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〕

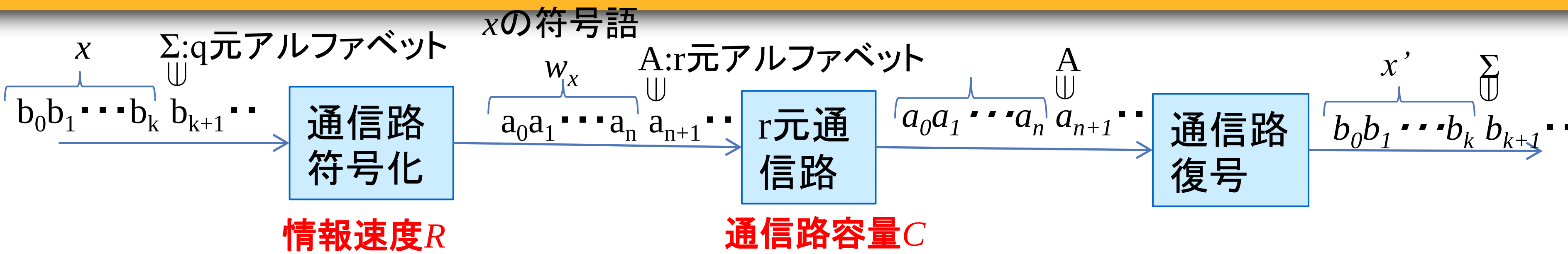
講義「情報理論」

第12回

第8章 通信路符号化法

8.1 線形符号の基礎

[復習]情報速度と通信路容量



符号Cの情報速度$R \overset{\text{def}}{\Longleftrightarrow}$

Cの符号語列に埋め込むことができる1通信路記号あたりの最大平均情報量

長さnの符号語M個の内，どれが起きたか知らせるのに使う符号の情報速度Rは

$$R = \frac{\log_2 M}{n} \quad \text{（ビット／記号）}$$

通信路の通信路容量$C \overset{\text{def}}{\Longleftrightarrow}$

通信路を介して伝送できる1通信路記号あたり平均情報量

長さnの入力系列X_nに対する定常通信路の出力列Y_nに対し

$$C = \lim_{n\to\infty} \max_{P_{X_n}} \frac{I(X_n;Y_n)}{n} \quad \text{(ビット/記号)}$$

無記憶定常通信路の場合は入力Xに対する出力Yに対し　$C = \max_{P_X} I(X;Y)$

[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。

情報速度Rと符号語の長さnを固定しても，情報速度Rで復号誤り率を最小にする符号の構成法は知られていない．

情報速度Rが速く復号誤り率が小さな符号の構成法は？⇒今日からの3回！

cf. 情報源符号化の場合
ブロック長を固定すると平均符号長を最小にする符号の構成法が知られている。
（ハフマン符号などのコンパクト符号の構成法)

単一パリティ検査符号の定義

[問題] 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$x_1x_2\cdots x_k$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を付け加えた符号語$\boldsymbol{w} = x_1x_2\cdots x_k\,c$に対応させる符号化

このとき、加える記号cは

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

と書ける。ただし、＋ は排他的論理和⊕
(つまり1+1=0. 剰余演算(mod 2) と考えてもよい)である。

$$\text{符号語}\boldsymbol{w} = \overbrace{x_1x_2\cdots x_k}^{\text{情報記号(information symbol)[2元の場合は情報ビット]}}\underbrace{c}_{\text{(パリティ)検査記号([parity] check symbol)[2元の場合は(パリティ)検査ビット]}}$$

符号語は$\boldsymbol{w} = (x_1, x_2, \cdots, x_k, c)$ とベクトルの形で表すこともある.

符号語長3の単一パリティ検査符号

長さ $k=2$ の単一パリティ検査符号の符号語 $w=(x_1, x_2, c)$：

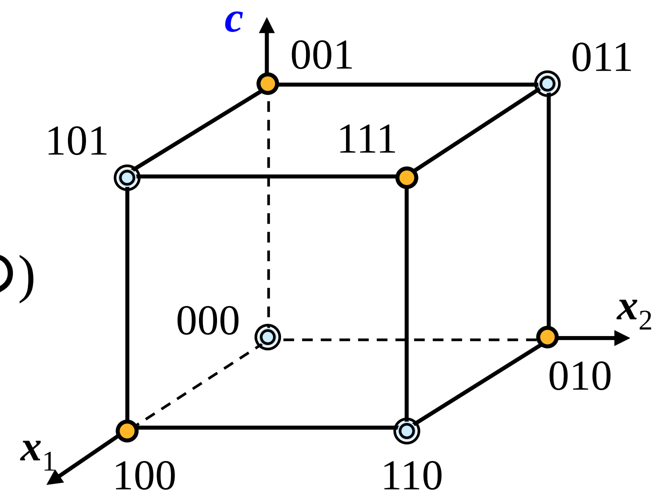
$$(0,0,0)、(0,1,1)、(1,0,1)、(1,1,0)$$

⇒長さ 3、符号語数 4 の符号

$$C=\{000, 011, 101, 110\}$$

を用いているとみなせる。

- 符号語はすべて1の数が偶数なので幾何的表現で隣のもの(1ビット違いのもの)はない。
⇒1ビット誤りは検出可能



符号語長3の 単一パリティ検査符号の幾何的表現

単一パリティ検査符号の性質

- 長さ k の系列を入力とする単一パリティ検査符号

  符号長 $n=k+1$、符号語数 $M=2^k$ ⇒情報速度$R=k/(k+1)$

- 一つの誤りの検出が可能な誤り検出符号（error-detecting code）

- $(k+1,k)$組織符号
- 線形符号

符号としては扱いやすい良い性質
(説明は次頁以降)

組織符号

組織符号（systematic code）

$\boldsymbol{k}$ 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

- (n, k) 二元符号の効率は、

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

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。

線形符号

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 $C = \{000, 011, 101, 110\}$

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

パリティ検査方程式

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$\boldsymbol{w}=(w_1, w_2, \cdots, w_{\boldsymbol{n}})$ 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.

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

（符号語に含まれる1の個数が偶数）

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$\boldsymbol{e}=\boldsymbol{w}+\boldsymbol{y}=(w_1+y_1, w_2+y_2,\cdots, w_n+y_n)$

(差だから$\boldsymbol{e}=\boldsymbol{w}-\boldsymbol{y}$なのでは？ 引き算は足し算の逆演算で定義されるので、$c-a$は$a+?=c$の?となる. 1+1=0の世界(2元のガロア体)では、$a+a=0$なので両辺にaを足すと$?=c+a$となり足し算と引き算が同じ$c-a=c+a$であることが確かめられる.)

- 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⇒$\boldsymbol{s}=0$,　 1個の誤り ⇒$\boldsymbol{s}=1$

ちょっと休憩

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。

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

$$\begin{array}{cc|c} x_1 & x_2 & c_1 \\ x_3 & x_4 & c_2 \\ \hline c_3 & c_4 & c_5 \end{array}$$

図: 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の性質

- 符号長 $n=9$、情報記号長4の(9,4)組織符号である.
- 符号語数$M=2^4$であるから情報速度$R=\frac{\log_2 M}{n}=\frac{4}{9}$, 効率$\eta=\frac{4}{9}$
- 検査記号が情報記号の線形式で計算できるので線形符号である.
- パリティ検査方程式
  長さ9の系列$\boldsymbol{w}=(w_1, w_2, \cdots, w_9)$ に対し

$$\begin{cases} w_1+w_2+w_5=0 \\ w_3+w_4+w_6=0 \\ w_1+w_3+w_7=0 \\ w_2+w_4+w_8=0 \\ w_1+w_2+w_3+w_4+w_9=0 \end{cases}$$

$$\begin{array}{cc|c} x_1 & x_2 & c_1 \\ x_3 & x_4 & c_2 \\ \hline c_3 & c_4 & c_5 \end{array}$$

- シンドローム
  受信語$\boldsymbol{y}=(y_1, y_2, \cdots, y_9)$に対するシンドローム$\boldsymbol{s}=(s_1, s_2, \cdots, s_5)$

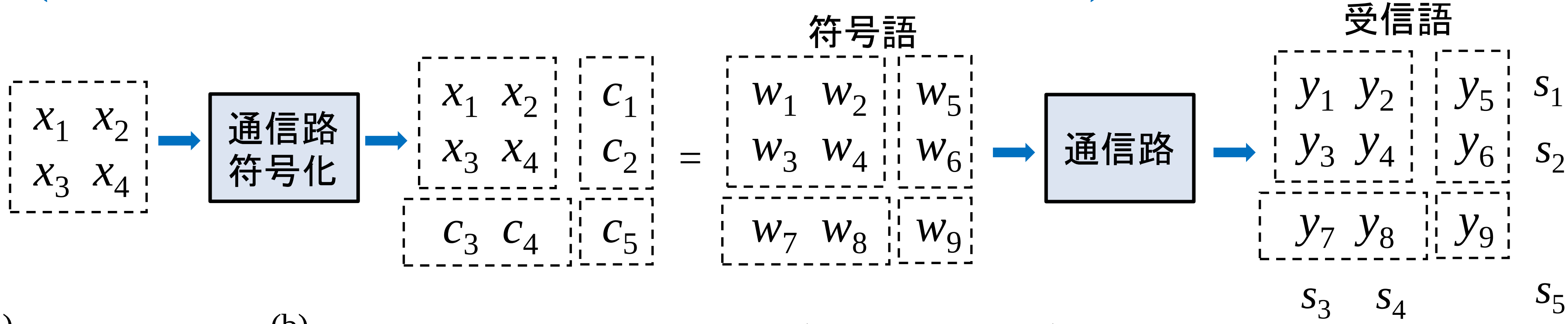
$$\begin{cases} s_1=y_1+y_2+y_5 \\ s_2=y_3+y_4+y_6 \\ s_3=y_1+y_3+y_7 \\ s_4=y_2+y_4+y_8 \\ s_5=y_1+y_2+y_3+y_4+y_9 \end{cases}$$

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。)



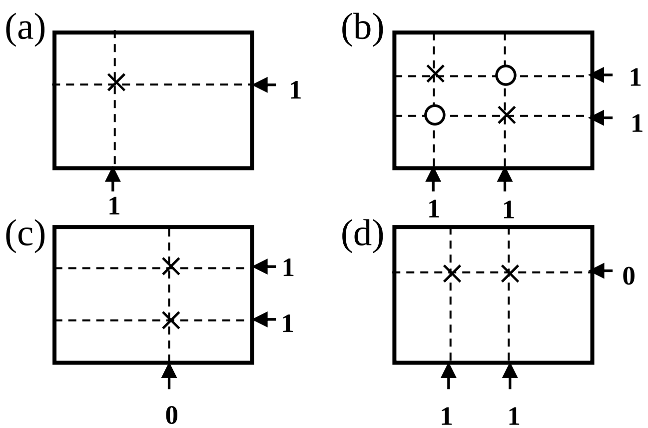


図: 単一誤りの訂正と2重誤りの検出

(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

一般の水平垂直パリティ検査符号

- 符号長$n=(m_1+1)(m_2+1)$, 情報記号長$k=m_1m_2$の
  $((m_1+1)(m_2+1), m_1m_2)$組織符号
- 符号語数$M=2^{m_1m_2}$,
  情報速度$R=\frac{\log_2 M}{n}=\frac{m_1m_2}{(m_1+1)(m_2+1)}$, 効率$\eta=\frac{m_1m_2}{(m_1+1)(m_2+1)}$
- 誤り(検出)訂正符号
  [訂正をする場合]1個の誤り訂正ができ、同時に2個の誤りまで検出可能
  [訂正しない場合]3個の誤りまで検出可能
  ※４個以上の誤りは検出できる場合もあるができない誤りもある

$$
\begin{array}{ccc|c}
x_1 & \cdots & x_{m_2} & c_1 \\
\vdots & & \vdots & \vdots \\
x_{(m_1-1)m_2+1} & \cdots & x_{m_1m_2} & c_{m_1} \\
\hline
c_{m_1+1} & \cdots & c_{m_1+m_2} & c_{m_1+m_2+1}
\end{array}
$$