Des. Codes Cryptogr. (2018) 86:841–859
https://doi.org/10.1007/s10623-017-0365-1

New constructions of MDS symbol-pair codes

Baokun Ding[1] · Gennian Ge[2] · Jun Zhang[3] · Tao Zhang[1] · Yiwei Zhang[3]

Received: 14 June 2016 / Revised: 19 April 2017 / Accepted: 26 April 2017 /
Published online: 8 May 2017

Abstract Motivated by the application of high-density data storage technologies, symbol-pair codes are proposed to protect against pair-errors in symbol-pair channels, whose outputs are overlapping pairs of symbols. The research of symbol-pair codes with the largest minimum pair-distance is interesting since such codes have the best possible error-correcting capability. A symbol-pair code attaining the maximal minimum pair-distance is called a maximum distance separable (MDS) symbol-pair code. In this paper, we focus on constructing linear MDS symbol-pair codes over the finite field $\mathbb{F}_q$. We show that a linear MDS symbol-pair code over $\mathbb{F}_q$ with pair-distance 5 exists if and only if the length n ranges from 5 to q^2+q+1. As for codes with pair-distance 6, length ranging from $q+2$ to q^2, we construct linear MDS symbol-pair codes by using a configuration called ovoid in projective geometry. With the help of elliptic curves, we present a construction of linear MDS symbol-pair codes for any pair-distance $d+2$ with length n satisfying $7 \leq d+2 \leq n \leq q+\lfloor 2\sqrt{q} \rfloor + \delta(q) - 3$, where $\delta(q) = 0$ or 1.

Keywords Symbol-pair read channels · MDS symbol-pair codes · Projective geometry · Elliptic curves

Mathematics Subject Classification 94B25 · 94B60

1 Introduction

With the development of high-density data storage technologies, while the codes are defined as usual over some discrete symbol alphabet, their reading from the channel is performed

Communicated by T. Etzion.

✉ Gennian Ge
gnge@zju.edu.cn

[1] School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China

[2] School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China

[3] School of Mathematical Sciences, Capital Normal University, Beijing 100048, China

as overlapping pairs of symbols. A channel whose outputs are overlapping pairs of symbols is called a symbol-pair channel. A pair-error is defined as a pair-read in which one or more of the symbols are read in error. The design of codes to protect efficiently against a certain number of pair-errors is significant.

Cassuto and Blaum first studied codes that protect against pair-errors in [2], as well as pair-error correctability conditions, code construction and decoding, and lower and upper bounds on code sizes. Later, Cassuto and Litsyn [3] gave algebraic cyclic code constructions of symbol-pair codes and asymptotic bounds on code rates. They also showed the existence of pair-error codes with rates strictly higher than those of the codes in the Hamming metric with the same relative distance. Yaakobi et al. proposed efficient decoding algorithms for cyclic symbol-pair codes in [14,15].

Chee et al. in [4] established a Singleton-type bound on symbol-pair codes and constructed infinite families of symbol-pair codes that meet the Singleton-type bound, which are called maximum distance separable symbol-pair codes or MDS symbol-pair codes for short. The construction of MDS symbol-pair codes is interesting since the codes have the best pair-error correcting capability for fixed length and dimension. The authors in [4] made use of interleaving and graph theoretic concepts as well as combinatorial configurations to construct MDS symbol-pair codes. Kai et al. [8] constructed MDS symbol-pair codes from cyclic and constacyclic codes.

Classical MDS codes are MDS symbol-pair codes [4] and other known families of MDS $(n, d)_q$ symbol-pair codes are shown in Table 1.

In this paper, we present new constructions of linear MDS symbol-pair codes over the finite field $\mathbb{F}_q$ and obtain the following three new families:

1. there exists a linear MDS $(n, 5)_q$ symbol-pair code if and only if $5 \le n \le q^2 + q + 1$;
2. there exists a linear MDS $(n, 6)_q$ symbol-pair code for $q \ge 3$ and $\max\{6, q+2\} \le n \le q^2$;
3. there exists a linear MDS $(n, d+2)_q$ symbol-pair code for general n, d satisfying $7 \le d + 2 \le n \le q + \lfloor 2\sqrt{q} \rfloor + \delta(q) - 3$, where

$$\delta(q) = \begin{cases} 0, \text{ if } q = p^a,\ a \ge 3,\ a \text{ is odd and } p \mid \lfloor 2\sqrt{q} \rfloor; \\ 1, \text{ otherwise.} \end{cases}$$

Table 1 Known families of MDS symbol-pair codes

d	q	n	Reference
2, 3	$q \ge 2$	$n \ge 2$	[4]
4	$q \ge 2$	$n \ge 2$	[4]
5	Even prime power	$n \le q + 2$	[4]
	Odd prime	$5 \le n \le 2q + 3$	[4]
	Prime power	$n \mid q^2 - 1, n > q + 1$	[8]
	Prime power	$n = q^2 + q + 1$	[8]
	Prime power, $q \equiv 1 \pmod 3$	$n = \frac{q^2+q+1}{3}$	[8]
6	Prime power	$n = q^2 + 1$	[8]
	Odd prime power	$n = \frac{q^2+1}{2}$	[8]
7	Odd prime	$n = 8$	[4]

Compared with the known MDS symbol-pair codes, the MDS symbol-pair codes constructed in this paper provide a much larger range of parameters.

This paper is organized as follows. Basic notations and definitions are given in Sect. 2. In Sect. 3, we construct MDS symbol-pair codes with pair-distance 5. And in Sect. 4 we derive MDS symbol-pair codes with pair-distance 6 from projective geometry. In Sect. 5, by using elliptic curves, we give the construction of MDS symbol-pair codes for any pair-distance satisfying certain conditions. Section 6 concludes the paper.

2 Preliminaries

Let Σ be the alphabet consisting of q elements. Each element in Σ is called a symbol. For a vector $\mathbf{u} = (u_0, u_1, \ldots, u_{n-1})$ in Σ^n, we define the symbol-pair read vector of $\mathbf{u}$ as

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

Throughout this paper, let q be a prime power and $\mathbb{F}_q$ be the finite field containing q elements. We will focus on vectors over $\mathbb{F}_q$, so $\Sigma = \mathbb{F}_q$. It is obvious that each vector $\mathbf{u}$ in $\mathbb{F}_q^n$ has a unique symbol-pair read vector $\pi(\mathbf{u})$ in $(\mathbb{F}_q \times \mathbb{F}_q)^n$. For two vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{F}_q^n$, the pair-distance between $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$d_p(\mathbf{u}, \mathbf{v}) := |\{0 \le i \le n-1 : (u_i, u_{i+1}) \ne (v_i, v_{i+1})\}|,$$

where the subscripts are reduced modulo n. For any vector $\mathbf{u}$ in $\mathbb{F}_q^n$, the pair-weight of $\mathbf{u}$ is defined as

$$w_p(\mathbf{u}) = |\{0 \le i \le n-1 : (u_i, u_{i+1}) \ne (0, 0)\}|,$$

where the subscripts are reduced modulo n.

The following relationship between the pair-distance and the Hamming distance was shown in [2].

Proposition 2.1 *Let $\mathbf{u}, \mathbf{v} \in \mathbb{F}_q^n$ be such that $0 < d_H(\mathbf{u}, \mathbf{v}) < n$, where d_H denotes the Hamming distance, we have*

$$d_H(\mathbf{u}, \mathbf{v}) + 1 \le d_p(\mathbf{u}, \mathbf{v}) \le 2d_H(\mathbf{u}, \mathbf{v}).$$

Meanwhile, the following relationship between the pair-distance and the pair-weight holds.

Proposition 2.2 *For all $\mathbf{u}, \mathbf{v} \in \mathbb{F}_q^n$, $d_p(\mathbf{u}, \mathbf{v}) = w_p(\mathbf{u} - \mathbf{v})$.*

A code $\mathcal{C}$ over $\mathbb{F}_q$ of length n is a nonempty subset of $\mathbb{F}_q^n$ and the elements of $\mathcal{C}$ are called codewords. The minimum pair-distance of $\mathcal{C}$ is defined as

$$d_p(\mathcal{C}) = \min\{d_p(\mathbf{u}, \mathbf{v}) : \mathbf{u}, \mathbf{v} \in \mathcal{C}, \mathbf{u} \ne \mathbf{v}\},$$

and the size of $\mathcal{C}$ is the number of codewords it contains. In general, a code $\mathcal{C}$ over $\mathbb{F}_q$ of length n, size M and minimum pair-distance d is called an $(n, M, d)_q$ symbol-pair code. Besides, if $\mathcal{C}$ is a subspace of $\mathbb{F}_q^n$, then $\mathcal{C}$ is called a linear symbol-pair code. When $\mathcal{C}$ is a linear code, the minimum pair-distance of $\mathcal{C}$ is the smallest pair-weight of nonzero codewords of $\mathcal{C}$. In this paper we consider linear symbol-pair codes over $\mathbb{F}_q$.

The minimum pair-distance d is an important parameter in determining the error-correcting capability of $\mathcal{C}$. Thus it is significant to find symbol-pair codes of fixed length n with pair-distance d as large as possible. In [4], the authors proved the following Singleton-type bound.

Theorem 2.3 (Singleton bound) *Let $q \geq 2$ and $2 \leq d \leq n$. If $\mathcal{C}$ is an $(n, M, d)_q$ symbol-pair code, then $M \leq q^{n-d+2}$.*

A symbol-pair code achieving the Singleton bound is a maximum distance separable (MDS) symbol-pair code. An MDS $(n, M, d)_q$ symbol-pair code is simply called an MDS $(n, d)_q$ symbol-pair code. In [8], the authors presented the following theorem.

Theorem 2.4 *Let $\mathcal{C}$ be an $[n, n - d_H, d_H]$ linear code over $\mathbb{F}_q$. If the pair-distance $d \geq d_H + 2$, then $\mathcal{C}$ is an MDS $(n, d_H + 2)_q$ symbol-pair code.*

Now we are ready to give a sufficient condition for the existence of linear MDS symbol-pair codes in the following theorem.

Theorem 2.5 *There exists a linear MDS $(n, d_H + 2)_q$ symbol-pair code $\mathcal{C}$ if there exists a matrix with d_H rows and $n \geq d_H + 2 \geq 4$ columns over $\mathbb{F}_q$, denoted by $H = [H_0, H_1, \ldots, H_{n-1}]$, where H_i $(0 \leq i \leq n - 1)$ is the i-th column of H, satisfying:*

1. *any $d_H - 1$ columns of H are linearly independent;*
2. *any d_H cyclically consecutive columns are linearly independent, i.e., $H_i, H_{i+1}, \ldots, H_{i+d_H-1}$ are linearly independent for $0 \leq i \leq n - 1$, where the subscripts are reduced modulo n.*

Proof Let $\mathcal{C}$ be the linear code with parity check matrix H. The first condition indicates that $\mathcal{C}$ is a linear code of length n, size q^{n-d_H} and minimum Hamming distance greater than or equal to d_H. If there exists a codeword $c \in \mathcal{C}$ with d_H nonzero coordinates, then the second condition ensures that the d_H nonzero coordinates are not in cyclically consecutive positions. Thus, from Propositions 2.1 and 2.2, we have $w_p(c) \geq d_H + 2$. For any other codeword $c' \in \mathcal{C}$ with Hamming weight $w_H(c') \geq d_H + 1$, it is easy to see that $w_p(c') \geq d_H + 2$. Hence the pair-distance $d \geq d_H + 2$ and $\mathcal{C}$ is an MDS $(n, d_H + 2)_q$ symbol-pair code. □

3 MDS symbol-pair codes with pair-distance 5

We first show a necessary condition for the existence of MDS $(n, 5)_q$ symbol-pair codes.

Lemma 3.1 *A linear MDS $(n, 5)_q$ symbol-pair code, where q is a prime power, exists only if the length n ranges from 5 to $q^2 + q + 1$.*

Proof The parity check matrix H of a linear MDS $(n, 5)_q$ symbol-pair code has three rows. From Proposition 2.1, we know that a symbol-pair code with the minimum pair-distance $d = 5$ must have the minimum Hamming distance $d_H \geq 3$. Therefore, any two columns in H must be linearly independent. In $\mathbb{F}_q$ the largest set of mutually linearly independent vectors of length three contains $q^2 + q + 1$ vectors. □

In this section we aim to show the existence of MDS $(n, 5)_q$ symbol-pair codes for every $5 \leq n \leq q^2 + q + 1$. According to Theorem 2.5, what we need is to construct a matrix H with 3 rows and n columns over $\mathbb{F}_q$ satisfying the following conditions:

1. any two columns of H are linearly independent;
2. any three cyclically consecutive columns are linearly independent.

We first describe how to construct a full matrix $H(q)$ of size $3 \times (q^2 + q + 1)$ and then we mention how to adjust $H(q)$ to get a matrix $H(q; n)$ of size $3 \times n$ for any n, $5 \leq n \leq$

q^2+q+1. Choose the column vectors of $H(q)$ from the following q^2+q+1 vectors: $\{(0,0,1)^{\mathrm{T}}, (1,a,b)^{\mathrm{T}}, (0,1,c)^{\mathrm{T}} : a,b,c \in \mathbb{F}_q\}$. We need to order these vectors in a proper way such that any three cyclically consecutive columns are linearly independent.

First we deal with the case when q is odd. Denote the elements in $\mathbb{F}_q$ in an arbitrary order $\{x_0, x_1, \ldots, x_{q-1}\}$. As a preparatory step, we partition the q^2 vectors of the form $\{(1,a,b)^{\mathrm{T}} : a,b \in \mathbb{F}_q\}$ into q disjoint blocks $B_i = \{(1,a,a^2+x_i)^{\mathrm{T}} : a \in \mathbb{F}_q\}$ for $0 \le i < q$. We give an order of the vectors within B_i as follows, where subscripts are reduced modulo q.

$$B_i = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_i & x_{i+1} & x_{i+2} & \cdots & x_{i+q-1} \\ x_i^2+x_i & x_{i+1}^2+x_i & x_{i+2}^2+x_i & \cdots & x_{i+q-1}^2+x_i \end{bmatrix}.$$

Then we construct the matrix $H(q)$ as follows. List all the blocks B_i defined above in the reverse order of their subscripts: $B_{q-1}, B_{q-2}, \ldots, B_1, B_0$. Between any pair of consecutive blocks B_{i+1} and B_i, insert a vector $(0,1,2x_i)^{\mathrm{T}}$. Note that the pair of B_0 and B_{q-1} is also considered, and the vector $(0,1,2x_{q-1})^{\mathrm{T}}$ should be inserted between them, which is further restricted to be the first column of $H(q)$. Finally the vector $(0,0,1)^{\mathrm{T}}$ could be placed anywhere and we just set it as the last column. That is,

$$H(q) = \begin{bmatrix} 0 & & 0 & & 0 & \ldots & 0 & \cdots & 0 & 0 \\ 1 & B_{q-1} & 1 & B_{q-2} & 1 & B_{q-3} \ldots B_{i+1} & 1 & B_i \cdots B_1 & 1 & B_0\ 0 \\ 2x_{q-1} & & 2x_{q-2} & & 2x_{q-3} & \ldots & 2x_i & \cdots & 2x_0 & 1 \end{bmatrix}.$$

Proposition 3.2 *When q is odd, every three cyclically consecutive columns of the matrix $H(q)$ constructed above are linearly independent over $\mathbb{F}_q$.*

Proof This can be easily checked by computing the determinants of any three cyclically consecutive columns. □

We now focus on the case when q is even and $q \neq 2, 4$. The general outline is similar. Let ω be a primitive element in $\mathbb{F}_q$. Denote the elements in $\mathbb{F}_q$ in an arbitrary order $\{x_0, x_1, \ldots, x_{q-1}\}$, with the only constraint that the first several elements are preset to be $x_0 = 0$, $x_1 = 1$, $x_2 = \omega$, $x_3 = \omega^2$, $x_4 = \omega+1$, $x_5 = \omega^2+\omega$. First define the blocks B_i in the same way as above and list all the blocks B_i in the reverse order of their subscripts: $B_{q-1}, B_{q-2}, \ldots, B_1, B_0$. Now we need to find out which vector of the form $(0,1,y)^{\mathrm{T}}$ can be inserted between the blocks B_{j+1} and B_j. We require that vectors $(0,1,y)^{\mathrm{T}}, (1,x_j,x_j^2+x_j)^{\mathrm{T}}, (1,x_{j+1},x_{j+1}^2+x_j)^{\mathrm{T}}$ are linearly independent and $(0,1,y)^{\mathrm{T}}, (1,x_j,x_j^2+x_{j+1})^{\mathrm{T}}, (1,x_{j-1},x_{j-1}^2+x_{j+1})^{\mathrm{T}}$ are linearly independent. It is easy to see y could be any value except for x_j+x_{j-1} and x_j+x_{j+1}.

Construct a bipartite graph. The left set of the vertices corresponds to $\mathbb{F}_q$. The right set of the vertices is $\{L_j : 0 \le j < q\}$, where the symbol L_j indicates the location between the blocks B_{j+1} and B_j. $y \in \mathbb{F}_q$ is connected to L_j if and only if the vector $(0,1,y)^{\mathrm{T}}$ could be inserted in the location L_j, i.e. $y \neq x_j+x_{j-1}$ and $y \neq x_j+x_{j+1}$. A perfect matching in this bipartite graph corresponds to a proper insertion scheme.

Following the analysis above, we can find that the degree of every vertex in the right part is exactly $q-2$. Recall that we have preset $x_0 = 0$, $x_1 = 1$, $x_2 = \omega$, $x_3 = \omega^2$, $x_4 = \omega+1$, $x_5 = \omega^2+\omega$. Thus we have:

- L_1 is connected to every $y \in \mathbb{F}_q$ except for 1 and $\omega+1$;
- L_2 is connected to every $y \in \mathbb{F}_q$ except for $\omega+1$ and $\omega^2+\omega$;
- L_3 is connected to every $y \in \mathbb{F}_q$ except for $\omega^2+\omega$ and $\omega^2+\omega+1$; and

- L_4 is connected to every $y \in \mathbb{F}_q$ except for $\omega^2+\omega+1$ and ω^2+1.
  So, even only among these four vertices, we can deduce that every $y \in \mathbb{F}_q$ is connected to at least two of them. So we have
- the neighbourhood of every $\Delta \le q-2$ vertices from the right part is of size at least $q-2 \ge \Delta$, since each vertex in the right part has degree $q-2$;
- the neighbourhood of every $q-1$ or q vertices from the right part is of size q.

Therefore the famous Hall's theorem [7] guarantees a perfect matching in this bipartite graph, which corresponds to a proper insertion scheme.

However, the case $q=4$ is listed as a separate case since the framework above using Hall's theorem would fail. To follow a similar framework, the order within a block needs some slight modifications and then a proper insertion scheme comes along. We shall just list the desired 3×21 matrix $H(4)$ instead of tedious explanations.

$$H(4)=\begin{bmatrix} 0&1&1&1&1&0&1&1&1&1&0&1&1&1&1&0&1&1&1&1&0\\ 1&0&1&\omega&\omega+1&1&\omega+1&\omega&1&0&1&0&\omega+1&\omega&1&1&1&\omega&\omega+1&0&0\\ 0&0&1&\omega+1&\omega&\omega+1&\omega+1&\omega&0&1&\omega&\omega&0&1&\omega+1&1&\omega&0&1&\omega+1&1 \end{bmatrix}.$$

Up till now we have constructed the matrix $H(q)$ of size $3\times(q^2+q+1)$ for every prime power $q\ge 3$. Next we discuss how to adjust $H(q)$ to get a $3\times n$ matrix $H(q;n)$ for every n, $5\le n\le q^2+q+1$. Denote $n=\alpha(q+1)+\beta$, where $0\le\beta\le q$. There are certainly lots of methods to get such a desired matrix and we propose one as follows.

- If $\beta\ne 2$, select the first $n-1$ columns of $H(q)$, then add the vector $(0,0,1)^{\mathrm{T}}$.
- If $\beta=2$, select the first $n-1$ columns of $H(q)$, then insert the vector $(0,0,1)^{\mathrm{T}}$ as the new third column.

The case $\beta=2$ is separated since if we still abide by the first rule then we will come across a triple of the form $\{(0,1,x)^{\mathrm{T}},(0,0,1)^{\mathrm{T}},(0,1,y)^{\mathrm{T}}\}$ which is certainly not independent.

The validity of the construction of the $3\times n$ matrix can be easily inferred from Proposition 3.2 plus some further simple checks on those triples containing the vector $(0,0,1)^{\mathrm{T}}$, and the two triples of the form $\{(0,1,a)^{\mathrm{T}},(0,1,b)^{\mathrm{T}},(1,c,d)^{\mathrm{T}}\}$ (in the $\beta=2$ case).

As illustrative examples, for $q=5$ we list the following matrices: the full matrix $H(5)$ of size 3×31, the adjusted matrix $H(5;13)$ (corresponding to $\beta\ne 2$) and $H(5;14)$ (corresponding to $\beta=2$).

$$H(5)=\begin{bmatrix} 0&1&1&1&1&1&0&1&1&1&1&1&0&1&1&1&1&1&0&1&1&1&1&1&0&1&1&1&1&1&0\\ 1&4&0&1&2&3&1&3&4&0&1&2&1&2&3&4&0&1&1&1&2&3&4&0&1&0&1&2&3&4&0\\ 3&0&4&0&3&3&1&2&4&3&4&2&4&1&1&3&2&3&2&2&0&0&2&1&0&0&1&4&4&1&1 \end{bmatrix},$$

$$H(5;13)=\begin{bmatrix} 0&1&1&1&1&1&0&1&1&1&1&1&0\\ 1&4&0&1&2&3&1&3&4&0&1&2&0\\ 3&0&4&0&3&3&1&2&4&3&4&2&1 \end{bmatrix},\ H(5;14)=\begin{bmatrix} 0&1&0&1&1&1&1&0&1&1&1&1&1&0\\ 1&4&0&0&1&2&3&1&3&4&0&1&2&1\\ 3&0&1&4&0&3&3&1&2&4&3&4&2&4 \end{bmatrix}.$$

Finally, for the case $q=2$, we list the matrices $H(2)$, $H(2;5)$, $H(2;6)$ as follows.

$$H(2)=\begin{bmatrix} 1&0&0&1&0&1&1\\ 0&1&0&1&1&1&0\\ 0&0&1&0&1&1&1 \end{bmatrix},\ H(2;5)=\begin{bmatrix} 1&0&0&1&1\\ 0&1&0&1&0\\ 0&0&1&1&1 \end{bmatrix},\ H(2;6)=\begin{bmatrix} 1&0&0&1&0&1\\ 0&1&0&1&1&0\\ 0&0&1&1&1&1 \end{bmatrix}.$$

So far we have finished the construction of MDS $(n,5)_q$ symbol-pair codes for any prime power $q\ge 2$ and $5\le n\le q^2+q+1$. The construction, together with Lemma 3.1, leads to the following theorem.

Theorem 3.3 *There exists a linear MDS* $(n, 5)_q$ *symbol-pair code, where* q *is a prime power, if and only if the length* n *ranges from* 5 *to* $q^2 + q + 1$.

4 MDS symbol-pair codes from projective geometry

Let $V(r+1, q)$ be a vector space of rank $r+1$ over $\mathbb{F}_q$. The projective space $PG(r, q)$ is the geometry whose points, lines, planes, ..., hyperplanes are the subspaces of $V(r+1, q)$ of rank $1, 2, 3, \ldots, r$, respectively. The dimension of a subspace of $PG(r, q)$ is one less than the rank of a subspace of $V(r+1, q)$. We label each point as $\langle(a_0, a_1, \ldots, a_r)\rangle$, the subspace spanned by a nonzero vector $(a_0, a_1, \ldots, a_r)$, where $a_i \in \mathbb{F}_q$ for $0 \leq i \leq r$. We refer to $a_0, a_1, \ldots, a_r$ as homogeneous coordinates for the point, since these coordinates are defined only up to multiplication by a nonzero scalar $\lambda \in \mathbb{F}_q$: here $\langle(\lambda a_0, \lambda a_1, \ldots, \lambda a_r)\rangle = \langle(a_0, a_1, \ldots, a_r)\rangle$. Thus, there are a total of $(q^{r+1} - 1)/(q - 1)$ points in $PG(r, q)$. For an integer $r \geq 2$, if we choose $n \geq r + 3$ points in $PG(r, q)$ and regard them as column vectors of a matrix H, then from Theorem 2.5 we have the following theorem.

Theorem 4.1 *There exists a linear MDS* $(n, r+3)_q$ *symbol-pair code if there exists a set* $\mathcal{S}$ *of* $n \geq r + 3 \geq 5$ *points of* $PG(r, q)$ *satisfying the following conditions:*

1. *any* r *points from* $\mathcal{S}$ *generate a hyperplane in* $PG(r, q)$*;*
2. *there exists a proper order* $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{n-1}$*, such that any* $r + 1$ *cyclically consecutive points do not lie on a hyperplane, i.e.,* $\mathcal{P}_i, \mathcal{P}_{i+1}, \ldots, \mathcal{P}_{i+r}$*, where the subscripts are reduced modulo* n*, do not lie on a hyperplane for* $0 \leq i \leq n - 1$.

Here we consider the case when $r = 3$ since there is a nice structure, namely the ovoid, in $PG(3, q)$. We first give the definition of the ovoid.

Definition 4.1 A set $\mathcal{O}$ of points of $PG(3, q)$ is called an ovoid if it satisfies the following conditions:

1. each line meets $\mathcal{O}$ in at most two points;
2. each point of $\mathcal{O}$ lies on exactly $q + 1$ tangent lines (a tangent line meets $\mathcal{O}$ in exactly one point), all of which lie on a plane.

The ovoid has been well studied and the following two lemmas can be found in [11].

Lemma 4.2 *Each ovoid has* $q^2 + 1$ *points.*

Lemma 4.3 *Each plane meets* $\mathcal{O}$ *either in one point or in* $q + 1$ *points.*

We can also easily derive the following lemma.

Lemma 4.4 *For an ovoid* $\mathcal{O}$ *in* $PG(3, q)$*, there exist* $q + 1$ *planes, each of which contains* $q + 1$ *points in* $\mathcal{O}$*. Moreover, these planes intersect in a common line in* $\mathcal{O}$ *and cover all the points of* $\mathcal{O}$.

Proof Fix two arbitrary points $A, B \in \mathcal{O}$, and then choose a point P from $\mathcal{O} \setminus \{A, B\}$. By Lemma 4.3, the plane formed by A, B, P, which we denote by ABP, must meet $\mathcal{O}$ in $q + 1$ points. Next, choose a point $Q \in \mathcal{O}$ which is not on ABP. Then, again, we get a plane ABQ which also meets $\mathcal{O}$ in $q + 1$ points. If we continue in this way, we can get $q + 1$ planes, each of which contains $q + 1$ points of $\mathcal{O}$. These planes intersect in a common line which meets $\mathcal{O}$ in the points A, B. □

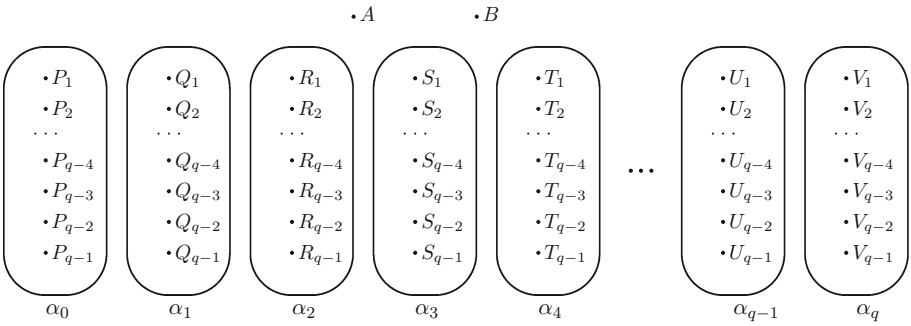

Fig. 1 The ovoid in PG(3,q)

Let $q \geq 5$ be a prime power. Suppose A and B are two points in the ovoid $\mathcal{O}$, and planes $\pi_0, \pi_1, \ldots, \pi_q$ intersect in the line AB and cover all the points of $\mathcal{O}$. Denote the set of points in $\pi_i \setminus \{A, B\}$ as α_i for $0 \leq i \leq q$, and there are $q-1$ points in each set. We illustrate the structure of the ovoid in Fig. 1.

Note that planes and hyperplanes are the same in $PG(3, q)$ and the points in the ovoid $\mathcal{O}$ satisfy the first condition in Theorem 4.1 inherently. Thus, in order to construct MDS symbol-pair codes we only need to order the points of $\mathcal{O}$ and make sure that any four cyclically consecutive points do not lie on a plane. In the rest of this section, we focus on the problem of ordering points in $\mathcal{O}$. Clearly, one can obtain the goal by many methods and we only propose one of them as follows.

We always choose A, B and arbitrary points $P_1 \in \alpha_0$, $Q_1 \in \alpha_1$ to be the first four points. It is obvious that the four points do not lie on a plane. Moreover, we always denote the last three points as X, Y, Z. For four ordered points P, Q, R, S, we say S is a *proper* point if S does not lie on the plane PQR. In other words, we say S is a *proper* point if S does not lie on the plane formed by the three points ordered right ahead of it. We first order the points in $\mathcal{O}$ and make sure that any four consecutive points do not lie on a plane. Then we ensure that X, Y, Z, A do not lie on a plane, nor do Y, Z, A, B and nor do Z, A, B, P_1.

We have the following observations that will be invoked multiple times in our proofs:

1. Two planes intersect in a line and a line meets $\mathcal{O}$ in at most two points. Therefore, two planes have at most two common points in $\mathcal{O}$.
2. Suppose we have ordered three points as P, Q, R, where $P \in \alpha_i$, $Q \in \alpha_j$, $R \in \alpha_k$ (at most two of i, j, k are equal), then the plane PQR intersects α_i in at most two points (one of which is P). If there are at least two points remaining in α_i at this moment, then we can always choose a *proper* point $P' \in \alpha_i$. The same conclusion also works for α_j and α_k.
3. We can always take *proper* points from two sets α_i and α_j or three sets α_i, α_j and α_k in turn until only one point remains in each set.
   This is an immediate conclusion from the last observation.
4. If exactly two of the three points X, Y, Z lie in the same set α_i, then these two points, together with A, form the plane π_i which does not contain the remaining point. Therefore, X, Y, Z, A do not lie on a plane.
5. If $Y \in \alpha_i$, $Z \in \alpha_j$, $i \neq j$, then Y, A, B form the plane π_i and Z, A, B form the plane π_j, and thus Y, Z, A, B do not lie on a plane.

6. If Z does not lie in the set α_0, then A, B, P_1 form a plane and A, B, Z form another, i.e., Z, A, B, P_1 do not lie on a plane.

Now we are ready to order n points in $\mathcal{O}$ such that any four cyclically consecutive points do not lie on a plane, and thus obtain MDS $(n, 6)_q$ symbol-pair codes. We restrict to the case when $q+2 \leq n \leq q^2$ since MDS $(n, 6)_q$ symbol-pair codes have already been constructed for $n \leq q+1$ [4] and $n = q^2+1$ [8]. We consider the following two cases.

4.1 The case when q is odd

We give different strategies for the three cases when $q+2 \leq n \leq 2q$, $2q < n \leq q^2-q$ and $q^2-q < n \leq q^2$.

Lemma 4.5 *Let $q \geq 5$ be an odd prime power. We can order n points such that any four cyclically consecutive points do not lie on a plane for $q+2 \leq n \leq 2q$.*

Proof After choosing A, B, P_1, Q_1, we choose a *proper* point $R_1 \in \alpha_2$ to be the fifth, and a *proper* point $S_1 \in \alpha_3$ to be the sixth. Take *proper* points from α_2 and α_3 in turn until we have ordered n $(q+2 \leq n \leq 2q)$ points.

By now we have ordered n points such that any four consecutive points do not lie on a plane. For the last three points X, Y, Z, we have X, Z lying in the same set α_i, $i = 2$ or 3, Y, Z lying in different sets and Z not lying in α_0. Thus, any four cyclically consecutive points do not lie on a plane. □

Lemma 4.6 *Let $q \geq 5$ be an odd prime power. We can order n points such that any four cyclically consecutive points do not lie on a plane for $2q < n \leq q^2-q$.*

Proof After choosing A, B, P_1, Q_1, we take *proper* points from α_0 and α_1 in turn until only one point remains in each set. Suppose that we have ordered the points as A, B, P_1, Q_1, P_2, Q_2, …, P_{q-2}, Q_{q-2}. Repeat the same process for α_2 and α_3, for α_4 and α_5, …. The number of such sets (i.e., the sets $\alpha_0, \alpha_1, \ldots, \alpha_q$) is $q+1$, an even number. Thus, we can always keep doing this until we have put n $(2q < n \leq q^2-q)$ points in order.

By now we have ordered n points such that any four consecutive points do not lie on a plane. For the last three points X, Y, Z, we have Y, Z lying in different sets and Z not lying in α_0. Therefore, we only need to make sure that X, Y, Z, A do not lie on a plane. The only special case is when the three points lie in three different sets $\alpha_i, \alpha_{i+1}, \alpha_{i+2}$ for some i. For example, suppose we have ordered the points as A, B, $P_1, \ldots, S_{q-3}$, R_{q-2}, S_{q-2}, T_1 and R_{q-2}, S_{q-2}, T_1 are the last three points. In this case, if R_{q-2}, S_{q-2}, T_1, A lie on a plane, then we find another point in α_4, which does not lie on the planes $R_{q-2}S_{q-3}S_{q-2}$ and $R_{q-2}S_{q-2}A$, to be the new last point. This can always succeed since the plane $R_{q-2}S_{q-3}S_{q-2}$ intersects α_4 in at most two points, the plane $R_{q-2}S_{q-2}A$ intersects α_4 in the point T_1 (this plane intersects π_4 in two points, one is the point A, another is the point T_1) and there are $q-1 \geq 4$ points in α_4. Therefore, we can order n points such that any four cyclically consecutive points do not lie on a plane. □

When $q^2-q < n \leq q^2$, we need to put more points in order. For a pair of sets α_i, α_{i+1}, $i = 0, 2, 4, \ldots, q-3$, we first take *proper* points from them alternatively until three points remain in each set. Then we give a strategy to order the remaining six points. We repeat this process until we have ordered all the points in $\alpha_0, \alpha_1, \ldots, \alpha_{q-2}$. After that, we take *proper* points from α_{q-1} and α_q in turn until we have ordered enough points. In the following lemma, we first restrict to the case when $q^2-q < n < q^2$ and then we discuss the case when $n = q^2$ separately.

Lemma 4.7 *Let $q \geq 5$ be an odd prime power. We can order n points such that any four cyclically consecutive points do not lie on a plane for $q^2 - q < n \leq q^2$.*

Proof After choosing A, B, P_1, Q_1, we take *proper* points from α_0 and α_1 in turn until three points remain in each set. Suppose we have ordered the points as $A, B, P_1, Q_1, P_2, Q_2, \ldots, P_{q-5}, Q_{q-5}, P_{q-4}, Q_{q-4}$. Then we choose a *proper* point $P_{q-3} \in \alpha_0$, which can always succeed (from the second observation and we will use this observation one more time in this proof), to be the next. After that, let the remaining two points $P_{q-2}, P_{q-1} \in \alpha_0$ and an arbitrary point $Q_{q-3} \in \alpha_1$ be the next three points. Take a *proper* point $Q_{q-2} \in \alpha_1$, and then let the remaining point $Q_{q-1} \in \alpha_1$ be the next. Let the order be $A, B, \ldots, P_{q-4}, Q_{q-4}, P_{q-3}, P_{q-2}, P_{q-1}, Q_{q-3}, Q_{q-2}, Q_{q-1}$. Points $P_{q-4}, Q_{q-4}, P_{q-3}, P_{q-2}$ do not lie on a plane, since $P_{q-4}, P_{q-3}, P_{q-2}$ form the plane π_0 while Q_{q-4} lies in π_1. The same reason works for points $Q_{q-4}, P_{q-3}, P_{q-2}, P_{q-1}$, points $P_{q-3}, P_{q-2}, P_{q-1}, Q_{q-3}$ and points $P_{q-1}, Q_{q-3}, Q_{q-2}, Q_{q-1}$.

So far we have ordered all the points in α_0 and α_1 such that any four consecutive points do not lie on a plane. Next, we take *proper* points from α_2 and α_3 in turn until three points remain in each set and then order the remaining points in α_2 and α_3 in the same way. Repeat the same process for α_4 and $\alpha_5, \ldots, \alpha_{q-3}$ and α_{q-2} until we have ordered all the points in them. After that, we take *proper* points from α_{q-1} and α_q in turn. We can always do this until we have put n ($q^2 - q < n \leq q^2 - 1$) points in order.

By now we have ordered n points such that any four consecutive points do not lie on a plane. Note that there are totally $q^2 - 2q + 1$ points in $\alpha_0, \alpha_1, \ldots, \alpha_{q-2}$, $n > q^2 - q$ and $q \geq 5$. Therefore, for the last three points X, Y, Z, we have X, Z lying in the same set α_i, $i = q - 1$ or q, Y, Z lying in different sets and Z not lying in α_0. Thus, any four cyclically consecutive points do not lie on a plane.

When $n = q^2 - 1$, suppose we have ordered the points as $A, B, P_1, Q_1, \ldots, U_{q-4}, V_{q-4}, U_{q-3}, V_{q-3}, U_{q-2}, V_{q-2}$. This indicates that $U_{q-4}, V_{q-4}, U_{q-3}, V_{q-3}$ do not lie on a plane, nor do $U_{q-3}, V_{q-3}, U_{q-2}, V_{q-2}$. For the case when $n = q^2$, we add one more point V_{q-1} and let the order be $U_{q-4}, V_{q-4}, U_{q-3}, V_{q-3}, V_{q-2}, U_{q-2}, V_{q-1}$. Clearly, points $V_{q-4}, U_{q-3}, V_{q-3}, V_{q-2}$ do not lie on a plane, nor do $V_{q-3}, V_{q-2}, U_{q-2}, V_{q-1}$. And it is easy to check that any four cyclically consecutive points do not lie on a plane by a similar discussion. □

We have ordered n points in $\mathcal{O}$ such that any four cyclically consecutive points do not lie on a plane for $q \geq 5$ and $q + 2 \leq n \leq q^2$. Therefore we obtain MDS $(n, 6)_q$ symbol-pair codes for all n, $q + 2 \leq n \leq q^2$. We exclude the case when $q = 3$ since there are not enough points on each plane π_i. We give MDS symbol-pair codes for $q = 3$ in the following example.

Example 4.8 There exists a linear MDS $(n, 6)_3$ symbol-pair code, $n \in \{6, 7, 8, 9, 10\}$, whose parity check matrix is formed by the first n columns of the matrix

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 1 & 2 & 2 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 & 2 & 0 & 2 & 2 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 2 & 1 & 0 & 2 & 0 \end{bmatrix}.$$

4.2 The case when q is even

The case when q is even is different from that when q is odd due to there being an odd number of planes $\pi_0, \pi_1, \ldots, \pi_q$. For $q + 2 \leq n < q^2 - q + 2$, we can order n points on q planes

$\pi_0, \pi_1, \ldots, \pi_{q-1}$ just as the case when q is odd. For $q^2 - q + 2 \le n \le q^2$, we first order all the points in the first three sets $\alpha_0, \alpha_1, \alpha_2$ by a similar method as in Lemma 4.7, then we can simply proceed as the case when q is odd since there are an even number of sets left.

Lemma 4.9 *Let $q \ge 8$ be an even prime power. We can order n points such that any four cyclically consecutive points do not lie on a plane for $q + 2 \le n \le q^2$.*

Proof When $q + 2 \le n < q^2 - q + 2$, we order n points on q planes $\pi_0, \pi_1, \ldots, \pi_{q-1}$ just as the case when q is odd. For the case when $q^2 - q + 2 \le n \le q^2$, the key step is to order all the points in $\alpha_0, \alpha_1, \alpha_2$.

Let a *proper* point $R_1 \in \alpha_2$ be the fifth and take *proper* points from $\alpha_0, \alpha_1, \alpha_2$ in turn until three points remain in each set. Suppose we have ordered the points as $A, B, P_1, Q_1, R_1, \ldots, P_{q-4}, Q_{q-4}, R_{q-4}$. Choose a *proper* point $P_{q-3} \in \alpha_0$ and a *proper* point $P_{q-2} \in \alpha_0$, which can always succeed (from the second observation, and we will use this multiple times in this proof), to be the next two points. After that, let the remaining point $P_{q-1} \in \alpha_0$ and an arbitrary point $Q_{q-3} \in \alpha_1$ be the next two points. Choose a *proper* point $Q_{q-2} \in \alpha_1$ to be the next, and then let the remaining point $Q_{q-1} \in \alpha_1$ and an arbitrary point $R_{q-3} \in \alpha_2$ be the next two points. Choose a *proper* point $R_{q-2} \in \alpha_2$ and then let the remaining point $R_{q-1} \in \alpha_2$ be the next. So far, We have ordered the points as $P_{q-4}, Q_{q-4}, R_{q-4}, P_{q-3}, P_{q-2}, P_{q-1}, Q_{q-3}, Q_{q-2}, Q_{q-1}, R_{q-3}, R_{q-2}, R_{q-1}$. Clearly, $R_{q-4}, P_{q-3}, P_{q-2}, P_{q-1}$ do not lie on a plane, nor do $P_{q-3}, P_{q-2}, P_{q-1}, Q_{q-3}$, nor do $P_{q-1}, Q_{q-3}, Q_{q-2}, Q_{q-1}$, nor do $Q_{q-3}, Q_{q-2}, Q_{q-1}, R_{q-3}$, and nor do $Q_{q-1}, R_{q-3}, R_{q-2}, R_{q-1}$.

By now we have ordered all the points in α_0, α_1 and α_2, and any four consecutive points do not lie on a plane. There are an even number of sets left. We then take *proper* points from α_3 and α_4 in turn and proceed as in Lemma 4.7. □

We give MDS symbol-pair codes for $q = 4$ in the following example.

Example 4.10 Denote the primitive element of $\mathbb{F}_4$ as w. Then there exists a linear MDS $(n, 6)_4$ symbol-pair code, $n \in \{6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17\}$, and its parity check matrix is formed by the first n columns of the matrix

$$\begin{bmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & w & 1+w & 1 & w & 1+w & w & w & 1+w & 1 & w & 1 & 1+w & 1+w & 1 \\
0 & 0 & 1 & 0 & w & 0 & 1+w & 0 & 1 & 1 & w & w & w+1 & w & 1+w & 1+w & 1 \\
0 & 0 & 0 & 1 & 0 & w & 0 & 1+w & 1 & w & 1 & w & w & 1+w & 1+w & 1 & 1+w
\end{bmatrix}.$$

There exists a linear MDS $(7, 6)$ symbol-pair code with parity check matrix

$$\begin{bmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & w & 1+w & 1 & w \\
0 & 0 & 1 & 0 & w & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & w & w
\end{bmatrix}.$$

Summing up the above, we can conclude the following theorem.

Theorem 4.11 *For any prime power q, $q \ge 3$, and any integer n, $\max\{6, q+2\} \le n \le q^2$, there exists a linear MDS $(n, 6)_q$ symbol-pair code.*

Remark 4.1 We can also obtain linear MDS $(n, 5)_q$ symbol-pair codes for $5 \le n \le q^2+q+1$ by ordering points in $PG(2, q)$ such that any three cyclically consecutive points do not lie on a line. Thus, this method deserves further investigation, which may derive MDS symbol-pair codes with larger pair-distance.

5 MDS symbol-pair codes from elliptic curves

The previous two sections construct MDS symbol-pair codes with pair-distance 5 and 6. In this section, we give a construction of MDS symbol-pair codes with general pair-distance ($\geq$7) from elliptic curve codes. We first briefly review some facts about elliptic curve codes.

Let $E/\mathbb{F}_q$ be an elliptic curve over $\mathbb{F}_q$ with function field $\mathbb{F}_q(E)$. Let $E(\mathbb{F}_q)$ be the set of all $\mathbb{F}_q$-rational points on E. Suppose $D = \{P_1, P_2, \ldots, P_n\}$ is a proper subset of rational points $E(\mathbb{F}_q)$, and G is a divisor of degree k ($0 < k < n$) with $\mathrm{Supp}(G) \cap D = \emptyset$. Without any confusion, we also write $D = P_1 + P_2 + \cdots + P_n$. Denote by $\mathscr{L}(G)$ the $\mathbb{F}_q$-vector space of all rational functions $f \in \mathbb{F}_q(E)$ with the principal divisor $\mathrm{div}(f) \geqslant -G$, together with the zero function (see [13]).

The functional AG code $C_{\mathscr{L}}(D, G)$ is defined to be the image of the following evaluation map:

$$ev : \mathscr{L}(G) \to \mathbb{F}_q^n;\ f \mapsto (f(P_1), f(P_2), \ldots, f(P_n)).$$

It is well-known that $C_{\mathscr{L}}(D, G)$ is a linear code with parameters $[n, k, d_H]$, where the minimum Hamming distance d_H has two choices:

$$d_H = n - k, \text{ or } d_H = n - k + 1.$$

A linear $[n, k, d_H]$ code is called an MDS code if $d_H = n - k + 1$ and is called an almost MDS code if $d_H = n - k$.

Suppose O is one of the $\mathbb{F}_q$-rational points on E. The set of rational points $E(\mathbb{F}_q)$ forms an abelian group with zero element O (for the definition of the sum of any two points, we refer to [12]), and it is isomorphic to the Picard group $\mathrm{div}^o(E)/\mathrm{Prin}(\mathbb{F}_q(E))$, where $\mathrm{Prin}(\mathbb{F}_q(E))$ is the subgroup consisting of all principal divisors. Denote by $\oplus$ and $\ominus$ the additive and minus operator in the group $E(\mathbb{F}_q)$, respectively.

To readers who are not familiar with the above abstract language, an elliptic curve E over $\mathbb{F}_q$ is defined by a non-singular Weierstrass equation

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \text{ for some } a_1, a_2, a_3, a_4, a_6 \in \mathbb{F}_q,$$

together with an extra point O at infinity. The set $E(\mathbb{F}_q)$ of $\mathbb{F}_q$-rational points on E is the union of the infinity point O and solutions (called finite points) of the Weierstrass equation over the finite field $\mathbb{F}_q$. That is,

$$E(\mathbb{F}_q) = \{(x, y) \in \mathbb{F}_q^2 : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\} \cup \{O\}.$$

It is easy to see that there are at most 3 intersection points of a line in the plane $\mathbb{F}_q^2$ with the cubic curve $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ over $\mathbb{F}_q$. Now, the above group structure on $E(\mathbb{F}_q)$ can be defined as follows.

- The infinity point O is the zero element. That is, for any $P \in E(\mathbb{F}_q)$, set $P \oplus O = P$.
- The opposite point $\ominus P$ of any finite point $P \in E(\mathbb{F}_q)$ is defined to be the finite point Q such that the line PQ intersects the elliptic curve E in only two points P and Q counting with multiplicity. If the two points coincide, then the line PQ is considered as the tangent line of E at the point P. Moreover, the opposite of O is itself.
- The sum $P \oplus Q$ for any finite points $P, Q \in E(\mathbb{F}_q)$ is defined to be the point $\ominus R \in E(\mathbb{F}_q)$ where R is the third intersection point of the line PQ with the elliptic curve E. If the two points P, Q coincide, then the line PQ is considered as the tangent line of E at the point P.

For simplicity but enough for our application, we take the divisor $G = mO$. Note that here mO is only a formal sum of m O's, but not the sum $\oplus$ defined above.

Proposition 5.1 ([5,16]) *Let E be an elliptic curve over $\mathbb{F}_q$ with an $\mathbb{F}_q$-rational point O, $D = \{P_1, P_2, \ldots, P_n\}$ a subset of $E(\mathbb{F}_q)$ such that $O \notin D$ and let $G = kO$ $(0 < k < n)$. Endow $E(\mathbb{F}_q)$ a group structure with the zero element O. Denote by*

$$N(k, O, D) = |\{S \subset D : |S| = k, \oplus_{P \in S} P = O\}|.$$

Then the AG code $C_{\mathscr{L}}(D, G)$ has the minimum Hamming distance $d_H = n - k + 1$ if and only if

$$N(k, O, D) = 0.$$

And the minimum Hamming distance $d_H = n - k$ if and only if

$$N(k, O, D) > 0.$$

Proof We have already seen that the minimum distance of $C_{\mathscr{L}}(D, G)$ has two choices: $n - k, n - k + 1$. So $C_{\mathscr{L}}(D, G)$ is not MDS, i.e., $d = n - k$ if and only if there is a function $f \in \mathscr{L}(G)$ such that the evaluation $ev(f)$ has weight $n - k$. This is equivalent to that f has k zeros in D, say $P_{i_1}, \ldots, P_{i_k}$. That is

$$\mathrm{div}(f) \geq -(k-1)O - P + (P_{i_1} + \cdots + P_{i_k}),$$

which is equivalent to

$$\mathrm{div}(f) = -(k-1)O - P + (P_{i_1} + \cdots + P_{i_k}).$$

The existence of such an f is equivalent to saying

$$P_{i_1} \oplus \cdots \oplus P_{i_k} = P.$$

Namely, $N(k, P, D) > 0$. It follows that the AG code $C_{\mathscr{L}}(D, G)$ has the minimum Hamming distance $n - k + 1$ if and only if $N(k, P, D) = 0$. □

We restrict to the case when $n > q + 1$, since for $n \leq q + 1$, MDS symbol-pair codes of length n can be constructed from Reed-Solomon codes. In this case, the minimum Hamming distance d_H of elliptic curve codes is related to the main conjecture of MDS codes which was affirmed for elliptic curve codes [9,10].

Proposition 5.2 ([9,10]) *Let $C_{\mathscr{L}}(D, G)$ be the elliptic curve code constructed in Proposition 5.1 with length $n > q + 1$. Then the subset sum problem always has solutions, i.e.,*

$$N(k, O, D) > 0.$$

And hence, elliptic curve codes with length $n > q + 1$ have deterministic minimum Hamming distance $d_H = n - k$.

That is, elliptic curve codes with length $n > q + 1$ are almost MDS codes. In order to make an almost MDS code have maximal minimum pair-distance, we need to separate the zeros in the minimal codewords. Thus, to construct MDS symbol-pair codes from elliptic curves, it is sufficient to make sure that there are no k cyclically consecutive zeros in the minimal codewords.

Lemma 5.3 *Let $C_{\mathscr{L}}(D, G)$ be the elliptic curve code constructed in Proposition 5.1 with length $n > q + 1$. If there are no k cyclically consecutive zeros in any codeword, then the code $C_{\mathscr{L}}(D, G)$ attains the maximal minimum pair-distance $n - k + 2$.*

To obtain long codes from elliptic curves, we need the following two well-known results of elliptic curves over finite fields.

Lemma 5.4 (Hasse-Weil Bound [12]) *Let E be an elliptic curve over $\mathbb{F}_q$. Then the number of $\mathbb{F}_q$-rational points on E is bounded by*

$$|E(\mathbb{F}_q)| \leq q + \lfloor 2\sqrt{q} \rfloor + 1.$$

Lemma 5.5 (Hasse-Deuring [6]) *The maximal number $N(\mathbb{F}_q)$ of $\mathbb{F}_q$-rational points on E, where E runs over all elliptic curves over $\mathbb{F}_q$, is*

$$N(\mathbb{F}_q) = \begin{cases} q + \lfloor 2\sqrt{q} \rfloor, & \text{if } q = p^a,\ a \geq 3,\ a \text{ is odd and } p | \lfloor 2\sqrt{q} \rfloor; \\ q + \lfloor 2\sqrt{q} \rfloor + 1, & \text{otherwise.} \end{cases}$$

Denote by

$$\delta(q) = \begin{cases} 0, \text{ if } q = p^a,\ a \geq 3, a \text{ is odd and } p \mid \lfloor 2\sqrt{q} \rfloor; \\ 1, \text{ otherwise.} \end{cases}$$

To construct an MDS symbol-pair code from classical error-correcting codes with large minimum Hamming distance, the key step is to find a way of ordering the coordinates. For general codes, this step seems very difficult. In the rest of this paper, we deal with the case of elliptic curve codes.

Theorem 5.6 *Let $N(\mathbb{F}_q) = q + \lfloor 2\sqrt{q} \rfloor + \delta(q)$. Then for any $7 \leq d + 2 \leq n \leq N(\mathbb{F}_q) - 3$, there exist linear MDS symbol-pair codes over $\mathbb{F}_q$ with parameters $(n, d + 2)_q$.*

Proof The existence of MDS symbol-pair codes with parameters $d+2 = n$ follows from [4]. Below we only consider the case when $7 \leq d + 2 < n \leq N(\mathbb{F}_q) - 3$. By Lemma 5.5, take E to be a maximal elliptic curve over $\mathbb{F}_q$ with an $\mathbb{F}_q$-rational point O, i.e.,

$$|E(\mathbb{F}_q)| = N(\mathbb{F}_q).$$

Take divisor $G = kO$ in the construction of elliptic curve codes.

Case (I): $N = N(\mathbb{F}_q)$ is odd, then there is no element of order 2 in $E(\mathbb{F}_q)$. Suppose

$$E(\mathbb{F}_q) = \{P_1, P_2, \ldots, P_{N-2}, P_{N-1}, O\},$$

where

$$P_1 \oplus P_2 = P_3 \oplus P_4 = \cdots = P_{N-2} \oplus P_{N-1} = O \tag{1}$$

1. For odd d and even n : $7 \leq d + 2 < n \leq N - 1$, in this case $k = N - 1 - d$ is odd. Take

$$D = \{P_1, P_2, \ldots, P_{N-2}, P_{N-1}\}.$$

Then it is easy to see from Eq. (1) that there are no k cyclically consecutive points whose sum is O. And hence, by Lemma 5.3 the elliptic curve code $C_{\mathscr{L}}(D, G)$ is an MDS symbol-pair code with parameters $(N - 1, d + 2)_q$. By deleting pairs (P_1, P_2), (P_3, P_4), etc., we can obtain MDS symbol-pair codes with parameters $(n, d + 2)_q$, where n runs over all even integers $7 \leq d + 2 < n \leq N - 1$.

2. For even d and odd n : $7 \leq d + 2 < n \leq N - 2$, in this case $k = N - 2 - d$ is odd. Take

$$D = \{P_1, P_2, \ldots, P_{N-2}\}.$$

Then it is easy to see from Eq. (1) that there are no k cyclically consecutive points whose sum is O. And hence, by Lemma 5.3 the elliptic curve code $C_{\mathscr{L}}(D, G)$ is an MDS symbol-pair code with parameters $(N-2, d+2)_q$. By deleting pairs (P_1, P_2), (P_3, P_4), etc., we can obtain MDS symbol-pair codes with parameters $(n, d+2)_q$ where n runs over all odd integers $d+2 < n \leq N-2$.

3. For even d and even n : $7 \leq d+2 < n \leq N-3$, in this case $k = N-3-d$ is even. Write $N-3 = (k+1)s + r$ for some integers $s \geq 1$ and $0 \leq r \leq k$. Take the pre-evaluation set

$$D_0 = \{P_1, P_2, \ldots, P_{N-5}, P_{N-4}, P_{N-2}\}$$

and arrange it by the following algorithm:
Step 1. For the set $\{P_1, P_2, \ldots, P_{sk+r-2}, P_{sk+r-1}\}$, we insert P_{N-i-4} between P_{ik-1} and P_{ik} for $1 \leq i \leq s-1$, insert P_{N-4} between P_{sk-1} and P_{sk}, and insert P_{N-2} behind P_{sk+r-1}. In other words, we arrange D_0 as follows

$$\begin{aligned} D_1 = \{&P_1, \ldots, P_{k-1}, P_{N-5}, P_k, \ldots, P_{(s-1)k-1}, P_{N-3-s}, P_{(s-1)k}, \\ &\cdots, P_{sk-1}, P_{N-4}, P_{sk}, P_{sk+1}, \ldots, P_{sk+r-1}, P_{N-2}\}. \end{aligned}$$

After this step, there are no k consecutive points whose sum is O in the sequence

$$\begin{aligned} &P_1, \ldots, P_{k-1}, P_{N-5}, P_k, \ldots, P_{(s-1)k-1}, P_{N-3-s}, P_{(s-1)k}, \ldots, P_{sk-1}, P_{N-4}, \\ &P_{sk}, P_{sk+1}, \ldots, P_{sk+r-1}. \end{aligned}$$

One can verify this exhaustively, for instance, $P_1 + \cdots + P_{k-1} + P_{N-5} = P_{k-1} + P_{N-5} \neq O$ since $P_{k-1} + P_k = O$ and $P_k \neq P_{N-5}$. And $P_k + P_{k+1} + \cdots + P_{2k-1} = P_k + P_{2k-1} \neq O$ since $P_{2k-1} + P_{2k} = O$ and $P_k \neq P_{2k}$.
But there may be some k cyclically consecutive points whose sum is O in the tail sequence

$$P_{(s-1)k+r+1}, \ldots, P_{sk-1}, P_{N-4}, P_{sk}, P_{sk+1}, \ldots, P_{sk+r-1}, P_{N-2}, P_1, \ldots, P_{k-r-1}.$$

For instance, $k = 6$, $N = 19$, by Step 1, we get

$$D_1 = P_1, \ldots, P_5, P_{14}, P_6, \ldots, P_{11}, P_{15}, P_{12}, P_{13}, P_{17}.$$

There are no 6 consecutive points whose sum is O in the sequence

$$P_1, \ldots, P_5, P_{14}, P_6, \ldots, P_{11}, P_{15}, P_{12}, P_{13}.$$

But there may be some 6 cyclically consecutive points whose sum is O in the tail sequence

$$P_{10}, P_{11}, P_{15}, P_{12}, P_{13}, P_{17}, P_1, P_2.$$

Step 2. In the case that r is even. It is easy to see that at most one of the following two equalities holds:

$$\begin{aligned} P_{(s-1)k+r+2} \oplus \cdots \oplus P_{N-4} \oplus \cdots \oplus P_{N-2} &= P_{(s-1)k+r+2} \oplus P_{N-4} \oplus P_{sk+r-1} \oplus P_{N-2} \\ &= O, \end{aligned}$$

and

$$P_{(s-1)k+r+3} \oplus \cdots \oplus P_{N-4} \oplus \cdots \oplus P_{N-2} \oplus P_1 = P_{N-4} \oplus P_{sk+r-1} \oplus P_{N-2} \oplus P_1 = O.$$

If the first one holds, then SWITCH $P_{(s-1)k+r+1}$ and $P_{(s-1)k+r+2}$; if the second one holds, then SWITCH P_1 and P_2; if neither of the two holds, then do nothing.

For any $i = 1, \ldots, \frac{k-r-2}{2}$, similarly at most one of the following two equalities holds:

$$\begin{aligned} &P_{(s-1)k+r+2i+2} \oplus \cdots \oplus P_{N-4} \oplus \cdots \oplus P_{N-2} \oplus P_1 \oplus \cdots \oplus P_{2i} \\ &= P_{(s-1)k+r+2i+2} \oplus P_{N-4} \oplus P_{sk+r-1} \oplus P_{N-2} = O, \end{aligned}$$

and

$$\begin{aligned} &P_{(s-1)k+r+2i+1} \oplus \cdots \oplus P_{N-4} \oplus \cdots \oplus P_{N-2} \oplus P_1 \oplus \cdots \oplus P_{2i-1} \\ &= P_{N-4} \oplus P_{sk+r-1} \oplus P_{N-2} \oplus P_{2i+1} = O. \end{aligned}$$

If the first one holds, then SWITCH $P_{(s-1)k+r+2i+1}$ and $P_{(s-1)k+r+2i+2}$; if the second one holds, then SWITCH P_{2i+1} and P_{2i+2}; if neither of the two holds, then do nothing. In the case that r is odd, the algorithm is the same as the even case, check the sum of k cyclically consecutive points and do the corresponding SWITCH operation.
Continue the above example, if

$$P_{10} \oplus P_{11} \oplus P_{15} \oplus P_{12} \oplus P_{13} \oplus P_{17} = P_{10} \oplus P_{15} \oplus P_{13} \oplus P_{17} = O,$$

then SWITCH P_9 and P_{10}; and in this case, it is immediate that

$$P_{11} \oplus P_{15} \oplus P_{12} \oplus P_{13} \oplus P_{17} \oplus P_1 = P_{15} \oplus P_{13} \oplus P_{17} \oplus P_1 \neq O,$$

so we do not need to reorder P_1 and P_2, and so on.
Using the above algorithm to rearrange the evaluation set to get a newly arranged evaluation set D, finally there are no k cyclically consecutive points whose sum is O. And hence, by Lemma 5.3 the elliptic curve code $C_{\mathscr{L}}(D, G)$ is an MDS symbol-pair code with parameters $(N-3, d+2)_q$. So, similarly as above, by deleting pairs from the pre-evaluation set, we can obtain MDS symbol-pair codes with parameters $(n, d+2)_q$ where n runs over all even integers $d+2 < n \leq N-3$.

4. For odd d and odd n : $7 \leq d+2 < n \leq N-2$, in this case $k = N-2-d$ is even. Write $N-2 = (k+1)s + r$ for some integers $s \geq 1$ and $0 \leq r \leq k$. Take the pre-evaluation set

$$D_0 = \{P_1, P_2, \ldots, P_{N-3}, P_{N-2}\}$$

and arrange it as follows

$$\begin{aligned} D = \{&P_1, \ldots, P_{k-1}, P_{N-3}, P_k, \ldots, P_{(s-1)k-1}, P_{N-1-s}, \\ &P_{(s-1)k}, \ldots, P_{sk-1}, P_{N-2}, P_{sk}, P_{sk+1}, \ldots, P_{sk+r}\}. \end{aligned}$$

If r is even. Moreover, if $r = k$, then replace P_{sk+r} by P_{N-1} in D, otherwise, keep the above D, then it is easy to see that there are no k cyclically consecutive points whose sum is O.
If r is odd, then similarly as the case when d and n are even, there may be some k cyclically consecutive points whose sum is O in the tail sequence. In this case, we just need process the same algorithm in the case 3 to obtain a rearranged evaluation set D such that there are no k cyclically consecutive points whose sum is O.
And hence, by Lemma 5.3 the elliptic curve code $C_{\mathscr{L}}(D, G)$ is an MDS symbol-pair code with parameters $(N-2, d+2)_q$. So, similarly as above, by deleting pairs from the pre-evaluation set, we can obtain MDS symbol-pair codes with parameters $(n, d+2)_q$ where n runs over all odd integers $7 \leq d+2 < n \leq N-2$.

In conclusion, in the case that $N = N(\mathbb{F}_q)$ is odd, for any $7 \leq d+2 \leq n \leq N(\mathbb{F}_q) - 3$, no matter whether d is odd or even, there exists an MDS symbol-pair code with parameters $(n, d+2)_q$.

Case (II): $N = N(\mathbb{F}_q)$ is even. The proof is the same. Note that there are one or three non-zero elements of order 2 in the group $E(\mathbb{F}_q)$. Using these elements in the setting of the pre-evaluation set, the remainder of the argument is analogous. We omit the details here.

So, by the discussion above, we complete the proof of the theorem. □

Remark 5.1 From the proof, we see that in some cases, the length of the MDS symbol-pair code constructed from elliptic curve can attain $N(\mathbb{F}_q) - 2$ or $N(\mathbb{F}_q) - 1$. We omit the detail of the statements in the theorem to get a clear description of our result. Also, note that there are other works devoted to constructing almost MDS codes using curves [1] besides elliptic curves. The advantage of construction upon elliptic curves is that we can translate the combinatorial problem on elliptic curve codes to that on a geometric object which becomes easier to deal with. To construct MDS symbol-pair codes using other almost MDS codes, how to arrange the evaluation set becomes the difficult step. We leave it as an open problem.

We finish this section by a toy example illustrating the above algorithm and the above remark.

Example 5.7 Let E be an elliptic curve over the finite field $\mathbb{F}_{13}$ defined by the equation

$$y^2 = x^3 + 9.$$

Using the software MAGMA or by direct computation, one can verify the elliptic curve E has $N = 21$ $\mathbb{F}_{13}$-rational points. They are $P_1 = (0, 3)$, $P_2 = (0, 10)$, $P_3 = (1, 6)$, $P_4 = (1, 7)$, $P_5 = (2, 2)$, $P_6 = (2, 11)$, $P_7 = (3, 6)$, $P_8 = (3, 7)$, $P_9 = (5, 2)$, $P_{10} = (5, 11)$, $P_{11} = (6, 2)$, $P_{12} = (6, 11)$, $P_{13} = (7, 1)$, $P_{14} = (7, 12)$, $P_{15} = (8, 1)$, $P_{16} = (8, 12)$, $P_{17} = (9, 6)$, $P_{18} = (9, 7)$, $P_{19} = (11, 1)$, $P_{20} = (11, 12)$ and the infinity point $P_{21} = O$. So it achieves the Hasse-Weil bound.

1. A construction of MDS $(20, 18)_{13}$ symbol-pair code.
   In this case, $n = 20, d = 16, k = 4$ are even, so it belongs to the case when d, n are even. The arranged evaluation set

   $$D_1 = \{P_1, P_2, P_3, P_{19}, P_4, P_5, P_6, P_7, P_{18}, P_8, P_9, P_{10}, P_{11}, P_{17}, P_{12}, P_{13}, P_{14}, P_{15}, P_{20}, P_{16}\}$$

   satisfies the property that any 4 cyclically consecutive points have nonzero sum. By Lemma 5.3 the elliptic curve code $C_{\mathscr{L}}(D_1, 4O)$ is an MDS symbol-pair code with parameters $(20, 18)_{13}$.
2. A construction of MDS $(19, 17)_{13}$ symbol-pair code.
   In this case, $n = 19, d = 15$ are odd, and $k = 4$ is even, so it belongs to the case when d, n are odd. The arranged evaluation set D is

   $$\{P_1, P_2, P_3, P_{18}, P_4, P_5, P_6, P_7, P_{17}, P_8, P_9, P_{10}, P_{11}, P_{19}, P_{12}, P_{13}, P_{14}, P_{15}, P_{16}\}.$$

   As $r = k$, in this case, we replace P_{16} by P_{20}. So we obtain the final evaluation set

   $$D' = \{P_1, P_2, P_3, P_{18}, P_4, P_5, P_6, P_7, P_{17}, P_8, P_9, P_{10}, P_{11}, P_{19}, P_{12}, P_{13}, P_{14}, P_{15}, P_{20}\}$$

   which satisfies the cyclically consecutive nonzero-sum property. By Lemma 5.3 the elliptic curve code $C_{\mathscr{L}}(D', 4O)$ is an MDS symbol-pair code with parameters $(19, 17)_{13}$.

6 Conclusion

In this paper, we first give a sufficient condition for the existence of linear MDS symbol-pair codes over $\mathbb{F}_q$. On this basis, we show that a linear MDS $(n, 5)_q$ symbol-pair code over $\mathbb{F}_q$ exists if and only if the length n ranges from 5 to $q^2 + q + 1$. Next, we introduce a special configuration in projective geometry called ovoid, which allows us to derive q-ary linear MDS symbol-pair codes with $d = 6$ and length ranging from $q + 2$ to q^2. This is an interesting method and deserves further investigation since it works well for both $d = 5$ and $d = 6$, and it may work for larger pair-distance. With the help of elliptic curves, we show that we can construct linear MDS $(n, d+2)_q$ symbol-pair codes for any n, d satisfying $7 \leq d + 2 \leq n \leq q + \lfloor 2\sqrt{q} \rfloor + \delta(q) - 3$. Compared with the results listed in Table 1, our results provide a much larger range of parameters.

Acknowledgements The authors express their gratitude to the two anonymous reviewers for their detailed and constructive comments which are very helpful to the improvement of this paper, and to Prof. Tuvi Etzion, the Associate Editor, for his insightful advice and excellent editorial job. The research of Gennian Ge is supported by the National Natural Science Foundation of China under Grant Nos. 11431003 and 61571310, Beijing Hundreds of Leading Talents Training Project of Science and Technology, and Beijing Municipal Natural Science Foundation. The research of Jun Zhang is supported by the National Natural Science Foundation of China under Grant No. 11601350, by Scientific Research Project of Beijing Municipal Education Commission under Grant No. KM201710028001, and by Beijing outstanding talent training program under Grant No. 2014000020124G140. Jun Zhang is supported by Chinese Scholarship Council during visiting the University of Oklahoma, USA.

References

1. Ballico E., Cossidente A.: Curves in projective spaces and almost MDS codes. Des. Codes Cryptogr. **24**(2), 233–237 (2001).
2. Cassuto Y., Blaum M.: Codes for symbol-pair read channels. IEEE Trans. Inf. Theory **57**(12), 8011–8020 (2011).
3. Cassuto, Y., Litsyn, S.: Symbol-pair codes: Algebraic constructions and asymptotic bounds. In: IEEE International Symposium on Information Theory, pp. 2348–2352 (2011).
4. Chee Y.M., Ji L., Kiah H.M., Wang C., Yin J.: Maximum distance separable codes for symbol-pair read channels. IEEE Trans. Inf. Theory **59**(11), 7259–7267 (2013).
5. Cheng Q.: Hard problems of algebraic geometry codes. IEEE Trans. Inf. Theory **54**(1), 402–406 (2008).
6. Deuring M.: Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. Abh. Math. Semin. Univ. Hamburg **14**(1), 197–272 (1941).
7. Hall P.: On representatives of subsets. J. Lond. Math. Soc. s1 **10**(1), 26–30 (1935).
8. Kai X., Zhu S., Li P.: A construction of new MDS symbol-pair codes. IEEE Trans. Inf. Theory **61**(11), 5828–5834 (2015).
9. Li, J., Wan, D., Zhang, J.: On the minimum distance of elliptic curve codes. In: IEEE International Symposium on Information Theory, pp. 2391–2395 (2015).
10. Munuera C.: On the main conjecture on geometric MDS codes. IEEE Trans. Inf. Theory **38**(5), 1573–1577 (1992).
11. Payne, S.: Topics in finite geometry: ovals, ovoids and generalized quadrangles. UC Denver Course Notes (2009).
12. Silverman, J.H.: The arithmetic of elliptic curves. In: Graduate Texts in Mathematics, 2nd edn., vol. 106. Springer, Dordrecht (2009).
13. Stichtenoth, H.: Algebraic function fields and codes. In: Graduate Texts in Mathematics, 2nd edn., vol. 254. Springer, Berlin (2009).
14. Yaakobi, E., Bruck, J., Siegel, P.H.: Decoding of cyclic codes over symbol-pair read channels. In: IEEE International Symposium on Information Theory, pp. 2891–2895 (2012).

15. Yaakobi E., Bruck J., Siegel P.H.: Constructions and decoding of cyclic codes over b-symbol read channels. IEEE Trans. Inf. Theory **62**(4), 1541–1551 (2016).
16. Zhang J., Fu F., Wan D.: Stopping sets of algebraic geometry codes. IEEE Trans. Inf. Theory **60**(3), 1488–1495 (2014).