On Encoding Symbol Degrees of Array BP-XOR Codes

Maura B. Paterson
Dept. Economics, Math. & Statistics
Birkbeck University of London
Email: m.paterson@bbk.ac.uk

Douglas R. Stinson
David R. Cheriton School of Computer Science
University of Waterloo, Canada
Email:dstinson@math.uwaterloo.ca

Yongge Wang
Dept. SIS, UNC Charlotte
Charlotte, NC 28223, USA
Email: yongge.wang@uncc.edu

***Abstract*—Low density parity check (LDPC) codes, LT codes and digital fountain techniques have received significant attention from both academics and industry in the past few years. By employing the underlying ideas of efficient Belief Propagation (BP) decoding process (also called iterative message passing decoding process) on binary erasure channels (BEC) in LDPC codes, Wang has recently introduced the concept of array BP-XOR codes and showed the necessary and sufficient conditions for MDS $[k+2,k]$ and $[n,2]$ array BP-XOR codes. In this paper, we analyze the encoding symbol degree requirements for array BP-XOR codes and present new necessary conditions for array BP-XOR codes. These new necessary conditions are used as a guideline for constructing several array BP-XOR codes and for presenting a complete characterization (necessary and sufficient conditions) of degree two array BP-XOR codes and for designing new edge-colored graphs. Meanwhile, these new necessary conditions are used to show that the codes by Feng, Deng, Bao, and Shen in IEEE Transactions on Computers are incorrect.**

I. Introduction

Low-density parity-check (LDPC) codes were invented by Gallager [11] in his PhD thesis. After being invented, they were largely forgotten and have been reinvented multiple times for the next 30 years (see, e.g., [30], [31], [1], [18], [19], [16], [29], [25], [21], [22], [20], [2]). For example, based on expander graph results by Lubotzky, Phillips and Sarnak [15] and Margulis [24], Sipser and Spielman [30], Spielman [31], Alon et al. [1], and others introduced asymptotically linear LDPC error-correcting and erasure codes. Luby et al. [18], [19] introduced LDPC Tornado codes, Luby [16] introduced LT-code, and Shokrollahi [29] introduced Raptor codes.

Array codes have been studied extensively for burst error correction in communication systems and storage systems (see, e.g., [3], [4], [5], [6], [8], [14], [36], [37]). Array codes are linear codes where information and parity data are placed in a two dimensional matrix array.

Formally, the array code is defined as follows: For fixed numbers n, k, t, l, and b where $n > \max\{k, t\}$, let $M = \{0,1\}^l$ be the message symbol set and $v_1, \cdots, v_{bk}$ be variables taking values from M, which are called information symbols. A t-erasure tolerating $[n,k]$ array code is a $b \times n$ matrix $\mathcal{C} = [a_{i,j}]_{1 \le i \le b, 1 \le j \le n}$ such that each encoding symbol $a_{i,j} \in \{0,1\}^l$ is the exclusive-or (XOR) of one or more information symbols from $v_1, \cdots, v_{bk}$ and $v_1, \cdots, v_{bk}$ could be recovered from any $n-t$ columns of the matrix. For an encoding symbol $a_{i,j} = v_{i_1} \oplus \cdots \oplus v_{i_\sigma}$, we call v_{i_j} $(1 \le j \le \sigma)$ a neighbor of $a_{i,j}$ and call σ the degree of $a_{i,j}$. A t-erasure tolerating $[n,k]$ $b \times n$ array code $\mathcal{C}$ is said to be maximum distance separable (MDS) if $k = n - t$. The $[n,k]$ array code $\mathcal{C}$ over the alphabet M can be considered as a linear code over the extension alphabet M^b of length n or a linear code over the alphabet M of length bn.

The Belief Propagation decoding process (also called message passing iterative decoding) for binary symmetric channels (BSC) is present in Gallager [11] and is also used in artificial intelligence community [27]. The BP decoding process for binary erasure channels (BEC) is described as follows:

> (Cf. [16], [17]) If there is at least one encoding symbol that has exactly one neighbor then the neighbor can be recovered immediately. The value of the recovered information symbol is XORed into any remaining encoding symbols that have this information symbol as a neighbor. The recovered information symbol is removed as a neighbor of these encoding symbols and the degree of each such encoding symbol is decreased by one to reflect this removal.

Wang [32], [33], [35] recently studied array codes that could be be decoded using BP decoding process: An $[n,k]$ array code $\mathcal{C} = [a_{i,j}]_{1 \le i \le b, 1 \le j \le n}$ is called a t-erasure tolerating $[n,k]$ array BP-XOR code if all information symbols $v_1, \cdots, v_{bk}$ can be recovered from any $n-t$ columns of the matrix using the BP-decoding process on the BEC.

In this paper, we analyze the encoding symbol degree requirements for array BP-XOR codes, present new necessary conditions for general array codes and array BP-XOR codes, and give a complete characterization of degree two BP- XOR codes. These necessary conditions are used as a guideline for constructing several array BP-XOR codes and the characterization of degree two BP-XOR codes are used to design new edge-colored graphs. Meanwhile, these necessary conditions are used to show that the codes by Feng, Deng, Bao, and Shen [9], [10] are incorrect.

The structure of the paper is as follows. Section II establishes the degree requirements for weakly systematic array codes. Section III proves necessary conditions for the existence of array BP-XOR codes. Section IV shows that the necessary conditions in Section III is sufficient for degree two encoding

symbol based array BP-XOR codes. Bounds for high degree encoding symbol based array BP-XOR codes are briefly discussed in Section V. Using the results in Section II, Section VI shows that the codes in [9] are incorrect.

II. Degree requirements for weakly systematic array codes

For each MDS $b \times n$ array code $\mathcal{C}$ (not necessarily a BP-XOR code) such that the original data could be recovered from any $k \leq n$ columns, let G_i be a $bk \times b$ binary matrix such that $(y_{i,1}, \cdots, y_{i,b}) = (x_1, \cdots, x_{bk})G_i$, where $(y_{i,1}, \cdots, y_{i,b})$ is the encoding symbols in the ith column of $\mathcal{C}$, $(x_1, \cdots, x_{bk})$ is the information symbols, and the addition of two strings in M is defined as the XOR on bits. In other words, we could consider G_i as the generator matrix for the ith column of $\mathcal{C}$. The generator matrix for $\mathcal{C}$ is defined as the $bk \times bn$ matrix $G_{\mathcal{C}} = [G_1, G_2, \cdots, G_n]$.

An array code $\mathcal{C}$ is called *systematic* if there exist $1 \leq i_1, \cdots, i_k \leq n$ such that $[G_{i_1}, \cdots, G_{i_k}]$ is the $kb \times kb$ identity matrix I_{kb}. An array code $\mathcal{C}$ is called *weakly systematic* if there exists a $kb \times kb$ permutation matrix P such that $G_{\mathcal{C}}P = [I_{kb}|A_{\mathcal{C}}]$ where $A_{\mathcal{C}}$ is a $kb \times (n-k)b$ binary matrix.

A $bt \times bn$ binary matrix matrix H is said to be a parity-check matrix of a $b \times n$ array code $\mathcal{C}$ if we have $H\mathbf{y}^T = 0$ where $\mathbf{y} = (a_{1,1}, \cdots, a_{b,1}, \cdots, a_{1,n}, \cdots a_{b,n})$, $\mathbf{x} = (v_1, \cdots, v_{bk})$. By [23], we have the following proposition.

Proposition 2.1: (MacWilliams and Sloane [23]) If $G_{\mathcal{C}} = [I_{kb}|A]$ is the generator matrix for a systematic array code $\mathcal{C}$, then $H_{\mathcal{C}} = [A^T|I_{(n-k)b}]$ is the parity check matrix for $\mathcal{C}$.

By Proposition 2.1, it is straightforward to get the following proposition.

Proposition 2.2: If the $G_{\mathcal{C}} = [I_{kb}|A]P^{-1}$ is the generator matrix for a weakly systematic array code $\mathcal{C}$, then $H_{\mathcal{C}} = [A^T|I_{(n-k)b}]P^T$ is the parity check matrix for $\mathcal{C}$.

For a weakly systematic $b \times n$ MDS array code $\mathcal{C}$ with generator matrix $G_{\mathcal{C}} = [G_1, G_2, \cdots, G_n] = [I_{kb}|A]P^{-1}$ and parity check matrix $H_{\mathcal{C}} = [A^T|I_{(n-k)b}]P^T$, the information symbols could be recovered from any k columns of encoding symbols in the array code $\mathcal{C}$. Thus for each $i \in [1, kb]$, there exist $j_1, \cdots, j_{n-k+1} \in [1, n]$ such that for each $j \in \{j_1, \cdots, j_{n-k+1}\}$, the ith row of G_j contains at least one non zero element.

The dual code of the weakly systematic $b \times n$ MDS array code $\mathcal{C}$ is a $b \times n$ MDS array code $\mathcal{C}^D$ with $H_{\mathcal{C}} = [A^T|I_{(n-k)b}]P^T$ as the generator matrix and all the information symbols could be recovered from any $n-k$ columns of encoding symbols in the array code $\mathcal{C}^D$. Thus it is straightforward to verify that each row of A^T should have at least k non zero elements. In other words, each column of A should have at least k non zero elements.

Combining the above discussion, we get the following Theorem 2.3. It should be noted that Blaum and Roth [6, page 52, Proposition 3.4] presented similar results for systematic array codes.

Theorem 2.3: For a weakly systematic $b \times n$ MDS array code $\mathcal{C}$ with generator matrix $G_{\mathcal{C}} = [G_1, G_2, \cdots, G_n] = [I_{kb}|A]P^{-1}$ and parity check matrix $H_{\mathcal{C}} = [A^T|I_{(n-k)b}]P^T$, each row of A contains at least $n-k$ non zero elements and each column of A contains at least k non zero elements.

The above discussion shows that for each weakly systematic MDS $b \times n$ array BP-XOR code $\mathcal{C}$, it contains either degree one encoding symbols or degree k' encoding symbols for $k' \geq k$. Our examples in Table VIII of Section IV show that the above requirements are not necessary for non weakly systematic array codes.

III. Necessary Conditions on degrees of array BP-XOR codes

Wang [33] showed the equivalence between edge-colored graphs and array BP-XOR codes with degree two encoding symbols. In particular, degree two encoding symbols are sufficient to construct $[n, 2]$ MDS $b \times n$ array BP-XOR codes. Generally, we are interested in $[n, k]$ MDS $b \times n$ array BP-XOR codes for any $k < n$.

For an $[n, k]$ MDS $b \times n$ array BP-XOR code, we assume that there are bk information symbols, each of which is a variable that takes value from $M = \{0, 1\}^l$. The following theorem provides a necessary condition for the existence of array BP-XOR codes.

Theorem 3.1: Let $\mathcal{C} = [a_{i,j}]_{1 \leq i \leq b, 1 \leq j \leq n}$ be an $[n, k]$ MDS $b \times n$ array BP-XOR code such that the degree of each encoding symbol $a_{i,j}$ is less than or equal to $\sigma < k + (k-1)/(b-1)$. Then we have

$$n \leq k + \sigma - 1 + \left\lfloor \frac{\sigma(\sigma-1)(b-1)}{(k-\sigma)b + \sigma - 1} \right\rfloor \tag{1}$$

Proof. By the fact that there are $n-k$ erasure columns, each information symbol must occur in at least $n-k+1$ columns. Since there are kb information symbols (data fragments) to encode, the total number of information symbol occurrences in the array BP-XOR code $\mathcal{C}$ is at least $kb(n-k+1)$.

In order for the BP decoding process to work, we must start from a degree one encoding symbol. Thus we need to have at least $n-k+1$ degree one encoding symbols in distinct columns of $\mathcal{C}$. This implies that we could use at most $bn - (n-k+1)$ cells to hold encoding symbols for degree two to σ. In other words, $\mathcal{C}$ contains at most $\sigma(bn - (n-k+1)) + n - k + 1$ occurrences of information symbols. By the above fact, we must have

$$kb(n-k+1) \leq \sigma(bn - (n-k+1)) + n - k + 1.$$

By rearranging the terms, we get

$$kbn - kb(k-1) \leq \sigma bn - (\sigma-1)(n-k+1).$$

If we move all terms to the right hand side and rewrite the inequality as variables of b and n, we get

$$k(k-1)b - ((k-\sigma)b + (\sigma-1))n + (\sigma-1)(k-1) \geq 0.$$

That is,

$$n((k-\sigma)b + \sigma - 1) \leq (k-1)(kb + \sigma - 1). \tag{2}$$

By $\sigma < k + (k-1)/(b-1)$, we have $(k-\sigma)b + \sigma - 1 > 0$. Since n must be an integer, (2) implies (3)

$$\begin{aligned} n &\leq \left\lfloor \frac{(k-1)(kb+\sigma-1)}{(k-\sigma)b+\sigma-1} \right\rfloor \\ &= \left\lfloor \frac{(k-\sigma)kb + k(\sigma-1) + (\sigma-1)kb - (\sigma-1)}{(k-\sigma)b+\sigma-1} \right\rfloor \\ &= k + \left\lfloor \frac{(\sigma-1)(kb-1)}{(k-\sigma)b+\sigma-1} \right\rfloor \\ &= k + \left\lfloor \frac{(\sigma-1)(kb-b\sigma+\sigma-1+b\sigma-\sigma)}{(k-\sigma)b+\sigma-1} \right\rfloor \\ &= k + \sigma - 1 + \left\lfloor \frac{\sigma(\sigma-1)(b-1)}{(k-\sigma)b+\sigma-1} \right\rfloor \end{aligned} \tag{3}$$

Thus (1) holds. □

It is easy to see that the hypotheses of Theorem 3.1 are satisfied if $k \geq \sigma \geq 2$. So we have the following corollary.

Corollary 3.2: Suppose that $k \geq \sigma \geq 2$. Then (1) holds.

Next, we observe that equation (1) can be strengthened if $\sigma > 2$.

Theorem 3.3: Suppose that $(k-\sigma)b + \sigma - 1 > 0$, $\sigma > 2$, and $\sigma(\sigma-1)(b-1)/((k-\sigma)b+\sigma-1)$ is an integer. Then equality cannot hold in (1).

Proof. If equality holds in (1), then the following conditions must be satisfied:

- There are $n-k+1$ encoding symbols having degree 1 and the remaining $bn-(n-k+1)$ encoding symbols all have degree σ.
- The encoding symbols of degree 1 occur in $n-k+1$ different columns of the array.

Suppose we choose k columns such that only one of these columns contains an encoding symbol of degree 1. Then within these k columns, all but one of the encoding symbols have degree 3 or greater. It therefore follows that the BP process cannot succeed. □

When $k = \sigma$, (1) can be simplified.

Corollary 3.4: 1) If $k = \sigma = 2$, then

$$n \leq 2b + 1. \tag{4}$$

2) If $k = \sigma > 2$, then

$$n \leq kb + k - 2. \tag{5}$$

Proof. The equation (4) follows from (1). The equation (5) follows from Theorem 3.3. □

As an example, the code in Table I shows that the equality can hold in (4).

TABLE I
ARRAY BP-XOR CODE FOR $b = 2, n = 5, k = 2, \sigma = 2$

v_1	v_2	v_3	v_4	$v_1 \oplus v_2$
$v_2 \oplus v_3$	$v_1 \oplus v_4$	$v_2 \oplus v_4$	$v_1 \oplus v_3$	$v_3 \oplus v_4$

IV. Degree two MDS array BP-XOR codes and edge-colored graphs

By Corollary 3.4 and Theorem 3.3, Table II lists the upper bounds of n for the existence of $[n,k]$ MDS array BP-XOR codes with $\sigma = 2$.

TABLE II
UPPER BOUNDS OF n FOR $[n,k]$ MDS ARRAY BP-XOR CODES WITH $\sigma = 2$

k	2	3	3	$[4,\infty]$
n	$2b+1$	4 if $b \leq 2$	5 if $b \geq 3$	$k+1$

In this section, we give a complete characterization of degree two MDS array BP-XOR codes by showing that the bounds in Table II are sufficient. We first describe the edge-colored graph model by Wang and Desmedt [34]. The reader should be reminded that the edge-colored graph model in [34] is slightly different from the edge-colored graph definition in most literatures. In most literatures, the coloring of the edges is required to meet the condition that no two adjacent edges have the same color. This condition is not required in the definition of [34].

Definition 4.1: (Wang and Desmedt [34]) An edge-colored graph is a tuple $G = (V, E, C, f)$, with V the node set, E the edge set, C the color set, and f a map from E onto C. For any set $Z \subseteq E$, let $f(Z) = \{f(e) : e \in Z\}$. The structure

$$\mathcal{Z}_{C,t} = \{Z : Z \subseteq E \text{ and } |f(Z)| \leq t\}.$$

is called a *t-color* adversary structure. Let $A, B \in V$ be distinct nodes of G. A, B are called $(t+1)$-*color connected* for $t \geq 1$ if for any color set $C_t \subseteq C$ of size t, there is a path p from A to B in G such that the edges on p do not contain any color in C_t. An edge-colored graph G is $(t+1)$-*color connected* if and only if for any two nodes A and B in G, they are $(t+1)$-color connected.

In [33], Wang showed the equivalence of degree two encoding symbol based array BP-XOR codes and edge-colored graphs.

A. $[n,2]$ MDS array BP-XOR codes with $\sigma = 2$ from [33]

By Theorem 3.1, a necessary condition for the existence of $[n,2]$ MDS array BP-XOR codes with $\sigma = 2$ is $n \leq 2b+1$. Wang [33] constructed $[n,2]$ MDS $b \times n$ array BP-XOR codes with $n = 2b+1$ using edge-colored graphs based on perfect one-factorization of complete graphs.

We first briefly review the construction of $[n,2]$ MDS array BP-XOR codes in Wang [33]. Let p be a prime number with $n \leq p$. Using perfect one-factorization of K_{p+1}, Wang [33] constructed the $(p-1)$-color connected edge-colored graph in Table III where edges in the i-th column have the color c_i.

The edge-colored graph in Table III is converted to the $b \times p$ array BP-XOR code in Table IV by mapping each edge to a degree two encoding symbol and removing the occurrence of node v_p, and the $[n,2]$ MDS $b \times n$ BP-XOR code is obtained by taking any of the n columns in Table IV, where $b = (p-1)/2$.

TABLE III
$(p-1)$-COLOR CONNECTED EDGE-COLORED GRAPHS

$\langle v_1, v_{p-1}\rangle$	$\cdots$	$\langle v_p, v_{p-2}\rangle$
$\langle v_2, v_{p-2}\rangle$	$\cdots$	$\langle v_1, v_{p-3}\rangle$
$\cdots$	$\cdots$	$\cdots$
$\langle v_{(p-1)/2}, v_{(p+1)/2}\rangle$	$\cdots$	$\langle v_{(p-3)/2}, v_{(p-1)/2}\rangle$

TABLE IV
$(p-1)/2 \times p$ BP-XOR CODE

$v_1 \oplus v_{p-1}$	$\cdots$	$v_{p-1} \oplus v_{p-3}$	v_{p-2}
$v_2 \oplus v_{p-2}$	$\cdots$	v_{p-4}	$v_1 \oplus v_{p-3}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$v_b \oplus v_{b+1}$	$\cdots$	$v_{b-2} \oplus v_{b-1}$	$v_{b-1} \oplus v_b$

In the following sections, we show the construction of degree two $[n,k]$ MDS array BP-XOR codes and the corresponding edge-colored graphs for $2 < k < n$ when such kind of codes exist.

B. $[n,k]$ MDS array BP-XOR codes with $\sigma = 2$ and $n = k+1$

Wang and Desmedt [34] constructed the 2-color connected edge-colored cycle graph in Table V. For $n = k+1$, the edge-colored graph in Table V could be used to obtain the $[n,k]$ MDS array BP-XOR codes with $\sigma = 2$ in Table VI.

TABLE V
2-COLORED CONNECTED EDGE-COLORED GRAPH

$\langle v_0, v_1\rangle$	$\langle v_1, v_2\rangle$	$\cdots$	$\langle v_{n-1}, v_n\rangle$	$\langle v_n, v_0\rangle$

TABLE VI
2-COLORED CONNECTED EDGE-COLORED GRAPH

v_1	$v_1 \oplus v_2$	$\cdots$	$v_{n-1} \oplus v_n$	v_n

Based on the construction in Wang and Desmedt [34], one can obtain general $[k+1,k]$ MDS $b \times n$ array BP-XOR codes with $\sigma = 2$ by gluing together the v_0 nodes of b copies of edge-colored cycle graphs. For the example of $b = 2$ and $n = 4$, the array code in Table VII is a $[4,3]$ MDS array BP-XOR code. The corresponding edge-colored graph is shown in Figure 1.

C. $[n,3]$ MDS array BP-XOR codes with $\sigma = 2$

By Theorem 2.3, there is no weakly systematic $[n,3]$ array BP-XOR codes for $\sigma = 2$. Theorem 3.1 shows that a necessary condition for the existence of $[n,3]$ MDS array BP-XOR codes with $\sigma = 2$ is $n = 4, b \geq 1$ or $n = 5, b \geq 3$.

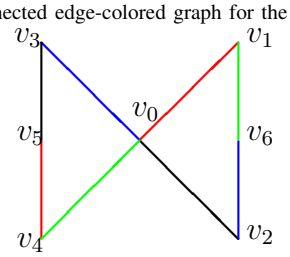
For the case of $n = 4, b \geq 1$, the codes in Section IV-B show that there exist $[4,3]$ MDS $b \times 4$ array BP-XOR codes.

For the case of $n = 5, b = 3$, Table VIII contains two $[5,3]$ MDS 3×5 array BP-XOR codes with $\sigma = 2$. The corresponding 3-color connected edge-colored graphs are shown in Figure 2 (removal of any two colors will not disconnect the graph).

TABLE VII
ARRAY BP-XOR CODE FOR $b = 2, n = 4, k = 3$

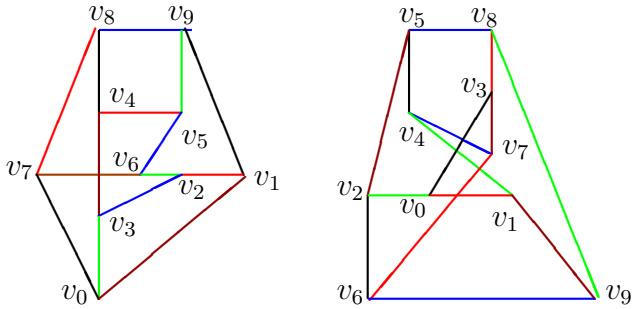
v_1	v_2	v_3	v_4
$v_5 \oplus v_4$	$v_5 \oplus v_3$	$v_6 \oplus v_2$	$v_6 \oplus v_1$

Fig. 1. 2-color connected edge-colored graph for the code in Table VII



The first graph in Figure 2 contains a four node cycle (v_4, v_5, v_9, v_8) while the second graph in Figure 2 does not contain any four node cycle. Thus the two $[5,3]$ MDS 3×5 array BP-XOR codes in Table VIII are not isomorphic.

Fig. 2. 3-color connected edged-colored graphs



For any integer $u \geq 1$, the graphs in Figure 2 could be used to construct 3-color connected edge-colored graphs with $9u+1$ nodes, 5 colors, and $15u$ edges by gluing together the v_0 nodes of u copies of the graphs in Figure 2.

D. $[n,k]$ MDS array BP-XOR codes with $\sigma = 2$ and $k \geq 4$

By Theorem 2.3, there is no weakly systematic $[n,k]$ array BP-XOR codes for $\sigma = 2$ and $k \geq 4$. Theorem 3.1 shows that a necessary condition for the existence of $[n,k]$ MDS array BP-XOR codes with $\sigma = 2$ and $k \geq 4$ is $n \leq k+1$. Since we also have $k < n$, it must be that $n = k+1$. The codes in Section IV-B show that there exist $[n,k]$ MDS $1 \times n$ array BP-XOR codes with $n = k+1$ and $\sigma = 2$.

TABLE VIII
TWO ARRAY BP-XOR CODES FOR $b = 3, n = 5, k = 3$

v_1	$v_1 \oplus v_2$	$v_2 \oplus v_3$	v_7	v_3
$v_3 \oplus v_4$	$v_4 \oplus v_5$	$v_5 \oplus v_6$	$v_9 \oplus v_1$	$v_2 \oplus v_6$
$v_6 \oplus v_7$	$v_7 \oplus v_8$	$v_8 \oplus v_9$	$v_4 \oplus v_8$	$v_9 \oplus v_5$

v_1	v_2	v_3	$v_2 \oplus v_5$	$v_5 \oplus v_8$
$v_6 \oplus v_7$	$v_1 \oplus v_4$	$v_4 \oplus v_5$	$v_3 \oplus v_7$	$v_4 \oplus v_7$
$v_3 \oplus v_8$	$v_8 \oplus v_9$	$v_2 \oplus v_6$	$v_1 \oplus v_9$	$v_6 \oplus v_9$

V. High degree MDS array BP-XOR codes

A. Upper bounds for higher degree MDS array BP-XOR codes

By Theorem 3.1, Theorem 3.3, and Corollary 3.4, Table IX lists the upper bounds of n for the existence of $[n,k]$ MDS array BP-XOR codes with $\sigma = 3,4,5$. It should be noted that the upper bounds in Table IX are obtained without any constraint on the values of b. In other words, we assume that b could take any values when necessary. When there are restrictions on the largest values that b could take, then Theorem 3.1 could be used to get stronger upper bounds on n. As an example, for $\sigma = 3, k = 4$, Theorem 3.1 gives $n \leq 12 - 18/(b+2)$. When $b \geq 17$, this gives $n \leq 11$ which is the bound in the table. However, for $b < 17$, the upper bound on n will be smaller than 11. We should also mention that the bounds in Table IX are upper bounds (necessary conditions). At present, it is not known whether any of these bounds could be achieved.

TABLE IX
Upper bounds of n for $[n,k]$ MDS array BP-XOR codes with $\sigma = 3,4,5$

$\sigma=3$		$\sigma=4$		$\sigma=5$	
k	n	k	n	k	n
3	$3b+1$	4	$4b+2$	5	$5b+3$
4	11	5	19	6	29
5	9	6	14	7	20
$[6,8]$	$k+3$	$[7,8]$	13	8	18
$[9,\infty]$	$k+2$	$[9,15]$	$k+4$	9	17
		$[16,\infty]$	$k+3$	$[10,11]$	$k+7$
				$[12,13]$	$k+6$
				$[14,24]$	$k+5$
				$[25,\infty]$	$k+4$

From Theorem 3.1, it is easy to show for any b and any $k \geq \sigma^2$ that the upper bound for the existence of $[n,k]$ MDS degree σ array BP-XOR codes is $n \leq k + \sigma - 1$.

B. Comparison with bounds for linear MDS codes

As mentioned in [6, Introduction], each $[n,k]$ MDS linear code over the finite field $GF(2^b)$ could be considered as an MDS $b \times n$ array code (not necessarily array BP-XOR code). However, the converse is not true (see Theorem 5.1 in Section V-C. Table X lists some known maximum value of n (see, e.g., [12], [28]) for the existence of $[n,k]$ MDS linear codes over $GF(2^b)$ with $b \geq 2$. For other values of $5 < k < 2^b - 1$, the well-known MDS conjecture states that the maximum value for n is $2^b + 1$. For $k = 2^b - 1$, the MDS conjecture states that the maximum value for n is $2^b + 2$. This conjecture was proved to be true for $b \leq 4$. Furthermore, Bush [7] showed that $n \leq 2^b + k - 1$ for $2 \leq k < 2^b$. This upper bound has been improved to $n \leq 2^b + k - 3$ for $k \geq 4$ in [13] (see also [26]). Comparing the analysis in the previous sections and the values in Table X, we see a big gap for the existence of MDS $b \times n$ array BP-XOR codes over $GF(2)$ and MDS linear codes over $GF(2^b)$.

TABLE X
Maximum value of n for $[n,k]$ MDS linear codes over $GF(2^b)$

k	2	3	4	5	$[2^b,\infty]$
n	2^b+1	2^b+2	2^b+1	2^b+2	$k+1$

C. Array codes (not necessarily array BP-XOR codes)

In the previous sections, we provided the upper bounds of n for $[n,k]$ MDS array BP-XOR codes with $\sigma \geq 3$. We do not know whether these bounds are sufficient. In the literature, there have been some constructions for high degree array codes though these codes are not BP-process decodable. For example, the authors of [14], [8] showed that if 2 is primitive in F_p, then one can construct $(p-1)/\sigma \times (p-1)$ array codes for $\sigma = 3$ and $\sigma = 4$ such that the information symbols could be recovered from any $k = \sigma$ columns of the encoding symbols.

As an example, we briefly describe the construction in [14], [8] . Let p be a prime such that 2 is primitive in F_p. In the finite field F_p, pick an element α of multiplicative order $\sigma (= 2,3,4)$ and an element β of multiplicative order $p-1$. Let $C_{-1} = \{0\}$ and

$$C_0 = \{\alpha^0, \alpha^1, \cdots, \alpha^{\sigma-1}\}$$

be the cyclic subgroup generated by α. For $1 \leq i < \frac{p-1}{\sigma}$, let $C_i = \beta^i C_0$ be the coset of C_0. Then $C_{-1}, C_0, \cdots, C_{\frac{p-1}{\sigma}-1}$ is a partition of $\{0,1,\cdots,p-1\}$. For $(i,j) \in [-1, \frac{p-1}{\sigma} - 1] \times [0, p-2]$, let

$$D'_{i,j} = \langle C_i + j \rangle_p$$

where $\langle C_i + j \rangle_p$ denotes the set that is obtained by adding j to the element of C_i modulo p.

It should be noted that exactly one of the sets $D'_{-1,j}$, $D_{0,j}$, $\cdots$, $D_{\frac{p-1}{\sigma}-1,j}$ contains $p-1$. For each $j \in [0, p-2]$, let $D_{0,j}, D_{1,j}, \cdots, D_{\frac{p-1}{\sigma}-1,j}$ be a list of the sets $D'_{i,j}$ such that $p - 1 \notin D'_{i,j}$.

Define the $(p-1)/\sigma \times (p-1)$ array code $\mathcal{C}_\sigma = [a_{i,j}]$ such that $a_{i,j}$ is the exclusive-or of all elements in $D_{i,j}$. It is shown in [14] that all of the information symbols could be recovered from any $k = \sigma$ columns of the encoding symbols of $\mathcal{C}$. For $p = 13$, $\sigma = 3$, $\alpha = 3$, and $\beta = 2$, we have

$$C_{-1} = \{0\} \quad C_0 = \{1,3,9\} \quad C_1 = \{2,6,5\}$$
$$C_2 = \{4,12,10\} \quad C_3 = \{8,7,11\}$$

Theorem 5.1: There is a $[12,4]$ MDS 3×12 array code $\mathcal{C}$ which is not a $[12,4]$ MDS linear code over $GF(2^3)$.

Proof. For the code from [14], [8] that we have just discussed, let $\sigma = 4$ and $p = 13$. Then we get a $[12,4]$ MDS 3×12 array code $\mathcal{C}$. By Table X, for $k = 4$ and $b = 3$, we have $n \leq 9$ for the existence of $[n,4]$ linear MDS code. Thus $\mathcal{C}$ is not a MDS linear code over $GF(2^3)$. □

VI. Incorrect Codes in [9]

Feng, Deng, Bao, and Shen [9], [10] introduced extended Reed-Solomon "MDS" array codes to tolerate three column faults [9] and multiple (≥ 4) column faults [10] respectively. In the following we show that the codes in [9] are incorrect.

Both [9] and [10] used similar techniques and analysis, and we believe the codes in [10] are incorrect as well. But we did not try to give the counter examples for [10].

Using circular permutation matrices as blocks, Vandermonde-like matrices are constructed as parity check matrices for extended Reed-Solomon codes to tolerate three columns faults in [9]. In particular, the authors used a sequence of Example 2.1 [9, pages 1072-1073], Examples 2.2 [9, pages 1073], Examples 2.3 [9, pages 1074], Examples 3.1 [9, pages 1075], and Examples 3.2 [9, pages 1076] to show how to construct a 4×8 array codes to tolerate three column erasure. After the code is constructed, a general decoding procedure is presented in [9, Section 4 on page 1076]. But the theoretical decoding procedure is not used to decode the code based on Examples 3.2 [9, pages 1076]. In the following, we show that the codes in Examples 3.2 [9, pages 1076] could not be decoded. Indeed, since all the codes in [9] do not meet the degree requirements for general array codes in Theorem 2.3, these codes will not decode.

The parity check matrix in Examples 3.2 [9, pages 1076] is defined as $H = [I|A]$ where I is $4 \cdot 3 \times 4 \cdot 3$ (i.e., 12×12) identity matrix and A is the following $4 \cdot 3 \times 4 \cdot 5$ (i.e., 12×20) matrix.

$$A = \left[\begin{array}{ccccc}
1000 & 1000 & 1000 & 1000 & 1000 \\
0100 & 0100 & 0100 & 0100 & 0100 \\
0010 & 0010 & 0010 & 0010 & 0010 \\
0001 & 0001 & 0001 & 0001 & 0001 \\
 & & & & \\
1000 & 0000 & 0001 & 0010 & 0100 \\
0100 & 1000 & 0000 & 0001 & 0010 \\
0010 & 0100 & 1000 & 0000 & 0001 \\
0001 & 0010 & 0100 & 1000 & 0000 \\
 & & & & \\
1000 & 0001 & 0100 & 0000 & 0010 \\
0100 & 0000 & 0010 & 1000 & 0001 \\
0010 & 1000 & 0001 & 0100 & 0000 \\
0001 & 0100 & 0000 & 0010 & 1000
\end{array}\right]$$

For the 4×8 array coded defined by the parity check matrix $H = [I|A]$, it is claimed that the code distance equals 4 (that is, $k = 5$) in [9]. That is, it will tolerate 3 erasure columns. By Theorem 2.3, each column of $H = [I|A]$ should contain at least 3 non zero elemenst. However, each of the columns in $7, 8, 9, 11, 14, 16, 17, 18$ contains 2 non-zero element. In other words, the code defined by the parity check matrix $H = [I|A]$ could not tolerate $k = 5$ erasure columns.

As an example, we show why the code could not be decoded. The code defined by the above parity check matrix $H = [I|A]$ could be represented in Table XI. It is straightforward to check that the variable v_7 only appears in columns $2, 6, 7$. Thus if we remove columns $2, 6$, and 7, then the variable v_7 could not be recovered from the remaining 5 columns (i.e., columns $1, 3, 4, 5, 8$). Similarly, each of the variables $v_8, v_9, v_{11}, v_{14}, v_{16}$, and v_{17} only appears in three columns. Thus these variables could not be recovered when the corresponding columns with their occurrences are missing.

TABLE XI
ARRAY CODE FOR $b = 4, n = 8, k = 5$ IN [9, EXAMPLES 3.2]

v_1	v_5	v_9	v_{13}	v_{17}	$v_1 \oplus v_5 \oplus v_9 \oplus v_{13} \oplus v_{17}$
v_2	v_6	v_{10}	v_{14}	v_{18}	$v_2 \oplus v_6 \oplus v_{10} \oplus v_{14} \oplus v_{18}$
v_3	v_7	v_{11}	v_{15}	v_{19}	$v_3 \oplus v_7 \oplus v_{11} \oplus v_{15} \oplus v_{19}$
v_4	v_8	v_{12}	v_{16}	v_{20}	$v_4 \oplus v_8 \oplus v_{12} \oplus v_{16} \oplus v_{20}$

$v_1 \oplus v_{12} \oplus v_{15} \oplus v_{18}$	$v_1 \oplus v_8 \oplus v_{10} \oplus v_{19}$
$v_2 \oplus v_5 \oplus v_{16} \oplus v_{19}$	$v_2 \oplus v_{11} \oplus v_{13} \oplus v_{20}$
$v_3 \oplus v_6 \oplus v_9 \oplus v_{20}$	$v_3 \oplus v_5 \oplus v_{12} \oplus v_{14}$
$v_4 \oplus v_7 \oplus v_{10} \oplus v_{13}$	$v_4 \oplus v_6 \oplus v_{15} \oplus v_{17}$

Similarly, the dual code of [9, Examples 3.2] in Table XI is a 4×8 array code which is shown in Table XII. It is also straightforward to check that the code in Table XII could not tolerate 5 column erasures. In other words, the original information symbols could not be recovered from any three columns. Specifically, each of the variables $v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}$, and v_{12} appears only in 5 columns. For example, v_5 only appears in columns $2, 4, 6, 7, 8$. Thus v_5 could not be recovered from columns $1, 3, 5$.

TABLE XII
DUAL ARRAY CODE OF [9, EXAMPLES 3.2] WITH $b = 4, n = 8, k = 3$

v_1	v_5	v_9	$v_1 \oplus v_5 \oplus v_9$	$v_1 \oplus v_6 \oplus v_{11}$
v_2	v_6	v_{10}	$v_2 \oplus v_6 \oplus v_{10}$	$v_2 \oplus v_7 \oplus v_{12}$
v_3	v_7	v_{11}	$v_3 \oplus v_7 \oplus v_{11}$	$v_3 \oplus v_8$
v_4	v_8	v_{12}	$v_4 \oplus v_8 \oplus v_{12}$	$v_4 \oplus v_9$

$v_1 \oplus v_7$	$v_1 \oplus v_8 \oplus v_{10}$	$v_1 \oplus v_{12}$
$v_2 \oplus v_8 \oplus v_9$	$v_2 \oplus v_{11}$	$v_2 \oplus v_5$
$v_3 \oplus v_{10}$	$v_3 \oplus v_5 \oplus v_{12}$	$v_3 \oplus v_6 \oplus v_9$
$v_4 \oplus v_5 \oplus v_{11}$	$v_4 \oplus v_6$	$v_4 \oplus v_7 \oplus v_{10}$

VII. CONCLUSION

In this paper, we presented new upper bounds for the existence of $[n, k]$ MDS array BP-XOR codes and showed that these bounds could be achieved for $k = 2$. It is an open question to show that these bounds are also achievable for other values of $k \in [3, n)$.

REFERENCES

[1] N. Alon, J. Edmonds, and M. Luby. Linear time erasure codes with nearly optimal recovery. In *Proc. 36th FOCS*, pages 512–. IEEE Computer Society, 1995.

[2] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: Turbo-codes. *Communications, IEEE Transactions on*, 44(10):1261–1271, 1996.

[3] M. Blaum, J. Brady, J. Bruck, and J. Menon. EVENODD: An efficient scheme for tolerating double disk failures in raid architectures. *IEEE Trans. Computers*, 44(2):192–202, 1995.

[4] M. Blaum, J. Bruck, and E. Vardy. MDS array codes with independent parity symbols. *IEEE Trans. on Information Theory*, 42:529–542, 1996.

[5] M. Blaum and R. M. Roth. New array codes for multiple phased burst correction. *IEEE Trans. on Information Theory*, 39(1):66–77, 1993.

[6] M. Blaum and R. M. Roth. On lowest-density MDS codes. *IEEE Trans. on Information Theory*, 45:46–59, 1999.

[7] K.A. Bush. Orthogonal arrays of index unity. *The Annals of Mathematical Statistics*, 23(3):426–434, 1952.

[8] Yuval Cassuto and Jehoshua Bruck. Cyclic lowest density mds array codes. *IEEE Trans. Inf. Theor.*, 55(4):1721–1729, April 2009.

[9] G.L. Feng, R.H. Deng, F. Bao, and J.C. Shen. New efficient MDS array codes for RAID. Part I. Reed-Solomon-like codes for tolerating three disk failures. *IEEE Trans. Computers*, 54(9):1071–1080, 2005.

[10] G.L. Feng, R.H. Deng, F. Bao, and J.C. Shen. New efficient MDS array codes for RAID. Part II. Rabin-like codes for tolerating multiple (≥ 4) disk failures. *IEEE Trans. Computers*, 54(12):1473–1483, 2005.

[11] R. G. Gallager. *Low density Parity Check Codes*. MIT Press, 1963.

[12] J.W.P. Hirschfeld, L. Storme, et al. The packing problem in statistics, coding theory and finite projective spaces: update 2001. *Developments in Mathematics*, 3:201–246, 2000.

[13] S. Kounias and CI Petros. Orthogonal arrays of strength three and four with index unity. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 228–240, 1975.

[14] E. Louidor and R. M. Roth. Lowest density MDS codes over extension alphabets. *IEEE Trans. Inf. Theor.*, 52(7):3186–3197, 2006.

[15] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[16] M. Luby. LT codes. In *Proc. FOCS*, pages 271–280, 2002.

[17] M. Luby, M. Mitzenmacher, M. Shokrollahi, and D. Spielman. Efficient erasure correcting codes. *IEEE Trans. Information Theory*, 47:569–584, 2001.

[18] M. Luby, M. Mitzenmacher, M. Shokrollahi, D. Spielman, and V. Stemann. Practical loss-resilient codes. In *Proc. 29th ACM STOC*, pages 150–159. ACM, 1997.

[19] M. G. Luby, M. Mitzenmacher, and M. Amin Shokrollahi. Analysis of random processes via and-or tree evaluation. In *In Proc. 9th Annual ACM-SIAM SODA*, pages 364–373, 1998.

[20] D. MacKay and R. Neal. Good codes based on very sparse matrices. *Cryptography and Coding*, pages 100–111, 1995.

[21] D.J.C. MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Trans. Infor. Theory*, 45(2):399–431, 1999.

[22] D.J.C. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[23] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes*. NH Pub. Company, 1978.

[24] G. Margulis. Explicit construction of concentrators. *Probl. Inform. transm.*, 9:325–332, 1975.

[25] G.A. Margulis. Explicit constructions of graphs without short cycles and low density codes. *Combinatorica*, 2(1):71–78, 1982.

[26] MinT. Bound for oas with index unity. *http://mint.sbg.ac.at/desc_CBoundT0.html*, 2012.

[27] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[28] R. Roth. *Introduction to coding theory*. Cambridge University Press, 2006.

[29] A. Shokrollahi. Raptor codes. *IEEE Trans. on Inform. Theory*, 52(6):2551–2567, 2006.

[30] M. Sipser and D. Spielman. Expander codes. *IEEE Trans. Infor. Theo.*, 42(6):1710–1722, 1996.

[31] D.A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Information Theory*, 42(6):1723–1731, 1996.

[32] Yongge Wang. Efficient LDPC code based secret sharing schemes and private data storage in cloud without encryption. Technical report, UNC Charlotte, 2012.

[33] Yongge Wang. LT codes for efficient and reliable distributed storage systems revisited. Technical report, UNC Charlotte, *submitted for publication*, 2012.

[34] Yongge Wang and Yvo Desmedt. Edge-colored graphs with applications to homogeneous faults. *Inf. Process. Lett.*, 111(13):634–641, 2011.

[35] Yongge Wang and Yvo Desmedt. Efficient secret sharing schemes achieving optimal information rate. Technical report, UNC Charlotte, 2012.

[36] L. Xu, V. Bohossian, J. Bruck, and D. Wagner. Low density mds codes and factors of complete graphs. *IEEE Trans. Inf. Theor.*, 45:1817–1826, 1998.

[37] L. Xu and J. Bruck. X-code: Mds array codes with optimal encoding. *IEEE Trans. on Information Theory*, 45:272–276, 1999.