Fundamentals of Information Theory

Channel Capacity

Yayu Gao
School of Electronic Information and Communications
Huazhong University of Science and Technology
Email: yayugao@hust.edu.cn

Outline

- Beautiful Mind: Overview of Channel Capacity Analysis
- Theory to Applications: Impact of Channel Capacity Analysis
- Classifications of Channels
- How to define channel capacity in math?
- How to calculate channel capacity?
- How to define channel capacity in operation?
- Why are these two equal?
- Shannon’s second theorem: channel coding theorem
- Channel capacity: from discrete, continuous to analog
- Most famous formula in IT: Shannon Formula

Outline

- Beautiful Mind: Overview of Channel Capacity Analysis
- Theory to Applications: Impact of Channel Capacity Analysis
- Classifications of Channels
- How to define channel capacity in math?
- How to calculate channel capacity?
- How to define channel capacity in operation?
- Why are these two equal?
- Shannon’s second theorem: channel coding theorem
- Channel capacity: from discrete, continuous to analog
- Most famous formula in IT: Shannon Formula

本节学习目标

1. 宏观理解信道容量分析问题的全貌
2. 信道容量分析的理论与应用意义
3. 说出信道的分类
4. 说出信道的定义
5. 说出信道矩阵的意义
6. 写出信道容量的定义式
7. 说出信道容量的物理意义
8. 计算一些简单信道的信道容量
9. 计算对称信道的信道容量

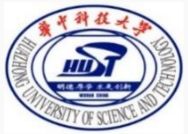
重难点：

- 信道容量问题概览
- 信道容量分析思维方法
- 信道容量定义
- 信道容量计算

01

Beautiful Mind:
Overview of Channel Capacity Analysis

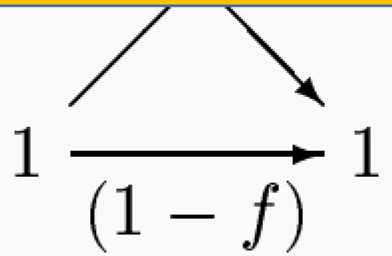
What is communication?



What is communication? Noisy channels…

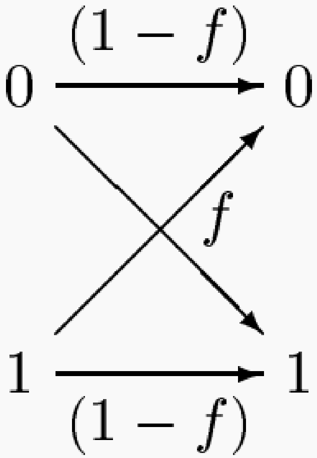
- Physical acts of *A* have induced a desired physical state in *B*.
- Subject to the uncontrollable ambient **noise** and imperfections of the physical **signaling process** itself.

Can we provide reliable communication over unreliable channels?



Binary Symmetric Channel

Questions were asked…



Binary Symmetric Channel

- How to design the codebook to reduce the error probability?
- Is there a tradeoff between the code rate and error probability?
- **What is the maximum code rate R for which arbitrarily small error probability Pe can be achieved?**

Information Theory: Basic theory of communication field

- Key question:
  - Can we transmit information through a noisy channel without error?
  - How much information can be transmitted at most over noisy channel?

- Claude E. Shannon, "A Mathematical Theory of Communications," Bell System Technical Journal, July & October 1948.

- One of the most important contributions
  - **Channel capacity analysis**

Claude Shannon

Overview of Channel Capacity Analysis

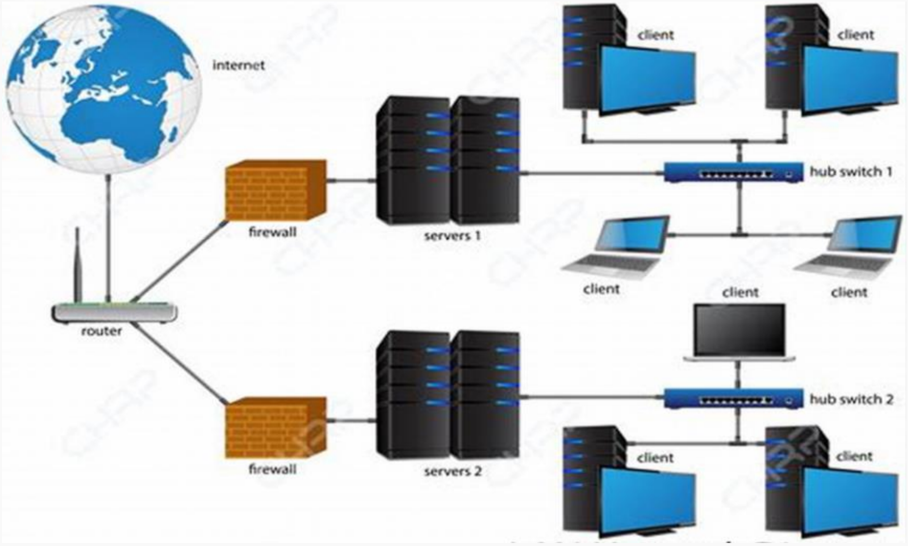
- A unified abstraction model of communication systems

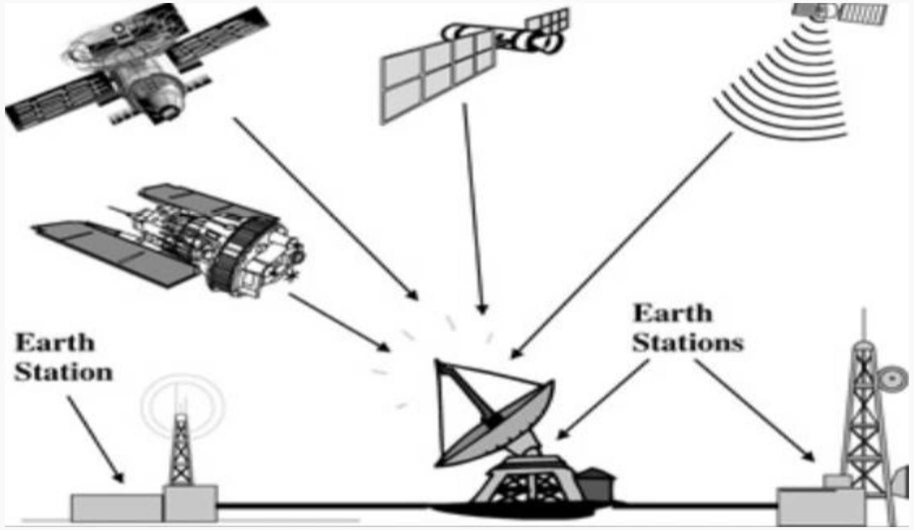
- Definition of the objective: What is channel?

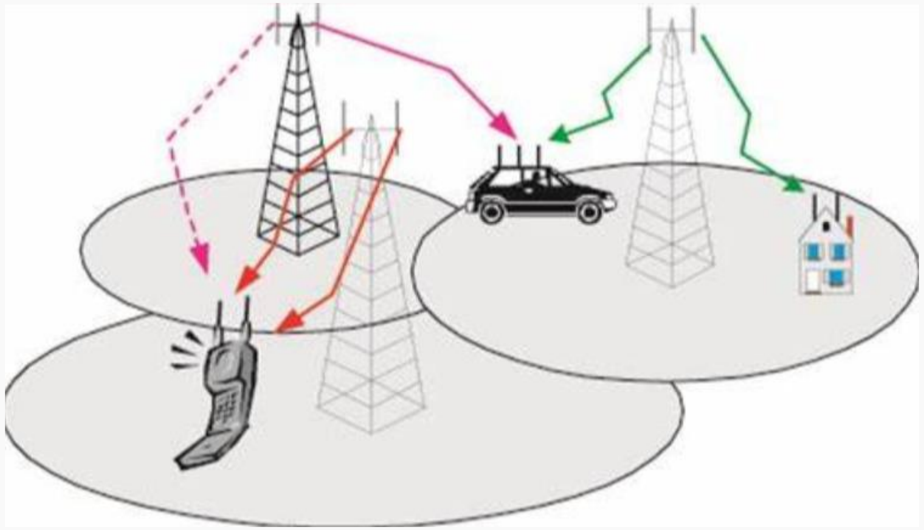
- Definition of channel capacity
  - Mathematical meaning (theory)
  - Operational meaning (applications)

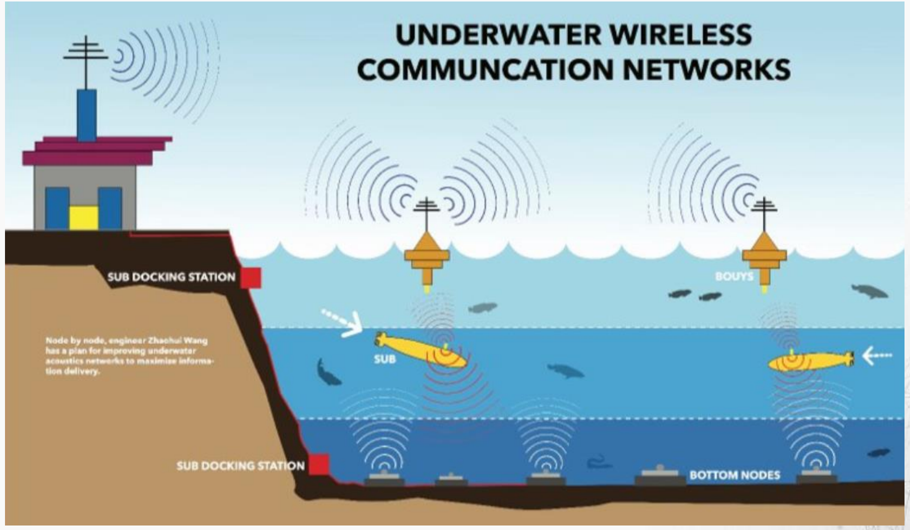
- Channel capacity analysis
  - Channel coding theorem

Varied communication systems…



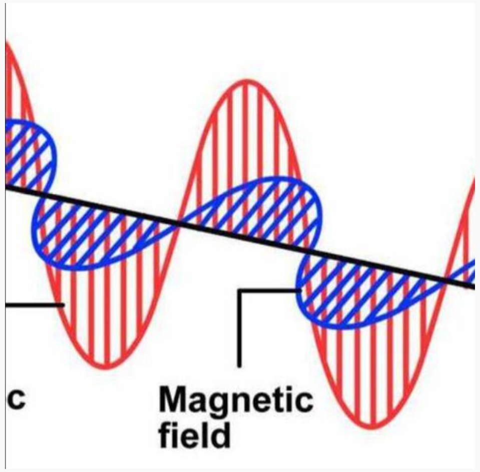




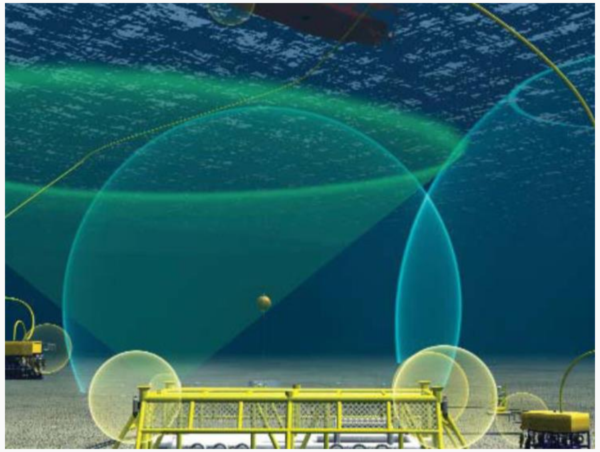


Varied communication channels…

Fiber



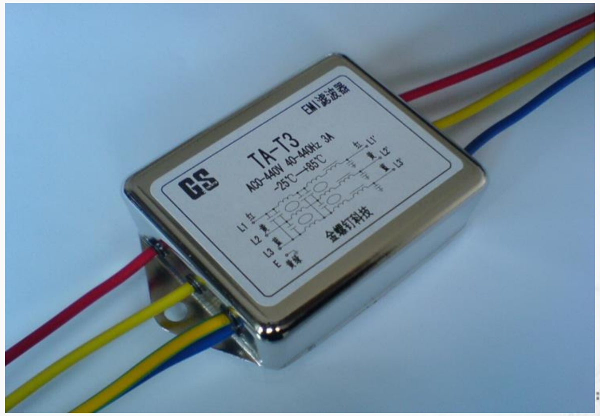
Electromagnetic wave



Water

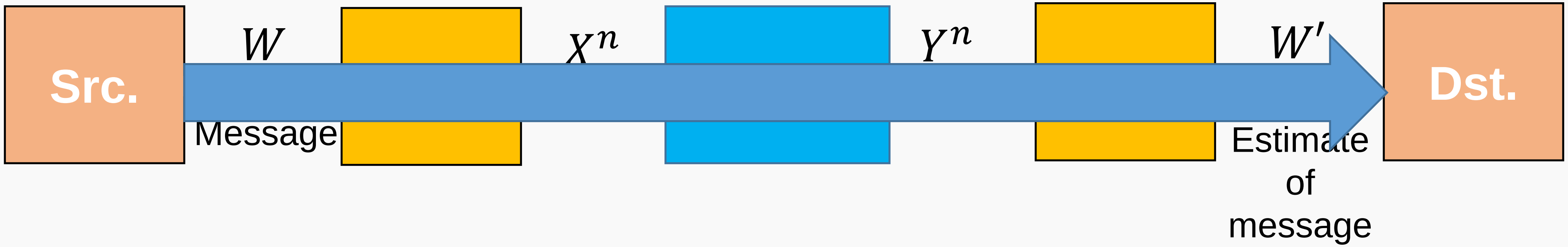
u-disk

CD



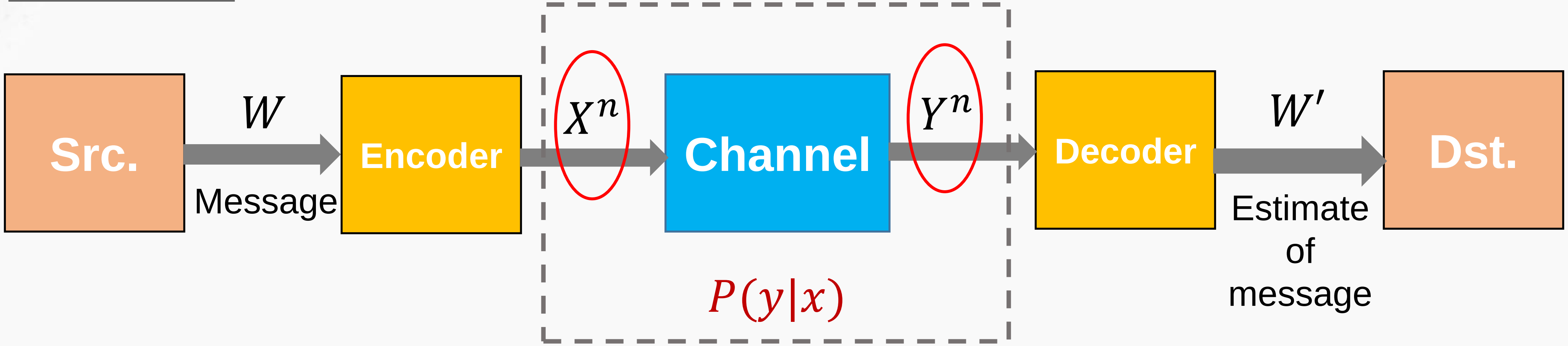
Filter

A unified abstraction model of communication systems



- Communication system model
  - Unified abstraction of various communication methods
  - Applied throughout the entire analysis
- Essence of communication in terms of information
  - Reconstruct the source message at the receiver without error

Definition of Objective: What is channel?



- How to describe the characteristics of the channel?
  - Input x, output y; the statistical behavior of a channel can be represented by a matrix consisting of conditional probabilities $P(y|x)$.

- What is channel?
  - A system consisting of an input alphabet and output alphabet and a probability transition matrix

Definition of channel capacity: mathematical meaning

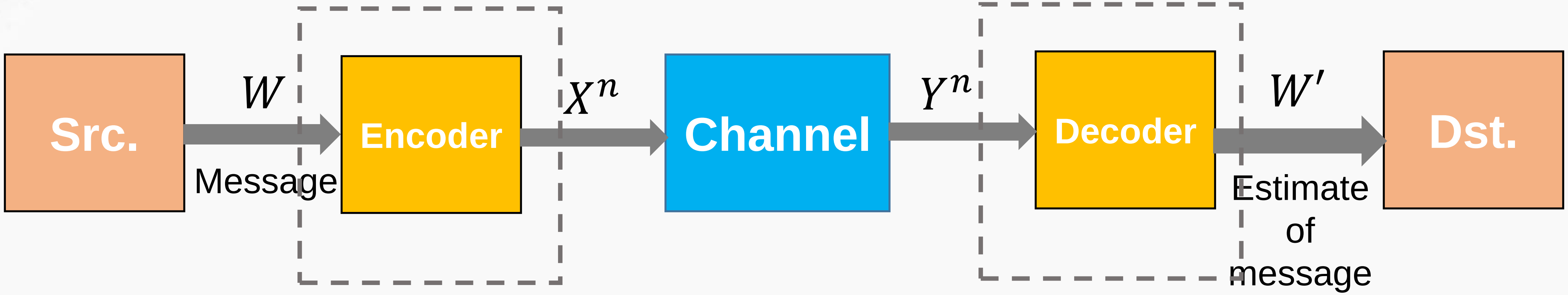
$$X \longrightarrow \boxed{\text{Channel}} \longrightarrow Y$$

$$P(Y|X)$$

- Question: For each usage of the channel, how much information can be transmitted?
- Channel capacity: Maximum mutual information of the channel

Mutual information $I(X;Y) \xrightarrow{\text{Optimization}}$ Channel capacity $\boxed{C = \max_{p(x)} \{I(X;Y)\}}$

Definition of channel capacity: operational meaning



- Provide an encoding/decoding algorithm, which can guarantee lossless data communication

- Channel capacity: the maximum transmission rate for reliable communication over the channel
  - Rates less than capacity yield arbitrarily small probability of error for information transmission.

Channel capacity analysis

Channel capacity: Math meaning

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

=

Channel capacity: operational meaning

maximum transmission rate for reliable communication over the channel

Theory | Easier to obtain | Unrelated to Reliable communication

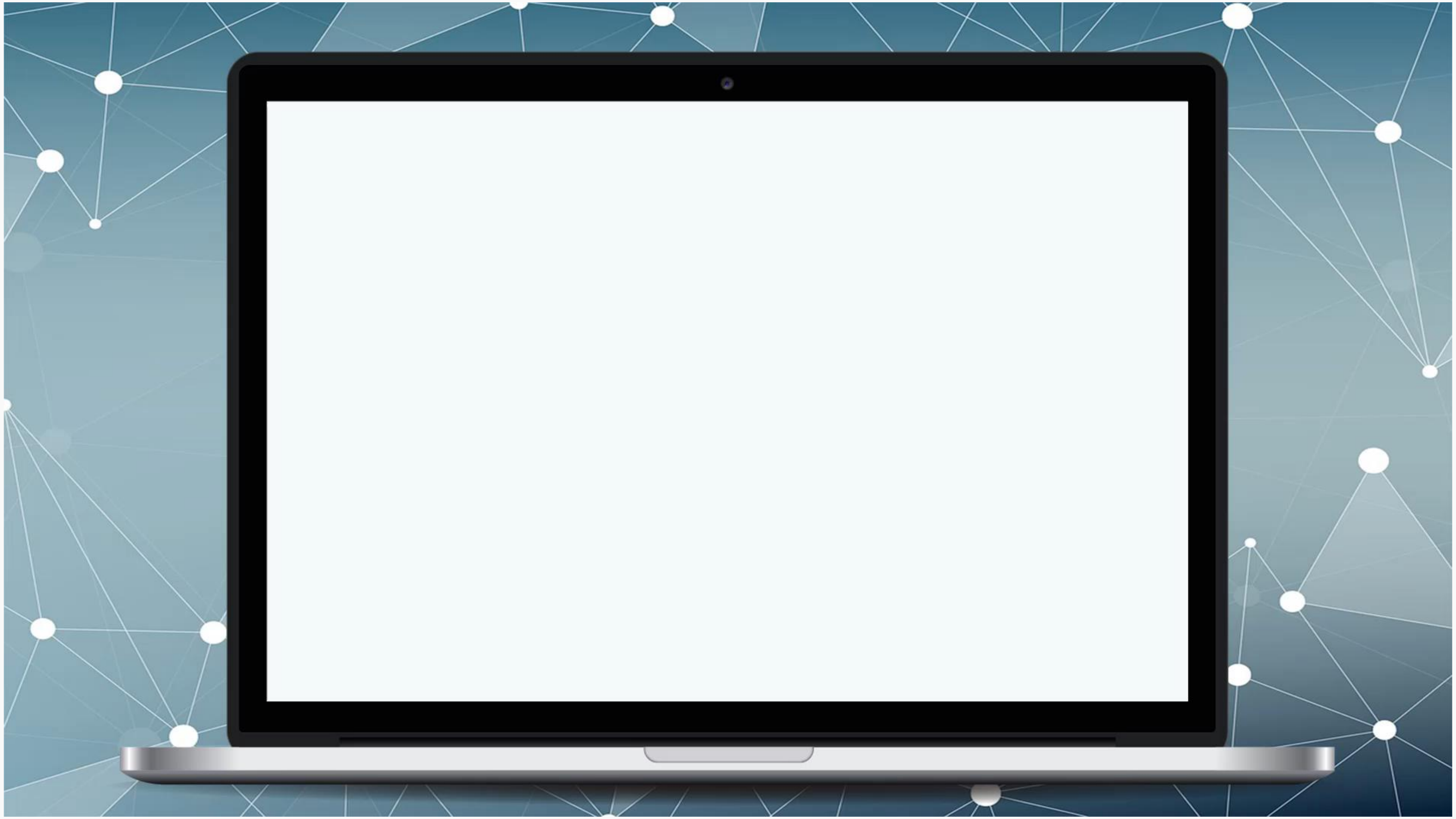
Practice | Difficult to obtain | Achieve Reliable communication

All rates below capacity *C* can achieve reliable communication.

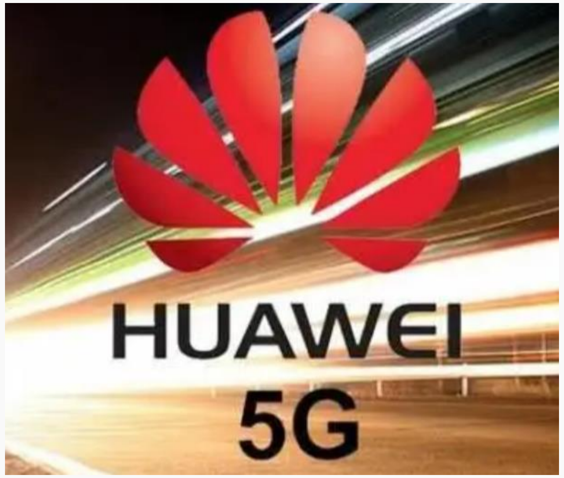
02

Theory to Applications:
Impact of Channel Capacity Analysis

Impact of Channel Capacity Analysis



Insights: Theory and Applications



任正非2020.5.21长篇访谈：5G标准源自土耳其科学家的一篇数学论文

“大家今天讲5G标准对人类社会有多么厉害，怎么会想到，5G标准是源于十多年前土耳其Arikan教授的一篇数学论文？

Arikan教授发表这篇论文两个月后，被我们发现了，我们就开始以这个论文为中心研究各种专利，一步步研究解体，共投入了数千人。

十年时间，我们就把土耳其教授数学论文变成技术和标准。我们的5G基本专利数量占世界27%左右，排第一位。”

Insights: Theory and Applications

中国要踏踏实实在数学、物理、化学、神经学、脑科学……各方面努力去改变，我们可能在这个世界上能站起来

“任正非如是说：以中国为中心建立理论基地要突破美国的重围，眼前这个方式比较难，因为中国在基础理论上不够，这些年好一些了。我曾在全国科学大会上讲了数学的重要性，听说现在数学毕业生比较好分配了。我们有几个人愿意读数学的?

我不是学数学的，我曾经说，我退休以后想找一个好大学，学数学。校长问我，学数学干什么？我说，想研究热力学第二定律。他问，研究用来做什么？我说，想研究宇宙起源。他说，我很欢迎你！但是我到现在还不能退休，还去不了。我们那时是工科学生，学的是高等数学，最浅的数学。”

“一切新产品和新工艺都不是突如其来、自我发育和自我生长起来的。它们皆源自新的科学原理和科学概念。新科学原理和科学概念则必须来自最纯粹科学领域持续不懈的艰难探索。如果一个国家最基础的前沿科学知识依赖他人，其产业进步必然异常缓慢，其产业和世界贸易竞争力必然极其孱弱。”

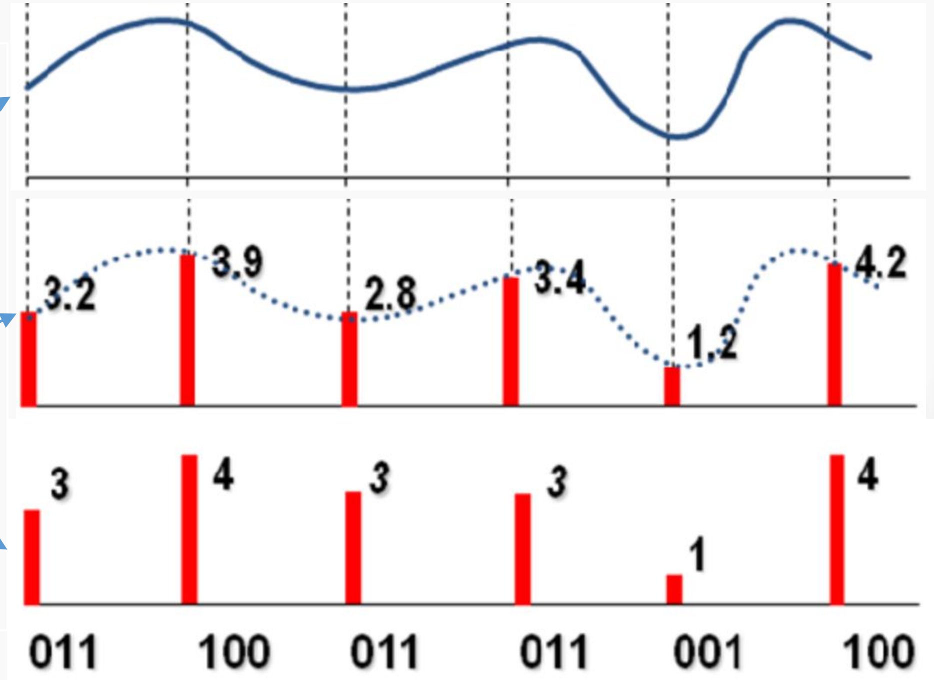
03

Classifications of Channels

Classification of channels

- According to input/output signals and time

Signal	Time	Channel
Continuous	Continuous	Analog channel
Continuous	Discrete	Continuous channel
Discrete	Discrete	Discrete channel
Discrete	Continuous	



Classification of channels

- According to properties of channel stochastic process
  - Memoryless channel: The output at each time *i* is only related to the input at time *i*

$$P(y|x) = P(y_1, y_2, \ldots, y_n | x_1, x_2, \ldots, x_n) = \prod_{i=1}^{N} P(y_i|x_i)$$

  - We often consider a stationary channel: the channel statistics does not vary with time.

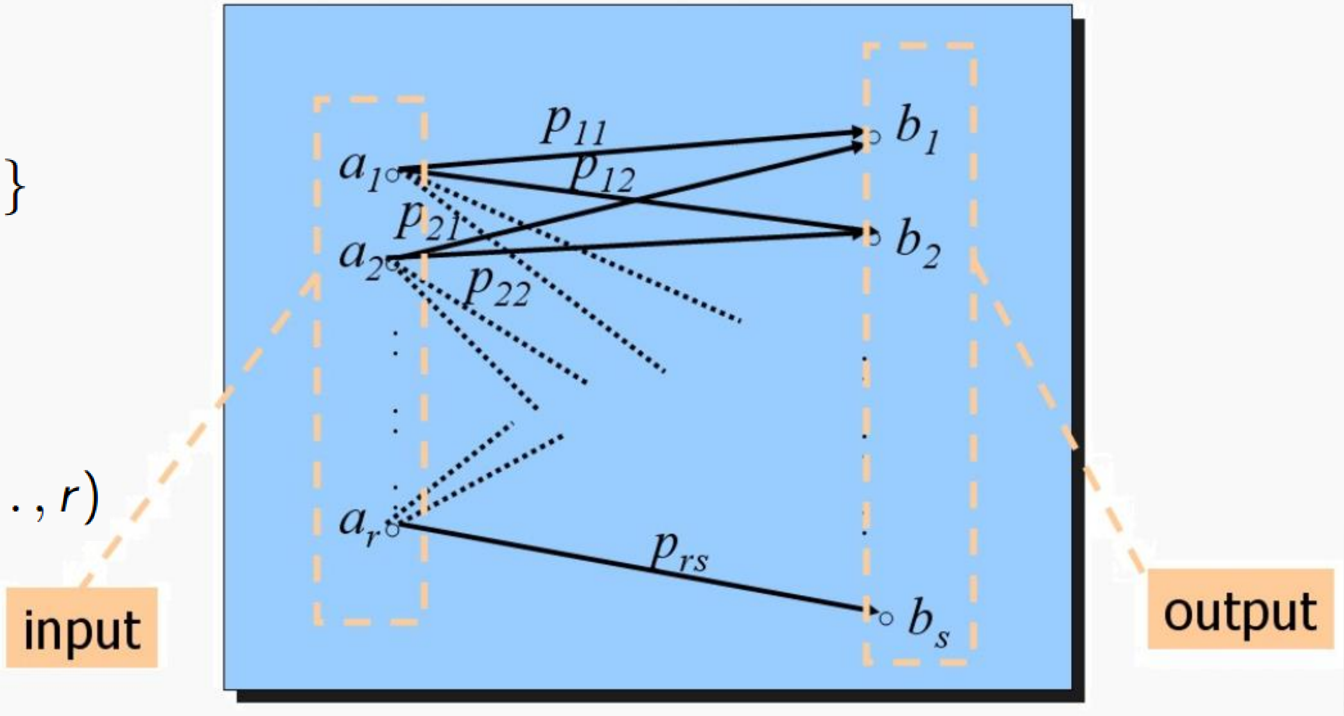
$$P(y|x) = P(y_i|x_i)$$

Channel model: discrete memoryless channel

- Information channel
  - The output is decided by some probability
  - Input alphabet $\mathcal{X}$: the set of input symbols
  - Output alphabet $\mathcal{Y}$: the set of output symbols
- Statistical behavior of an information channel
  - Input $r.v.$ X
  - Output $r.v.$ Y
  - Probability transition (matrix) $P(Y|X)$
- A discrete channel can be denoted by $(\mathcal{X}, p(y|x), \mathcal{Y})$.

Channel matrix

- Assume a channel with
  - Input alphabet $A = \{a_1, a_2, \ldots, a_r\}$
  - Output alphabet $B = \{b_1, b_2, \ldots, b_s\}$
- Channel matrix
  - $p_{ij} = P_r(Y = b_j | X = a_i)$
  - For each row the following holds:
    $p_{i1} + p_{i2} + \ldots + p_{is} = 1 \quad (i = 1, 2, \ldots, r)$



$$T = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \begin{array}{cccc} b_1 & b_2 & \cdots & b_s \end{array} & \\ \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix} & \begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_r \end{array} \end{array}$$

The statistical behavior of the channel is completely defined by the channel matrix.

04

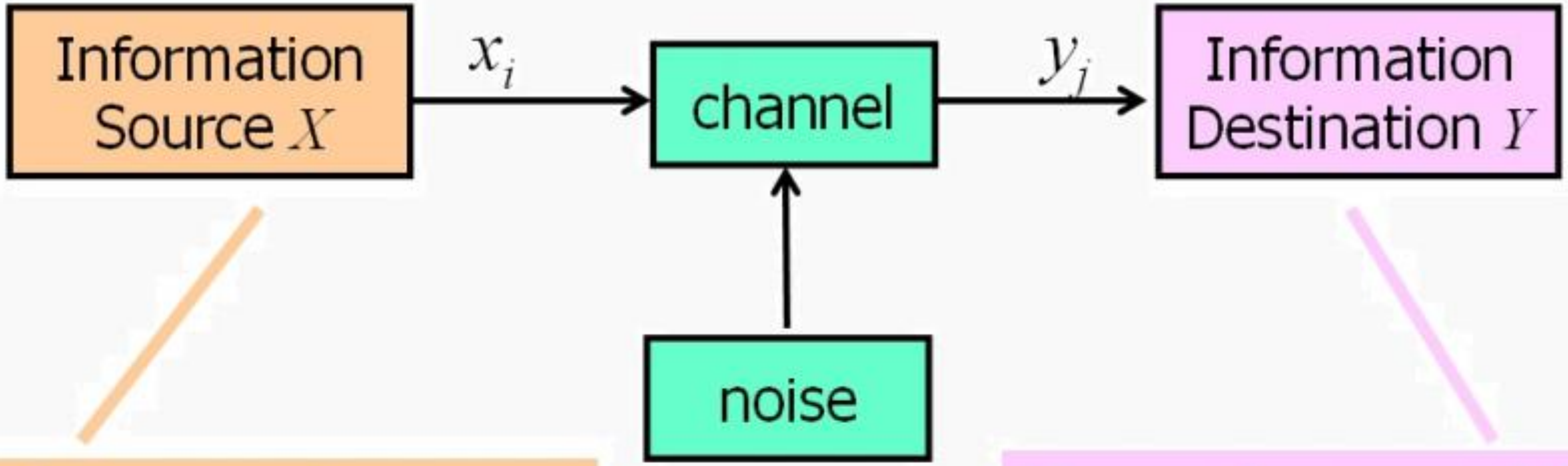
How to define channel capacity in math?

How much information is transmitted from src. to dst.?

Source → channel → Destination

Noise

Review: Entropies in information channel



From the view of source

source information

$= loss + sent$

$$H(X) = H(X|Y) + I(X;Y)$$

From the view of destination

destination information

$= noise + received$

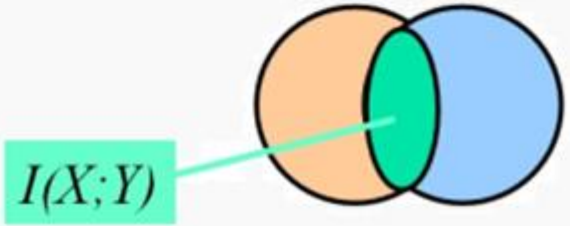
$$H(Y) = H(Y|X) + I(X;Y)$$

- Loss entropy $H(X|Y)$
- Noise entropy $H(Y|X)$

Review: Mutual information

$$X \longrightarrow \boxed{\text{channel}} \longrightarrow Y$$

$$P(Y|X)$$



- Mutual information: information transferred from the source to the destination

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x)p(y|x) \log \left[\frac{p(y|x)}{\sum_x p(x)p(y|x)} \right]$$

- A function of the input $p(x)$ and the channel characteristics $p(y|x)$

Information transmission rate

- Information transmission rate

$$R = \boxed{I(X;Y)} = \boxed{H(X)} - \boxed{H(X|Y)} \quad \text{(bits/symbol)}$$

Information successfully transmitted

Source information

Loss information

- Physical meaning
  - the average information per symbol that can be transmitted through the channel

Channel capacity: definition

- Channel capacity: the maximum information transmission rate.

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

The maximum is taken over ALL possible input distributions.

$$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x)p(y|x) \log \left[\frac{p(y|x)}{\sum_x p(x)p(y|x)} \right]$$

- A function of the input *p(x)* and the channel characteristics *p(y|x)*

Channel capacity: discussions

- Channel capacity: the maximum information transmission rate.

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

- Channel capacity is independent of the input source.
- Channel capacity only depends on the statistical behaviors of the channel.

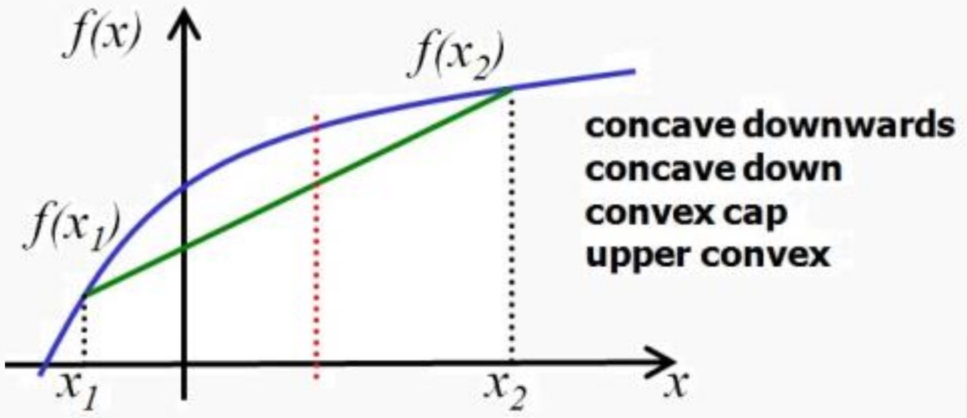
- Physical meaning
  - For each channel usage, how much information can be transmitted as most?
  - The maximum capability of a DMC to convey information.

Channel capacity: concavity of I(X;Y)

- Channel capacity: the maximum information transmission rate.

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

- $I(X;Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(y|x)$ for fixed $p(x)$. (Chapter 2.)
  - The local maximum is a global one.



Channel capacity: properties

- $C \geq 0$
  - since $I(X;Y) \geq 0$.
- $C \leq \log(|X|)$.
  - since $C = \max\{I(X;Y)\} \leq \max\{H(X)\} = \log(|X|)$.
- $C \leq \log(|Y|)$
  - Same to the above.
- $I(X;Y)$ is a continuous function of $p(x)$.
- $I(X;Y)$ is a concave function of $p(x)$.
  - Local maximum is a global one and finite.
  - Convex optimization can be used to find the maximum value C.

05

How to calculate channel capacity?

Discrete memoryless channel: classification

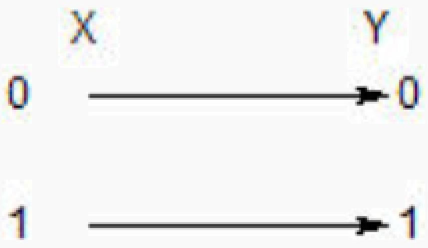
- Discrete channel
  - Single symbol channel
  - Symbol sequence channel

Single symbol channel: classification

- Classified by noise
  - Noiseless channel
    - Noiseless binary channel
  - Noisy channel
    - Noisy channel with non-overlapping
    - Noisy Typewriter
  - Noise and loss
    - Discussions on noise and loss
- Classified by matrix
  - Symmetric channel

Noiseless binary channel

- One error-free bit can be transmitted per channel use.
- Intuitively, channel capacity is 1 bit per channel use.
- Information capacity



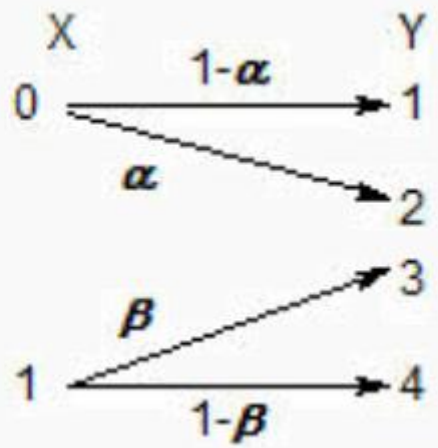
$$C = \max_{p(x)} \{I(X;Y)\} = \max_{p(x)} \left\{ \sum_X \sum_Y p(x,y) \log \left[\frac{p(x,y)}{p(x)p(y)} \right] \right\}$$

$$= \max_{p(x)} \left\{ \sum_X p(x) \log \left[\frac{1}{p(x)} \right] \right\} = \max_{p(x)} H(X) = 1$$

$$\Rightarrow p(x) = \left(\frac{1}{2}, \frac{1}{2} \right)$$

Noisy channel with non-overlapping

- Channels usually have noise.
- Does noise always lead to errors?



- Channel is noisy, therefore the output is uncertain. $(H(Y|X) > 0)$
- Input can be determined from the output. $(H(X|Y) = 0)$
  $\Rightarrow$ Channel capacity is 1 bit per channel use.
- $X \rightarrow Y$: one-to-many mapping, yet the output can be divided into non-overlapping sets.
  $\Rightarrow$ Each column in the channel matrix has one and only one non-zero element.
- Information capacity

$$C = \max_{p(x)}[I(X;Y)] = \max_{p(x)} H(X) - H(X|Y) \Rightarrow p(x) = \left(\frac{1}{2}, \frac{1}{2}\right)$$

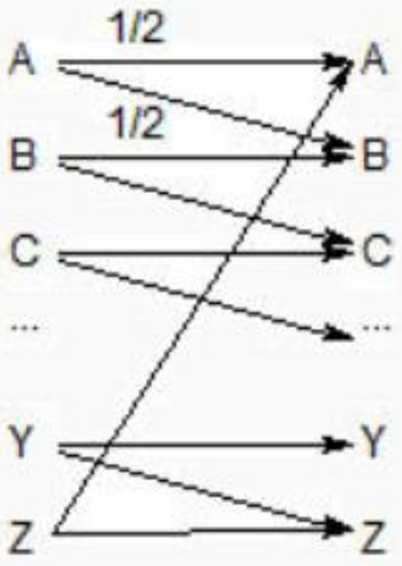
- We can still have reliable transmission over a noisy channel.

Noisy typewriter channel

- The channel input is either received unchanged at the output with probability ½ or is transformed into the next letter with probability ½.
- Consider the keyboard as input, the typewriter as channel, the printed characters as output.
- Information capacity

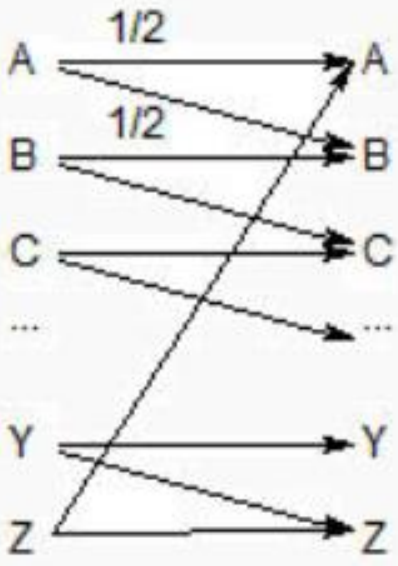
$$C = \max_{p(x)}[H(Y) - H(Y|X)] = \log(26) - 1 = \log(13)$$

$$\Rightarrow \text{ C is achieved when the input symbols follow uniform distribution.}$$



- **Is this capacity achievable?**

Noisy typewriter channel



- The channel input is either received unchanged at the output with probability ½ or is transformed into the next letter with probability ½.
- Consider the keyboard as input, the typewriter as channel, the printed characters as output.

- A simple coding solution
  - Use only every second of the 26 possible input symbols
  - Input can be determined from the output
  - Channel capacity is log(13) bits per channel

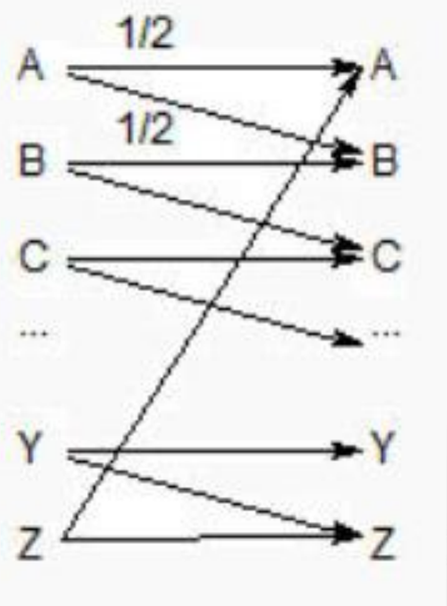
- Does it mean the channel capacity is achievable? $C = \max_{p(x)}[I(X;Y)]$

Noisy typewriter channel: insights

- Any channel can be seen as a noisy typewriter channel.

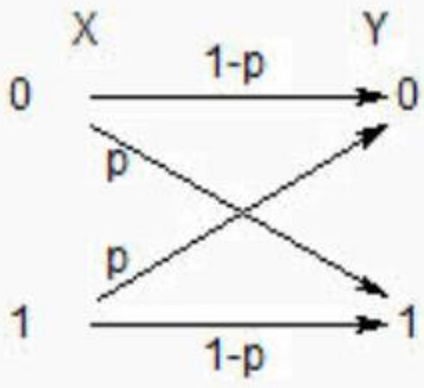
- The channel capacity defined in terms of mathematical meaning is achievable with some kind of coding solution.

- Key idea:
  - Find a set of input symbols, the output of each symbol is **non-overlapped** with each other.
  - In this case, the decoder can decode the input symbols **without error**, indicating reliable communication.
  - The **maximum distinguishable number of input symbols** will be the same as the channel capacity.



Binary symmetric channel (BSC)

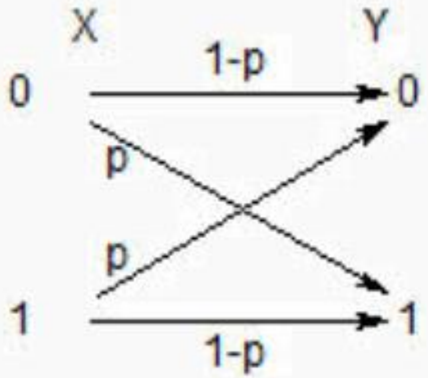
- Simplest channel model with errors
- No simple procedure to achieve reliable communication
  - All bits are potentially unreliable



BSC

Can we provide a reliable transmission over an unreliable channel?

Error probability of BSC



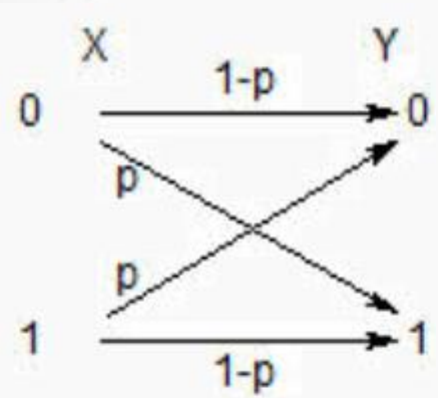
BSC

- Assume $p < 0.5$.
- Repeat bits once: $0 \to 0$; $1 \to 1$.

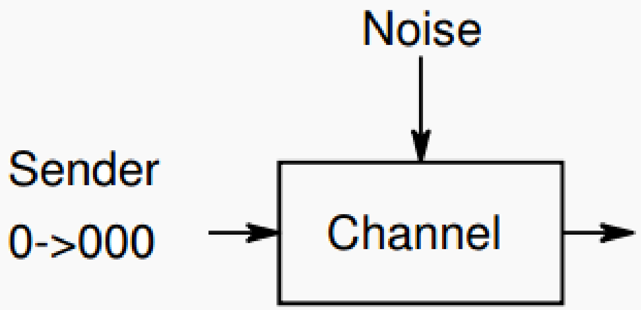
$$
\begin{aligned}
p_e &= P(X=0)P(Y=1|X=0) + P(X=1)P(Y=0|X=1) \\
&= \pi p + (1-\pi)p \\
&= p
\end{aligned}
$$

Can we reduce the error probability with certain coding method?

Repeated code for BSC



- Assume $p < 0.5$.
- Repeat 3 times: $0 \to 000$; $1 \to 111$.
- Apply the simple majority decoding rule.



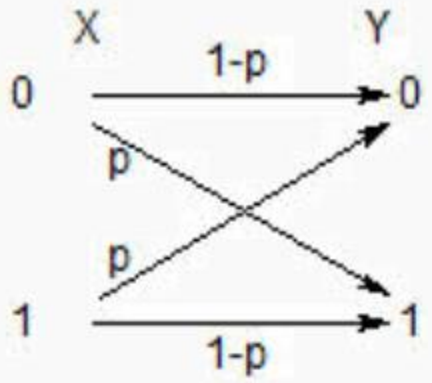
Receiver
000 -> 0
001 -> 0
010 -> 0
011 -> 1
100 -> 0
101 -> 1
110 -> 1
111 -> 1

$$\begin{aligned} p_e &= C_3^2 p^2(1-p) + C_3^3 p^3 \\ &= 3p^2 - 2p^3 \\ &= 3p^2 - 2p^3 - p + p \\ &= -p(1-p)(1-2p) + p \\ &< p \end{aligned}$$

What if $n \to \infty$?

Binary symmetric channel (BSC)

- Simplest channel model with errors
- No simple procedure to achieve reliable communication
  - All bits are potentially unreliable
- Information capacity



BSC

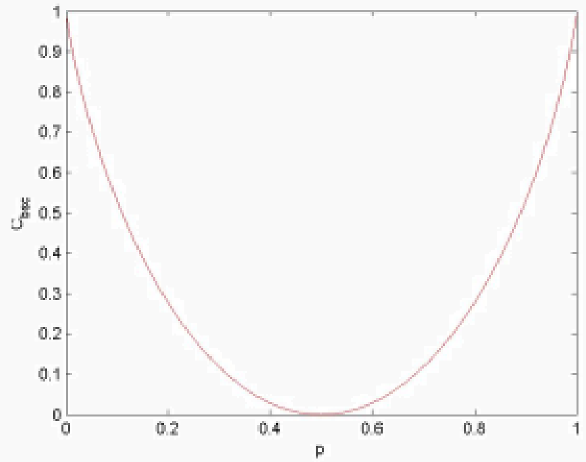
$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum p(x) H(Y|X=x)$$

$$= H(Y) - \sum p(x) H(p) = H(Y) - H(p) \leq 1 - H(p), \qquad \Rightarrow C \sim U(p).$$

$$\Rightarrow C = \max_{p(x)} [I(X;Y)] = \max_{p(x)} [H(Y) - H(p)]$$

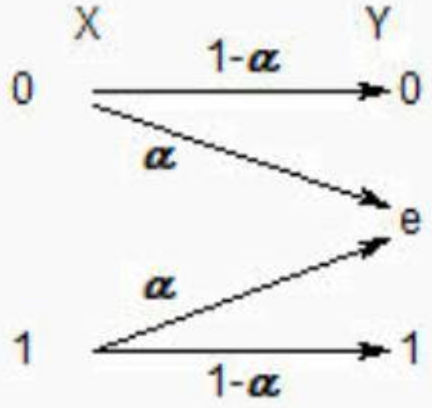
$$= \max_{p(x)} [H(Y)] - H(p) = 1 - H(p),$$

where $H(p) = -p \log p - (1-p) \log (1-p).$



Binary erasure channel

- Erasure recognized by the receiver.
- Compared with BSC: bits lost, not corrupted.
- Information capacity



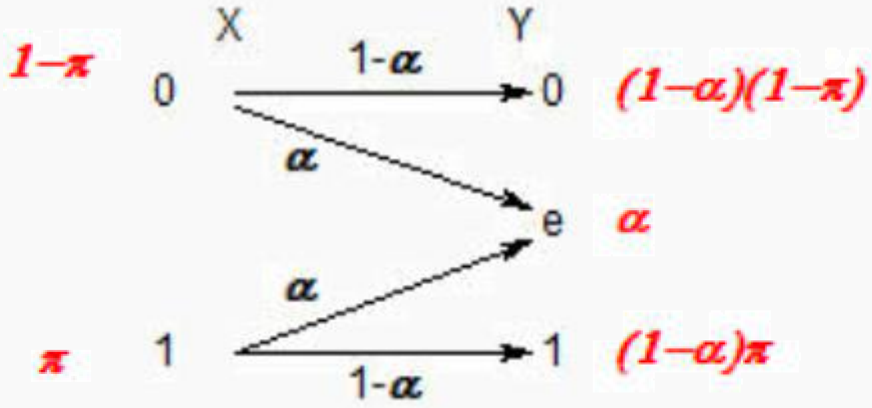
$$C = \max_{p(x)}[I(X;Y)] = \max_{p(x)}[H(Y)] - H(\alpha) = 1 - \alpha$$

Proportion α of bits lost $\Rightarrow$ at most proportion $1-\alpha$ recovered

- How to achieve the capacity?
  - If feedback available: request retransmission
  - However, the feedback does not increase the capacity.

Binary erasure channel

$$
\begin{aligned}
C &= \max_{p(x)}[I(X;Y)] \\
&= \max_{p(x)}[H(Y) - H(Y|X)] \\
&= \boxed{\max_{p(x)}[H(Y)]} - H(\alpha)
\end{aligned}
$$



$$
\begin{aligned}
H(Y) &= H\left((1-\pi)(1-\alpha), \alpha, \pi(1-\alpha)\right) \\
&= H(\alpha) + (1-\alpha)H(\pi);
\end{aligned}
$$

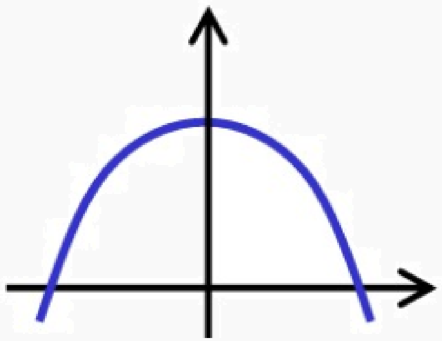
$$
\begin{aligned}
C &= \max_{p(x)}[H(Y)] - H(\alpha) \\
&= \max_{\pi}(1-\alpha)H(\pi) + H(\alpha) - H(\alpha) \\
&= \max_{\pi}(1-\alpha)H(\pi) \\
&= 1 - \alpha
\end{aligned}
$$

The maximum is taken over ALL possible input distributions.

Computing channel capacity: generalization

Review of the properties of channel capacity

- $I(X;Y)$ is a concave function of $p(x)$.
- $C \leq \log(|X|)$ and $C \leq \log(|Y|)$.



$\Rightarrow$ The local maximum is a global one and finite.

$\Rightarrow$ We are justified in using the term maximum, rather than supremum in the definition of capacity.

- The maximum can be found using the standard nonlinear optimization techniques, e.g., gradient search.
- However, in general, there are NO closed-form solutions for computing channel capacity.

Discrete memoryless channel: classification

- Classified by noise
  - Noiseless channel
    - Noiseless binary channel
  - Noisy channel
    - Noisy channel with non-overlapping
    - Noisy Typewriter
  - Noise and loss
    - Discussions on noise and loss
- Classified by matrix
  - Symmetric channel
  - Strong symmetric

Computing channel capacity: method

- Simplified methods for special channels
  - Using properties like symmetry
  - Constructing input distribution

- General methods
  - Constrained maximization using calculus and the Kuhn-Tucker conditions
  - The Frank-Wolfe gradient search algorithm
  - An iterative algorithm developed by Arimoto and Blahut

Special case: symmetric channel

- The rows of the channel transition matrix are permutations of each other and the columns are permutations of each other.
  - Every row is formed by the elements in the same set $\{p'_1, p'_2, \ldots, p'_s\}$
  - Every column is formed by the elements in the same set $\{q'_1, q'_2, \ldots, q'_r\}$
    $\Rightarrow$ if $r = s$, $\{p'_i\} = \{q'_i\}$; if $r < s$, $\{q'_i\} \subseteq \{p'_i\}$.
  - Property: when X is uniform, Y is also uniform.
  - Example:

$$p(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}, \quad [p(y|x)]_{ij} = p(y_j|x_i)$$

Symmetric channel

- Which of the following are the channel matrix of a symmetric channel?

$$p(y|x) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix},$$

$$p(y|x) = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix},$$

$$p(y|x) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{bmatrix},$$

$$p(y|x) = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$$

Symmetric channel: information capacity

- Input alphabet $A = \{a_1, a_2, \ldots, a_r\}$.
- Output alphabet $B = \{b_1, b_2, \ldots, b_s\}$.

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) + \sum_{i=1}^{r} p(x_i) \sum_{j=1}^{s} p(y_j|x_i) \log p(y_j|x_i)$$

$$= H(Y) - \sum_{i=1}^{r} p(x_i) H_{si} = H(Y) - H_{si} = H(Y) - H(p_1', p_2', \ldots, p_s'),$$

where $H_{si} = -\sum_{j=1}^{s} p(y_j|x_i) \log p(y_j|x_i)$. ($H_{si}$ is a constant and independent of X.)

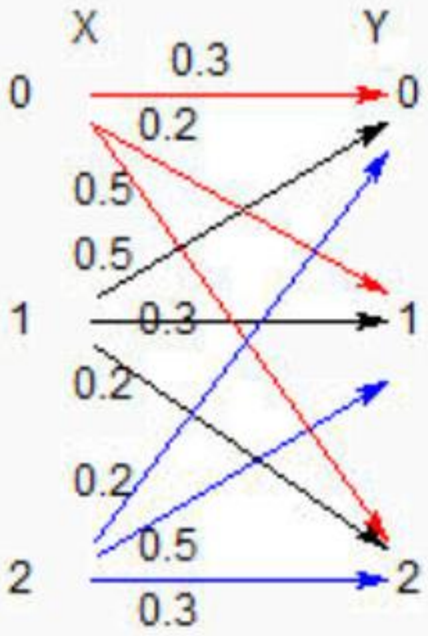
The information capacity is achieved when Y is uniformly distributed.

$$C = \max_{p(x_i)} \{H(Y) - H_{si}\} = \log_2 s - H(p_1', p_2', \ldots, p_s')$$

- s: the number of possible output symbols
- $\{p_1', p_2', \ldots, p_s'\}$: the row vector of channel matrix

Symmetric channel: example

$$p(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$



$$H_{si} = -\sum_{i=1}^{s} p(y_j|x_i) \log p(y_j|x_i)$$

- x_0: $H_{s0} = -(0.3 \times \log 0.3 + 0.2 \times \log 0.2 + 0.5 \times \log 0.5)$
- x_1: $H_{s1} = -(0.5 \times \log 0.5 + 0.3 \times \log 0.3 + 0.2 \times \log 0.2)$
- x_2: $H_{s2} = -(0.2 \times \log 0.2 + 0.5 \times \log 0.5 + 0.3 \times \log 0.3)$

$$\therefore H_{s0} = H_{s1} = H_{s2}$$

Symmetric channel: example

Find the channel capacity of

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{bmatrix}.$$

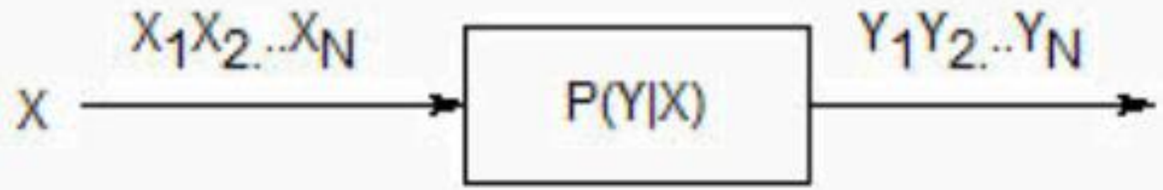
Since it is a symmetric channel, we have

$$\begin{aligned} C &= \log_2 s - H(p_1', p_2', \ldots, p_s') \\ &= \log 4 - H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}\right) \\ &= 2 + \left[\frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3} + \frac{1}{6}\log\frac{1}{6} + \frac{1}{6}\log\frac{1}{6}\right] \\ &= 0.0817(\text{bit/symbol}). \end{aligned}$$

Discrete channel: classification

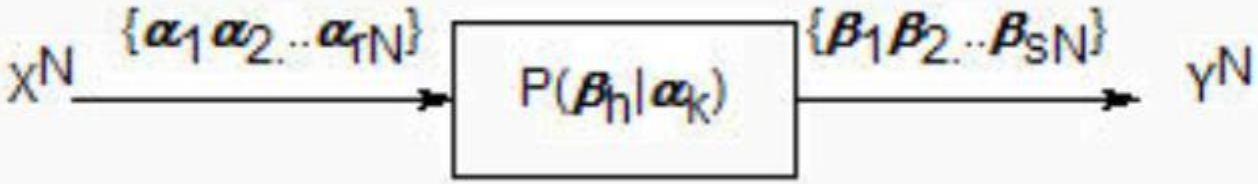
- Discrete channel
  - Single symbol channel
    - Classified by noise
      - Noiseless channel
      - Noisy channel
      - Noise and loss
    - Classified by matrix
      - Symmetric channel
  - Symbol sequence channel
    - Extension of single symbol channel
    - Independent parallel channel
    - Independent series channel

Symbol sequence channel



- Symbol sequence discrete channel
  - The channel input and output are random sequences
- Typical channels
  - Extension of single symbol channel
  - Independent parallel channel
  - Independent series channel

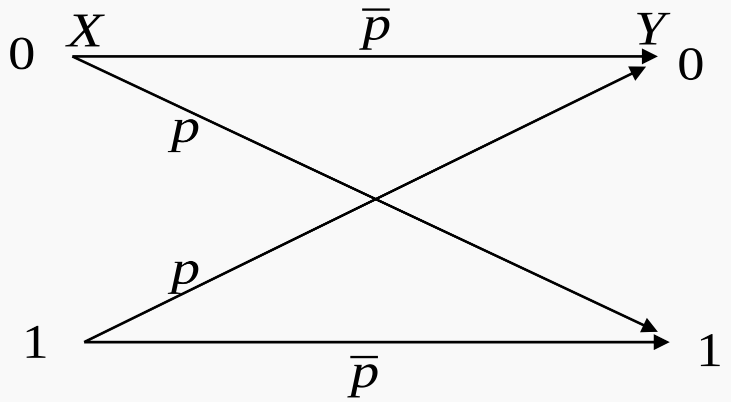
Extension of single symbol channel



- Original discrete memoryless channel
  - Input: $A = \{a_i\}$, output: $B = \{b_j\}$ $(i = 1, \ldots, r; j = 1, \ldots, s)$
  - Matrix: $P = \{p_{ij}\} = P_r(b_j|a_i)$ $(i = 1, \ldots, r; j = 1, \ldots, s)$
- N-extension discrete memoryless channel
  - Input: $A^N = \{\alpha_k\}$, output: $B^N = \{\beta_h\}$ $(k = 1, \ldots, r^N, h = 1, \ldots, s^N)$
    $\alpha_k = (a_{k1}a_{k2}\ldots a_{kN}), \beta_h = (b_{h1}b_{h2}\ldots b_{hN})$
    $a_{ki} \in A, b_{hi} \in B, i = 1, \ldots, N$
  - Matrix: $\pi = P(\beta_h|\alpha_k) = P(b_{h1}b_{h2}\ldots b_{hN}|a_{k1}a_{k2}\ldots a_{kN})$
    $= \prod P(b_{hi}|a_{ki})$

Extension of single symbol channel: example

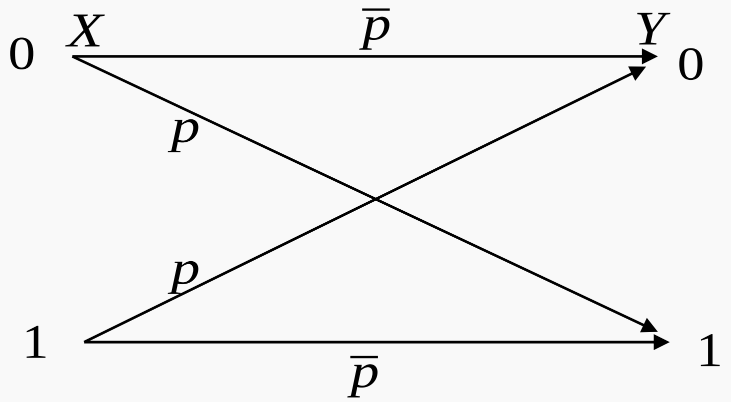
- Consider 2-extension of a binary symmetric channel
  - BSC: X={0,1}, Y={0,1}
- For the 2-extension channel,
  - Input: X^2={00,01,10,11}
  - Output: Y^2={00,01,10,11}
  - Channel matrix:



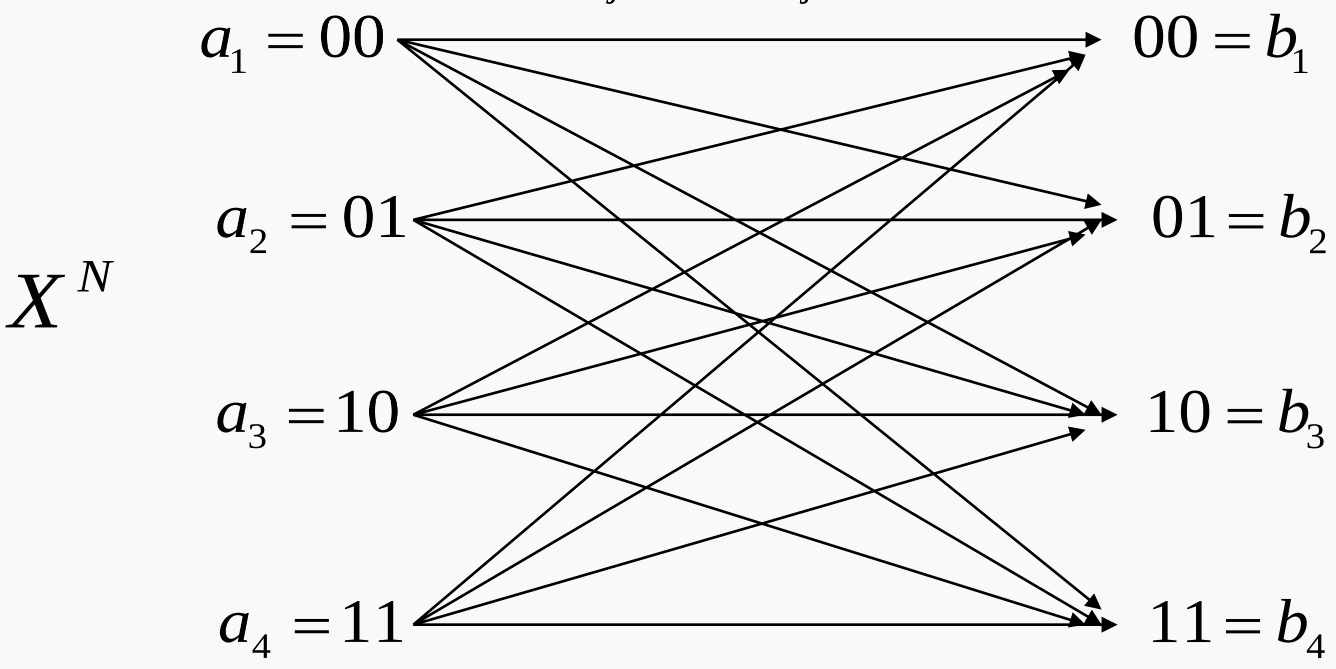
$$P = \begin{bmatrix} \bar{p}^2 & \bar{p}p & p\bar{p} & p^2 \\ \bar{p}p & \bar{p}^2 & p^2 & p\bar{p} \\ p\bar{p} & p^2 & \bar{p}^2 & \bar{p}p \\ p^2 & p\bar{p} & \bar{p}p & \bar{p}^2 \end{bmatrix}$$

Extension of single symbol channel: example

- Consider 2-extension of a binary symmetric channel
  - BSC: X={0,1}, Y={0,1}
- Channel diagram



$$p_{ij} = p(b_j / a_i)$$



Extension of single symbol channel

- For N-extension discrete memoryless channel $(p(y^N|x^N) = \prod_{i=1}^{N} p(y_i|x_i))$

$$I(\mathbf{X};\mathbf{Y}) \leq \sum_{i=1}^{N} I(X_i; Y_i)$$

- If both the input source and the channel are memoryless $(p(x^N) = \prod_{i=1}^{N} p(x_i))$, then

$$I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{N} I(X_i; Y_i).$$

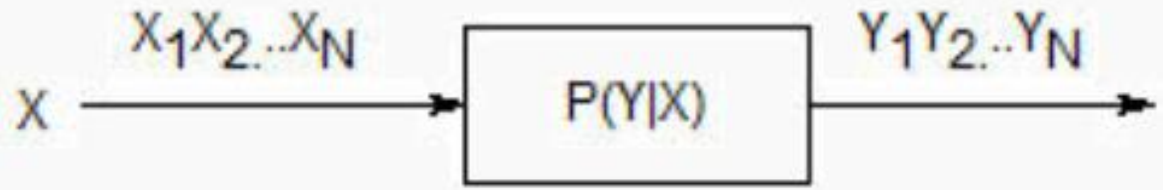
- If X_i are from the same space and have the identical distribution, then

$$I(X_1; Y_1) = I(X_2; Y_2) = \ldots = I(X_N; Y_N) = I(X; Y).$$

$$\Rightarrow I(\mathbf{X};\mathbf{Y}) \leq \sum_{i=1}^{N} I(X_i; Y_i) = NI(X; Y).$$

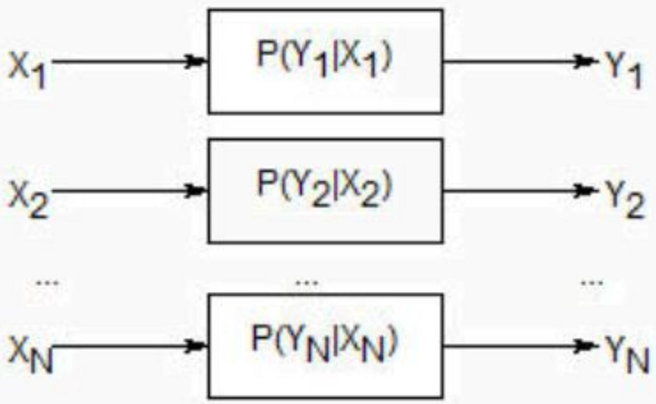
$$\boxed{C^N = \max_{P(x)} I(\mathbf{X};\mathbf{Y}) = \max_{P(x)} \sum_{i=1}^{N} I(X_i; Y_i) = \sum_{i=1}^{N} C_i = NC}$$

Symbol sequence channel



- Symbol sequence discrete channel
  - The channel input and output are random sequences
- Typical channels
  - Extension of single symbol channel
  - Independent parallel channel
  - Independent series channel

Independent parallel channel



- The output of each channel Y_i only depends on its own input X_i.
- $p(y_1 y_2 \ldots y_N | x_1 x_2 \ldots x_N) = p(y_1|x_1)p(y_2|x_2)\ldots p(y_N|x_N) \Rightarrow$ Apply the results of N-extension DMC

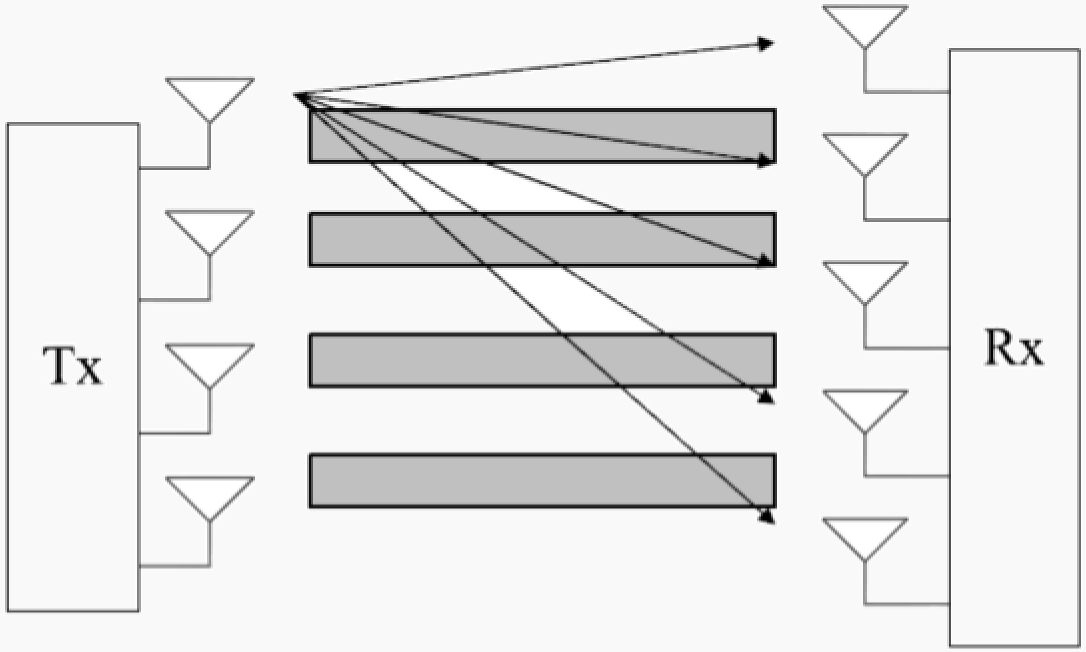
- Channel capacity $C_{1,2,\ldots,N} = \max\limits_{P(x_1\ldots x_n)} I(X_1\ldots X_N; Y_1\ldots Y_N) = \sum\limits_{i=1}^{N} C_i,$

  where C_i is the capacity of each independent channel $C_i = \max\limits_{P(x_i)} I(X_i; Y_i)$.

- When X_i are independent and achieve the optimal distribution to maintain its channel capacity, it achieves the capacity.

- Modern Communication Technologies: Multiple Input Multiple Output (MIMO)

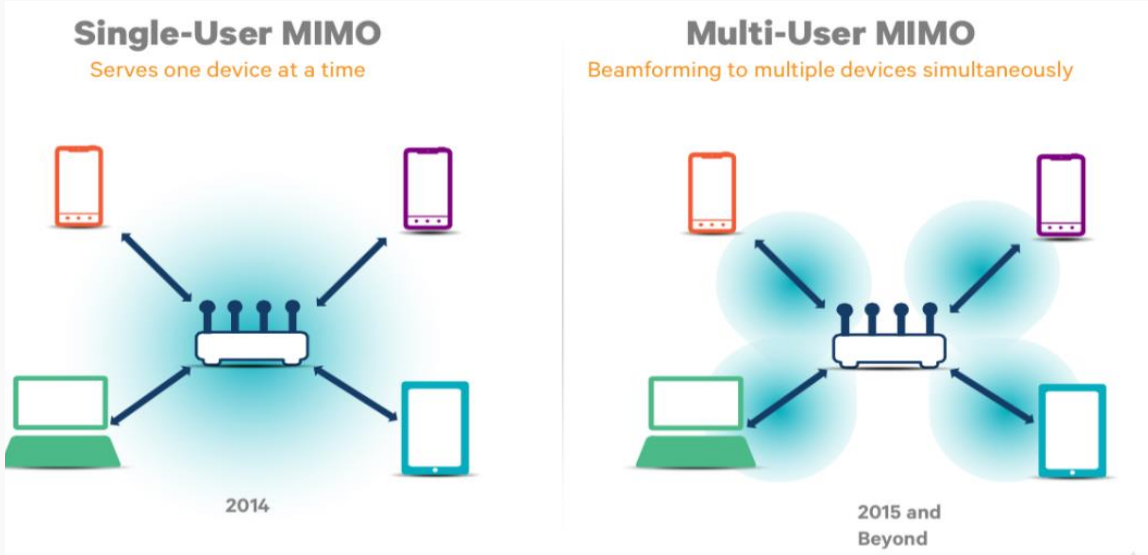
Multiple Input Multiple Output (MIMO)



- Multiple antennas at the transmitter and receiver side.
- Provide parallel channels to increase the channel capacity
- Higher data rate

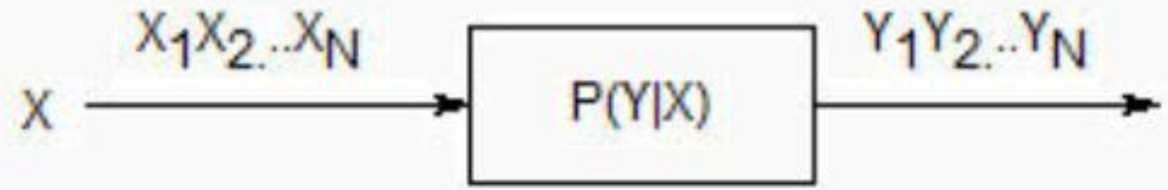
Development of MIMO technology

- Applications: WiFi (11n/ac/ax), cellular (4G, 5G)
- MIMO to Massive MIMO
- Single-User MIMO to Multi-User MIMO



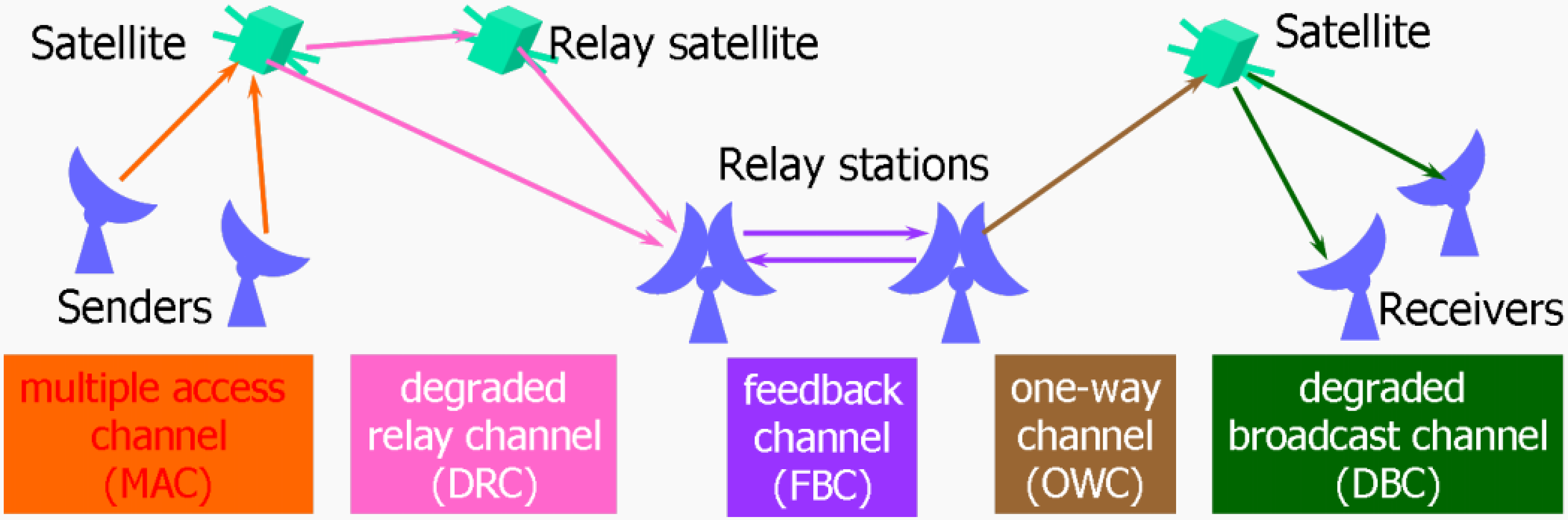
Development of MIMO technology: Antenna

Symbol sequence channel



- Symbol sequence discrete channel
  - The channel input and output are random sequences
- Typical channels
  - Extension of single symbol channel
  - Independent parallel channel
  - Independent series channel

Independent series channel



- End-to-end communication

Independent series channel

- Data processing theorem

$$H(X) \geq I(X;Y) \geq I(X;Z) \geq I(X;W) \geq \dots$$

- Channel capacity of series channel

$$C_{\text{series}}(I, II) = \max_{P(x)} I(X;Z)$$

$$C_{\text{series}}(I, II, III) = \max_{P(x)} I(X;W)$$

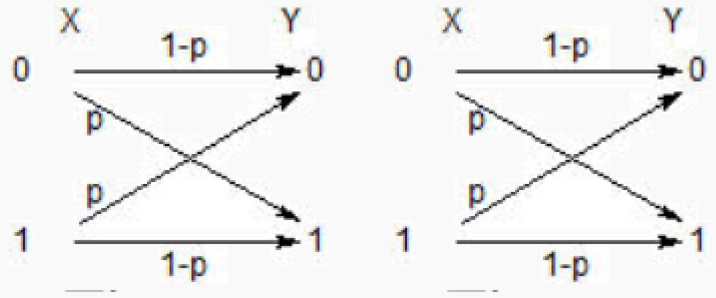
- Comments
  - The more channels in series, the smaller the capacity may become.
  - If # of channels $\to \infty$, series capacity may $\to 0$.

Independent series channel: example

Two 2-ary discrete symmetric channel, the input space is

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

The channel diagram of two series channels are



Connect the above two channels in series.

$$P_1 = P_2 = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

$$P = P_1 P_2 = \begin{bmatrix} (1-p)^2 + p^2 & 2p(1-p) \\ 2p(1-p) & (1-p)^2 + p^2 \end{bmatrix}$$

Mutual information

$$I(X;Y) = 1 - H(p)$$

$$I(X;Z) = 1 - H[2p(1-p)]$$

Channel capacity

$$C_{\text{series}}(I, II) = \max_{P(x)} I(X;Z) = 1 - H[2p(1-p)]$$

Summary

- Classification of channels
  - Simplest one: discrete memoryless channel

- Channel capacity (math. meaning)
  - maximum information transmission rate

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

- Calculate channel capacity
  - Single symbol channel
  - Symbol sequence channel

- How is this channel capacity related to coding?
  - Channel coding theorem

本节学习目标

1. 宏观理解信道容量分析问题的全貌
2. 信道容量分析的理论与应用意义
3. 说出信道的分类
4. 说出信道的定义
5. 说出信道矩阵的意义
6. 写出信道容量的定义式
7. 说出信道容量的物理意义
8. 计算一些简单信道的信道容量
9. 计算对称信道的信道容量

重难点：

- 信道容量问题概览
- 信道容量分析思维方法
- 信道容量定义
- 信道容量计算

My Homepage

Thank you!

Yayu Gao

School of Electronic Information and Communications
Huazhong University of Science and Technology
Email: yayugao@hust.edu.cn