On the rate of the binary puzzle

Putranto Hadi Utomo[a,b,*], Ruud Pellikaan[a]

[a]*Eindhoven University of Technology, P.O. Box 513. 5600MB Eindhoven*
[b]*Sebelas Maret University, Ir. Sutami No.36A Surakarta 57126*

Abstract

International Conference on Graph Theory and Information Security ICGTIS, 7-9 August 2017, Universitas Indonesia, Depok, Indonesia

A binary puzzle is a Sudoku-like puzzle consisting of a square array where each entry consists of a zero, a one, or a blank. Let $n \geq 4$ be an even integer. A solved binary puzzle is an $n \times n$ binary array that satisfies the following conditions: (1) no three consecutive ones and no three consecutive zeros in each row and each column; (2) the number of ones and zeros must be equal in each row and in each column; (3) there are no repeated rows and no repeated columns.

We break down the binary puzzle problem into three sub problems according to the constraints.
The problem of the first constraint is the constrained array problem, that is to find the rate of the arrays with the run-length restriction (1) on the columns and on the rows.
The second constraint problem deals with finding the rate of the arrays such that its columns and rows satisfy condition (2).
The third constraint problem is to find the rate of the arrays that have mutually distinct columns and mutually distinct rows.

*Corresponding author
Email addresses: p.h.utomo@tue.nl (Putranto Hadi Utomo), g.r.pellikaan@tue.nl (Ruud Pellikaan)

Keywords: constrained array, spectrum of an array, rate of a code, binary puzzle.
2010 MSC: 00-01, 99-00

1. Introduction

A binary puzzle is a Sudoku-like puzzle consisting of a square array where each entry consists of a zero, a one, or a blank. Let $n \geq 4$ be an even integer. A solved binary puzzle is an $n \times n$ binary array that satisfies the following conditions: (1) no three consecutive ones and no three consecutive zeros in each row and each column; (2) the number of ones and zeros must be equal in each row and in each column; (3) there are no repeated rows and no repeated columns.

We outline several mathematical problems related to the binary puzzle in [1]. We refer to [2, 3, 4] for further results on techniques in solving a binary puzzle. In this paper, we give an improvement of the result in [5], that is about the rate of the code based on the binary puzzle.

We break down the binary puzzle problem into three problems according to the constraints.
The problem of the first constraint is the constrained array problem, that is to find the rate of the arrays with the run-length restriction (1) on the columns and on the rows. We approach this problem with the theory of eigenvalues [6].
The second constraint problem, that is on the rate of the arrays that satisfy condition (2) is shown to have asymptotic rate equal to 1 [7]. Moreover, we also have a good approximation of the number of such arrays [8].
The third constraint problem, which is to find the rate of the arrays that have mutually distinct columns and have mutually distinct rows is approached using enumerable combinatorial techniques [9].

Let

$A_{m\times n} = \{X \in \mathbb{F}_2^{m\times n} \mid X \text{ satisfies (1) }\};$

$B_{m\times n} = \{X \in \mathbb{F}_2^{m\times n} \mid X \text{ satisfies (2) }\};$

$C_{m\times n} = \{X \in \mathbb{F}_2^{m\times n} \mid X \text{ satisfies (3) }\};$

$D_{m\times n} = \{X \in \mathbb{F}_2^{m\times n} \mid X \text{ satisfies (1), (2) and (3) }\};$

$E_{m\times n} = \{X \in \mathbb{F}_2^{m\times n} \mid X \text{ satisfies (1) and (2) for the columns and (1) for the rows.}\},$

where $\mathbb{F}_2^{m\times n}$ is the set of all $m \times n$ binary arrays.

Definition 1.1. *Let C be a code in Q^n, where the alphabet Q has q elements. Recall that the* (information) rate *of C is defined by*

$$R(C) = \frac{\log_q |C|}{n}.$$

Let $\mathcal{C}$ be a collection of codes over a fixed alphabet Q with q elements. The length of $C \subseteq Q^n$ is n and is denoted by $n(C)$. Suppose that the length $n(C)$ goes to infinity for $C \in \mathcal{C}$. The upper and lower asymptotic rate *or* capacity *of $\mathcal{C}$ are defined by*

$$\bar{R}(\mathcal{C}) = \limsup_{C\in\mathcal{C}} \frac{\log_q |C|}{n(C)}, \qquad \underline{R}(\mathcal{C}) = \liminf_{C\in\mathcal{C}} \frac{\log_q |C|}{n(C)}$$

and in case the limit exists, that means $\bar{R}(\mathcal{C}) = \underline{R}(\mathcal{C})$ it will be denoted by $R(\mathcal{C})$.

Let $\mathcal{A}_m$ be the collection of all codes $A_{m\times n}$ for fixed m and all positive integers n and let $\mathcal{A}$ be the collection of all codes $A_{m\times n}$ for all positive integers $m,\ n$. Similarly the collections $\mathcal{B}_m$, $\mathcal{C}_m$, $\mathcal{D}_m$, $\mathcal{E}_m$ and $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$ are defined.

In the sequel we need the following results

Proposition 1.2. *Let $\{r_{m,n}\}_{m,n=1}^{\infty}$ be an array of real numbers in $[0,1]$ such that for all l, m and $n \in \mathbb{N}$:*

$$(l+m)r_{l+m,n} \le l r_{l,n} + m r_{m,n} \tag{1}$$

$$(l+n)r_{m,l+n} \le l r_{m,l} + n r_{m,n} \tag{2}$$

Then $\rho = \lim_{m,n\to\infty}\{r_{m,n}\}$ exist and is equal to $\inf r_{m,n}$. In particular, $\rho \le r_{m,n}$ for all m, n.

Proof. See the appendix of [10]. □

Proposition 1.3. *Let* $\{r_{m,n}\}_{m,n=1}^{\infty}$ *be an array of real numbers in* $[0,1]$ *such that there are a constants* c *and* d *and for all* $m_1, m_2, n_1, n_2 \in \mathbb{N}$*:*

$$m_1 r_{m_1,n} + m_2 r_{m_2,n} \leq (m_1 + m_2 + c) r_{m_1+m_2+c,n} \tag{3}$$

$$n_1 r_{m,n_1} + n_2 r_{m,n_2} \leq (n_1 + n_2 + d) r_{m,n_1+n_2+d} \tag{4}$$

Then $\rho = \lim_{m,n\to\infty} \frac{mn}{(m+c)(n+d)} r_{m,n}$ *exist and is equal to* $\sup \frac{mn}{(m+c)(n+d)} r_{m,n}$. *In particular,* $\frac{mn}{(m+c)(n+d)} r_{m,n} \leq \rho$ *for all* m, n.

Corollary 1.4. *Let* $\{r_{m,n}\}_{m,n=1}^{\infty}$ *be an array of real numbers such that the inequalities (1) - (4) hold. Then* $\rho = \lim r_{m,n}$ *exist and for all* m, n*:*

$$\frac{mn}{(m+c)(n+d)} r_{m,n} \leq \rho \leq r_{m,n}$$

2. Runlength-constrained array

Run length constrained systems have an important role for data recording. The idea is to avoid patterns that are more prone to errors. For example, in a one-dimensional system of magnetic recording, a '1' is recorded as a change in the magnetic polarity. When reading, if the system detects a change in the magnetization, it will read as '1', otherwise it will read as '0'.

Since a magnetic recording is using a clock cycle and usually the clock is not perfect, it needs to re-synchronize with the data pattern and hence we do not want to have too many consecutive zeroes. On the other hand, if the ones are to close to each other, that is the magnetization poles could interfere with each other. Hence in this case, we want to have the ones more frequently, but not too often.

The theory of constrained sequences, that is for $m = 1$, is well established and uses the theory of graphs and the eigenvalues of the incidence matrix of the graph to give the solution of a linear recurrence. See [11, 12, 13]. An explicit formula for the number of such sequences of a given length n can be expressed in terms of the eigenvalues. The asymptotical rate is equal to $\log_q(\lambda_{max})$, where λ_{max} is the largest eigenvalue of the incidence matrix.

Remark 2.1. *Shannon [14] showed already that the following Fibonacci relation holds for $n \geq 1$:*

$$|A_{1\times(n+2)}| = |A_{1\times(n+1)}| + |A_{1\times n}|.$$

Asymptotically this gives

$$R(\mathcal{A}_1) = \lim_{n\to\infty} R(A_{1\times n}) = \log_2\left(\frac{1}{2} + \frac{1}{2}\sqrt{5}\right) = 0.69424\cdots$$

Furthermore $|A_{2\times n}| = |A_{1\times n}|^2$, since there is no constraint for the columns in case $m = 2$. Therefore also $R(\mathcal{A}_2) = \log_2\left(\frac{1}{2} + \frac{1}{2}\sqrt{5}\right)$.

Now, we want to find the capacity $R(\mathcal{A}_m)$ where the code $A_{m\times n}$ of $\mathcal{A}_m$ is viewed as a code of constrained sequences of length n in the alphabet $Q_m = \mathbb{F}_2^m$. The limit $R(\mathcal{A}_m)$ exists, since if we have $m \times 2$ array, we could always extent to $m \times 4$ by appending a compatible $m \times 2$ array. Therefore, there exist a linear recurrence formula for the number of an arbitrary $m \times n$ array using the same idea as before. Let Γ_m be a graph with $A_{m\times 2}$ as set of vertices. A directed edge in Γ_m is from vertex X to vertex Y such that $(X|Y)$ is in $A_{m\times 4}$. So the degree of the characteristic polynomial is $|A_{m\times 1}|^2$, where $|A_{m\times 1}|$ is given by the Fibonacci sequence $2,\ 4,\ 6,\ 10, \ldots, 110$ for $m = 1, 2, 3, 4, \ldots, 9$. So in particular, the degree of the characteristic polynomial $p_m(\lambda)$ in case of $m = 9$ is equal to $110^2 = 12,100$. The largest eigenvalue λ_m of the incidence matrix of Γ_m, is not computed as the largest root of $p_m(\lambda)$, but approximated numerically by Rayleigh quotients [15], that is by the power method with entry scaling. This straight forward idea is due-able for $m \leq 9$ with capacities given in Table 1.

Proposition 2.2. *The limit $R(\mathcal{A})$ exist and for all m*

$$\frac{m}{m+2} R(\mathcal{A}_m) \leq R(\mathcal{A}) \leq R(\mathcal{A}_m).$$

Proof. We can split an array Z in $A_{(l+m)\times n}$ into two sub arrays X and Y in $A_{l\times n}$ and $A_{m\times n}$, respectively, where X consists of the first l rows of Z, and Y

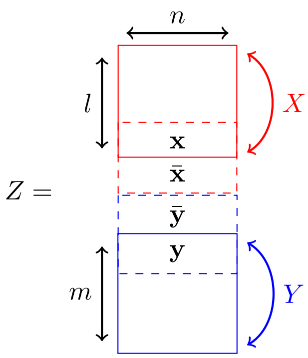

Figure 1: Illustration for the proof of Proposition 2.2

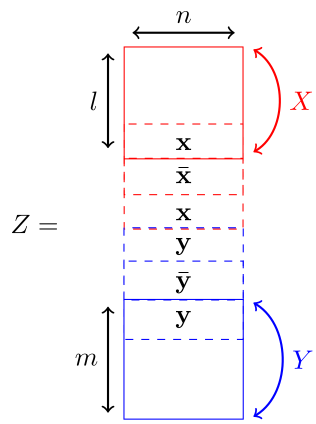

Figure 2: Illustration for the proof of Proposition **??**

of the last m rows of Z. Hence

$$|A_{(l+m)\times n}| \leq |A_{l\times n}| \cdot |A_{m\times n}|. \tag{5}$$

Following the idea in [16], suppose $X \in A_{l\times n}$ with last row $\mathbf{x}$ and $Y \in A_{m\times n}$ with first row $\mathbf{y}$. Let Z be the $(l+m+2)\times n$ array with X in the first l rows of Z, and Y in the last m rows of Z, and $\bar{\mathbf{x}}$ in row $l+1$ and $\bar{\mathbf{y}}$ in row $l+2$, where $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are the complementary rows of $\mathbf{x}$ and $\mathbf{y}$, respectively, by changing every zero in a one and every one in a zero. An illustration is shown in Figure 1. Then Z is an element of $A_{(l+m+2)\times n}$. Hence

$$|A_{l\times n}| \cdot |A_{m\times n}| \leq |A_{(l+m+2)\times n}|. \tag{6}$$

From 5, we have

$$\begin{aligned}
\log(|A_{(l+m)\times n}|) &\leq \log(|A_{l\times n}|) + \log(|A_{m\times n}|) \\
R(A_{(l+m)\times n}) &\leq \frac{R(A_{l\times n})}{1+m/l} + \frac{R(A_{m\times n})}{1+l/m} \\
(l+m)R(A_{(l+m)\times n}) &\leq lR(A_{l\times n}) + mR(A_{m\times n})
\end{aligned}$$

It also follow that $(l+n)R(A_{m\times(l+n)}) \leq lR(A_{m\times l}) + nR(A_{m\times n})$.

Using similar argument, we derive also the following inequalities from 6:

$$lR(A_{l\times n}) + mR(A_{m\times n})(l+m+2) \leq R(A_{(l+m+2)\times n})$$
$$lR(A_{m\times l}) + nR(A_{m\times n})(l+m+2) \leq R(A_{n\times(l+n+2)})$$

Then we apply the Corollary 1.4 with $r_{m,n} = R(A_{m\times n})$ to derive the result. $\square$

Remark 2.3. *Looking at the Table 1, we conjecture that $R(\mathcal{A}_m)$ is decreasing in m and $\frac{m}{m+2}R(\mathcal{A}_m)$ is increasing in m for $m > 4$.*

In contrast to the case for constrained sequences, there is so far no theory available that gives a closed formula for the capacity of constrained arrays.

m	λ_{max}	$R(\mathcal{A}_m)$	$\frac{m}{m+2}R(\mathcal{A}_m)$
1	1.6180	0.69424	0.23141
2	6.8541	0.69424	0.34712
3	11.7793	0.59303	0.35581
4	23.5755	0.56990	0.37993
5	44.3167	0.54697	0.39069
6	85.3928	0.53467	0.40100
7	162.9352	0.52486	0.40822
8	312.1198	0.51787	0.41429
9	596.9673	0.51230	0.41915

Table 1: Capacity of $\mathcal{A}_m$

3. Balancing a constrained array

Using the same idea as before we can also find $R(\mathcal{E}_{2m})$. The corresponding graph Λ_m has $E_{2m\times 2}$ as set of vertices, which is the subset of $A_{2m\times 2}$ of arrays that have balanced columns. So Λ_m is a subgraph of Γ_{2m}. The corresponding result is shown in Table 2. The number of vertices of Λ_5 is $7,056$.

Remark 3.1. *Using similar argument as in Proposition 2.2, we have that*

$$\frac{m}{m+4}R(\mathcal{E}_{2m}) \leq R(\mathcal{E})$$

m	λ_{max}	$R(\mathcal{E}_{2m})$	$\frac{m}{m+4}R(\mathcal{E}_{2m})$
1	2.6180	0.347120	0.06942
2	10.0125	0.415465	0.13848
3	29.3321	0.406200	0.17408
4	89.7965	0.405536	0.20276
5	284.9148	0.407719	0.22651

Table 2: Capacity of $\mathcal{E}_{2m}$

Remark 3.2. *For constrained sequences it is known that the capacity does not change if one adds the condition that the sequence is balanced [17, 18]. But, it is still an open problem whether a similar balancing argument for higher dimension as in [19] is true for our constrained arrays. From our research we only have*

$$A_{2m\times 2m} \cap B_{2m\times 2m} \subseteq E_{2m\times 2m} \subseteq A_{2m\times 2m}.$$

Remark 3.3. *Let $AB_{2m\times 2}$ be the collection of array satisfying (1) and (2) conditions. Now lets construct $F_{2m\times 2n}$ using the arrays in $AB_{2m\times 2}$ as the building blocks. Let $\mathcal{F}_{2m}$ be the collection of $F_{2m\times 2n}$ for fixed m and an arbitrary n. Clearly that $R(\mathcal{F}_{2m}) \leq R(\mathcal{A}_{2m} \cap \mathcal{B}_{2m})$.*

We also interested in finding the value of $R(\mathcal{F}_m)$. Since there is no restriction for the rows, $|\mathcal{F}_{2m\times 2n}|$ is equal to $|\mathcal{AB}_{2m\times 2}|^n$. Moreover, because the second column is the complement of the first column and the fact that the capacity does not change if we add the balanced condition to the sequence [17, 18], asymptotically the $R(\mathcal{F}_{2m})$ is equal to $R(\mathcal{A}_1)/2 = 0.34712$ as m goes to infinity. Table 3 give some result for m up to 8.

4. Weight-constrained array

Referring [20], balanced code have interesting applications such as detecting all unidirectional errors.

A sequence of even length is called balanced, if the number of zeros is equal to the number ones. The number of balanced sequences of length $2m$ is $\binom{2m}{m}$.

m	λ_{max}	$R(\mathcal{F}_{2m})$
1	2	0.2500000
2	6	0.3231203
3	14	0.3172795
4	34	0.3179664
5	84	0.3196158
6	208	0.3208516
7	518	0.3220288
8	1296	0.3231203

Table 3: Capacity of $\mathcal{F}_{2m}$

The rows of an $2 \times 2m$ array in $B_{2\times 2m}$ are complementary to each other. Hence $|B_{2\times 2m}| = \binom{2m}{m}$ and $R(\mathcal{B}_2) = \frac{1}{2}$.

The number of all $l \times 2m$ binary matrices such that all the rows are balanced is equal to $\binom{2m}{m}^l$. If X is an $l \times 2m$ binary matrix such that all its rows are balanced, then the $2l \times 2m$ binary matrix that is obtained by adding $\bar{X}$ under X has all its rows and columns balanced, where $\bar{X}$ is the complement of X. Hence

$$\binom{2m}{m}^l \leq |B_{2l\times 2m}| \leq \binom{2m}{m}^{2l}.$$

From these inequalities it follows that, asymptotically: $\frac{1}{2} \leq R(\mathcal{B}_{2l}) \leq 1$ for all l.

l	$R(\mathcal{B}_{2l})$
1	0.5
2	0.6462
3	0.7203
4	0.7259
5	0.7977
6	0.8210
7	0.8389

Table 4: Capacity of $\mathcal{B}_{2l}$

m	$R(B_{2m\times 2m})$
1	0.25000
2	0.40574
3	0.50502
4	0.57448
5	0.62546
6	0.66453
7	0.69550

Table 5: Rate of $B_{2m\times 2m}$

Four arbitrary elements of $B_{2m\times 2m}$ give an element of $B_{4m\times 4m}$. So $|B_{4m\times 4m}| \geq |B_{2m\times 2m}|^4$. Therefore $R(B_{2m\times 2m})$ is increasing in m.

Remark 4.1. *Referring to [8], we have a good approximation of $|B_{2m\times 2m}|$:*

$$|B_{2l\times 2m}| = \frac{\binom{2l}{l}^{2m}\binom{2m}{m}^{2l}}{\binom{4lm}{2lm}}\left(\frac{2l-1}{2l}\right)^{(2l-1)/2}\left(\frac{2m-1}{2m}\right)^{(2m-1)/2}\exp(\tfrac{1}{2}+o(1))$$

for $l, m \to \infty$. From this it follows that

$$R(\mathcal{B}_{2l}) = \frac{1}{2l}\log_2\binom{2l}{l}$$

and therefore $R(\mathcal{B}) = 1$, which was already shown in [7].

Proposition 4.2. *Let $b_{2l\times 2m} = |B_{2l\times 2m}|$. The exact number of $b_{2l\times 2m}$ can be obtained recursively as follow.*

$$b_{l\times(m+n)}(\mathbf{i};[\mathbf{j}\ \mathbf{k}]) = \sum_{\mathbf{x}+\mathbf{y}=\mathbf{i}} b_{l\times m}(\mathbf{x};\mathbf{j})\cdot b_{l\times n}(\mathbf{y};\mathbf{k}),$$

where $\mathbf{k} = (k_1,\ldots,k_n)$, $\mathbf{x} = (x_1,\ldots,x_l)$ and $\mathbf{y} = (y_1,\ldots,y_l)$. Similar formula also hold for $b_{(l+m)\times n}([ij];k)$.

Proof. Let $\mathbf{i} = (i_1,\ldots,i_l)$ and $\mathbf{j} = (j_i,\ldots,j_m)$. Denote by $b_{l\times m}(\mathbf{i};\mathbf{j})$ the collection of all $l\times m$ arrays with 0/1's such that there are i_s ones in row s, for all $s = 1,\ldots,l$, and there are j_t ones in column t, for all $t = 1,\ldots,m$. Then $b_{2l\times 2m} = b_{2l\times 2m}(\mathbf{i};\mathbf{j})$ where $i_s = m$ for all $s = 1,\ldots,2l$ and $j_t = l$ for all $t = 1,\ldots,2m$

Suppose we know the number of $b_{l\times m}(\mathbf{x};\mathbf{j})$ and $b_{l\times m}(\mathbf{y};\mathbf{j})$ for all $\mathbf{x}$ and $\mathbf{y}$. It is clear that $b_{l\times(m+n)}(\mathbf{i};[\mathbf{j}\ \mathbf{k}])$ is equal to summation of $b_{l\times m}(\mathbf{x};\mathbf{j})\cdot b_{l\times m}(\mathbf{y};\mathbf{j})$ for all possible $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x}+\mathbf{y} = \mathbf{i}$. $\square$

5. Mutually distinct vectors

In order to have a better approximation of the rate of binary puzzle, we look at the third constraint .

Consider $\tilde{C}_{m\times n}$, the set of all $m\times n$ binary arrays that have mutually distinct columns. Let $Q_m = \mathbb{F}_2^m$. Then $\tilde{C}_{m\times n}$ can de identified with all words in Q_m^n with mutually distinct entries. Hence $|\tilde{C}_{m\times n}| = 2^m(2^m-1)\cdots(2^m-n+1)$. Now $C_{m\times n}$ is a subcode of $\tilde{C}_{m\times n}$. Therefore

$$|C_{m\times n}| \leq 2^m(2^m-1)\cdots(2^m-n+1).$$

In particular $|C_{m\times n}| = 0$ for all $n > 2^m$. So $R(\mathcal{C}_m) = 0$. Therefore $\underline{R}(\mathcal{C}) = 0$.

Let X in $C_{m\times m}$. Then all the rows of X are distinct and all its columns are distinct. We can extend X to an $(m+1)\times(m+1)$ array in $C_{(m+1)\times(m+1)}$ by appending a column y and a row z to X, such that still all rows are distinct and all columns are distinct. Then y is an $m\times 1$ array and z is an $(m+1)\times 1$ array. There are $2^m - m$ possibilities for y and $2^{m+1}-m$ possibilities for z. Hence, if $m=n$, we have

$$|C_{(m+1)\times(m+1)}| \geq |C_{m\times m}|\cdot(2^{2m+1} - 3m2^m + m^2).$$

Proposition 5.1. *The following relation hold:*

$$R(C_{(m+l)\times(m+l)}) \geq \frac{m^2}{(m+l)^2}.$$

Therefore $\lim_{m\to\infty} R(C_{m\times m}) = 1$, *since* $\lim_{m\to\infty}\frac{l}{m} = 0$.
Moreover, $\bar{R}(\mathcal{C}) = 1$ *is not equal to* $\underline{R}(\mathcal{C}) = 0$*, and* $R(\mathcal{C})$ *is not defined.*

Proof. Suppose we have an arbitrary $m\times m$ binary array, say X. Let $l = \lceil \log_2(m)\rceil + 1$. Let Y be an $l\times m$ binary array such that all its columns are mutually distinct and have weight not equal to one. Such an array exists, since $m \leq 2^l - l$. Then $(Y|I_l)$ is an $l\times(m+l)$ array such that all its columns are mutually distinct. Let Z be the $(m+l)\times(m+l)$ array such that X is the upper $m\times m$ subarray of Z in the first m rows and columns, with $(Y|I_l)$ in its last l rows and $(Y|I_l)^T$ in its last l columns. Then Z is an array with mutually distinct columns and mutually distinct rows. Hence $|C_{(m+l)\times(m+l)}| \geq 2^{m^2}$. $\square$

A *partition* of a set of m elements is a collection of non-empty subsets that are mutually disjoint and their union is the whole set. The collection of all partitions of t non-empty subsets of the set of $\{1, \dots, m\}$ is denoted by $\mathcal{S}(m,t)$. The number $|\mathcal{S}(m,t)|$ is called the *Stirling number* of the *second kind* and is denoted by $S(m,t)$. Then $S(m,0) = 0$, $S(m,1) = S(m,m) = 1$ and the following recurrence relation holds [9]:

$$S(m+1,t) = tS(m,t) + S(m,t-1).$$

The following explicit formula holds [21]:

$$S(m,t) = \frac{1}{t!}\sum_{j=0}^{t}(-1)^{t-j}\binom{t}{j}j^m.$$

Proposition 5.2. *The numbers* $|C_{m\times n}|$ *satisfy the following recurrence:*

$$\sum_{t=1}^{m}|C_{t\times n}|\cdot S(m,t) = 2^m(2^m-1)\cdots(2^m-n+1).$$

Proof. Let $C^t_{m\times n}$ be the collection of all $m\times n$ binary arrays that have mutually distinct columns and have exactly t mutually distinct rows. Then in particular $|C^1_{m\times n}| = |C_{1\times n}| = |\tilde{C}_{1\times n}|$ and $|C^m_{m\times n}| = |C_{m\times n}|$.

Since the $C^k_{m\times n}$ are mutually distinct for $k = 1, \dots, m$, and give a partitioning of $\tilde{C}_{m\times n}$, we have that $\sum_{t=1}^m |C^t_{m\times n}| = |\tilde{C}_{m\times n}|$.

We have seen before that $|\tilde{C}_{m\times n}| = 2^m(2^m-1)\cdots(2^m-n+1)$.

Let $X \in C^t_{m\times n}$ and let $\mathbf{x}_i$ be the i-th row of X. Define the sequence $\mathbf{y}_1, \dots, \mathbf{y}_t$ by induction as follows $\mathbf{y}_1 = \mathbf{x}_1$. Suppose that $\mathbf{y}_1, \dots, \mathbf{y}_i$ are defined. Then $\mathbf{y}_{i+1}$ is the first row in X that is distinct from $\mathbf{y}_1, \dots, \mathbf{y}_i$. Then $\mathbf{y}_1, \dots, \mathbf{y}_t$ are mutually distinct and give a $t\times n$ array $Y \in C_{t\times n}$ such that $\mathbf{y}_i$ is the i-th row of Y. Let $I_j = \{i | \mathbf{x}_i = \mathbf{y}_j$ for $j = 1, \dots, t$. Then $\mathcal{I} = \{I_1, \dots, I_t\}$ is a partitioning of $\{1, \dots, m\}$ with t non-empty subsets.

Conversely, let Y be in $C_{t\times n}$ with rows $\mathbf{y}_1, \dots, \mathbf{y}_t$. Let $\mathcal{I} = \{I_1, \dots, I_t\}$ be a partitioning of $\{1, \dots, m\}$ with t non-empty subsets. Without loss of generality we may reorder $I_1, \dots, I_t$ such that $1 \in I_1$ and the minimal $j \in \{1, \dots, m\}$ that is not in $I_1 \cup \dots \cup I_i$, is in I_{i+1} for all $i < t$. Let X be the $m\times n$ matrix such that the i-row of X is equal to $\mathbf{y}_j$ is $i \in I_j$. Then $X \in C^t_{m\times n}$.

In this way we have obtained a bijection between $C^t_{m\times n}$ and $C_{t\times n} \times \mathcal{S}(m,t)$. $\square$

m	$R(C_{m\times m})$
1	1.00000
3	0.89382
5	0.96765
8	0.99527
11	0.99936
12	0.99967
15	0.99995
17	0.99998
25	0.99999

Table 6: Rate of $C_{m\times m}$

6. Conclusion: The rate of binary puzzle

Proposition 6.1. *The asymptotic rate of $D_{2m\times 2m}$ is the same as $A_{2m\times 2m} \cap B_{2m\times 2m}$ as m goes to infinity*

Proof. Following the work by De Biasi [22] and similar to our proof in Proposition 5.1, suppose we have a particular array X in $E_{2m\times 2m}$. Define the blocks $I_2 = \left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$ and $O_2 = \left[\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right]$. Let $l = 2\lceil\frac{\log_2(2m)}{4}\rceil$. Let Y be an array of blocks I_2 and O_2 such that it has $m \times l$ blocks and all its rows are mutually distinct. From the construction, the array will not break the first and second condition. Now, let Z be the $2(m+l)\times 2(m+l)$ array such that X be the upper right subarray of Z, Y be the upper right subarray of Z, Y^T be the lower left subarray of Z, and for the lower right part of Z, put an identity like array with I_2 in the diagonal entry and O_2 otherwise. Then Z is an array satisfying all the three conditions. The result follows from the fact that l is relatively small compared to m. $\square$

From Proposition 2.2, Remark 3.3 and Table 1 we conclude that the rate of $D_{2m\times 2m}$ is bounded by

$$0.34712 \leq R(\mathcal{D}) \leq 0.51230$$

By brute force computation and the idea of building blocks, we could give only a few values in Table 7.

It is also an interesting question to know weather $R(\mathcal{D}) = R(\mathcal{F})$. Therefore more research is needed to determine $R(\mathcal{D})$, that is the capacity of Binary Puzzles.

m	$R(D_{2m\times 2m})$
1	0.2500
2	0.3856
3	0.3337
4	0.3432

Table 7: Rate of $D_{2m\times 2m}$

.

7. Acknowledgement

Parts of this paper have been presented at IICMA 2015 [1] and ICGTIS 2017 [23]

References

[1] P. Utomo, R. Pellikaan, Binary Puzzles as an Erasure Decoding Problem, in: J. Roland, F. Horlin (Eds.), Proceedings of the 36th WIC Symposium on Information Theory in the Benelux, 2015, pp. 129–134, `http://www.w-i-c.org/proceedings/proceedings_SITB2015.pdf`.

[2] P. H. Utomo, R. H. Makarim, Solving a binary puzzle, Mathematics in Computer Science 11 (2017) 515 – 526.

[3] P. Utomo, Satisfiability modulo theory and binary puzzle, Journal of Physics: Conference Series 855 (1) (2017) 012057.

[4] P. Utomo, R. Pellikaan, Binary puzzles as a SAT problem, in: R. Heusdens, J. H. Weber (Eds.), Proc. 2017 Symposium on Information Theory and Signal Processing in the Benelux, 2017, pp. 223–229, `http://www.w-i-c.org/proceedings/proceedings_SITB2017.pdf`.

[5] P. Utomo, R. Pellikaan, On the rate of constrained arrays, in: Proceedings IndoMS International Conference on Mathematics and its Applications (IICMA), 2015, `http://www.win.tue.nl/~ruudp/paper/75.pdf`.

[6] F. Chatelin, Eigenvalues of matrices, John Wiley & Sons, Ltd., Chichester, 1993, with exercises by Mario Ahués and the author, Translated from the French and with additional material by Walter Ledermann.

[7] E. Ordentlich, R. M. Roth, Two-dimensional weight-constrained codes through enumeration bounds, IEEE Transactions on Information Theory 46 (4) (2000) 1292–1301.

[8] E. R. Canfield, B. D. McKay, Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums, the electronic journal of combinatorics 12 (1) (2005) R29.

[9] R. P. Stanley, Enumerative combinatorics, Cambridge Studies in Advanced Mathematics ; 49, 62, Cambridge University Press, Cambridge ; New York, 1997.

[10] A. Kato, K. Zeger, On the capacity of two-dimensional run-length constrained channels, IEEE Transactions on Information Theory 45 (5) (1999) 1527–1540. `doi:10.1109/18.771152`.

[11] H. Hollmann, Modulation Codes, Philips Electronics N.V., 1996, `http://books.google.nl/books?id=7w1BAAAACAAJ`.

[12] K. S. Immink, P. Siegel, J. Wolf, Codes for digital recorders, IEEE Transactions on Information Theory 44 (6) (1998) 2260–2299.

[13] D. Lind, B. Marcus, An introduction to symbolic dynamics and coding, Cambridge University Press, 1995.

[14] C. Shannon, A mathematical theory of communication, j-BELL-SYST-TECH-J 27 (3) (1948) 379–423.

[15] E. Louidor, B. H. Marcus, Improved lower bounds on capacities of symmetric 2d constraints using Rayleigh quotients, IEEE Transactions on Information Theory 56 (4) (2010) 1624–1639.

[16] W. Weeks, R. Blahut, The capacity and coding gain of certain checkerboard codes, in: Proceedings. 1998 IEEE International Symposium on Information Theory., 1998, p. 322.

[17] P. Henry, Zero disparity coding/decoding system, CA Patent 1,146,672. http://www.google.com/patents/CA1146672A1?cl=en (May 17 1983).

[18] D. Knuth, Efficient balanced codes, IEEE Transactions on Information Theory 32 (1) (1986) 51–53.

[19] A. Vardy, M. Blaum, P. Siegel, G. T. Sincerbox, Conservative arrays: multidimensional modulation codes for holographic recording, Information Theory, IEEE Transactions on 42 (1) (1996) 227–230.

[20] S. Al-Bassam, B. Bose, On balanced codes, IEEE Transactions on Information Theory 36 (2) (1990) 406–408.

[21] H. Sharp, Cardinality of finite topologies, Journal of Combinatorial Theory 5 (1) (1968) 82 – 86.

[22] M. D. Biasi, Binary puzzle is NP-complete, http://nearly42.org (July 2012).

[23] International Conference on Graph Theory and Information Security, http://www.math.ui.ac.id/icgtis2017, accessed: 2017-09-19.