EECE 580B
Modern Coding Theory

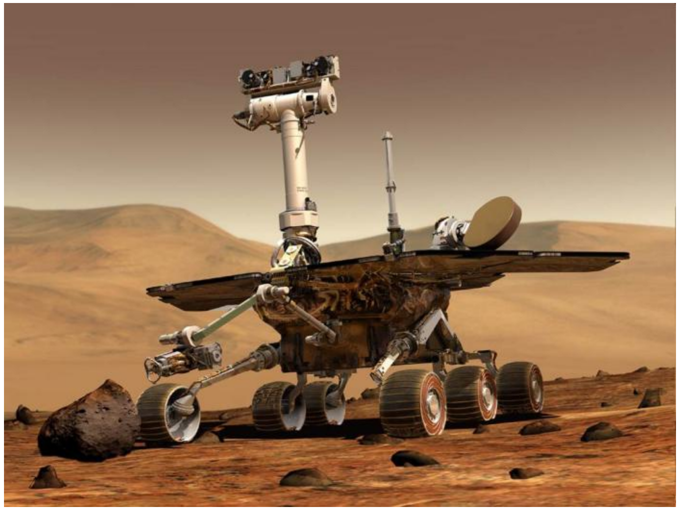
First lecture

Tomas Filler
(tomas.filler@binghamton.edu)
Jessica Fridrich
(fridrich@binghamton.edu)

BINGHAMTON
UNIVERSITY

State University of New York

Why coding theory?



EECE 580B – Modern Coding Theory

CRN# 95028 **Credits:** 3

Lectures: Tue, Thu @ 2:50 pm - 4:15 pm, Fine Arts 352

Office hours:
J. Fridrich: Mon @ 1:00 pm - 3:00pm, EB Q16
T. Filler: Wed @ 1:00 pm - 3:00pm, EB P7

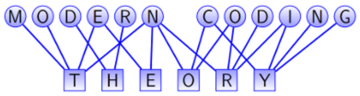
Grading:

Homework assignments	letter grade, weight 40%
In-class midterm exam	letter grade, weight 30%
Take home final exam	letter grade, weight 30%

Final grade is weighted average of all grades.

Course material: Will be posted at the course web page

http://dde.binghamton.edu/filler/mct



Course website

Course Prerequisites

Students are expected to be familiar with basic concepts from calculus and elementary statistics.

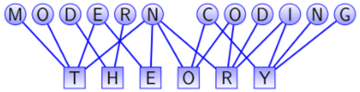
Knowledge of Matlab is **essential**. Although I will explain selected Matlab commands relevant to this course in the lecture, students not familiar with Matlab should study tutorials to become comfortable when programming in Matlab.

There are a number of Matlab tutorials available on-line, see

http://www.mathworks.com/academia/student_center/tutorials/launchpad.html

or see the course website

http://dde.binghamton.edu/filler/mct



Course website

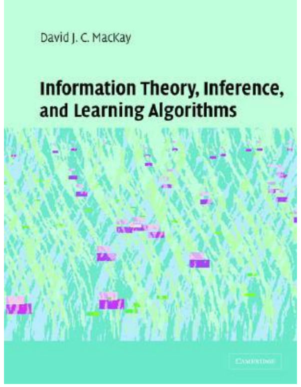
MODERN CODING THEORY

See course website

Course Material

There are many textbooks covering coding theory (both algebraic and modern). We will use the following books occasionally:

Available online!



- Information Theory, Inference & Learning Algorithms by David J. C. MacKay

  Available at Science libr. Q360 .M23 Google "itila"

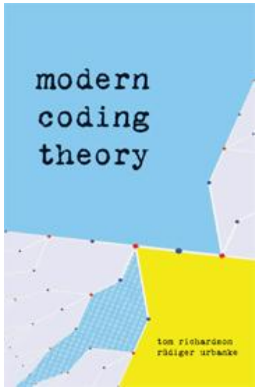
- A First Course in Coding Theory by Raymond Hill

  Available at Bartle libr. QA268 .H55

All other necessary material will be provided on the course web site.

Very nice (but quite technical) text is:

- Modern Coding Theory by Tom Richardson and Ruediger Urbanke



Course Topic

Coding theory (channel coding) deals with the problem of sending data over noisy channel (WiFi network) or storing data to unreliable media (CD, hard disk).

Modern approach is characterized by codes based on sparse random structures with low-complexity encoding end decoding algorithms. These codes allow transmission of data at nearly optimal rate and are thus considered for practical applications and standards (3G, DVB, deep space communication, ...).

Course Objective

Ensure students understand modern principles of error correction codes via **hands-on experience**.

Course will put strong emphasis on gaining **practical knowledge** with the following topics:

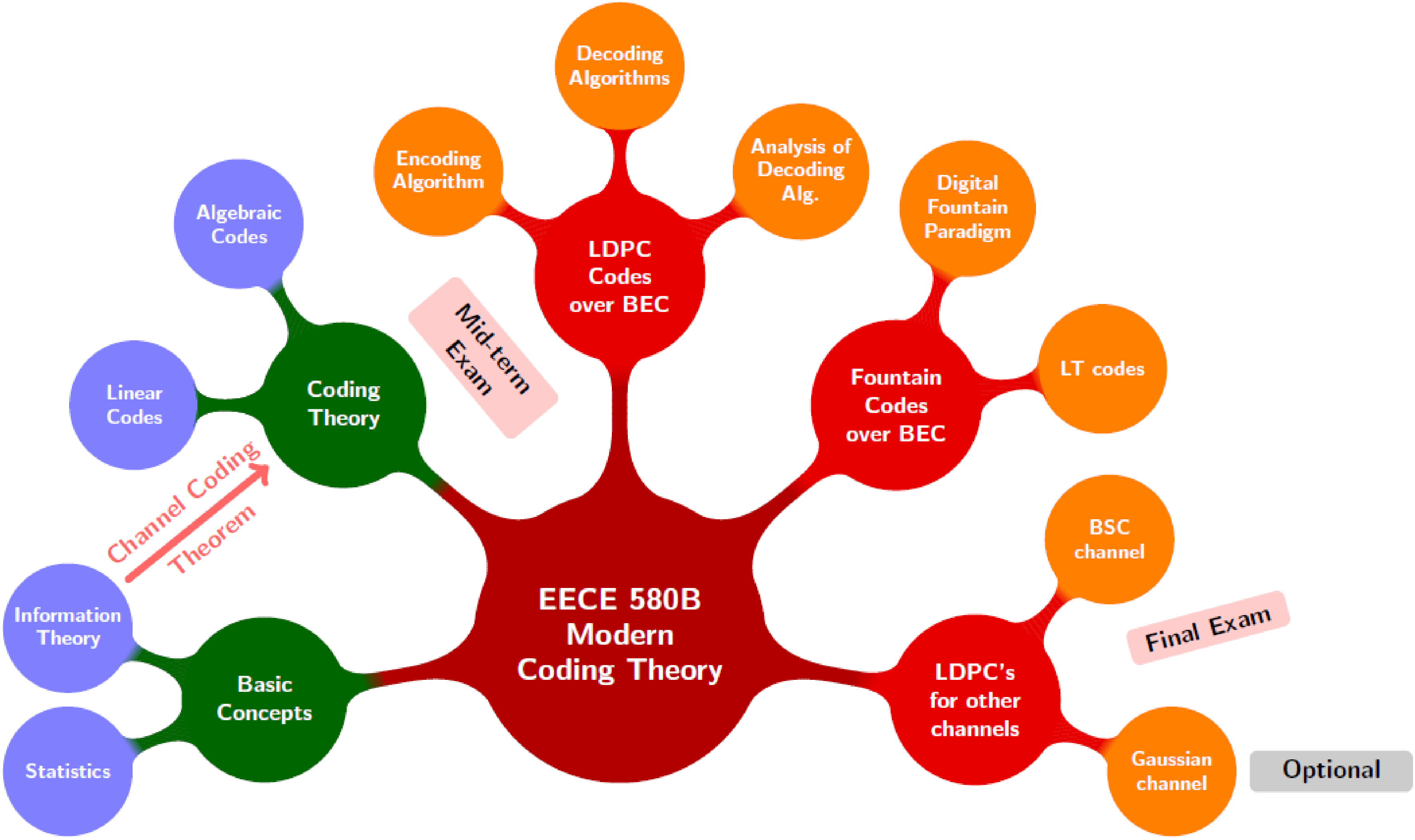
- Basic models of communication channels and their capacities
- Linear block codes over finite fields
  - representation, encoding, and decoding problems
- Low Density Parity Check codes
  - linear time encoding and decoding
  - design and analysis for different channels
- Selected special topics, such as Fountain codes (LT codes).

Academic Honesty

- All students must adhere to the Student Academic Honesty Code of the University and the Watson School.
- Student are expected to work on assignments **individually.**
- First instance of academic dishonesty in the class:
  - No credit for the assignment/exam/quiz/etc. on which the offense was committed
  - **AND… a reduction in course grade by one full letter grade**
  - **AND… record of offense will be reported to university administration**
- Second instance of academic dishonesty in the class:
  - Failure of course… and possible further action (e.g., suspension)

Syllabus

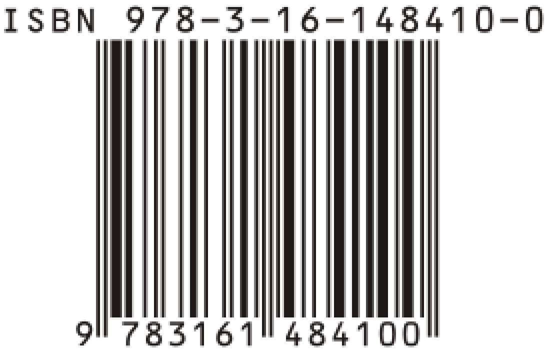
See course website

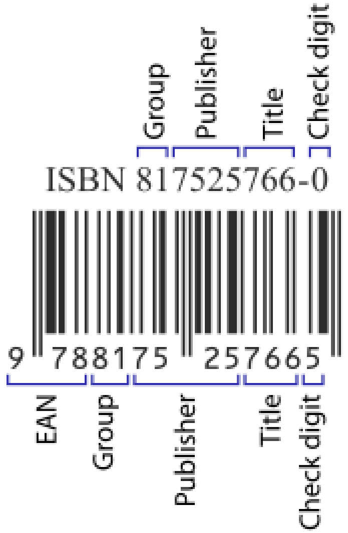


Simple Applications of Coding Theory

International Standard Book Num.

- Every book has its number, called ISBN.
- Every number contains (10 or 13 digits).





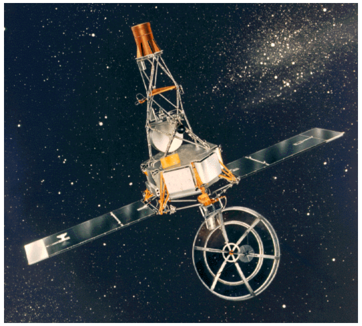
- Last digit is used as check digit to correct errors.

$$x_{10} = \sum_{i=1}^{9} i x_i \bmod 11$$

Source: http://en.wikipedia.org/wiki/International_Standard_Book_Number

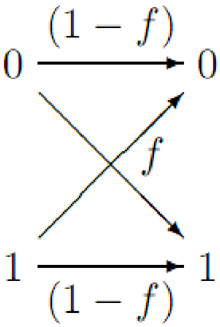
Perfect Communication over Imperfect Channels

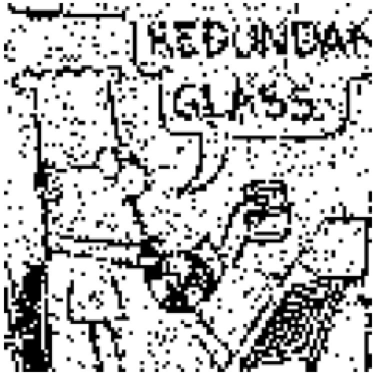
Many (if not all) real channels we use to send (store) data are imperfect (noisy).



- modem → phone line → modem
- Mariner (Mars) → radio waves → Earth
- comp. memory → hard drive → comp. memory

REDUNDAN
GLASS.





Noisy Hard Drive

We have an unreliable hard drive.

Drive stores and reads the bits with f=10% error, i.e., on average, every 10th bit is read incorrectly.

But we want the drive to be reliable with Pr(bit error)≈10^{-15}.

If we have Pr(bit error)≈10^{-15}, then we can expect 1 wrong bit in ≈ 113TB of data. This should be enough to safely read and write 1GB per day for 10 years.

What can we do to achieve reliable communication or data storage?

Physical Solution to Perfect Communication

Improve physical characteristics of the system:

- Use more reliable components.
- Use more power at the transmitter side.
- Use larger antenna at receiver side.
- …

Advantages:

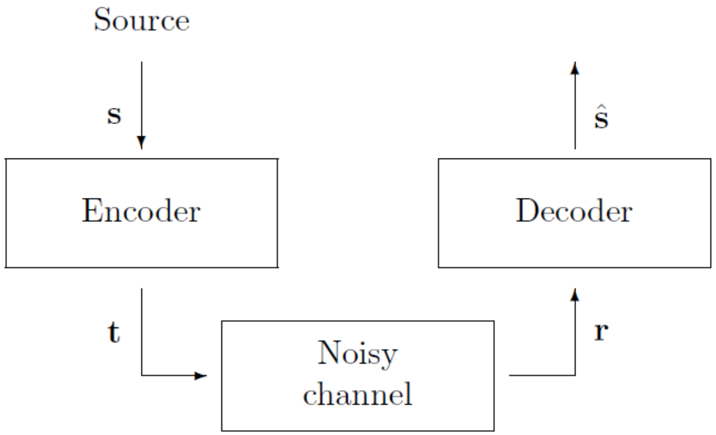
- simple to implement (in the past)
- no need for further data processing

Disadvantages:

- more expensive solution
- not suitable when limited resources (such as power)
- ever-increasing costs

The ‘system’ Solution to Perfect Communication

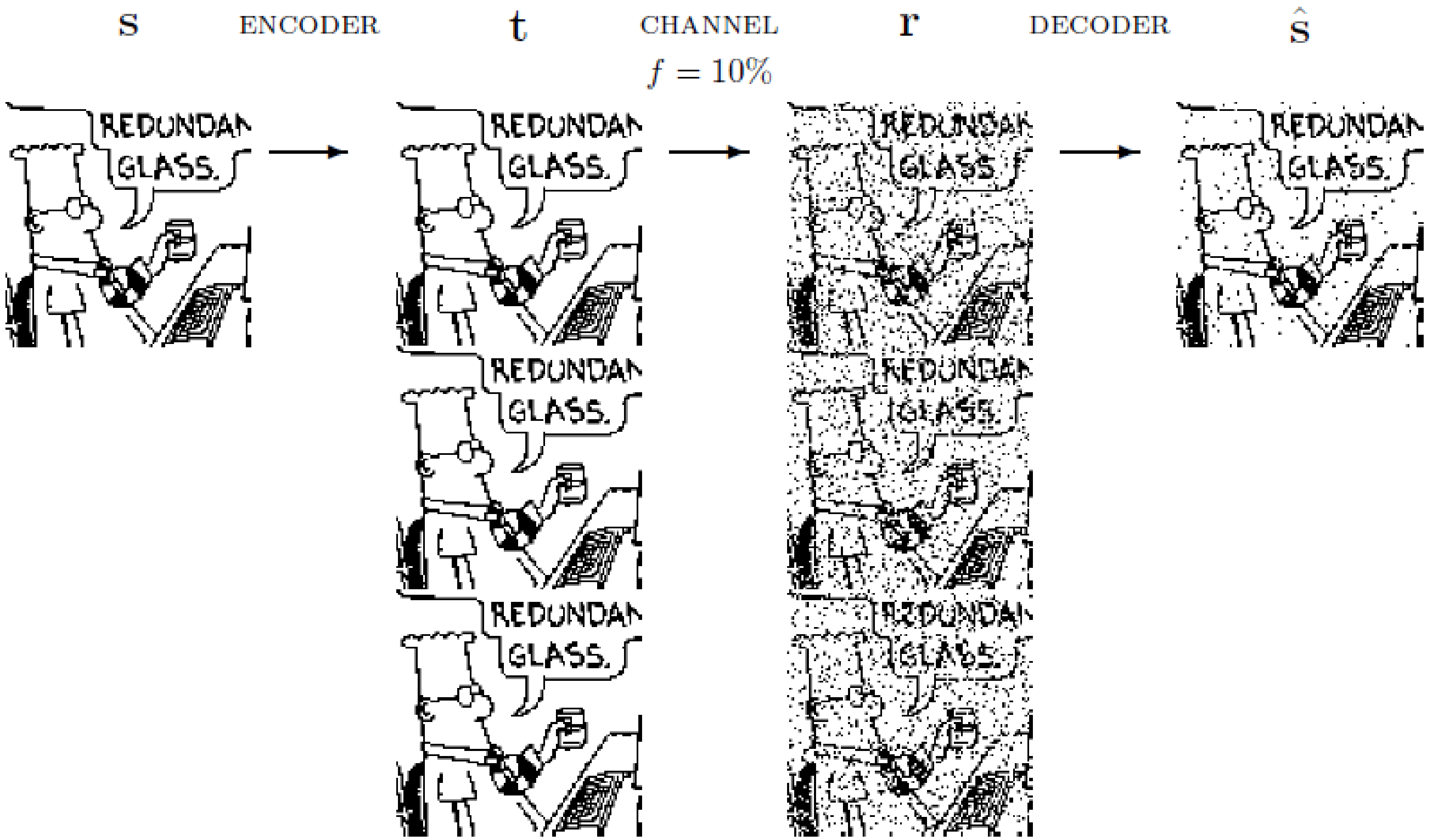
- Use **information theory** and **coding theory**.
- We accept the channel as is, no change necessary.
- Construct an encoder and decoder that achieves reliable communication by putting redundancy into original message.
- Can achieve reliable communication by increasing computational requirements on the system.



Example with Repetition Code

See Chapter 1 in [MacKay].

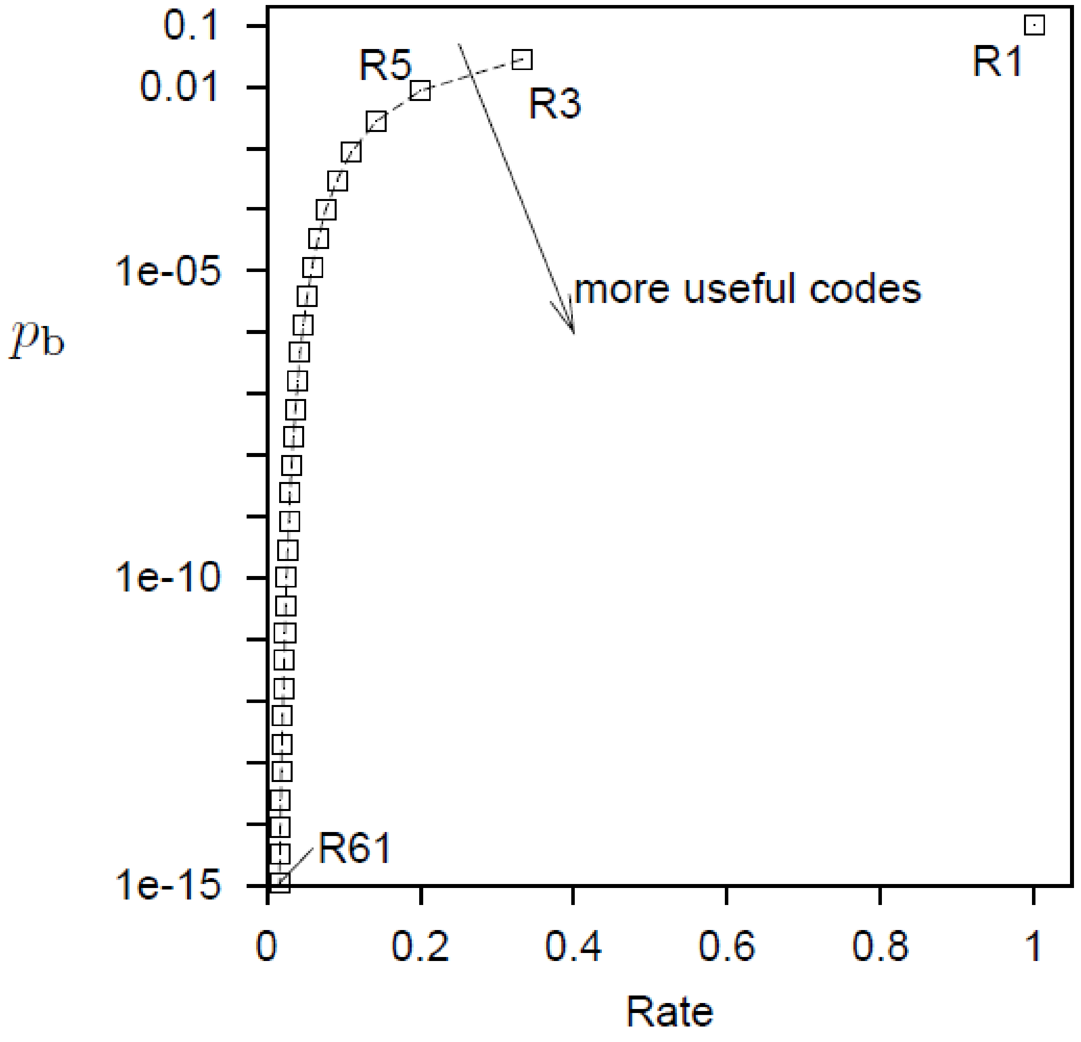
Repetition Code R_3



$\mathbf{s}$ ENCODER $\mathbf{t}$ CHANNEL $\mathbf{r}$ DECODER $\hat{\mathbf{s}}$

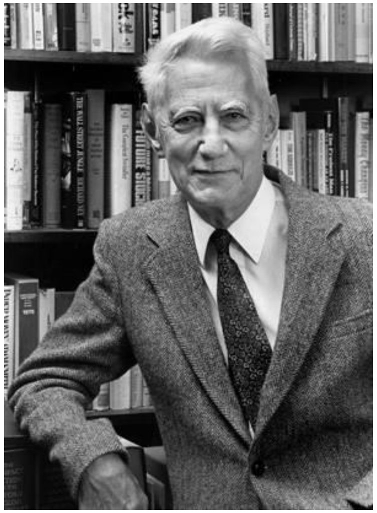
$f = 10\%$

Bit Error Probability vs. Rate for R_3



Do we really need 61 unreliable drives to assemble 1 reliable?

Shannon’s Information Theory



Claude Elwood Shannon
(1916-2001)

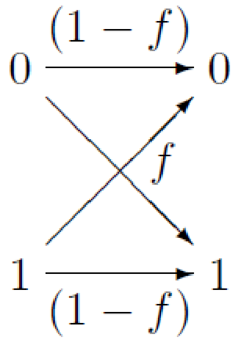
1948: A mathematical theory of communication.

Channel Coding Theorem:
“For any channel, there exist codes that make it possible to communicate with arbitrarily small probability of error at non-zero rates. The maximum rate at which communication is possible with arbitrarily small error is called **the capacity** of the channel. ” (MacKay 2003).

Source Coding Theorem:
“N i.i.d. random variables each with entropy H(X) can be compressed into more than NH(X) bits with negligible risk of information loss, as N tends to infinity; but conversely, if they are compressed into fewer than NH(X) bits it is virtually certain that information will be lost.” (MacKay 2003).

Bit Error Probability vs. Rate for R_3

Binary symmetric channel:



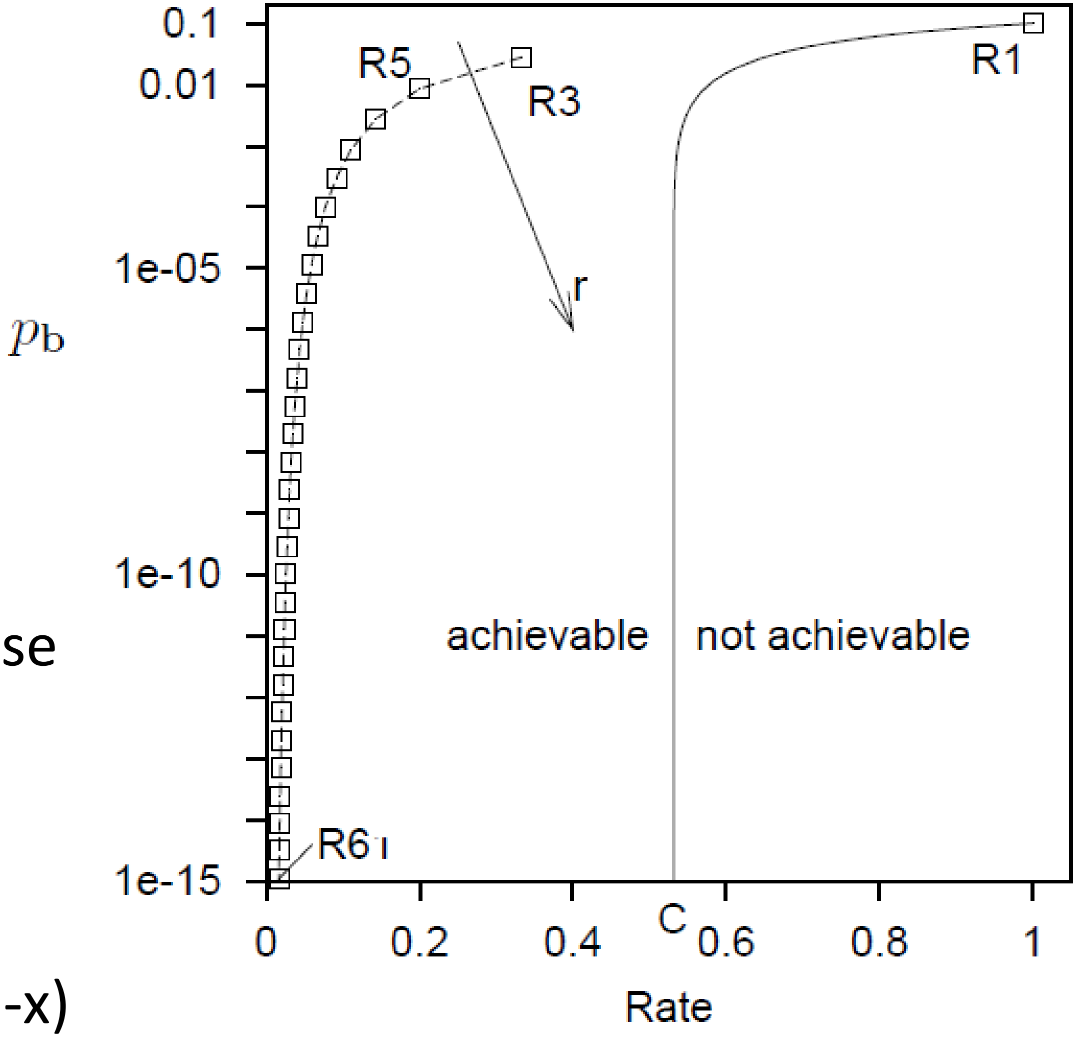
Capacity:

C(f) = 1-H^{-1}(f) bits/channel use

C(0.1) = 1-0.469 = 0.531

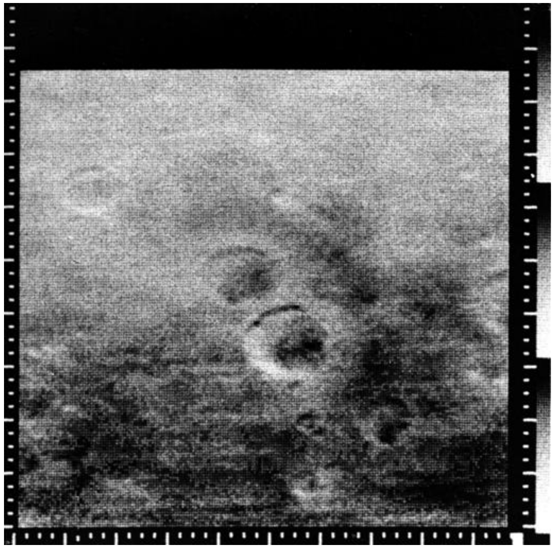
Binary entropy function:

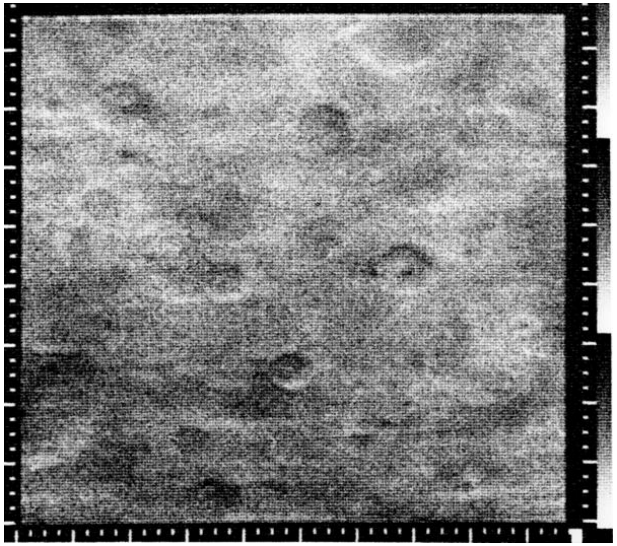
H(x) = -x·log$_2$(x) – (1-x)·log$_2$(1-x)



Mariner 4 – First Images From Space

- Launched in 1964, arrived at Mars in 1965 (after 7.5 months).
- Equipped with TV camera and 5.2Mb magnetic tape recorder.
- Can send @ rate 8⅓ and receive @ rate 33⅓ bps.
- In total, 22 200×200 pixel 6-level grayscale images were transferred (5.2Mb ≈ 650KB). Every image was transferred twice.





Hard Drive Failures

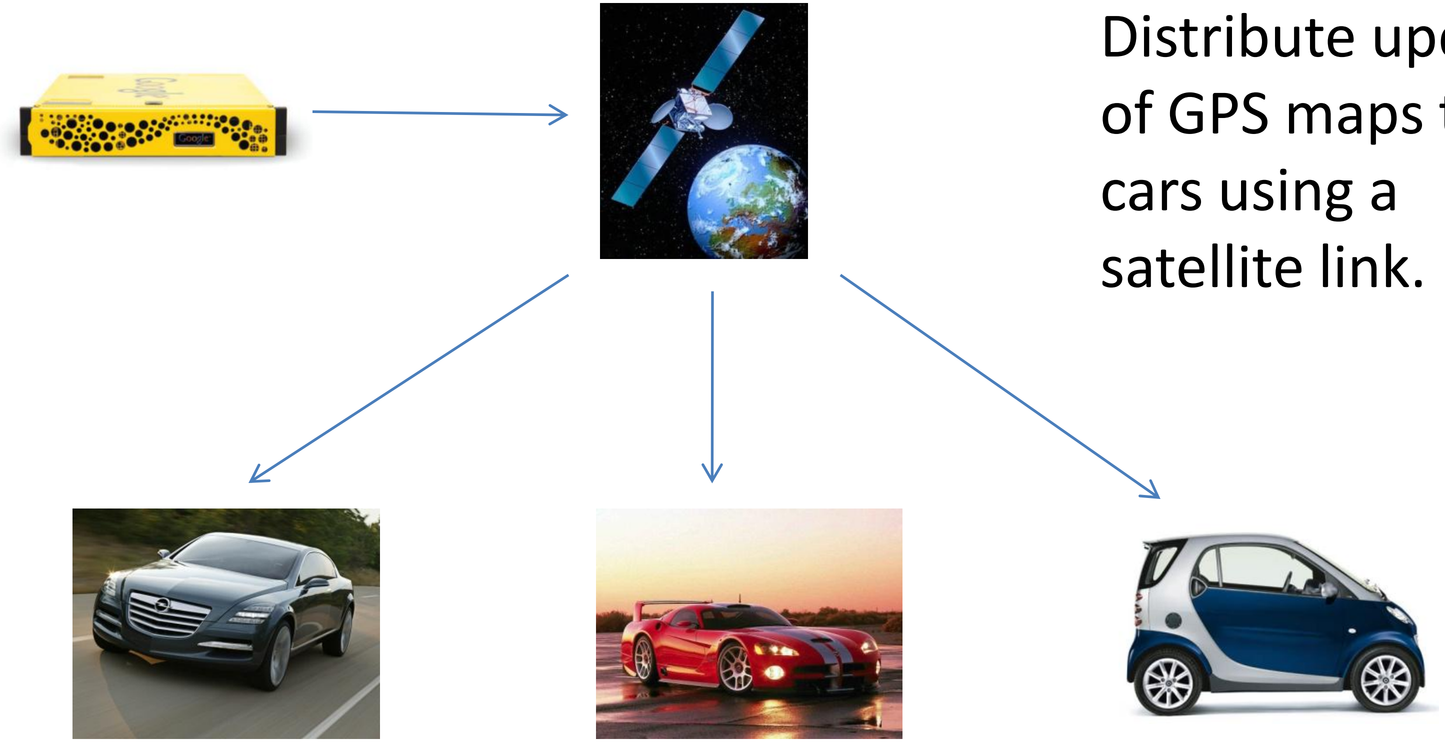
Redundant Array of Inexpensive Disks - RAID

- Used in today's PCs, servers, to increase
  - data throughput – data are distributed across disks
  - data reliability against hard drive failure (complete loss)
- Used in different configurations:
  - RAID 0 – striped disks – need 2 disks - speed
  - RAID 1 – mirroring – need 2 disks - reliability
  - RAID 5 – stripped disks with parity – need 4+1 = 5 disks

Source: http://en.wikipedia.org/wiki/RAID

Point-to-Multipoint Communication

Point-to-Multipoint Communication



Distribute updates of GPS maps to cars using a satellite link.

Cars see the satellite at random times and experience different losses. There is no feedback between car and satellite.

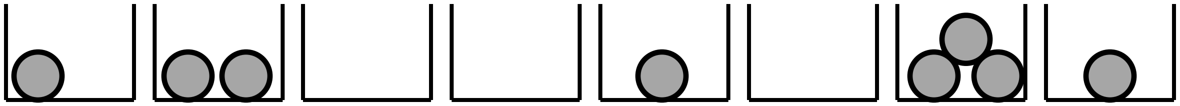
Point-to-Multipoint Communication

Trivial solution:

- Send the original data several times in a carousel manner.
- Original file consists of k packets; cars tune in at a random times, and each time they receive b packets.
- Assume that a complete transmission of k packets takes one day.
- Every car tunes in 2 times per day. How many days d of transmission are needed to ensure that 99.99% of the cars have received all the packets? (minimum is k/2b)

Model:

- throw dk balls at random into k bins. For a given bin, what is the probability that it has received at least one ball?



k bins

Point-to-Multipoint Communication

Each day, every bin receives a ball with probability $2b/k$.

Probability that the bin is empty after d days is

$$(1 - 2b/k)^d \approx \exp(-2bd/k)$$

Want this quantity to be less than 0.0001; so d is roughly $4.6k/2b$, that means every car receives $9.4k$ packets (instead of only k) of which many duplicate.

file = k packets
car receives b packets in one day
d = number of days needed

There is an elegant solution to this problem that needs only little bit more that k packets!