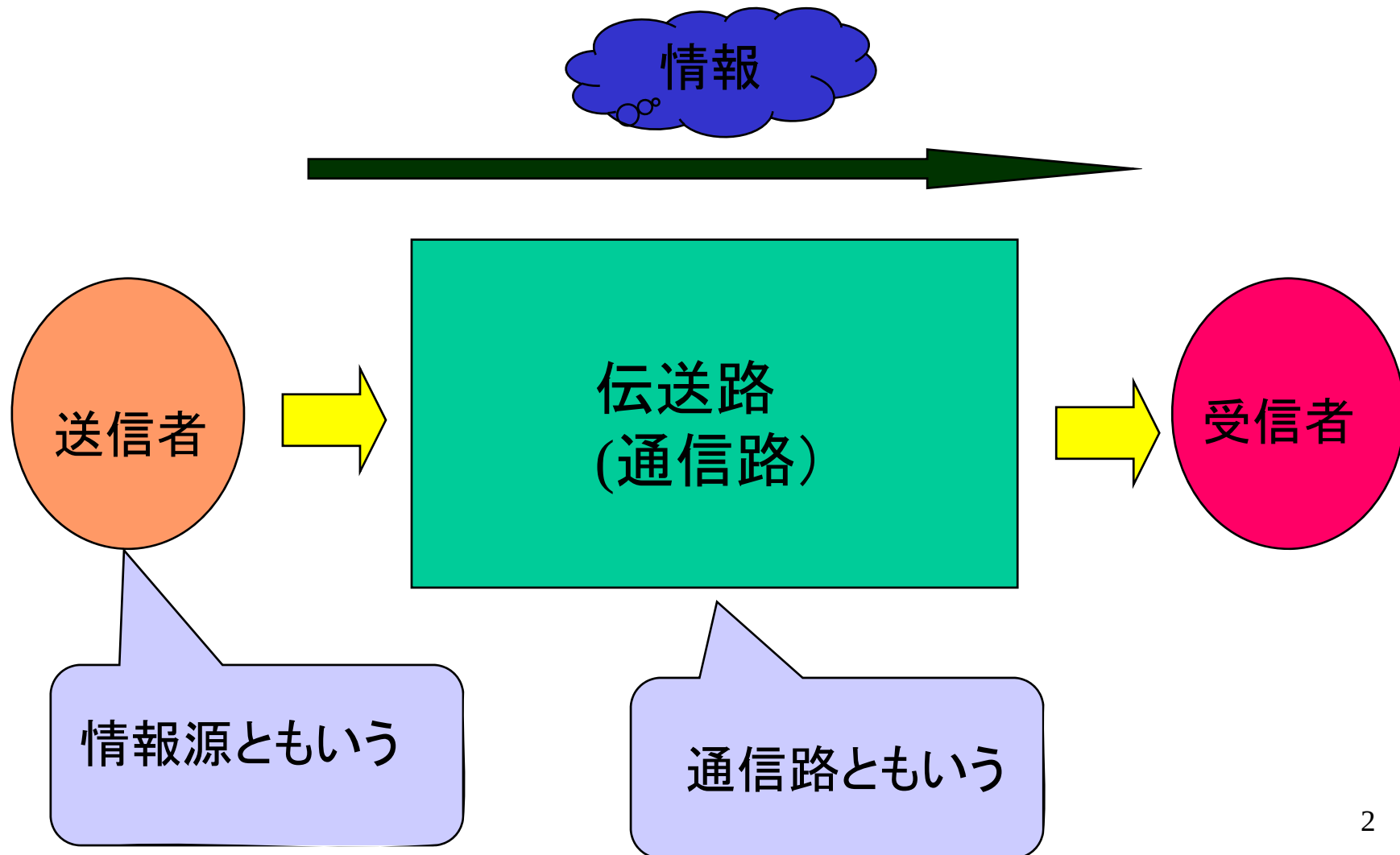
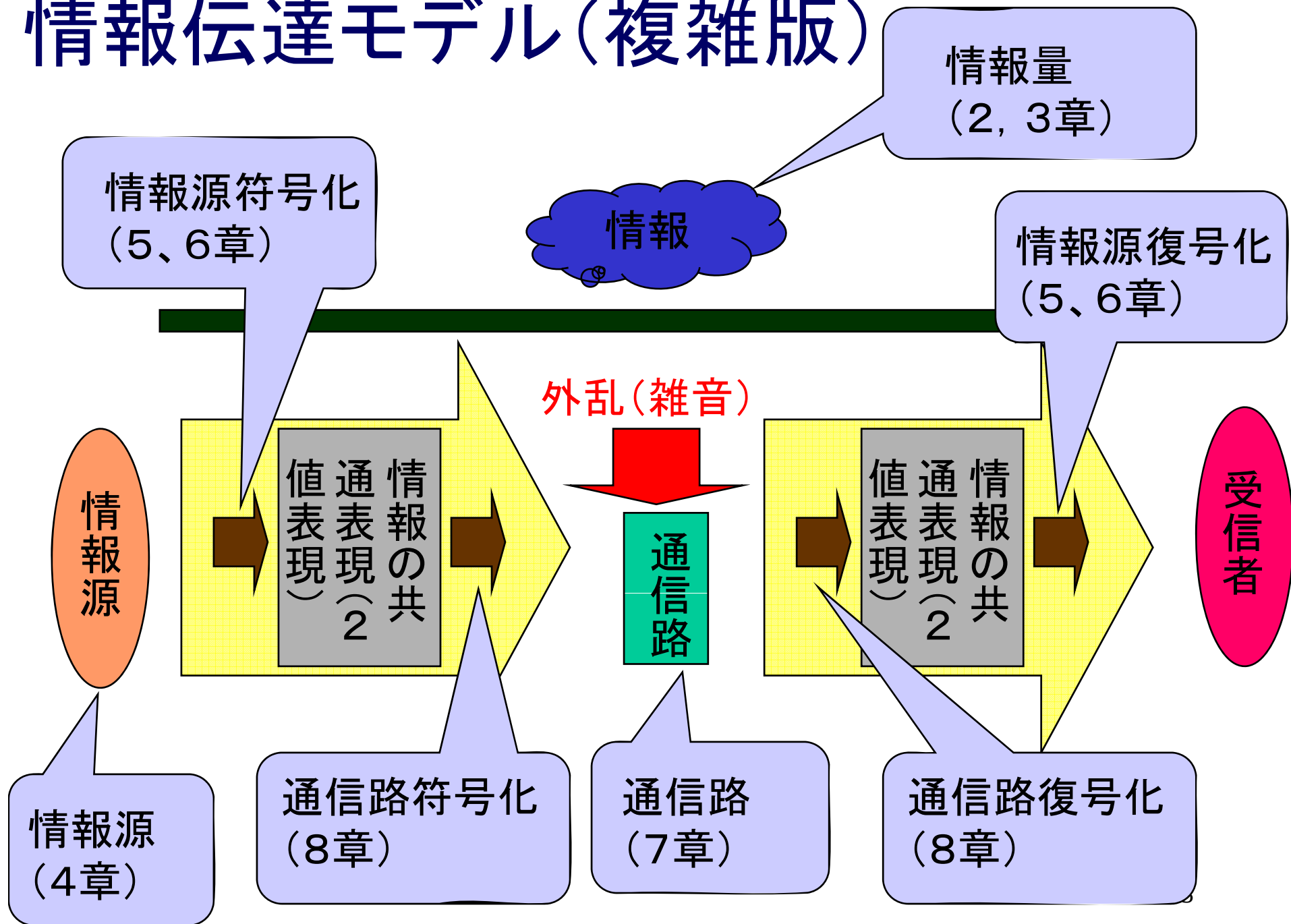


通信路符号化(8章)

情報伝達モデル(簡易版)



情報伝達モデル(複雑版)



情報源符号化と通信路符号化の比較

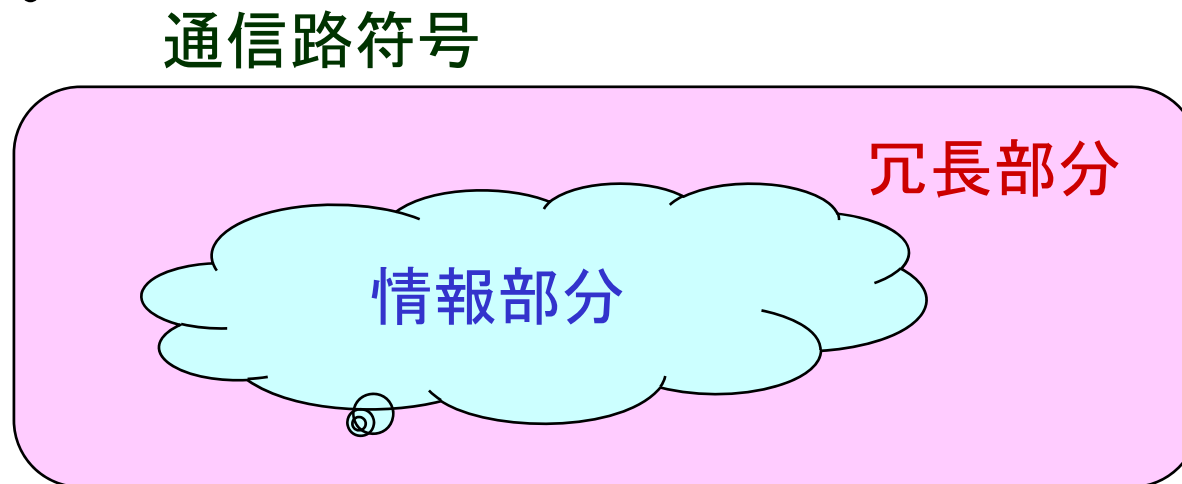
符号化	最終目標	実現状況	符号長	定理
情報源符号化	効率化	エネルギー・時間の節約	最短符号の実現	情報源符号化定理(シャノンの第1基本定理)
通信路符号化	信頼性の向上	誤りの検出・訂正	冗長性を付加	通信路符号化定理(シャノンの第2基本定理)

通信路符号化の基礎概念

通信路符号化は信頼性の向上を達成するのが目的であり、そのために情報部分の他に冗長部分を加える。すなわち、以下のように表せる。

通信路符号＝情報部分＋冗長部分

符号化の際には、いかにして冗長部分を作り出すかが重要となる。



通信路における記号の系列

情報源アルファベットを $A = \{a_1, a_2, \dots, a_r\}$ とする。

A の元からなる長さ n の系列すべての集合を A^n と書く。

このとき、 A^n には $|A|^n = r^n$ 個の要素が含まれる。

例

$$B = \{0, 1\}$$

$$B^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

$$|B| = 2$$

$$|B^3| = |B|^3 = 2^3 = 8$$

通信路符号化では、送信記号の系列すべてを符号として用いるのではなくて、その中の特定の系列だけを符号として扱う。このことが冗長性を生み出す。

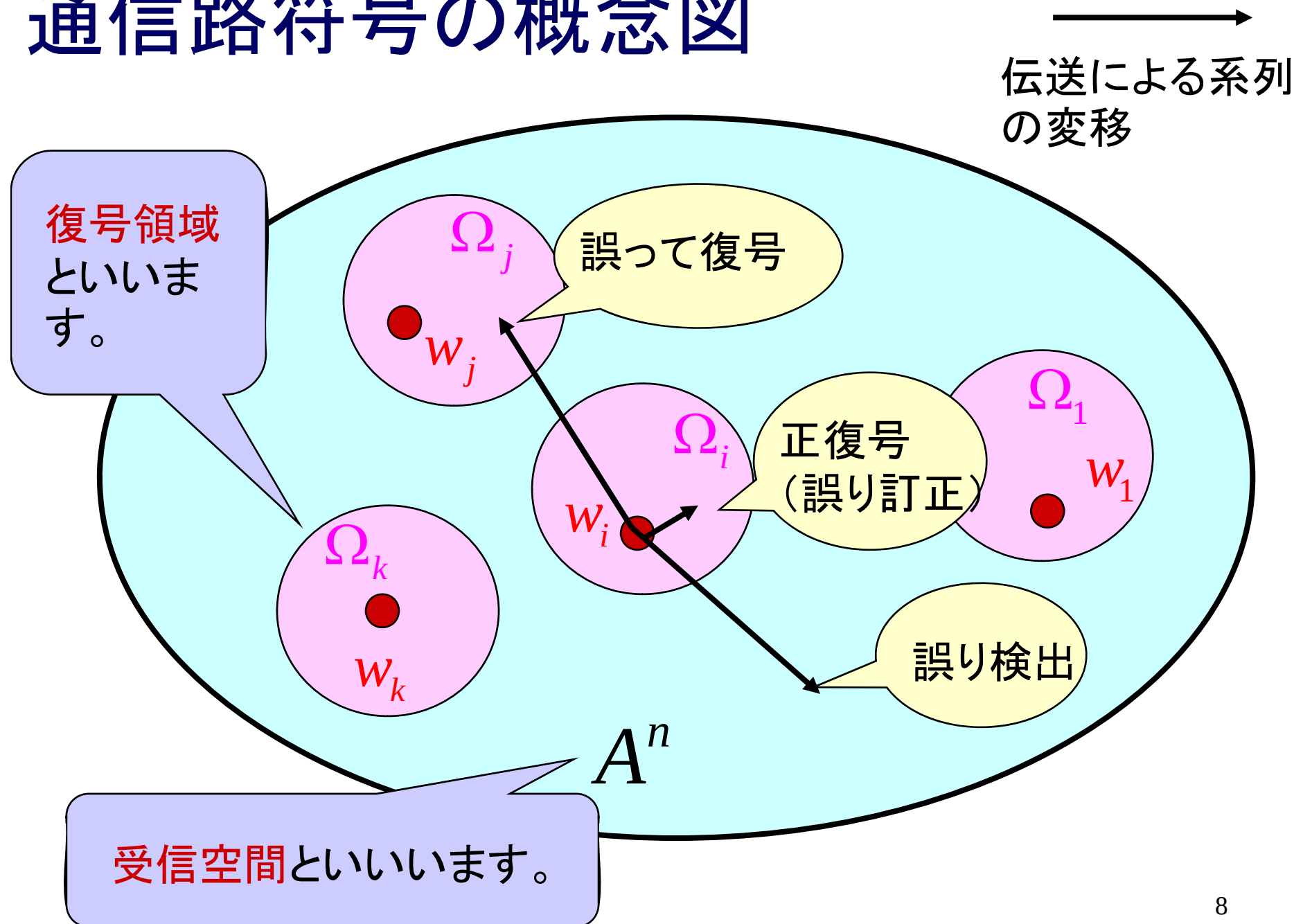
例

$$B^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

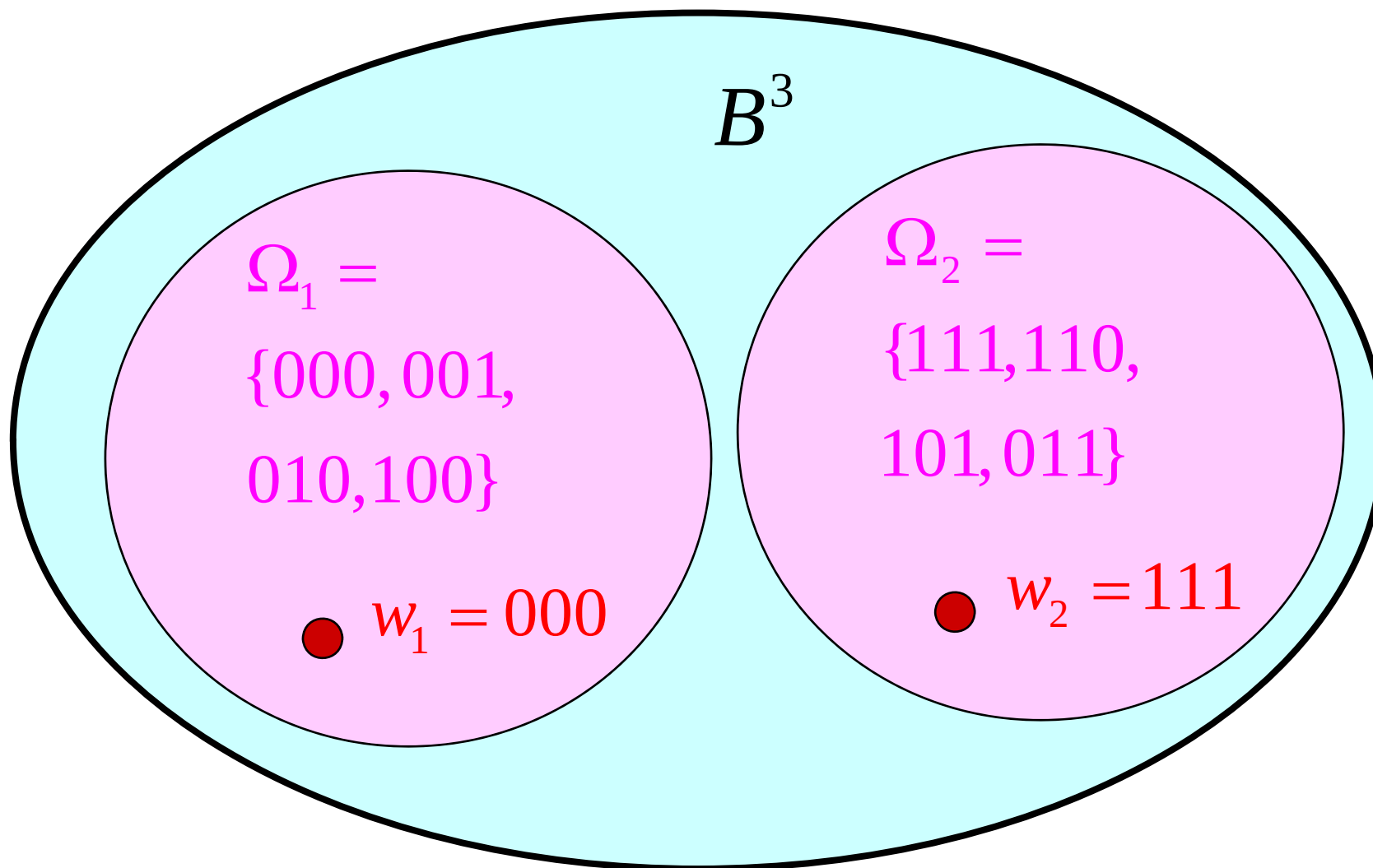
の部分集合である $C = \{000, 111\}$ のなかの要素だけを通信路符号として用いる。

このように、送信記号系列の部分集合 $C \subseteq A^n$ を(通信路)符号という。また、通信路符号に含まれる各系列 $w \in C$ を(通信路)符号語という。なお、通信路符号としては等長符号が持ちられることが多い。

通信路符号の概念図



例



代表的系列

代表的系列

情報源 $A = \left\{ \begin{array}{cccc} a_1 & , & a_2 & , & \cdots & , & a_r \\ p_1 & , & p_2 & , & \cdots & , & p_r \end{array} \right\}$ の長さ n

の系列の中で、各記号 a_i の出現個数 n_i と n の比が生成確率 p_i に等しいような系列を代表的系列という。

例

情報源 $A = \left\{ \begin{matrix} a & , & b & , & c \\ 1/6 & , & 2/6 & , & 3/6 \end{matrix} \right\}$ の長さ12の代表的系列。

$$p_a \simeq \frac{n_a}{n}, \therefore n_a = np_a = 12 \times \frac{1}{6} = 2$$

$$p_b \simeq \frac{n_b}{n}, \therefore n_b = np_b = 12 \times \frac{2}{6} = 4$$

$$p_c \simeq \frac{n_c}{n}, \therefore n_c = np_c = 12 \times \frac{3}{6} = 6$$

bbaccccabccb

cbccbabcabc

⋮

練習

次の各情報源と系列の長さに対して、代表的系列をそれぞれ3個示せ。

$$(1) \quad B = \left\{ \begin{array}{cc} 0 & , \quad 1 \\ 1/3 & , \quad 2/3 \end{array} \right\} \quad n = 24$$

$$(2) \quad S = \left\{ \begin{array}{cccc} a & , & b & , & c & , & d \\ 1/2 & , & 1/4 & , & 1/6 & , & 1/12 \end{array} \right\} \quad n = 24$$

代表的系列の個数

代表的系列の個数とエントロピー

情報源 $A = \{a_1, a_2, \dots, a_r\}$ の十分な長さ n
の代表的系列の個数 N は、次式で表される。

ここで、 $N = 2^{nH(A)}$ は情報源 A のエントロピー。

証明

長さ n の代表的系列を $\mathbf{w} = w_1 \cdots w_n$ とし、その発生確率を $P(\mathbf{w})$ とする。

記号 a_i の発生確率が p_i であり、それが $\mathbf{w}$ に $n_i = np_i$ 個含まれているので次式が成り立つ。

$$P(\mathbf{w}) = \prod_{a_i \in A} p_i^{n_i} = \prod_{i=1}^r p_i^{np_i}$$

底の変換と指数法則より、次のように計算できる。

$$\begin{aligned} P(\mathbf{w}) &= \prod_{i=1}^r 2^{np_i \log p_i} = 2^{n \sum_{i=1}^r p_i \log p_i} \\ &= 2^{-nH(A)} \end{aligned}$$

n が十分大きければ、代表的系列以外の発生確率は0に漸近する。したがって、代表的系列の個数 N は発生確率の逆数である。すなわち、次式が成り立つ。

$$N = \frac{1}{P(\mathbf{w})} = 2^{nH(A)}$$

QED

情報(伝送)速度

1 記号あたりに伝送される情報量を、
情報速度あるいは情報伝送速度という。

情報速度は主に

$$R \text{ [bit / 記号]}$$

の記号を用いる。

物理的な伝送速度との関係

1 秒(単位時間)あたりに伝送可能な記号 k [記号 / 秒] とする。
このとき、1 秒あたりで伝送される情報量 R^* [bit / 秒] は、
次式で与えられる。

$$R^* \text{ [bit / 秒]} = k \text{ [記号 / 秒]} \times R \text{ [bit / 記号]}$$

情報速度と符号語数

情報源 A の長さ n の代表的系列の中から M 個の系列を選んだとする。これらの系列(符号語)を代表的系列の中から等しい確率で用いるとする。

$$\left. \begin{array}{l} \bullet \mathbf{w}_1 = w_1^1 \cdots w_n^1 \\ \bullet \mathbf{w}_2 = w_1^2 \cdots w_n^2 \\ \vdots \\ \bullet \mathbf{w}_N = w_1^N \cdots w_n^N \end{array} \right\} N = 2^{nH(A)}$$

M

このとき、各符号を伝送したときの情報量 R_n は、次式である。

$$R_n = -\log_2 \frac{1}{M} = \log_2 M \text{ [bit]}$$

自己情報量の式

この情報の伝送に n 記号用いているので、情報速度は以下の式で与えられる。

$$R = \frac{\log_2 M}{n} \text{ [bit / 記号]}$$

この式を逆に用いることにより、符号語数は情報速度で表すことができる。すなわち、次式が成り立つ。

$$M = 2^{nR} \text{ [個]}$$

この式は通信路符号化定理を導くときに用いられる。

通信路符号化定理(重要)

通信路符号化定理

通信路容量 C の通信路 T において、情報速度 R が

$$R < C$$

を満たすとき、任意の正数 $\varepsilon > 0$ に対して、復号誤り率 $P_e(T)$

$$P_e(T) < \varepsilon$$

限りなく0にできるという意味。

を満たす(通信路)符号 $\mathbf{w} = \{w_1, \dots, w_M\}$ が存在する。
もし、

$$R > C$$

なら、そのような符号は存在しない。

証明

A_0 を通信路に接続すれば、伝送される情報量が C となる情報源

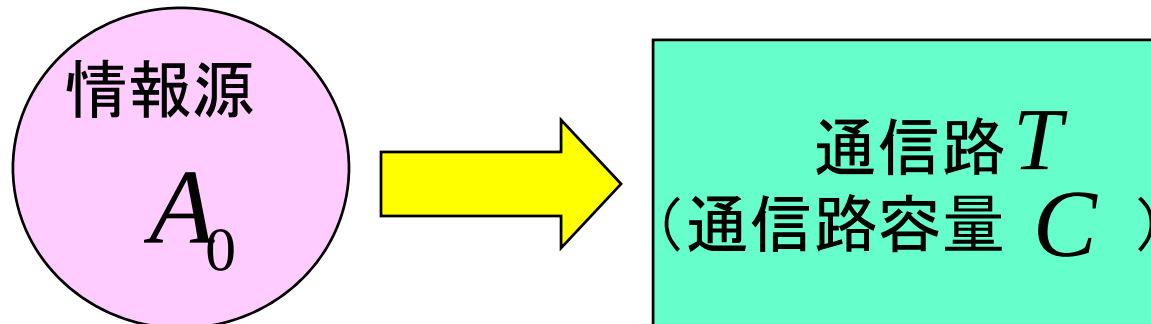
証明の方針

- (1) 通信路容量 C を達成する情報源を A_0 とする。
- (2) A_0 から発生する長さ n の代表的系列の中から、 $M = 2^{nR}$ 個の符号語をランダムに選び、その集合を符号

$$W = \{w_1, w_2, \dots, w_M\}$$

とする。

- (3) これらより、誤り確率を求める。



受信される情報源を B_0 とする。

このとき、 A_0 は通信路容量を達成する情報源なので、通信路容量の定義より次式が成り立つ。

$$C = H(A_0) - H(A_0 | B_0)$$

一方、以前に示したように、長さ n の代表的系列の数は

$$N = 2^{nH(A_0)}$$

ランダム符号化という。

である。これらの中から代表的系列を、 $M = 2^{nR}$ 個ランダムに選ぶ。(もちろん $R < H(A_0)$ である。)

したがって、ある代表的系列が符号語として選ばれる確率は、

$$p_s = M / N = 2^{nR} / 2^{nH(A_0)} \quad \text{である。}$$

出力系列1つあたりの
入力系列数

$$2^{nH(A_0|B_0)}$$

$$M = 2^{nR}$$

符号語数

代表的系列数

$$N = 2^{nH(A_0)}$$

w_i

w_{i+1}

入力系列

出力系列

$$2^{nH(B_0)}$$

y

出力系列に1
つの符号語し
かなければ誤
らない。

符号語 W_i を通信路 T を通して送り、受信系列 Y が得られたとする。通信路には一般に誤りがあり、送信符号(送信系列)は確率的にしかわからない。

しかし、条件付エントロピー $H(A_0 | B_0)$ が平均情報量を意味することから、受信系列が y として受信される長さ n の代表的系列の個数 n_y は平均すれば次式表される。

$$n_y = 2^{nH(A_0 | B_0)}$$

条件付エントロピーが1記号あたりの情報量であり、[bit/記号]の単位を持つことに注意する。

もし、 $n_y = 2^{nH(A_0|B_0)}$ 個の代表的系列の中に W_i 以外に $W = \{w_1, \dots, w_M\}$ の符号語がなければ、送られた符号語が W_i 受信語 y から一意に確定し、誤りなく復号できる。

n_y 個の代表的系列を選んだとき、 W_i 以外の W の符号語を一つも選ばれない確率 $\overline{p_e}$ は、先ほど求めた p_s を用いて以下のように表せる。

$$\overline{p_e} = (1 - p_s)^{n_y}$$

連続して n_y 回符号語を選ばない確率。

これらまでの式を元にこの確率を計算する。

$$\overline{p_e} = (1 - p_s)^{n_y}$$

$$= \left(1 - \frac{2^{nR}}{2^{nH(A_0)}} \right)^{2^{nH(A_0|B_0)}}$$

$$\simeq 1 - 2^{nH(A_0|B_0)} \cdot \left(\frac{2^{nR}}{2^{nH(A_0)}} \right)$$

$$= 1 - 2^{-n(H(A_0) - H(A_0|B_0))} \cdot 2^{nR}$$

$$= 1 - 2^{-n(C-R)}$$

$R < C$ なので
 $C - R > 0$

テイラー展開
の1次の項

これは誤りなく復号できる
確率なので、誤り確率 p_e
は次式で表される。

$$p_e = 1 - \overline{p_e} = 2^{-n(C-R)}$$

以上より、誤り確率は
~~を~~大きくすればいくらで
も小さくできる。

(前半の証明終)

(後半の証明)

$R > C$ でしかも $p_e \rightarrow 0$ とできたとする。

このときには、誤りなく復号されるのだから、通信路を通じて実際に R [bit/記号]の情報が伝送されたことになる。

しかし、これは通信路容量の定義に矛盾する。

QED

情報源符号化と通信路符号化のまとめ2

符号化	定理	意味
情報源符号化	シャノンの第一定理 (情報源符号化定理)	平均符号長はエントロピー以上 $\frac{H(S)}{\log r} < \bar{L}$
通信路符号化	シャノンの第二定理 (通信路符号化定理)	情報速度は通信路容量以下 $R < C$ $= \max\{H(A) - H(A B)\}$

情報源符号化定理と雑音の無い通信路

平均符号長 $\bar{L}$ の r 元符号を、雑音の無い通信路を通して伝送することを考える。

符号を構成する記号を1個伝送するために必要な時間を秒とする。

$$\tau [\text{秒} / \text{記号}] = \frac{1}{k [\text{記号} / \text{秒}]}$$

このとき、伝送速度 $\hat{R}$ [符号 / 秒] は次式で表される。

$$\hat{R} = \frac{1}{\tau L}$$

符号語を伝送する時間(秒)

一方、雑音の無い通信路では、通信路容量 C [bit / 記号] は次式で表せる。

$$C = \max \{H(A) - H(A|B)\}$$

$$= \max H(A)$$

$$= \log r$$

雑音がないので、
 $H(A|B) = 0$

ここで、通信路容量の単位を変換する。

$$C [\text{bit} / \text{記号}] \rightarrow C^* [\text{bit} / \text{秒}]$$

$$C^* [\text{bit} / \text{秒}] = \frac{C [\text{bit} / \text{記号}]}{\tau [\text{秒} / \text{記号}]} = \frac{\log r}{\tau}$$

$$\therefore \frac{1}{\tau} = \frac{C^*}{\log r}$$

よって、 $\hat{R} = \frac{1}{\tau L} = \frac{C^*}{L \log r}$

分子は変化しない

r 元符号なので
 $\log r$ も変化しない。

伝送速度をあげるには、
平均符号長を小さくする。

シャノンの第1定理より、

$$\bar{L} \rightarrow \frac{H(A)}{\log r}$$

2元符号の場合

$$\bar{L} \rightarrow H(A)$$

とできる。

したがって、雑音のない通信路では、伝送速度を以下のようにすることができる。

$$\hat{R} = \frac{C^*}{L \log r} \rightarrow \frac{C^*}{H(A)} \quad [\text{符号 / 秒}]$$

すなわち、雑音の無い通信路では、シャノンの第1定理と第2定理は同値となる。