

# Constructions of Partial MDS Codes over Small Fields

Ryan Gabrys\*, Eitan Yaakobi<sup>†</sup>, Mario Blaum<sup>‡</sup> and Paul H. Siegel<sup>§</sup>

\* Code 53227, Spawar Systems Center, San Diego, San Diego, CA 92115, USA

<sup>†</sup> Technion — Israel Institute of Technology, Haifa, 32000 Israel

<sup>‡</sup> IBM Research Division, Almaden Research Center, San Jose, CA 95120, USA

<sup>§</sup> University of California, San Diego, La Jolla, CA 92093, USA

Emails: ryan.gabrys@navy.mil, yaakobi@cs.technion.ac.il, mmblauch@us.ibm.com, psiegel@ucsd.edu

**Abstract**—Partial MDS (PMDS) codes are a class of erasure-correcting array codes which combine local correction of the rows with global correction of the array. An  $m \times n$  array code is called an  $(r; s)$  PMDS code if each row belongs to an  $[n, n - r, r + 1]$  MDS code and the code can correct erasure patterns consisting of  $r$  erasures in each row together with  $s$  more erasures anywhere in the array. While a recent construction by Calis and Koyluoglu generates  $(r; s)$  PMDS codes for all  $r$  and  $s$ , its field size is exponentially large. In this paper, a family of PMDS codes with field size  $\mathcal{O}(\max\{m, n^{r+s}\}^s)$  is presented.

## I. INTRODUCTION

Erasure-correcting array codes are a class of codes mainly used to protect data in a Redundant Array of Independent Disks (RAID) architecture against catastrophic disk failures. Every column in the array corresponds to sectors from the same disk, such that a disk failure is modeled by a column erasure. Column failures are the widely studied model in the literature and different solutions such as RAID 5 and RAID 6 are designed specifically for this failure model. However, introducing solid state drives (SSDs) to the enterprise industry has brought new challenges in the design of a RAID architecture. Namely, existing solutions for RAID architectures are no longer adequate since SSDs may experience both disk failures and hard errors. To overcome this problem, Partial MDS (PMDS) codes were proposed in [1].

PMDS codes are designed to tolerate this mixed type of failures. We say that a code consisting of  $m \times n$  arrays is an  $(r; s)$  PMDS code if each row in the codeword arrays belongs in an  $[n, n - r, r + 1]$  MDS code, and furthermore, it is possible to correct  $s$  more arbitrary erasures in the array. The construction of PMDS codes for all  $r$  and  $s$  is a challenging task. In [1] the problem was solved when either  $s = 1$  or  $r = 1$ . Then, in [2] it was solved for  $s = 2$  and arbitrary  $r$  and recently in [3] it was solved for all  $r$  and  $s$  using Gabidulin codes. However, this construction requires a field of size  $\mathcal{O}(n^{mn})$ . Our goal is constructing PMDS codes over fields of relatively small size. In fact, [5] gives a construction of PMDS codes over smaller fields for the case  $r = 1$ . In this paper, we extend the work of [5] to the case where  $r > 1$ . Our  $(r; s)$  PMDS codes have field size  $\mathcal{O}(\max\{m, n^{r+s}\}^s)$ .

The rest of the paper is organized as follows. In Section II, we formally define PMDS codes. In Section III, we present a code, which we denote by  $\mathcal{C}(\Gamma, \beta)$ , using its parity check matrix. In Section IV, we state necessary and sufficient conditions for  $\mathcal{C}(\Gamma, \beta)$  to be a PMDS code. Lastly, in Section V we present a construction of matrices satisfying these conditions while using small fields.

## II. DEFINITIONS AND PRELIMINARIES

For a prime power  $q$ ,  $\mathbb{F}_q$  is the finite field of size  $q$ . A linear code of length  $n$  and dimension  $k$  over  $\mathbb{F}_q$  will be denoted by

$[n, k]_q$  or  $[n, k, d]_q$ , where  $d$  denotes its minimum distance. For an integer  $n \geq 1$ , the set  $\{1, 2, \dots, n\}$  will be denoted by  $[n]$ . We begin with the definition of PMDS codes from [1].

**Definition 1.** Let  $\mathcal{C}$  be a linear  $[mn, m(n - r) - s]$  code over a field such that when codewords are taken row-wise as  $m \times n$  arrays, each row belongs to an  $[n, n - r, r + 1]$  MDS code. Given  $(\sigma_1, \sigma_2, \dots, \sigma_t)$  such that for  $j \in [t]$ ,  $\sigma_j \geq 1$ , we say that  $\mathcal{C}$  is an  $(r; \sigma_1, \sigma_2, \dots, \sigma_t)$ -erasure correcting code if, for any  $1 \leq i_1 < i_2 < \dots < i_t \leq m$ ,  $\mathcal{C}$  can correct up to  $\sigma_j + r$  erasures in row  $i_j$  of an array in  $\mathcal{C}$ . We say that  $\mathcal{C}$  is an  $(r; s)$  partial-MDS (PMDS) code if, for every  $(\sigma_1, \sigma_2, \dots, \sigma_t)$  where  $\sum_{j=1}^t \sigma_j = s$ ,  $\mathcal{C}$  is an  $(r; \sigma_1, \sigma_2, \dots, \sigma_t)$ -erasure correcting code.

We refer to a set  $E_{(r;s)} \subseteq [m] \times [n]$  as an  $(r; s)$ -erasure set if it corresponds to an erasure pattern that can be corrected by an  $(r; s)$  PMDS code. The next example illustrates a  $(2; 2)$  PMDS code and a corresponding erasure set.

**Example 1.** Assume  $m = 3, n = 5, r = 2$ , and  $s = 2$ . Then we can interpret our codewords as arrays having the following form:

$$\mathbf{x} = \begin{pmatrix} c_1 & c_2 & c_3 & p_1^{(1)} & p_2^{(1)} \\ c_4 & c_5 & c_6 & p_3^{(1)} & p_4^{(1)} \\ c_7 & p_1^{(2)} & p_2^{(2)} & p_5^{(1)} & p_6^{(1)} \end{pmatrix}, \quad (1)$$

where there are 7 information symbols and 8 parity symbols. Given this setup, a PMDS code with these parameters is able to correct at most 2 erasures in any row using only the symbols from that row. Furthermore, it can correct 2 more erasures occurring anywhere in the codeword matrix.

Suppose  $\mathbf{x}$  was stored where  $\mathbf{x}$  belongs to a  $(2; 2)$  PMDS code. Suppose further, that  $\mathbf{x}$  experiences erasures resulting in the vector  $\mathbf{y}$  shown below (the symbols in bold correspond to the erasures):

$$\mathbf{y} = \begin{pmatrix} c_1 & c_2 & c_3 & p_1^{(1)} & \mathbf{p}_2^{(1)} \\ c_4 & \mathbf{c}_5 & c_6 & \mathbf{p}_3^{(1)} & p_4^{(1)} \\ c_7 & p_1^{(2)} & p_2^{(2)} & p_5^{(1)} & p_6^{(1)} \end{pmatrix}.$$

The erasures in  $\mathbf{x}$  can be described by the following  $(2; 2)$ -erasure set:

$$E_{(2;2)} = \{(1, 1), (1, 5), (2, 1), (2, 2), (2, 3), (2, 4), (3, 2), (3, 4)\}.$$

Let  $\mathbb{F}_q$  be a field of size  $q$  and  $M$  be a positive integer. For a given basis of  $\mathbb{F}_{q^M}$  over  $\mathbb{F}_q$ , every element  $v \in \mathbb{F}_{q^M}$  can also be represented as a length- $M$  vector over  $\mathbb{F}_q$ . A set of elements  $v_1, \dots, v_N \in \mathbb{F}_{q^M}$  is said to be *linearly independent* over  $\mathbb{F}_q$  if the length- $M$  vectors representing these  $N$  elements are linearly independent over  $\mathbb{F}_q$ .

### III. GENERAL CONSTRUCTION OF PMDS CODES

A PMDS code can be seen either as an  $m \times n$  array code or as a code of length  $mn$ . We will interchangeably use these two options to represent the code. From the context, it will be clear which option we use. We assume that  $q$  is a prime power such that  $q \geq n$  and  $\mathbb{F}_{q^M}$  is an extension field of  $\mathbb{F}_q$ . Let  $\beta \in \mathbb{F}_q$  be a primitive element. Let  $\mathcal{H}$  be the  $(m \cdot r + s) \times mn$  matrix given by (2), which will be the parity check matrix of the code  $\mathcal{C}$ , i.e.,

$$\mathcal{C} = \{\mathbf{x} \in (\mathbb{F}_{q^M})^{mn} \mid \mathcal{H} \cdot \mathbf{x}^T = \mathbf{0}\}.$$

Let  $\Gamma = (\alpha_{1,1}, \alpha_{1,2}, \dots, \alpha_{m,n}) \in (\mathbb{F}_{q^M})^{mn}$  be a sequence of  $mn$  distinct elements. For  $i \in [m], j \in [n], \ell \in [s]$ , we set

$$v_{i,j}^{(\ell)} = \alpha_{i,j}^{q^{\ell-1}}.$$

We denote the code generated by this construction by  $\mathcal{C}(\Gamma, \beta)$ . Next we discuss how to select the elements in the sequence  $\Gamma$  such that the code  $\mathcal{C}(\Gamma, \beta)$  given by the parity check matrix  $\mathcal{H}$  is a PMDS code. This selection will determine also the size of the extension field  $\mathbb{F}_{q^M}$ .

Notice that we can interpret our constraints as follows. Let  $\mathbf{x} = (x_{1,1}, x_{1,2}, \dots, x_{m,n}) \in \mathcal{C}(\Gamma, \beta) \subseteq (\mathbb{F}_{q^M})^{mn}$  be a codeword, where  $x_{i,j}$  denotes the element in  $\mathbf{x}$  that is located in row  $i$  and column  $j$ . From (2),

$$0^{k-1} \cdot x_{i,1} + \sum_{j=2}^n \beta^{(k-1)(j-1)} x_{i,j} = 0, \text{ for } i \in [m], k \in [r] \quad (3)$$

$$\sum_{i=1}^m \sum_{j=1}^n \alpha_{i,j}^{q^{\ell-1}} x_{i,j} = 0, \text{ for } \ell \in [s]. \quad (4)$$

In the rest of this paper, we will refer to an  $(r; 0)$ -erasure set as an  $(r; 0)$ -erasure vector  $S \in ([n]^r)^m$  and denote it by  $S = (s_1, \dots, s_m)$ , where for  $i \in [m]$ ,  $s_i = (s_{i,1}, \dots, s_{i,r}) \in [n]^r$ , and  $s_{i,1} < s_{i,2} < \dots < s_{i,r}$ . For example, suppose we store the vector  $\mathbf{x}$  from (1) and we receive

$$\mathbf{y} = \begin{pmatrix} c_1 & c_2 & c_3 & p_1^{(1)} & p_2^{(1)} \\ c_4 & c_5 & c_6 & p_3^{(1)} & p_4^{(1)} \\ c_7 & p_1^{(2)} & p_2^{(2)} & p_5^{(1)} & p_6^{(1)} \end{pmatrix},$$

so that a  $(2; 0)$ -erasure vector occurred, marked by the symbols in bold. Then we represent this erasure pattern using the vector  $S = (s_1, s_2, s_3) = ((1, 2), (3, 4), (1, 5))$ .

For a matrix  $A$  with  $n$  columns and a vector  $\mathbf{s} = (s_1, \dots, s_r) \in [n]^r$ , where  $s_1 < s_2 < \dots < s_r$ , let  $A_{\mathbf{s}}$  be the submatrix of  $A$  given by the columns in the set  $\{s_1, s_2, \dots, s_r\}$ . Similarly, for a vector  $\mathbf{u}$  of length  $n$ ,  $\mathbf{u}_{\mathbf{s}}$  is the sub-vector of  $\mathbf{u}$  given by the indices of the set  $\{s_1, s_2, \dots, s_r\}$ . By a slight abuse of notation, for a vector  $\mathbf{s} = (s_1, \dots, s_r) \in [n]^r$ , we denote by  $\bar{\mathbf{s}} \in [n]^{n-r}$  the vector

$$\bar{\mathbf{s}} = (\bar{s}_1, \bar{s}_2, \dots, \bar{s}_{n-r}), \quad (5)$$

which contains, in increasing order, all the values in  $[n]$  that do not appear in  $\mathbf{s}$ . For example, if  $\mathbf{s} = (1, 3, 7, 8) \in [9]^4$ , then  $\bar{\mathbf{s}} = (2, 4, 5, 6, 9)$ . Lastly, we denote by  $H^{(\beta)}$  the matrix

$$H^{(\beta)} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & \beta & \beta^2 & \dots & \beta^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \beta^{r-1} & \beta^{2(r-1)} & \dots & \beta^{(r-1)(n-1)} \end{pmatrix} \in (\mathbb{F}_q)^{r \times n}.$$

### IV. NECESSARY AND SUFFICIENT CONDITIONS FOR PMDS CODES

In this section we state necessary and sufficient conditions for the construction from Section III to generate PMDS codes. Here we use the notation  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$  to represent an  $m \times n$  codeword array, where for  $i \in [m]$ ,  $\mathbf{x}_i$  is a length- $n$  vector. We also assume that  $\beta$  and the sequence  $\Gamma$  are given so they determine the code  $\mathcal{C}(\Gamma, \beta)$ . Unless stated otherwise we assume that all fields used in the paper have characteristics 2. We start with the following claim.

**Claim 1** *If  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_m) \in \mathcal{C}(\Gamma, \beta)$ , then for any  $(r; 0)$ -erasure vector  $S = (s_1, \dots, s_m) \in ([n]^r)^m$ , the following equality holds for all  $i \in [m]$ :*

$$(\mathbf{x}_i)_{s_i}^T = ((H_{s_i}^{(\beta)})^{-1} \cdot H_{\bar{s}_i}^{(\beta)}) \cdot (\mathbf{x}_i)_{\bar{s}_i}^T.$$

*Proof:* Since  $H^{(\beta)}$  is a parity check matrix of an  $[n, n-r]$  MDS code, any  $r$  columns of  $H^{(\beta)}$  are linearly independent, and thus the matrix  $H_{s_i}^{(\beta)} \in \mathbb{F}_q^{r \times r}$  has an inverse matrix  $(H_{s_i}^{(\beta)})^{-1}$ . According to (3), we have

$$\mathbf{0}^T = H^{(\beta)} \cdot \mathbf{x}_i^T = H_{s_i}^{(\beta)} \cdot (\mathbf{x}_i)_{s_i}^T + H_{\bar{s}_i}^{(\beta)} \cdot (\mathbf{x}_i)_{\bar{s}_i}^T.$$

After multiplying both sides on the left by the matrix  $(H_{s_i}^{(\beta)})^{-1}$  we get

$$\begin{aligned} \mathbf{0}^T &= (H_{s_i}^{(\beta)})^{-1} \cdot H_{s_i}^{(\beta)} \cdot (\mathbf{x}_i)_{s_i}^T + (H_{s_i}^{(\beta)})^{-1} \cdot H_{\bar{s}_i}^{(\beta)} \cdot (\mathbf{x}_i)_{\bar{s}_i}^T \\ &= (\mathbf{x}_i)_{s_i}^T + ((H_{s_i}^{(\beta)})^{-1} \cdot H_{\bar{s}_i}^{(\beta)}) \cdot (\mathbf{x}_i)_{\bar{s}_i}^T \end{aligned}$$

which implies the statement in the claim.  $\blacksquare$

Next, we recall the notion of  $t$ -wise independence [5].

**Definition 2.** *Let  $\mathbb{F}$  be a field. A multiset  $S \subseteq \mathbb{F}$  is  $t$ -wise independent over a subfield  $\mathbb{F}' \subseteq \mathbb{F}$  if for every  $T \subseteq S$  such that  $|T| \leq t$ , the elements of  $T$  are linearly independent over  $\mathbb{F}'$ .*

For any  $(r; 0)$ -erasure vector  $S = (s_1, \dots, s_m) \in ([n]^r)^m$ , let  $\mathcal{C}^S$  be the code obtained by puncturing  $\mathcal{C}$  in the positions corresponding to the erasure locations referenced by the vector  $S$ . The next claim is straightforward.

**Claim 2** *Code  $\mathcal{C}$  is an  $(r; s)$  PMDS code if and only if  $\mathcal{C}^S$  is an  $[m(n-r), m(n-r) - s]$  MDS code for any  $(r; 0)$ -erasure vector  $S$ .*

The following lemma is well known:

**Lemma 3.** (cf. [6]) *Let  $\mathbb{F}_{q^M}$  be an extension field of  $\mathbb{F}_q$ , and  $H \in (\mathbb{F}_{q^M})^{s \times s}$  be the following matrix*

$$H = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_s \\ \alpha_1^q & \alpha_2^q & \dots & \alpha_s^q \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{q^{s-1}} & \alpha_2^{q^{s-1}} & \dots & \alpha_s^{q^{s-1}} \end{pmatrix},$$

where  $\alpha_1, \alpha_2, \dots, \alpha_s \in \mathbb{F}_{q^M}$ . Then  $H$  has rank  $s$  if and only if the elements  $\alpha_1, \alpha_2, \dots, \alpha_s$  are linearly independent over  $\mathbb{F}_q$ .

Let  $S = (s_1, \dots, s_m) \in ([n]^r)^m$  be an  $(r; 0)$ -erasure vector. For  $i \in [m]$ , let  $A_{s_i}^{(\beta)} \in (\mathbb{F}_q)^{r \times (n-r)}$  be the matrix

$$A_{s_i}^{(\beta)} = (H_{s_i}^{(\beta)})^{-1} \cdot H_{\bar{s}_i}^{(\beta)}.$$

Let  $\alpha_i$  be the vector  $\alpha_i = (\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n})$ , and denote  $\alpha_{s_i} = (\alpha_{i,s_{i,1}}, \alpha_{i,s_{i,2}}, \dots, \alpha_{i,s_{i,r}})$ , where  $s_i = (s_{i,1}, s_{i,2}, \dots, s_{i,r}) \in [n]^r$ , and  $\alpha_{\bar{s}_i} = (\alpha_{i,\bar{s}_{i,1}}, \dots, \alpha_{i,\bar{s}_{i,n-r}})$ ,



$$\begin{pmatrix} \gamma_{s_1,1}^1 & \gamma_{s_1,2}^1 & \gamma_{s_1,3}^1 & \cdots & \gamma_{s_1,n-r}^1 & \gamma_{s_2,1}^1 & \gamma_{s_2,2}^1 & \cdots & \gamma_{s_2,n-r}^1 & \cdots & \cdots & \cdots & \gamma_{s_m,1}^1 & \cdots & \gamma_{s_m,n-r}^1 \\ \gamma_{s_1,1}^q & \gamma_{s_1,2}^q & \gamma_{s_1,3}^q & \cdots & \gamma_{s_1,n-r}^q & \gamma_{s_2,1}^q & \gamma_{s_2,2}^q & \cdots & \gamma_{s_2,n-r}^q & \cdots & \cdots & \cdots & \gamma_{s_m,1}^q & \cdots & \gamma_{s_m,n-r}^q \\ \gamma_{s_1,1}^{q^2} & \gamma_{s_1,2}^{q^2} & \gamma_{s_1,3}^{q^2} & \cdots & \gamma_{s_1,n-r}^{q^2} & \gamma_{s_2,1}^{q^2} & \gamma_{s_2,2}^{q^2} & \cdots & \gamma_{s_2,n-r}^{q^2} & \cdots & \cdots & \cdots & \gamma_{s_m,1}^{q^2} & \cdots & \gamma_{s_m,n-r}^{q^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \cdots & \vdots & \ddots & \vdots \\ \gamma_{s_1,1}^{q^{s-1}} & \gamma_{s_1,2}^{q^{s-1}} & \gamma_{s_1,3}^{q^{s-1}} & \cdots & \gamma_{s_1,n-r}^{q^{s-1}} & \gamma_{s_2,1}^{q^{s-1}} & \gamma_{s_2,2}^{q^{s-1}} & \cdots & \gamma_{s_2,n-r}^{q^{s-1}} & \cdots & \cdots & \cdots & \gamma_{s_m,1}^{q^{s-1}} & \cdots & \gamma_{s_m,n-r}^{q^{s-1}} \end{pmatrix}. \quad (7)$$

## V. PMDS CODES OVER SMALL FIELDS

Note again that the condition in Corollary 6 is a condition only on the sequence  $\Gamma$ . Thus we say that a sequence  $\Gamma$  is an  $(r; s)$ -PMDS sequence of size  $m \times n$  if it satisfies the condition from Corollary 6. As mentioned above, our goal is to find  $(r; s)$ -PMDS sequences of size  $m \times n$  over a field with the smallest possible size. We denote by  $r(n, d, q)$  the minimum redundancy of an  $[n, k, d]_q$  code. We will make use of the following useful lemma, based upon BCH codes, in subsequent derivations.

**Lemma 7.** (cf. [9, Problem 8.12]) *The value  $r(n, d, q)$  satisfies*

$$r(n, d, q) \leq \min \left\{ 1 + \left\lceil \left( 1 - \frac{1}{q} \right) (d - 2) \right\rceil \cdot \lceil \log_q(n) \rceil, \right. \\ \left. \left\lceil \left( 1 - \frac{1}{q} \right) (d - 1) \right\rceil \cdot \lceil \log_q(n) \rceil \right\}.$$

### A. First Construction of PMDS Sequences over a Small Field

Our first result on PMDS codes with small field size is stated in the next lemma.

**Lemma 8.** *If the set of elements in the sequence  $\Gamma = (\alpha_{1,1}, \dots, \alpha_{m,n})$  is  $(r+1)s$ -wise independent over  $\mathbb{F}_q$ , then it is an  $(r; s)$ -PMDS sequence. Hence, there exists an  $(r; s)$ -PMDS sequence over a field  $\mathbb{F}_{q^M}$ , where*

$$M = r(mn, (r+1)s + 1, q), \quad (9)$$

and  $q$  is the smallest prime power larger than  $n$ . In particular, there exists an  $(r; s)$ -PMDS code with field size  $\mathcal{O}(n(mn)^{(r+1)s-1})$ .

*Proof:* The result follows by noting that according to (8), for any  $(r; 0)$ -erasure vector  $S = (s_1, \dots, s_m) \in ([n]^r)^m$  and any  $m$  vectors  $\mathbf{v}_1, \dots, \mathbf{v}_m \in \mathbb{F}_q^r$ , each element in the set  $\hat{T}(S)$  from Corollary 6 can be written as a linear combination over  $\mathbb{F}_q$  of exactly  $r+1$  elements from  $\Gamma$ . That is, we can write any element  $\gamma \in \hat{T}(S)$  as  $\gamma = \mathbf{v} \cdot \mathbf{u}^T$ , where  $\mathbf{v} \in \mathbb{F}_q^{r+1}$  and  $\mathbf{u} \in \Gamma^{r+1}$ , while all the entries in  $\mathbf{u}$  are all distinct. Therefore, any  $s$  elements in  $\hat{T}(S)$  are linear combinations over  $\mathbb{F}_q$  of at most  $(r+1)s$  elements from  $\Gamma$ . Thus, if the set of elements in the sequence  $\Gamma$  is  $(r+1)s$ -wise independent over  $\mathbb{F}_q$ , it follows that the set  $\hat{T}(S)$  is  $s$ -wise independent over  $\mathbb{F}_q$  as well, and hence  $\Gamma$  is an  $(r; s)$ -PMDS sequence.

A construction of such a sequence  $\Gamma$  whose elements are  $(r+1)s$ -wise independent over  $\mathbb{F}_q$  is as follows. Let  $H$  be a parity check matrix of an  $[mn, k, (r+1)s+1]_q$  code  $C$  with redundancy  $r = mn - k$ , where  $q$  is the smallest prime power larger than  $n$ . We set  $\Gamma$  to consist of the  $mn$  columns of  $H$ , which are  $mn$  elements over the field  $\mathbb{F}_{q^r}$ . Since the minimum distance of the code  $C$  is  $(r+1)s+1$ , every  $(r+1)s$  columns of  $H$  are linearly independent over  $\mathbb{F}_q$  and so are every  $(r+1)s$  elements from the sequence  $\Gamma$ . By choosing a code  $C$  where  $r = r(mn, (r+1)s+1, q)$  we get the result stated in (9), and by the

bound from Lemma 7, we obtain  $q^M = \mathcal{O}(n(mn)^{(r+1)s-1})$ .  $\blacksquare$

According to Lemma 8, we already deduce that there exists an  $(r; s)$ -PMDS code over a field which is polynomial in  $mn$  and not exponential as reported in [3]. We present next a further improvement on the field size.

### B. Second Construction of PMDS Sequences over a Small Field

In this section, we give a construction of  $(r; s)$ -PMDS sequences with a smaller field size than the one achieved in Lemma 8. Our primary tool towards achieving this goal is the construction of tensor product codes, which were first proposed in [10]. As will be shown shortly, we will use the columns of the parity check matrix of a tensor product code as the elements of the sequence  $\Gamma$ . Let us first review the definition of tensor product codes, while focusing only on the case of erasure correction. A code  $\mathcal{C} \subseteq (\mathbb{F}_q)^{m \times n}$  is called an  $[m, n; t_1, t_2]$  erasure-correcting code if it can correct any erasure pattern of the following form  $\mathbf{E} = (E_1, \dots, E_m)$ , where for  $i \in [m]$ ,  $E_i \subseteq [n]$  and

- 1)  $|\{i : E_i \neq \emptyset\}| \leq t_1$ ,
- 2) for  $i \in [m]$ ,  $|E_i| \leq t_2$ .

More explicitly, such a code is required to correct erasures in at most  $t_1$  rows, while in each row there are at most  $t_2$  erasures. Such an erasure pattern will be called a  $(t_1; t_2)$ -erasure pattern.

The following theorem draws a connection between  $[m, n; t_1, t_2]$  erasure-correcting codes and  $(r; s)$ -PMDS sequences and gives the main result of this section. Note that in Corollary 11, we give an upper bound on  $q$  which is derived as a consequence of the theorem below.

**Theorem 9.** *Let  $\mathcal{C}_{TP}$  be an  $[m, n; s, r+s]$  erasure-correcting code over  $\mathbb{F}_q$  with redundancy  $\rho$  and parity check matrix  $H_{TP} = (\alpha_{1,1}, \dots, \alpha_{m,n}) \in (\mathbb{F}_{q^\rho})^{mn}$ . Then, the sequence  $\Gamma_{TP} = (\alpha_{1,1}, \dots, \alpha_{m,n}) \in (\mathbb{F}_{q^\rho})^{mn}$  is an  $(r; s)$ -PMDS sequence of size  $m \times n$ .*

*Proof:* Assume  $S = (s_1, \dots, s_m) \in ([n]^r)^m$  is an  $(r; 0)$ -erasure vector and  $\mathbf{v}_1, \dots, \mathbf{v}_m \in \mathbb{F}_q^r$  are  $m$  vectors which determine the set  $\hat{T}(S) = \{\hat{\gamma}_{s_i, j}\}_{i \in [m], j \in [n-r]}$  as specified in (8). We will show that  $\hat{T}(S)$  is  $s$ -wise independent over  $\mathbb{F}_q$ . Any  $s$  elements from the set  $\hat{T}(S)$  can be expressed as a linear combination of elements from the sequence  $\Gamma_{TP}$ . For  $i \in [m]$ , we denote by  $E_i \subseteq [n]$  the locations of elements which belong to this linear combination from the  $i$ -th row of  $\Gamma_{TP}$ , that is, from the elements  $\alpha_{i,1}, \dots, \alpha_{i,n}$ . Consider the vector  $\mathbf{E} = (E_1, \dots, E_m)$  and note that it is an  $(s; r+s)$ -erasure pattern. Since the code  $\mathcal{C}_{TP}$  is an  $[m, n; s, r+s]$  erasure-correcting code, every collection of columns from the parity matrix  $H_{TP}$  which correspond to an  $(s; r+s)$ -erasure pattern is linearly independent. Hence, the set of elements in  $\cup_{i=1}^m E_i \subseteq \Gamma_{TP}$  is linearly independent over  $\mathbb{F}_q$  and, therefore, so is every set of  $s$  elements from  $\hat{T}(S)$ .  $\blacksquare$

According to Theorem 9, the task of finding  $(r; s)$ -PMDS sequences can now be translated to the construction of  $[m, n; s, r + s]$  erasure-correcting codes over  $\mathbb{F}_q$  with small redundancy  $\rho$ . We review the construction of such codes as presented in [10]. Specifically, the parity check matrix for an  $[m, n; s, r + s]$  erasure-correcting code  $\mathcal{C}_{TP}$  over  $\mathbb{F}_q$  (where  $q = n$ ) is given as follows:

- 1) Assume that  $H' \in \mathbb{F}_q^{(r+s) \times n}$  is a parity check matrix for an  $[n, n - r - s]_q$  MDS code.
- 2) Assume that  $H'' \in \mathbb{F}_q^{R \times m}$  is a parity check matrix for an  $[m, m - R, s + 1]_{q^{r+s}}$  code.
- 3) Representing every column in  $H'$  as an element in  $\mathbb{F}_{q^{r+s}}$ , we let  $\mathcal{C}_{TP}$  be a code over  $\mathbb{F}_q$  with the parity check matrix  $H'' \otimes H'$  formed by taking the tensor product of  $H''$  and  $H'$ .

**Theorem 10.** (cf. [10]) *The code  $\mathcal{C}_{TP}$  is an  $[m, n; s, r + s]$  erasure-correcting code over  $\mathbb{F}_q$  with redundancy  $\rho = (r + s)R$ .*

We are now ready to provide an upper bound on the field size for  $(r; s)$  PMDS codes.

**Corollary 11.** *There exists an  $(r; s)$  PMDS code with field size  $q^{(r+s)r(m, s+1, q^{r+s})}$ , where  $q$  is the smallest prime power larger than  $n$ , which is  $\mathcal{O}(\max\{m, n^{r+s}\}^s)$ .*

*Proof:* According to Theorem 9 and Theorem 10, the field size of the constructed  $(r; s)$ -PMDS sequence and thus the  $(r; s)$ -PMDS code is  $q^{R(r+s)}$ , where  $q$  is the smallest prime power larger than  $n$  and  $R$  was defined above. Choosing  $R = r(m, s + 1, q^{r+s})$ , we get an  $(r; s)$  PMDS code with field size  $q^{(r+s)r(m, s+1, q^{r+s})}$ . If  $m \leq q^{r+s}$  then  $R = s$  and the field size becomes  $q^{s(r+s)} = \mathcal{O}(n^{s(r+s)})$ . Otherwise, according to the bound in Lemma 7, we choose  $R = 1 + \left\lceil \left(1 - \frac{1}{q^{r+s}}\right) (s - 1) \right\rceil \cdot \lceil \log_{q^{r+s}}(m) \rceil$  and thus the field size becomes

$$q^{(r+s) \left(1 + \left\lceil \left(1 - \frac{1}{q^{r+s}}\right) (s - 1) \right\rceil \cdot \lceil \log_{q^{r+s}}(m) \rceil\right)} = q^{r+s} m^{s-1} = \mathcal{O}(m^s).$$

In conclusion, it follows that the field size of the construction is  $\mathcal{O}(\max\{m, n^{r+s}\}^s)$ . ■

Corollary 11 states our best result for the field size of PMDS codes for arbitrary fixed  $r$  and  $s$ . In the rest of this section we consider some special cases of  $r$  and  $s$ . Let us revisit again the case  $r = 1$ . Recall from Corollary 5 that for  $r = 1$  the set  $T(S)$  (defined in Corollary 5) needs to be  $s$ -wise independent only over  $\mathbb{F}_2$ . Consequently, repeating the same ideas as in Theorem 9, we have the following corollary which holds for the case  $r = 1$ .

**Corollary 12.** *Let  $\mathcal{C}_{TP}$  be an  $[m, n; s, s + 1]$  erasure-correcting code over  $\mathbb{F}_2$  with redundancy  $\rho$  and parity check matrix  $H_{TP} = (\alpha_{1,1}, \dots, \alpha_{m,n}) \in \mathbb{F}_2^{\rho m}$ . Then, the sequence  $\Gamma_{TP} = (\alpha_{1,1}, \dots, \alpha_{m,n})$  is a  $(1; s)$ -PMDS sequence of size  $m \times n$ .*

The next corollary bounds the field size for the case  $r = 1$ . Similar to before, we first give the construction of an  $[m, n; s, s + 1]$  erasure-correcting code adopted from [10]:

- 1) Suppose  $H' \in \mathbb{F}_2^{\ell \times n}$  is a parity check matrix for an  $[n, n - \ell, s + 2]_2$  code.
- 2) Suppose  $H'' \in \mathbb{F}_2^{R \times m}$  is a parity check matrix for an  $[m, m - R, s + 1]_{2^\ell}$  code.

- 3) Representing every column in  $H'$  as an element in  $\mathbb{F}_{2^\ell}$ , we let  $\mathcal{C}_{TP}$  be a code over  $\mathbb{F}_2$  with the parity check matrix  $H'' \otimes H'$  formed by taking the tensor product of  $H''$  and  $H'$ , and redundancy  $\ell R$ .

Using Corollary 12, we get the following analog of Corollary 11 for the case  $r = 1$ . While this upper bound on the field size is loose in many cases, it is still possible to show that in several cases this result improves upon the bound from [5], where the authors showed an achievable field size of  $\mathcal{O}((mn)^{(s-1) \cdot (1 - \frac{1}{2^\ell})})$ .

**Corollary 13.** *There exists a  $(1; s)$  PMDS code with field size at most  $(2n)^{s \lceil \frac{s+1}{2} \rceil} m^{s-1}$ .*

*Proof:* Applying Lemma 7, we see that  $\ell \leq \lceil \frac{s+1}{2} \rceil \lceil \log_2(n) \rceil$ . As before, we can choose  $R = r(m, s + 1, 2^\ell)$ , and thus the code redundancy is given by

$$\begin{aligned} \ell \cdot r(m, s + 1, 2^\ell) &= \ell \cdot \left(1 + \left\lceil \left(1 - \frac{1}{2^\ell}\right) (s - 1) \right\rceil \cdot \lceil \log_2(m) \rceil\right) \\ &\leq \ell \cdot (1 + (s - 1) \cdot \lceil \log_2(m) \rceil), \end{aligned}$$

and the field size is at most

$$2^{\ell \cdot (1 + (s-1) \cdot \lceil \log_2(m) \rceil)} \leq (2n)^{s \lceil \frac{s+1}{2} \rceil} m^{s-1}. \quad \blacksquare$$

For the case of  $s = 3$ , we have the following corollary, which follows by combining the tensor product construction with the non-binary codes from [4]. A similar claim can be proved for the case  $s = 4$ .

**Corollary 14.** *There exists an  $(r; 3)$  PMDS code with field size  $\mathcal{O}(n^{3(r+3)})$  if  $m < n^3$  and otherwise  $\mathcal{O}(n^{r+3} m^{1.5})$ .*

*Proof:* Following the proof of Corollary 11, we deduce that the field size of the constructed  $(r; 3)$  PMDS code is given by  $q^{(r+3)r(m, 4, q^{r+3})}$ . Here we use the result from [4], which states that  $r(m, 4, q^{r+3}) = 1.5 \log_{q^{r+3}}(m) + 1$  for  $m > q^{r+3}$ , and  $r(m, 4, q^{r+3}) = 3$ , otherwise. In the former case, we get

$$\begin{aligned} q^{(r+3)r(m, 4, q^{r+3})} &= q^{(r+3)(1.5 \log_{q^{r+3}}(m) + 1)} = q^{r+3} m^{1.5} \\ &= \mathcal{O}(n^{r+3} m^{1.5}). \end{aligned}$$

#### ACKNOWLEDGMENT

This work was supported in part by NSF Grant CCF-1619053 and by the NISE program at SSC Pacific.

#### REFERENCES

- [1] M. Blaum, J.L. Hafner, and S. Hetzler, "Partial-MDS codes and their application to RAID type of architectures," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4510–4519, Mar. 2013.
- [2] M. Blaum, J.S. Plank, M. Schwartz, and E. Yaakobi, "Construction of partial MDS and sector-disk codes with two global parity symbols," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2673–2681, May 2016.
- [3] G. Calis and O. Koyluoglu, "A general construction for PMDS codes," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 452–455, Mar. 2017.
- [4] I. Dumer, "Nonbinary double-error-correcting codes designed by means of algebraic varieties," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1657–1666, Nov. 1995.
- [5] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin, "Explicit maximally recoverable codes with locality," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5245–5256, Jun. 2014.
- [6] E. M. Gabidulin, "Theory of codes with maximum rank distance," *Probl. Peredachi Inf.*, vol. 21, no. 1, pp. 3–16, Jul. 1985.
- [7] P. Gopalan, G. Hu, S. Kopparty, S. Saraf, C. Wang, and S. Yekhanin, "Maximally recoverable codes for grid-like topologies," *arXiv* at <https://arxiv.org/abs/1605.05412>, 2016.
- [8] G. Hu and S. Yekhanin, "New constructions of SD and MR codes over small finite fields," *arXiv* at <https://arxiv.org/abs/1605.02290>, 2016.
- [9] R. Roth, *Introduction to Coding Theory*, Cambridge University Press, 2006.
- [10] J.K. Wolf, "On codes derivable from the tensor product of check matrices," *IEEE Trans. Inf. Theory*, vol. 11, no. 2, pp. 281–284, Apr. 1965.