# Asymptotically MDS Array BP-XOR Codes 

Suayb S. Arslan<br>Department of Computer Engineering, MEF University<br>34099, Maslak, Istanbul-Turkey<br>Email: arslans@mef.edu.tr


#### Abstract

Belief propagation (BP) on binary erasure channels (BEC) is a low complexity decoding algorithm that allows the recovery of message symbols based on bipartite graph pruning process. Recently, array XOR codes have attracted attention for storage systems due to their burst error recovery performance and easy arithmetic based on Exclusive OR (XOR)-only logic operations. Array BP-XOR codes are a subclass of array XOR codes that can be decoded using BP under BEC. Requiring the capability of BP-decodability in addition to Maximum Distance Separability (MDS) constraint on the code construction process is observed to put an upper bound on the achievable code blocklength, which leads to the code construction process to become a hard problem. In this study, we introduce asymptotically MDS array BP-XOR codes that are alternative to exact MDS array BPXOR codes to allow for easier code constructions while keeping the decoding complexity low with an asymptotically vanishing coding overhead. We finally provide a code construction method that is based on discrete geometry to fulfill the requirements of the class of asymptotically MDS array BP-XOR codes.


A full version of this paper is accessible at: https://arxiv. org/abs/1709.07949

## I. Introduction

Array codes are linear codes defined for two dimensional data structures in which both data and parity values organized in a matrix form. These codes are quite attractive candidates for burst error recovery in communication and distributed storage systems [1] and provide data reliability with optimal time/space consumption while maintaining Maximum Distance Separability (MDS) constraint in the code construction process. Moreover, a great deal of work has been done for these codes [2], [3] to secure simpler and low-complexity math quite desirable from the implementation point of view.

Typically, any linear code can be represented using a bipartite graph either using the parity check matrix or the generator matrix of the code [4]. Using the generator matrix representation, the corresponding bipartite graph has two types of nodes: Nodes that are used to decode (check or coded nodes) and nodes that are decoded (information nodes). Nodes in bipartite graph representation are connected with edges to represent node adjacency. The neighbors of node $j$ (neighbor set), denoted by $\mathcal{N}_{j}$, is the set of all nodes connected to node $j$. The cardinality of the neighbour set is called the degree of node $j$. The Belief Propagation (BP) algorithm a.k.a. message passing algorithm is an iterative process (updating nodes and edges) to decode data from coded nodes over symmetric erasure channels using the bipartite representation of the code. At the onset of the BP algorithm, we begin by setting all the contents of information nodes to NULL that
need to be decoded. Then, we look for a degree-one coded node and copy the content to its neighbor information node by replacing NULL. Next, we update all the coded nodes that are connected to the this neighbor and eliminate the edges that established neighborhood relationship. This completes the first step, and in the next iteration we continue applying the same methodology until there remains no information node with NULL content. If algorithm stops prematurely during iteration, we claim a decoding failure, otherwise we report a decoding success.

Array codes have recently been studied under BP decoding [5] and useful upper bounds are derived in [6] that theoretically establishes the relationship between the blocklength (and hence the rate of the code), decodability and sparsity of the generator matrix which is directly related to the encoding/decoding computation complexity of the code. In this study, we shall demonstrate by relaxing the MDS constraint on the code construction process, we shall be able to dramatically relax the previously found bounds on the code blocklength [6] while keeping low complexity BP algorithm successfully decode the whole data block. Such an observation shall yield easier and more powerful code constructions. For instance, we shall consider one of the discrete geometry based code constructions based on Mojette transform [7] that are recently studied within the context of low density parity check codes and are shown to reduce the node repair complexity [8].
The rest of the paper is organized as follows. In Section II, we provide the basics of array MDS BP-XOR codes and give some known results as well as state the main result of the paper. In Section III, we provide a discrete geometry construction of an asymptotically-MDS array BP-XOR codes. In Section IV, we validate our theoretical results by numerically plotting rate and code blocklength for discrete geometry construction. Finally, we conclude our paper in Section V.

## II. Asymptotically MDS Array BP-XOR Codes

Before defining the class of Asymptotically MDS (AMDS) array BP-XOR codes, let us provide the conventional definition of MDS BP-XOR codes using the notation of reference [6].

## A. Background

Let $l$ be the symbol size in bits and $M=\{0,1\}^{l}$ be the symbol set from which we select our information as well as coded symbols. The fundamental operation we use is the Exclusive OR (XOR) which helps add symbols logically bit by bit in binary domain. In our study, nodes represent blocks
of data that contains one or more symbols in it. Symbols are the smallest data unit over which XOR operations are defined.

An $[n, k, t, b]$ array BP-XOR code is a $b \times n$ two dimensional rate $r=k / n$ binary linear code $\mathcal{C}=\left[a_{i, j}\right]_{1 \leq i \leq b, 1 \leq j \leq n}$ in which the coding symbol $a_{i, j} \in M$ is the XOR of a subset of source symbols $I=\left\{v_{1}, \ldots, v_{b k}\right\}$, typically structured as a $b \times k$ data matrix, and $I$ can be reconstructed from any $n-t$ columns of the linear code $\mathcal{C}$ using BP algorithm for an appropriate integer $t \leq n-k$. The degree of a coded symbol $a_{i, j}$, denoted as $\sigma_{i, j}$, is the number of information symbols that participate in logical XOR operation i.e., $a_{i, j}=$ $v_{z_{1}} \oplus \cdots \oplus v_{z_{\sigma_{i, j}}}$ such that $v_{z_{s}} \in I$ for all $s \in\left\{1, \ldots, \sigma_{i, j}\right\}$. A $t$-erasure correcting array BP-XOR code is MDS if the source symbols can be reconstructed from $k=n-t$ columns of $\mathcal{C}$.
For a given positive number $b^{\prime}$ satisfying $b^{\prime}>b$, a [ $\left.n, k, t, b, b^{\prime}\right]$ AMDS array BP-XOR code $\mathcal{C}^{a}$ is a linear code with $i$-th column $\left(y_{i, 1}, \ldots, y_{i, b_{i}}\right)=\left(x_{1}, \ldots, x_{b k}\right) G_{i}$ for a $b k \times b_{i}$ generator matrix $G_{i}, i \in\{1, \ldots, n\}$ such that $b^{\prime}=$ $(1 / n) \sum_{i} b_{i}$. Therefore, the generator matrix for $\mathcal{C}^{a}$ is given by the following matrix of size $b k \times \sum_{i} b_{i}$,

$$
\begin{equation*}
G_{\mathcal{C}^{a}}=\left[G_{1}\left|G_{2}\right| \ldots \mid G_{n}\right] . \tag{1}
\end{equation*}
$$

What makes this code asymptotically MDS is that it is possible to perfectly reconstruct user data matrix $I$ from any $k$ column combinations of $\mathcal{C}^{a}$ using BP decoding and as $b \rightarrow \infty$ we have $b^{\prime} \rightarrow b$. Note that the raw source data needs not to be in standard $b \times k$ form. For any positive integer $g$ satisfying $b \mid g$ and $k \mid g$, the matrix $G_{\mathcal{C}^{a}}$ will still be a valid generator matrix for different arrangements of the data block such as $b / g \times k g$. We finally note that the code $\mathcal{C}^{a}$ is not in two dimensional standard rectangle form as in $\mathcal{C}$. However, we introduced parameter $b^{\prime}$ to be able to make AMDS array BP-XOR codes analogous to standard MDS array code representations through regular binary matrices. For a given fixed code rate $r$ and $n$, let us define $\epsilon(b, n)$ to be the maximum coding overhead ${ }^{1}$ of $\mathcal{C}^{a}$ satisfying $b^{\prime}=(1+\epsilon(b, n)) b$. The asymptotically optimal overhead property implies that as $b \rightarrow \infty$ we have $\epsilon(b, n) \rightarrow 0$.

Letting $\sigma$ denote the maximum check node degree of a given array BP-XOR code, we note from [6] that if $k=\sigma$ it is not hard to show that

$$
\begin{equation*}
n \leq k b+1+\max \{k-3,0\} \tag{2}
\end{equation*}
$$

where the upper bound can be arbitrarily large (i.e., for $b \gg 1$ ) which in turn allows any arbitrarily small $r$ to be a practical choice. However, for $k>\sigma$ it is observed that the array code blocklength $n$ is upper bounded based on a specific choice of $k$ [6]. In addition, we observe from the same study that for $b \gg 1$ and large enough $k$ i.e., $k>\sigma^{2}$ we have $n \leq k+\sigma-1$. This also implies that for large enough information blocklength $k$, the achievable rate will be close to 1 , putting a constraint on the code design rate.

[^0]
## B. Main Result

We begin with providing the following theorem that sets the necessary condition/s on the code parameters for the existence of AMDS array BP-XOR codes.
Theorem 2.1. Let $\mathcal{C}^{a}$ be $a\left[n, k, t, b, b^{\prime}\right]$ AMDS array BPXOR code such that the maximum coded node degree satisfies $2<\sigma<(b k-1) /\left(b^{\prime}-1\right)$. Then, we have

$$
\begin{align*}
& n \leq k+\sigma-1+  \tag{3}\\
& \left\lfloor\frac{b\left(k\left(\sigma^{\prime}-\sigma\right)+(\sigma-1) \sigma^{\prime}\right)-(\sigma-1)(3 \sigma / 2-1)}{b\left(k-\sigma^{\prime}\right)+\sigma-1}\right\rfloor
\end{align*}
$$

where $\sigma^{\prime}=\sigma(1+\epsilon(b, n))$ and $\epsilon(b, n)$ is the coding overhead.
Proof. The proof can be found in the full version [9].
Note that if $b \rightarrow \infty$ we will have $\sigma^{\prime} \rightarrow \sigma$ and hence equation (3) becomes identical to equation (2) of [6] except the term $(\sigma-1)(\sigma / 2-1)$ that appears when we set $\sigma^{\prime}=\sigma$. This term is essentially what makes the upper bound improved (tighter).
There are two cases that are interesting to consider for understanding the asymptotical performance. First, if $b$ tends large we will have $\sigma^{\prime} \rightarrow \sigma$. Hence,

$$
n \leq k+\sigma-1+\left\lfloor\frac{(\sigma-1) \sigma}{k-\sigma}\right\rfloor-\mathbf{1}_{(k-\sigma) \mid(\sigma-1) \sigma}
$$

where $\mathbf{1}_{A}$ is logical one if $A$ is true, otherwise it is zero. This indicator function is used due to the flooring operation and $\sigma$ only equals to $\sigma^{\prime}$ in the limit. Thus, if the code becomes array MDS in the limit, there remains no dependence of $n$ on $b$. On the other hand, if we let large but fixed $b \leq k$, and if $k$ tends large, we shall have

$$
\begin{equation*}
n \leq k+\sigma^{\prime}-1=k+\sigma(1+\epsilon(b, n))-1 \tag{4}
\end{equation*}
$$

which can be made arbitrarily large if we choose $\epsilon(b, n) \rightarrow \infty$ for a fixed $b$ and large $n$. This quick observation demonstrates that as the array BP-XOR code becomes near-optimal in terms of recovery performance, the upper bound on the number of columns i.e., the blocklength $n$ can dramatically be improved.
Although desirable properties of the coding overhead are specified, we still need specific constructions to quantify tighter bounds on $n$ (and $r$ ). Based on this argument, we shall present a code construction method that uses the result of Theorem 2.1 and has an appropriate $\epsilon(b, n)$ with the required properties as summarized below.
(1) For fixed $k$ and rate $r$, as $b \rightarrow \infty$ we have vanishing coding overhead i.e., $\epsilon(b, n) \rightarrow 0$.
(2) For fixed $b$ and rate $r$, as $k, n \rightarrow \infty$ we have a diverging coding overhead i.e., $\epsilon(b, n) \rightarrow \infty$.

## III. Discrete Geometry Constructions of ASYMPTOTICALLY-MDS ARRAY BP-XOR CODES

In this section, we will introduce a particular code construction method based on discrete geometry [7] and show that they can be regarded as a special type of AMDS array BP-XOR codes.


Fig. 1. A simple illustration of the projection concept and Mojette coding.

The discrete geometry construction is known as Mojette transform which is based on discrete version of Radon transform [10], and can be used to generate redundancy, not just for rectangle two dimensional data grid but also for any convex shape data grid. In our study, we consider matrix (rectangle) data and let encoder compute a linear set of projections at angles specified by a couple of coprime integers $(p, q)$ from a $b \times k$ discrete data structure $f:(z, l) \rightarrow \mathbb{N}$ as shown in Fig. 1. Suppose that we generate $n$ projections with parameters $\left\{\left(p_{i}, q_{i}\right), 0 \leq i \leq n-1\right\}$. The length of the projection $i$, denoted by $b_{i}$, is a function of the number of projections $n$, the angle parameters $\left(p_{i}, q_{i}\right)$ and the data grid size $b \times k$. It can be expressed in a closed form as follows [7],

$$
\begin{equation*}
b_{i}=\left|p_{i}\right|(k-1)+\left|q_{i}\right|(b-1)+1 \tag{5}
\end{equation*}
$$

Note that in this construction, generated projections can be treated as the columns of the asymptotically-MDS BP-XOR code. An example code with parameters $k=3, b=4$ with $n=$ 3 projections with parameters $(-1,1),(1,0),(1,1)$ is shown in Fig. 1. Each bin or symbol of the $i$-th projection, based on ( $p_{i}, q_{i}$ ), can be computed as given by the following compact formulation

$$
\begin{align*}
M_{\left(p_{i}, q_{i}\right)} f\left(m+(b-1) q_{i} u\left(q_{i}\right)+(k-1) p_{i} u\left(p_{i}\right)\right)  \tag{6}\\
=\bigoplus_{z=0}^{b-1} \bigoplus_{l=0}^{k-1} f(z, l) \delta_{m+z q_{i}+l p_{i}} \tag{7}
\end{align*}
$$

for all $m$ values satisfying the inequality,

$$
\begin{gathered}
-(b-1) q_{i} u\left(q_{i}\right)-(k-1) p_{i} u\left(p_{i}\right) \leq m \leq \\
b_{i}-(b-1) q_{i} u\left(q_{i}\right)-(k-1) p_{i} u\left(p_{i}\right)-1
\end{gathered}
$$

where $\bigoplus$ stands for Boolean XOR operation, $u($.$) is the$ discrete unit function and $\delta_{i}$ is Kronecker delta function which are given by

$$
u(s)=\left\{\begin{array}{ll}
1, & \text { if } s>0 \\
0, & \text { Otherwise }
\end{array}, \quad \delta_{i}= \begin{cases}0, & \text { if } i \neq 0 \\
1, & \text { if } i=0\end{cases}\right.
$$

Mojette transform codes can be decoded using BP algorithm and the exact reconstruction of user data matrix is possible if the projection parameters $\left(p_{i}, q_{i}\right)$ are selected judiciously according to the following Katz criterion.

Theorem 3.1. For a given AMDS array BP-XOR code defined by $n$ projections with parameters $\left(p_{i}, q_{i}\right)$ on $a b \times k$ data matrix, exact data reconstruction is possible using iterative BP if $\sum_{i=0}^{n-1}\left|p_{i}\right| \geq b$ or $\sum_{i=0}^{n-1}\left|q_{i}\right| \geq k$.
Proof. The proof can be found in [11].
According to Theorem 2.1, the maximum degree of the coded symbols plays the key role in the attainable blocklength of the BP-XOR codes. Therefore, next we find the maximum degree number in the case of Mojette transform codes and see that this parameter can be adjusted based on the selection of projection parameters $\left(p_{i}, q_{i}\right)$. The following theorem quantifies this number.

Theorem 3.2. Let us use $\sigma_{i}, i \in\{1,2, \ldots, n\}$ to denote the maximum degree of the ith projection with parameters $\left(p_{i}, q_{i}\right)$. We have $\sigma_{i}=\min \left\{\left\lceil b /\left|p_{i}\right|\right\rceil,\left\lceil k /\left|q_{i}\right|\right\rceil\right\}$ and hence $\sigma=\max _{i}\left\{\sigma_{i}\right\}$.
Proof. The proof can be found in the full version [9].
Next, let us quantify the coding overhead for Mojette transform based AMDS array BP-XOR codes by considering $k=\sigma$ and $k>\sigma$ cases separately.

## A. Case $k=\sigma$

First of all, note that depending on the choices of $\left(p_{i}, q_{i}\right)$, the coding overhead as well as the maximum degree of the code can change. Although, there are multiple choices for $k=\sigma$, we provide the typical choice below that also ensures small coding overhead.

Construction 3.3. Let us consider the following choice of coprime integers,
$q_{i}=1, p_{i} \in \mathfrak{T}=\left\{-\left\lfloor\frac{n-1}{2}\right\rfloor, \ldots,-1,0,1,2, \ldots,\left\lceil\frac{n-1}{2}\right\rceil\right\}$
where $\mathfrak{T}$ is known as canonical enumeration of integers [12] that goes with the name A 007306 and satisfies $\operatorname{gcd}\left(p_{i}, q_{i}\right)=1$ for $i=0, \ldots, n-1$.

Note that this construction satisfies the Katz criterion simply because collecting any $k$ projections will lead us to have $\sum\left|q_{i}\right|=k$. If we use coprime integers as given by the Construction 3.3, we have $q_{i}$ which never equals to zero and $\sigma_{i}=\min \{\lceil b /\lceil(n-1) / 2\rceil, k\}$. We note that we have $\sigma=k$ for $b \gg 1$. We next quantify the coding overhead for this particular construction and show the asymptotically optimal property.
Theorem 3.4. For the AMDS BP-XOR code based on Mojette Transform with parameters as given in Construction 3.3, for $b \gg 1$, we have

$$
\begin{equation*}
\epsilon(b, n) \approx \frac{n(2-r)(n r-1)}{4 b} \tag{9}
\end{equation*}
$$

where $r$ is the fixed rate of the array $B P-X O R$ code.
Proof. The proof can be found in the full version [9].

For fixed $r$ and $k$ (i.e., fixed $n$ ), if $b \rightarrow \infty$ then it is clear that $\epsilon(b, n) \rightarrow 0$ proving the asymptotical property. On the other hand, for fixed $r$ and $b$, if $n \rightarrow \infty$ then we have $\epsilon(b, n) \rightarrow \infty$. In fact, it is not hard to see that $\epsilon(b, n)=O\left(n^{2}\right)$. Therefore, due to these desirable properties of the overhead and considering the inequality (4), we can make $n$ arbitrarily large. Particularly we can find the following lower bound on $n$ for $k=r n=\sigma$ and $r>0.5$,

$$
\begin{equation*}
n \leq r n+r n\left(1+\frac{n(2-r)(n r-1)}{4 b}\right)-1 \tag{10}
\end{equation*}
$$

which yields the inequality

$$
\begin{equation*}
n-2 n r \leq \frac{n^{3} r^{2}(2-r)}{4 b} \Rightarrow n \geq \sqrt{\frac{4 b(1-2 r)}{r^{2}(2-r)}} \tag{11}
\end{equation*}
$$

This final lower bound shows that the value for the blocklength $n$ can be arbitrarily large for judiciously selected large $b$. Note that the case $k=\sigma$ has the least constraint on the code blocklength for any MDS array BP-XOR code. The case $k>\sigma$ is more interesting which is considered next.

## B. Case $k>\sigma$

With classical array BP-XOR codes, the code blocklength $n$ is constrained by the following upper bound for $b \gg 1$,

$$
\begin{equation*}
n \leq k+\sigma-1+\left\lfloor\frac{\sigma(\sigma-1)}{k-\sigma}\right\rfloor-\mathbf{1}_{(k-\sigma) \mid(\sigma-1) \sigma} \tag{12}
\end{equation*}
$$

which is the same for AMDS array BP-XOR codes as mentioned in Section II. However, as the blocklength gets large, we should no longer have constraints on the size of the blocklength for AMDS array BP-XOR codes which can be achieved by selecting an appropriate set of parameters.

Let us provide another set of parameters that shall satisfy $k>\sigma$. The possibilities of the pair $\left(p_{i}, q_{i}\right)$ selection for making $k>\sigma$ is not unique. We will consider the typical class as given in construction 3.5.
Construction 3.5. Let us consider the following choice of coprime integers for $n$ projections,

$$
\begin{align*}
& q_{i}=q_{e}>0 \\
& p_{i} \in \mathfrak{U}=\left\{\lceil-n+1\rceil_{\text {odd }}, \ldots,-1,1,3, \ldots,\lceil n-1\rceil_{\text {odd }}\right\} \tag{13}
\end{align*}
$$

where $q_{e}$ is a positive even number, and $\lceil.\rceil_{\text {odd }}$ rounds to the next biggest odd integer of the argument, respectively.

Note that using construction 3.5, it is easy to verify that we have $G C D\left(p_{i}, q_{i}\right)=1$. Also, we have $k>\sigma=$ $\max _{i}\left\{\min \left\{\left\lceil b /\left|p_{i}\right|\right\rceil,\left\lceil k /\left|q_{i}\right|\right\rceil\right\}\right\}=\left\lceil k / q_{e}\right\rceil$. It is of interest to quantify the coding overhead to be able to find the upper bounds on the code blocklength.


Fig. 2. Upper bounds on $n$ as a function of $k$ for $b=10000$.

Theorem 3.6. For the AMDS array BP-XOR code based on Mojette transform with parameters as given in construction 3.5, for $b \gg 1$, we have
$\epsilon(b, n) \approx$

$$
\begin{equation*}
\frac{\left\lceil k / q_{e}\right\rceil}{k b}\left((k-1)\left(n-\frac{\left\lceil k / q_{e}\right\rceil}{2}\right)+(b-1) q_{e}+1\right)-1 \tag{14}
\end{equation*}
$$

where $q_{e}$ is a positive even number, and $\lceil.\rceil_{\text {odd }}$ rounds to the next biggest odd integer of the argument, respectively.

Proof. The proof can be found in the full version [9].
Note that as long as $q_{e} \mid k$, we have $\epsilon(b, n) \rightarrow 0$ for large $b$ demonstrating the asymptotically optimal overhead property. Similarly, for fixed $r$ and $b$, if $k, n \rightarrow \infty$ then we have $\epsilon(b, n) \rightarrow \infty$ satisfying the second desirable property.

Finally, using equation (4) we can express the upper bound on $n$ as follows,
$n \leq k+\frac{\sigma\left\lceil k / q_{e}\right\rceil}{k b}\left((k-1)\left(n-\frac{\left\lceil k / q_{e}\right\rceil}{2}\right)+(b-1) q_{e}+1\right)-1$
Since it might be non-trivial to see with this result that we dramatically improve the upper bounds on the code blocklength, in the next section, we provide some numerical results that quantify the upper bounds in order to make numerical comparisons easier.

## IV. Numerical Results

Let us consider $q_{e}=2$ and a large $b$ value, such as $b=10000$ (this choice is completely arbitrary) and compare the upper bounds on $n$ with using classical exact MDS array BP-XOR codes and their asymptotically optimal version proposed in our study, abbreviated as AMDS (asym). We present results in Fig. 2 and Fig. 3 each corresponding to two different rates $3 / 4,1 / 2$, respectively as example use cases. These results demonstrate that as the code rate decreases, classical MDS array BP-XOR codes are only possible for very


Fig. 3. Upper bounds on $n$ as a function of $k$ for $b=10000$.


Fig. 4. Upper bounds on $n$ as a function of $k$ for $b=10000$.
small values of $k$. On the other hand, although the same is true for AMDS array BP-XOR codes for small $k$, it is also observed that for large enough $k$ our bounds are larger than the required $n$ (fixed by the code rate), allowing possible constructions to achieve the corresponding rate AMDS array BP-XOR code such as Mojette construction we have provided in previous sections. These figures also present the upper bound behavior for small $k$ on the left corner of each plot. Plots include a curve "Required $n$ " to specify the required value for $n$ for the corresponding code rate $r=k / n$.

In order to see clearly the range of rates that are possible with both constructions provided in previous sections, Fig. 4 depicts the minimum rate that is possible as a function of the assumed rate. Note that with asymptotically MDS array BP-XOR codes, the upper bound on $n$ depends on the coding overhead which is a function of code rate. Thus, the minimum code rate changes as the assumed code rate changes. For
each assumed rate, we calculate the upper bound and then compute the minimum code rate possible. With respect to classical MDS BP-XOR codes, since the upper bound does not change with varying assumed rate (since the coding overhead is always zero), the curves turns out to be flat.

According to Fig. 4, the region that lies above the curves represent all possible code rates. However, there is no guarantee each assumed rate is achievable. However as $k$ tends large, it becomes impossible to construct classical MDS array BPXOR codes with rate smaller than 1 . In contrast, by relaxing the exact MDS condition (such as adapting asymptotically MDS constructions), we can improve the the region of possibilities for better achievability. With this study, we have just provided a simple construction based on discrete geometry (with judicious selection of parameters) that helps improve the upper bounds on the code blocklength $n$. Other constructions as a future work may prove useful to improve the results presented in this subsection.

## V. Conclusion

Array BP-XOR codes are attractive data protection schemes for low-complexity and optimal reliability. Their finite versions are shown to have limitations on the maximum blocklength when the coding symbol degree is particularly lower than the data size. We have shown in this study, this limitation can greatly be relaxed by extending the original optimal class to asymptotically optimal class. We have also have shown one particular code construction based on discrete geometry that satisfies all the requirements of being AMDS array BP-XOR codes. These codes can be encoded and decoded in linear time with the blocklength and the achievable bound on the blocklength is far from that of the finite counterpart.

## References

[1] P. G. Farrell, "A survey of array error control codes," preprint, 1990.
[2] M. Blaum and R. M. Roth, "New Array Codes for Multiple Phased Burst Correction," IEEE Trans. on Information Theory, 339(1):66-77, 1993.
[3] Y. Cassuto and A. Shokrollahi, "LDPC codes for 2D arrays", IEEE Trans. Inf. Theory, vol. 60, no. 6, pp. 3279-3291, 2014.
[4] S. Lin, and D. J. Costello, Jr., Error Control Coding: Fundamentals and Applications. Prentice-Hall. 1983.
[5] Y. Wang, "Array BP-XOR codes for reliable cloud storage systems," In Proc. of IEEE ISIT, pp. 326-330, 2013.
[6] M. B. Paterson, D. R. Stinson and Y. Wang, "On Encoding Symbol Degrees of Array BP-XOR Codes," Cryptography and Communications, vol. 8, no. 1, pp. 19-32, 2016.
[7] J. P. Guedon and N. Normand, "The Mojette transform: The first ten years," Discrete Geometry for Computer Imagery, series Lecture Notes in Computer Science, vol. 3429, pp. 79-91, 2005.
[8] S. S. Arslan, B. Parrein and N. Normand, "Mojette transform based LDPC erasure correction codes for distributed storage systems," 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, pp. 1-4, 2017.
[9] S. S. Arslan, "Asymptotically MDS Array BP-XOR Codes," Available online: https://arxiv.org/abs/1709.07949 [cs], Sept. 2017.
[10] J. Guedon, D. Barba, N. Burger, "Psychovisual image coding via an exact discrete Radon transform," In: Wu, L. (ed.) Proc. Visual Communications and Image Processing 1995 (VCIP95), Taipei, Taiwan, CORESA, pp. 562-572, 1995.
[11] M. B. Katz, "Questions of uniqueness and resolution in reconstruction from projections," In: Levin, S. (Ed.), Lecture Notes in Biomathematics, vol. 26, Springer-Verlag, New York, 1978.
[12] The On-Line Encyclopedia of Integer Sequences. Available online: https://oeis.org/


[^0]:    ${ }^{1}$ Since columns of $\mathcal{C}^{a}$ may have different sizes, the coding overhead depends on which $k$ columns are used for reconstruction. Also note that the coding overhead also depends on the number of columns $n$ in the code, so called array code blocklength.

