581286-6 Three concepts:Information Spring 2006

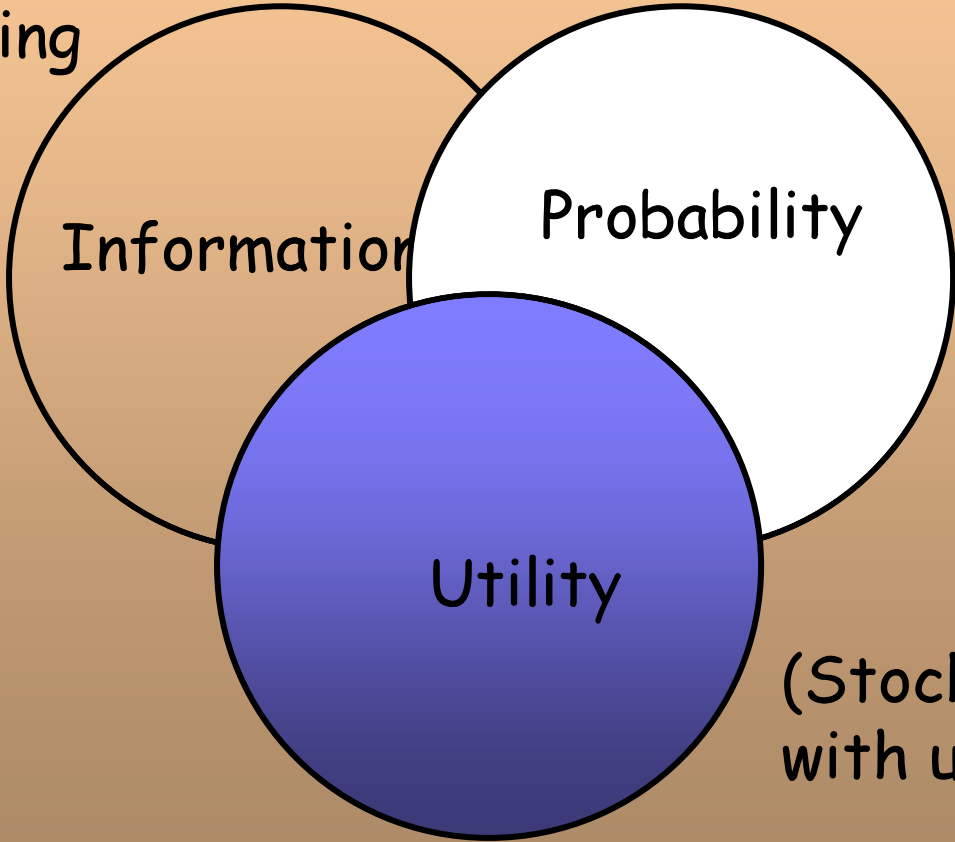
http://www.cs.helsinki.fi/group/cosco/Teaching/Information/2006/

Petri Myllymäki
Complex Systems Computation Group
Department of Computer Science, University of Helsinki
http://www.cs.helsinki.fi/petri.myllymaki/

Three concepts

Compression,
coding, modeling

Uncertain
reasoning



(Stochastic) search
with utility functions

Why information theory?

- "Educational argument"
  - ✓ general background
- "Employment argument"
  - ✓ information theory is **the** theory of data (tele)communication
- "Intelligent systems argument"
  - ✓ information theoretical concepts are deeply related to learning and adaptation

Information theory for Intelligent systems?

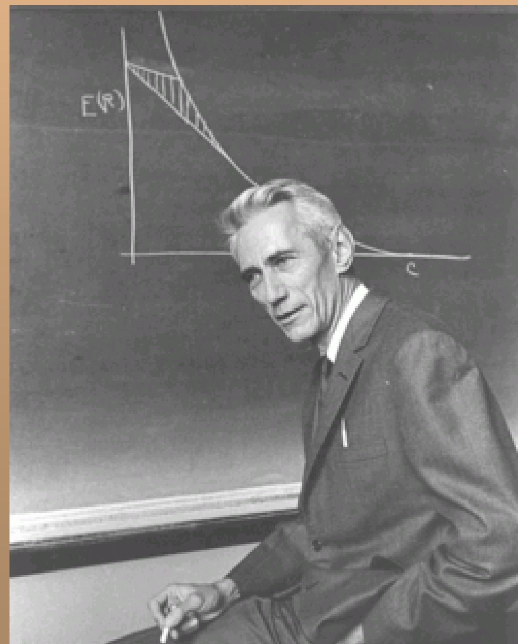
- Many problems are the same
  - ✓data compression and error correcting codes are based on modeling and inference
  - ✓"reliable communication over unreliable channels" vs. "reliable computation with unreliable hardware" (e.g., neural networks)
  - ✓working with probability distributions in high dimensional spaces

What do we learn?

- Central results by Shannon and their consequences
  - ✓ the source coding theorem
  - ✓ the noisy channel coding theorem
- "The legend of Minimum Description Length (MDL) Principle"

What is Information theory?

Claude Shannon, “A mathematical Theory of Communication”.
Bell Syst. Tech. Journal, 27: 379-423,623-656, 1948.



Simply put

- The problem of representing the source alphabet symbols s in terms of another system of symbols (0,1)
  - ✓ Channel encoding: how to represent the source symbols so that their representations are far apart in some suitable sense (“error-correction”)
  - ✓ Source encoding: How to represent the source symbols in a minimal form for purposes of efficiency (“compression”)

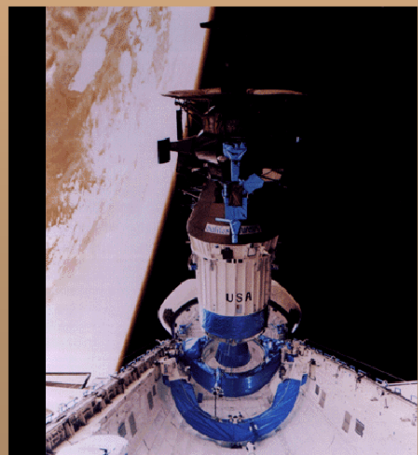
The course focus

- we will address source encoding as it has deep relationship to modeling
- (by the end of the course) abstract from actual codes to code lengths
- discuss information-theoretic principles that can be used as a foundation of statistical modeling

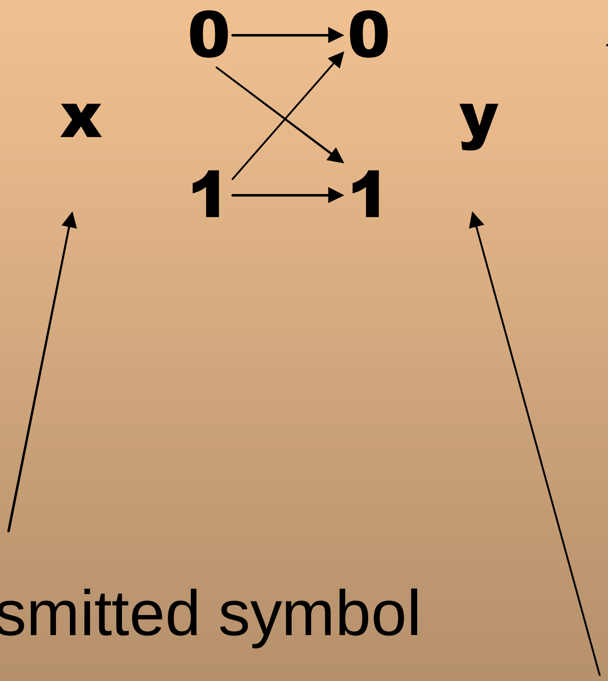
What we will NOT discuss ....

Noisy communication channels

- An analogue telephone line used by modems (to transmit digital information)
- DVB-T transmissions
- the radio communication link from Galileo to earth
- a disk drive



Binary symmetric channel



$$P(y=0 \mid x=0) = 1 - f$$

$$P(y=1 \mid x=0) = f$$

$$P(y=0 \mid x=1) = f$$

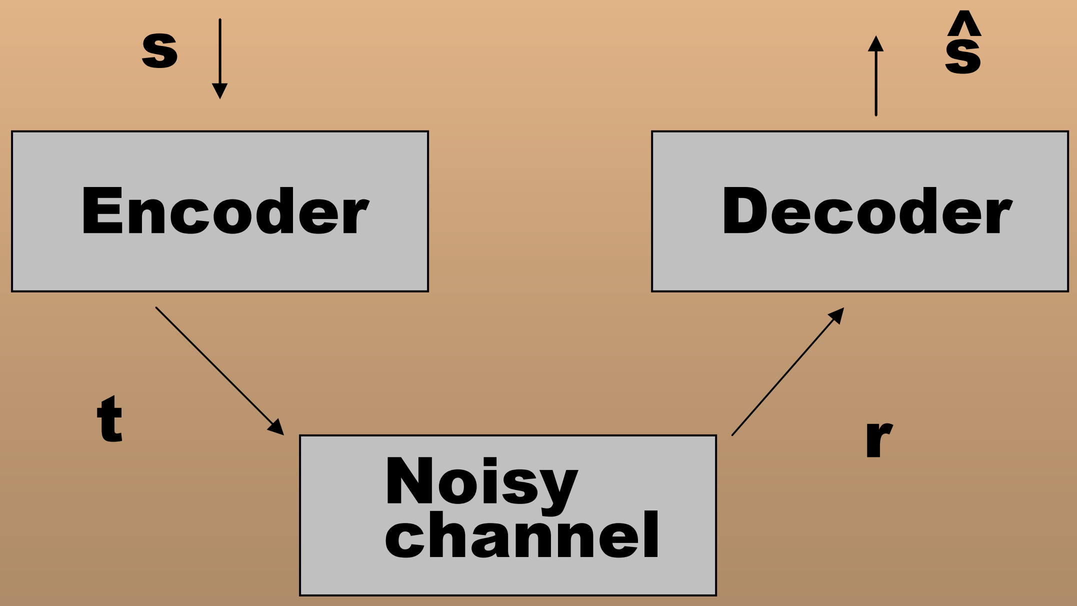
$$P(y=1 \mid x=1) = 1 - f$$

transmitted symbol

received symbol

How to reach error probabilities of order 10^{-15}?

- The physical solution
- The system solution



“To be more precise”

- Information theory answers questions about the theoretical limitations of such systems
- Coding theory discusses how to build practical encoding and decoding systems

Repetition codes

Encoding

s	t
0	000
1	111

Decoding

r	000	001	010	100	101	110	011	111
$\hat{s}$	0	0	0	0	1	1	1	1

---

s	0	0	1	0	1	1	0
t	000	000	111	000	111	111	000
n	000	001	000	000	101	000	000
r	000	001	111	000	010	111	000
$\hat{s}$	0	$\boxed{0}$	1	0	$\boxed{0}$	1	0

Think!

- What is the error probability for the previous repetition code for a binary symmetric channel with noise level f?

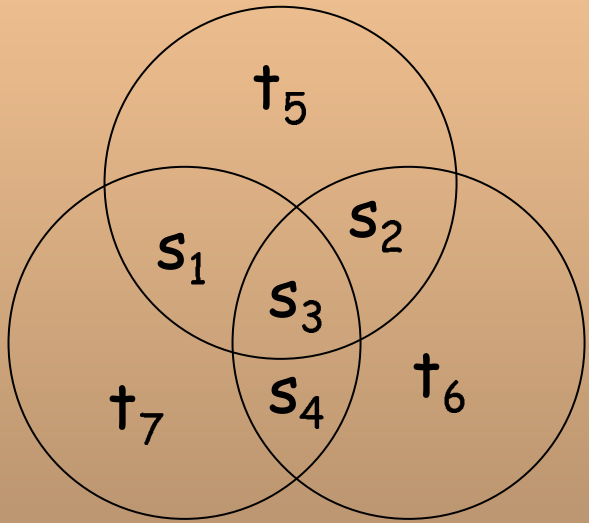
Some analysis

- For f=0.1 the error probability is $p_b = 3f^2(1-f)+f^3 \sim 0.03$
- What did we loose?
  - ✓information transmission rate reduced by factor of three!
- Good?
  - ✓assume we want a probability of error close to 10^{-15}. What would be the rate of the repetition code? (~1/60)

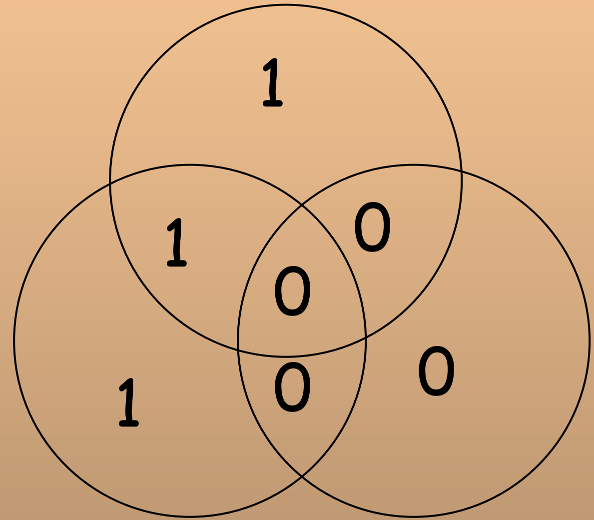
Block codes

- Goal: (very) small probability of error and a good transmission rate
- Idea: add redundancy to blocks instead of encoding one bit at a time (the origin of "parity")
- Solution: (N,K) block code adds (N-K) redundant bits to the end of the sequence of K source bits

(7,4) Hamming encoding

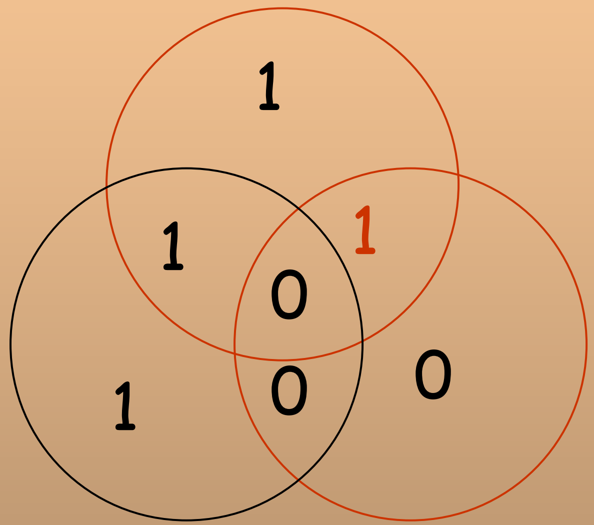


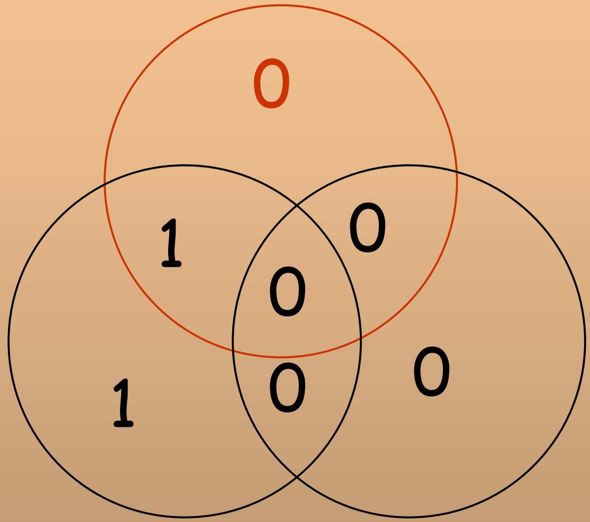
S = 1000



Rule: parity in each circle is even

(7,4) Hamming decoding





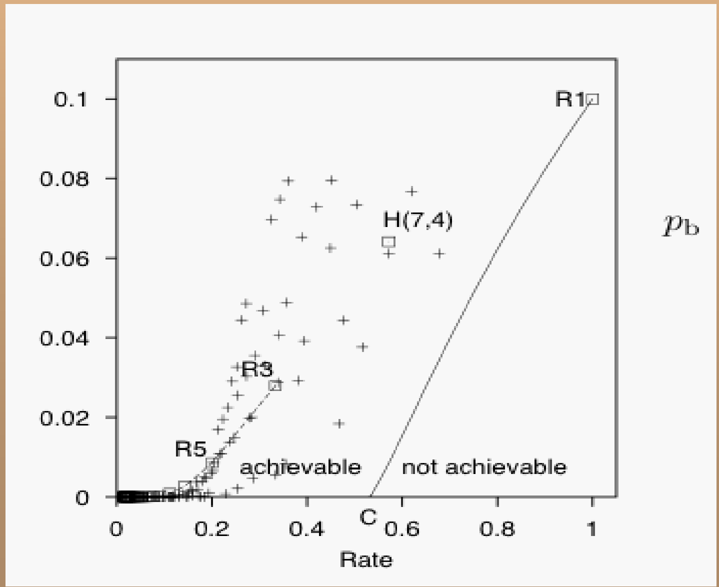
Rule: for the received vector check that the parity in each circle is even; identify the most likely cause

Performance of the best codes

- We want
  - ✓ small error probability p_b
  - ✓ large (transmission) rate R
- What points in the (p_b, R)-plane are achievable?
- A good guess: boundary passes through the origin (0,0)

Wrong!
(The noisy channel theorem)

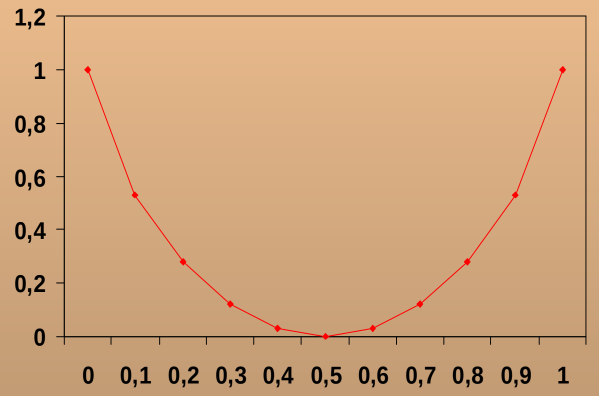
- Shannon proved that for any given channel, the boundary meets the R axis at a non-zero value R=C



Channel capacity

- The channel capacity C for binary symmetric channel is

$$C(f) = 1 - \left[f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{1-f} \right]$$



- Generally, the channel capacity is the maximal mutual information between input X and output Y

So how many disks?

- For f=0.1 we have $C \cong 0.53$
- Repetition code R_3 gave us R=1/3 with p_b=0.03 (3 noisy gigabyte disk drives)
- To reach $p_b=10^{-15}$ we needed 60 noisy gigabyte disk drives
- Shannon says:
  - ✓ to reach $p_b=10^{-15}$ you can achieve with 2 disk drives (2 > 1/0.53)
  - ✓ and to reach $p_b=10^{-24}$ you still need only 2 disk drives!