講義「情報理論」

第14回　通信路符号化法(3)

情報理工学部門 情報知識ネットワーク研究室
喜田拓也

講義資料

2019/7/24

演習第2回解答例　http://www-ikn.ist.hokudai.ac.jp/~kida/lecture/IT2019_EX2ans.pdf

一般のハミング符号(おさらい)

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

検査行列の転置$\mathbf{H}^T$の作成：$\mathbf{H}^T$の上部k行には，各行が異なるビットパターンP^Tを，下部 m 行は $m \times m$ 単位行列E_mを配置する

生成行列を$\mathbf{G} = [E_k\ P^T]$とする

P^Tをコピー

$$\text{検査行列 } \mathbf{H}^T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix}1\\1\\1\\0\end{matrix}}\right\} P^T \\ \left.\vphantom{\begin{matrix}1\\0\\0\end{matrix}}\right\} E_m \end{array} \Rightarrow \text{生成行列 } \mathbf{G} = \Bigg[\underbrace{\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}}_{E_k} \quad \begin{matrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{matrix}\Bigg]$$

この部分は固定的

情報ビット

$\boldsymbol{x} = x_1, x_2, \dots, x_k$ → 符号化 $\boldsymbol{w} = (x_1, x_2, \dots, x_k, c_1, c_2, \dots, c_m) = \boldsymbol{x}\mathbf{G}$ → $\boldsymbol{w}$ → 通信路 → $\boldsymbol{y}$

復号 $\boldsymbol{s} = \boldsymbol{y}\mathbf{H}^T$ を計算．もし$\boldsymbol{s} \neq \mathbf{0}$かつ$\boldsymbol{s}$が$\mathbf{H}^T$の第$i$行目と一致なら，$y_i = y_i + 1$と訂正 → $\boldsymbol{y}'$ → $\boldsymbol{x}$

訂正済の系列

最小距離と誤り訂正検出能力(おさらい)

符号Cの最小ハミング距離(最小距離)$d_{\min}$の定義:

$$d_{\min} = \min_{\boldsymbol{u} \neq \boldsymbol{v};\, \boldsymbol{u},\boldsymbol{v} \in \mathrm{C}} \{d_{\mathrm{H}}(\boldsymbol{u}, \boldsymbol{v})\} .$$

定理8.5

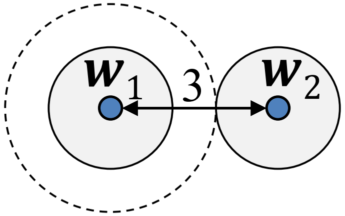
線形符号の最小距離は，符号の最小ハミング重みに一致する.

$$d_{\min} = \min_{\substack{\boldsymbol{u} \neq \boldsymbol{v};\\ \boldsymbol{u},\boldsymbol{v} \in \mathrm{C}}} d_{\mathrm{H}}(\boldsymbol{u}, \boldsymbol{v}) = \min_{\substack{\boldsymbol{u} \neq \boldsymbol{v};\\ \boldsymbol{u},\boldsymbol{v} \in \mathrm{C}}} w_{\mathrm{H}}(\boldsymbol{u} - \boldsymbol{v}) = \min_{\boldsymbol{w} \in \mathrm{C}} w_{\mathrm{H}}(\boldsymbol{w}) .$$

<ins>ハミング符号の場合</ins>

最小距離$d_{\min} = 3$，誤り訂正能力 $t_0 = 1$

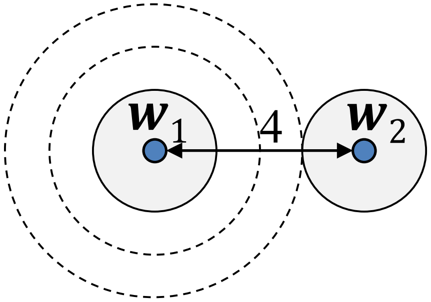
(7,4)ハミング符号の場合，最小距離 $d_{\min}$＝最小ハミング重み＝3



<ins>(9,4)水平垂直パリティ検査符号の場合</ins>

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



今日の内容

8.3 巡回符号

2元系列の多項式表現

巡回符号では，系列長 n の2元系列を，0,1の2値を係数とする多項式に対応付け，そのような多項式の演算に基づいて符号化や復号を行う

0,1を成分とするn次元ベクトル$\boldsymbol{v} = (v_{n-1}, v_{n-2}, \cdots, v_1, v_0)$を

$$F(x) = v_{n-1}x^{n-1} + v_{n-2}x^{n-2} + \cdots + v_1 x + v_0$$

で表す．これを2元系列の多項式表現という

符号長nの符号は，$n-1$次以下の多項式の集合として表せる

このとき，各符号語に対応する多項式を符号多項式と呼ぶ

例8.7）　$\boldsymbol{v} = (1,0,1,1,0)$

$$\Updownarrow$$

$$F(x) = x^4 + x^2 + x$$

$1 \cdot x^4 + 0 \cdot x^3 + 1 \cdot x^2 + 1 \cdot x + 0$

変数 x に大きな意味はない
単に係数を区別するためのもの

Try 問8.14

2元系列の多項式表現の演算

二つの多項式の加算は，通常の多項式の計算と同様，対応する次数どうしをそれぞれ加算すればよい．ただし，係数は mod 2 の計算なので，係数どうしの排他的論理和演算となる

例）$(x^4+x^2+1)+(x^4+x^3+x+1)=x^3+x^2+x$

mod 2の計算では，加算と減算は同じ意味を持つ

例）$x^3-x^2+1=x^3+x^2+1$

変数 x どうしの乗算は，次数は通常通りに加算される

例）$x\cdot x=x^2$, $(x+1)(x+1)=x^2+1$

（$x+x=0$に注意）

多項式を x 倍すると，係数は1ビット左にシフトする

例）$(x^3+x^2+1)\times x=x^4+x^3+x$

(0,1,1,0,1)　　(1,1,0,1,0)

Try 問8.15

巡回符号とは？

定義8.14

最大次数m($m > 0$)で定数項が 1 の任意の多項式$G(x)$を選ぶ.

$$G(x) = x^m + g_{m-1}x^{m-1} + \cdots + g_1 x + 1\,.$$

$g_1, \cdots, g_{m-1}$は0 か 1

長さn ($> m$) のすべての2元系列に対応する 2^n 通りの多項式のうち, $G(x)$ で割り切れる多項式だけをすべて取り出し, それらを符号語とした符号のことを巡回符号と呼ぶ. 検査ビット長は m, 情報ビット長は$n - m$となる. このとき$G(x)$を生成多項式とよぶ.

巡回符号では, 任意の符号語は $W(x) = Q(x)G(x)$ という形の多項式(符号多項式)に対応づけられる

$Q(x)$は$n - m - 1$ 次以下の任意の多項式

巡回符号の特徴:

- 線形符号である
- 符号長が長くても符号化･シンドローム計算の装置化が比較的容易
- 誤り検出能力に優れる. 特にバースト誤りに対する理論的保証がある

$G(x)$の符号多項式の例（例8.9）

$n = 7, m = 4$，次の多項式

$$G(x) = x^4 + x^2 + x + 1$$

を生成多項式とする巡回符号の符号語を求めてみよう．

このとき，$G(x)$により作られる符号Cの符号多項式は，符号長が $n = 7$ なので，

$$W(x) = w_6 x^6 + \cdots + w_1 x + w_0$$

と書ける．

項は7つ．最大次数は6

このうち，$G(x)$で割り切れる多項式は，$Q(x)$から逆算すると，右の表8.3のとおり．

足す項が偶数個だと消える

表8.3. $G(x) = x^4 + x^2 + x + 1$ の倍多項式と対応する符号語

$Q(x)$	$W(x) = Q(x)G(x)$	$\boldsymbol{w}$
0	0	0000000
1	$x^4 + x^2 + x + 1$	0010111
x	$x^5 + x^3 + x^2 + x$	0101110
$x + 1$	$x^5 + x^4 + x^3 + 1$	0111001
x^2	$x^6 + x^4 + x^3 + x^2$	1011100
$x^2 + 1$	$x^6 + x^3 + x + 1$	1001011
$x^2 + x$	$x^6 + x^5 + x^4 + x$	1110010
$x^2 + x + 1$	$x^6 + x^5 + x^2 + 1$	1100101

0,1 を係数とする多項式の乗算

(a)

$$\begin{array}{rrrrrrrr}
 & & & x^4 & & +x^2 & +x & +1 \\
\times) & & & & & x^2 & +x & +1 \\ \hline
 & & & x^4 & & +x^2 & +x & +1 \\
 & & x^5 & & +x^3 & +x^2 & +x & \\
 & x^6 & & +x^4 & +x^3 & +x^2 & & \\ \hline
 & x^6 & +x^5 & & & +x^2 & & +1
\end{array}$$

(b)

$$\begin{array}{rr}
 & 10111 \\
\times) & 111 \\ \hline
 & 10111 \\
 & 10111 \\
 & 10111 \\ \hline
 & 1100101
\end{array}$$

巡回符号は線形符号

[証明]
　任意の二つの符号多項式$W_1(x)$と$W_2(x)$の和を考える.
いま,

$$W_1(x) = Q_1(x)G(x),$$

$$W_2(x) = Q_2(x)G(x)$$

とおくと,

$$W_1(x) + W_2(x) = [Q_1(x) + Q_2(x)]G(x)$$

となるので, これは$G(x)$の倍多項式である.

$G(x)$で割れるってこと

よって, $W_1(x) + W_2(x)$も符号多項式となるので, 対応する系列も符号語となる. すなわち, 任意のふたつの符号語の和が符号語となるので, 巡回符号は線形符号である.【証明終】

線形符号の必要十分条件

巡回符合の符号化方法

$n-m$個の情報ビット列$\boldsymbol{v}=(v_{n-m-1},\cdots,v_0)$を長さ$n$の符号語に符号化する．まず，情報ビットを係数とする多項式

$$V(x)=v_{n-m-1}x^{n-m-1}+\cdots+v_1x^1+v_0$$

に x^m を掛け，生成多項式$G(x)$で割る．

$G(x)$の次数はmなので，剰余多項式を

$$C(x)=c_{m-1}x^{m-1}+\cdots+c_1x+c_0$$

$m-1$次

とおき，また，$Q(x)$ を商多項式とすると，

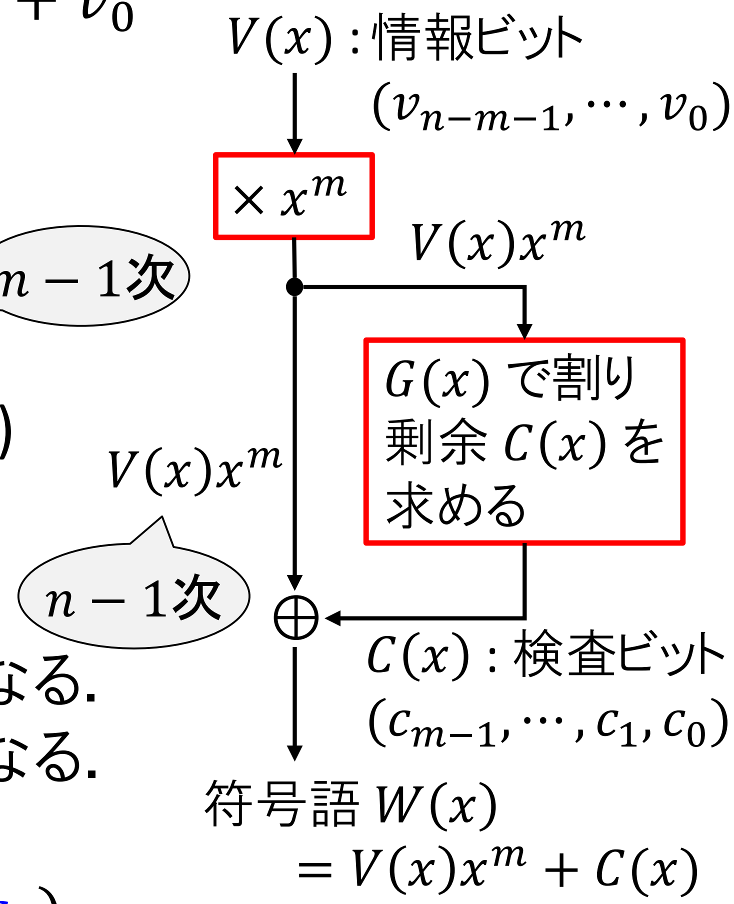
$$V(x)x^m=Q(x)G(x)+C(x)\;.\cdots(1)$$

ここで，

$$W(x)=V(x)x^m+C(x)$$

とおくと，式(1)から$W(x)=Q(x)G(x)$となる．
よって，$W(x)$は符号C の符号多項式となる．
$W(x)$をベクトルの形で表すと，

$$\boldsymbol{w}=(v_{n-m-1},\cdots,v_1,v_0,c_{m-1},\cdots,c_1,c_0).$$



巡回符号の符号化の例（例題8.2）

生成多項式が$G(x) = x^4 + x^2 + x + 1$，符号長$n = 7$の巡回符号において，情報ビット (1,1,0) を符号化せよ．

生成多項式は 4次なので，検査ビット数は 4.
情報ビットを係数とする多項式は $V(x) = x^2 + x$で，これに x^4 を掛けると，$V(x)x^4 = x^6 + x^5$ となる．
これを$G(x)$で割ると剰余は$C(x) = x^2 + 1$ となる．

よって，符号多項式は $W(x) = V(x)x^4 + C(x) = x^6 + x^5 + x^2 + 1$ となり，符号語は (1,1,0,0,1,0,1) となる．

表．0,1を係数とする多項式の割り算

(a)

$$
\begin{array}{r|l}
 & x^2 + x + 1 \\ \hline
x^4 + x^2 + x + 1 & x^6 + x^5 \\
 & x^6 \quad + x^4 + x^3 + x^2 \\ \cline{2-2}
 & \quad x^5 + x^4 + x^3 + x^2 \\
 & \quad x^5 \quad + x^3 + x^2 + x \\ \cline{2-2}
 & \qquad x^4 \qquad + x \\
 & \qquad x^4 \quad + x^2 + x + 1 \\ \cline{2-2}
 & \qquad\qquad x^2 \quad + 1
\end{array}
$$

(b)

$$
\begin{array}{r|r}
 & 111 \\ \hline
10111 & 1100000 \\
 & 10111 \\ \cline{2-2}
 & 11110 \\
 & 10111 \\ \cline{2-2}
 & 10010 \\
 & 10111 \\ \cline{2-2}
 & 101
\end{array}
$$

Try 練習問題8.2

ちょっと休憩

なぜ「巡回」符号と呼ばれるのか？

定理8.7

ある生成多項式$G(x)$から構成した符号長nの巡回符号において，多項式$x^n - 1$が$G(x)$で割り切れるとする．このとき，この巡回符号の任意の符号語

$$\boldsymbol{w} = (w_{n-1}, w_{n-2}, \cdots, w_1, w_0)$$

を左に1ビット巡回させた系列

$$\boldsymbol{w}' = (w_{n-2}, w_{n-3} \cdots, w_0, w_{n-1})$$

もまた，この符号の符号語に含まれている．

$$\boldsymbol{w} = (w_{n-1}, w_{n-2}, w_{n-3}, \cdots, w_1, w_0)$$

$$\boldsymbol{w}' = (w_{n-2}, w_{n-3}, \cdots, w_1, w_0, w_{n-1})$$

【証明は教科書参照】

本来の巡回符号は，多項式 $x^n - 1$ が $G(x)$ で割り切れなければならない．これが成立しないものを擬巡回符号と呼ぶ

$G(x)$で生成される符号は，この条件が成立していなくてもほとんど同様に扱えるため，擬巡回符号も含めて単に巡回符号と呼ぶ

巡回符号の誤り検出・訂正能力

右の表8.3に示した巡回符号の例では，全ゼロ以外のすべての符号語のハミング重みが 4 であることから，この符号の最小距離は 4 であることが分かる

よって，この巡回符号は単一誤り訂正と2重誤り検出可能

訂正を行わない場合には，3重誤り検出可能

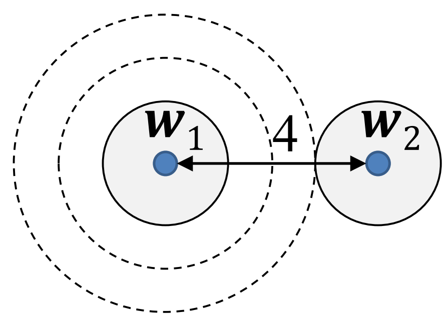
では，一般の巡回符号はどのような誤り検出・訂正能力をもっているのだろうか？

表8.3. $G(x) = x^4 + x^2 + x + 1$ の倍多項式と対応する符号語

$Q(x)$	$W(x) = Q(x)G(x)$	$\boldsymbol{w}$
0	0	0000000
1	$x^4 + x^2 + x + 1$	0010111
x	$x^5 + x^3 + x^2 + x$	0101110
$x + 1$	$x^5 + x^4 + x^3 + 1$	0111001
x^2	$x^6 + x^4 + x^3 + x^2$	1011100
$x^2 + 1$	$x^6 + x^3 + x + 1$	1001011
$x^2 + x$	$x^6 + x^5 + x^4 + x$	1110010
$x^2 + x + 1$	$x^6 + x^5 + x^2 + 1$	1100101

⇩

最小ハミング重み = 4 ⇔ 符号の最小距離 = 4



生成多項式$G(x)$の周期について

定義8.15

ある生成多項式$G(x)$が与えられたときに，$x^n - 1$ $(n = 1, 2, 3, \dots)$ という形の多項式が $G(x)$ で割り切れるかどうかを調べ，これが割り切れるような最小の n を，多項式 $G(x)$ の周期と呼ぶ．

【例8.10】生成多項式 $G(x) = x^4 + x^2 + x + 1$ の周期を調べる．
$G(x)$ は4次式なので，$n = 4$ 以下では明らかに割り切れない．$n = 5, 6$ のときを計算すると，

$$x^5 + 1 = x(x^4 + x^2 + x + 1) + (x^3 + x^2 + x + 1)$$
$$x^6 + 1 = (x^2 + 1)(x^4 + x^2 + x + 1) + (x^3 + x)$$

となり，やはり割り切れない．しかし，$n = 7$ のときに初めて

$$x^7 + 1 = (x^3 + x + 1)(x^4 + x^2 + x + 1)$$

となって割り切れる．よって，この多項式$G(x)$の周期は 7 である．

Try 問8.17

$G(x)$の周期と符号の最小距離の関係

定理8.8

符号長 n の巡回符号において，n より短い周期の生成多項式 $G(x)$ を用いると，符号の最小距離が 2 になってしまう．$G(x)$の周期が n 以上であれば，最小距離は 3 より小さくならない．

【証明は教科書参照】

周期pの生成多項式を選べば，符号長pまでの良い符号が作れる

生成多項式 $G(x) = x^4 + x^2 + x + 1$ の周期は 7．したがって，符号語長 n を 8 以上にすると，最小距離が 2 となり誤り訂正ができない．【例8.10】の例では，符号語長を 7 としており，実際，最小距離は 4 である．

$G(x)$の項数と符号の最小距離の関係

定理8.9

巡回符号において，生成多項式$G(x)$の項数をdとすると，符号の最小距離を d より大きくはできない．さらに d が偶数ならば，すべての符号語のハミング重みは必ず偶数となる.

【証明は教科書参照】

巡回符号では $G(x)$ 自体が符号語に含まれることと，$G(1) = 0$ ならば $W(1) = Q(1)G(1) = 0$ であることから証明できる.

【例8.11】先の例では，生成多項式 $G(x) = x^4 + x^2 + x + 1$ の周期は 7 で符号長と同じである．よって，（定理8.7より）符号の最小距離は 3 以上である．$G(x)$ の項数は偶数なので，符号語のハミング重みは必ず偶数であり，最小距離は 4 以下となる. しかも $G(x)$ の項数がちょうど 4 であるため，符号の最小距離もちょうど 4 となる.

バースト誤りの検出と訂正

長さ n のブロック内に1回のバースト誤りがある場合を考える

l

00・・・・・・・001＊＊・・・＊100・・・・・・・00

n

ギルバートモデルや
フリッチマンモデルも含む

＊は0,1任意であることを示す

ある長さまでのバースト誤りを，すべて訂正（あるいは検出）するような符号をバースト誤り訂正（検出）符号という

ある符号 C が訂正（検出）できる最大のバースト長を，符号 C のバースト誤り訂正（検出）能力とよぶ

バースト誤り検出能力が l_0 ⇔ 任意の符号語w_1に長さ l_0 以下の任意のバースト誤りパターンe を加えた$w_1 + e$が，別の符号語w_2にならない

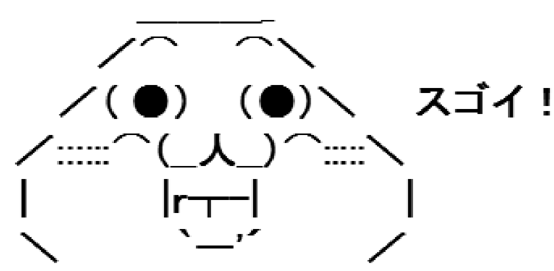
バースト誤り訂正能力が l_0 ⇔ 任意の符号語w_1に長さ l_0 以下の任意のバースト誤りパターンe_1を加えた$w_1 + e_1$が，別の符号語w_2に任意のバースト誤りパターンe_2を加えた$w_2 + e_2$と一致しない

バースト誤りに対する理論的保証

定理8.10

巡回符号において，生成多項式$G(x)$の次数をmとすると，長さmの区間内で発生する多重誤りはすべて検出可能である．

連続して発生する誤りなら，長さmまでのどんな誤りでも検出できる！



定理8.10の証明

符号長をnとする．長さmの多重誤りパターンを多項式で表現すると，ある整数 j（$0 \leq j$ かつ $j + m < n$)に対して，

$$E(x) = x^j(e_{m-1}x^{m-1} + e_{m-2}x^{m-2} + \cdots + e_1x + e_0)$$

となる．$e_i \in \{0,1\}$は多重誤りパターンを表す係数である．符号語$W(x)$に対する受信語$Y(x) = W(x) + E(x)$が別の符号語になっていなければ誤りを検出できる．すなわち，$E(x)$が$G(x)$で割り切れなければよい．$G(x)$はxを因数として持たないので，$E(x)$が$G(x)$で割り切るのは，二つ目の項である

定数項が1だから！

$$e_{m-1}x^{m-1} + e_{m-2}x^{m-2} + \cdots + e_1x + e_0$$

が$G(x)$で割り切れるときのみである．いま，$G(x)$の次数がmとすると，この部分は最大次数が$m-1$であるため割り切れることはない．したがって，$Y(x)$は符号語とならないので必ず$E(x)$のような誤りは検出できる．

CRC方式

巡回符号による誤り検出方式は，CRC（cyclic redundancy check; 巡回冗長検査）方式と呼ばれ，広く実用に用いられている

CRC方式には，**CCITT**（国際電信電話諮問委員会）の勧告による

$$G(x) = x^{16} + x^{12} + x^5 + 1$$

という生成多項式がよく用いられる

この生成多項式の周期は 32767 なので，符号長 32767 以下の符号の場合，定理8.8と定理8.9より最小距離は 4 となる．したがって，3 個以下の任意の誤りを検出できる

さらに，定理8.10より，長さ 16 以下の任意のバースト誤りは検出可能となる．長さ17以上のバースト誤りには検出不可能なものも混じるが，その大部分は検出可能であることがわかっている

イーサネットの規格でのCRC方式

480～12112 bit	32 bit
ヘッダー ＋ データ	検査ビット

図8.8 イーサネットのパケット構成

イーサネットの規格（IEEE802.3）でもCRCが使われている（CRC-32）

イーサネットでは約 500～12000 ビットで1つのパケットを構成し，パケットの末尾に 32 ビットの検査ビットが追加されている

生成多項式は次のとおり.

$$G(x) = x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10}$$
$$+x^{8} + x^{7} + x^{5} + x^{4} + x^{2} + x + 1\,.$$

パケット長の範囲内では符号の最小距離は 4. したがって，任意の3重誤りまではすべて検出可能

長さ 32 までの連続区間内で発生した多重誤りを全て検出可能

同じ符号化が，MPEG-2, zlib, PNG などでも利用されている

巡回ハミング符号

0,1 を係数とするm次多項式の周期は，最大$2^m - 1$であることが知られている．この最大周期を持つ多項式を原始多項式といい，各次数について原始多項式が存在することが証明されている．

m次の原始多項式を生成多項式とする符号長$n = 2^m - 1$の巡回符号を考えよう！

符号長が周期と一致するので，この符号の最小距離は3以上．また，ちょうど3になることも簡単に確認できる．この巡回符号は，符号長$n = 2^m - 1$，情報ビット数$2^m - 1 - m$，検査ビット数mのハミング符号になることが知られている．このようなハミング符号を巡回ハミング符号と呼ぶ．

表．20次までの原始多項式の例

次数	原始多項式	次数	原始多項式
1	$x+1$	11	$x^{11}+x^2+1$
2	x^2+x+1	12	$x^{12}+x^6+x^4+x+1$
3	x^3+x+1	13	$x^{13}+x^4+x^3+x+1$
4	x^4+x+1	14	$x^{14}+x^{10}+x^6+x+1$
5	x^5+x^2+1	15	$x^{15}+x+1$
6	x^6+x+1	16	$x^{16}+x^{12}+x^3+x+1$
7	x^7+x+1	17	$x^{17}+x^3+1$
8	$x^8+x^4+x^3+x^2+1$	18	$x^{18}+x^7+1$
9	x^9+x^4+1	19	$x^{19}+x^5+x^2+x+1$
10	$x^{10}+x^3+1$	20	$x^{20}+x^3+1$

今日のまとめ

8.2.6 バースト誤りの検出と訂正

8.3 巡回符号

- 8.3.1 2元系列の多項式表現
- 8.3.2 巡回符号の構成法
- 8.3.4 巡回符号による誤り検出･訂正能力
  - $G(x)$ の項数，周期，次数と符号の最小距離の関係
  - CRC方式
- 8.3.5 巡回ハミング符号

補足資料：

- 巡回符号の符号器のための回路構成 (8.3.3節の補足)