CHAPTER 1: INTRODUCTION TO INFORMATION THEORY

By
Dr. Ebenezer Anohah
Department of Computer Science and Engineering
University of Mines and Technology

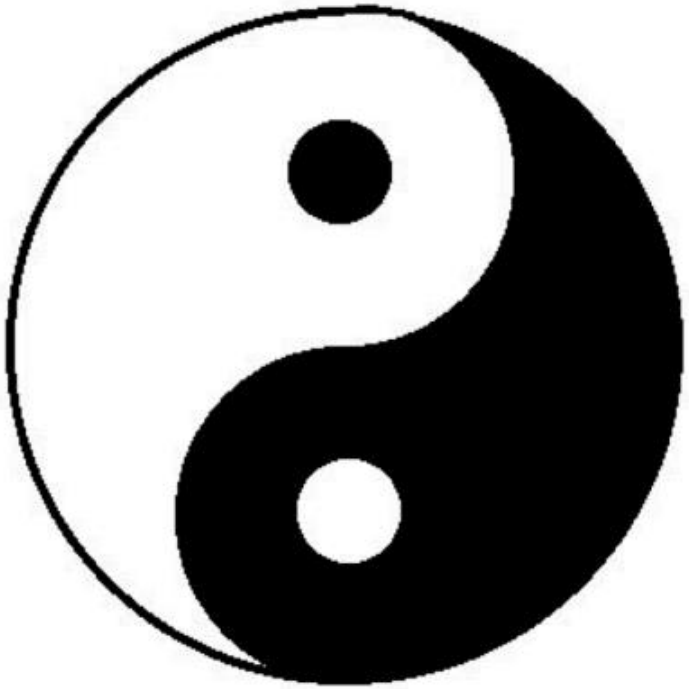
What is this theoretical course about?

- The purpose of this theoretical course is to enable students have fundamental insights into modern communication systems
- Students are system designers and design codewords for communication systems
- This course of information theory demands that students have basic knowledge in algebra, whilst logarithm and probabilities are used extensively.
- This is because algebra, calculus, logarithm functions and probabilities play important role in understanding and explaining communications.

What is information?

- We can say that information is fundamental, abstract notion.
- Information has sense only when involves two correspondents: one generating it (information source S) and another receiving it (the destination D, or the user U)
- Information is an abstract concept that can be transmitted through physical form (current, voltage, electromagnetic wave), that is by a signal, at distance or stored (memorized) for later reading.
- The physical medium e.g. fiber optic cables, including telephone and computer networks, that ensures remote transmission of the information from S to D, is called transmission channel; in the case of storage systems such as magnetic and optical disk drives the channel is replaced by the storage medium, e.g. CD, DVD, Tape etc.

How to quantify information?



Small information content

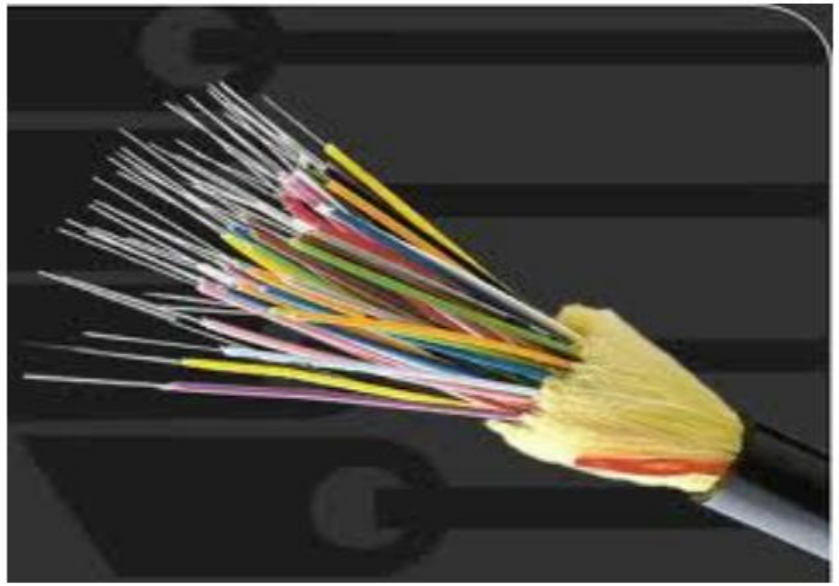
Large information content

Assuming we accidently throw a laptop or compact disc in this black hole, which has higher information loss?

What is the fundamental limit of information transfer rate?

- As indicated below, every communication channel has different information capacity to limit information transfer rate

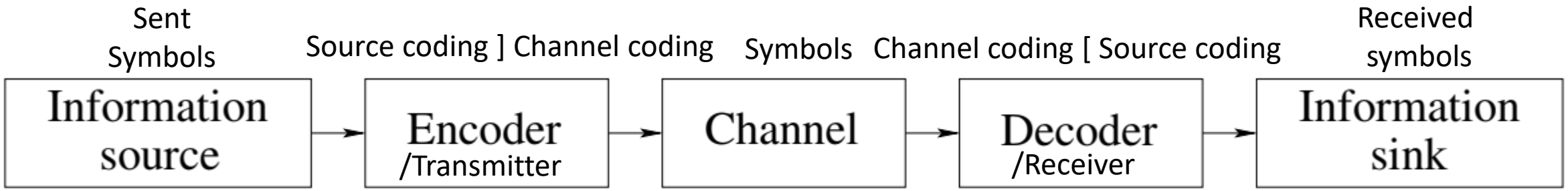
WIFI: Information rate$\sim Mbit/s$



Fiber Optics: Information rate $\sim Tbit/s$

Model and basic operations of information processing systems

- Communication and storage systems can be regarded as examples of information processing systems and may be represented abstractly by the model depicted in the figure below;



- Compression (source coding), (that is, removes inherent redundancy), (in practicality uses compression algorithm like ZIP).
- Source Entropy (maximum information to be transferred)
- Transmission Rate (within channel capacity) vs Distortion (losses in bits)
- Error Correction(adds bits to combat channel noise effect)
- Channel Capacity (Maximum amount of transmission rate in bits per channel)
- Capacity vs Efficiency
- Decompression (that is, adds controlled redundancy)

What is Information source?

- Information source can be discrete (digital source), or continuous (signal source).
- The discrete source generates a finite number of symbols (e.g. 0 and 1 used in digital communications)
- The continuous source, includes an infinite number of symbols (e.g. voice, television signal, measurement and control signals)
- Generally, digital sources in contrast to continuous sources are widely used as information source because of their two levels corresponding “0” and “1” they have high noise immunity.

What is digital Encoding?

- The information source output is processed by an encoder to facilitates the transmission of the information.
- There basic operations can be executed in the encoder: source coding, channel coding, and modulation.
- For the source coding, encoder maps the source output into digital format.
- For the channel coding, the encoder introduces extra redundant data to combat the noisy environment (caused by imperfect hardware, lightening, voltage fluctuations, old-high-resistance wires, sudden surge-in-temperature, interference from machines that generate strong electromagnetic fields) in which information must be transmitted or stored.

What is Encoding?

- We need proper modulation to convert the data after source and channel coding to waveforms that are suitable for transmission or recording.
- In communication systems, the encoder function is often called a transmitter, while in storage systems we usually speak of a recorder

What are communication channels?

- The output of the encoder is then transmitted through some physical communication channel e.g. fiber optic cables or copper cables, wireless radio transmission based on electromagnetic waves (in the case of a communication system) or stored in some physical storage medium e.g. magnetic tape, a hard drive and optical storage disks such as CD-ROM or DVD (in the case of a storage system)

Some other communication channels?

modem $\rightarrow$ phone line $\rightarrow$ modem

Galileo $\rightarrow$ radio waves $\rightarrow$ Earth

finger tips $\rightarrow$ nerves $\rightarrow$ brain

parent cell $\rightarrow$ daughter cells

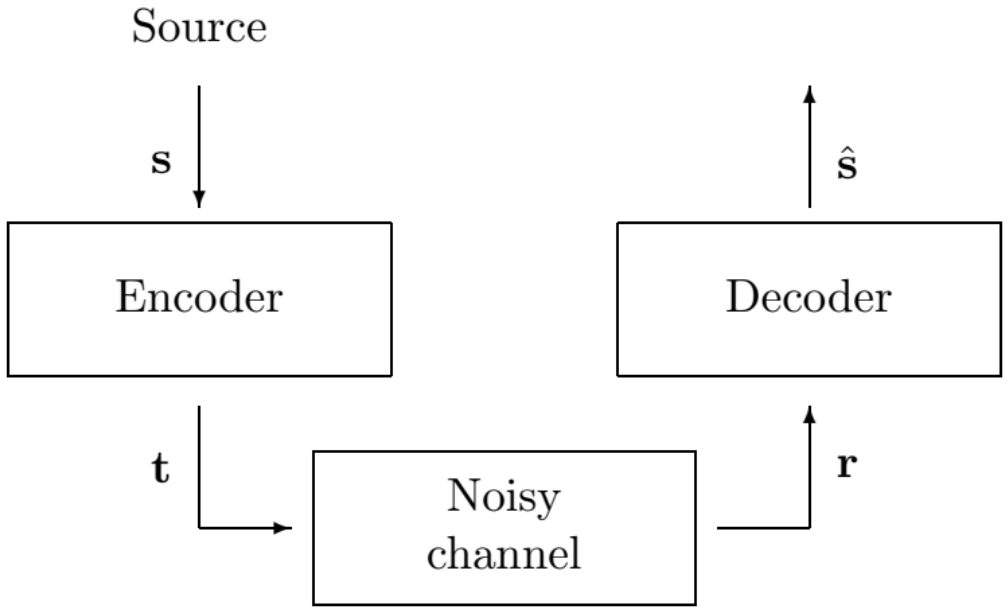
computer memory $\rightarrow$ disk drive $\rightarrow$ computer memory

Ways of improving communication noisy channels by physical characteristics

- Using more reliable hard disc drives
- Using higher-power signals or cooling circuitry systems in order to reduce thermal noise
- These physical modifications typically increase the cost of the communication channel

Ways of improving communication noisy channels by system characteristics

- Adding communication system to channels so that we can detect and correct errors introduced by the channels
- As shown in figure below we add an encoder before the channel and a decoder after it. This computational requirement doesn't increase cost of channels.



What is decoding?

- Information conveyed through (or stored in) the channel must be recovered at the decoder and processed to restore its original form.
- The signal processing performed by the decoder can be viewed as the inverse of the function performed by the encoder.
- The physical channel usually produces a received signal which differs from the original input signal. This is because of signal distortion and noise introduced by the channel. Consequently, the decoder can only produce an estimate of the original information message.

Information sink

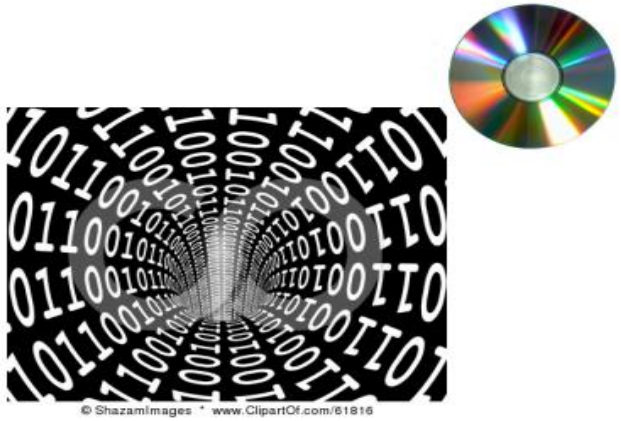
- The output of the decoder is then presented to the final user, which we call the information sink.

Definitions of Information theory?

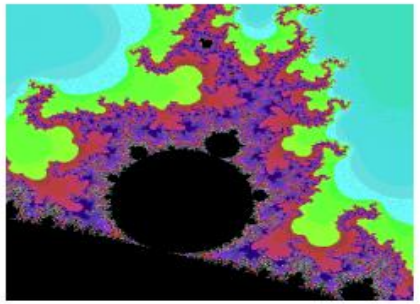
- Information theory is concerned with the transmission of information from a sender (termed a source), through a communication channel, to a receiver.
- The information is sent in units called symbols (normally bits) and all set of all possible data symbols is a source alphabet.
- Generally speaking, an information theory is an abstract system of concepts with indications of relationships among these concepts that help us to understand, explain and predict something in communication systems.
- These concepts are words or terms that label the most important elements in information theory.
- These concepts may be nominal or real. Nominal concepts are those that are not observable, such as binary digits in communication systems. Real concepts are observable, such as sent or received messages in communication systems.
- These relationships specify the ways in which the concepts in the information theory are combined.

Information theory is about these

Data Compression



Coding



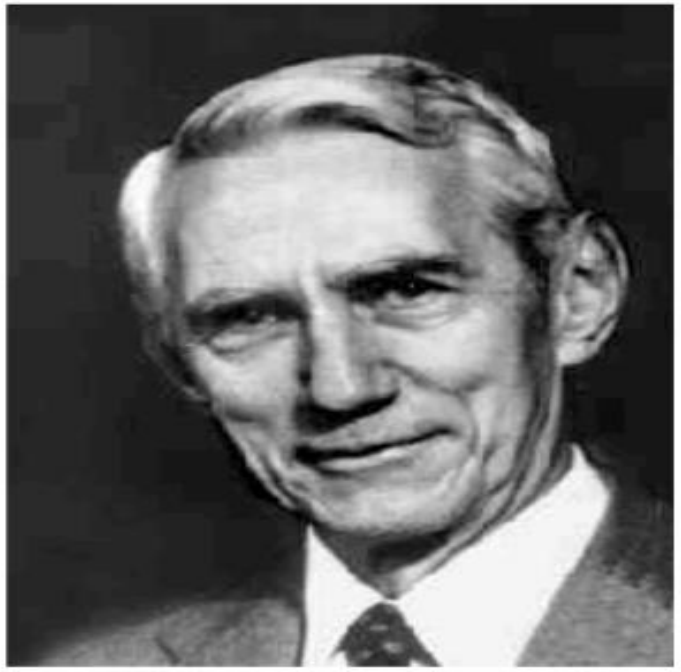
Computation: Kolmogorov Complexity

Data Communication

Information theory addresses and answers what two fundamental questions?

- What is the ultimate data compression?
- (answer: the entropy of the data, H, is its compression limit.)
- What is the ultimate rate of reliable communication?
- (answer: the channel capacity, C, is its transmission rate limit.)

Where Information Theory begun?



Shannon, 1916 - 2001

Claude Shannon's information theory

- Claude Shannon's information theory deals with limits on data compression (source coding) and reliable data transmission (channel coding)
- Source coding theorem and Channel coding theorem (1948)
- In the past, information theory was considered an esoteric theory with no apparent relation to the "real world"
- However, recent advances in technology such as algorithms, hardware and software are practical schemes for
  - data compression
  - transmission and modulation
  - error correcting coding
  - compressed sensing techniques
  - Information security

Relationship of information theory to other fields of studies