Maximum Distance Separable Symbol-Pair Codes

Yeow Meng Chee, Han Mao Kiah, and Chengmin Wang
School of Physical and Mathematical Sciences, Nanyang Technological University
21 Nanyang Link, Singapore 637371

***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 $\Omega(q^2)$. We also construct equidistant cyclic MDS symbol-pair codes from Mendelsohn designs.**

I. Introduction

Symbol-pair coding theory has recently been introduced by Cassuto and Blaum [1], [2] 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 rather rudimentary stage. Cassuto and Blaum [1], [2] 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 [3] 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.

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 completely settle the existence of MDS symbol-pair codes of length n with pair-distance d, for $2 \leq d \leq 4$ and $d = n$. We also 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 provide a new construction for equidistant cyclic MDS symbol-pair codes based on Mendelsohn designs.

II. Preliminaries

Throughout this paper, Σ is a set of q elements, called *symbols*. For a positive integer n, $\mathbb{Z}_n$ 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 $\mathbf{u} = (u_0, u_1, \cdots, u_{n-1})$.

A *pair-vector* over Σ is a vector in $(\Sigma \times \Sigma)^n$. For any $\mathbf{u} = (u_0, u_1, \cdots, u_{n-1}) \in \Sigma^n$, the *symbol-pair read vector* of $\mathbf{u}$ is the pair-vector (over Σ)

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

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

$$D_{\mathrm{p}}(\mathbf{u}, \mathbf{v}) = |\{i \in \mathbb{Z}_n : u_i \neq v_i\}|.$$

The pair-distance between two vectors in Σ^n is the pair-distance between their corresponding symbol-pair read vectors, and if $\mathbf{u}, \mathbf{v} \in \Sigma^n$, we write $D_{\mathrm{p}}(\mathbf{u}, \mathbf{v})$ to mean $D_{\mathrm{p}}(\pi(\mathbf{u}), \pi(\mathbf{v}))$. Cassuto and Blaum [2] proved that $(\Sigma^n, D_{\mathrm{p}})$ is a metric space, and showed the following relationship between pair-distance and Hamming distance D_{H}.

Proposition 2.1 (Cassuto and Blaum [2]): 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 $D_{\mathrm{p}}(\mathbf{u}, \mathbf{v}) \geq d$ for all distinct $\mathbf{u}, \mathbf{v} \in \mathcal{C}$, 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. The maximum size of an $(n, d)_q$-symbol-pair code is denoted $A_q^{\mathrm{p}}(n, d)$. An $(n, d)_q$-symbol-pair code having $A_q^{\mathrm{p}}(n, d)$ codewords is said to be *optimal*. The size of an optimal symbol-pair code satisfies the following Singleton-type bound.

Theorem 2.1: (Singleton Bound) Let $q \geq 2$ and $2 \leq d \leq n$. Then $A_q^{\mathrm{p}}(n, d) \leq q^{n-d+2}$.

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

978-1-4673-2579-0/12/$31.00 ©2012 IEEE

length $n-d+2$ over an alphabet of size q is q^{n-d+2}. Hence, $A_q^{\mathrm{p}}(n,d) \leq q^{n-d+2}$. ∎

Hence, an $(n,d)_q$-symbol-pair code of size q^{n-d+2} is optimal by Theorem 2.1 and we call such a symbol-pair code *maximum distance separable* (MDS). In this paper, we construct new infinite classes of MDS symbol-pair codes.

III. Constructions of MDS Symbol-Pair Codes

In this section, we give several methods for deriving MDS symbol-pair codes from classical MDS codes. We also provide direct constructions for MDS symbol-pair codes. Note that $\mathcal{C} = \Sigma^n$ is an MDS $(n,2)_q$-symbol-pair code for all $n \geq 2$ and $q \geq 2$, 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. ∎

Existence of MDS $(n,d)_q$-codes with $d < n$ is provided below (see [4]). These MDS codes arise mainly from Reed-Solomon codes and their extensions.

Theorem 3.1:

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

The following corollary is an immediate consequence of both Theorem 3.1 and Proposition 3.1.

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 $n = 2^m$, $m \geq 1$ 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 $(n,3)_q$-symbol-pair codes.

Blanchard [5]–[7] (see also [8, chap. XI, §8]) proved the following asymptotic result.

Theorem 3.2 (Blanchard [5]–[7]): Let $2 \leq d \leq n$. Then there exists an MDS $(n,d)_q$-code for all q sufficiently large.

This implies that for $2 \leq d \leq n$, MDS $(n,d)_q$-symbol-pair codes exist for all q sufficiently large.

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

We use the interleaving method of Cassuto and Blaum [2] to obtain MDS symbol-pair codes. Cassuto and Blaum showed that a symbol-pair code with high pair-distance can be obtained by interleaving two classical codes of the same length and distance.

Theorem 3.3 (Cassuto and Blaum [2]): If there exist 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.

Applying Theorem 3.3 with classical MDS codes gives the following.

Corollary 3.2: If there exists an MDS $(n,d)_q$-code, then there exists an MDS $(2n,2d)_q$-symbol-pair code.

Hence, the following is an immediate consequence of Theorem 3.1 and Corollary 3.2.

Corollary 3.3:

(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 $n = 2^m$, $m \geq 1$ 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 $n \geq 2$, $q \geq 2$.

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 subsection, 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 [9] and assume readers' familiarity.

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 an eulerian tour $T = x_0e_1x_1e_2x_2\cdots e_mx_m$, where $x_m = x_0$, $x_i \in V(G)$, and $e_i \in E(G)$, for $1 \leq i \leq m$. 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 $D_{\mathrm{H}}(\mathbf{u}, \mathbf{v}) \geq d$, we have $|\{x \in V(G) : u_x = v_x\}| \leq n-d$. It follows that

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

since on the contrary there would exist at least $n-d$ edges $\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_{n-d}, y_{n-d}\}$ in $E(G)$ such that $u_{x_j} = v_{x_j}$ and $u_{y_j} = v_{y_j}$ for all $1 \leq j \leq n-d$. 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, $D_{\mathrm{p}}(\mathbf{u}, \mathbf{v}) \geq m-n+d+1$. 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. ∎

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 [10], [11] 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 a connected even graph of order n (of girth three), and those of girth four.

Proposition 3.3: Let $n \geq 3$ and $M = n\lfloor (n-1)/2 \rfloor$. Then there exists an eulerian graph of order n and size m, for $n \leq m \leq M$, 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 $n \geq 6$. Then there exists an eulerian graph of order n, size m, and girth at least four, for all $m \equiv n \bmod 2$, $n \leq m \leq M(n)$, except when $m = M(n) - 2$.

Idea of proofs for Proposition 3.3 & 3.4: An extremal eulerian graph G with girth at least three (or four) of size M (respectively, $M(n)$) can be constructed directly. Hamiltonian cycles and cycles of appropriate lengths are then removed from G to obtain eulerian graphs of the required sizes. It remains to ensure that the subgraph remains connected after the cycles are removed. Details are deferred to the full version of this paper. ∎

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

Proposition 3.5:

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

Idea of proof: For $n \in \{9, 10\}$, eulerian graphs with girth four and the required sizes can be constructed directly. Denote these graphs by $H_{n,m}$, where n is the order and m is the size.

For $n \geq 11$, let $n' = 2\lfloor n/2 \rfloor$. Consider the complete bipartite graph $G = K_{2\lfloor n'/4 \rfloor, 2\lceil n'/4 \rceil}$. Then G is a graph of order n', girth four and size $M(n')$ and in addition, G contains an induced subgraph $K_{4,4}$. Replacing the induced 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.

Constructions of $H_{9,14}$, $H_{9,16}$, $H_{10,15}$ and $H_{10,17}$ are given in the full version of this paper. ∎

Corollary 3.4: Let q be a prime power, $q \geq 4$. Then there exists an MDS $(n, n-1)_q$-symbol-pair code whenever

(i) $2 \leq n \leq (q^2-7)/2$ or $n = (q^2-1)/2$, for q odd;
(ii) $2 \leq n \leq (q-2)(q+3)/2$ or $n = q(q+1)/2$, for q even.

Proof: Follows from Corollary 3.1, Proposition 3.2, and Proposition 3.3. ∎

Define

$$N(q) = 2\left\lceil \frac{(q-1)^2}{8} \right\rceil.$$

Corollary 3.5: Let q be a prime power, $q \geq 5$. Then there exists an MDS $(n, n-2)_q$-symbol-pair code whenever

(i) $2 \leq n \leq N(q)$, or $N(q) \leq n \leq N(q+2)$ and n even, for q odd;
(ii) $2 \leq n \leq q^2/4 + 1$, $n \neq q^2/4 - 1$, for q even.

Proof: Follows from Corollary 3.1, Proposition 3.2, Proposition 3.4, and Proposition 3.5. ∎

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

D. $\mathbb{Z}_q$-linear MDS Symbol-Pair Codes with Pair-Distance Four, Five and n

We give direct constructions for MDS $(n, d)_q$-symbol-pair codes for $d \in \{4, 5, n\}$. We remark that for even n, MDS $(n, 4)_q$-symbol-pair codes have been constructed in Corollary 3.3, 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 subsection, we assume $\Sigma = \mathbb{Z}_q$. Besides being MDS, the codes constructed have *$\mathbb{Z}_q$-linearity*.

Definition 3.1: 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 3.2: The *pair-weight* of $\mathbf{u} \in \Sigma^n$ is

$$\mathrm{wt}_{\mathrm{p}}(\mathbf{u}) = D_{\mathrm{p}}(\mathbf{u}, \mathbf{0}).$$

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

Lemma 3.1: Let $\mathcal{C}$ be a $\mathbb{Z}_q$-linear code. Then $\mathcal{C}$ has pair-distance $\min_{\mathbf{u}\in\mathcal{C}\setminus\{\mathbf{0}\}} \mathrm{wt_p}(\mathbf{u})$.

Proposition 3.6: Let $n \geq 4$ and define f and g as follows:

$$\begin{aligned} f : \mathbb{Z}_q^{n-2} &\longrightarrow \mathbb{Z}_q \\ \mathbf{u} &\longmapsto \sum_{i=0}^{n-3} (i+1)u_i, \\ g : \mathbb{Z}_q^{n-2} &\longrightarrow \mathbb{Z}_q \\ \mathbf{u} &\longmapsto \sum_{i=0}^{n-3} u_i. \end{aligned}$$

Let $\mathcal{C} = \{(u_0, u_1, \ldots, u_{n-3}, f(\mathbf{u}), g(\mathbf{u})) : \mathbf{u} \in \mathbb{Z}_q^{n-2}\}$. Then $\mathcal{C}$ is a $\mathbb{Z}_q$-linear MDS $(n, 4)_q$-symbol-pair code.

Proof: It is readily verified that $\mathcal{C}$ is $\mathbb{Z}_q$-linear of size q^{n-2}. Hence, by Lemma 3.1, it suffices to show that for all $\mathbf{u} \in \mathbb{Z}_q^{n-2} \setminus \{\mathbf{0}\}$,

$$\mathrm{wt_p}((u_0, u_1, \ldots, u_{n-3}, f(\mathbf{u}), g(\mathbf{u}))) \geq 4.$$

Write $\tilde{\mathbf{u}} = (u_0, u_1, \ldots, u_{n-3}, f(\mathbf{u}), g(\mathbf{u}))$ and let

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

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 $f(\mathbf{u})$ or $g(\mathbf{u})$ is nonzero. Otherwise,

$$\begin{aligned} (j+1)u_j + (j+2)u_{j+1} &= 0, \\ u_j + u_{j+1} &= 0, \end{aligned}$$

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

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

■

Proposition 3.7: Suppose that q is odd prime and $5 \leq n \leq 2q+3$. Define f, g and h as follows:

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

Let $\mathcal{C} = \{(u_0, u_1, \ldots, u_{n-4}, f(\mathbf{u}), g(\mathbf{u}), h(\mathbf{u})) : \mathbf{u} \in \mathbb{Z}_q^{n-3}\}$. Then $\mathcal{C}$ is a $\mathbb{Z}_q$-linear MDS $(n, 5)_q$-symbol-pair code.

Proof: Similar to the proof for Proposition 3.6. ■

Proposition 3.8: Let $n \geq 2$ and let

$$\mathcal{C} = \begin{cases} \{(i, j, i, j, \ldots, i, j) : (i, j) \in \mathbb{Z}_q^2\}, & \text{if } n \text{ is even} \\ \{(i, j, i, j, \ldots, i, j, i+j) : (i, j) \in \mathbb{Z}_q^2\}, & \text{if } n \text{ is odd.} \end{cases}$$

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

Proof: It is readily verified that $\mathcal{C}$ is a $\mathbb{Z}_q$-linear code of size q^2. Hence, by Lemma 3.1, it is also easy to see that the pair-weight of all nonzero vectors in $\mathcal{C}$ is n. ■

IV. MDS Cyclic Symbol-Pair Codes from Mendelsohn Designs

A code $\mathcal{C} \subseteq \Sigma^n$ is *cyclic* if its automorphism group contains a cyclic group of order n. In other words, $\mathcal{C}$ contains a codeword $(u_0, u_1, \ldots, u_{n-1})$ if and only if it also contains $(u_1, u_2, \ldots, u_{n-1}, u_0)$ as a codeword. In this section, we present a construction for cyclic MDS symbol-pair codes. The constructed codes turned out to be also equidistant.

Let $\Sigma_*^n = \{\mathbf{u} \in \Sigma^n : u_0, u_1, \ldots, u_{n-1} \text{ are all distinct}\}$. A vector $(x_0, x_1, x_2, \cdots, x_{n-1}) \in \Sigma_*^n$ is said to *cyclically contain* the ordered pairs $(x_0, x_1), (x_1, x_2), \cdots, (x_{n-1}, x_0) \in \Sigma_*^2$, and no others.

Definition 4.1: A *Mendelsohn design* $\mathrm{M}(q, n)$ is a pair $(\Sigma, \mathcal{B})$ ($|\Sigma| = q$), where $\mathcal{B} \subseteq \Sigma_*^n$, such that each element of Σ_*^2 is cyclically contained in exactly one vector in $\mathcal{B}$. Elements of $\mathcal{B}$ are called *blocks*.

Mendelsohn designs were introduced by Mendelsohn [12] and constitutes a central topic in combinatorial design theory (see [13]). A necessary condition for $\mathrm{M}(q, n)$ to exist is that $n|q(q-1)$. This necessary condition is also asymptotically sufficient.

Theorem 4.1: (Mendelsohn [14], Bennett et al. *[15], Zhang [16]):* Let $q \geq n$. Then there exists an $\mathrm{M}(q, n)$ for all $q \equiv 0, 1 \bmod n$, provided q is sufficiently large.

Complete and near-complete solutions to the existence of $\mathrm{M}(q, n)$ have been obtained for $n \in \{3, 4, 5, 6, 7\}$ [12], [17]–[20], and the following result is known.

Theorem 4.2 (see [13]): There exists an $\mathrm{M}(p^2, p)$ for all odd primes p, and there exists an $\mathrm{M}(p^r, n)$ for all $r > 1$ and odd primes $p \equiv 1 \bmod n$.

We now establish the connection between Mendelsohn designs and cyclic symbol-pair codes.

Proposition 4.1: If there exists an $\mathrm{M}(q, n)$, then there exists a cyclic MDS $(n, n)_q$-symbol-pair code.

Proof: Let $(\Sigma, \mathcal{B})$ be an $\mathrm{M}(q, n)$. Simple counting shows that $|\mathcal{B}| = q(q-1)/n$. For each $\mathbf{u} \in \mathcal{B}$, let $\tau_i(\mathbf{u}) = (u_i, u_{i+1}, \ldots, u_{i+n-1})$. Now, let

$$\mathcal{C} = \left(\bigcup_{\mathbf{u} \in \mathcal{B}} \bigcup_{i=0}^{n-1} \tau_i(\mathbf{u}) \right) \cup \{(i, i, \ldots, i) \in \Sigma^n : i \in \mathbb{Z}_q\}.$$

We claim that $\mathcal{C}$ is a cyclic MDS $(n, n)_q$-symbol-pair code. It is easy to see that $\mathcal{C} \subseteq \Sigma^n$ has size q^2, and is cyclic. It remains to show that $\mathcal{C}$ has pair-distance n.

First, observe that

$$\begin{aligned} D_{\mathrm{H}}(\tau_i(\mathbf{u}), \tau_j(\mathbf{u})) &= n, \\ D_{\mathrm{H}}((i, i, \ldots, i), (j, j, \ldots, j)) &= n, \\ D_{\mathrm{H}}((i, i, \ldots, i), \tau_k(\mathbf{u})) &= n - 1, \end{aligned}$$

for $0 \le i < j \le n-1$, $0 \le k \le n-1$, and $\mathbf{u} \in \mathcal{B}$. By Proposition 2.1, we have

$$\begin{aligned} D_{\mathrm{p}}(\tau_i(\mathbf{u}), \tau_j(\mathbf{u})) &= n, \\ D_{\mathrm{p}}((i, i, \ldots, i), (j, j, \ldots, j)) &= n, \\ D_{\mathrm{p}}((i, i, \ldots, i), \tau_k(\mathbf{u})) &\ge n. \end{aligned} \tag{1}$$

It is, in fact, easy to see that equality always holds in inequality (1). Also, no pair of distinct blocks in $\mathcal{B}$ cyclically contain a common element of $\Sigma \times \Sigma$. Hence $D_{\mathrm{p}}(\tau_i(\mathbf{u}), \tau_j(\mathbf{v})) = n$ for all $0 \le i, j < n$ and distinct $\mathbf{u}, \mathbf{v} \in \mathcal{B}$. This shows that the code $\mathcal{C}$ has pair-distance n. ∎

The proof of Proposition 4.1 actually shows that a Mendelsohn design $\mathrm{M}(q, n)$ gives rise to a cyclic MDS $(n, n)_q$-symbol-pair code that is *equidistant*, one in which every pair of distinct codewords is at pair-distance exactly n. Applying Proposition 4.1 with Theorem 4.1 and Theorem 4.2 gives the following.

Theorem 4.3:

(i) There exists an equidistant cyclic MDS $(n, n)_q$-symbol-pair code for all $q \equiv 0, 1 \bmod n$, as long as q is sufficiently large.

(ii) There exists an equidistant cyclic MDS $(p, p)_{p^2}$-symbol-pair code for all odd primes p.

(iii) There exists an equidistant cyclic MDS $(n, n)_{p^r}$-symbol-pair code for all $r > 1$ and odd primes $p \equiv 1 \bmod n$.

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 $\Omega(q^2)$.

We also give constructions of equidistant cyclic MDS symbol-pair codes based on Mendelsohn designs.

Acknowledgment

Research of the authors is supported in part by the Singapore National Research Foundation under Research Grant NRF-CRP2-2007-03. C. Wang is also supported in part by NSFC under Grant 10801064. The authors thank the anonymous reviewers, whose comments improved the presentation of the paper.

References

[1] Y. Cassuto and M. Blaum, "Codes for symbol-pair read channels," in *ISIT 2010 – Proceedings of the 2010 IEEE International Symposium on Information Theory*. Austin, Texas: IEEE Press, June 2010, pp. 988–992.

[2] ——, "Codes for symbol-pair read channels," *IEEE Trans. Inform. Theory*, vol. 57, no. 12, pp. 8011–8020, 2011.

[3] Y. Cassuto and S. Litsyn, "Symbol-pair codes: algebraic constructions and asymptotic bounds," in *ISIT 2011 – Proceedings of the 2011 IEEE International Symposium on Information Theory*. St Petersburg, Russia: IEEE Press, July 2011, pp. 2348–2352.

[4] A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal Arrays*, ser. Springer Series in Statistics. New York: Springer-Verlag, 1999.

[5] J. L. Blanchard, "The existence of orthogonal arrays of strength three with large order," 1993, unpublished manuscript.

[6] ——, "The existence of orthgonal arrays of any strength with large order," 1994, unpublished manuscript.

[7] ——, "A representation of large integers from combinatorial sieves," *J. Number Theory*, vol. 54, no. 2, pp. 287–296, 1995.

[8] T. Beth, D. Jungnickel, and H. Lenz, *Design Theory*, 2nd ed. Cambridge University Press, 1999.

[9] J. A. Bondy and U. S. R. Murty, *Graph Theory*, ser. Graduate Texts in Mathematics. Springer, 2008.

[10] J. Novák, "Eulerovské grafy bez trojúhelníků s maximálním počtem hran," *Sborník vědeckých prací VŠST, Liberec*, 1971.

[11] ——, "Edge bases of complete uniform hypergraphs," *Mat. Časopis Sloven. Akad. Vied*, vol. 24, pp. 43–57, 1974.

[12] N. S. Mendelsohn, "A natural generalization of Steiner triple systems," in *Computers in number theory (Proc. Sci. Res. Council Atlas Sympos. No. 2, Oxford, 1969)*. London: Academic Press, 1971, pp. 323–338.

[13] E. Mendelsohn, "Mendelsohn designs," in *The CRC Handbook of Combinatorial Designs*, 2nd ed., C. J. Colbourn and J. H. Dinitz, Eds. Boca Raton, FL: CRC Press, 2007, pp. 528–534.

[14] N. S. Mendelsohn, "Perfect cyclic designs," *Discrete Math.*, vol. 20, no. 1, pp. 63–68, 1977/78.

[15] F. E. Bennett, E. Mendelsohn, and N. S. Mendelsohn, "Resolvable perfect cyclic designs," *J. Combin. Theory Ser. A*, vol. 29, no. 2, pp. 142–150, 1980.

[16] X. B. Zhang, "On the existence of $(v, 4, 1)$-RPMD," *Ars Combin.*, vol. 42, pp. 3–31, 1996.

[17] F. E. Bennett and X. Zhang, "Resolvable Mendelsohn designs with block size 4," *Aequationes Math.*, vol. 40, no. 2–3, pp. 248–260, 1990.

[18] Y. Miao and L. Zhu, "Perfect Mendelsohn designs with block size six," *Discrete Math.*, vol. 143, no. 1–3, pp. 189–207, 1995.

[19] R. J. R. Abel, F. E. Bennett, and H. Zhang, "Perfect Mendelsohn designs with block size six," *J. Statist. Plann. Inference*, vol. 86, no. 2, pp. 287–319, 2000.

[20] F. E. Bennett, "Recent progress on the existence of perfect Mendelsohn designs," *J. Statist. Plann. Inference*, vol. 94, no. 2, pp. 121–138, 2001.