Maximum Distance Separable Codes for Symbol-Pair Read Channels

Yeow Meng Chee, *Senior Member, IEEE*, Lijun Ji, Han Mao Kiah, Chengmin Wang, and Jianxing Yin

***Abstract*—We study (symbol-pair) codes for symbol-pair read channels introduced recently by Cassuto and Blaum (2010). A Singleton-type bound on symbol-pair codes is established and infinite families of optimal symbol-pair codes are constructed. These codes are maximum distance separable (MDS) in the sense that they meet the Singleton-type bound. In contrast to classical codes, where all known q-ary MDS codes have length $O(q)$, we show that q-ary MDS symbol-pair codes can have length $\Omega(q^2)$. In addition, we completely determine the existence of MDS symbol-pair codes for certain parameters.**

***Index Terms*—Codes for magnetic storage, maximal distance separable, Singleton-type bound, symbol-pair read channels.**

I. Introduction

SYMBOL-PAIR coding theory has recently been introduced by Cassuto and Blaum [2], [3] to address channels with high write resolution but low read resolution, so that individual symbols cannot be read off due to physical limitations. An example of such channels is magnetic-storage, where information may be written via a high resolution process such as lithography and then read off by a low resolution technology such as magnetic head.

The theory of symbol-pair codes is at a rudimentary stage. Cassuto and Blaum [2], [3] laid out a framework for combating pair-errors, relating pair-error correction capability to a new metric called pair-distance. They also provided code constructions and studied decoding methods. Bounds and asymptotics on the size of optimal symbol-pair codes are obtained. More recently, Cassuto and Litsyn [4] constructed cyclic symbol-pair codes using algebraic methods, and showed that there exist symbol-pair codes whose rates are strictly higher, compared to codes for the Hamming metric with the same relative distance. Yaakobi *et al.* [5] presented efficient algorithms for decoding of cyclic symbol-pair codes.

Manuscript received November 08, 2012; revised May 23, 2013; accepted July 22, 2013. Date of publication August 15, 2013; date of current version October 16, 2013. Y. M. Chee, H. M. Kiah, and C. Wang were supported in part by the National Research Foundation of Singapore under Research Grant NRF-CRP2-2007-03. L. Ji was supported by the NSFC under Grant 11222113. J. Yin and C. Wang were supported by the NSFC under Grant 11271280. This paper was presented in part at the 2012 IEEE International Symposium on Information Theory.

Y. M. Chee and H. M. Kiah are with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371 (e-mail: ymchee@ntu.edu.sg; hmkiah@ntu.edu.sg).

L. Ji and J. Yin are with the Department of Mathematics, Suzhou University, Suzhou 215006, China (e-mail: jilijun@suda.edu.cn; jxyin@suda.edu.cn).

C. Wang is with School of Science, Jiangnan University, Wuxi 214122, China (e-mail:wcm@jiangnan.edu.cn).

Communicated by N. Kashyap, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2013.2276615

This paper continues the investigation of codes for symbol-pair channels. We establish a Singleton-type bound for symbol-pair codes and construct MDS symbol-pair codes (codes meeting this Singleton-type bound). In particular, we construct q-ary MDS symbol-pair codes of length n and pair-distance $n-1$ and $n-2$, where n can be as large as $\Omega(q^2)$. In contrast, the lengths of nontrivial classical q-ary MDS codes are conjectured to be $O(q)$. In addition, we completely settle the existence of MDS symbol-pair codes of length n with pair-distance d, for certain parameters.

The rest of the paper is organized as follows. In Section II, we introduce basic notation and definitions and derive a Singleton-type bound for symbol-pair codes. In Section III, we make use of interleaving and graph theoretic concepts to construct MDS symbol-pair codes from classical MDS codes. Unfortunately, such methods are inadequate to determine completely the existence of MDS symbol-pair codes. In Section IV, we introduce other construction methods and give complete solutions in certain instances. Technical proofs are deferred to the Appendix.

II. Preliminaries

Throughout this paper, Σ is a set of q elements, called *symbols*. For a positive integer $n \geq 2$, $\mathbb{Z}_n$ denotes the ring $\mathbb{Z}/n\mathbb{Z}$. The coordinates of $\mathbf{u} \in \Sigma^n$ are indexed by elements of $\mathbb{Z}_n$, so that $\mathbf{u} = (u_0, u_1, \ldots, u_{n-1})$.

A *pair-vector* over Σ is a vector in $(\Sigma \times \Sigma)^n$. We emphasize that a vector is a pair-vector through the notation $(\Sigma \times \Sigma)^n$, in lieu of $(\Sigma^2)^n$. For any $\mathbf{u} = (u_0, u_1, \ldots, u_{n-1}) \in \Sigma^n$, the *symbol-pair read vector* of $\mathbf{u}$ is the pair-vector (over Σ)

$$\pi(\mathbf{u}) = ((u_0, u_1), (u_1, u_2), \ldots, (u_{n-2}, u_{n-1}), (u_{n-1}, u_0)).$$

Obviously, each vector $\mathbf{u} \in \Sigma^n$ has a unique symbol-pair read vector $\pi(\mathbf{u}) \in (\Sigma \times \Sigma)^n$. However, not all pair-vectors over Σ have a corresponding vector in Σ^n.

Let $\mathbf{u}, \mathbf{v} \in \Sigma^n$. The *pair-distance* between vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\begin{aligned} D_{\mathrm{p}}(\mathbf{u}, \mathbf{v}) &:= D_{\mathrm{H}}(\pi(\mathbf{u}), \pi(\mathbf{v})) \\ &= |\{i \in \mathbb{Z}_n : (u_i, u_{i+1}) \neq (v_i, v_{i+1})\}|, \end{aligned}$$

where D_{H} denotes the usual Hamming distance. Cassuto and Blaum [3] proved that $(\Sigma^n, D_{\mathrm{p}})$ is a metric space, and showed the following relationship between pair-distance and Hamming distance.

Proposition 2.1 (Cassuto and Blaum [3]): For $\mathbf{u}, \mathbf{v} \in \Sigma^n$ such that $0 < D_{\mathrm{H}}(\mathbf{u}, \mathbf{v}) < n$, we have

$$D_{\mathrm{H}}(\mathbf{u}, \mathbf{v}) + 1 \leq D_{\mathrm{p}}(\mathbf{u}, \mathbf{v}) \leq 2D_{\mathrm{H}}(\mathbf{u}, \mathbf{v}).$$

In the extreme cases in which $D_{\mathrm{H}}(\mathbf{u},\mathbf{v}) = 0$ or n, we have $D_{\mathrm{p}}(\mathbf{u},\mathbf{v}) = D_{\mathrm{H}}(\mathbf{u},\mathbf{v})$.

A (q-*ary*) code of length n is a set $\mathcal{C} \subseteq \Sigma^n$. Elements of $\mathcal{C}$ are called *codewords*. The code $\mathcal{C}$ is said to have *pair-distance* d if $d = \min\{D_{\mathrm{p}}(\mathbf{u},\mathbf{v}) : \mathbf{u},\mathbf{v} \in \mathcal{C}, \mathbf{u} \neq \mathbf{v}\}$ and we denote such a code by $(n,d)_q$*-symbol-pair code*. The *size* of an $(n,d)_q$-symbol-pair code is the number of codewords it contains and the size of a symbol-pair code satisfies the following Singleton-type bound.

Theorem 2.1. (Singleton Bound): Let $q \geq 2$ and $2 \leq d \leq n$. If $\mathcal{C}$ is an $(n,d)_q$-symbol-pair code, then $|\mathcal{C}| \leq q^{n-d+2}$.

Proof: Let $\mathcal{C}$ be an $(n,d)_q$-symbol-pair code with $q \geq 2$ and $2 \leq d \leq n$. Delete the last $d-2$ coordinates from all the codewords of $\mathcal{C}$. Observe that any $d-2$ consecutive coordinates contribute at most $d-1$ to the pair-distance. Since $\mathcal{C}$ has pair-distance d, the resulting vectors of length $n-d+2$ remain distinct after deleting the last $d-2$ coordinates from all codewords. The maximum number of distinct vectors of length $n-d+2$ over an alphabet of size q is q^{n-d+2}. Hence, $|\mathcal{C}| \leq q^{n-d+2}$. ■

We call an $(n,d)_q$-symbol-pair code of size q^{n-d+2} *maximum distance separable* (MDS). In this paper, we construct new infinite classes of MDS symbol-pair codes and completely determine the existence of MDS symbol-pair codes for certain parameters.

III. MDS Symbol-Pair Codes From Classical MDS Codes

In this section, we give several methods for deriving MDS symbol-pair codes from classical MDS codes.

Note that $\mathcal{C} = \Sigma^n$ is trivially an MDS $(n,2)_q$-symbol-pair code for all $n \geq 2$ and $q \geq 2$ and so, we consider codes of pair-distance at least three.

A. MDS Symbol-Pair Codes and Classical MDS Codes

Recall that a classical MDS $(n,d)_q$-code, is a q-ary code of length n with Hamming distance d and size q^{n-d+1}. Exploiting the relationship between pair-distance and Hamming distance, we develop some general constructions for MDS symbol-pair codes and determine the existence of all such codes with pair-distance three.

Proposition 3.1: An MDS $(n,d)_q$-code with $d < n$ is an MDS $(n,d+1)_q$-symbol-pair code.

Proof: Let $\mathcal{C}$ be an MDS $(n,d)_q$-code of size q^{n-d+1}. By Proposition 2.1, $\mathcal{C}$ has pair-distance at least $d+1$. Therefore, $\mathcal{C}$ meets the Singleton bound of Theorem 2.1. ■

The following corollary follows immediately from classical MDS codes, mainly, Reed–Solomon codes and their extensions (see [6]).

Corollary 3.1:

i) There exists an MDS $(n,n-1)_q$-symbol-pair code for all $q = 2^m$, $m \geq 1$, and $n \leq q+2$.
ii) There exists an MDS $(n,5)_q$-symbol-pair code for all $q = 2^m$, $m \geq 1$, and $n \leq q+2$.
iii) There exists an MDS $(n,d)_q$-symbol-pair code whenever q is a prime power, $4 \leq d \leq n$ and $n \leq q+1$.
iv) There exists an MDS $(n,3)_q$-symbol-pair code for all $n \geq 2$, $q \geq 2$.

In particular, Corollary 3.1(iv) settles completely the existence of MDS $(n,3)_q$-symbol-pair codes.

B. MDS Symbol-Pair Codes From Interleaving Classical MDS Codes

We use the interleaving method of Cassuto and Blaum [3] to obtain MDS symbol-pair codes. Cassuto and Blaum showed that a symbol-pair code with even pair-distance can be obtained by interleaving two classical codes of the same length and distance.

Theorem 3.1 (Cassuto and Blaum [3]): If there exists an $(n,d)_q$-code of size M_1 and an $(n,d)_q$-code of size M_2, then there exists a $(2n,2d)_q$-symbol-pair code of size M_1M_2.

Interleaving classical MDS codes yield the following corollary.

Corollary 3.2:

i) There exists an MDS $(2n,2n-4)_q$-symbol-pair code for all $q = 2^m$, $m \geq 1$, and $n \leq q+2$.
ii) There exists an MDS $(2n,8)_q$-symbol-pair code for all $q = 2^m$, $m \geq 1$, and $n \leq q+2$.
iii) There exists an MDS $(2n,2d)_q$-symbol-pair code whenever q is a prime power, $3 \leq d \leq n-1$ and $n \leq q+1$.
iv) There exists an MDS $(2n,4)_q$-symbol-pair code for all $n \geq 2$, $q \geq 2$.
v) There exists an MDS $(2n,2n)_q$-symbol-pair code for all $n \geq 2$, $q \geq 2$.

Corollary 3.2 (iv) and (v) settle the existence of MDS $(n,4)_q$-symbol-pair codes and MDS $(n,n)_q$-symbol-pair codes for even n. In Section IV, we exhibit that such MDS codes indeed exist for all $n \geq 2$ and $q \geq 2$.

C. MDS Symbol-Pair Codes From Extending Classical MDS Codes

MDS symbol-pair codes obtained by interleaving necessarily have even length and distance. Furthermore, the length of symbol-pair codes obtained is only a factor of two longer than that of the input classical codes. In this section, we use graph theoretical concepts to extend classical MDS codes of length n to MDS symbol-pair codes of length up to $n(n-1)/2$.

We use standard concepts of graph theory presented by Bondy and Murty [7, Ch. 1–3]. Namely, a graph is a pair $G = (V,E)$, where V is a set of *vertices* and E is a set of unordered pairs of V, called *edges*. The *order* of G is $|V|$, the number of vertices, while the *size* of G is $|E|$, the number of edges.

A *trail* of length k in G is a list of vertices $v_0, v_1, \ldots, v_k$ such that $\{v_i, v_{i+1}\} \in E$ for $0 \leq i \leq k-1$, and $\{v_i, v_{i+1}\} \neq \{v_j, v_{j+1}\}$ for $0 \leq i < j \leq k-1$. The trail is said to be *closed* if $v_0 = v_k$. A closed trail $v_0, v_1, \ldots, v_k$ is a *cycle* if $v_i \neq v_j$ for $0 \leq i < j \leq k-1$. The length of a shortest cycle in a graph is called its *girth*.

On the other hand, a trail that transverses all edges in G is called an *eulerian trail*. If G admits a closed eulerian trail, then G is said to be *eulerian*. Equipped with the concepts of girth and eulerian trails, we introduce the next construction.

Proposition 3.2: Suppose there exists an MDS $(n,d)_q$-code and there exists an eulerian graph of order n, size m and girth at least $n-d+1$. Then, there exists an MDS $(m, m-n+d+1)_q$-symbol-pair code.

Proof: Let G be an eulerian graph of order n, size m and girth at least $n-d+1$, where $V(G)=\mathbb{Z}_n$. Consider a closed eulerian trail $T = x_0e_1x_1e_2x_2\cdots e_mx_m$, where $x_m = x_0$, $x_i \in V(G)$, and $e_i \in E(G)$, for $1 \le i \le m$. Let $\mathcal{C}$ be an MDS $(n,d)_q$-code and consider the q-ary code of length m,

$$\mathcal{C}' = \{(u_{x_0}, u_{x_1}, \ldots, u_{x_{m-1}}) : \mathbf{u} \in \mathcal{C}\}.$$

We claim that $\mathcal{C}'$ has pair-distance at least $m-n+d+1$. Indeed, pick any $\mathbf{u},\mathbf{v} \in \mathcal{C}$. Since $D_{\mathrm{H}}(\mathbf{u},\mathbf{v}) \ge d$, we have $|\{x \in V(G) : u_x = v_x\}| \le n-d$. It follows that

$$|\{i : (u_{x_i}, u_{x_{i+1}}) = (v_{x_i}, v_{x_{i+1}}), 0 \le i \le m-1\}| \le n-d-1,$$

since on the contrary there would exist at least $n-d$ edges $\{y_1,z_1\},\{y_2,z_2\},\ldots,\{y_{n-d},z_{n-d}\}$ in $E(G)$ such that $u_{y_j} = v_{y_j}$ and $u_{z_j} = v_{z_j}$ for all $1 \le j \le n-d$. But since the number of vertices $x \in V(G)$ such that $u_x = v_x$ is at most $n-d$, these $n-d$ edges must induce a subgraph (of order $n-d$) that contains a cycle of length at most $n-d$. This contradicts our assumption that G has girth at least $n-d+1$.

Consequently, $D_{\mathrm{p}}(\mathbf{u},\mathbf{v}) \ge m-n+d+1$. Finally, observe that $|\mathcal{C}'| = |\mathcal{C}| = q^{n-d+1}$, and hence $\mathcal{C}'$ is an MDS symbol-pair code by Theorem 2.1. ■

Example 3.1: Consider the complete graph K_5 of order five, whose vertex set is of $\mathbb{Z}_5$. Hence, its edge set comprises all ten unordered pairs of $\mathbb{Z}_5$. The graph K_5 is eulerian as it admits the closed eulerian trail, 01234024130. Trivially, the girth of K_5 is three.

Hence, given an MDS $(5,3)_q$-code $\mathcal{C}$ and since K_5 satisfies the conditions of Proposition 3.2, we have an MDS $(10,9)_q$-symbol-pair code.

More concretely, an MDS $(10,9)_q$-symbol-pair code is given by

$$\mathcal{C}' = \{(u_0,u_1,u_2,u_3,u_4,u_0,u_2,u_4,u_1,u_3) : \mathbf{u} \in \mathcal{C}\}.$$

Observe that when $q=4$, an MDS $(10,9)_4$-symbol-pair code cannot be obtained via Corollary 3.1 or Corollary 3.2.

To apply Proposition 3.2, we need eulerian graphs of specified order, size, and girth. However, little is known about how many edges an eulerian graph with a given number of vertices and given girth can have. Novák [8], [9] proved tight upper bounds on the number of edges in an eulerian graph of girth four. Below, we establish the following results on the size of an eulerian graph of order n (of girth three), and those of girth four.

Proposition 3.3: Let $n \ge 3$ and $M = n\lfloor (n-1)/2 \rfloor$. Then, there exists an eulerian graph of order n and size m, for $n \le m \le M$, except when $m \in \{M-1, M-2\}$.

Define

$$M(n) = \begin{cases} 2\lfloor n^2/8 \rfloor, & \text{if } n \text{ even} \\ 2\lfloor (n-1)^2/8 \rfloor + 1, & \text{if } n \text{ odd.} \end{cases}$$

Proposition 3.4: Let $n \ge 6$. Then, there exists an eulerian graph of order n, size m, and girth at least four, for all $m \equiv n \bmod 2$, $n \le m \le M(n)$, except when $m = M(n)-2$ and $n \ge 8$.

For $m \not\equiv n \bmod 2$, we have the following proposition.

Proposition 3.5:

i) For even $n \ge 10$, there exists an eulerian graph of order n, girth at least four, and size $m \in \{M(n-2)-1, M(n-2)+1\}$.
ii) For odd $n \ge 9$, there exists an eulerian graph of order n, girth at least four, and size $m \in \{M(n)-1, M(n)-3\}$.

We remark that Novák [8], [9] established the existence of eulerian graph of order n and girth at least four with size exactly $M(n)$. In contrast, Propositions 3.4 and 3.5 provide an eulerian graph of order n and girth at least four for a spectrum of sizes. Proofs for Propositions 3.3, 3.4, and 3.5 are deferred to Appendix A.

Corollary 3.3: Let q be a prime power, $q \ge 4$. Then, there exists an MDS $(n,n-1)_q$-symbol-pair code whenever

1) $2 \le n \le (q^2-1)/2-3$ or $n = (q^2-1)/2$, for q odd;
2) $2 \le n \le q(q+2)/2-3$ or $n = q(q+2)/2$, for q even.

Proof: Apply Propositions 3.2 and 3.3 to classical MDS codes. ■

Corollary 3.4: Let q be a prime power, $q \ge 5$. Then, there exists an MDS $(n,n-2)_q$-symbol-pair code whenever

1) $2 \le n \le M(q)+1$, or $M(q)+1 \le n \le M(q+1)$ and n even and $n \ne M(q+1)-2$, for q odd;
2) $2 \le n \le q^2/4+1$, $n \ne q^2/4-1$, for q even.

Proof: Apply Propositions 3.2, 3.4, and 3.5 to classical MDS codes. ■

These results show that in contrast to classical q-ary MDS codes of length n, where it is conjectured that $n \le q+2$, we can have q-ary MDS symbol-pair codes of length n with $n = \Omega(q^2)$.

IV. Construction of MDS Symbol-Pair Codes With Specific Lengths and Pair-Distances and the Existence of MDS Symbol-Pair Codes

Observe that while Section III constructs MDS symbol-pair codes from classical MDS codes, the latter is usually defined over a finite field, whose size is necessarily a prime power. Unfortunately, the set of prime powers has density zero in the set of positive integers.

In contrast, for fixed n and d, we conjecture that the set of alphabet sizes where an MDS $(n,d)_q$-symbol-pair code exists has density one. Specifically, we conjecture the following.

Conjecture: Fix $2 \le d \le n$. There exists a q_0 such that an MDS $(n,d)_q$-symbol-pair exists for all $q \ge q_0$.

In this section, we verify the conjecture for the following set of parameters.

1) $2 \le d \le 4$ and $d = n$, for all n,
2) $d = n-1$, for $6 \le n \le 8$, and,
3) $d = n-2$, for $7 \le n \le 10$.

To this end, we utilize a recursive method that builds an MDS symbol-pair code over a larger alphabet using MDS symbol-pair codes defined over smaller alphabets. This recursive construction is introduced formally in Section IV-C. However, to seed this recursion, the MDS symbol-pair codes given in Section III are insufficient. Therefore, we need additional MDS $(n,d)_q$-symbol-pair codes, particularly when q is not a prime power. These codes are given in Sections IV-A and IV-B.

A. $\mathbb{Z}_q$-Linear MDS Symbol-Pair Codes

We provide constructions for MDS $(n,d)_q$-symbol-pair codes for $d \in \{4,5,n\}$ and for certain small values of n, d and q. We remark that for even n, MDS $(n,4)_q$-symbol-pair codes have been constructed in Corollary 3.2, and MDS $(n,n)_q$-symbol-pair codes can be constructed by interleaving classical repetition codes. Here, we construct MDS $(n,4)_q$-symbol-pair codes and MDS $(n,n)_q$-symbol-pair codes for all n.

Throughout this section, we assume $\Sigma = \mathbb{Z}_q$. Besides being MDS, the codes constructed have $\mathbb{Z}_q$-*linearity*.

Definition 4.1: A code $\mathcal{C} \subseteq \Sigma^n$ is said to be $\mathbb{Z}_q$-*linear* if $\mathbf{u}+\mathbf{v}, \lambda\mathbf{u} \in \mathcal{C}$ for all $\mathbf{u},\mathbf{v} \in \mathcal{C}$, and $\lambda \in \mathbb{Z}_q$.

As with classical codes, a $\mathbb{Z}_q$-linear code must contain the zero vector $\mathbf{0}$. In addition, determining the minimum pair-distance of a $\mathbb{Z}_q$-linear code is equivalent to determining the minimum *pair-weight* of a nonzero codeword.

Definition 4.2: The *pair-weight* of $\mathbf{u} \in \Sigma^n$ is $\mathrm{wt_p}(\mathbf{u}) = D_\mathrm{p}(\mathbf{u},\mathbf{0})$.

The proof of the following lemma is similar to the classical case.

Lemma 4.1: Let $\mathcal{C}$ be a $\mathbb{Z}_q$-linear code. Then, $\mathcal{C}$ has pair-distance $\min_{\mathbf{u}\in\mathcal{C}\setminus\{\mathbf{0}\}} \mathrm{wt_p}(\mathbf{u})$.

In the rest of the section, the $\mathbb{Z}_q$-linear codes we construct are of size q^k. We describe such a code via a generator matrix in standard form, that is, a $k \times n$ matrix over $\mathbb{Z}_q$ of the form

$$G = (I_k | X),$$

so that each codeword is given by $\mathbf{u}G$, where $\mathbf{u} \in \mathbb{Z}_q^k$.

Proposition 4.1: Let $n \geq 4$ and $q \geq 2$. Let $\mathcal{C}$ be a $\mathbb{Z}_q$-linear code with generator matrix,

$$G = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 & 1 \\ 0 & 1 & \cdots & 0 & 2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & n-2 & 1 \end{pmatrix}.$$

Then, $\mathcal{C}$ is a $\mathbb{Z}_q$-linear MDS $(n,4)_q$-symbol-pair code.

Proof: It is readily verified that $\mathcal{C}$ has size q^{n-2}. Hence, by Lemma 4.1, it suffices to show that for all $\mathbf{u} \in \mathbb{Z}_q^{n-2} \setminus \{\mathbf{0}\}$,

$$\mathrm{wt_p}(\mathbf{u}G) \geq 4.$$

Write $\tilde{\mathbf{u}} = \left(u_0, u_1, \ldots, u_{n-3}, \sum_{i=0}^{n-3}(i+1)u_i, \sum_{i=0}^{n-3} u_i\right)$ and let

$$\Delta = \{i : \ 0 \leq i \leq n-3 \text{ and } u_i \neq 0\},$$
$$\Delta_\mathrm{p} = \{i : \ 0 \leq i \leq n-4 \text{ or } i = n-1, \text{ and } (u_i, u_{i+1}) \neq \mathbf{0}\}.$$

We have the following cases.

i) *The case* $|\Delta| \geq 3$:
Then $|\Delta_\mathrm{p}| \geq 4$, and so $\mathrm{wt_p}(\tilde{\mathbf{u}}) \geq 4$.

ii) *The case* $|\Delta| = 2$:
If $\Delta \neq \{j, j+1\}$ for all $0 \leq j \leq n-4$, then $|\Delta_\mathrm{p}| \geq 4$, and so $\mathrm{wt_p}(\tilde{\mathbf{u}}) \geq 4$. If $\Delta = \{j, j+1\}$ for some j, $0 \leq j \leq n-3$, then either $\tilde{\mathbf{u}}_{n-2}$ or $\tilde{\mathbf{u}}_{n-1}$ is nonzero. Otherwise,

$$(j+1)u_j + (j+2)u_{j+1} = 0,$$
$$u_j + u_{j+1} = 0,$$

which implies $u_{j+1} = 0$, a contradiction. Hence, $|\Delta_\mathrm{p}| \geq 3$, and since $\tilde{\mathbf{u}}_{n-2}$ or $\tilde{\mathbf{u}}_{n-1}$ is nonzero, $\mathrm{wt_p}(\tilde{\mathbf{u}}) \geq 4$.

iii) *The case* $|\Delta| = 1$:
If $u_0 \neq 0$, then both $\tilde{\mathbf{u}}_{n-2}$ and $\tilde{\mathbf{u}}_{n-1}$ are nonzero. Hence, $\mathrm{wt_p}(\tilde{\mathbf{u}}) \geq 4$. If $u_j \neq 0$ for some j, $1 \leq j \leq n-3$, then $\tilde{\mathbf{u}}_{n-1}$ is nonzero and $\{j-1, j, n-2, n-1\} \subseteq \{i : (u_i, u_{i+1}) \neq \mathbf{0}\}$ and hence, $\mathrm{wt_p}(\tilde{\mathbf{u}}) \geq 4$.

This completes the proof. ∎

Proposition 4.2: Let $n \geq 2$ and let $\mathcal{C}$ be a $\mathbb{Z}_q$-linear code with generator matrix

$$G = \begin{cases} \begin{pmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \end{pmatrix}, & \text{if } n \text{ is even,} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 & 1 \end{pmatrix}, & \text{otherwise.} \end{cases}$$

Then $\mathcal{C}$ is an MDS $(n,n)_q$-symbol-pair code.

Proof: It is readily verified that $\mathcal{C}$ has size q^2. Hence, by Lemma 4.1, it is also easy to see that the pair-weight of all nonzero vectors in $\mathcal{C}$ is n. ∎

Propositions 4.1 and 4.2 settle completely the existence of MDS $(n,4)$- and (n,n)-symbol-pair codes, respectively. When $5 \leq d \leq n-1$, the task is complex and hence, we determine the existence only for a certain set of parameters.

The next two propositions provide an infinite class and some small MDS symbol-pair codes required to seed the recursive method in Section IV-C.

Proposition 4.3: Suppose that q is odd prime and $5 \leq n \leq 2q+3$. Let $\mathcal{C}$ be a $\mathbb{Z}_q$-linear code with generator matrix,

$$G = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & \cdots & 0 & 2 & 1 & -1 \\ 0 & 0 & 1 & \cdots & 0 & 3 & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & n-3 & 1 & (-1)^{n-4} \end{pmatrix}.$$

Then, $\mathcal{C}$ is an MDS $(n,5)_q$-symbol-pair code.

Proof: It is readily verified that $\mathcal{C}$ has size q^{n-3}. Hence, by Lemma 4.1, it suffices to show that for all $\mathbf{u} \in \mathbb{Z}_q^{n-3} \setminus \{\mathbf{0}\}$,

$$\mathrm{wt_p}(\mathbf{u}G) \geq 5.$$

Define f, g, and h as follows:

$$f : \mathbb{Z}_q^{n-3} \longrightarrow \mathbb{Z}_q, \quad \mathbf{u} \longmapsto \sum_{i=0}^{n-4}(i+1)u_i,$$
$$g : \mathbb{Z}_q^{n-3} \longrightarrow \mathbb{Z}_q, \quad \mathbf{u} \longmapsto \sum_{i=0}^{n-4} u_i,$$
$$h : \mathbb{Z}_q^{n-3} \longrightarrow \mathbb{Z}_q, \quad \mathbf{u} \longmapsto \sum_{i=0}^{n-4}(-1)^i u_i.$$

Write $\tilde{\mathbf{u}} = (u_0, u_1, \ldots, u_{n-4}, f(\mathbf{u}), g(\mathbf{u}), h(\mathbf{u}))$ and let

$$\Delta = \{i : 0 \le i \le n-4, u_i \ne 0\},$$
$$\Delta_{\mathrm{p}} = \{i : i \in \mathbb{Z}_n, (\tilde{\mathbf{u}}_i, \tilde{\mathbf{u}}_{i+1}) \ne \mathbf{0}\}.$$

We have the following cases.

1) *The case* $|\Delta| \ge 4$:
Then $|\Delta_{\mathrm{p}}| \ge 5$ and so, $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$.
2) *The case* $|\Delta| = 3$:
If $\Delta \ne \{j, j+1, j+2\}$ for all $0 \le j \le n-6$, then $|\Delta_{\mathrm{p}}| \ge 5$ and so $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$. Otherwise, $\Delta = \{j, j+1, j+2\}$ for some $0 \le j \le n-6$. Then either $g(\mathbf{u})$ or $h(\mathbf{u})$ is nonzero. Otherwise,

$$u_j + u_{j+1} + u_{j+2} = 0,$$
$$u_j - u_{j+1} + u_{j+2} = 0,$$

implies that $2u_{j+1} = 0$. Since q is odd, $u_{j+1} = 0$, a contradiction. Hence, $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$.
3) *The case* $|\Delta| = 2$:
   a) Suppose that $\Delta = \{i, j\}$ with $j - i > 1$.
   If $j - i \equiv 1 \pmod 2$, then either $g(\mathbf{u})$ or $h(\mathbf{u})$ is nonzero, so $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$. Otherwise,

$$u_i + u_j = 0,$$
$$u_i - u_j = 0,$$

   implies that $2u_i = 0$. Since q is odd, $u_i = 0$, a contradiction.
   If $j - i \equiv 0 \pmod 2$, then either $f(\mathbf{u})$ or $g(\mathbf{u})$ is nonzero, so $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$. Otherwise,

$$(i+1)u_i + (j+1)u_j = 0,$$
$$u_i + u_j = 0$$

   implies that $(j-i)u_j = 0$. Since $j - i \le n-4 \le 2q-1$ is even and q is prime, $u_j = 0$, a contradiction.
   b) Suppose that $\Delta = \{j, j+1\}$ for some $0 \le j \le n-5$.
   If $j = 0$, then either $f(\mathbf{u})$ or $g(\mathbf{u}) = 0$ and hence, $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$. Otherwise, $j > 0$, then either $g(\mathbf{u})$ or $h(\mathbf{u}) = 0$ and so, $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$.
4) *The case* $|\Delta| = 1$:
   If $u_0 \ne 0$, then both $f(\mathbf{u})$ and $g(\mathbf{u})$ are nonzero. So, $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$. Otherwise, $u_j \ne 0$ for some $1 \le j \le n-4$. Then both $g(\mathbf{u})$ and $h(\mathbf{u})$ are nonzero and hence, $\mathrm{wt}_{\mathrm{p}}(\tilde{\mathbf{u}}) \ge 5$.

This completes the proof. ■

Proposition 4.4: There exist $\mathbb{Z}_q$-linear MDS $(n,d)_q$-symbol-pair codes for the following set of parameters:

i) $q = 2$, $(n,d) \in \{(6,5), (7,6), (7,5), (8,6), (9,7), (10,8)\}$,
ii) $q = 3$, $(n,d) \in \{(7,6), (8,7), (9,7), (10,8)\}$,
iii) $q = 5$, $(n,d) = (9,7)$.

Proof: Generator matrices for the respective codes are given in Table I. ■

TABLE I
GENERATOR MATRICES FOR $\mathbb{Z}_q$-LINEAR MDS SYMBOL-PAIR CODES

q	n	d	Generator matrix for a $\mathbb{Z}_q$-linear MDS $(n,d)_q$-symbol-pair code
2	6	5	$\begin{pmatrix} 1&0&0&1&0&1 \\ 0&1&0&1&1&0 \\ 0&0&1&1&1&1 \end{pmatrix}$
	7	5	$\begin{pmatrix} 1&0&0&0&1&0&1 \\ 0&1&0&0&1&1&1 \\ 0&0&1&0&1&1&0 \\ 0&0&0&1&0&1&1 \end{pmatrix}$
	9	7	$\begin{pmatrix} 1&0&0&0&1&0&1&1&0 \\ 0&1&0&0&0&1&0&1&1 \\ 0&0&1&0&1&1&0&1&0 \\ 0&0&0&1&1&1&1&1&1 \end{pmatrix}$
3	7	6	$\begin{pmatrix} 1&0&0&2&2&1&1 \\ 0&1&0&0&1&0&1 \\ 0&0&1&0&1&1&2 \end{pmatrix}$
	9	7	$\begin{pmatrix} 1&0&0&0&2&2&1&0&1 \\ 0&1&0&0&2&0&1&1&1 \\ 0&0&1&0&2&2&0&0&2 \\ 0&0&0&1&1&0&2&1&1 \end{pmatrix}$
5	9	7	$\begin{pmatrix} 1&0&0&0&1&0&1&0&1 \\ 0&1&0&0&0&1&0&1&1 \\ 0&0&1&0&1&0&2&0&3 \\ 0&0&0&1&0&1&0&2&3 \end{pmatrix}$
2	7	6	$\begin{pmatrix} 1&0&0&1&0&1&1 \\ 0&1&0&1&1&1&0 \\ 0&0&1&0&1&1&1 \end{pmatrix}$
	8	6	$\begin{pmatrix} 1&0&0&0&1&0&1&0 \\ 0&1&0&0&0&1&0&1 \\ 0&0&1&0&1&0&0&1 \\ 0&0&0&1&0&1&1&1 \end{pmatrix}$
	10	8	$\begin{pmatrix} 1&0&0&0&1&0&1&0&1&1 \\ 0&1&0&0&1&1&0&1&1&0 \\ 0&0&1&0&1&0&0&1&0&1 \\ 0&0&0&1&1&1&1&1&1&1 \end{pmatrix}$
3	8	7	$\begin{pmatrix} 1&0&0&1&1&1&2&0 \\ 0&1&0&0&1&1&1&2 \\ 0&0&1&1&1&2&0&1 \end{pmatrix}$
	10	8	$\begin{pmatrix} 1&0&0&0&1&1&1&2&2&0 \\ 0&1&0&0&0&1&1&1&1&2 \\ 0&0&1&0&0&2&0&1&2&2 \\ 0&0&0&1&1&1&2&2&1&2 \end{pmatrix}$

B. Family of MDS Symbol-Pair Codes via Development

We construct an MDS $(8,7)_{2p}$-symbol-pair code for all odd primes p. Similar to the concept of generator matrices, we obtain a full set of codewords by *developing* a smaller subset of codewords over some group. The concept of development is ubiquitous in combinatorial design theory (see [10, Chs. VI and VII],) and we construct the required MDS symbol-pair codes via this method.

We define the notion of development formally. Proofs in this section are deferred to Appendix B.

Definition 4.3: Let n be even and Γ be an Abelian additive group. A Γ^2-*development* $(n, n-1)$-*symbol-pair code* is a set of q codewords in Γ^n such that for distinct codewords $\mathbf{u}$, $\mathbf{v}$, the following hold.

i) $(u_i - u_j, u_{i+1} - u_{j+1}) \ne (v_i - v_j, v_{i+1} - v_{j+1})$ for $i, j \in \mathbb{Z}_n$, $i \equiv j \bmod 2$, and,
ii) $(u_i - u_{j+1}, u_{i+1} - u_j) \ne (v_i - v_{j+1}, v_{i+1} - v_j)$ for $i, j \in \mathbb{Z}_n$, $i \not\equiv j \bmod 2$.

Proposition 4.5: Let n be even. Suppose $\mathcal{C}_0$ is a Γ^2-development $(n, n-1)$-symbol-pair code with $|\Gamma| = q$.

For $\mathbf{u} \in \mathcal{C}_0$, $\alpha, \alpha' \in \Gamma$, let

$$\phi(\mathbf{u}, \alpha, \alpha') = (u_0 + \alpha, u_1 + \alpha', u_2 + \alpha, u_3 + \alpha', \ldots, u_{n-2} + \alpha, u_{n-1} + \alpha'). \quad (1)$$

Then, $\mathcal{C} = \{\phi(\mathbf{u}, \alpha, \alpha') : \mathbf{u} \in \mathcal{C}_0, \alpha, \alpha' \in \Gamma\}$ is an MDS $(n, n-1)_q$-symbol-pair code.

Therefore, to construct an MDS $(n, n-1)_q$-symbol-pair code, it suffices to construct a set of q codewords, instead of a set of q^2 codewords. Hence, for certain values of n and q, a computer search is effective to construct MDS symbol-pair codes. In the instance when $n = 8$, we have the following collection of Γ^2-development MDS $(8,7)_q$-symbol-pair codes.

Proposition 4.6: Let p be prime with $p \ge 5$ and $\Gamma = \mathbb{Z}_p \times \mathbb{Z}_2$. Let $\mathcal{C}_0$ consist of the following four codewords:

$$((0,0),(0,0),(0,0),(1,0),(0,0),(1,1),(0,0),(0,1)),$$
$$((0,0),(0,0),(0,1),(1,1),(2,0),(0,1),(2,1),(2,0)),$$
$$((0,0),(0,0),(1,0),(0,0),(1,1),(0,0),(0,1),(0,0)),$$
$$((0,0),(0,0),(1,1),(0,1),(0,1),(2,0),(2,0),(2,1)).$$

Let $\mathcal{C}_1$ be the following set of $2p-4$ codewords:

$$((0,0),(0,0),(a,0),(\hat{a},1),(3a,1),(0,1),(2a,1),(2\hat{a},0)),$$
$$((0,0),(0,0),(a,1),(a,0),(0,1),(3a,1),(2a,0),(2a,1)),$$

where $a \in \{2,3,\ldots,p-1\}$ and

$$\hat{a} = \begin{cases} p-1, & \text{if } a = 2, \\ a-1, & \text{otherwise.} \end{cases}$$

Then, $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1$ is a Γ^2-development $(8,7)$-symbol-pair code.

In addition, when $p = 3$, a $\mathbb{Z}_6^2$-development $(8,7)$-symbol-pair code is given by the following six codewords:

$$(0,0,0,0,0,0,0,0),$$
$$(0,0,1,1,0,5,1,2),$$
$$(0,0,2,2,4,5,3,4),$$
$$(0,0,3,3,0,4,2,5),$$
$$(0,0,4,4,2,3,5,1),$$
$$(0,0,5,5,0,1,4,3).$$

Therefore, applying Propositions 4.5 and 4.6, we have the following existence result.

Corollary 4.1: There exists an MDS $(8,7)_{2p}$-symbol-pair code for odd primes p.

C. Complete Solution of the Existence of MDS Symbol-Pair Codes for Certain Parameters

We settle completely the existence of MDS symbol-pair codes for certain parameters.

In particular, define

$$q(n,d) = \min\{q_0 : \text{an MDS } (n,d)_q\text{-symbol-pair code exists for all } q \geq q_0\},$$

and we establish the following.

Theorem 4.1: The following hold.

i) $q(n,d) = 2$ for $2 \leq d \leq 4$ and $n \geq d$, or $d = n$,
ii) $q(n,n-1) = 2$ for $n \in \{6,7\}$, $q(8,7) = 3$ and,
iii) $q(n,n-2) = 2$ for $7 \leq n \leq 10$.

Observe that Theorem 4.1(i) follows from the opening remark in Section III, Corollary 3.1(iv), and Propositions 4.1 and 4.2. For Theorem 4.1(ii) and Theorem 4.1(iii), we require the following recursive construction.

Proposition 4.7 (Product Construction): If there exists an MDS $(n,d)_{q_1}$-symbol-pair code and an MDS $(n,d)_{q_2}$-symbol-pair code, then there exists an MDS $(n,d)_{q_1 q_2}$-symbol-pair code.

Proof: Let $\mathcal{C}_i$ be an MDS $(n,d)_{q_i}$-symbol-pair code over Σ_i for $i = 1, 2$. For $\mathbf{u} \in \mathcal{C}_1$ and $\mathbf{v} \in \mathcal{C}_2$, let $\mathbf{u} \times \mathbf{v} = ((u_0,v_0),(u_1,v_1),\ldots,(u_{n-1},v_{n-1})) \in (\Sigma_1 \times \Sigma_2)^n$.

Consider the code $\mathcal{C}$ over $\Sigma_1 \times \Sigma_2$,

$$\mathcal{C} = \{\mathbf{u} \times \mathbf{v} : \mathbf{u} \in \mathcal{C}_1, \mathbf{v} \in \mathcal{C}_2\}.$$

It is readily verified that $|\mathcal{C}| = (q_1 q_2)^{n-d+2}$ and it remains to verify that the minimum pair-distance is at least d.

TABLE II
SOME MDS SYMBOL-PAIR CODES

n	d	q	Authority
6	5	2	Proposition 4.4
		p odd prime	Proposition 4.3
7	6	2,3	Proposition 4.4
		$p \geq 5$, odd prime	Corollary 3.3
8	7	3	Proposition 4.4
		$p \geq 5$, odd prime	Corollary 3.3
		$2p$, p odd prime	Corollary 4.1
		2^r, $r \geq 2$	Corollary 3.3
7	5	2	Proposition 4.4
		p, p odd prime	Proposition 4.3
8	6	2	Proposition 4.4
		p, p odd prime	Corollary 3.2
9	7	2,3,5	Proposition 4.4
		$p \geq 7$, p odd prime	Corollary 3.4
10	8	2,3	Proposition 4.4
		$p \geq 5$, p odd prime	Corollary 3.2

Indeed for distinct $(\mathbf{u} \times \mathbf{v}), (\mathbf{u}' \times \mathbf{v}') \in \mathcal{C}$,

$$D_{\mathrm{p}}(\mathbf{u} \times \mathbf{v}, \mathbf{u}' \times \mathbf{v}') \geq \max\{D_{\mathrm{p}}(\mathbf{u},\mathbf{u}'), D_{\mathrm{p}}(\mathbf{v},\mathbf{v}')\} \geq d.$$

■

Proof of Theorem 4.1(ii) and (iii): Define

$$Q(2) = \{p : p \text{ prime}\},$$
$$Q(3) = \{p : p \geq 3 \text{ prime}\} \cup \{2p : p \geq 3 \text{ prime}\} \cup \{2^r : r \geq 2\}.$$

To show that $q(n,d) \leq q_0 (q_0 \in \{2,3\})$, it suffices by Proposition 4.7 to construct MDS $(n,d)_q$-symbol-pair codes for $q \in Q(q_0)$. The required MDS $(n,d)_q$-symbol-pair codes are constructed in Section III, IV-A, and IV-B. We summarize the results in Table II.

Observe that $q(n,d) \geq 2$ trivially. However, when $(n,d) = (8,7)$, regard an $(8,7)_2$-symbol-pair code as a (classical) $(8,7)_4$-code, whose size is at most seven by Plotkin bound. Hence, an MDS $(8,7)_2$-symbol-pair code whose size is eight cannot exist and so, $q(8,7) \geq 3$. ■

V. CONCLUSION

In this paper, we established a Singleton-type bound for symbol-pair codes and constructed infinite families of optimal symbol-pair codes. All these codes are of the maximum distance separable (MDS) type in that they meet the Singleton-type bound. We also show how classical MDS codes can be extended to MDS symbol-pair codes using eulerian graphs of specified girth. In contrast with q-ary classical MDS codes, where all known such codes have length $O(q)$, we establish that q-ary MDS symbol-pair codes can have length $\Omega(q^2)$. In addition, we gave complete solutions to the existence of MDS symbol-pair codes for certain parameters.

APPENDIX A
EULERIAN GRAPHS OF SPECIFIED SIZE AND GIRTH

We give detailed proofs of Propositions 3.3, 3.4, and 3.5. In particular, we construct eulerian graphs with girth at least three and four and specified sizes.

A graph $G = (V,E)$ is said to be *even* if the degree of each vertex is even. Hence, we have the following characterization of eulerian graphs due to Euler.

TABLE III
EULERIAN GRAPHS OF SMALL SIZE WITH ORDER n, GIRTH AT LEAST THREE

$n = 2k+1$, $V = \mathbb{Z}_{2k} \cup \{\infty\}$

m	m_1	m_2	H_m
$2l+1$ for $1 \le l \le k-1$	0	$k-l$	$(\infty, 0, -1, 1, -2, \ldots, -l)$
$2l$ for $2 \le l \le k$	0	$l-1$	$(\infty, 0, -1, 1, -2, \ldots, l-1)$
$2k+1$	0	1	Φ_0
$2n-2l-1$ for $1 \le l \le k-1$	0	$k-l$	$\Phi_0 \cup \Phi_{k-l} \setminus (\infty, 0, -1, 1, -2, \ldots, -l)$
$2n-2l$ for $2 \le l \le k$	0	$l-1$	$\Phi_0 \cup \Phi_{l-1} \setminus (\infty, 0, -1, 1, -2, \ldots, l-1)$

$n = 2k+2$, $V = \mathbb{Z}_{2k+1} \cup \{\infty\}$

m	m_1	m_2	H_m
3	0	1	$(0, -1, 1)$
$2l+1$ for $2 \le l \le k$	0	$k-l+1$	$(\infty, 0, -1, 1, -2, \ldots, -l)$
$2l$ for $2 \le l \le k$	0	$l-1$	$(\infty, 0, -1, 1, -2, \ldots, l-1)$
$2k+2$	0	1	Φ_0
$2n-3$	0	1	$\Phi_0 \cup \Phi_1 \setminus (0, -1, 1)$
$2n-2l-1$ for $2 \le l \le k-1$	0	$k-l+1$	$\Phi_0 \cup \Phi_{k-l} \setminus (\infty, 0, -1, 1, -2, \ldots, -l)$
$2n-2l$ for $2 \le l \le k$	0	$l-1$	$\Phi_0 \cup \Phi_{l-1} \setminus (\infty, 0, -1, 1, -2, \ldots, l-1)$

TABLE IV
EULERIAN GRAPHS OF SMALL SIZE WITH ORDER n, GIRTH AT LEAST FOUR

$n = 4k$ or $n' = 2k$, $V = \mathbb{Z}_{n'} \cup \{\bullet, \circ\}$

m	m_1	m_2	H_m
$4l$ for $1 \le l \le k-1$	0	l	$(0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l-1)_\bullet, (2l-1)_\circ)$
$4l+2$ for $1 \le l \le k-1$	0	l	$(0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l)_\bullet, (2l)_\circ)$
$4k$	0	1	Φ_0
$2n-4l$ for $1 \le l \le k-1$	0	l	$\Phi_0 \cup \Phi_l \setminus (0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l-1)_\bullet, (2l-1)_\circ)$
$2n-4l-2$ for $1 \le l \le k-1$	0	l	$\Phi_0 \cup \Phi_l \setminus (0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l)_\bullet, (2l)_\circ)$

$n = 4k+2$ or $n' = 2k+1$, $V = \mathbb{Z}_{n'} \cup \{\bullet, \circ\}$

m	m_1	m_2	H_m
$4l$ for $1 \le l \le k-1$	0	l	$(0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l-1)_\bullet, (2l-1)_\circ)$
$4l+2$ for $1 \le l \le k-1$	0	l	$(0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l)_\bullet, (2l)_\circ)$
$4k$	0	1	$(0_\bullet, 2_\circ, 1_\bullet, 3_\circ, \ldots, (n'-2)_\bullet, 0_\circ)$
$4k+2$	0	1	Φ_0
$4k+4$	0	1	$\Phi_0 \cup \Phi_1 \setminus (0_\bullet, 2_\circ, 1_\bullet, 3_\circ, \ldots, (n'-2)_\bullet, 0_\circ)$
$2n-4l$ for $1 \le l \le k-1$	0	l	$\Phi_0 \cup \Phi_l \setminus (0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l-1)_\bullet, (2l-1)_\circ)$
$2n-4l-2$ for $1 \le l \le k-1$	0	l	$\Phi_0 \cup \Phi_l \setminus (0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (2l)_\bullet, (2l)_\circ)$

Theorem A.1. (See [7, Th. 3.5]): Let G be a connected graph. Then G is eulerian if and only if G is an even graph.

Next, we define certain operations on graphs which aid us in constructing even graphs.

i) Let G, H be graphs defined on the same vertex set V. We denote the graph $(V, E(G) \cup E(H))$ by $G \cup H$ and the graph $(V, E(G) \setminus E(H))$ by $G \setminus H$. Suppose G and H are even graphs. If G and H are edge-disjoint, then $G \cup H$ is even and if in addition, $G \cup H$ is connected, then eulerian by Theorem A.1. Similarly, if $E(G) \supset E(H)$, then $G \setminus H$ is even and eulerian (if $G \setminus H$ is connected).

ii) Let $G = (V, E)$ be a graph with vertices u, v and edge $e = \{u, v\}$. We *subdivide edge* e (see [7, Sec. 2.3]) to obtain the graph $(V \cup \{x\}, E \setminus \{e\} \cup \{\{u, x\}, \{v, x\}\})$. In other words, we add the vertex x and replace the edge $\{u, v\}$ with the edges $\{u, x\}$ and $\{v, x\}$. Suppose G is an eulerian graph with order n, size m, and girth g. Then subdividing any edge of G, we obtain an eulerian graph with order $n+1$, size $m+1$ and girth at least g.

With these operations, we prove the stated propositions.

1) Proof of Proposition 3.3: The proposition is readily verified for $n \in \{3, 4\}$. When $n \ge 5$, let $k = \lfloor (n-1)/2 \rfloor$ and we prove the proposition by induction. We first construct eulerian graphs of small sizes and then inductively add edge-disjoint Hamilton cycles to obtain eulerian graphs of the desired sizes.

Define the following collection of k edge-disjoint Hamilton cycles in K_n.

TABLE V
DIFFERENCES $u_i - u_{i+2}$ FOR $\mathbf{u} \in \mathcal{C}$, $i \in \mathbb{Z}_8$

i	$\mathcal{C}_0$	$u_i - u_{i+2}$ $\mathcal{C}_1$
0	$\{(0,0),(0,1),(-1,0),(-1,1)\}$	$\{(-a,0),(-a,1)\}$ for $a \in \{2,3,\ldots,p-1\}$
1	$\{(-1,0),(-1,1),(0,0),(0,1)\}$	$\{(-\hat{a},1),(-a,0)\}$ for $a \in \{2,3,\ldots,p-1\}$
2	$\{(0,0),(-2,1),(0,1),(1,0)\}$	$\{(-2a,1),(a,0)\}$ for $a \in \{2,3,\ldots,p-1\}$
3	$\{(0,1),(1,0),(0,0),(-2,1)\}$	$\{(\hat{a},0),(-2a,1)\}$ for $a \in \{2,3,\ldots,p-1\}$
4	$\{(0,0),(0,1),(1,0),(-2,1)\}$	$\{(a,0),(-2a,1)\}$ for $a \in \{2,3,\ldots,p-1\}$
5	$\{(1,0),(-2,1),(0,0),(0,1)\}$	$\{(-2\hat{a},1),(a,0)\}$ for $a \in \{2,3,\ldots,p-1\}$
6	$\{(0,0),(2,1),(0,1),(2,0)\}$	$\{(2a,1),(2a,0)\}$ for $a \in \{2,3,\ldots,p-1\}$
7	$\{(0,1),(2,0),(0,0),(2,1)\}$	$\{(2\hat{a},0),(2a,1)\}$ for $a \in \{2,3,\ldots,p-1\}$

a) When $n = 2k+1$, let $V = \mathbb{Z}_{2k} \cup \{\infty\}$. For $0 \le i \le k-1$, the Hamilton cycle Φ_i is given by

$$\Phi_i = (\infty, i, i-1, i+1, \ldots, i-k+1, i+k-1, i-k).$$

b) When $n = 2k+2$, let $V = \mathbb{Z}_{2k+1} \cup \{\infty\}$. For $0 \le i \le k-1$, the Hamilton cycle Φ_i is given by

$$\Phi_i = (\infty, i, i-1, i+1, \ldots, i-k, i+k).$$

For $3 \le m \le 2n-3$, there exists two Hamilton cycles Φ_{m_1}, Φ_{m_2}, and a subgraph H_m such that the following holds:

i) H_m is a subgraph of $\Phi_{m_1} \cup \Phi_{m_2}$,

ii) H_m is even with size m and when $m \ge n$, H_m is connected and hence, eulerian.

We give explicit constructions of Φ_{m_1}, Φ_{m_2}, H_m in Table III.

Then, for $2n-3 < m \le kn-3$, choose $1 \le r \le k-2$ such that $3 \le m - rn \le 2n-3$. Let $m' = m - rn$ and choose r Hamilton cycles $\Phi_{j_1}, \Phi_{j_2}, \ldots, \Phi_{j_r}$ such that $j_s \notin \{m'_1, m'_2\}$. Then, $H_{m'} \cup (\bigcup_{s=1}^{r} \Phi_{j_s})$ is an eulerian graph of size m since H_m is even, contains a Hamilton cycle and is hence connected. ■

2) Proof of Proposition 3.4: The proposition can be readily verified for $n \in \{6, 7\}$.

First, we prove for the case n even.

Let $n' = n/2$ and $k = \lfloor n'/2 \rfloor$ and we show that there exists an eulerian graph of order n, girth at least four and size m, for $n \le m \le nk$ and m even, except for $m = nk-2$. The proof for n even is similar to proof of Proposition 3.3.

Consider the following collection of k edge-disjoint Hamilton cycles in $K_{n',n'}$ due to Dirac [11].

Let the vertex set $V = (\mathbb{Z}_{n'} \times \{\bullet, \circ\})$ and the partitions be $\mathbb{Z}_{n'} \times \{\bullet\}$ and $\mathbb{Z}_{n'} \times \{\circ\}$). Write (a, b) as a_b and for $0 \le i \le k-1$, consider the Hamilton cycle Φ_i given by

$$\Phi_i = (0_\bullet, (2i)_\circ, 1_\bullet, (1+2i)_\circ, \ldots, (n'-1)_\bullet, (n'-1+2i)_\circ).$$

As in Proposition 3.3, for $4 \le m \le 2n-4$, there exists two Hamilton cycles Φ_{m_1} and Φ_{m_2} and a subgraph H_m such that the following holds:

i) H_m is a subgraph of $\Phi_{m_1} \cup \Phi_{m_2}$,

ii) H_m is even with size m and when $m \le n$, H_m is connected and hence, eulerian.

We give explicit constructions of Φ_{m_1}, Φ_{m_2}, and H_m in Table IV and the rest of the proof proceeds in the same manner. Since the graphs constructed are subgraphs of $K_{n',n'}$, their girths are at least four.

Recall that $M(n) = 2\lfloor n^2/8 \rfloor$ when n is even. When $n = 4k$, $M(n) = 4k^2 = nk$ and hence, the stated graphs are constructed.

TABLE VI
DIFFERENCES $(u_i - u_j, u_{i+1} - u_{j+1})$ FOR $\mathbf{u} \in \mathcal{C}$ AND $i - j \equiv 4 \bmod 8$

	$(u_i - u_j, u_{i+1} - u_{j+1})$	
(i,j)	$\mathcal{C}_0$	$\mathcal{C}_1$
(0,4)	$\{((0,0),(-1,1)),((-2,0),(0,1)),((-1,1),(0,0)),((0,1),(-2,0))\}$	$\{((-3a,1),(0,1)),((0,1),(-3a,1))\}$ for $a \in \{2,3,\ldots,p-1\}$
(1,5)	$\{((-1,1),(0,0)),((0,1),(-2,0)),((0,0),(1,1)),((-2,0),(-1,1))\}$	$\{((0,1),(-a,1)),((-3a,1),(-a,1))\}$ for $a \in \{2,3,\ldots,p-1\}$
(2,6)	$\{((0,0),(1,1)),((-2,0),(-1,1)),((1,1),(0,0)),((-1,1),(-2,0))\}$	$\{((-a,1),(-\hat{a},1)),((-a,1),(-a,1))\}$ for $a \in \{2,3,\ldots,p-1\}$
(3,7)	$\{((1,1),(0,0)),((-1,1),(2,0)),((0,0),(1,1)),((-2,0),(0,1))\}$	$\{((-\hat{a},1),(3a,1)),((-a,1),(0,1))\}$ for $a \in \{2,3,\ldots,p-1\}$

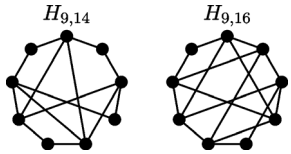

Fig. 1. Eulerian Graphs of order 9 and size 14, 16.

When $n = 4k + 2$, note that $K_{2k,2k+2}$ (defined on partitions $\mathbb{Z}_{2k} \times \{\bullet\}$, $\mathbb{Z}_{2k+2} \times \{\circ\}$) is an eulerian graph with size $M(n) = 4k^2+4k$ and girth at least four. Observe that $K_{2k,2k+2}$ contains cycles of even length $4 \leq m' \leq 2k + 2$, namely, $(0_\bullet, 0_\circ, 1_\bullet, 1_\circ, \ldots, (m'/2-1)_\bullet, (m'/2-1)_\circ)$. Hence, removing a cycle of length m', we obtain eulerian graphs with order n and girth at least four with size m, $nk - 2 \leq m \leq M(n) - 4$.

Finally, when n is odd, let m be odd, with $n \leq m \leq M(n)$ and $m \neq M(n) - 2$. Then, there exists an eulerian graph H with order $n - 1$, size $m - 1$, and girth at least four. Pick any edge in H and subdivide the edge to obtain an eulerian graph with order n, size m, and girth at least four. This completes the proof. ■

3) Proof of Proposition 3.5: Eulerian graphs with order nine, girth four, and sizes 14, 16 are given in Fig. 1. For each graph of order nine, subdivide any edge to obtain an eulerian graph of order ten, girth four, and orders 15, 17. Denote these graphs by $H_{n,m}$, where n is the order and m is the size.

For $n \geq 11$, let $n' = 2\lfloor (n-1)/2 \rfloor$. Then, $K_{2\lfloor n'/4 \rfloor, 2\lceil n'/4 \rceil}$ is a graph of order n', girth four, and size $M(n')$, containing a subgraph $K_{4,4}$. Replacing the subgraph $K_{4,4}$ with

$$\begin{cases} H_{9,14} \text{ or } H_{9,16}, & \text{if } n \text{ is odd,} \\ H_{10,15} \text{ or } H_{10,17}, & \text{otherwise,} \end{cases}$$

yields an eulerian graph of order n, girth at least four with the desired sizes. ■

APPENDIX B
MDS SYMBOL-PAIR CODES VIA DEVELOPMENT

We provide detailed proofs of propositions given in Section IV-B.

1) Proof of Proposition 4.5: It is readily verified that $|\mathcal{C}| = q^3$ and so, it remains to show that $\mathcal{C}$ has minimum pair-distance $n - 1$.

Suppose otherwise that there exist distinct codewords $\phi(\mathbf{u}, \alpha, \alpha')$ and $\phi(\mathbf{v}, \beta, \beta')$ in $\mathcal{C}$ with

$$D_{\mathrm{p}}(\phi(\mathbf{u}, \alpha, \alpha'), \phi(\mathbf{v}, \beta, \beta')) < n - 1.$$

Then, there exist $i, j \in \mathbb{Z}_n$, $i \neq j$, such that

$$(\phi(\mathbf{u}, \alpha, \alpha')_i, \phi(\mathbf{u}, \alpha, \alpha')_{i+1}) = (\phi(\mathbf{v}, \beta, \beta')_i, \phi(\mathbf{v}, \beta, \beta')_{i+1}),$$
$$(\phi(\mathbf{u}, \alpha, \alpha')_j, \phi(\mathbf{u}, \alpha, \alpha')_{j+1}) = (\phi(\mathbf{v}, \beta, \beta')_j, \phi(\mathbf{v}, \beta, \beta')_{j+1}).$$

Without loss of generality, assume $i \equiv 0 \bmod 2$. Suppose $j \equiv 0 \bmod 2$. Then,

$$(u_i + \alpha, u_{i+1} + \alpha') = (v_i + \beta, v_{i+1} + \beta'),$$
$$(u_j + \alpha, u_{j+1} + \alpha') = (v_j + \beta, v_{j+1} + \beta').$$

Hence,

$$(u_i - u_j, u_{i+1} - u_{j+1}) = (v_i - v_j, v_{i+1} - v_{j+1}),$$

contradicting Condition (i) in Definition 4.3.

Similarly, when $j \equiv 1 \bmod 2$,

$$(u_i + \alpha, u_{i+1} + \alpha') = (v_i + \beta, v_{i+1} + \beta'),$$
$$(u_j + \alpha', u_{j+1} + \alpha) = (v_j + \beta', v_{j+1} + \beta),$$

and so,

$$(u_i - u_{j+1}, u_{i+1} - u_j) = (v_i - v_{j+1}, v_{i+1} - v_j).$$

We derive a contradiction to Condition (ii) in Definition 4.3. ■

2) Proof of Proposition 4.6: We exhibit that $\mathcal{C}$ is a $(\mathbb{Z}_p \times \mathbb{Z}_2)^2$-development $(8, 7)$-symbol-pair code, by checking the conditions of Definition 4.3.

The values of $u_i - u_{i+2}$ for $\mathbf{u} \in \mathcal{C}, i \in \mathbb{Z}_8$ are given in Table V and we verify that for $i \in \mathbb{Z}_8$

$$u_i - u_{i+2} \neq v_i - v_{i+2} \text{ for } \mathbf{u},\ \mathbf{v} \in \mathcal{C}. \tag{2}$$

For Condition (i), note that when $j = i + 2$, (2) ensures that the differences $(u_i - u_{i+2}, u_{i+1} - u_{i+3})$ are distinct. Hence, it remains to check when $i - j \equiv 4 \bmod 8$ and these values are given in Table VI.

For Condition (ii), if $i \not\equiv j \bmod 2$, then either $j + 1 = i + 2$, $i + 1 = j + 2$, $j = i + 3$, or $i = j + 3$ since $n = 8$. Equation (2) ensures that the values $(u_i - u_{j+1}, u_{i+1} - u_j)$ are distinct. ■

ACKNOWLEDGMENT

We are grateful to the anonymous reviewers and Associate Editor Professor Kashyap for their constructive comments, which improved greatly the presentation of this paper.

REFERENCES

[1] Y. M. Chee, H. Kiah, and C. Wang, "Maximum distance separable symbol-pair codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, 2012, pp. 2896–2900.
[2] Y. Cassuto and M. Blaum, "Codes for symbol-pair read channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 988–992.
[3] Y. Cassuto and M. Blaum, "Codes for symbol-pair read channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 8011–8020, Dec. 2011.
[4] Y. Cassuto and S. Litsyn, "Symbol-pair codes: Algebraic constructions and asymptotic bounds," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 2348–2352.
[5] E. Yaakobi, J. Bruck, and P. H. Siegel, "Decoding of cyclic codes over symbol-pair read channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, 2012, pp. 2891–2895.

[6] A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal Arrays*, ser. Springer Series in Statistics. New York, NY, USA: Springer-Verlag, 1999.
[7] J. A. Bondy and U. S. R. Murty, *Graph Theory*, ser. Graduate Texts in Mathematics. New York, NY, USA: Springer-Verlag, 2008.
[8] J. Novák, "Eulerovské grafy bez trojúhelníků s maximálním počtem hran," *Sborník vědeckých prací VŠST, Liberec*, 1971.
[9] J. Novák, "Edge bases of complete uniform hypergraphs," *Mat. Časopis Sloven. Akad. Vied*, vol. 24, pp. 43–57, 1974.
[10] T. Beth, D. Jungnickel, and H. Lenz, *Design Theory*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[11] G. A. Dirac, "On Hamilton circuits and Hamilton paths," *Math. Ann.*, vol. 197, pp. 57–70, 1972.

Yeow Meng Chee (SM'08) received the B.Math. degree in computer science and combinatorics and optimization and the M.Math. and Ph.D. degrees in computer science, from the University of Waterloo, Waterloo, ON, Canada, in 1988, 1989, and 1996, respectively.

Currently, he is an Associate Professor at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. Prior to this, he was Program Director of Interactive Digital Media R&D in the Media Development Authority of Singapore, Postdoctoral Fellow at the University of Waterloo and IBM's Zürich Research Laboratory, General Manager of the Singapore Computer Emergency Response Team, and Deputy Director of Strategic Programs at the Infocomm Development Authority, Singapore. His research interest lies in the interplay between combinatorics and computer science/engineering, particularly combinatorial design theory, coding theory, extremal set systems, and electronic design automation.

Lijun Ji received the Ph.D. degree in applied mathematics from Suzhou University, China, in 2003. He is currently a Professor at Department of Mathematics in Suzhou University. His research interests include combinatorial design theory and coding theory.

Han Mao Kiah received the B.Sc. (Hon) degree in mathematics from the National University of Singapore, Singapore in 2006. Currently, he is working towards his Ph.D. degree at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. His research interest lies in the application of combinatorics to engineering problems in information theory. In particular, his interests include combinatorial design theory, coding theory and power line communications.

Chengmin Wang received the B.Math. and Ph.D. degrees in mathematics from Suzhou University, China in 2002 and 2007, respectively. Currently, he is an Associate Professor at the School of Science, Jiangnan University, China. Prior to this, he was a Visiting Scholar at the School of Computing, Informatics and Decision Systems Engineering, Arizona State University, USA, from 2010 to 2011 and was a Research Fellow at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore from 2011 to 2012. His research interests include combinatorial design theory and its applications in coding theory and cryptography.

Jianxing Yin graduated from Suzhou University, Suzhou, China, in 1977. Since 1977, he has been a Teacher in the Department of Mathematics, Suzhou University. Currently, he is a Full Professor there. He is an editorial board member of Journal of Combinatorial Designs and an associate editor of Discrete Mathematics, Algorithms and Applications. He has held various grants of the National Natural Science Foundation of China (NSFC) as Project Leader. His research interests include combinatorial design theory, combinatorial coding theory and the application of combinatorics to software testing.