

## The Check Positions of Hamming Codes and the Construction of a 2EC-AUED Code

<sup>1</sup>HOW GUAN AUN AND <sup>2</sup>ANG MIIN HUEY

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Pulau Pinang, Malaysia

<sup>1</sup>gahow@cs.usm.my, <sup>2</sup>ang\_miin\_huey@hotmail.com

**Abstract.** Hamming Code is the oldest and the most commonly used single error correcting and double errors detecting code. For implication, it is constructed over the field  $GF(2)$ . For each  $r \geq 2$ , there is a  $[n, k, 3]$  Hamming Code where  $n = 2^r - 1$  and  $k = 2^r - r - 1$ . A message word of length  $k$  is encoded using a generating matrix  $G$  into a codeword of length  $n$ . This amounts to inserting  $r$  parity check digits into the message word. The positions of the parity check digits in the codeword are called the *check positions* of the code (with respect to  $G$ ). A received word is then decoded using a parity check matrix  $H$ . If the check positions of the code are in the  $i_1^{\text{th}}, i_2^{\text{th}}, \dots, i_r^{\text{th}}$  coordinates of the codeword, then the  $i_1^{\text{th}}, i_2^{\text{th}}, \dots, i_r^{\text{th}}$  rows of  $H$  are called the *check rows* of the code. We proved in this paper that for any parity check matrix of a Hamming Code, there exists a generating matrix  $G$  of the code, such that the check rows of the code are linearly independent. We believe that this fact is contained implicitly in a paper of Hamming [7] but we cannot find any explicit proof in existing literature. Using the above fact, we construct a 2EC-AUED code.

### 1. Introduction

More than 40 years, error control coding theory has been proved to have increased the reliability of computer or communication systems against errors [17]. Different system may be vulnerable to different types of errors. Error that changes digit 1 to 0 is called a *1-error* whereas error that changes digit 0 to 1 is called a *0-error*. If both 0-errors and 1-errors occur with equal probability in each received word, the errors are classified as *symmetry type*. If 0-errors and 1-errors are both likely to occur but not simultaneously in each received word of a system, the errors are classified as *unidirectional type*. It has been found that the most likely errors that occurred in VLSI memories are not of symmetric type but of unidirectional type [1, 6, 12, 13, 14]. As a result, a receiver usually gets limited number of symmetry errors while the number of unidirectional errors can be very large. Therefore for the late 10 years, the aim of most research works in coding theory [2, 3, 4, 5, 11, 16] is to design codes for correcting up to  $t$  symmetry errors and detecting all occurrences of unidirectional error. These codes are called  $t$  EC-AUED codes. In [5], Bose and Rao have shown that constant weight codes of minimal distance  $2t + 2$  are  $t$  EC-AUED codes. In this paper, we shall construct a 2 EC-AUED code,

which is not of constant weight type. Before this, we need to show a property regarding the check rows of  $[n, k, 3]$  Hamming Codes. The basic theories used in this paper can be found in [8, 9, 15].

## 2. The check rows of Hamming Code

Let  $F$  be any finite field  $GF(p^\mu)$ . A subset  $V$  of  $F^r$  is said to be a *pair-wise independent subset* if and only if any two distinct elements in  $V$  are linearly independent.  $V$  is a *maximal pair-wise independent subset* if and only if there does not exist any other pair-wise independent subset in  $F^r$  that contain  $V$ . Proposition below gives a property of the maximal pair-wise independent subset.

**Proposition.** *Let  $V$  be a pair-wise independent subset of  $F^r$ ,  $r \geq 2$ .  $V$  is a maximal pair-wise independent subset of  $F^r$  if and only if  $|V| = \frac{|F|^r - 1}{|F| - 1}$ .*

*Proof.* Let  $K$  be a relation define on  $F^r - \{0\}$  such that  $\forall a, b \in F^r - \{0\}$ ,  $aKb$  if and only if  $\exists k \in F - \{0\}$  such that  $b = ka$ , i.e., if and only if  $\{a, b\}$  is linearly dependent. Clearly,  $K$  is an equivalent relation. Let  $a \in F^r - \{0\}$  and  $E_a$  be the equivalent class containing  $a$ . Thus,  $E_a = \{b \in F^r \mid aKb\} = \{ka \mid k \in F - \{0\}\}$ .  $|E_a| = |F| - 1$  and thus  $K$  has  $\frac{|F|^r - 1}{|F| - 1}$  distinct equivalent classes.

Let  $V$  be a pair-wise independent subset of  $F^r$ . If  $|V| > \frac{|F|^r - 1}{|F| - 1}$ ,  $\exists b \in F^r - \{0\}$  such that  $|V \cap E_b| \geq 2$ , contradicting the pair-wise independence of  $V$ . If  $|V| < \frac{|F|^r - 1}{|F| - 1}$ , then  $\exists b \in F^r - \{0\}$  such that  $V \cap E_b = \{\}$ . Choose any  $a_j \in E_b$  then  $V \cup \{a_j\}$  is also a pair-wise independent subset of  $F^r$ . This shows that  $V$  is not a maximal pair-wise independent subset of  $F^r$ . Thus we have proved the theorem.

In term of maximal pair-wise independent subset, the definition of Hamming Codes over  $F$  is given as follows

**Definition.** *A linear code over  $F$  such that the rows of its parity check matrix forms a maximal pair-wise independent subset of  $F^r$ ,  $r \geq 2$ , is called a  $[n, k, 3]$  Hamming Code where  $n = \frac{|F|^r - 1}{|F| - 1}$  and  $k = \frac{|F|^r - 1}{|F| - 1} - r$ .*

The only maximal pair-wise independent subset of  $GF(2)^r$ , is  $V = GF(2)^r - \{0\}$  with  $|V| = 2^r - 1$ . Thus over  $GF(2)$ , a parity check matrix of any  $[n, k, 3]$  Hamming Code is simply a matrix whose rows are non-zero elements of

$GF(2)^r$ . From now on, we assume  $F = GF(2)$ . Algorithm below gives a method of getting a generating matrix  $G$  for a  $[n, k, 3]$  Hamming Code from its parity check matrix.

**Algorithm.** Let  $H$  be any parity check matrix of  $[n, k, 3]$  Hamming Code.

- (i) Permute the row of  $H$  to get  $H'$  of the form  $\begin{pmatrix} X \\ I_r \end{pmatrix}$ . Thus  $H' = P_f H$  where  $f$  is some element in  $S_n$ .
- (ii) Let  $G' = (I_k \ X)$ .
- (iii) Permute the column of  $G'$  according to  $f^{-1}$  to get  $G$ . Hence,  $G = G'P_f$ .

Obviously,

$$r(G) = r(G') = k \quad \text{and} \quad GH = (G'P_f)(P_f^{-1}H') = (I_k \ X) \begin{pmatrix} X \\ I_r \end{pmatrix} = X + X = \mathbf{0}.$$

Thus  $G$  is a generating matrix of the  $[n, k, 3]$  Hamming Code.

**Example 1.** Let  $GF(2^4)$  be constructed using the irreducible polynomial  $1 + x + x^4$  over  $F$ . Thus  $GF(2^4) = F[x] / \langle 1 + x + x^4 \rangle = \{0, \beta, \beta^2, \dots, \beta^{15} = 1\}$  and  $1 + x + x^4$  is the minimal polynomial of  $\beta$ . We write  $m_\beta(x) = 1 + x + x^4$ .

Let  $H$  be a parity check matrix of  $[3, 11, 15]$  cyclic Hamming Code having  $m_\beta(x) = 1 + x + x^4$  as its generator polynomial. Permute the rows of  $H$  according to

$$f = (1 \ 5 \ 9 \ 13 \ 2 \ 6 \ 10 \ 14 \ 3 \ 7 \ 11 \ 15 \ 4 \ 8 \ 12),$$

we get  $H' = \begin{pmatrix} X \\ I_4 \end{pmatrix}$ , where

$$X = \begin{pmatrix} \beta^4 \\ \beta^5 \\ \beta^6 \\ \vdots \\ \beta^{14} \end{pmatrix}.$$

Then we get  $G' = (I_{11} \ X)$ . Permute the column of  $G'$  according to  $f^{-1}$  and thus we get  $G = (X \ I_{11})$ . Note that  $r(G) = 11$  and  $GH = \mathbf{0}$ . Hence,  $G$  is a generating matrix

of the  $[3,11,15]$  Hamming code having  $H$  as parity check matrix. Therefore  $\forall m = (a_0, a_1, \dots, a_{10}) \in F^{11}$ ,  $m$  is encoded into

$$mG = \left( \sum_{i=0}^{10} a_i \beta^{i+4}, m \right).$$

Now we shall prove our main theorem regarding the check rows of  $[n, k, 3]$  Hamming Code.

**Theorem.** Assume  $r \geq 2$  and  $H$  is a parity check matrix for a  $[n, k, 3]$  Hamming Code. Let the rows of  $H$  be  $r_1, r_2, \dots, r_n$ . Then there is a generating matrix  $G$  for the code such that when encode using  $G$ , if the check positions of the codeword are in the  $i_1^{\text{th}}, i_2^{\text{th}}, \dots, i_r^{\text{th}}$  coordinates, then  $\{r_{i_1}, r_{i_2}, \dots, r_{i_r}\} = \{e_1, e_2, \dots, e_r\}$ .

*Proof.* Consider a parity check matrix of  $[n, k, 3]$  Hamming Code with the form  $H' = \begin{pmatrix} X \\ I_r \end{pmatrix}$ , where  $X$  is a  $k \times r$  matrix. Apparently  $G' = (I_k \ X)$  is a generating matrix of the code. When encoding using  $G'$ , the check positions are in the last  $r$  coordinates and the theorem is obviously true for this particular case.

Let  $H$  be any parity check matrix of the  $[n, k, 3]$  Hamming Code  $C$ . Then  $H$  could be obtained from  $H'$  by permuting the rows of  $H'$  according to some permutation in  $S_n$ . That is  $H = P_\theta H'$ , for some  $\theta \in S_n$  and thus  $G = G' P_{\theta^{-1}}$  is a generating matrix of  $C$ . While encoding using  $G$ , every message word  $m \in F^k$  is encoded into  $mG$ . Hence,

$$\begin{aligned} mG &= (mG') P_{\theta^{-1}} \\ &= (a_1, a_2, \dots, a_k, a_{k+1}, \dots, a_n) P_{\theta^{-1}} \\ &= (a_{\theta(1)}, a_{\theta(2)}, \dots, a_{\theta(k)}, a_{\theta(k+1)}, \dots, a_{\theta(n)}). \end{aligned}$$

$i_k$  is a check position (with respect to the generating matrix  $G$ )

$\Rightarrow a_{\theta(i_k)}$  is a parity check digit

$\Rightarrow a_{\theta(i_k)} = a_y$  for some  $y$ ,  $k+1 \leq y \leq n$

$\Rightarrow \mathcal{G}(i_k) = y$  for some  $y$ ,  $k+1 \leq y \leq n$

$\Rightarrow i_k^{\text{th}}$  row of  $H$  is the  $y^{\text{th}}$  row of  $H'$  for some  $y$ ,  $k+1 \leq y \leq n$

(as  $H = P_\theta H' \Rightarrow i_k^{\text{th}}$  row of  $H$  is the  $\mathcal{G}(i_k)^{\text{th}}$  row of  $H'$ )

$\Rightarrow i_k^{\text{th}}$  row of  $H$  has the form of  $e_i$  for some  $i$ ,  $1 \leq i \leq r$ .

Thus, we have proved the theorem.

Let  $H$  be a parity check matrix and  $G$  be a generating matrix of a  $[n, k, d]$  linear code where  $n - k = r$ . We named the  $i_1^{\text{th}}, i_2^{\text{th}}, \dots, i_r^{\text{th}}$  rows of  $H$  as the *check rows* if the  $i_1^{\text{th}}, i_2^{\text{th}}, \dots, i_r^{\text{th}}$  coordinates are the check positions of the code (with respect to  $G$ ). Therefore theorem above shows that given any parity check matrix  $H$  of a Hamming Code we can find a generating matrix such that when encoding using  $G$ , the check rows of the code are linearly independent.

### 3. The construction of a 2EC-AUED code

We shall now construct a code  $C$  of minimal distance 5. The message set  $M$  is a constant weight code of length  $k$  and each message word  $\mathbf{m}$  is encoded into a codeword of the form  $(\mathbf{y}, \mathbf{m}, \mathbf{u})$  where  $\mathbf{y}$  and  $\mathbf{u}$  are elements in  $F^t$  and  $F^p$  respectively.

Let us start by explaining how to get  $\mathbf{u}$  from  $\mathbf{m}$ . Let  $Q = \{\omega_0, \omega_1, \dots, \omega_{k-1}\}$  be an additive abelian group of order  $k$  with  $\omega_0$  its identity element and  $N$  be a constant weight code of length  $p$ . We assume  $|N| \geq k$ . For every  $\mathbf{m} = (m_0, m_1, \dots, m_{k-1}) \in M$ , we calculate  $\sum_{i=0}^{k-1} m_i \omega_i$  where

$$m_i \omega_i = \begin{cases} \omega_i & \text{if } m_i = 1 \\ \omega_0 & \text{if } m_i = 0. \end{cases}$$

Let

$$g : Q \longrightarrow N$$

be any one to one function. Then we define  $\mathbf{u} = g(\sum_{i=0}^{k-1} m_i \omega_i)$ . This method is due to Rao and Bose [5].

**Example 2.** Let  $M$  be a 3 out of 7 code;  $Q = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}\}$  be the additive group of  $GF(7)$  and  $N = \{\alpha_0 = 11000, \alpha_1 = 01100, \alpha_2 = 00110, \alpha_3 = 00011, \alpha_4 = 10001, \alpha_5 = 10100, \alpha_6 = 10010\}$ . a subset of the 2 out of 5 code. Define

$$g : Q \longrightarrow N$$

such that  $g(\bar{r}) = \alpha_r$ . Then for  $\mathbf{m} = 0101010 \in M$ , we get

$$\mathbf{u} = g\left(\sum_{i=0}^6 m_i \bar{i}\right) = g(\bar{0} + \bar{1} + \bar{0} + \bar{3} + \bar{0} + \bar{5} + \bar{0}) = g(\bar{2}) = \alpha_2 = 00110.$$

Next we describe how  $y$  is obtained from  $m$ . Let  $G$  be a generating matrix and  $H$  be a parity check matrix of a linear code having independent check rows. For every  $m = (m_0, m_1, \dots, m_{k-1}) \in M$ , we then permute the coordinates of  $mG$  so that the resulting word has the form  $(y, m)$ . Thus we get  $y$ .

**Example 3.** Let

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

be a generating matrix and parity check matrix of  $[3, 4, 7]$  Hamming Code having independent check rows. Let  $M = \{1100, 0110, 0011, 1001\}$ . Then  $\forall m = (a, b, c, d) \in M$ ,  $mG = (a+c+d, a+b+c, c, a+b+d, d, b, a)$ . Permute the coordinates of  $mG$  according to  $f^{-1} = (3 \ 4 \ 7 \ 5 \ 6)$  or compute  $(mG)P_f$  to get  $y$

$$\begin{array}{ccc} \begin{matrix} m \\ (a, b, c, d) \end{matrix} & \longrightarrow & \begin{matrix} mG \\ (a+c+d, a+b+c, c, a+b+d, d, b, a) \end{matrix} \\ & & \begin{matrix} \downarrow f^{-1} \\ (mG)P_f \\ (a+c+d, a+b+c, a+b+d, a, b, c, d) = (y, m) \end{matrix} \end{array}$$

Hence if  $m = 0101 \in M$ , then  $(0101)G = 1100110$  and thus  $(1100110)P_f = 1100101$  and we get  $y = 110$ .

Now we choose suitable  $Q, N, g, M, G, H$  to construct our code  $C$ . We take  $Q$  to be the additive group of  $GF(11)$ , write  $Q = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}, \bar{7}, \bar{8}, \bar{9}, \bar{10}\}$ ;  $N = \{\alpha_0 = 110000, \alpha_1 = 011000, \alpha_2 = 001100, \alpha_3 = 000110, \alpha_4 = 000011, \alpha_5 = 101000, \alpha_6 = 100100, \alpha_7 = 100010, \alpha_8 = 100001, \alpha_9 = 101000, \alpha_{10} = 010010\}$  a subset of the 2 out of 6 code and the one to one map  $g$  is defined as follow:

$$g : Q \longrightarrow N$$

such that  $g : (\bar{r}) = \alpha_r$ .

Let  $G$  and  $H$  be respectively the generating matrix and the parity check matrix of the [3,11,15] cyclic Hamming Code,  $C'$  given in Example 1. Let  $M'$  be the 5 out of 11 code.  $m \in M'$  is then encoded into a codeword  $c = (y, m, u)$  as explained before. For our chosen generating matrix  $G = (X \ I_{11})$ , we get  $y = mX$  as  $mG = (y, m)$ . Hence  $\forall c = (y, m, u) \in C$ ,  $(y, m) \in C'$ . It is clear from the choice of  $Q$  that  $\forall m = (m_0, m_1, \dots, m_{10}) \in M'$ , if  $\sum_{i=0}^{10} m_i \bar{i} = \bar{r}$ , then  $u = g(\bar{r}) = \alpha_r$ .

Unfortunately the code  $C$  encoded from  $M'$  has minimal distance less than 5. To increase its distance we choose the set of message words to be  $M$ , a subset of  $M'$  that satisfies a further condition, namely if  $a, a' \in M$ ,  $a \neq a'$ , which are encoded into  $(y, a, u)$  and  $(y', a', u')$  respectively, then  $u = u'$  and  $y = y'$  implies  $d(a, a') \geq 6$ . There exist many such subsets  $M$ , we exhibit one in appendix. Now let  $C$  be the code encoded from  $M$  instead of from  $M'$  and we shall prove  $d(C) = 5$ .

Let  $c = (y, a, u)$ ,  $c' = (y', a', u') \in C$ ,  $c \neq c'$  and thus  $a \neq a'$ .

**Case 1.**  $y \neq y'$  and  $u \neq u'$ :  $u \neq u'$  implies  $d(u, u') \geq 2$ . As  $(y, a), (y', a') \in C'$ ,  $d((y, a), (y', a')) \geq 3$ . Thus,  $d(c, c') \geq 3 + 2 = 5$ .

**Case 2.**  $y = y'$  and  $u \neq u'$ : Similar to Case 1,  $d(u, u') \geq 2$  and  $d((y, a), (y', a')) \geq 3$ . Thus  $d(c, c') \geq 3 + 2 = 5$ .

**Case 3.**  $y \neq y'$  and  $u = u'$ : We claim that  $d(a, a') \geq 4$ . Let  $a = (a_0, a_1, \dots, a_{10})$  and  $a' = (a'_0, a'_1, \dots, a'_{10})$ . Assume that  $d(a, a') = 2$  when  $u = u'$ . Then we have  $\sum_{i=0}^{10} a_i \bar{i} = \sum_{i=0}^{10} a'_i \bar{i}$ , where at position  $j$  and  $k$ ,  $j \neq k$ , we have  $a_j = 1$ ,  $a'_j = 0$  and  $a_k = 0$ ,  $a'_k = 1$ . This results in  $\bar{j} = \bar{k}$  for  $j \neq k$ , which is impossible. Hence if  $u = u'$  then  $d(a, a') \geq 4$ . Therefore  $d(c, c') \geq 4 + 1 = 5$  as  $d(y, y') \geq 1$ .

**Case 4.**  $y = y'$  and  $u = u'$ : Apparently  $d(c, c') \geq 5$  by the further condition satisfied by  $M$ .

Exhausting all possible cases, we see that  $C$  is of minimal distance five and thus is a 2 EC code.

Let  $w = (y', a, u')$  be any received word. As  $N$  and  $M$  are constant weight codes, the occurrences of unidirectional errors in  $a$  and  $u'$  are always detected by  $C$ . On the other hand, if unidirectional errors occur only in  $y'$  which is also the check positions of  $C'$ , then the errors can be detected by computing  $(y', a)H$ , since the check rows of  $C'$  are linearly independent. Therefore  $C$  is a 2 EC-AUED code. Below is the decoding algorithm of  $C$ .

**Decoding algorithm.** Assume  $w = (y', a, u')$  is received, where  $a = (a_0, a_1, \dots, a_{10})$ .

Let  $w(x)$  be the check polynomial of  $(y', a)$  and  $u'' = g(\sum_{i=0}^{10} a_i \bar{i})$ .

**Step 1.**  $\text{wt}(u') \neq 2$ .

- (i)  $\text{wt}(a) = 5$ :  $a$  is the transmitted message word.
- (ii)  $\text{wt}(a) \neq 5$ : If  $w(\beta) = \beta^i$ ,  $4 \leq i \leq 14$  then  $a + e_{i-3}$  is the decoded message word. Else we detect an uncorrectable error pattern in  $w$ .

**Step 2.**  $\text{wt}(u') \neq 2$ .

- (i)  $\text{wt}(a) = 5$  and  $u' = u''$ :  $a$  is the transmitted message word.
- (ii)  $\text{wt}(a) = 5$  and  $u' \neq u''$ :
  - (a)  $w(\beta) = 0$ :  $a$  is the transmitted message word,
  - (b)  $w(\beta) \neq 0$ : Find

$$Q = \left\{ m = (m_0, m_1, \dots, m_{10}) \in M \mid g\left(\sum_{i=0}^{10} m_i \bar{i}\right) = u' \text{ and } mX = y' \right\}.$$

If  $\exists m \in Q$  such that  $d(m, a) = 2$ , then  $m$  is the decoded message word.

Else we detect an uncorrectable error pattern in  $w$ .

- (iii)  $\text{wt}(a) = 5 \pm 1$  and  $g^{-1}(u') \neq \{\}$ : Compute

$$s = \pm \left[ \left( \sum_{i=0}^{10} a_i \bar{i} \right) - g^{-1}(u') \right].$$

If  $s = \bar{i} - 1$ , then  $a + e_i$  is the decoded message word. Else an uncorrectable error pattern detected in  $w$ .

- (iv)  $\text{wt}(a) = 5 \pm 2$ : Find

$$Q = \left\{ m = (m_0, m_1, \dots, m_{10}) \in M \mid g\left(\sum_{i=0}^{10} m_i \bar{i}\right) = u' \text{ and } mX = y' \right\}.$$

If  $\exists m \in Q$  such that  $d(m, a) = 2$ , then  $m$  is the decoded message word. Else we detect an uncorrectable error pattern in  $w$ .



**Step 3.** Other conditions besides Steps 1 and 2, we detect uncorrectable error pattern in  $w$ .

The decoding algorithm given above is capable of correcting  $t$  errors if  $t \leq 2$ . Suppose no error occurred in  $w$ . Then  $\text{wt}(\mathbf{a}) = 5$ ,  $\text{wt}(\mathbf{u}') = 2$  and  $\mathbf{u}' = \mathbf{u}''$ . Hence Step 1 of the decoding algorithm fails and we go on to Step 2. We get  $\mathbf{a}$  as the decoded message word according to Step 2(i).

Assume that an error has occurred in  $w$ . If the error is in  $\mathbf{y}'$ , then  $\text{wt}(\mathbf{a}) = 5$ ,  $\text{wt}(\mathbf{u}') = 2$  and  $\mathbf{u}' = \mathbf{u}''$ . Again Step 1 fail and we get  $\mathbf{a}$  as the decoded message word according to Step 2(i). However if the error is in the  $k^{\text{th}}$  position of  $\mathbf{a}$ , then  $\text{wt}(\mathbf{a}) = 5 \pm 1$  with  $\text{wt}(\mathbf{u}') = 2$ . Obviously

$$s = \pm \left[ \left( \sum_{i=0}^{10} a_i \bar{i} \right) - g^{-1}(\mathbf{u}') \right] = \overline{k-1}.$$

Thus, according to Step 2(iii),  $\mathbf{a} + \mathbf{e}_k$  is the decoded message word. If the error happened to be in  $\mathbf{u}'$ , we get  $\text{wt}(\mathbf{u}') = 2 \pm 1$  with  $\text{wt}(\mathbf{a}) = 5$ . Then Step 1(i) in the decoding algorithm propose  $\mathbf{a}$  to be the decoded message word.

Assume that double errors have occurred in  $w$  during the transmission. If the errors occurred in

- (i)  $\mathbf{y}'$  and the  $j^{\text{th}}$  position of  $\mathbf{a}$ , then  $\text{wt}(\mathbf{u}') = 2$  and  $\text{wt}(\mathbf{a}) = 5 \pm 1$ . Apparently  $s = \pm \left[ \left( \sum_{i=0}^{10} a_i \bar{i} \right) - g^{-1}(\mathbf{u}') \right] = \overline{j-1}$  and thus  $\mathbf{a} + \mathbf{e}_j$  is the decoded message word according to Step 2(iii).
- (ii)  $\mathbf{y}'$  and  $\mathbf{u}'$ , then  $\text{wt}(\mathbf{u}') = 2 \pm 1$  together with  $\text{wt}(\mathbf{a}) = 5$ . According to Step 1(i),  $\mathbf{a}$  is the decoded message word.
- (iii)  $\mathbf{u}'$  and the  $j^{\text{th}}$  position of  $\mathbf{a}$  (which is the  $(j+4)^{\text{th}}$  position of  $(\mathbf{y}', \mathbf{a}) \in C'$ ), then  $\text{wt}(\mathbf{u}') = 2 \pm 1$ ,  $\text{wt}(\mathbf{a}) = 5 \pm 1$  and  $w(\beta) = \beta^{(j+4)-1}$ . Let  $\beta^i = \beta^{(j+4)-1}$ . Apparently  $4 \leq i \leq 14$ . Thus  $\mathbf{a} + \mathbf{e}_{i-3}$  is the decoded message word according to Step 1(ii).

For cases where double errors have occurred simultaneously in  $\mathbf{y}'$ ,  $\mathbf{a}$  or  $\mathbf{u}'$ , the errors involved might be of unidirectional or symmetry types. Assume that two errors have occurred in  $\mathbf{y}'$ . Whether the errors are of unidirectional or symmetry type, we get  $\text{wt}(\mathbf{a}) = 5$ ,  $\text{wt}(\mathbf{u}') = 2$  and  $\mathbf{u}' = \mathbf{u}''$ . Hence for both cases, we take  $\mathbf{a}$  as the decoded message word as proposed in Step 1(i).

If the errors are in  $\mathbf{u}'$ , then  $\text{wt}(\mathbf{u}') = 2 \pm 2$  if the errors are of unidirectional type or  $\text{wt}(\mathbf{u}') = 2$  if the errors are of symmetry type with  $\text{wt}(\mathbf{a}) = 5$ . In the first case,  $\mathbf{a}$  will be taken as decoded message word as given in Step 1(i). For the later case, the two

symmetry errors in  $\mathbf{u}'$  will cause  $\mathbf{u}' \neq \mathbf{u}''$  with  $w(\beta) = \mathbf{0}$ . Hence by Step 2(ii)(a) in the decoding algorithm  $\mathbf{a}$  is the decoded message word.

Suppose both errors are in  $\mathbf{a}$ , then  $\text{wt}(\mathbf{a}) = 5 \pm 2$  if the errors are unidirectional errors or  $\text{wt}(\mathbf{a}) = 5$  if the errors are symmetry errors with  $\text{wt}(\mathbf{u}') = 2$ . The two symmetry errors in  $\mathbf{a}$  will cause  $\mathbf{u}' \neq \mathbf{u}''$  with  $w(\beta) \neq \mathbf{0}$ . Thus the errors can be corrected by Step 2(ii)(b) in the decoding algorithm. For the remaining case, we will use Step 2(iv) which is similar to Step 2(ii)(b) to correct the errors.

Refer to Steps 2(ii)(b) and 2(iv), we now describe how to construct the set

$$Q = \left\{ \mathbf{m} = (m_0, m_1, \dots, m_{10}) \in M \mid g \left( \sum_{i=0}^{10} m_i \bar{i} \right) = \mathbf{u}' \text{ and } \mathbf{m}X = \mathbf{y}' \right\}.$$

We partition  $M$  into a number of equivalent classes, each denoted by  $V_{ij}$  for  $i = 0, 1, 2, \dots, 15$  and  $j = 0, 1, 2, \dots, 10$  using two equivalent relations,  $S$  and  $Z$  as given below

$$\forall \mathbf{a}, \mathbf{b} \in M, \quad \mathbf{a}S\mathbf{b} \quad \text{if and only if} \quad \mathbf{a}X = \mathbf{b}X.$$

Obviously  $S$  is an equivalent relation and thus  $M$  is partitioned into 16 equivalent classes, denoted by  $V_0, V_1, \dots, V_{15}$ , where

$$V_i = \{ \mathbf{a} \in M \mid \mathbf{a}X \text{ is the binary representation of integer } i \}.$$

Let  $Z$  be an equivalent relation defined on  $V_i$  such that  $\forall \mathbf{a} = (a_0, a_1, \dots, a_{10}), \mathbf{b} = (b_0, b_1, \dots, b_{10}) \in V_i$ ,

$$\mathbf{a}Z\mathbf{b} \quad \text{if and only if} \quad \sum_{k=0}^{10} a_k \bar{k} = \sum_{k=0}^{10} b_k \bar{k}.$$

Apparently  $Z$  is an equivalent relation on  $V_i \forall i$ . Thus, each  $V_i$  can be further partitioned into 11 equivalent classes, denoted by  $V_{ij}$ ,  $j = 0, 1, 2, \dots, 10$  where  $V_{ij} = \{ (a_0, a_0, \dots, a_{10}) \in V_i \mid \sum_{k=0}^{10} a_k \bar{k} = \bar{j} \}$ . Therefore

$$V_i = \bigcup_{j=0}^{10} V_{ij} \quad \text{and} \quad M = \bigcup_{i=0}^{15} V_i = \bigcup_{i=0}^{15} \bigcup_{j=0}^{10} V_{ij}.$$

Thus for a received word  $w = (y', a, u')$ , if  $y'$  is the binary form of integer  $i$  and  $g^{-1}(u') = \bar{j}$ , then  $Q = V_{ij}$ . The list of all elements in each  $V_{ij}$ ,  $i = 0, 1, 2, \dots, 15$  and  $j = 0, 1, 2, \dots, 10$  is given in appendix.

Assume that there exist  $m_r, m_s \in V_i$ ,  $m_r \neq m_s$  such that  $d(m_r, a) = d(m_s, a) = 2$ . Then  $d(m_r, m_s) \leq d(m_s, a) + d(m_r, a) = 4$ , which is a contradiction as each  $V_{ij}$  is chosen to be a distance 6 constant weight code. Hence if two errors have occurred in  $a$ , there is an unique  $m \in V_{ij}$  such that  $d(a, m) = 2$ .

**Example 4.** Assume that  $w = 0000\ 01100111000\ 011000$  is received. Let  $y' = 0000$ ,  $a = 01100111000$  and  $u' = 011000$ . Note that  $wt(a) = 5$ ,  $wt(u') = 2$  and  $u' \neq u''$  as  $u'' = \alpha_{10} \left( \sum_{i=0}^{10} a_i \bar{i} = \bar{0} + \bar{1} + \bar{2} + \bar{0} + \bar{0} + \bar{5} + \bar{6} + \bar{7} + \bar{0} + \bar{0} + \bar{0} = \bar{10} \right)$  and  $u' = \alpha_1$ . Since

$$\begin{aligned} w(\beta) &= \beta^5 + \beta^6 + \beta^9 + \beta^{10} + \beta^{11} \\ &= 0110 + 0011 + 0101 + 1110 + 0111 \\ &= 1001 \neq 0, \end{aligned}$$

$a$  is compared to each message word in  $V_{01}$  by Step 2(ii)(b) (as  $y'$  is the binary representation of integer 0 and  $g^{-1}(u') = \bar{1}$ ). From appendix, we get  $d(01001111000, a) = 2$  and thus 01001111000 is the decoded message word according to Step 2(ii)(b).

Assume that  $w = 0000\ 01011111000\ 011000$  is received. Let  $y' = 0000$ , then  $a = 01011111000$  and  $u' = 011000$ . Note that  $wt(a) = 6$  and  $wt(u') = 2$ . By Step 2(iii), compute  $s = \pm \left[ \left( \sum_{i=0}^{10} a_i \bar{i} \right) - g^{-1}(u') \right]$ . As  $g^{-1}(u') = \bar{1}$  and  $\sum_{i=0}^{10} a_i \bar{i} = \bar{0} + \bar{1} + \bar{0} + \bar{3} + \bar{4} + \bar{5} + \bar{6} + \bar{7} + \bar{0} + \bar{0} + \bar{0} = \bar{4}$ ,  $s = \bar{4} - \bar{1} = \bar{3}$ . Thus  $a + e_4 = 01001111000$  is the decoded message word according to Step 2(iii).

Assume that  $w = 0000\ 01001111010\ 010000$  is received. Let  $y' = 0000$ , then  $a = 01001111010$  and  $u' = 010000$ . Note that  $wt(a) = 6$  and  $wt(u') = 1$ . Obviously two errors have occurred separately in  $a$  and  $u'$ . As

$$\begin{aligned} w(\beta) &= \beta^5 + \beta^8 + \beta^9 + \beta^{10} + \beta^{11} + \beta^{13} \\ &= 0110 + 1010 + 0101 + 1110 + 0111 + 1011 = 1011 \\ &= 1011 = \beta^{13}. \end{aligned}$$

Then  $a + e_{10} = 01001111000$  is the decoded message word according to Step 1(ii).

## 4. Conclusion

We make a few remarks to conclude this paper.

- (i) The information rate of the code we constructed is 0.4048, which is good compare to most commonly used codes.
- (ii) The main theorem we proved is also true over any arbitrary finite field.
- (iii) The check rows of Golay code,  $C_{23}$ , are also independent. Using our method, a 4EC-AUED code could be constructed.
- (iv) We do not have an efficient algorithm to compute  $V_{ij}$ . This may be a future research problem.

## References

1. D.A. Anderson, *Design of Self-checking Digital Networks Using Coding Techniques*, Univ. Illinois, Urbana, CSL Rep. R-527, 1971.
2. M. Blaum and H.C.A. Van Tilborg, On  $t$ -error correcting/all unidirectional error detecting codes, *IEEE Trans. Comp.* **C-38** (1989), 1493-1501.
3. F.J.H. Bonick and H.C.A. Van Tilborg, Constructions and bounds for systematic  $t$ -EC/AUED codes, *IEEE Trans. Inf. Theory* **IT-36** (1990), 1381-1390.
4. B. Bose and D.K. Pradhan, Optimum unidirectional error detecting/correcting codes, *IEEE Trans. Comp.* **C-31** (1982), 521-530.
5. B. Bose and T.R.N. Rao, Theory of unidirectional error correcting/detecting codes, *IEEE Trans. Comp.* **C-31** (1982), 564-568.
6. R.W. Cook, W.H. Sisson, T.G. Stoney and W.N. Toy, Design of self-checking microprogram control, *IEEE Trans. Comp.* **C-22** (1973), 225-262.
7. R.W. Hamming, Error detecting and error correcting codes, *Bell Syst. Tech. J.* **29** (1950), 147-160.
8. I.N. Herstein, *Topic in Algebra*, Wiley, New York, 1975.
9. G.A. How, *Algebra with Coding Theory*, Tung Hua Book Co., Taiwan, 1996. (in Chinese)
10. M.C. Lin, Constant weight codes for correcting symmetry errors and detecting unidirectional errors, *IEEE Trans. Comp.* **42** (1993), 1294-1302.
11. D. Nilolos, N. Gaitanis, and G. Philokyprou, Systematic  $t$ -error correcting/all unidirectional error detecting codes, *IEEE Trans. Comp.* **C-35** (1986), 394-402.
12. B. Parhami and A. Avizienis, Detection of storage errors in mass memories using low-cost arithmetic error codes, *IEEE Trans. Comp.* **C-27** (1978), 302-308.
13. D.K. Pradhan and J.J. Stiffler, Error correcting codes and self-checking circuits in fault-tolerant computers, *Computer* (1980), 27-37.
14. R.W. Sahni, Reliability of integration circuits, *Proceedings of the IEEE International Computers Group Conference*, Washington (1970), 213-219.
15. N.J.A. Sloane and F.J. MacWilliam, *The Theory of Error Correcting Code*, North-Holland, Amsterdam, The Netherlands, 1978.
16. D.L. Tao, C.R.P. Hartmann, and P.K. Lala, An efficient class of unidirectional error detecting/correcting codes, *IEEE Trans. Comp.* **C-37** (1988), 879-882.
17. T.M. Thompson, From error-correcting codes through sphere packings to simple group, *The Carus Math. Monographs* **21**, The Math. Assoc. of America, 1983.

**Appendix**

$$M = \bigcup_{i=0}^{15} \bigcup_{j=0}^{10} V_{ij}, \quad i=0,1,\dots,15 \quad \text{and} \quad j=0,1,\dots,10. \quad V_i = \bigcup_{j=0}^{10} V_{ij}, \quad i=0,1,\dots,15.$$

<i>j</i>	<b>V<sub>0j</sub></b>	4	10010011001	10	-
0	00010011110	5	01010010110	<i>j</i>	<b>V<sub>3j</sub></b>
1	01001111000	6	01000111010	0	11000101010
2	11000010110		00011011100		10011001100
	10110000011		10001100011		01101100001
3	10001011100	7	11100001100	1	00001110011
	10010100110		10000110101		01101001010
4	01010101001	8	01110011000	2	10101000101
	00111000110		00101010101		11110001000
5	00001001111		10000011110		10010110001
	11100011000	9	10011100100	3	10010011010
6	00100111100	10	01101100010		00111010001
	10010001101				00110101100
7	01100001011	<i>j</i>	<b>V<sub>2j</sub></b>	4	00011110100
	10011110000	0	10100101100	5	11100100100
8	01000101110		00011001011		10100010011
	01001010011		10101010001		01011000011
9	00011100011	1	00110111000	6	10000110110
	11010010001		10001110100		01110110000
10	01110001100	2	11001000011	7	10111000010
	00010110101	3	01010101010		01000111001
	10101010010		00001100111	8	10001001011
			11100110000		10110010100
<i>j</i>	<b>V<sub>1j</sub></b>	4	10010100101	9	10100111000
0	11010000101	5	01100100011		01011101000
	01100110100		00111000101	10	11010000110
1	11000101001		10010001110	<i>j</i>	<b>V<sub>4j</sub></b>
	10010110010	6	00101101001	0	01110001010
	10101000110	7	10100000111		00101000111
2	01101001001	8	01110100100		00010110011
	00111010010		00110010011		11001000110
3	01001101100	9	01000101101	1	10110000101
	00001011011		00010110110		01011010010
			10011011000		

## Appendix (Cont'd)

2	10100101001	8	10000110011	2	10010011100
3	01100100110		11100001010		01100110001
	11101001000	9	00101010011	3	10101000011
	10001110001		01101100100		01100011010
4	00101101100	10	00001110110	4	10001100110
	10001011010		10110010001		10100010101
5	-				01011000101
					00110101010
6	00110010110	$j$	$V_{6j}$	5	00011110010
	01100001101	0	10110000110		01001101001
7	00100111010	1	01110001001	6	10111000100
	10010001011		11001000101		11000000111
	01001010101		10100101010	7	10001001101
8	00011100101	2	00010011011		01010010011
			01010101100		10101101000
9	10101010100		10001110010	8	00011011001
	00011001110	3	-		
10	01110100001	4	01100100101	9	10110010010
	10000100111		11000010011		11100001001
			10111010000	10	10000011011
$j$	$V_{5j}$	5	10001011001		11000101100
0	10010110100		00000110111	$j$	$V_{8j}$
			01100001110	0	11000110001
1	00111010100	6	00111000011		
	01000010111		01001010110	1	01100101100
	01100110010	7	00110010101		00100011011
2	00101111000		00011100110		11000011010
	10011001001	8	00100111001	2	00000111110
	00010100111				11011100000
3	01011000110	9	01010000111	3	00111001010
	10100010110		01101011000	4	00110011100
4	01001101010	10	01000101011		10100100011
	01100011001		00111101000	5	01010100101
5	10001100101	$j$	$V_{7j}$		11101000010
	01000111100	0	00001110101	6	01010001110
	00110101001		01010111000		00101100110
6	10001001110		01101001100		11100010100
	01111001000	1	10011001010	7	10110100100
	00011110001		00001011110		
7	00011011010				
	11100100001				

Appendix (Cont'd)

8	10000101101 01100000111	2	00110110100 00001101011 11000011001	8	11010100001
9	11001001100 00010111001	3	01110000011 00000111101	9	10001000111 01000011110 00110001011
10	01011011000	4	00111001001 01010100110 01001011100	10	10100110100 00010101110 00011010011
<i>j</i>	$V_{9j}$	<i>j</i>	$V_{11j}$	<i>j</i>	$V_{12j}$
0	00010101101 10101001010	0	11100000011	0	00001101110 10110001001 11000011100
1	10100011100 01011001100	1	10101001001 11000100110	1	01110000110 10010101100 10011010001
2	11000100101	2	10001101100 01010110010 01101000110	2	01100101010 00111001100
3	01101000101 11000001110 01010110001	3	11110000100 00011111000	3	01001110010 10100100101
4	01010011010 00101110010 10011000011	4	10010010110 11100101000	4	10100001110 00110110001
5	10010010101	5	10000111010 01010011001 00101110001	5	00110011010 10001010110
6	10000111001 01100010011	6	00101011010	6	01010100011 01001011001
7	00101011001 01000110110 00110100011	7	10110011000 01001100011	7	11100010010 00011101001
8	00001111100 01111000010	8	01001101100 01010110010 01101000110	8	00100110110 10010000111 10101011000
9	01001001011 01110010100	9	10010010110 11100101000	9	10000101011
10	01100111000 11010001001 10101100001	10	10000111010 01010011001 00101110001	10	00101001011
<i>j</i>	$V_{10j}$	<i>j</i>	$V_{13j}$	<i>j</i>	$V_{13j}$
0	10001111000 01101010010	0	001101011000 01000011011	0	00111011000 01000011011
1	00011000111 11110010000	1	10110011000 01001100011	1	01011100001 10100110001

## Appendix (Cont'd)

		$j$	$V_{14j}$	$j$	$V_{15j}$
2	01011001010 11101010000 10100011010	0	01000100111 11001001001 00111100100 10011010010	0	01110010001 01011100010 10100110010
3	00101110100 10011000101 11000100011 01010011100	1	00001101101	1	01010110100
4	10001101001 01101000011	2	01110000101 10100100110	2	10011000110
5	01001100110 11110000001	3	00010010111 01100101001	3	01011001001 10100011001 10001101010
6	11010100100 10010010011	4	01001110001	4	01100010110 10000111100
7	01111000100	5	10100001101 01001011010	5	11000001011 10111001000 00110100110
8	01101101000 00011010110	6	00110011001 10001010101 00011101010	6	01001100101
9	00001111010 11100000110	7	00010111100	7	01001001110
10	10101001100 01110010010 10010111000	8	01010001011 00101100011 11100010001	8	11010001100 10101100100
		9	10110100001 11000110100	9	01000110011 00111110000
		10	00100011110 11111000000	10	11100000101 00001111001