# Generation of Maximum Distance Separable Codes

G. Solomon
Communications Systems Research Section

*A new set of e-error-correcting Maximum Distance Separable codes of lengths $(2^m+1)$ over GF$(2^m)$ are generated as binary codes of length $m(2^m+1)$ shortened from certain cyclic codes of length $(2^{2m}-1)$ with $2me$ parity bits. Encoding uses a binary division shift register of size $2me$ and an auxiliary computation involving $2me$ binary additions. Decoding can be done by standard Reed-Solomon decoding techniques. In fact, the new codes can be considered extended Reed-Solomon codes.*

## I. Introduction

Maximum Distance Separable (MDS) codes have a minimum distance of the code length minus the code dimension plus one. In short, there are two redundant symbols for each of the $e$ errors to be corrected. The most famous and practical examples are the Reed-Solomon codes over GF$(2^m)$ extendable up to length $2^m + 1$. Presented here is a new set of MDS codes over GF$(2^m)$ up to lengths $2^m + 1$ that are more easily encodable using a binary division shift register of length $2me$ plus an auxiliary computation involving $2me$ binary additions. Similarly, the syndrome generation may be reduced to a simple division shift register reencoding. The actual decoding can be done by the available techniques used in Reed-Solomon decoding. In fact, these codes can be considered extended Reed-Solomon codes. The news of interest here is that these codes are seen as subsets of certain shortened binary cyclic codes, and in this light, the Hamming quaternary codes along with a simpler algebraic decoding can be obtained.

## II. An Illustrative Example

MDS codes of length $(2^m + 1)$ over GF$(2^m)$ can be generated by shortening certain binary cyclic codes of length $(2^{2m} - 1)$ to the length $m(2^m + 1)$ and identifying the symbols as $m$-tuples corresponding to the coordinate values at $\{j + i(2^m + 1);\ i = 0, 1, 2, \ldots, m - 1\}$ for $j = 0, 1, 2, \ldots, 2^m$. The generator polynomial for an $e$-symbol error-correcting code is chosen to give the syndromes $S_{1+j(2^m-1)}$, for $j = 1, 2, 3, \ldots, e$.

**Example ($m = 3$, $e = 2$).** A (9,5;5) MDS code over GF(8) is obtained from the (63,51;?) cyclic code with syndromes $S_8$ and $S_{15}$, which are actually conjugates of $S_1$ and $S_{-3}$. So the generator polynomial for the code is $f_1(x)f_{-3}(x)$, where $f_1(x)$ is a primitive polynomial of degree 6 with root $\alpha$, and $f_{-3}(x)$ is irreducible of degree 6 with root $\alpha^{-3}$. The code is shortened to the $(27, 15; 5)$ binary code. With the identification of the symbols with the values at the triples (0,9,18), (1,10,19), (2,11,20),..., (8,17,26), this is now a (9,5;5) MDS code over GF(8).

## III. Encoding

The $((2^m+1),\ (2^m+1)-2e;\ 2e+1)$ MDS codes over GF($2^m$) obtainable this way may be encoded in two stages. The first stage uses a binary division shift register of length $2me$ with the $m$-bit information symbols and the $2e$ zero $m$-tuples sent through the register to obtain a $2me$-bit parity sequence $\{p_i\}$. (The information symbols and fake $2e$ zero $m$-tuples have been interleaved $2^m+1$ positions apart as described above.) The second stage of the encoding takes the obtained $2me$-bit parity sequence $\{p_i\}$ and forms $\sum p_i\mathbf{u}_i$ where the $\mathbf{u}_i$ are a stored set of $2me$ binary sequences of length $2me$. These are the $2em$-bit parity symbols to be sent. The bits to be transmitted are assumed demultiplexed and sent in the order of $m$-binary-bit symbols to take advantage of the burst error-correcting properties of this code.

**Example ($m = 2$, $e = 1$).** The $(5,3;3)$ MDS code over GF(4) is generated from the $(15,11;3)$ Hamming code when shortened to the $(10,6;3)$ code. The encoding takes the three doubles, say $a_1a_2b_1b_2c_1c_2$, and encodes the binary sequence $a_1b_1c_100a_2b_2c_200$ via the division shift register $x^4+x+1$ to get $p_1p_2p_3p_4$. The four zeros above will be replaced in transmission by the sum $\sum p_i\mathbf{u}_i$, where $\mathbf{u}_i$ are the respective rows of the matrix $\mathbf{U}$:

$$\mathbf{U} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$\mathbf{U}$ is the inverse of the matrix $\mathbf{Q}$ whose four rows are the parity sequences $\{q_i\}$ obtained by sending ones in the consecutive parity positions and zeros in the information:

$$\mathbf{Q} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

## IV. The Construction

Construction is performed by the following steps.

1. Consider a binary cyclic code of length $2^{2m}-1$. The coordinates may be represented by powers of a primitive element $\beta$. Note that $\beta^s$ for $s = 2^m+1$ is a primitive generator of the subgroup of $(2^m-1)$ roots of unity. So if $\beta^s = \alpha$, and $\gamma^i = \beta^i$ for $i = 0,1,2,\ldots,2^m$, there is an alternative representation $\gamma^i\alpha^j$ for these coordinates; this alternate will be used.
2. Note that $\alpha^j$ for $j = 0,1,2,\ldots,2^m-1$ is a basis for GF($2^m$). Therefore, for any element $\delta \in$ GF($2^m$), a subfield of GF($2^{2m}$), $\delta = c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{m-1}\alpha^{m-1}$ for $c_i \in$ GF(2). Let $J = j(2^m-1)+1$. It is easy to see that $(\gamma^i\alpha^j)^J = \gamma^{iJ}\alpha^j$. Therefore, if the element $\gamma^i(c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{m-1}\alpha^{m-1})$ is raised to the $J$th power, $\gamma^{iJ}(c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{m-1}\alpha^{m-1})$ is obtained.
3. If a cyclic code is shortened to the first $m(2^m+1)$ coordinates, elements like $\gamma^i(c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{m-1}\alpha^{m-1})$ represent a coordinate not in the shortened code when two or more of the $c_i$ are nonzero. Furthermore, if the code is defined by the syndromes $S_J$ for $J = j(2^m-1)+1$, a set of errors in two or more of the coordinates $\gamma^i, \gamma^i\alpha, \gamma^i\alpha^2, \ldots, \gamma^i\alpha^{m-1}$ looks like a single error somewhere in the nonshortened part of the code. This property is crucial to the generation of MDS codes of length $(2^m+1)$ over GF($2^m$).
4. Let $J = j(2^m-1)+1$. Define the syndromes $T_j = S_J$ as usual. Note that $j = 0$ and $J = 1$ give rise to conjugate syndromes, since $T_1 = T_0^{2^m}$. Note too that $T_2$ and $T_{-1}$, $T_3$ and $T_{-2}$ are also conjugates and share the same properties. Proof is by calculation.
5. First consider the binary cyclic code of length $2^{2m}-1$ with the generator polynomial that gives the syndromes $\{T_i;\ i = 1,2,\ldots,e\}$ for $e < (2^m+2)/2$. This code has dimension $2^{2m}-1-2me$. If this code is shortened to length $m(2^m+1)$, a code of binary dimension $m(2^m+1)-2me$ is obtained. Now identify as $m$-tuples the values at coordinates $j,\ j+(2^m+1),\ j+2(2^m+1),\ldots,j+(m-1)(2^m+1)$, for $j = 0,1,\ldots,2^m$, to obtain a code of dimension $(2^m+1)-2e$ over $m$-tuples with minimum distance $2e+1$. This is an MDS code of length $(2^m+1)$.

**Proof:** With the syndromes $\{T_i;\ i = 1,2,\ldots,e\}$, the syndromes $\{T_{-i+1};\ i = 1,2,\ldots,e\}$ are also available. This is a consecutive set of $2e$ syndromes for the values $\gamma^{is}$, $i = 0,1,\ldots,2^m$ for $s = 2^m-1$. Using standard Reed-Solomon decoding techniques for error locations and error values, up to $e$ symbol errors can be corrected. Note, for example, that $T_1$ gives the value $\gamma^i e_i$, where $e_i = \sum_{j=0}^{m-1} a_{ji}\alpha^i$; $a_{ji} \in$ GF(2) are the binary error values in the $i$th symbol.

**Example ($m = 3$, $e = 2$).** There is a $(9,5;5)$ MDS code associated with the $(27,15;5)$ binary code, which is shortened from the $(63,51;?)$ cyclic code with syndromes

$T_1 = S_0$ and $T_2 = S_{15}$. $T_1 = T_0^8$ and $T_{-1} = T_2^8$. For error positions $\gamma^i$, $i = 1, 2$, error values $e_i$ (where $e_i = \sum c_i \alpha^i$ for $i = 0, 1, \ldots, m-1$), and $c_i \in \mathrm{GF}(2)$, there are consecutive syndromes $\{T_i;\ i = -1, 0, 1, 2\}$. In particular

$$T_{-1} = \gamma^{-6} e_1 + \gamma^{-12} e_2$$

$$T_0 = \gamma^1 e_1 + \gamma^2 e_2$$

$$T_1 = \gamma^8 e_1 + \gamma^{16} e_2$$

$$T_2 = \gamma^{15} e_1 + \gamma^{30} e_2$$

Note that the error positions for these syndromes are given by $\gamma^7$ and $\gamma^{14}$. Because these are seventh powers of the original $\gamma^i$, $\gamma$ and $\gamma^2$ are obtained. The error patterns are, of course, given by the values $e_i$ and are solvable in many ways.

**Corollary: Quaternary codes encoded as shortened binary cyclic codes.** The $((4^p - 1)/3,\ (4^p - 1)/3 - p;\ 3)$ Hamming code over GF(4) can be viewed as a shortened quasi-cyclic code that comes from shortening the Bose-Chaudhuri-Hocquenghem (BCH) Hamming cyclic code of length $4^p - 1$ to two-thirds of its length and identifying the $j$th quaternary pairs with the values at the pairs of coordinates $(j, j + (4^p - 1)/3)$.

**Proof:** Take the Hamming binary cyclic code of length $4^p - 1$ and set the last $(4^p - 1)/3$ coordinates to zero. Identify the quaternary symbols as pairs of positions $j$ and $j + (4^p - 1)/3$. If $S_1$ is the error-correcting syndrome, it will correct symbol-error patterns of weight one as per prescription. Note that an error pattern of weight two, i.e., an error in both $i$ and $i + (4^p - 1)/3$ positions, will have the syndrome give the error as $i + 2(4^p - 1)/3$. This is not in the code, so it clearly identifies the pair $i$ and $i + (4^p - 1)/3$ as being in error.