北海道大学
Hokkaido University

情報エレクトロニクス学科共通科目・2年次・夏ターム〔必修科目〕

講義「情報理論」

第13回

第8章 通信路符号化法

8.2 ハミング符号

[復習]単一パリティ検査符号

単一パリティ検査符号

系列$x_1x_2\cdots x_k$を、含まれる1の個数が偶数になるように、もう一つ記号

$$c = x_1 + x_2 + \cdots + x_k$$

を付け加えた符号語$w = x_1x_2\cdots x_k\, c$に対応させる符号

情報記号（information symbol）[2元の場合は情報ビット]

符号語$w = \overbrace{x_1x_2\cdots x_k}\, \underbrace{c}$

（パリティ）検査記号（[parity] check symbol）[2元の場合は（パリティ）検査ビット]

符号語のベクトル表現：$w = (x_1, x_2, \cdots, x_k, c)$

[復習]組織符号

組織符号（systematic code）

k 個の情報記号から、$n-k$ 個の検査記号を一定の方法で求め、付加することにより符号化される符号長 n の符号

$$w = \underbrace{x_1\, x_2 \cdots x_k\; c_1\, c_2 \cdots c_{n-k}}_{n}$$

(n, k) 符号

符号長 n、情報記号数 k の組織符号

[復習]線形符号とパリティ検査方程式

線形符号（linear code）

$c = x_1 + x_2 + \cdots + x_k$ のように、
検査記号が情報記号の線形な式で与えられる符号

パリティ検査方程式

＝0 という形で線形符号の符号語となるための必要十分条件を与える式（または式の組）

[単一パリティ検査符号のパリティ検査方程式]

長さnの系列$w=(w_1, w_2, \cdots, w_n)$ が単一パリティ検査符号の符号語となるための必要十分条件は以下のパリティ検査方程式を満たすことである.

$$w_1 + w_2 + \cdots + w_{n-1} + w_n = 0$$

（符号語に含まれる1の個数が偶数）

[復習]誤りパターンとシンドローム

誤りパターン（error pattern）

送信した符号語 $\boldsymbol{w}=(w_1, w_2, \cdots, w_n)$と
受信語$\boldsymbol{y}=(y_1, y_2, \cdots, y_n)$との差$\boldsymbol{e}=\boldsymbol{w}+\boldsymbol{y}=(w_1+y_1, w_2+y_2, \cdots, w_n+y_n)$

- 誤りパターン$\boldsymbol{e}=(e_1, e_2, \cdots, e_n)$の各ビットは以下のような値になる

$$e_i=\begin{cases} 1 & (\text{第i成分に誤りが生じたとき}) \\ 0 & (\text{そうでないとき}) \end{cases}$$

- 誤りパターン$\boldsymbol{e}$を用いると$\boldsymbol{y}=\boldsymbol{w}+\boldsymbol{e}$と表現できる

シンドローム（syndrome）

受信語 $\boldsymbol{y}$ をパリティ検査方程式の左辺に代入した結果 $\boldsymbol{s}=y_1+y_2+\cdots+y_n$

- 符号語 $\boldsymbol{w}$ はパリティ検査方程式を満たすので、

$$\boldsymbol{s}=w_1+e_1+w_2+e_2+\cdots+w_n+e_n=e_1+e_2+\cdots+e_n$$

誤りがない$\Rightarrow \boldsymbol{s}=0$,　1個の誤り $\Rightarrow \boldsymbol{s}=1$

[復習]2×2水平垂直パリティ検査符号

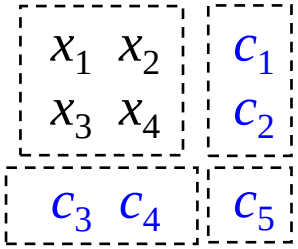
2×2水平垂直パリティ検査符号

与えられた長さ4の系列$x_1x_2x_3x_4$に対し、長さ9の符号語$(x_1,x_2,x_3,x_4,c_1,c_2,c_3,c_4,c_5)$を、以下のように生成する符号。右図のように4個の情報ビットを 2×2 の配列に並べ、各行各列に一つずつ検査ビットを付け加える。すなわち、

$$c_1=x_1+x_2 \qquad c_2=x_3+x_4$$
$$c_3=x_1+x_3 \qquad c_4=x_2+x_4$$

さらに、検査ビットの行の1の数が偶数になるように、検査ビットの検査ビットを右隅におく。

$$c_5=c_1+c_2=x_1+x_2+x_3+x_4=c_3+c_4$$



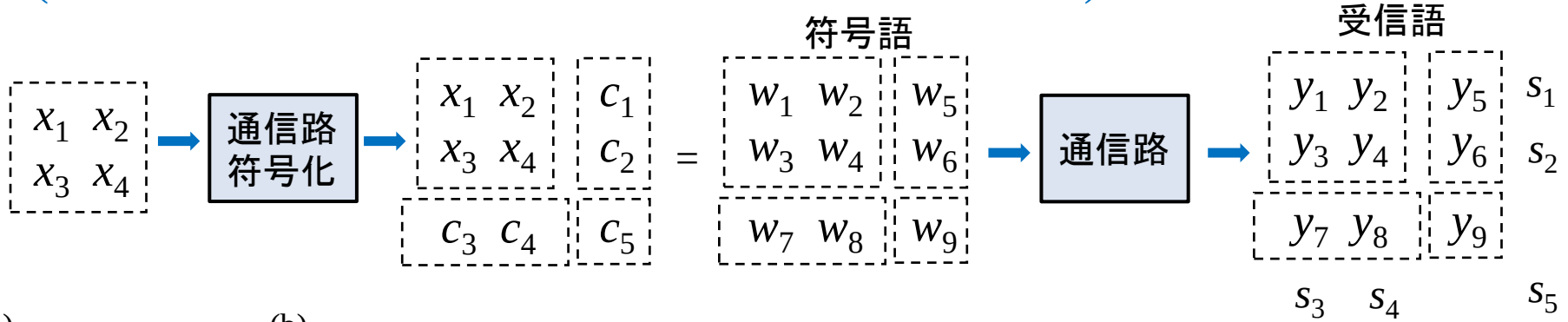
図：水平垂直パリティ検査符号

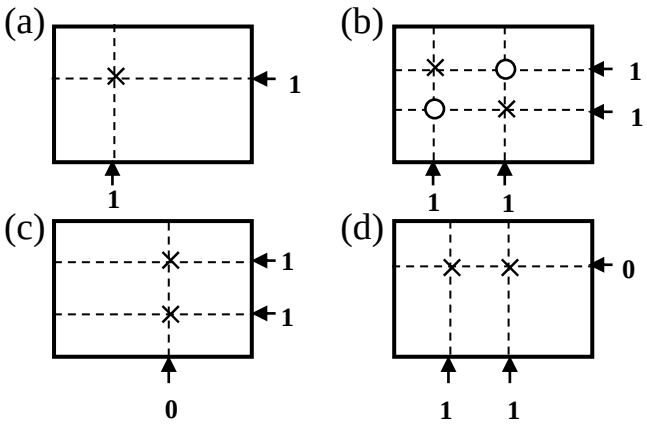
[復習]2×2水平垂直パリティ検査符号の誤り訂正能力

誤りの訂正まで行える誤り(検出)訂正符号（error-correcting code）

1個の誤りが訂正でき、同時に2個の誤りを検出することができる。

(検出のみに徹すれば、3個の誤りまで検出することが可能。)





図：単一誤りの訂正と2重誤りの検出

(シンドロームが0の位置は検出できない)

(a) 情報ビットのみに1ビットの誤りが生じた場合

(s_1, s_2の一方が1) and (s_3, s_4の一方が1) and s_5=1

誤りが生じた行と列が特定可能⇒訂正可能

検査ビットのみの1ビット誤りも訂正可能

(b)(c)(d)2ビットの誤りが生じた場合

s_1, s_2, s_3, s_4, s_5のどれかは1

ビットの特定は不可能⇒検出可能、訂正不可能

(7,4)ハミング符号(1)

- 水平垂直パリティ検査符号の問題点
  (9,4)符号であり、情報ビットよりも検査ビットが多く、効率がよくない。

- (7,4)ハミング符号
  4個の情報ビット x_1, x_2, x_3, x_4 に対し、

$$c_1 = x_1 + x_2 + x_3$$
$$c_2 = x_2 + x_3 + x_4$$
$$c_3 = x_1 + x_2 + x_4$$

  により、検査ビット c_1, c_2, c_3 を作り、

$$w = (x_1, x_2, x_3, x_4, c_1, c_2, c_3)$$

  という符号語に符号化する符号
  (検査ビットの作り方にはいくつかのバリエーションがあることに注意!)

情報ビットが4個であるから、符号語は $2^4 = 16$ 個

表:(7,4)ハミング符号

x_1	x_2	x_3	x_4	c_1	c_2	c_3
0	0	0	0	0	0	0
1	0	0	0	1	0	1
0	1	0	0	1	1	1
1	1	0	0	0	1	0
0	0	1	0	1	1	0
1	0	1	0	0	1	1
0	1	1	0	0	0	1
1	1	1	0	1	0	0
0	0	0	1	0	1	1
1	0	0	1	1	1	0
0	1	0	1	1	0	0
1	1	0	1	0	0	1
0	0	1	1	1	0	1
1	0	1	1	0	0	0
0	1	1	1	0	1	0
1	1	1	1	1	1	1

(7,4)ハミング符号(2)

- 符号語を $\boldsymbol{w}=(w_1, w_2, \cdots, w_7)$ する。

(7,4)ハミング符号のパリティ検査方程式

$$w_1+w_2+w_3 \quad +w_5 \quad =0$$
$$w_2+w_3+w_4 \quad +w_6 \quad =0$$
$$w_1+w_2 \quad +w_4 \quad +w_7=0$$

受信語 $\boldsymbol{y}=(y_1, y_2, \cdots, y_7)$ に対するシンドローム

$$\boldsymbol{s}_1=y_1+y_2+y_3 \quad +y_5$$
$$\boldsymbol{s}_2= \quad y_2+y_3+y_4 \quad +y_6$$
$$\boldsymbol{s}_3=y_1+y_2 \quad +y_4 \quad +y_7$$

$y_i=w_i+e_i$

誤りパターン を$\mathrm{e}=(e_1, e_2, \cdots, e_7)$とすると

$$\boldsymbol{s}_1=e_1+e_2+e_3 \quad +e_5$$
$$\boldsymbol{s}_2= \quad e_2+e_3+e_4 \quad +e_6$$
$$\boldsymbol{s}_3=e_1+e_2 \quad +e_4 \quad +e_7$$

表：単一誤りに対するシンドローム

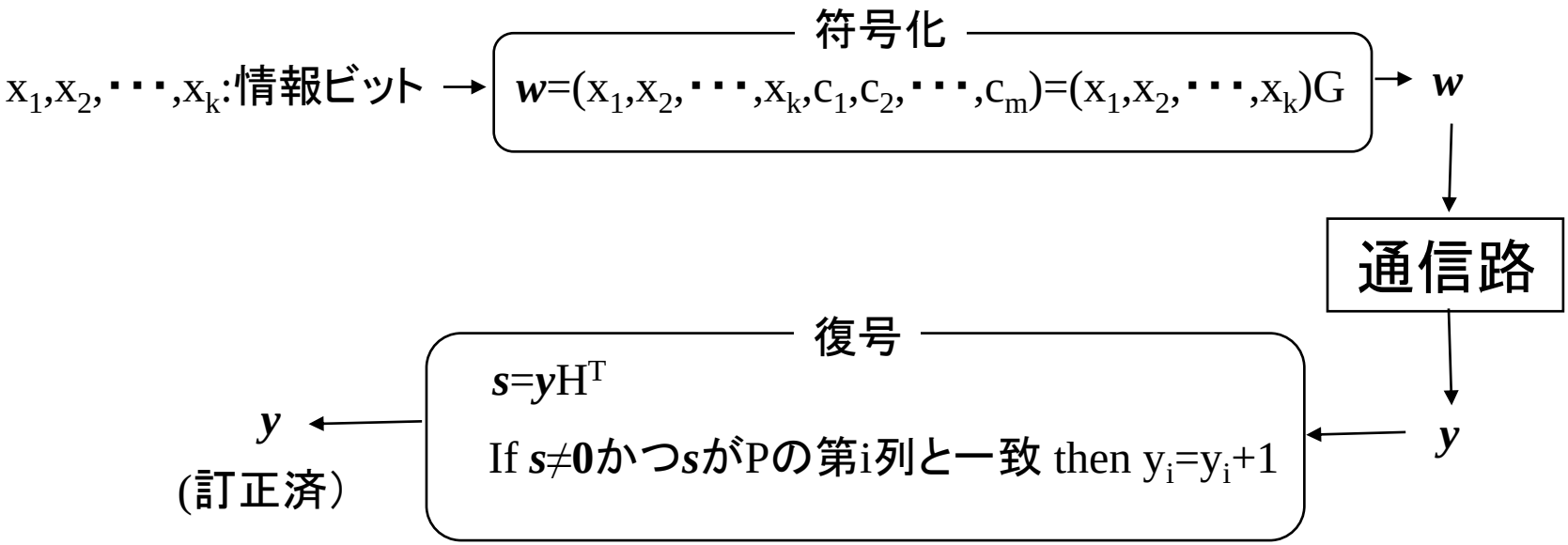
誤りパターン							シンドローム		
e_1	e_2	e_3	e_4	e_5	e_6	e_7	s_1	s_2	s_3
1	0	0	0	0	0	0	1	0	1
0	1	0	0	0	0	0	1	1	1
0	0	1	0	0	0	0	1	1	0
0	0	0	1	0	0	0	0	1	1
0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0

シンドロームのパターンから
1個の誤りパターンが判る！

ハミング符号の符号化と復号

検査行列 $H=\left(P\ E_m\right)$ $(m \times n)$　　生成行列 $G=\left(E_k\ P^T\right)$ $(k \times n)$

(P: $m \times k$行列,　E_m: $m \times m$単位行列,　$n=k+m$)



各ビットを並列に処理することが可能→並列符号器、並列復号器

生成行列

- (7,4)ハミング符号の符号語 w は情報記号のみで表すと

$$w=(x_1, x_2, x_3, x_4,\ x_1+x_2+x_3,\ \ x_2+x_3+x_4,\ \ x_1+x_2+x_4)$$

という形で書ける。ここで、

$$G=\left(\begin{array}{cccc:ccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right) \begin{array}{l} \leftarrow\!\text{--}\ x_1 \\ \leftarrow\!\text{--}\ x_2 \\ \leftarrow\!\text{--}\ x_3 \\ \leftarrow\!\text{--}\ x_4 \end{array}$$

という行列を考えれば、符号語 w を

$$w=x\,G$$

と書くことができる。ただし、$x=(x_1, x_2, x_3, x_4)$ である。

生成行列（generator matrix）

k 個の情報記号からなるベクトル x をかけたとき、
対応する符号語が生成されるような行列

(n, k)線形符号の生成行列は $k \times n$ 行列となる。

検査行列

- （パリティ）検査行列（parity check matrix）
  (7,4)ハミング符号のパリティ検査方程式の係数行列 $\boldsymbol{H}$

$$\boldsymbol{H}=\begin{array}{c}\begin{matrix}x_1 & x_2 & x_3 & x_4 & c_1 & c_2 & c_3\end{matrix}\\ \left(\begin{array}{ccccccc}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&1&0&1&0&0&1\end{array}\right)\end{array}$$

$$\boldsymbol{H}^{\mathrm{T}}=\begin{pmatrix}1&0&1\\1&1&1\\1&1&0\\0&1&1\\1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

これを用いれば、パリティ検査方程式は

$$\boldsymbol{w}\,\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{0}$$

($\boldsymbol{H}^{\mathrm{T}}$: $\boldsymbol{H}$ の転置行列、　$\boldsymbol{0}$: 全成分が0のベクトル)

と書ける。

(n, k)線形符号のパリティ検査行列は $(n-k)\times n$ 行列

- 検査行列 $\boldsymbol{H}$ を用いれば、(7,4)ハミング符号のシンドロームの計算式は

$$\boldsymbol{s}=\boldsymbol{y}\,\boldsymbol{H}^{\mathrm{T}}$$

と書ける。ここに $\boldsymbol{s}$ は $\boldsymbol{s}=(s_1, s_2, s_3)$ であり、シンドロームパターンまたは単にシンドロームと呼ばれる。$\boldsymbol{s}=(\boldsymbol{w}+\boldsymbol{e})\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{w}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{e}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{e}\boldsymbol{H}^{\mathrm{T}}$

一般のハミング符号

表：単一誤りに対するシンドローム

誤りパターン							シンドローム		
e_1	e_2	e_3	e_4	e_5	e_6	e_7	s_1	s_2	s_3
1	0	0	0	0	0	0	1	0	1
0	1	0	0	0	0	0	1	1	1
0	0	1	0	0	0	0	1	1	0
0	0	0	1	0	0	0	0	1	1
0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0

検査行列

$$\left(\begin{array}{cccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

単一誤りに対するシンドロームの行 ⟷ 検査行列の列

すべて異なる計2^m-1行
=0以外のm次元2元ベクトル　(m: 検査ビット数)

一般のハミング符号

検査行列 $\left(\begin{array}{c|c} \boldsymbol{p}_1\ \boldsymbol{p}_2 \cdots \boldsymbol{p}_k & \boldsymbol{e}_1\ \boldsymbol{e}_2 \cdots \boldsymbol{e}_m \end{array}\right)\Big\}$ m

0以外のm次元2元ベクトル($n=2^m-1$)

符号長：　$n=2^m-1$
情報ビット数：$k=2^m-1-m$
検査ビット数：$n-k=m$

ちょっと休憩

ハミング距離とハミング重み（1）

2つのn次元ベクトル$\boldsymbol{u}=(u_1,u_2,\cdots,u_n)$, $\boldsymbol{v}=(v_1,v_2,\cdots,v_n)$の間の**ハミング距離**$d_H(\boldsymbol{u},\boldsymbol{v})$

$$\overset{\text{def}}{\Leftrightarrow} \quad d_H(u,v)=\sum_{i=1}^{n}\delta(u_i,v_i) \qquad \text{ただし} \quad \delta(u,v)=\begin{cases} 0 & \text{if } u=v \\ 1 & \text{if } u\neq v \end{cases}$$

$d_H(u,v)$はuとvの対応する位置にある成分の対のうち、互いに異なるものの数

ハミング距離は**距離の3公理**を満たす。

距離の3公理

任意のn次元ベクトル$\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$に対して以下のことが成り立つ。

(i) $d_H(\boldsymbol{v}_1,\boldsymbol{v}_2)\geqq 0$であり、等号が成立するのは$\boldsymbol{v}_1=\boldsymbol{v}_2$のときに限る。

(ii) $d_H(\boldsymbol{v}_1,\boldsymbol{v}_2)=d_H(\boldsymbol{v}_2,\boldsymbol{v}_1)$

(iii) $d_H(\boldsymbol{v}_1,\boldsymbol{v}_2)+d_H(\boldsymbol{v}_2,\boldsymbol{v}_3)\geqq d_H(\boldsymbol{v}_1,\boldsymbol{v}_3)$ (三角不等式)

ハミング距離とハミング重み（2）

n次元ベクトル$\boldsymbol{v}$の**ハミング重み**または**重み** $w_H(\boldsymbol{v}) \overset{\text{def}}{\Leftrightarrow} w_H(\boldsymbol{v})=d_H(\boldsymbol{v},\boldsymbol{0})$

$w_H(\boldsymbol{v})$は$\boldsymbol{v}$の0でない成分の数

ハミング距離はハミング重みを用いて次のように表せる。

$$d_H(\boldsymbol{u},\boldsymbol{v})=w_H(\boldsymbol{u}-\boldsymbol{v})$$

（例） 符号語$\boldsymbol{w}$を送りt個の誤りが生じて$\boldsymbol{y}=\boldsymbol{w}+\boldsymbol{e}$が受信された場合

$$d_H(\boldsymbol{w},\boldsymbol{y})=w_H(\boldsymbol{e})=t$$

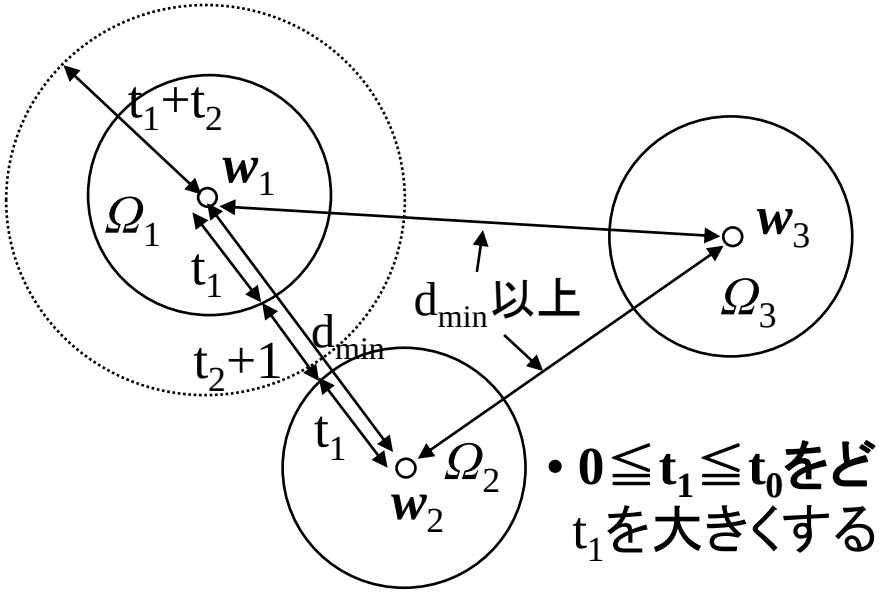
最小距離と誤り訂正能力(1)

符号Cの**最小ハミング距離**または**最小距離**(minimum distance) d_{min}

$$\overset{\text{def}}{\Leftrightarrow} \quad d_{min} = \min_{\boldsymbol{u}\neq\boldsymbol{v},\ \boldsymbol{u},\boldsymbol{v}\in C} d_H(\boldsymbol{u},\boldsymbol{v})$$

限界距離復号法

式 $d_{min} \geqq 2t_1+1$ を満たす整数t_1を定め、t_1以下の誤り訂正を行う復号法



- t_1の最大値$t_0 = \lfloor (d_{min}-1)/2 \rfloor$ を **誤り訂正能力**という。
- $t_2 = d_{min} - 2t_1 - 1$とおけば $t_1+1 \leqq t \leqq t_1+t_2$個の誤りは訂正はできないが検出は可能

- **$0 \leqq t_1 \leqq t_0$をどのように選ぶかは重要な問題**

t_1を大きくする→正しく復号される確率は増大するが誤って復号される確率も増大

検出さえできれば、再送要求などの救済措置が可能

最小距離と誤り訂正能力(2)

【例】$d_{min}=5$の符号による誤りの訂正と検出

t_1	訂正可能な誤り	訂正できないが検出可能な誤り
0	―	1～4個
1	1個	2～3個
2	2個	―

線形符号の最小距離＝**0**でない符号語のハミング重みの最小値

最小ハミング重みまたは**重み**

何故ならば $d_{min} = \min_{\boldsymbol{u}\neq\boldsymbol{v},\ \boldsymbol{u},\boldsymbol{v}\in C} d_H(\boldsymbol{u},\boldsymbol{v}) = \min_{\boldsymbol{u}\neq\boldsymbol{v},\ \boldsymbol{u},\boldsymbol{v}\in C} w_H(\boldsymbol{u}-\boldsymbol{v}) = \min_{\boldsymbol{w}\in C,\boldsymbol{w}\neq 0} w_H(\boldsymbol{w})$

[ハミング符号]　最小距離d_{min}＝3、 誤り訂正能力$t_0=1$

（例）(7,4)ハミング符号の場合　　最小距離d_{min}＝最小ハミング重み=3

[**水平垂直パリティ検査符号**]　最小距離$d_{min}=4$、 誤り訂正能力$t_0=1$

単一誤り訂正・2重誤り検出符号

(single-error-correcting/double-error-detecting code;SEC/DED符号)