of length $n-1$ preceded by 1 ($A_i = 1w$ where $w \in D_{n-1}$ and $w \neq u$).

- **type 7**: a Dyck word of length $n-1$ followed by 0 ($A_i = w0$ where $w \in D_{n-1}$ and $w \neq u$).

- **type 8**: a Dyck word of length $n-1$ preceded by 0 ($A_i = 0w$ where $w \in D_{n-1}$).

  We refer to a row A_i ($i = 2, 3, \ldots, m-1$) as a **type k** row, with **k** $= 6, 7, 8$. Note that for **type 6** and **type 7** rows it is required $w \neq u$, where u is the Dyck word used to generate A_1.

- The last row A_m is a **type 6** or **type 7** row (**type 8** is not allowed in the last row).

In Figure 5 an element of the set $L_{6 \times 7}$ is given. The *italic* entries correspond to the ones needed to obtain a `type 6`, `type 7` or `type 8` row.

1	**1**	**1**	**1**	**0**	**0**	**0**	type α
1	**1**	**0**	**1**	**0**	**1**	**0**	type 6
1	**0**	**1**	**0**	**1**	**0**	*0*	type 7
0	**1**	**1**	**1**	**0**	**0**	**0**	type 8
1	**1**	**1**	**0**	**0**	**1**	**0**	type 6
1	**1**	**0**	**0**	**1**	**0**	*0*	type 7

Figure 5: An element of $L_{6 \times 7}$.

Proposition 3.4. *$L_{m \times n}$ with n odd is a non-overlapping set.*

Proof. Let A and B be two matrices in $L_{m \times n}$ (possibly the same). The argument is very similar to the one used in the proof of the Proposition 3.2. Since we have to check if A and B overlap, we immediately notice that the first row of A or B is always involved in the control window.

We distinguish two different kinds of overlaps: *vertical overlap* and *skew overlap*. A vertical overlap can not occur since, by Definition 3.3, the j-th row, $j = 2, \ldots, m$, of any matrix in $L_{m \times n}$ is different from the common first row of all the matrices in the set. If A and B skew overlap, then two cases must be considered: the control window contains a prefix of A_1 or B_1 or the control windows contains a suffix of A_1 or B_1.

- Let γ be a proper prefix of A_1 which is involved in the control window: it is superimposed with a proper suffix δ of B_j, $j = 1, 2, \ldots, m$. Depending on the type of the row B_j, we have different cases:

  - B_j is a `type 6` or a `type 8` row: since $|\gamma|_1 > |\gamma|_0$ and $|\delta|_0 \geqslant |\delta|_1$, the prefix γ and the suffix δ do not match, then rows A_1 and B_j, and consequently the matrices A and B, do not overlap.

  - B_j is a `type 7` row: since $|\gamma|_1 > |\gamma|_0$ and $|\delta|_0 > |\delta|_1$, the prefix γ and the suffix δ do not match, then rows A_1 and B_j, and consequently the matrices A and B, do not overlap.

  Notice that in the above bullet, the roles of A and B can be interchanged, leading to the same conclusion.

- Let δ be a proper suffix of A_1 which is involved in the control window: it is superimposed with a proper prefix γ of B_j, $j = 1, 2, \ldots, m$. Depending on the type of the row B_j, we have different cases:

  - B_j is a `type 6` row: since $|\delta|_0 \geqslant |\delta|_1$ ($|\delta|_0 > |\delta|_1$ if $|\delta| \leqslant n-2$) and $|\gamma|_1 > |\gamma|_0$, the prefix γ and the suffix δ do not match, then rows A_1 and B_j, and consequently the matrices A and B, do not overlap.

  - B_j is a `type 7` row: if $|\delta| \leqslant n-2$, since $|\delta|_0 > |\delta|_1$ and $|\gamma|_1 \geqslant |\gamma|_0$, the prefix γ and the suffix δ do not match, then rows A_1 and B_j, and consequently the matrices A and B, do not overlap. If $|\delta| = n-1$, then $\delta = u$, where $u \in D_{n-1}$ and $\gamma = b_{j,1}b_{j,2} \ldots b_{j,n-1} = w$ where $w \in D_{n-1}$ so that γ and δ do not match since, by Definition 3.4, it is $u \neq w$. Therefore A and B can not overlap.

  - B_j is a `type 8` row: in this case $j \neq m$. If $|\gamma| = 1$, then $\gamma = b_{1,j} = 0$ and $\delta = a_{1,n} = 0$ (each row of a matrix in the set ends with 0) form a match. Nevertheless, there exists an index i (with $2 \leqslant i \leqslant m-1$) such that $b_{1,m} = 1$ is over the entry $a_{i,n}$ which is 0. Therefore, no overlapping between A and B is possible (Figure 4). If $|\gamma| > 1$, then $\gamma = 0w$ where w is a proper prefix of a Dyck word,

so that $|\gamma|_1 \geqslant |\gamma|_0$ (if $|w|_1 > |w|_0$) or $|\gamma|_1 + 1 = |\gamma|_0$ (if $|w|_1 = |w|_0$). In the first case γ and δ do not match since it is $|\delta|_0 > |\delta|_1$. In the latter, the string w would be a suffix of A_1. This is not possible since A_1 is a type α row. We conclude that also in this case A and B do not overlap.

Notice that, similarly to Proposition 3.2, from the above proof we can deduce that A_1 and any other row of type 6 or 7 are non-overlapping strings. □

4. The enumeration of $L_{m \times n}$

In this section we are going to enumerate the set $L_{m \times n}$. The first row A_1 of all the matrices in $L_{m \times n}$ is the same, so that it is easy to realize that the cardinality of $L_{m \times n}$ depends on the number of the remaining $m - 1$ rows satisfying the constraints of Definition 3.1 or 3.3. We recall that the Dyck words of length $2n$ are enumerated by the n-th Catalan number C_n given by

$$C_n = \frac{1}{n+1} \binom{2n}{n} .$$

Following the constraints of Definition 3.1, in the case n even and reminding that $A_1 = 1u0$ with $u \in D_{n-2}$, we observe that if the row A_i, $i = 2, 3, \ldots, m - 1$, is a type 2,3,4 row, then it can be constructed by using a Dyck word $w \in D_{n-2}$ and $w \neq u$ and adding the prefixes 11 or 01 or the suffix 00, so that there are $C_{\frac{n-2}{2}} - 1$ possibilities for each type of row. If A_i is a type 5 row, then it can be constructed in $C_{\frac{n-2}{2}}$ different ways by using whichever Dyck word $w \in D_{n-2}$ and adding the prefix and the suffix 0 ($A_i = 0w0$). Finally, if A_i is a type 1 row, then $A_i \in D_n$ and $A_i \neq A_1$. The foregoing is summarized in the first term of the following formula where the exponent $m - 2$ tracks the number of rows from the second to the second to last one. The second term counts the possibilities for the last row which is a type 1, 2 or type 3 row.

$$|L_{m \times n}| = \left(\underbrace{C_{\frac{n}{2}} - 1}_{\text{type 1}} + \underbrace{3C_{\frac{n-2}{2}} - 3}_{\text{type 2, 3, 4}} + \underbrace{C_{\frac{n-2}{2}}}_{\text{type 5}} \right)^{(m-2)} \left(\underbrace{C_{\frac{n}{2}} - 1}_{\text{type 1}} + \underbrace{2C_{\frac{n-2}{2}} - 2}_{\text{type 2, 3}} \right) . \tag{1}$$

With a similar argument, following Definition 3.3 in the case n odd, we obtain:

$$|L_{m \times n}| = \left(\underbrace{2C_{\frac{n-1}{2}} - 2}_{\text{type 6, 7}} + \underbrace{C_{\frac{n-1}{2}}}_{\text{type 8}} \right)^{(m-2)} \left(\underbrace{2C_{\frac{n-2}{2}} - 2}_{\text{type 6, 7}} \right) . \tag{2}$$

Table 1 shows some values of the cardinality $L_{m \times n}$ with some estimated values of m and n up to 10. A careful analysis of them, reveals that if $n > m$, then the cardinality of $L_{m \times n}$ is greater then the cardinality of $L_{n \times m}$. So, fixed n and m, in the case $n > m$, it is convenient to define the matrices by columns (according to Definition 3.1 and 3.3) in order to obtain the set with the highest cardinality. In other words, if the columns are longer than the rows, then Dyck words should be used to construct the columns of the matrices of the set.

5. A comparison with $S_{m \times n}^{(k)}$

In [6] a remarkable family of non-overlapping matrices was defined. Here we briefly recall its construction.

Definition 5.1. Let $3 \leq k \leq \lfloor \frac{n}{2} \rfloor$ and $m \geq 2$. We denote by $\mathcal{S}_{m \times n}^{(k)} \subset \mathcal{M}_{m \times n}$ the set of the matrices $A = (a_{i,j})$ satisfying the following conditions:

- $A_1 = 1^{k-1}0w_110^{k-1}$, where $v_1 = 0w_11$ is a binary string of length $n - 2k + 2$ avoiding both 0^k and 1^k (explicitly, the string v_1 can contain neither k consecutive 0's nor k consecutive 1's);

- for $i = 2, \ldots, m - 1$, $A_i = w_i0 = v_i$, where v_i is a binary string of length n avoiding both 0^k and 1^k;

- $A_m = 1^k v_m 0^k$, where v_m is a binary string of length $n - 2k$ avoiding both 0^k and 1^k.

  (With A_1, A_i and A_m we denote the first, the i-th and the m-th row of the matrix A.)

$m\backslash n$	4	5	6	7	8	9	10
2	1	2	6	8	21	26	67
3	2	4	54	104	630	1040	$6.2 \cdot 10^3$
4	4	8	486	1352	$1.9 \cdot 10^4$	$4.1 \cdot 10^4$	$5.9 \cdot 10^5$
5	8	16	4374	$1.7 \cdot 10^4$	$5.7 \cdot 10^5$	$1.6 \cdot 10^6$	$5.5 \cdot 10^7$
6	16	32	$3.9 \cdot 10^4$	$2.2 \cdot 10^5$	$1.7 \cdot 10^7$	$6.6 \cdot 10^7$	$5.2 \cdot 10^9$
7	32	64	$3.5 \cdot 10^5$	$3.0 \cdot 10^6$	$5.1 \cdot 10^8$	$2.7 \cdot 10^9$	$4.9 \cdot 10^{11}$
8	64	128	$3.1 \cdot 10^6$	$3.8 \cdot 10^7$	$1.5 \cdot 10^{10}$	$1.1 \cdot 10^{11}$	$4.6 \cdot 10^{13}$
9	128	256	$2.8 \cdot 10^7$	$5.0 \cdot 10^8$	$4.6 \cdot 10^{11}$	$4.2 \cdot 10^{12}$	$4.3 \cdot 10^{15}$
10	256	512	$2.6 \cdot 10^8$	$6.5 \cdot 10^9$	$1.4 \cdot 10^{13}$	$1.7 \cdot 10^{14}$	$4.1 \cdot 10^{17}$

Table 1: Some values of the cardinality of $L_{m \times n}$

$$\begin{pmatrix} 1 & 1 & 0 & * & * & * & * & 1 & 0 & 0 \\ * & * & * & * & * & * & * & * & * & 0 \\ * & * & * & * & * & * & * & * & * & 0 \\ * & * & * & * & * & * & * & * & * & 0 \\ * & * & * & * & * & * & * & * & * & 0 \\ 1 & 1 & 1 & * & * & * & * & 0 & 0 & 0 \end{pmatrix}$$

Figure 6: The structure of the matrices in $\mathcal{S}_{6 \times 10}^{(3)}$.

In other words, some entries on the frame of a matrix in $\mathcal{S}_{m \times n}^{(k)}$ are fixed. For example, the matrices in $\mathcal{S}_{6 \times 10}^{(3)}$ are represented in Figure 6 where the generic entries $* \in \{0, 1\}$ are chosen so that the conditions of Definition 5.1 are satisfied.

The cardinality of $S_{m \times n}^{(k)}$ can be computed [6] in terms of k-generalized Fibonacci numbers as

$$|\mathcal{S}_{m \times n}^{(k)}| = \begin{cases} \left(f_{2k-1}^{(k-1)} \right)^{m-2} & \text{if } n = 2k \\ f_{n-2k-1}^{(k-1)} \cdot \left(f_{n-2k+1}^{(k-1)} + d_{n-2k+2}^{(k)} \right) \cdot \left(f_{n-1}^{(k-1)} \right)^{m-2} & \text{if } n > 2k \ , \end{cases} \tag{3}$$

where

$$d_\ell^{(k)} = \begin{cases} 1 & \text{if } (\ell \bmod k) = 0 \\ -1 & \text{if } (\ell \bmod k) = 1 \\ 0 & \text{if } (\ell \bmod k) \geq 2 \end{cases}$$

(explicitly $\left\{ d_\ell^{(k)} \right\}_{\ell \geq 0} = \left\{ 1, -1, \underbrace{0, 0, \ldots, 0}_{k-2}, 1, -1, \underbrace{0, 0, \ldots, 0}_{k-2}, 1, -1, 0, \ldots \right\}$), and ($k$-generalized Fibonacci numbers sequence)

$$f_\ell^{(k)} = \begin{cases} 2^\ell & \text{if } 0 \leq \ell \leq k - 1 \\ f_{\ell-1}^{(k)} + f_{\ell-2}^{(k)} + \ldots + f_{\ell-k}^{(k)} & \text{if } \ell \geq k \ . \end{cases}$$

The cardinality of $L_{m \times n}$ involves Catalan numbers (formulas (1) and (2)), while equation (3) involves the k-generalized Fibonacci numbers. Even if the first ones grow faster than the second ones, it is not true in general that $|L_{m \times n}| \geq |S_{m \times n}|$ for $n \geq n_0$, since $|L_{m \times n}|$ involves $C_{n/2}$ rather than C_n. Actually, an analysis of the two cardinalities reveals that $|L_{m \times n}| \leq |S_{m \times n}|$, for large values of n and m. Nevertheless, in the following we present a brief inspection of formulas (1) and (3), showing which are the cases when $|L_{m \times n}| \geq |S_{m \times n}|$.

We start by recalling that, using Stirling's approximation for $n!$, it is

$$C_n \sim \frac{4^n}{n^{3/2} \sqrt{\pi}} \ ,$$

so that, after some simple calculation, formula (1) boils down to

$$|L_{m \times n}| \simeq 3 \cdot 2^{m-3} \left(\sqrt{\frac{2}{\pi}} \right)^{m-1} \left(\frac{2^{n+1}}{n^{3/2}} \right)^{m-1} \ . \tag{4}$$

From [14] it is known that

$$f_n^{(k)} = \left\lfloor \frac{\alpha - 1}{2 + (k+1)(\alpha - 2)} \cdot \alpha^{n+1} + 1/2 \right\rfloor \tag{5}$$

where α is the unique real root of $x^k - x^{k-1} - \ldots - x - 1 = 0$ and, moreover,

$$2 - \frac{1}{3k} < \alpha < 2 \quad \text{if} \quad k \geq 4 . \tag{6}$$

First, we consider the case $n = 2k$, so that

$$|\mathcal{S}_{m \times n}^{(k)}| = \left(f_{2k-1}^{(k-1)} \right)^{m-2} .$$

Supposing $k \geq 4$, thanks to (5) and (6) and some calculation, we have:

$$f_{2k-1}^{(k-1)} = \left\lfloor \frac{(\alpha - 1)\alpha^{2k}}{2 + k(\alpha - 2)} + 1/2 \right\rfloor < \left\lfloor \frac{3 \cdot 2^{2k+1} + 5}{10} \right\rfloor < \frac{3}{5} \left(2^{2k} + 1 \right) ,$$

so that

$$|\mathcal{S}_{m \times n}^{(k)}| < \left(\frac{3}{5} \left(2^{2k} + 1 \right) \right)^{m-2} .$$

Moreover, being $n = 2k$, it is

$$|L_{m \times n}| \simeq \frac{3 \cdot 2^{2k(m-1)+m-3}}{\left(\sqrt{\pi} \right)^{m-1} k^{\frac{3(m-1)}{2}}} .$$

Solving

$$\left(\frac{3}{5} \left(2^{2k} + 1 \right) \right)^{m-2} < \frac{3 \cdot 2^{2k(m-1)+m-3}}{\left(\sqrt{\pi} \right)^{m-1} k^{\frac{3(m-1)}{2}}} ,$$

it is possible to prove that

$$|\mathcal{S}_{m \times n}^{(k)}| \leq |L_{m \times n}| \quad \text{if} \quad m < 1 + \frac{2k + \log_2 \frac{9}{10}}{\log_2 \left(\frac{3\sqrt{\pi}}{5} k^{3/2} \right)} ,$$

or, showing the number of columns n and after a little adjustment,

$$|\mathcal{S}_{m \times n}^{(k)}| \leq |L_{m \times n}| \quad \text{if} \quad m \leq 1 + \left\lfloor \frac{n-1}{\frac{3}{2} \log_2 n - 1} \right\rfloor .$$

This last formula shows that if the number of rows is opportunely bounded (by a quantity depending on the number of columns), then we get a larger number of non overlapping matrices.

If $n > 2k$, then

$$|\mathcal{S}_{m \times n}^{(k)}| = f_{n-2k-1}^{(k-1)} \cdot \left(f_{n-2k+1}^{(k-1)} + d_{n-2k+2}^{(k)} \right) \cdot \left(f_{n-1}^{(k-1)} \right)^{m-2} .$$

We have:

$$f_{n-2k-1}^{(k-1)} \leq \frac{3}{5} \left(2^{n-2k} + 1 \right) ,$$

$$f_{n-2k+1}^{(k-1)} \leq \frac{3}{5} \left(2^{n-2k+2} + 1 \right) ,$$

$$f_{n-1}^{(k-1)} \leq \frac{3}{5} \left(2^n + 1 \right) ,$$

$$d_{n-2k+2} \leq 1 ,$$

and

$$|\mathcal{S}_{m \times n}^{(k)}| \leq \frac{3}{5} \left(2^{n-2k} + 1 \right) \left(\frac{3}{5} \left(2^{n-2k+2} + 1 \right) \right) \left(\frac{3}{5} (2^n + 1) \right)^{m-2}$$

$$\leq 4 \left(\frac{3}{5} \right)^m \left(2^{n-2k} + 1 \right)^2 \left(2^n + 1 \right)^{m-2} .$$

The expression

$$4 \left(\frac{3}{5}\right)^m \left(2^{n-2k} + 1\right)^2 \left(2^n + 1\right)^{m-2} \leq 3 \cdot 2^{m-3} \left(\sqrt{\frac{2}{\pi}}\right)^{m-1} \left(\frac{2^{n+1}}{n^{3/2}}\right)^{m-1}$$

is satisfied when

$$m \leq \frac{4k - 8 + \log_2 3 - \log_2 \sqrt{\frac{2}{\pi}} - n + \frac{3}{2} \log_2 n}{\log_2 \frac{3}{5} - 1 - \log_2 \sqrt{\frac{2}{\pi}} + \frac{3}{2} \log_2 n} = \theta(n) .$$

In other words, if $m \leq \theta(n)$, then $|\mathcal{S}_{m \times n}^{(k)}| \leq |L_{m \times n}|$. Since $m > 0$, also $\theta(n) > 0$ must hold. It is possible to prove that

$$\theta(n) > 0 \qquad \text{if} \quad 17 \leq n \leq 4k .$$

Concluding, if the number of columns is bounded ($\max\{17, 2k\} \leq n \leq 4k$), and the number of rows is not so large ($m \leq \theta(n)$), also in this case we get a larger number of non overlapping matrices.

6.   On the non-expandability of $L_{m \times n}$

As well as in the theory of cross-fix-free codes, an interesting study concerns their non-expandability. We recall the following definition, adapted to matrices.

Definition 6.1. Let N be a set of non-overlapping matrices and let U be the set of all self-non-overlapping matrices. N is *non-expandable* if for each $A \in U \setminus N$ there exists at least one matrix $B \in N$ such that A and B are not non-overlapping matrices.

On the contrary, the set N is *expandable* if there exists a matrix $A \in U \setminus N$ such that A and B are non-overlapping matrices, for any $B \in N$.

We will prove that $L_{m \times n}$ is not a non-expandable set of matrices. The reason why it is not, maybe lies on the fact that, for such kinds of sets, some constraints on the columns of the matrices are needed, as it seems to come out from [1]. The columns of the matrices of $L_{m \times n}$ are not subject to obvious constraints. We recall that $L_{m \times n}$ is constructed by using Dyck words which form the rows of the matrices (adding some suitable prefixes or suffixes). This does not induces explicit restrictions over the columns.

Fixed a set $L_{m \times n}$, denote A_1 the first row of each $A \in L_{m \times n}$. Suppose that there exist a binary strings X with the following properties:

1. X is not a `type 1, 2, 3, 4` or `type 5` string in the n even case (`type 6, 7` or `type 8` in the n odd case);

2. X and A_1 are non-overlapping strings.

We observe that, from the second property, X begins with 1 and ends with 0. Now, we define the matrix $Z \in M_{n \times n}$ as follows:

- $Z_1 = A_1$;

- $Z_i = X \quad$ for $i = 2, \ldots, m$.

It is not difficult to realize that $Z \notin L_{m \times n}$: its rows from the second to the last one do not match the requirements of Definitions 3.1 or 3.3. Moreover, Z is a self non-overlapping matrix since any prefix (suffix) of Z_1 (involved in any vertical or skew overlap) does not match any suffix (prefix) of $Z_i = X$ for the second property of the string X.

We conclude with the following

Proposition 6.2. $Z \cup L_{m \times n}$ *is a non-overlapping set.*

$$
A = \begin{array}{|ccc|ccccc|}
\hline
\multicolumn{8}{c}{\begin{array}{cccccccc} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{array}} \\
\hline
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\
\hline
\end{array} \qquad Z
$$

$$
Z = \begin{array}{|ccc|ccccc|}
\hline
\multicolumn{8}{c}{\begin{array}{cccccccc} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{array}} \\
\hline
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
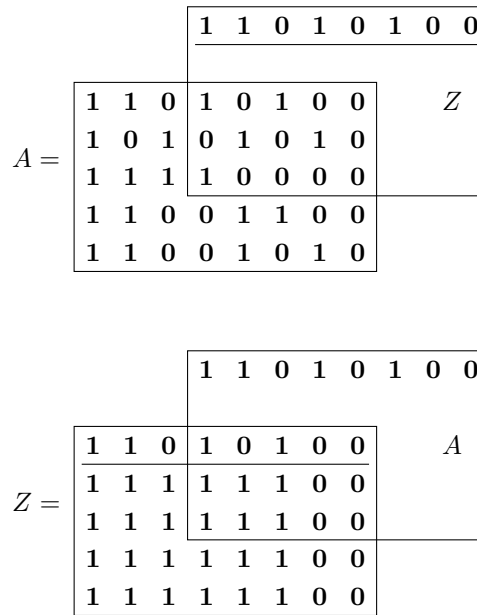\hline
\end{array} \qquad A
$$

Figure 7: Two skew overlaps between Z and an element $A \in L_{m \times n}$

Proof. Let $A \in L_{m \times n}$. Vertical overlaps between A and Z do not create control windows with identical corresponding entries since, for $i = 2, 3, \ldots, m$, it is $A_1 \neq Z_i$ and $Z_1 \neq A_i$ according to Definitions 3.1 (or 3.3) and the construction of Z. In case of skew overlaps (see Figure 7), we observe that a prefix or a suffix of $A_1 = Z_1$ is always involved in the control window and it is superimposed to a suffix or a prefix of X or A_i ($i = 1, 2, \ldots, m$).

If the involved rows are $A_1 = Z_1$ and X, no their prefix or suffix can match since A_1 and X are non-overlapping strings. If the involved row are Z_1 and A_i, then the proof is similar to the one of Proposition 3.2 (or 3.4) recalling that X ends with 0. □

The construction of the string X strictly depends on the first row A_1 of an element of $L_{m \times n}$. For example, if $A_1 = 11010100$, the string $X = 11111100$ can be used to construct the matrix Z in order to expand the set $L_{m \times n}$:

$$
Z = \begin{array}{|cccccccc|}
\hline
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
\hline
\end{array}
$$

Nevertheless, if $A_1 = 11100010$, then the same $X = 11111100$ can not be used to construct Z since its suffix 11100 matches the prefix of length 5 of A_1.

Unfortunately, once fixed a string A_1, we are not able to find a general criterion to construct a suitable string X.

7. Further developments

A first attempt to improve this study could start from the last hint of the preceding section. If $A_1 = 1u0$ with $u \in D_{n-2}$ is a type α Dyck word of length n, is there always a string X with the required properties? In the positive case, is there a general procedure to construct it? If this procedure were found, it could shed light on the possibility of developing the set $L_{m \times n}$ in order to make it a non-expandable set.

As far as the expandability is concerned, we notice that the types of rows of a matrix from the second to the last one depends on the first row A_1. For example, if $A_1 = 1^{n/2}0^{n/2}$, then a row A_i ($i \neq 1$) can be the following: $A_i = 1^4 v$ with $v \in D_{n-4}$ and $v \neq 1^{(n-4)/2}0^{(n-4)/2}$. This is a different row type with respect to type

`1, 2, 3, 4, 5`. It is not difficult to show that this new row and A_1 are non overlapping strings, so that it can be used to generate new matrices in $L_{m \times n}$. Consequently, we could ask to find all the possible strings which are non-overlapping with A_1. Again, is there a general and efficient procedure which, fixed A_1, lists all the non-overlapping strings with it?

Finally, an attempt aimed to the construction of variable-dimension matrices which are non-overlapping could be carried on, following up what already done [10] for strings set.

Acknowledgements

References

[1] M. Anselmo, D. Giammarresi and M. Madonia, *Non-expandable non-overlapping sets of pictures*, Theoret. Comput. Sci. 657 (2017), 127–136.

[2] D. Bajic, *On Construction of Cross-Bifix-Free Kernel Sets*, 2nd MCM COST 2100 TD(07)237(2007).

[3] D. Bajic and T. Loncar-Turukalo, *A simple suboptimal construction of cross-bifix-free codes*, Cryptogr. Commun. 6 (2014), 27–37.

[4] E. Barcucci, A. Bernini, S. Bilotta and R. Pinzani, *A 2D non-overlapping code over a q-ary alphabet*, Cryptogr. Commun. 10 (2018), 667–683.

[5] E. Barcucci, A. Bernini, S. Bilotta and R. Pinzani, *Cross-bifix-free sets in two dimensions*, Theoret. Comput. Sci. 664 (2017), 29–38.

[6] E. Barcucci, A. Bernini, S. Bilotta and R. Pinzani, *Non-overlapping matrices*, Theoret. Comput. Sci. 658 (2017), 36–45.

[7] E. Barcucci, S. Bilotta, E. Pergola, R. Pinzani and J. Succi, *Cross-bifix-free sets via Motzkin paths generation*, RAIRO - Theor. Inform. Appl. 50 (2016), 81–92.

[8] A. Bernini, S. Bilotta, R. Pinzani, A. Sabri and V. Vajinovszki, *Prefix partitioned Gray codes for particular cross-bifix-free sets*, Cryptogr. Commun. 6 (2014), 359–369.

[9] A. Bernini, S. Bilotta, R. Pinzani and V. Vajinovszki, *A Gray code for cross-bifix-free sets*, Math. Structures Comput. Sci. 27 (2017), 184–196.

[10] S. Bilotta, *Variable-Length Non-Overlapping Codes*, IEEE Trans. Inform. Theory 63 (2017), 6530–6537.

[11] S. Bilotta, E. Pergola and R. Pinzani, *A new approach to cross-bifix-free sets*, IEEE Trans. Inform. Theory 58 (2012), 4058–4063.

[12] S. Blackburn, *Non-overlapping codes*, IEEE Trans. Inform. Theory 61 (2015), 4890–4894.

[13] Y. M. Chee, H. M. Kiah, P. Purkayastha and C. Wang, *Cross-bifix-free codes within a constant factor of optimality*, IEEE Trans. Inform. Theory 59 (2013), 4668–4674.

[14] G. P. Dresden and Z. Du, *A Simplified Binet Formula for k-generalized Fibonacci Numbers*, J. Integer Seq. 17 (2014), Article 14.4.7.

[15] E. N. Gilbert and E. F. Moore, *Variable-length binary encodings*, Bell Syst. tech. 58 (1959), 933–967.

[16] S. W. Golomb, G. Basil and L. R. Welch, *Comma-Free Codes*, Canad. J. Math. 10 (1958), 202–209.

[17] P. T. Nielsen, *A note on bifix-free sequences*, IEEE Trans. Inform. Theory IT-29 (1973), 704–706.

[18] M. P. Schützenberger, *On an application of semi-group methods to some problem in coding*, IEEE Trans. Inform. Theory IT-2 (1956), 47–60.

[19] S. M. Hossein Tabatabaei Yazdi, H. M. Kiah, E. Garcia-Ruiz, J. Ma, H. Zhao and O. Milenkovich, *DNA-based storage: trends and methods*, IEEE T-MBMSC 1 (2015), 230–248.

[20] S. M. Hossein Tabatabaei Yazdi, Y. Yuan, J. Ma, H. Zhao and O. Milenkovich, *A rewritable, random-access DNA-based storage system*, Sci. Rep. 5 (2015), Article 14138.