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11回

第8章 通信路符号化法

8.1 線形符号の基礎

8.2 ハミング符号

[復習]通信路容量

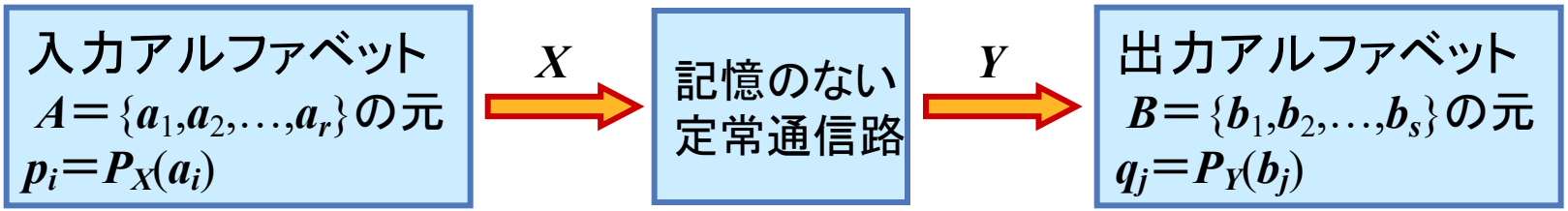


図6.1 記憶のない定常通信路のモデル

- 入力確率分布を $\boldsymbol{p}=(p_1, p_2,\ldots, p_r)$ と置く。このとき、

$$C=\max_{\boldsymbol{p}}\{\, \boldsymbol{I}(\mathrm{X};\mathrm{Y})\,\}$$

をこの通信路の通信路容量（channel capacity）と呼ぶ。

- 通信路に記憶がある場合には、エントロピーの場合と同様、拡大の手法を用いれば求められる。すなわち、長さ n の入力系列を X_n、出力系列を Y_n とし、$\mathbf{p}_n$ を X_n の確率分布とすれば

$$\boldsymbol{C}=\lim_{n\to\infty}\ \max_{\mathbf{p}_n}\{\boldsymbol{I}(\boldsymbol{X}_n;\boldsymbol{Y}_n)/\boldsymbol{n}\,\}$$

により定義される。

[復習]通信路容量の例

- 記憶のないs元出力、通信路行列$[p_{ij}]$の２重に一様な定常通信路

$$C = \log_2 s + \sum_{j=1}^{s} p_{1j} \log_2 p_{1j}$$

- ビット誤り率がpの2元対称通信路

$$C = 1 - \mathcal{H}(p)$$

- 誤り源S_{E}で表される加法的2元通信路

$$C = 1 - H(S_{\mathrm{E}})$$

[復習]通信路符号化定理

定理 7.4 [通信路符号化定理(Shannonの第2符号化定理)]

通信路容量がCである通信路に対し、$R<C$であれば、情報速度Rの符号で復号誤り率がいくらでも小さいものが存在する。$R>C$であれば、そのような符号は存在しない。

単一パリティ検査符号(1)

- 0,1 からなる長さ k の系列 $x_1x_2\cdots x_k$ を2元通信路を介して伝送したい。
  1個の誤りが生じた場合、それを検出（検知）するにはどうすればよいか？
- 単一パリティ検査符号
  系列に含まれる「1」の個数が偶数になるように、
  もう一つ記号を付け加えて送る符号化。
  付け加える記号を c とすると、

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

＋ は排他的論理和⊕

と書ける。この演算はまた、mod 2の演算と考えてもよい。

含まれる1の個数が
偶数の場合は 0
奇数の場合は 1

すなわち、通信路を通して送られる系列は

$$w = x_1x_2\cdots x_k\, c \quad \text{----------(1)}$$

となる。$x_1x_2\cdots x_k$ は情報を伝達するために用いられる記号なので、
情報記号（information symbol）あるいは2元記号の場合は情報ビットと呼ぶ。
c は誤り検出のために付加された記号なので、（パリティ）検査記号
（[parity] check symbol）、または（パリティ）検査ビットという。

- 式(1)は、 $w = (x_1, x_2, \cdots, x_k, c)$ と表すことがある。

単一パリティ検査符号(2)

- たとえば長さ $k=2$ の単一パリティ検査符号は、$w=(x_1, x_2, c)$ が
  $(0,0,0)$、$(0,1,1)$、$(1,0,1)$、$(1,1,0)$
  となるので、長さ 3、符号語数 4 の符号
  $\mathbf{C}=\{000, 011, 101, 110\}$
  を用いているとみなせる。
- 一般には、長さ k の系列に対し、長さが $k+1$ で、1の個数が偶数であるような全ての2元系列からなる符号長 $n=k+1$、符号語数 $M=2^k$ の符号を用いているとみなせる。
- 単一パリティ検査符号は、一つの誤りを検出できる。このように誤りの検出に用いられる符号を誤り検出符号（error-detecting code）と呼ぶ。

含まれる1の個数が偶数個

符号語に隣接する点は符号語になっていない

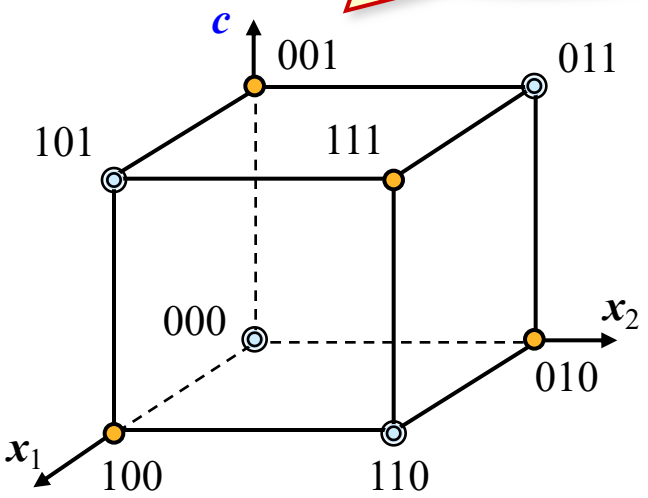


図7.1 単一パリティ検査符号の幾何的表現

組織符号

- 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 の符号を組織符号（systematic code）と呼ぶ。

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

- 符号長 n、情報記号数 k の組織符号を (n, k) 符号と書く。
  (n, k) 符号の効率は、

$$\eta = R / R_{\max} = (\log_2 M)/n = (\log_2 2^k)/n = k / n$$

  である。
- 単一パリティ検査符号は $(k+1, k)$ 符号であり、効率は $\eta = k / (k+1)$ である。

長さ n の系列で、情報記号数 k を送ることができるから

k を大きくとれば効率は上がるが、冗長度が低くなり信頼性は小さくなる

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

  線形符号の最も基本的な性質

  「任意の二つの符号語について、その成分ごとの和をとると、それがまた符号語になる」（線形符号となるための必要十分条件）

  [例]符号長3の単一パリティ検査符号 $\mathbf{C} = \{000, 011, 101, 110\}$

  $(0,1,1)+(1,0,1)=(0+1,1+0,1+1)=(1,1,0)$

- 長さnの系列$\boldsymbol{w}=(w_1, w_2, \cdots, w_n)$ 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:

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

  （符号語に含まれる1の個数が偶数）

  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）

シンドローム

- 符号語 w を送ったとき、

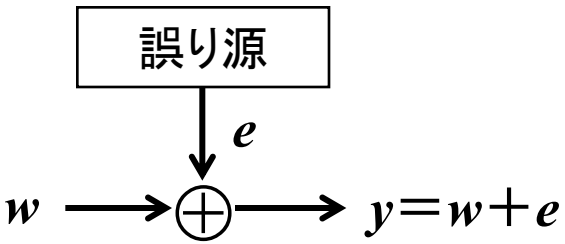
$$y=(y_1, y_2, \cdots, y_n)$$

が受信されたとする。これは右図のように w に

誤りパターン（error pattern）$e=(e_1, e_2, \cdots, e_n)$

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

が加わったものと見ることができる。すなわち、



図：誤りパターン e を用いた通信路のモデル

$$y=w+e$$
$$=(w_1+e_1, w_2+e_2, \cdots, w_n+e_n)$$

- シンドローム（syndrome）

受信語 y をパリティ検査方程式の左辺に代入した結果 s 。すなわち、

$$s=y_1+y_2+\cdots+y_n$$

符号語 w はパリティ検査方程式を満たすので、

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

誤りがない⇒$s=0$

1個の誤り⇒$s=1$

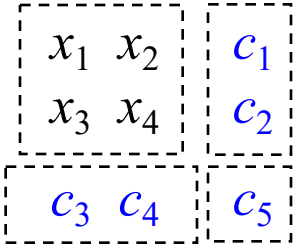
水平垂直パリティ検査符号

- 単一パリティ検査符号の問題点

  誤りの検出はできるが、どの情報ビットが誤っているのか、あるいは検査ビットが誤っているのかは判らない。⇒誤り訂正ができない！

- 水平垂直パリティ検査符号

  符号語が$(x_1,x_2,x_3,x_4,c_1,c_2,c_3,c_4,c_5)$の$(9,4)$組織符号を、以下のように生成する。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、



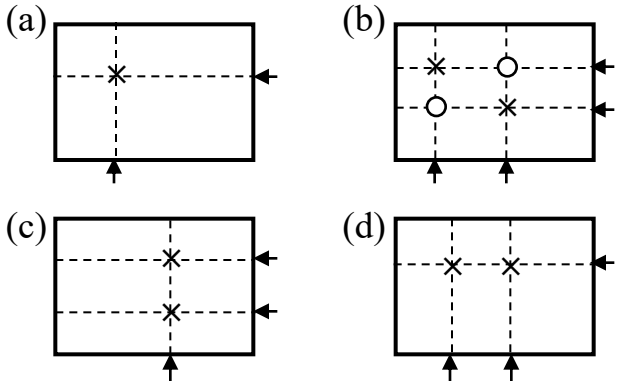
図：水平垂直パリティ検査符号

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

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。

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

この符号化により、1個の誤りが訂正できる。



図：単一誤りの訂正と2重誤りの検出

また、2個の誤りを検出することができる。このような符号を、誤り訂正検出符号、あるいは簡単に誤り訂正符号（error-correcting code）と呼ぶ。

(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。この符号は、情報ビットが4個であるから、符号語は $2^4 = 16$ 個ある。

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

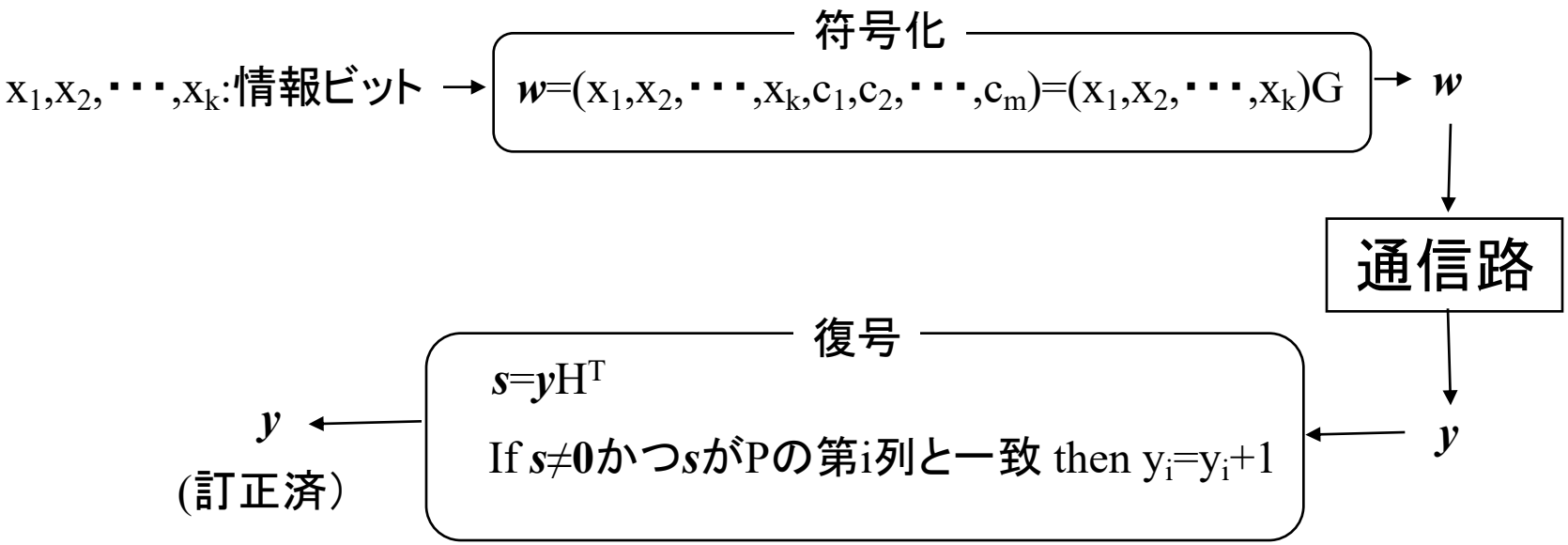
誤りパターン							シンドローム		
e_1	e_2	e_3	e_4	e_5	e_6	e_7	s_1	s_2	s_3
1	0	0	0	0	0	0	1	0	1
0	1	0	0	0	0	0	1	1	1
0	0	1	0	0	0	0	1	1	0
0	0	0	1	0	0	0	0	1	1
0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0

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×k行列,　E_m: m×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、

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

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。

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

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

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：$n-k=m$