INFORMATION AND CONTROL 8, 163–169 (1965)

On an Extended Class of Error-Locating Codes*

JACK K. WOLF

Department of Electrical Engineering, New York University, New York, New York

LIST OF SYMBOLS

a_j	constant
b	width of burst of errors
C_1	binary error-detecting code
C_2	nonbinary error-correcting code
d	minimum distance
e	number of random errors
$\mathcal{E}_c$	class of correctable errors for C_2
$\mathcal{E}_d$	class of detectable errors for C_1
$\mathbf{H}$	parity check matrix for resultant code
k	number of binary information digits in resultant code
m	number of check symbols in C_2
n	block length of resultant code
$\mathbf{P}$	parity check matrix for C_1
p_l	lth column of $\mathbf{P}$
r	number of binary check digits in resultant code
S	syndrome
S_i	ith component of syndrome (considered as an element of $GF(2^\rho)$)
s	number of sub-blocks in resultant code: also, total number of symbols in code C_2
t	number of digits per sub-block in resultant code: also, total number of digits in code C_1
W	primitive element of $GF(2^3)$
x_i	unknown element of $GF(2^\rho)$
α	element of $GF(2^6)$
$\mathbf{\Gamma}$	parity check matrix for C_2
γ_{ij}	i-jth element of $\mathbf{\Gamma}$
ρ	number of check digits in code C_1

* This study was supported by the United States Air Force Office of Scientific Research under grant number AF-AFOSR-499-65.

An explicit method of constructing error-location (EL) codes is presented for the case where errors occur in multiple sub-blocks. The procedure is applicable both when the erroneous sub-blocks occur randomly throughout the message and when they occur in bursts. The method allows for a wide range in redundancy and error-location capability. The method of construction is outlined for an EL code which locates errors occurring within 2 or fewer random sub-blocks. A general decoding procedure for the codes is presented.

INTRODUCTION

A previous paper (Wolf and Elspas, 1963) introduced the notion of error-locating (EL) codes: codes whose error control capability lies midway between error-correcting (EC) codes and error-detecting (ED) codes. In such codes, each block of received digits is regarded as subdivided into mutually exclusive sub-blocks. A class of codes was described which permitted the detection of errors occurring within a single sub-block and in addition, the identification of that sub-block containing errors.

This paper describes a construction procedure for generating parity check matrices for a much broader class of EL codes than previously reported (Wolf and Elspas, 1963). Codes in this class permit the location of digit errors to within several sub-blocks of the received message (without specifying the precise location of the erroneous digit positions). Codes are described for the location of clustered sub-blocks containing errors (burst sub-block EL codes) and for the location of random sub-blocks containing errors (random sub-block EL codes).

Following the notation of Wolf and Elspas (1963) each block of n binary digits, of which r are check digits and $k = n - r$ are information digits, is subdivided into s mutually exclusive sub-blocks. Each sub-block contains $t = n/s$ binary digits. Parity checks, in general, extend over more than one sub-block with n being the relevant constraint length of the code.

MULTIPLE SUB-BLOCK EL CODES

The following theorem yields an extended class of random sub-block and burst sub-block EL codes.

THEOREM. *Let C_1 be a binary $(t, t - \rho)$ group code with parity check matrix* $\mathbf{P}$ *which* detects *the class of error patterns $\mathcal{E}_d$. Let C_2 be a nonbinary $(s, s - m)$ group code with transmission digits being elements of $GF(2^p)$. Let C_2 have a parity check matrix $\mathbf{\Gamma}$, the i-jth element of this matrix, γ_{ij} ,*

being an element in $GF(2^\rho)$. *Let* C_2 correct *the class of errors* $\mathcal{E}_c$, *where* $\mathcal{E}_c$ *contains either* (1) *the set of all patterns with* e *or fewer random errors or* (2) *the set of all patterns containing a burst of errors spanning* b *or fewer digits* (*these digits being elements of* $GF(2^\rho)$). *Then, there exists a binary* $(n, n - r)$ *group code with* $n = ts$ *and* $r = m\rho$ *such that if*: (1) *the error patterns within each sub-block containing errors are within the class* $\mathcal{E}_d$; *and,* (2) *the erroneous sub-blocks form a pattern of errors which fall in the class* $\mathcal{E}_c$; *then, the errors will be detected and the erroneous sub-blocks identified.*

The parameters t and s in the above theorem are, as previously defined, the number of binary digits per sub-block and the number of sub-blocks, respectively.

Proof: The method of proof is constructive and demonstrates that the matrix $\mathbf{H}$, equal to the Kronecker product[1] of the two matrices $\mathbf{\Gamma}$ and $\mathbf{P}$; that is,

$$\mathbf{H} = \mathbf{\Gamma} \times \mathbf{P} = \overset{\longleftarrow\ n = st\ \longrightarrow}{\begin{bmatrix} \gamma_{11}\mathbf{P} & \cdots & \gamma_{1s}\mathbf{P} \\ & \cdots & \\ \gamma_{m1}\mathbf{P} & \cdots & \gamma_{ms}\mathbf{P} \end{bmatrix}} \begin{matrix} \uparrow \\ m \\ (GF(2^\rho)) \\ \downarrow \end{matrix} \quad \text{or} \quad \begin{matrix} \uparrow \\ r = \rho m \\ (\text{binary}) \\ \downarrow \end{matrix},$$

$$\overset{\longleftarrow t \longrightarrow}{}$$

is indeed the parity check matrix for the desired EL code. The columns of $\mathbf{P}$ are treated as elements of $GF(2^\rho)$ and the indicated products are formed in accordance with the rules of multiplication for elements in $GF(2^\rho)$. The syndrome S is considered as an m-component vector, the ith component, denoted S_i, being an element of $GF(2^\rho)$.

Consider the situation where errors only occur in the jth sub-block, the errors belonging to the class $\mathcal{E}_d$. Denoting the lth component of $\mathbf{P}$ as p_l, the resulting syndrome will contain the components

$$S_i = \Big(\sum_{l \epsilon \mathcal{E}_d} p_l\Big)\gamma_{ij} = a_j\,\gamma_{ij}\,, \qquad i = 1, 2, \cdots, m.$$

The constants a_j will be a nonzero element of $GF(2^\rho)$ since the errors in the sub-block are in the class of detectable errors $\mathcal{E}_d$.

If errors occur within several sub-blocks, say $j_1, j_2, \cdots, j_v$, and if the errors within each sub-block are contained in $\mathcal{E}_d$, the resulting syn-

[1] Slepian (1960) introduced the concept of the Kronecker product of generator matrices. The resulting codes, which Slepian termed product codes, do not seem to be related to the codes discussed in this paper.

drome will contain the components

$$S_i = a_{j_1}\gamma_{ij_1} + a_{j_2}\gamma_{ij_2} + \cdots + a_{j_v}\gamma_{ij_v}, \qquad i = 1, 2, \cdots, m$$

where the a_{j_u} are nonzero elements of $GF(2^p)$.

It must now be shown that the syndrome resulting from detectable errors occurring within one pattern of sub-blocks in $\mathcal{E}_c$ is distinct from syndromes resulting from detectable errors occurring within any other pattern of sub-blocks in $\mathcal{E}_c$. This condition is first shown for the case where $\mathcal{E}_c$ is the set of all patterns with e or fewer random errors. For this case all sets of $2e$ or fewer columns of the matrix $\mathbf{\Gamma}$ are linearly independent. Thus if x_i are elements of $GF(2^p)$ the only solution to the equations

$$\sum_{i=1}^{2e} x_i \gamma_{ij_i} = 0 \qquad i = 1, 2, \cdots, m$$

is the trivial solution $x_1 = x_2 = \cdots = x_{2e} = 0$. However, if a syndrome resulting from detectable errors occurring within e or fewer sub-blocks was equal to the syndrome resulting from detectable errors occurring within another set of e or fewer sub-blocks, this equation would have a nontrivial solution for the x_i. Thus the syndromes resulting from detectable errors occurring in e or fewer sub-blocks are distinct if the erroneous blocks are not identical.

A similar argument holds when $\mathcal{E}_c$ is the set of all patterns containing a single burst of b or fewer digits. Now, the linear independence of the columns results in the equations

$$\sum_{i=1}^{b} x_i \gamma_{i(j_1+i)} + \sum_{i=1}^{b} x_{b+i} \gamma_{i(j_2+i)} = 0 \qquad \begin{matrix} i = 1, 2, \cdots, m \\ j_1 \neq j_2 \end{matrix}$$

having only the trivial solution $x_1 = x_2 = \cdots = x_{2b} = 0$ for the x_i elements of $GF(2^p)$. Such an equation would have to have a nontrivial solution for the x_i if the syndromes in question were not distinct. Thus the theorem is proved.

The codes constructed from this theorem will be random sub-block EL codes or burst sub-block EL codes in accordance with whether C_2 is chosen as a random EC code or a burst EC code.

It is important to note that the set of error patterns within the sub-blocks which can be detected (and then located) is exactly the set of error patterns which can be detected by the code C_1. Thus if C_1 is a code

with minimum distance d, all error patterns within a sub-block containing $d - 1$ or fewer errors will always be detected. However, many other error patterns also will be detected and in fact only $2^{t-\rho} - 1$ of the 2^t possible error patterns within a sub-block will be undetected. The number of undetected errors can be made zero by choosing C_1 to be the trivial code ($\rho = t$): this choice, resulting in a nontrivial EL code.

The following decoding procedure for the EL codes differs only slightly from the decoding procedure for the EC code C_2. The syndrome is calculated as a vector with components in $GF(2)$ and is then ex-expressed as a vector (of lower dimension) with components in $GF(2^\rho)$. Decoding is then accomplished as for code C_2, with errors being specified in a hypothetical received word having s digits which are elements of $GF(2^\rho)$. Only the location of the errors in this hypothetical received word is noted and not the "amount" of each error. The resultant pattern of errors for this hypothetical received word is the pattern of errors for the sub-blocks in the actual received word.

EXAMPLE

As an example, the construction procedure is outlined for an EL code which locates errors occurring within 2 or fewer random sub-blocks. Let $\mathbf{P}$ be the parity check matrix for a binary (7, 4) single EC code. That is:

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

The matrix $\mathbf{P}$ can then be rewritten as

$$\mathbf{P} = [1 \; W^1 \; W^2 \; W^3 \; W^4 \; W^5 \; W^6]$$

where W is an element of $GF(2^3)$ satisfying the equation $W^3 + W + 1 = 0$. Let α be an element of $GF(2^6)$ satisfying the equation $\alpha^2 + W\alpha + 1 = 0$. Using the Bose-Chaudhuri-Hocquenghem (Bose and Ray-Chaudhuri, 1960; Hocquenghem, 1959) procedure for constructing a 2-EC code for elements in $GF(2^3)$ yields the parity check matrix

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & \alpha^1 & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 & \alpha^6 & \alpha^7 & \alpha^8 \\ 1 & \alpha^2 & \alpha^4 & \alpha^6 & \alpha^8 & \alpha^1 & \alpha^3 & \alpha^5 & \alpha^7 \\ 1 & \alpha^3 & \alpha^6 & \alpha^0 & \alpha^3 & \alpha^6 & \alpha^0 & \alpha^3 & \alpha^6 \\ 1 & \alpha^4 & \alpha^8 & \alpha^3 & \alpha^7 & \alpha^2 & \alpha^6 & \alpha^1 & \alpha^5 \end{bmatrix}$$

Writing α^i in terms of the elements of $GF(2^3)$, $\mathbf{\Gamma}$ becomes

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & 1 & W & W^6 & W^3 & W^3 & W^6 & W \\ 0 & 1 & W & W^6 & W^3 & W^3 & W^6 & W & 1 \\ 1 & 1 & W^6 & W^3 & W & 0 & W & W^3 & W^6 \\ 0 & W & W^3 & W^6 & 1 & 1 & W^6 & W^3 & W \\ 1 & W & W^3 & 1 & W & W^3 & 1 & W & W^3 \\ 0 & W^6 & W^6 & 0 & W^6 & W^6 & 0 & W^6 & W^6 \\ 1 & W^6 & W & W & W^6 & 1 & W^3 & 0 & W^3 \\ 0 & W^3 & 1 & W^6 & W & W & W^6 & 1 & W^3 \end{bmatrix}$$

The desired parity check matrix $\mathbf{H}$ would then be the Kronecker product of $\mathbf{\Gamma}$ and $\mathbf{P}$. The multiplication would follow the rule $W^iW^j = W^{i+j}$, the exponents being taken modulo 7. The resultant matrix $\mathbf{H}$ would finally be written as a 24(row) by 63(column) array of elements from $GF(2)$. This EL code has 63 digits of which 24 are check digits and 39 are information digits. These 63 digits are considered as subdivided into 9 sub-blocks containing 7 digits each. If errors occur within 2 or fewer sub-blocks and if the error patterns within these sub-blocks are detectable by the code with parity check matrix $\mathbf{P}$ (this code will detect 113 out of the 128 possible error patterns) then the errant sub-blocks will be identified.

A burst sub-block EL code is not illustrated but would be constructed in like manner. A wide range in error location capability and redundancy is possible depending upon the choice of the codes C_1 and C_2.

OPTIMUM SINGLE SUB-BLOCK EL CODES

The burst sub-block EL codes and the random sub-block EL codes reduce to the same class of codes for the case where errors can occur only within a single sub-block. It can be verified that a suitable choice for $\mathbf{\Gamma}$ (for $m = 2$) would then be

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \alpha^1 & \alpha^2 & \cdots & \alpha^{2^\rho - 2} \end{bmatrix}$$

where α is a primitive element of $GF(2^\rho)$. For the case of $m > 2$, $\mathbf{\Gamma}$ is constructed of columns being all m-tuples containing i zeros ($i = 0, 1, \cdots, m - 1$) the remaining elements being all combinations of nonzero elements of $GF(2^\rho)$ with the restriction that the uppermost nonzero element of each column is 1. Such a matrix will have $s = 2^{\rho m} - 1/2^\rho - 1$ columns. The resultant single sub-block EL code would have the parameters: $n = (2^{\rho m} - 1)t/2^\rho - 1$, $r = \rho m$.

In the previous paper (Wolf and Elspas, 1963) it was shown that the number of check digits required for a single sub-block EL code which locates e or fewer errors occurring in a single sub-block, is bounded from below by

$$r \geqq \log_2 \left[1 + s \sum_{1}^{[e/2]} \binom{t}{i} \right]$$

where $[x]$ denotes the integer value of x. Codes meeting this lower bound were termed optimum EL codes. Substituting the values of r and s found above, this inequality becomes

$$2^\rho \geqq \sum_{0}^{[e/2]} \binom{t}{i} = \sum_{0}^{[(d-1)/2]} \binom{t}{i}$$

where d is the minimum distance of the code. This inequality is an equality if and only if C_1 is a perfect binary EC code (Peterson, 1961). The only (binary) perfect EC codes known are the Hamming SEC codes (Hamming, 1950), the Golay code (Golay, 1949) and all multiple EC codes having only one information digit repeated $n - 1$ times. Thus these codes and only these codes will result in an optimum single sub-block EL code using the procedure described above.

ACKNOWLEDGMENT

The author is deeply indebted to Dr. Bernard Elspas of the Stanford Research Institute, Menlo Park, California, whose ideas are reflected throughout this paper.

RECEIVED: July 13, 1964

REFERENCES

BOSE, R. C., AND RAY-CHAUDHURI, C. K. (1960), On a class of error-correcting binary group codes. *Inform. Control* **3**, 68–79.

GOLAY, M. J. E. (1949), Notes on digital coding. *Proc. IRE* **37**, 657.

HAMMING, R. W. (1950), Error detecting and error correcting codes. *Bell System Tech. J.* **29**, 147–160.

HOCQUENGHEM, A. (1959), Codes correcteurs d'erreurs. *Chiffres.* **2**, 147–156.

PETERSON, W. W. (1961), "Error Correcting Codes." The MIT Press, Cambridge, Mass.

SLEPIAN, D. (1960), Some further theory of group codes. *Bell System Tech. J.* **39**, 1219–1252.

WOLF, J. K., AND ELSPAS, B. (1963), Error-locating codes—a new concept in error control. *IEEE Trans. Inform. Theory* **IT-9**, 20–28.