ELEC 519A
Selected Topics in Digital Communications: Information Theory

Channel Capacity and Coding

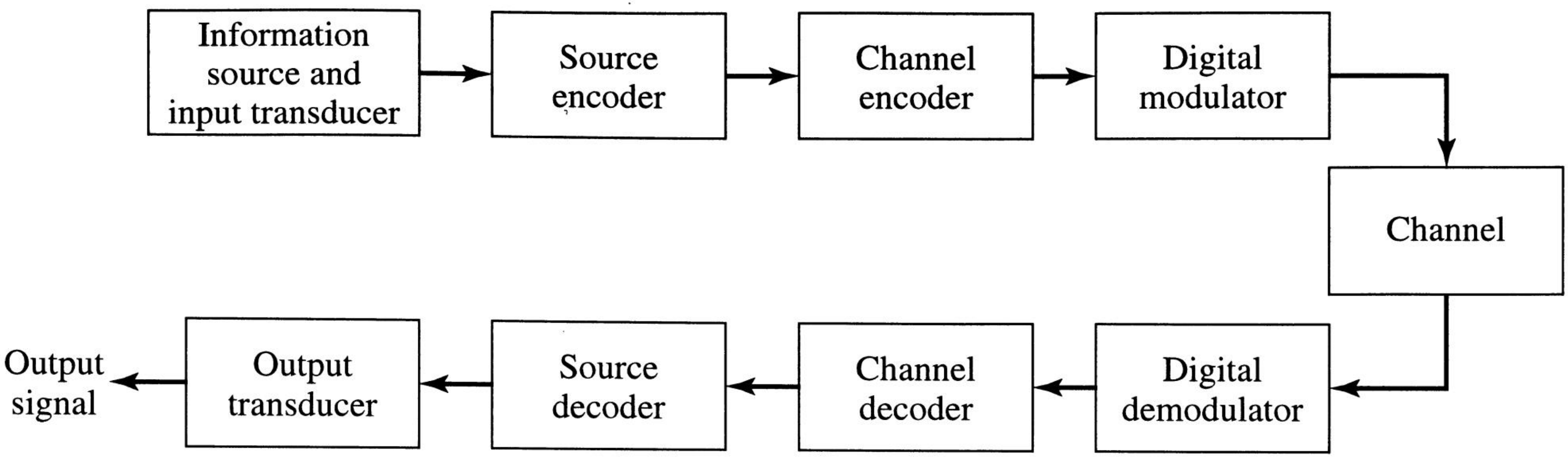


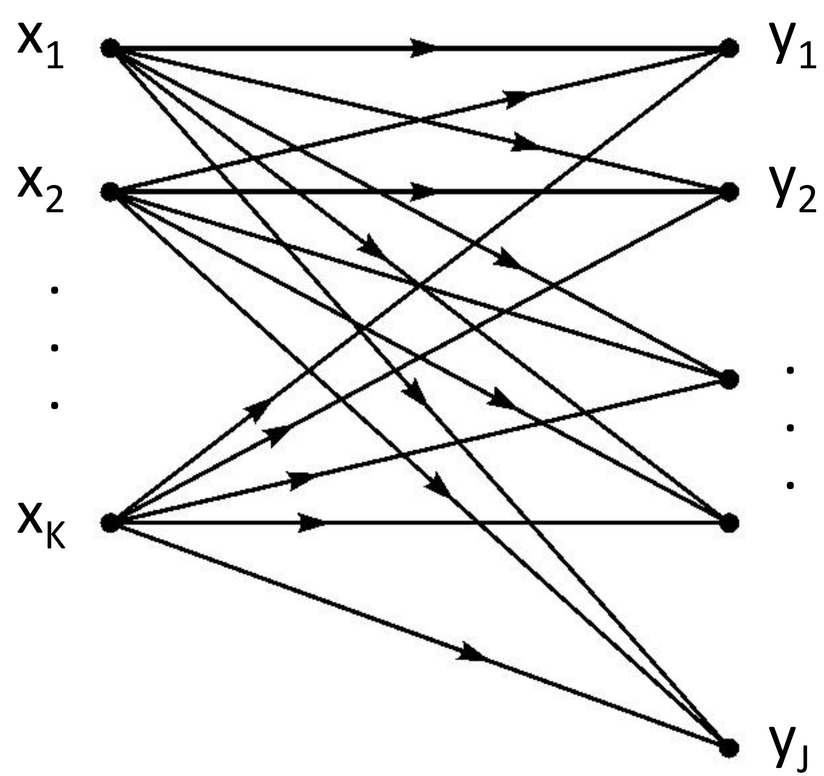
Figure 1.2 Basic elements of a digital communication system.

Channel

- ``The channel is merely the medium used to the signal from transmitter to receiver. It may be a pair of wires, a coaxial cable, a band of radio frequencies, a beam of light, etc.''
- More abstractly, the channel is

  ``that part of the communication system that is either unwilling or unable to change''

Discrete Memoryless Channel

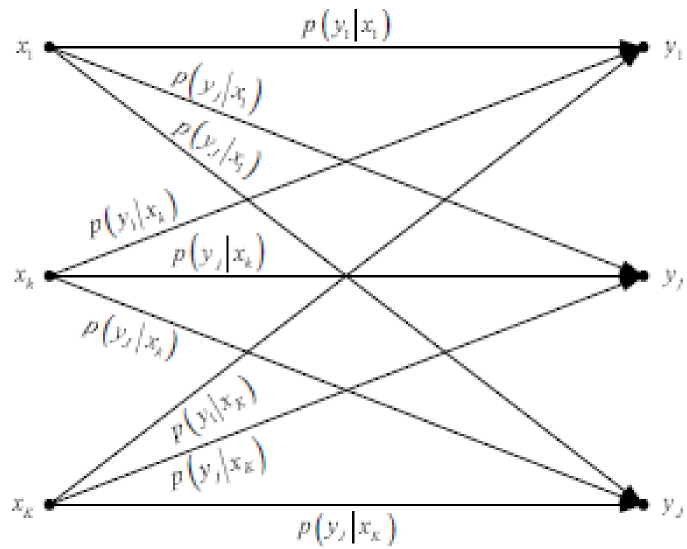


J≥K

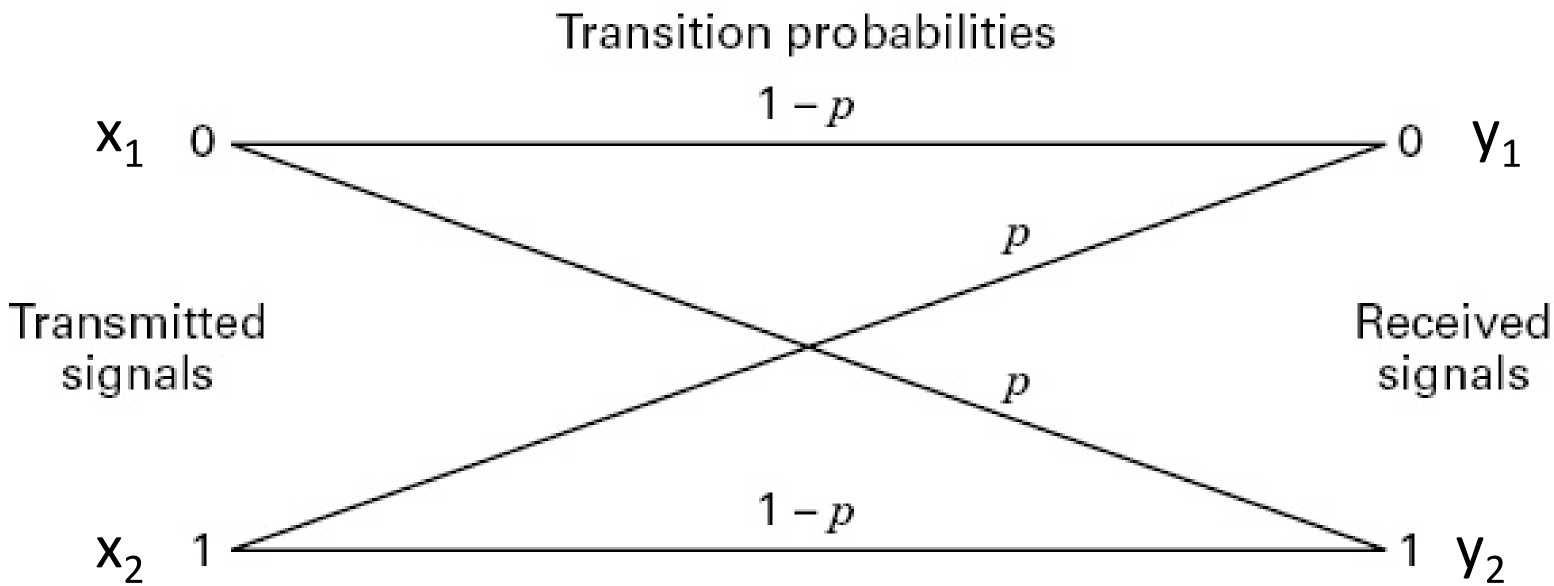
Discrete Memoryless Channel

Channel Transition Matrix

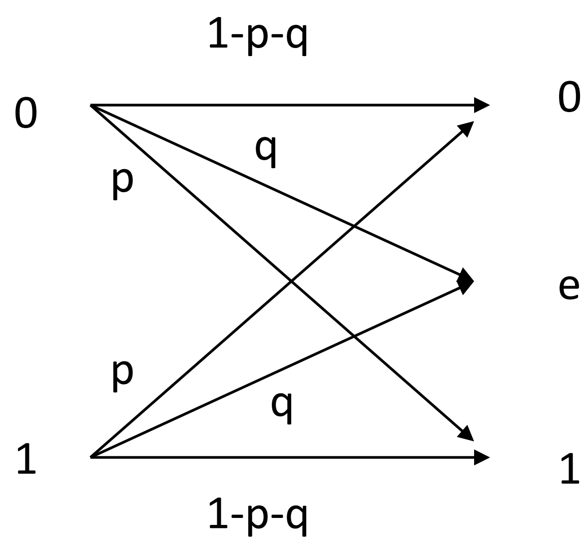
$$P = \begin{bmatrix} p(y_1 \mid x_1) & \cdots & p(y_1 \mid x_k) & \cdots & p(y_1 \mid x_K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p(y_j \mid x_1) & \cdots & p(y_j \mid x_k) & \cdots & p(y_j \mid x_K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p(y_J \mid x_1) & \cdots & p(y_J \mid x_k)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



Binary Symmetric Channel



Binary Errors with Erasure Channel

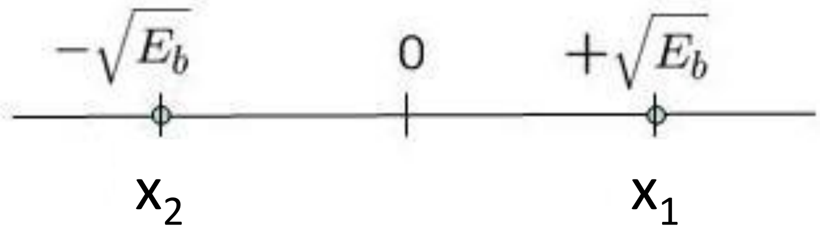


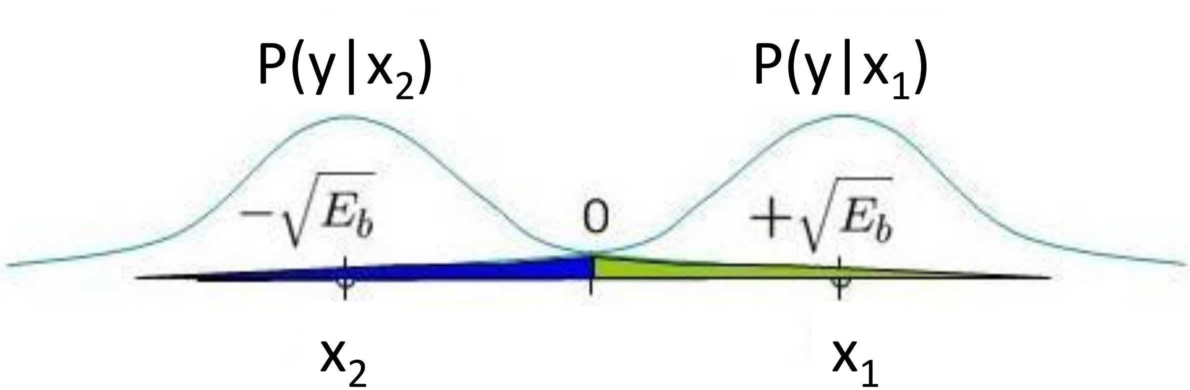
BPSK Modulation

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

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

BPSK Modulation





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

Mutual Information for a BSC

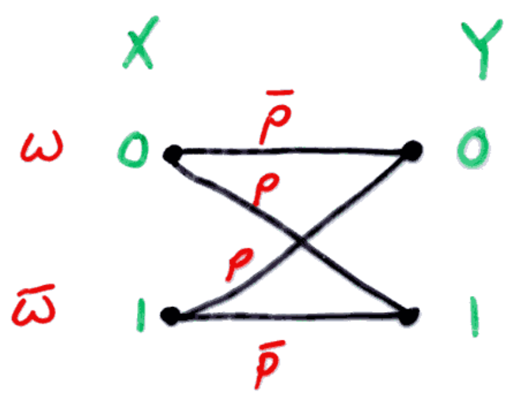
$X \to \boxed{BSC} \to Y$

crossover probability p

$\bar{p} = 1 - p$

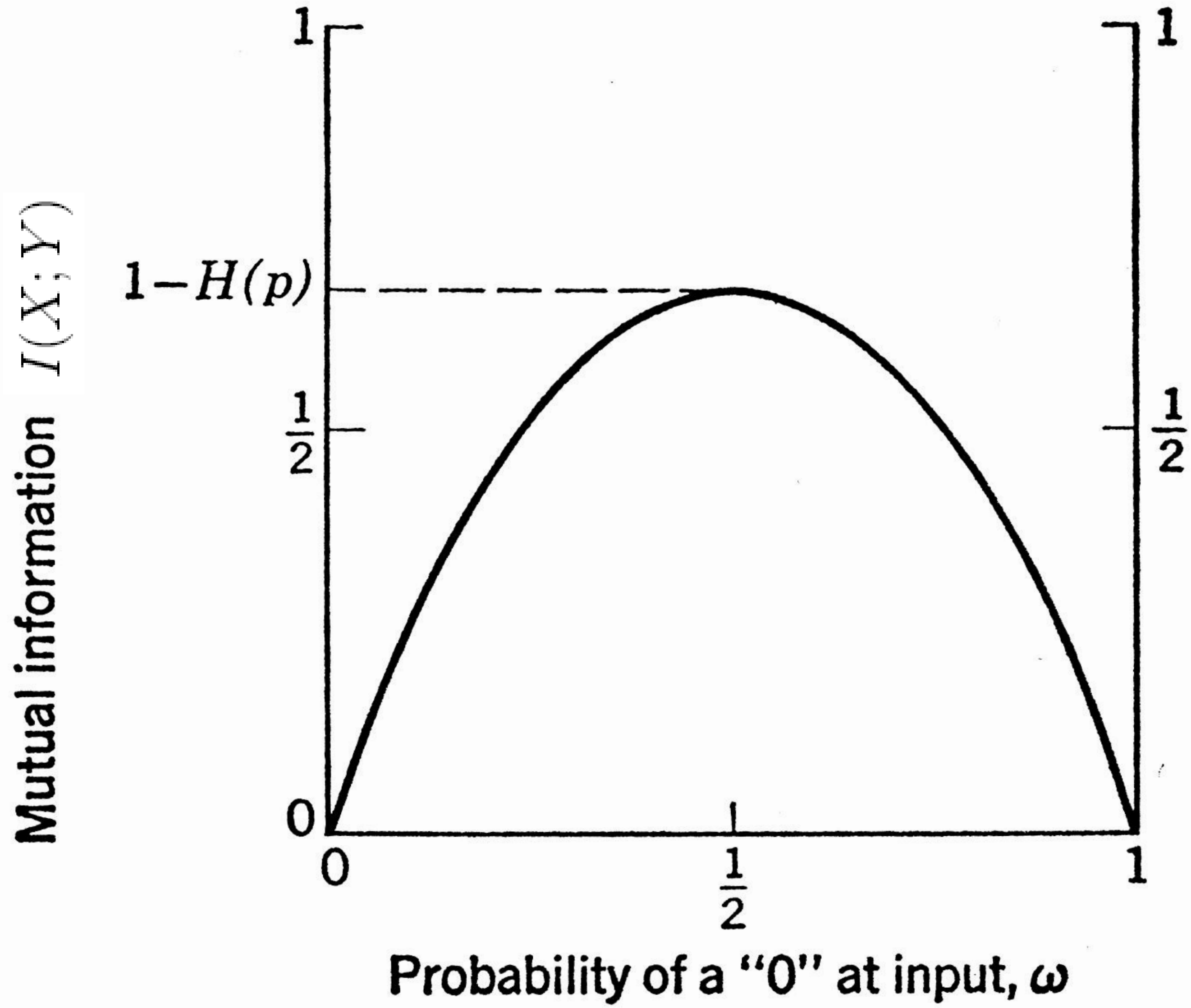
channel matrix

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



$p(X=0) = \omega$

$p(X=1) = 1 - \omega = \bar{\omega}$


1
1
1−H(p)
1/2
1/2
0
0
1/2
1
Mutual information I(X;Y)
Probability of a "0" at input, ω

Convex Combination

- A convex combination is a linear combination of points where all coefficients are non-negative and sum to 1.
- Given a finite number of points $x_1, x_2, \ldots, x_N$ in a real vector space, a convex combination of these points is given by

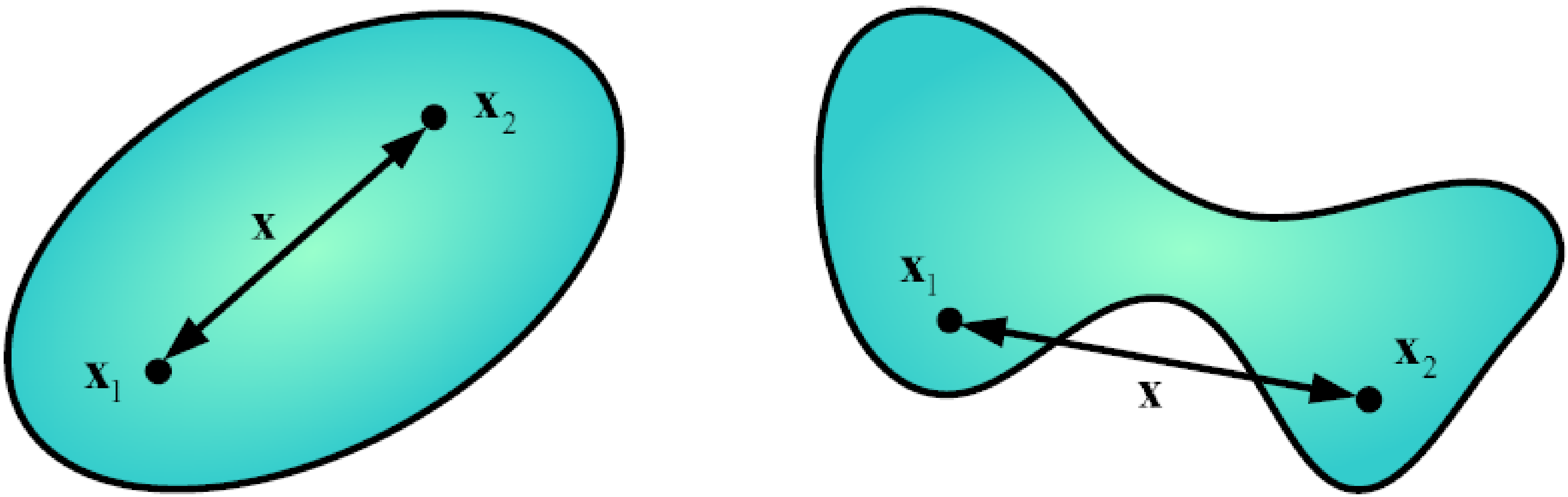
  $$\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_N x_N$$

  where the λ_i satisfy $\lambda_i \geq 0$ and $\lambda_1 + \lambda_2 + \ldots + \lambda_N = 1$.
- Every convex combination of two points lies on the line segment between the points.

Definition *(Convex set):*

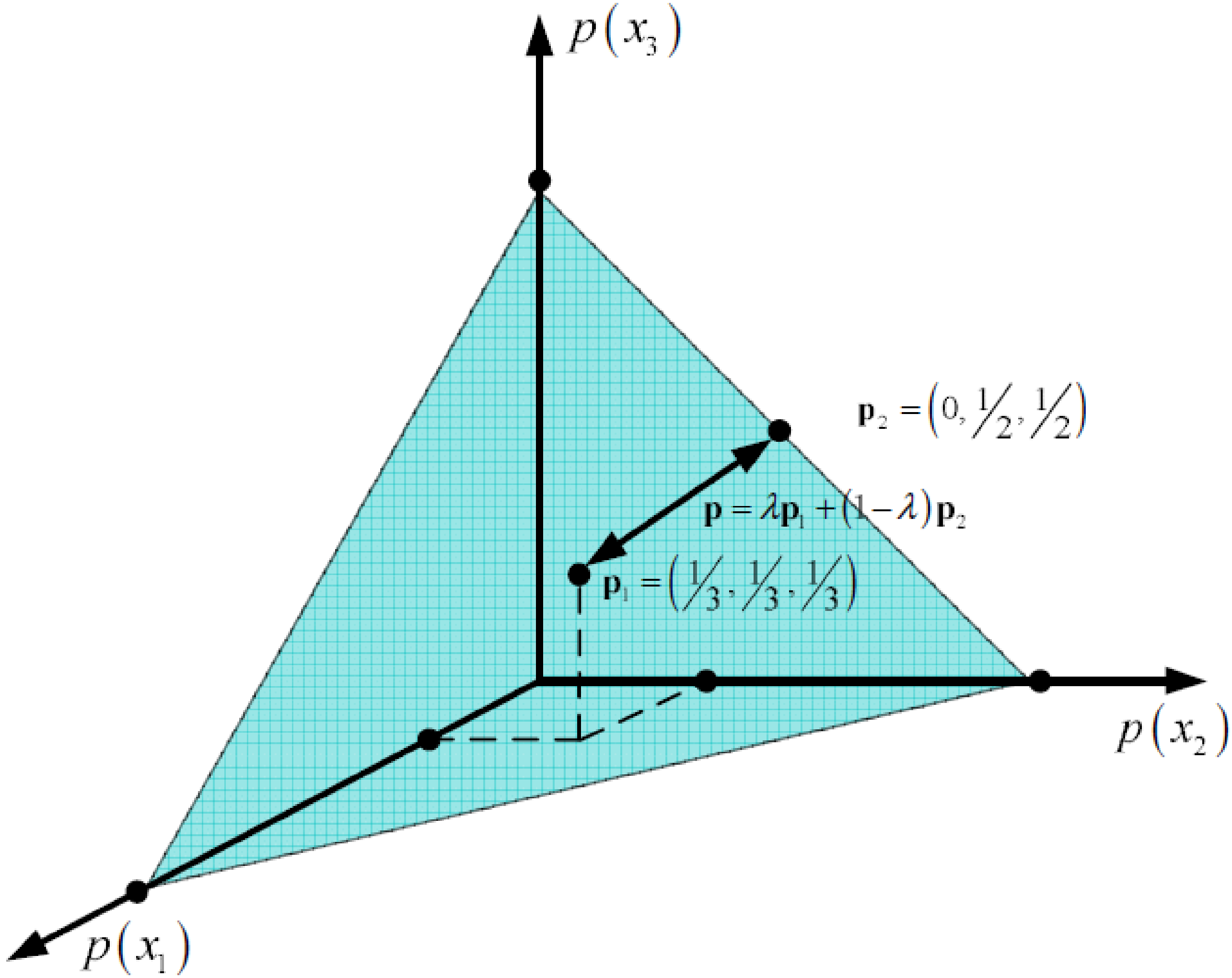
A set of points $\mathcal{S}$ is *convex* if, for any pair of points $p_1 \in \mathcal{S}$ and $p_2 \in \mathcal{S}$, any point p on the *straight line* connecting p_1 and p_2 will be also in the set $\mathcal{S}$.

Convex and Non-convex Sets



$$x = \lambda_1 x_1 + (1-\lambda_1)x_2$$

3-Dimensional Probability Distribution



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* $\cup$).

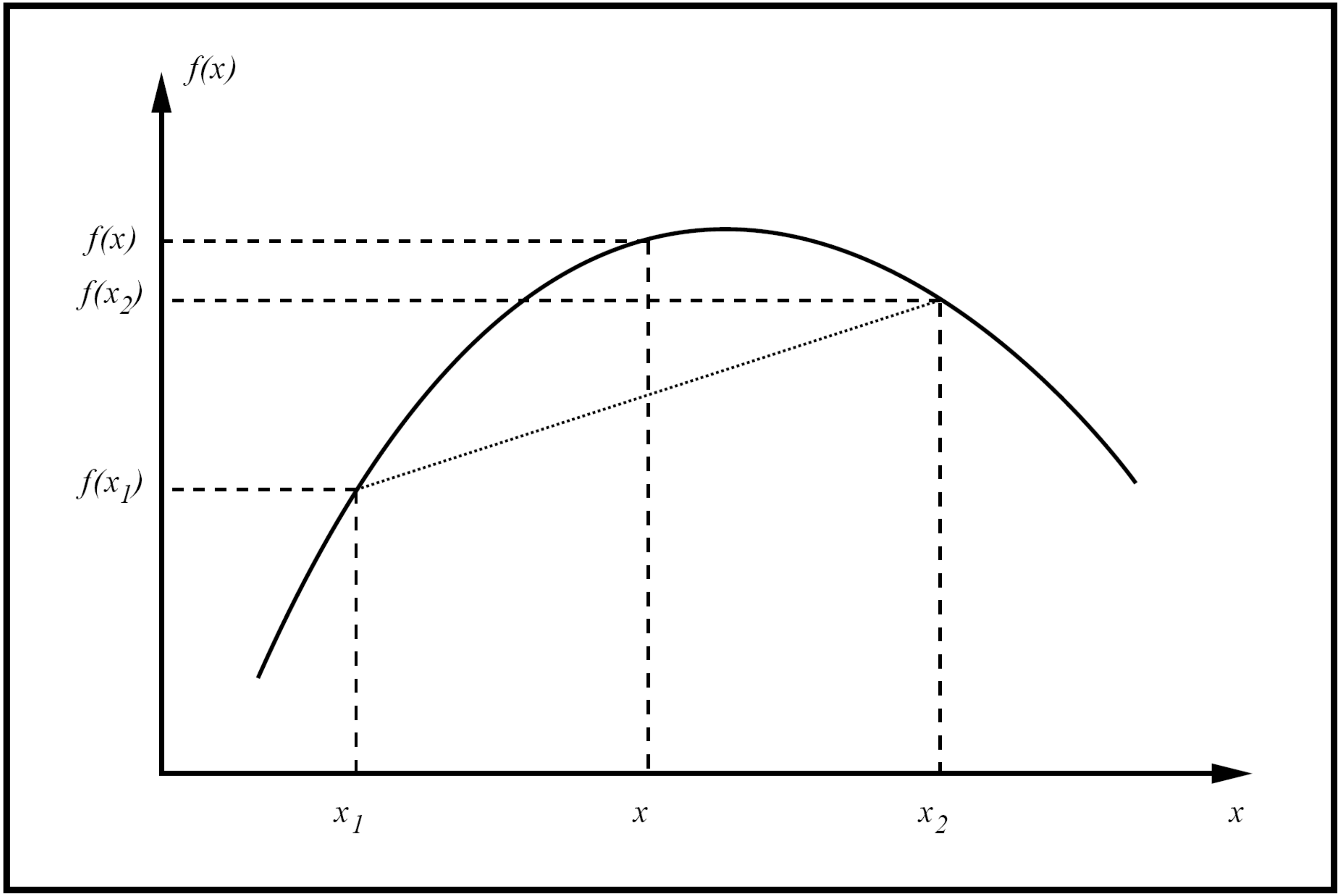


Figure 3.3: Convex $\cap$ (*convex down* or *convex "cap"*) function.

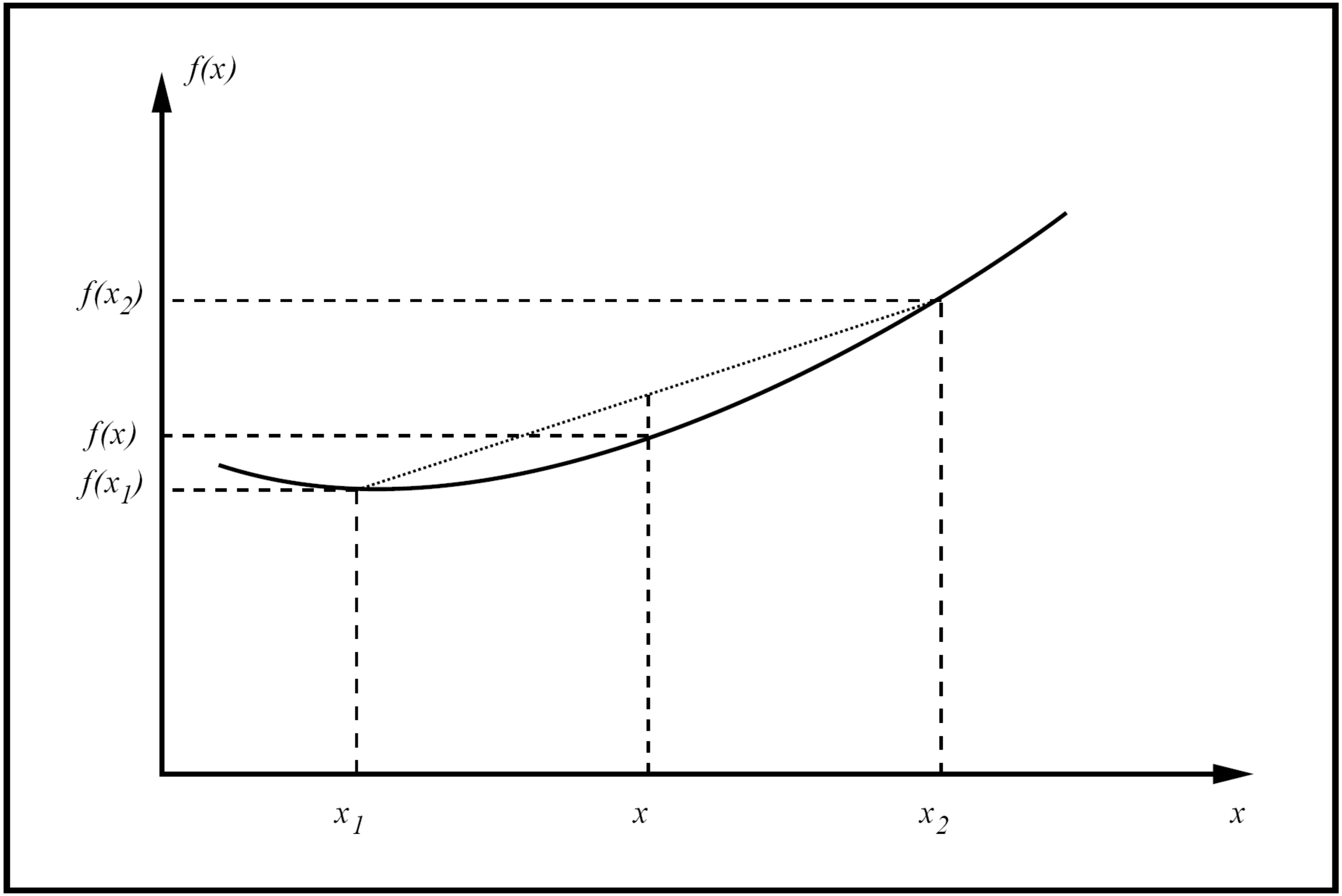


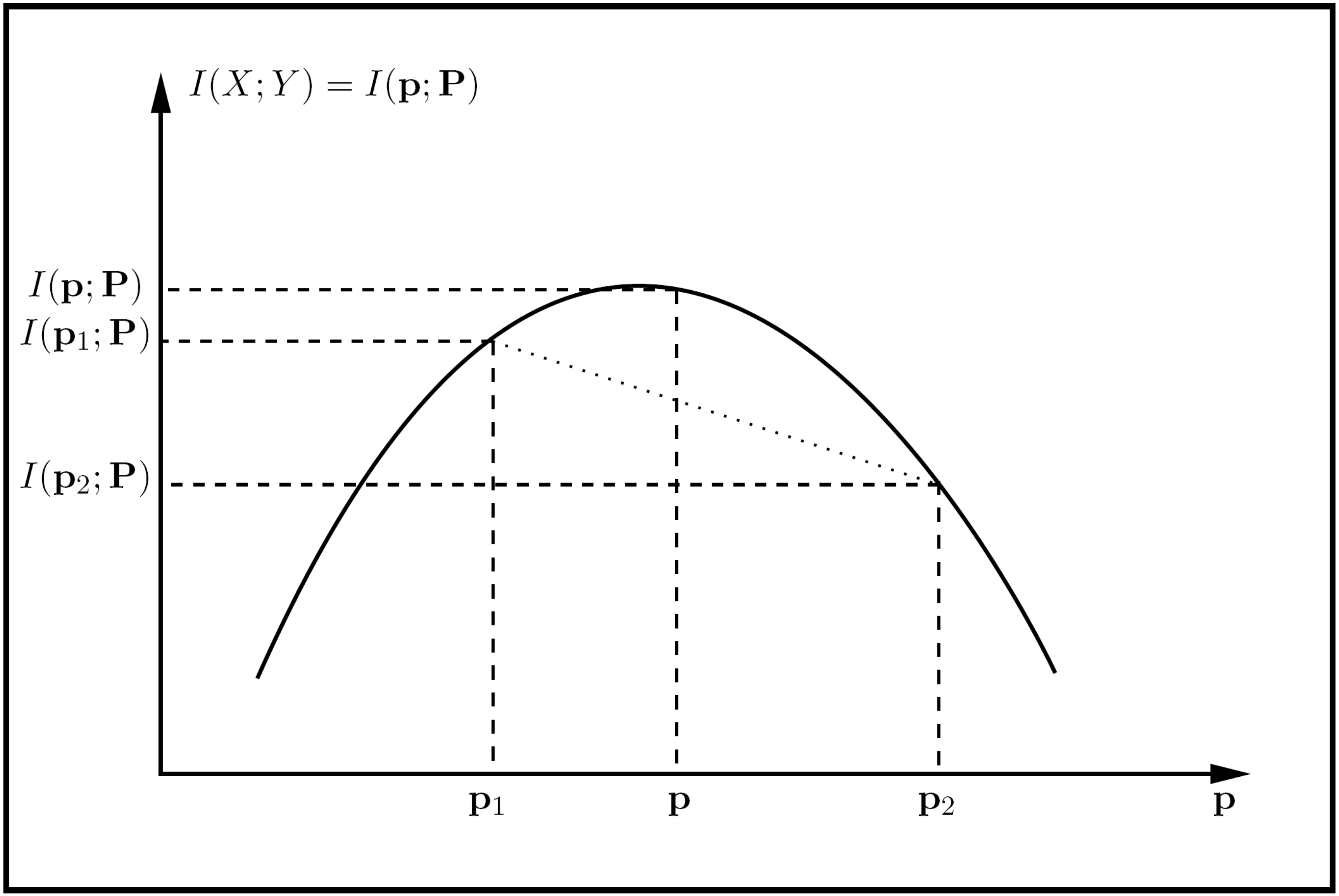
Figure 3.4: Convex ∪ (*convex up* or *convex “cup”*) function.

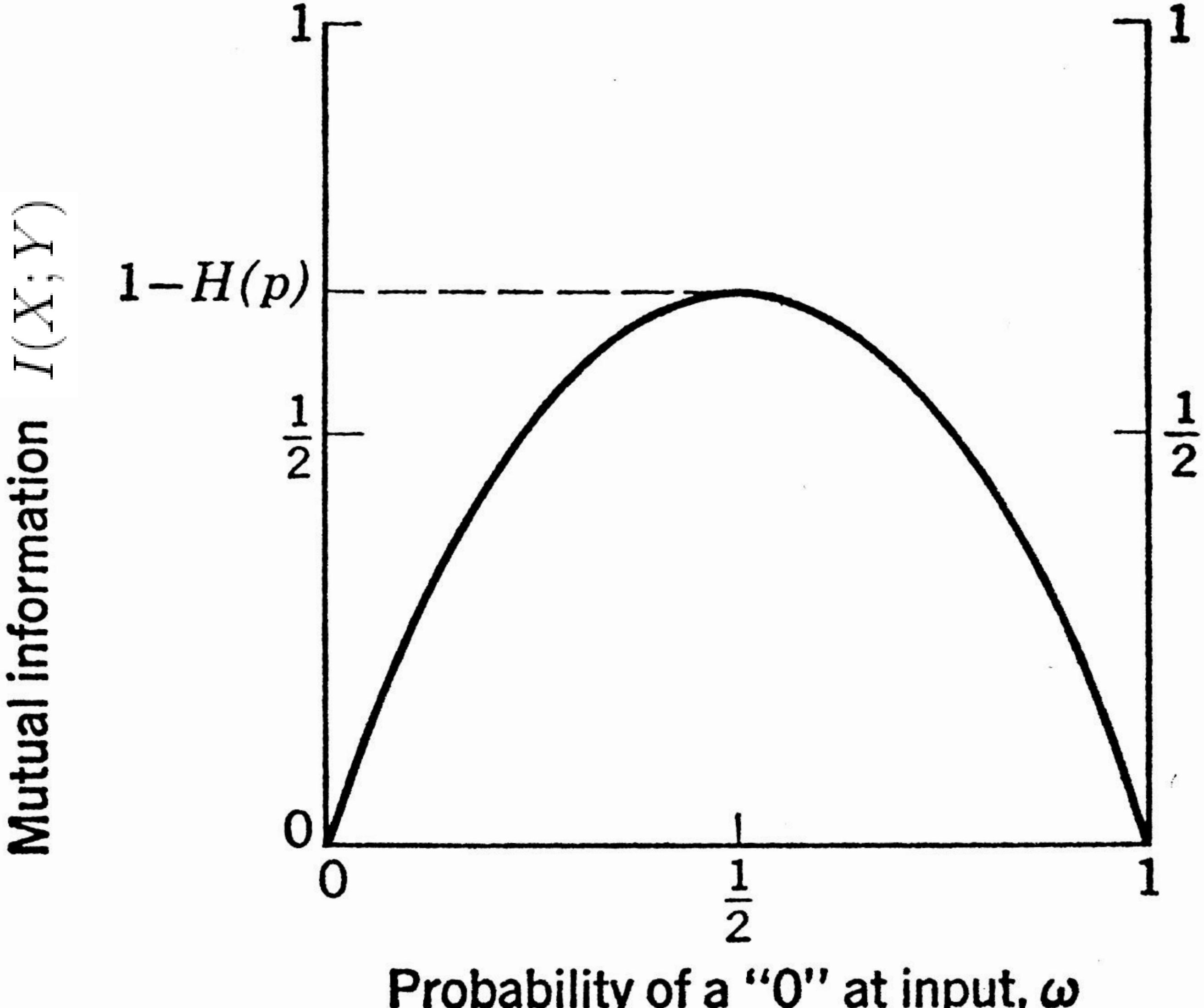
Mutual Information

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

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

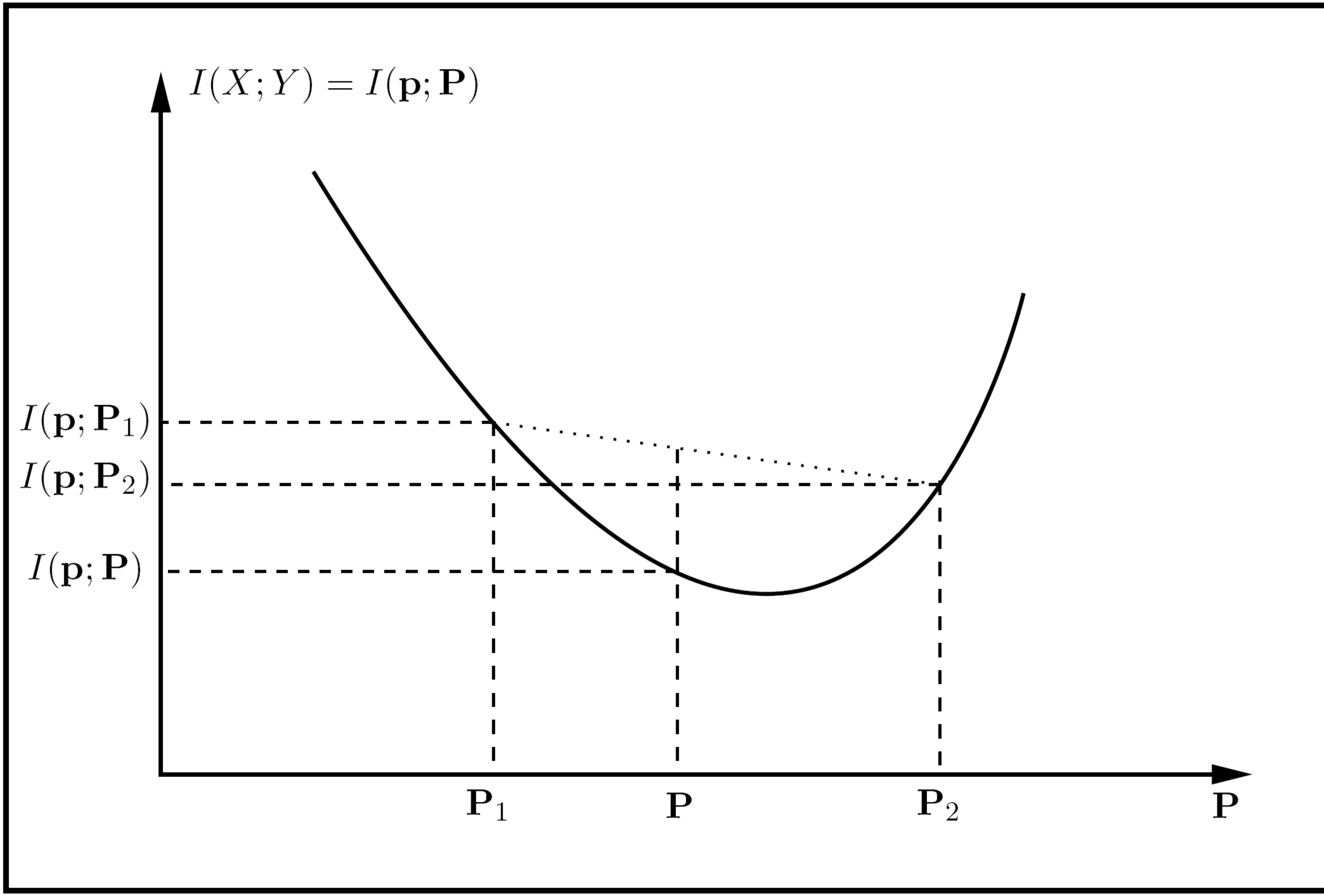
The (average) mutual information $I(X;Y)$ is a *concave* (or *convex "cap"*, or *convex* $\cap$) function over the *convex set* $\mathcal{S}_{\mathbf{p}}$ of all possible input distributions $\{\mathbf{p}\}$.


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


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

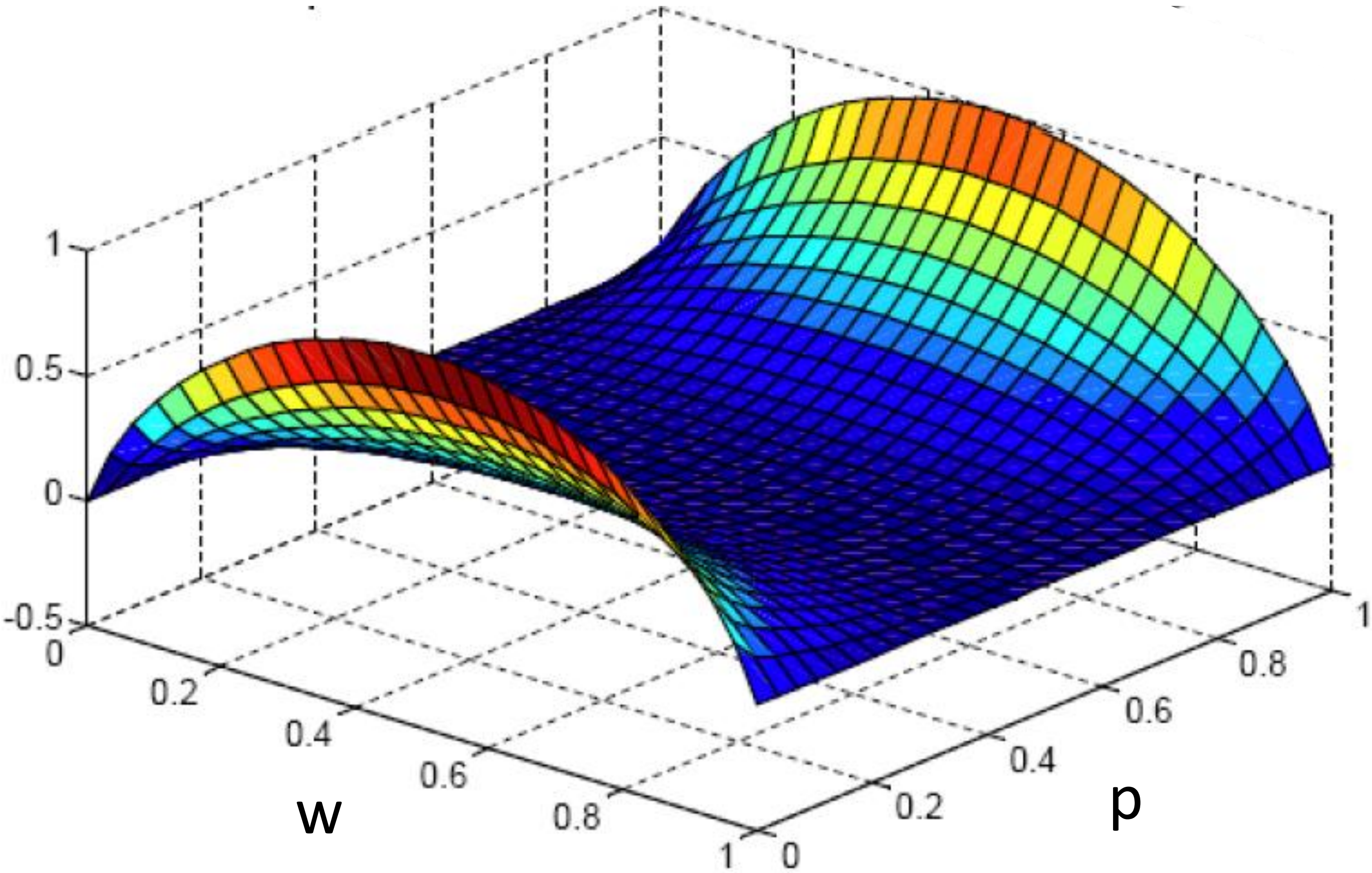
Theorem *(Convexity of the mutual information function):*

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 $\{\mathbf{P}\}$.



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

I(X;Y)



Channel Capacity

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

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

Symmetric Channels

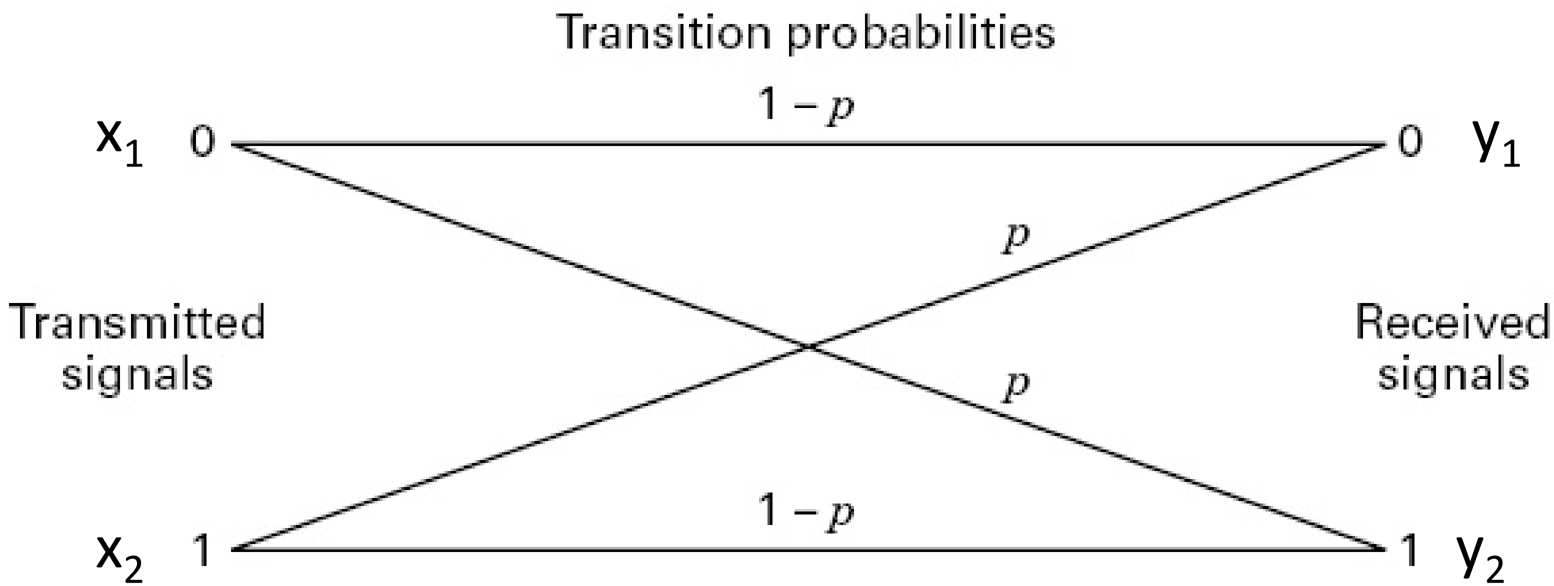
A discrete memoryless channel is said to be *symmetric* if the set of output symbols $\{y_j\}_{j=1,\ldots,J}$ can be partitionned into subsets such that for each subset of the matrix of transition probabilities, each column is a permutation of each other and each row is also a permutation of each other row.

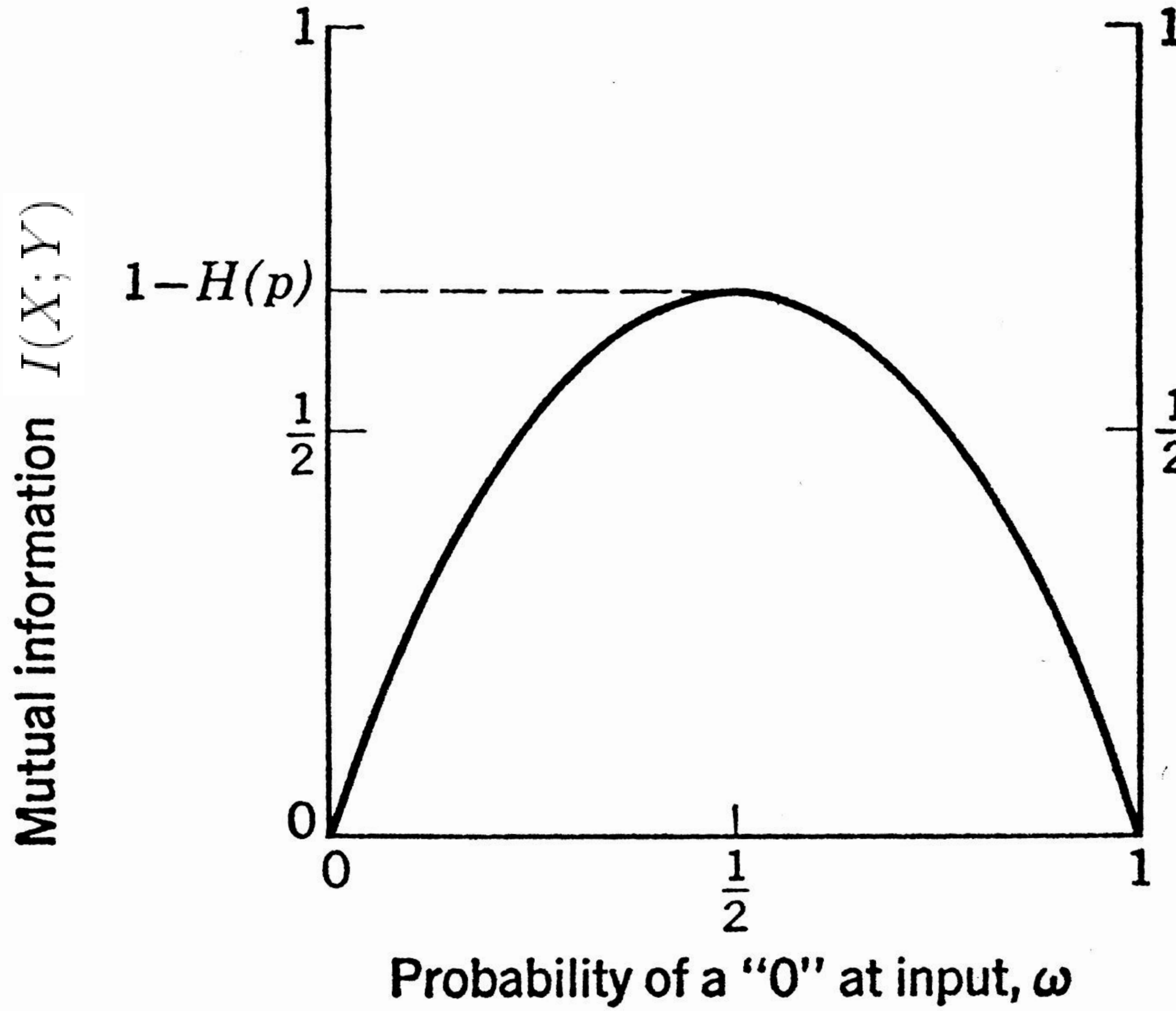
Theorem *(Capacity of a symmetric channel):*

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:

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

Binary Symmetric Channel

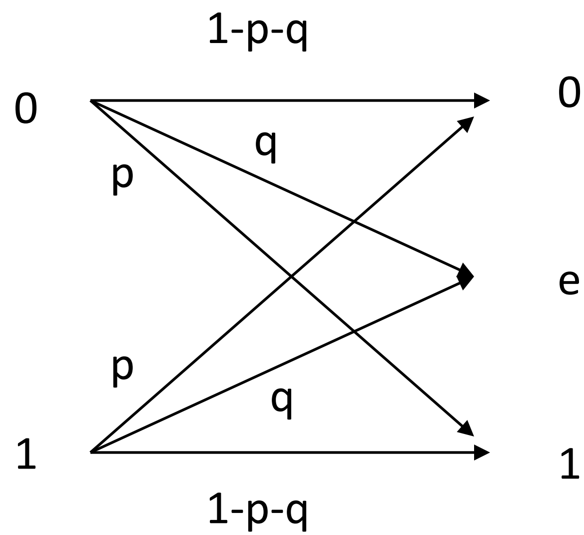



1
1
$I(X;Y)$
Mutual information
$1-H(p)$
$\frac{1}{2}$
$\frac{1}{2}$
0
0
$\frac{1}{2}$
1
Probability of a "0" at input, ω

r-ary Symmetric Channel

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

Binary Errors with Erasure Channel

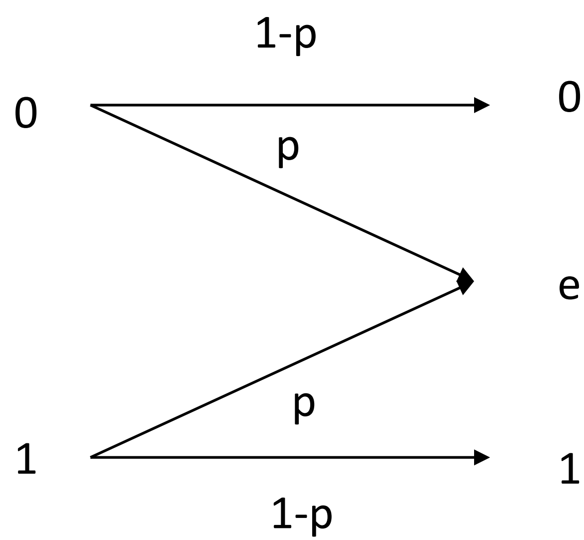


Binary Errors with Erasure Channel

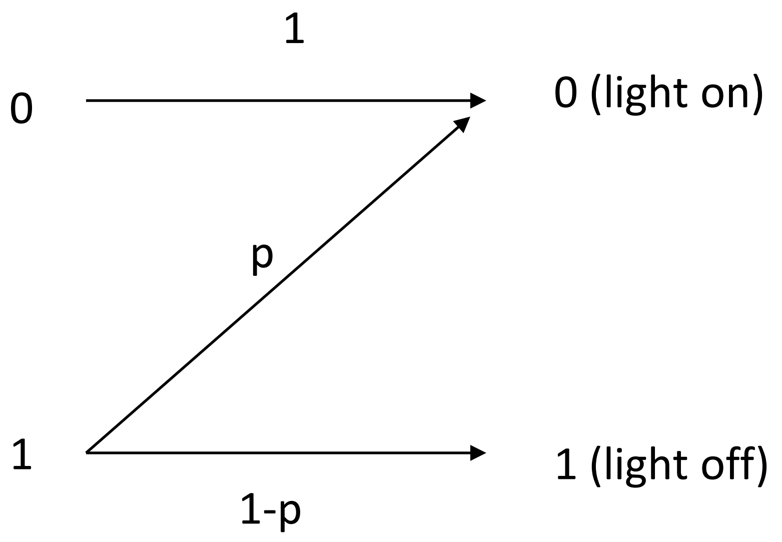
$$P_{Y|X} = \begin{bmatrix} .8 & .05 \\ .15 & .15 \\ .05 & .8 \end{bmatrix}$$

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

Binary Erasure Channel



Z Channel (Optical)

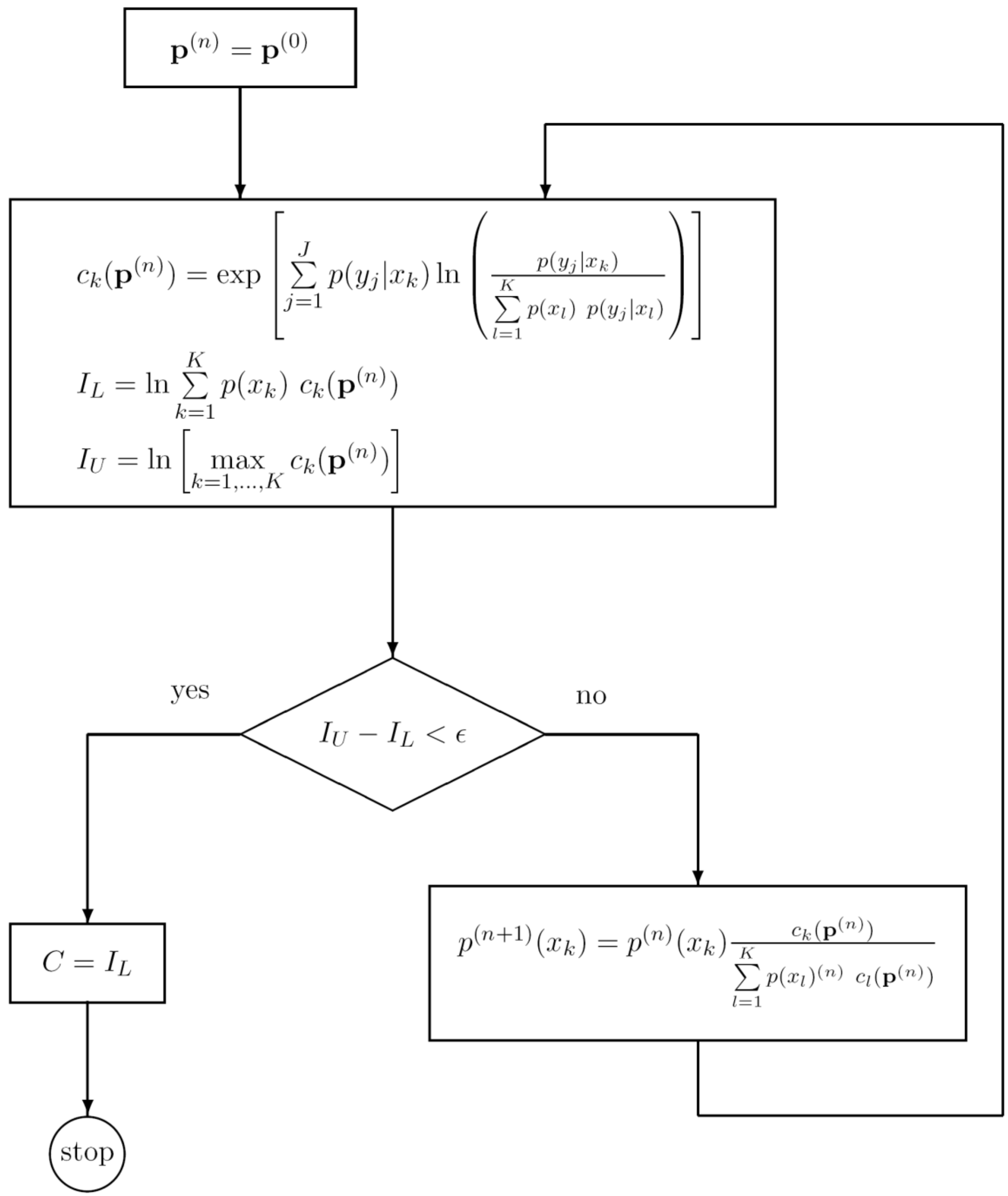


Blahut-Arimoto Algorithm

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

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

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



$\mathbf{p}^{(n)} = \mathbf{p}^{(0)}$
$c_k(\mathbf{p}^{(n)}) = \exp\left[\sum_{j=1}^{J} p(y_j|x_k) \ln\left(\frac{p(y_j|x_k)}{\sum_{l=1}^{K} p(x_l)\ p(y_j|x_l)}\right)\right]$
$I_L = \ln \sum_{k=1}^{K} p(x_k)\ c_k(\mathbf{p}^{(n)})$
$I_U = \ln\left[\max_{k=1,\ldots,K} c_k(\mathbf{p}^{(n)})\right]$
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

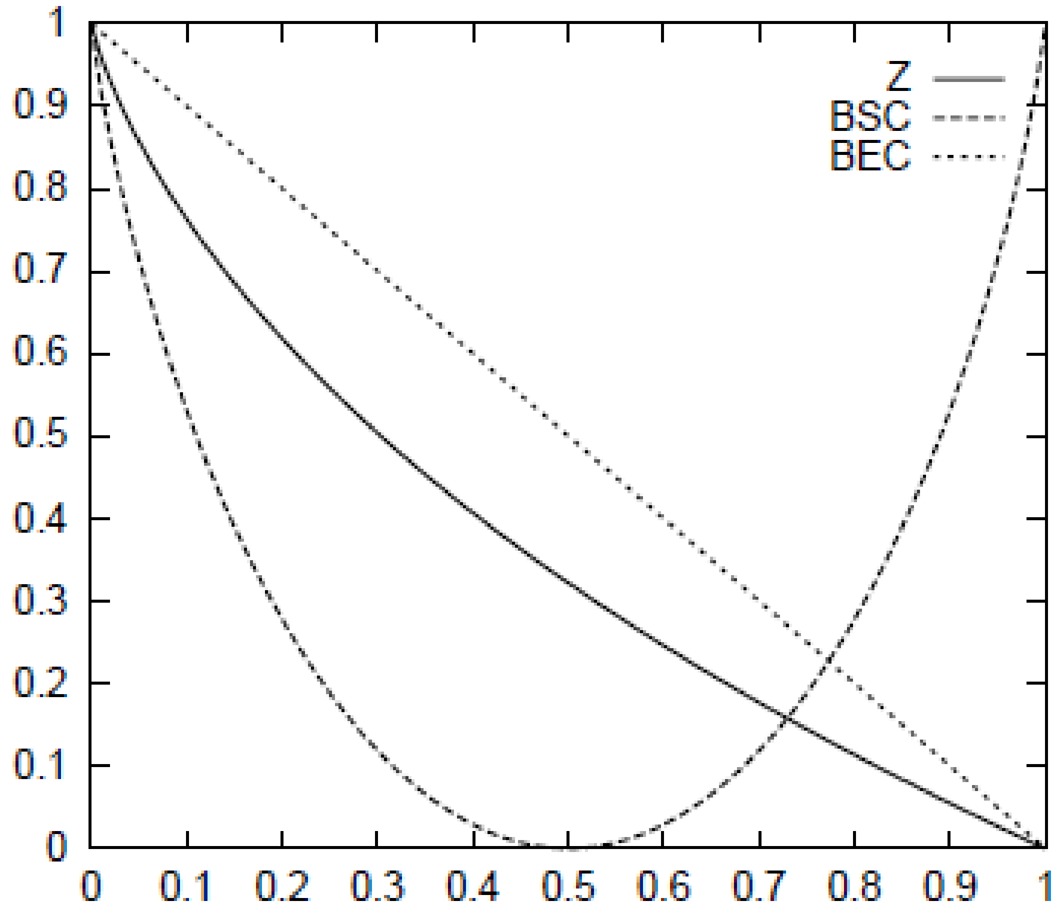
$$P_1 = \begin{bmatrix} .98 & .05 \\ .02 & .95 \end{bmatrix} \quad P_2 = \begin{bmatrix} .80 & .05 \\ .20 & .95 \end{bmatrix}$$

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

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

	C	$\mathbf{p}^*$	$I(X;Y)_{\mathrm{u}}$
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

Channel Capacity for the Z, BSC and BEC



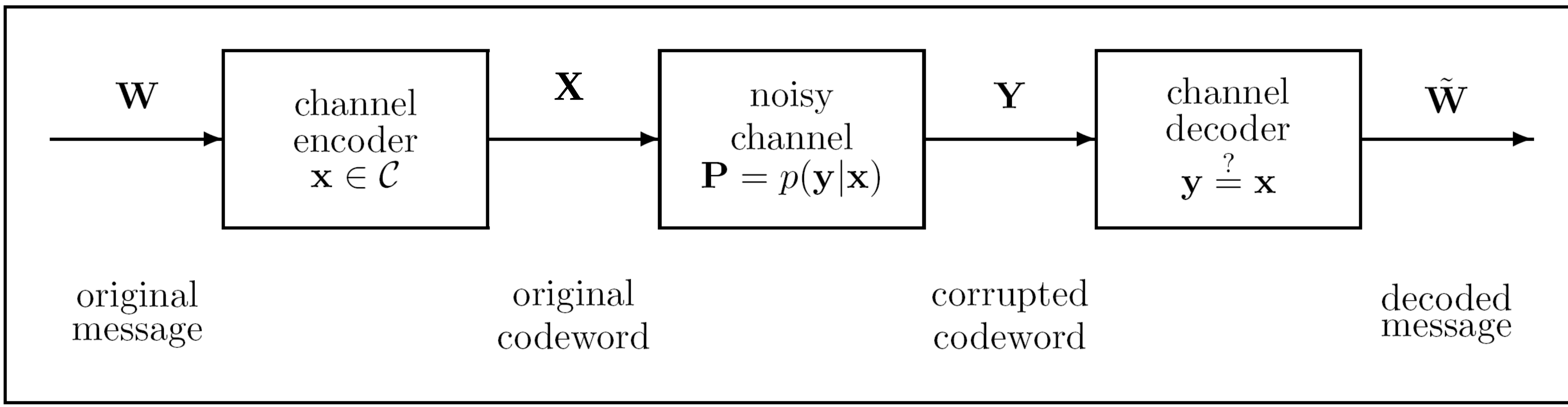
Channel Capacity

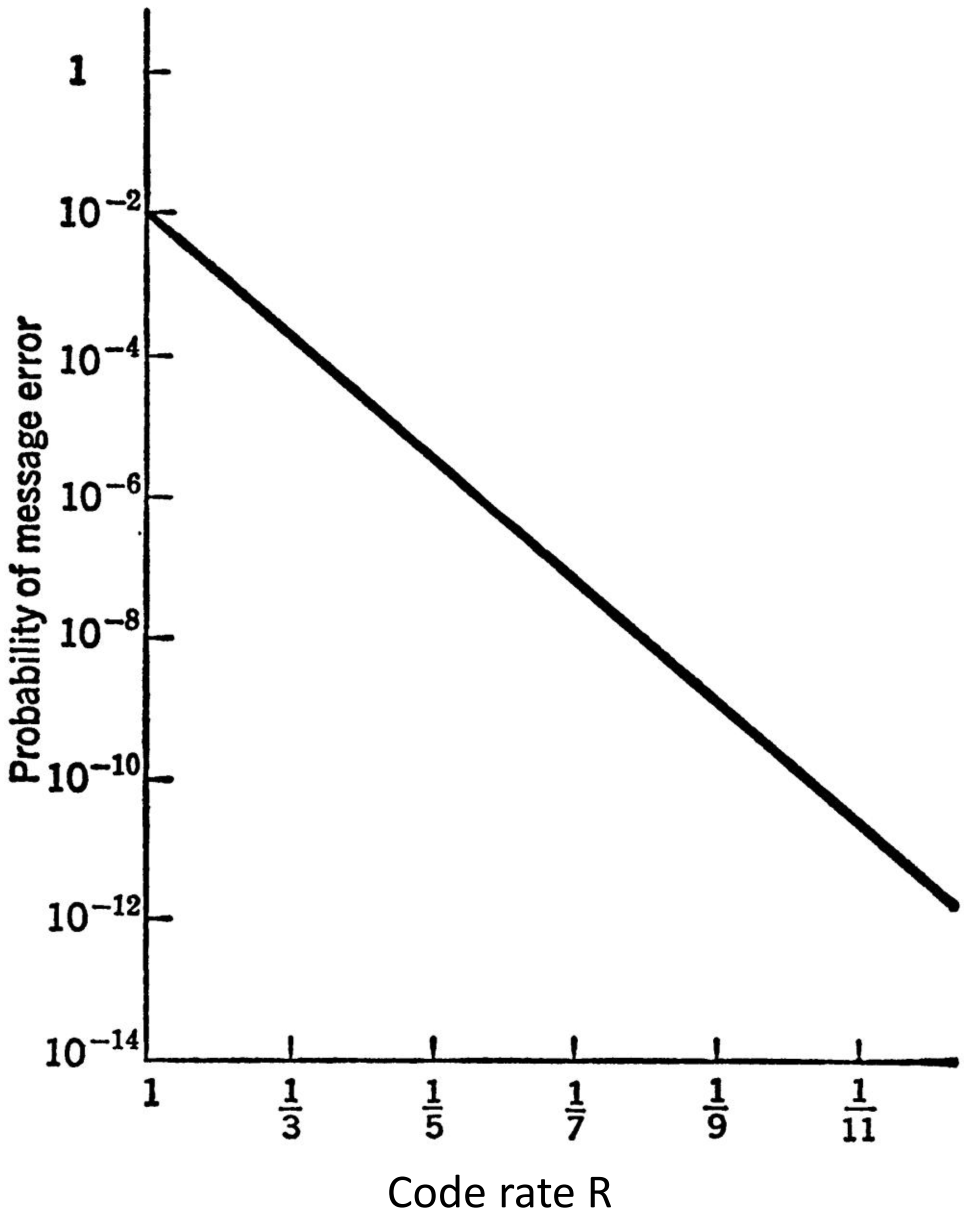
- The channel capacity is the maximum information rate that can be supported by a channel
- Each time a symbol is sent it is said to use the channel
- Thus capacity has the units bits per channel use

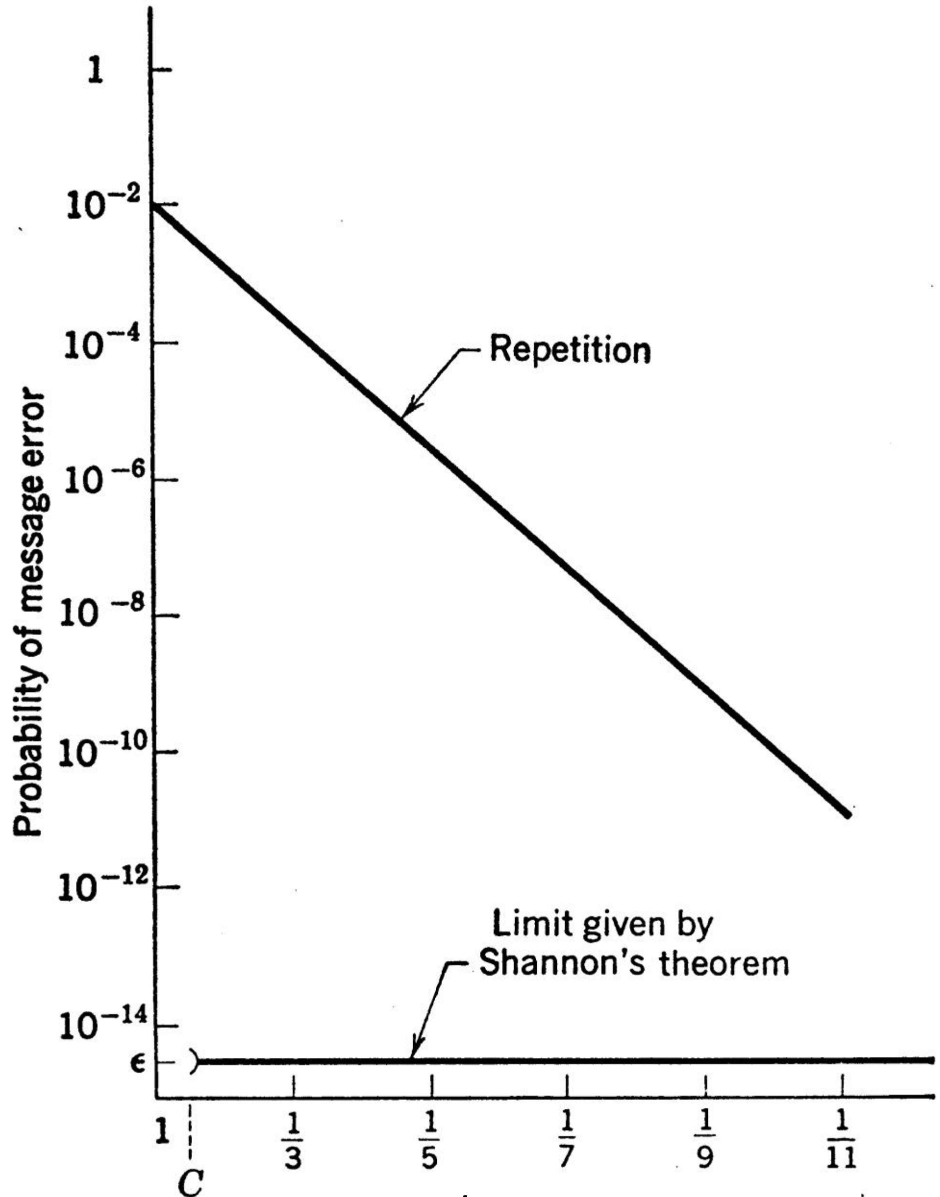
A Mathematical Theory of Communications, BSTJ July, 1948

``If the channel is noisy it is not in general possible to reconstruct the original message or the transmitted signal with certainty by any operation on the received signal.''

Noisy Communication System




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


1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
ϵ
Probability of message error
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

- BSC $p = 0.01$ $N = 3$

Code Rate R	P_e	$M=2^{NR}$
1	0.0297	8
1/3	0.000298	2

- Tradeoff between data rate and error rate

Channel Code

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

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., $N \geq N_0$).

Jointly Typical Pairs of Sequences

$$\mathcal{T}_X(\delta) = \left\{\mathbf{x} \text{ such that: } \left|-\frac{1}{N}\log_b p(\mathbf{x}) - H(X)\right| < \delta\right\}$$

$$\mathcal{T}_Y(\delta) = \left\{\mathbf{y} \text{ such that: } \left|-\frac{1}{N}\log_b p(\mathbf{y}) - H(Y)\right| < \delta\right\}$$

$$\boxed{\mathcal{T}_{XY}(\delta) \equiv \left\{(\mathbf{x},\mathbf{y}) : \left|-\frac{1}{N}\log_b p(\mathbf{x},\mathbf{y}) - H(XY)\right| < \delta\right\}}$$

Theorem *(Shannon-McMillan theorem for jointly typical pairs of sequences):*

Given a dependent pair of memoryless sources of joint entropy $H(XY)$, for a blocklength N sufficiently large (i.e. $N \geq N_0$), the set of pairs of vectors $\{(\mathbf{x}, \mathbf{y})\}$ can be partitioned into a set of jointly typical pairs of sequences $\mathcal{T}_{XY}(\delta)$ and a set of jointly atypical pairs of sequences $\mathcal{T}^c_{XY}(\delta)$ for which:

a) The probability of atypical sequences is upperbounded as:

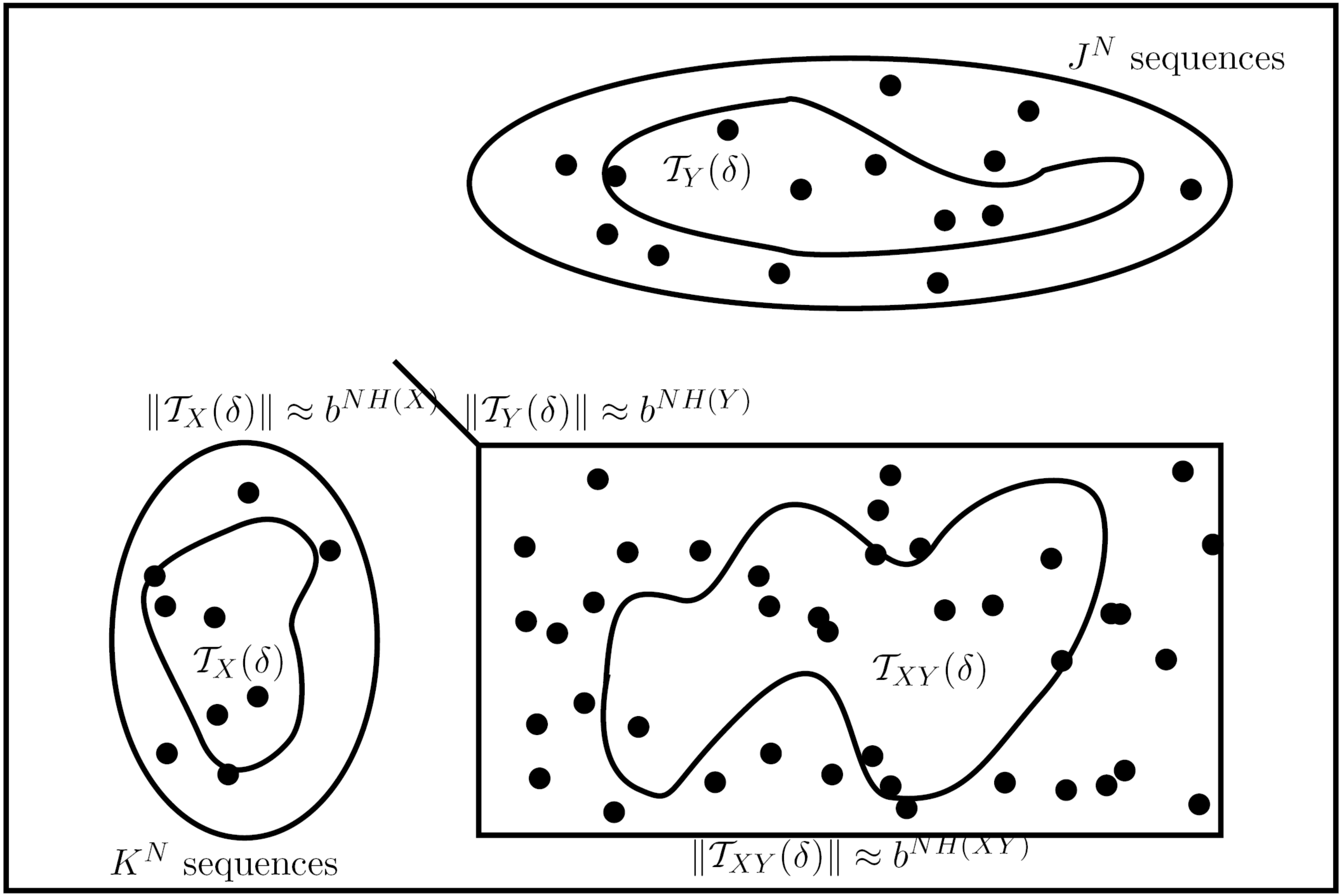
$$\boxed{Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}^c_{XY}(\delta)\right] < \epsilon} \tag{3.30}$$

b) If the pair of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)$ then:

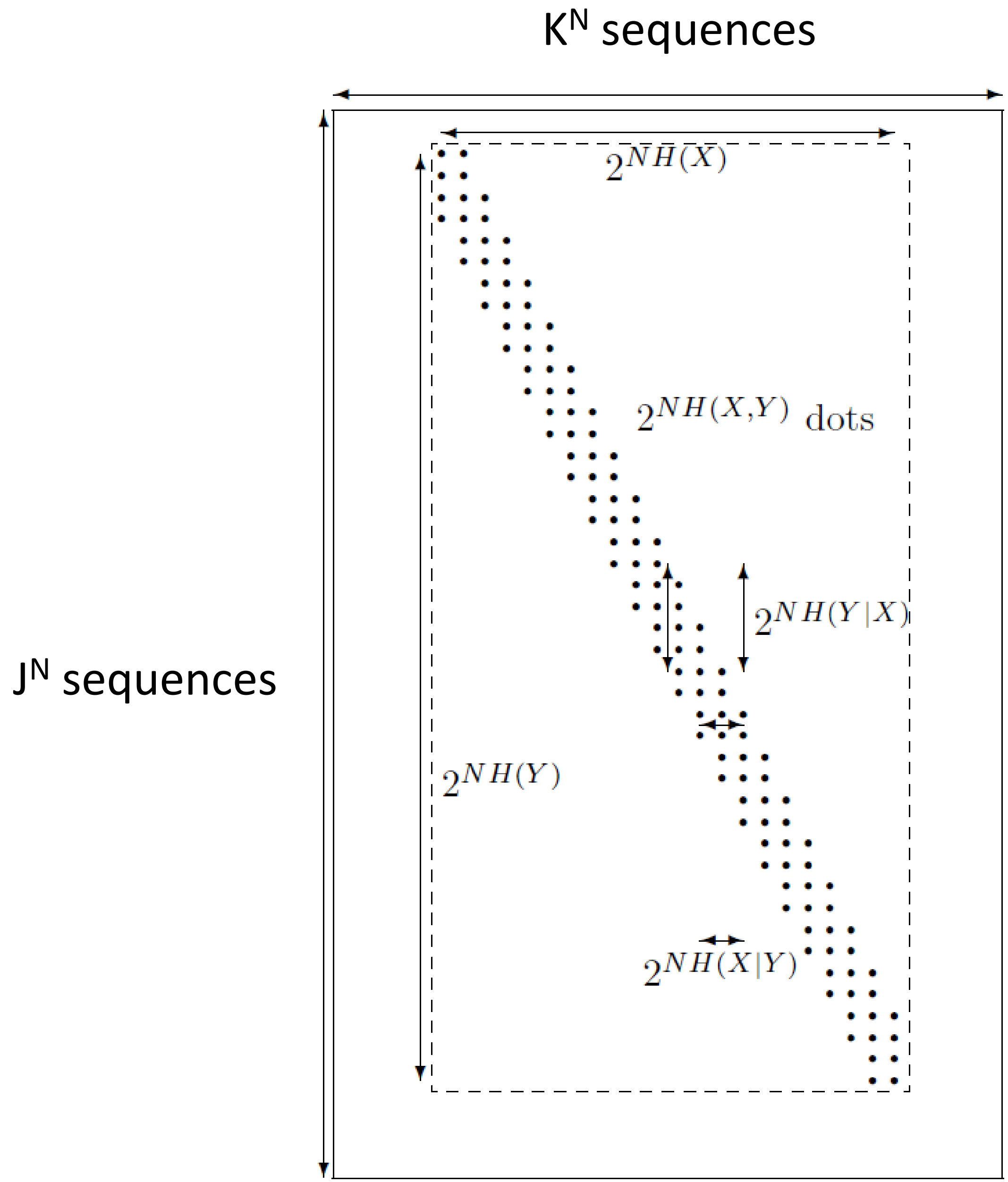
$$\boxed{b^{-N[H(XY)+\delta]} < p(\mathbf{x}, \mathbf{y}) < b^{-N[H(XY)-\delta]}} \tag{3.31}$$

c) The number of elements in the set of jointly typical pairs of sequences $\mathcal{T}_{XY}(\delta)$ is upperbounded by:

$$\boxed{\|\mathcal{T}_{XY}(\delta)\| < b^{N[H(XY)+\delta]}} \tag{3.32}$$

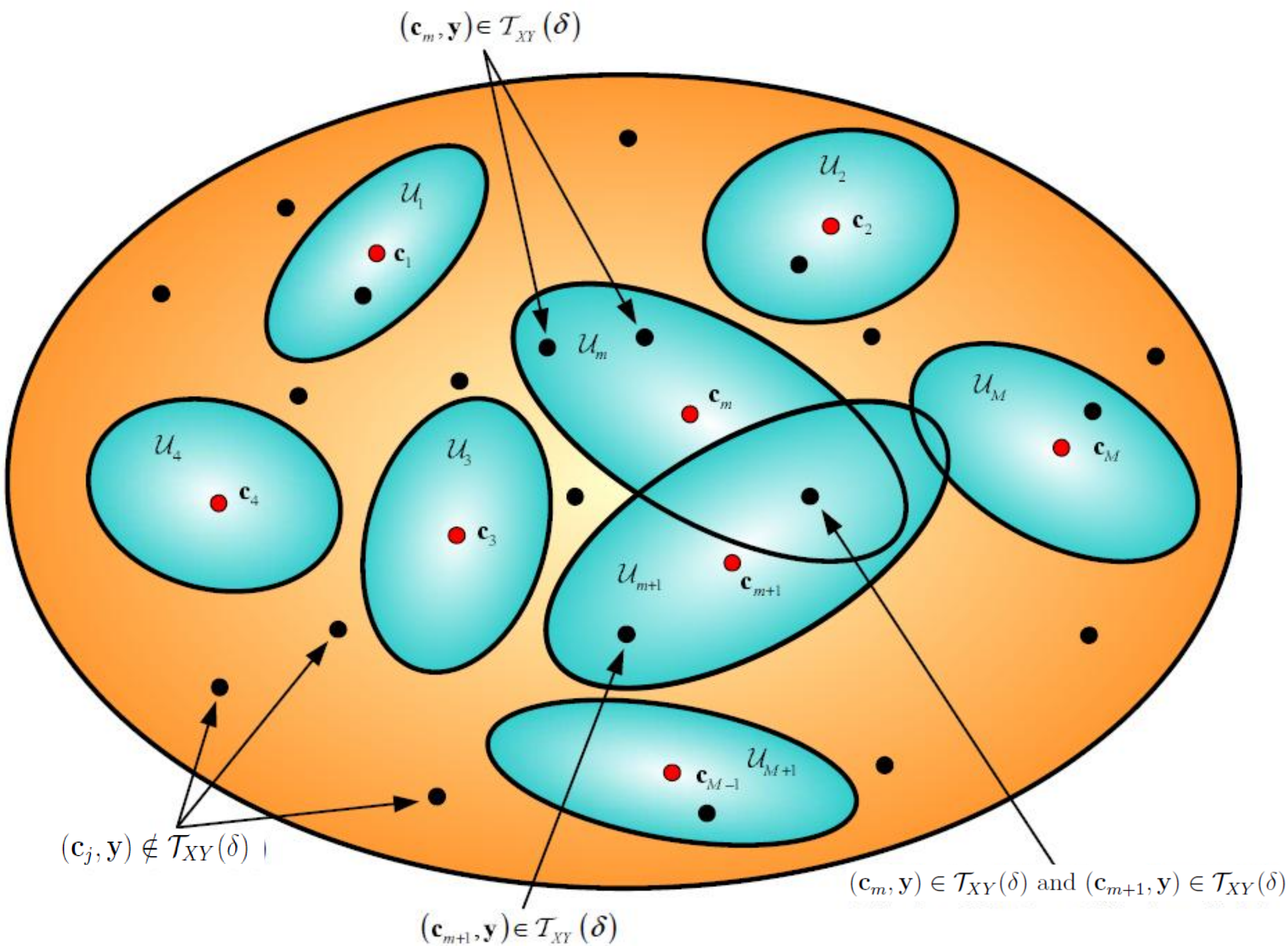

J^N sequences
$\mathcal{T}_Y(\delta)$
$\|\mathcal{T}_X(\delta)\| \approx b^{NH(X)}$
$\|\mathcal{T}_Y(\delta)\| \approx b^{NH(Y)}$
$\mathcal{T}_X(\delta)$
$\mathcal{T}_{XY}(\delta)$
K^N sequences
$\|\mathcal{T}_{XY}(\delta)\| \approx b^{NH(XY)}$

Jointly Typical Sequences



Codeword Elements Chosen from X

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


$(\mathbf{c}_m, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)$
$\mathcal{U}_1$
$\mathbf{c}_1$
$\mathcal{U}_2$
$\mathbf{c}_2$
$\mathcal{U}_m$
$\mathbf{c}_m$
$\mathcal{U}_M$
$\mathbf{c}_M$
$\mathcal{U}_4$
$\mathbf{c}_4$
$\mathcal{U}_3$
$\mathbf{c}_3$
$\mathcal{U}_{m+1}$
$\mathbf{c}_{m+1}$
$\mathbf{c}_{M-1}$
$\mathcal{U}_{M+1}$
$(\mathbf{c}_j, \mathbf{y}) \notin \mathcal{T}_{XY}(\delta)$
$(\mathbf{c}_m, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)$ and $(\mathbf{c}_{m+1}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)$
$(\mathbf{c}_{m+1}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)$

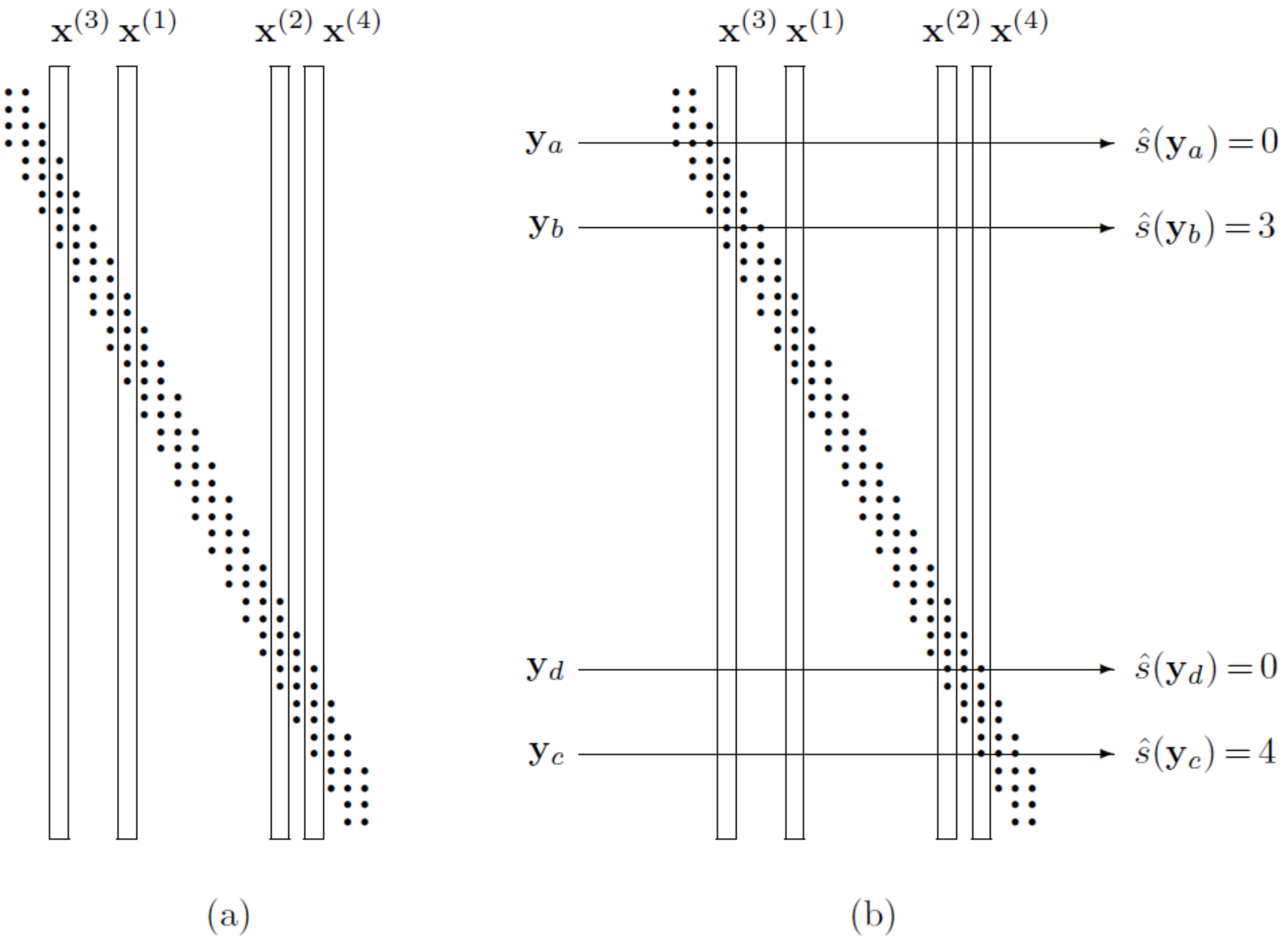


Figure 10.4. (a) A random code. (b) Example decodings by the typical set decoder. A sequence that is not jointly typical with any of the codewords, such as $\mathbf{y}_a$, is decoded as $\hat{s} = 0$. A sequence that is jointly typical with codeword $\mathbf{x}^{(3)}$ alone, $\mathbf{y}_b$, is decoded as $\hat{s} = 3$. Similarly, $\mathbf{y}_c$ is decoded as $\hat{s} = 4$. A sequence that is jointly typical with more than one codeword, such as $\mathbf{y}_d$, is decoded as $\hat{s} = 0$.

$$
\begin{aligned}
P_{e|m} &= Pr\left\{[(\mathbf{c}_m,\mathbf{y}) \notin \mathcal{T}_{XY}(\delta)] \bigcup \bigcup_{\substack{m'=1\\ m'\neq m}}^{M} [(\mathbf{c}_{m'},\mathbf{y}) \in \mathcal{T}_{XY}(\delta)]\right\} \\
P_{e|m} &\leq Pr\left[(\mathbf{c}_m,\mathbf{y}) \notin \mathcal{T}_{XY}(\delta)\right] + \sum_{\substack{m'=1\\ m'\neq m}}^{M} Pr\left[(\mathbf{c}_{m'},\mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right]
\end{aligned}
$$

$$\sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[P_{e|m}\right] \leq \sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[\epsilon_1 + \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \sum_{\mathbf{y}} \phi(\mathbf{c}_{m'}, \mathbf{y}) \; p(\mathbf{y}|\mathbf{c}_m)\right]$$

$$\sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[P_{e|m}\right] \leq \sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) [\epsilon_1] + \sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[\sum_{\substack{m'=1 \\ m' \neq m}}^{M} \sum_{\mathbf{y}} \phi(\mathbf{c}_{m'}, \mathbf{y}) \; p(\mathbf{y}|\mathbf{c}_m)\right]$$

$$\sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[P_{e|m}\right] \leq \epsilon_1 + \sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[\sum_{\substack{m'=1 \\ m' \neq m}}^{M} \sum_{\mathbf{y}} \phi(\mathbf{c}_{m'}, \mathbf{y}) \; p(\mathbf{y}|\mathbf{c}_m)\right]$$

$$\sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[P_{e|m}\right] \leq \epsilon_1 + \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[\sum_{\mathbf{y}} \phi(\mathbf{c}_{m'}, \mathbf{y}) \; p(\mathbf{y}|\mathbf{c}_m)\right]$$

$$\sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[P_{e|m}\right] \;\; \leq \;\; \epsilon_1 + \sum_{\substack{m'=1 \\ m' \neq m}}^{M} Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right]$$

$$\sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[P_{e|m}\right] \;\; \leq \;\; \epsilon_1 + (M-1)\; Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right]$$

$$P_e = \sum_{m=1}^{M} p(\mathbf{c}_m)\; P_{e|m}$$

$$\sum_{\mathcal{S}_{\mathcal{C}}} Pr(\mathcal{C}) \left[P_e\right] \leq \epsilon_1 + (M-1)\; Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right]$$

$$
\begin{aligned}
\mathbf{x} \in \mathcal{T}_X(\delta) &\rightarrow b^{-N[H(X)+\delta]} \leq p(\mathbf{x}) \leq b^{-N[H(X)-\delta]} \\
\mathbf{y} \in \mathcal{T}_Y(\delta) &\rightarrow b^{-N[H(Y)+\delta]} \leq p(\mathbf{y}) \leq b^{-N[H(Y)-\delta]} \\
(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta) &\rightarrow \|\mathcal{T}_{XY}(\delta)\| \leq b^{N[H(XY)+\delta]}
\end{aligned}
$$

$$
\begin{aligned}
Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right] &= \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{T}_{XY}(\delta)} p(\mathbf{x}, \mathbf{y}) \\
Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right] &= \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{T}_{XY}(\delta)} p(\mathbf{x})p(\mathbf{y}) \\
Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right] &= \|\mathcal{T}_{XY}(\delta)\| p(\mathbf{x})p(\mathbf{y}) \\
Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right] &\leq \|\mathcal{T}_{XY}(\delta)\| \; b^{-N[H(X)-\delta]} \; b^{-N[H(Y)-\delta]} \\
Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right] &\leq b^{N[H(XY)+\delta]} \; b^{-N[H(X)-\delta]} \; b^{-N[H(Y)-\delta]} \\
Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right] &\leq b^{-N[H(X)+H(Y)-H(XY)-3\delta]} \\
Pr\left[(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right] &\leq b^{-N[I(X;Y)-3\delta]}
\end{aligned}
$$

$$\sum_{\mathcal{S}_\mathcal{C}} Pr(\mathcal{C})\,[P_e] \;\leq\; \epsilon_1 + (M-1)\; Pr\left[(\mathbf{x},\mathbf{y}) \in \mathcal{T}_{XY}(\delta)\right]$$

$$\sum_{\mathcal{S}_\mathcal{C}} Pr(\mathcal{C})\,[P_e] \;\leq\; \epsilon_1 + (M-1)\; b^{-N[I(X;Y)-3\delta]}$$

$$\sum_{\mathcal{S}_\mathcal{C}} Pr(\mathcal{C})\,[P_e] \;\leq\; \epsilon_1 + M\; b^{-N[I(X;Y)-3\delta]}$$

$$\sum_{\mathcal{S}_\mathcal{C}} Pr(\mathcal{C})\,[P_e] \;\leq\; \epsilon_1 + b^{NR}\; b^{-N[I(X;Y)-3\delta]}$$

$$\sum_{\mathcal{S}_\mathcal{C}} Pr(\mathcal{C})\,[P_e] \;\leq\; \epsilon_1 + b^{-N[I(X;Y)-R-3\delta]}$$

$$\sum_{\mathcal{S}_\mathcal{C}} Pr(\mathcal{C})\,[P_e] \leq \epsilon_1 + \epsilon_2$$

$$P_e \leq \epsilon$$

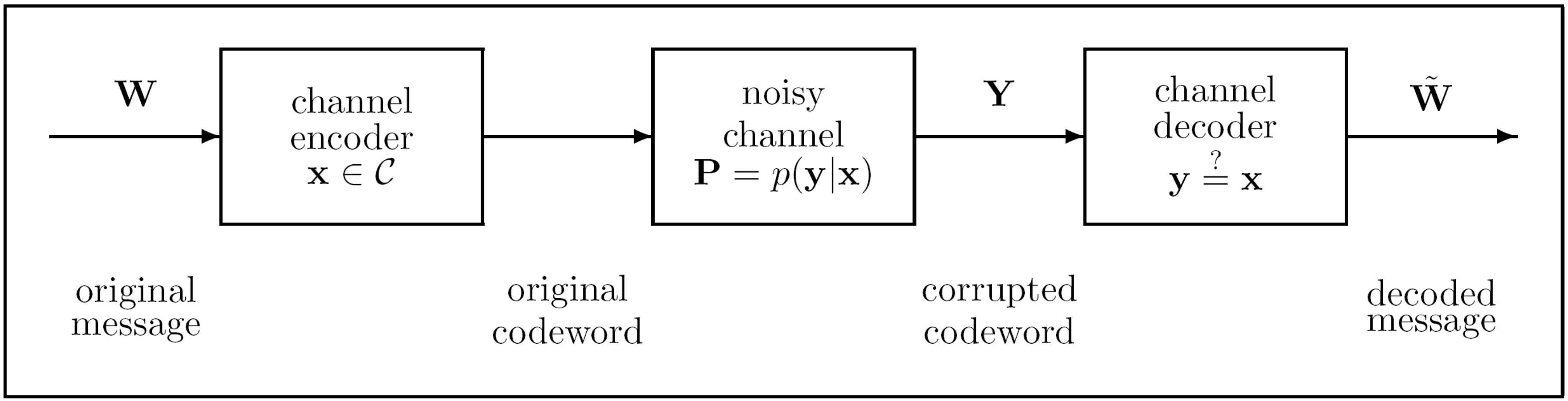
Theorem *(Shannon's channel coding theorem):*

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;

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

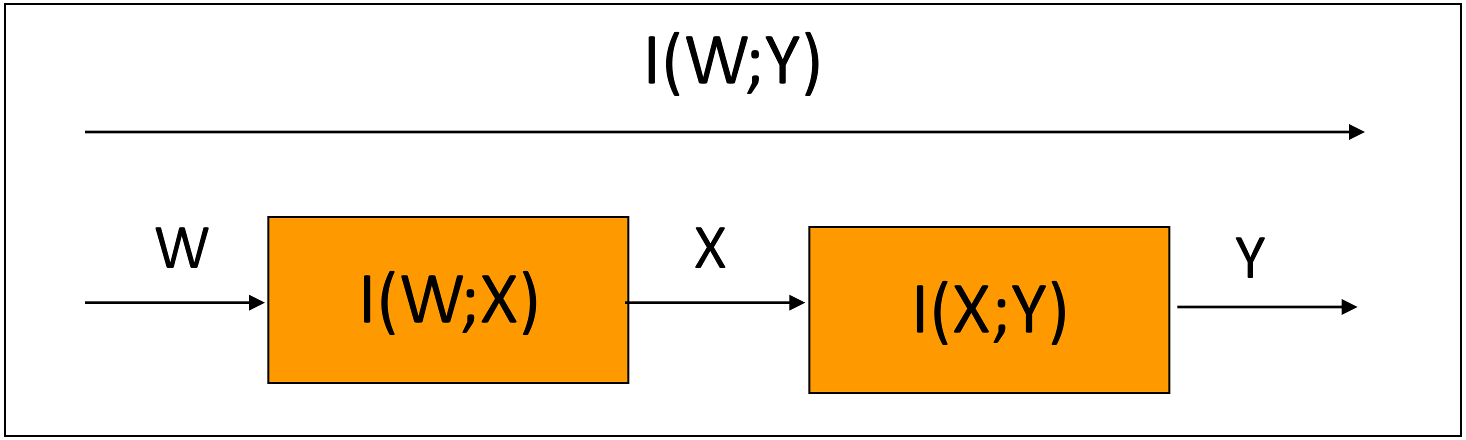
provided that the blocklength N is sufficiently large (i.e., $N \geq N_0$).

Noisy Communication System



The Data Processing Lemma

Cascaded Channels



The overall transmission rate $I(W;Y)$ for the cascade cannot be larger than $I(W;X)$ or $I(X;Y)$, so that

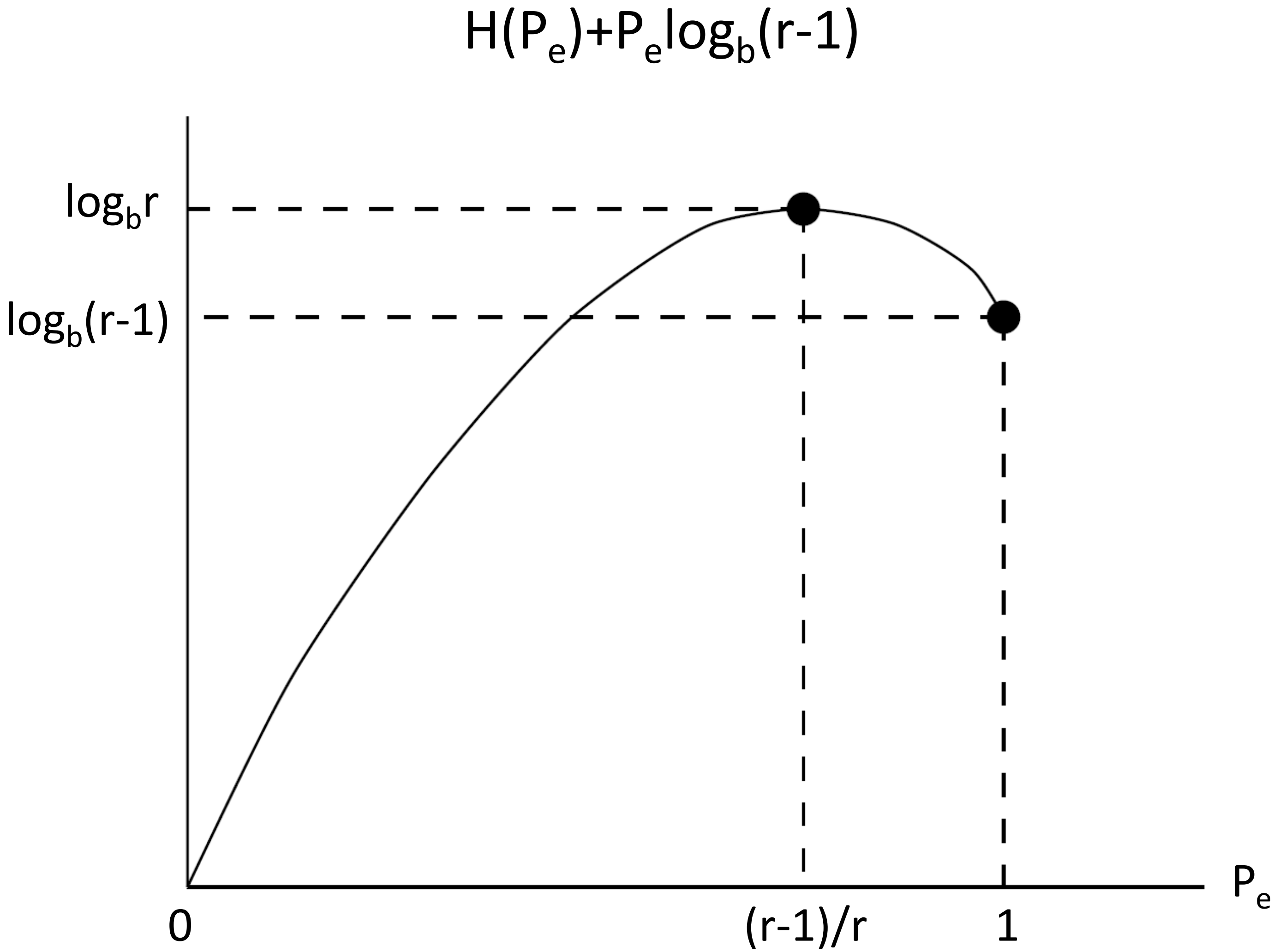
$$I(W;Y) \leq I(W;X) \qquad I(W;Y) \leq I(X;Y)$$

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.



H(Pe)+Pelogb(r-1)
logbr
logb(r-1)
Pe
0
(r-1)/r
1

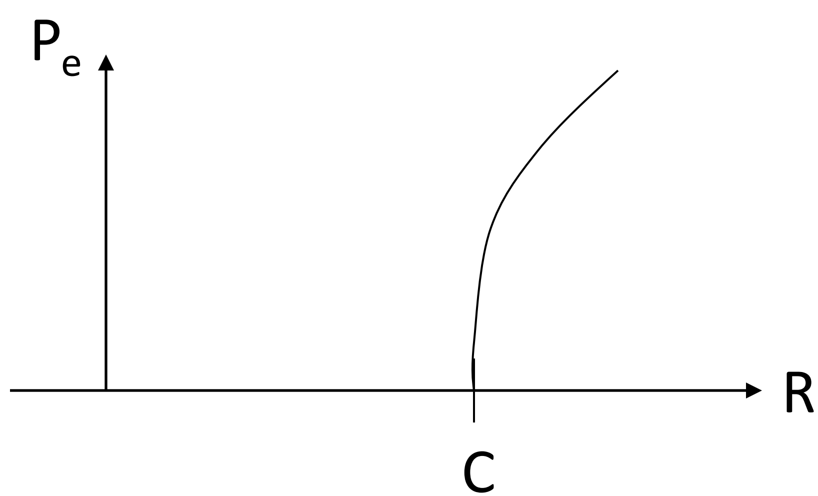
Fano's Inequality

$$H(\mathbf{X}|\mathbf{Y}) \leq H(P_e)+P_e\log_b(r-1)$$

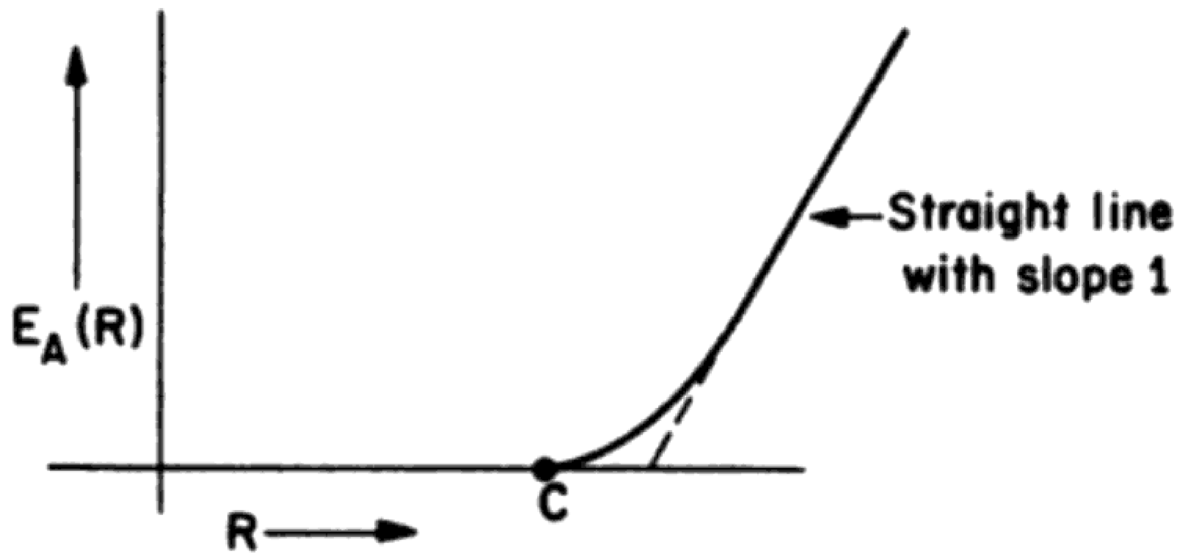
$$H(\mathbf{X}|\mathbf{Y}) \leq 1+NRP_e(N,R)$$

Channel Capacity: Converse

For R > C the decoding error probability > 0



Arimoto’s Error Exponent $E_A(R)$



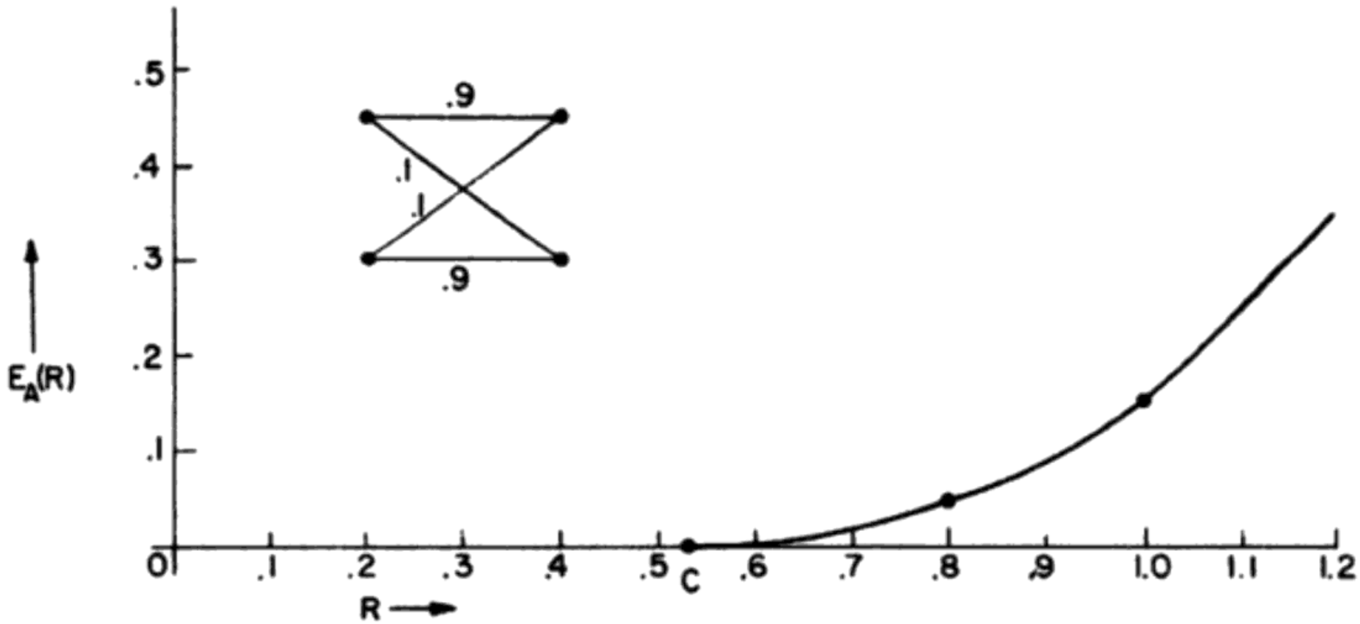


Figure 6.7 $E_A(R)$ for a BSC with $p = 0.1$.

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

Let C be the information transfer capacity of a memoryless channel defined by its transition probabilities matrix $\mathbf{P}$. 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* as:

$$\boxed{P_e(N,R) \leq \min_{s \in [0,1]} \min_{\mathcal{S}_{\mathbf{p}}} \left[b^{sNR} \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p(\mathbf{x}) p(\mathbf{y}|\mathbf{x})^{\frac{1}{(1+s)}} \right)^{1+s} \right]}$$

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

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \prod_{n=1}^{N} p(y_j|x_k) \\ p(\mathbf{x}) &= \prod_{n=1}^{N} p(x_k) \end{aligned}$$

$$P_e \leq b^{NRs} \left[\sum_{j=1}^{J} \left[\sum_{k=1}^{K} p(x_k) p(y_j|x_k)^{\left(\frac{1}{1+s}\right)} \right]^{(s+1)} \right]^{N}$$

$$P_e(N,R) \leq \min_{s\in[0,1]} \min_{\mathcal{S}_{\mathbf{p}}} \left[b^{NRs} \left[\sum_{j=1}^{J} \left[\sum_{k=1}^{K} p(x_k) p(y_j|x_k)^{\left(\frac{1}{1+s}\right)} \right]^{(s+1)} \right]^{N} \right]$$

Random Coding Exponent

$$E_r(R) \equiv \max_{s\in[0,1]} \max_{\mathcal{S}_{\mathbf{p}}} \left[-sR - \log_b \sum_{j=1}^{J} \left(\sum_{k=1}^{K} p(x_k) p(y_j|x_k)^{\frac{1}{(1+s)}} \right)^{1+s} \right]$$

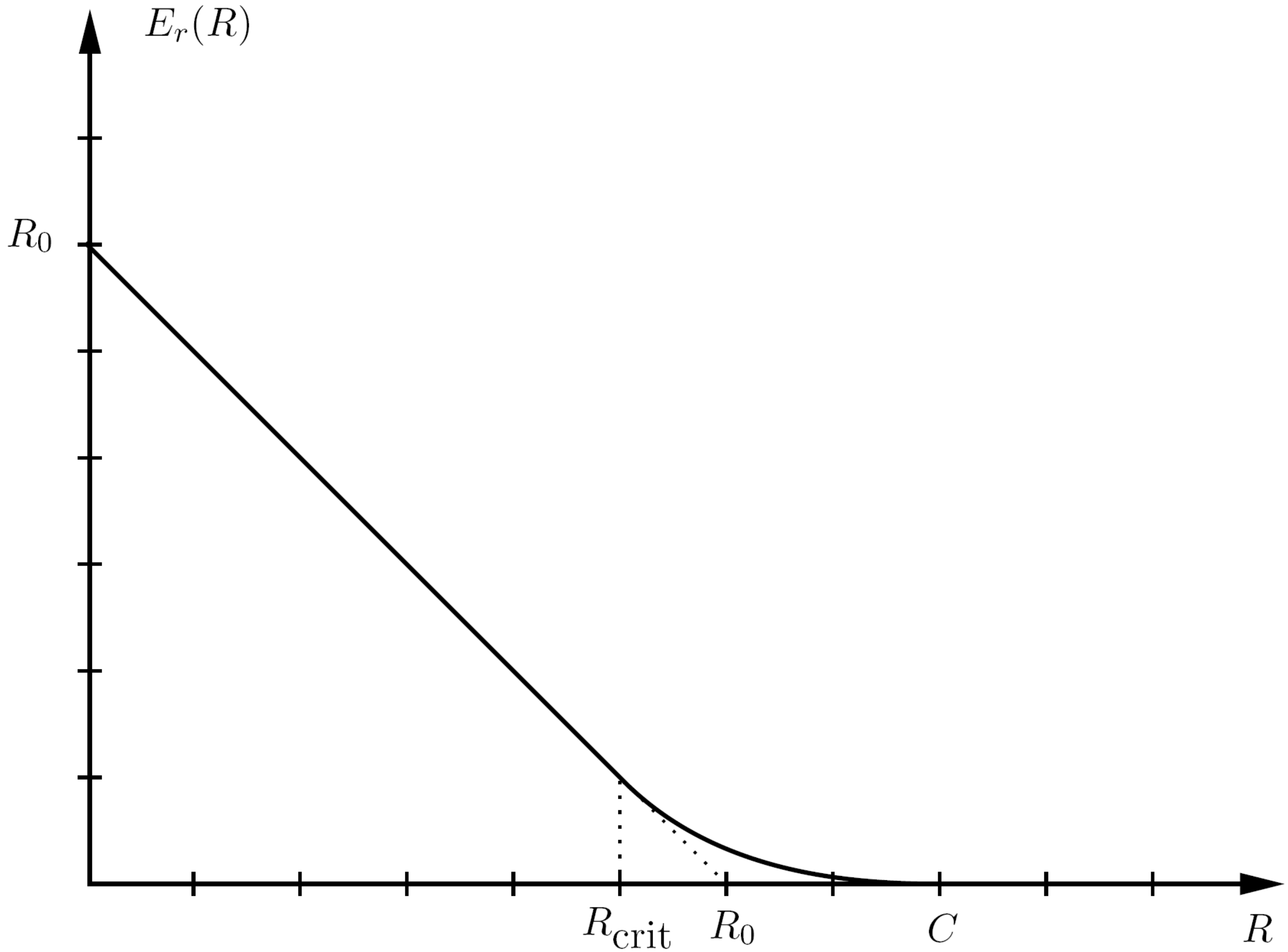
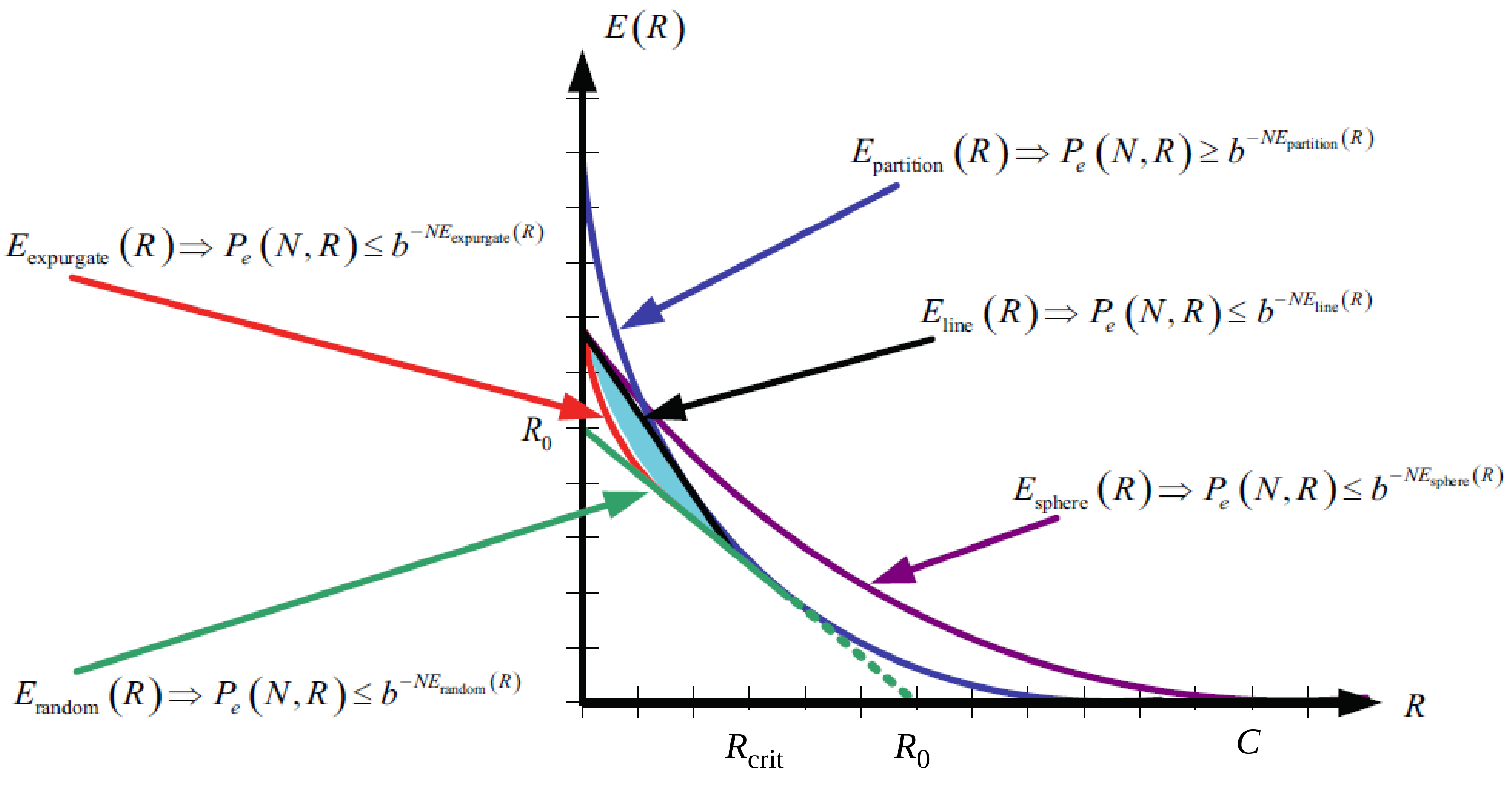


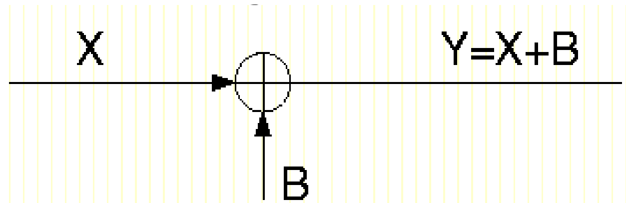
Figure 3.13: Random coding exponent $E_r(R)$ for block codes for BSC with $\epsilon = 10^{-2}$.

Error Bounds

- Random coding bound $P_e(N,R) \le b^{-NE_r(R)}$
- Expurgated bound $P_e(N,R) \le b^{-NE_x(R)}$
- Space-partitioning bound $P_e(N,R) \ge b^{-NE_p(R)}$
- Sphere-packing bound $P_e(N,R) \ge b^{-NE_s(R)}$
- Straight-line bound $P_e(N,R) \ge b^{-NE_l(R)}$

Channel Reliability Function

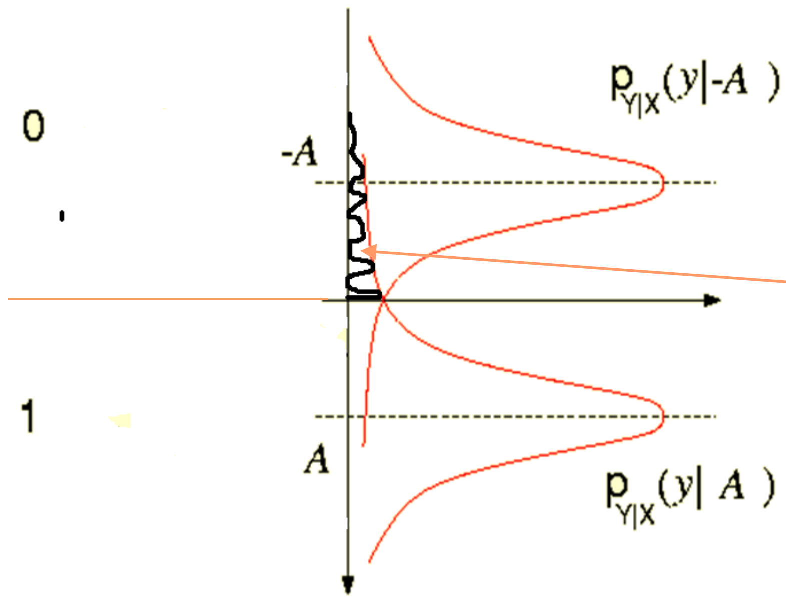


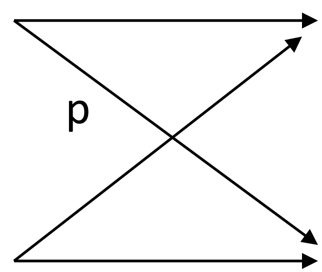


$$X = \{x_1, x_2, \ldots, x_k\}$$

$$B \sim \mathcal{N}(0, \sigma^2)$$

$$Y \in \mathbb{R}$$





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