Principles of Communications

Meixia Tao
Shanghai Jiao Tong University

Chapter 9: Information Theory

Selected from: Chapter 12 of textbook

Reference:《通信原理》韩声栋: 第1.4节

Communication Systems Engineering: Ch 6.1, Ch 9.1~9.2

1916-2001

Claude Elwood Shannon

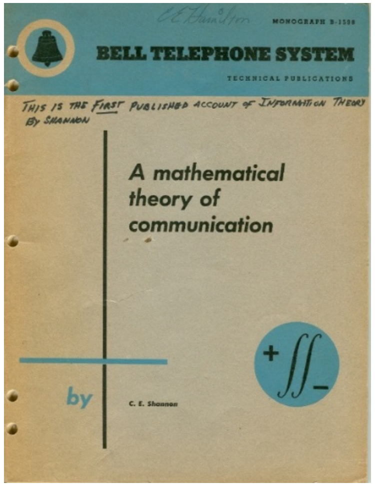
Father of Information Theory

Electrical engineer, mathematician, and native son of Gaylord. His creation of information theory, the mathematical theory of communication, in the 1940s and 1950s inspired the revolutionary advances in digital communications and information storage that have shaped the modern world.

This statue was donated by the Information Theory Society of the Institute of Electrical and Electronics Engineers, whose members follow gratefully in his footsteps.

Dedicated October 6, 2000
Eugene Daub, Sculptor

Information Theory



"A Mathematical Theory of Communication", *Bell System Technical Journal*. 1948

All about max-min problems in communications

- Information theory deals with fundamental limits on comm.
  - Channel transmission rate: What is the maximum rate at which information can be reliably transmitted over a communication channel?
  - Source compression rate: What is the minimum rate at which information can be compressed and still be retrievable with small or no error?
  - What is the complexity of such optimal schemes?

Topics to Discuss

- Modeling of information source
- Source coding theorem
- Modeling of communication channel
- Channel coding theorem

9.1 Modeling of Information Source

- Information source can be modeled by random process
- Discrete memoryless source (DMS)
  - A discrete-time, discrete-amplitude random process with *i.i.d* random variables
- A full description of DMS:
  - Alphabet set $\mathcal{A} = \{a_1, a_1, \ldots, a_N\}$ where the random variable X takes its values
  - Probabilities $\{p_i\}_{i=1}^{N}$

Information

- How to measure information?
- Examples:
  - "The sun will rise"
  - "It will rain tomorrow"
  - "Every one got 'A+' in the mid-term test"

  - The smaller the probability of an event is, the more information the occurrence of that event will convey

Measure of Information

- The information I that a source event X can convey and the probability of the event $P(X)$ satisfy:
  I. $I = I[P(X)]$
  II. $P(X)$ decreases $\rightarrow$ I increases, vice versa
  $P(X) = 1, I = 0$
  III. Consider multiple independent events X_1, X_2 , …
  $$I[P(X_1)P(X_2)...] = I[P(X_1)] + I[P(X_2)] + ...$$

- Information of symbol X

$$I = \log_a \frac{1}{P(X)} = -\log_a P(X)$$

  - $a = e$, nat
  - $a = 2$, bit

Entropy (熵)

- Consider a discrete source with N possible symbols
- Entropy $H(.)$: average amount of information conveyed per symbol

$$H(X) \overset{\Delta}{=} E\left[I(x_j)\right] = \sum_{j=1}^{N} P(x_j)\log_2 \frac{1}{P(x_j)} \left(\text{bit/symbol}\right)$$

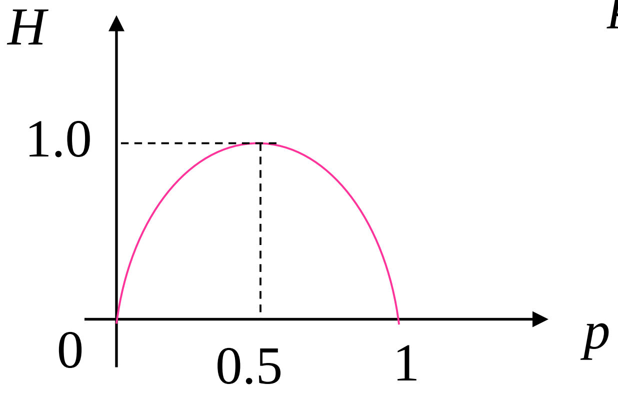
- ***Example***: Consider a discrete memoryless source having 3 symbols alphabet where $P(x_1) = \frac{1}{2}, P(x_2) = P(x_3) = \frac{1}{4}$. Determine the entropy of the source.
- Solution: $$H = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + p_3 \log_2 \frac{1}{p_3}$$

$$= \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{4}\times 2 = 1.5\text{bit/Symbol}$$

Entropy (Cont'd)

- What is the maximum entropy?
- Consider binary case $\{0, 1\}$ with $P(1) = p$ $P(0) = 1 - p$

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



Entropy is maximized when all the symbols are equiprobable

- N symbols: $H = \sum_{n=1}^{N} \frac{1}{N} \log_2 N = \log_2 N$ bit/symbol

Exercise

- A source with bandwidth $4KHz$ is sampled at the Nyquist rate
- Assuming that the resulting sequence can be modeled by a discrete memoryless source $\{-2,-1,0,1,2\}$ with probabilities $\left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right\}$
- What is the information rate of the source in bit/sec?

Solution

- We have

$$H(X) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + 2\times\frac{1}{16}\log_2 16$$
$$= \frac{15}{8} \text{ bits/sample}$$

- Since we have 8000 samples/sec, the source produces information at a rate of 15k bits/sec.

- Can it be reliably transmitted over a channel with rate 10 kbps?

Joint and Conditional Entropy

- The **joint entropy** of (X, Y) is defined as

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$$

- The **conditional entropy** of X given Y is defined as

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y)$$

- Using chain rule, it can be shown that

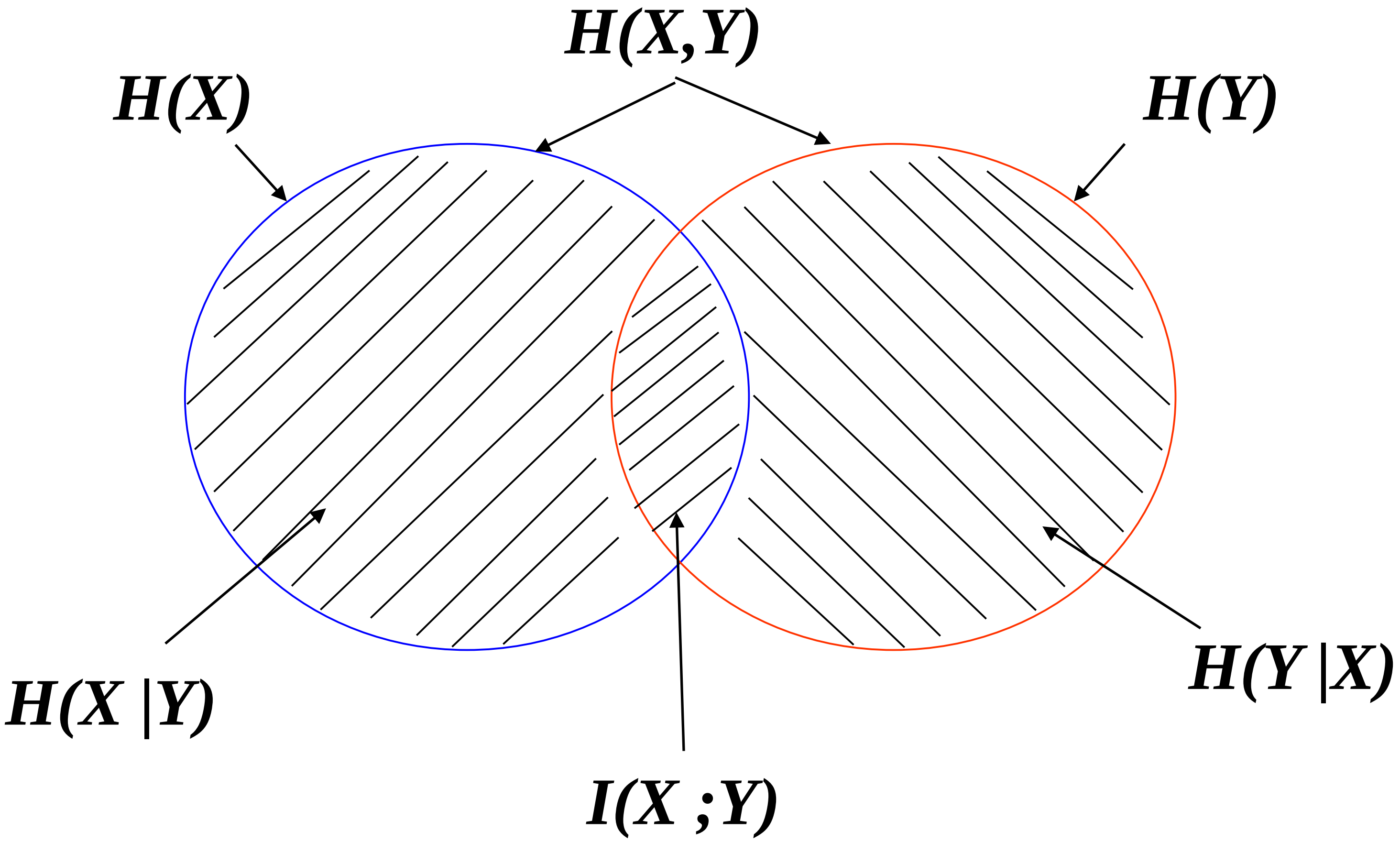
$$H(X,Y) = H(X|Y) + H(Y)$$

Mutual Information (互信息)

- Given that
  - $H(X)$ denotes the uncertainty of the random variable X
  - $H(X|Y)$ denotes the uncertainty of random variable X after random variable Y is known
- Then, $H(X) - H(X|Y)$
  - Denotes the amount of uncertainty of X that has been removed given Y is known
  - In other words, it is the amount of information provided by random variable Y about random variable X
- Definition of **mutual information**

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

Entropy, Conditional Entropy and Mutual Information



Differential Entropy

- The **differential entropy** of a discrete-time continuous alphabet source X with pdf f(x) is defined as:

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx$$

- Example: the differential entropy of $X \sim N(0, \sigma^2)$ is

$$h(X) = \frac{1}{2} \log_2 \left(2\pi e \sigma^2\right) \text{ bits}$$

- Mutual information between two continuous random variables X and Y:

$$I(X;Y) = h(X) - h(X|Y)$$

9.2 Source Coding Theorem

- **Source coding theorem**:
  - A source with entropy (or entropy rate) H can be encoded with an arbitrarily small error probability at any rate R (bits/source output) as long as $R > H$.
  - Conversely, if $R < H$, the error probability will be bounded away from zero, independent of the complexity of the encoder and decoder employed

H: the minimum rate at which an information source can be compressed for reliable reconstruction.

Huffman Source Coding

- Huffman coding is a variable-length binary coding.
- The idea is to map the more probable source sequences to shorter binary codewords
- Synchronization is a problem in variable-length coding
- Example:

Letter	Probability	Codewords			
		Code 1	Code 2	Code 3	Code 4
a_1	$p_1 = \frac{1}{2}$	1	1	0	00
a_2	$p_2 = \frac{1}{4}$	01	10	10	01
a_3	$p_3 = \frac{1}{8}$	001	100	110	10
a_4	$p_4 = \frac{1}{16}$	0001	1000	1110	11
a_5	$p_5 = \frac{1}{16}$	00001	10000	1111	110

Huffman code

9.3 Modeling of Communication Channel

- Recall that a communication channel is any medium over which information can be transmitted
- It is characterized by a relationship between its input and output, which is generally a stochastic relation due to the presence of fading and noise

Waveform (continuous-time) channel

↓ Sampling theorem

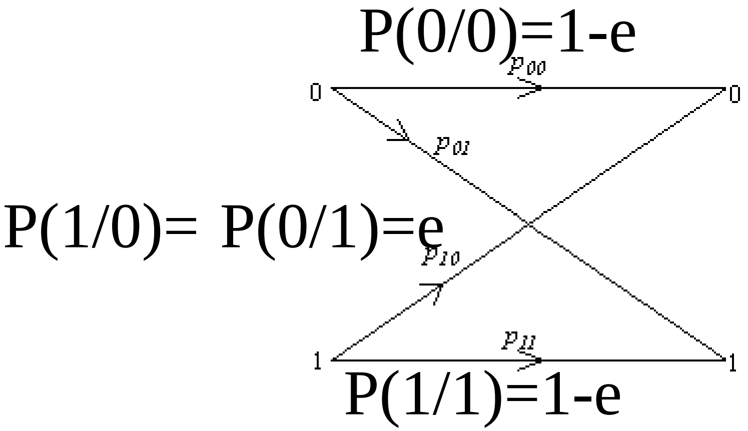
Discrete-time channel

Discrete-input discrete-output channel

continuous alphabet channel

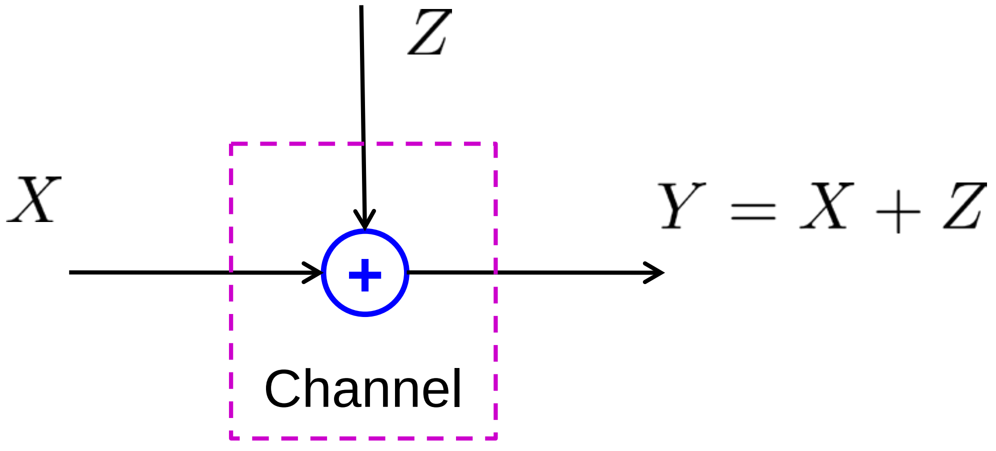
Binary-Symmetric Channel

- BSC channel is characterized by the crossover probability e=P(0|1)=P(1|0)
- For instance, $e = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$



AWGN Channel

- Both input and output are real numbers



- The input satisfies some power constraint

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \le P$$

9.4 Channel Capacity

- Channel capacity: a maximum rate, C in bits/sec of a channel
  - If $R \leq C$, theoretically guarantee almost error free transmission
  - If $R > C$, reliable transmission is impossible

- The capacity of a discrete-memoryless channel:

$$C = \max_{p(x)} I(X;Y)$$ (max over all possible input distribution)

The Noisy Channel Coding Theorem

Binary Symmetric Channel Capacity

- Since $I(X;Y) = H(Y) - H(Y|X)$

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

$$= H(Y) - \sum p(x) H(P_e)$$

$$= H(Y) - H(P_e)$$

$$\leq 1 - H(P_e)$$

- Here, $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$

  $H(Y) \leq 1$ Equality holds when X is equal probably

- Thus, the capacity of a BSC is

$$C = 1 - H(P_e)$$

Capacity of a BSC

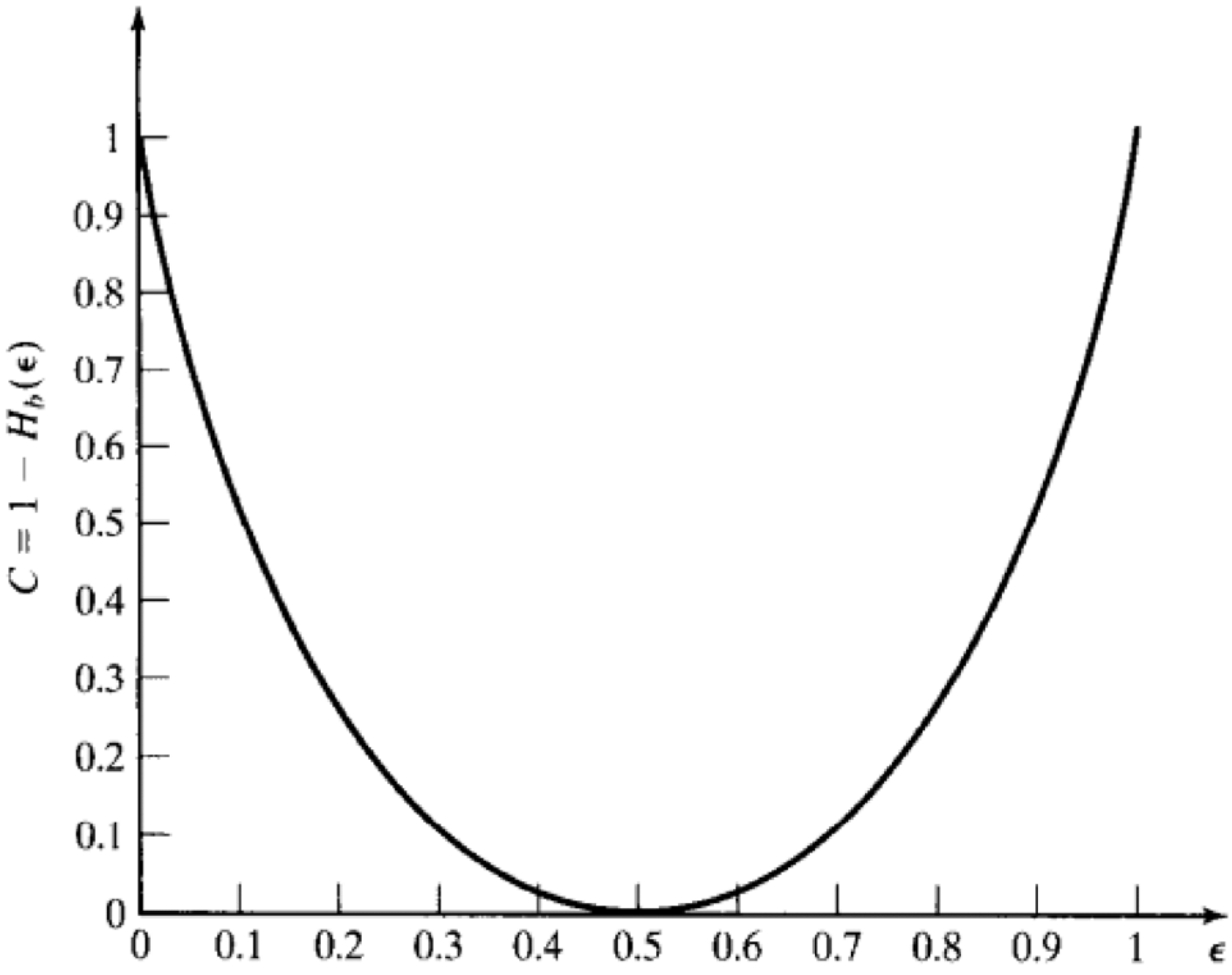


Figure 12.13 The capacity of a BSC.

Example

- A binary source with $P(X=0)=\frac{1}{4}$, $P(X=1)=\frac{3}{4}$ is to be transmitted over a BSC channel with a crossover probability p_e. Assume that the channel can be used once per symbol output.
- Determine the range of p_e for reliable communication of the source.

Gaussian Channel Capacity

- Consider a discrete-time Gaussian channel with

$$Y = X + Z$$

  - Input power constraint: $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$
  - $Z \sim N(0, P_N)$

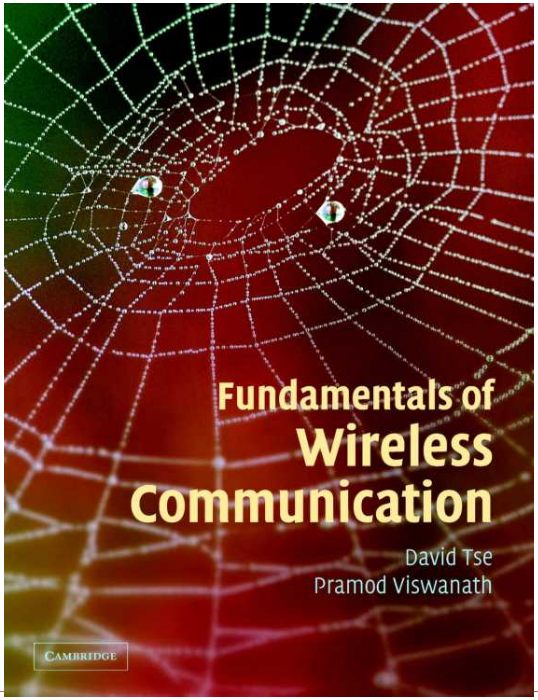
- Its capacity is given by

$$C = \frac{1}{2}\log\left(1+\frac{P}{P_N}\right) \quad \text{bits/channel use}$$

Proof of AWGN Channel Capacity

- A sketch from
  - http://www.eecs.berkeley.edu/~dtse/book.html

Fundamentals of Wireless Communications
Cambridge University Press, 2005



David Tse
UC Berkeley

Pramod Viswanath
UIUC

Proof

- AWGN channel model

$$y[m] = x[m] + w[m]$$

where m is the discrete time index and $w[m]$ is $\mathcal{N}(0, \sigma^2)$

- Using uncoded BPSK $x[m] = \pm\sqrt{P}$, the error prob. Is $Q\left(\sqrt{P/\sigma^2}\right)$
- Repetition code of block length N (average out the randomness of noise):

$$\boldsymbol{x}_A = \sqrt{P}[1, \cdots, 1] \qquad \boldsymbol{x}_B = \sqrt{P}[-1, \cdots, -1]$$

- The error prob is, which decays exponentially with the block length N

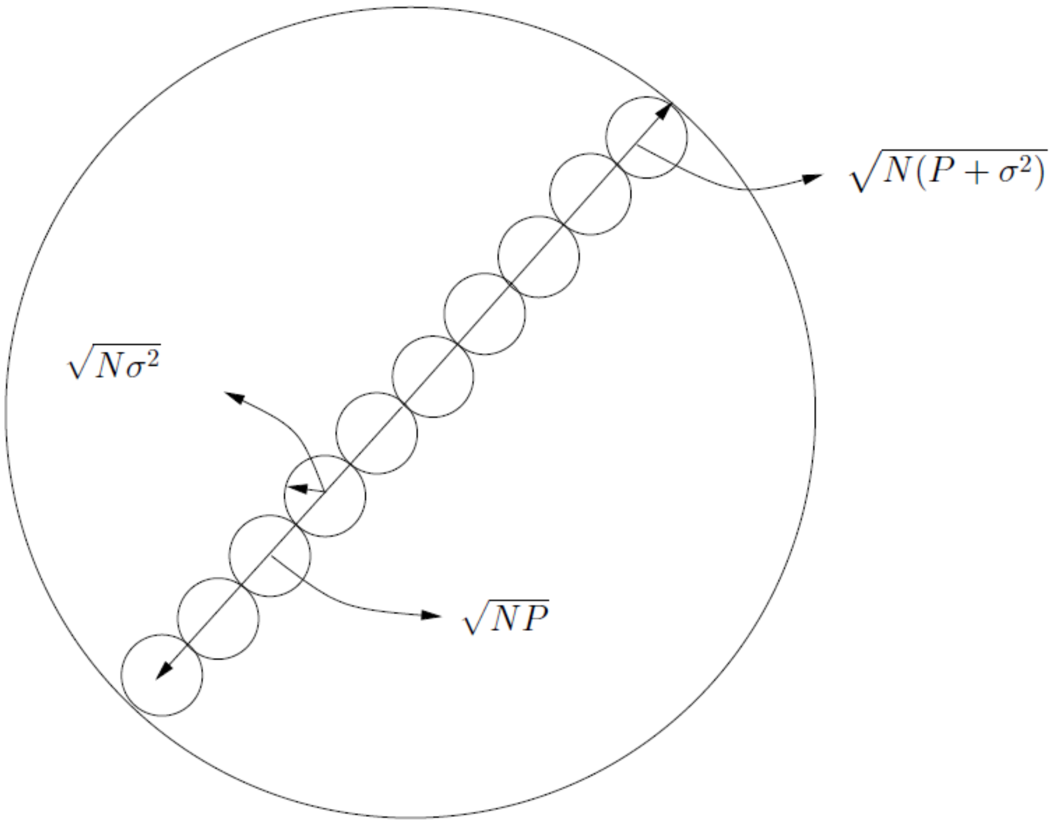
$$Q\left(\frac{\|\mathbf{x}_A - \mathbf{x}_B\|}{2\sigma}\right) = Q\left(\sqrt{\frac{NP}{\sigma^2}}\right)$$

☺ arbitrary reliability by choosing a large enough N

☹ data rate is only 1/N

Proof (cont'd)

- Repetition codes are on the same line
- Inefficient!

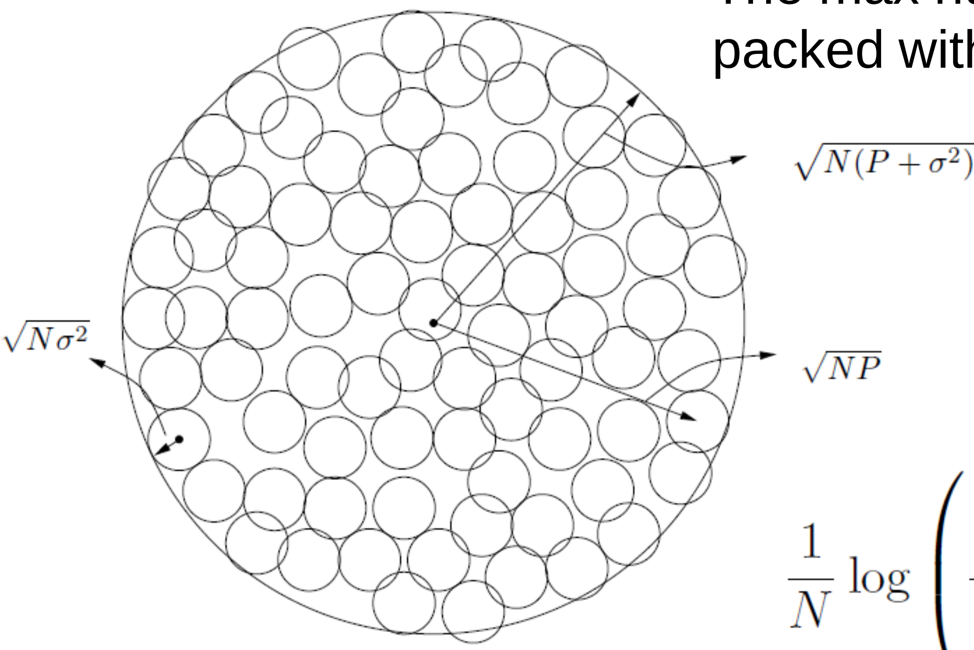


Proof (cont'd) – Sphere Packing

- By law of large number

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \xrightarrow{\text{with high prob}} \text{N-dim sphere of radius } \sqrt{N(P+\sigma^2)}$$

- y lies near the surface of a noise sphere of radius $\sqrt{N\sigma^2}$ around x



The max number of codewords that can be packed with no-overlap is: $\dfrac{\left(\sqrt{N(P+\sigma^2)}\right)^N}{\left(\sqrt{N\sigma^2}\right)^N}$

Thus, capacity is

$$\frac{1}{N}\log\left(\frac{\left(\sqrt{N(P+\sigma^2)}\right)^N}{\left(\sqrt{N\sigma^2}\right)^N}\right) = \frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right)$$

Notes

- The sphere-packing argument only yields the maximum number of codewords that can be packed while ensuring reliable communication
- How to construct codes to achieve the promised rate is another story
- From an engineering stand point, the essential problem is to identify easily encodable and decodable codes that have performance close to the capacity

Capacity of Bandlimited AWGN channel

- Consider a continuous-time, bandlimited AWGN channel with noise PSD = $N_0/2$, input power constraint P, bandwidth W.
- Sample it at Nyquist rate and obtain a discrete-time channel. The power/sample will be P and the noise power/sample will be

$$P_N = \int_{-W}^{W} \frac{N_0}{2} df = WN_0$$

- Thus,

$$C = \frac{1}{2}\log\left(1+\frac{P}{N_0 W}\right) \text{ bits/transmission}$$

- Since the number of transmissions/sec is 2W, we obtain the channel capacity in bits/sec

$$C = W\log\left(1+\frac{P}{N_0 W}\right) \text{ bits/sec}$$

(Shannon Formula)

Example

- Find the capacity of a telephone channel with bandwidth W=3000Hz, and SNR of 39dB

- Solution:
  - The SNR of 39 dB is equivalent to 7943. Using Shannon formula, we have

$$C = 3000 \log(1 + 7943) \approx \sim 38{,}867 \text{ bits/sec}$$

Insights from Shannon Formula

1. Increasing signal power P increases the capacity C
   - When SNR is high enough, every doubling of P adds additional B bits/s in capacity
   - When P approaches infinity, so is C

$$\begin{aligned} \log_2(1+x) &\approx x\log_2 e \qquad \text{when } x \approx 0, \\ \log_2(1+x) &\approx \log_2 x \qquad \text{when } x \gg 1. \end{aligned}$$

2. Increasing channel bandwidth W can increase C, but cannot increase infinitely (as noise power also increases)

$$\lim_{W\to\infty} C = \lim_{W\to\infty}\left[\frac{WN_0}{P}\log\left(1+\frac{P}{N_0 W}\right)\right]\frac{P}{N_0}$$

$$= \frac{P}{N_0}\log e = 1.44\frac{P}{N_0}$$

3. Bandwidth efficiency – energy efficiency tradeoff

- In any practical system, we must have

$$R \le W \log_2\left(1+\frac{P}{N_0 W}\right)$$

- Defining r=R/W, the spectral bit rate

$$r = \frac{R}{W} \le \log_2(1+\frac{P}{N_0 W})$$

- Let Eb be the energy per bit, $E_b = \frac{P}{R}$

- Then, $$r \le \log_2\left(1 + r\frac{E_b}{N_0}\right)$$

E_b/N_0 = SNR per bit

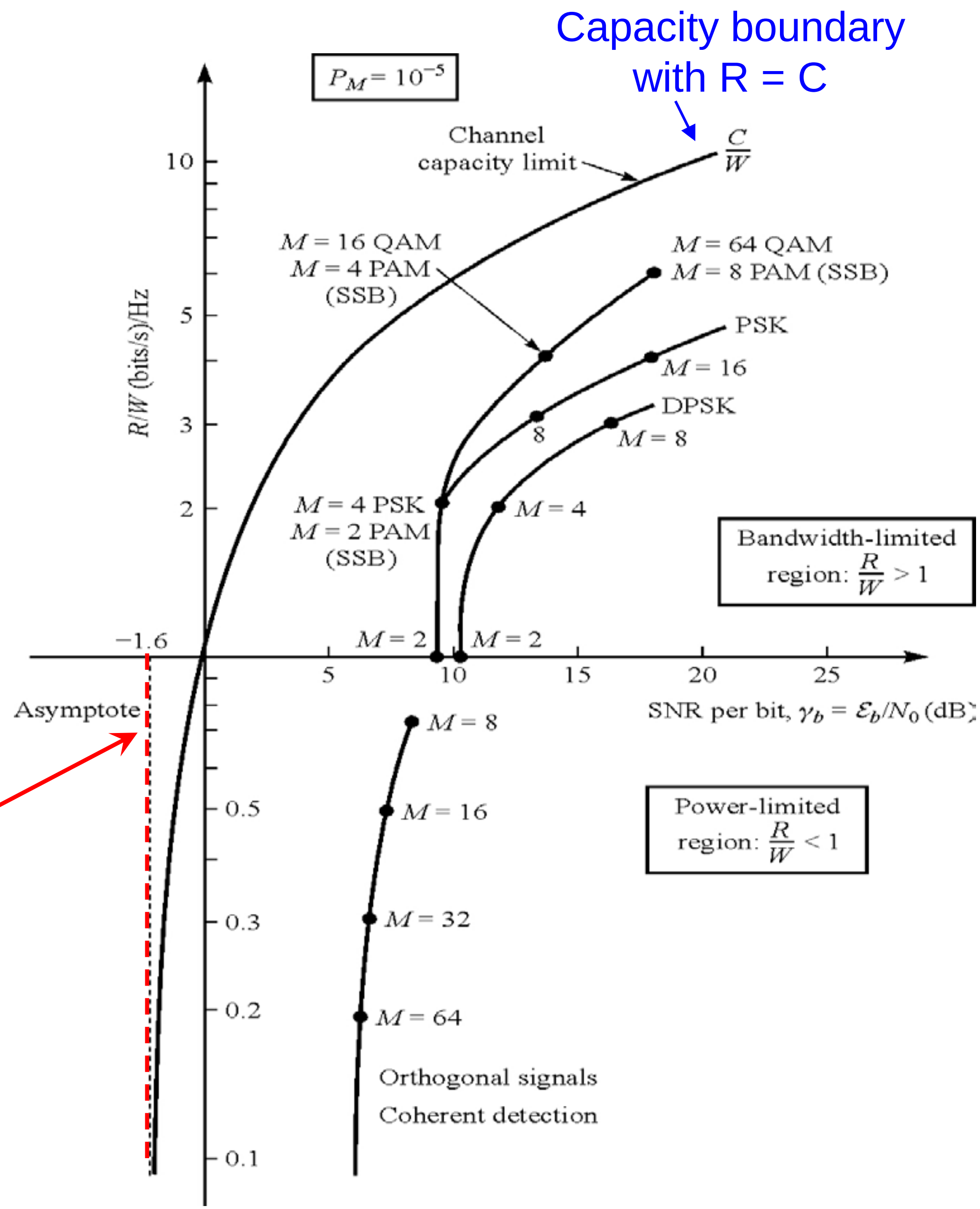
r = spectral efficiency

$$\frac{E_b}{N_0} = \frac{2^r - 1}{r}$$

- As $r = \frac{R}{B} \to 0$

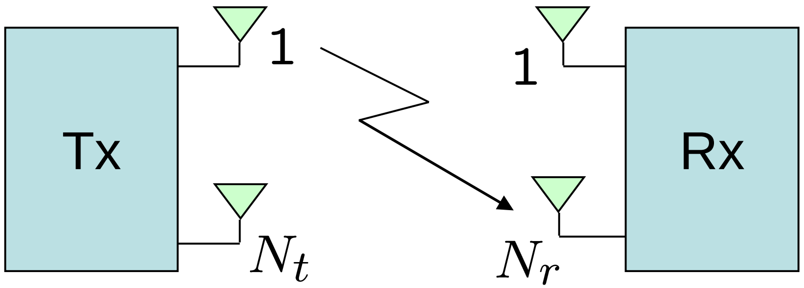
$$\left.\frac{E_b}{N_0}\right|_{r\to 0} = \lim_{r\to 0}\frac{1}{r}(2^r - 1)$$
$$= \ln 2$$
$$= 0.693$$
$$= -1.59\text{dB}$$

Shannon Limit, an absolute minimum for reliable communication



Picture from: Proakis J G, Salehi M. Fundamentals of communication systems[M]. Pearson Education India, 2007.

Capacity of MIMO channel



- Channel capacity [Foschini-Gans-98]

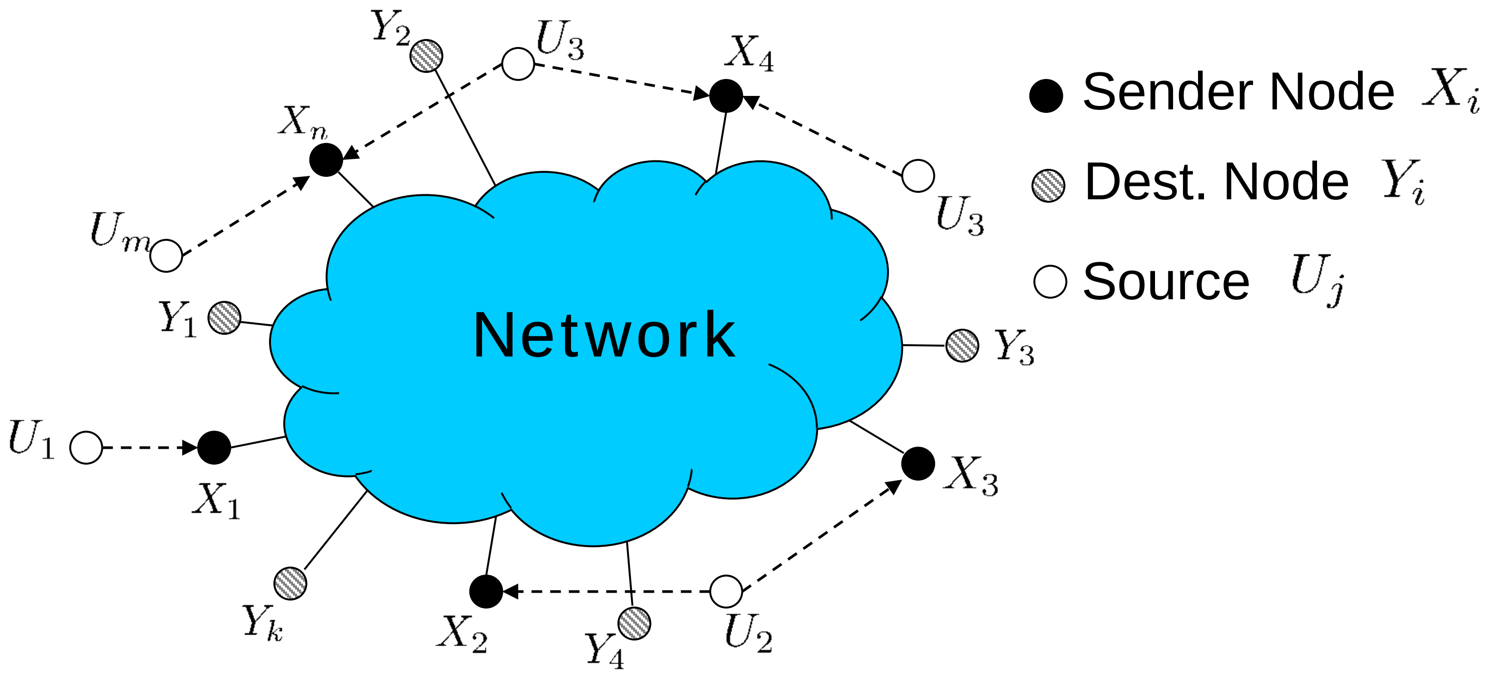
$$C = \max_{\mathrm{tr}(\mathbf{Q}) \leq \mathbf{1}} \log\det\left(\mathbf{I} + \mathrm{SNR} \cdot \mathbf{H}^{\mathbf{H}}\mathbf{Q}\mathbf{H}\right)$$

$$\sim \min(N_t, N_r)\log(\mathrm{SNR})$$

- Degrees of freedom (DoF)

$$d = \lim_{\mathrm{SNR} \to \infty} \frac{R(\mathrm{SNR})}{\log \mathrm{SNR}} = \min(N_t, N_r)$$

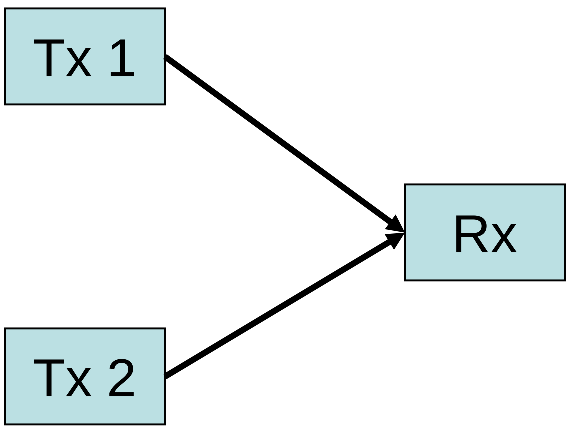
Network Information Theory

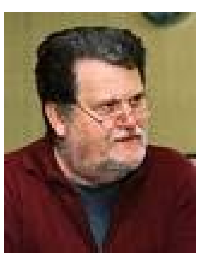


- Noisy channel coding
- Lossy source coding
- Joint source-channel coding

Basic Elements

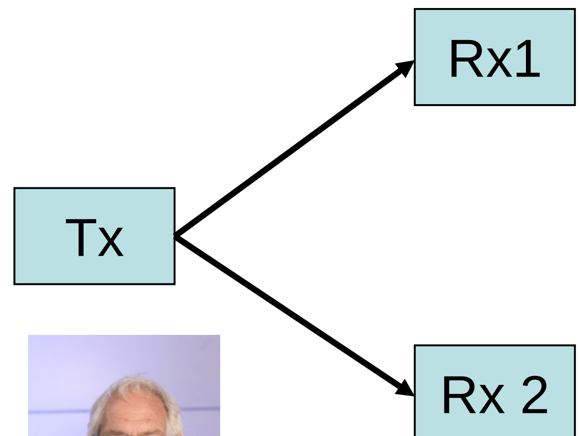
- Multiple Access Channel
  - Capacity is known for any number of users and for general channel models.





Ahlswede 71
Liao 72

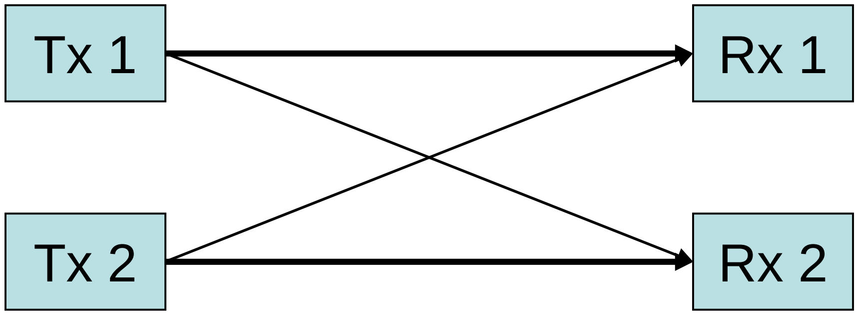
- Broadcast Channel
  - Partly known.



Cover 72

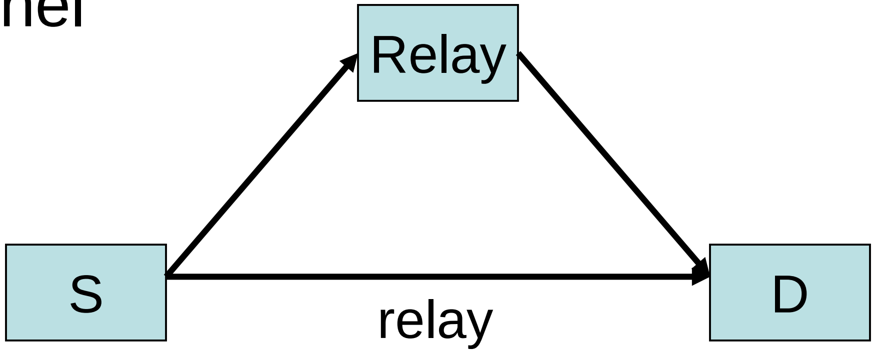
Basic Elements

- Interference channel



Best known achievable region: Han & Kobayashi 81

- Relay Channel



Best known achievable strategy: Cover & El Gamal 79