

# S-Code: New Distance-3 MDS Array Codes

Rajendra Katti and Xiaoyu Ruan

Department of Electrical and Computer Engineering  
North Dakota State University, Fargo, ND 58105

E-mail: {rajendra.katti,xiaoyu.ruan}@ndsu.edu

**Abstract**— In this paper, we present a new description of the X-Code, a class of MDS array code, using skewness, named S-Code. The X-Codes result in codewords that are arrays of size  $n \times n$ , where  $n$  is prime. Our new description does not require  $n$  to be prime but requires  $n$  to be an odd number with smallest prime factor greater than 3. We prove that the S-Codes result in a distance-3 MDS code and give a description of which slopes other than 1 and  $-1$  can be used to construct S-Codes. Encoding and decoding procedures of the S-Code are also introduced.

## I. INTRODUCTION

Array codes [1] have applications in communications and storage systems [2]. Array codes use only XOR and cyclic shift operations for encoding and decoding procedures and are hence more efficient than Reed-Solomon Codes in terms of computational complexity [3].

Xu and Bruck [4] proposed a class of distance-3 MDS array codes called X-Code. The construction of X-Code is given below.

In X-Code, information symbols are placed in an array of size  $(n - 2) \times n$ . Symbols are defined over any Abelian group with an addition operation  $+$ . Parity symbols are constructed from the information symbols along several parity check diagonals with the addition operation  $+$ . The parity symbols are placed in the bottommost two rows of the array. So the array is of size  $n \times n$  where rows 0 through  $n - 3$  contain information symbols while rows  $n - 2$  and  $n - 1$  contain parity symbols. Each column has information symbols as well as parity symbols.

Let  $C(i, j)$  be the symbol at row  $i$  and column  $j$ . The parity symbols are computed according to the following encoding rules:

$$C(n-2, i) = \sum_{k=0}^{n-3} C(k, (i+k+2) \bmod n) \quad (1)$$

$$C(n-1, i) = \sum_{k=0}^{n-3} C(k, (i-k-2) \bmod n)$$

where  $i = 0, 1, \dots, n - 1$ . Geometrically, the two parity rows are checksums along diagonals of slopes 1 and  $-1$  respectively. Let us see an example.

## Example 1.

A  $5 \times 5$  X-Code array is constructed as follows. In the two schemes shown in Figures 1 and 2, every block in the array is numbered. The symbols in the blocks of the same number are added to form a parity symbol.

	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Figure 1

	0	1	2	3	4
0	0	4	3	2	1
1	1	0	4	3	2
2	2	1	0	4	3
3	3	2	1	0	4
4	4	3	2	1	0

Figure 2

Note that the last rows of both schemes are not used. An example codeword is shown in Figure 3.

	0	1	2	3	4
0	1	0	0	1	1
1	0	1	0	1	1
2	0	0	1	0	1
3	0	0	1	1	0
4	1	1	0	1	1

Figure 3

Parity symbols in row 3 correspond to row 3 of the scheme in Figure 1. Similarly, parity symbols in row 4 correspond to row 3 of the scheme in Figure 2.

The X-Codes have optimal encoding/update complexity, i.e., a change of any single data symbol affects exactly  $d$  parity symbols.

X-Code is an  $(n, k)$  code where  $k$  is the number of information rows in the codeword. A code is MDS if the code distance,  $d$ , meets the Singleton bound [5]  $d \leq n - k + 1$  with equality. The X-Code is MDS because  $k = n - 2$  and it is shown in [4] that the X-Code has a column distance of 3. i.e.,  $d = 3$ . Distance-3 implies that either 2 column erasures or 1 column error can be corrected. Refer to [4] for decoding procedures of X-Codes.

The remainder of the paper is organized as follows. Section II proposes the encoding procedure for an S-Code. Section III states some lemmas and a theorem describing properties of code distance and skews. Section IV discusses which skews can be used to construct an S-Code. Section V presents the decoding procedures for correcting two column erasures and one column error, respectively. Finally Section VI concludes the paper.

## II. ENCODING PROCEDURE FOR S-CODES

In this section we use another approach to describe the construction of X-Codes. We name the code under the new construction rule S-Code. This alternative approach uses skews  $(1, 1)$  and  $(1, n - 1)$ .

An information symbol in row  $i$  and column  $j$  of the array is referred to as  $d(i, j)$ , where  $0 \leq i \leq n - 3$  and  $0 \leq j \leq n - 1$ . A parity symbol  $p(n - 2, k)$  in row  $(n - 2)$  where  $0 \leq k \leq n - 1$  is given by the following equation:

$$p(n-2, k) = \sum_{i+j \equiv x \pmod{n}} d(i, j) \quad (2)$$

where  $(n - 2) + k \equiv x \pmod{n}$ . The  $k^{\text{th}}$  parity symbol in row  $(n - 2)$  is the sum of the information symbols in position  $(i, j)$  such that  $i$  and  $j$  satisfy the equation  $i + j \equiv x \pmod{n}$  where  $x$  is given by  $(n - 2) + k \equiv x \pmod{n}$ . Similarly the  $k^{\text{th}}$  parity symbol in row  $(n - 1)$  is given by the following equation:

$$p(n-1, k) = \sum_{i+(n-1)j \equiv y \pmod{n}} d(i, j) \quad (3)$$

where  $y$  is given by  $(n - 2) + (n - 1)k \equiv y \pmod{n}$ .

In [4] it was stated that  $n$  must be a prime number. However here it is only required that  $n$  be such that the smallest prime factor of  $n$  is at least 5. i.e.,  $n$  must be an odd number that does not have a prime factor of 3.

For a  $5 \times 5$  S-Code array, using the above construction rule we have two schemes shown in Figures 1 and 2, respectively. In each scheme, rows 0 through 2 correspond to information symbols. One of rows 3 and 4 corresponds to the parity bits but not both. According to the construction rule, row 3 corresponds to parity bits. For illustration purposes consider the entry  $(3, 1)$  in Figure 1. This entry is 4. This means that the co-ordinates  $(i, j)$  of the information symbols that will be added to compute  $p(3, 1)$  are specified by the co-ordinates of the occurrences of 4's in the first three rows of Figure 1. This implies that

$$p(3, 1) = d(0, 4) + d(1, 3) + d(2, 2) \quad (4)$$

Note again that rows  $n - 1$  of both schemes are not used. Row  $n - 2$  of each array is respectively stored in one of the last two rows of a codeword.

## III. THE MDS PROPERTY OF S-CODES

In the following lemmas and theorems we let the  $(i, j)^{\text{th}}$  entry in one of the schemes resulting from (2) and (3) be  $e_a(i, j)$  and that of the other scheme be  $e_b(i, j)$ . Note that  $0 \leq i \leq n - 1$  and  $0 \leq j \leq n - 1$  and the same range applies to other variables denoting a row or a column. No proofs for lemmas and theorems are given in this paper because of length limitations.

**Lemma 1.** *The scheme resulting from (2) or (3) is a Latin square of order  $n$ . i.e., no row or column contains the same entry twice.*

**Lemma 2.** *There do not exist  $(i_1, j_1)$  and  $(i_2, j_2)$  where  $(i_1, j_1) \neq (i_2, j_2)$  such that  $e_a(i_1, j_1) = e_b(i_1, j_1)$  and  $e_a(i_2, j_2) = e_b(i_2, j_2)$ .*

**Lemma 3.** *If  $e_a(i_1, j_1) = e_a(i_2, j_2)$  where  $(i_1, j_1) \neq (i_2, j_2)$  then there do not exist  $(i_1, j_1)$  and  $(i_2, j_2)$  such that  $e_b(i_1, j_1) = e_b(n - 2, j_2)$  and  $e_b(i_2, j_2) = e_b(n - 2, j_1)$ .*

**Lemma 4.** *If  $e_a(i_1, j_1) = e_a(i_2, j_2)$  and  $e_a(i_3, j_2) = e_a(i_4, j_1)$  where  $(i_1, j_1)$ ,  $(i_2, j_2)$ ,  $(i_3, j_2)$ , and  $(i_4, j_1)$  are all distinct, then there do not exist  $(i_1, j_1)$ ,  $(i_2, j_2)$ ,  $(i_3, j_2)$ , and  $(i_4, j_1)$  such that  $e_b(i_1, j_1) = e_b(i_3, j_2)$  and  $e_b(i_2, j_2) = e_b(i_4, j_1)$ .*

**Theorem 1.** *The S-Codes have a code distance of 3.*

## IV. S-CODES WITH SLOPES OTHER THAN $1/1$

Recall that the S-Code was constructed using skews  $(1, 1)$  and  $(1, n - 1)$  in Section II. We now reconstruct the S-Code using skews  $(1, a)$  and  $(1, b)$  such that  $1 \leq a \leq n - 1$  and  $1 \leq b \leq n - 1$ . The skew  $(p, q)$  corresponds to slope  $q/p$ . i.e., we will construct the X-Code with slopes  $a$  and  $b$ .

Using skews  $(1, a)$  and  $(1, b)$ , construction rule (2) and (3) would be modified to (5) and (6).

$$p(n-2, k) = \sum_{i+aj \equiv x \pmod{n}} d(i, j) \quad (5)$$

$$p(n-1, k) = \sum_{i+bj \equiv y \pmod{n}} d(i, j) \quad (6)$$

where  $0 \leq i \leq n - 3$ ,  $0 \leq j \leq n - 1$ ,  $(n - 2) + ak \equiv x \pmod{n}$ , and  $(n - 2) + bk \equiv y \pmod{n}$ .

Though we have ignored the proofs for lemmas in Section III, we need to state some conditions regarding these lemmas here. Lemma 1 requires that  $\gcd(a, n) = \gcd(b, n) = 1$ . Lemma 2 requires that  $\gcd(b - a, n) = \gcd(a - b, n) = 1$ . Lemma 3 requires that  $\gcd(a, b) = \gcd(2a - b, n) = \gcd(2b - a, n) = 1$ , and  $n$  must not have prime factors 2 or 3. If all these conditions are satisfied then the MDS property remains.

Theorem 2 is a direct consequence of Lemmas 1 through 4 and Theorem 1.

**Theorem 2.** If  $\gcd(a, n) = \gcd(b, n) = \gcd(b - a, n) = \gcd(a - b, n) = \gcd(a, b) = \gcd(2a - b, n) = \gcd(2b - a, n) = 1$  and the smallest prime factor of  $n$  is neither 2 nor 3, then the S-Code constructed using skews  $(1, a)$  and  $(1, b)$  has a code distance of 3.

**Example 2.**

We use skews  $(1, 2)$  and  $(1, 3)$  to construct a  $5 \times 5$  S-Code. i.e.,  $a = 2$  and  $b = 3$ , which satisfy the condition  $\gcd(a, n) = \gcd(b, n) = \gcd(b - a, n) = \gcd(a - b, n) = \gcd(a, b) = \gcd(2a - b, n) = \gcd(2b - a, n) = 1$ . The  $k^{\text{th}}$  ( $0 \leq k \leq n - 1$ ) entry in row 3 is given by (7):

$$p(3, k) = \sum_{i+2j \equiv k \pmod{5}} d(i, j) \quad (7)$$

where  $3 + 2k \equiv x \pmod{5}$ . Similarly the  $k^{\text{th}}$  entry in row 4 is given by (8):

$$p(4, k) = \sum_{i+3j \equiv y \pmod{5}} d(i, j) \quad (8)$$

where  $3 + 3k \equiv y \pmod{5}$ . The two resulting schemes are shown in Figures 4 and 5 respectively. An example codeword is shown in Figure 6.

	0	1	2	3	4
0	0	2	4	1	3
1	1	3	0	2	4
2	2	4	1	3	0
3	3	0	2	4	1
4	4	1	3	0	2

Figure 4

	0	1	2	3	4
0	0	3	1	4	2
1	1	4	2	0	3
2	2	0	3	1	4
3	3	1	4	2	0
4	4	2	0	3	1

Figure 5

	0	1	2	3	4
0	1	0	0	1	1
1	0	1	0	1	1
2	0	0	1	0	1
3	0	0	1	1	0
4	0	0	1	1	0

Figure 6

## V. DECODING PROCEDURES FOR S-CODES

It is shown in [4] that a distance-3 MDS array code can correct either two column erasures or one column error. In this section we give such decoding procedures for S-Codes.

### A. Correcting Two Column Erasures

From the construction rule of the S-Code it is easy to see that every parity bit is the sum of a unique set of bits, none of which are in the same column as the parity bit.

If two columns are erasures, then there are  $2 \times (n - 2)$  data symbols that are unknowns and there are  $2 \times (n - 2)$  parity bits that are not erased. These surviving parity bits result in  $2 \times (n - 2)$  equations. By Theorem 1 the S-Code has a distance of 3, thus these  $2 \times (n - 2)$  equations are linearly independent. They can be solved to obtain the  $2 \times (n - 2)$  erased data symbols.

From the code construction there exist two parity check symbols (in each row) such that only one of the data symbols affecting these parity checks is erased. The recovery of data symbols from parity checks in row  $(n - 2)$  results in the recovery of more data symbols that affect parity checks in row  $(n - 1)$ . This in turn results in more data symbols being recovered due to parity checks in row  $(n - 2)$ . Thus a chain of checks can be solved one-by-one to recover all erased data symbols.

We consider the same array code as in Example 2.

Figure 4 shows the values of  $(i + 2j) \pmod{5}$  for all  $(i, j)$ . Note that the sum of all  $d(i, j)$  which have the same value of  $(i + 2j) \pmod{5}$  is a parity bit in row 3. Figure 5 shows the values of  $(i + 3j) \pmod{5}$  for all  $(i, j)$ . Again the sum of all  $d(i, j)$  which have the same value of  $(i + 3j) \pmod{5}$  is a parity bit in row 4.

Example 3 illustrates the decoding procedure for correcting two column erasures.

**Example 3.**

Suppose that columns 2 and 3 in the array code of Example 2 are the erased columns. According to Figure 4 we have

$$p(3, 1) = d(0, 0) + d(1, 2) + d(2, 4)$$

$$p(3, 0) = d(1, 1) + d(2, 3) + d(0, 4)$$

Note that row 3 parity bits of a codeword are shown in row 3 of Figure 4, and row 4 parity bits of a codeword are shown in row 3 of Figure 5. In both the above equations there is only one data symbol in column 2 or 3 and these are  $d(1, 2)$  in  $p(3, 1)$  and  $d(2, 3)$  in  $p(3, 0)$ . Since all the other quantities in the above two equations are known,  $d(1, 2)$  and  $d(2, 3)$  can be solved.

We now look for parity symbols in row 4 that depend on two symbols in columns 2 and 3 such that one of which is either  $d(1, 2)$  or  $d(2, 3)$ . We find that

$$p(4, 3) = d(2, 0) + d(1, 2) + d(0, 4)$$

$p(4, 3)$  depends on  $d(1, 2)$  and other known data symbols. Since  $d(1, 2)$  was just calculated  $p(4, 3)$  can be found. Similarly,

$$p(4, 1) = d(1, 0) + d(0, 2) + d(2, 3)$$

$p(4, 1)$  depends on  $d(2, 3)$  and other known data symbols. Therefore  $d(0, 2)$  can be computed since  $d(2, 3)$  is known.

We continue by looking for parity symbols in row 3 that depend on two symbols in columns 2 and 3 such that one of

which is either  $d(0,2)$  or  $d(2,3)$ . We can thus go back and forth along parity symbols in rows 3 and 4 and compute all erased data symbols.

Note that another such chain can be set up by starting from parity symbols in row 4 that depend only on one data symbol in the erased columns. In this example  $p(4,0)$  and  $p(4,4)$  depend only on  $d(2,2)$  and  $d(1,3)$  respectively. Therefore  $d(2,2)$  and  $d(1,3)$  can be found since all the other data symbols that determine  $p(4,0)$  and  $p(4,4)$  are known.

### B. Correcting One Column Error

The main task of correcting one column error is to locate the position of the column. The error data can then be computed using either skew. We first recall some properties from the construction of S-Codes and then introduce the decoding procedure.

#### Proposition.

An S-Code array has the following properties:

1) For each parity bit computed from a skew, there exists one and only one column that is not involved in computing this parity bit.

2) Let the  $(i,j)^{th}$  ( $0 \leq i \leq n-1$ ,  $0 \leq j \leq n-1$ ) entry in Figure 4 be  $r_{ij}$  and the  $(i,j)^{th}$  entry in Figure 5 be  $s_{ij}$ , then every ordered pair  $(r_{ij}, s_{ij})$  is unique among all positions in the array.

Above propositions are direct consequences from Lemmas 1 through 4. The decoding procedure is described step by step below.

**Step 1.** Compute all parity bits using (5) and (6). Compare the results with the parity bits stored in the last two rows of the array. If they exactly match, then there is no error or more than one column has error bits.

**Step 2.** For the two skews, mark each mismatched parity bit and the corresponding data symbols that are used to compute it. i.e., they are with the same entry in Figure 4 or 5. Call those marked symbols "suspects".

**Step 3.** Compare the suspect locations of data symbols in schemes of two skews. No matter how many errors the error column has, for this column the suspect locations of rows 0 through  $n-3$  should exactly match. Such a column is unique in the array according to Proposition 2).

Example 4 uses the same S-Code as previous examples.

#### Example 4.

Suppose symbols  $d(0,1)$ ,  $d(1,1)$ , and  $d(2,1)$  of column 1 are error bits. Figures 7 and 8 show the suspects from skews (1,2) and (1,3) respectively. A question mark denotes a suspect. Entries of row 4 in both schemes are not shown because they are not used.

We find that the suspect locations of data symbols in column 1 of the two schemes exactly match. This is the only column that the suspect locations of data symbols exactly match. Therefore it is determined that column 1 is the error column we are looking for. The suspects not in column 1 are correct. The data symbols in other columns can then be used to compute data symbols in the error column.

	0	1	2	3	4
0	0	?	?	1	?
1	1	?	0	?	?
2	?	?	1	?	0
3	?	0	?	?	1
4					

Figure 7

	0	1	2	3	4
0	?	?	1	?	2
1	1	?	2	?	?
2	2	?	?	1	?
3	?	1	?	2	?
4					

Figure 8

## VI. CONCLUSION

We have presented an  $n \times n$  code called S-Code that has distance 3 and requires  $n$  to be such that its smallest prime factor is at least 5. Our codes are formed based on skews. The computation of encoding and decoding of S-Codes is simple and equally distributed across all columns.

## ACKNOWLEDGMENT

This work has been supported in part by the US National Science Foundation under grant CCR-0429523.

## REFERENCES

- [1] M. Blaum, J. Bruck and A. Vardy, "MDS array codes with independent parity symbols", *IEEE Trans. Inform. Theory*, 42(2), 529-542, Mar. 1996.
- [2] P. G. Farrell, "A survey of array error control Codes", *Europ. Trans. Telecommun.*, 3(5), 441-454, 1992.
- [3] M. Blaum, P. G. Farrell and H.C. A. van Tilborg, "Chapter on array codes", *Handbook of Coding Theory*, VS Pless and WC Huffman Eds., Elsevier, North-Holland, 1998.
- [4] L. Xu and J. Bruck, "X-Code: MDS array codes with optimal encoding", *IEEE Trans. Inform. Theory*, 45(1), 272-276, Jan. 1999.
- [5] F. J. MacWilliams and N. J. A. Sloane, *The theory of error correcting codes*, Amsterdam, North-Holland, 1977.