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?
- 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. 写出什么是(M, n) code
2. 说出可操作意义上的信道容量定义
3. 理解两个信道容量为什么是相等的
4. 写出香农第二定理及其意义
5. 获得高斯信道的信道容量
6. 写出香农公式
7. 说出≥3种从香农公式能观察到的特性

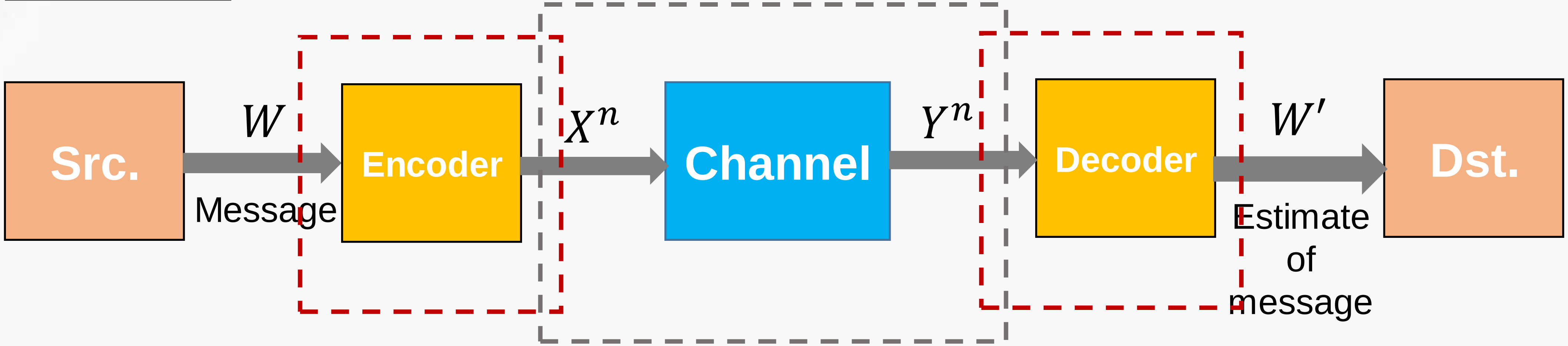
重难点：

- 香农第二定理及其意义
- 香农公式及其意义

06

How to define channel capacity in operation?

Revisiting: Channel capacity analysis



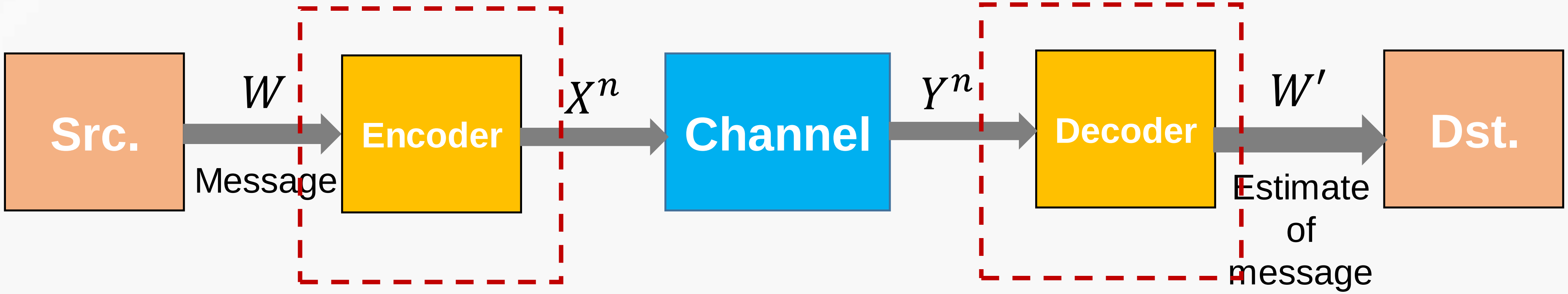
Math meaning

Channel capacity: maximum information transmission rate.

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

- **Is the channel capacity achievable?**
- **Is there an encoder/decoder to achieve reliable communication?**

System overview



- Message W(finite set of possible messages $\mathcal{W} = \{1, 2 \ldots, M\}$) is encoded by the encoder into a sequence of n channel input symbols, denoted by $X^n(W)$.
- At the other end of the channel another (random) sequence of channel output symbols Y^n is received, distributed according to $P(y^n|x^n(W))$.
- The sequence Y^n is then decoded by the decoder , who chooses an element $W'(Y^n) \in \mathcal{W}$,the receiver makes an error if $W'(Y^n) \neq W$.
- We suppose that the encoder and the decoder operate in a deterministic fashion:
  - $x^n(W)$ is the encoding rule (or function);
  - $W'(Y^n)$ is the decoding rule (or function);

(M, n) code: definition

An (M, n) code for a channel $(\mathcal{X}, P(y|x), \mathcal{Y})$ is defined by

1. An index set $\{1, \ldots, M\}$;
2. An encoding function $X^n(\square) : \{1, \ldots, M\} \to \mathcal{X}$, yielding codewords $x^n(1), \ldots, x^n(M)$. The set of codewords is called the codebook.
3. A decoding function $g(\square) : \mathcal{Y}^n \to \{1, \ldots, M\}$, which is a deterministic rule that assigns a guess to each possible received string.

***M* possible messages coded using a sequence of *n* input symbols**

(M, n) code: example

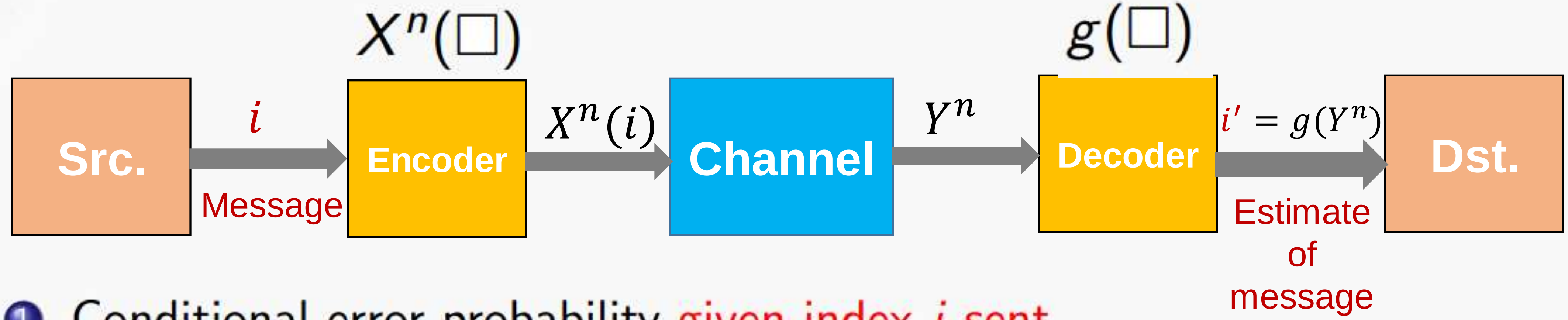
M possible messages coded using a sequence of n input symbols

- Transmit the outcome of tossing a fair coin
- M = 2: “head on”; “tail on”

M possible messages (M=2)	Channel coding rule 1: n=1	Channel coding rule 2: n=3
head on	1	111
tail on	0	000

When would an error occur?

(M, n) code: decoding error rate



1. Conditional error probability given index i sent

$$\boxed{\lambda_i = \Pr\left(g\left(Y^n\right) \neq i | X^n = x^n(i)\right) = \sum_{y^n} p\left(y^n | x^n(i)\right) I\left(g(y^n) \neq i\right)}$$

- Maximal error probability

$$\lambda^{(n)} = \max_{i \in \{1, \ldots, M\}} \lambda_i$$

- Arithmetic average error probability:

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$$

(M, n) code: communication rate

- R: the communication rate of an (M, n) code, which is defined as

$$R = \frac{\log M}{n}$$

  - The maximum information each bit of the codeword can carry for a channel code.

- Example
  - Transmit the outcome of tossing a fair coin. M=2
  - Channel coding rule 1: n=1. R=1.
  - Channel coding rule 2: n=3. R=1/3.

(M, n) code: achievable rate

- R is said to be **achievable** if there exists a sequence of $(M(n), n)$ codes such that

  1. $M(n) = \lceil 2^{nR} \rceil$;
  2. $\lim_{n \to \infty} \lambda^{(n)} = 0$.

  - Maximal error probability tends to zero as n goes to infinity.

Another definition of channel capacity: operational meaning

- *Definition:* The capacity of a channel is the supremum of all achievable rates.

- Physical meaning
  - Rates less than capacity yield arbitrarily small probability of error for sufficiently large block lengths (when using very long codewords).
  - For any rates below the above channel capacity, we can transmit reliably over the channel.
  - It shows the existence of lossless channel code, which provides plenty of insights to the communication engineering.

Channel capacity analysis

Channel capacity: Math meaning

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

Channel capacity: operational meaning

the supremum of all achievable rates $R = \frac{\log M}{n}$

How are they related?

Theory	Easier to obtain	Unrelated to Reliable communication

Practice	Difficult to obtain	Achieve Reliable communication

07

Shannon’s second theorem: channel coding theorem

Channel coding theorem: objective

Channel capacity: Math meaning

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

=

Channel capacity: operational meaning

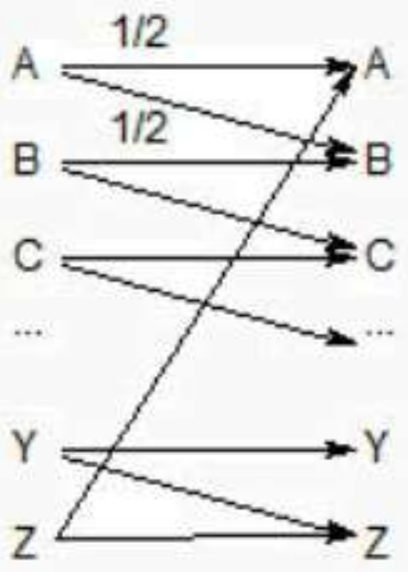
the supremum of all achievable rates

Theory	Easier to obtain	Unrelated to Reliable communication
Practice	**Difficult to obtain**	**Achieve Reliable communication**

Prove that the information capacity C is equal to the operational capacity.

Review: Insights of the noisy typewriter

- Question:
  - How to transmit without errors over a noisy channel?
- 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.
  - Find the **maximum distinguishable number of input symbols.**
- Solution:
  - Use only every second of the 26 possible input symbols
  - Channel capacity C=log(M)/n=log(13).



Special Case: Binary Symmetric Channel

At the transmitter

000000000

000000101

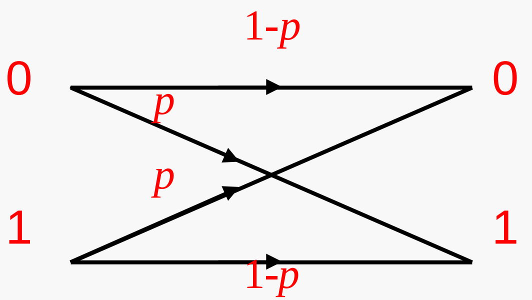
000001010

000010111

……

Code length n

Discrete-memoryless
Binary Symmetric Channel



At the receiver

For a given transmitted codeword x_i, it could be received in error, and associated with multiple error codewords.

- Is it possible to achieve zero error probability?

  ✓ Yes, if every received codeword is only associated with a single transmitted codeword.

Special Case: Binary Symmetric Channel

At the transmitter

000000000

000000101

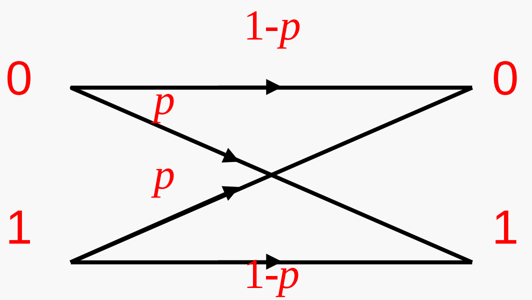
000001010

000010111

……

Code length n

Discrete-memoryless
Binary Symmetric Channel



At the receiver

For a given transmitted codeword x_i, it could be received in error, and associated with multiple error codewords.

- How to achieve zero error probability?
  - ✓ If the error pattern of each transmitted codeword is known, then we can properly select the transmitted codewords to ensure that their associated error codewords are all different from each other.

Special Case: Binary Symmetric Channel

At the transmitter

000000000

000000101

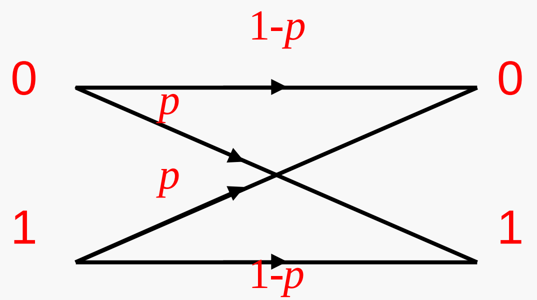
000001010

000010111

……

Code length n

Discrete-memoryless
Binary Symmetric Channel



At the receiver

For a given transmitted codeword x_i, it could be received in error, and associated with multiple error codewords.

- How to obtain the error pattern of each transmitted codeword?
- What is the maximum number of selected transmitted codewords?

Channel capacity: Binary Symmetric Channel

At the transmitter

000000000

000000101

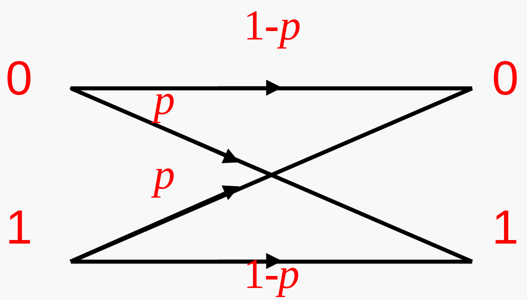
000001010

000010111

……

Code length n

Discrete-memoryless
Binary Symmetric Channel



At the receiver

- For any codeword x_i, ***np*** bits will be received in error with high probability, if n is large.
- The number of possible error codewords corresponding to x_i is

$$\binom{n}{np} = \frac{n!}{(np)!(n(1-p))!} \approx 2^{nH_b(p)}$$

$$H_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

Choose a subset of all possible codewords, so that the possible error codewords for each element of this subset is NOT overlapping!

- The maximum size of the subset: $M = \frac{2^n}{2^{nH_b(p)}} = 2^{n(1-H_b(p))}$
- The maximum rate that can be reliably communicated:

$$\boxed{C = \frac{1}{n}\log_2 M = 1 - H_b(p)}$$

(bit/transmission)

What is AEP?

- AEP: Asymptotic Equipartition Property 渐近均分性

Law of Large Number

- In a Bernoulli experiment sequence with the probability p.

$$X_i = \begin{cases} 1, & A \text{ occurs at } i\text{-th experiment.} \\ 0, & A \text{ does not occur at } i\text{-th experiment.} \end{cases}$$

$$S_n = \sum_{i=1}^{n} X_i, n = 1, 2, \ldots$$

- Weak law of large number: $\frac{S_n}{n} \xrightarrow{P} p$
- A general law of large numbers is as follows:
- If X_i is i.i.d,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} EX_i.$$

AEP property for i.i.d. r.v.

- AEP: Asymptotic Equipartition Property

$$\boxed{\begin{array}{l}\text{If } X_1, X_2, \ldots, X_n \text{ are } i.i.d. \sim p(x)\text{, then}\\ -\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \to H(X) \text{ in probability.}\end{array}}$$

- AEP was first stated by Shannon in his original 1948 paper, where he proved the results for i.i.d. processes and stated the result for stationary ergodic processes.

Proof:

$$\begin{aligned}-\frac{1}{n}\log\left[p(X_1, X_2, \ldots, X_n)\right] &= -\frac{1}{n}\log\left[p(X_1)p(X_2)\ldots p(X_n)\right]\\ &= -\frac{1}{n}\left(\log\left(p(X_1)\right) + \log\left(p(X_2)\right) + \ldots + \log\left(p(X_n)\right)\right)\\ &= -\frac{1}{n}\sum_i \log p(X_i) \to -E\left[\log p(X)\right] \text{ in probability}\\ &= H(X)\end{aligned}$$

AEP property for i.i.d. r.v.

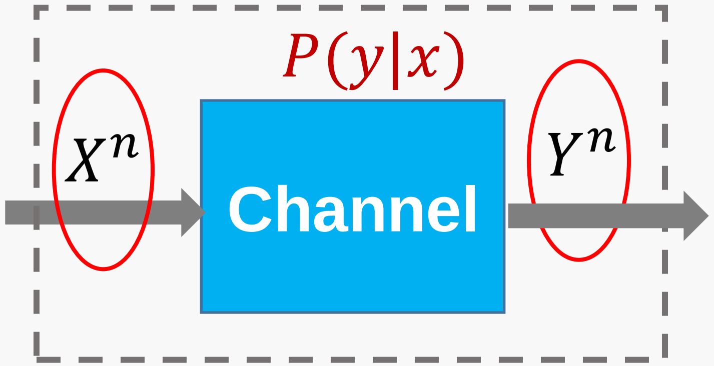
- AEP: Asymptotic Equipartition Property

> If $X_1, X_2, \ldots, X_n$ are $i.i.d. \sim p(x)$, then
> $-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(X)$ in probability.

- The probability $p(X_1, X_2, \ldots, X_n)$ assigned to an observed sequence will be close to $2^{-nH(X)}$.
- Almost all events are almost equally surprising.

An intuitive idea: general case

- For each input n-sequence, there are approximately $2^{nH(Y|X)}$ possible Y sequences, all of them equally likely (AEP).
- We wish to ensure that no two X sequences produce the same output sequence.
- The total number of possible Y sequences is $\approx 2^{nH(Y)}$ (AEP). This set has to be divided into sets of size $2^{nH(Y|X)}$.
- The total number of disjoint sets is less than or equal to $2^{n(H(Y)-H(Y|X))} = 2^{nI(X;Y)}$.
- We can send at most $2^{nI(X;Y)}$ distinguishable sequences of length n.



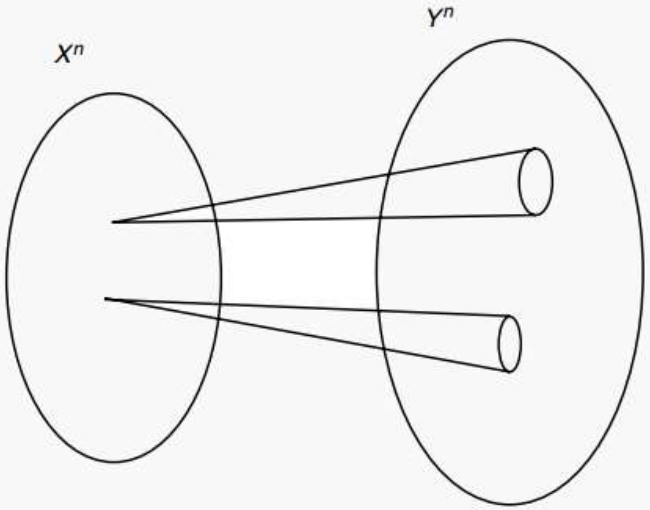


Figure. Channels after n uses.

The maximum rate that can be reliably communicated: $C = \frac{\log(2^{nI(X;Y)})}{n} = I(X;Y)$

Channel coding theorem

Channel code theorem

For a discrete memoryless channel, all rates below capacity C are achievable.
Specifically, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.
Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.

- Known as Shannon's second theorem.
- Provide lossless data transmission limit over a channel

Channel coding theorem: discussions

- The theorem statements have two parts:
  - Forwards: Any rates below C are achievable (zero error, reliable communication)
  - Backwards: Any rates above C are not achievable (errors will occur, unreliable communication)

- Provide the lossless data transmission limit over a channel
- Prove the existence of ideal channel codes to achieve the limit

- **We can provide reliable transmission over unreliable channels!**

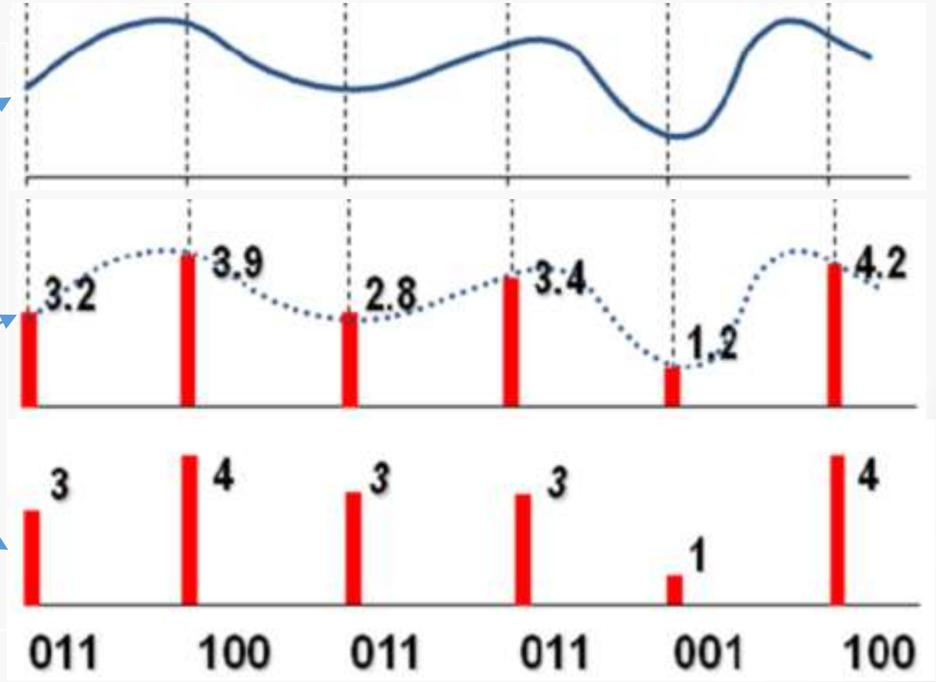
08

Channel capacity: from discrete to continuous

Revisiting: Classification of channels

- According to input/output signal value and time

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

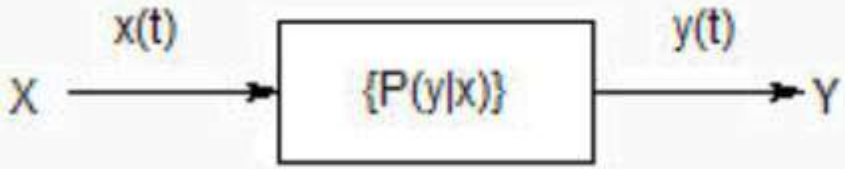


$$C = \max_{p(x)} \{I(X;Y)\}$$

What is AWGN channel?

Analog channel (Waveform channel)

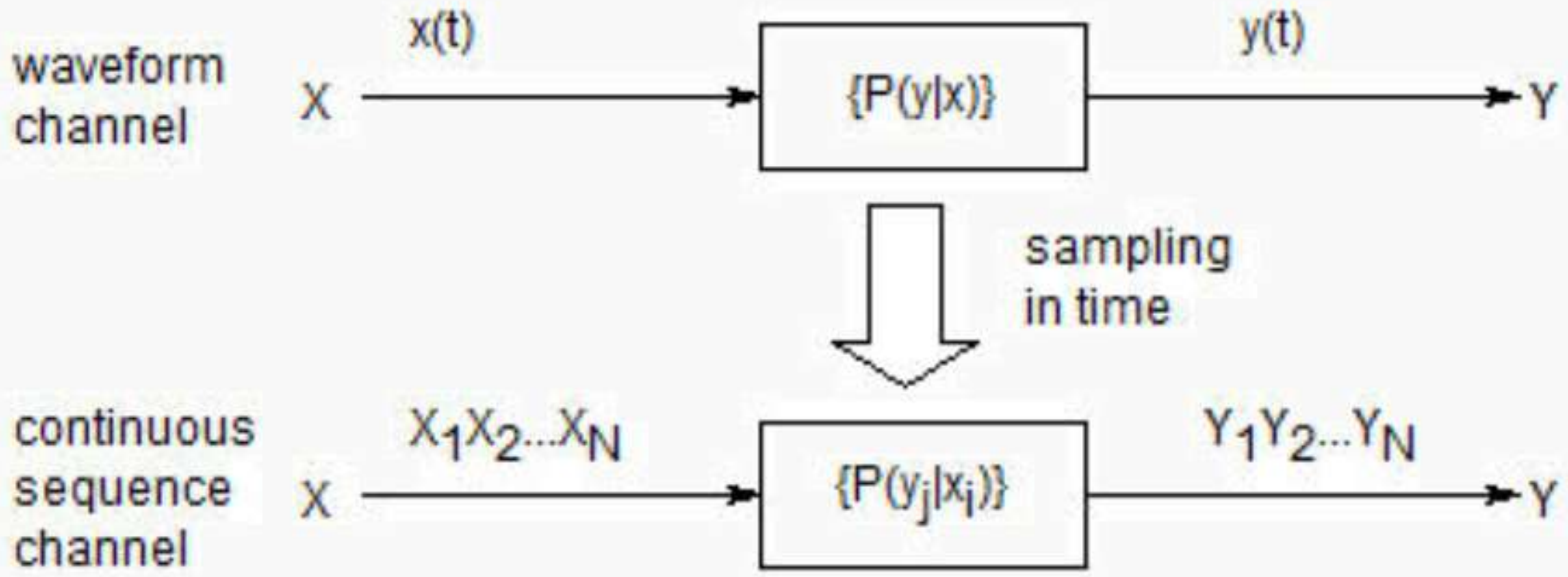
- Definition: if the input $\{x(t)\}$ and the output $\{y(t)\}$ are both stochastic processes with time t.



- Widely used in analog telecommunication systems.
- Characterized by the noise types, such as Gaussian, white noise, colored noise, etc..
  - Gaussian noise: p.d.f. follows Gaussian distribution;
  - White noise: power spectral density is uniformly distributed.

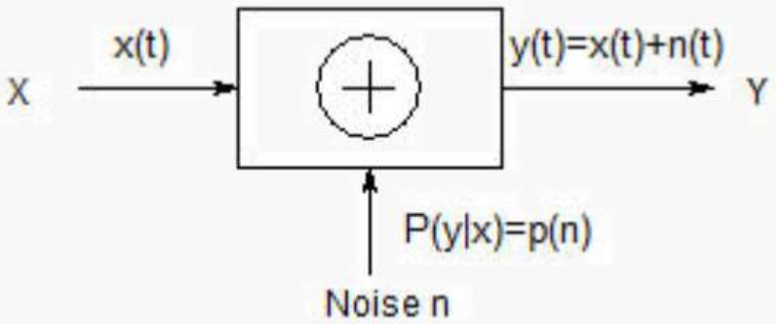
Continuous channel

- Definition: Both input X and output Y are continuous in value but discrete in time, can be sampled from waveform channels.

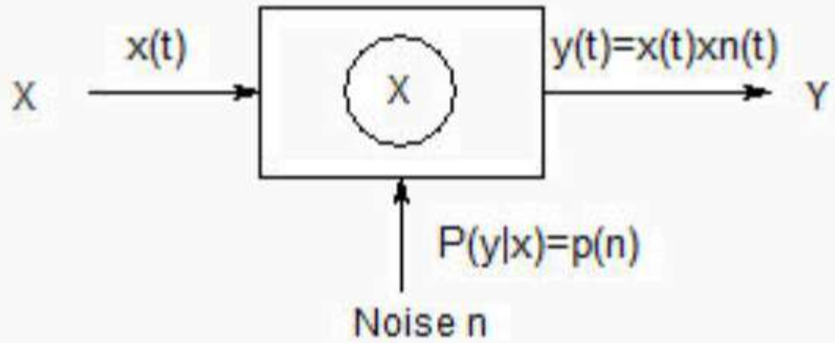


Continuous channel: special cases

- Additive noise channel
  - The noise and input are independent, and $Y = X + n$.



- Multiplicative noise channel
  - The noise and input are independent, and $Y = X \times n$.



Continuous channel: channel capacity

- Mutual information

$$\boxed{I(X;Y) = h(Y) - h(Y|X)}$$

- Channel capacity
  - continuous channel

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

  - additive noise channel

$$\boxed{C = \max_{p(x)} \{I(X;Y)\} = \max_{p(x)} \{\color{red}{h(Y) - h(n)}\}}$$

  - waveform channel

$$\boxed{C_t = \max_{p(x)} \left\{ \lim_{T\to\infty} [I(X;Y)] \right\} = \max_{p(x)} \left\{ \lim_{T\to\infty} [h(Y) - h(Y|X)] \right\}}$$

Revisiting: Differential entropy

- A continuous random variable contains **infinite information**.

$$\lim_{n\to\infty,\Delta\to 0} H(X) = -\int_a^b [f(x)\log f(x)]dx - \lim_{\Delta\to 0}\log\Delta$$

- Define **differential entropy** as the information measure of a continuous random variable.

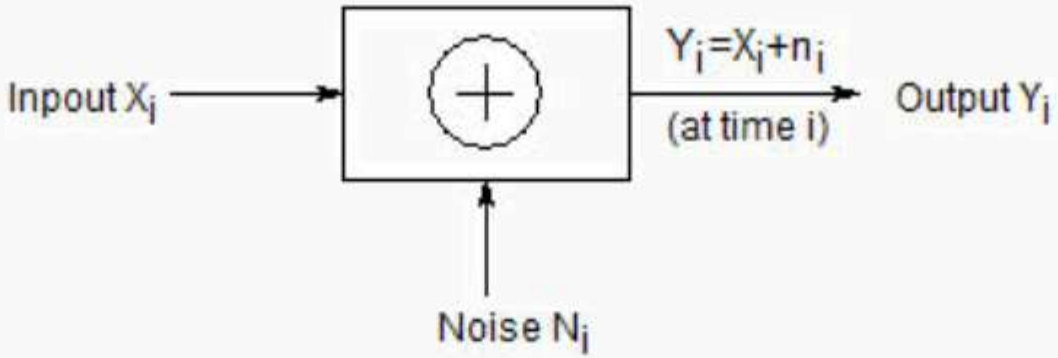
$$\boxed{h(X) = h(f) = -\int_S f(x)\log f(x)dx}$$

- It is not the absolute entropy of a continuous source.
- It cannot represent the average uncertainty/information of the source.
- It is a relative value with the reference point $-\lim_{\Delta\to 0}\log\Delta$
- It represents the difference between former and later source entropy

09

Most important continuous channel:
Gaussian channel

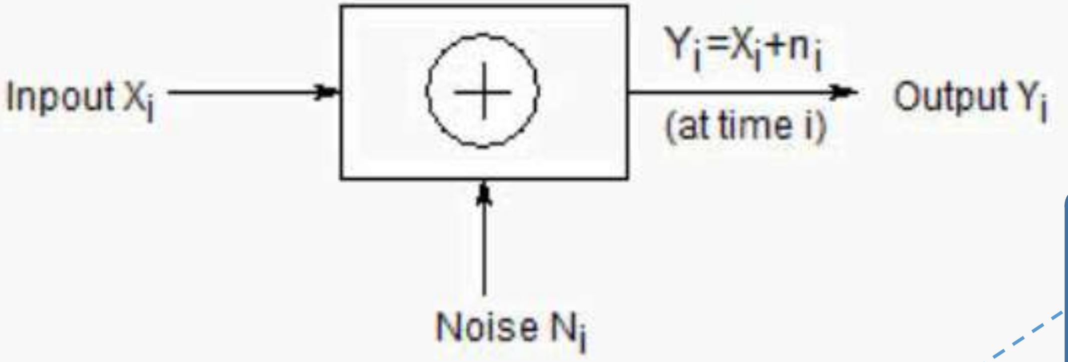
Gaussian channel



- Most important continuous channel.
- Time discrete channel $Y_i = X_i + N_i$.
- N_i: additive noise.
- N_i: $i.i.d.$ from a Gaussian distribution with zero mean and variance σ^2. $\Rightarrow$ The average power $P_N = E[n^2(t)] = \sigma^2$.
- N_i: assume to be independent of the signal X_i.

What is the channel capacity of a Gaussian channel?

Gaussian channel: channel capacity

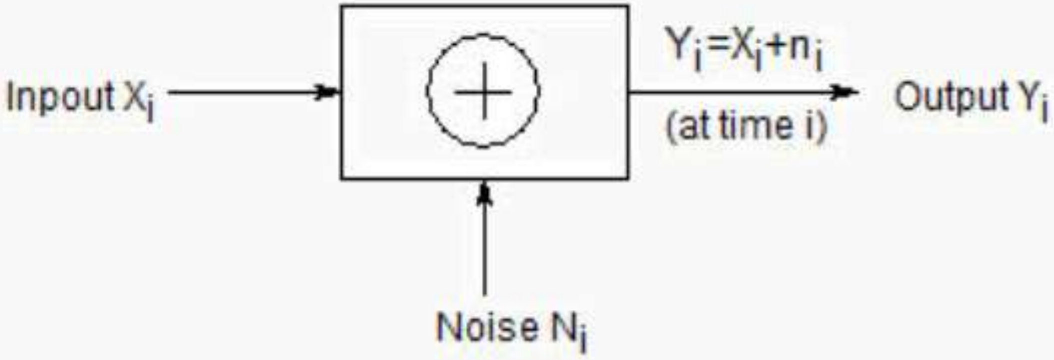


$$h(X) = \frac{1}{2}\ln\left(2\pi e\sigma^2\right) nats$$

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

- e.g. If the noise variance is zero, the receiver receives the transmitted symbol perfectly.
- The channel can transmit an arbitrary real number with no error.
- Without further conditions, the channel capacity may be infinite.

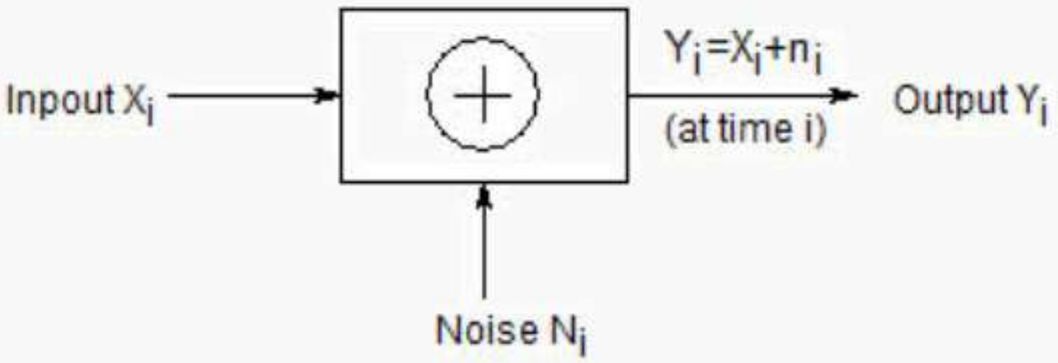
Gaussian channel: infinite channel capacity?



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

- Continuous *r.v.* contains infinite information (irrational has infinite details)
- However, we cannot produce an information source with infinite information in practice.
- Common limitation: **energy or power constraint on the input**

Gaussian channel: channel capacity

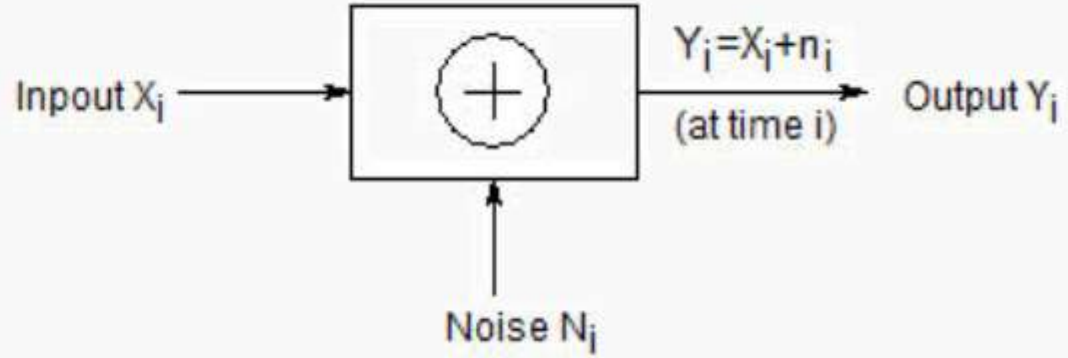


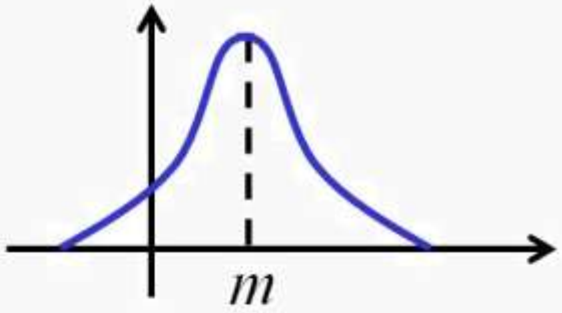
- The optimization of mutual information of input and output for a continuous channel is conducted under these constraints, such as the average power constraint on the input.

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

$$\boxed{s.t.\ \ E[X^2] = P_X}$$

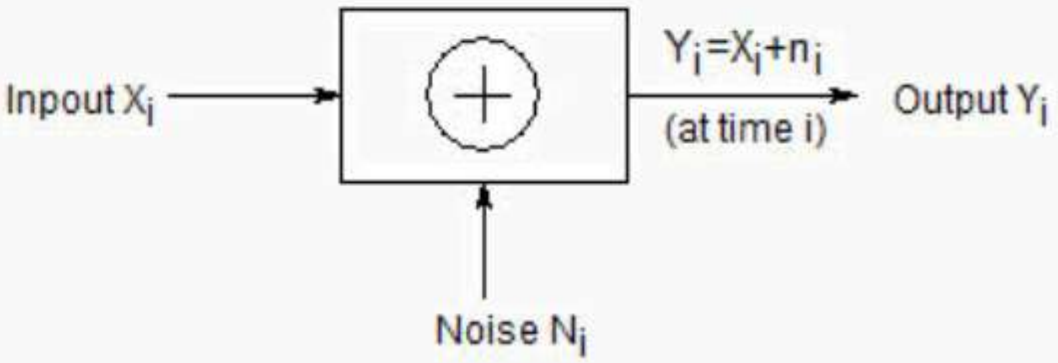
Gaussian channel: noise entropy





$$
\begin{aligned}
h(Y|X) = h(X+N|X) = h(N) &= -\int_N p(n)\log_2 p(n)dn \\
&= -\int_N p(n)\log_2[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{n^2}{2\sigma^2}}]dn \\
&= \frac{1}{2}\log[2\pi\sigma^2] + (\log_2 e)\cdot\int_N [p(n)\frac{n^2}{2\sigma^2}]dn \\
&= \frac{1}{2}\log[2\pi e\sigma^2]
\end{aligned}
$$

Gaussian channel: channel capacity



- Capacity of the Gaussian channel with the input power constraint P_X and the noise variance $P_N = \sigma^2$:

$$C = \max_{p(x)}\{h(Y)\} - h(N) = \max_{p(x)}\{h(Y)\} - \frac{1}{2}\log_2[2\pi e\sigma^2]$$

$$s.t.\ \ E[X^2] = P_X$$

Gaussian channel: channel capacity

$$C = \max_{p(x)}\{h(Y)\} - h(N) = \max_{p(x)}\{h(Y)\} - \frac{1}{2}\log_2[2\pi e\sigma^2]$$

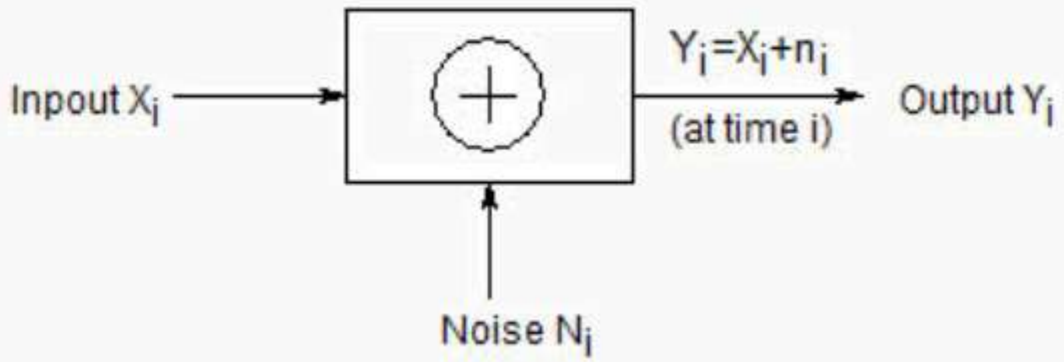
$$s.t.\;\; E[X^2] = P_X$$

- X and Z are independent and $EZ = 0 \Rightarrow$
  $EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2 = P_X + P_N$

With the variance constraint, the entropy of Y is maximized if Y is Gaussian distributed.

$$\begin{aligned} C &= \max_{p(x)}[\{h(Y)\} - h(N)] \\ &= \frac{1}{2}\log_2[2\pi e(P_X + P_N)] - \frac{1}{2}\log_2[2\pi e P_N] \\ &= \frac{1}{2}\log_2\left(1 + \frac{P_X}{P_N}\right) \text{ (bits per transmission).} \end{aligned}$$

Gaussian channel: channel capacity



- Capacity of the Gaussian channel with the input power constraint P_X and the noise variance $P_N = \sigma^2$:

$$C = \frac{1}{2}\log_2\left(1 + \frac{P_X}{P_N}\right) \text{ (bits per transmission).}$$

  - P_X / P_N is called as SNR (signal noise ratio).
  - The capacity is achieved when $X \sim \mathcal{N}(0, P_X)$.

$\Rightarrow$ Given continuous $r.v.$ X with mean m and variance σ^2, the differential entropy is maximized when it follows Gaussian distribution.

Another interpretation of capacity: Sphere packing

- $\mathbf{x}$ is an input sequence with power constraint: $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \le P$

$$\mathbf{y} = \mathbf{x} + \mathbf{z}$$

- Noise z_i is a zero-mean Gaussian random variable with variance σ^2.

for large n

$$\frac{1}{n}\sum_{i=1}^{n} z_i^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - x_i)^2 \to \sigma^2$$

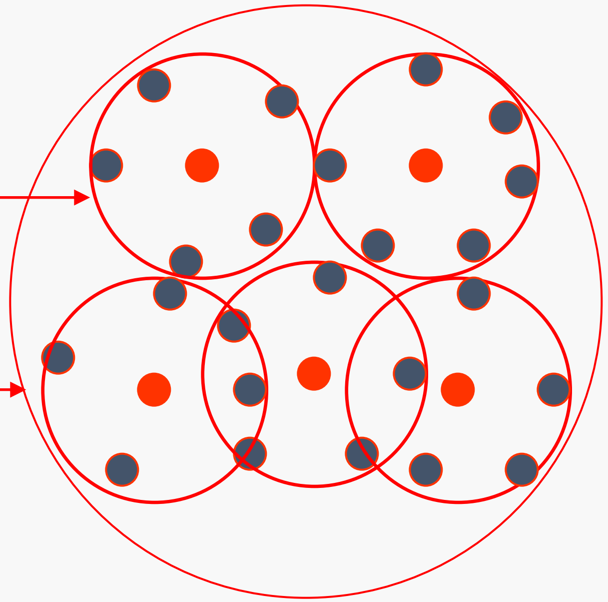
$$\frac{1}{n}\sum_{i=1}^{n} y_i^2 \le P + \sigma^2$$

$\| \mathbf{y} - \mathbf{x} \|^2 = n\sigma^2$

n-dimensional hyper-sphere with radius $\sqrt{n\sigma^2}$

$\| \mathbf{y} \|^2 \le n(P + \sigma^2)$

n-dimensional hyper-sphere with radius $\sqrt{n(P+\sigma^2)}$



- How many input sequences can we transmit at most over this channel such that the hyperspheres do not overlap?

$$M = (\sqrt{P+\sigma^2})^n / (\sqrt{\sigma^2})^n$$

- The maximum rate that can be reliably communicated ：

$$\boxed{C = \frac{1}{n}\log_2 M = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right) \quad \text{(bit/transmission)}}$$

10

Most famous formula in IT:

Shannon Formula

Bandlimited channel with Gaussian noise

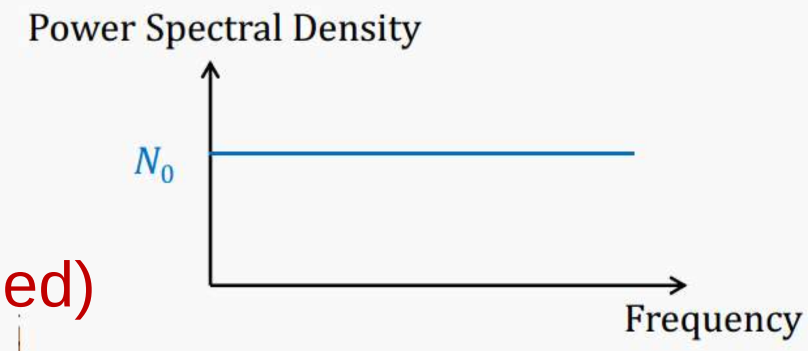
- A common model for communication over a radio network or a telephone line.
- Analog/Waveform channel
  - Continuous in both value and time.

- The output $Y(t) = (X(t) + N(t)) * h(t)$
  - $X(t)$: the signal waveform.
  - $N(t)$: the waveform of the white Gaussian noise with the power spectral density $\frac{N_0}{2}$ W/Hz $(-\infty < w < \infty)$.
  - $h(t)$: the impulse response of an ideal bandpass filter, which cuts out all frequencies greater than W.
  - $X(t)$, $N(t)$ and $Y(t)$ are all bandlimited signals.

Bandlimited AWGN channel

- Additive White Gaussian Noise
  - A basic noise model used in Information Theory.

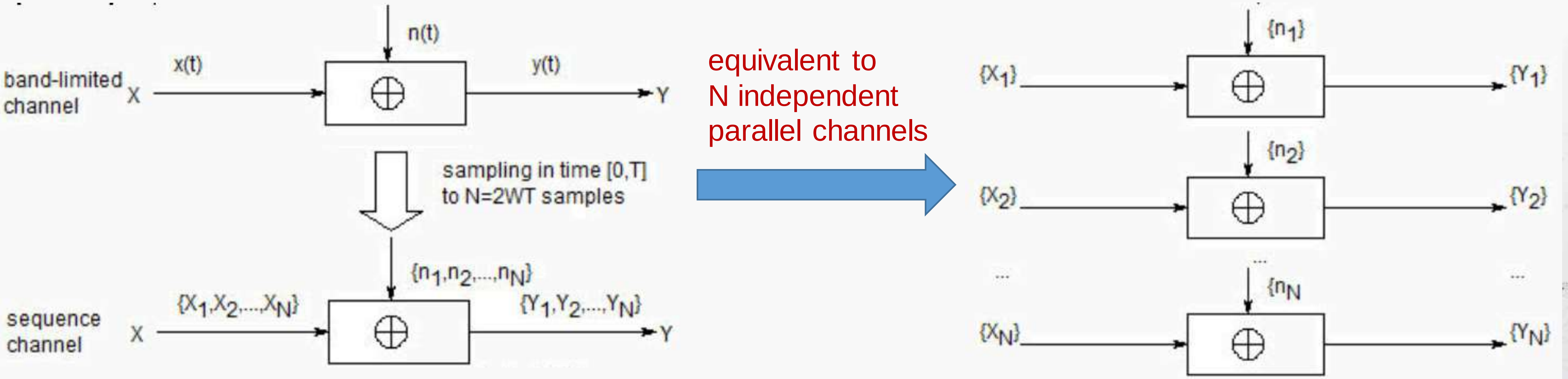
- **White Noise**
  - It has uniform power across the frequency band.
  - Modeled as a random signal with constant (one-sided) power spectral density, denoted by N_0 (Watts/Hz).
  - If the channel bandwidth is W Hz, then the noise power is given by $\boldsymbol{N_0 W}$ (Watts).



- If it is one-sided $(0 < w < \infty)$, N_0 W/Hz.
- If it is double-sided $(-\infty < w < \infty)$, $\frac{N_0}{2}$ W/Hz.

From Continuous-time to Discrete-time

- Nyquist's sampling theorem
  - Sampling a bandlimited signal at a **sampling rate 1/(2*W*)** is sufficient to reconstruct the signal from the samples for a bandlimited signal of bandwidth *W*.
    - i.e. 2*W* samples per second.



From Continuous-time to Discrete-time

- Let the channel be used in time interval [0, T], then bandlimited channel is equivalent to the parallel of ***2WT*** independent sample channels.
- Input power P_x
  - Power (variance) per sample: $\frac{P_X \cdot T}{2WT} = \frac{P_X}{2W}$.
- Noise power N_0W
  - Noise power (variance) per sample: $\frac{N_0 WT}{2WT} = \frac{N_0}{2}$.
- Capacity per sample

$$C = \frac{1}{2}\log_2\left(1 + \frac{P_X}{P_N}\right) = \frac{1}{2}\log_2\left(1 + \frac{P_X}{N_0 W}\right) \text{ (bits/transmission).}$$

Shannon Formula: Channel capacity for AWGN channel

- Since there are $2W$ samples each second, the capacity of bandlimited AWGN channel is

$$C_t = \frac{1}{2}\log_2\left(1 + \frac{P_X}{N_0 W}\right) * 2W = W\log_2\left(1 + \frac{P_X}{N_0 W}\right) \text{ (bits/s)}.$$

- **Shannon formula**
  - One of the most famous formulae in information theory.
  - It can be used to evaluate the performance of practical coding schemes.
  - It provides a high-level way of thinking how the performance of a communication system depends on the basic resources available in the channel.

Shannon Formula: Insights for Communication System Design

$$C_t = W \log_2\left(1 + \frac{P_X}{N_0 W}\right) \text{ (bits/s)}.$$

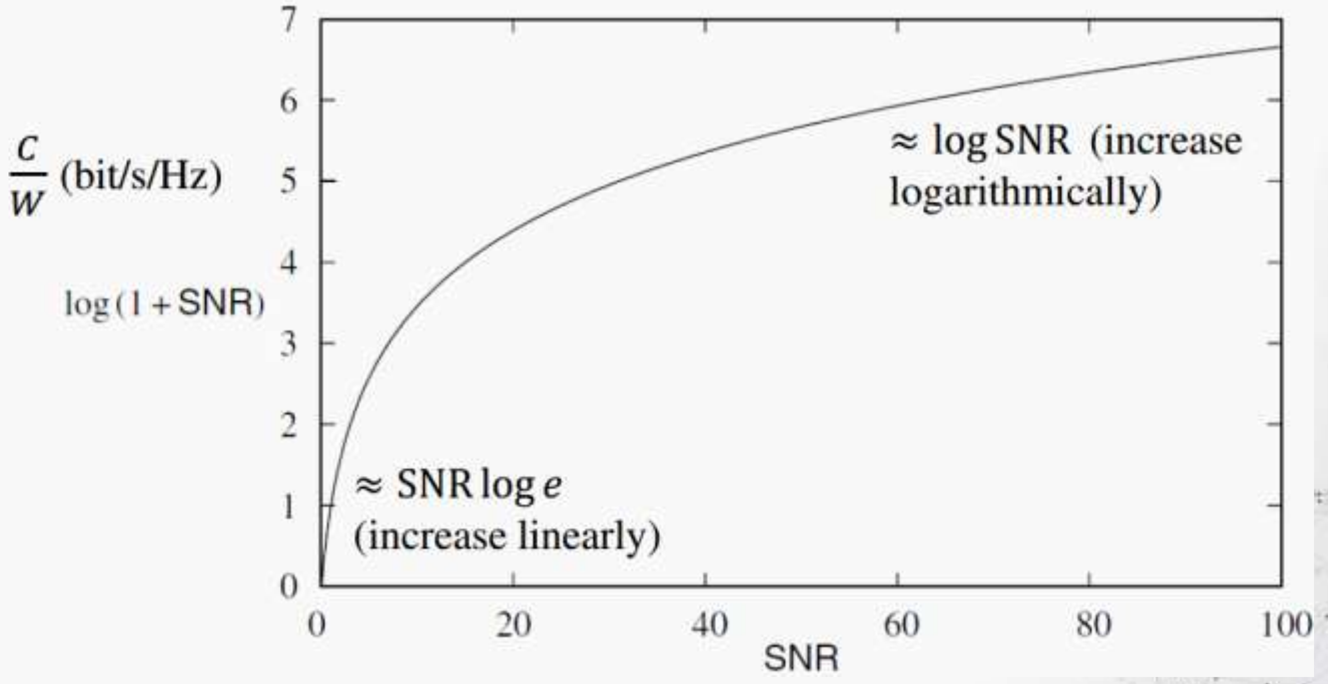
- Two most important resources for communications systems:
  - Transmission power *P* and bandwidth *W*.
- Given capacity *C*, bandwidth and SNR are interchangeable.

Insights: capacity as a function of SNR

$$C_t = W \log_2\left(1 + \frac{P_X}{N_0 W}\right) \text{ (bits/s)}.$$

The (normalized) capacity function log(1 + SNR) is a concave function.

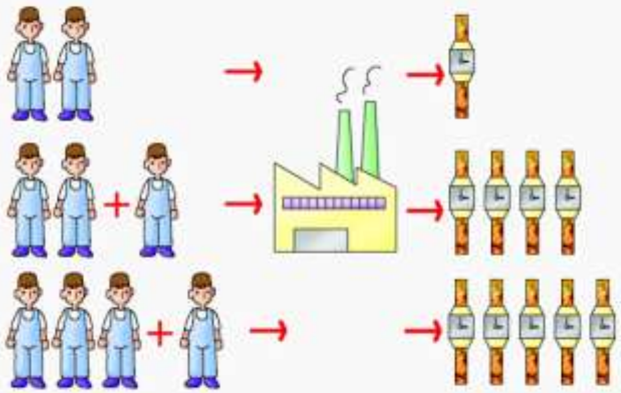
- The capacity increases with SNR. (unlimitedly)
- The capacity increases **linearly** with SNR when SNR is small.
- The capacity increases **logarithmically** with SNR when SNR is large.



Insights: capacity as a function of SNR

$$C_t = W \log_2 \left(1 + \frac{P_X}{N_0 W} \right) \ \text{(bits/s)}.$$

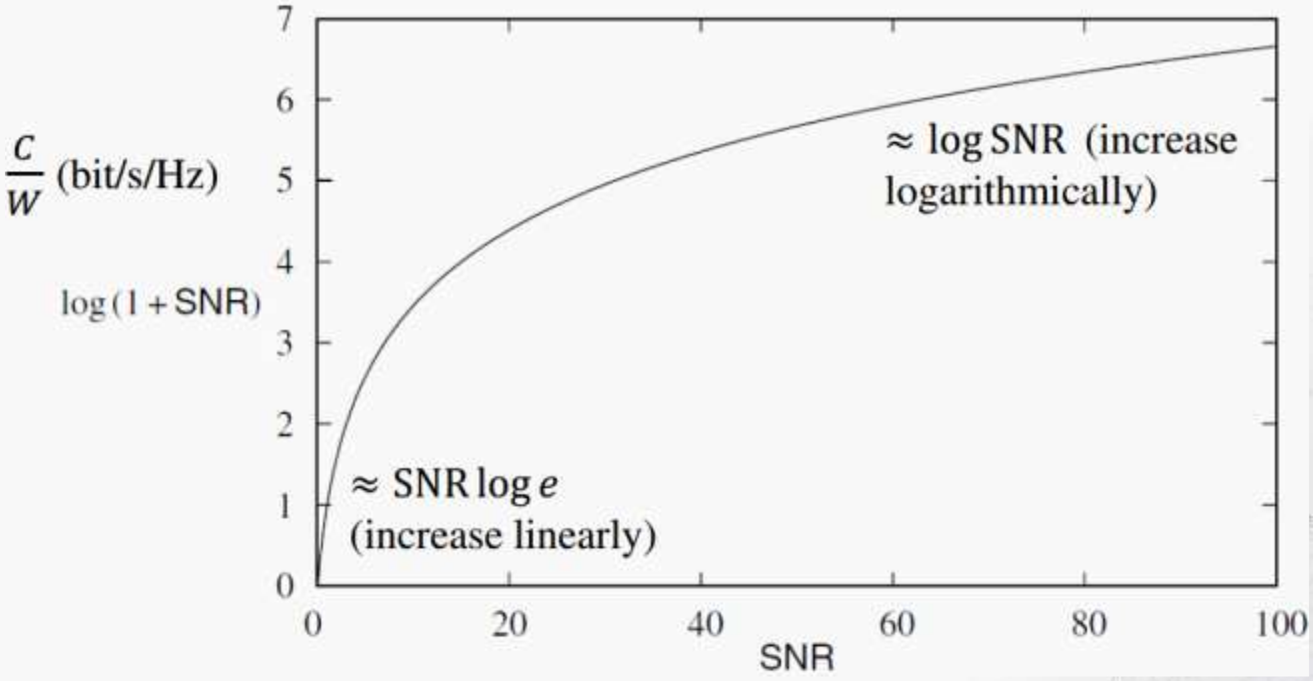
- Economic principle: Law of diminishing marginal utility (边际效益递减规律)
  - There will be a decrease in the marginal (incremental) output of a production process as the amount of a single factor of production is incrementally increased, while the amounts of all other factors of production stay constant.



Insights: capacity as a function of SNR

$$C_t = W \log_2\left(1 + \frac{P_X}{N_0 W}\right) \text{ (bits/s)}.$$

- Diminishing marginal utility in bit rates: concave function in *Px*
- Adding more units of power yields lower incremental per-unit returns (bit rate)
- When SNR is low, the function is linear.
    - Every 3 dB (i.e. doubling) increase in power doubles the capacity.
- When SNR is high, the function is a log function.
    - Every 3 dB increase in power increases the capacity by 1 bit.



Insights: capacity as a function of bandwidth W

Two conflicting effects when W increases.

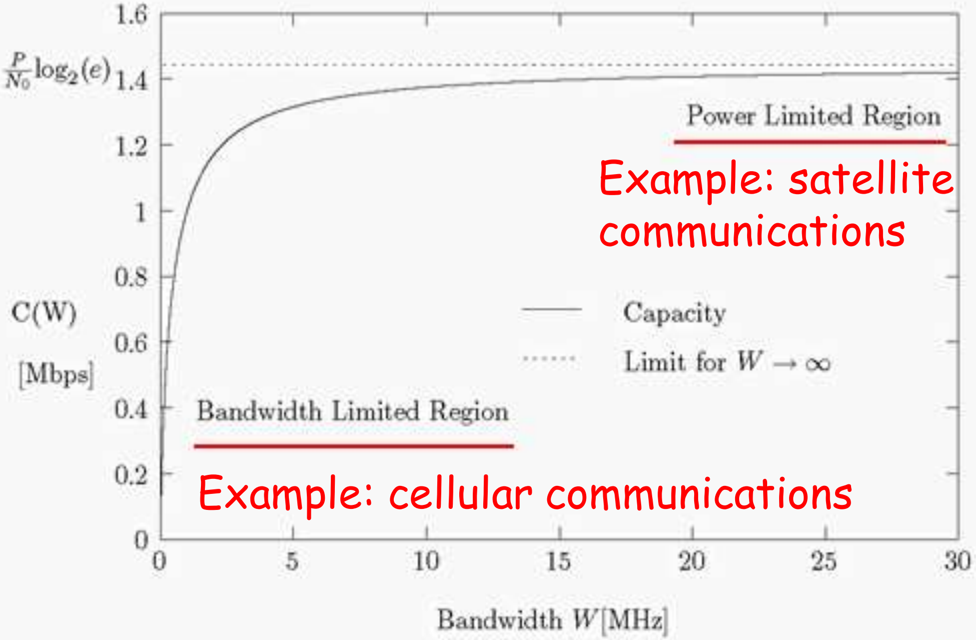
$$C_t = W \log_2\left(1 + \frac{P_X}{N_0 W}\right) \text{ (bits/s)}.$$

Increases the degree of freedom

Increases the noise power, (or equivalently, decreases the SNR)

- When the bandwidth W is small, the capacity can be significantly improved by increasing W.

- If we increase the bandwidth W without limit, can we get an infinitely large channel capacity?
  - **No.** $\lim_{W\to\infty} C = \frac{P}{N_0}\log_2 e$



Insights: Infinite bandwidth limit

$$C_t = W \log_2\left(1 + \frac{P_X}{N_0 W}\right) \text{ (bits/s)}.$$

$$\lim_{W\to\infty} C_t = \lim_{W\to\infty} W \log_2\left[1 + \frac{P_X}{P_N}\right] = \lim_{W\to\infty} W \log_2\left[1 + \frac{P_X}{N_0 W}\right]$$

$$= \lim_{W\to\infty} \frac{P_X}{N_0} \cdot \frac{N_0 W}{P_X} \log_2\left[1 + \frac{P_X}{N_0 W}\right]$$

Let $\alpha = \frac{P_X}{N_0 W}$. $\lim_{\alpha\to 0}(1+\alpha)^{\frac{1}{\alpha}} = e.$

$$\lim_{W\to\infty} C_t = \lim_{\alpha\to 0} \frac{P_X}{N_0} \log_2 (1+\alpha)^{\frac{1}{\alpha}} = \frac{P_X}{N_0} \log_2 e$$

Continuous channel: example of telephone line

- Telephone signals are bandlimited to 3300 Hz.
- Using a bandwidth of 3300 Hz and a SNR of 20 db, we find the capacity of the telephone channel to be about 22000 bits/sec.
- Practical modems achieve transmission rates up to 19200 bits/sec.
- In real telephone lines, other factors, such as crosstalk, interference etc., also reduce the transmission rate.

It's only a beginning...

- Fading Channel
  - What if the channel is time varying?
- Multi-Input-Multi-Output (MIMO) Channel
  - What if there are multiple antennas?
- Multiple Access Channel (MAC)
  - What if there are multiple users?
- Packet Transmission
  - How to incorporate delay?

Summary

- Mutual information

$$\boxed{I(X;Y) = h(Y) - h(Y|X)}$$

- Continuous channel

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

- Additive noise channel

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

- Waveform channel

$$\boxed{C_t = \max_{p(x)} \left\{ \lim_{T\to\infty} [I(X;Y)] \right\} = \max_{p(x)} \left\{ \lim_{T\to\infty} [h(Y) - h(Y|X)] \right\}}$$

- Shannon formula

$$\boxed{C_t = W \log\left(1 + \frac{P_X}{P_N}\right)}$$

My Homepage

Thank you!

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