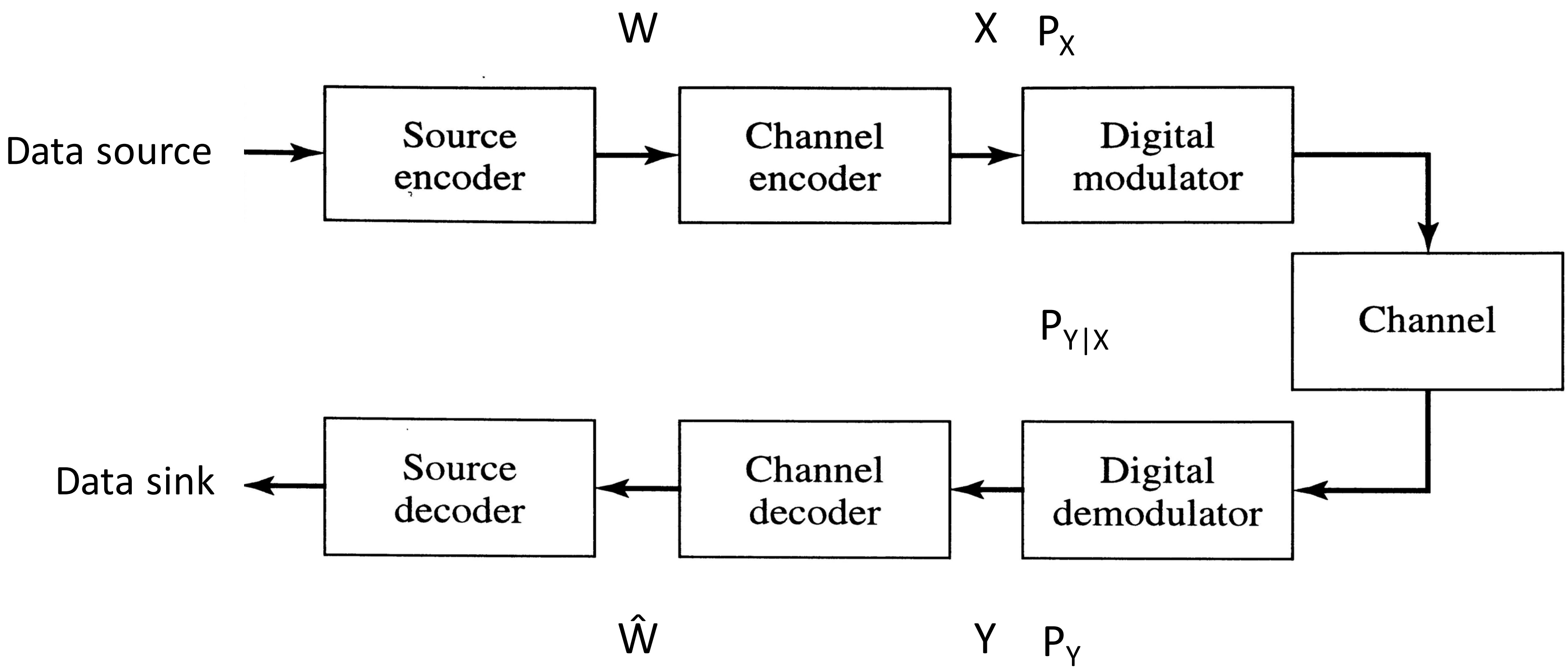
ECE 515
Information Theory

Channel Capacity and Coding

Information Theory Problems

- How to transmit or store information as efficiently as possible.
- What is the maximum amount of information that can be transmitted or stored reliably?
- How can information be kept secure?

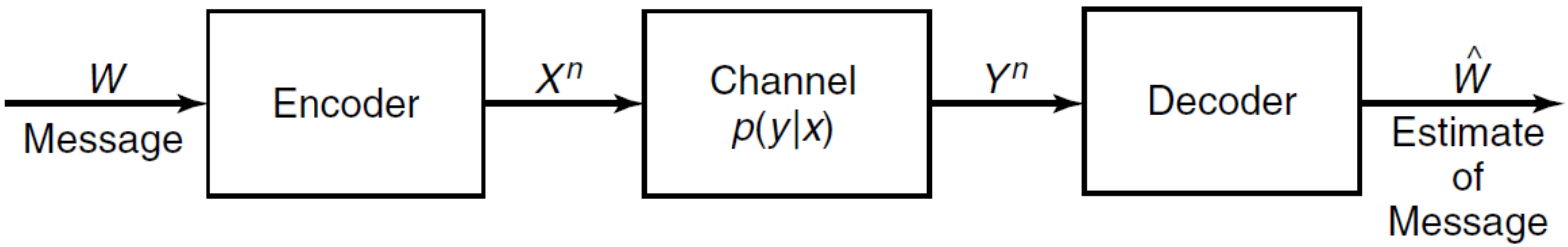
Digital Communications System



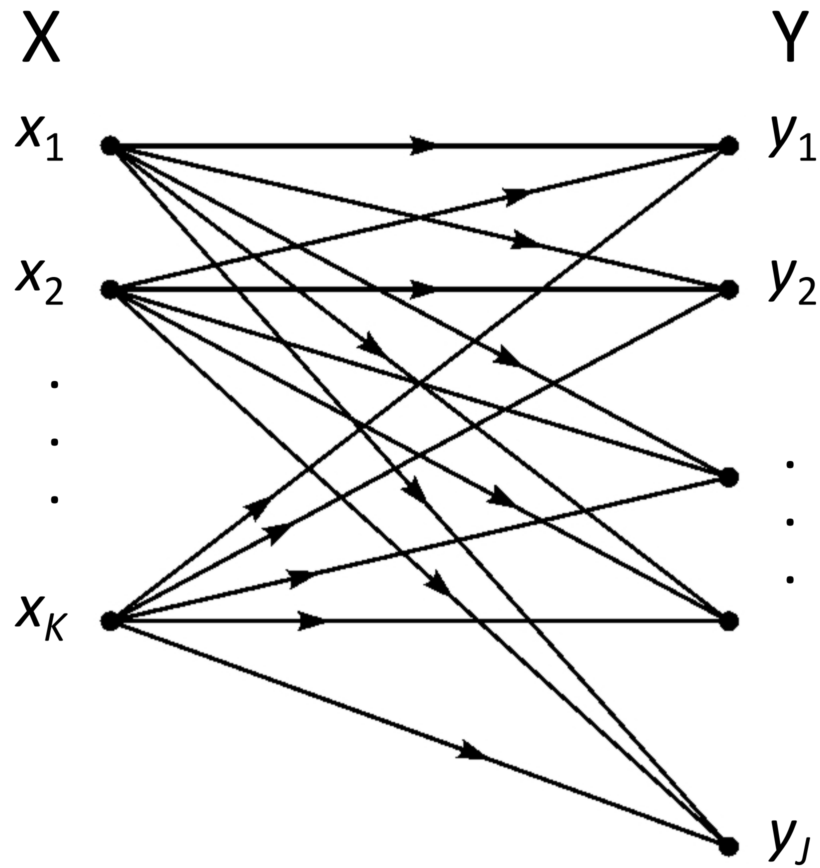
$\hat{W}$ is an estimate of W

Communication Channel

- Cover and Thomas Chapter 7



Discrete Memoryless Channel

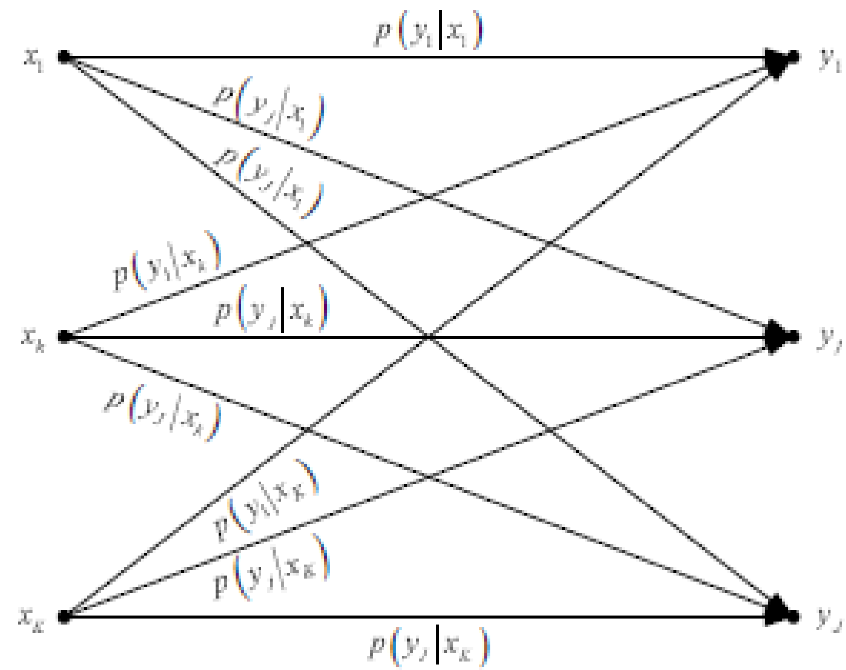


$J \geq K \geq 2$

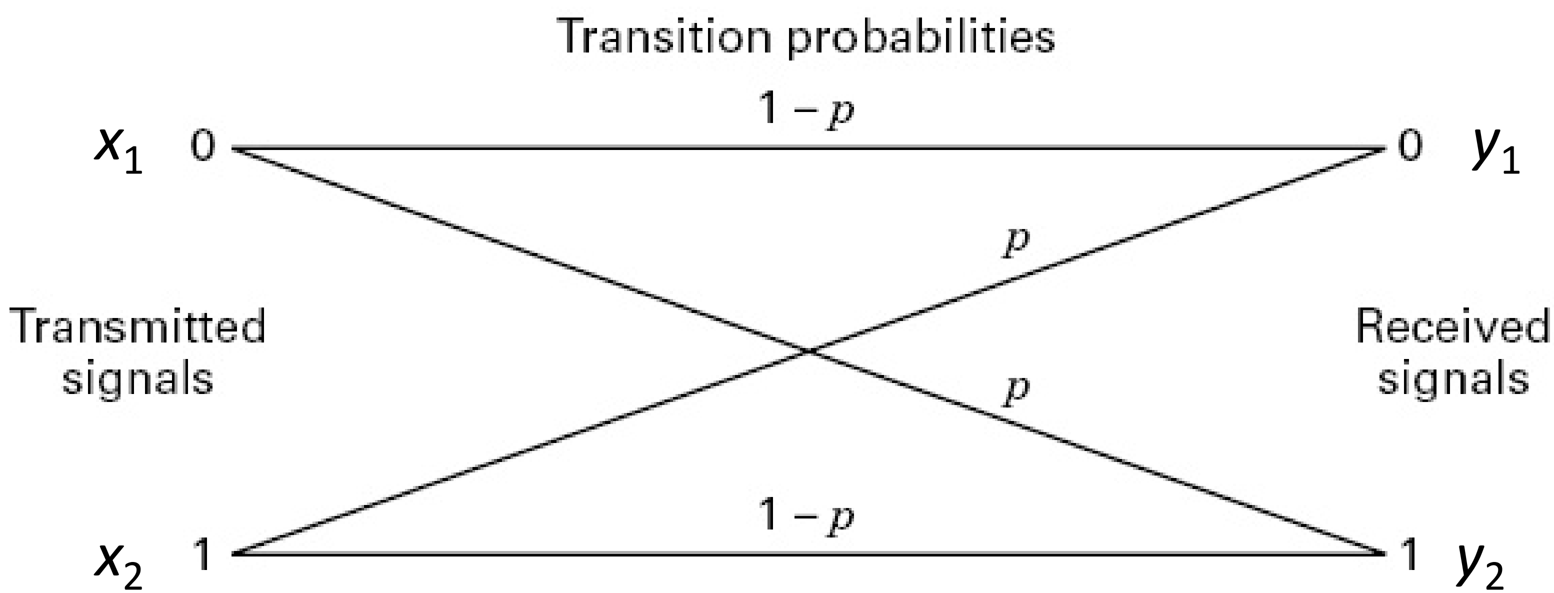
Discrete Memoryless Channel

Channel Transition Matrix

$$P = \begin{bmatrix} p(y_1 | x_1) & \cdots & p(y_1 | x_k) & \cdots & p(y_1 | x_K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p(y_j | x_1) & \cdots & p(y_j | x_k) & \cdots & p(y_j | x_K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p(y_J | x_1) & \cdots & p(y_J | x_k) & \cdots & p(y_J | x_K) \end{bmatrix}$$



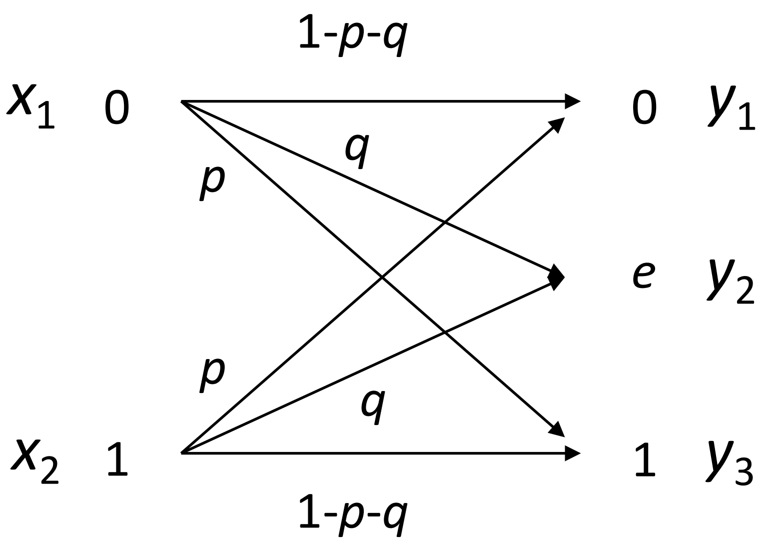
Binary Symmetric Channel



channel matrix

$$\begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$

Binary Errors with Erasure Channel

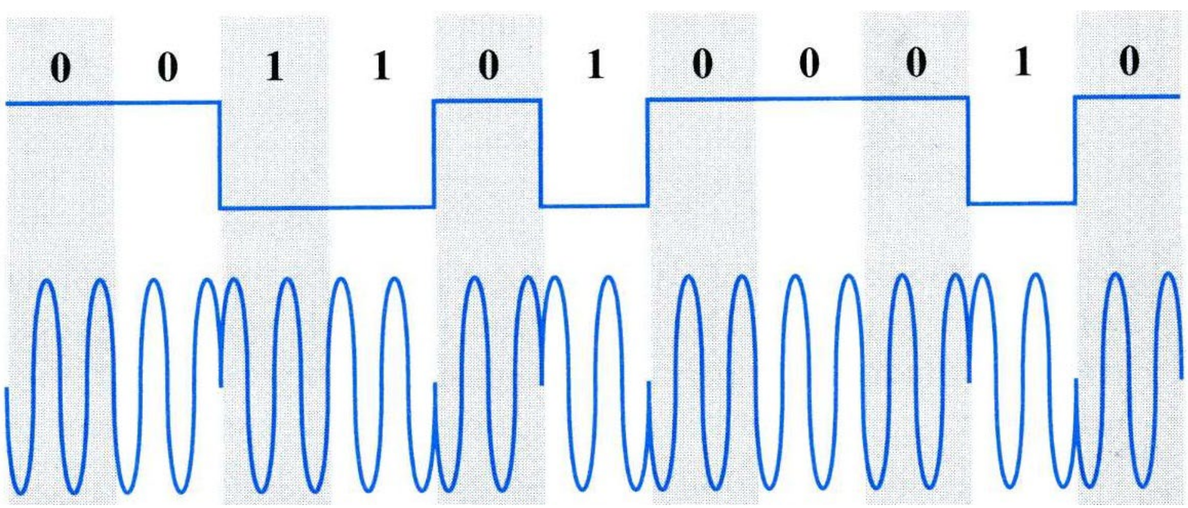


BPSK Modulation K=2

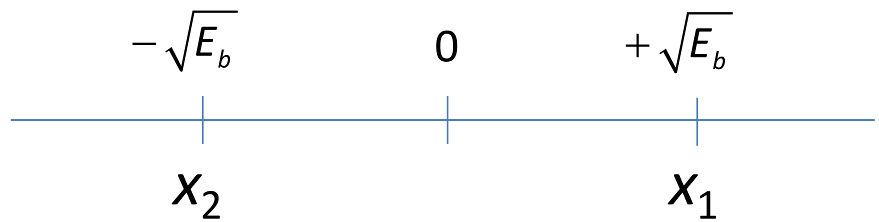
$$x_1(t) = +\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t)$$ 0 bit

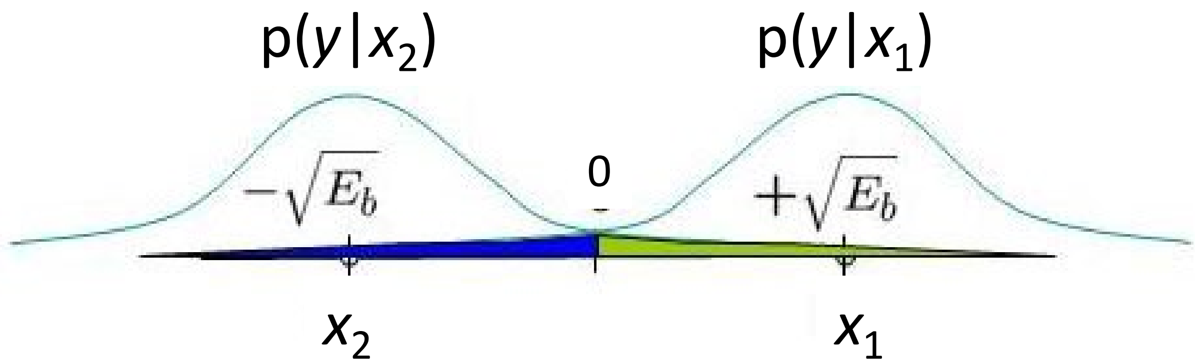
$$x_2(t) = -\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t)$$ 1 bit

E_b is the energy per bit
T_b is the bit duration
f_c is the carrier frequency



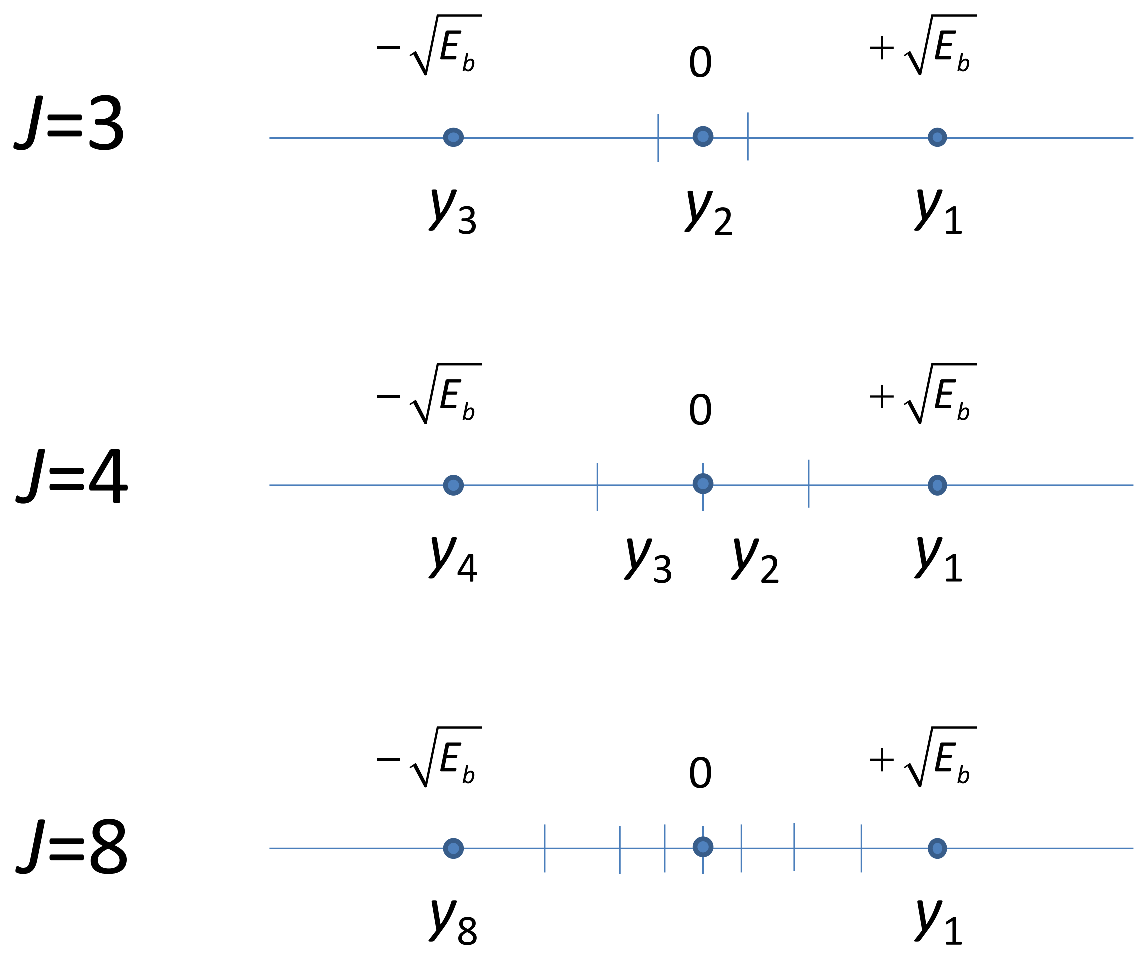
BPSK Demodulation in AWGN



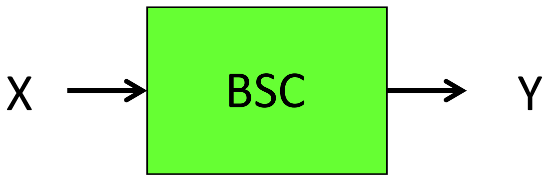


$$\mathrm{p}_{Y|X}(y \mid x = x_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - x_k)^2}{2\sigma^2} \right]$$

BPSK Demodulation K=2



Mutual Information for a BSC

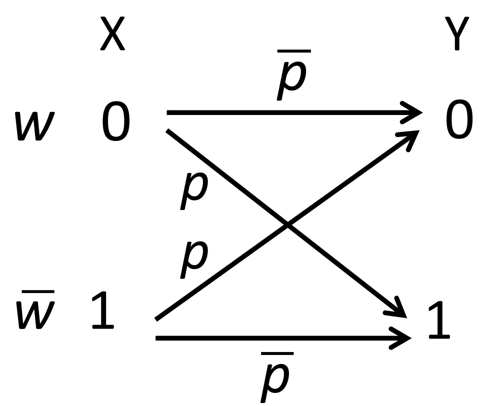


crossover probability p

$$\bar{p} = 1 - p$$

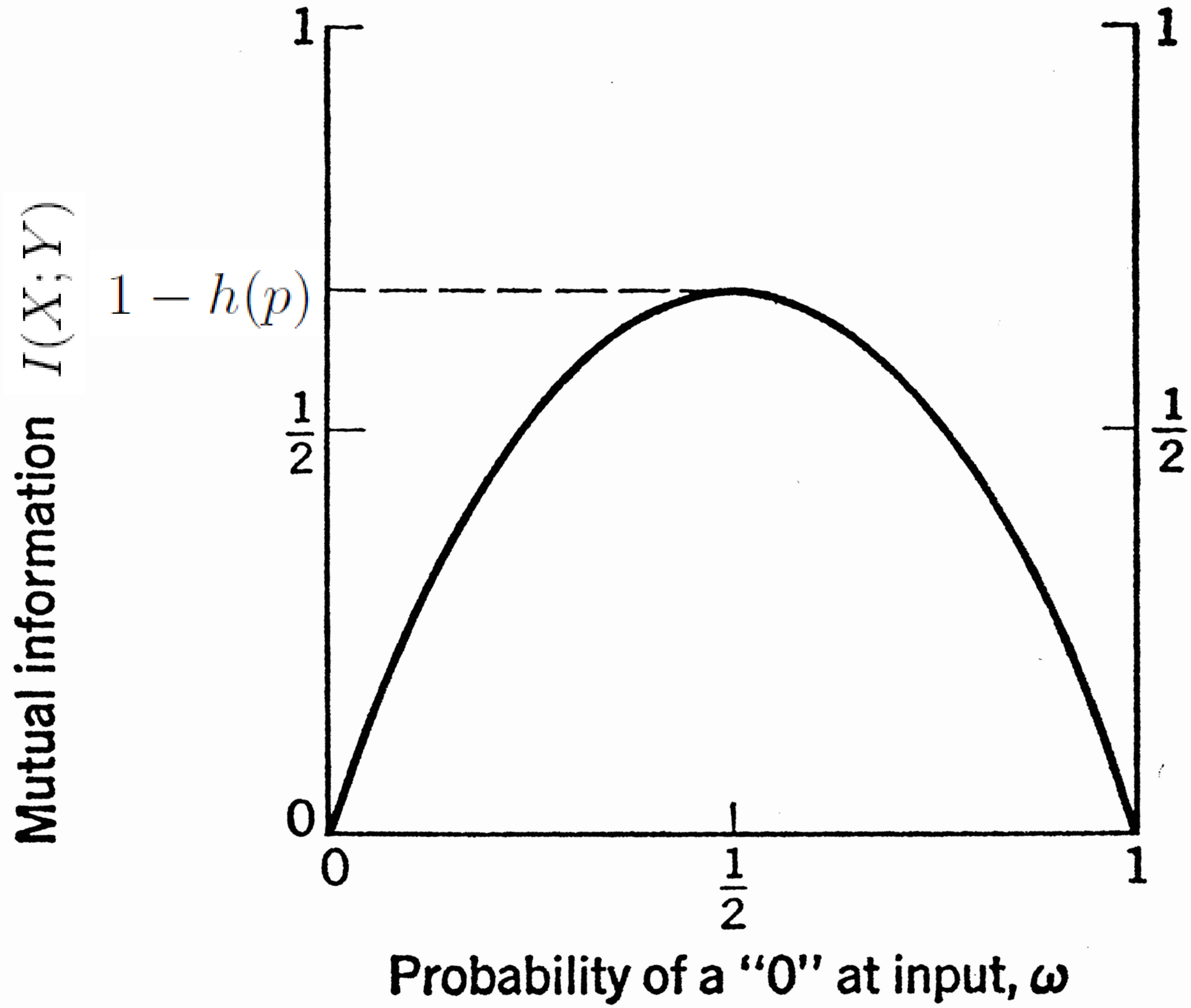
channel matrix

$$P_{Y|X} = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$



$$p(x = 0) = w$$

$$p(x = 1) = 1 - w = \bar{w}$$



Mutual information $I(X;Y)$
$1 - h(p)$
1
$\frac{1}{2}$
0
0
$\frac{1}{2}$
1
1
$\frac{1}{2}$
Probability of a "0" at input, ω

Convex Functions

Definition *(Convex function):*

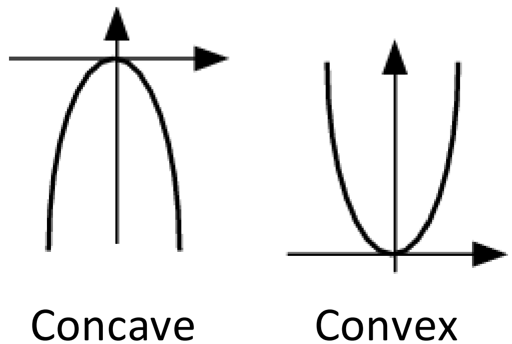
A real function $f(x)$, defined on a convex set $\mathcal{S}$ (e.g., input symbol distributions), is *concave* (*convex down*, *convex "cap"* or *convex* $\cap$) if, for any point x on the straight line between the pair of points x_1 and x_2, i.e., $x = \lambda x_1 + (1-\lambda)x_2$ ($\lambda \in [0,1]$), in the convex set $\mathcal{S}$:

$$\boxed{f(x) \geq \lambda f(x_1) + (1-\lambda) f(x_2)}$$

otherwise, if:

$$\boxed{f(x) \leq \lambda f(x_1) + (1-\lambda) f(x_2)}$$

then the function is said to be simply *convex* (*convex up*, *convex "cup"* or *convex* $\cup$).



Concave Function

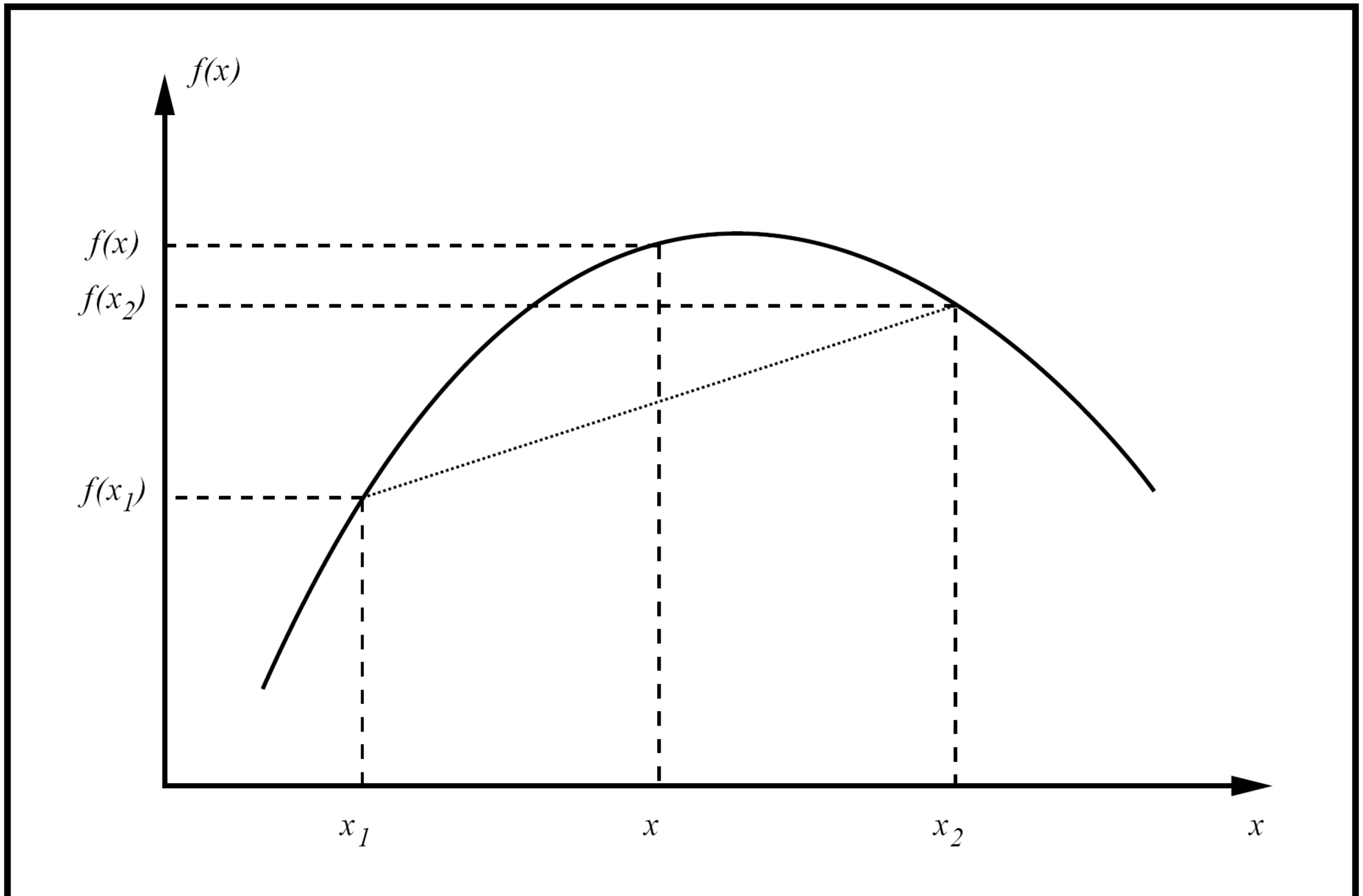


Figure 3.3: Convex ∩ (*convex down* or *convex "cap"*) function.

Convex Function

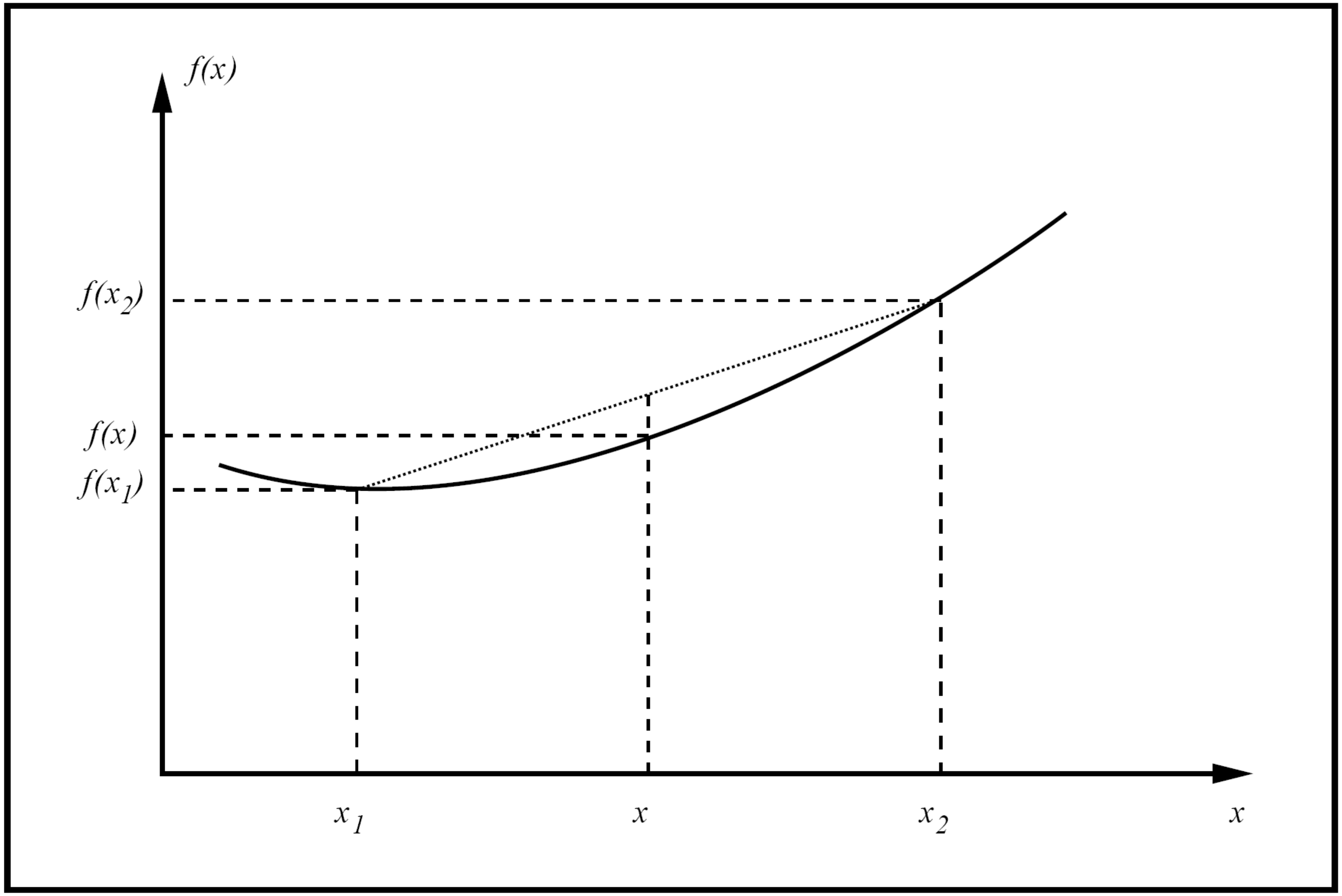


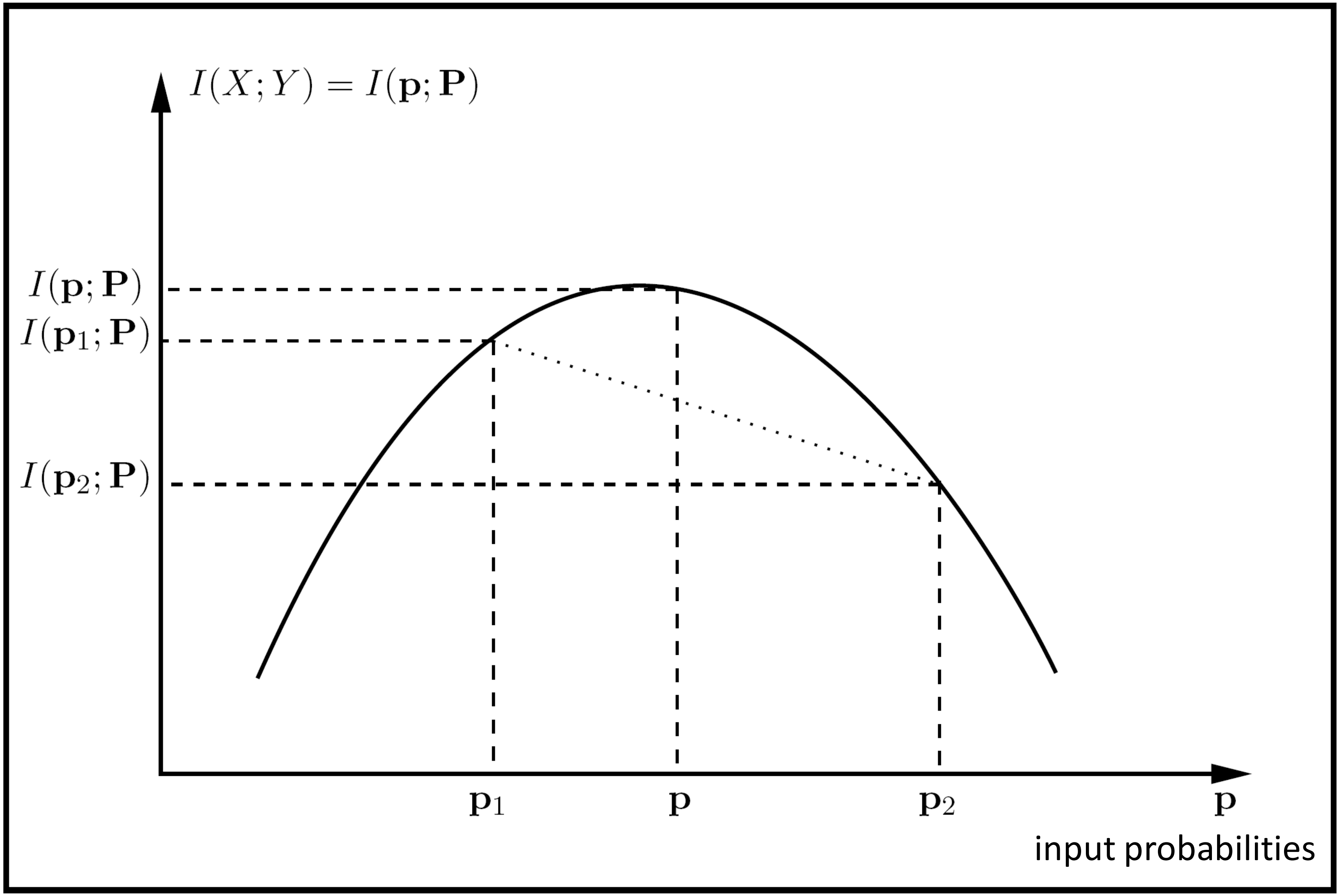
Figure 3.4: Convex ∪ (*convex up* or *convex "cup"*) function.

Mutual Information

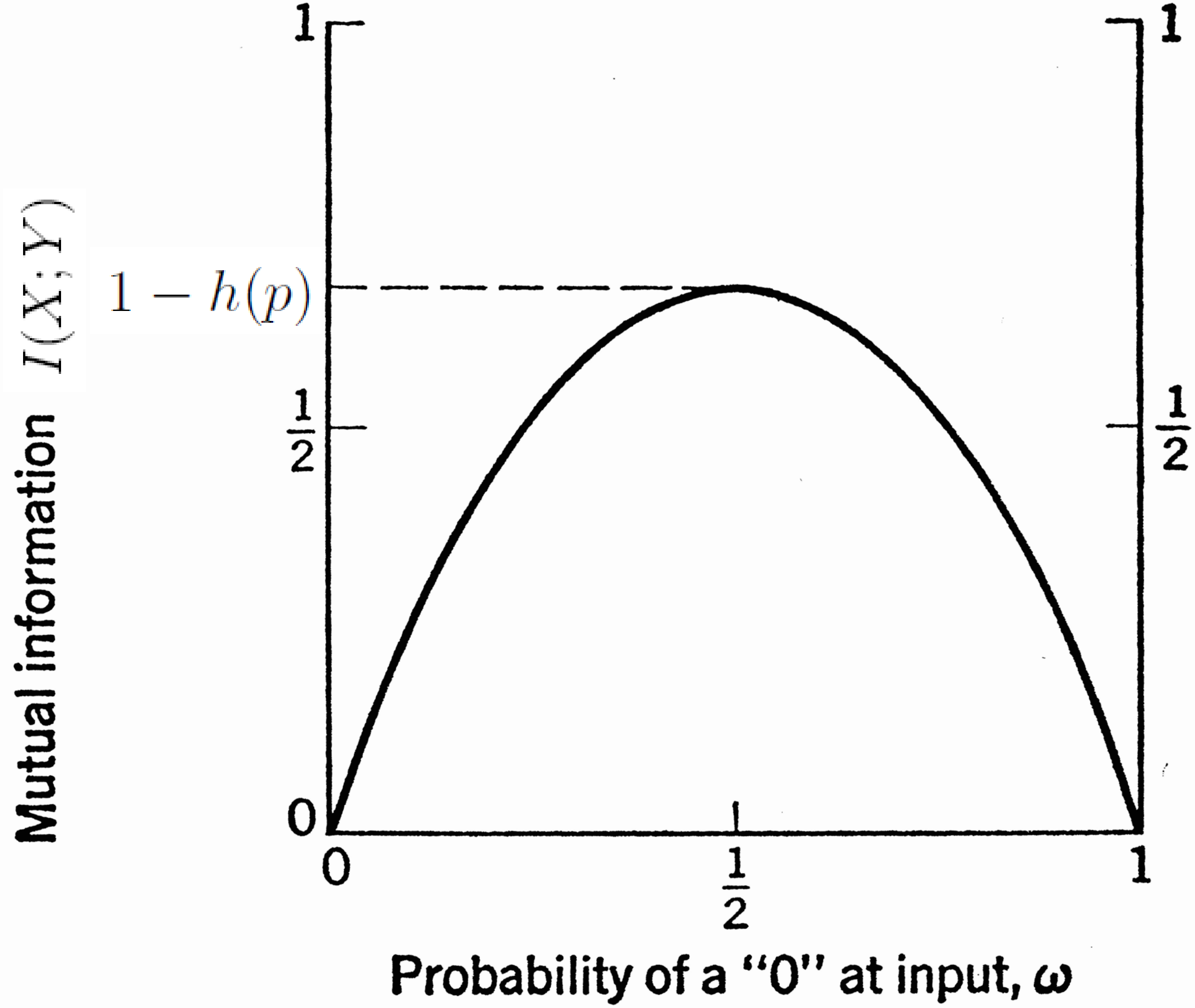
$$
\begin{aligned}
I(X;Y) &= \sum_{k=1}^{K}\sum_{j=1}^{J} p(x_k)p(y_j|x_k)\log_b\left[\frac{p(y_j|x_k)}{\sum_{l=1}^{K} p(x_l)p(y_j|x_l)}\right] \\
&= f\left[p(x_k), p(y_j|x_k)\right] \\
I(X;Y) &= f\left(\mathbf{p}, \mathbf{P}\right)
\end{aligned}
$$

Theorem *(Convexity of the mutual information function):*

The (average) mutual information $I(X;Y)$ is a *concave* (or *convex "cap"*, or *convex* $\cap$) function over the *convex set* $\mathcal{S}_{\mathbf{p}}$ of all possible input distributions $\{\mathbf{p}\}$.



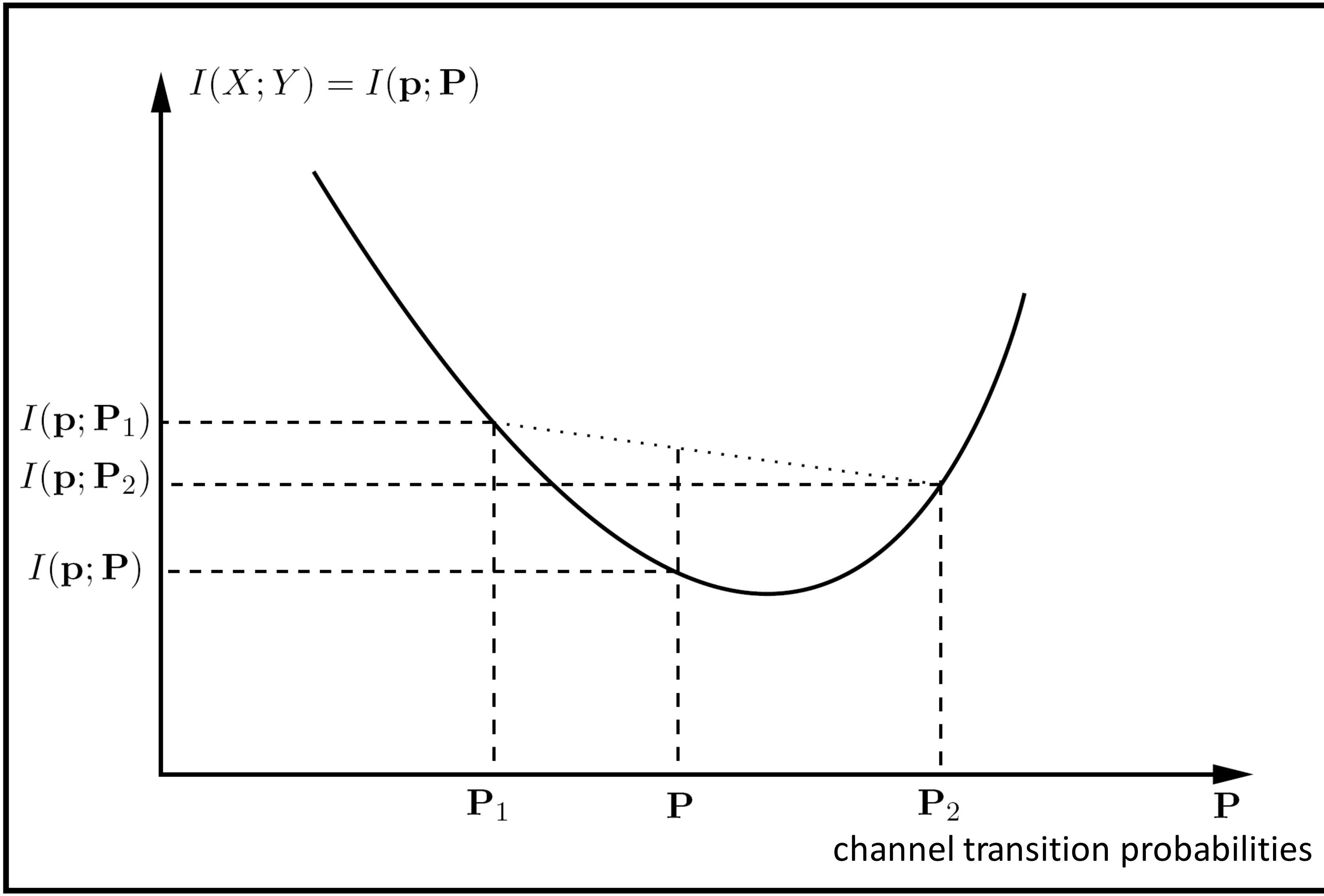
$I(X;Y) = I(\mathbf{p};\mathbf{P})$
$I(\mathbf{p};\mathbf{P})$
$I(\mathbf{p}_1;\mathbf{P})$
$I(\mathbf{p}_2;\mathbf{P})$
$\mathbf{p}_1$
$\mathbf{p}$
$\mathbf{p}_2$
$\mathbf{p}$
input probabilities



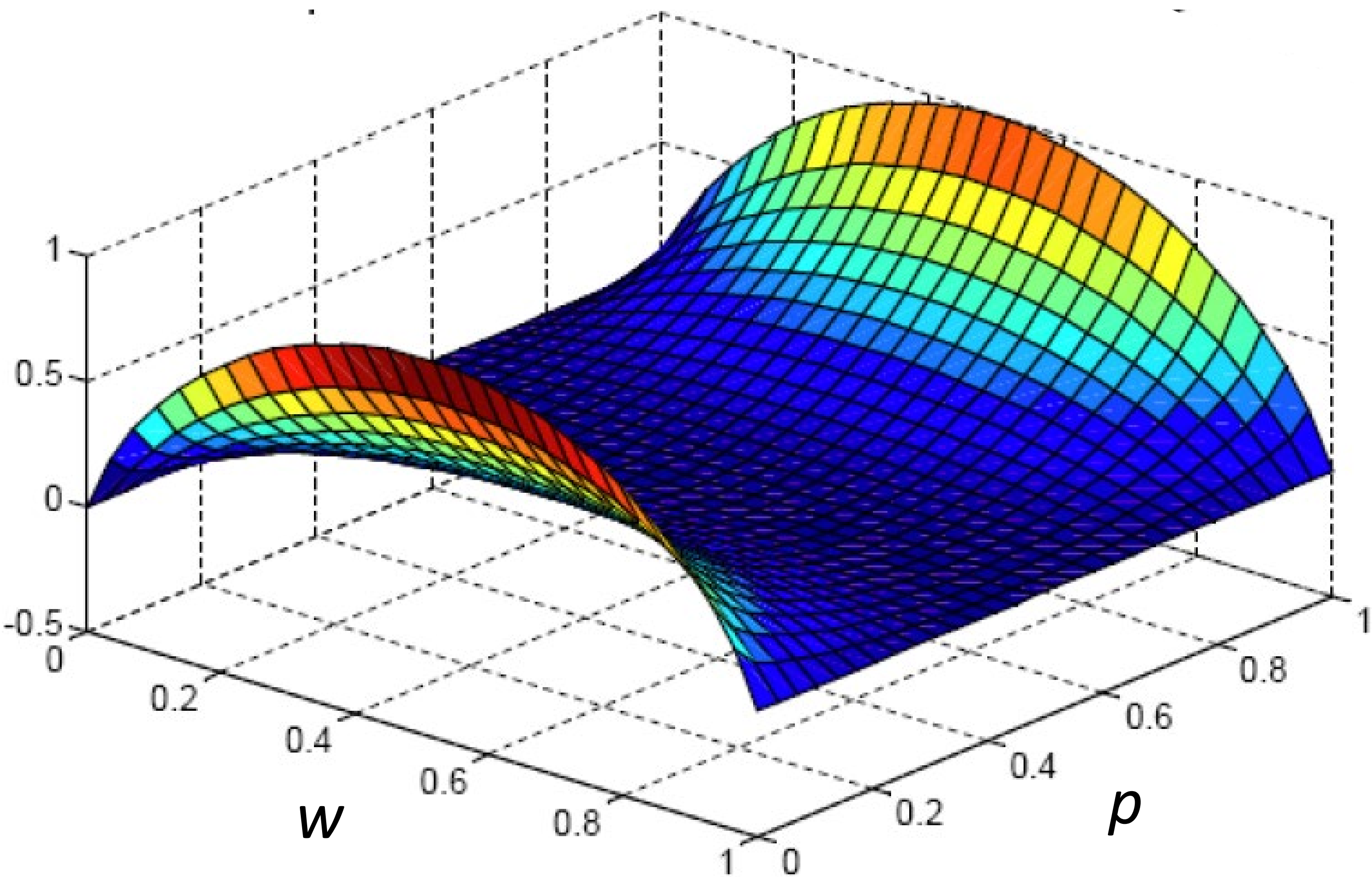
Mutual information $I(X;Y)$
$1 - h(p)$
1
$\frac{1}{2}$
0
0
$\frac{1}{2}$
1
1
$\frac{1}{2}$
Probability of a "0" at input, ω

Theorem *(Convexity of the mutual information function):*

The (average) mutual information $I(X;Y)$ is a *convex* (or *convex "cup"*, or *convex* $\cup$) function over the *convex set* $\mathcal{S}_{\mathbf{P}}$ of all possible transition probability matrices $\{\mathbf{P}\}$.


$I(X;Y) = I(\mathbf{p};\mathbf{P})$
$I(\mathbf{p};\mathbf{P}_1)$
$I(\mathbf{p};\mathbf{P}_2)$
$I(\mathbf{p};\mathbf{P})$
$\mathbf{P}_1$
$\mathbf{P}$
$\mathbf{P}_2$
$\mathbf{P}$
channel transition probabilities

BSC I(X;Y)



Properties of the Channel Capacity

- $C \geq 0$ since $I(X;Y) \geq 0$
- $C \leq \log|X| = \log(K)$ since

$$C = \max I(X;Y) \leq \max H(X) = \log(K)$$

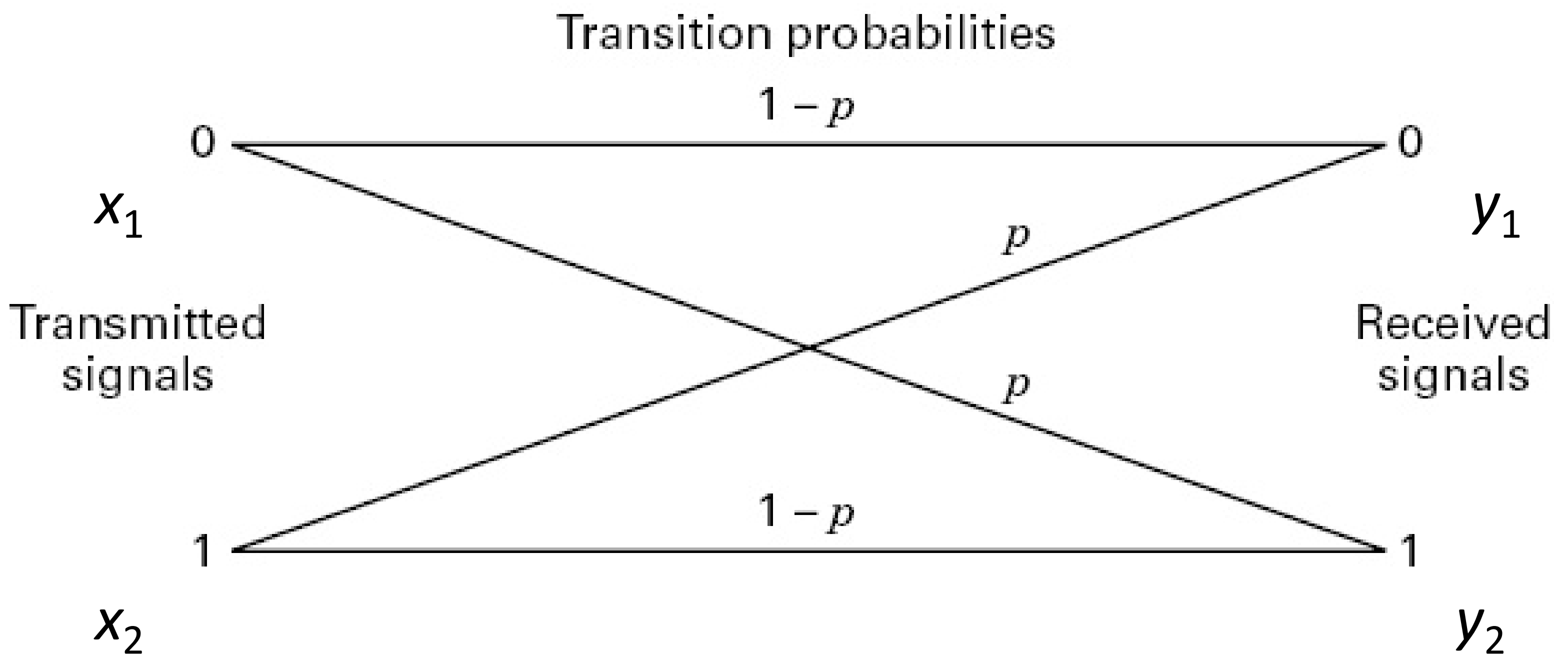
- $C \leq \log|Y| = \log(J)$ for the same reason
- $I(X;Y)$ is a concave function of $p(X)$, so a local maximum is a global maximum

Channel Capacity

The *maximum* value of $I(X;Y)$ as the input probabilities $p(x_i)$ are varied is called the Channel Capacity

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

Binary Symmetric Channel



$$C = 1 - h(p) \text{ for } w = \bar{w} = 1/2$$

Symmetric Channels

A discrete memoryless channel is said to be **symmetric** if the set of output symbols

$$\{y_j\}, j = 1, 2, \ldots, J,$$

can be partitioned into subsets such that for each subset of the matrix of transition probabilities

- each column is a permutation of the other columns
- each row is a permutation of the other rows.

Binary Channels

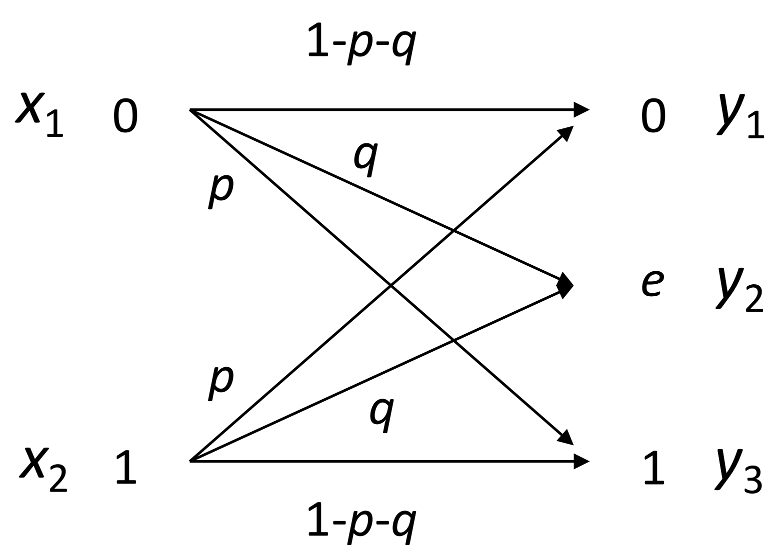
Symmetric channel matrix

$$P = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

Non-symmetric channel matrix

$$P = \begin{bmatrix} 1-p_1 & p_2 \\ p_1 & 1-p_2 \end{bmatrix} \quad p_1 \neq p_2$$

Binary Errors with Erasure Channel



Binary Errors with Erasure Channel

$$P = \begin{bmatrix} 1-p-q & p \\ q & q \\ p & 1-p-q \end{bmatrix}$$

$$P_1 = \begin{bmatrix} 1-p-q & p \\ p & 1-p-q \end{bmatrix}$$

$$P_2 = \begin{bmatrix} q & q \end{bmatrix}$$

Symmetric Channels

- No partition required → strongly symmetric
- Partition required → weakly symmetric

Capacity of a Strongly Symmetric Channel

Theorem

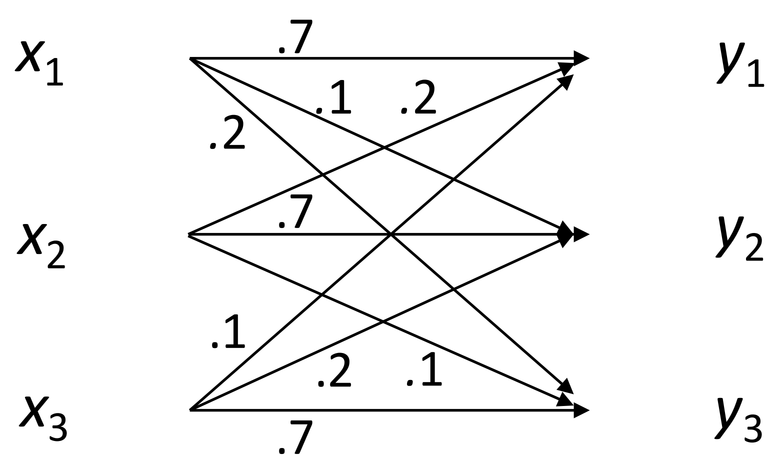
For a discrete symmetric channel, the channel capacity C is achieved with an equiprobable input distribution, i.e., $p(x_k) = \frac{1}{K}, \forall k$, and is given by:

$$\boxed{C = \left[\sum_{j=1}^{J} p(y_j|x_k) \log_b p(y_j|x_k)\right] + \log_b J}$$

$$\mathrm{I}(X;Y) = \sum_{k=1}^{K} \mathrm{p}(x_k) \sum_{j=1}^{J} \mathrm{p}(y_j \mid x_k) \log \mathrm{p}(y_j \mid x_k) + \mathrm{H}(Y)$$

$$= \sum_{j=1}^{J} \mathrm{p}(y_j \mid x_k) \log \mathrm{p}(y_j \mid x_k) + \mathrm{H}(Y)$$

Example $J = K = 3$



Example

$$P_{\mathrm{Y|X}} = \begin{bmatrix} .7 & .2 & .1 \\ .1 & .7 & .2 \\ .2 & .1 & .7 \end{bmatrix}$$

$$\sum_{k=1}^{K} \mathrm{p}(x_k) \sum_{j=1}^{J} \mathrm{p}(y_j \mid x_k) \log \mathrm{p}(y_j \mid x_k)$$

$$= \mathrm{p}(x_1)\left[.7\log.7 + .1\log.1 + .2\log.2\right]$$

$$+\mathrm{p}(x_2)\left[.2\log.2 + .7\log.7 + .1\log.1\right]$$

$$+\mathrm{p}(x_3)\left[.1\log.1 + .2\log.2 + .7\log.7\right]$$

$$= .7\log.7 + .2\log.2 + .1\log.1$$

$$= \sum_{j=1}^{J} \mathrm{p}(y_j \mid x_k) \log \mathrm{p}(y_j \mid x_k)$$

Example

$$\mathrm{H}(Y) = -\sum_{j=1}^{J} \mathrm{p}(y_j)\log \mathrm{p}(y_j)$$

$$\mathrm{p}(y_1) = \sum_{k=1}^{K} \mathrm{p}(y_1 \mid x_k)\mathrm{p}(x_k)$$

$$\mathrm{p}(y_2) = \sum_{k=1}^{K} \mathrm{p}(y_2 \mid x_k)\mathrm{p}(x_k)$$

$$\vdots$$

$$\mathrm{p}(y_J) = \sum_{k=1}^{K} \mathrm{p}(y_J \mid x_k)\mathrm{p}(x_k)$$

$$\mathrm{p}(y_1) = .7\mathrm{p}(x_1) + .2\mathrm{p}(x_2) + .1\mathrm{p}(x_3)$$

$$\mathrm{p}(y_2) = .1\mathrm{p}(x_1) + .7\mathrm{p}(x_2) + .2\mathrm{p}(x_3)$$

$$\mathrm{p}(y_3) = .2\mathrm{p}(x_1) + .1\mathrm{p}(x_2) + .7\mathrm{p}(x_3)$$

r-ary Symmetric Channel

$$P = \begin{bmatrix} 1-p & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & 1-p & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & \frac{p}{r-1} & 1-p & \cdots & \frac{p}{r-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & 1-p \end{bmatrix}$$

r-ary Symmetric Channel

$$C = (1-p)\log(1-p) + (r-1)\frac{p}{r-1}\log(\frac{p}{r-1}) + \log r$$

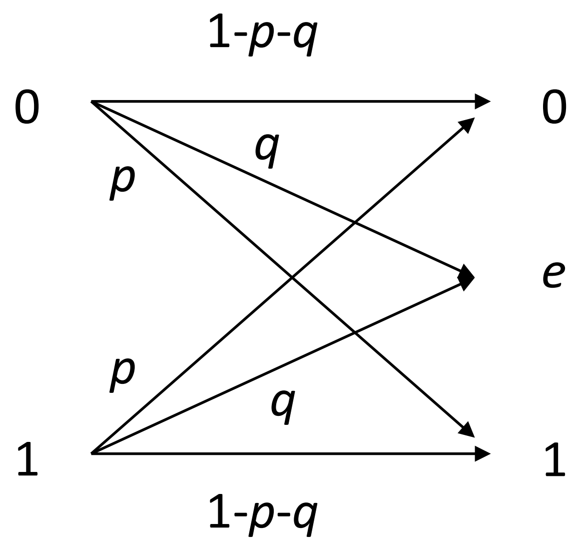
$$= \log r + (1-p)\log(1-p) + p\log(\frac{p}{r-1})$$

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

$$= \log r - h(p) - p\log(r-1)$$

- $r = 2$ C = 1 – h(p)
- $r = 3$ C = $\log_2 3$ – h(p) – p
- $r = 4$ C = 2 – h(p) – $p\log_2 3$

Binary Errors with Erasure Channel



Binary Errors with Erasure Channel

$$P_{Y|X} = \begin{bmatrix} .8 & .05 \\ .15 & .15 \\ .05 & .8 \end{bmatrix}$$

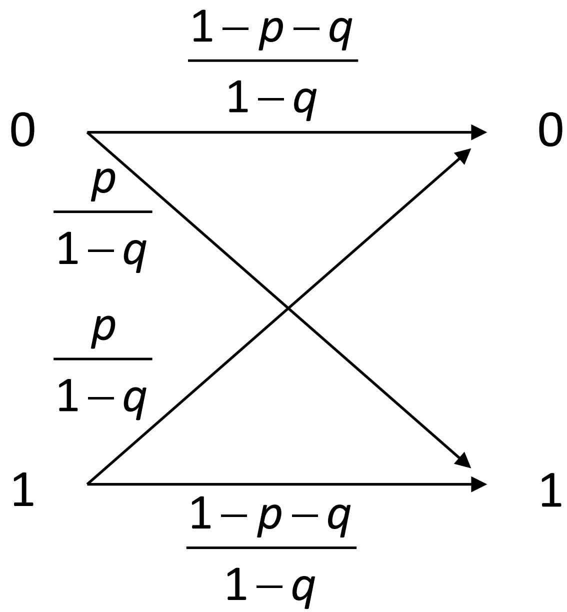
$$P_X = \begin{bmatrix} .5 \\ .5 \end{bmatrix} \qquad P_Y = P_{Y|X} \times P_X = \begin{bmatrix} .425 \\ .15 \\ .425 \end{bmatrix}$$

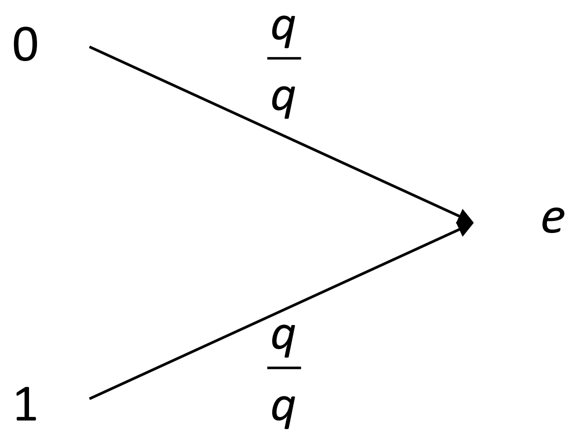
Capacity of a Weakly Symmetric Channel

$$\mathrm{C} = \sum_{i=1}^{L} q_i \mathrm{C}_i$$

- q_i – probability of channel *i*
- C_i – capacity of channel *i*

Binary Errors with Erasure Channel





$$P_1 = \begin{bmatrix} .9412 & .0588 \\ .0588 & .9412 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 1.0 & 1.0 \end{bmatrix}$$

Binary Errors with Erasure Channel

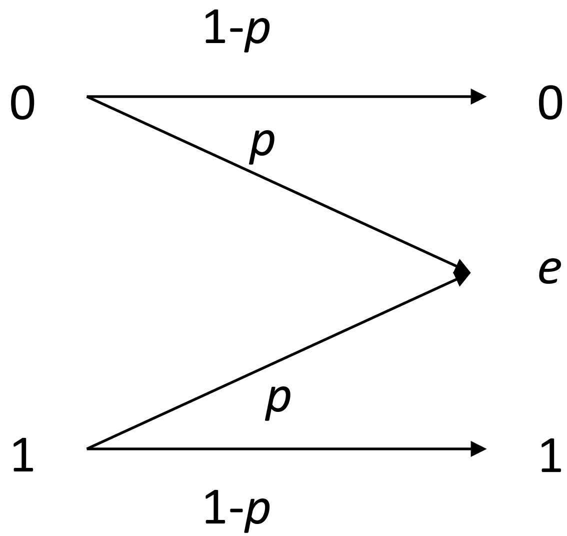
$$\mathrm{C} = \sum_{i=1}^{L} q_i \mathrm{C}_i$$

$$\mathrm{C}_1 = .9412 \log .9412 + .0588 \log .0588 + \log 2 = .6773$$

$$\mathrm{C}_2 = 1 \log 1 + \log 1 = 0.0$$

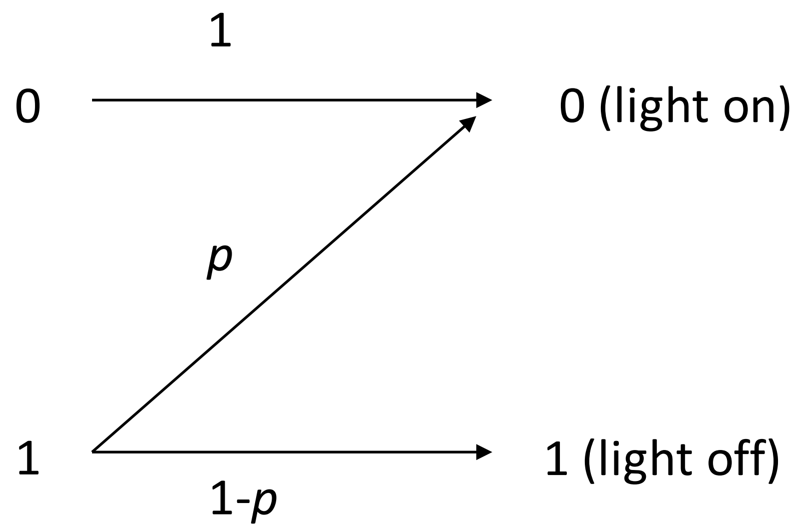
$$\mathrm{C} = .85(.6773) + .15(0.0) = .5757$$

Binary Erasure Channel



$$C = 1 - p$$

Z Channel (Optical)



$$p(x = 0) = w$$

$$p(x = 1) = 1 - w = \bar{w}$$

Z Channel (Optical)

$$\mathrm{I(X;Y)} = \sum_{k=1}^{2}\sum_{j=1}^{2} \mathrm{p}(x_k)\mathrm{p}(y_j \mid x_k)\log\left[\frac{\mathrm{p}(y_j \mid x_k)}{\mathrm{p}(y_j)}\right]$$

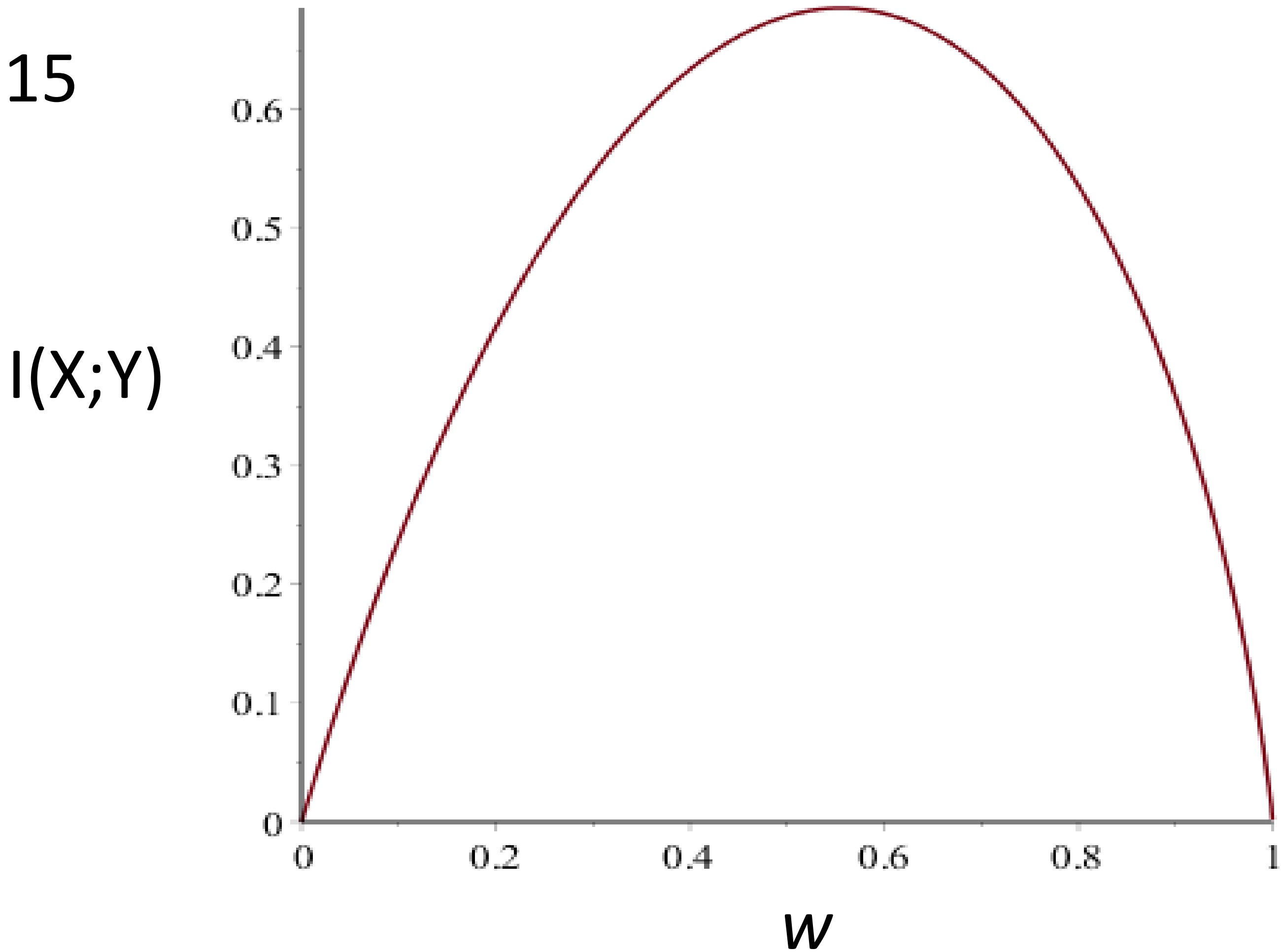
$$\mathrm{I}(x_1;\mathrm{Y}) = \log\left(\frac{1}{w + p\overline{\overline{w}}}\right)$$

$$\mathrm{I}(x_2;\mathrm{Y}) = p\log\left(\frac{p}{w + p\overline{\overline{w}}}\right) + (1-p)\log\left(\frac{1}{\overline{\overline{w}}}\right)$$

$$\mathrm{I(X;Y)} = \mathrm{w} \times \mathrm{I}(x_1;\mathrm{Y}) + \overline{\mathrm{w}} \times \mathrm{I}(x_2;\mathrm{Y})$$

Mutual Information for the Z Channel

- p = 0.15



Z Channel (Optical)

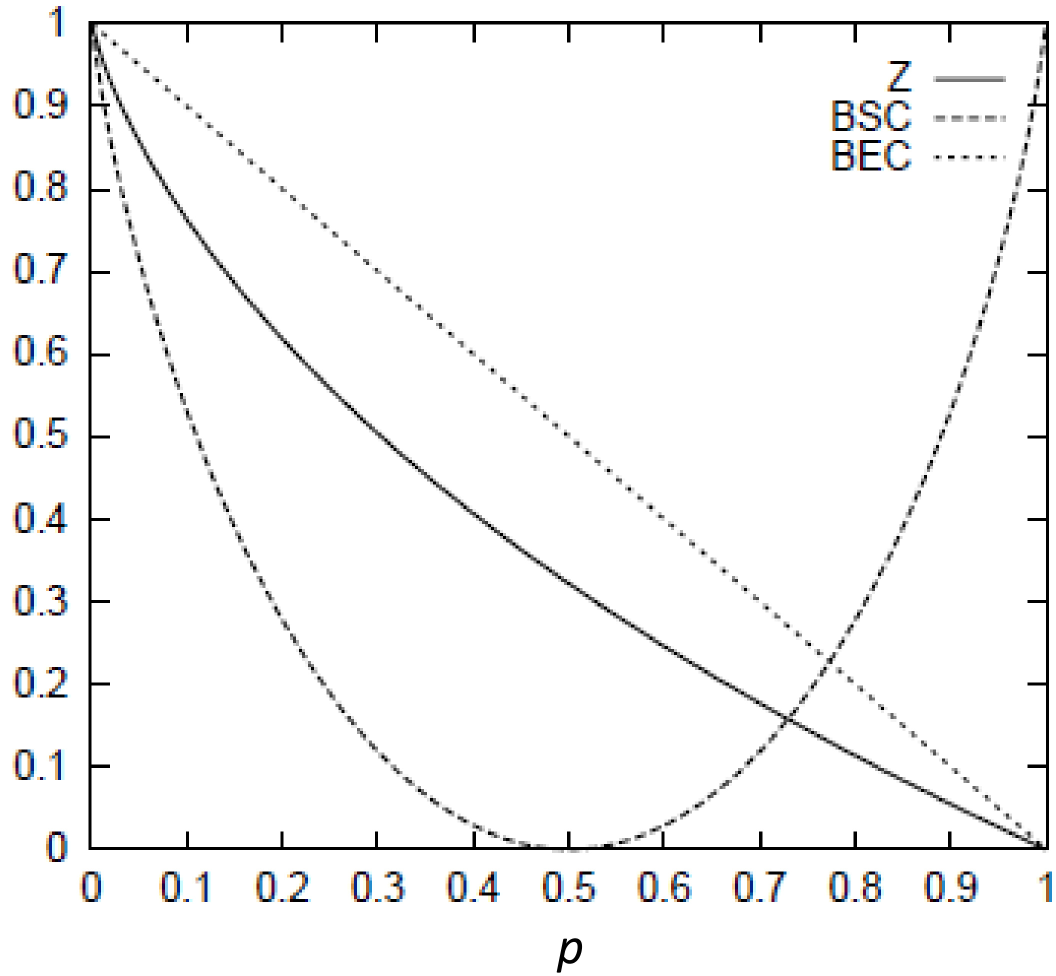
$$\mathrm{I(X;Y)} = \mathrm{w} \times \mathrm{I}(x_1;\mathrm{Y}) + \overline{\mathrm{w}} \times \mathrm{I}(x_2;\mathrm{Y})$$

$$w^* = 1 - \frac{1}{(1-p)(1+2^{\mathrm{h}(p)/(1-p)})}$$

$$\mathrm{C} = \log_2\left(1 + (1-p)p^{p/(1-p)}\right)$$

$$p = 0.15 \quad w^* = 0.555 \quad \mathrm{C} = 0.685$$

Channel Capacity for the Z, BSC and BEC



Blahut-Arimoto Algorithm

$$\mathrm{I}(X;Y)=\sum_{k=1}^{K}\mathrm{p}(x_k)\sum_{j=1}^{J}\mathrm{p}(y_j\mid x_k)\log\left[\frac{\mathrm{p}(y_j\mid x_k)}{\sum_{l=1}^{K}p(x_l)p(y_j\mid x_l)}\right]$$

- An analytic solution for the capacity can be very difficult to obtain
- The alternative is a numerical solution
  - Arimoto Jan. 1972
  - Blahut Jul. 1972
- Exploits the fact that I(X;Y) is a concave function of p(x_k)

Blahut-Arimoto Algorithm

$$c_k = \exp\left[\sum_{j=1}^{J} p(y_j|x_k) \ln\left(\frac{p(y_j|x_k)}{\sum_{l=1}^{K} p(x_l)\; p(y_j|x_l)}\right)\right] \qquad \text{for } k = 1, \ldots, K$$

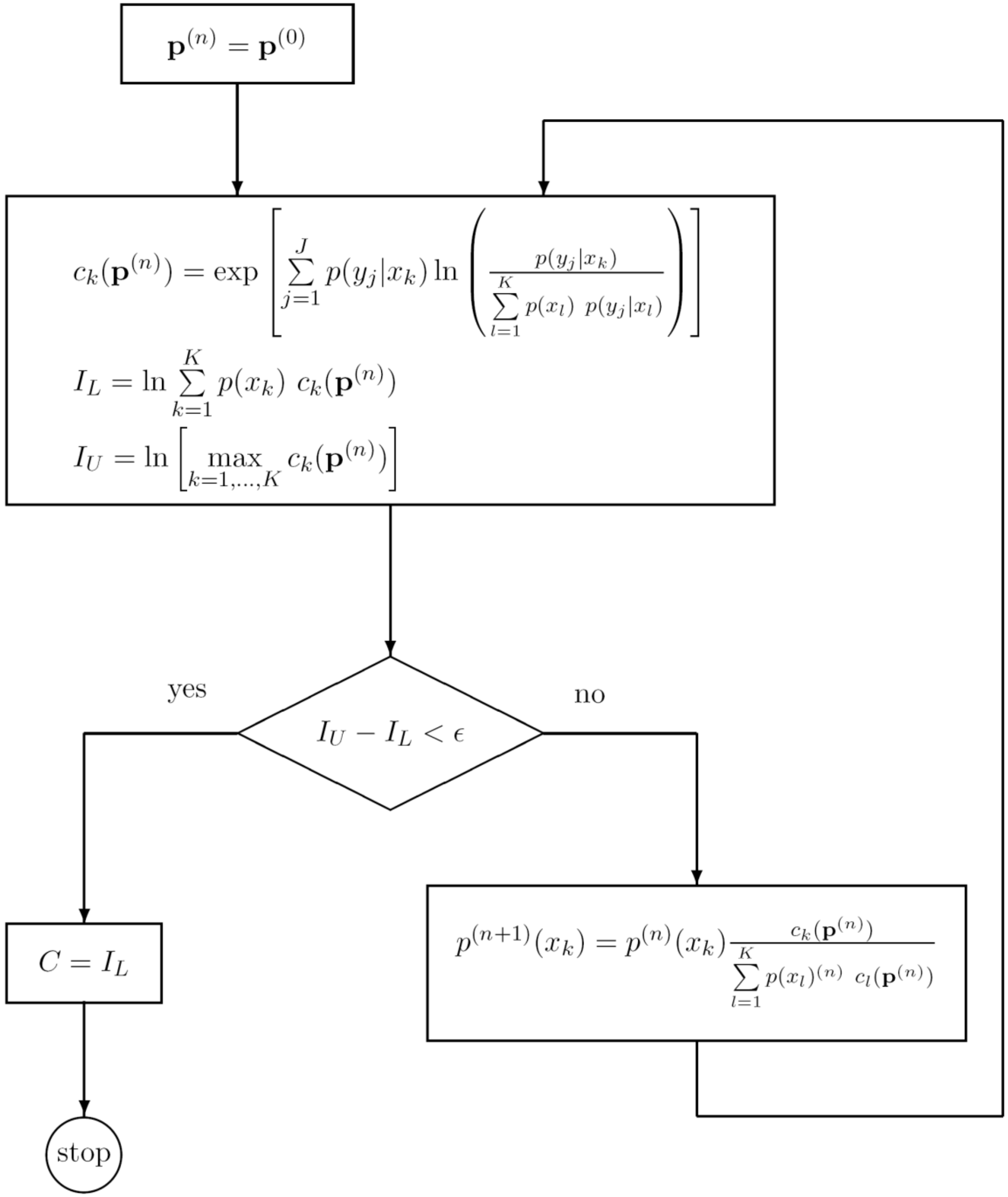
$$I_L = \ln \sum_{k=1}^{K} p(x_k)\; c_k$$

$$I_U = \ln\left(\max_{k=1,\ldots,K} c_k\right)$$

Blahut-Arimoto Algorithm

- Update the probabilities

$$p^{(n+1)}(x_k) = \frac{p^{(n)}(x_k)\ c_k}{\sum_{l=1}^{K} p(x_l)^{(n)}\ c_l}$$



$\mathbf{p}^{(n)} = \mathbf{p}^{(0)}$
$c_k(\mathbf{p}^{(n)}) = \exp\left[\sum_{j=1}^{J} p(y_j|x_k) \ln\left(\frac{p(y_j|x_k)}{\sum_{l=1}^{K} p(x_l)\ p(y_j|x_l)}\right)\right]$
$I_L = \ln \sum_{k=1}^{K} p(x_k)\ c_k(\mathbf{p}^{(n)})$
$I_U = \ln\left[\max_{k=1,\ldots,K} c_k(\mathbf{p}^{(n)})\right]$
yes
no
$I_U - I_L < \epsilon$
$C = I_L$
$p^{(n+1)}(x_k) = p^{(n)}(x_k) \frac{c_k(\mathbf{p}^{(n)})}{\sum_{l=1}^{K} p(x_l)^{(n)}\ c_l(\mathbf{p}^{(n)})}$
stop

Symmetric Channel Example

$$\mathbf{P}_1 = \begin{pmatrix} 0.4000 & 0.3000 & 0.2000 & 0.1000 \\ 0.1000 & 0.4000 & 0.3000 & 0.2000 \\ 0.3000 & 0.2000 & 0.1000 & 0.4000 \\ 0.2000 & 0.1000 & 0.4000 & 0.3000 \end{pmatrix}$$

n	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	I_U	I_L	ϵ
1	0.2500	0.2500	0.2500	0.2500	0.1064	0.1064	0.0000

n	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	I_U	I_L	ϵ
1	0.1000	0.6000	0.2000	0.1000	0.1953	0.0847	0.1106
2	0.1073	0.5663	0.2155	0.1126	0.1834	0.0885	0.0949
3	0.1141	0.5348	0.2287	0.1249	0.1735	0.0916	0.0819
4	0.1204	0.5061	0.2394	0.1369	0.1650	0.0942	0.0708
5	0.1264	0.4802	0.2480	0.1484	0.1576	0.0963	0.0613
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
10	0.1524	0.3867	0.2668	0.1963	0.1326	0.1024	0.0302
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
20	0.1912	0.3037	0.2604	0.2448	0.1219	0.1057	0.0162
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
40	0.2320	0.2622	0.2476	0.2578	0.1110	0.1064	0.0046
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
80	0.2482	0.2515	0.2486	0.2516	0.1068	0.1064	0.0004
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
202	0.2500	0.2500	0.2500	0.2500	0.1064	0.1064	0.0000

Non-Symmetric Channel Example

$$\mathbf{P}_2 = \begin{pmatrix} 0.1000 & 0.2500 & 0.2000 & 0.1000 \\ 0.1000 & 0.2500 & 0.6000 & 0.2000 \\ \color{red}{0.7000} & 0.2500 & 0.1000 & 0.2000 \\ 0.1000 & 0.2500 & 0.1000 & 0.5000 \end{pmatrix}$$

n	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	I_U	I_L	ϵ
1	0.2500	0.2500	0.2500	0.2500	0.4498	0.2336	0.2162
2	0.3103	0.1861	0.2545	0.2228	0.4098	0.2504	0.1595
3	0.3640	0.1428	0.2653	0.2036	0.3592	0.2594	0.0998
4	0.4022	0.1118	0.2804	0.1899	0.3289	0.2647	0.0642
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
8	0.4522	0.0450	0.3389	0.1639	0.2988	0.2763	0.0225
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
16	0.4629	0.0076	0.3732	0.1565	0.2848	0.2830	0.0018
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
32	0.4641	0.0002	0.3769	0.1588	0.2846	0.2844	0.0003
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
64	0.4640	0.0000	0.3768	0.1592	0.2844	0.2844	0.0000
65	0.4640	0.0000	0.3768	0.1592	0.2844	0.2844	0.0000

$$P_1 = \begin{bmatrix} .98 & .05 \\ .02 & .95 \end{bmatrix} \quad P_2 = \begin{bmatrix} .80 & .05 \\ .20 & .95 \end{bmatrix}$$

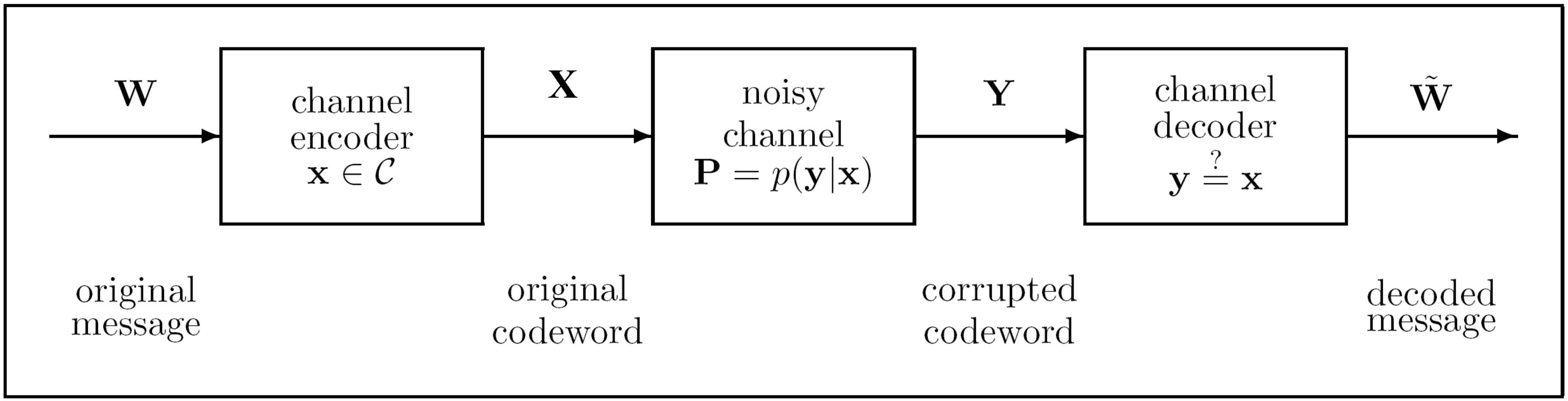
$$P_3 = \begin{bmatrix} .80 & .10 \\ .20 & .90 \end{bmatrix} \quad P_4 = \begin{bmatrix} .60 & .01 \\ .40 & .99 \end{bmatrix}$$

$$P_5 = \begin{bmatrix} .80 & .30 \\ .20 & .70 \end{bmatrix}$$

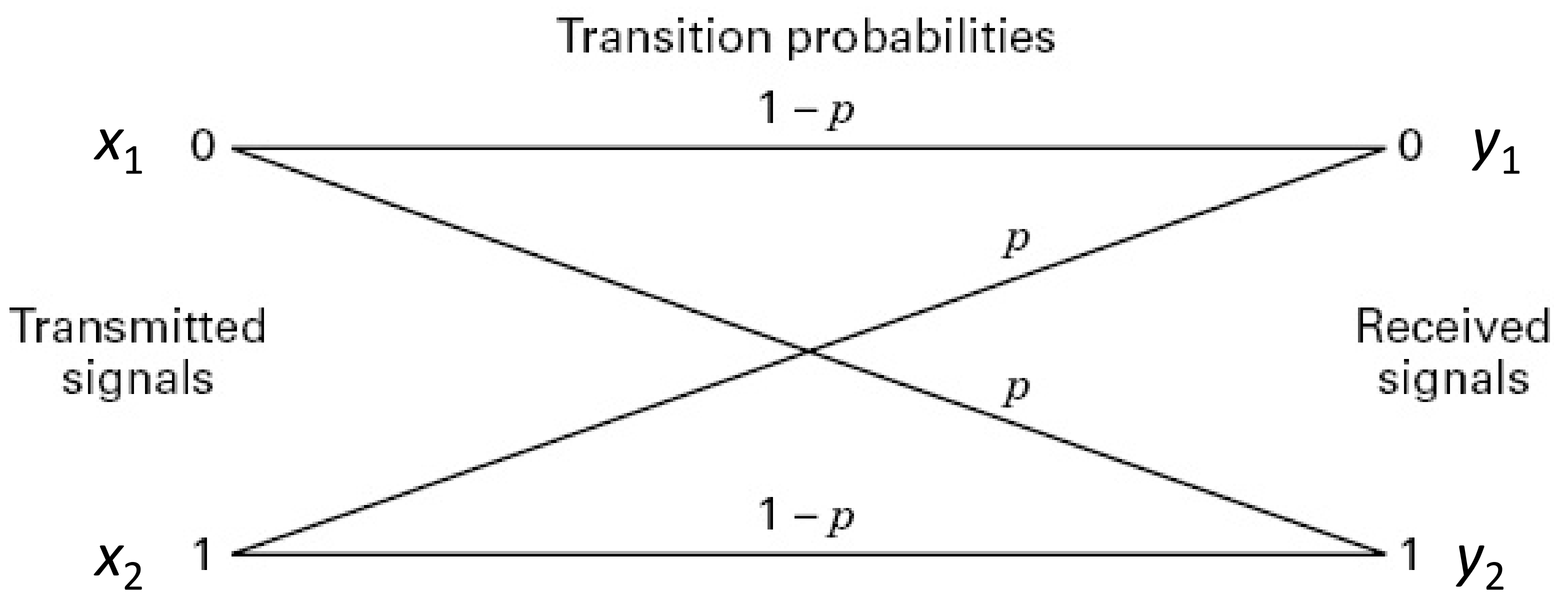
	C	$\mathbf{p}^*$
P_1	.7859	(.5129 .4871)
P_2	.4813	(.4676 .5324)
P_3	.3976	(.4824 .5176)
P_4	.3688	(.4238 .5762)
P_5	.1912	(.5100 .4900)

	C	$\mathbf{p}^*$	$I(X;Y)_{\mathrm{u}}$
P_1	.7859	(.5129 .4871)	.7854
P_2	.4813	(.4676 .5324)	.4796
P_3	.3976	(.4824 .5176)	.3973
P_4	.3688	(.4238 .5762)	.3615
P_5	.1912	(.5100 .4900)	.1912

Communication over Noisy Channels



Binary Symmetric Channel



channel matrix

$$\begin{bmatrix} \overline{p} & p \\ p & \overline{p} \end{bmatrix} \qquad \overline{p} = 1 - p$$

Binary Symmetric Channel

- Consider a block of N = 1000 bits
  - if p = 0, 1000 bits are received correctly
  - if p = 0.01, 990 bits are received correctly
  - if p = 0.5, 500 bits are received correctly
- When $p > 0$, we do not know which bits are in error
  - if p = 0.01, C = .919 bit
  - if p = 0.5, C = 0 bit

Triple Repetition Code

- $N = 3$

message w	codeword c
0	000
1	111

Binary Symmetric Channel Errors

- If N bits are transmitted, the probability of an m bit error pattern is

$$p^m \left(1-p\right)^{N-m}$$

- The probability of exactly m errors is

$$\binom{N}{m} p^m \left(1-p\right)^{N-m}$$

- The probability of m or more errors is

$$\sum_{i=m}^{N} \binom{N}{i} p^i \left(1-p\right)^{N-i}$$

Triple Repetition Code

- $N = 3$
- The probability of 0 errors is $(1-p)^3$
- The probability of 1 error is $3\text{p}(1-p)^2$
- The probability of 2 errors is $3p^2(1-p)$
- The probability of 3 errors is p^3

Triple Repetition Code

- For $p = 0.01$
  - The probability of 0 errors is .970
  - The probability of 1 error is 2.94×10^{-2}
  - The probability of 2 errors is 2.97×10^{-4}
  - The probability of 3 errors is 10^{-6}

- If $p \ll \frac{1}{2}$

  $$\text{p(0 errors)} \gg \text{p(1 error)} \gg \text{p(2 errors)} \gg \text{p(3 errors)}$$

Triple Repetition Code – Decoding

Received Word	Codeword	Error Pattern
0 0 0	0 0 0	0 0 0
0 0 1	0 0 0	0 0 1
0 1 0	0 0 0	0 1 0
1 0 0	0 0 0	1 0 0
1 1 1	1 1 1	0 0 0
1 1 0	1 1 1	0 0 1
1 0 1	1 1 1	0 1 0
0 1 1	1 1 1	1 0 0

Triple Repetition Code

- Majority vote or nearest neighbor decoding will correct all single errors

$$000, 001, 010, 100 \rightarrow 000$$

$$111, 110, 101, 011 \rightarrow 111$$

- The probability of a decoding error is then

$$\mathrm{P_e} = 3p^2(1-p) + p^3 = 3p^2 - 2p^3 < p$$

- If p = 0.01, then $\mathrm{P_e}$ = 0.000298 and only one word in 3356 will be in error after decoding.
- A reduction by a factor of 33.

Code Rate

- After compression, the data is (almost) memoryless and uniformly distributed (equiprobable)
- Thus the entropy of the messages (codewords) is

$$\mathrm{H}(W) = \log_b M$$

- The blocklength of a codeword is N

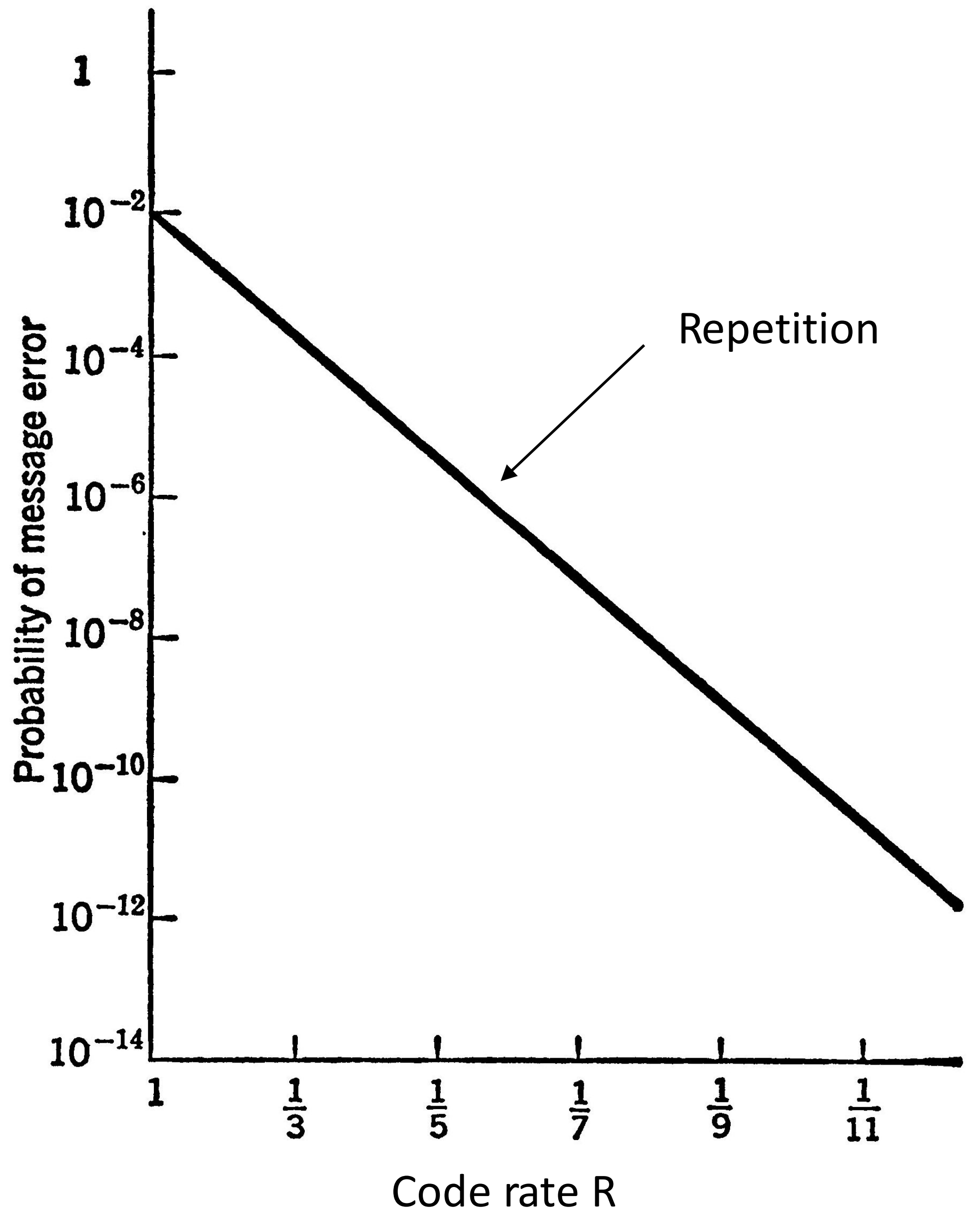
Code Rate

- The code rate is given by

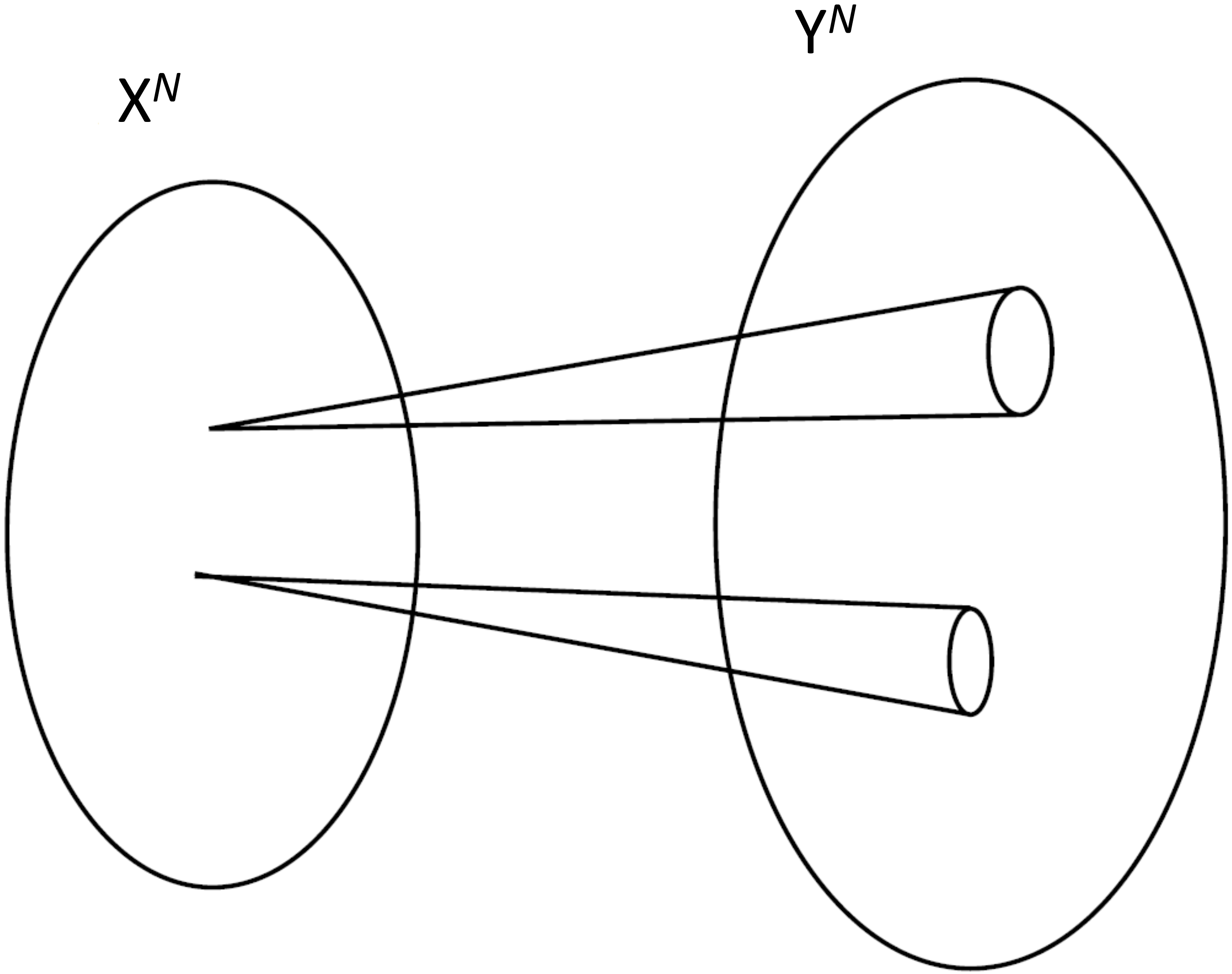
$$R = \frac{\log_2 M}{N} \text{ bits per channel use}$$

- M is the number of codewords
- N is the block length
- For the triple repetition code

$$R = \frac{\log_2 2}{3} = \frac{1}{3}$$



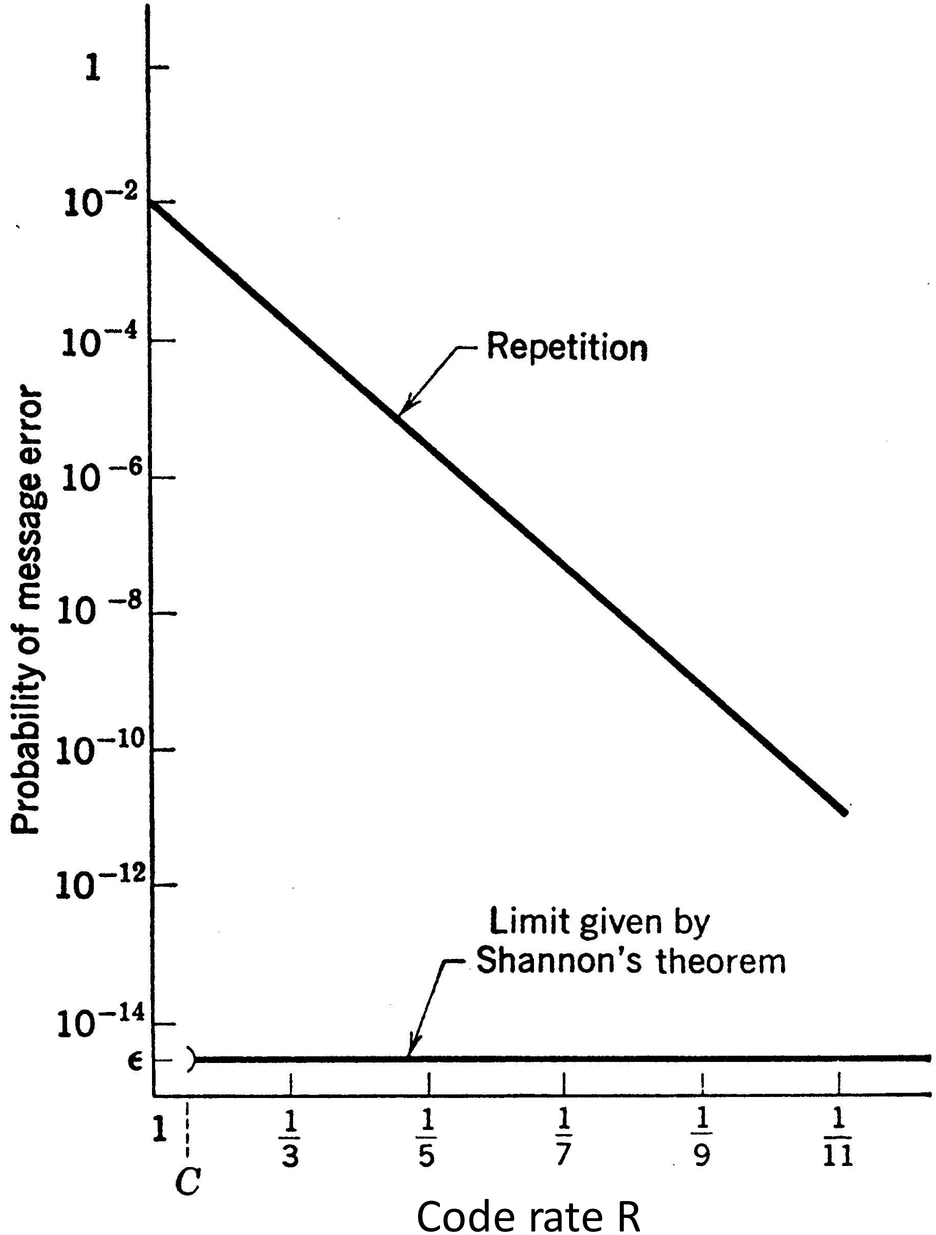
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
Probability of message error
Repetition
1
$\frac{1}{3}$
$\frac{1}{5}$
$\frac{1}{7}$
$\frac{1}{9}$
$\frac{1}{11}$
Code rate R



Y^N
X^N

Shannon's Noisy Coding Theorem

For any $\varepsilon > 0$ and for any rate R less than the channel capacity C, there is an encoding and decoding scheme that can be used to ensure that the probability of decoding error P_e is less than ε for a sufficiently large block length N.



Probability of message error
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
ϵ
Repetition
Limit given by
Shannon's theorem
1
C
$\frac{1}{3}$
$\frac{1}{5}$
$\frac{1}{7}$
$\frac{1}{9}$
$\frac{1}{11}$
Code rate R

Error Correction Coding $N = 3$

- R = 1/3 $M = 2$

  0 → 000

  1 → 111

- R = 1 $M = 8$

  000 → 000   001 → 001   010 → 010   011 → 011

  111 → 111   110 → 110   101 → 101   100 → 100

- Another choice R = 2/3 $M = 4$

  00 → 000   01 → 011

  10 → 101   11 → 110

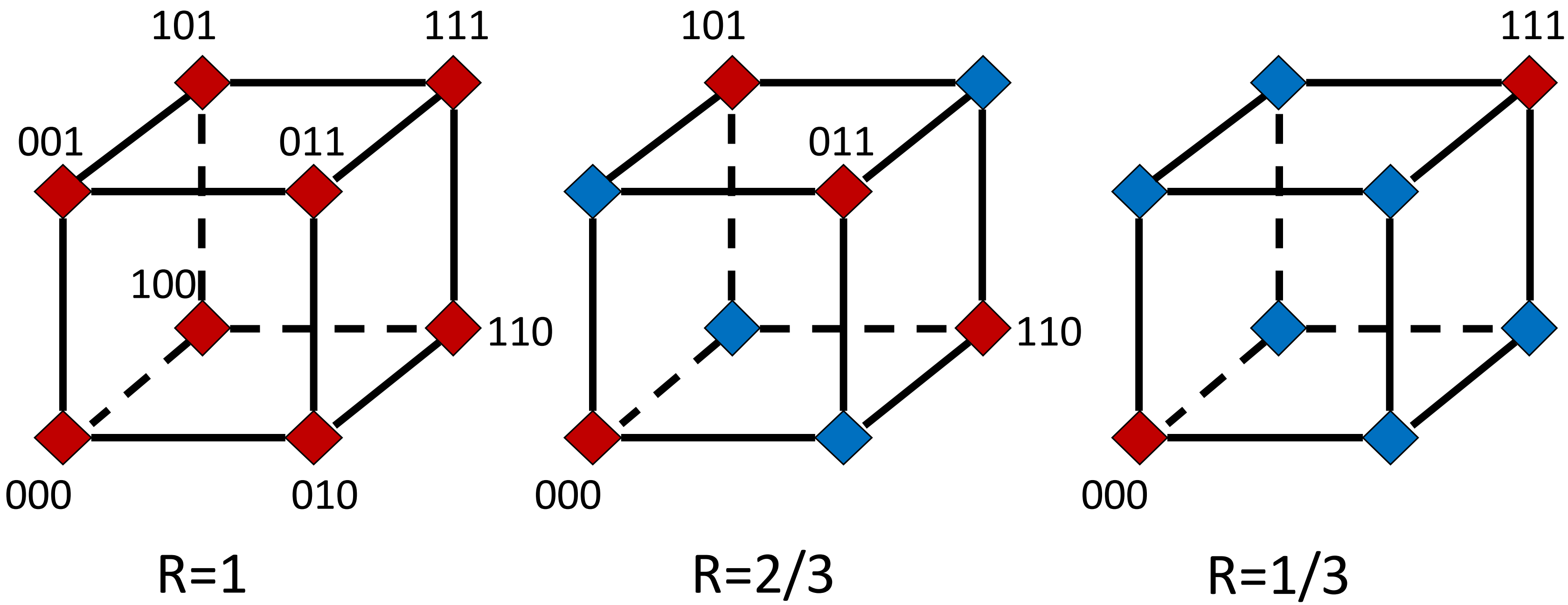
Error Correction Coding $N = 3$

- BSC $p = 0.01$
- M is the number of codewords

Code Rate R	P_e	$M=2^{NR}$
1	0.0297	8
2/3	0.0199	4
1/3	2.98×10^{-4}	2

- Tradeoff between code rate and error rate

Codes for N=3


101
111
001
011
100
110
000
010
101
011
110
000
111
000
R=1
R=2/3
R=1/3

Error Correction Coding $N = 5$

- BSC $p = 0.01$

Code Rate R	P_e	$M=2^{NR}$
1	0.0490	32
4/5	0.0394	16
3/5	0.0297	8
2/5	9.80×10^{-4}	4
1/5	9.85×10^{-6}	2

- Tradeoff between code rate and error rate

Error Correction Coding $N = 7$

- BSC $p = 0.01$ $N = 7$

Code Rate R	P_e	$M=2^{NR}$
1	0.0679	128
6/7	0.0585	64
5/7	0.0490	32
4/7	2.03×10^{-3}	16
3/7	1.46×10^{-3}	8
2/7	9.80×10^{-4}	4
1/7	3.40×10^{-7}	2

- Tradeoff between code rate and error rate

Best Codes Comparison

- BSC $p = 0.01$ R = 2/3 $M = 2^{NR}$

N	P_e	$\log_2 M$
3	1.99×10^{-2}	2
12	6.17×10^{-3}	8
30	3.32×10^{-3}	20
51	1.72×10^{-3}	34
81	1.36×10^{-3}	54

- For fixed R, P_e can be decreased by increasing N

Code Matrix

$$\mathcal{C} = \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_m \\ \vdots \\ \mathbf{c}_M \end{bmatrix} = \begin{bmatrix} c_{1,1} & \cdots & c_{1,n} & \cdots & c_{1,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{m,1} & \cdots & c_{m,n} & \cdots & c_{m,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{M,1} & \cdots & c_{M,n} & \cdots & c_{M,N} \end{bmatrix}$$

Binary Codes

- For given values of M and N, there are

  $$2^{MN}$$

  possible binary codes.
- Of these, some will be bad, some will be best (optimal), and some will be good, in terms of P_e
- An average code will be good.

Theorem *(Shannon's channel coding theorem):*

Let C be the information transfer capacity of a memoryless channel defined by its transition probabilities matrix $\mathbf{P} = \{p(\mathbf{y}|\mathbf{x})\}$. If the code rate $R < C$, then there *exists* a channel code $\mathcal{C}$ of size M and blocklength N, such that the probability of decoding error P_e is *upperbounded* by an arbitrarily small number ϵ;

$$\boxed{P_e \leq \epsilon}$$

provided that the blocklength N is sufficiently large (i.e., $N \geq N_0$).

Channel Capacity

- To prove that information can be transmitted reliably over a noisy channel at rates up to the capacity, Shannon used a number of new concepts
  - Allowing an arbitrarily small but nonzero probability of error
  - Using long codewords
  - Calculating the average probability of error over a random choice of codes to show that at least one good code exists

Channel Coding Theorem

- Random coding used in the proof
- Joint typicality used as the decoding rule
- Shows that good codes exist which provide an arbitrarily small probability of error
- Does not provide an explicit way of constructing good codes
- If a long code (large N) is generated randomly, the code is likely to be good but is difficult to decode

Theorem *(Converse of the channel coding theorem):*

Let a memoryless channel with capacity C be used to transmit codewords of blocklength N and input information R. Then the error decoding probability P_e satisfies the following inequality:

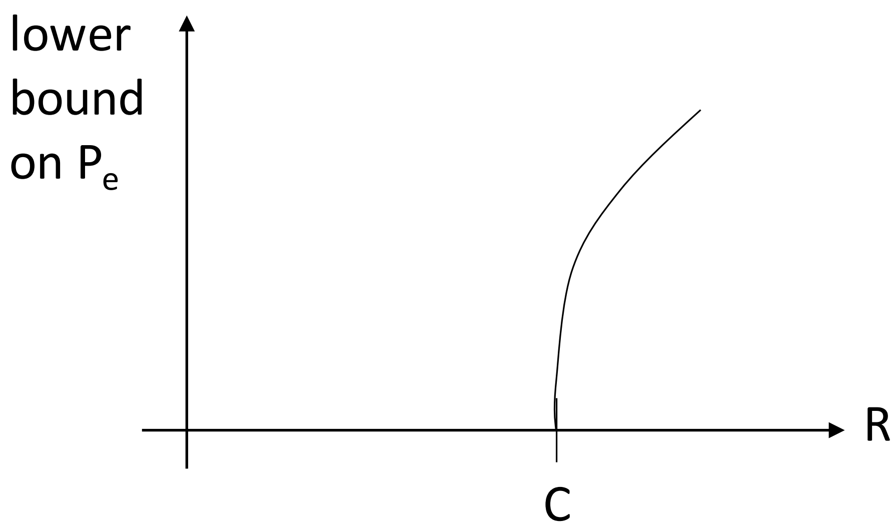
$$\boxed{P_e(N,R) \geq 1 - \frac{C}{R} - \frac{1}{NR}}$$

If the rate $R > C$, then the error decoding probability P_e is bounded away from zero.

Channel Capacity: Weak Converse

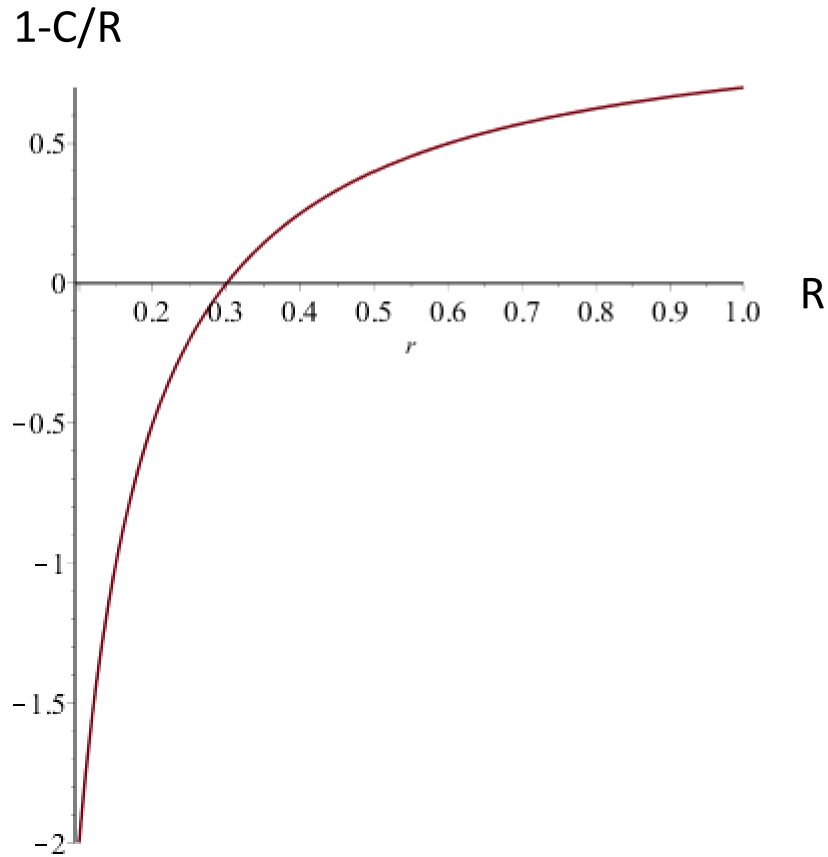
$$P_e(N, R) \geq 1 - \frac{C}{R} - \frac{1}{NR}$$

For R > C, the decoding error probability is bounded away from 0



Channel Capacity: Weak Converse

- C = 0.3



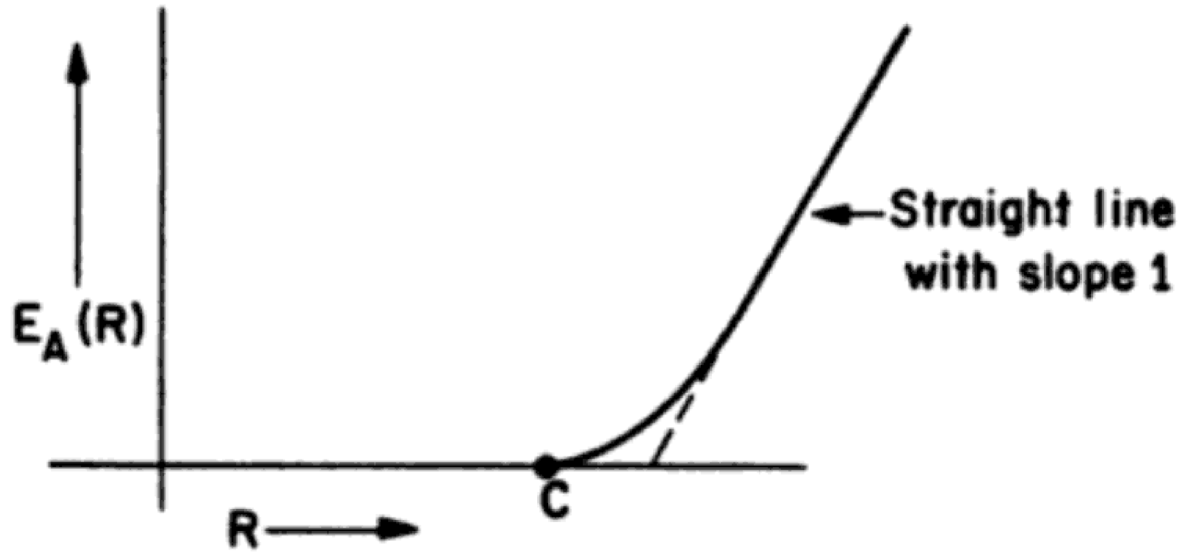
Channel Capacity: Strong Converse

- For rates above capacity (R > C)

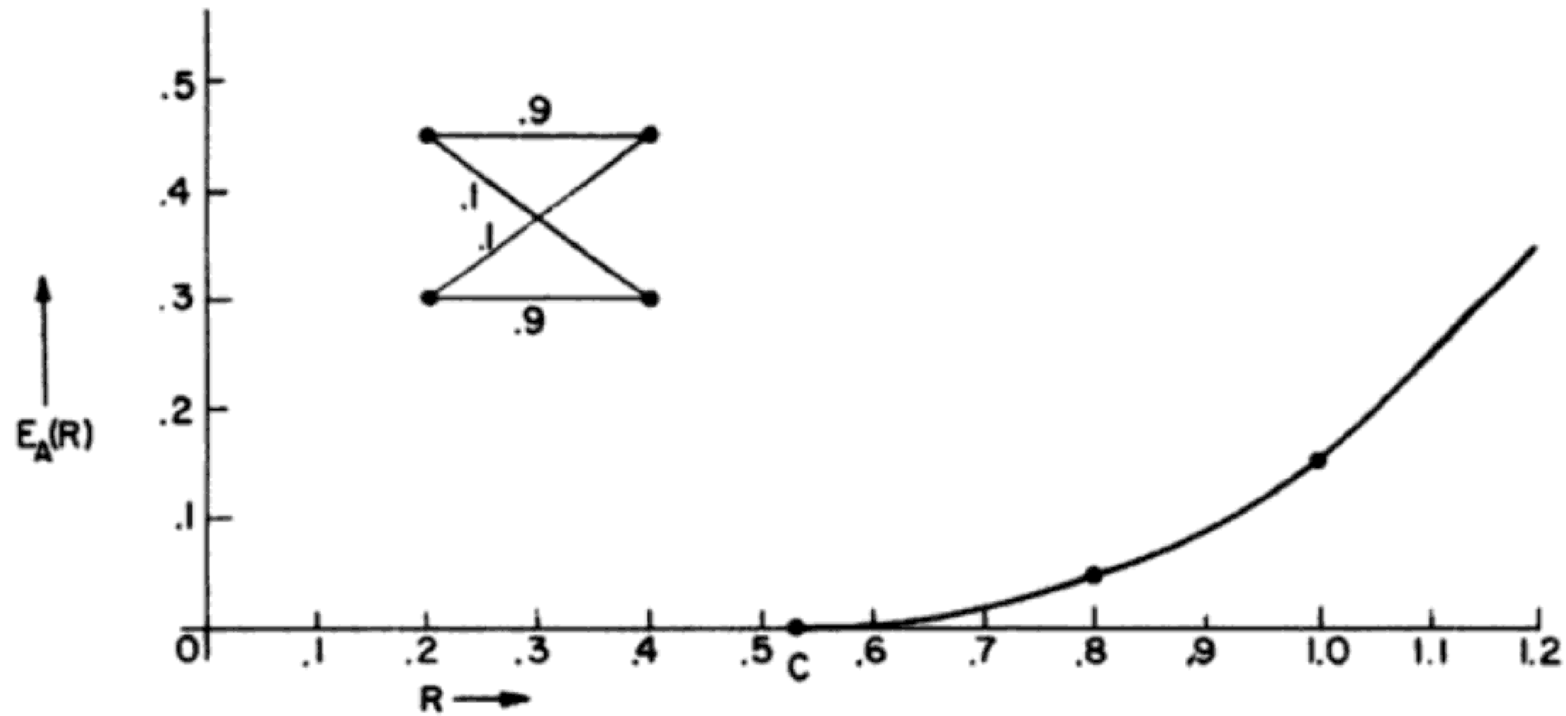
$$P_e(N,\mathrm{R}) \geq 1 - 2^{-N\mathrm{E_A}(\mathrm{R})}$$

- where $\mathrm{E_A}(\mathrm{R})$ is Arimoto's error exponent and $\mathrm{E_A}(\mathrm{R}) > 0$

Arimoto's Error Exponent $E_A(R)$



$E_A(R)$ for a BSC with p=0.1



- The capacity is a very clear dividing point
- At rates below capacity, $P_e \to 0$ exponentially as $N \to \infty$
- At rates above capacity, $P_e \to 1$ exponentially as $N \to \infty$