This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

New Families of MDS Symbol-Pair Codes from Matrix-Product Codes

Gaojun Luo, Martianus Frederic Ezerman, San Ling, and Xu Pan

***Abstract*—In emerging storage technologies, the outputs of the channels consist of overlapping pairs of symbols. The errors are no longer individual symbols. Controlling them calls for a different approach. Symbol-pair codes have been proposed as a solution. The error-correcting capability of such a code depends on its minimum pair distance instead of the usual minimum Hamming distance. Longer codes can be conveniently constructed from known shorter ones by a matrix-product approach. The parameters of a matrix-product code can be determined from the parameters of the ingredient codes. We construct a new family of maximum distance separable (MDS) symbol-pair matrix-product codes.**

Codes which are permutation equivalent to matrix-product codes may have improved minimum pair distances. We present four new families of MDS symbol-pair codes and a new family of almost MDS symbol-pair codes. The codes in these five new families are permutation equivalent to matrix-product codes. Each of our five constructions identifies permutations that can increase the minimum pair distances. We situate the new families among previously known families of MDS symbol-pair codes to highlight the versatility of our matrix-product construction route.

***Index Terms*—Matrix-product code, maximum distance separable code, symbol-pair code.**

I. Introduction

The traditional model for information transmission over noisy channels divides a message into individual information units. The writing and reading processes are done on individual symbols. Emerging storage technologies, however, come with a high write resolution and a low read resolution. Writing and reading individual symbols cannot be carried out consistently in channels whose outputs consist of overlapping pairs of symbols. To overcome the physical limitations on such channels, Casutto and Blaum in [3], followed by Casutto and Litsyn in [4], introduced and gave early constructions of symbol-pair codes. In this new framework, the outputs and the errors are no longer individual symbols. They are *overlapping pairs of adjacent symbols*.

G. Luo, M. F. Ezerman, and S. Ling are with the School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, e-mails: {gaojun.luo, fredezerman, lingsan}@ntu.edu.sg.

M. F. Ezerman is also with Sandhiguna, Suite 707 Graha Pena, Jalan Raya Batam Center, Teluk Tering, Kota Batam, Kepulauan Riau, 29461, Indonesia, e-mail: fred@sandhiguna.com.

X. Pan is with the School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China, e-mail: panxu@mails.ccnu.edu.cn.

G. Luo, M. F. Ezerman, and S. Ling are supported by Nanyang Technological University Research Grant No. 04INS000047C230GRT01. National Natural Science Foundation of China (Grant No. 11971175) partially supports S. Ling. X. Pan is supported by the State Scholarship Fund of China Scholarship Council (CSC).

A. Symbol-Pair Codes

Let Θ denote a code alphabet that consists of q elements. The *symbol-pair read* of a codeword $\mathbf{a} = (a_0, \ldots, a_{n-1}) \in \Theta^n$ in a *symbol-pair read channel* is

$$((a_0, a_1), \ldots, (a_{n-1}, a_0)).$$

Hence, each codeword $\mathbf{a} \in \Theta^n$ has a unique symbol-pair representation in $(\Theta, \Theta)^n$. To characterize the symbol-pair error-correcting capability, the *pair distance* between any two codewords $\mathbf{a} = (a_0, \ldots, a_{n-1})$ and $\mathbf{b} = (b_0, \ldots, b_{n-1})$ in Θ^n is defined as

$$d_P(\mathbf{a}, \mathbf{b}) = |\{0 \le i \le n-1 : (a_i, a_{i+1}) \ne (b_i, b_{i+1})\}|, \quad (1)$$

with the subscripts taken modulo n. The *pair weight* $w_P(\mathbf{a})$ of $\mathbf{a}$ is $d_P(\mathbf{a}, \mathbf{0})$. The set Θ^n equipped with the pair distance is a metric space [3, Section II.A]. We denote a symbol-pair code $\mathcal{C}$ by $(n, \kappa, d_P)_q$ when $\mathcal{C} \subseteq \Theta^n$ has size κ and *minimum pair distance*

$$d_P := \min\{d_P(\mathbf{a}, \mathbf{b}) : \mathbf{a}, \mathbf{b} \in \mathcal{C}, \mathbf{a} \ne \mathbf{b}\}. \quad (2)$$

The error-correcting capability of a symbol-pair code closely resembles that of a classical code as it is measured in terms of the minimum pair distance. It was established in [3] that an $(n, \kappa, d_P)_q$-code $\mathcal{C}$ can correct up to t pair errors if and only if $d_P \ge 2t + 1$. For fixed n and κ, our general aim is to construct an $(n, \kappa, d_P)_q$-symbol-pair code with as large d_P as possible. Chee *et al.* in [6] derived a Singleton-type bound for an $(n, \kappa, d_P)_q$-code that says

$$\kappa \le q^{n-d_P+2} \text{ for } q \ge 2 \text{ and } 2 \le d_P \le n. \quad (3)$$

A *maximum distance separable (MDS) symbol-pair code* is an $(n, \kappa, d_P)_q$-code with $\kappa = q^{n-d_P+2}$. Since κ is completely determined by q, n, and d_P, we often omit κ and write $(n, d_P)_q$-code instead of $(n, \kappa, d_P)_q$-code. The *Singleton defect* of a code is the difference $(n - d_P + 2) - \log_q \kappa$. A code is an *almost maximum distance separable (almost MDS) symbol-pair code* if its Singleton defect is 1.

B. Known Results

Explicit constructions of MDS symbol-pair codes are both theoretically and practically significant. We know from [6] that each classical MDS code of dimension greater than 1 leads to an MDS symbol-pair code. The same work gave several constructions of MDS symbol-pair codes by interleaving and extending classical MDS codes. Using tools from projective geometry, Ding *et al.* proposed another construction of MDS symbol-pair codes in [18]. Cyclic and constacyclic codes are

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

powerful ingredients in the construction of MDS symbol-pair codes because of their rich algebraic structures. Kai, Zhu, and Li in [26] constructed MDS symbol-pair codes of minimum pair distances $d_P \in \{5, 6\}$ from almost-MDS constacyclic codes. Following their idea, Li and Ge in [21] presented three new classes of MDS symbol-pair codes, also with minimum pair distances $d_P \in \{5, 6\}$. Using repeated root cyclic or constacyclic codes, many families of MDS symbol-pair codes with minimum pair distances $d_P \leq 12$ were built in [7], [9]–[12], [27], [30], [31]. In addition to giving construction methods, the respective authors of [9]–[12] completely determined the distance distributions of the constructed codes. Some families of MDS symbol-pair codes over finite rings derived from repeated root cyclic codes were presented in [13]–[16], [20].

Aside from the design of MDS symbol-pair codes, deriving good lower bounds for minimum pair distances of well-known families of linear codes is an important topic in the theory of symbol-pair codes. It plays a crucial role in our understanding of the error-correcting capability of the codes. Combining Discrete Fourier Transform (DFT) and the BCH bound, Cassuto and Blaum proved in [3] that a simple-root cyclic code with at least ℓ roots has $d_P \geq \ell + 2$. This lower bound improves to $\ell + 3$ whenever the Hartmann-Tzeng bound is applicable. Yaakobi, Bruck, and Siegel have shown in [36] that a binary cyclic code of dimension $k \geq 2$ and minimum Hamming distance d has minimum pair distance $d_P \geq d + \lceil d/2 \rceil$. Chen, Lin and Liu generalized the results of [3] by providing two lower bounds for the d_P of constacyclic codes and another lower bound for the d_P of repeated-root cyclic codes in [7]. Elishco, Gabrys, and Yaakobi in [17] employed a linear code with minimum Hamming distance d to construct codes with minimum pair distance $d_P \geq \lceil 3d/2 \rceil$. In [34], lower and upper bounds on pair weights of q-ary cyclic codes were provided by a geometric approach.

C. Our Contributions and Techniques

We construct new families of MDS symbol-pair codes. The techniques and results can be summarized as follows.

1. Matrix-product codes, which we will review in Section II, were introduced by Blackmore and Norton in [1] to construct longer codes from known shorter ones. Another early treatment was done by Özbudak and Stichtenoth in [32]. Some quasi-cyclic codes as well as generalized Reed-Muller codes can be written as matrix-product codes. We utilize matrix-product codes to construct MDS symbol-pair codes. This approach, to the best of our knowledge, has never been attempted before.
   Based on the minimum pair distance of shorter linear codes, we derive a lower bound for the minimum pair distance of matrix-product codes. Measured against this lower bound, we construct long symbol-pair codes with good minimum pair distance. For $q = 5$, Table II lists numerous symbol-pair codes with Singleton defect 0 or 1 that we construct in this paper. The total number of such q-ary symbol-pair codes increases superlinearly as q grows.
   We compile and present the parameters of known MDS symbol-pair codes in Table I with our results included. We put as Entry 6 a family of MDS symbol-pair codes of minimum pair distance $d_P = 6$ that we obtain based on this lower bound. The parameters of this family of codes are covered by the union of Entries 2 and 21 in Table I. Our construction method for Entry 6 is new if the length is smaller than $q + 2$, since it is based on classical *non-MDS codes*, whereas known codes of comparable parameters in Entry 21 came from classical MDS codes.
2. Very recently, Liu and Pan in [23] demonstrated that permuting the positions of the entries in a code can sometimes increase the pair distance. It is, therefore, natural to try using permutations on a code to improve its minimum pair distance. We start with matrix-product codes with large minimum pair distances. By identifying permutations, on the coordinates of the codes, that improve on the minimum pair distances, we present four families of MDS symbol-pair codes which are permutation equivalent to matrix-product codes. The resulting parameters are listed in five entries as Entries 7, 11, 12, 17, 22, and 23 in Table I. The table makes it clear that the parameters of these MDS symbol-pair codes are new. The parameters given in Entries 7, 11, and 17 outperform previously known ones, listed as Entries 4, 8 and 16, in terms of their alphabet sizes and on the flexible choices of their length.

After this introduction, Section II introduces preliminary concepts and collects useful known results on generalized Reed-Solomon codes and matrix-product codes. Section III establishes a lower bound on the minimum pair distance of matrix-product codes and, further, on that of a family of MDS symbol-pair codes. We use the bound to propose a direct construction of a new family of MDS symbol-pair MP codes. In Section IV we construct four new families of MDS symbol-pair codes and a new family of almost MDS symbol-pair codes from codes which are permutation equivalent to matrix-product codes. Section V concludes the paper. All computations are done in MAGMA V2.26-10 [2].

II. Preliminaries

Let $\mathbb{F}_q$ be the finite field with $q = p^t$ elements, where p is a prime and t is a positive integer, and let $\mathbb{F}_q^*$ be the multiplicative group of $\mathbb{F}_q$. We use the shorthand notation $[a] := \{1, 2, \ldots, a\}$ and $[a, b] := \{a, a+1, \ldots, b\}$, for integers $a < b$. A linear code $\mathcal{C}$ with parameters $[n, k, d]_q$ is a k-dimensional subspace of $\mathbb{F}_q^n$ with minimum Hamming distance d. Its dual code is

$$\mathcal{C}^\perp = \Bigg\{ \mathbf{x} = (x_0, \ldots, x_{n-1}) \in \mathbb{F}_q^n : \sum_{i=0}^{n-1} x_i c_i = 0 \quad \text{for all } \mathbf{c} = (c_0, \ldots, c_{n-1}) \in \mathcal{C} \Bigg\}. \quad (4)$$

Let $\mathbf{a} = (\theta_0, \ldots, \theta_{n-1})$ be a vector of length n, where $\theta_0, \ldots, \theta_{n-1}$ are distinct elements of $\mathbb{F}_q$, and let $\mathbf{u} = (u_0, \ldots, u_{n-1}) \in (\mathbb{F}_q^*)^n$. For $0 \leq k \leq n$, the *generalized*

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

TABLE I
PARAMETERS $(n, d_P)_q$ OF KNOWN $\left(n, q^{n-d_P+2}, d_P\right)_q$-MDS SYMBOL-PAIR CODES WITH $q = p^t$ FOR A PRIME p AND A POSITIVE INTEGER t.

No.	$(n, d_P)_q$	Constraints	Reference(s)
1	$(n, 5)_q$	$5 \le n \le q^2 + q + 1$	[18], [26]
2	$(n, 6)_q$	$\max\{6, q+2\} \le n \le q^2$	[7], [18]
3		$n = q^2 + 1$	[26]
4		$q = p$, $n = p^2 + p$, and $2 \nmid p$	[27]
5		$q = p$, $n = 2p^2 - 2p$, and $2 \nmid p$	[27]
6		$n = 3m$ and $m \in [3, q]$	Corollary 1
7		$n = q^2 + q$ and $q \neq 4, 5$, $q \neq 2^t$ with odd integer t	Theorem 15
8	$(n, 7)_q$	$q = p \ge 5$ and $n = 3p$	[7]
9		$q = p \ge 5$ and $n = 4p$	[27], [31]
10		$q = p$, $5 \mid (p-1)$, and $n = 5p$	[31]
11		$3 \mid (q-1)$, $2 \nmid q$, and $n = 3mp$, with $m \in [1, q/p]$	Theorem 10
12		$2 \mid q$ and $n = 2q + 2$	Theorem 7
13	$(n, 8)_q$	$q = p$, $3 \mid (p-1)$, and $n = 3p$	[7]
14		$q = p$, $5 \mid (p-1)$, and $n = 5p$	[31]
15		$2 \mid q$ and $n = 2m$, with $m \le q + 2$	[6]
16	$(n, 10)_q$	$q = p$, $3 \mid (p-1)$, and $n = 3p$	[30]
17		$3 \mid (q-1)$ and $n = 3mp$, with $m \in [1, q/p]$	Theorem 12
18	$(n, 12)_q$	$q = p$, $3 \mid (p-1)$, and $n = 3p$	[30]
19	$(2n, 2\ell)_q$	$3 \le \ell \le n - 1 \le q$	[6]
20	$(2n, 2n-4)_q$	$2 \mid q$ and $n \le q + 2$	[6]
21	$(n, \ell)_q$	$4 \le \ell \le n \le q + 1$	[6]
22	$(2n, 2\ell+1)_q$	$1 \le \ell \le n - 1 < q$	Theorem 7
23	$(2q+2, 2q-1)_q$	$2 \mid q$	Theorem 7

Reed-Solomon (GRS) code $GRS_k(\mathbf{a}, \mathbf{u})$ is given by

$$GRS_k(\mathbf{a}, \mathbf{u}) := \{(u_0 f(\theta_0), \ldots, u_{n-1} f(\theta_{n-1})) : f(x) \in \mathbb{F}_q[x], \text{ with } \deg(f(x)) < k\}, \quad (5)$$

We know, *e.g.*, from [24, Chapter 9] that $GRS_k(\mathbf{a}, \mathbf{u})$ is an $[n, k, n-k+1]_q$-MDS code with parity-check matrix

$$H = \begin{pmatrix} u'_0 & u'_1 & \cdots & u'_{n-1} \\ u'_0\theta_0 & u'_1\theta_1 & \cdots & u'_{n-1}\theta_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ u'_0\theta_0^{n-k-1} & u'_1\theta_1^{n-k-1} & \cdots & u'_{n-1}\theta_{n-1}^{n-k-1} \end{pmatrix},$$

for some $\mathbf{u}' = (u'_0, \ldots, u'_{n-1}) \in (\mathbb{F}_q^*)^n$. The dual code of $GRS_k(\mathbf{a}, \mathbf{u})$ is $GRS_k(\mathbf{a}, \mathbf{u})^\perp := GRS_{n-k}(\mathbf{a}, \mathbf{u}')$.

Let ω be an element with multiplicative order s in $\mathbb{F}_q^*$. An $[n, k, d]_q$-code $\mathcal{C}$, with $\gcd(n, q) = 1$, is *ω-constacyclic* if $(\omega c_{n-1}, c_0, \ldots, c_{n-2}) \in \mathcal{C}$ for each $(c_0, c_1, \ldots, c_{n-1}) \in \mathcal{C}$. We associate each codeword $(c_0, c_1, \ldots, c_{n-1}) \in \mathcal{C}$ with the polynomial $c(x) = c_0 + c_1 x + \ldots + c_{n-1}x^{n-1} \in \mathbb{F}_q[x]$. This representation allows for the identification of $\mathcal{C}$ as an ideal in $\mathbb{F}_q[x]/\langle x^n - \omega \rangle$. The monic polynomial $g(x)$ of degree $n - k$ in the ideal is a divisor of $x^n - \omega$ and is called the *generator polynomial* of $\mathcal{C}$. If $r = \mathrm{ord}_{sn}(q)$ and γ is a primitive n^{th} root of unity in $\mathbb{F}_{q^r}$, then the roots of $x^n - \omega$ are the elements $\beta\gamma^j$ for $j \in [0, n-1]$ with $\beta^n = \omega$. The collection $D = \{i \in [0, n-1] : g(\beta\gamma^i) = 0\}$ is the *defining set* of the ω-constacyclic code $\mathcal{C}$. If $n = q + 1$, $s = q - 1$, and β is a primitive element of $\mathbb{F}_{q^2}$, then we know from [28] that the ω-constacyclic code with generator polynomial $g(x) = (x - \beta)(x - \beta\gamma)$ has parameters $[q+1, q-1, 3]_q$ and a parity-check matrix

$$\begin{pmatrix} 1 & \beta & \cdots & \beta^q \\ 1 & \beta\gamma & \cdots & (\beta\gamma)^q \end{pmatrix}. \quad (6)$$

Let $A = (\alpha_{i,j})_{i\in[K],\, j\in[N]}$ be a $K \times N$ matrix over $\mathbb{F}_q$, with $K \le N$. For every $i \in [K]$, let $\mathcal{C}_i$ be an $[n, k_i, d_i]_q$-linear code. The *matrix-product (MP) code* $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ with *constituent codes* $\mathcal{C}_1, \cdots, \mathcal{C}_K$ is

$$\{(\mathbf{c}_1, \cdots, \mathbf{c}_K) \cdot A : \mathbf{c}_1 \in \mathcal{C}_1, \cdots, \mathbf{c}_K \in \mathcal{C}_K\} = \left\{ \left(\sum_{\ell=1}^{K} \mathbf{c}_\ell \, \alpha_{\ell,1}, \cdots, \sum_{\ell=1}^{K} \mathbf{c}_\ell \, \alpha_{\ell,N} \right) : \mathbf{c}_1 \in \mathcal{C}_1, \cdots, \mathbf{c}_K \in \mathcal{C}_K \right\}. \quad (7)$$

If G_i is a generator matrix of $\mathcal{C}_i$ for each $i \in [K]$, then the MP code $\mathcal{C}$ in (7) is $\mathbb{F}_q$-linear of length Nn with generator matrix

$$G = \begin{pmatrix} \alpha_{1,1}\, G_1 & \alpha_{1,2}\, G_1 & \cdots & \alpha_{1,N}\, G_1 \\ \alpha_{2,1}\, G_2 & \alpha_{2,2}\, G_2 & \cdots & \alpha_{2,N}\, G_2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{K,1}\, G_K & \alpha_{K,2}\, G_K & \cdots & \alpha_{K,N}\, G_K \end{pmatrix}. \quad (8)$$

Let $\ell \in [K]$ and let $1 \le i_1 < \ldots < i_\ell \le N$ be ℓ indices. For a $K \times N$ matrix A, the $\ell \times \ell$ submatrix $A\{[\ell]; \{i_1, \ldots, i_\ell\}\}$ of A is formed by the first ℓ rows and ℓ selected columns of A with indices $i_1, \ldots, i_\ell$, The matrix A is *nonsingular by columns* (NSC) if $A\{[\ell]; \{i_1, \ldots, i_\ell\}\}$ is nonsingular for each $\ell \in [K]$ and for each selection of column indices $1 \le i_1 < \ldots < i_\ell \le N$. Given a $K \times N$ NSC matrix A, there is no limitation on the numbers of columns if $K = 1$. For $K > 1$,

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

however, it was shown in [1] that there exists a $K \times N$ NSC matrix over $\mathbb{F}_q$ if and only if $K \leq N \leq q$.

Example 1. Let $\beta_1, \ldots, \beta_N$ be N distinct elements of $\mathbb{F}_q$. For $1 \leq K \leq N \leq q$, a well-known example of an NSC matrix is the Vandermonde matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_N \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1^{K-1} & \beta_2^{K-1} & \cdots & \beta_N^{K-1} \end{pmatrix}. \tag{9}$$

□

The next lemma presents the minimum Hamming distance and the dimension of an MP code.

Lemma 1. ([1], [32]) *Let $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ be an MP code. If $\mathrm{rank}(A) = K$, then $\dim(\mathcal{C}) = \sum_{i=1}^{K} k_i$. For each $i \in [K]$, let A_i be the submatrix formed by the first i rows of A and let $\mathcal{D}_i$ be the linear code with generator matrix A_i. If λ_i is the minimum Hamming distance of $\mathcal{D}_i$, then the minimum Hamming distance of $\mathcal{C}$ is $d(\mathcal{C}) \geq \min_{i \in [K]} \{d_i \lambda_i\}$, where d_i is the minimum Hamming distance of $\mathcal{C}_i$. Equality holds if $\mathcal{C}_1, \ldots, \mathcal{C}_K$ are nested codes, that is, $\mathcal{C}_K \subseteq \cdots \subseteq \mathcal{C}_1$. In addition, $\lambda_i = N - i + 1$, for each $i \in [K]$, provided that the matrix A is NSC.*

When A is a square matrix, the next lemma gives a known characterization of the dual $\mathcal{C}^{\perp}$ of an MP code $\mathcal{C}$.

Lemma 2. ([1]) *Let $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ be an MP code. Let $\mathcal{C}_i^{\perp}$ be the dual of $\mathcal{C}_i$, for each $i \in [K]$. Let $(A^{-1})^{\top}$ be the transpose of the inverse of A. If A is a nonsingular $N \times N$ matrix, then $\mathcal{C}^{\perp}$ is an MP code given by*

$$(\mathcal{C}_1^{\perp}, \cdots, \mathcal{C}_K^{\perp}) \cdot (A^{-1})^{\top}. \tag{10}$$

III. A Lower Bound on the Minimum Pair Distance of a Given MP Code

In this section we prove a lower bound on the minimum pair distance of any MP code based on the minimum pair distances of its constituent codes. We also present a family of MDS symbol-pair MP codes.

For a vector $\mathbf{a} \in \mathbb{F}_q^n$, recall that $w_H(\mathbf{a})$ and $w_P(\mathbf{a})$ denote the Hamming weight and the pair weight of $\mathbf{a}$, respectively. Cassuto and Blaum calculated in [3] that $w_P(\mathbf{a})$ is equal to the sum of $w_H(\mathbf{a})$ and another parameter, denoted by $L(S(\mathbf{a}))$, whose definition we now recall.

Definition 1. If B is a proper subset of $[0, n-1]$, then B can be partitioned into subsets $B_1, B_2, \ldots, B_k$ such that all elements of B_j, for $1 \leq j \leq k$, are consecutive modulo n. The *partition* of B is *minimal* if k is the smallest possible. The minimal partition of B is clearly unique, up to relabeling of the indices. Hence, we let $L(B)$ be the smallest integer that makes the partition of B minimal.

Example 2. For $n = 10$ with $B = \{0, 1, 2, 5, 6, 7, 9\}$, we have $L(B) = 2$ with $B = \{9, 0, 1, 2\} \cup \{5, 6, 7\}$. □

For a vector $\mathbf{a} = (a_0, \ldots, a_{n-1}) \in \mathbb{F}_q^n$ with $0 < w_H(\mathbf{a}) < n$, the *support* of $\mathbf{a}$ is $S(\mathbf{a}) := \{0 \leq i < n : a_i \neq 0\}$. We have $w_P(\mathbf{a}) = w_H(\mathbf{a}) + L(S(\mathbf{a}))$ from [3, Theorem 2]. It is also evident that $1 \leq L(S(\mathbf{a})) \leq w_H(\mathbf{a})$. For an $[n, k, d]_q$-code $\mathcal{C}$ with $d < n$, its minimum pair weight $d_P(\mathcal{C})$ is in $[d+1, \min\{n, 2d\}]$.

The following lemma helps in estimating the minimum pair distance of an MP code. Let $B = \{b_1, \ldots, b_e\}$ be a set of integers and let r be an integer. We use the notation $B + r$ for the set $\{b_1 + r, \ldots, b_e + r\}$.

Lemma 3. *Let B be a proper subset of $[0, m-1]$ with cardinality $|B| = g > 0$. Let $A = (a_{i,j})_{i \in [0,m-1], j \in [0,n-1]}$ be an $m \times n$ matrix over $\mathbb{F}_q$ such that the i^{th} row has ℓ nonzero elements for each $i \in B$ and the other entries are zero. If*

$$\mathbf{a} = (a_{0,0}, \ldots, a_{m-1,0}, \ldots, a_{0,n-1}, \ldots, a_{m-1,n-1})$$

is a vector of length mn, then $w_P(\mathbf{a}) \geq \ell(g + L(B))$.

Proof: Let $\mathbf{a}$ be an arbitrary vector of length mn that satisfies the conditions in the lemma. We can verify that $w_H(\mathbf{a}) = \ell g$ and $w_P(\mathbf{a}) = \ell g + L(S(\mathbf{a}))$. Our goal is to determine the minimum value of $L(S(\mathbf{a}))$. Let $B := \{h_1, \ldots, h_g\}$ and let the subsets $B_1, \ldots, B_k$ form the minimal partition of B as in Definition 1. We write $a_{i,j} = a_{i+mj}$ for each $i \in [0, m-1]$ and $j \in [0, n-1]$. If $h_1 > 0$, then the indices that signify the positions of nonzero elements in $\mathbf{a} = (a_0, a_1, \ldots, a_{mn-1})$ belong to the sets $B_i + jm$, again, with $i \in [k]$ and $j \in [0, n-1]$. We study the support of $\mathbf{a}$ by noting the possible positions of its nonzero entries among the entries written as $*$ in the expression

$$\mathbf{a} = (0, \ldots, 0, \overbrace{*, \ldots, *}^{B_1}, 0, \ldots, 0, \overbrace{*, \ldots, *}^{B_i + jm}, \\ 0, \ldots, 0, \overbrace{*, \ldots, *}^{B_k + (n-1)m}, \overbrace{0, \ldots, 0}^{(m-1-h_g)}). \tag{11}$$

We instantiate a set F. For each i and each j in the stipulated range, we include the set $B_i + jm$ as an element in F whenever $B_i + jm$ intersects the support of $\mathbf{a}$ nontrivially. Hence, we deduce that $L(S(\mathbf{a})) \geq |F|$. To determine the smallest feasible $|F|$, we need to establish the existence of a nonzero vector $\mathbf{a}$ whose support is covered by a minimum number of sets $B_i + jm$. We observe that $B_i, B_i + m, \ldots, B_i + (n-1)m$ cover the positions of $|B_i|\ell$ nonzero components of $\mathbf{a}$ for any $i \in [k]$. In addition, we need only ℓ out of the n sets $B_i, B_i + m, \ldots, B_i + (n-1)m$ to cover the $|B_i|\ell$ positions. Since $w_H(\mathbf{a}) = \ell g$ and $\sum_{i=1}^{k} |B_i| = g$, we get $|F| \geq \ell k$. Equality holds if $S(\mathbf{a}) = \bigcup_{i=1}^{k} \bigcup_{j=0}^{\ell-1} (B_i + jm)$. Thus, $w_P(\mathbf{a}) \geq \ell(g + L(B))$.

The case of $h_1 = 0$ can be similarly proven. The detail is omitted for brevity. ■

By Lemma 1, the minimum Hamming distance of an MP code is related to the minimum Hamming distances of its constituent codes. We obtain a similar result for the minimum pair distance of an MP code.

Theorem 4. *Let K and N be positive integers with $K \leq N$. For each $i \in [K]$, let $\mathcal{C}_i$ be an $[n, k_i, d_i]_q$-code with $d_i < n$ and minimum pair distance $d_P(\mathcal{C}_i)$. Let A be a $K \times N$ matrix of (full) rank K over $\mathbb{F}_q$ and let $\mathcal{C} = (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ be an MP code. Let ρ_i be the minimum Hamming distance of the*

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

$\mathbb{F}_q$-linear code generated by the first i rows of A. Then the minimum pair distance of $\mathcal{C}$ is

$$d_P(\mathcal{C}) \geq \min_{i\in[K]}\{d_P(\mathcal{C}_i)\rho_i\}. \tag{12}$$

If A is an NSC matrix, then $\rho_i = N-i+1$ for each $i \in [K]$, making $d_P(\mathcal{C}) \geq \min_{i\in[K]}\{d_P(\mathcal{C}_i)(N-i+1)\}$.

Proof: Let $A = (a_{i,j})_{i\in[K],j\in[N]}$ be given. By the definition of an MP code, each codeword $\mathbf{c} \in \mathcal{C}$ is of the form $\left(\sum_{\ell=1}^{K}\mathbf{c}_\ell a_{\ell,1},\ldots,\sum_{\ell=1}^{K}\mathbf{c}_\ell a_{\ell,N}\right)$, where $\mathbf{c}_\ell \in \mathcal{C}_\ell$ for each $\ell \in [K]$. Let $\mathbf{c}_t \neq \mathbf{0}$ and $\mathbf{c}_r = \mathbf{0}$ for any $r > t$. Writing $\mathbf{c}_\ell = (c_{\ell,1},\ldots,c_{\ell,n})$, we obtain $\mathbf{c} = \left(\sum_{\ell=1}^{t}c_{\ell,1}a_{\ell,1},\ldots,\sum_{\ell=1}^{t}c_{\ell,n}a_{\ell,1},\ldots,\sum_{\ell=1}^{t}c_{\ell,n}a_{\ell,N}\right)$, which, in an $n\times N$ matrix form, becomes

$$M_{\mathbf{c}} = \begin{pmatrix} \sum_{\ell=1}^{t}c_{\ell,1}a_{\ell,1} & \sum_{\ell=1}^{t}c_{\ell,1}a_{\ell,2} & \cdots & \sum_{\ell=1}^{t}c_{\ell,1}a_{\ell,N} \\ \sum_{\ell=1}^{t}c_{\ell,2}a_{\ell,1} & \sum_{\ell=1}^{t}c_{\ell,2}a_{\ell,2} & \cdots & \sum_{\ell=1}^{t}c_{\ell,2}a_{\ell,N} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{\ell=1}^{t}c_{\ell,n}a_{\ell,1} & \sum_{\ell=1}^{t}c_{\ell,n}a_{\ell,2} & \cdots & \sum_{\ell=1}^{t}c_{\ell,n}a_{\ell,N} \end{pmatrix}$$

We confirm that each row of $M_{\mathbf{c}}$ is a codeword of the linear code $\mathcal{D}_t$ whose generator matrix

$$G_t = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{t,1} & a_{t,1} & \cdots & a_{t,N} \end{pmatrix}$$

is formed by the first t rows of A. If $(c_{1,i},\ldots,c_{t,i}) \neq \mathbf{0}$, then the i^{th} row of $M_{\mathbf{c}}$ is a nonzero codeword of $\mathcal{D}_t$, since $\mathrm{rank}(G_t) = t$. Let us assume that the minimum Hamming distance of $\mathcal{D}_t$ is ρ_t. Since $\mathbf{c}_t = (c_{t,1},\ldots,c_{t,n}) \neq \mathbf{0}$, there are at least $w_H(\mathbf{c}_t)$ rows of $M_{\mathbf{c}}$ with the property that each row contains at least ρ_t nonzero elements. Applying Lemma 3, we derive $w_P(\mathbf{c}) \geq \rho_t(w_H(\mathbf{c}_t) + L(S(\mathbf{c}_t)))$. Since

$$d_P(\mathcal{C}_t) = \min_{\mathbf{c}_t\in\mathcal{C}_t\setminus\{\mathbf{0}\}}\{w_P(\mathbf{c}_t) = w_H(\mathbf{c}_t) + L(S(\mathbf{c}_t))\},$$

we arrive at $d_P(\mathcal{C}) \geq \rho_t\, d_P(\mathcal{C}_t)$ and, finally, conclude that $d_P(\mathcal{C}) \geq \min_{i\in[K]}\{d_P(\mathcal{C}_i)\,\rho_i\}$.

If the matrix A is NSC, then $\mathcal{D}_t$ is an $[N, t, \rho_t = N-t+1]_q$-MDS code by the linear independence of any t columns of G_t. This completes the proof. ∎

Remark 1. Choosing constituent codes with large minimum pair distances and using Theorem 4, we construct many long linear codes with large minimum pair distances. Let $q = 5$ and let A be a $K \times N$ NSC matrix. Choosing $\mathcal{C}_1,\ldots,\mathcal{C}_K$ to be random $[n, k_i, n-k_i+1]$-GRS codes with $1 < k_i \leq n \leq 5$ and $K \leq N \leq 5$, we obtain 64 symbol-pair codes which are optimal or almost optimal with respect to the Singleton-type bound in (3). The parameters of the constructed MP codes are listed in Table II and the lower bound on d_P follows immediately from Theorem 4. We see in the table that different (K, N) choices may lead to the same set of parameters n, k, and d_P. Deciding if the codes with the same parameters are equivalent lies outside the scope of our present work. More generally, we can let q grow and we use classical MDS codes over $\mathbb{F}_q$ as the constituent codes to derive MP codes with excellent pair distance properties. Figure 1 counts the number $\mathcal{N}$ of *distinct* sets of parameters n, k, and d_P for $q \in \{4, 5, 7, 8, 9\}$ to exhibit their growth. Asymptotically, the total number of optimal or almost optimal symbol-pair codes that we can construct by Theorem 4 increases superlinearly in q as the values of $\log_q(\mathcal{N})$ show. □

Example 3. Using binary codes

$$\mathcal{C}_1 = \{(000),(100),(011),(111)\},\ \mathcal{C}_2 = \{(000),(011)\},$$

and $A = \begin{pmatrix}1&0&1\\0&1&1\end{pmatrix}$, the MP code $\mathcal{C} = (\mathcal{C}_1, \mathcal{C}_2)\cdot A$ is

$$\begin{aligned}\mathcal{C} = \ & \{(000000000),(100011111),(000011011),\\ & (111000111),(011000011),(111011100),\\ & (100000100),(011011000)\}.\end{aligned}$$

We quickly determine that $d_P(\mathcal{C}_1) = 2$ and $d_P(\mathcal{C}_2) = 3$. The respective codes generated by A and $(1, 0, 1)$ have parameters $[3, 2, 2]_2$ and $[3, 1, 2]_2$. By Theorem 4, we know that $d_P(\mathcal{C}) \geq 4$. By direct observation, we confirm $d_P(\mathcal{C}) = 4$. □

Example 4. We use the $[5, 3, 3]_5$ and $[5, 4, 2]_5$-codes $\mathcal{C}_1$ and $\mathcal{C}_2$, with respective generator matrices

$$G_1 = \begin{pmatrix}1&0&0&1&2\\0&1&0&4&0\\0&0&1&4&3\end{pmatrix} \text{ and } G_2 = \begin{pmatrix}1&0&0&0&4\\0&1&0&0&1\\0&0&1&0&3\\0&0&0&1&1\end{pmatrix},$$

and the NSC matrix $A = \begin{pmatrix}1&1&1\\1&2&3\end{pmatrix}$ to construct the MP code $\mathcal{C} = (\mathcal{C}_1, \mathcal{C}_2)\cdot A$ whose generator matrix is

$$G = \begin{pmatrix} 1&0&0&1&2&1&0&0&1&2&1&0&0&1&2\\ 0&1&0&4&0&0&1&0&4&0&0&1&0&4&0\\ 0&0&1&4&3&0&0&1&4&3&0&0&1&4&3\\ 1&0&0&0&4&2&0&0&0&3&3&0&0&0&2\\ 0&1&0&0&1&0&2&0&0&2&0&3&0&0&3\\ 0&0&1&0&3&0&0&2&0&1&0&0&3&0&4\\ 0&0&0&1&1&0&0&0&2&2&0&0&0&2&2 \end{pmatrix}.$$

We have $d_P(\mathcal{C}_1) = 4$, $d_P(\mathcal{C}_2) = 3$, and $d_P(\mathcal{C}) = 8 > \min\{4(3-1+1), 3(3-2+1)\} = 6$. □

The following corollary to Theorem 4 gives a new family of MDS symbol-pair codes.

Corollary 1. *Let $q > 2$ be a prime power. Then there exists a $(3N, 6)_q$-MDS symbol-pair code for each $N \in [3, q]$.*

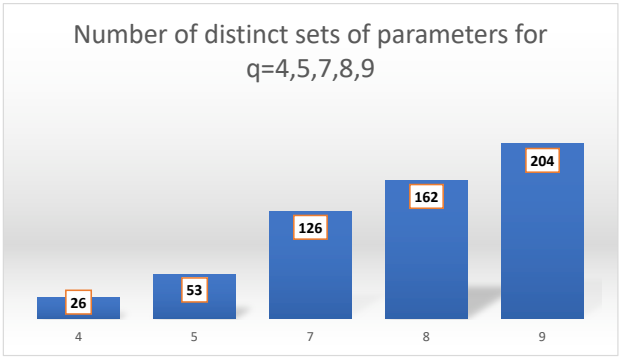
Proof: Let $\mathcal{C}_1 = \cdots = \mathcal{C}_{N-2}$ be the $[3, 3, 1]_q$-GRS code $GRS_3(\mathbf{a}, \mathbf{v})$ and let $\mathcal{C}_{N-1}$ be the $[3, 2, 2]_q$-GRS code $GRS_2(\mathbf{a}, \mathbf{v})$. The minimum pair distances of $GRS_3(\mathbf{a}, \mathbf{v})$ and $GRS_2(\mathbf{a}, \mathbf{v})$ are 2 and 3, respectively [6]. Let A be an $(N-1)\times N$ NSC matrix over $\mathbb{F}_q$. We can then define an MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_{N-1})\cdot A$. By Lemma 1, the code $\mathcal{C}$ has length $3N$ and dimension $3N-4$. It follows from Theorem 4 that $d_P(\mathcal{C}) \geq 6$. By the Singleton-type bound in (3), we conclude that $d_P(\mathcal{C}) = 6$. Thus, $\mathcal{C}$ is a $(3N, 6)_q$-MDS symbol-pair code. ∎

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

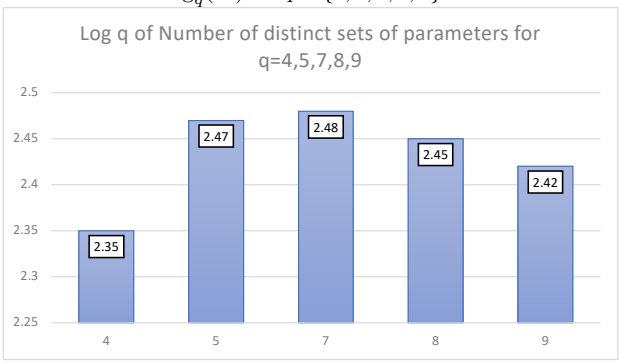
TABLE II
LIST OF SYMBOL-PAIR CODES OF LENGTH Nn, SIZE $\kappa = 5^k$, MINIMUM PAIR DISTANCE d_P, AND SINGLETON DEFECT δ FOR $q = 5$ FROM THEOREM 4.

No.	K	N	Nn	k	$d_P \geq$	δ	No.	K	N	Nn	k	$d_P \geq$	δ
1	2	2	6	4	3	≤ 1	33	3	4	12	8	6	0
2			6	5	3	0	34			12	9	4	≤ 1
3			8	5	4	≤ 1	35			16	9	8	≤ 1
4			8	6	3	≤ 1	36			16	11	6	≤ 1
5			8	6	4	0	37	3	5	15	7	9	≤ 1
6			8	7	3	0	38			15	8	8	≤ 1
7			10	6	5	≤ 1	39	4	4	12	10	3	≤ 1
8			10	7	4	≤ 1	40			12	11	3	0
9			10	8	3	≤ 1	41			16	13	4	≤ 1
10			10	8	4	0	42			16	14	3	≤ 1
11			10	9	3	0	43			16	14	4	0
12	2	3	9	5	6	0	44			16	15	3	0
13			9	4	6	≤ 1	45			20	16	5	≤ 1
14			12	5	8	≤ 1	46			20	17	4	≤ 1
15			12	7	6	≤ 1	47			20	18	3	≤ 1
16	2	4	12	4	9	≤ 1	48			20	19	3	0
17			12	5	8	≤ 1	49	4	5	15	10	6	≤ 1
18			16	5	12	≤ 1	50			15	11	6	0
19			15	4	12	≤ 1	51			15	12	4	≤ 1
20	3	3	9	6	4	≤ 1	52			20	13	8	≤ 1
21			9	7	3	≤ 1	53			20	15	6	≤ 1
22			9	8	3	0	54	5	5	15	13	3	≤ 1
23			12	9	4	≤ 1	55			15	14	3	0
24			12	10	3	≤ 1	56			20	17	4	≤ 1
25			12	10	4	0	57			20	18	3	≤ 1
26			12	11	3	0	58			20	18	4	0
27			15	11	5	≤ 1	59			20	19	3	0
28			15	12	4	≤ 1	60			25	21	5	≤ 1
29			15	13	3	≤ 1	61			25	22	4	≤ 1
30			15	13	4	0	62			25	23	3	≤ 1
31			15	14	3	0	63			25	23	4	0
32	3	4	12	7	6	≤ 1	64			25	24	3	0

Fig. 1. The numbers $\mathcal{N}$ of distinct sets of parameters obtained by Theorem 4 and the values of $\log_q(\mathcal{N})$ for $q \in \{4, 5, 7, 8, 9\}$.



(a) Numbers $\mathcal{N}$



(b) Values of $\log_q(\mathcal{N})$

We give an example that illustrates Corollary 1.

Example 5. Let $N = 3$ and $q = 4$ and let w be a primitive element of $\mathbb{F}_4$. Let

$$\mathcal{C}_1 = GRS_3((0, 1, w), (1, 1, 1)),$$
$$\mathcal{C}_2 = GRS_2((0, 1, w), (1, 1, 1)), \text{ and}$$
$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & w & w^2 \end{pmatrix}.$$

The MP code $(\mathcal{C}_1, \mathcal{C}_2) \cdot A$ is a $[9, 5, 3]_4$-code with minimum pair distance $d_P = 6$ and generator matrix

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & w & 0 & 1 & w & 0 & 1 & w \\ 0 & 1 & w^2 & 0 & 1 & w^2 & 0 & 1 & w^2 \\ 1 & 1 & 1 & w & w & w & w^2 & w^2 & w^2 \\ 0 & 1 & w & 0 & w & w^2 & 0 & w^2 & 1 \end{pmatrix}.$$

Hence, the code $(\mathcal{C}_1, \mathcal{C}_2) \cdot A$ is a $(9, 6)_4$-MDS symbol-pair code. □

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

IV. MDS Symbol-Pair Codes from permutation equivalent Matrix-Product codes

We now investigate the minimum pair distance of codes that are permutation equivalent to MP codes to construct four new families of MDS symbol-pair codes and a new family of almost MDS symbol-pair codes.

Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two $\mathbb{F}_q$-linear codes. If there is a permutation τ on the coordinates of $\mathcal{C}_1$ that makes $\tau(\mathcal{C}_1) = \mathcal{C}_2$, then $\mathcal{C}_1$ and $\mathcal{C}_2$ are *permutation equivalent*. Equivalent codes have the same length, dimension, and minimum Hamming distance. Permutations, however, do not necessarily preserve pair distances [23]. Given a linear code with a small Singleton defect, that is, its minimum pair distance nearly meets the Singleton-type bound, an equivalent MDS symbol-pair code may be derivable by permuting the code's coordinate positions.

Cassuto and Blaum introduced an interleaving approach in the construction of new symbol-pair codes in [3].

Proposition 5. *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be, respectively, codes with parameters $[n, k_1, d]_q$ and $[n, k_2, d]_q$. Then the code*

$$\mathcal{C} := \{(c_{1,0}, c_{2,0}, \ldots, c_{1,n-1}, c_{2,n-1}) : (c_{i,0}, \ldots, c_{i,n-1}) \in \mathcal{C}_i \text{ for } i = 1, 2\} \quad (13)$$

has length $2n$, dimension $k_1 + k_2$ and minimum pair distance $2d$.

Applying Proposition 5 on classical MDS codes $\mathcal{C}_1$ and $\mathcal{C}_2$ results in the five families of MDS symbol-pair codes due to Chee *et al.* in [6]. We note that the code $\mathcal{C}$ in (13) is equivalent to the MP code $(\mathcal{C}_1, \mathcal{C}_2) \cdot A$ under a permutation τ when A is the identity matrix of order 2. Viewing a codeword $\mathbf{c}$ of $(\mathcal{C}_1, \mathcal{C}_2) \cdot A$ as a $2 \times n$ matrix, the codeword $\tau(\mathbf{c})$ is obtained from $\mathbf{c}$ by reading the matrix off by columns.

Motivated by the above construction, we study linear codes that are permutation equivalent to a given MP code. The following lemma is useful in determining the supports of the codewords of an MP code.

Lemma 6. *Let $\mathcal{C}_K \subseteq \cdots \subseteq \mathcal{C}_1$ be nested $\mathbb{F}_q$-linear codes of length n and let A be a $K \times N$ NSC matrix. We write a codeword $\mathbf{b}$ of the MP code $(\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ as $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_N)$, where each $\mathbf{b}_i$ is a vector of length n. For each $s \in [0, K-1]$, if there are exactly s zero vectors among the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_N$, then $\mathbf{b}_j$ is a codeword of $\mathcal{C}_{s+1}$, for every $j \in [N]$. If the number of zero vectors among $\mathbf{b}_1, \ldots, \mathbf{b}_N$ is greater than $K - 1$, then $\mathbf{b} = \mathbf{0}$.*

Proof: We prove for the case that $\mathcal{C}_K \neq \mathcal{C}_{K-1} \neq \cdots \neq \mathcal{C}_1$. The respective proofs of the other cases follow a similar route and are omitted for brevity.

Let the basis of $\mathcal{C}_i$ be $\{\mathbf{g}_1, \ldots, \mathbf{g}_{k_i}\}$ for $i \in [K]$. Since $\mathcal{C}_K \subsetneq \mathcal{C}_{K-1} \subsetneq \cdots \subsetneq \mathcal{C}_1$, we have

$$\mathcal{C}_i = \{u_{i,1}\mathbf{g}_1 + \ldots + u_{i,k_i}\mathbf{g}_{k_i} : u_{i,1}, \ldots, u_{i,k_i} \in \mathbb{F}_q\},$$

with $1 \leq k_K < k_{K-1} < \cdots < k_1$. Given $A = (a_{i,j})_{i \in [K], j \in [N]}$, for any codeword $\mathbf{b} = (\mathbf{b}_1, \cdots, \mathbf{b}_N)$ in the MP code $(\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$, the block $\mathbf{b}_i$ can be expressed as

$$\mathbf{b}_i = \sum_{z=1}^{K} \sum_{\ell=1+k_{z+1}}^{k_z} \left(\sum_{r=1}^{z} a_{r,i} u_{r,\ell} \right) \mathbf{g}_\ell, \text{ with } k_{K+1} = 0. \quad (14)$$

It is easy to see that $\mathbf{b}_i$ is a codeword of $\mathcal{C}_1$ for each $i \in [N]$. Recall that $A\{[t]; \{i_1, \ldots, i_t\}\}$ is a $t \times t$ submatrix of A formed by the first t rows and t selected columns of A with indices $i_1, \ldots, i_t$. If $\mathbf{b}_{i_1} = \cdots = \mathbf{b}_{i_s} = \mathbf{0}$, then

$$(u_{1,\ell}, \ldots, u_{g,\ell})\, A\{[g]; \{i_1, \ldots, i_g\}\} = \mathbf{0}, \quad \text{for } g \in [s] \text{ and } \ell \in [k_{g+1} + 1, k_g]. \quad (15)$$

Since A is NSC, we deduce that $(u_{1,\ell}, \ldots, u_{g,\ell}) = \mathbf{0}$ for $g \in [s]$ and $\ell \in [k_{g+1} + 1, k_g]$. The nonzero blocks of $\mathbf{b}$ are, therefore, in $\mathcal{C}_{s+1}$. If there are at least K zero vectors among $\mathbf{b}_1, \ldots, \mathbf{b}_N$, then, by the definition of an NSC matrix, we obtain $\mathbf{b}_j = \mathbf{0}$ for any $j \in [N]$. ∎

Using Lemma 6, we propose a new construction of symbol-pair codes whose parameters are different from those of the symbol-pair codes in Proposition 5.

Theorem 7. *Let $\mathcal{C}_2 \subseteq \mathcal{C}_1$ be MDS codes with respective parameters $[n, n-\ell, \ell+1]_q$ and $[n, n-\ell+1, \ell]_q$. If $q > 2$, then there exists a $(2n, 2\ell+1)_q$-MDS symbol-pair code.*

Proof: We select an $\alpha \in \mathbb{F}_q^*$ with $\alpha \neq 1$. We use the NSC matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & \alpha \end{pmatrix}$ over $\mathbb{F}_q$ to construct the MP code $\mathcal{C} := (\mathcal{C}_1, \mathcal{C}_2) \cdot A$. By Lemma 1, $\mathcal{C}$ has length $2n$ and dimension $2n - 2\ell + 1$. Let $\mathbf{c} = (\mathbf{c}_0, \mathbf{c}_1)$ be a codeword in $\mathcal{C}$, with $\mathbf{c}_0 = (c_{0,0}, \ldots, c_{0,n-1})$ and $\mathbf{c}_1 = (c_{1,0}, \ldots, c_{1,n-1})$. By Lemma 6, $\mathbf{c}_0 \in \mathcal{C}_2$ if $\mathbf{c}_1 = \mathbf{0}$ and $\mathbf{c}_1 \in \mathcal{C}_2$ if $\mathbf{c}_0 = \mathbf{0}$. Furthermore, $\mathbf{c}_0, \mathbf{c}_1 \in \mathcal{C}_1$ if $\mathbf{c}_0$ and $\mathbf{c}_1$ are nonzero vectors. Let τ be the permutation on the indices of $\mathbf{c}$ such that $\tau(\mathbf{c}) = (c_{0,0}, c_{1,0}, \ldots, c_{0,n-1}, c_{1,n-1})$. If exactly one of $\mathbf{c}_0$ and $\mathbf{c}_1$ is a zero vector, then $d_P(\tau(\mathbf{c})) \geq 2\ell+2$. If both $\mathbf{c}_0$ and $\mathbf{c}_1$ are nonzero, then $d_P(\tau(\mathbf{c})) \geq 2\ell + 1$ since $d(\tau(\mathbf{c})) \geq 2\ell$. Hence, $d_P(\tau(\mathcal{C})) \geq 2\ell+1$. By the Singleton-type bound, $\tau(\mathcal{C})$ is a $(2n, 2\ell+1)_q$-MDS symbol-pair code. ∎

Remark 2. We can always select nested $GRS_{n-\ell}(\mathbf{a}, \mathbf{v}) \subsetneq GRS_{n-\ell+1}(\mathbf{a}, \mathbf{v})$-codes of length $n \leq q$ to ensure the existence of a $(2n, 2\ell+1)_q$-MDS symbol-pair code whenever q is a prime power, $1 \leq \ell \leq n-1$, and $n \leq q$. For classical MDS codes over $\mathbb{F}_q$ of length $q+1$, we know, *e.g.*, from [19], that nested MDS codes $\mathcal{C}_1$ and $\mathcal{C}_2$ in Theorem 7 exist only when q is even and $\ell = q-1$ or $\ell = 3$. Thus, there exist MDS symbol-pair codes with parameters $(2q+2, 7)_q$ and $(2q+2, 2q-1)_q$. □

Example 6. Let $q = 5$, $\mathbf{a} := (0, 1, 2, 3, 4)$, and $\mathbf{u} := (1, 1, 1, 1, 1)$. Let $\mathcal{C}_1$ be the $[5, 3, 3]_5$-GRS code $GRS_3(\mathbf{a}, \mathbf{u})$ and let $\mathcal{C}_2$ be the $[5, 2, 4]_5$-GRS code $GRS_2(\mathbf{a}, \mathbf{u})$, which is a subset of $\mathcal{C}_1$. Using the NSC matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$, the code $\tau(\mathcal{C})$, as defined in Theorem 7, is permutation equivalent to an MP code. The code $\tau(\mathcal{C})$ is a $[10, 5, 4]_5$-code with generator

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ 0 & 0 & 1 & 1 & 4 & 4 & 4 & 4 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 & 2 & 4 & 3 & 1 & 4 & 3 \end{pmatrix}.$$

Since $d_P(\tau(\mathcal{C})) = 7$, we have here a $(10, 7)_5$-MDS symbol-pair code. □

As stated in the introduction, most known constructions of MDS symbol-pair codes in the literature were based on cyclic or constacyclic codes. These two families of codes have rich algebraic structures that greatly help in ensuring the MDS property. In the case of MP codes, we still have to check whether an MP code achieves the Singleton-type bound in (3) from its parity-check matrix. Lemma 2 gives us a parity-check matrix of an MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ when A is a square matrix. Using MDS codes as the constituent codes of the MP code $\mathcal{C} := (\mathcal{C}_1, \mathcal{C}_2) \cdot A$ in Theorem 7 ensures that $\mathcal{C}$ is an MDS symbol-pair code. It is natural to consider MP codes $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ with A being a square matrix of order $K > 2$. In such cases, we have more flexibility in choosing the constituent codes of the corresponding MP code $\mathcal{C}$. To construct MDS symbol-pair codes via the matrix product construction route, we would like to determine all feasible constituent codes of the MP codes.

Lemma 8. *Let $\mathcal{C}_i$ be a linear code with parameters $[n, k_i, d_i]_q$ for each $i \in [K]$ such that $\mathcal{C}_K \subseteq \cdots \subseteq \mathcal{C}_1$ and $3 \leq K \leq q$. Let A be a square matrix of order K over $\mathbb{F}_q$ with the NSC property. We use A and $\mathcal{C}_1, \cdots, \mathcal{C}_K$ to define an MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$. If a permutation equivalent code $\mathcal{E}$ of $\mathcal{C}$ is an MDS symbol-pair code with minimum pair distance $d_P \geq 6$, then the relationship between K and d_P is listed in Table III.*

Proof: Let $\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_K)$ be a codeword of $\mathcal{C}$. By Lemma 6, if there are $i-1$ zero vectors among the vectors $\mathbf{c}_1, \ldots, \mathbf{c}_K$ for each $i \in [K]$, then the Hamming weight $w_H(\mathbf{c}) \geq (K - i + 1)\, d_i$. Since $\mathcal{E}$ is permutation equivalent to $\mathcal{C}$ and is an MDS symbol-pair code with minimum pair distance d_P, it follows from the definition of pair weight that $d_P \leq 2(K-i+1)\, d_i$, which is equivalent to $d_i \geq \left\lceil \frac{d_P}{2(K-i+1)} \right\rceil$, for each $i \in [K]$. By the Singleton bound on the Hamming distance of linear codes, we have $k_i \leq n - \left\lceil \frac{d_P}{2(K-i+1)} \right\rceil + 1$. We note $\mathcal{E}$ has length Kn and dimension $\sum_{i=1}^{K} k_i$. By the bound in (3), we arrive at

$$\sum_{i=1}^{K} k_i = Kn - d_P + 2 \leq Kn + K - \sum_{i=1}^{K} \left\lceil \frac{d_P}{2(K-i+1)} \right\rceil, \tag{16}$$

from which we get

$$\sum_{i=1}^{K} \left\lceil \frac{d_P}{2(K-i+1)} \right\rceil - K \leq d_P - 2. \tag{17}$$

If $K \geq 5$, then the left hand side of (17) is greater than or equal to $\left(\sum_{j=1}^{5} \frac{d_P}{2j}\right) - 5$, which implies that $d_P \leq 21 + \frac{3}{17}$. If $K \geq 4$, then the left hand side of (17) is greater than or equal to $\left(\sum_{j=1}^{4} \frac{d_P}{2j}\right) - 4$, which implies that $d_P \leq 48$. If $K = 3$, the inequality (17) holds for each $d_P \geq 6$. The entries in Table III are the outcomes of substituting relevant values of d_P and K to (16). ■

Based on Lemma 8, one can write a heuristic that identifies the parameters of feasible constituent codes to produce all the possible MP codes which may meet the Singleton-type bound in (3). And then, we check whether these MP codes or their permutation equivalent codes indeed achieve the Singleton-type bound from their parity-check matrices. We propose the following heuristic to generate an MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ with the requirement that the nested constituent codes $\mathcal{C}_K \subseteq \cdots \subseteq \mathcal{C}_1$ are classical MDS codes with parameters $[n, k_i, d_i]_q$ for each $i \in [K]$ and A is a square matrix of order K over $\mathbb{F}_q$ with NSC property.

Step 1: Choose q, $K \in [3, q]$, and a target $d_P = Kn - \sum_{i=1}^{K} k_i + 2$. If q is odd, then the length n of a constituent code is $2 \leq n \leq q + 1$. If $q > 2$ is even, then the length n of a constituent code is $2 \leq n \leq q + 2$, with some well-known restriction on its dimension when $n = q + 2$.

Step 2: Guided by Lemma 8, we infer the inequalities

$$d_K \geq \left\lceil \frac{d_P}{2} \right\rceil, \quad d_{K-1} \geq \left\lceil \frac{d_P}{4} \right\rceil, \quad \ldots, \quad d_1 \geq \left\lceil \frac{d_P}{2K} \right\rceil. \tag{18}$$

Step 3: Since the constituent codes are all classical MDS codes, the sum of their dimensions is

$$\sum_{i=1}^{K} k_i = Kn - \sum_{i=1}^{K} d_i + K = Kn - d_P + 2. \tag{19}$$

Step 4: List tuples $(d_1, d_2, \ldots, d_K)$ of feasible minimum distances of the constituent codes such that both (18) and (19) are met. Since the codes are classical MDS, each tuple gives a set of parameters of the constituent codes.

Example 7. We choose $q = 11$, $n \geq 6$, $K = 3$, and $d_P = 10$. Hence, $(d_1 \geq 2, d_2 \geq 3, d_3 \geq 5)$ by (18). Taking (19) into account, the feasible tuples (d_1, d_2, d_3) are $(2, 3, 6)$, $(2, 4, 5)$, and $(3, 3, 5)$. The following are the three options for the respective parameters of the nested constituent codes.

- $[n, n-1, 2]_{11}$, $[n, n-2, 3]_{11}$, $[n, n-5, 6]_{11}$.
- $[n, n-1, 2]_{11}$, $[n, n-3, 4]_{11}$, $[n, n-4, 5]_{11}$.
- $[n, n-2, 3]_{11}$, $[n, n-2, 3]_{11}$, $[n, n-4, 5]_{11}$.

We will show in Theorem 12 that a permutation equivalent MP code from the third option generates an MDS symbol-pair code. □

Example 8. We choose $q = 11$, $n \geq 5$, $K = 4$, and $d_P = 10$. Hence, $(d_1 \geq 2, d_2 \geq 2, d_3 \geq 3, d_4 \geq 5)$ by (18). Taking (19) into account, the only feasible tuple is $(2, 2, 3, 5)$. The following is the choice of the parameters of the nested constituent codes.

$$[n, n-1, 2]_{11}, [n, n-1, 2]_{11}, [n, n-2, 3]_{11}, [n, n-4, 5]_{11}.$$

One can proceed with the construction of the desired MP codes or their permutation equivalent codes and check whether they can produce MDS symbol-pair codes or not. □

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

TABLE III
VALUES THAT d_P CAN TAKE, GIVEN $3 \leq K \leq q$, FOR A FIXED q, AS DEFINED IN LEMMA 8.

No.	K	Range for d_P
1	3	$d_P \geq 6$
2	4	$d_P \in \{6, 7, 8, 10, 11, 12, 14, 15, 16, 18, 20, 22, 23, 24, 48\}$
3	5	$d_P \in \{6, 7, 8, 10, 12, 16\}$
4	$6 \leq K \leq q$	$d_P \in \{6, 7, 8, 10, 12\}$

A. Constructions from MP codes with a square matrix A of order 3

We begin our treatment by considering a specific case of $d_P = 7$. By Lemma 8, if one wants to construct an MDS symbol-pair MP code $\mathcal{C} := (\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3) \cdot A$, then the constituent codes must be $\mathcal{C}_1 = \mathcal{C}_2$ with parameters $[n, n-1, 2]_q$ and $\mathcal{C}_3$ with parameters $[n, n-3, 4]_q$.

Next, we consider an MP code $(\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3) \cdot A$ over $\mathbb{F}_q$, where $3 \mid (q-1)$ and A is a 3×3 NSC matrix. Let $\mathbf{a} = (\beta_0, \ldots, \beta_{N-1})$ be a vector of length N, where $\beta_0, \ldots, \beta_{N-1}$ are distinct elements of $\mathbb{F}_q$. We use as $\mathcal{C}_1 = \mathcal{C}_2$ the GRS code $GRS_{N-1}(\mathbf{a}, \mathbf{v})$ with parity-check matrix $H_1 = (1, 1, \cdots, 1)$. Our $\mathcal{C}_3$ is the GRS code $GRS_{N-3}(\mathbf{a}, \mathbf{v})$ with parity-check matrix

$$H_3 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta_0 & \beta_1 & \cdots & \beta_{N-1} \\ \beta_0^2 & \beta_1^2 & \cdots & \beta_{N-1}^2 \end{pmatrix}.$$

Let α be an element of multiplicative order 3 in $\mathbb{F}_q$. We define the MP code

$$\mathcal{C} := (\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3) \cdot A, \text{ with } A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 \\ 1 & \alpha^2 & \alpha^4 \end{pmatrix}. \tag{20}$$

For a set $\mathcal{D}$ of vectors, we denote by $S(\mathcal{D})$ the set $\{S(\mathbf{a}) : \mathbf{a} \in \mathcal{D}\}$ that contains the support $S(\mathbf{a})$ of $\mathbf{a} \in \mathcal{D}$. To calculate the minimum pair distances of $\mathcal{C}$ and its permutation equivalent code $\mathcal{E}$, we need the following lemma.

Lemma 9. *Let $3 \mid (q-1)$ and let α be an element of multiplicative order 3 in $\mathbb{F}_q$. Let $\mathcal{C}$ be the MP code in (20). Let $\mathcal{D}_i := \{\mathbf{c} \in \mathcal{C} : w_H(\mathbf{c}) = i\}$, that is, $\mathcal{D}_i$ is the set of all codewords of Hamming weight i in $\mathcal{C}$. Let the coordinates of every codeword of $\mathcal{C}$ be indexed by the set $[0, 3N-1]$. Then we have the following results.*

1) *The support set $S(\mathcal{D}_4)$ is the union of two sets $\{\{i_1, i_2, i_3, i_4\} : jN \leq i_1 < i_2 < i_3 < i_4 < (j+1)N$ with $j = 0, 1, 2\}$ and $\{\{i_1, i_2, i_3, i_4\} : j_1 N \leq i_1 < i_2 < (j_1+1)N, j_2 N < i_3 < i_4 < (j_2+1)N$, and $\beta_{i_1 \bmod N} + \beta_{i_2 \bmod N} = \beta_{i_3 \bmod N} + \beta_{i_4 \bmod N}$ for $0 \leq j_1 \neq j_2 \leq 2\}$.*
2) *The support set $S(\mathcal{D}_5)$ is a subset of the union of two sets $\{\{i_1, i_2, i_3, i_4, i_5\} : jN \leq i_1 < i_2 < i_3 < i_4 < i_5 < (j+1)N$ for $j = 0, 1, 2\}$ and $\{\{i_1, i_2, i_3, i_4, i_5\} : j_1 N \leq i_1 < i_2 < i_3 < (j_1+1)N$ and $j_2 N < i_3 < i_4 < (j_2+1)N$ for $0 \leq j_1 \neq j_2 \leq 2\}$.*

Proof: By Lemma 1, we know that $\mathcal{C}$ is a $[3N, 3N-5, 4]_q$-code. Since α is of order 3, we have

$$(A^{-1})^\top = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \alpha^2 & \alpha \\ 1 & \alpha & \alpha^2 \end{pmatrix}.$$

By Lemma 2, a parity-check matrix of $\mathcal{C}$ is

$$H = \begin{pmatrix} H_1 & H_1 & H_1 \\ H_1 & \alpha^2 H_1 & \alpha H_1 \\ H_3 & \alpha H_3 & \alpha^2 H_3 \end{pmatrix} = \left(\mathbf{g}_{0,0}^\top, \cdots, \mathbf{g}_{0,N-1}^\top, \mathbf{g}_{1,0}^\top, \cdots, \mathbf{g}_{1,N-1}^\top, \mathbf{g}_{2,0}^\top, \cdots, \mathbf{g}_{2,N-1}^\top\right). \tag{21}$$

Let $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3)$ be a codeword of $\mathcal{C}$. If there is exactly one nonzero block among $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$, then it follows from Lemma 6 that $\mathbf{c}_j \in \mathcal{C}_3$ for each $j \in [3]$. We note that $\mathcal{C}_3$ is an $[N, N-3, 4]_q$-code. By computing the determinants, we infer that $\mathbf{g}_{j,t_1}^\top, \mathbf{g}_{j,t_2}^\top, \mathbf{g}_{j,t_3}^\top, \mathbf{g}_{j,t_4}^\top$ are $\mathbb{F}_q$-linearly dependent, for each $j \in [0, 2]$ and $0 \leq t_1 < t_2 < t_3 < t_4 < N$. If there exists exactly one block $\mathbf{c}_i = \mathbf{0}$, then, by Lemma 6, we know that $\mathbf{c}_j \in \mathcal{C}_2$ for each $j \in [3]$. Since $\mathcal{C}_2$ is an $[N, N-1, 2]_q$-code, by computing the determinants, we confirm that $\mathbf{g}_{j_1,t_1}^\top, \mathbf{g}_{j_1,t_2}^\top, \mathbf{g}_{j_2,t_3}^\top, \mathbf{g}_{j_2,t_4}^\top$ are $\mathbb{F}_q$-linearly dependent for each $j_1 \neq j_2 \in [0, 2]$, $0 \leq t_1 < t_2 < N$, and $0 \leq t_3 < t_4 < N$ if and only if $\beta_{t_1} + \beta_{t_2} = \beta_{t_3} + \beta_{t_4}$. Since column $\mathbf{g}_{j,t}^\top$ corresponds to the index $t + Nj$ of $\mathbf{c}$, Condition 1) holds. The fact that Condition 2) is met can be demonstrated in a similar way. ∎

By Lemma 1, the MP code $\mathcal{C}$ constructed in Lemma 9 has parameters $[3N, 3N-5, 4]_q$. By Theorem 4 and Lemma 9 we obtain $d_P(\mathcal{C}) = 5$, which is not optimal with respect to the Singleton-type bound. Based on $S(\mathcal{D}_4)$ and $S(\mathcal{D}_5)$ in Lemma 9, we identify suitable permutations on the coordinates of $\mathcal{C}$ that increase the minimum pair distance in such a way that we obtain a family of MDS symbol-pair codes. Each code $\mathcal{E}$ in this family is permutation equivalent to a code $\mathcal{C}$ from Lemma 9.

Theorem 10. *There exists a $(3mp, 7)_q$-MDS symbol-pair code for each $m \in [1, q/p]$ whenever q is a power of an odd prime p such that $3 \mid (q-1)$.*

Proof: We begin by proving the case of $m = 2$ and $q > p$. Let $\mathcal{C}$ be the MP code in (20). Let $N = 2p$ and let α be an element of multiplicative order 3 in $\mathbb{F}_q$. Let $x \in \mathbb{F}_q \setminus \mathbb{F}_p$ and let

$$\mathbf{a} = (0, \ldots, p-1, x, \ldots, x+p-1) \tag{22}$$

be a vector of length $2p$, that is, $\beta_i = i$ and $\beta_{i+p} = x + i$ for $i \in [0, p-1]$. By (21) in the proof of Lemma 9, $\mathcal{C}$ has a

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

parity-check matrix

$$H = \begin{pmatrix} H_1 & H_1 & H_1 \\ H_1 & \alpha^2 H_1 & \alpha H_1 \\ H_3 & \alpha H_3 & \alpha^2 H_3 \end{pmatrix} = \left(\mathbf{g}_{0,0}^\top, \cdots, \mathbf{g}_{0,2p-1}^\top, \mathbf{g}_{1,0}^\top, \cdots, \mathbf{g}_{1,2p-1}^\top, \mathbf{g}_{2,0}^\top, \cdots, \mathbf{g}_{2,2p-1}^\top\right). \tag{23}$$

Let the coordinates of a codeword $\mathbf{c}$ of $\mathcal{C}$ be indexed by $[0, 6p-1]$. For every $\ell \in [0, 6p-1]$, we write $\ell = i + 2pj$ with $i \in [0, 2p-1]$ and $j \in [0, 2]$. We define a permutation τ on $[0, 6p-1]$ as $\tau(i + 2pj) = j + 3i$. To be more specific, $\tau(\mathcal{C})$ has a parity-check matrix

$$\tau(H) = \left(\mathbf{g}_{0,0}^\top, \mathbf{g}_{1,0}^\top, \mathbf{g}_{2,0}^\top, \cdots, \mathbf{g}_{0,2p-1}^\top, \mathbf{g}_{1,2p-1}^\top, \mathbf{g}_{2,2p-1}^\top\right). \tag{24}$$

For any codeword $\tau(\mathbf{c})$ with $w_H(\tau(\mathbf{c})) = 4$, Lemma 9 says that the four nonzero terms of $\tau(\mathbf{c})$ cannot appear in four consecutive coordinates. This implies that $L(S(\tau(\mathbf{c}))) > 1$, which is equivalent to $d_P(\tau(\mathcal{C})) \geq 6$. Since the column vectors $\mathbf{g}_{0,0}^\top$, $\mathbf{g}_{1,0}^\top$, $\mathbf{g}_{0,1}^\top$, $\mathbf{g}_{1,1}^\top$ are $\mathbb{F}_q$-linearly dependent and the minimum Hamming distance of $\tau(\mathcal{C})$ is greater than or equal to 4, there exists a codeword in $\tau(\mathbf{c})$ of the form $(a, b, 0, c, d, 0, \ldots, 0)$, with $a, b, c, d \in \mathbb{F}_q^*$, which implies that $d_P(\tau(\mathcal{C})) = 6$.

Next, we consider a code $\mathcal{E}_\rho$ that is equivalent to $\tau(\mathcal{C})$ via a permutation ρ on $[0, 6p-1]$ defined by

$$\rho(j + 3i) = \begin{cases} j + 3((i+j) \bmod p), & \text{if } i \in [0, p-1], \\ j + 3\left(((i - p + j) \bmod p) + p\right), & \text{if } i \in [p, 2p-1]. \end{cases} \tag{25}$$

Let $\tau(\mathbf{c})$, written as

$$(c_{0,0}, c_{1,0}, c_{2,0}, c_{0,1}, c_{1,1}, c_{2,1}, \ldots, c_{0,p-1}, c_{1,p-1}, c_{2,p-1}, b_{0,p}, b_{1,p}, b_{2,p}, b_{0,p+1}, b_{1,p+1}, b_{2,p+1}, \ldots, b_{0,2p-1} b_{1,2p-1} b_{2,2p-1}),$$

be a codeword of $\tau(\mathcal{C})$ whose index (j, i) corresponds to $j+3i$ for each $j \in [0, 2]$ and $i \in [0, 2p-1]$. Hence, we can write

$$\rho(\tau(\mathbf{c})) = (c_{0,0}, c_{1,p-1}, c_{2,p-2}, c_{0,1}, c_{1,0}, c_{2,p-1}, \ldots, c_{0,p-1}, c_{1,p-2}, c_{2,p-3}, b_{0,p}, b_{1,2p-1}, b_{2,2p-2}, b_{0,p+1}, b_{1,p}, b_{2,2p-1}, \ldots, b_{0,2p-1} b_{1,2p-2} b_{2,2p-3}).$$

Using Lemma 9, we can confirm that $\rho(\tau(\mathcal{C}))$ does not contain any codeword of Hamming weight 4 or 5 whose nonzero entries appear in consecutive coordinates.

We can now prove that there is no codeword $\mathbf{u} \in \rho(\tau(\mathbf{c}))$ of Hamming weight 4 and $L(S(\mathbf{u})) = 2$. Thanks to the proof of Lemma 9, we know that $\mathbf{g}_{j_1,t_1}^\top$, $\mathbf{g}_{j_1,t_2}^\top$, $\mathbf{g}_{j_2,t_3}^\top$, $\mathbf{g}_{j_2,t_4}^\top$ are $\mathbb{F}_q$-linearly dependent for each $j_1 \neq j_2 \in [0, 2]$, $0 \leq t_1 < t_2 < 2p$, and $0 \leq t_3 < t_4 < 2p$, if and only if $\beta_{t_1} + \beta_{t_2} = \beta_{t_3} + \beta_{t_4}$, with $\beta_i = i$ and $\beta_{i+p} = x + i$ for $i \in [0, p-1]$. We divide our justification into the following nine cases.

Case 1: If the four nonzero components are $c_{0,i}$, $c_{1,(i-1) \bmod p}$, $c_{0,j}$ and $c_{1,(j-1) \bmod p}$ with $i \neq j$, then $i + j = i + j - 2$, which is impossible since p is odd.

Case 2: If the four nonzero components are $c_{1,(i-1) \bmod p}$, $c_{2,(i-2) \bmod p}$, $c_{1,(j-1) \bmod p}$ and $c_{2,(j-2) \bmod p}$ with $i \neq j$, then $i + j - 2 = i + j - 4$, which is a contradiction.

Case 3: If the four nonzero components are $c_{2,(i-2) \bmod p}$, $c_{0,i+1}$, $c_{2,(j-2) \bmod p}$ and $c_{0,j+1}$ with $i \neq j$, then $i + j - 4 = i + j + 2$, which is impossible since p is odd and $\gcd(3, p) = 1$.

Case 4: If the four nonzero components are $c_{0,i}$, $c_{1,(i-1) \bmod p}$, $b_{0,j}$ and $b_{1,(j-1) \bmod p+p}$, then $i + j + x = i + j + x - 2$, which is absurd.

Case 5: If the four nonzero components are $c_{1,(i-1) \bmod p}$, $c_{2,(i-2) \bmod p}$, $b_{1,(j-1) \bmod p+p}$ and $b_{2,(j-2) \bmod p+p}$, then we obtain the contradiction $i + j - 2 + x = i + j - 4 + x$.

Case 6: If the four nonzero components are $c_{2,(i-2) \bmod p}$, $c_{0,i+1}$, $b_{2,(j-2) \bmod p+p}$ and $b_{0,j+1}$, then $i + j - 4 + x = i + j + 2 + x$, which is impossible.

Case 7: If the four nonzero components are $b_{0,i}$, $b_{1,(i-1) \bmod p+p}$, $b_{0,j}$ and $b_{1,(j-1) \bmod p+p}$ with $i \neq j$, then $i + j + 2x = i + j + 2x - 2$, which is impossible.

Case 8: If the four nonzero components are $b_{1,(i-1) \bmod p+p}$, $b_{2,(i-2) \bmod p+p}$, $b_{1,(j-1) \bmod p+p}$ and $b_{2,(j-2) \bmod p+p}$ with $i \neq j$, then we have the contradiction $i + j + 2x - 2 = i + j + 2x - 4$.

Case 9: If the four nonzero components are $b_{2,(i-2) \bmod p+p}$, $b_{0,i+1}$, $b_{2,(j-2) \bmod p+p}$ and $b_{0,j+1}$ with $i \neq j$, then $i + j + 2x - 4 = i + j + 2x + 2$, which is absurd.

We can then conclude that the pair distance of the codewords of Hamming weight 4 or 5 in $\rho(\tau(\mathcal{C}))$ is ≥ 7. Since $\rho(\tau(\mathcal{C}))$ is a $[6p, 6p-5, 4]_q$-code, its codewords with Hamming weight 6 have pair distance $d_P \geq 7$. Thus, $d_P(\rho(\tau(\mathcal{C}))) \geq 7$. By the Singleton-type bound, $\rho(\tau(\mathcal{C}))$ is a $(6p, 7)_q$-MDS symbol-pair code.

Since $\mathbb{F}_q$ can be written as the union $\mathbb{F}_q := \bigcup_{i=1}^{q/p} (x_i + \mathbb{F}_p)$, one can select a vector $\mathbf{a}$ of length mp whose coordinates are chosen from the m distinct additive cosets in $\mathbb{F}_q$ arranged in order as in (22). Using a method analogous to the one above, we can show that there exists a $(3mp, 7)_q$-MDS symbol-pair code for each $m \in [1, q/p]$. ■

The following example illustrates Theorem 10.

Example 9. Let $p = 5$, $q = 25$, and $m = 2$. Let w be a primitive element of $\mathbb{F}_{25}$. Let $\mathcal{C}_1 = \mathcal{C}_2 := GRS_9(\mathbf{a}, \mathbf{u})$ be a $[10, 9, 2]_{25}$-code with parity-check matrix $H_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$ and let $\mathcal{C}_3 := GRS_7(\mathbf{a}, \mathbf{u})$ be a $[10, 7, 4]_{25}$-code with parity-check matrix

$$H_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & w & w+1 & w+2 & w+3 & w+4 \\ 0 & 1 & 2^2 & 3^2 & 4^2 & w^2 & (w+1)^2 & (w+2)^2 & (w+3)^2 & (w+4)^2 \end{pmatrix}.$$

Suppose that A is an NSC matrix with

$$(A^{-1})^\top = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & w^{16} & w^8 \\ 1 & w^8 & w^{16} \end{pmatrix}.$$

The MP code $\mathcal{C}$ constructed in Theorem 10 is a $[30, 25, 4]_{25}$-code with parity-check matrix in (26). We use colors to highlight our permutations τ and ρ on the coordinates $[0, 29]$ of $\mathcal{C}$ in (27).

Thus, we have $d_P(\rho(\tau(\mathcal{C}))) = 7$ and $\rho(\tau(\mathcal{C}))$ is a $(30, 7)_{25}$-MDS symbol-pair code. □

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

$$H = \begin{pmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1&1&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{8}&w^{16}&w^{16}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8} \\ 1&1&1&1&1&1&1&1&1&1&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{8}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16} \\ 0&1&2&3&4&w&w^{22}&w^{15}&w^{2}&w^{17}&0&w^{8}&w^{14}&w^{2}&w^{20}&w^{9}&2&w^{23}&w^{10}&w&0&w^{16}&w^{22}&w^{10}&w^{4}&w^{17}&w^{14}&w^{7}&3&w^{9} \\ 0&1&4&4&1&w^{2}&w^{20}&2&w^{4}&w^{10}&0&w^{8}&w^{20}&w^{20}&w^{8}&w^{10}&w^{4}&w^{14}&4&3&0&w^{16}&w^{4}&w^{4}&w^{16}&3&4&w^{22}&w^{20}&w^{2} \end{pmatrix}. \tag{26}$$

$$\begin{array}{c} 0,1,2,3,4,5,6,7,8,9,\ 10,11,12,13,14,15,16,17,18,19,\ 20,21,22,23,24,25,26,27,28,29, \\ \downarrow \tau \\ 0,10,20,\ 1,11,21,\ 2,12,22,\ 3,13,23,\ 4,14,24,\ \vdots\ 5,15,25,\ 6,16,26,\ 7,17,27,\ 8,18,28,\ 9,19,29, \\ \downarrow \rho \\ 0,14,23,\ 1,10,24,\ 2,11,20,\ 3,12,21,\ 4,13,22,\ \vdots\ 5,19,28,\ 6,15,29,\ 7,16,25,\ 8,17,26,\ 9,18,27. \end{array} \tag{27}$$

Shifting our attention to $d_P = 10$, for the the MP code $\mathcal{C} := (\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3) \cdot A$ we have three choices for the parameters of feasible constituent MDS codes.

- $\mathcal{C}_1 = \mathcal{C}_2$ is an $[n, n-2, 3]_q$-code and $\mathcal{C}_3$ is an $[n, n-4, 5]_q$-code.
- $\mathcal{C}_1$ is an $[n, n-1, 2]_q$-code, $\mathcal{C}_2$ is an $[n, n-3, 4]_q$-code, and $\mathcal{C}_3$ is an $[n, n-4, 5]_q$-code.
- $\mathcal{C}_1$ is an $[n, n-1, 2]_q$-code, $\mathcal{C}_2$ is an $[n, n-2, 3]_q$-codes, and $\mathcal{C}_3$ is an $[n, n-5, 6]_q$-code.

Our third new construction of MDS symbol-pair codes is for $d_P = 10$ and $K = 3$ as stated in Theorem 12. Let $\beta_0, \ldots, \beta_{N-1}$ be distinct elements of $\mathbb{F}_q$. Let $\mathbf{a} = (\beta_0, \ldots, \beta_{N-1})$ be a vector of length N over $\mathbb{F}_q$ and let $3 \mid (q-1)$. We use the GRS code $GRS_{N-2}(\mathbf{a}, \mathbf{v})$ as $\mathcal{B}_1 = \mathcal{B}_2$ and the GRS code $GRS_{N-4}(\mathbf{a}, \mathbf{v})$ as $\mathcal{B}_3$. Their respective parity-check matrices over $\mathbb{F}_q$ are

$$B_1 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta_0 & \beta_1 & \cdots & \beta_{N-1} \end{pmatrix} \text{ and } \quad B_3 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta_0 & \beta_1 & \cdots & \beta_{N-1} \\ \beta_0^2 & \beta_1^2 & \cdots & \beta_{N-1}^2 \\ \beta_0^3 & \beta_1^3 & \cdots & \beta_{N-1}^3 \end{pmatrix}.$$

Fixing an element α of multiplicative order 3 in $\mathbb{F}_q$, we define the MP code

$$\mathcal{B} := (\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3) \cdot A, \text{ with } A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 \\ 1 & \alpha^2 & \alpha \end{pmatrix}. \tag{28}$$

The following lemma determines the support of the codewords with Hamming weights up to 8 in $\mathcal{B}$.

Lemma 11. *Let q be a prime power with $3 \mid (q-1)$. Let $\mathcal{B}$ be the MP code in (28). Let $\mathcal{D}_i = \{\mathbf{c} \in \mathcal{B} : w_H(\mathbf{c}) = i\}$ and let the coordinates of each codeword of $\mathcal{B}$ be indexed by the elements in $[0, 3N-1]$. Then we obtain the following results.*

1) *The support set $S(\mathcal{D}_5)$ is $\{\{i_1, i_2, i_3, i_4, i_5\} : jN \le i_1 < i_2 < i_3 < i_4 < i_5 < (j+1)N$ and $j \in \{0, 1, 2\}\}$.*
2) *The support set $S(\mathcal{D}_6)$ is the union of two sets $\{\{i_1, i_2, i_3, i_4, i_5, i_6\} : jN \le i_1 < i_2 < i_3 < i_4 < i_5 < i_6 < (j+1)N$ for $j = 0, 1, 2\}$ and $\{\{i_1, i_2, i_3, i_4, i_5, i_6\} : j_1 N \le i_1 < i_2 < i_3 < (j_1+1)N, \quad j_2 N < i_4 < i_5 < i_6 < (j_2+1)N$, and $\beta_{i_1 \bmod N} + \beta_{i_2 \bmod N} + \beta_{i_3 \bmod N} = \beta_{i_4 \bmod N} + \beta_{i_5 \bmod N} + \beta_{i_6 \bmod N}$ for $0 \le j_1 \ne j_2 \le 2\}$.*
3) *The support set $S(\mathcal{D}_7)$ is a subset of the union of two sets $\{\{i_1, i_2, i_3, i_4, i_5, i_6, i_7\} : jN \le i_1 < i_2 < i_3 < i_4 < i_5 < i_6 < i_7 < (j+1)N$ for $j = 0, 1, 2\}$ and $\{\{i_1, i_2, i_3, i_4, i_5, i_6, i_7\} : j_1 N \le i_1 < i_2 < i_3 < (j_1+1)N$ and $j_2 N < i_4 < i_5 < i_6 < i_7 < (j_2+1)N$ for $0 \le j_1 \ne j_2 \le 2\}$.*
4) *The support set $S(\mathcal{D}_8)$ is a subset of the union of three sets $\{\{i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8\} : jN \le i_1 < i_2 < i_3 < i_4 < i_5 < i_6 < i_7 < i_8 < (j+1)N, j = 0, 1, 2\}$, $\{\{i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8\} : j_1 N \le i_1 < i_2 < i_3 < (j_1+1)N$ and $j_2 N < i_4 < i_5 < i_6 < i_7 < i_8 < (j_2+1)N$ for $0 \le j_1 \ne j_2 \le 2\}$, and $\{\{i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8\} : j_1 N \le i_1 < i_2 < i_3 < i_4 < (j_1+1)N$ and $j_2 N < i_5 < i_6 < i_7 < i_8 < (j_2+1)N$ for $0 \le j_1 \ne j_2 \le 2\}$.*

Proof: The proof can be completed by a method analogous to the one used in proving Lemma 9. ∎

Choosing suitable permutations on the coordinates of $\mathcal{B}$ gives us a new family of MDS symbol-pair codes.

Theorem 12. *Let $q > 4$ be a power of a prime p with $3 \mid (q-1)$. We let $p > 4$ if $q = p$. Then there exists a $(3mp, 10)_q$-MDS symbol-pair code for each $m \in [1, q/p]$.*

Proof: We start with the case of $m = 2$ and $q > p$. Let $\mathcal{B}$ be the MP code in (28). Let $x \in \mathbb{F}_q \setminus \mathbb{F}_p$ and let

$$\mathbf{a} = (0, \ldots, p-1, x, \ldots, x+p-1) \tag{29}$$

be a vector of length $2p$. By Lemma 1, $\mathcal{B}$ is a $[6p, 6p-8, 5]$-code. We retain the permutations τ and ρ from the proof of Theorem 10. For a codeword $\mathbf{c} = (c_{j,0}, \ldots, c_{j,p-1}, b_{j,p}, \ldots, b_{j,2p-1})_{j \in [0,2]}$ in $\mathcal{B}$, we infer that $\rho(\tau(\mathbf{c})) = (\widehat{\mathbf{c}}, \widehat{\mathbf{b}})$, where

$$\widehat{\mathbf{c}} = (c_{0,i}, c_{1,(i-1) \bmod p}, c_{2,(i-2) \bmod p})_{i \in [0,p-1]} \text{ and}$$
$$\widehat{\mathbf{b}} = (b_{0,i}, b_{1,(i-1) \bmod p+p}, b_{2,(i-2) \bmod p+p})_{i \in [p,2p-1]}.$$

Using Lemma 11 and the same argument as in the proof of Theorem 10, we get $d_P(\rho(\tau(\mathbf{c}))) \ge 10$ if the Hamming weight of $\mathbf{c}$ is an element in $[5, 8]$. Hence, $d_P(\rho(\tau(\mathcal{B}))) \ge 10$. The

Singleton-type bound confirms that $\rho(\tau(\mathcal{B}))$ is a $(6p, 10)_q$-MDS symbol-pair code.

Since we can write, in terms of cosets, $\mathbb{F}_q := \bigcup_{i=1}^{q/p} (x_i + \mathbb{F}_p)$, the proof for the case of $m = 1$ or $m > 2$ can be completed by letting $\mathbf{a}$ be a vector of length mp as in (29) whose coordinates are chosen from the m distinct additive cosets in $\mathbb{F}_q$. ∎

Example 10. Let $p = 5$, $q = 25$, $m = 2$ and let w be a primitive element of $\mathbb{F}_{25}$. Let $\mathcal{B}_1 = \mathcal{B}_2 := GRS_8(\mathbf{a}, \mathbf{u})$ be a $[10, 8, 3]_{25}$-code with parity-check matrix

$$H_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & w & w+1 & w+2 & w+3 & w+4 \end{pmatrix}$$

and let $\mathcal{B}_3 := GRS_7(\mathbf{a}, \mathbf{u})$ be a $[10, 6, 5]_{25}$-code with parity-check matrix

$$H_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & w & w+1 & w+2 & w+3 & w+4 \\ 0 & 1 & 2^2 & 3^2 & 4^2 & w^2 & (w+1)^2 & (w+2)^2 & (w+3)^2 & (w+4)^2 \\ 0 & 1 & 2^3 & 3^3 & 4^3 & w^3 & (w+1)^3 & (w+2)^3 & (w+3)^3 & (w+4)^3 \end{pmatrix}.$$

We use the NSC matrix A with

$$(A^{-1})^\top = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & w^{16} & w^8 \\ 1 & w^8 & w^{16} \end{pmatrix}$$

to build the $[30, 22, 5]_{25}$-code $\mathcal{C}$ with parity-check matrix in (30)

Let τ and ρ be the permutations defined in Example 9. After some computation, we verify that $\rho(\tau(\mathcal{C}))$ is a $(30, 10)_{25}$-MDS symbol-pair code. □

B. Constructions from MP codes with a square matrix A of order $p-1$

By Lemma 8, if one wants to construct an MDS symbol-pair code with $d_P = 7$, then the corresponding MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ with $K \geq 3$ must have constituent codes $\mathcal{C}_i$ with parameters $[n, n, 1]_q$ for $i \in [K-3]$, $\mathcal{C}_i$ with parameters $[n, n-1, 2]_q$ for $i \in \{K-2, K-1\}$, and $\mathcal{C}_K$ with parameters $[n, n-3, 4]_q$. Let the matrix A be defined by

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \alpha & \cdots & \alpha^{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{K-1} & \cdots & \alpha^{(K-1)(K-1)} \end{pmatrix},$$

where $K \mid (q-1)$ and α is an element of multiplicative order K in $\mathbb{F}_q$. Let $\mathbf{a} = (\beta_0, \ldots, \beta_{N-1})$ be a vector of length N whose entries $\beta_0, \ldots, \beta_{N-1}$ are distinct elements of $\mathbb{F}_q$. Let $\mathcal{C}_i$ be the $GRS_N(\mathbf{a}, \mathbf{v})$ for each $i \in [K-3]$. We use as $\mathcal{C}_{K-2} = \mathcal{C}_{K-1}$ the GRS code $GRS_{N-1}(\mathbf{a}, \mathbf{v})$ with parity-check matrix $(1, 1, \cdots, 1)$. The code $\mathcal{C}_K$ is the GRS code $GRS_{N-3}(\mathbf{a}, \mathbf{v})$ with parity-check matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta_0 & \beta_1 & \cdots & \beta_{N-1} \\ \beta_0^2 & \beta_1^2 & \cdots & \beta_{N-1}^2 \end{pmatrix}.$$

In Theorem 10, we construct an MDS symbol-pair code with $d_P = 7$ by using the MP code $(\mathcal{C}_{K-2}, \mathcal{C}_{K-1}, \mathcal{C}_K) \cdot A$ when $K = 3$. The MDS property of the code is proved by the $\mathbb{F}_q$-linear independence of some columns of a parity-check matrix of $\mathcal{C}$. Using a similar technique as in Theorem 10, we can obtain a $(Kmp, q^{Kmp-5}, 6)_q$ symbol-pair code for each $m \in [1, q/p]$ whenever q is a power of an odd prime p and $K > 2$ is even. This code is an almost MDS symbol-pair code. However, this code is not optimal as there exists better $(n, 6)_q$-MDS symbol-pair codes with $\max\{6, q+2\} \leq n \leq q^2$ in Table I. If $K > 3$ is odd, we cannot determine the minimum pair distance of the permutation equivalent code $\mathcal{E}$ of $\mathcal{C}$ since the $\mathbb{F}_q$-linear independence of some columns of a parity-check matrix of $\mathcal{C}$ is generally unknown. The searches that we have performed in MAGMA do not yield any MDS symbol-pair code with $d_P = 7$ for odd $K > 3$.

In what follows, we provide a construction of almost MDS symbol-pair codes with length $(p-1)q$ and $d_P = 7$. This construction produces good alternative codes to MDS symbol-pair codes due to the non existence of q-ary MDS symbol-pair codes of length $(p-1)q$ and minimum pair distance 7.

Theorem 13. *Let $p > 3$ be a prime and let q be a power of p. Then there exists an $\left(m(p-1)p, q^{m(p-1)p-6}, 7\right)_q$-almost MDS symbol-pair code for each $m \in [1, q/p]$.*

Proof: Let $\beta_0, \ldots, \beta_{N-1}$ be distinct elements of $\mathbb{F}_q$ and define a vector $\mathbf{a} = (\beta_0, \ldots, \beta_{N-1})$ of length N. Let $\mathcal{C}_i$ be the $GRS_N(\mathbf{a}, \mathbf{v})$ for each $i \in [p-5]$. Let $\mathcal{C}_{p-4} = \mathcal{C}_{p-3} = \mathcal{C}_{p-2}$ be the GRS code $GRS_{N-1}(\mathbf{a}, \mathbf{v})$ with parity-check matrix $H_1 = (1, 1, \cdots, 1)$ and let $\mathcal{C}_{p-1}$ be the GRS code $GRS_{N-3}(\mathbf{a}, \mathbf{v})$ with parity-check matrix

$$H_2 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta_0 & \beta_1 & \cdots & \beta_{N-1} \\ \beta_0^2 & \beta_1^2 & \cdots & \beta_{N-1}^2 \end{pmatrix}.$$

Let α be a primitive element of $\mathbb{F}_p$ and

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \alpha & \cdots & \alpha^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{p-2} & \cdots & \alpha^{(p-2)(p-2)} \end{pmatrix}.$$

We construct an MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_{p-1}) \cdot A$ of length $(p-1)N$ and dimension $(p-1)N - 6$. Since

$$A^{-1} = \frac{1}{p-1} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \alpha^{-1} & \cdots & \alpha^{-(p-2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{-(p-2)} & \cdots & \alpha^{-(p-2)(p-2)} \end{pmatrix},$$

Lemma 2 confirms that the code $\mathcal{C}$ has parity-check matrix

$$H = \begin{pmatrix} H_1 & \alpha^4 H_1 & \cdots & \alpha^{4(p-2)} H_1 \\ H_1 & \alpha^3 H_1 & \cdots & \alpha^{3(p-2)} H_1 \\ H_1 & \alpha^2 H_1 & \cdots & \alpha^{2(p-2)} H_1 \\ H_2 & \alpha H_2 & \cdots & \alpha^{p-2} H_2 \end{pmatrix}. \tag{31}$$

We begin by proving the case $m = 2$. Let

$$\mathbf{a} = (0, \ldots, p-1, x, \ldots, x+p-1) \tag{32}$$

be a vector of length $2p$ with $x \in \mathbb{F}_q \setminus \mathbb{F}_p$. We express H as

$$\left(\mathbf{g}_{0,0}^\top, \cdots, \mathbf{g}_{0,2p-1}^\top, \cdots, \mathbf{g}_{p-2,0}^\top, \cdots, \mathbf{g}_{p-2,2p-1}^\top\right),$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

$$H = \begin{pmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
0&1&2&3&4&w&w^{22}&w^{15}&w^2&w^{17}&0&1&2&3&4&w&w^{22}&w^{15}&w^2&w^{17}&0&1&2&3&4&w&w^{22}&w^{15}&w^2&w^{17}\\
1&1&1&1&1&1&1&1&1&1&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^8\\
0&1&2&3&4&w&w^{22}&w^{15}&w^2&w^{17}&0&w^{16}&w^{22}&w^{10}&w^4&w^{17}&w^{14}&w^7&3&w^9&0&w^8&w^{14}&w^2&w^{20}&w^9&2&w^{23}&w^{10}&w\\
1&1&1&1&1&1&1&1&1&1&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^8&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}&w^{16}\\
0&1&2&3&4&w&w^{22}&w^{15}&w^2&w^{17}&0&w^8&w^{14}&w^2&w^{20}&w^9&2&w^{23}&w^{10}&w&0&w^{16}&w^{22}&w^{10}&w^4&w^{17}&w^{14}&w^7&3&w^9\\
0&1&4&4&1&w^2&w^{20}&2&w^4&w^{10}&0&w^8&w^{20}&w^{20}&w^8&w^{10}&w^4&w^{14}&4&3&0&w^{16}&w^4&w^4&w^{16}&3&4&w^{22}&w^{20}&w^2\\
0&1&3&2&4&w^3&3&w^{21}&2&w^3&0&w^8&w^2&w^{14}&w^{20}&w^{11}&w^2&w^5&w^{14}&w^{11}&0&w^{16}&w^{10}&w^{22}&w^4&w^{19}&w^{10}&w^{13}&w^{22}&w^{19}
\end{pmatrix}. \quad (30)$$

with

$$\mathbf{g}_{i,j} = (\alpha^{4i}, \alpha^{3i}, \alpha^{2i}, \alpha^i, \alpha^i j, \alpha^i j^2) \text{ and}$$
$$\mathbf{g}_{i,j+p} = (\alpha^{4i}, \alpha^{3i}, \alpha^{2i}, \alpha^i, \alpha^i(x+j), \alpha^i(x+j)^2)$$

for each $i \in [0, p-2]$ and $j \in [0, p-1]$. Let $\mathbf{c} = (\mathbf{c}_0, \cdots, \mathbf{c}_{p-2})$ be a codeword of $\mathcal{C}$ with $\mathbf{c}_i = (c_{i,0}, \cdots, c_{i,p-1}, b_{i,0}, \cdots, b_{i,p-1})$. One immediately confirms that the coordinates $c_{i,j}$ and $b_{i,j}$ correspond to respective columns $\mathbf{g}_{i,j}^\top$ and $\mathbf{g}_{i,j+p}^\top$ of H for each $i \in [0, p-2]$ and $j \in [0, p-1]$. Extending the permutations τ and ρ previously defined in Theorem 10, we get $\rho(\tau(\mathbf{c})) = (\widehat{\mathbf{c}}, \widehat{\mathbf{b}})$ with

$$\widehat{\mathbf{c}} = \left(c_{0,i \bmod p}, \ldots, c_{p-2,(i-p+2) \bmod p}\right)_{i \in [0,p-1]} \text{ and}$$
$$\widehat{\mathbf{b}} = \left(b_{0,i \bmod p}, \ldots, b_{p-2,(i-p+2) \bmod p}\right)_{i \in [0,p-1]}. \quad (33)$$

Next, we show that $d_P(\rho(\tau(\mathcal{C}))) = 7$ by considering the number of nonzero vectors $\mathbf{c}_i$ in a codeword $\mathbf{c} = (\mathbf{c}_0, \ldots, \mathbf{c}_{p-2})$.

Case 1: There is exactly one nonzero vector $\mathbf{c}_i$ among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{p-2}$. By Lemma 6, we know that $\mathbf{c}_i \in \mathcal{C}_{p-1}$. The corresponding codeword $\mathbf{c}$ of $\mathcal{C}$ has Hamming weight $w_H(\mathbf{c}) \geq 4$. It is clear from (33) that $L(S(\rho(\tau(\mathbf{c}))) \geq 4$, which is equivalent to $w_P(\rho(\tau(\mathbf{c}))) \geq 8$.

Case 2: There are exactly two nonzero vectors $\mathbf{c}_i$ and $\mathbf{c}_j$ among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{p-2}$. By Lemma 6, we infer that $\mathbf{c}_i, \mathbf{c}_j \in \mathcal{C}_{p-2}$. The Hamming weight of the corresponding codeword $\mathbf{c}$ of $\mathcal{C}$ is greater than or equal to 4. One can verify by (33) that $L(S(\rho(\tau(\mathbf{c})))) > 1$. If $w_H(\mathbf{c}) \geq 5$, then it is clear that $w_P(\rho(\tau(\mathbf{c}))) \geq 7$. Our next task is to prove, by contradiction, that there is no codeword $\mathbf{c}$ with $w_H(\mathbf{c}) = 4$ and $L(S(\rho(\tau(\mathbf{c})))) = 2$. Our justification is divided into three cases.

Case 2.1: If $w_H(\mathbf{c}) = 4$ and the four nonzero components of $\rho(\tau(\mathbf{c}))$ are

$$c_{i,j_1 \bmod p},\ c_{i+1,(j_1-1) \bmod p},\ c_{i,j_2 \bmod p},\ c_{i+1,(j_2-1) \bmod p},$$

then the corresponding columns

$$\mathbf{g}_{i,j_1 \bmod p}^\top,\ \mathbf{g}_{i+1,(j_1-1) \bmod p}^\top,\ \mathbf{g}_{i,j_2 \bmod p}^\top,\ \mathbf{g}_{i+1,(j_2-1) \bmod p}^\top$$

are $\mathbb{F}_q$-linearly dependent. This is contrary to the fact that

$$\begin{vmatrix} \alpha^i & \alpha^{i+1} & \alpha^i & \alpha^{i+1} \\ 1 & 1 & 1 & 1 \\ j_1 & j_1-1 & j_2 & j_2-1 \\ j_1^2 & (j_1-1)^2 & j_2^2 & (j_2-1)^2 \end{vmatrix} = 2\alpha^i(\alpha-1)(j_2-j_1)^2 \neq 0.$$

Case 2.2: If $w_H(\mathbf{c}) = 4$ and the four nonzero components of $\rho(\tau(\mathbf{c}))$ are

$$c_{p-2,1},\ b_{0,0},\ b_{p-2,j \bmod p},\ b_{0,(j-1) \bmod p},$$

then the corresponding columns

$$\mathbf{g}_{p-2,1}^\top,\ \mathbf{g}_{0,p}^\top,\ \mathbf{g}_{p-2,j \bmod p+p}^\top,\ \mathbf{g}_{0,(j-1) \bmod p+p}^\top$$

are $\mathbb{F}_q$-linearly dependent, which contradicts

$$\begin{vmatrix} \alpha^{p-2} & 1 & \alpha^{p-2} & 1 \\ 1 & 1 & 1 & 1 \\ 1 & x & j+x & j-1+x \\ 1 & x^2 & (j+x)^2 & (j-1+x)^2 \end{vmatrix} =$$
$$(\alpha^{p-2}-1)(j-1)(j+x-1)(x-2) \neq 0.$$

Case 2.3: For the other cases of $w_H(\mathbf{c}) = 4$, using the same argument as in Cases 2.1 and 2.2, we reach a contradiction. In addition, the codeword $\rho(\tau(\mathbf{c}))$ with four nonzero components $c_{0,0}$, $c_{1,p-1}$, $c_{0,p-1}$ and $c_{1,0}$ confirms that $w_P(\rho(\tau(\mathbf{c}))) = 7$.

Case 3: There are three or four nonzero vectors among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{p-2}$. By Lemma 6, we know that these nonzero vectors belong to $\mathcal{C}_{p-3} = \mathcal{C}_{p-4}$. This implies that the corresponding codeword $\mathbf{c}$ of $\mathcal{C}$ has Hamming weight $w_H(\mathbf{c}) \geq 6$. Hence, we conclude that $w_P(\rho(\tau(\mathbf{c}))) \geq 7$.

Case 4: There are exactly five nonzero vectors among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{p-2}$. By Lemma 6, these nonzero vectors are in the $[n, n, 1]$-code $\mathcal{C}_{p-5}$. The corresponding codeword $\mathbf{c}$ of $\mathcal{C}$ has Hamming weight $w_H(\mathbf{c}) \geq 5$. If $w_H(\mathbf{c}) \geq 6$, then $w_P(\rho(\tau(\mathbf{c}))) \geq 7$. We show that there does not exist any codeword of Hamming weight 5 in $\rho(\tau(\mathcal{C}))$ with five consecutive nonzero entries, as the following four cases reveal.

Case 4.1: If $w_H(\mathbf{c}) = 5$ and the five nonzero components of $\rho(\tau(\mathbf{c}))$ are

$$c_{i,j \bmod p},\ c_{i+1,(j-1) \bmod p},\ c_{i+2,(j-2) \bmod p},$$
$$c_{i+3,(j-3) \bmod p},\ c_{i+4,(j-4) \bmod p},$$

then the corresponding columns

$$\mathbf{g}_{i,j \bmod p}^\top,\ \mathbf{g}_{i+1,(j-1) \bmod p}^\top,\ \mathbf{g}_{i+2,(j-2) \bmod p}^\top,$$
$$\mathbf{g}_{i+3,(j-3) \bmod p}^\top,\ \mathbf{g}_{i+4,(j-4) \bmod p}^\top$$

are $\mathbb{F}_q$-linearly dependent, which contradicts

$$\begin{vmatrix} \alpha^{3i} & \alpha^{3(i+1)} & \alpha^{3(i+2)} & \alpha^{3(i+3)} & \alpha^{3(i+4)} \\ \alpha^{2i} & \alpha^{2(i+1)} & \alpha^{2(i+2)} & \alpha^{2(i+3)} & \alpha^{2(i+4)} \\ \alpha^i & \alpha^{i+1} & \alpha^{i+2} & \alpha^{i+3} & \alpha^{i+4} \\ 1 & 1 & 1 & 1 & 1 \\ j & j-1 & j-2 & j-3 & j-4 \end{vmatrix} =$$
$$\alpha^{6i+4}(\alpha-1)^4(\alpha^2-1)^3(\alpha^3-1)^2 \neq 0. \quad (34)$$

Case 4.2: If $w_H(\mathbf{c}) = 5$ and the five nonzero components of $\rho(\tau(\mathbf{c}))$ are

$$c_{p-5,4},\ c_{p-4,3},\ c_{p-3,2},\ c_{p-2,1},\ b_{0,0},$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

then the corresponding columns

$$\mathbf{g}_{p-5,4}^{\top},\ \mathbf{g}_{p-4,3}^{\top},\ \mathbf{g}_{p-3,2}^{\top},\ \mathbf{g}_{p-2,1}^{\top},\ \mathbf{g}_{0,p}^{\top}$$

are $\mathbb{F}_q$-linearly dependent. This contradicts

$$\begin{vmatrix} \alpha^{3(p-5)} & \alpha^{3(p-4)} & \alpha^{3(p-3)} & \alpha^{3(p-2)} & \alpha^{3(p-1)} \\ \alpha^{2(p-5)} & \alpha^{2(p-4)} & \alpha^{2(p-3)} & \alpha^{2(p-2)} & \alpha^{2(p-1)} \\ \alpha^{p-5} & \alpha^{p-4} & \alpha^{p-3} & \alpha^{p-2} & \alpha^{p-1} \\ 1 & 1 & 1 & 1 & 1 \\ 4 & 3 & 2 & 1 & x \end{vmatrix} = \alpha^{6(p-5)+4}(\alpha-1)^4(\alpha^2-1)^2(\alpha^3-1)((\alpha^2-1)(\alpha^3-1)+x) \neq 0. \tag{35}$$

Case 4.3: In all other cases when $w_H(\mathbf{c}) = 5$, we proceed as in the proof of Cases 4.1 and 4.2 to derive a contradiction.

Case 5: There are more than five vectors among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{p-2}$. One can verify that the pair distance of the codeword $\rho(\tau(\mathbf{c}))$ is ≥ 7 since its Hamming weight is > 5.

In summary, in all of the above cases, we can show that the minimum pair distance of $\rho(\tau(\mathcal{C}))$ is 7 when $\mathcal{C}$ has length $2(p-1)p$ and dimension $2(p-1)p-6$. Since $\mathbb{F}_q$ can be expressed as $\mathbb{F}_q := \bigcup_{i=1}^{q/p}(x_i + \mathbb{F}_p)$, one can construct a vector $\mathbf{a}$ of length mp whose coordinates are chosen from the m distinct additive cosets in $\mathbb{F}_q$, arranged in order as in (32). In a similar manner, we can demonstrate the existence of an $\left(m(p-1)p, q^{m(p-1)p-6}, 7\right)_q$-almost MDS symbol-pair code for each $m \in [1, q/p]$. ■

Example 11. Let $q = p = 5$. Let $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{C}_3 := GRS_4(\mathbf{a}, \mathbf{u})$ be a $[5, 4, 2]_5$-code with parity-check matrix

$$H_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

and let $\mathcal{C}_4 := GRS_2(\mathbf{a}, \mathbf{u})$ be a $[5, 2, 4]_5$-code with parity-check matrix

$$H_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2^2 & 3^2 & 4^2 \end{pmatrix}.$$

We can define an NSC matrix A such that

$$(A^{-1})^{\top} = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 3^2 & 3^3 \\ 1 & 3^2 & 3^4 & 3^6 \\ 1 & 3^3 & 3^6 & 3^9 \end{pmatrix}.$$

The MP code $\mathcal{C}$ constructed in Theorem 13 has parameters $[20, 14, 4]_5$ and parity-check matrix

$$H = \begin{pmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&3&3&3&3&3&4&4&4&4&4&2&2&2&2&2 \\ 1&1&1&1&1&4&4&4&4&4&1&1&1&1&1&4&4&4&4&4 \\ 1&1&1&1&1&2&2&2&2&2&4&4&4&4&4&3&3&3&3&3 \\ 0&1&2&3&4&0&2&4&1&3&0&4&3&2&1&0&3&1&4&2 \\ 0&1&4&4&1&0&2&3&3&2&0&4&1&1&4&0&3&2&2&3 \end{pmatrix}.$$

The permutations τ and ρ on the coordinates $[0, 19]$ of $\mathcal{C}$ defined in Theorem 13 are as follows.

$$\begin{gathered} 0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19 \\ \downarrow \tau \\ 0,5,10,15,\quad 1,6,11,16,\quad 2,7,12,17,\quad 3,8,13,18,\quad 4,9,14,19 \\ \downarrow \rho \\ 0,9,13,17,\quad 1,5,14,18,\quad 2,6,10,19,\quad 3,7,11,15,\quad 4,8,12,16. \end{gathered}$$

After some computation, we confirm that $\rho(\tau(\mathcal{C}))$ is a $(20, 5^{14}, 7)_5$-almost MDS symbol-pair code. □

Remark 3. In Theorem 13, we construct an $\left(m(p-1)p, q^{m(p-1)p-6}, 7\right)_q$ symbol-pair code, for each $m \in [1, q/p]$, by an MP code $(\mathcal{C}_1, \cdots, \mathcal{C}_{p-1}) \cdot A$, that is , with $K = p-1$. For some $K \neq p-1$ such that $K \mid (q-1)$, we have confirmed by computation in MAGMA that there also exist some symbol-pair codes with parameters $\left(mpK, q^{mpK-6}, 7\right)_q$, *e.g.*, $\left(52, 13^{46}, 7\right)_{13}$ with $K = 4$ and $q = p = 13$. However, we cannot prove the existence of almost MDS symbol-pair codes when $K \neq p-1$. Whether the corresponding determinants in the form of (34) and (35) are nonzero or not is unknown in the general case.

C. Constructions from MP codes using a square matrix A of order q

For an MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_K) \cdot A$ with $K \geq 3$ to have $d_P = 6$, its constituent MDS codes must come in one of the following possibilities.

- $\mathcal{C}_i$ is an $[n, n, 1]_q$-code for each $i \in [K-2]$ and $\mathcal{C}_{K-1} = \mathcal{C}_K$ is an $[n, n-2, 3]_q$-code.
- $\mathcal{C}_i$ is an $[n, n, 1]_q$-code for each $i \in [K-2]$, $\mathcal{C}_{K-1}$ is an $[n, n-1, 2]_q$-code, and $\mathcal{C}_K$ is an $[n, n-3, 4]_q$-code.
- $\mathcal{C}_i$ is an $[n, n, 1]_q$-code for each $i \in [K-3]$, $\mathcal{C}_{K-2} = \mathcal{C}_{K-1}$ is an $[n, n-1, 2]_q$-code, and $\mathcal{C}_K$ is an $[n, n-2, 3]_q$-code.

We show that the first possibility leads to MDS symbol-pair codes when $K = q$.

Proposition 14. *Let q be a prime power. Then there exists a $\left(q^2+q, 6\right)_q$-MDS symbol-pair code if one of the following statements holds.*

1) *q is odd and there exists a primitive element α of $\mathbb{F}_q$ such that*

$$(\alpha+1)(\alpha-3),\ \alpha(\alpha+4),\ \alpha^{q-2}(\alpha^{q-2}+4),\ (\alpha^{q-2}+1)(\alpha^{q-2}-3)$$

are either 0 or square elements of $\mathbb{F}_q$.

2) *q is even and there exists a primitive element α of $\mathbb{F}_q$ such that*

$$\mathrm{Tr}((1+\alpha)^{-1}) = \mathrm{Tr}(\alpha^{q-2}) = \mathrm{Tr}(\alpha) = \mathrm{Tr}((1+\alpha^{q-2})^{-1}) = 0,$$

where Tr *is the trace function from $\mathbb{F}_q$ to $\mathbb{F}_2$.*

Proof: 1) Let q be an odd prime power. Let α be a primitive element of $\mathbb{F}_q$ satisfying Statement 1) and let β be a primitive element of $\mathbb{F}_{q^2}$. Let $\mathcal{C}_i$ be a $[q+1, q+1, 1]_q$-code for each $i \in [q-2]$ and let $\mathcal{C}_{q-1} = \mathcal{C}_q$ be the β^{q+1}-constacyclic code with parameters $[q+1, q-1, 3]_q$ and parity-check matrix

$$H_1 = \begin{pmatrix} 1 & \beta & \cdots & \beta^q \\ 1 & \beta\gamma & \cdots & (\beta\gamma)^q \end{pmatrix},$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

defined in (6) with $\gamma = \beta^{q-1}$. The primitive element α may or may not be equal to β^{q+1}. Let

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & \alpha & \cdots & \alpha^{q-2} & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & \alpha^{q-2} & \cdots & \alpha^{(q-2)(q-2)} & 0 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

be an NSC matrix of order q. We define an MP code $\mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_q) \cdot A$ of length q^2+q and dimension q^2+q-4 and immediately confirm that

$$A^{-1} = \frac{1}{q-1} \begin{pmatrix} 0 & 1 & \cdots & 1 & 1 \\ 0 & \alpha^{-1} & \cdots & \alpha^{-(q-2)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & \alpha^{-(q-2)} & \cdots & \alpha^{-(q-2)(q-2)} & 1 \\ q-1 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

By Lemma 2, the code $\mathcal{C}$ has parity-check matrix

$$H = \begin{pmatrix} H_1 & \alpha H_1 & \cdots & \alpha^{q-2} H_1 & O \\ H_1 & H_1 & \cdots & H_1 & H_1 \end{pmatrix},$$

with O being the $2 \times (q+1)$ zero matrix. (36)

We denote the columns of H as

$$\left(\mathbf{g}_{0,0}^\top, \cdots, \mathbf{g}_{0,q}^\top, \cdots, \mathbf{g}_{q-1,0}^\top, \cdots, \mathbf{g}_{q-1,q}^\top\right)$$

with

$$\mathbf{g}_{i,j}^\top = \begin{cases} (\alpha^i\beta^j, \alpha^i(\beta\gamma)^j, \beta^j, (\beta\gamma)^j)^\top, & \text{if } i \in [0, q-2], j \in [0,q], \\ (0, 0, \beta^j, (\beta\gamma)^j)^\top, & \text{if } i = q-1, j \in [0,q]. \end{cases}$$

Let $\mathbf{c} = (\mathbf{c}_i)_{i\in[0,q-1]}$ be a codeword of $\mathcal{C}$ with $\mathbf{c}_i = (c_{i,0}, \ldots, c_{i,q})$. It is clear that $\sum_{i=0}^{q-1}\sum_{j=0}^{q} c_{i,j}\mathbf{g}_{i,j}^\top = \mathbf{0}$. Let ϕ be a permutation on the coordinates of $\mathcal{C}$ such that

$$\phi(\mathbf{c}) = (c_{0,j \bmod (q+1)}, c_{1,(j-1) \bmod (q+1)}, \cdots, c_{q-1,(j-q+1) \bmod (q+1)})_{j\in[0,q]}. \quad (37)$$

We establish $d_P(\phi(\mathcal{C})) = 6$ based on the number of nonzero vectors $\mathbf{c}_i$ in a codeword $\mathbf{c} = (\mathbf{c}_i)_{i\in[0,q-1]}$.

Case 1: There is exactly one nonzero vector $\mathbf{c}_i$ among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{q-1}$. By Lemma 6, we have $\mathbf{c}_i \in \mathcal{C}_q$. The corresponding codeword $\mathbf{c} \in \mathcal{C}$ has Hamming weight $w_H(\mathbf{c}) \geq 3$. It is clear, by (37), that $L(S(\phi(\mathbf{c}))) \geq 3$, which is equivalent to $w_P(\phi(\mathbf{c})) \geq 6$.

Case 2: There are exactly two nonzero vectors $\mathbf{c}_i$ and $\mathbf{c}_j$ among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{q-1}$. We infer by Lemma 6 that $\mathbf{c}_i, \mathbf{c}_j \in \mathcal{C}_{q-1}$. The Hamming weight of the corresponding codeword $\mathbf{c} \in \mathcal{C}$ is ≥ 6. This implies $w_P(\phi(\mathbf{c})) \geq 7$.

Case 3: There are exactly three nonzero vectors among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{q-1}$. By Lemma 6, the three nonzero vectors are in $\mathcal{C}_{q-2}$ whose parameters are $[q+1, q+1, 1]_q$. The corresponding codeword $\mathbf{c} \in \mathcal{C}$ has Hamming weight $w_H(\mathbf{c}) \geq 3$. If $w_H(\mathbf{c}) \geq 4$, then the fact that $w_P(\phi(\mathbf{c})) \geq 6$ follows from (37). We argue that there is no codeword $\mathbf{c}$ with $w_H(\mathbf{c}) = 3$ and $L(S(\phi(\mathbf{c}))) \leq 2$ by considering the following two cases.

Case 3.1: If $w_H(\mathbf{c}) = 3$ and the three nonzero components of $\phi(\mathbf{c})$ are

$$c_{i_1,j_1 \bmod (q+1)},\ c_{i_1+1,(j_1-1) \bmod (q+1)},\ c_{i_2,j_2},$$
with $i_1 \in [0, q-3]$, $i_2 \in [0, q-2]$, and $i_2 \neq i_1, i_1+1$,

then the corresponding columns

$$\mathbf{g}_{i_1,j_1 \bmod (q+1)}^\top,\ \mathbf{g}_{i_1+1,(j_1-1) \bmod (q+1)}^\top,\ \mathbf{g}_{i_2,j_2}^\top$$

are $\mathbb{F}_q$-linearly dependent. Thus,

$$\begin{aligned} &\begin{vmatrix} \alpha^{i_1}\beta^{j_1} & \alpha^{i_1+1}\beta^{j_1-1} & \alpha^{i_2}\beta^{j_2} \\ \beta^{j_1} & \beta^{j_1-1} & \beta^{j_2} \\ (\beta\gamma)^{j_1} & (\beta\gamma)^{j_1-1} & (\beta\gamma)^{j_2} \end{vmatrix} \\ &= \beta^{2j_1-1+j_2}((\alpha^{i_1+1}-\alpha^{i_1})(\gamma^{j_2}-\gamma^{j_1}) \\ &\quad -(\alpha^{i_2}-\alpha^{i_1})(\gamma^{j_1-1}-\gamma^{j_1})) \\ &= 0. \end{aligned} \quad (38)$$

This implies that

$$\frac{\gamma^{j_2}-\gamma^{j_1}}{\gamma^{j_1-1}-\gamma^{j_1}} \in \mathbb{F}_q^*,$$

which, since $\gamma^q = \gamma^{-1}$, is equivalent to

$$\frac{\gamma^{j_2}-\gamma^{j_1}}{\gamma^{j_1-1}-\gamma^{j_1}} = \frac{\gamma^{-j_2}-\gamma^{-j_1}}{\gamma^{-j_1+1}-\gamma^{-j_1}}.$$

Simplifying the above equation, we confirm that

$$(\gamma-1)(\gamma^{j_2-j_1+1}-1)(\gamma^{j_2-j_1}-1) = 0. \quad (39)$$

Noticing that $(j_2-j_1) \bmod (q+1) \in [0,q]$, we have two scenarios. If $(j_2-j_1) \bmod (q+1) \in [0, q-1]$, then we have a contradiction as (39) fails to hold. If $(j_2-j_1) \bmod (q+1) = q$, then we get that

$$\begin{aligned} &(\alpha^{i_1+1}-\alpha^{i_1})(\gamma^{j_2}-\gamma^{j_1}) - (\alpha^{i_2}-\alpha^{i_1})(\gamma^{j_1-1}-\gamma^{j_1}) \\ &\qquad = (\alpha^{i_1+1}-\alpha^{i_2})(\gamma^{j_1-1}-\gamma^{j_1}) \neq 0, \end{aligned}$$

which contradicts to (38).

Case 3.2: For other cases of $w_H(\mathbf{c}) = 3$, we can use the same argument as in Case 3.1 to reach a contradiction.

Case 4: There are exactly four nonzero vectors among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{q-1}$. By Lemma 6, the four nonzero vectors are in $\mathcal{C}_{q-3}$, which is an $[q+1, q+1, 1]_q$-code. The corresponding codeword $\mathbf{c} \in \mathcal{C}$ has Hamming weight $w_H(\mathbf{c}) \geq 4$. If $w_H(\mathbf{c}) \geq 5$, then $w_P(\phi(\mathbf{c})) \geq 6$. It suffices to show that the codeword $\phi(\mathbf{c})$ with Hamming weigh 4 does not have four consecutive nonzero components. We prove it by contradiction and divide the proof into the following five cases.

Case 4.1: If $w_H(\mathbf{c}) = 4$ and the four nonzero components of $\phi(\mathbf{c})$ are

$$c_{i,j},\ c_{i+1,(j-1) \bmod (q+1)},\ c_{i+2,(j-2) \bmod (q+1)},\ c_{i+3,(j-3) \bmod (q+1)},$$

with $i \in [0, q-5]$, then the corresponding columns

$$\mathbf{g}_{i,j}^\top,\ \mathbf{g}_{i+1,(j-1) \bmod (q+1)}^\top,\ \mathbf{g}_{i+2,(j-2) \bmod (q+1)}^\top,\ \mathbf{g}_{i+3,(j-3) \bmod (q+1)}^\top$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

are $\mathbb{F}_q$-linearly dependent, which contradicts

$$\begin{vmatrix} \alpha^i\beta^j & \alpha^{i+1}\beta^{j-1} & \alpha^{i+2}\beta^{j-2} & \alpha^{i+3}\beta^{j-3} \\ \alpha^i(\beta\gamma)^j & \alpha^{i+1}(\beta\gamma)^{j-1} & \alpha^{i+2}(\beta\gamma)^{j-2} & \alpha^{i+3}(\beta\gamma)^{j-3} \\ \beta^j & \beta^{j-1} & \beta^{j-2} & \beta^{j-3} \\ (\beta\gamma)^j & (\beta\gamma)^{j-1} & (\beta\gamma)^{j-2} & (\beta\gamma)^{j-3} \end{vmatrix} = \beta^{4j-6}\gamma^{2j-5}\alpha(\alpha-1)^2(\gamma-1)^2(\alpha\gamma-1)(\gamma-\alpha) \neq 0.$$

Case 4.2: If $w_H(\mathbf{c}) = 4$ and the four nonzero components of $\phi(\mathbf{c})$ are

$$c_{q-1,j},\ c_{0,(j-1) \bmod (q+1)},\ c_{1,(j-2) \bmod (q+1)},\ c_{2,(j-3) \bmod (q+1)},$$

then the corresponding columns

$$\mathbf{g}_{q-1,j}^\top,\ \mathbf{g}_{0,(j-1) \bmod (q+1)}^\top,\ \mathbf{g}_{1,(j-2) \bmod (q+1)}^\top,\ \mathbf{g}_{2,(j-3) \bmod (q+1)}^\top$$

are $\mathbb{F}_q$-linearly dependent. Thus, we have

$$\begin{vmatrix} 0 & \beta^{j-1} & \alpha\beta^{j-2} & \alpha^2\beta^{j-3} \\ 0 & (\beta\gamma)^{j-1} & \alpha(\beta\gamma)^{j-2} & \alpha^2(\beta\gamma)^{j-3} \\ \beta^j & \beta^{j-1} & \beta^{j-2} & \beta^{j-3} \\ (\beta\gamma)^j & (\beta\gamma)^{j-1} & (\beta\gamma)^{j-2} & (\beta\gamma)^{j-3} \end{vmatrix} = \beta^{4j-6}\gamma^{2j-5}\alpha(\alpha-1)(\gamma-1)^2(\gamma^2+(1-\alpha)\gamma+1) = 0,$$

which is equivalent to

$$\gamma^2 + (1-\alpha)\gamma + 1 = 0. \tag{40}$$

By (40), we know that $\left(\gamma + \frac{1-\alpha}{2}\right)^2 = \frac{(\alpha+1)(\alpha-3)}{4}$. Since $(\alpha-1)(\alpha-3)$ is either 0 or a square element of $\mathbb{F}_q$, we get $\gamma + \frac{1-\alpha}{2} \in \mathbb{F}_q$, which is a contradiction.

Case 4.3: If $w_H(\mathbf{c}) = 4$ and the four nonzero components of $\phi(\mathbf{c})$ are

$$c_{q-2,j},\ c_{q-1,(j-1) \bmod (q+1)},\ c_{0,(j-2) \bmod (q+1)},\ c_{1,(j-3) \bmod (q+1)},$$

then the corresponding columns

$$\mathbf{g}_{q-2,j}^\top,\ \mathbf{g}_{q-1,(j-1) \bmod (q+1)}^\top,\ \mathbf{g}_{0,(j-2) \bmod (q+1)}^\top,\ \mathbf{g}_{1,(j-3) \bmod (q+1)}^\top$$

are $\mathbb{F}_q$-linearly dependent. Thus we have

$$\begin{vmatrix} \alpha^{q-2}\beta^j & 0 & \beta^{j-2} & \alpha\beta^{j-3} \\ \alpha^{q-2}(\beta\gamma)^j & 0 & (\beta\gamma)^{j-2} & \alpha(\beta\gamma)^{j-3} \\ \beta^j & \beta^{j-1} & \beta^{j-2} & \beta^{j-3} \\ (\beta\gamma)^j & (\beta\gamma)^{j-1} & (\beta\gamma)^{j-2} & (\beta\gamma)^{j-3} \end{vmatrix} = -\frac{\beta^{4j-6}\gamma^{2j-6}(\gamma-1)^2}{\alpha^{q-4}-\alpha^{q-3}}(\gamma^2+(\alpha+2)\gamma+1) = 0,$$

which is equivalent to

$$\gamma^2 + (\alpha+2)\gamma + 1 = 0. \tag{41}$$

We obtain a contradiction in the same way as in Case 4.2.

Case 4.4: If $w_H(\mathbf{c}) = 4$ and the four nonzero components of $\phi(\mathbf{c})$ are

$$c_{q-3,j},\ c_{q-2,(j-1) \bmod (q+1)},\ c_{q-1,(j-2) \bmod (q+1)},\ c_{0,(j-3) \bmod (q+1)},$$

then the corresponding columns

$$\mathbf{g}_{q-3,j}^\top,\ \mathbf{g}_{q-2,(j-1) \bmod (q+1)}^\top,\ \mathbf{g}_{q-1,(j-2) \bmod (q+1)}^\top,\ \mathbf{g}_{0,(j-3) \bmod (q+1)}^\top$$

are $\mathbb{F}_q$-linearly dependent. This implies

$$\begin{vmatrix} \alpha^{q-3}\beta^j & \alpha^{q-2}\beta^{j-1} & 0 & \beta^{j-3} \\ \alpha^{q-3}(\beta\gamma)^j & \alpha^{q-2}(\beta\gamma)^{j-1} & 0 & (\beta\gamma)^{j-3} \\ \beta^j & \beta^{j-1} & \beta^{j-2} & \beta^{j-3} \\ (\beta\gamma)^j & (\beta\gamma)^{j-1} & (\beta\gamma)^{j-2} & (\beta\gamma)^{j-3} \end{vmatrix} = \frac{\beta^{4j-6}\gamma^{2j-5}(\gamma-1)^2\alpha^{q-3}}{1-\alpha}\left(\gamma^2+(\alpha^{q-2}+2)\gamma+1\right) = 0,$$

which is equivalent to

$$\gamma^2 + (\alpha^{q-2}+2)\gamma + 1 = 0. \tag{42}$$

Using a similar method as in Case 4.2, we reach a contradiction.

Case 4.5: If $w_H(\mathbf{c}) = 4$ and the four nonzero components of $\phi(\mathbf{c})$ are

$$c_{q-4,j},\ c_{q-3,(j-1) \bmod (q+1)},\ c_{q-2,(j-2) \bmod (q+1)},\ c_{q-1,(j-3) \bmod (q+1)},$$

then the corresponding columns

$$\mathbf{g}_{q-4,j}^\top,\ \mathbf{g}_{q-3,(j-1) \bmod (q+1)}^\top,\ \mathbf{g}_{q-2,(j-2) \bmod (q+1)}^\top,\ \mathbf{g}_{q-1,(j-3) \bmod (q+1)}^\top$$

are $\mathbb{F}_q$-linearly dependent. We then get

$$\begin{vmatrix} \alpha^{q-4}\beta^j & \alpha^{q-3}\beta^{j-1} & \alpha^{q-2}\beta^{j-2} & 0 \\ \alpha^{q-4}(\beta\gamma)^j & \alpha^{q-3}(\beta\gamma)^{j-1} & \alpha^{q-2}(\beta\gamma)^{j-2} & 0 \\ \beta^j & \beta^{j-1} & \beta^{j-2} & \beta^{j-3} \\ (\beta\gamma)^j & (\beta\gamma)^{j-1} & (\beta\gamma)^{j-2} & (\beta\gamma)^{j-3} \end{vmatrix} = \beta^{4j-6}\gamma^{2j-5}(\gamma-1)^2\alpha^{2q-6}(1-\alpha)\left(\gamma^2+(1-\alpha^{q-2})\gamma+1\right) = 0,$$

which is equivalent to

$$\gamma^2 + (1-\alpha^{q-2})\gamma + 1 = 0. \tag{43}$$

A contradiction follows in the same way as in Case 4.2.

Case 5: There are more than four nonzero vectors among the vectors $\mathbf{c}_0, \ldots, \mathbf{c}_{q-1}$. By Lemma 6, the nonzero vectors are in a $[q+1, q+1, 1]_q$-code $\mathcal{C}_{q-4}$. The corresponding codeword $\phi(\mathbf{c})$ of $\phi(\mathcal{C})$ has Hamming weight $w_H(\phi(\mathbf{c})) \geq 5$, which implies that $w_P(\phi(\mathbf{c})) \geq 6$.

Taking all of the above cases into consideration, we have shown that $\phi(\mathcal{C})$ is a $[q^2+q, 6]_q$-MDS symbol-pair code.

2) We have just concluded that $\phi(\mathcal{C})$ is a $[q^2+q, 6]_q$-MDS symbol-pair code if and only the quadratic equations in (40), (41), (42), and (43) fail to hold.

We now prove that $\gamma^2 + u\gamma + 1 \neq 0$ if $\mathrm{Tr}(u^{-1}) = 0$, for $u \in \mathbb{F}_q^*$ when $q = 2^m$ is even. Let us assume, for a contradiction,

that $\gamma^2+u\gamma+1=0$, which is equivalent to $\gamma+\gamma^{-1}=u$. Hence, we obtain

$$\frac{1}{u}=\frac{1}{\gamma+\gamma^{-1}}=\frac{1}{1+\gamma}+\frac{1}{(1+\gamma)^2}.$$

We get a contradiction to $\mathrm{Tr}(u^{-1})=0$ because

$$\begin{aligned}\mathrm{Tr}\left(\frac{1}{u}\right)&=\mathrm{Tr}\left(\frac{1}{1+\gamma}+\frac{1}{(1+\gamma)^2}\right)\\&=\sum_{i=0}^{m-1}\left(\frac{1}{1+\gamma}\right)^{2^i}+\sum_{i=0}^{m-1}\left(\frac{1}{1+\gamma}\right)^{2^{i+1}}\\&=\frac{1}{1+\gamma}+\frac{1}{(1+\gamma)^{2^m}}\\&=\frac{1}{1+\gamma}+\frac{1}{1+\gamma^{-1}}\\&=1.\end{aligned}$$

Thus,

$$\gamma^2+u\gamma+1\neq 0 \text{ if } \mathrm{Tr}(u^{-1})=0 \text{ for } u\in\mathbb{F}_q^*.$$

The quadratic equations in (40), (41), (42), and (43) fail to hold because

$$\mathrm{Tr}((1+\alpha)^{-1})=\mathrm{Tr}(\alpha^{-1})=\mathrm{Tr}(\alpha)=\mathrm{Tr}((1+\alpha^{q-2})^{-1})=0.$$

We can finally conclude that the code $\phi(\mathcal{C})$ constructed in 1) is a $[q^2+q,6]_q$-MDS symbol-pair code. ■

Due to the nonexistence of cyclic codes with parameters $[q+1,q-1,3]_q$ when q is odd, we use constacyclic codes as constituent codes to construct the MP code in Proposition 14. In the proof of Proposition 14, we show that $\phi(\mathcal{C})$ is a $[q^2+q,6]_q$-MDS symbol-pair code if and only if there exists a primitive element α of $\mathbb{F}_q$ such that the quadratic equations in (40), (41), (42), and (43) do not hold. In general, it is hard to propose a sufficient and necessary condition such that the quadratic equations in (40), (41), (42), and (43) fail to hold. We propose two sufficient conditions in Statements 1) and 2) of Proposition 14. The proof of the existence of special primitive elements requires number theoretic tools and is given in the Appendix. The existence of primitive elements with properties which are prescribed in Statements 1) and 2) leads to the following theorem.

Theorem 15. *Let q be a prime power such that $q\notin\{4,5\}$. If $q=2^m$ with m being even or q is odd, then there exists a $\left(q^2+q,6\right)_q$-MDS symbol-pair code.*

Proof: By Proposition 20 in the Appendix, if q is odd and $q\geq 3^{17}$, then there exists a primitive element α of $\mathbb{F}_q$ such that $(\alpha+1)(\alpha-3)$, $\alpha(\alpha+4)$, $\alpha^{q-2}(\alpha^{q-2}+4)$, and $(\alpha^{q-2}+1)(\alpha^{q-2}-3)$ are square elements in $\mathbb{F}_q$. Using Proposition 21 in the Appendix, if $q=2^m$ with $m\geq 22$ being even, then there exists a primitive element α of $\mathbb{F}_q$ such that

$$\mathrm{Tr}((1+\alpha)^{-1})=\mathrm{Tr}(\alpha^{q-2})=\mathrm{Tr}(\alpha)=\mathrm{Tr}((1+\alpha^{q-2})^{-1})=0.$$

By Proposition 14, there exists a $\left(q^2+q,6\right)_q$-MDS symbol-pair code if q is odd and $q\geq 3^{17}$ or $q=2^m$ with even integer $m\geq 22$.

We recall that the code $\phi(\mathcal{C})$ constructed in Proposition 14 is a $[q^2+q,6]_q$-MDS symbol-pair code if and only if there exists a primitive element α of $\mathbb{F}_q$ such that the quadratic equations in (40), (41), (42), and (43) fail to hold. By exhaustive computations in MAGMA, we verify that there exists a primitive element α in $\mathbb{F}_q$ such that the quadratic equations in (40), (41), (42), and (43) do not hold if $q\in[3,3^{17}]\backslash\{4,5\}$. ■

Example 12. Let $q=3$ and let β be a primitive element of $\mathbb{F}_9$. Let $\mathcal{C}_1$ be the 2-constacyclic code with generator polynomial $g_1(x)=1$ and let $\mathcal{C}_2=\mathcal{C}_3$ be the 2-constacyclic code with generator polynomial $g_1(x)=(x-\beta)(x-\beta^3)$ and parity-check matrix

$$H_1=\begin{pmatrix}1&0&1&1\\0&1&1&2\end{pmatrix}.$$

Using the NSC matrix A with

$$(A^{-1})^\top=\frac{1}{2}\begin{pmatrix}0&0&2\\1&2&0\\1&1&1\end{pmatrix},$$

we construct the $[12,8,3]_3$-code $\mathcal{C}$ defined by Theorem 15 with parity-check matrix

$$H=\begin{pmatrix}1&0&1&1&2&0&2&2&0&0&0&0\\0&1&1&2&0&2&2&1&0&0&0&0\\1&0&1&1&1&0&1&1&1&0&1&1\\0&1&1&2&0&1&1&2&0&1&1&2\end{pmatrix}.$$

The permutation ϕ on the coordinates $[0,11]$ of $\mathcal{C}$ is defined by

$$\begin{array}{c}0,1,2,3,4,5,6,7,8,9,10,11\\\downarrow\phi\\0,7,10,\quad 1,4,11,\quad 2,5,8,\quad 3,6,9.\end{array}$$

After some computation, we deduce that $\phi(\mathcal{C})$ is a $(12,6)_3$-MDS symbol-pair code. □

Remark 4. Since we cannot determine the existence of a primitive element α of $\mathbb{F}_{2^m}$ such that

$$\mathrm{Tr}((1+\alpha)^{-1})=\mathrm{Tr}(\alpha^{q-2})=\mathrm{Tr}(\alpha)=\mathrm{Tr}((1+\alpha^{q-2})^{-1})=0$$

when $m>1$ is odd, we do not know in general whether there exists a $\left(2^{2m}+2^m,6\right)_q$-MDS symbol-pair code for odd $m>1$. We verify computationally in MAGMA that the code $\phi(\mathcal{C})$ constructed in Proposition 14 is a $[2^{2m}+2^m,6]_q$-MDS symbol-pair code for each odd integer $m\in[3,49]$. We conjecture that such an MDS symbol-pair code exists for all odd integer $m>1$.

V. Concluding Remarks

As demand for data storage continues to increase, new technologies emerge to address crucial challenges that naturally arise. Symbol-pair codes were originally proposed to better control errors in channels whose outputs are overlapping pairs of symbols. An important performance measure for a symbol-pair code is its minimum pair distance, instead of the minimum Hamming distance for the classical error-correcting code. A Singleton-type bound leads to the definition of a maximum distance separable (MDS) symbol-pair code.

Many families of MDS symbol-pair codes have been explicitly built prior to our work. There are surely more families

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

of interesting symbol-pair codes to be discovered. Here, we construct matrix-product symbol-pair codes. Our approach is simpler than most prior constructions and, to the best of our knowledge, has never been attempted before. We have shown how to use generalized Reed-Solomon codes of much smaller length to construct families of MDS symbol-pair codes, either directly or with additional well-chosen permutations on the coordinate positions of the matrix-product codes. Finding more efficient computational methods to determine the minimum pair distance of a symbol-pair code is an interesting direction to investigate.

For our current construction purposes, it suffices to exhibit a pair of permutations τ and ρ that always works. Our choice is inspired by the approach from cyclic codes with repeated roots. It allows us to determine the parameters of the MP code under consideration by looking at a repeated-root cyclic code which is permutation equivalent to the MP code. The exact formulation of τ and ρ is based on the supports of the codewords, as stated before Theorem 10.

There are suitable alternatives to the τ and ρ that we have chosen in Theorems 10 and 12. Keeping the permutation τ, for instance, we can replace ρ by ρ_ℓ, for $\ell = 1, 2$, which is given by

$$\rho_\ell(j+3i) = \begin{cases} j + 3((i + (r \bmod 3)) \bmod p), & \text{if } i \in [0, p-1], \\ j + 3\left(((s + (r \bmod 3)) \bmod p) + p\right), & \text{if } i \in [p, 2p-1], \end{cases} \tag{44}$$

where $r = j+\ell$ and $s = i-p$. One can then show, by following a similar route to the one in the respective proofs of Theorems 10 and 12, that the codes $\rho_\ell(\tau(\mathcal{C}))$ and $\rho_\ell(\tau(\mathcal{B}))$ are still MDS symbol-pair codes. In [23, Corollary V.2], a necessary and sufficient condition for permutations to preserve the pair weights of any codeword in a linear code was proposed. Generalizing this observation to some families of symbol-pair codes, *e.g.*, by finding all permutations which preserve the MDS property of the codes in Theorems 10 and 12, deserves a separate treatment beyond this work. The scope can of course be widened to determining all permutations that preserve the MDS property of any family of MDS symbol-pair codes.

In Lemma 8, we have provided a heuristic to generate MP codes or their permutation equivalent codes which may achieve the Singleton-type bound in (3). Based on the heuristic and suitable choices of constituents codes, we have constructed three families of MDS symbol-pair codes in Theorems 10, 12, and 15. A particularly interesting direction is to study if we can use the general heuristic to explicitly build more new families of MDS symbol-pair codes.

ACKNOWLEDGMENT

The authors would like to thank Professors Xiwang Cao and Guangkui Xu for helpful discussions on the existence of special primitive elements in Proposition 14 and Theorem 15. The authors would also like to thank the anonymous referees and the associate editor, Professor Eitan Yaakobi for their useful comments and suggestions.

APPENDIX

Several proofs of the results in Section IV require the existence of primitive elements in $\mathbb{F}_q$ with specific properties. This Appendix demonstrate that those elements indeed exist. We recall some basic results about character sums over finite fields.

Let p be a prime and let q be a power of p. Let $\mathrm{Tr}(\cdot)$ denote the trace mapping from $\mathbb{F}_q$ to $\mathbb{F}_p$. For every $a \in \mathbb{F}_q$, an *additive character* of $\mathbb{F}_q$ is defined as $\chi_a(x) = \zeta_p^{\mathrm{Tr}(ax)}$, with $x \in \mathbb{F}_q$ and $\zeta_p = e^{2\pi\sqrt{-1}/p}$. The additive character $\chi_1(x)$ is *canonical* whereas the additive character $\chi_0(x)$ is *trivial*. Let α be a primitive element of $\mathbb{F}_q$. For each $i \in [0, q-2]$, a *multiplicative character* of $\mathbb{F}_q$ is defined by $\psi_i(\alpha^j) = \zeta_{q-1}^{ij}$, with $j \in [0, q-2]$ and $\zeta_{q-1} = e^{2\pi\sqrt{-1}/(q-1)}$. In addition, we assume that $\psi_i(0) = 0$ for $i \in [0, q-2]$. The *order* of a multiplicative character ψ_i is defined to be the smallest positive integer e such that $(\psi_i(\alpha^j))^e = 1$ for each $j \in [0, q-2]$. The multiplicative character $\psi_{\frac{q-1}{2}}$ of order 2 is called the *quadratic character* of $\mathbb{F}_q$. The respective orthogonality relations of additive characters and multiplicative characters are given by

$$\sum_{g \in \mathbb{F}_q} \chi_a(g) = \begin{cases} 0, & \text{if } a \neq 0 \\ q, & \text{if } a = 0 \end{cases} \text{ and}$$

$$\sum_{j=0}^{q-2} \psi_i(\alpha^j) = \begin{cases} 0, & \text{if } i \neq 0 \\ q-1, & \text{if } i = 0 \end{cases}.$$

Let $\mathbb{F}_q(x)$ be the rational function field. Given a non-trivial additive character χ and a multiplicative character ϕ of $\mathbb{F}_q$, we define a *mixed character sum*

$$A(\chi, f; \phi, g) = \sum_{x \in \mathbb{F}_q \setminus S} \chi(f(x))\phi(g(x)), \tag{45}$$

where $f(x)$ and $g(x)$ are rational functions of $\mathbb{F}_q(x)$ and S is the set of poles of f and g. This sum is *degenerate* if both $f(x) = h(x)^p - h(x) + c$, for some $h(x) \in \mathbb{F}_q(x)$ and $c \in \mathbb{F}_q$, and $g(x) = br(x)^n$, for some $r(x) \in \mathbb{F}_q(x)$ and $b \in \mathbb{F}_q$, where n is the order of ϕ.

Castro and Moreno [5] proposed the following bound for the character sum $A(\chi, f; \phi, g)$.

Lemma 16. [5] *Let $f(x)$ and $g(x)$ be rational functions of $\mathbb{F}_q(x)$. Let ℓ_g be the number of distinct zeros and non-infinite poles of g. Let ℓ_1 be the number of poles including the infinite pole of f and let ℓ_2 be the sum of the multiplicities of these poles. Let ℓ_3 be the number of non-infinite poles of f which are zeros or poles of g. Let χ and ϕ be, respectively, a non-trivial additive character and a multiplicative character of $\mathbb{F}_q$. If $A(\chi, f; \phi, g)$ is the non-degenerate character sum in (45), then*

$$|A(\chi, f; \phi, g)| \leq (\ell_g + \ell_1 + \ell_2 - \ell_3 - 2)\sqrt{q}.$$

The following new version of the Weil bound was provided in [35].

Lemma 17. [35] *Let $f_1(x), \cdots, f_s(x) \in \mathbb{F}_q[x]$ be s monic, distinct, and irreducible polynomials. Let ℓ be the number of distinct roots of $\prod_{i=1}^{s} f_i(x)$ in its splitting field of*

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

$\mathbb{F}_q$. *Let* $\phi_1, \cdots, \phi_s$ *be multiplicative characters of* $\mathbb{F}_q$. *If* $\prod_{i=1}^{s} \phi_i(f_i(z)) \neq 1$ *for some* $z \in \mathbb{F}_q$, *then, for each* $a_i \in \mathbb{F}_q^*$ *and* $i \in [s]$, *we have*

$$\left| \sum_{z \in \mathbb{F}_q} \prod_{i=1}^{s} \phi_i(a_i f_i(z)) \right| \leq (\ell - 1)\sqrt{q}.$$

Let $\varphi(t)$ be the Euler function and let $\mu(t)$ be the Möbius function given by

$$\mu(t) = \begin{cases} 1, & \text{if } t = 1, \\ (-1)^{\ell}, & \text{if } t \text{ is the product of } \ell \text{ distinct primes}, \\ 0, & \text{if } t \text{ is divisible by the square of a prime.} \end{cases}$$

We know from [22, Lemma 3.23] that

$$\sum_{t|r} \mu(t) = \begin{cases} 1, & \text{if } r = 1, \\ 0, & \text{if } r > 1. \end{cases} \tag{46}$$

The following lemma gives a method to determine whether an element of $\mathbb{F}_q$ is primitive.

Lemma 18. [8] *Let* z *be an element of* $\mathbb{F}_q$. *If*

$$P(z) = \frac{\varphi(q-1)}{q-1} \sum_{\ell|(q-1)} \frac{\mu(\ell)}{\varphi(\ell)} \sum_{\phi_\ell} \phi_\ell(z),$$

where $\sum_{\phi_\ell}$ *is defined over all multiplicative characters* ϕ_ℓ *of order* ℓ, *then*

$$P(z) = \begin{cases} 1, & \text{if } z \text{ is a primitive element of } \mathbb{F}_q, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\omega(t)$ denote the number of distinct primes factors of t. Robin in [33] presented an upper bound for $\omega(t)$.

Lemma 19. [33] *Given a positive integer* $t > 2$, *we have* $\omega(t) \leq \frac{1.385 \log t}{\log \log t}$.

After we have recalled the above four essential lemmas, we proceed to demonstrate the existence of special primitive elements that are used in the proof of Proposition 15.

Proposition 20. *Let* p *be an odd prime and let* $q = p^m$. *If* $m \geq 17$, *then there always exists a primitive element* α *of* $\mathbb{F}_q$ *such that* $(\alpha+1)(\alpha-3)$, $\alpha(\alpha+4)$, $\alpha^{q-2}(\alpha^{q-2}+4)$, *and* $(\alpha^{q-2}+1)(\alpha^{q-2}-3)$ *are square elements of* $\mathbb{F}_q$.

Proof: Let α be a primitive element of $\mathbb{F}_q$. It is easy to see that α and α^{q-2} are not square elements. The statement that $(\alpha+1)(\alpha-3)$, $\alpha(\alpha+4)$, $\alpha^{q-2}(\alpha^{q-2}+4)$, and $(\alpha^{q-2}+1)(\alpha^{q-2}-3)$ are square elements of $\mathbb{F}_q$ is equivalent to the statement that $(\alpha+1)(\alpha-3)$, $1+4\alpha$, and $(1+\alpha)(1-3\alpha)$ are square elements and $\alpha+4$ is not a square element. Let η be the quadratic character of $\mathbb{F}_q$. It is easy to check that $\eta(z) = 1$ if z is a square element of $\mathbb{F}_q$ and $\eta(z) = -1$ if z is not a square element of $\mathbb{F}_q$. We define the set $U := \{z \in \mathbb{F}_q : z \text{ is a primitive element}, \eta((z+1)(z-3)) = \eta(1+4z) = \eta((1+z)(1-3z)) = 1, \text{ and } \eta(z+4) = -1\}$. If the set U is not empty, then there exists a primitive element α of $\mathbb{F}_q$ such that $(\alpha+1)(\alpha-3)$, $\alpha(\alpha+4)$, $\alpha^{q-2}(\alpha^{q-2}+4)$, and $(\alpha^{q-2}+1)(\alpha^{q-2}-3)$ are square elements of $\mathbb{F}_q$. Let N denote the cardinality of U. For each $z \in \mathbb{F}_q$, we write

$$\begin{aligned} Q(z) = (\eta((z+1)(z-3))+1)(\eta(1+4z)+1) \\ (\eta((1+z)(1-3z))+1)(1-\eta(z+4)). \end{aligned}$$

The element $z \in \mathbb{F}_q$ has the properties that $(z+1)(z-3)$, $1+4z$, and $(1+z)(1-3z)$ are square elements and $z+4$ is not a square element if and only if $Q(z) = 16$. By Lemma 18, we deduce that

$$\begin{aligned} N &= \sum_{z \in \mathbb{F}_q} \frac{\varphi(q-1)}{q-1} \sum_{\ell|(q-1)} \frac{\mu(\ell)}{\varphi(\ell)} \sum_{\phi_\ell} \phi_\ell(z) \frac{Q(z)}{16} \\ &= \frac{\varphi(q-1)}{16(q-1)} \sum_{\ell|(q-1)} \frac{\mu(\ell)}{\varphi(\ell)} \sum_{\phi_\ell} \sum_{z \in \mathbb{F}_q} \phi_\ell(z) Q(z) \\ &= \frac{\varphi(q-1)}{16(q-1)} \sum_{\ell|(q-1)} P(\ell), \end{aligned}$$

where

$$P(\ell) = \frac{\mu(\ell)}{\varphi(\ell)} \sum_{\phi_\ell} \sum_{z \in \mathbb{F}_q} \phi_\ell(z) Q(z).$$

One can then confirm that 1 is the only constant term in $Q(z)$ and

$$P(1) = \sum_{z \in \mathbb{F}_q} Q(z) = q - 1 + \sum_{z \in \mathbb{F}_q} (Q(z) - 1).$$

By Lemma 17, we have $\left| \sum_{z \in \mathbb{F}_q} (Q(z) - 1) \right| \leq 33\sqrt{q}$. If ℓ is divisible by the square of a prime, then by the definition of the Möbius function, we obtain $P(\ell) = 0$. If $\ell \neq 1$ is the product of $\omega(\ell)$ distinct primes, then, by Lemma 17, we get

$$\begin{aligned} |P(\ell)| \leq \frac{1}{\varphi(\ell)} \sum_{\phi_\ell} \left| \sum_{z \in \mathbb{F}_q} \phi_\ell(z) Q(z) \right| \leq \\ \frac{1}{\varphi(\ell)} \sum_{\phi_\ell} 48\sqrt{q} = 48\sqrt{q}. \end{aligned}$$

Summarizing all of the above cases, we arrive at

$$\begin{aligned} \left| N - \frac{\varphi(q-1)(q-1)}{16(q-1)} \right| \leq \\ \frac{\varphi(q-1)}{16(q-1)} \left(33\sqrt{q} + 48\sqrt{q}(2^{\omega(q-1)} - 1) \right). \end{aligned} \tag{47}$$

We can, by using (47), infer that

$$N \geq \frac{\varphi(q-1)(q-1)}{16(q-1)} - \frac{\varphi(q-1)}{16(q-1)} \left(48\sqrt{q} 2^{\omega(q-1)} - 15\sqrt{q} \right).$$

By Lemma 19, we have

$$\omega(q-1) \leq \frac{1.385 \log(q-1)}{\log \log(q-1)}.$$

The set U is not empty if

$$\sqrt{q} - \frac{1}{\sqrt{q}} + 15 > 48 \cdot 2^{\frac{1.385 \log(q-1)}{\log \log(q-1)}}. \tag{48}$$

Since p is an odd prime and $q = p^m$, we check that (48) holds for each $p \geq 3$ and $m \geq 17$. This completes the proof. ■

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3220638

Proposition 21. *Let $m > 0$ be an even integer and let $q = 2^m$. If $m \geq 22$, then there exists a primitive element α of $\mathbb{F}_q$ such that*

$$\mathrm{Tr}((1+\alpha)^{-1}) = \mathrm{Tr}(\alpha^{q-2}) = \mathrm{Tr}(\alpha) = \mathrm{Tr}((1+\alpha^{q-2})^{-1}) = 0.$$

Proof: We define the set $V := \{z \in \mathbb{F}_{2^m} : z$ is a primitive element with $\mathrm{Tr}((1+z)^{-1}) = \mathrm{Tr}(z^{-1}) = \mathrm{Tr}(z) = \mathrm{Tr}((1+z^{-1})^{-1}) = 0\}$ of cardinality $|V| = N$. Let χ_1 be the canonical additive character of $\mathbb{F}_{2^m}$. By the orthogonality property of additive characters, an element $z \in \mathbb{F}_{2^m}$ satisfies

$$\mathrm{Tr}((1+z)^{-1}) = \mathrm{Tr}(z^{-1}) = \mathrm{Tr}(z) = \mathrm{Tr}((1+z^{-1})^{-1}) = 0$$

if and only if

$$T(z) = \frac{1}{16}\sum_{a\in\mathbb{F}_2}(-1)^{a\,\mathrm{Tr}((1+z)^{-1})}\sum_{b\in\mathbb{F}_2}(-1)^{b\,\mathrm{Tr}(z^{-1})} \sum_{c\in\mathbb{F}_2}(-1)^{c\,\mathrm{Tr}(z)}\sum_{d\in\mathbb{F}_2}(-1)^{d\,\mathrm{Tr}((1+z^{-1})^{-1})} = 1. \quad (49)$$

By Lemma 18, we deduce that

$$\begin{aligned} N &= \sum_{z\in\mathbb{F}_{2^m}\setminus\{0,1\}} \frac{\varphi(2^m-1)}{2^m-1}\sum_{\ell|(2^m-1)}\frac{\mu(\ell)}{\varphi(\ell)}\sum_{\phi_\ell}\phi_\ell(z)T(z) \\ &= \frac{\varphi(2^m-1)}{16(2^m-1)}\sum_{\ell|(2^m-1)}\sum_{a,b,c,d\in\mathbb{F}_2}\frac{\mu(\ell)}{\varphi(\ell)} \sum_{\phi_\ell}\sum_{z\in\mathbb{F}_{2^m}\setminus\{0,1\}}\phi_\ell(z)A(z;a,b,c,d) \\ &= \frac{\varphi(2^m-1)}{16(2^m-1)}\sum_{\ell|(2^m-1)}\sum_{a,b,c,d\in\mathbb{F}_2}P(\ell;a,b,c,d), \quad (50) \end{aligned}$$

with

$$A(z;a,b,c,d) = \chi_1\left(\frac{cz^3+(c+d)z^2+(a+b)z+b}{z(1+z)}\right),$$

$$P(\ell;a,b,c,d) = \frac{\mu(\ell)}{\varphi(\ell)}\sum_{\phi_\ell}\sum_{z\in\mathbb{F}_{2^m}\setminus\{0,1\}}\phi_\ell(z)A(z;a,b,c,d).$$

Since m is even, we have $\chi_1(1) = 1$. It is in the end easy to check that

$$P(1;0,0,0,0) = 2^m - 2 \text{ and}$$
$$P(1;1,0,0,1) = \chi_1(1)(2^m-2) = 2^m - 2.$$

Due to the orthogonality relation of multiplicative characters and based on (46), we have

$$\begin{aligned} &\sum_{\ell|(2^m-1),\ell\neq1} P(1;0,0,0,0) \\ &= \sum_{\ell|(2^m-1),\ell\neq1}\frac{\mu(\ell)}{\varphi(\ell)}\sum_{\phi_\ell}\sum_{z\in\mathbb{F}_{2^m}\setminus\{0,1\}}\phi_\ell(z) \\ &= \sum_{\ell|(2^m-1),\ell\neq1}\frac{\mu(\ell)}{\varphi(\ell)}\sum_{\phi_\ell}(-1) \\ &= -\sum_{\ell|(2^m-1),\ell\neq1}\mu(\ell) \\ &= 1. \end{aligned}$$

Using the same argument as shown above, we obtain

$$\sum_{\ell|(2^m-1),\ell\neq1} P(1;1,0,0,1) = 1.$$

For the other cases of ℓ, a, b, c, and d, we can use Lemma 16 to deduce an upper bound of $|P(\ell;a,b,c,d)|$. By Lemma 16 and (50), we obtain

$$\begin{aligned} &\left|N - \frac{\varphi(2^m-1)}{16(2^m-1)}\sum_{\ell|(2^m-1)}(P(\ell;0,0,0,0)+P(\ell;1,0,0,1))\right| \\ &= \left|N - \frac{2\varphi(2^m-1)}{16(2^m-1)}(2^m-1)\right| \quad (51) \\ &\leq \frac{\varphi(2^m-1)}{16(2^m-1)}\sum_{\ell|(2^m-1)}\sum_{\substack{a,b,c,d\in\mathbb{F}_2 \\ (a,b,c,d)\neq(0,0,0,0) \\ (a,b,c,d)\neq(1,0,0,1)}}|P(\ell;a,b,c,d)| \\ &\leq \frac{\varphi(2^m-1)}{16(2^m-1)}29\cdot 2^{m/2}2^{\omega(2^m-1)}. \quad (52) \end{aligned}$$

From (52), we obtain

$$N \geq \frac{\varphi(2^m-1)}{16(2^m-1)}\left(2(2^m-1) - 29\cdot 2^{m/2}2^{\omega(2^m-1)}\right).$$

By Lemma 19, we have

$$\omega(2^m-1) \leq \frac{1.385\log(2^m-1)}{\log\log(2^m-1)}.$$

The set V is not empty if

$$2^{\frac{m}{2}} > \frac{29}{2}2^{\frac{1.385\log(2^m-1)}{\log\log(2^m-1)}} + 2^{-\frac{m}{2}}. \quad (53)$$

Finally, we confirm that (53) holds for each $m \geq 22$. ∎

Remark 5. If m is odd in Proposition 21, then

$$\sum_{\ell|(2^m-1)}(P(\ell;0,0,0,0)+P(\ell;1,0,0,1)) = 0$$

and we cannot use (51) and (52) to estimate the value of N.

REFERENCES

[1] T. Blackmore and G.H. Norton, "Matrix-product codes over $\mathbb{F}_q$," Appl. Algebra Eng. Comm. Comput., vol. 12, pp. 477–500, 2001.

[2] W. Bosma, J. Cannon and C. Playoust, "The Magma algebra system I: The user language," J. Symb. Comput., vol. 24, pp 235–265, 1997.

[3] Y. Cassuto and M. Blaum, "Codes for symbol-pair read channels," IEEE Trans. Inf. Theory, vol. 57, no. 12, pp. 8011–8020, 2011.

[4] Y. Cassuto and S. Litsyn, "Symbol-pair codes: Algebraic constructions and asymptotic bounds," in Proc. IEEE Int. Symp. Inf. Theory (ISIT2011), Saint Petersburg, Russia, Jul./Aug. 2011, pp. 2348–2352.

[5] F. Castro and C. Moreno, "Mixed exponential sums over finite fields," Proc. Amer. Math. Soc., vol. 128, pp. 2529-2537, 2000.

[6] Y. M. Chee, L. Ji, H. M. Kiah, C. Wang, and J. Yin, "Maximum distance separable codes for symbol-pair read channels," IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7259–7267, 2013.

[7] B. Chen, L. Lin, and H. Liu, "Constacyclic symbol-pair codes: Lower bounds and optimal constructions," IEEE Trans. Inf. Theory, vol. 63, no. 12, pp. 7661–7666, 2017.

[8] S. D. Cohen, "Primitive polynomials with a prescribed coefficient." Finite Fields Their Appl. vol. 12, pp. 425-491. 2006.

[9] H. Q. Dinh, B. T. Nguyen, A. K. Singh, and S. Sriboonchitta, "On the symbol-pair distance of repeated-root constacyclic codes of prime power lengths," IEEE Trans. Inf. Theory, vol. 64, no. 4, pp. 2417–2430, 2017.

[10] H. Q. Dinh, B. T. Nguyen, and S. Sriboonchitta, "MDS symbol-pair cyclic codes of length $2p^s$ over $\mathbb{F}_{p^m}$," IEEE Trans. Inf. Theory, vol. 66, no. 1, pp. 240–262, 2020.

[11] H. Q. Dinh, X. Wang, H. Liu, and S. Sriboonchitta, "On the symbol-pair distances of repeated-root constacyclic codes of length $2p^s$," Discrete Math., vol. 342, no. 11, pp. 3062–3078, 2019.

[12] H. Q. Dinh, B. T. Nguyen, A. K. Singh, and W. Yamaka, "MDS constacyclic codes and MDS symbol-pair constacyclic codes," IEEE Access, vol. 9, pp. 137970–137990, 2021.

[13] H. Q. Dinh, B. T. Nguyen, A. K. Singh, and S. Sriboonchitta, "Hamming and symbol-pair distances of repeated-root constacyclic codes of prime power lengths over $\mathbb{F}_{p^m} + u\mathbb{F}_{p^m}$," IEEE Commun. Letters, vol. 22, no. 2, pp. 2400–2403, 2018.

[14] H. Q. Dinh, P. Kumam, P. Kumar, S. Satpati, A. K. Singh, and W. Yamaka, "MDS symbol-pair repeated-root constacylic codes of prime power lengths over $\mathbb{F}_{p^m} + u\mathbb{F}_{p^m}$," IEEE Access, vol. 9, pp. 145039–145048, 2019.

[15] H. Q. Dinh, N. Kumar, A. K. Singh, M. K. Singh, I. Gupta, and P. Maneejuk, "On the symbol-pair distance of some classes of repeated-root constacyclic codes over Galois ring," Appl. Algebra Eng. Commun. Comput., Early Access, DOI: 10.1007/s00200-020-00472-6, 2021.

[16] H. Q. Dinh, A. K. Singh, and M. K. Thakur, "On symbol-pair distances of repeated-root constacyclic codes of length $2p^s$ over $\mathbb{F}_{p^m} + u\mathbb{F}_{p^m}$ and MDS symbol-pair codes," Appl. Algebra Eng. Commun. Comput., Early Access, DOI: 10.1007/s00200-021-00534-3, 2021.

[17] O. Elishco, R. Gabrys, and E. Yaakobi "Bounds and constructions of codes over symbol-pair read channels," IEEE Trans. Inf. Theory, vol. 66, no. 3, pp. 1385–1395, 2020.

[18] B. Ding, G. Ge, J. Zhang, T. Zhang, and Y. Zhang, "New constructions of MDS symbol-pair codes," Des. Codes Cryptogr., vol. 86, no. 4, pp. 841–859, 2018.

[19] E. M. Gabidulin and T. Kløve, "The Newton radius of MDS codes," in Proc. IEEE Inf. Theory Workshop (ITW) 1998, Killarney, Ireland, 1998, pp. 50–51.

[20] J. Laaouine, H. Q. Dinh, M. E. Charkani, and W. Yamaka, "MDS symbol-pair repeated-root constacylic codes of prime power lengths over $\mathbb{F}_q + u\mathbb{F}_q + u^2\mathbb{F}_q$," J. Appl. Math. Comput., Early Access, DOI: 10.1007/s12190-022-01738-7, 2022.

[21] S. Li and G. Ge, "Constructions of maximum distance separable symbol-pair codes using cyclic and constacyclic codes," Des. Codes Cryptogr., vol. 84, no. 3, pp. 359–372, 2017.

[22] R. Lidl and H. Niederreiter, Finite Fields. Cambridge university press, 1997.

[23] H. Liu and X. Pan, "Generalized pair weights of linear codes and linear isomorphisms preserving pair weights," IEEE Trans. Inf. Theory, vol. 68, no. 1, pp. 105–117, 2022.

[24] S. Ling and C. Xing, Coding Theory: A First Course. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[25] S. Ling and P. Solé, "On the algebraic structure of quasi-cyclic codes I: Finite fields," IEEE Trans. Inf. Theory, vol. 47, no. 7, pp. 2751–2760, 2001.

[26] X. Kai, S. Zhu, and P. Li, "A construction of new MDS symbol-pair codes," IEEE Trans. Inf. Theory, vol. 61, no. 11, pp. 5828–5834, 2015.

[27] X. Kai, S. Zhu, Y. Zhao, H. Luo, and Z. Chen, "New MDS symbol-pair codes from repeated root codes," IEEE Commun. Lett., vol. 22, no. 3, pp. 462–465, 2018.

[28] A. Krishna and D. V. Sarwate, "Psudocyclic maximum-distance separable codes,"' IEEE Trans. Inf. Theory, vol. 36, no. 4, pp. 880-884, Jul. 1990.

[29] F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes. Amsterdam: North-Holland Pub. Co., 1983.

[30] J. Ma and J. Luo, "MDS symbol-pair codes from repeated-root cyclic codes," Des. Codes Cryptogr., vol. 90, pp. 121–137, 2021.

[31] J. Ma and J. Luo, "Constructions of MDS symbol-pair codes with minimum distance seven or eight," Des. Codes Cryptogr., Early Access, DOI: 10.1007/s10623-022-01081-9, 2022.

[32] F. Özbudak and H. Stichtenoth, "Note on Niederreiter-Xing's propagation rule for linear codes," Appl. Algebra Eng. Commun. Comput., vol. 13, pp. 53–56, 2002.

[33] G. Robin, "Estimation de la fonction de Tchebychef θ sur le k-ième nombre premier et grandes valeurs de la fonction $\omega(n)$ nombre de diviseurs premiers de n," Acta Arith., vol. 42, no. 4, pp. 367-389, 1983.

[34] M. Shi, F. Özbudak and P. Solé, "Geometric approach to b-symbol hamming weights of cyclic codes," IEEE Trans. Inf. Theory, vol. 67, no. 6, pp. 3735–3751, 2021.

[35] D. Wan, "Generators and irreducible polynomials over finite fields," Math. Comput., vol. 66, no. 219, pp. 1195-1212, Jul. 1997.

[36] E. Yaakobi, J. Bruck, and P. H. Siegel, "Constructions and decoding of cyclic codes over b-symbol read channels," IEEE Trans. Inf. Theory, vol. 62, no. 4, pp. 1541–1551, 2016.

Gaojun Luo received the Ph.D. degree in mathematics from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2019.

He is currently a Research Fellow with Nanyang Technological University. His research interests include sequences, quantum information theory, and coding theory.

Martianus Frederic Ezerman received the double B.A. in philosophy and B.Sc. in mathematics degrees and the M.Sc. degree in mathematics from Ateneo de Manila University, Philippines, in 2005 and 2007, respectively, He obtained the Ph.D. degree in mathematical sciences from Nanyang Technological University (NTU), Singapore, in 2011.

He is currently an Adjunct Assistant Professor at NTU and serves as the chief executive of Sandhiguna, a company that develops key management systems and cryptographic services on trusted execution environment. His research interests include coding theory, cryptography, holographic data representations, and quantum information theory.

San Ling received the B.A. degree in mathematics from the University of Cambridge and the Ph.D. degree in mathematics from the University of California, Berkeley.

He is currently President's Chair in Mathematical Sciences, at the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, which he joined in April 2005. Prior to that, he was with the Department of Mathematics, National University of Singapore. Since August 2022, he also holds the concurrent appointment as the Chief Scientific Advisor of the National Research Foundation, Singapore.

His research fields include: arithmetic of modular curves and application of number theory to combinatorial designs, coding theory, cryptography and sequences.

Xu Pan received the B.Sc. degree in Mathematics from Shanxi University in 2016 and the M.Sc. degrees in Pure Mathematics from Central China Normal University in 2019. Currently, he is a doctoral student with Central China Normal University. His current research area is algebraic coding theory.