Fundamentals of Information Theory

Youlong Wu

ShanghaiTech University

wuyl1@shanghaitech.edu.cn

Reference Text books:

- T. M. Cover, J. A. Thomas, "Elements of Information Theory"
- A. B. Gamal, Y-H. Kim, "Network Information Theory"
- Claude E. Shannon, "A Mathematical Theory of Communication,"

Reference Materials:

- TV series, *Silicon Valley*, HBO, 2014
- David Tse, *The spirit of Information Theory*, 2017

Background knowledge:

Probability and stochastic process

Grades:
20% homework, 30% project, 50% final exam

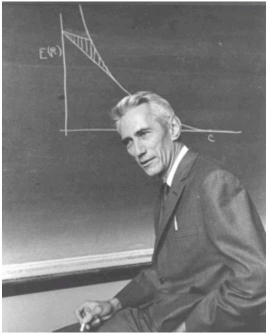
Office Hour: Room SIST building 2, 302D, Thursday, 3:00-5:00pm

Information Theory

Two fundamental questions:

- Q1: How to measure the quantity of information?
- Q2: How to send the information?

Ans: Information Theory (created by Claude E. Shannon in 1948)



The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

Why Study IT

- Help us better understand the fundamentals of communication.
- Tell us what we can do and what remains to do.
- Provide insights on scheme design.
- Have strong connections with other fields and provide deep insights
  - Computer science (e.g., machine learning, cross entropy, InfoGAN)
  - Game theory (e.g., horse race)
  - Bioinformatic (e.g, signal transmission among cells)
  - ...

What's Communications?

- Convey intended meanings from one entity or group to another through the use of mutually understood signs and semiotic rules
- e.g., Phone Call, Personal Talk, Lecture, WiFi transferring

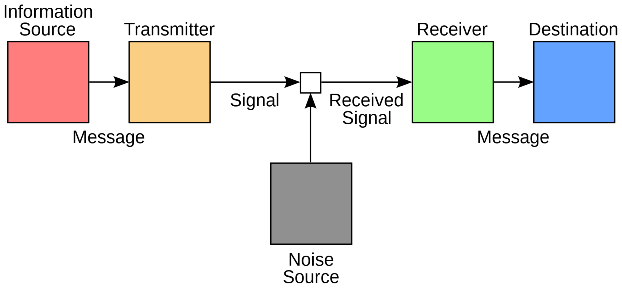
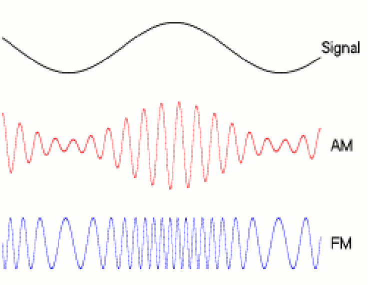


Figure: Communication System

Analog Communications

- Convey information using a continuous signal which varies in amplitude, phase, or some other property in proportion to that information.
- Methods: amplitude modulation (AM), frequency modulation (FM), ...

What's Digital Communications?

Use a digital sequence as an interface between the source and the channel

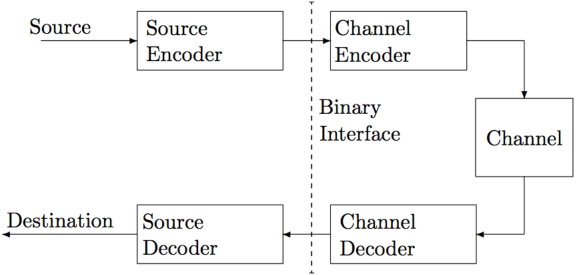


Figure: Separation of source and channel coding [Gallagar'Book]

Digital Communication Architechture

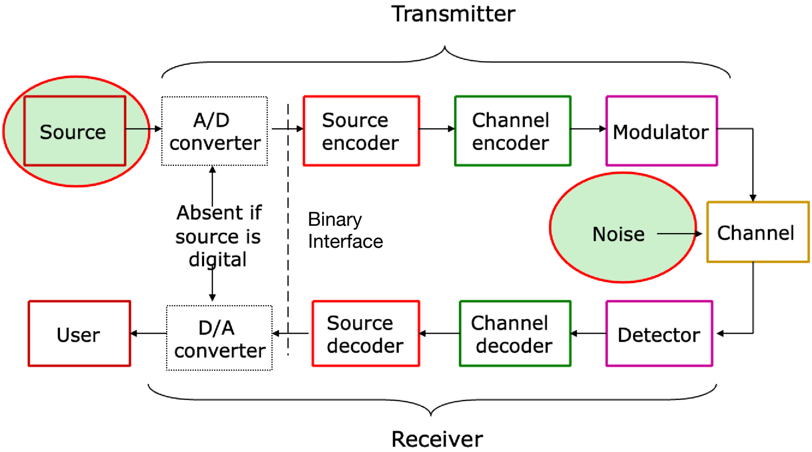
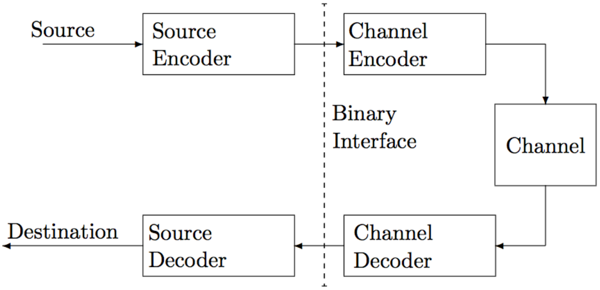


Figure: Architechture of digital communication

Why Need Digital Communications?

- Digital hardware has become so cheap, reliable and miniaturized
- Simplify implementation and understanding
- Security
- Doing this way won't decrease rate performance

Digital Communications System



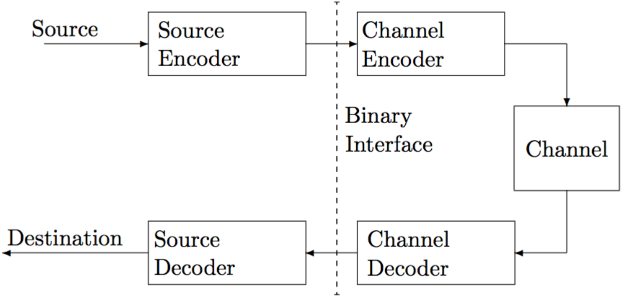
- Source
- Source coding
- Channel coding
- Binary/Digital interface
- Channel

Part 1: Source

Important Classes of Sources:

- Analog sources. E.g., voice, music, video and images etc. (We restrict to wave form sources, i.e. voice and music)
- Discrete sources: A sequence of symbols from a known discrete alphabet. E.g. English letters, Chinese characters, binary digits etc.

Part 2: Source Coding



Source Coding

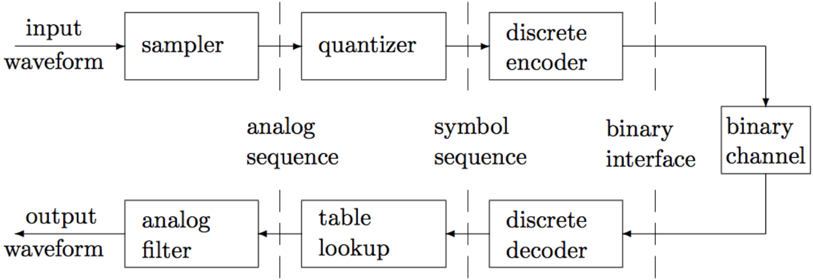
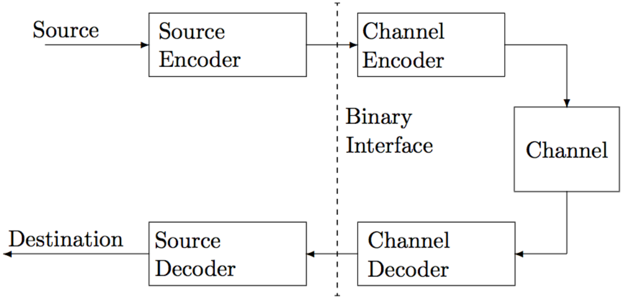


Figure: Layering of Source coding [Gallagar'Book]

- Sampling: Analog signal to sequence $u(t) \to u(mT)$
- Quantizer: Analog sequence into a symbols
- Encoder: Symbols to bits

Part 3: Channel Coding



Channel Coding

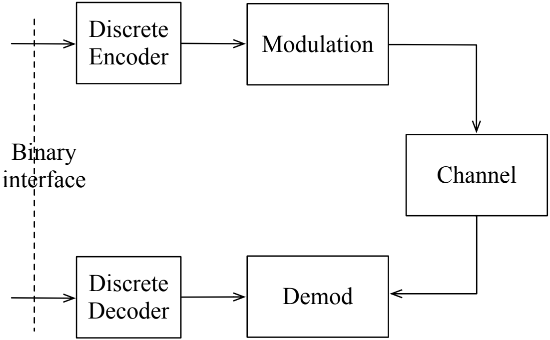
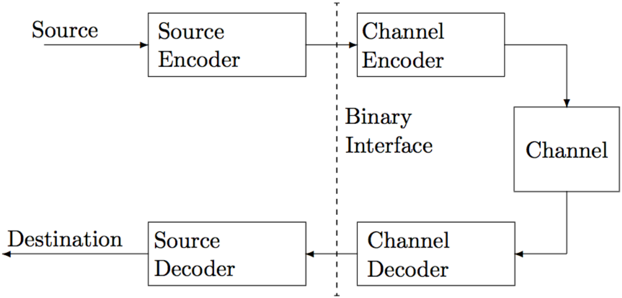


Figure: Layering of channel coding

- Discrete Enc: Add redundancy to improve reliability of communication
- Modulation: Maps the binary sequence at the source/channel interface into a channel waveform

Part 4: Digital/Binary Interface



Network Aspects of Digital Interface

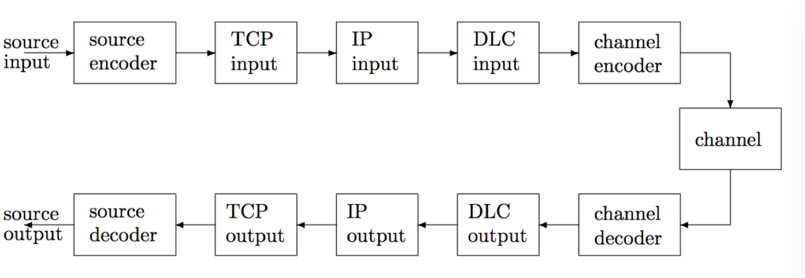
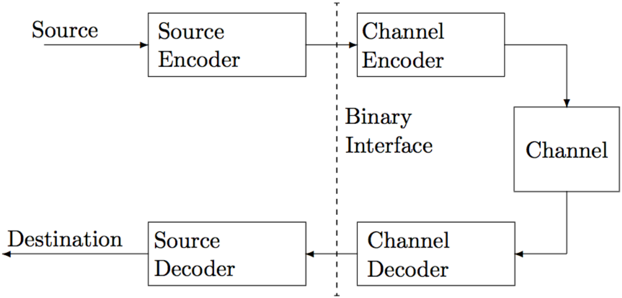


Figure: Binary interface for internet

DLC header	IP header	TCP header	Source encoded packet	DLC trailer

Figure: Structure of data frame

Part 5: Channel



Channel

Properties on channel:

- Channel is the part between the transmitter and receiver
- Channel is given (not under control of designer)
- Given the inputs, and outputs, the channel is a description of how the input affect the output. The description is usually probabilistic.

Types of channel:

- Wireless (e.g., 5G, WiFi) v.s. Wire (e.g., fiber channel)
- Memoryless (main focus) v.s. Memory
- Discrete v.s. Continuous

Discrete Memoryless Channel (DMC)

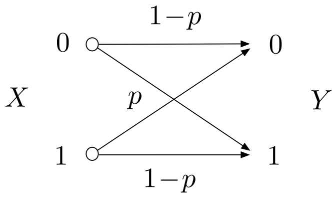


Figure: Binary symmetry channel

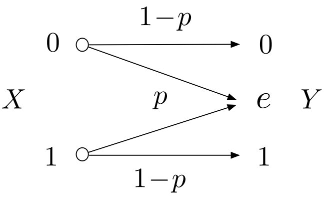
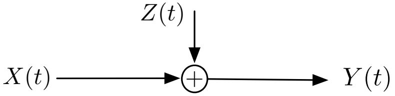


Figure: Binary erasure channel

Continuous Channel

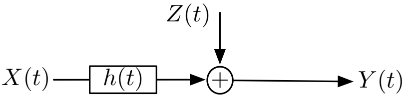
Given Gaussian noise $Z(t)$:

- Additive white Gaussian noise (AWGN) channel:



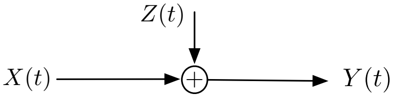
$$Y(t) = X(t) + Z(t)$$

- Linear Gaussian channel (with linear filter $h(t)$):



$$Y(t) = X(t) * h(t) + Z(t)$$

AWGN Channel



$$Y(t) = X(t) + Z(t)$$

For the AWGN channel with bandwidth W, the capacity (in bps) is

$$C = W \log_2 \left(1 + \frac{P}{N_0 W}\right)$$

- This is the ultimate, but it is essential achievable in practice
- Wireless channels have added complications
- Multiple physical paths from input to output
- the relative path lengths change, causing multipath fading

Outline

- Probability and Stochastic Process
- Entropy, Relative Entropy, and Mutual Information
- Asymptotic Equipartition Property
- Entropy Rates of a Stochastic Process
- Data Compression
- Channel Capacity
- Differential Entropy
- Gaussian Channel
- Rate Distortion
- Network Information Theory
- Projects and Presentation

Preliminary: Probability

Given events A, B

- $P(A), P(\bar{A})$,
- $P(A \cap B)$ or $P(A, B)$, $P(A \cup B)$
- $P(A|B) = P(A \cap B)/P(B) = P(A, B)/P(B)$

Preliminary: Probability

Given events A, B

- $P(A), P(\bar{A})$,
- $P(A \cap B)$ or $P(A, B)$, $P(A \cup B)$
- $P(A|B) = P(A \cap B)/P(B) = P(A, B)/P(B)$

Two important properties

- $P(A \cup B) = P(A) + P(B)$, **only if** $A \cap B = \emptyset$
- $P(A|B) = P(A)$, **only if** A and B are *independent*

Preliminary: Probability

Given events A, B

- $P(A), P(\bar{A})$,
- $P(A \cap B)$ or $P(A, B)$, $P(A \cup B)$
- $P(A|B) = P(A \cap B)/P(B) = P(A, B)/P(B)$

Two important properties

- $P(A \cup B) = P(A) + P(B)$, **only if** $A \cap B = \emptyset$
- $P(A|B) = P(A)$, **only if** A and B are *independent*

Questions:

- $A \cap B = \emptyset$ v.s. $A \perp B$
- For $P(A|B)$ and $P(A)$, which one is larger?
  (hint: check when $A \cap B = \emptyset$ and $A \cap B = B$)

Q1: How to measure the quantity of information?

Example (*Football Games*):

China sucks at playing soccer, while France and Brazil both are very good at it. Which game result below contains more uncertainty?

- China V.S. Brazil
- France V.S. Brazil

$\boxed{\text{The more uncertain an event is, the more information it contains.}}$

Q1: How to measure quantity of information?

Consider binary RVs X_1 and X_2, with

$$P(X_1 = 0) = 1, \quad P(X_1 = 1) = 0$$
$$P(X_2 = 1) = P(X_2 = 0) = 1/2$$

X_1 is deterministic (no uncertainty), X_2 has uncertainty.

In communication, we send information X_1 and X_2

- How to measure the quantity of information?

Entropy

Assume $X \in \mathcal{X}$, and $|\mathcal{X}| = M$

$$\boxed{H(X) = \sum_{x \in \mathcal{X}} -p(x) \log_2(p(x))}$$

- $H(X) \geq 0$. Equality holds if X is constant
- $H(X) \leq \log_2 M$. Equality holds if X is equiprobable

Small test:

- Given $P(X=0) = 1/4, P(X=1) = 1/4, P(X=3) = 1/2$, $H(X) =?$
- Which distribution $\{P_X(x_i)\}$ can maximize $H(X)$?