Chapter 2 Channel Capacity

- 2.1 Introduction
- 2.2 Binary Symmetric Channel
- 2.3 Binary Erasure Channel
- 2.4 AWGN Channel
- 2.5 Fading Channels

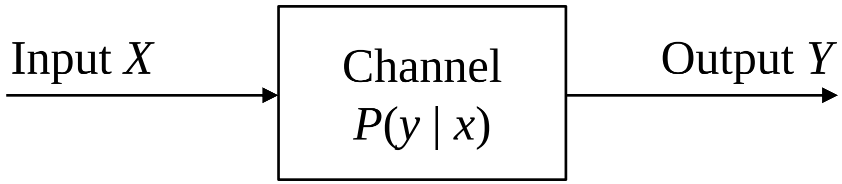
§ 2.1 Introduction

- In a communication system, with the observation of Y, we aim to recover X.
- Mutual Information $I(X,Y) = H(X) - H(X|Y)$
$$= H(Y) - H(Y|X)$$

  It defines the amount of uncertainty about X that has been reduced by knowing Y, and vise versa. This uncertainty discrepancy is introduced by the channel.

- Channel capacity describes the channel’s best capability in reducing the uncertainty.

§ 2.1 Introduction

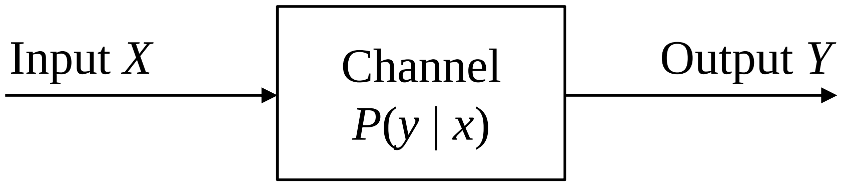


- Let the realizations of input X and output Y be x and y, respectively.
- Channel transition probability $P(y \mid x)$: knowing x was transmitted, the probability of observing y. It defines the channel quality.
- Mutual information between X and Y

$$I(X, Y) = \mathbb{E}\left[\log_2 \frac{P(y \mid x)}{P(y)}\right]$$
$$= \mathbb{E}\left[\log_2 \frac{P(y \mid x)}{\sum_x P(y \mid x) P(x)}\right]$$

$P(y \mid x)$: channel quality ; $P(x)$: input distribution

§ 2.1 Introduction



- **Channel Capacity**

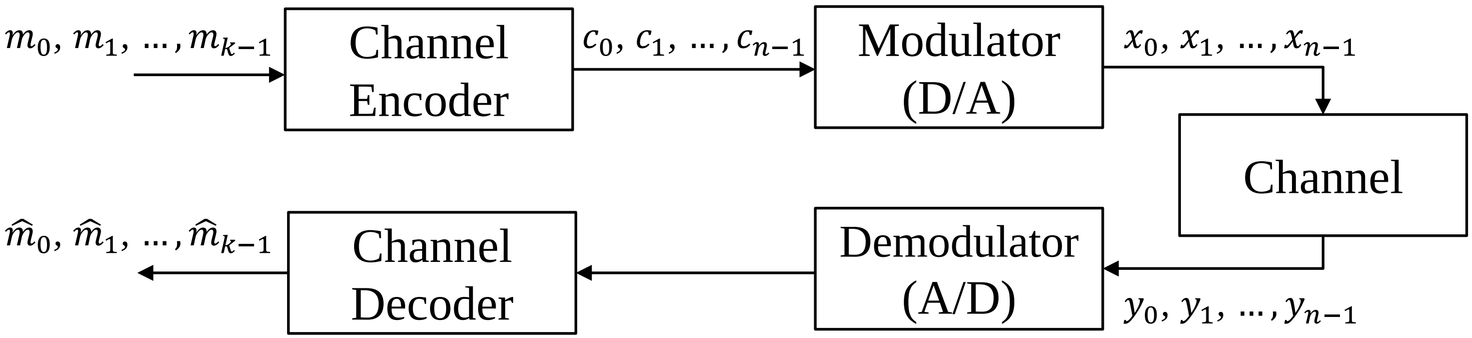
$$C = \max_{P(x)}\{I(X,Y)\}$$

The maximum mutual information $I(X,Y)$ that can be realized over all input distribution $P(x)$.

- In a wireless communication system, it is the maximum number of information bits that can be carried by a modulated symbol such that the information can be recovered with an arbitrarily low probability of error.

§ 2.1 Introduction

- A wireless communication system



- Channel coding is needed to realize this reliable communications
- Given k information symbols (or bits), redundancy is added to obtain n ($n > k$) codeword symbols (or bits), and the coding rate is $r = \frac{k}{n}$.
- Using binary modulation, e.g., BPSK, reliable communications is possible if $r < C$. This property will be proved in Shannon's Channel Coding Theorem (Chapter 4).

§ 2.1 Introduction

- Why input distribution $P(x)$ matters?
- ***Example 2.1:*** Consider the data rate as the human flow measured by capitals per hour (that cross the border from Shenzhen to Hong Kong). There are 200k people wanting to make the cross on the day.

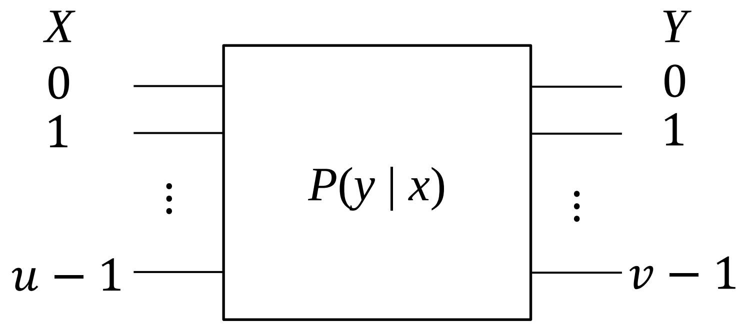
LW's capacity: 10k/hr

SZB' capacity: 30k/hr

LW only	SZB only	{ LW takes 100k, SZB takes 100k	{ LW takes 50k, SZB takes 150k	{ LW takes 60k, SZB takes 140k
20 hrs	6.67 hrs	10 hrs	5 hrs	6 hrs

§ 2.1 Introduction

- Discrete Memoryless Channel (DMC)



- Discrete Input: $X \in \{0, 1, \ldots, u-1\}$

  Discrete Output: $Y \in \{0, 1, \ldots, v-1\}$

- Channel transition probability $P(y|x)$. Note that

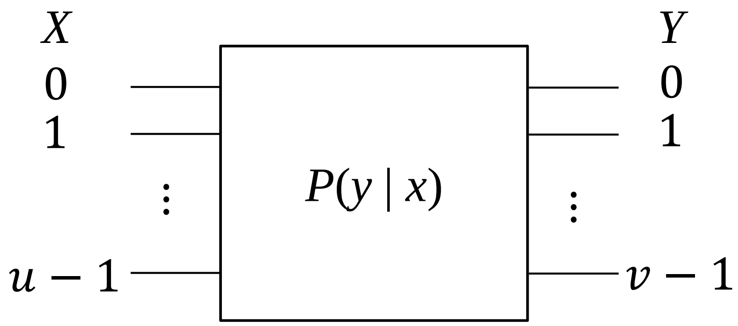
$$P(y|x) \geq 0$$

$$\sum_y P(y|x) = 1, \forall x$$

- Memoryless: Given input and output sequences $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$

$$P(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} P(y_i|x_i)$$

§ 2.1 Introduction

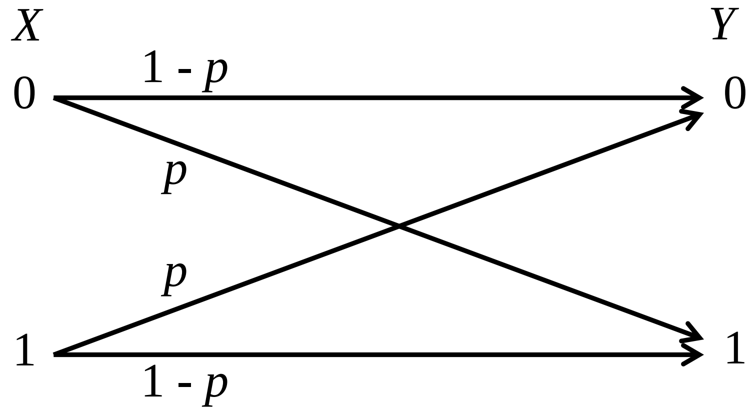


- The channel can often be described by a transition probability matrix

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cccc} Y: \; 0 & 1 & \cdots & v-1 \end{array} \\ \begin{array}{c} X: \; 0 \\ \vdots \\ u-1 \end{array} & \begin{bmatrix} P(0|0) & \cdots & P(v-1|0) \\ P(0|1) & \cdots & P(v-1|1) \\ \vdots & \ddots & \vdots \\ P(0|u-1) & \cdots & P(v-1|u-1) \end{bmatrix} \end{array}$$

- Classical DMCs : BSC, BEC.

§ 2.2 Binary Symmetric Channel (BSC)



- Input: 0 1 0 0 0 1 1 0 1 0 ...

  Output: 0 1 1 1 0 0 1 0 0 0 ...
- Input and output are discrete
- The channel condition

  $P(y=1|x=0) = P(y=0|x=1) = p$

  $P(y=0|x=0) = P(y=1|x=1) = 1-p$

$$P = \begin{array}{c|cc} X \ddots Y & 0 & 1 \\ 0 & 1-p & p \\ 1 & p & 1-p \end{array}$$

- It is the simplest wireless communication channel model.

§ 2.2 Binary Symmetric Channel (BSC)

- Analytic intuition

  $I(X,Y) = H(Y) - H(Y|X)$

  $I(X,Y)$ will be maximized if $H(Y)$ is maximized and $H(Y|X)$ is minimized.

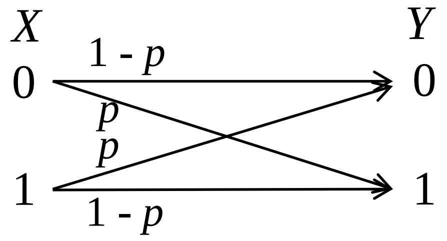
- Analysis

  $H(Y) \leq 1.$

  $$H(Y|X) = -\sum_{x\in X}\sum_{y\in Y} P(x,y)\log_2 P(y|x)$$
  $$= -\sum_{x\in X}\sum_{y\in Y} P(y|x)P(x)\log_2 P(y|x)$$
  $$= -P(x=0)\sum_{y\in\{0,1\}} P(y|x=0)\log_2 P(y|x=0)$$
  $$-P(x=1)\sum_{y\in\{0,1\}} P(y|x=1)\log_2 P(y|x=1)$$
  $$= -P(x=0)((1-p)\log_2(1-p) + p\log_2 p)$$
  $$-P(x=1)(p\log_2 p + (1-p)\log_2(1-p))$$
  $$= -(1-p)\log_2(1-p) - p\log_2 p \text{ bits/sym.}$$

  When $P(x=0) = P(x=1) = \frac{1}{2}$, $H(Y) = 1$ and

  $C = 1 - H(Y|X)$ bits/sym.

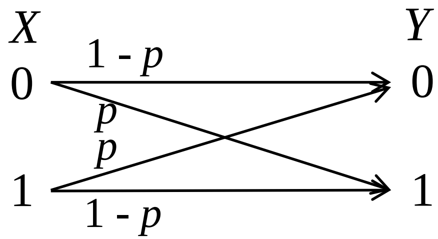


§ 2.2 Binary Symmetric Channel (BSC)

- Intuition: If 0 and 1 experience the same degree of channel impairment, i.e., $P(y=1|x=0)=P(y=0|x=1)$, there is no need to prioritize either 0 or 1 for transmission and $P(x=0)=P(x=1)=\frac{1}{2}$.
- If $P(x=0)=P(x=1)=\frac{1}{2}$, $P(y=0)=P(y=1)=\frac{1}{2}$ and $H(Y)=1$.

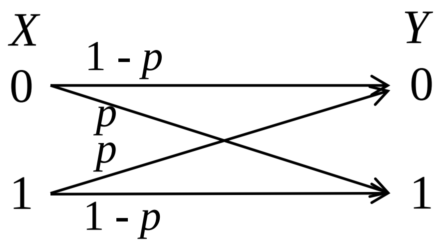
$$
\begin{aligned}
H(Y|X) &= -P(y=0|x=0)\cdot\frac{1}{2}\cdot\log_2 P(y=0|x=0)\\
&= -P(y=1|x=0)\cdot\frac{1}{2}\cdot\log_2 P(y=1|x=0)\\
&= -P(y=0|x=1)\cdot\frac{1}{2}\cdot\log_2 P(y=0|x=1)\\
&= -P(y=1|x=1)\cdot\frac{1}{2}\cdot\log_2 P(y=1|x=1)\\
&= -p\log_2 p-(1-p)\log_2(1-p)\ \text{bits/sym.}
\end{aligned}
$$

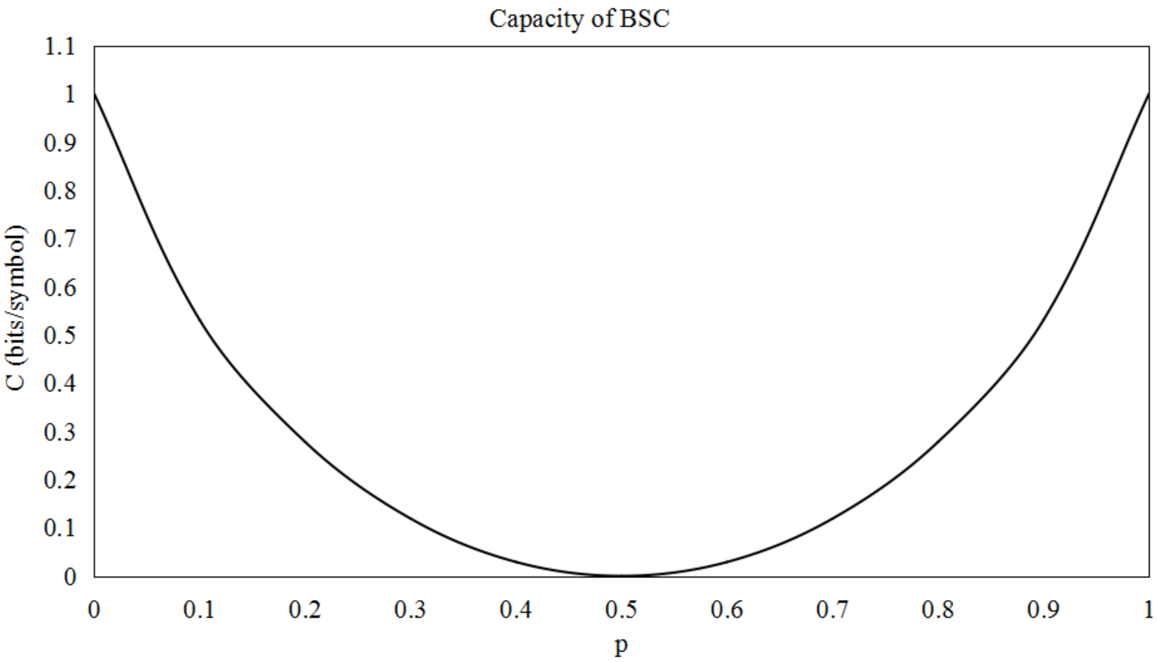
- $C=1+p\log_2 p+(1-p)\log_2(1-p)$ bits/sym.



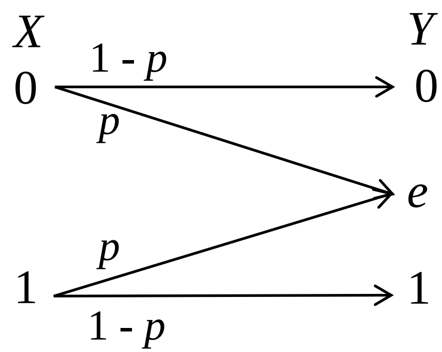
§ 2.2 Binary Symmetric Channel (BSC)

- $C = 1 + p\log_2 p + (1-p)\log_2(1-p)$ bits/symbol





§ 2.3 Binary Erasure Channel (BEC)



- Input: 1 0 0 1 0 1 0 1 0 0 …

  Output: 1 *e* 0 1 0 *e* 0 *e* 0 *e* …

- The channel condition

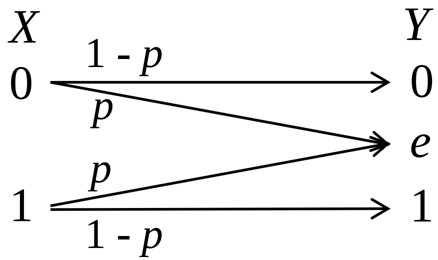
  $P(y = e|x = 0) = P(y = e|x = 1) = p$

  $P(y = 0|x = 0) = P(y = 1|x = 1) = 1 - p$

$$P = \begin{array}{c} X \ddots Y \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & e \end{array} \\ \begin{bmatrix} 1-p & 0 & p \\ 0 & 1-p & p \end{bmatrix} \end{array}$$

- It is a channel model often used in computer networks. Data packets are either perfectly received or lost.

§ 2.3 Binary Erasure Channel (BEC)



- Similar to the analytic intuition of BSC, channel capacity is reached when $P(x=0) = P(x=1) = \frac{1}{2}$.
- $C = H(Y) - H(Y|X)$ bits/sym.
- Since $P(y=0) = P(y=1) = \frac{1}{2}(1-p)$ and $P(y=e) = p$

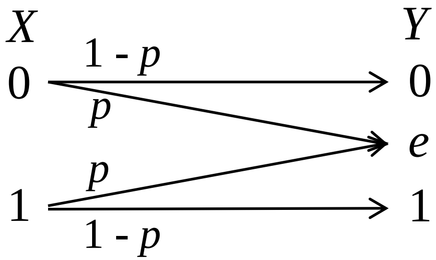
$$H(Y) = -\sum_{y \in Y} P(y) \log_2 P(y)$$

$$= -P(y=0)\log_2 P(y=0) - P(y=e)\log_2 P(y=e) - P(y=1)\log_2 P(y=1)$$

$$= -\frac{1}{2}(1-p)\log_2 \frac{1}{2}(1-p) - p\log_2 p - \frac{1}{2}(1-p)\log_2 \frac{1}{2}(1-p)$$

$$= -(1-p)\log_2 \frac{1}{2}(1-p) - p\log_2 p \text{ bits/sym.}$$

§ 2.3 Binary Erasure Channel (BEC)



- $H(Y|X) = -\sum_{x\in X}\sum_{y\in Y} P(y|x)P(x)\log_2 P(y|x)$

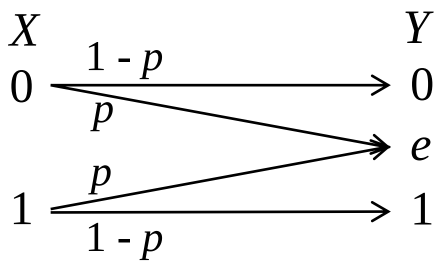
$$= -P(y=0|x=0)\cdot\frac{1}{2}\cdot\log_2 P(y=0|x=0)$$
$$-P(y=1|x=1)\cdot\frac{1}{2}\cdot\log_2 P(y=1|x=1)$$
$$-P(y=e|x=0)\cdot\frac{1}{2}\cdot\log_2 P(y=e|x=0)$$
$$-P(y=e|x=1)\cdot\frac{1}{2}\cdot\log_2 P(y=e|x=1)$$
$$= -(1-p)\log_2(1-p) - p\log_2 p \text{ bits/sym.}$$

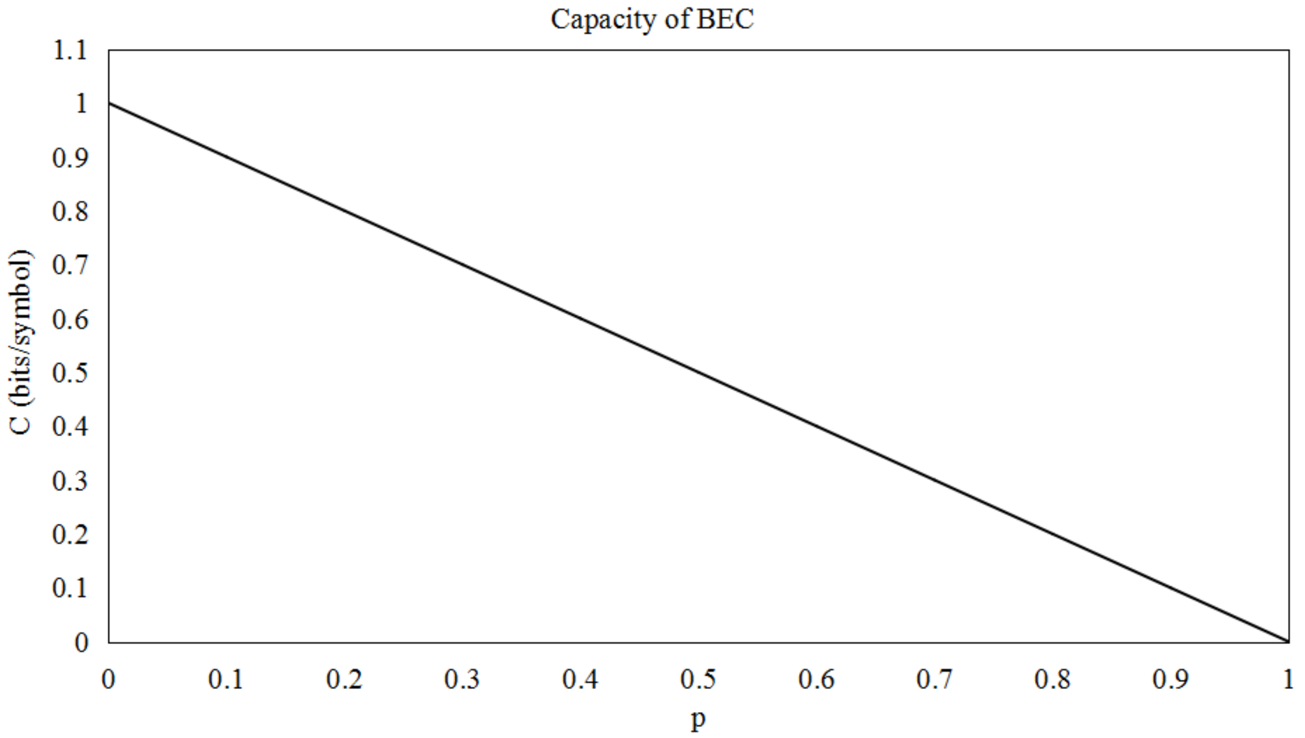
- $C = H(Y) - H(Y|X)$

$$= 1 - p \text{ bits/sym.}$$

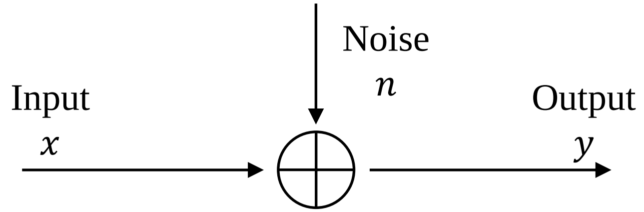
§ 2.3 Binary Erasure Channel (BEC)

- $C = 1 - p$ bits/sym.





§ 2.4 AWGN Channel



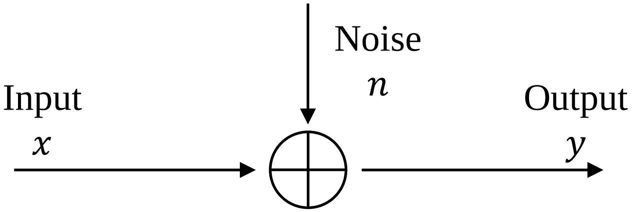
- The additive white Gaussian noise (AWGN) channel model $y = x + n$

  x: discrete input signal, a modulated signal

  n: white Gaussian noise as $\mathcal{N}(0, \sigma_N^2)$, independent of x

  y: continuous output signal, a variation of x
- It is a more realistic wireless channel model where the transmitted signal is impaired by noise.
- It is adopted to represent the space communication channel where light-of-sight (LoS) transmission can be ensured.
- It is also often used as a common platform for evaluating channel codes.

§ 2.4 AWGN Channel



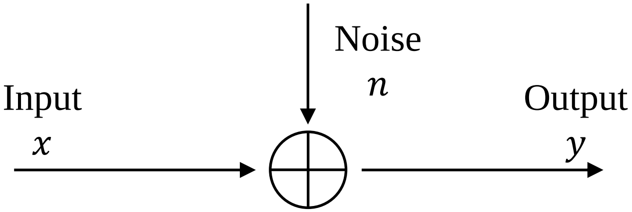
- Channel model $y = x + n$
- Mutual Information: $I(X,Y) = H(Y) - H(Y|X)$

$$\begin{aligned} &= H(Y) - H(X+N|X) \\ &= H(Y) - H(N|X) \\ &= H(Y) - H(N) \end{aligned}$$

- Capacity: $C = \max_{P(x)}\{I(X,Y)\}$

$$= \max_{P(x)}\{H(Y) - H(N)\}$$

§ 2.4 AWGN Channel



- For AWGN, $N\colon \mathcal{N}(0,\sigma_N^2)$, its pdf is

$$P(n)=\frac{1}{\sqrt{2\pi}\sigma_N}\exp\left(-\frac{n^2}{2\sigma_N^2}\right)$$

$$H(N)=-\int_{-\infty}^{+\infty}P(n)\log_2 P(n)\mathrm{d}n$$

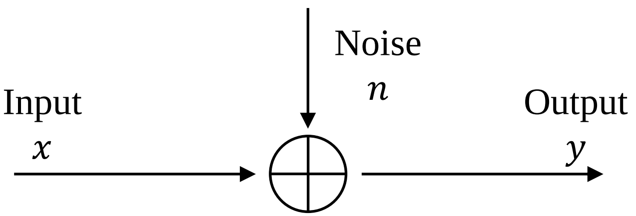
$$=-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_N}\exp\left(-\frac{n^2}{2\sigma_N^2}\right)\log_2\left(\frac{1}{\sqrt{2\pi}\sigma_N}\exp\left(-\frac{n^2}{2\sigma_N^2}\right)\right)\mathrm{d}n$$

$$=\frac{1}{2}\log_2(2\pi e\sigma_N^2)\ \text{bits/sym.}$$

- If input X is a continuous signal and Gaussian distributed as $\mathcal{N}(\mu_X,\sigma_X^2)$, $I(X,Y)$ will be maximized and

$$C=H(Y)-H(N)$$

§ 2.4 AWGN Channel



- For input $X : \mathcal{N}(\mu_X, \sigma_X^2)$, its pdf is

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)$$

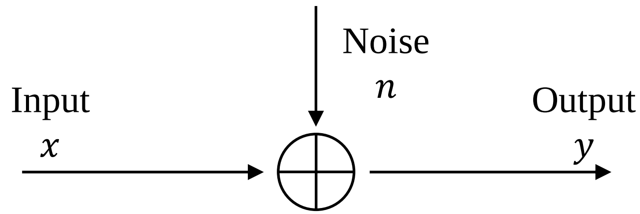
$$\begin{aligned} H(X) &= -\int_{-\infty}^{+\infty} P(x)\log_2 P(x)\mathrm{d}x \\ &= -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)\log_2\left(\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)\right)\mathrm{d}x \\ &= \frac{1}{2}\log_2(2\pi e\sigma_X^2) \ \text{bits/sym.} \end{aligned}$$

- Since $Y = X + N$ and X and N are independent

  Output $Y : \mathcal{N}(\mu_X, \sigma_X^2 + \sigma_N^2) = \mathcal{N}(\mu_X, \sigma_Y^2)$

$$H(Y) = \frac{1}{2}\log_2(2\pi e(\sigma_X^2 + \sigma_N^2)) \ \text{bits/sym.}$$

§ 2.4 AWGN Channel



- Channel model: $y = x + n$
- Capacity: $$\begin{aligned} C &= H(Y) - H(N) \\ &= \frac{1}{2}\log_2\left(2\pi e(\sigma_X^2 + \sigma_N^2)\right) - \frac{1}{2}\log_2(2\pi e \sigma_N^2) \\ &= \frac{1}{2}\log_2\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \text{bits/sym.} \end{aligned}$$
- σ_X^2 is the power of the transmitted signal, while σ_N^2 is the power of noise. Hence, $\frac{\sigma_X^2}{\sigma_N^2}$ is often defined as the channel signal-to-noise ratio (SNR).
- This only defines the unachievable transmission limit since in a practical communication system, X will not be normal distributed.

§ 2.4 AWGN Channel

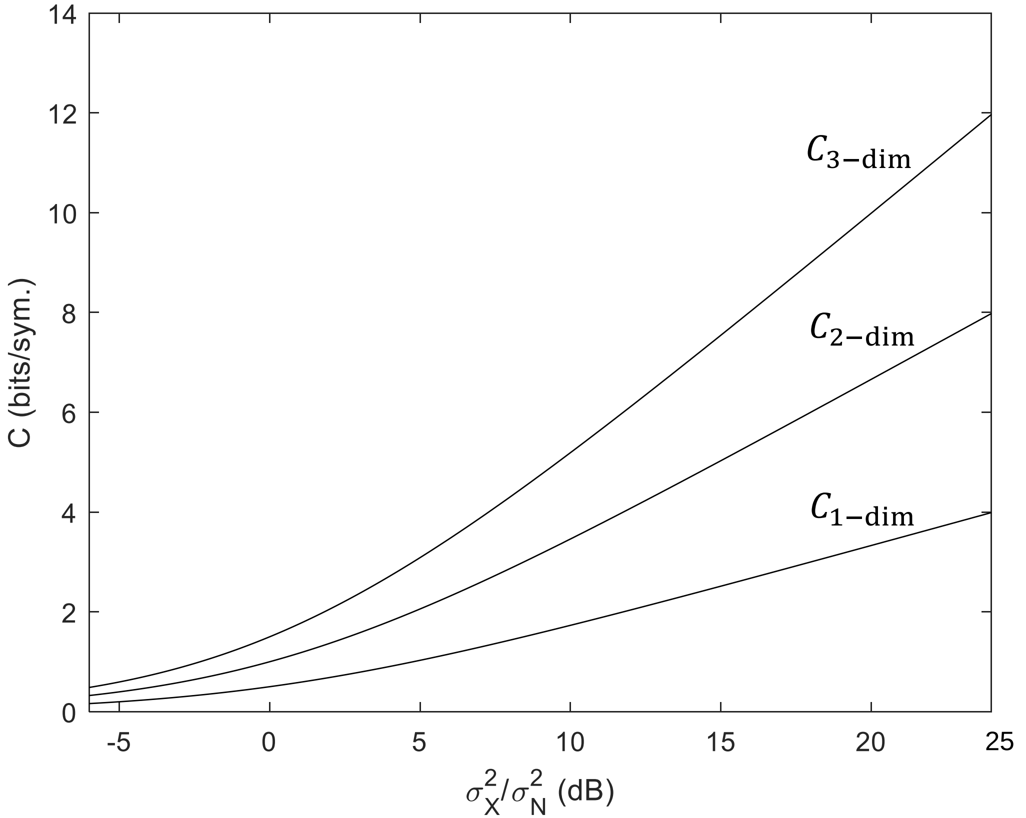
- Transmission resource (dimension) in an electromagnetic wave: amplitude, frequency, phase.

$$C_{1-\text{dim}} = \frac{1}{2} \cdot \log_2\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \text{ bits/sym.}$$

$$C_{2-\text{dim}} = 1 \cdot \log_2\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \text{ bits/sym.}$$

$$C_{3-\text{dim}} = \frac{3}{2} \cdot \log_2\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \text{ bits/sym.}$$

Note: $\frac{\sigma_X^2}{\sigma_N^2}$ in dB $= 10\log_{10}\frac{\sigma_X^2}{\sigma_N^2}$



§ 2.4 AWGN Channel

- <u>Band Limited AWGN Channel</u>
- In a practical system, sampling is needed at the receiver to reconstruct the received signal as Fig. 1.

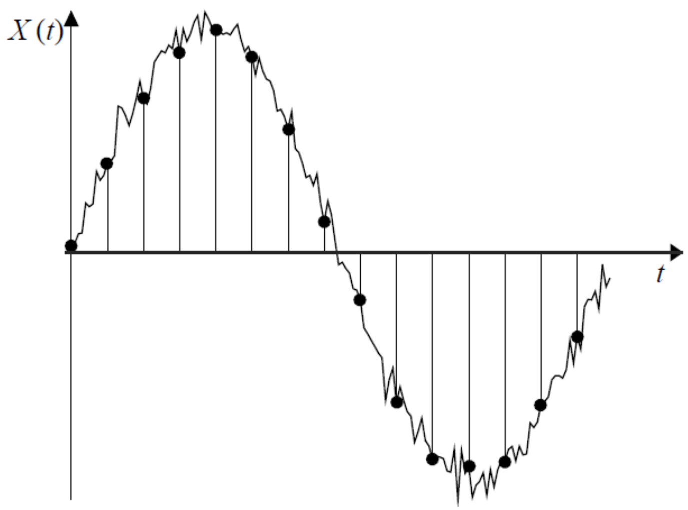


Fig. 1 Received Signal and Sampling

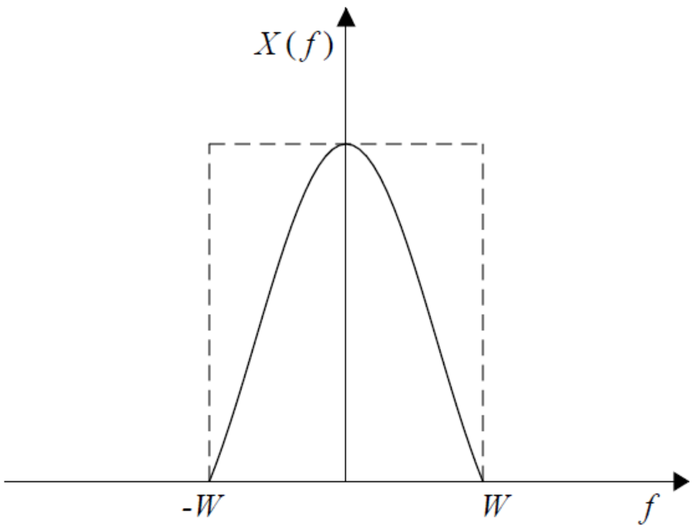


Fig. 2 Signal Sampling in frequency domain

- If the signal has a frequency of W, the sampling frequency should be at least $2W$ for perfect signal reconstruction, as in Fig. 2.

§ 2.4 AWGN Channel

- <u>Band Limited AWGN Channel</u>
- With the sampling, we now have a series of time discrete Gaussian samples and the channel model becomes

$$y\left(t=\frac{s}{2W}\right)=x\left(t=\frac{s}{2W}\right)+n\left(t=\frac{s}{2W}\right), s=1,2,\cdots$$

- Signal $x\left(t=\frac{s}{2W}\right)$ has variance σ_X^2

  Noise $n\left(t=\frac{s}{2W}\right)$ has variance σ_N^2

- Capacity for each time discrete Gaussian channel

$$C_s=\frac{1}{2}\log_2\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right) \quad \text{bits/sample}$$

§ 2.4 AWGN Channel

- <u>Band Limited AWGN Channel</u>
- σ_N^2: Power spectral density of noise samples
- N_0: Power spectral density of noise symbols
- σ_X^2: Power of signal samples
- E: Power of signal symbols
- Over a period of T

$$\sigma_N^2 \cdot 2W \cdot T = N_0 \cdot W \cdot T$$

$$\Downarrow$$

$$\boxed{\sigma_N^2 = \frac{N_0}{2}}$$

$$\sigma_X^2 \cdot 2W \cdot T = E \cdot T$$

$$\Downarrow$$

$$\boxed{E = 2W\sigma_X^2}$$

§ 2.4 AWGN Channel

- Band Limited AWGN Channel
- Since

$$C_s = \frac{1}{2}\log_2\left(1 + \frac{E}{WN_0}\right) \quad \text{bits/sample}$$

- Capacity of this band limited AWGN channel can be determined by

$$C = \frac{\sum_{s=1}^{2WT} C_s}{T}, \; T\text{–sampling duration}$$

$$C = \frac{2WT \cdot \frac{1}{2}\log_2\left(1 + \frac{E}{WN_0}\right)}{T} = W\log_2\left(1 + \frac{E}{WN_0}\right) \quad \text{bits/sec.}$$

§ 2.4 AWGN Channel

- ***Example 2.2 (Shannon Limit):*** Error free transmission over the Gaussian channel is possible if the signal-to-noise ratio $\frac{E_b}{N_0}$ is at least -1.6 dB.

Proof:
- This possibility is sealed by the use of a channel code (information length k, codeword length n).
- Let E_b and E_c denote the energy of each information symbol and each coded symbol, respectively. It is required

$$k \cdot E_b = n \cdot E_c$$

such that adding redundancy does not increase the transmission energy.

- Consider the binary modulation (BPSK) with a modulated symbol energy of E, e.g.,

$$E = E_c = \frac{E_b \cdot k}{n} = E_b \cdot r$$

§ 2.4 AWGN Channel

Continue the Proof

- Assume the signal frequency $W \to \infty$

$$C = \lim_{W \to \infty} W \log_2 \left(1 + \frac{E}{N_0 W}\right)$$
$$= \frac{E}{N_0 \ln 2}$$
$$= \frac{E_b \cdot r}{N_0 \ln 2} \text{ bits/sec.}$$

- For error free transmission, it is required

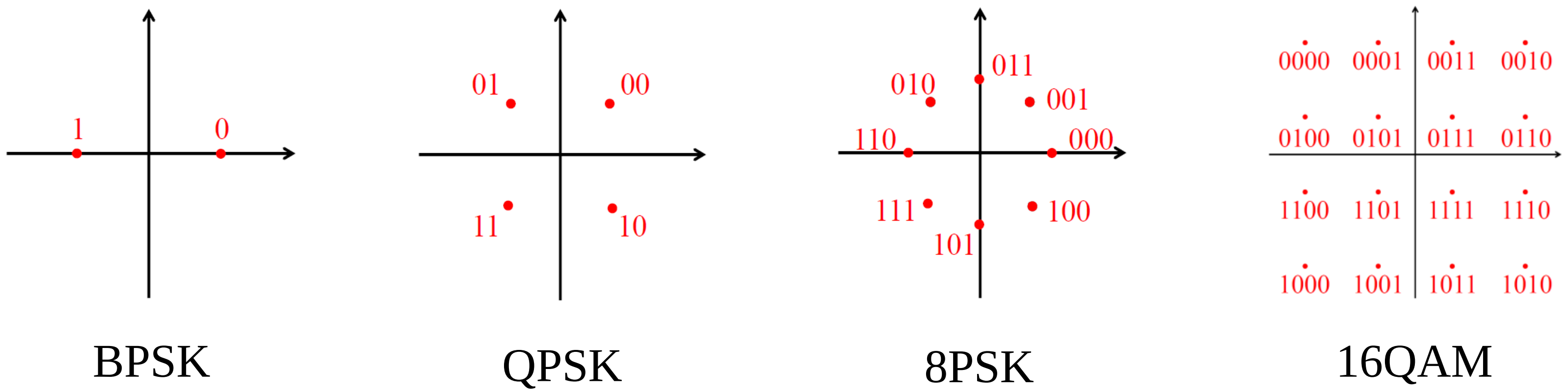
$$r < C \Longrightarrow \frac{E_b}{N_0} > \ln 2 = 0.69$$

- The pinch-off SNR would be

$$\mathrm{SNR_{off}} = 10 \log_{10} 0.69 = -1.6 \text{ dB}$$

§ 2.4 AWGN Channel

- Finite Modulation Alphabets
- In a wireless communication system, digital signals are modulated (mapped) to analog signals for transmission.
- Commonly used modulation schemes include:



- The AWGN channel has finite inputs but continuous outputs.

§ 2.4 AWGN Channel

- <u>Finite Modulation Alphabets</u>
- Input $X \in \{x_1, x_2, \ldots, x_M\}$, e.g., BPSK M = 2, QPSK M = 4, 16QAM M = 16.
- Output $y = x + n$, where n is AWGN.
- Channel Capacity

$$C = \max_{P(x_i)} \left\{ \sum_{i=1}^{M} \int_{y:-\infty}^{+\infty} P(x_i, y) \log_2 \frac{P(y|x_i)}{P(y)} \mathrm{d}y \right\}$$

Since

$$P(x_i, y) = P(y|x_i)P(x_i) \quad \text{and} \quad P(y) = \sum_{i'=1}^{M} P(y|x_{i'})P(x_{i'})$$

$$C = \max_{P(x_i)} \left\{ \sum_{i=1}^{M} \int_{y:-\infty}^{+\infty} P(y|x_i)P(x_i) \log_2 \frac{P(y|x_i)}{P(y)} \mathrm{d}y \right\}$$

§ 2.4 AWGN Channel

- Finite Modulation Alphabets

$$C = \max_{P(x_i)} \left\{ \sum_{i=1}^{M} P(x_i) \int_{y:-\infty}^{+\infty} P(y|x_i) \log_2 \frac{P(y|x_i)}{\sum_{i'=1}^{M} P(x_{i'}) P(y|x_{i'})} \mathrm{d}y \right\}$$

- Assume each modulated symbol is equally likely to be transmitted

$$P(x_i) = P(x_{i'}) = \frac{1}{M}.$$

- Capacity:

$$C = \frac{1}{M} \sum_{i=1}^{M} \int_{y:-\infty}^{+\infty} P(y|x_i) \log_2 \frac{P(y|x_i)}{\frac{1}{M} \sum_{i'=1}^{M} P(y|x_{i'})} \mathrm{d}y \quad \text{bits/sym.}$$

§ 2.4 AWGN Channel

- <u>Finite Modulation Alphabets</u>
- Over the AWGN Channel with noise power σ_N^2

$$P(y|x_i) = \frac{1}{\sqrt{2\pi}\sigma_N}\exp\left(-\frac{|y-x_i|^2}{2\sigma_N^2}\right)$$

and likewise for $P(y|x_{i'})$.

- Capacity:

$$C = \frac{1}{M}\sum_{i=1}^{M}\int_{y:-\infty}^{+\infty} P(y|x_i)\log_2\frac{P(y|x_i)}{\frac{1}{M}\sum_{i'=1}^{M}P(y|x_{i'})}\mathrm{d}y$$

$$= \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\left[\log_2\frac{P(y|x_i)}{\frac{1}{M}\sum_{i'=1}^{M}P(y|x_{i'})}\right]$$

§ 2.4 AWGN Channel

Continue

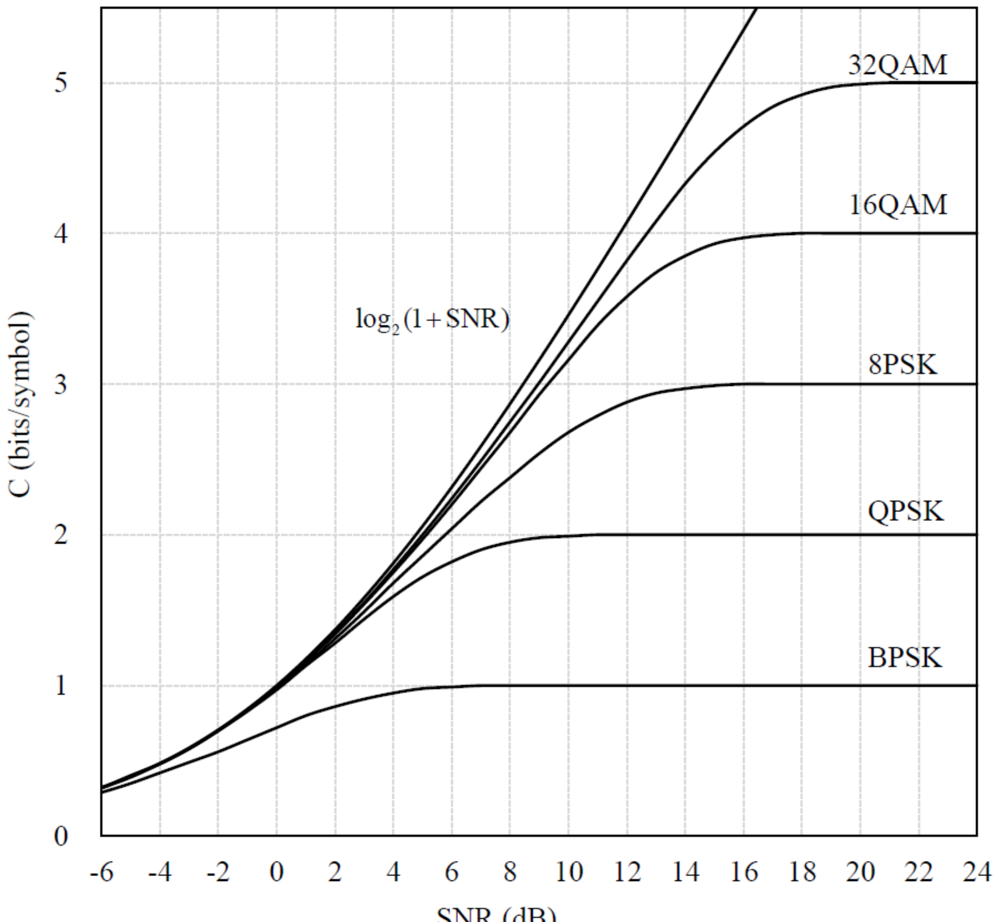
$$= \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}\left[\log_2 M + \log_2 \frac{P(y|x_i)}{\sum_{i'=1}^{M} P(y|x_{i'})}\right]$$

$$= \log_2 M - \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}\left[\log_2 \frac{\sum_{i'=1}^{M} P(y|x_{i'})}{P(y|x_i)}\right]$$

$$= \log_2 M - \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}\left[\log_2 \sum_{i'=1}^{M} \exp\left(\frac{|y-x_i|^2 - |y-x_{i'}|^2}{2\sigma_N^2}\right)\right] \quad \text{bits/sym.}$$

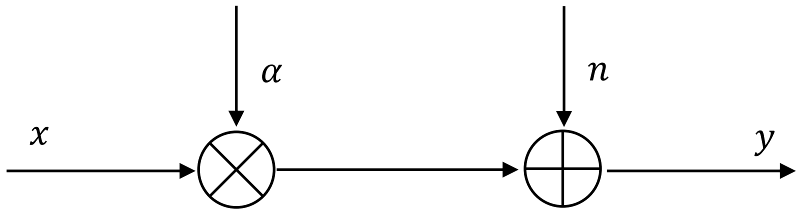
§ 2.4 AWGN Channel

- Finite Modulation Alphabets



$$\mathrm{SNR} = \frac{\sigma_X^2}{\sigma_N^2}$$

§ 2.5 Fading Channels



- Channel Model: $y = \alpha \cdot x + n$
- Fading coefficients α further represent the effect of signal attenuation, signal scattering, path loss and multi-path accumulation.
- It is a channel model often used for urban communications.
- If α is Rayleigh distributed following $\alpha = |\alpha| \exp(j\varphi)$, it is called the Rayleigh fading channel.
- Fading types:

(1) Fast fading: α changes independently for every x.

(2) Quasi-static fading: α remains unchanged during the transmission of a codeword and changes independently from one codeword to another.

(3) Block fading: α changes independently block by block.

§ 2.5 Fading Channels

- Assume fading coefficients α are known by both the transmitter and receiver.

- Instantaneous capacity:

$$C(\alpha_i) = W\log_2\left(1 + \frac{\alpha_i^2 \cdot E(\alpha_i)}{WN_0}\right)$$

$E(\alpha_i)$: the signal power depending on α_i.

It is the maximal achievable transmission rate defined by a particular fading realization α_i.

- Ergodic Capacity:

$$C = \max_{E(\alpha_i)} \mathbb{E}\left[W\log_2\left(1 + \frac{\alpha_i^2 \cdot E(\alpha_i)}{WN_0}\right)\right]$$

It is the average transmission rate that can be realized over all fading realizations.

References:

[1] Elements of Information Theory, by T. Cover and J. Thomas.

[2] Scriptum for the lectures, Applied Information Theory, by M. Bossert.