講 座

符號理論의 槪念

線形符號篇

李 晩 榮
漢陽大學校 敎授

符號理論(coding theory)은 그 內容이 매우 廣範圍할 뿐 아니라 群論, 環論, 体論 等 抽象代數學과 確率論 및 數理統計學을 背景으로 發展된 學問이기 때문에 一般 讀者를 相對로 論述하기에는 적지 않은 難點이 있다. 그렇다고 單純히 用語羅列에만 그칠 수도 없고, 理論에 置重한 論文式으로 쓸수도 없으므로 大學 4年生을 爲한 講義水準으로 紹介하겠으며 3回에 걸쳐 線形符號(linear code), 巡回符號(cyclic code), 길쌈符號(convolutional code)의 順으로 連載하기로 하겠다.

Ⅰ. 序 論

高度로 發達된 現代社會에서는 보다 效率的이고 信賴할 수 있는 情報傳達시스템이 絶對的으로 必要하다. 그러나 通信系에서 避할 수 없는 問題中의 하나가 通信路上의 雜音에 依한 誤謬(error)이다.

그림 1은 誤謬訂正符號를 利用한 디지탈通信系의 典型的인 모델이다. 情報源에서 나오는 情報는 2元系列(binary sequence)로 變換되어 符號器(encoder) 에 入力되며, 符號器는 情報의 保護와 誤謬의 檢出및 訂正을 爲한 檢査長(parity-check bit), 即 冗長(redundancy)을 情報에 添加함으로써 情報를 符號化한다. 이렇게 符號化된 符號語는 變調器에 依해 傳送路에 適合한 "1"과 "0"에 對應하는 波形(pulse signal)으로 變換되어 傳送路로 送出된다. 通信路라 함은 變調器, 傳送路, 復調器를 包含한 그림 1의 點線部分을 뜻하며 普通 "1"의 波形은 $s_1(t)=A\cos\omega_0 t$, "0"은 $s_0(t)=-s_1(t)=A\cos(\omega_0 t+180°)$로 表示되는 coherent PSK 傳送方式이 使用된다. 通信路에는 固有雜音(fixed noise)과 干涉(interference) 및 歪曲(distortion) 等이 存在하며 이런 것들이 正確한 情報傳送에 妨害가 되는데 이런 妨害를 總稱해서 雜音이라 한다. 雜音이 섞인 受信波形은 復調器에 依해서 冗長이 包含된 2元系列로 復調되며 復號器(decoder)는 受信系列로부터 얻어지는 誤謬症候, 即 誤症(syndrome)을 使用하여 誤謬를 檢出, 訂正하게 된다. 이렇게 誤謬가 除去된 2元系列中 冗長部分을 除外한 情報系列만을 使用者(data sink)에게 供給함으로써 信賴性 있는 情報傳達이 可能하게 된다.

符號는 크게 나누어 block 符號와 convolutional 符號로 大別되는데 이 두 種類의 符號에 對한 符號器間의 顯著한 差異는 記憶(memory)의 存在有無에 있다. 槪念的으로 이 差異를 살펴보면 block 符號의 符號器

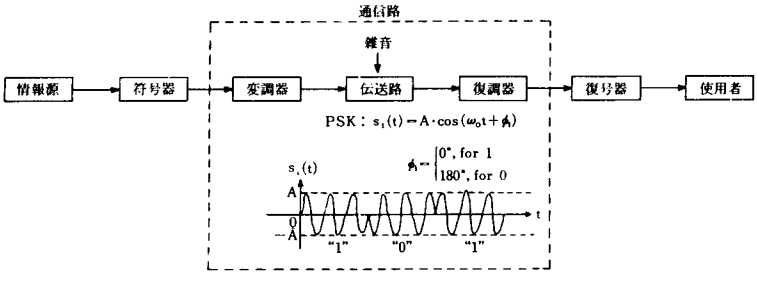

그림 1. 디지탈 通信系

는 k個 비트 情報를 n個 비트 符號語로 만들어 내는 無記憶裝置인 反面에 convolutional 符號의 符號器는 그 出力系列이 現在뿐만 아니라 그 以前 入力系列의 影響을 받아 決定되므로 記憶이 있는 裝置로 看做된다.

Block 符號는 線形符號와 巡回符號로 나누어지는데, 一般的으로 n, k, R=k/n 및 d_{min} 等의 媒介變數(parameter)가 있다. n은 符號長(code length), 即 符號器의 出力系列의 비트數를 말하고, 情報長(information bits), k는 情報비트의 數를 뜻하는데 그 實用値는 3에서부터 數百비트까지 이르고 符號比率(code rate), R=k/n은 大槪 1/4≦R≦7/8의 限界內에 있으며 傳送能率(transmission rate)을 指稱한다. 그리고 n−k 비트의 檢査長은 情報를 保護하기 爲해 挿入한 冗長이며 d_{min}은 符號의 誤謬檢出 및 訂正可能性을 알려주는 Hamming 最小距離를 말한다.

Ⅱ. Galois体와 符號

線形符號는 block 符號中 極히 一部를 차지하는 符號를 말하는 것으로 符號語(code word)의 集合이 線形벡터空間의 構造를 形成하기 때문에 符號語를 符號벡터, 또는 符號系列이라고도 한다.

一般的으로 Galois体 GF(q)의 元 q를 記號(symbol)로 擇하여 符號를 構成할 수 있으나 q=2인 2元体, GF(2)의 元 "0" 또는 "1"을 記號로 삼아 生成된 2進符號 或은 2元符號(binary code)가 널리 使用되고 있다. 그 理由는 現在 데이타通信系 및 컴퓨터記憶系에 收用되는 情報는 모두 2元符號로 表示되기 때문이다.

그러나 Galois体 GF(q)는 任意의 位數 q에 對해 恒常 存在하는 것이 아니고 q가 어떤 素數 p의 m乘인 p^m으로 表示될 때만 存在하므로 GF(p^m)으로 表示할 수 있고 特히 元의 數 또는 位數가 q=2일때 線形符號를 群符號(group code)라 부르기도 한다.

2元体 GF(2)上의 m次 原始多項式(primitive polynomial) p(x)의 根을 α라 하면 그 擴大体 GF(2^m)의 元은 $\{0, 1, \alpha, \cdots, \alpha^{2^m-2}\}$가 으로 되며, 이 講述에서는 2元体 GF(2)의 元으로 構成되는 2元符號와 그 擴大体 GF(2^m)의 元으로 構成되는 群符號에 對해서 說明하기로 한다.

情報源의 出力系列을 情報長이 k가 되도록 同一한 블럭으로 나누면 總 2^k個의 서로 다른 情報블럭(information block) $\bar{d}$를 얻게 되는데 이 情報 $\bar{d}$를 符號化(encoding)하여 符號語 $\bar{c}$를 얻기 爲해 符號器(encoder)를 使用한다. 符號器는 情報長 k인 入力情報 $\bar{d}$를 符號長 n인 符號語 $\bar{c}$로 變換시킴으로써 block符號를 形成하게 되는데, 一般的으로 k<n과 R<1의 關係가 있다. $\bar{d}$와 $\bar{c}$는 1對1로 對應되므로 2^k個의 入力情報는 符號器에 依해 亦是 2^k個의 符號語로 만들어진다. 即 符號器는 各各의 情報블럭 $\bar{d}$에 n−k個의 패리티檢査비트(parity-check bit)를 添加함으로써 符號벡터 $\bar{c}$를 生成해 내는 것이다. 이 檢査비트에는 情報가 全혀 包含되어 있지 않고 다만 傳送途中 符號에 發生하는 誤謬를 檢出, 訂正하는 能力만 있을 뿐이다.

이렇게 해서 生成된 符號長 n인 符號語 $\bar{c}=(c_0, c_1, c_2, \cdots, c_{n-1})$의 集合 C를 符號(code)라 한다. 一定한 길이 k를 가진 情報를 block 單位로 符號化하였으므로 線形 block符號(linear block code)라 하며 各 符號語가 同一한 길이 n을 가지므로 等長符號라고도 한다.

Ⅲ. 符 號 化

情報, $\bar{d}=(d_0, d_1, \cdots, d_{k-1})$를 符號語, $\bar{c}=(c_0, c_1, \cdots, c_{n-k-1}, c_{n-k}, \cdots, c_{n-1})$로 符號化하는 경우를 생각해보자.

符號語 $\bar{c}$를 다음과 같이 羅列하면

$$\bar{c}=(c_0, c_1, \cdots\cdots, c_{n-k-1}, c_{n-k}, \cdots\cdots, c_{n-1})$$
$$=(\gamma_0, \gamma_1, \cdots\cdots, \gamma_{n-k-1}, d_0, d_1, \cdots\cdots, d_{k-1})$$
$$\quad |\leftarrow n-k \text{ 檢査長} \rightarrow| \leftarrow k \text{ 情報長} \rightarrow|$$

情報 $\bar{d}$를 그대로 나타내는 右側 k비트와 檢査長을 이루는 左側 n−k비트로 完全 分離되어 組織構造를 이루게 되므로 이와 같이 符號化된 符號를 (n, k) 組織符號(systematic code)라 한다.

情報 $\bar{d}$는 生成行列(generator matrix) $\bar{G}$에 依해서 符號化되며 符號語는

$$\bar{c}=\bar{d}\cdot\bar{G}. \qquad (1)$$

로 表示된다. 여기서 生成行列 $\bar{G}$는 GF(2) 上의 k×n 2元行列이며 다음과 같이 表現된다.

$$\bar{G}=[\bar{P}_{k\times(n-k)} : \bar{I}_k]$$
$$=\begin{bmatrix} p_{00} & p_{01} & \cdots p_{0,(n-k-1)} & 1\,0\,0\,0\cdots 0 \\ p_{10} & p_{11} & \cdots p_{1,n-k-1} & 0\,1\,0\,0\cdots 0 \\ \vdots & & & \\ p_{k-1,0} & p_{k-1,1} \cdots & p_{k-1,n-k-1} & 0\,0\,0\,0\cdots 1 \end{bmatrix}$$

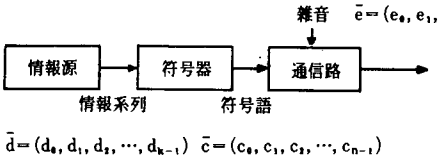

그림 2. 符號化 過程

$$= \begin{bmatrix} \bar{g}_0 \\ \bar{g}_1 \\ \vdots \\ \bar{g}_{n-k} \end{bmatrix} \qquad (2)$$

여기서 $\bar{I}_k$는 k×k 單位行列(identity matrix)이다. 따라서 (1)式은

$$\bar{c} = (d_0, d_1, d_2, \cdots, d_{k-1}) \cdot \begin{bmatrix} \bar{g}_0 \\ \bar{g}_1 \\ \vdots \\ \bar{g}_{k-1} \end{bmatrix}$$

$$= d_0\bar{g}_0 + d_1\bar{g}_1 + \cdots\cdots + d_{k-1}\bar{g}_{k-1} \qquad (3)$$

로 쓸 수 있으므로 (n, k)線形符號는 生成行列 $\bar{G}$의 各 行 $\bar{g}_i, 0 \leqq i \leqq k-1$, 의 線形結合에 依해서 生成됨을 알 수 있다.

例 1. (7, 4)線形符號가 다음과 같은 生成行列 $\bar{G}$를 갖는다고 하자.

$$\bar{G} = \begin{bmatrix} \bar{g}_0 \\ \bar{g}_1 \\ \bar{g}_2 \\ \bar{g}_3 \end{bmatrix} = \begin{bmatrix} 1\,1\,0\,1\,0\,0\,0 \\ 0\,1\,1\,0\,1\,0\,0 \\ 1\,1\,1\,0\,0\,1\,0 \\ 1\,0\,1\,0\,0\,0\,1 \end{bmatrix}$$

符號化할 情報가 $\bar{d} = (0110)$이면 그에 對應하는 符號語는 (3)式에 依해

$$\bar{c} = d_0\bar{g}_0 + d_1\bar{g}_1 + d_2\bar{g}_2 + d_3\bar{g}_3$$
$$= 0\cdot\bar{g}_0 + 1\cdot\bar{g}_1 + 1\cdot\bar{g}_2 + 0\cdot\bar{g}_3$$
$$= (0110100) + (1110010)$$
$$= (1000110)$$

가 되며 또 情報가 $\bar{d} = (1010)$일 때는

$$\bar{c} = d_0\bar{g}_0 + d_2\bar{g}_2$$
$$= (1101000) + (1110010)$$
$$= (0011010)$$

로 된다. 이와 같은 方法으로 $2^4 = 16$個의 情報와 그에 對應되는 符號語를 모두 求하면 표 1과 같다.

(7, 4)線形符號는 符號長이 n=7이고 情報長이 k=4이므로 (1)式에 依해

$$\bar{c} = (d_0, d_1, d_2, d_3) \cdot \begin{bmatrix} 1\,1\,0\,1\,0\,0\,0 \\ 0\,1\,1\,0\,1\,0\,0 \\ 1\,1\,1\,0\,0\,1\,0 \\ 1\,0\,1\,0\,0\,0\,1 \end{bmatrix}$$

$$= (c_0, c_1, c_2, c_3, c_4, c_5, c_6)$$

로 表示되며 符號語의 各 비트를 다음과 같이 情報및 檢査비트로 나누어 求할 수 있다.

$$\begin{aligned} \text{檢査비트}: c_0 &= d_0 + d_2 + d_3 \\ c_1 &= d_0 + d_1 + d_2 \\ c_2 &= d_1 + d_2 + d_3 \end{aligned} \qquad (4)$$

표 1. (7, 4)線形符號의 符號語

情報	符號語
0 0 0 0	0 0 0 0 0 0 0
0 0 0 1	1 0 1 0 0 0 1
0 0 1 0	1 1 1 0 0 1 0
0 0 1 1	0 1 0 0 0 1 1
0 1 0 0	0 1 1 0 1 0 0
0 1 0 1	1 1 0 0 1 0 1
0 1 1 0	1 0 0 0 1 1 0
0 1 1 1	0 0 1 0 1 1 1
1 0 0 0	1 1 0 1 0 0 0
1 0 0 1	0 1 1 1 0 0 1
1 0 1 0	0 0 1 1 0 1 0
1 0 1 1	1 0 0 1 0 1 1
1 1 0 0	1 0 1 1 1 0 0
1 1 0 1	0 0 0 1 1 0 1
1 1 1 0	0 1 0 1 1 1 0
1 1 1 1	1 1 1 1 1 1 1

$$\begin{aligned} \text{情報비트}: c_3 &= d_0 \\ c_4 &= d_1 \\ c_5 &= d_2 \\ c_6 &= d_3 \end{aligned} \qquad (5)$$

檢査長이 n−k=3이므로 (4)式의 檢査비트 c_0, c_1, c_2는 情報비트의 線形結合으로 이루어짐을 알 수 있으며 이 세 個의 方程式을 주어진 (7, 4)符號의 패리티檢査方程式(parity check equation)이라 한다.

이 符號는 組織構造를 가지므로 符號語는 $\bar{c} = (d_0 + d_2 + d_3,\ d_0 + d_1 + d_2,\ d_1 + d_2 + d_3,\ d_0, d_1, d_2, d_3)$로 된다. 情報 $\bar{d} = (0110)$ 即 $d_0 = 0$, $d_1 = 1$, $d_2 = 1$, $d_3 = 0$에 對應하는 符號語를 구하면 (4)式과 (5)式에 依해 檢査비트는 $c_0 = 1$, $c_1 = 0$, $c_2 = 0$, 情報비트는 $c_3 = 0$, $c_4 = 1$, $c_5 = 1$, $c_6 = 0$가 되므로 完全한 符號語는 $\bar{c} = (1000110)$가 된다.

以上의 說明으로부터 어느 符號化 方法을 使用하여도 同一한 符號語를 얻을 수 있음을 알 수 있다.

一次獨立인 k個의 行(row)을 갖는 任意의 k×n 生成行列 $\bar{G}$에 對해 n−k個의 一次獨立인 行을 갖는 (n−k)×n行列 $\bar{H}$가 存在한다. 그런데, $\bar{G}$의 行空間(row space)內의 모든 벡터는 $\bar{H}$의 行에 直交(orthogonal)하며 이를 式으로 表示하면 다음과 같다.

$$\bar{G} \cdot \bar{H}^T = \bar{0} \qquad (6)$$

다시 말해서 $\bar{H}$의 行에 直交하는 任意의 벡터는 $\bar{G}$의 行空間內에 存在한다는 것이다. 이 行列 $\bar{H}$를 符號 C의 패리티檢査行列(parity-check matrix) 或은 檢査

行列이라 한다.

따라서 (n, k)組織符號의 生成行列 $\bar{G}$에 對應하는 檢査行列은 다음과 같은 形態를 갖는다.

$$\bar{H}=[\bar{I}_{n-k}:\bar{P}^{T}_{(n-k)\times k}]$$

$$=\begin{bmatrix} 1\,0\,0\cdots 0 & p_{00} & p_{10} & \cdots p_{k-1,0} \\ 0\,1\,0\quad 0 & p_{01} & p_{11} & \cdots p_{k-1,1} \\ \vdots & & & \\ 0\,0\,0\cdots 1 & p_{0,n-k-1} & p_{1,n-k-1} & \cdots p_{k-1,n-k-1} \end{bmatrix}$$

$$=\begin{bmatrix} \bar{h}_0 \\ \bar{h}_1 \\ \vdots \\ \bar{h}_{n-k-1} \end{bmatrix} \qquad (7)$$

여기서 $\bar{P}^{T}$는 (2)式에 表現된 行列 $\bar{P}$의 轉置行列 (transpose matrix)이다.

그리고 符號벡터 $\bar{c}=(c_0, c_1, c_2, \cdots, c_{n-1})$가 檢査行列 $\bar{H}$를 갖는 線形符號의 符號語가 되기 爲해서는 다음과 같은 關係式을 滿足하여야 한다.

$$\bar{c}\cdot\bar{H}^{T}=\bar{0} \qquad (8)$$

例 2. (6, 3)線形符號의 生成行列이

$$\bar{G}=[\bar{P}_{3\times 3}:\bar{I}_{3\times 3}]$$

$$=\begin{bmatrix} 0\,1\,1\,1\,0\,0 \\ 1\,0\,1\,0\,1\,0 \\ 1\,1\,0\,0\,0\,1 \end{bmatrix}$$

일 때 이에 對應하는 檢査行列은

$$\bar{H}=[\bar{I}_{3\times 3}:\bar{P}^{T}_{3\times 3}]$$

$$=\begin{bmatrix} 1\,0\,0\,0\,1\,1 \\ 0\,1\,0\,1\,0\,1 \\ 0\,0\,1\,1\,1\,0 \end{bmatrix}$$

가 되며 $\bar{H}$의 轉置行列은 다음과 같다.

$$\bar{H}^{T}=\begin{bmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \\ 0\,1\,1 \\ 1\,0\,1 \\ 1\,1\,0 \end{bmatrix}$$

그리고 이 符號의 符號器는 그림 3과 같다.

Ⅳ. Hamming 距離와 重

符號理論에서는 誤謬의 檢出및 訂正特性을 나타내는 重要한 媒介變數로서 距離(distance)라는 槪念을 導入한다.

定義 1. 符號語 $\bar{c}=(c_0, c_1, c_2, \cdots\cdots, c_{n-1})$가 n次元벡터(n-tuple)라면 Hamming 重(weight) $w(\bar{c})$는 符號벡터 內에서의 零이 아닌 成分의 個數로 定義한다.

例 3. 符號벡터 $\bar{c}=(0101100)$의 Hamming重은 3이다.

한편, 두 個의 n次元 벡터 $\bar{u}=(u_0, u_1, \cdots\cdots, u_{n-1})$와 $\bar{v}=(v_0, v_1, \cdots\cdots, v_{n-1})$間의 Hamming 距離 $d(\bar{u}, \bar{v})$는 $\bar{u}$와 $\bar{v}$의 對應되는 成分中 서로 다른 成分雙의 個數를 말한다.

例 4. 符號語 $\bar{u}=(1110010)$와 $\bar{v}=(1010001)$間의 Hamming 距離는 3이다. 그리고 $\bar{u}+\bar{v}=(0100011)$이므로 이 벡터의 Hamming重 $w(\bar{u}+\bar{v})$는 3이다.

따라서 $\bar{u}+\bar{v}$에 對한 Hamming重의 값은 벡터 $\bar{u}$와 $\bar{v}$間의 Hamming 距離 $d(\bar{u}, \bar{v})$와 같음을 알 수 있다.

$$d(\bar{u}, \bar{v})=w(\bar{u}+\bar{v}) \qquad (9)$$

定義 2. 線形符號 C에서 서로 다른 두 符號語間의 Hamming 距離中 最小値를 符號 C의 Hamming 最小距離, 또는 簡略하게 最小距離(minimum distance)라 부르며 d_{min}으로 表示한다.

$$d_{min}=\min\{d(\bar{u}, \bar{v});\bar{u}\neq\bar{v}, \bar{u}, \bar{v}\in C\} \qquad (10)$$

Ⅴ. 誤 症

生成行列 $\bar{G}$와 檢査行列 $\bar{H}$를 갖는 (n, k)線形符號를 생각해 보자. 傳送된 符號語를 $\bar{c}=(c_0, c_1, \cdots\cdots, c_{n-1})$라 하고 受信벡터를 $\bar{r}=(r_0, r_1, \cdots\cdots, r_{n-1})$이라 하면 傳送路(transmission channel)에서 發生한 誤謬形態(error pattern), $\bar{e}=(e_0, e_1, \cdots\cdots, e_{n-1})$는 다음과 같은 關係가 있다.

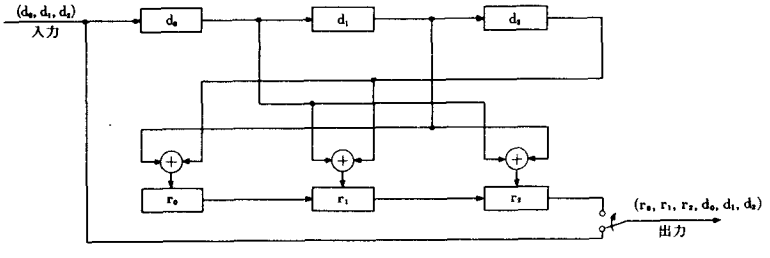

그림 3. (6, 3)線形符號의 符號器

$$\bar{r}=\bar{c}+\bar{e}$$

또는 $$\bar{c}=\bar{r}+\bar{e} \qquad (11)$$

여기서 $\begin{cases} r_i \neq c_i \text{ 일때 } e_i=1 \\ r_i=c_i \text{ 일때 } e_i=0 \end{cases}$

誤謬벡터 $\bar{e}$에 있는 $e_i=1, 0\leq i\leq n-1$ 은 傳送路上에서 나타난 雜音에 依해 發生한 傳送誤謬(transmission error)이다. 復號器는 $\bar{r}$에 傳送誤謬가 包含되어 있는가를 알아보고 誤謬의 存在가 確認되면 誤謬位置(error position)를 發見하여 訂正해야 된다.

誤謬의 存在與否를 檢出하기 爲해 誤症, $\bar{s}$를 다음과 같이 定義 한다.

$$\bar{s}=\bar{r}\cdot\bar{H}^T$$

$$=(s_0, s_1, s_2, \cdots\cdots, s_{n-k-1}) \qquad (12)$$

即, 誤症은 受信系列 $\bar{r}$과 檢查行列 $\bar{H}$의 轉置行列 $\bar{H}^T$와의 內積으로 求한다. 受信系列 $\bar{r}$에 誤謬가 包含되어 있을 때는 $\bar{s}\neq\bar{0}$이며 誤謬가 없을 때는 $\bar{s}=\bar{0}$이 된다.

例 5. (7, 4)線形符號에서 패리티檢查行列이 다음과 같으면

$$\bar{H}=\begin{bmatrix} 1&0&0&1&0&1&1 \\ 0&1&0&1&1&0&1 \\ 0&0&1&1&1&1&0 \end{bmatrix}$$

轉置行列, $\bar{H}^T$는

$$\bar{H}^T=\begin{bmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&1 \\ 1&1&1 \\ 0&1&1 \\ 1&0&1 \\ 1&1&0 \end{bmatrix}$$

가 된다. 受信벡터를 $\bar{r}=(r_0, r_1, r_2, r_3, r_4, r_5, r_6)$ 라 하면 誤症 $\bar{s}$는 (12)式으로부터 다음과 같이 求할 수 있다.

$$\bar{s}=(s_0, s_1, s_2)=(r_0, r_1, r_2, r_3, r_4, r_5, r_6)\cdot\begin{bmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&1 \\ 1&1&1 \\ 0&1&1 \\ 1&0&1 \\ 1&1&0 \end{bmatrix}$$

$$=(r_0+r_3+r_5+r_6,\ r_1+r_3+r_4+r_6,\ r_2+r_3+r_4+r_5)$$

이 符號에 對한 誤症回路(syndrome circuit)를 그림 4에 나타내었다.

(12)式에 (11)式을 代入 하면

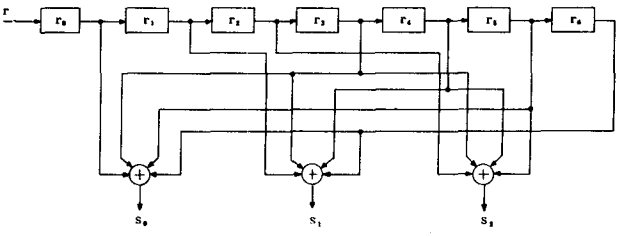

그림 4. (7, 4)線形符號의 誤症回路

$$\bar{s}=\bar{r}\cdot\bar{H}^T=(\bar{c}+\bar{e})\cdot\bar{H}^T=\bar{c}\cdot\bar{H}^T+\bar{e}\cdot\bar{H}^T$$

이 되고 (8)式을 適用하면 誤症은

$$\bar{s}=\bar{e}\cdot\bar{H}^T \qquad (13)$$

로 簡單히 表現된다.

이런 事實로 미루어 誤謬形態와 誤症은 서로 密接한 關係가 있음을 알 수 있으며, 앞에서 말한 바와 같이 誤謬가 發生하지 않았을 境遇에는 $\bar{e}=\bar{o}$이므로 誤症이 零이다. 그리고, 이 誤症을 土台로 하여 線形符號의 復號(decoding)가 이루어진다. 即, 誤症으로부터 誤謬形態가 決定되면 (11)式과 같이 受信벡터를 더함으로써 實際傳送된 符號벡터를 얻을 수 있다.

Ⅵ. 誤謬訂正能力과 限界式

符號長(code length), n인 符號語의 어느 한 비트에 誤謬가 發生한 것을 單一誤謬(single error)라 하며, 任意의 두 비트에 誤謬가 發生한 것을 2重誤謬(double error)라 한다. 이와 마찬가지로 t個의 誤謬가 發生했을 境遇를 t重誤謬(t-tuple error)라 말한다.

符號 C의 最小距離가 3以上, 即 $d_{min}\geq 3$이면 單一誤謬를 訂正할 수 있으며, 이런 條件을 滿足하는 符號를 單一誤謬訂正符號(single-error correcting code)라 한다. 例를 들면, 이 部類에 屬하는 代表的인 符號로서 最小距離가 3인 (7, 4) Hamming符號가 있다. 一般的으로 符號의 最小距離가 不等式, $d_{min}\geq 2t+1$ 을 滿足할 때 이 符號를 t重誤謬訂正符號(t-error correcting code)라 부른다. 또, 符號의 最小距離가 $d_{min}\geq t+1$의 不等條件을 滿足하면 이 符號는 t個 以下의 誤謬를 檢出할 수 있다. 例를 들면 $d_{min}=2$인 符號는 誤謬訂正이 不可能하다. 다만 한個의 誤謬를 檢出할 수 있을 뿐이다. $d_{min}=3$인 符號는 한個의 誤謬를 訂正할 수 있고, 두個의 誤謬를 檢出할 수 있다. 다시 말해서 $d_{min}=3$인 符號는 하나 또는 두個의 誤謬까지 檢出할 수 있으나 訂正에 있어서는 單一誤謬만이 可能한 것이다. 또 $d_{min}=5$일 때는 두個의 誤謬를 訂正할 수 있으며, 4個의 誤謬까지 檢出 可能하다.

그러면 여기서 一般的으로 t重誤謬를 訂正할 수 있는 符號의 情報長 k와 符號長 n의 값은 어떻게 定해야 하는지에 對해 알아보자. 이에 對해서는 여러가지 限界式(bound)이 提案되고 硏究되어 왔다. 그 中 代表的인 것 하나만을 紹介하기로 한다.

線形符號 C가 t重誤謬訂正이 可能하려면 重이 t 以下인 두個의 誤謬形態(error pattern)가 同一한 剩餘類(coset)內에 存在해서는 안된다. (n, k)符號內에는 2^{n-k}個의 剩餘類가 있고, $\binom{n}{i}$個의 i重 n次元符號語(n-tuple code word)가 存在하므로 다음의 限界式이 成立된다.

$$2^{n-k} \geq \sum_{i=0}^{t} \binom{n}{i} \tag{14}$$

이 不等式을 Hamming限界式(Hamming bound, or sphere-packing bound)이라 한다. 例를 들어 t=1, 即 單一誤謬訂正時에는 (14)式은 $2^{n-k} \geq n+1$이 된다. 따라서 檢査비트數 n−k는 $n-k > \log_2 n$의 不等式을 滿足해야 할 것이다.

Ⅶ. 標準配列과 復號

(n, k)線形符號의 符號語 $\bar{c}_1, \bar{c}_2, \cdots\cdots, \bar{c}_{2^k}$中에서 어떤 符號語를 傳送해도 受信벡터 $\bar{r}$은 GF(2)의 元素{0, 1}로 表示되는 2^n個의 n次元벡터 中의 하나로 나타난다. 受信側에서는 復號하기 爲해 各各의 符號語 $\bar{c}_i$가 集合 Ω_i에 1對1 對應되도록 2^n個 受信벡터를 2^k個의 서로 다른 集合 $\Omega_1, \Omega_2, \cdots\cdots, \Omega_{2^k}$로 나눈다. 受信벡터 $\bar{r}$이 Ω_i 內에 存在하면 $\bar{r}$은 $\bar{c}_i$로 復號되며 올바른 復號는 $\bar{r}$이 實際 符號語에 對應되는 集合 Ω_i 內에 存在할 境遇에 限해 이루어진다. 2^n個의 受信벡터를 2^k 個의 相異한 集合으로 나누는 方法은 다음과 같다.

표 2. (n, k)線形符號의 標準配列

$\bar{c}_1=\bar{0}$	$\bar{c}_2$ ……	$\bar{c}_j$ ……	$\bar{c}_{2^k}$
$\bar{e}_2$	$\bar{e}_2+\bar{c}_2$ ……	$\bar{e}_2+\bar{c}_j$ ……	$\bar{e}_2+\bar{c}_{2^k}$
$\bar{e}_3$	$\bar{e}_3+\bar{c}_2$ ……	$\bar{e}_3+\bar{c}_j$ ……	$\bar{e}_3+\bar{c}_{2^k}$
⋮	⋮	⋮	⋮
$\bar{e}_i$	$\bar{e}_i+\bar{c}_2$ ……	$\bar{e}_i+\bar{c}_j$ ……	$\bar{e}_i+\bar{c}_{2^k}$
⋮	⋮	⋮	⋮
$\bar{e}_{2^{n-k}}$	$\bar{e}_{2^{n-k}}+\bar{c}_2$ ……	$\bar{e}_{2^{n-k}}+\bar{c}_j$ ……	$\bar{e}_{2^{n-k}}+\bar{c}_{2^k}$

먼저 첫번째 行에는 2^k個의 符號語를 配列하는데 가장 왼쪽 位置에는 零符號語(all-zero code vector) $\bar{c}_1$을 둔다. 첫번째 行을 構成하고 난 나머지, 2^n-2^k個의 n次元 벡터 中에서 誤謬形態 $\bar{e}_2$를 選擇하여 $\bar{c}_1$ 아래에 配列하고 各各의 符號語 $\bar{c}_i$에 $\bar{e}_2$를 더하여 $\bar{c}_i$ 밑에 配列함으로써 두번째 行이 形成된다. 두 번째 行까지를 形成하고 남은, 나머지 n次元 벡터 中, $\bar{e}_3$를 選擇하여 $\bar{e}_2$아래에 두고 위와 같은 方法으로 세 번째 行을 構成한다. 이 過程을 모든 n次元 벡터가 使用될 때까지 繼續하면 표2와 같은 配列이 되며 이와 같은 配列을 標準配列(standard array)이라 부른다. 標準配列에는 $2^n/2^k=2^{n-k}$個의 行이 있고, 各各의 行은 2^k 個의 서로 다른 n次元 벡터로 이루어진다. 이런 各各의 行을 剩餘類(coset), 各 行의 가장 왼쪽에 있는 n次元 벡터를 剩餘類首(coset leader)라 한다.

定理 1. 標準配列에서 두個의 같은 n次元 벡터는 같은 行에 存在할 수 없다. 各 n次元 벡터는 오직 하나의 行에만 存在한다.

例 6. 生成行列이 아래와 같은 (6, 3)線形符號의 標準配列을 생각해 보자.

$$\bar{G}=\begin{bmatrix} 0&1&1&1&0&0 \\ 1&0&1&0&1&0 \\ 1&1&0&0&0&1 \end{bmatrix}$$

情報長이 3이므로 符號벡터는 모두 $2^3=8$ 個이며 이것을 求해 보면

$$\bar{c}=\bar{d}\cdot\bar{G}=\begin{bmatrix} 0&0&0 \\ 0&0&1 \\ 0&1&0 \\ 0&1&1 \\ 1&0&0 \\ 1&0&1 \\ 1&1&0 \\ 1&1&1 \end{bmatrix}\cdot\begin{bmatrix} 0&1&1&1&0&0 \\ 1&0&1&0&1&0 \\ 1&1&0&0&0&1 \end{bmatrix}$$

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

가 된다.

위에서 求한 符號벡터로부터 標準配列을 構成하면 다음 표3과 같다.

(n, k)線形符號의 標準配列은 2^k個의 列(column)로 이루어지며, 各各의 列은 2^{n-k}個 n次元 벡터로 構成되고 첫 번째 行은 符號벡터이다. Ω_j를 標準配列의 j번째 列이라 하면,

$$\Omega_j=\{\bar{c}_j, \bar{e}_2+\bar{c}_j, \bar{e}_3+\bar{c}_j, \cdots, \bar{e}_{2^{n-k}}+\bar{c}_j\}$$

표 3. (6, 3)線形符號의 標準配列

000000	110001	101010	011011	011100	101101	110110	000111
000001	110000	101011	011010	011101	101100	110111	000110
000010	110011	101000	011001	011110	101111	110100	000101
000100	110101	101110	011111	011000	101001	110010	000011
001000	111001	100010	010011	010100	100101	111110	001111
010000	100001	111010	001011	001100	111101	100110	010111
100000	010001	001010	111011	111100	001101	010110	100111
001001	111000	100011	010010	010101	100100	111111	001110

이다. 여기서 $\bar{c}_j$는 符號벡터이고, $\bar{e}_i$, $2 \leqq i \leqq 2^{n-k}$는剩餘類首를 말한다.

$\bar{c}_j$를 傳送된 符號벡터, $\bar{r}$을 受信벡터라 할 때, 誤謬形態 $\bar{e}$가 剩餘類首이고 $\bar{r}$이 Ω_j안에 있을 때는 受信벡터 $\bar{r}$은 $\bar{c}_j$로 正確하게 復號된다. 그러나 誤謬形態가 剩餘類首가 아닐 境遇에는 訂正된 結果는 $\bar{c}_j$ 가 아니다. 그 理由는 다음과 같다. 剩餘類首가 아닌 誤謬形態 $\bar{v}$를 符號벡터 $\bar{c}_j$ 아래의 1번째 行에 있는 벡터라 생각하면,

$$\bar{v}=\bar{e}_i+\bar{c}_i$$

이며

$$\bar{r}=\bar{c}_j+\bar{v}=\bar{e}_i+(\bar{c}_i+\bar{c}_j)=\bar{e}_i+\bar{c}_q$$

가 되어 受信벡터, $\bar{r}$은 Ω_q 內에 存在하고 $\bar{c}_q$로 復號된다. 이것은 傳送된 符號벡터, $\bar{c}_j$가 아니므로 傳送路上에서 發生한 誤謬形態가 剩餘類首일 때만 올바른 復號가 이루어진다는 것을 알 수 있다. 따라서 零을 包含하는 2^{n-k}個의 剩餘類首는 訂正 可能한 誤謬形態이다.

剩餘類首가 $\bar{e}_i$인 剩餘類의 誤症은

$$\begin{aligned}\bar{s}&=\bar{r}\bar{H}^T\\&=(\bar{e}_i+\bar{c}_i)\bar{H}^T\\&=\bar{e}_i\bar{H}^T+\bar{c}_i\bar{H}^T\\&=\bar{e}_i\bar{H}^T\end{aligned}$$

로 된다. 即, 剩餘類에 있는 어떤 벡터의 誤症은 剩餘類首의 誤症과 同一하다. 따라서 한 剩餘類內의 모든 벡터는 같은 誤症을 갖는다. $\bar{e}_j$와 $\bar{e}_i$이 各各 j 번째, 1번째 剩餘類首이고, $j<1$일 때 두 剩餘類의 誤症이 같다고 하면,

$$\bar{e}_j\bar{H}^T=\bar{e}_i\bar{H}^T$$

$$(\bar{e}_j+\bar{e}_i)\bar{H}^T=0$$

가 되는데, 이 式은 $\bar{e}_j+\bar{e}_i$이 符號벡터임을 表示한다. 即, $\bar{e}_j+\bar{e}_i=\bar{c}_i$

$$\bar{e}_j+\bar{c}_i=\bar{e}_i$$

이 式은 $\bar{e}_i$이 j번째 剩餘類內에 存在함을 意味하므로, 앞에서 說明한 標準配列의 構成法에 矛盾된다. 이런 事實로부터 어떤 두個의 剩餘類도 同一한 誤症을 가질 수 없음을 안다.

例7. 檢査行列이 다음과 같은 (8, 4)線形符號의 復號에 대해 알아보자.

$$\bar{H}=\begin{bmatrix}1&0&0&0&1&1&0&1\\0&1&0&0&1&0&1&1\\0&0&1&0&1&1&1&0\\0&0&0&1&0&1&1&1\end{bmatrix}$$

이 符號의 剩餘類 個數는 모두 $2^{n-k}=2^4=16$이므로, 訂正可能한 誤謬形態는 16가지이다. 그러므로 重이 1 或은 0인 모든 誤謬形態 9個外에 重이 2인 誤謬形態의 一部인 7個가 剩餘類首로 使用된다. 이들 16個 誤謬形態에 對한 誤症을 (12)式에 依해 求하면 표4와 같다.

표 4. (8, 4)線形符號의 復號表

誤症, $\bar{s}_i$	剩餘類首, $\bar{e}_i$
0 0 0 0	0 0 0 0 0 0 0 0
0 0 0 1	0 0 0 1 0 0 0 0
0 0 1 0	0 0 1 0 0 0 0 0
0 0 1 1	0 0 1 1 0 0 0 0
0 1 0 0	0 1 0 0 0 0 0 0
0 1 0 1	0 1 0 1 0 0 0 0
0 1 1 0	0 1 1 0 0 0 0 0
0 1 1 1	0 0 0 0 0 0 1 0
1 0 0 0	1 0 0 0 0 0 0 0
1 0 0 1	1 0 0 1 0 0 0 0
1 0 1 0	1 0 1 0 0 0 0 0
1 0 1 1	0 0 0 0 0 1 0 0
1 1 0 0	1 1 0 0 0 0 0 0
1 1 0 1	0 0 0 0 0 0 0 1
1 1 1 0	0 0 0 0 1 0 0 0
1 1 1 1	0 0 0 1 1 0 0 0

符號語, $\bar{c}=(10011010)$가 傳送되어 $\bar{r}=(10011110)$가 受信되었다면 $\bar{r}$의 誤症은

$$\bar{s}=\bar{r}\cdot\bar{H}^T=(10011110)\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\1&1&1&0\\1&0&1&1\\0&1&1&1\\1&1&0&1\end{bmatrix}=(1011)$$

로 計算된다. 표 4 에서 $\bar{s}=(1011)$ 은 剩餘類首, $\bar{e}=(00000100)$ 의 誤症이므로, 이것이 바로 傳送路上 에서 發生한 誤謬形態이다. 結局 $\bar{r}$은 다음 式과 같이

$$\bar{c}=\bar{r}+\bar{e}$$
$$=(10011110)+(00000100)$$
$$=(10011010)$$

로 復號되며, 이것은 實際로 送信한 符號語이다. 한편 2 重誤謬가 發生한 $\bar{r}=(00001010)$ 이 受信되었을 境遇를 생각해 보면, 위와 같은 方法으로 $\bar{s}=(1001)$ 와 $\bar{e}=(10010000)$ 를 求할 수 있고 $\bar{r}$은 (10011010) 로 옳게 復號된다. 이 境遇는 誤謬形態가 剩餘類首에 使用되었기 때문에 옳게 復號되었으나 誤謬形態가 $\bar{e}=(10000001)$ 이면 $\bar{r}$은 $\bar{c}$로 復號되지 않는다. 그 理由는 이 誤謬形態가 표 4 에서 剩餘類首로 使用되지 않았기 때문이다.

例 8. 例 6 과 같은 生成行列을 갖는 (6, 3) 線形符號의 復號器를 設計해 보자. 이 符號에는 모든 成分이 零인 零벡터를 包含하여 7 個의 訂正可能한 誤謬形態가 있으며, 이들의 誤症은 표 5 와 같다.

표 5. (6, 3) 線形符號의 誤症과 그의 對應하는 誤謬形態

誤症	誤謬形態
$s_0\ s_1\ s_2$	$e_0\ e_1\ e_2\ e_3\ e_4\ e_5$
0 0 0	0 0 0 0 0 0
1 0 0	1 0 0 0 0 0
0 1 0	0 1 0 0 0 0
0 0 1	0 0 1 0 0 0
0 1 1	0 0 0 1 0 0
1 0 1	0 0 0 0 1 0
1 1 0	0 0 0 0 0 1

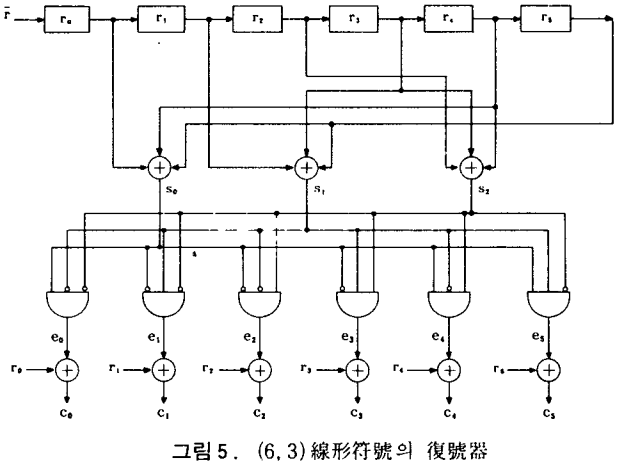

그림 5. (6, 3) 線形符號의 復號器

어떤 受信벡터에 對한 誤症이 計算되고 나면, 이에 對應하는 誤謬形態를 알 수 있으므로 復號를 할 수 있지만 誤症이 (111) 일 境遇에는 受信벡터에 訂正不可能한 誤謬形態가 發生하여 復號를 할 수 없다. 표 5 를 利用하여 復號器를 그림 5 와 같이 設計할 수 있다.

Ⅷ. 寸 評

現代 通信工學을 大別해서 情報理論, 統計通信理論 및 符號理論으로 區分하는데 通信系 全体를 놓고 보면 各其의 內容과 特徵에는 많은 差異가 있다. 1948年에 C. E. Shannon이 提唱한 通信路 符號化 定理 以來 情報에 確率 槪念을 導入해서 情報를 비트로 表現하고, 計算된 情報量 및 情報傳送速度가 通信路容量을 넘지 않는 範圍內에서는 誤謬없이 情報傳達이 可能하다는 抽象的인 情報理論이 發展되어 왔으나, 符號化와 復號에 關한 具体的인 方法은 提示되지 않았다. 따라서 符號理論은 通信史上 大革命을 일으킨 應用分野로서 C. E. Shannon의 情報理論에 基礎를 두고 發展되었으며, 이야말로 現代 디지탈 데이타 通信의 核心이고 現在 디지탈 通信系와 컴퓨터 記憶系의 信賴性 向上에 至大한 功獻을 해왔다 해도 過言이 아니다.

한편 統計通信理論은 아나로그 通信方式이든 디지탈 通信方式이든 그림 1 에 나타낸 點線內의 現象을 檢討하는 것으로, 變調器－傳送路－復調器(MODEM)를 놓고 信號設計(signal design), 濾過(filtering), 檢波(signal detection) 및 信號推定(signal estimation) 等 主로 通信路內에서의 雜音, 干涉, 歪曲 等이 信號에 미치는 影響과 改善策을 究明하는 分野로서 現在 우리나라의 大學이나 研究所에서도 活發히 研究하고 있는 것으로 안다. 그러나 符號器－通信路－復號器(CODEC)를 一括해서 디지탈 데이타 通信系 全体로 놓고 볼 때는 아직도 不毛地인 印象을 준다. 우리 나라도 大學과 研究所를 包含한 電子·通信界에서 좀 더 符號理論의 重要性을 認識해야 할 때가 온 것으로 안다. 따라서 많은 大學에서 이 分野를 研究, 設講함으로써 우리 나라 現在의 通信界에 革新이 이루어졌으면 하는 마음 懇切하다.