S-CODE: NEW MDS ARRAY CODES WITH OPTIMAL ENCODING

Rajendra Katti and Xiaoyu Ruan

Department of Electrical and Computer Engineering
North Dakota State University
Fargo, North Dakota 58105, U.S.A.

ABSTRACT

In this paper, we present a new description of the X-Code, a class of MDS array code, using skews, 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. We also give a description of which slopes other than 1 and –1 can be used to construct S-Codes.

1. X-CODE

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:

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

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

Example 1

A 5 × 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] $d \le n - k + 1$ 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.

2. S-CODE

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 $0 \le i \le n - 3$ and $0 \le j \le n - 1$. A parity symbol $p(n - 2, k)$ in row $(n - 2)$ where $0 \le k \le n - 1$ is given by the following equation:

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

where $(n-2)+k \equiv x(\bmod n)$. The k^{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 $i + j \equiv x(\bmod n)$ where x is given by $(n-2)+k \equiv x(\bmod n)$. Similarly the k^{th} parity symbol in row $(n - 1)$ is given by the following equation:

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

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

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 × 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

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

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.

3. THE MDS PROPERTY OF S-CODE

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 $0 \le i \le n - 1$ and $0 \le j \le n - 1$ and the same range applies to other variables denoting a row or a column. No proofs of lemmas are given in this manuscript because of space 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 $(i_1,j_1) \ne (i_2,j_2)$ 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 $(i_1,j_1) \ne (i_2,j_2)$ 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.*

Proof: Observe that S-Code is a linear code, thus proving that the code has distance 3 is equivalent to proving that a valid non-zero codeword has minimum column weight of 3, i.e., among the n columns at least 3 are non-zero. A column is non-zero if at least one symbol in it is non-zero. Let the number of non-zero columns be w. We will show that $w \ge 3$.

Suppose that in the information array, only one column, j_d $(0 \le j_d \le n - 1)$, is non-zero. We consider two cases:

(1) Column j_d contains only one non-zero information symbol. By Lemma 1, exactly one parity symbol in row n

– 2 is non-zero. Let it be in column j_{p1} ($0 \le j_{p1} \le n - 1$) then $j_{p1} \ne j_d$. Similarly exactly one parity symbol in row $n - 1$ is non-zero. Let it be in column j_{p2} ($0 \le j_{p2} \le n - 1$) then $j_{p2} \ne j_d$. By Lemma 2, $j_{p1} \ne j_{p2}$. Therefore there exist three non-zero columns, j_d, j_{p1}, and j_{p2}, i.e., $w = 3$.

(2) Column j_d contains r ($n - 2 \ge r \ge 2$) non-zero information symbols. By Lemma 1, the r non-zero information symbols do not add up to form any parity symbols. The non-zero parity symbols distribute in at least r columns. By Lemma 1 and Lemma 2, column j_d is not among these r columns. Thus $w \ge r + 1$ where $r \ge 2$. i.e., $w \ge 3$.

Now suppose that in the information array, two columns, j_{d1} and j_{d2} ($0 \le j_{d1} \le n - 1$, $0 \le j_{d2} \le n - 1$, $j_{d1} \ne j_{d2}$), are non-zero. We consider four cases:

(1) Column j_{d1} contains only one non-zero information symbol at (i_1,j_{d1}) and column j_{d2} contains only one non-zero information symbol at (i_2,j_{d2}).

If $e_a(i_1,j_{d1}) \ne e_a(i_2,j_{d2})$ and $e_b(i_1,j_{d1}) \ne e_b(i_2,j_{d2})$ then $w \ge 3$ due to obvious reasons.

By Lemma 2 it is impossible that $e_a(i_1,j_{d1}) = e_a(i_2,j_{d2})$ and $e_b(i_1,j_{d1}) = e_b(i_2,j_{d2})$.

If $e_a(i_1,j_{d1}) = e_a(i_2,j_{d2})$ and $e_b(i_1,j_{d1}) \ne e_b(i_2,j_{d2})$, then the array will contain exactly 2 non-zero columns if and only if $e_b(i_1,j_{d1}) = e_b(n - 2,j_{d2})$ *and* $e_b(i_2,j_{d2}) = e_b(n - 2,j_{d1})$. By Lemma 3 such (i_1,j_{d1}) and (i_2,j_{d2}) do not exist. Therefore there will be at least 3 non-zero columns. i.e., $w \ge 3$.

(2) Column j_{d1} contains only one non-zero information symbol at (i_1,j_{d1}) and column j_{d2} contains 2 non-zero information symbols at (i_2,j_{d2}) and (i_3,j_{d2}). Then the array will contain exactly 2 non-zero columns if and only if $e_a(i_1,j_{d1}) = e_a(i_2,j_{d2})$, $e_a(i_3,j_{d2}) = e_a(n - 2,j_{d1})$, $e_b(i_1,j_{d1}) = e_b(i_3,j_{d2})$, and $e_b(i_2,j_{d2}) = e_b(n - 2,j_{d1})$. By Lemma 4 such (i_1,j_{d1}), (i_2,j_{d2}), and (i_3,j_{d2}) do not exist. Therefore there will be at least 3 non-zero columns. i.e., $w \ge 3$.

(3) Column j_{d1} contains 2 non-zero information symbol at (i_1,j_{d1}) and (i_4,j_{d1}) and column j_{d2} contains 2 non-zero information symbols at (i_2,j_{d2}) and (i_3,j_{d2}). Then the array will contain exactly 2 non-zero columns if and only if $e_a(i_1,j_{d1}) = e_a(i_2,j_{d2})$, $e_a(i_3,j_{d2}) = e_a(i_4,j_{d1})$, $e_b(i_1,j_{d1}) = e_b(i_3,j_{d2})$, and $e_b(i_2,j_{d2}) = e_b(i_4,j_{d1})$. By Lemma 4 such (i_1,j_{d1}), (i_2,j_{d2}), (i_3,j_{d2}), and (i_4,j_{d1}) do not exist. Therefore there will be at least 3 non-zero columns. i.e., $w \ge 3$.

(4) Both Columns j_{d1} and j_{d2} contain more than 2 non-zero information symbols. Since more that 2 pairs of sums will be generated among different columns, $w \ge 3$.

Lastly, suppose that in the information array three or more columns are non-zero. Then $w \ge 3$.

□

We have shown $w \ge 3$ in all cases. Therefore the S-Code has a code distance of 3.

4. S-CODE 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 $1 \le a \le n - 1$ and $1 \le b \le n - 1$. 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).

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

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

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

It is worth revisiting the lemmas introduced in Section 3. 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 follows.

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.*

Proof: This is a direct consequence of Lemmas 1 through 4 and Theorem 1.

□

Example 2

We use skews (1,2) and (1,3) to construct a 5 × 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^{th} ($0 \le k \le n - 1$) *entry in row 3 is given by (7):*

$$p(3,k) = \sum_{i+2j \equiv x (\mathrm{mod}\, 5)} d(i,j) \tag{7}$$

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

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

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

5. CONCLUSION

We have presented an alternative description of the X-Code of size $n \times n$ using skews, called S-Codes. We have shown that n does not need to be a prime number. The constraint on n is that the smallest prime factor of n be at least 5. A general condition is proposed for using other slopes to construct the S-Code. Future research will be to extend the code distance from 3 to d ($d \geq 4$). Our preliminary research shows that, if n and the skews are carefully chosen, then applying more skews (thus more parity rows) would result in larger distances.

6. ACKNOWLEDGEMENT

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

7. 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", ETT, 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*.

[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.

	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