情報理論13

藤田 一寿

津山工業高等専門学校情報工学科 講師
電気通信大学先進理工学科 協力研究員

ver.20150722

雑音のある通信路を用いた誤りのない情報伝送

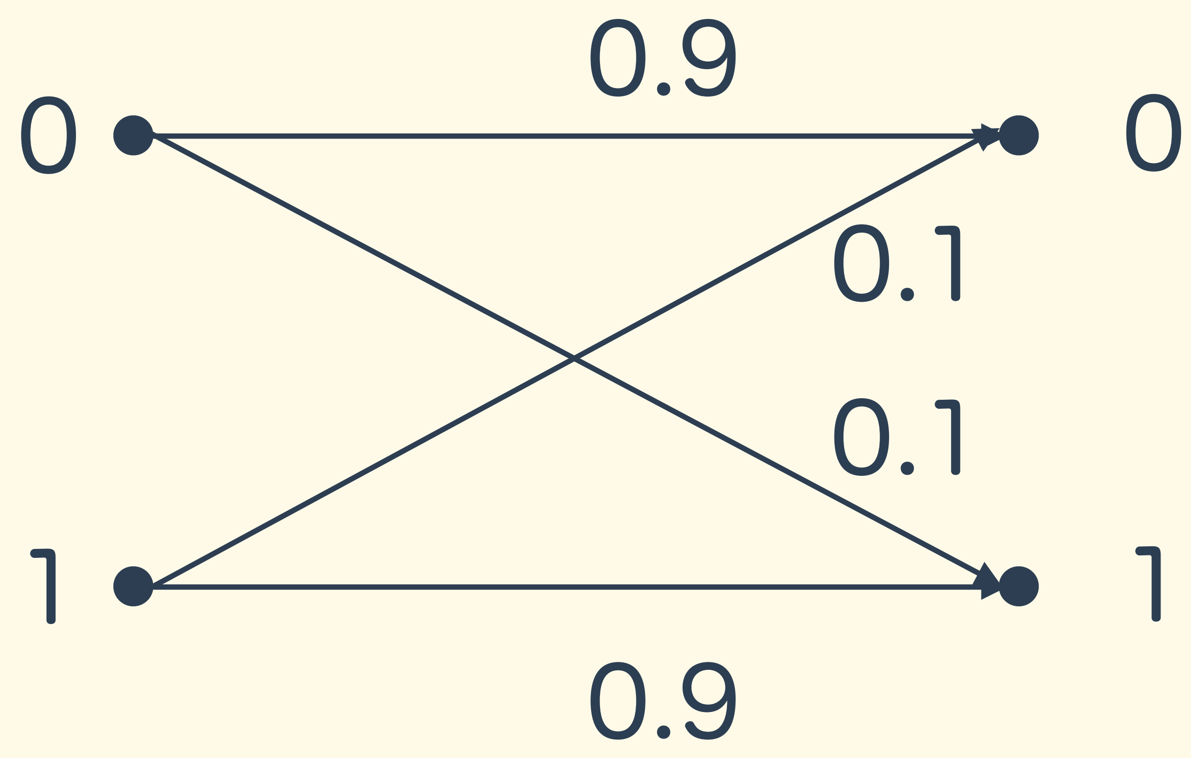
- 雑音がなければ毎秒1000ビット送れる通信路があるが，ノイズのため毎秒531ビットしか送れない場合がある．

- 逆に考えてこの通信路は毎秒531ビットのデータを正確に送ることが出来ると考えられないか．

例

- 1秒に０，１を１つづつ送り0.1の確率で信号が変わる2元対称通信路がある.

- この通信路を使ってより確実に信号を送るにはどうするか.



0を3つの信号で表す.

3つのうち2つ以上０ならば信号は０とすると,

000, 100は0を表す.

信号が誤って送られる確率は

$$(0.1)^3 + {}_3C_2(0.1)^2 \times 0.9 = 0.028$$

１つの信号では0.1の確率で間違った信号を受信していたが,
３つの信号を使うと間違える確率が約0.03まで減る.

ただし，信号1つ送るために3つ分の伝送時間がかかる.

- 使う信号の数を多くすれば確実に信号を送れるのではないか？
    - 確かに信号が誤る確率は大きく減る.
    - しかし，信号を送る時間は使う信号が多くなればなるほど遅くなる.
- 531ビットの通信路容量を持つ通信路を用い，間違いなく531ビット送る事はできないか.

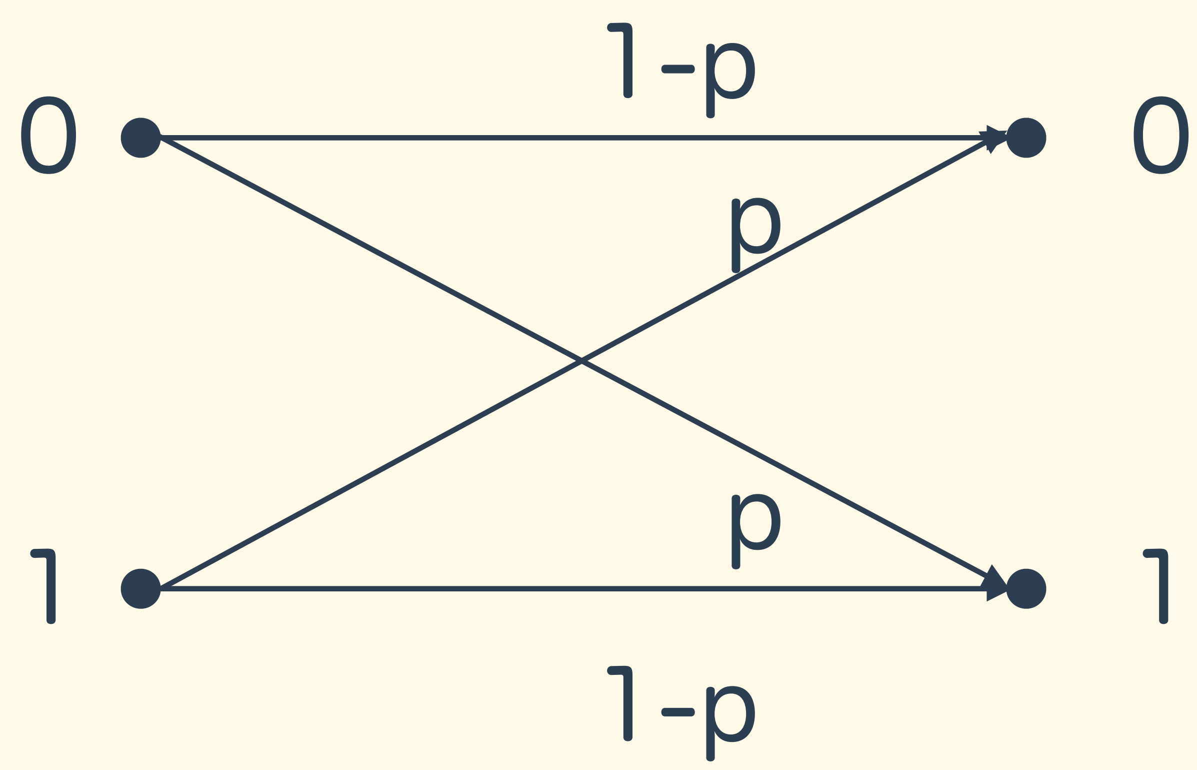
シャノンの雑音のある通信路の基本定理

容量Cの通信路と，1秒当たりRのエントロピーを持つ情報源とがあるとき

$R < C$

ならば，情報をこの通信路を通して，任意に小さい誤り確率で送ることが出来る符号化が存在する．

誤り訂正符号化



1秒ごとに1文字通す通信路がある．この通信路容量は

$$C = 1 + p\log p + (1-p)\log(1-p)$$

である．

n桁の0, 1から成る信号 $x_1 x_2 x_3 \ldots x_n$

雑音のためnC ビットしか送れない．

はじめからnCビットを確実に送ることを考える．

nを大きくすることで誤り率を小さくする．

n 桁の信号を用いてk桁の信号を正確に送ることを考える.

情報信号(初めのk桁の部分) $x_1x_2x_3 \ldots x_k$

検査信号(残り $m = n - k$ の部分) $x_{k+1}x_{k+2}\ldots x_n$

送信信号 = $\underbrace{x_1x_2\ldots x_k}_{\text{情報信号}}\ \underbrace{x_{k+1}x_{k+2}\ldots x_n}_{\text{検査信号}}$

もし誤りがすべて訂正できるとしたら

k/nビットの情報が伝送できる.

パリティチェック（偶奇性検査）

k=3, n=4の場合

情報信号は3桁，検査信号は1桁

1の数が偶数個になるように検査信号を決める.

もし，1文字だけ数字が変わるとすると，1の数が奇数となり，信号が起こったことを誤りを検出できる.

しかし，誤りを訂正できない.

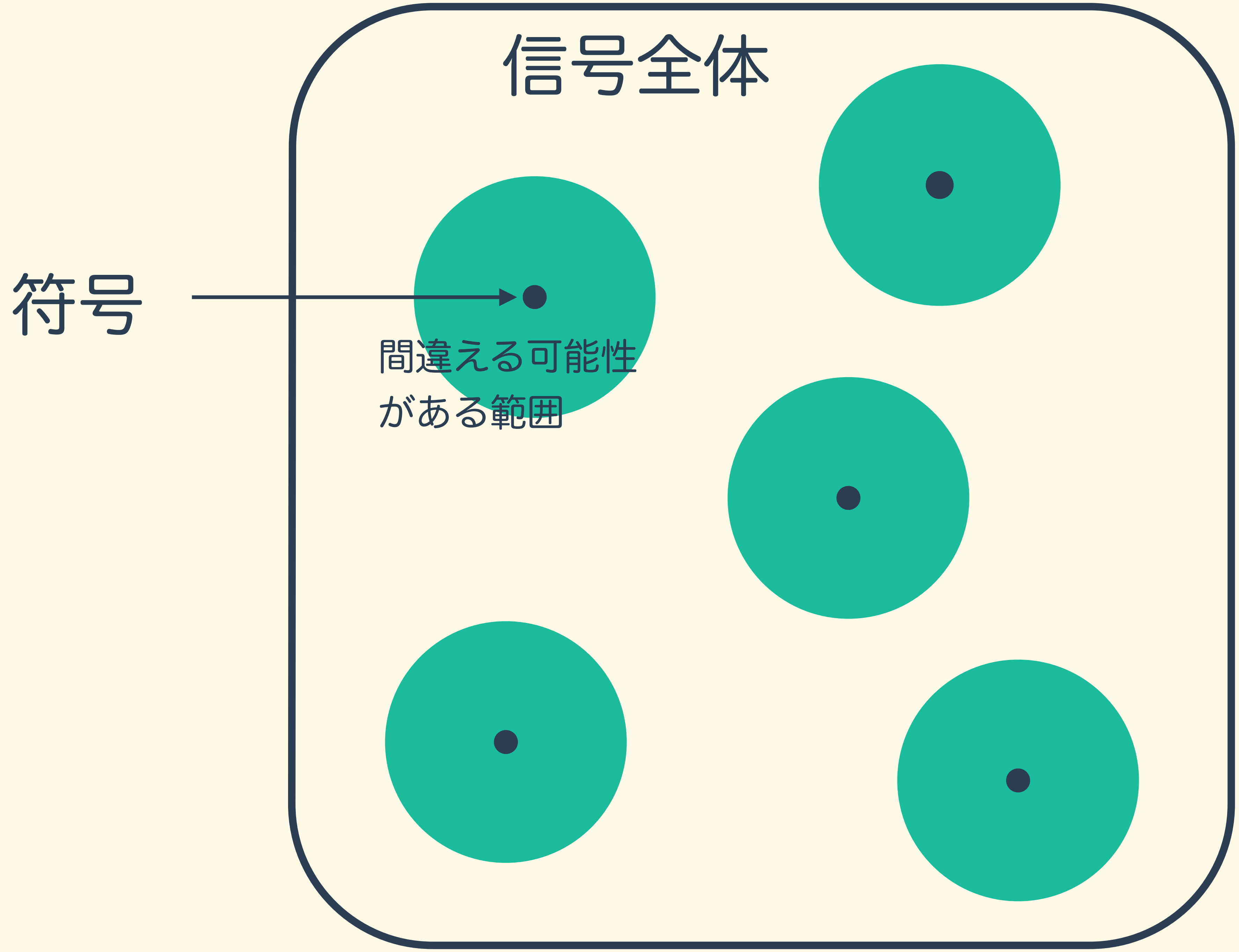
情報信号＋検査信号＝送信信号

000	0	0000
001	1	0011
010	1	0101
011	0	0110
100	1	1001
101	0	1010
110	0	1100
111	1	1111

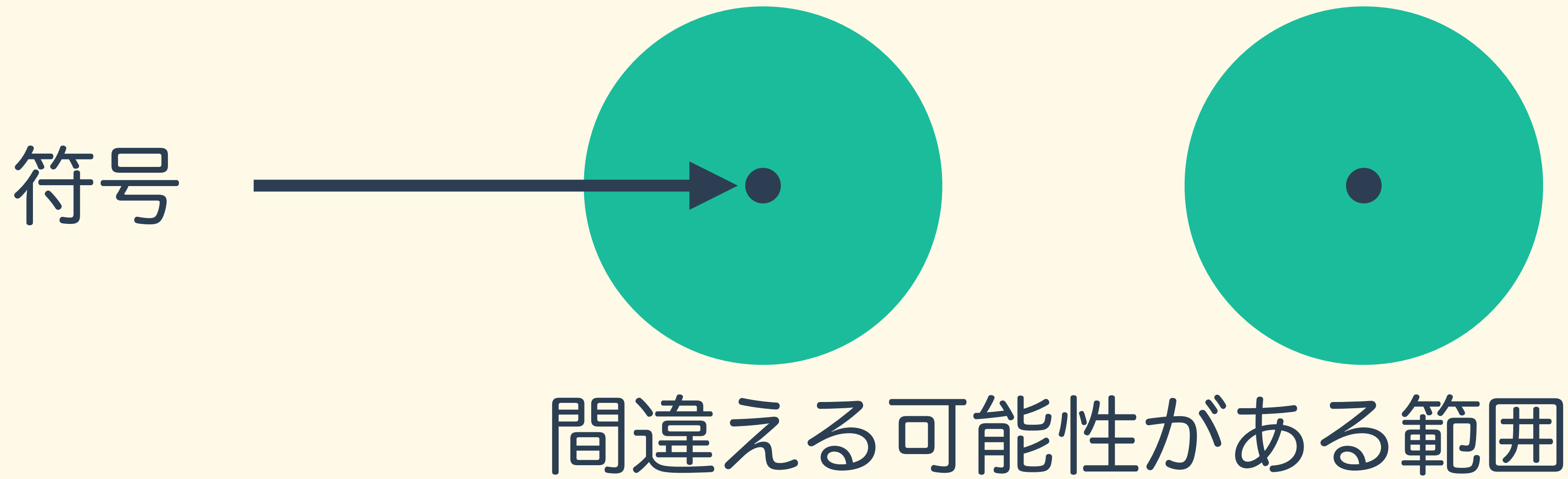
誤り訂正の仕組み

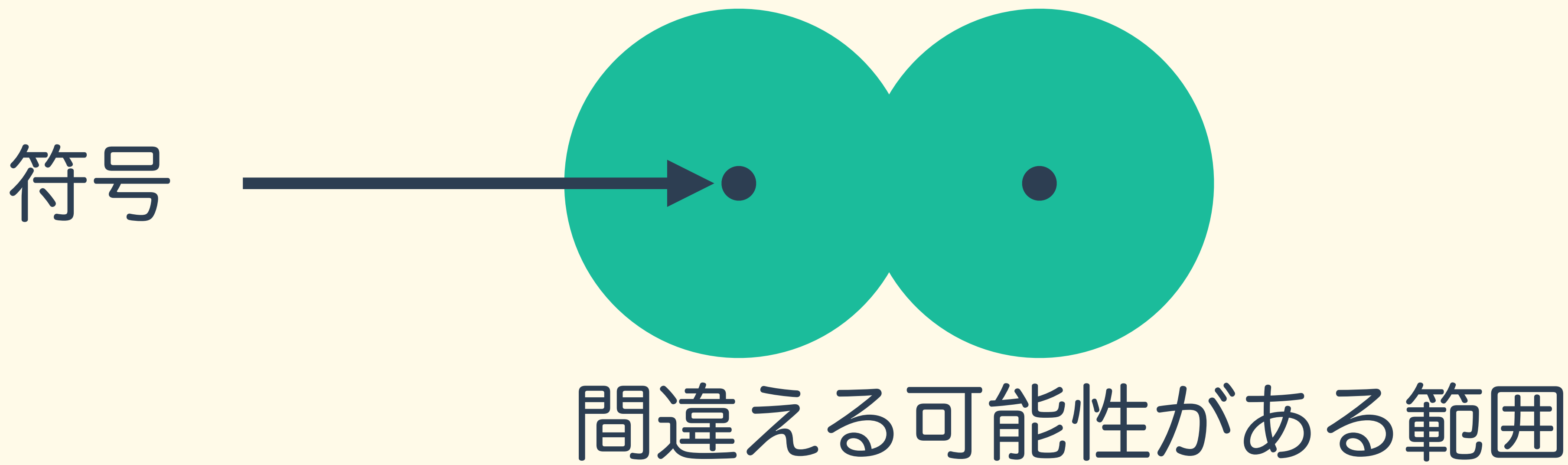
- 送信信号 x1x2…xn
- 信号の総数は2^n個
- 検査信号を除いた情報信号は2^k個
- 各桁で誤りが起こる確率をpとすると，n桁のうち誤りが発生した桁の数はnp個
- しかし，実際にはばらつきがあるのでr=n(p+ε)個を間違える.
- 大数の法則から，nが大きければ大きいほど誤って信号が送られる個数はnp個に近づく（(1-n)p個確実に送られる）
- r個間違えても訂正できる方法を考える.

信号間の距離



符号がr個の桁が異なる信号の範囲が重なり合わなければ，訂正可能である．





この間違える範囲がかぶるとどちらの
間違いか分からなくなる.

Hamming距離

- n桁の信号をベクトルで表す.

$$\boldsymbol{x} = (x_1, x_2, ..., x_n)$$

- 信号間の距離を次のように定義

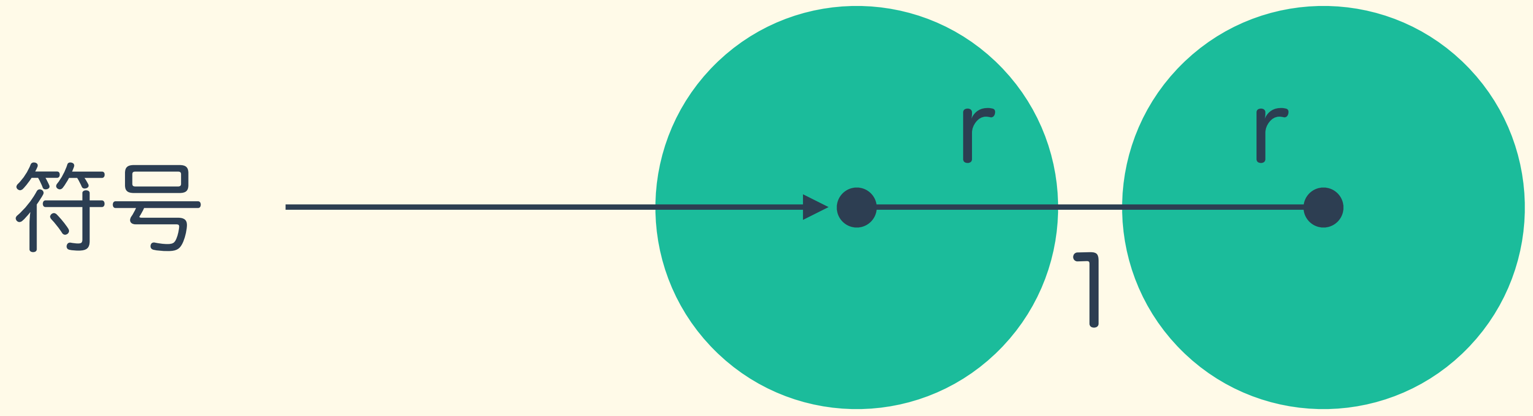
$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

- これは距離の公理を満たす.

符号wと符号wが雑音により変化してしまったあとのxの範囲は

$$d(\boldsymbol{w}, \boldsymbol{x}) \leq r$$

と表せる．雑音があっても符号同士区別可能にするためには，符号同士が2r+1離れている必要がある．



2rだと符号1つ分重なってしまい、それがどちらの符号なのか区別ができない

Hamming符号化

- 桁数$n=2^m-1$の符号
- 検査信号の長さm
- 情報信号の長さ$k=2^m-m-1$
- 1つの誤りを訂正する能力がある.

$n=7$, $k=4$, $m=3$を例に考える

符号

- 2進数で1から7まで縦に書く.

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

- 信号ベクトルxは $\boldsymbol{x} = (x_1, x_2, ..., x_7)^{\mathrm{T}}$ で表せる.
- 7桁の信号のうち $H\boldsymbol{x} = 0$ となる信号のみを符号として採用するのがHamming符号である. 計算方法は次のものを採用する.

- $0 + 0 = 1 + 1 = 0$,
- $0 + 1 = 1 + 0 = 1$,
- $0 \times 0 = 1 \times 0 = 0 \times 1 = 0$,
- $1 \times 1 = 1$
- $1 = -1$

$\boldsymbol{x} = (0,0,0,1,1,1,1)^{\mathrm{T}}$ が符号化どうか確かめてみよう

$$H\boldsymbol{x} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

符号化法

- 4桁の信号を送るときに検査信号として何を加えるか
- 情報信号 $x_1x_2x_3x_4$
- 検査信号 c_1, c_2, c_3
- 符号 $\boldsymbol{w} = (c_1, c_2, x_1, c_3, x_2, x_3, x_4)^{\mathrm{T}}$

$$
\begin{aligned}
H\boldsymbol{w} &= \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \cdot (c_1, c_2, x_1, c_3, x_2, x_3, x_4)^{\mathrm{T}} \\
&= \begin{pmatrix} c_3 + x_2 + x_3 + x_4 \\ c_2 + x_1 + x_3 + x_4 \\ c_1 + x_1 + x_2 + x_4 \end{pmatrix} = 0
\end{aligned}
$$

$$
\begin{aligned}
c_1 &= x_1 + x_2 + x_4 \\
c_2 &= x_1 + x_3 + x_4 \\
c_3 &= x_2 + x_3 + x_4
\end{aligned}
$$

誤り訂正のやり方

雑音により誤りが起こった場合，どこかに1足されたと考えられる．誤りが起こったかどうかを表すベクトルを

$$\boldsymbol{e} = (e_1, e_2, .., e_7)^{\mathrm{T}}$$

とする．ei = 1ならi桁に誤りが起こったことを意味し，ei=0ならi桁に誤りが怒らなかったことを意味する．誤りが起こったあとの信号は $\boldsymbol{z} = \boldsymbol{w} + \boldsymbol{e}$ と書け，Hzは

$$H\boldsymbol{z} = H(\boldsymbol{w} + \boldsymbol{e}) = H\boldsymbol{e} = \boldsymbol{s}$$

となる．誤りがなければ当然Hz=He=s=0となる．

Hz=0でも誤りがある可能性があるが，それは3桁以上間違える必要があり考えない．

誤りが1箇所とする(2箇所あると対応できない)

j桁に誤りがある場合ejのみ1となる．よって $H\boldsymbol{e} = \boldsymbol{h}_j$

となる.

これは，間違っている桁を表している．よって，hj桁を反転させれば下の符号に戻り，誤り訂正が可能となる.

例

4桁目が間違えたとすると，eは

$$\boldsymbol{e} = (0, 0, 0, 1, 0, 0, 0)^{\mathrm{T}}$$

と書ける．Heを計算するとHe = (1 0 0) となる．この結果から4桁目が間違えたことが分かる．

$$H\boldsymbol{e} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

期末試験

- 試験範囲
  - 情報源
  - 通信路
  - 符号化による冗長度の削減
  - 誤り訂正符号化
- 持ち込み可能物品
  - 電卓
- レポート
  - 試験の開始前に提出
  - レポート課題はslideshareもしくはgithubにおいてある.