基本情報技術 I

6. 通信と符号理論

菊池浩明

講義目標

- 教科書
  - 1-4 情報・通信に関する理論
    - » 1. 情報理論
    - » 2. 符号理論

1. 情報理論

- 1. 情報圧縮
  - ファイルの圧縮（file.zip, JPEG)
  - 地上波ディジタル放送 (MPEG, MP3)
  - ハフマン符号
- 2. 誤り訂正技術
  - メディア: CD, DVD, Blu-ray Disk
  - BS放送
  - パリティ検査符号

http://gigazine.jp/img/2007/09/27/sony_oneseg_walkman

http://ja.wikipedia.org/wiki/Blu-ray_Disc

生起表

- 生起表

情報源符号	確率	C1	C2
A	½	00	0
B	¼	01	10
C	¼	10	110

- どちらが「よい」符号化か？

例1

- 文字A-Eを符号化した時のビット表記と出現頻度である．一文字当たりの平均ビット数はいくらか？

文字	ビット表記	出現確率
A	0	0.5
B	10	0.3
C	110	0.1
D	1110	0.05
E	1111	0.05

ハフマン(Huffman)符号

- 可変長符号化
  - 生起確率の高いものは短く
  - エントロピーを最大化
- コンパクト符号
  - 平均符号長を最小化する符号
- 瞬時符号
  - 語の受信により即復号化できる

平均符号長

- 定義
  - 情報源記号 $a_1,a_2,\ldots,a_M$ がそれぞれ確率 $p_1,p_2,\ldots,p_M$ で生起するとき，

    $$L=\sum_{i=1}^{M}\ell_i p_i$$

    を情報源Sの平均符号長という．ただし，$\ell_1, \ell_2,\ldots, \ell_M$は各符号語の長さである．
- 例
  - $L = p_A \times 1 + p_B \times 2 + p_C \times 3$

    $= 1/2 + 2/4 + 3/4 = 1.75$ [bit/symbol]

例1の解

- 平均符号長
  - $L = \Sigma \ell_i P_i = 1*0.5 + 2*0.3 + 3*0.1 + 4*(0.05+0.05) = 1.8$ bit/文字

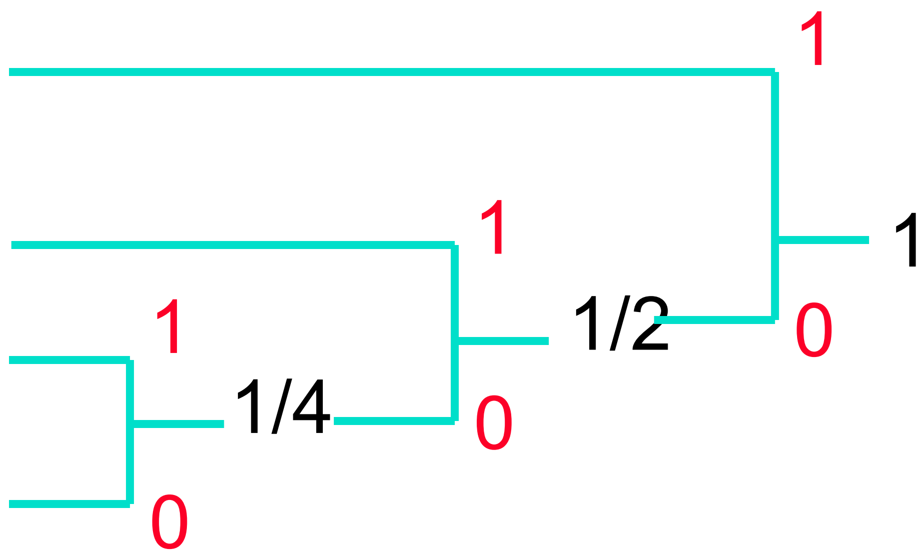
文字	ビット表記	長さ ℓ	出現確率 P
A	0	1	0.5
B	10	2	0.3
C	110	3	0.1
D	1110	4	0.05
E	1111	4	0.05

ハフマン符号

- Step 1
  - 最も小さな生起確率のシンボルを二つ選ぶ
- Step 2:
  - 二つをあわせて新しいシンボルを作る. その生起確率は両方の和.
- Step 3
  - シンボルがなくなるまで2を繰り返す

例

シンボル	確率
A	1/2
B	1/4
C	1/8
D	1/8



符号の木

1

01

001

000

符号語

圧縮率

- メッセージ (M=4)
  - AAAABBCD (= ABAACBDA)
- 固定長符号化
  - (log 4 = 2 bit) x 8文字 = 16 bit
- ハフマン符号化
  - $L = 1 \text{ bit} \times \frac{1}{2} + 2 \times \frac{1}{4} + 3\ (1/8+1/8) = 14/8$
  - $4 + 2 \times 2 + 3 \times 1 + 3 \times 1 = 14 \text{ bit} = L \times 8$
- 圧縮率
  - ハフマン／固定長 = 14/16 = 0.875

モールス信号

A	・—	I	・・	Q	——・—	Y	—・・—
B	—・・・	J	・———	R	・—・	Z	——・・
C	—・—・	K	—・—	S	・・・	1	・————
D	—・・	L	・—・・	T	—	2	・・———
E	・	M	——	U	・・—	3	・・・——
F	・・—・	N	—・	V	・・・—	4	・・・・—
G	——・	O	———	W	・——	5	・・・・・
H	・・・・	P	・——・	X	—・・—	6	—・・・・

例2

- 次の生起表で与えられる情報源Sがある.
  - ハフマン符号C_2を求めよ
  - C_2の平均符号長L_2を求めよ
  - (C_2の圧縮率 Rを求めよ)

シンボル	A	B	C	D	E	F
確率	0.6	0.12	0.07	0.03	0.08	0.1

例2.2)

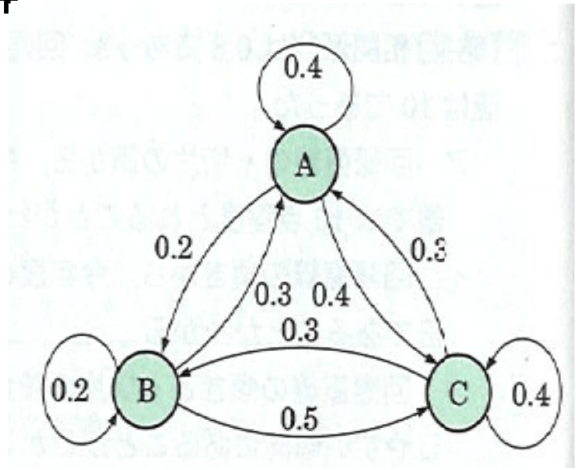
- (1)次の表で与えられる情報源をハフマン符号で符号化した．文字と符号語の組み合わせとして正しいものを全て選べ．
  - ❑ア A=1　C = 10
  - ❑イ A=1　C = 01
  - ❑ウ A=1　C = 001
  - ❑エ A=1　C = 110

文字	A	B	C	D
生起確率	0.4	0.1	0.3	0.2

- (1)のハフマン符号の平均符号長を求めよ

マルコフ情報源 (低頻度)

- マルコフ情報源(一次)
  - ある時点で発生した事象の生起確率がそれ以前(1回)の事象の状態に依存する記憶のある情報源.
  - P(次|前) = P(A|A) = 0.4
  - P(B|A) = 0.2
  - P(C|A) = 0.4
    (0.4+0.2+0.4 = 1.0)



例3

- ある地方の天気の移り変わりが次の表の単純マルコフ過程であるとする．雨の二日後が晴れである確率を求めよ．雨の二日後が曇りの確率を求めよ．

単位　%

	翌日晴れ	翌日曇り	翌日雨
晴れ	40	40	20
曇り	30	40	30
雨	30	50	20

2. 符号理論

- 誤り検出符号
  - CRC符号
    （パケット通信）
  - パリティ検査符号
    （シリアル通信）
- 誤り訂正符号
  - 水平垂直パリティ検査符号
  - ハミング符号
    （RAIDディスク)
  - RS符号
    (CS放送，DVD)

例4

- 7ビットの文字コード 30, 3F, 7Aに偶数パリティビットを付加せよ.

文字	16進	2進	パリティ
0	30	0011 0000	**0**011 0000
1	31	0011 0001	**1**011 0001
?	3F	0011 1111	
z	7A	0111 1010	

単一パリティ検査符号　頻出★

- 検査方程式
  - $c = x_1 + x_2 \bmod 2$
  - 線形符号
- 符号語
  - **w** $=(w_1,w_2,w_3)$
    $= (x_1, x_2, c)$
  - x_1, x_2: 情報ビット
  - c: 検査ビット
  - n = 3: 符号長
  - k = 2: 情報ビット長
- 組織符号
  - (n, k)符号
  - 符号化率 : 情報ビットの割合

    $$\eta = k/n$$
  - 例）
    - 単一パリティ検査符号は, (k+1,k)符号
    - 符号化率 $\eta=2/3$

剰余の演算

- mod 2の演算
  - 2で割った余り

0+0 = 0 mod 2

0+1 = 1 mod 2

1+0 = 1 mod 2

1+1 = **0 mod 2**

1+1+0+1 = 1　1の数が奇数

0　1の数が偶数

- 性質
  - 足し算と引き算が同じ
  - x + x = 0
  - x – y = x + y
  - x + 0 = x
  - x + 1 = ~x

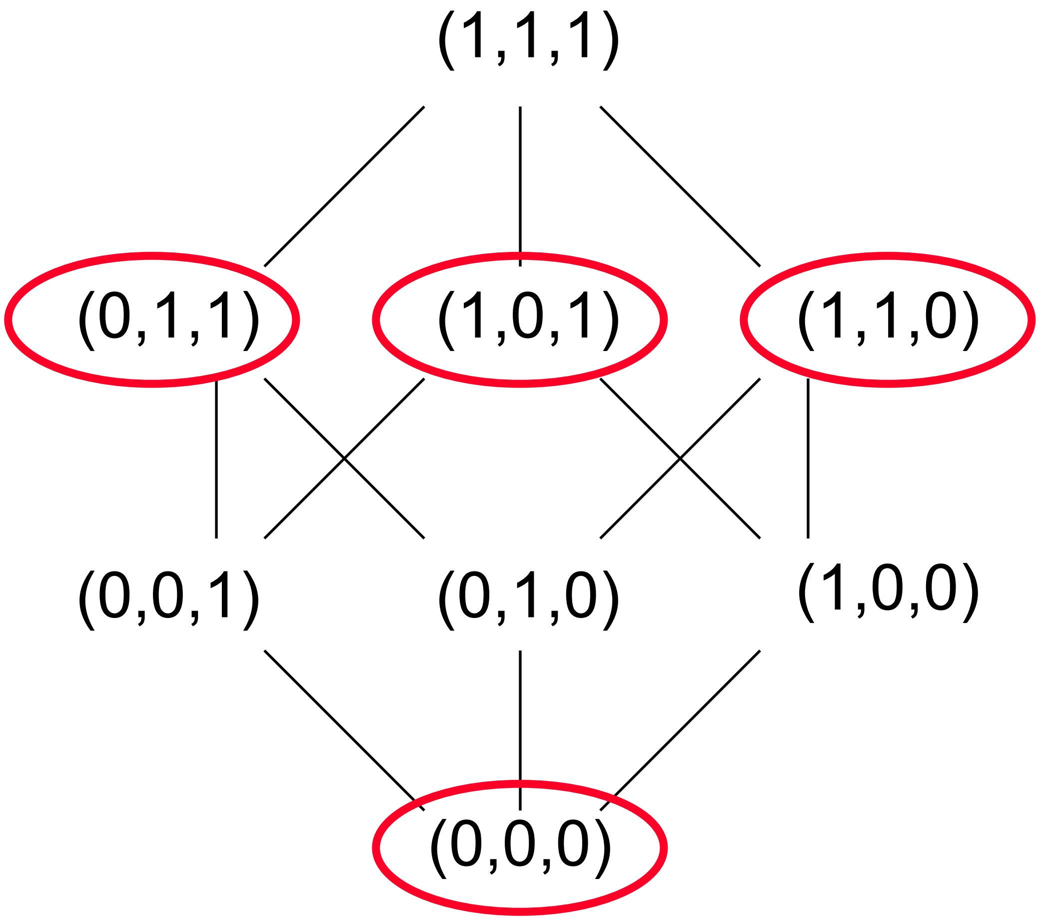
誤りの検出

- 正しい符号語は？　$w = (x_1, x_2, c)$
  - ❑(0,0,0)
  - ❑(0,1,0)
  - ❑(1,0,1)
  - ❑(1,1,1)

1の個数が偶数のものはOK

「*偶数パリティ*」

単一パリティ検査符号 C



単一誤り検査符号の性質

- 線形性
  - $w_i + w_j$ in C
  - (1,1,0)+(0,1,1) = (1,0,1) in C
- 限界
  - 誤り検出．訂正はできない．
  - 二重誤りは検出できない
    - » (1,1,0) → (**0**,**0**,0) は正しい符号語

例5

- 8ビットのレジスタ (d_0 d_1 … d_7)とパリティビットpがある．奇数パリティで常に成立する式はどれか？

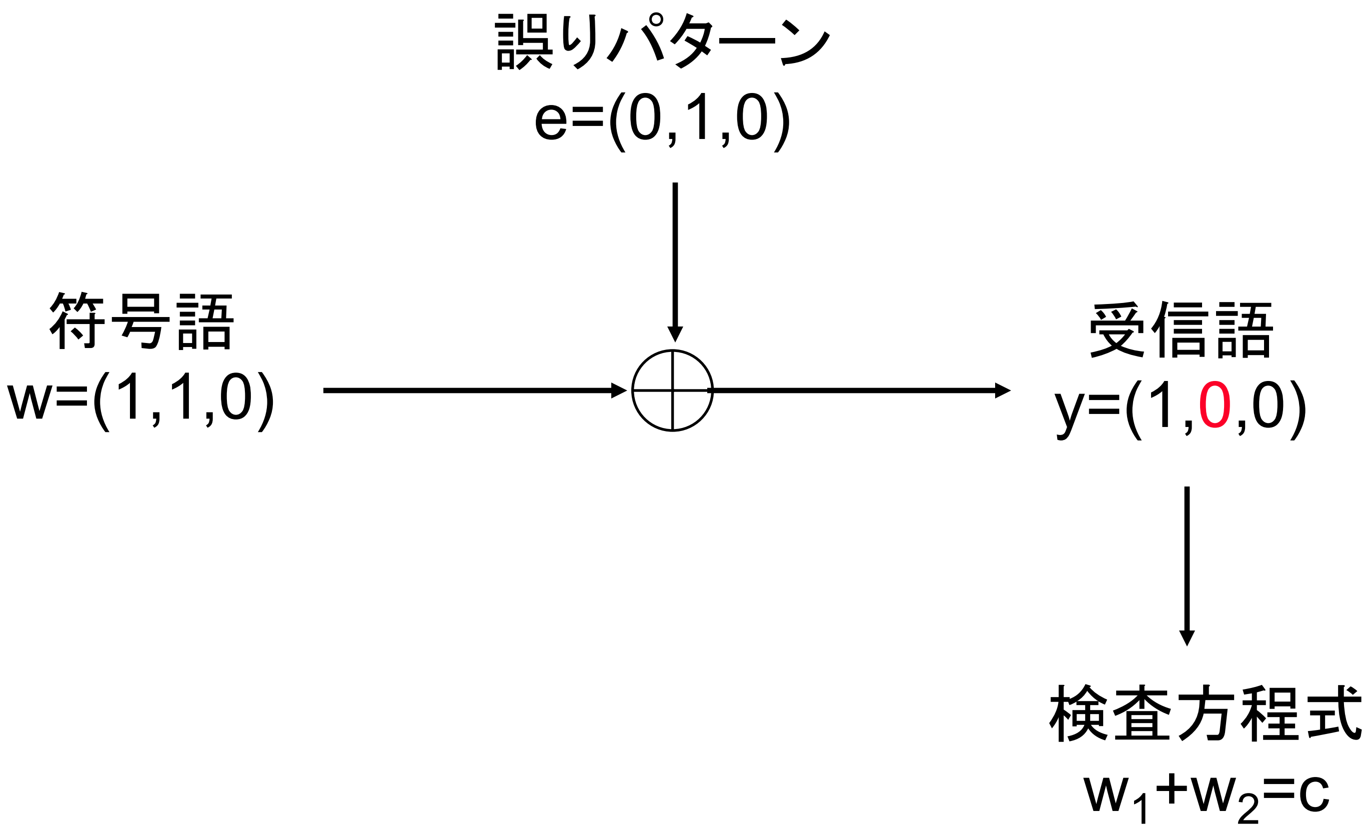
  ア $0 \oplus d0 \oplus d1 \oplus \ldots \oplus d7 = p$

  イ $d0 \oplus d1 \oplus \ldots \oplus d7 = p$

  ウ $d0 \oplus d1 \oplus \ldots \oplus d7 \oplus p = 0$

  エ $d0 \oplus d1 \oplus \ldots \oplus d7 \oplus p = 1$

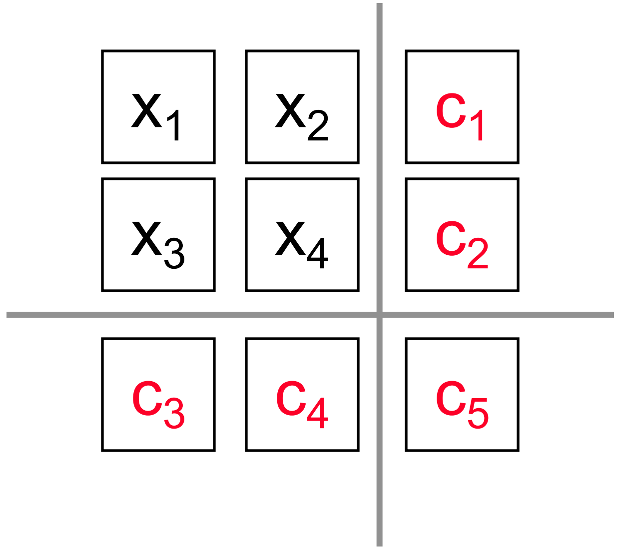
通信路モデル



水平垂直パリティ検査符号

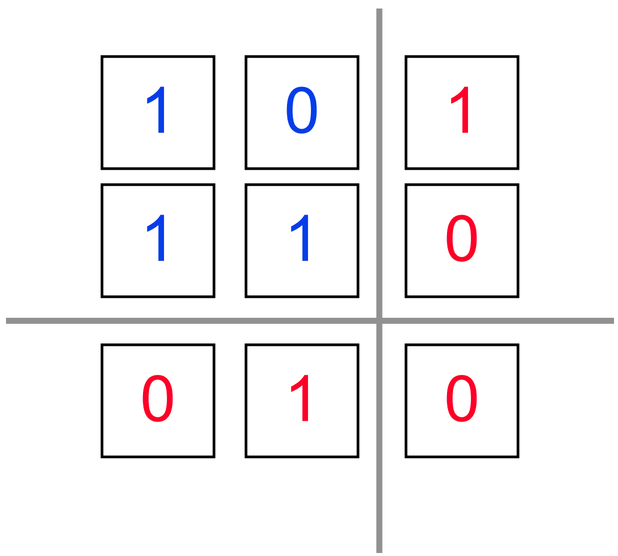
- 検査方程式
  - $c_1=x_1+x_2$
  - $c_2=x_3+x_4$
  - $c_3=x_1+x_3$
  - $c_4=x_2+x_4$
  - $c_5=c_1+c_2+c_3+c_4$
- (9,4)符号
- 符号語

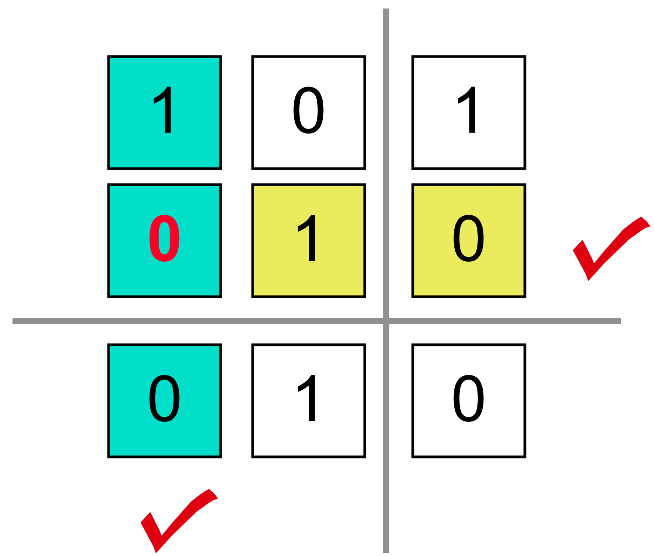
  $w=(x_1, x_2, x_3, x_4, c_1, c_2, c_3, c_4, c_5)$



誤り訂正の原理

- 正しい受信語
- 誤りのある受信語





$y=(1,0,1,1,\ 1,0,0,1,0)$

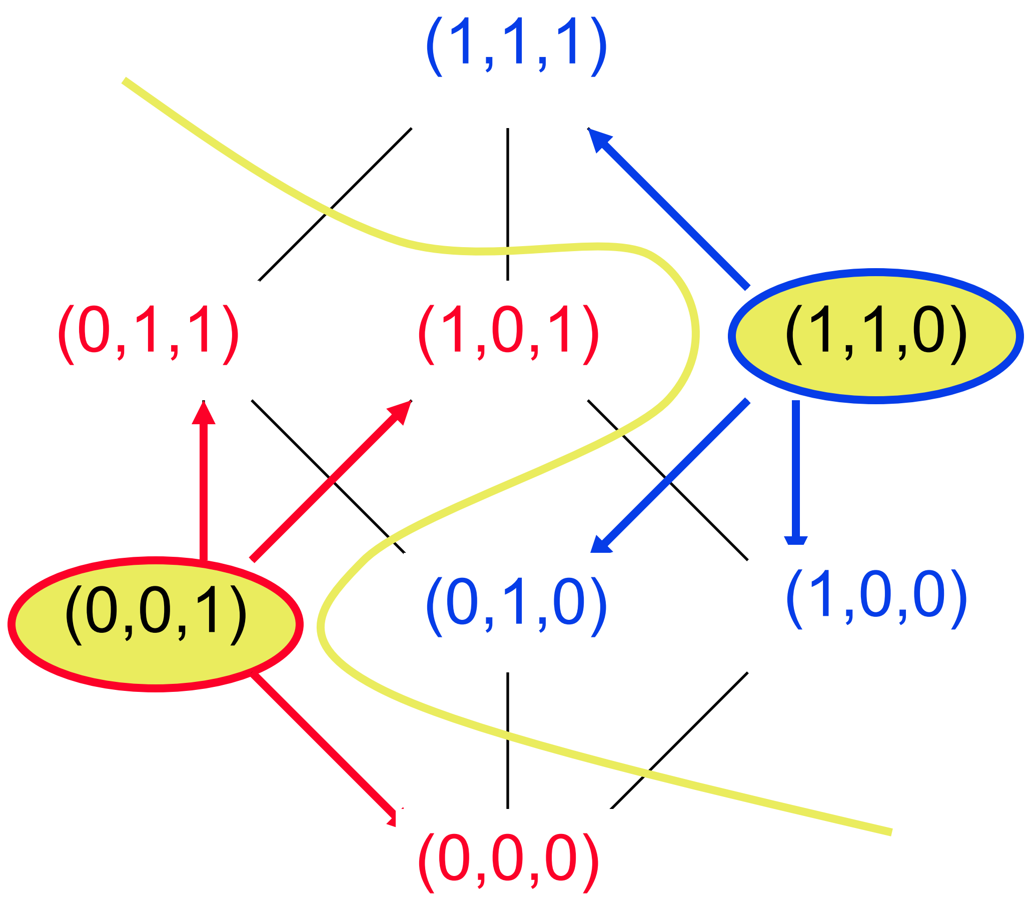
$y=(1,0,\mathbf{0},1,\ 1,0,0,1,0)$

$w=(1,0,\mathbf{1},1,\ 1,0,0,1,0)$

２値の誤り訂正

- 誤っているビットがわかる
  - 1→0
  - 0→1
  - 誤りが訂正できる

誤り検出能力



例5

- (9,4)水平垂直パリティ検査符号を考える
  $(w=(x_1,x_2,x_3,x_4,c_1,c_2,c_3,c_4,c_5))$

1. $\mathbf{x_1}$=(1,1,0,1)の時，符号語$\mathbf{w_1}$を求めよ.
2. $\mathbf{y_2}$=(1,1,1,0,1,1,0,0,0)の誤りパターン$\mathbf{e_2}$を求め，受信語を訂正せよ.
3. $\mathbf{y_3}$=(0,1,0,0,1,0,0,0,0)の誤りパターン$\mathbf{e_3}$を求め，受信語を訂正せよ.
4. $\mathbf{y_4}$=(0,1,0,0,0,1,1,1,0)の誤りパターン$\mathbf{e_4}$を求め，受信語を訂正せよ.

巡回符号　CRC *(難)

- Cyclic Redundancy Check
  - 復号化が容易
    - » ハードウェアによる実現
  - パケット通信
    - » 誤り検出→再送
    - » 訂正しなくてもよい
  - バースト誤り
    - » 0101010010101

バースト誤り

- 多重誤り
  - ランダム誤り
    - » e=(0,1,0,0,1,0,1,0,0)誤りは独立
  - バースト誤り
    - » e=(0,0,1,1,1,1,0,0)の長さ４
    - » 連続したビットの誤り．通信路の障害．
- 目的
  - 誤り検出，符号語の情報ビット大きく．

バースト誤り訂正能力

- m次の生成多項式G(x)
  - 受信語　Y(x)=E(x)+W(x)
  - 検出失敗
    - = Y(x)が別の符号語W'(x)になるとき
    - = Y(x)がG(x)で割り切れる
    - = E(x)がG(x)で割り切れる
    - = E(x)の次数が少なくともm以上
  - 長さm以下の任意のバースト誤りを検出可能
    - » $E(x) = x^{m-1} + .. + 1$ (次数m-1)
    - » 誤り訂正能力 b = m（Gの次数）

まとめ

- 情報理論には（　　　　　　）符号の様な情報を圧縮する技術と（　　　　　　）符号の様な誤りを検出，訂正する技術がある．
- ひとつ前の状態に依存してシンボルを生成する情報源を（　　　　　　　）情報源という．
- ハフマン符号では出現頻度の高い語を（　　　）ビットで符号化し，一語の平均ビットである（　　　　　　）を最小化する．
- 偶数パリティは単一の誤りを検出（　　　），二重の誤りを検出（　　　），単一の誤りを訂正（　　　）．CRC符号はバースト誤りを（　　　）する．