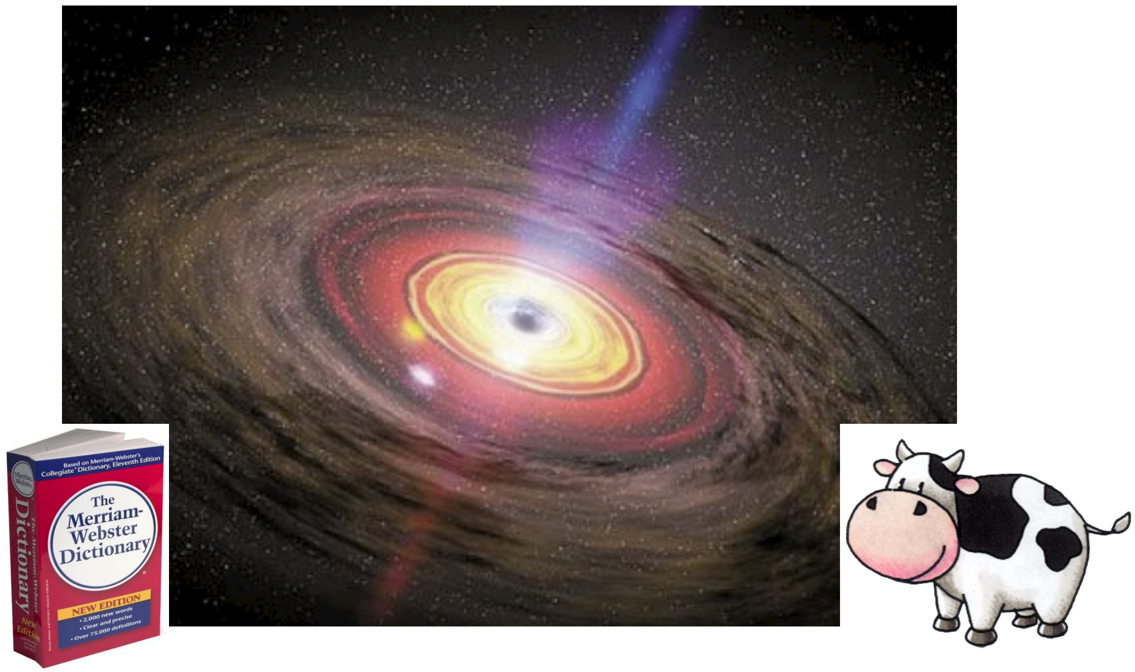
Lecture 1. Introduction

- What is the course about?
- Logistics
- Questionnaire

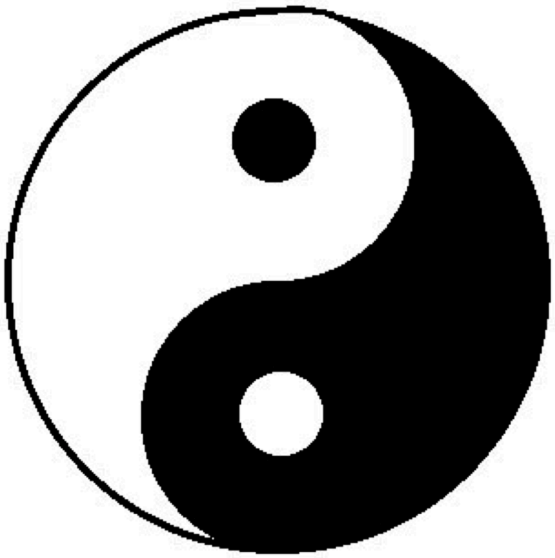
What is information?

Throw a cow or a dictionary into a black hole, which has higher information loss?

- Tom Cover



How to quantify information?

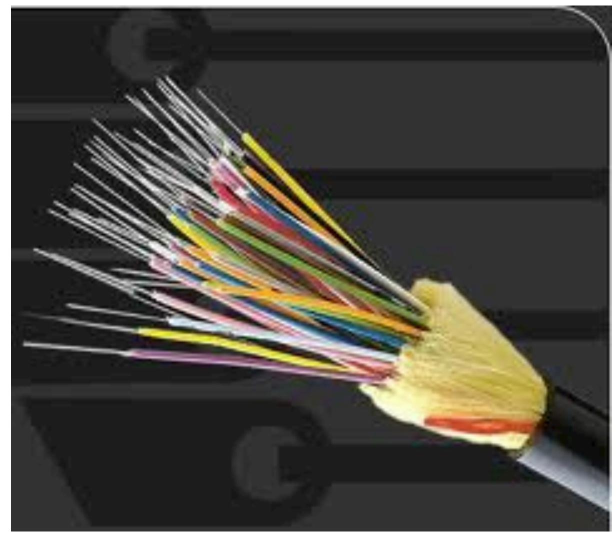


Small information content

Large information content

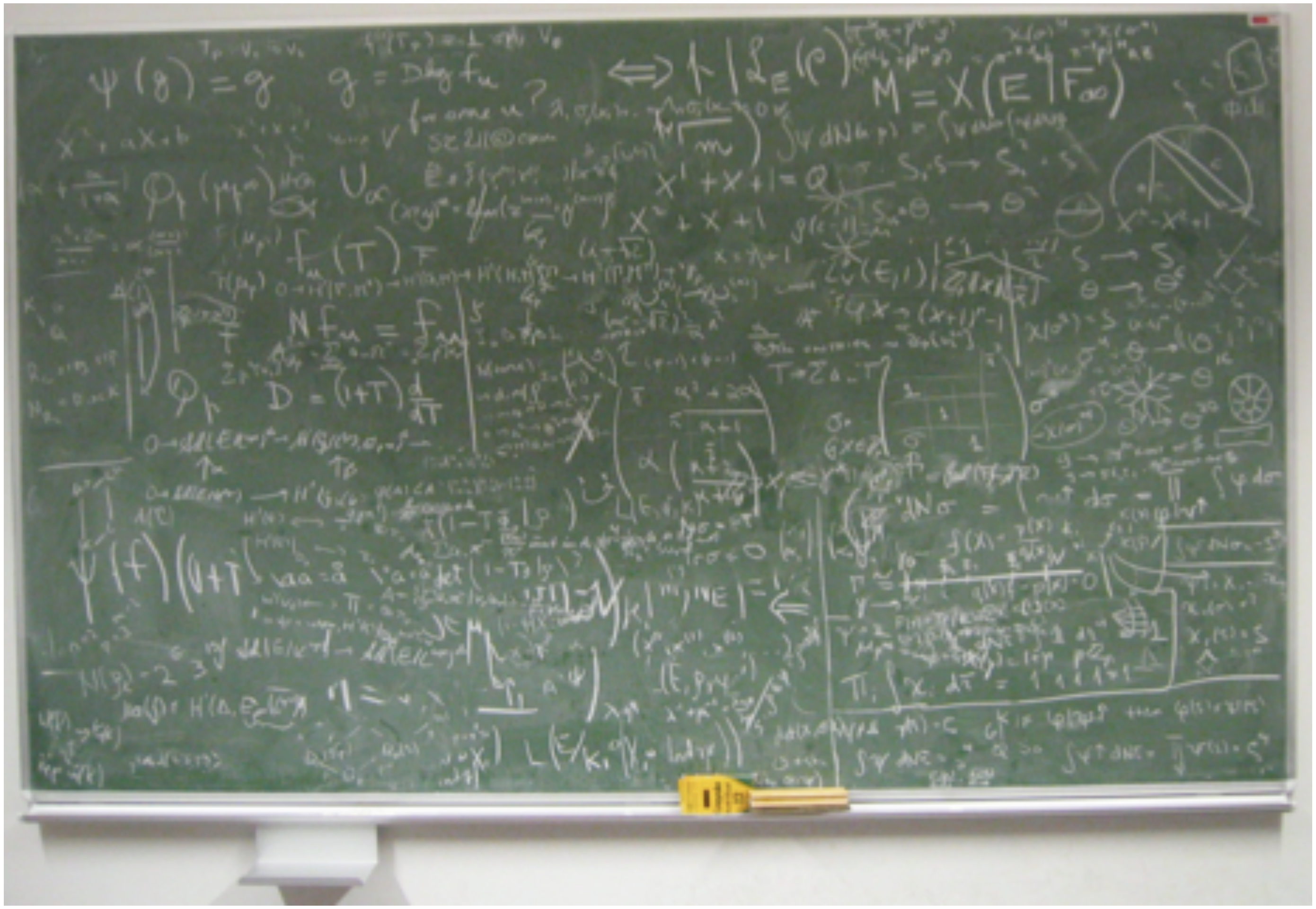
What is the fundamental limit of data transfer rate?

WiFi: data rate ~ Mbit/s

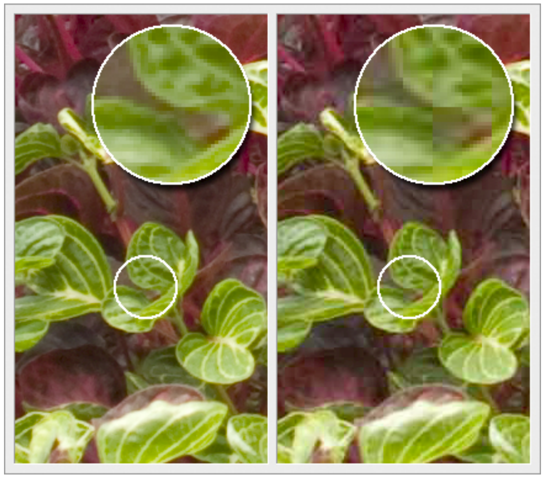


Fiber Optics: data rate ~ Tbit/s

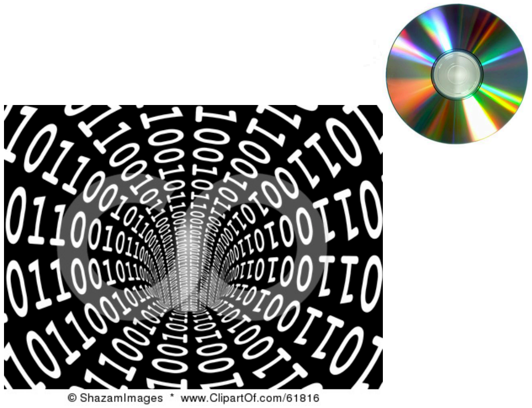
Some people think information theory (IT) is about...



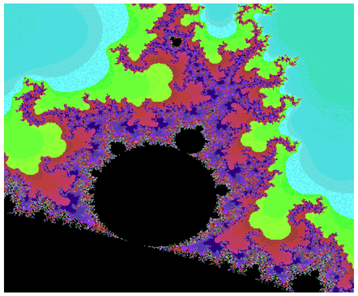
But IT is also about these...



Data Compression



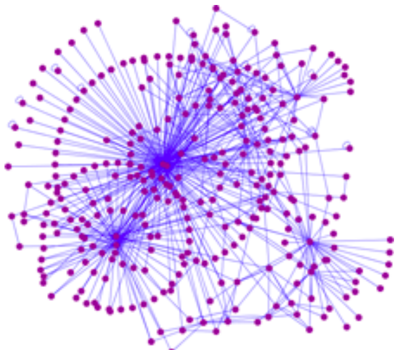
Coding



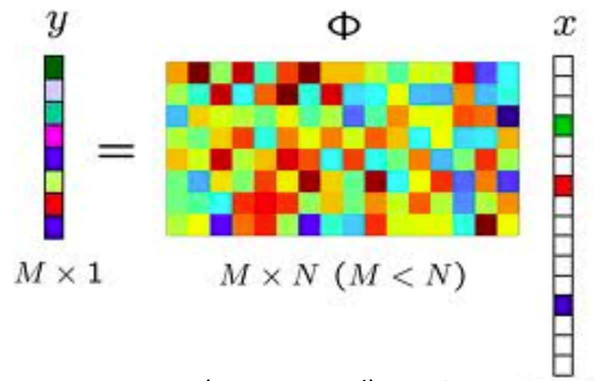
Computation: Kolmogorov Complexity

Data Communication

And even these...



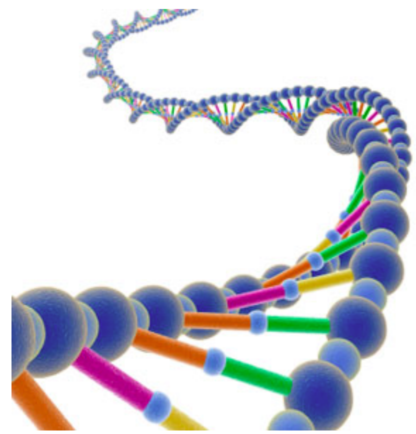
Network



(Compressed) Sensing

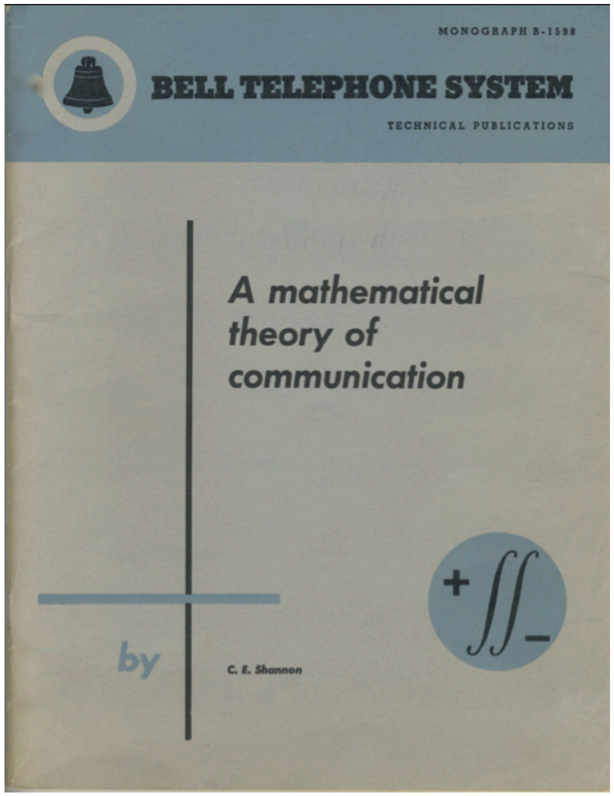
$N \times 1$

Investment, gambling

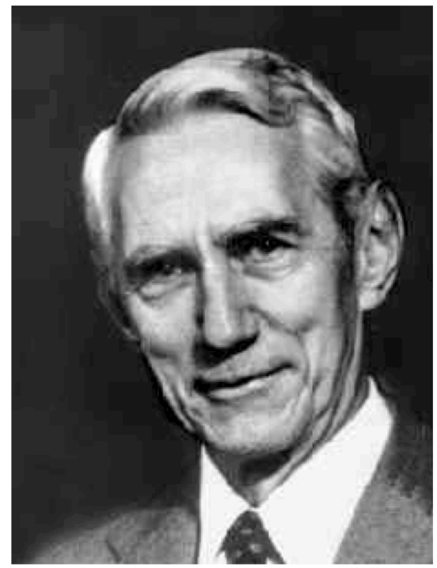


Bioinformatics

Where IT all begins...



1948, Bell Sys. Tech. Journal



Shannon, 1916 - 2001

Information Theory

- Shannon's information theory deals with limits on data compression (source coding) and reliable data transmission (channel coding)
  - How much can data can be compressed?
  - How fast can data be reliably transmitted over a noisy channel?

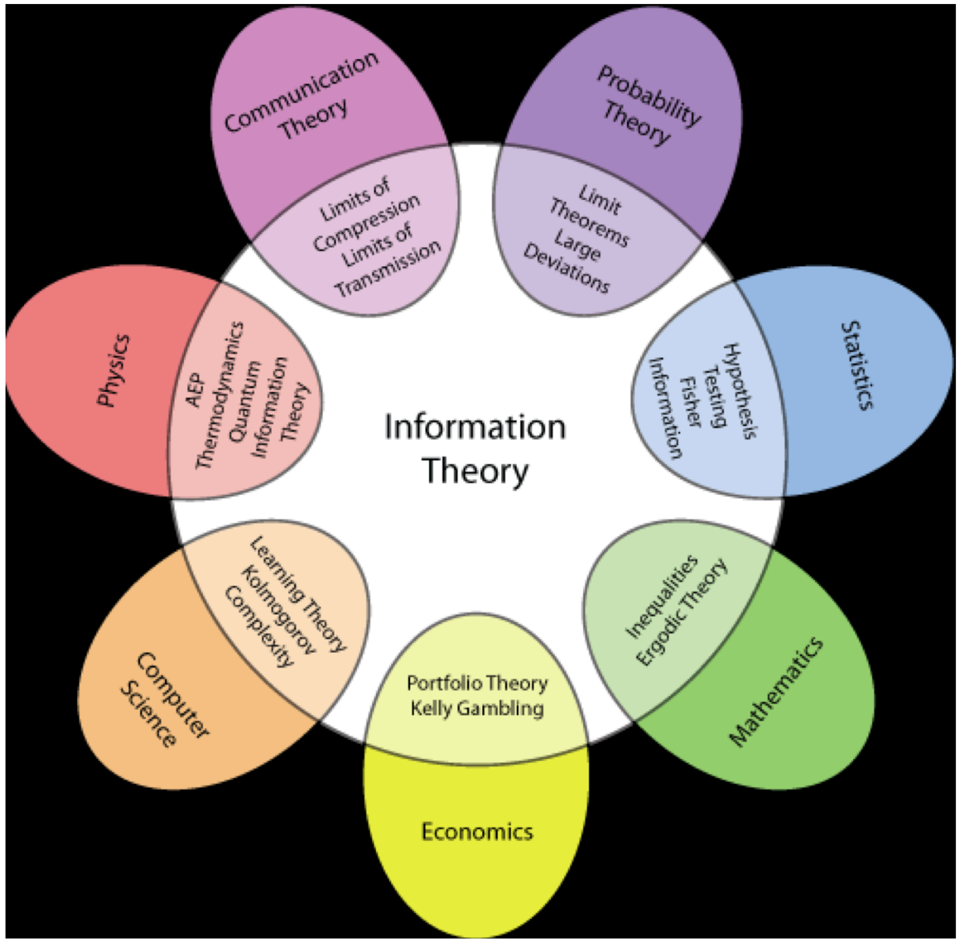
- Two basic "point-to-point" communication theorems (Shannon 1948)
  - Source coding theorem: the minimum rate at which data can be *compressed losslessly* is the *entropy rate* of the source
  - Channel coding theorem: The maximum rate at which data can be *reliably transmitted* is the *channel capacity* of the channel

- Since Shannon's 1948 paper, many extensions
  - Rate distortion theory
  - Source coding and channel capacity for more complex sources
  - Capacity for more complex channels (multiuser networks)

- Information theory was considered (by most) an esoteric theory with no apparent relation to the "real world"

- Recently, advances in technology (algorithms, hardware, software) today there are practical schemes for
  - data compression
  - transmission and modulation
  - error correcting coding
  - compressed sensing techniques
  - information security $\cdots$

IT encompasses many fields



In this class we will cover the basics

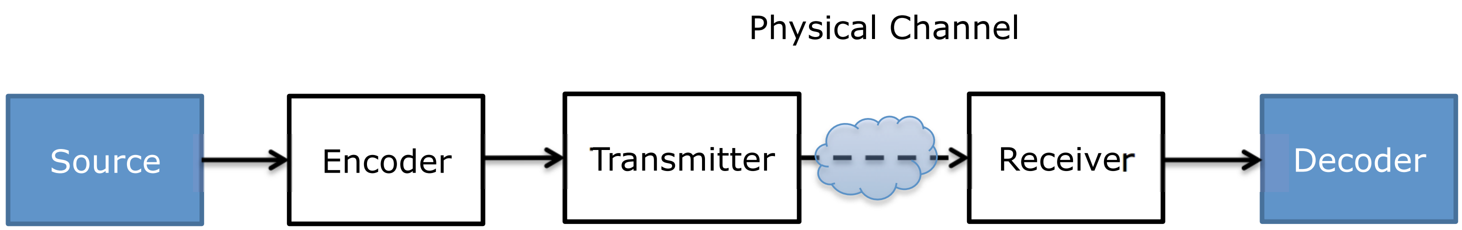
- Nuts and Bolts
  - Entropy: uncertainty of a single random variable

$$H(X) = -\sum_x p(x) \log_2 p(x) (\text{bits})$$

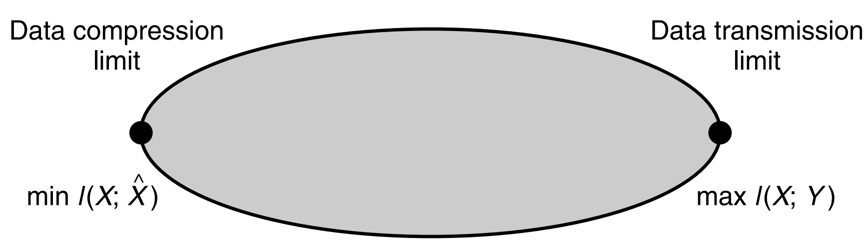
  - Conditional Entropy: $H(X|Y)$
  - Mutual information: reduction in uncertainty due to another random variable

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

  - Channel capacity $C = \max_{p(x)} I(X;Y)$
  - Relative entropy: $D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$



- – Data compression limit (lossless source coding)
  – Data transmission limit (channel capacity)
  – Tradeoff between rate and distortion (lossy compression)



Important Funcationals

- Upper case $X, Y, ...$ refer to random variables
- $\mathcal{X}$, $\mathcal{Y}$, alphabet of random variables
- $p(x) = \mathbb{P}(X = x)$
- $p(x, y) = \mathbb{P}(X = x, Y = y)$
- Probability density function $f(x)$

- Expectation: $\mu = \mathbb{E}\{X\} = \sum xp(x)$

- Why is this of particular interest? It appears in Law of Large Number (LLN): If x_n independent and identically distributed,

$$\frac{1}{N}\sum_{n=1}^{N} x_n \rightarrow \mathbb{E}\{X\}, \text{w.p.}1$$

- Variance: $\sigma^2 = \mathbb{E}\{(X-\mu)^2\} = \mathbb{E}\{X^2\} - \mu^2$

- Why is this of particular interest? It appears in Central Limit Theorem (CLT):

$$\frac{1}{\sqrt{N\sigma^2}}\sum_{n=1}^{N}(x_n - \mu) \rightarrow \mathcal{N}(0,1)$$

Information theory: is it all about theory?

Yes and No.

- Yes, it's theory. We will see many proofs. But it's also in preparation for other subjects
  - Coding theory (Prof. R. Calderbank)
  - Wireless communications
  - Compressed sensing
  - Stochastic network
  - Many proof ideas come in handy in other areas of research

- No. Hopefully you will walk out of this classroom understanding
  - Basic concepts people talk on the streets:
    entropy, mutual information ...
  - Channel capacity - all wireless guys should know
  - Huffman code (the optimal lossless code)
  - Hamming code (commonly used single error correction code)
  - "Water-filling" - power allocation in all communication systems
  - Rate-distortion function - if you want to talk with data compression guy

Course Logistics

- Schedule: T, Th, 1:25-2:40pm, Hudson 207
- TA: Miao Liu, Email: miao.liu@duke.edu
- Homework: out Thurs in class, due next Thurs in class
- Grading: homework 30%, two in-class Midterms (each 20%), one Final 30%
- Midterms: $10/4$, $11/6$ in class
- Final: $11/29$ in class
- Prerequisite: basic probability and statistics