

# New Constructions of MDS Symbol-Pair Codes

Baokun Ding<sup>a</sup>, Gennian Ge<sup>b,c,\*</sup>, Jun Zhang<sup>b</sup>, Tao Zhang<sup>a</sup> and Yiwei Zhang<sup>b</sup>

<sup>a</sup> School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China

<sup>b</sup> School of Mathematical Sciences, Capital Normal University, Beijing 100048, China

<sup>c</sup> Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing 100048, China.

## 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 6 to  $q^2 + 1$ , 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  $1$ .

*Key words:* Symbol-pair read channels, MDS symbol-pair codes, projective geometry, elliptic curves.

*Mathematics subject classifications:* 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 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.

---

\*Corresponding author (e-mail: gnge@zju.edu.cn). Research supported by the National Natural Science Foundation of China under Grant Nos. 11431003 and 61571310.

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.

Table 1: Known families of MDS symbol-pair codes

$d$	$q$	$n$	Reference
2, 3	$q \geq 2$	$n \geq 2$	[4]
4	$q \geq 2$	$n \geq 2$	[4]
5	even prime power	$n \leq q + 2$	[4]
	odd prime	$5 \leq n \leq 2q + 3$	[4]
	prime power	$n   q^2 - 1, n > q + 1$	[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]

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  $5 \leq n \leq q^2 + q + 1$ ;
2. there exists a linear MDS  $(n, 6)_q$  symbol-pair code for  $q \geq 3$  and  $6 \leq n \leq q^2 + 1$ ;

3. there exists a linear MDS  $(n, d+2)_q$  symbol-pair code for general  $n, d$  satisfying  $7 \leq d+2 \leq n \leq q + \lfloor 2\sqrt{q} \rfloor + \delta(q) - 3$ , where

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

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 Section 2. In Section 3, we construct MDS symbol-pair codes with pair-distance 5. And in Section 4 we derive MDS symbol-pair codes with pair-distance 6 from projective geometry. In Section 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  $\Sigma$  be the alphabet consisting of  $q$  elements. Each element in  $\Sigma$  is called a symbol. For a vector  $\mathbf{u} = (u_0, u_1, \dots, u_{n-1})$  in  $\Sigma^n$ , we define the symbol-pair read vector of  $\mathbf{u}$  as

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

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

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

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

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

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*

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

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

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

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  $\mathcal{C}$ . And 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  $H = [H_0, H_1, \dots, H_{n-1}]$ , 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. there exist  $d_H$  linearly dependent columns;
3. any  $d_H$  cyclically consecutive columns are linearly independent, i.e.,  $H_i, H_{i+1}, \dots, H_{i+d-1}$  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 two conditions indicate that  $\mathcal{C}$  is an  $[n, n - d_H, d_H]$  linear code with size  $q^{n-d_H}$ . Consider any codeword  $c \in \mathcal{C}$  with  $d_H$  nonzero coordinates. From Propositions 2.1, 2.2 and the third condition, we can see that the  $d_H$  nonzero coordinates are not in cyclically consecutive positions, and thus  $w_p(c) \geq d_H + 2$ . For any other codeword  $c' \in \mathcal{C}$ , we must have the Hamming weight  $w_H(c') \geq d_H + 1$  and  $w_p(c') \geq d_H + 2$ . Hence the pair-distance  $d \geq d_H + 2$  and  $\mathcal{C}$  is an MDS  $(n, d_H + 2)_q$  symbol-pair code.  $\square$

### 3 MDS symbol-pair codes with pair-distance 5

We construct MDS  $(n, 5)_q$  symbol-pair codes in this section. 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. there exist three linearly dependent columns;
3. any three cyclically consecutive columns are linearly independent.

**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.* From Proposition 2.1, we know that a symbol-pair code with the minimum pair-distance  $d = 5$  must have the minimum Hamming distance  $d_H \geq 3$ . Thus the parity check matrix of the code must satisfy the first condition above and the conclusion follows.  $\square$

In this section we aim to show the existence of MDS  $(n, 5)_q$  symbol-pair codes for every  $5 \leq n \leq q^2 + q + 1$ . 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:  $\{(0, 0, 1)^T, (0, 1, c)^T \text{ for } c \in \mathbb{F}_q, (1, a, b)^T \text{ for } a, b \in \mathbb{F}_q\}$ . In this way the first two conditions above are guaranteed, and we only need to order these vectors in a proper way to meet the third condition.

First we deal with the case when  $q$  is odd. Denote the elements in  $\mathbb{F}_q$  in an arbitrary order  $\{x_0, x_1, \dots, x_{q-1}\}$ . As a preparatory step, we partition the  $q^2$  vectors of the form  $\{(1, a, b)^T, a, b \in \mathbb{F}_q\}$  into  $q$  disjoint blocks  $B_i = \{(1, a, a^2 + x_i)^T, a \in \mathbb{F}_q\}$  for  $0 \leq i < q$ . We give an order of the vectors within  $B_i$ . Set the first vector to be  $(1, x_i, x_i^2 + x_i)^T$ , the next to be  $(1, x_{i+1}, x_{i+1}^2 + x_i)^T$ , and then the next to be  $(1, x_{i+2}, x_{i+2}^2 + x_i)^T \dots$  until finally the vector  $(1, x_{i+q-1}, x_{i+q-1}^2 + x_i)^T$ , where subscripts are reduced modulo  $q$ . That is,

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

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

That is,

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

**Proposition 3.2.** *Every three cyclically consecutive columns of  $H(q)$  are linearly independent over  $\mathbb{F}_q$ .*

*Proof.* For three consecutive columns within a block  $B_i$ ,  $0 \leq i \leq q-1$ , we have

$$\begin{vmatrix} 1 & 1 & 1 \\ x_{a-1} & x_a & x_{a+1} \\ x_{a-1}^2 + x_i & x_a^2 + x_i & x_{a+1}^2 + x_i \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ x_{a-1} & x_a & x_{a+1} \\ x_{a-1}^2 & x_a^2 & x_{a+1}^2 \end{vmatrix} = (x_{a-1} - x_a)(x_a - x_{a+1})(x_{a+1} - x_{a-1}) \neq 0.$$

For three consecutive columns with a vector  $(0, 1, 2x_j)^T$  in the middle, we have

$$\begin{vmatrix} 1 & 0 & 1 \\ x_j & 1 & x_j \\ x_j^2 + x_{j+1} & 2x_j & x_j^2 + x_j \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ x_j & 1 & 0 \\ x_j^2 + x_{j+1} & 2x_j & x_j - x_{j+1} \end{vmatrix} = x_j - x_{j+1} \neq 0.$$

For three consecutive columns containing a vector  $(0, 1, 2x_j)^T$ , which is not in the middle, we have either

$$\begin{vmatrix} 0 & 1 & 1 \\ 1 & x_j & x_{j+1} \\ 2x_j & x_j^2 + x_j & x_{j+1}^2 + x_j \end{vmatrix} = -(x_j - x_{j+1})^2 \neq 0, \quad (1)$$

or

$$\begin{vmatrix} 1 & 1 & 0 \\ x_{j-1} & x_j & 1 \\ x_{j-1}^2 + x_{j+1} & x_j^2 + x_{j+1} & 2x_j \end{vmatrix} = (x_j - x_{j-1})^2 \neq 0. \quad (2)$$

Finally, it is easy to see that every three consecutive columns in  $H(q)$  containing the vector  $(0, 0, 1)^T$  are linearly independent over  $\mathbb{F}_q$ .  $\square$

We now focus on the case when  $q$  is even and  $q \neq 2, 4$ . The general outline is similar. Let  $\omega$  be a primitive element in  $\mathbb{F}_q$ . Denote the elements in  $\mathbb{F}_q$  in an arbitrary order  $\{x_0, x_1, \dots, x_{q-1}\}$ , 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:  $B_{q-1}, B_{q-2}, \dots, B_1, B_0$ . Now we need to find out which vector of the form  $(0, 1, y)^T$  can be inserted between the blocks  $B_{j+1}$  and  $B_j$ . Recall the proof of Proposition 3.2. It can be checked that the choice of the value  $y$  only affects equations (1) and (2). So for the validity of that proof we only require that

$$\begin{vmatrix} 0 & 1 & 1 \\ 1 & x_j & x_{j+1} \\ y & x_j^2 + x_j & x_{j+1}^2 + x_j \end{vmatrix} = (x_{j+1} - x_j)(y - x_j - x_{j+1}) \neq 0,$$

and

$$\begin{vmatrix} 1 & 1 & 0 \\ x_{j-1} & x_j & 1 \\ x_{j-1}^2 + x_{j+1} & x_j^2 + x_{j+1} & y \end{vmatrix} = (x_j - x_{j-1})(y - x_j - x_{j-1}) \neq 0.$$

That is,  $y$  could be any value except for  $x_j + x_{j-1}$  and  $x_j + x_{j+1}$ . An explicit insertion scheme seems hard to be expressed in an easy form, however, we can show that a proper insertion scheme surely exists. Construct a bipartite graph. The first part of the vertices corresponds to  $\mathbb{F}_q$ . The second part of the vertices is the set  $\{L_j : 0 \leq j < q\}$ , 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  $(0, 1, y)^T$  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 second 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  $\omega^2 + 1$ .

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 no more than  $q - 2$  vertices from the second part is of size at least  $q - 2$ ;
- the neighbourhood of every  $q - 1$  or  $q$  vertices from the second 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 separated 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  $3 \times 21$  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  $q \geq 3$ . Next we discuss how to adjust  $H(q)$  to get a  $3 \times n$  matrix  $H(q; n)$  for every  $n$ ,  $5 \leq n \leq q^2 + q + 1$ . Denote  $n = \alpha(q + 1) + \beta$ , where  $0 \leq \beta \leq q$ . There are certainly lots of methods to get such a desired matrix and we offer one as follows.

- If  $\beta \neq 2$ , select the first  $n - 1$  columns of  $H(q)$ , then add the vector  $(0, 0, 1)^T$ .
- If  $\beta = 2$ , select the first  $n - 1$  columns of  $H(q)$ , then insert the vector  $(0, 0, 1)^T$  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  $\{(0, 1, x)^T, (0, 0, 1)^T, (0, 1, y)^T\}$  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  $(0, 0, 1)^T$ , and the two triples of the form  $\{(0, 1, a)^T, (0, 1, b)^T, (1, c, d)^T\}$  (in the  $\beta = 2$  case).

As illustrative examples, for  $q = 5$  we list the following matrices: the full matrix  $H(5)$  of size  $3 \times 31$ , the adjusted matrix  $H(5; 13)$  (corresponding to  $\beta \neq 2$ ) 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  $q \geq 2$  and  $5 \leq n \leq q^2 + q + 1$ . 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,  $\dots$ , hyperplanes are the subspaces of  $V(r+1, q)$  of rank  $1, 2, 3, \dots, r$ , respectively. The dimension of a subspace of  $PG(r, q)$  is one less than the rank of a subspace of  $V(r+1, q)$ .

Label each point of  $PG(r, q)$  as  $\langle (a_0, a_1, \dots, a_r) \rangle$ , the subspace spanned by a nonzero vector  $(a_0, a_1, \dots, a_r)$ , where  $a_i \in \mathbb{F}_q$  for  $0 \leq i \leq r$ . Since these coordinates are defined only up to multiplication by a nonzero scalar  $\lambda \in \mathbb{F}_q$  (here  $\langle (\lambda a_0, \lambda a_1, \dots, \lambda a_r) \rangle = \langle (a_0, a_1, \dots, a_r) \rangle$ ), we refer to  $a_0, a_1, \dots, a_r$  as homogeneous coordinates. 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 exist  $r+1$  points in  $\mathcal{S}$  lying on a hyperplane;*
3. *if the  $n$  points are ordered, say  $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_{n-1}$ , then any  $r+1$  cyclically consecutive points do not lie on a hyperplane, i.e.,  $\mathcal{P}_i, \mathcal{P}_{i+1}, \dots, \mathcal{P}_{i+r}$ , 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  $r = 3$ .

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

1. *each line meets  $\mathcal{O}$  in at most two points;*
2. *through each point of  $\mathcal{O}$  there are  $q+1$  lines, each of which meets  $\mathcal{O}$  in exactly one point, and all of them lie on a plane.*

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 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 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  $A, B$ .  $\square$

We can now state our construction.

**Theorem 4.5.** *Let  $q \geq 5$  be an odd prime power. Then there exist linear MDS  $(n, 6)_q$  symbol-pair codes for all  $n$ ,  $6 \leq n \leq q^2 + 1$ .*

*Proof.* Let  $\mathcal{O}$  be an ovoid in  $PG(3, q)$  and  $\pi_0, \pi_1, \dots, \pi_q$  be the planes described in Lemma 4.4. Moreover, let the intersection of  $\pi_0, \pi_1, \dots, \pi_q$  meet  $\mathcal{O}$  in the points  $A$  and  $B$ . For convenience, denote the plane formed by points  $P, Q, R$  by  $PQR$  and denote the set of the points lying in a set, say  $\Omega$ , but not on the plane  $PQR$  by  $\Omega \setminus PQR$ . 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 now consider the three conditions stated in Theorem 4.1. It is clear that, for the points of  $\mathcal{O}$ , the first condition is inherently satisfied and the second condition can be easily satisfied. Thus, the points of  $\mathcal{O}$  simply need to be ordered such that any four cyclically consecutive points do not lie on a plane. To attain this goal, we discuss it in two parts. First we order  $n$  ( $6 \leq n \leq q^2 + 1$ ) points of  $\mathcal{O}$  as  $\mathcal{P}_0, \dots, \mathcal{P}_{n-1}$  and make sure that any four consecutive points do not lie on a plane, i.e.,  $\mathcal{P}_i, \mathcal{P}_{i+1}, \mathcal{P}_{i+2}, \mathcal{P}_{i+3}$  do not lie on a plane for  $0 \leq i \leq n - 4$ . On this basis, we then adjust the order to make sure that any four cyclically consecutive points do not lie on a plane, i.e.,  $\mathcal{P}_i, \mathcal{P}_{i+1}, \mathcal{P}_{i+2}, \mathcal{P}_{i+3}$  do not lie on a plane for  $0 \leq i \leq n - 1$ .

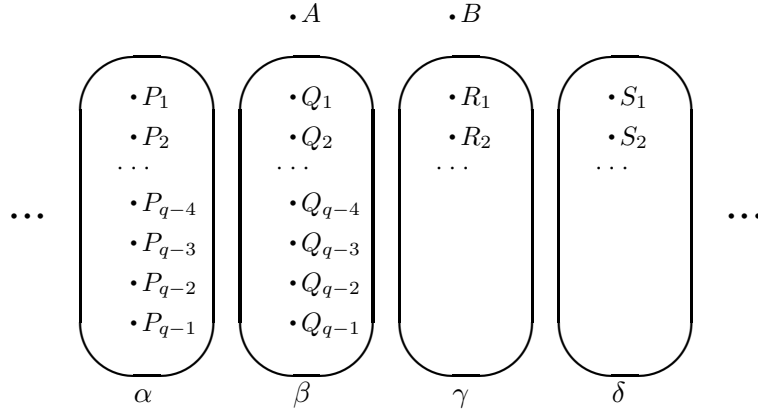


Figure 1: The sets  $\pi_i \setminus \{A, B\}$  when  $q$  is an odd prime power.

First, let  $\alpha, \beta, \gamma$  and  $\delta$  denote the sets  $\pi_0 \setminus \{A, B\}, \pi_1 \setminus \{A, B\}, \pi_2 \setminus \{A, B\}, \pi_3 \setminus \{A, B\}$  respectively, as illustrated in Figure 1. Let  $A, B$  be the first and the second points. Choose an arbitrary point

$P_1$  from  $\alpha$  to be the third and an arbitrary point  $Q_1$  from  $\beta$  to be the fourth. It is obvious that  $A, B, P_1, Q_1$  do not lie on a plane. Next, choose  $P_2 \in \alpha \setminus BP_1Q_1$  to be the fifth and  $Q_2 \in \beta \setminus P_1P_2Q_1$  to be the sixth. Two planes intersect in a line and a line meets  $\mathcal{O}$  in at most two points. Thus, we can continue in this way, i.e., take *proper* points from  $\alpha$  and  $\beta$  in turn, until only one point remains in  $\alpha$ .

Now suppose this has been done so that the point  $P_{q-1}$  remains, i.e., we have ordered the points as  $A, B, P_1, Q_1, \dots, P_{q-2}, Q_{q-2}$ . Then we have that  $P_{q-4}, Q_{q-4}, P_{q-3}, Q_{q-3}$  do not lie on a plane, nor do  $Q_{q-4}, P_{q-3}, Q_{q-3}, P_{q-2}$ , and nor do  $P_{q-3}, Q_{q-3}, P_{q-2}, Q_{q-2}$ . Next we order the two points  $P_{q-1}$  and  $Q_{q-1}$ . We consider the following three cases:

**Case 1:**  $P_{q-1} \notin P_{q-2}Q_{q-3}Q_{q-2}$ ,  $Q_{q-1} \notin P_{q-2}P_{q-1}Q_{q-2}$ .

Note that this situation is ideal. Let the order be  $P_{q-4}, Q_{q-4}, P_{q-3}, Q_{q-3}, P_{q-2}, Q_{q-2}, P_{q-1}, Q_{q-1}$ .

**Case 2:**  $P_{q-1} \notin P_{q-2}Q_{q-3}Q_{q-2}$ , but  $Q_{q-1} \in P_{q-2}P_{q-1}Q_{q-2}$ .

Change the order to be  $P_{q-4}, Q_{q-4}, P_{q-3}, P_{q-2}, Q_{q-3}, Q_{q-2}, P_{q-1}, Q_{q-1}$ .

**Case 3:**  $P_{q-1} \in P_{q-2}Q_{q-3}Q_{q-2}$ .

Change the order to be  $P_{q-4}, Q_{q-4}, P_{q-3}, Q_{q-3}, P_{q-2}, Q_{q-2}, Q_{q-1}, P_{q-1}$ .

Next, we find a *proper* point  $R_1 \in \gamma$  to be the next point, as well as *proper* points  $S_1 \in \delta$  and  $R_2 \in \gamma$ . Then order the remaining points in  $\gamma$  and  $\delta$  just as what we have done for the points in  $\alpha$  and  $\beta$ . Repeat the procedure until we have covered  $n$  ( $6 \leq n \leq q^2 + 1$ ) points in  $\mathcal{O}$ . By now, we have got  $n$  ordered points such that any four consecutive points do not lie on a plane.

Note that we have finished our first part. Denote the last four points by  $W, X, Y$  and  $Z$ . To make sure that any four cyclically consecutive points do not lie on a plane, we still need to 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 discuss in the following cases.

**Case a:**  $X, Y, Z$  and  $A$  lie on a plane.

This happens only when  $X \in \pi_i$ ,  $Y \in \pi_{i+1}$  and  $Z \in \pi_{i+2}$ , for some  $i$ ,  $0 \leq i \leq q-2$ . For example,  $P_{q-1}, Q_{q-1}$  and  $R_1$  in Figure 1. Otherwise, we always have exactly two of  $X, Y, Z$  belonging to the same set  $\pi_j \setminus \{A, B\}$ , which ensures  $X, Y, Z, A$  do not lie on a plane.

Note that  $WXY$  intersects  $\pi_{i+2}$  in at most two points and also  $XYA$  intersects  $\pi_{i+2}$  in at most two points, one of which is the point  $A$ . Thus, in this case, we find a new point  $Z'$  in  $\pi_{i+2} \setminus \{A, B\}$ , not lying on planes  $WXY$  and  $XYA$  to be the new last point. We can always do this since there are totally  $q+1 \geq 6$  points on  $\pi_{i+2}$ .

**Case b:**  $Y, Z, A$  and  $B$  lie on a plane.

This happens when the last two points lie in the same  $\pi_i \setminus \{A, B\}$ , which occurs in Cases 2 and 3 above. Note that  $\alpha$  and  $\beta$  can be any  $\pi_i \setminus \{A, B\}$  and  $\pi_{i+1} \setminus \{A, B\}$  respectively for  $i = 0, 2, 4, \dots, q-1$  in the following discussion. In Case 2, if the last three points are  $Q_{q-4}, P_{q-3}$  and  $P_{q-2}$ , then we replace them by  $Q_{q-4}, P_{q-3}$  and  $Q_{q-3}$ . If the last three points are  $P_{q-2}, Q_{q-3}$  and  $Q_{q-2}$ , then we replace them by  $Q_{q-3}, P_{q-2}$  and  $Q_{q-2}$ . In Case 3, if the last three points are  $P_{q-2}, Q_{q-2}$  and  $Q_{q-1}$ , then we replace them by  $P_{q-2}, P_{q-1}$  and  $Q_{q-1}$ .

**Case c:**  $Z, A, B$  and  $P_1$  lie on a plane.

This happens when  $Z$  lies in  $\pi_0 \setminus \{A, B\}$ , i.e.,  $7 \leq n \leq 2q-1$  and  $n$  is odd. In this case, after choosing the first three points  $A, B, P_1$ , we choose *proper* points from  $\pi_2 \setminus \{A, B\}$  and  $\pi_3 \setminus \{A, B\}$  in turn.  $\square$

**Remark 4.1.** We use the condition that points  $P_{q-4}, P_{q-3}$  and  $P_{q-2}$  are on the same plane  $\pi_i$ ,  $0 \leq i \leq q$ , in Theorem 4.5. Thus we exclude the case when  $q = 3$  since there are not enough points on each plane  $\pi_i$ . We give the MDS symbol-pair codes directly for  $q = 3$ . 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

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

**Theorem 4.6.** Let  $q \geq 8$  be an even prime power. Then there exist linear MDS  $(n, 6)_q$  symbol-pair codes for all  $n$ ,  $6 \leq n \leq q^2 + 1$ .

*Proof.* Let the notations be defined as in Theorem 4.5. Note that the case when  $q$  is even is different from that when  $q$  is odd due to there being an odd number of planes. For  $6 \leq n \leq q^2 - q + 2$ , we can order  $n$  points on  $\pi_0, \pi_1, \dots, \pi_{q-1}$  just as in Theorem 4.5 since the number of planes is even. The key step of this proof is to put the remaining  $q - 1$  points in order. To attain this goal, we first order all the points of the first three planes, and then we can just proceed as the case when  $q$  is odd.

Let  $\alpha, \beta, \gamma, \delta, \zeta$  denote the sets  $\pi_0 \setminus \{A, B\}, \pi_1 \setminus \{A, B\}, \pi_2 \setminus \{A, B\}, \pi_3 \setminus \{A, B\}, \pi_4 \setminus \{A, B\}$  respectively, as illustrated in Figure 2. Again, let  $A$  and  $B$  be the first two points and choose arbitrary  $P_1$  and  $Q_1$  from  $\alpha$  and  $\beta$  respectively. Choose the next point  $R_1 \in \gamma \setminus BP_1Q_1$ , and then  $P_2 \in \alpha \setminus P_1Q_1R_1$  and  $Q_2 \in \beta \setminus P_2Q_1R_1$ , i.e., take *proper* points from  $\alpha, \beta$  and  $\gamma$  in turn. We can continue in this way until only one point remains in  $\alpha$ .

Suppose this has been done so that  $P_{q-1}$  remains, i.e., we have ordered the points as  $A, B, P_1, Q_1, R_1, \dots, P_{q-2}, Q_{q-2}, R_{q-2}$ . Note that the intersection of two planes meets  $\mathcal{O}$  in at most two points. We can always find a point  $S_1$  in  $\delta$  that does not lie on the planes  $P_{q-2}Q_{q-2}R_{q-2}$  and  $P_{q-1}Q_{q-2}R_{q-2}$  since the two planes intersect  $\delta$  in at most four points and there are  $q - 1 \geq 7$  points in  $\delta$ . Similarly, we can find  $T_1 \in \zeta$  not lying on planes  $P_{q-1}R_{q-2}S_1$  and  $P_{q-1}Q_{q-1}S_1$ ,  $S_2 \in \delta$  not lying on planes  $P_{q-1}Q_{q-1}T_1$  and  $Q_{q-1}R_{q-1}T_1$ . Next find  $T_2 \in \zeta \setminus Q_{q-1}R_{q-1}S_2$ ,  $S_3 \in \delta \setminus R_{q-1}S_2T_2$  and  $T_3 \in \zeta \setminus R_{q-1}S_3T_2$ . Let the order of points be  $P_{q-2}, Q_{q-2}, R_{q-2}, S_1, P_{q-1}, T_1, Q_{q-1}, S_2, R_{q-1}, T_2, S_3, T_3$ . Note that we have ordered all the points in  $\alpha, \beta$  and  $\gamma$ , and any four consecutive points do not lie on a plane. There are an even number of planes left. We can then simply proceed as in Theorem 4.5, and also the similar discussion follows that of Theorem 4.5.  $\square$

**Remark 4.2.** When  $q = 4$ , since there are only five points on each plane  $\pi_i$ ,  $0 \leq i \leq q$ , we discuss it as a special case. Denote the primitive element of  $\mathbb{F}_4$  as  $w$ . Then there is 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

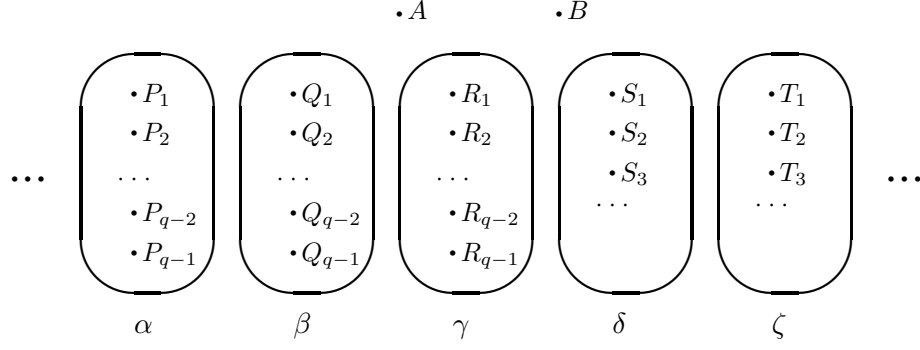


Figure 2: The sets  $\pi_i \setminus \{A, B\}$  when  $q$  is an even prime power.

first  $n$  columns of the matrix

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

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 4.7.** *For any prime power  $q$ ,  $q \geq 3$ , and any integer  $n$ ,  $6 \leq n \leq q^2 + 1$ , there exists a linear MDS  $(n, 6)_q$  symbol-pair code.*

**Remark 4.3.** *Compare the cases  $r = 2$  and  $r = 3$  in Theorem 4.1, if we consider the set of all the points instead of the ovoid, all the lines through a fixed point instead of the planes described in Lemma 4.4, then we can also get linear MDS  $(n, 5)_q$  symbol-pair codes for  $5 \leq n \leq q^2 + q + 1$  with  $q$  being a prime power in a similar way. Thus, this method deserves further investigation for larger  $r$ , 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  $D = \{P_1, P_2, \dots, P_n\}$  is a proper subset of rational points  $E(\mathbb{F}_q)$ , and  $G$  is a divisor of degree  $k$  ( $2g - 2 < k < n$ ) with  $\text{Supp}(G) \cap D = \emptyset$ . Without any confusion, we also write  $D = P_1 + P_2 + \dots + P_n$ . Denote by  $\mathcal{L}(G)$  the  $\mathbb{F}_q$ -vector space of all rational functions  $f \in \mathbb{F}_q(E)$  with the principal divisor  $\text{div}(f) \geq -G$ , together with the zero function ([13]).

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

$$ev : \mathcal{L}(G) \rightarrow \mathbb{F}_q^n; f \mapsto (f(P_1), f(P_2), \dots, f(P_n)).$$

It is well-known that  $C_{\mathcal{L}}(D, G)$  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  $\text{div}^o(E)/\text{Prin}(\mathbb{F}_q(E))$ , where  $\text{Prin}(\mathbb{F}_q(E))$  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.

**Proposition 5.1** ([5, 16]). *Let  $E$  be an elliptic curve over  $\mathbb{F}_q$  with an  $\mathbb{F}_q$ -rational point  $O$ ,  $D = \{P_1, P_2, \dots, P_n\}$  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  $C_{\mathcal{L}}(D, G)$  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  $C_{\mathcal{L}}(D, G)$  has two choices:  $n - k, n - k + 1$ . So  $C_{\mathcal{L}}(D, G)$  is not MDS, i.e.,  $d = n - k$  if and only if there is a function  $f \in \mathcal{L}(G)$  such that the evaluation  $ev(f)$  has weight  $n - k$ . This is equivalent to that  $f$  has  $k$  zeros in  $D$ , say  $P_{i_1}, \dots, P_{i_k}$ . That is

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

which is equivalent to

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

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  $C_{\mathcal{L}}(D, G)$  has the minimum Hamming distance  $n - k + 1$  if and only if  $N(k, P, D) = 0$ .  $\square$

We restrict to the case  $n > q + 1$ , since for every length  $n \leq q + 1$ , 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  $C_{\mathcal{L}}(D, G)$  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. To obtain long codes from elliptic curves, we need the following two well-known results of elliptic curves over finite fields.

**Lemma 5.3** (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 5.4** (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*

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

Denote by

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

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 5.5.** *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  $7 \leq d+2 < n \leq N(\mathbb{F}_q) - 3$ . By Lemma 5.4, 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

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

where  $P_1 \oplus P_2 = P_3 \oplus P_4 = \dots = P_{N-2} \oplus P_{N-1} = O$ .

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

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

Then by Proposition 5.2, there are no  $k$  cyclically consecutive points whose sum is  $O$ . And hence, the elliptic curve code  $C_{\mathcal{L}}(D, G)$  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

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

Then by Proposition 5.2, there are no  $k$  cyclically consecutive points whose sum is  $O$ . And hence, the elliptic curve code  $C_{\mathcal{L}}(D, G)$  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

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

and arrange it by the following algorithm:

**Step 1.** Arrange  $D_0$  as following

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



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

$$P_1, \dots, P_{k-1}, P_{N-5}, P_k, \dots, P_{(s-1)k-1}, P_{N-3-s}, P_{(s-1)k}, \dots, P_{sk-1}, P_{N-4}, P_{sk}, P_{sk+1}, \dots, P_{sk+r-2}.$$

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

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

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

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

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

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

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:

$$P_{(s-1)k+r+2} \oplus \dots \oplus P_{N-4} \oplus \dots \oplus P_{N-2} = P_{(s-1)k+r+2} \oplus P_{N-4} \oplus P_{sk+r-1} \oplus P_{N-2} = O,$$

and

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

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

$$P_{(s-1)k+r+2i+2} \oplus \dots \oplus P_{N-4} \oplus \dots \oplus P_{N-2} \oplus P_1 \oplus \dots \oplus P_{2i} = P_{(s-1)k+r+2i+2} \oplus P_{N-4} \oplus P_{sk+r-1} \oplus P_{N-2} = O,$$

and

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

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  $D$ , by Proposition 5.2, there are no  $k$  cyclically consecutive points whose sum is  $O$ . And hence, the elliptic curve code  $C_{\mathcal{L}}(D, G)$  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

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

and arrange it as following

$$D = \{P_1, \dots, P_{k-1}, P_{N-3}, P_k, \dots, P_{(s-1)k-1}, \\ P_{N-1-s}, P_{(s-1)k}, \dots, P_{sk-1}, P_{N-2}, P_{sk}, P_{sk+1}, \dots, P_{sk+r}\}.$$

If  $r$  is even, then it is easy to see that by Proposition 5.2 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, the elliptic curve code  $C_{\mathcal{L}}(D, G)$  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.  $\square$

**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. To construct MDS symbol-pair codes using these almost MDS codes, how to arrange the evaluation set becomes the difficult step.

## 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  $q^2 + q + 1$ . Later, 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 6 to  $q^2 + 1$ . 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.

## References

- [1] E. Ballico and A. Cossidente. Curves in projective spaces and almost MDS codes. *Des. Codes Cryptogr.*, 24(2):233–237, 2001.
- [2] Y. Cassuto and M. Blaum. Codes for symbol-pair read channels. *IEEE Trans. Inform. Theory*, 57(12):8011–8020, 2011.
- [3] Y. Cassuto and S. Litsyn. Symbol-pair codes: Algebraic constructions and asymptotic bounds. In *IEEE Int. Symp. Inf. Theory*, pages 2348–2352, 2011.
- [4] Y. M. Chee, L. Ji, H. M. Kiah, C. Wang, and J. Yin. Maximum distance separable codes for symbol-pair read channels. *IEEE Trans. Inform. Theory*, 59(11):7259–7267, 2013.
- [5] Q. Cheng. Hard problems of algebraic geometry codes. *IEEE Trans. Inform. Theory*, 54(1):402–406, 2008.
- [6] M. Deuring. Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. *Abh. Math. Sem. Univ. Hamburg*, 14(1):197–272, 1941.
- [7] P. Hall. On representatives of subsets. *J. Lond. Math. Soc.*, s1-10(1):26–30, 1935.
- [8] X. Kai, S. Zhu, and P. Li. A construction of new MDS symbol-pair codes. *IEEE Trans. Inform. Theory*, 61(11):5828–5834, 2015.
- [9] J. Li, D. Wan, and J. Zhang. On the minimum distance of elliptic curve codes. In *IEEE Int. Symp. Inf. Theory*, pages 2391–2395, 2015.
- [10] C. Munuera. On the main conjecture on geometric MDS codes. *IEEE Trans. Inform. Theory*, 38(5):1573–1577, 1992.
- [11] S. Payne. Topics in finite geometry: ovals, ovoids and generalized quadrangles. *UC Denver Course Notes*, 2009.

- [12] J. H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009.
- [13] H. Stichtenoth. *Algebraic function fields and codes*, volume 254 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, second edition, 2009.
- [14] E. Yaakobi, J. Bruck, and P. H. Siegel. Decoding of cyclic codes over symbol-pair read channels. In *IEEE Int. Symp. Inf. Theory*, pages 2891–2895, 2012.
- [15] E. Yaakobi, J. Bruck, and P. H. Siegel. Constructions and decoding of cyclic codes over  $b$ -symbol read channels. *IEEE Trans. Inform. Theory*, 62(4):1541–1551, 2016.
- [16] J. Zhang, F. Fu, and D. Wan. Stopping sets of algebraic geometry codes. *IEEE Trans. Inform. Theory*, 60(3):1488–1495, 2014.