混沌系工学特論 #5

情報科学研究科　井上純一

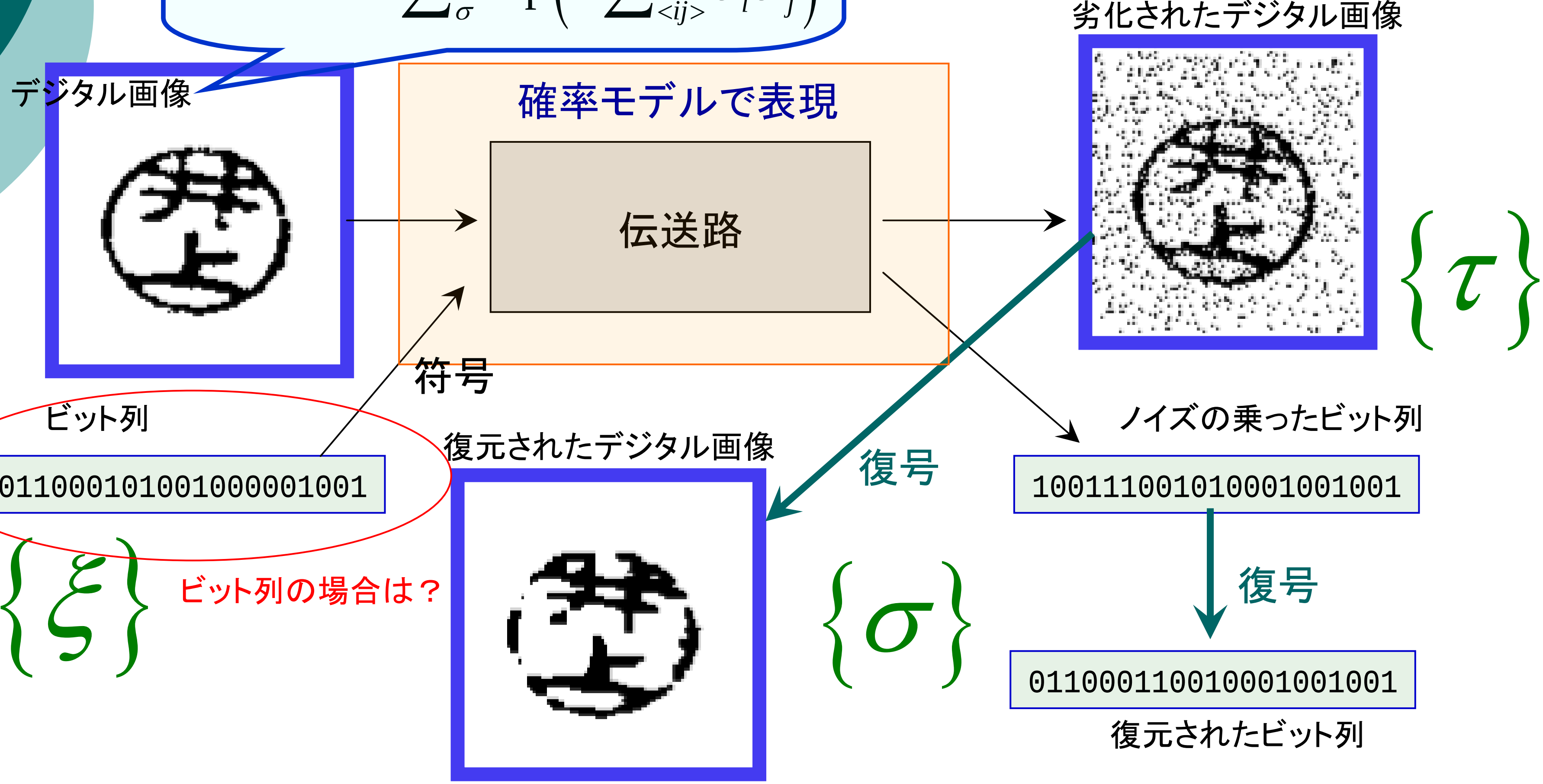
URL : http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/
Mirror : http://www005.upp.so-net.ne.jp/j_inoue/index.html

平成17年11月14日　第5回講義

デジタルデータの転送と復元再考

$$P\left(\{\sigma\}\right)=\frac{\exp\left(J\sum_{<ij>}\sigma_i\sigma_j\right)}{\sum_{\sigma}\exp\left(J\sum_{<ij>}\sigma_i\sigma_j\right)}$$

画像の復元では画像の特質を
事前分布で表現できる



雑音のある状況でのデータ送信

誤り訂正符号

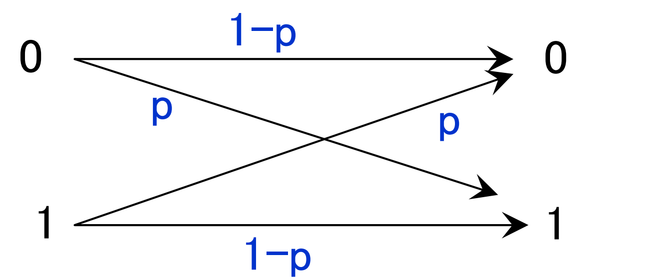
誤り率 p の2元対称通信路

ビット反転率

ただ1度だけでなく、何回か同じ記号を繰り返し送信する

受信者側は**多数決**で送信された記号を確定する

5回送信した場合：00101 ⇒ 0、 01111 ⇒ 1 等



2元対称通信路

送信回数 n = 5 のときの誤り確率は

$$f_e^{(5)}(p) = {}_5C_3\, p^3(1-p)^2 + {}_5C_4\, p^4(1-p) + p^5$$
$$= 10p^3(1-p)^2 + 5p^4(1-p) + p^5$$

3ビットの誤りが5ビットの何番目にくるかという場合の数

繰り返し回数nを
大きくすると
どうなるか？

伝送速度と誤り確率

$$R = \frac{\log M}{n} = \frac{1}{n}$$

送信記号の種類の数
1，０ ならば2

伝送速度（レート）

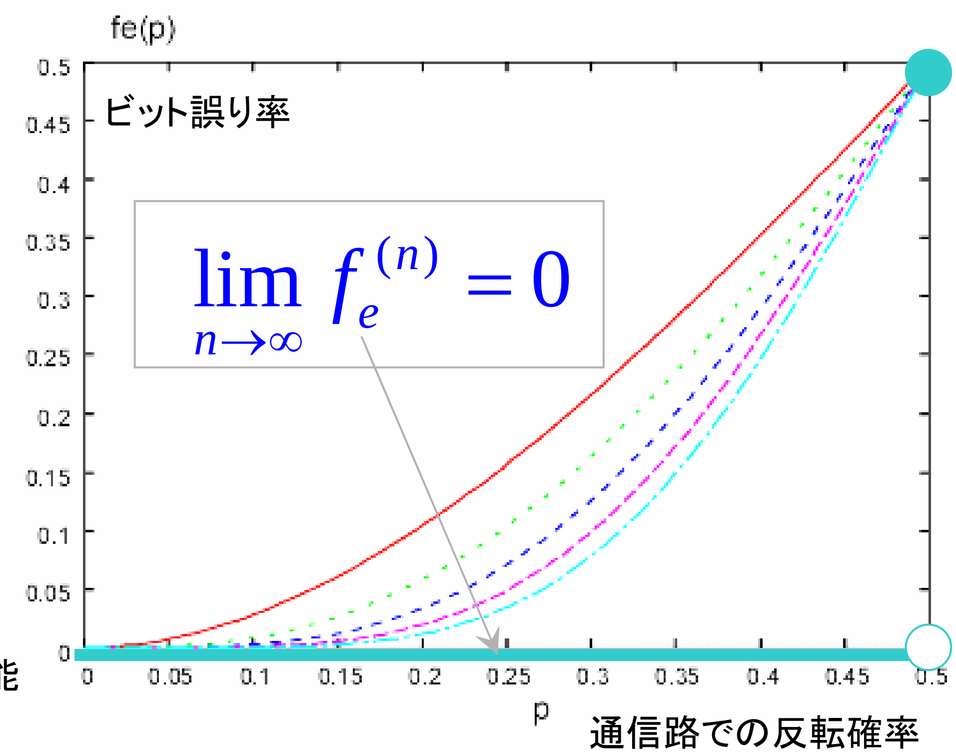
繰り返し回数を限りなく多くとると、
誤り確率は限りなくゼロに近づくが、
同時に伝送速度もゼロになってしまう

⇒　実用的ではない。しかし

$$R < C$$

通信路容量

ならば、限りなく小さな誤り確率での情報伝送が可能



通信路容量

$$C = \max_{P_X} I(X;Y)$$

$$I(X;Y) = H(Y) - H(Y \mid X)$$

通信路容量の定義

X： 入力

Y： 出力

相互情報量：入力Xを知ったときに、出力Yに対して得られる知識の増加分

入力の確率分布に関して、相互情報量を最大化したものが通信路容量

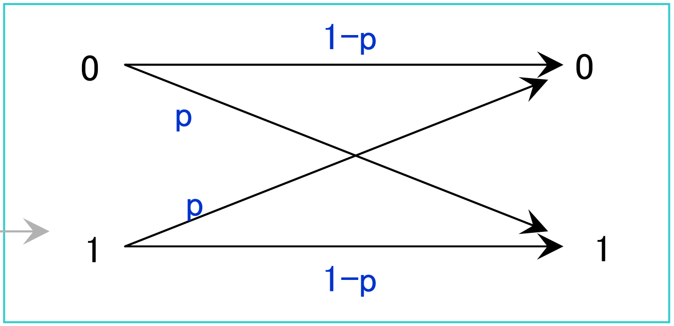
通信路容量は、通信路が伝送できる最大の情報量という意味を持つ

通信路容量の計算例

誤りの無い2元対称通信路

$$P_{Y|X}(0|0) = P_{Y|X}(1|1) = 1 - p$$
$$P_{Y|X}(0|1) = P_{Y|X}(1|0) = p$$

条件付き確率での表現



グラフ表現

$$H(Y|X) = -p\log p - (1-p)\log(1-p)$$ 入力分布によらない

$$C = I(X;Y) = \max_{P_X}\{H(Y) - H(Y|X)\} = 1 - h(p)$$

$$= \boxed{1 + p\log p + (1-p)\log(1-p)}$$ BSCの通信路容量

同様にガウス通信路に対して

通信路容量は $$C = \max_p I(X;Y) = \frac{1}{2}\log\left(1 + \frac{a_0^2}{a^2}\right)$$

通信路符号化定理

通信路符号化定理

i) $R < C$なる任意の速度$R = {}^{\log M}\!/_{n}$に対し

任意に小さな復号誤り率p_Eの符号が存在する。

ii) $R > C$なるRに対し、任意に小さなp_Eの符号は存在しない。

ゼロでなくてもよい。しかし、符号長 n を無限大ととることは必要

情報源の記号数 $M = 2^{nR}$

この定理を「ランダム符号」と呼ばれる符号に対して証明していく

ランダム符号

情報源の記号: $S_1, S_2, \ldots, S_M$　（等確率で生成される）

符号化　:

$S_1, S_2, \ldots, S_M$ の各々にランダムに0,1をn個並べた系列を割りあてる

$$S_1 \leftarrow 0010 \cdots 000$$

$$S_2 \leftarrow 0100 \cdots 001$$

$$S_3 \leftarrow 0010 \cdots 100$$

$$\cdots\cdots\cdots$$

$$S_M \leftarrow \boxed{1000 \cdots 010}$$

符号長 n

各ビットに ½ の確率で 0,1 を割り振る

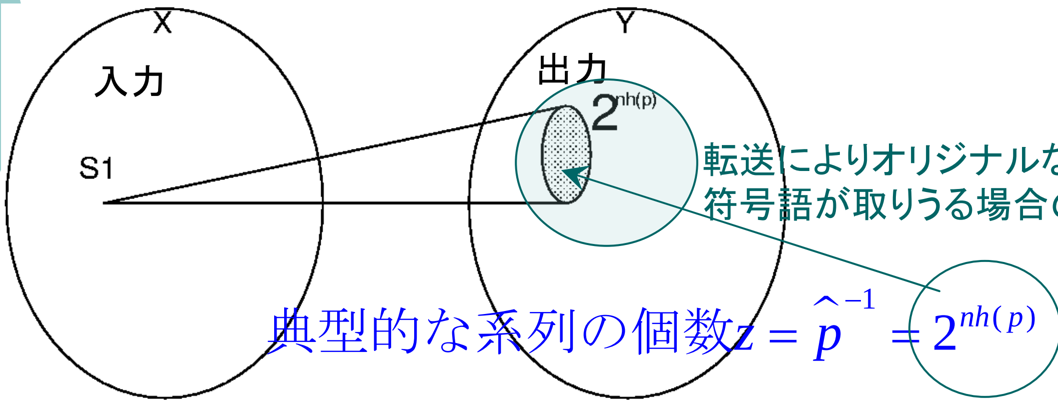
$M = 2^{nR} < 2^n$

$R \leq 1$　として議論を進める

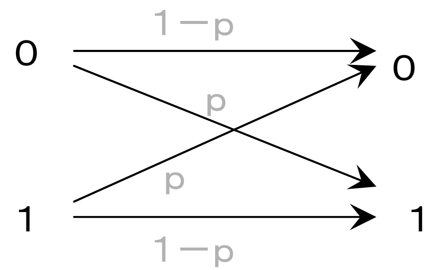
これらを2元対称通信路を介して転送する状況下で定理を証明していく

通信路符号化定理の証明#1

一つの符号語が転送により拡大する大きさの評価



典型的な系列の個数 $z = \hat{p}^{-1} = 2^{nh(p)}$



符号語は2元対称通信路を介して転送される

ある符号語 $S_1 = 000\cdots000$ は通信路の雑音により $010\cdots101$ として受信される

全 n ビットのなかで誤りの生じるビット数はおおよそ np と見積もられる

$$\underbrace{(0100\cdots011)}_{\text{np ビットの誤り}},\ \underbrace{(0111\cdots100)}_{\text{np ビットの誤り}},\ \cdots,\ \underbrace{(0101\cdots111)}_{\text{np ビットの誤り}}$$

通信路出力の典型的な系列

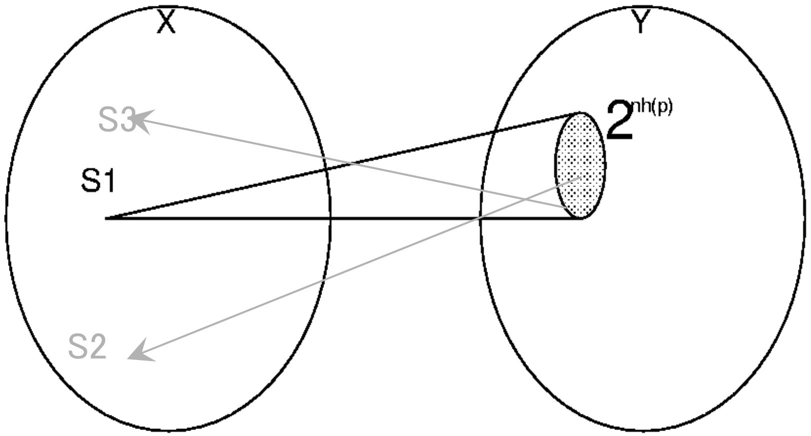
典型的な系列のなかの一つが現れる確率

$$\hat{p} = p^{np}\cdot(1-p)^{n(1-p)} = 2^{np\log n + n(1-p)\log(1-p)} = 2^{-nh(p)}$$

2値エントロピー関数

通信路符号化定理の証明#2

復号誤り率の評価



復号誤りが生じるのは・・・

受信される可能性のある$2^{nh(p)}$個の各々が、実際に送信された符号語S_1以外の$S_2, S_3, \ldots, S_M$の$(M-1)$個のいずれかに間違って復号されるとき

$$p_E = \frac{(M-1)}{2^n} \times 2^{nh(p)} \simeq \frac{M}{2^n} \cdot 2^{nh(p)} = 2^{n(R-1+h(p))} = 2^{n(R-C)}$$

(M−1)個のいずれかが選ばれる確率

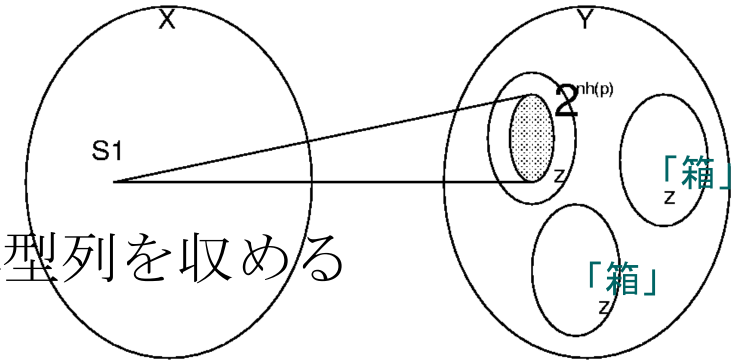
受信される可能性のある系列の個数

$\to 0 \quad (C = 1 - h(p) > R)$

i) の証明終わ

通信路符号化定理の証明#3

ii) の証明



受信系列の全ての可能性2^n

1つの符号語に対する通信路出力の典型列を収める

「箱」の数 $= z = 2^n / M$

2元対称通信路により実際に得られる典型列の数$2^{nh(p)}$

は全てこの「箱」に収まらなければならない

$$\frac{2^n}{M} > 2^{nh(p)} \Rightarrow 2^{n(C-R)} > 1$$

これは$R > C$では満たされない

ii)の証明終わり

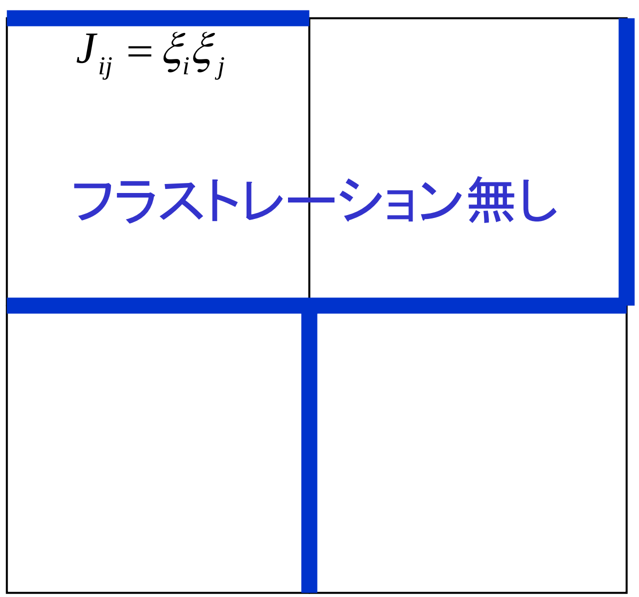
スピンモデルを用いた符号/復号化

ビット列を送るのではなく、結合 $J_{ij} = \xi_i \xi_j$ を送信

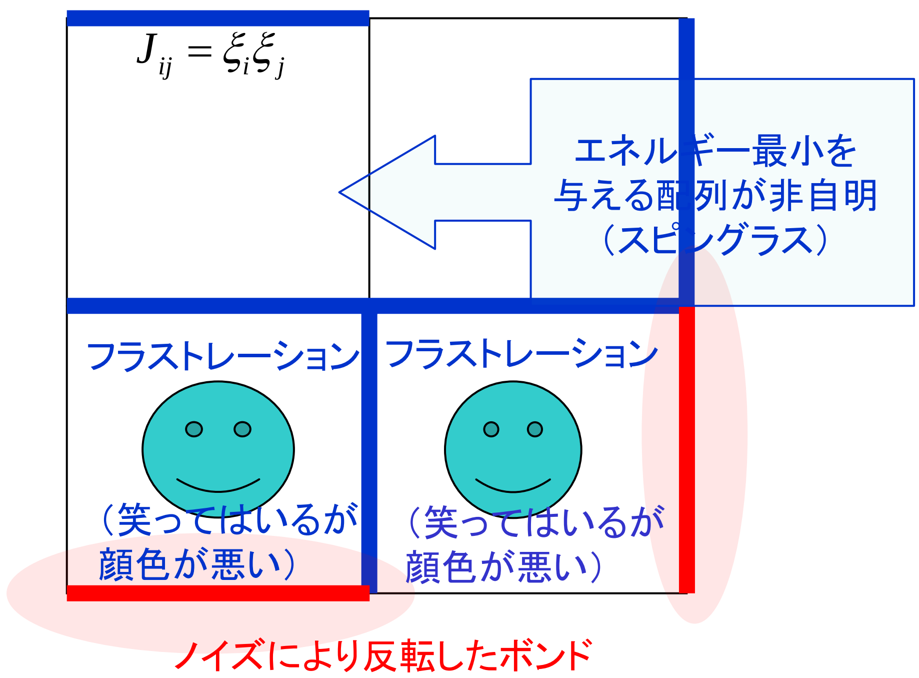
エネルギー関数 $$H = -\sum_{ij} \xi_i \xi_j \sigma_i \sigma_j$$

エネルギー関数の最小状態を選ぶことで復号する

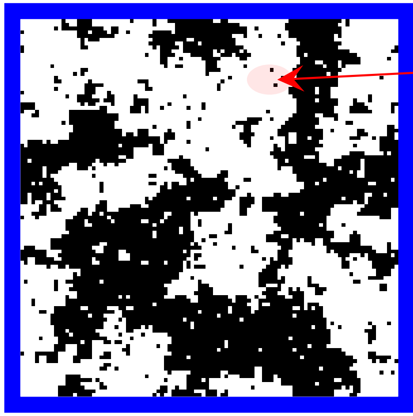
ノイズが無い場合



ノイズのある場合



パリティと画素の同時送信



$$J_{ij} = \xi_i \xi_j$$

画素だけではなく、隣接画素対
も同時に送信する(正方格子で2N個ある)

尤度

パリティ送信部分　画素送信部分

$$P\left(\{J\},\{\tau\} \mid \{\sigma\}\right) = \frac{\mathrm{e}^{\beta_J \sum_{ij} J_{ij}\sigma_i\sigma_j + h\sum_i \tau_i\sigma_i}}{[2\cosh(\beta_J)]^{N_B}[2\cosh(h)]^N}$$

事後分布とエネルギー関数

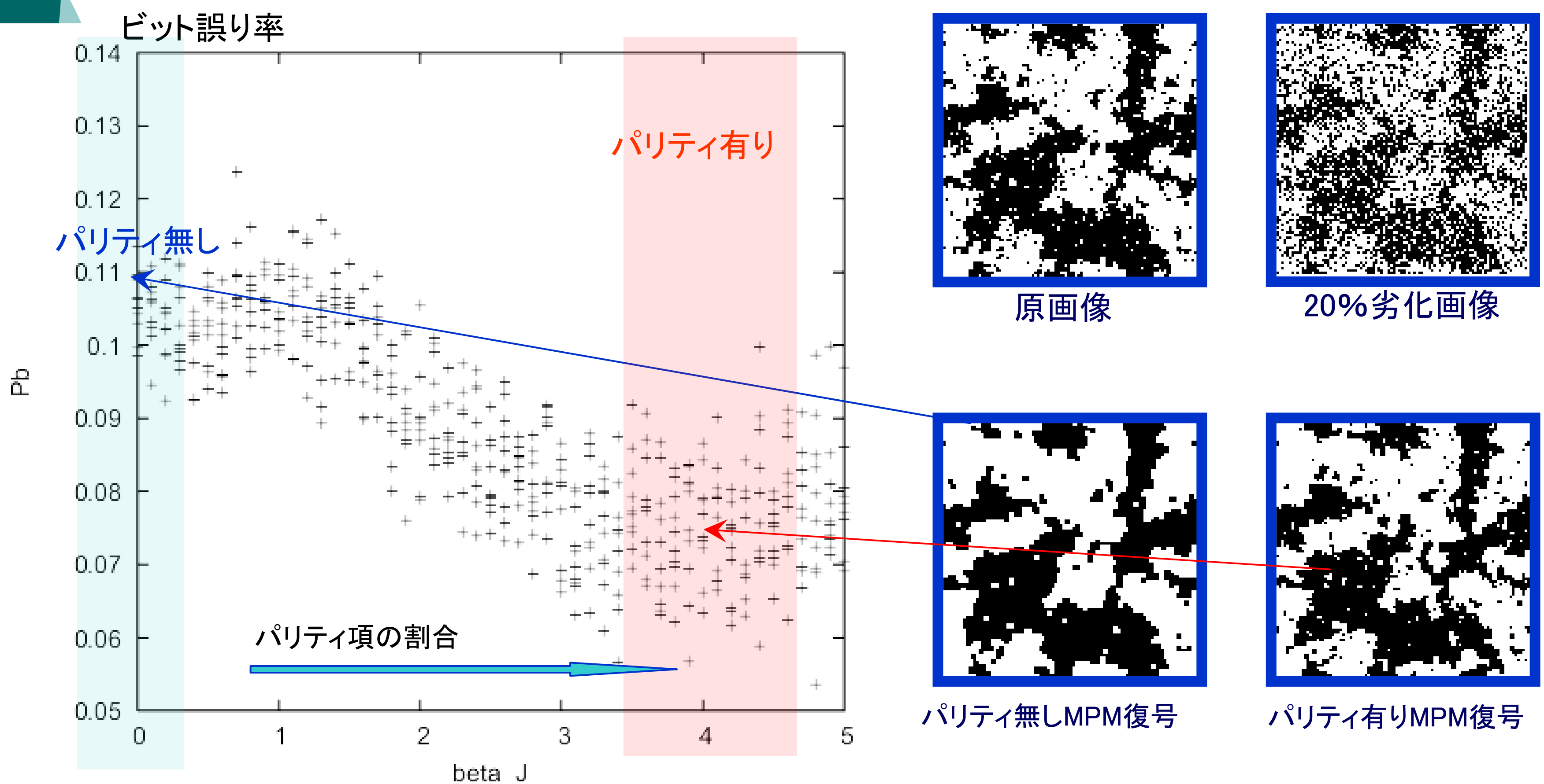
$$P\left(\{\sigma\} \mid \{\tau\}\right) = \frac{\mathrm{e}^{-H_{\mathrm{eff}}}}{\sum_{\{\sigma\}} \mathrm{e}^{-H_{\mathrm{eff}}}}$$

事前分布の部分

$$H_{eff} = -\beta_J \sum_{ij} J_{ij}\sigma_i\sigma_j - J\sum_{ij}\sigma_i\sigma_j - h\sum_i \tau_i\sigma_j$$

MAP復元、MPM復元を行う

復元結果



冗長性を付加することにより、より効果的な復元が実現される

Sourlas符号

N個のビットの中からp個をピックアップして送信する

$$J_{i_1 \cdots i_p} = \xi_{i_1} \cdots \xi_{i_p}$$

オリジナル・ビットのp個の積を送る

速度

$$R = \frac{N}{{}_N C_p} \simeq \frac{p!}{N^{p-1}}$$

ガウス通信路で送信

$$P(\{J\} \mid \{\xi\}) = \left(\frac{N^{p-1}}{J^2 \pi p!}\right) \exp\left[-\frac{N^{p-1}}{J^2 p!} \sum_{i_i < \cdots < i_p} \left(J_{i_1 \cdots i_p} - \frac{J_0 p!}{N^{p-1}} \xi_{i_1} \cdots \xi_{i_p} \right)^2 \right]$$

通信路

$$C = \frac{1}{2} \log\left(1 + \frac{a_0^2}{a^2}\right) \simeq \frac{J_0^2 p!}{J^2 N^{p-1} \log 2}$$

シャノンの限界式

$$\frac{R}{C} = \left(\frac{J_0}{J}\right)^2 \frac{1}{\log 2}$$

平均場理論による性能評価

事後確率

$$P(\{\sigma\}|\{J\}) = \frac{e^{\beta_J \sum_{i_1<\cdots<i_p} J_{i_1\cdots i_p}\sigma_{i_1}\cdots\sigma_{i_p}}}{\sum_{\{\sigma\}} e^{\beta_J \sum_{i_1<\cdots<i_p} J_{i_1\cdots i_p}\sigma_{i_1}\cdots\sigma_{i_p}}}$$

エネルギー関数

$$H_{\mathrm{eff}} = -\beta_J \sum_{i_1<\cdots<i_p} J_{i_1\cdots i_p}\sigma_{i_1}\cdots\sigma_{i_p}$$

レプリカ法から得られる状態方程式

$$m = \left[\langle\sigma_i\rangle\right] = \int_{-\infty}^{\infty} Dx \tanh(G)$$

磁化

$$q = \left[\langle\sigma_i\rangle^2\right] = \int_{-\infty}^{\infty} Dx \tanh^2(G)$$

スピングラス秩序変数

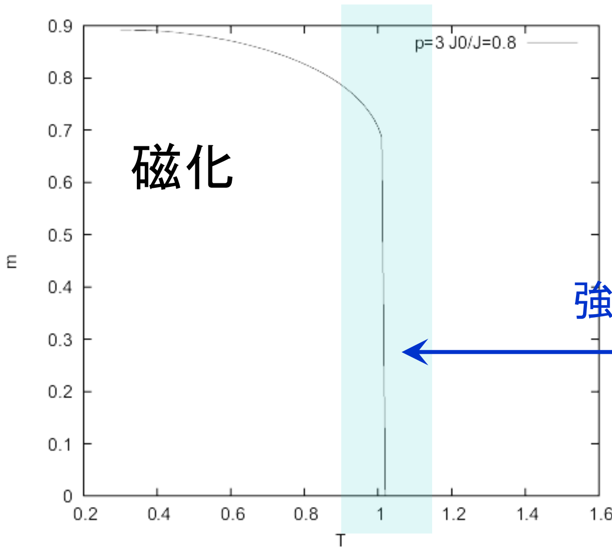
セルフ・コンシステントに解く

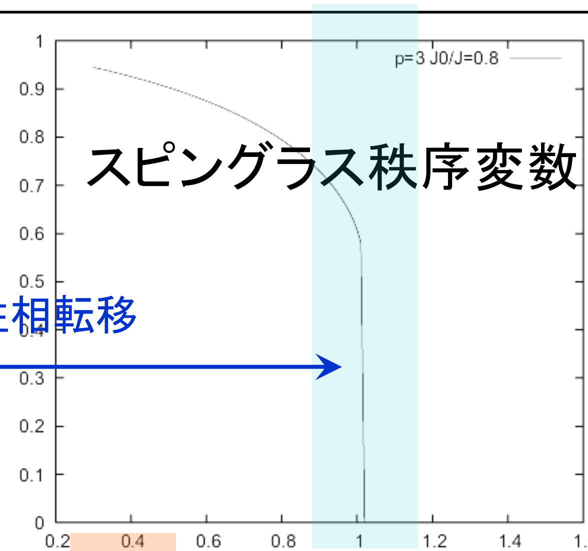
$$G = x\sqrt{\frac{p\beta_J^2 J^2 q^{p-1}}{2}} + \beta_J J_0 p m^{p-1}$$

$$P_b = \tfrac{1}{2}\left(1 - \left[\xi_1 \operatorname{sgn}\left(\langle\sigma\rangle_i\right)\right]\right) = \tfrac{1}{2}\left(1 - \int_{-\infty}^{\infty} Dx \operatorname{sgn}(G)\right)$$

ビット誤り率

解析結果 #1

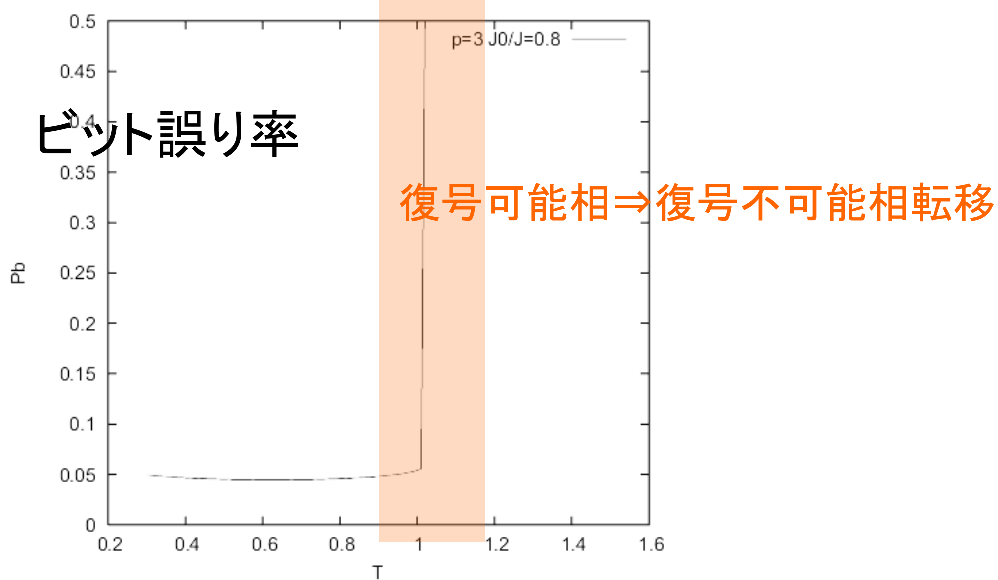




$$p = 3$$

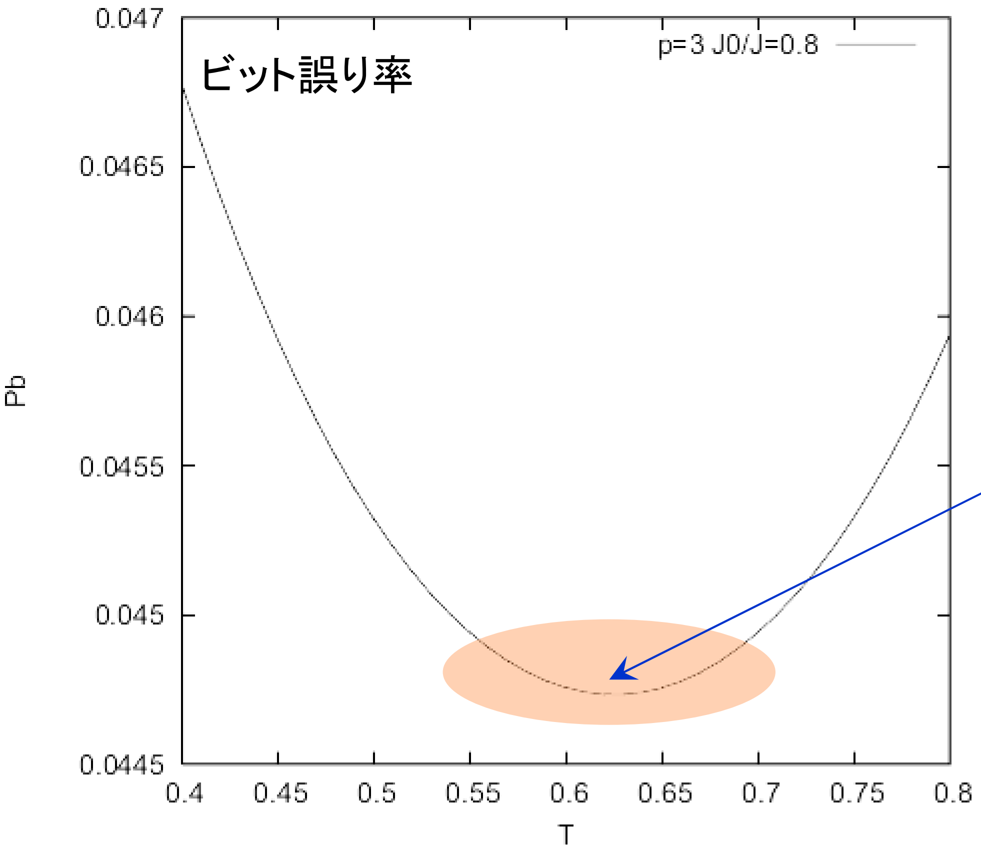
$$\boxed{J_0 / J} = 0.8$$

シグナルノイズ比



解析結果 #2

温度に関する最良性



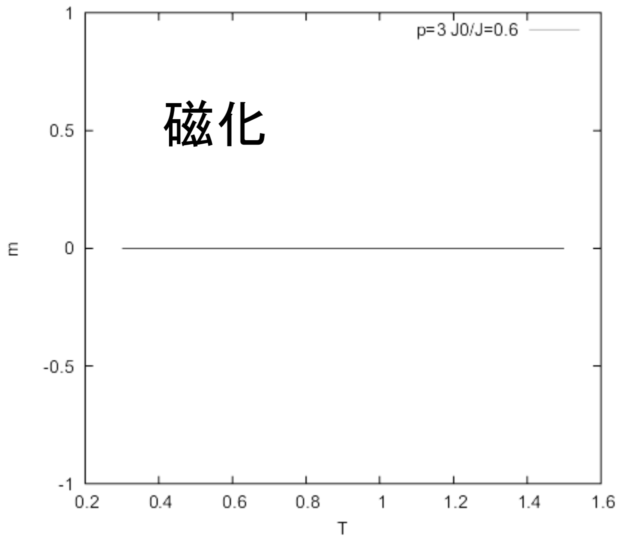
$$p = 3$$

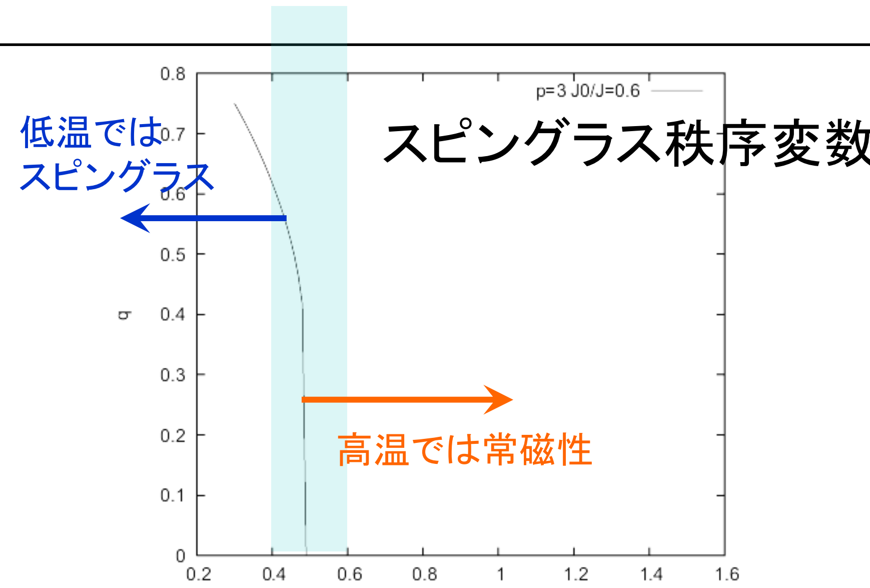
$$J_0 / J = 0.8$$

西森温度

$$T = \frac{J^2}{2J_0^2}$$

解析結果 #3

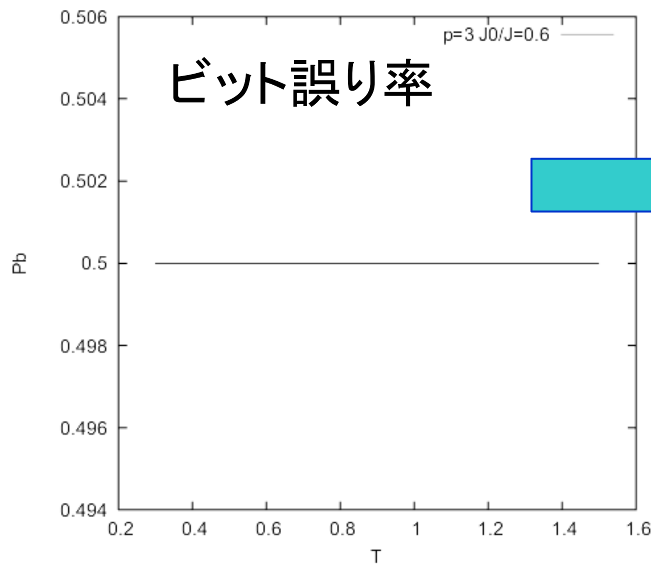




$$p = 3$$

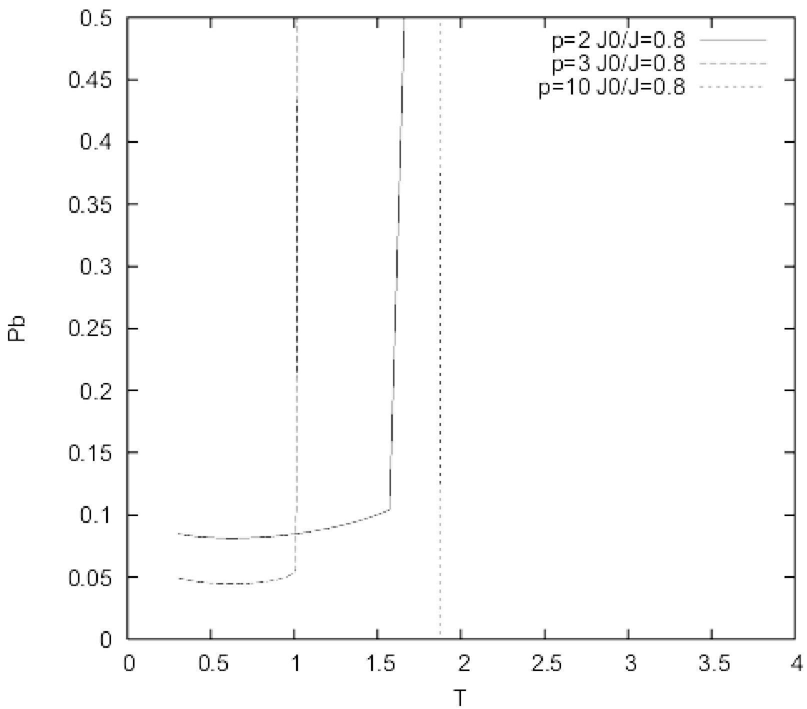
$$J_0/J = 0.6$$

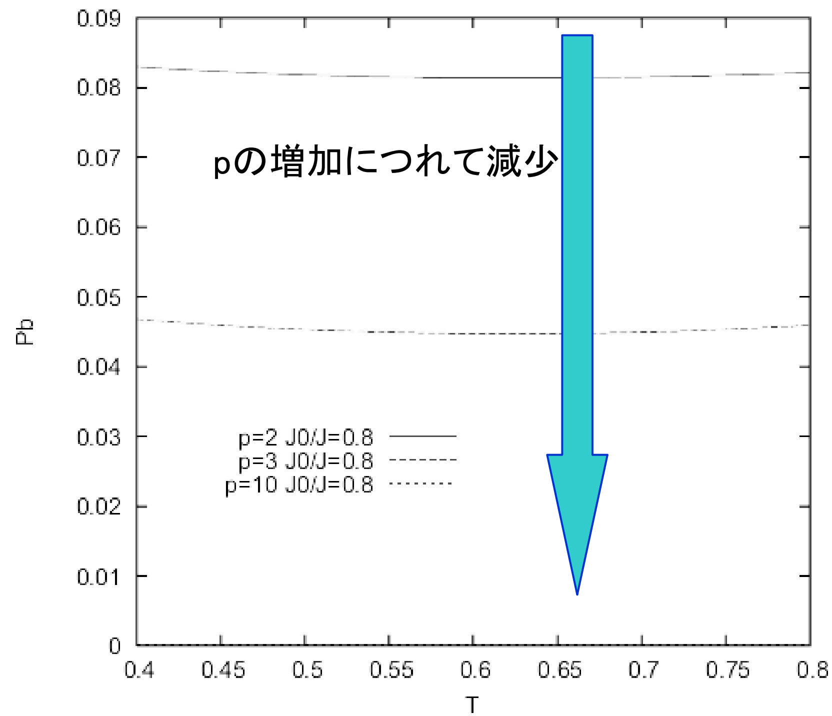
シグナルノイズ比



誤り訂正不可能

解析結果 #4





$$p \to \infty :$$

$$\left(J_0 / J \right) \geq \sqrt{\log 2}, \quad R = C$$

強磁性相の存在限界

Sourlas符号は漸近的にシャノン限界を達成する

次回はCDMAに移ります