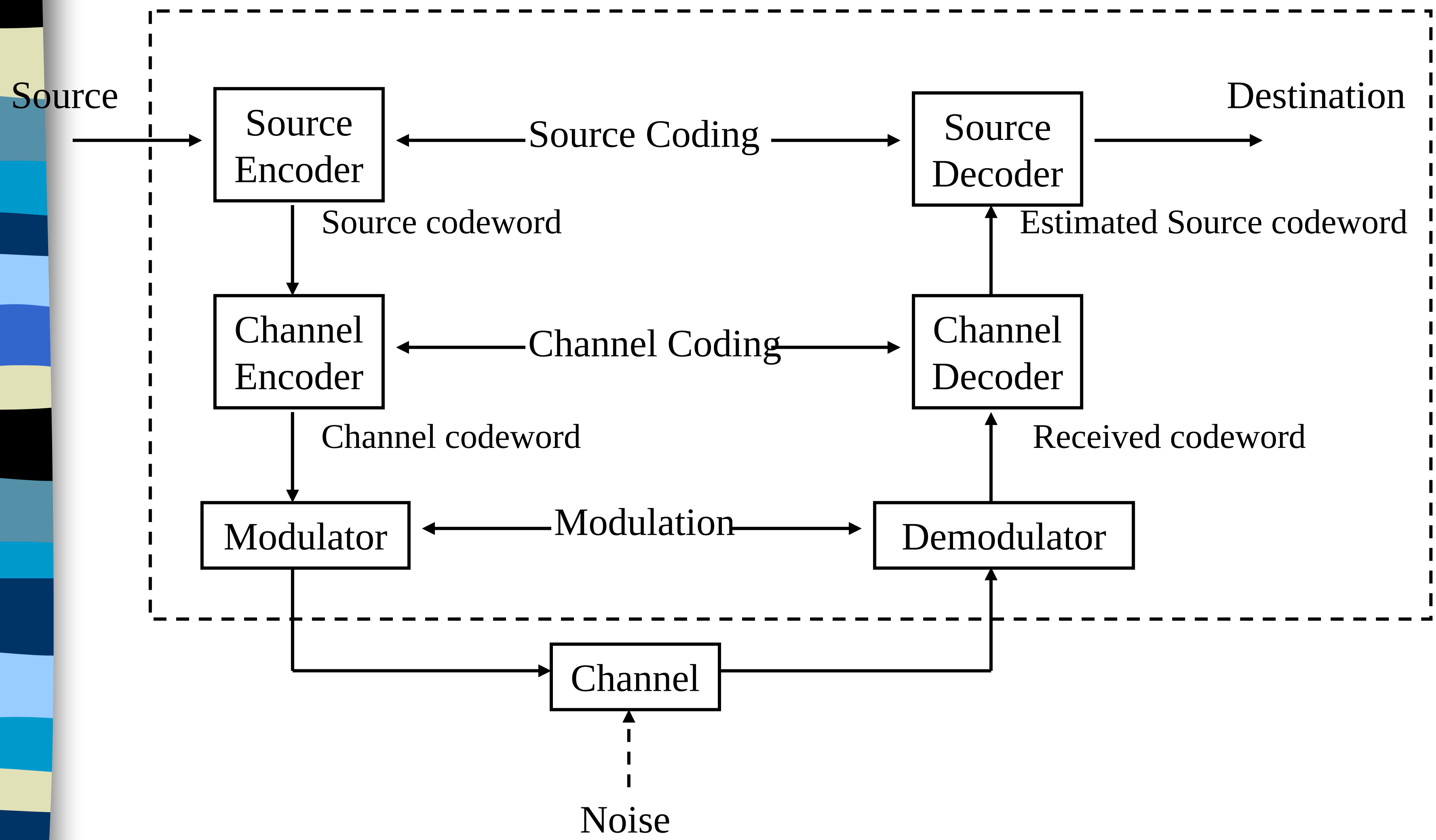
Information Theory and Coding Techniques

Prof. Ja-Ling Wu

Department of Computer Science
and Information Engineering
National Taiwan University

Digital Communication System



Source Coding

Based on characteristics/features of a source, Source Encoder-Decoder pair is designate to reduce the source output to a Minimal Representation.

[Shannon 1948]

How to model a signal source? ← Entropy, Random process

How to measure the content of a source?

How to represent a source? Code-design

How to model the behavior of a channel?

Stochastic mapping

channel capacity

Source Coding (cont.)

Redundancy Reduction $\rightarrow$ Data Compression

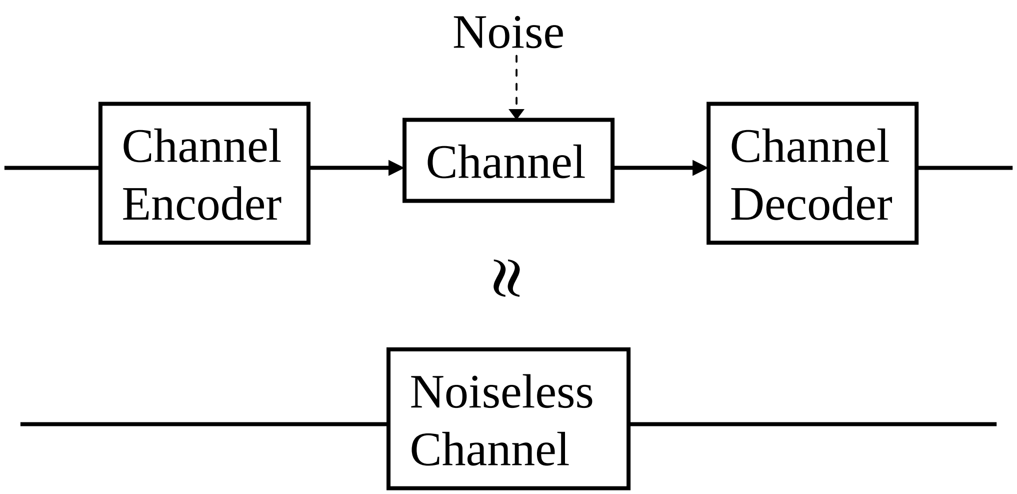
Data Compaction

Modalities of Sources:

- Text
- Image
- Speech/Audio
- Video
- Hybrid

Channel coding

Introduction redundancy into the channel encoder and using this redundancy at the decoder to reconstitute the input sequences as accurately as possible, i.e., channel coding is designate to minimize the effect of the channel noise.



Modulation

Physical channels can require electrical signals, radio signals, or optical signals. The modulator takes in the channel encoder/source encoder outputs and output waveforms that suit the physical nature of the channel and are also chosen to yield either system simplicity or optimal detection performance.

What is information?

- **What is meant by the "information" contained in an event?**

  If we are formally to defined a quantitative measure of information contained in an event, this measure should have some intuitive properties such as:

  1. **Information contained in events ought to be defined in terms of some measure of the uncertainty of the events.**
  2. **Less certain events ought to contain more information than more certain events.**
  3. **The information of unrelated/independent events taken as a single event should equal the sum of the information of the unrelated events.**

A nature measure of the uncertainty of an event α is the probability of α denoted $P(\alpha)$.

Once we agree to define the information of an event α in terms of $P(\alpha)$, the properties (2) and (3) will be satisfied if the information in α is defined as

$$I(\alpha) = -\log P(\alpha)$$

Self-information

∗ The base of the logarithm depends on the unit of information to be used.

Information (Source)

$S_1 \quad S_2 \quad \cdot \quad \cdot \quad \cdot \quad S_q$: Source alphabet

$P_1 \quad P_2 \quad \cdot \quad \cdot \quad \cdot \quad P_q$: Probability

Facts:

1) The information content (surprise) is somewhat inversely related to the probability of occurrence.
2) The information content from two different independent symbols is the sum of the information content from each separately. Since the probability of two independent choices are multiplied together to get the probability of the compound event, it is natural to define the amount of information as

$$I(S_i) = \log \frac{1}{P_i} \qquad \left(\text{or} - \log P_i\right)$$

As a result, we have

$$I(S_1) + I(S_2) = \log \frac{1}{P_1 P_2} = I(S_1, S_2)$$

Information Unit:

$\log_2$: bit

$\log_e$: nat

$\log_{10}$: Hartley

base conversions:

$\log_{10}2 = 0.30103$, $\log_2 10 = 3.3219$

$\log_{10}e = 0.43429$, $\log_e 10 = 2.30259$

$\log_e 2 = 0.69315$, $\log_2 e = 1.44270$

$$\log_a X = \frac{\log_b X}{\log_b a} = (\log_a b)\log_b X$$

Entropy: Average information content over the whole alphabet of symbols

$$H_r(S) = \sum_{i=1}^{q} P_i \log_r\left(\frac{1}{P_i}\right) \qquad \begin{Bmatrix} S_1 & S_2 & \cdots & S_q \\ P_1 & P_2 & \cdots & P_q \end{Bmatrix}$$

$$= -\sum_{i=1}^{q} P_i \log_r P_i$$

$$H_r(S) = H_2(S)(\log_r 2)$$

- Consider the entropy of the Source can have no meaning unless a model of the Source is included. For a sequence of numbers and if we cannot recognize that they are pseudo-random numbers, then we would probably compute the entropy based on the <u>frequency of occurrence</u> of the individual numbers.

* The entropy function involves only the distribution of the probabilities — it is a function of a probability Distribution P_i and does not involve the S_i

Ex: Weather of Taipei

$X = \{Rain, fine, cloudy, snow\} = \{R, f, c, s\}$

$P(R) = ¼, P(F) = ½, P(C) = ¼, P(S) = 0$

$H_2(X) = 1.5$ bits/symbol

If ¼ for each P(i) ⇒ $H_2(X) = 2$ bits/symbol. (>1.5)

(equal probability event)

$H(X) = 0$ for a certain event P(ai)=0 or P(ai)=1

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number.

2. It is nearer to the feeling of a humanbody

   Intensity — eve

   volume — ear

3. It is mathematically more suitable

   $\log_2$ — bits

   $\log_{10}$ — decimal digits

   $\log_e$ — natural unit

Change from the base a to base b merely requires multiplication by $\log_b a$

Course contents

- Basic Information Theory:
  - Entropy, Relative Entropy and Mutual Information
- Data Compression / Compaction:
  - Kraft Inequality, the prefix condition and Instantaneous decodable codes.
- Variable Length Codes
  - Huffman code, Arithmetic code and L-Z code.
- Coding Techniques
  - DPCM (predictive coding)
  - Transform coding
  - Motion Estimation
  - JPEG
  - MPEG-1, 2, 4
  - H.26P

  …
- Steganography and Information Hiding
  - Digital Watermarking