Introduction to Information Theory

Ref.: Hsiao-feng Francis Lu ,Dept. of Comm Eng., National Chung-Cheng Univ.

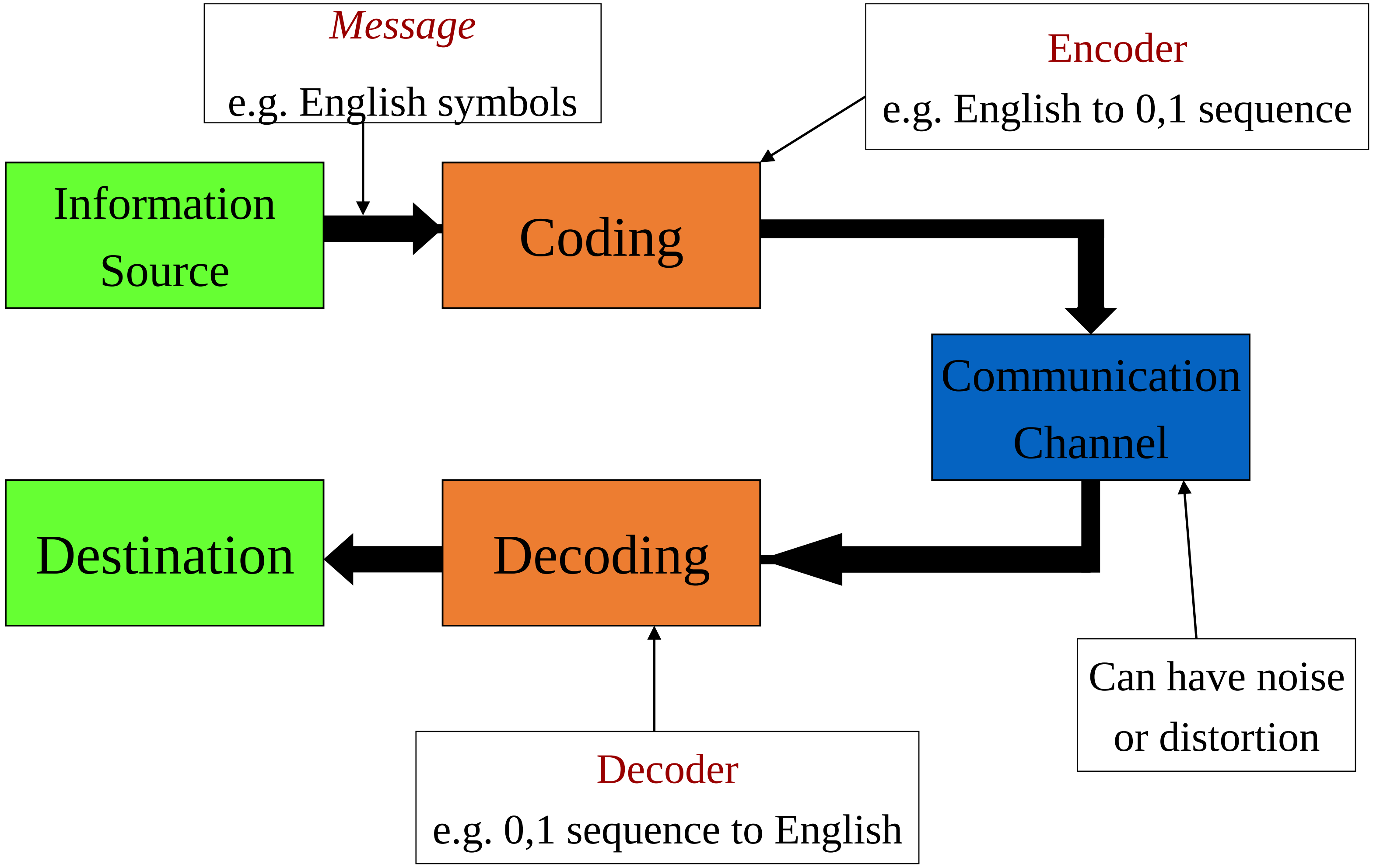
Edited by:

Dr. Maher Abdelrasoul

Father of Digital Communication

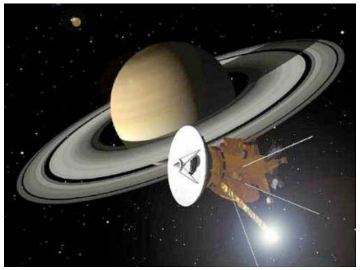
The roots of modern digital communication stem from the ground-breaking paper "A Mathematical Theory of Communication" by **Claude Elwood Shannon** in 1948.

Model of a Digital Communication System



Communication Channel Includes





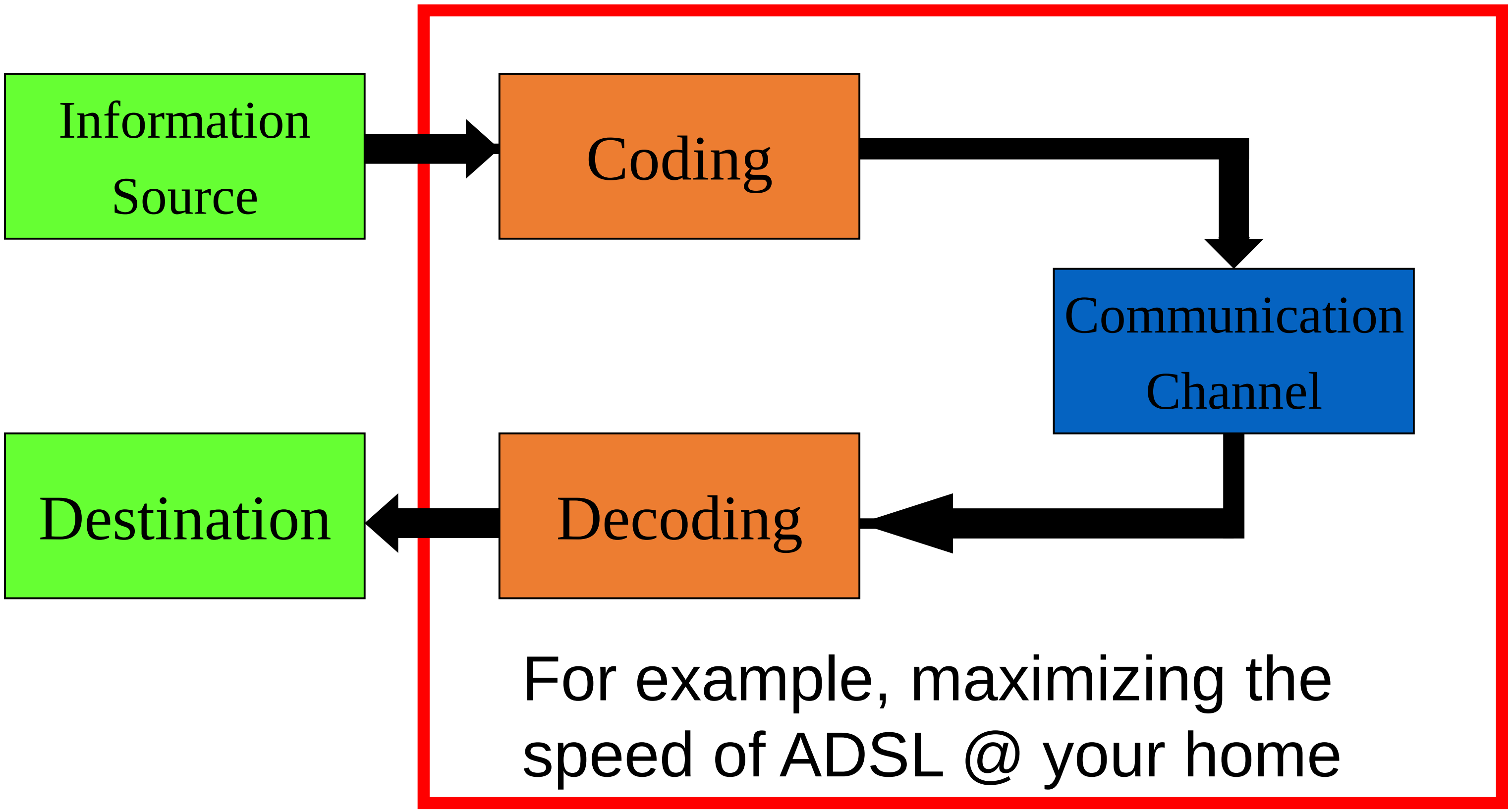

Shannon's Definition of Communication

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

"Frequently the messages have *meaning*"

Shannon Wants to…

- Shannon wants to find a way for "**reliably**" transmitting data throughout the channel at "**maximal**" possible rate.

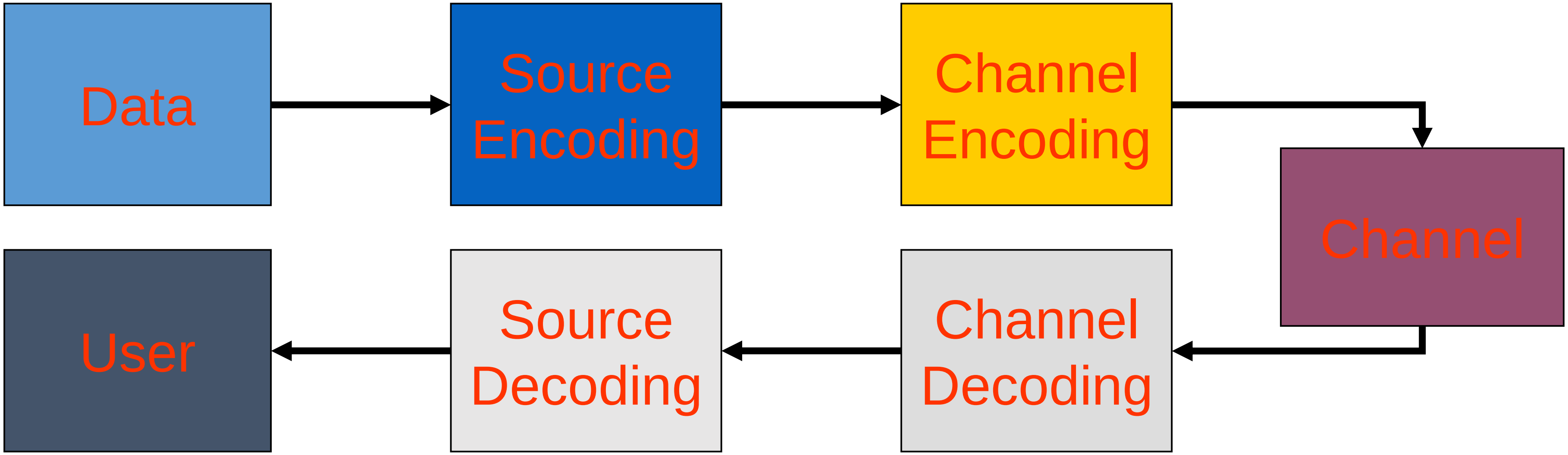


And he thought about this problem for a while…

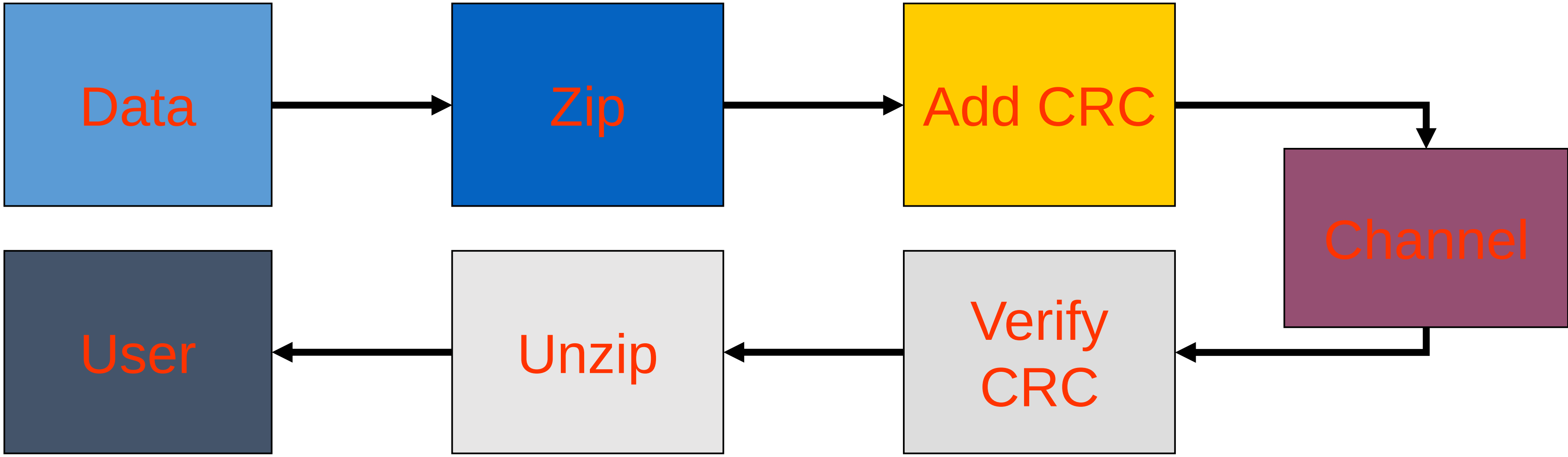
He later on found a solution and published in this 1948 paper.

In his 1948 paper he build a rich theory to the problem of reliable communication, now called "Information Theory" or "The Shannon Theory" in honor of him.

Shannon's Vision



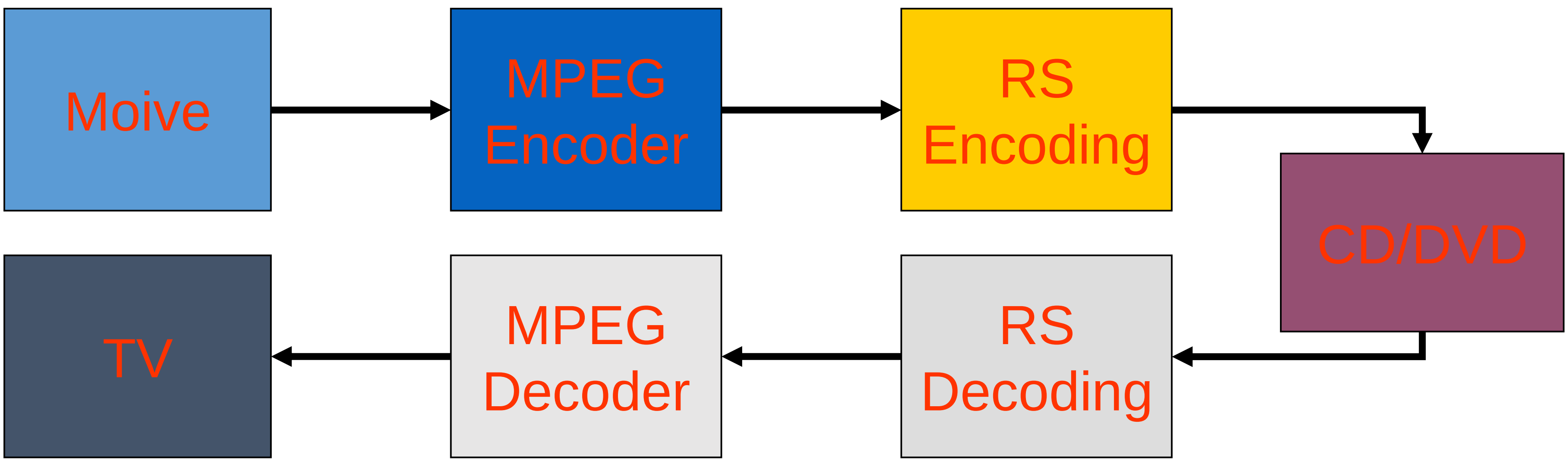
Example: Disk Storage



In terms of Information Theory Terminology

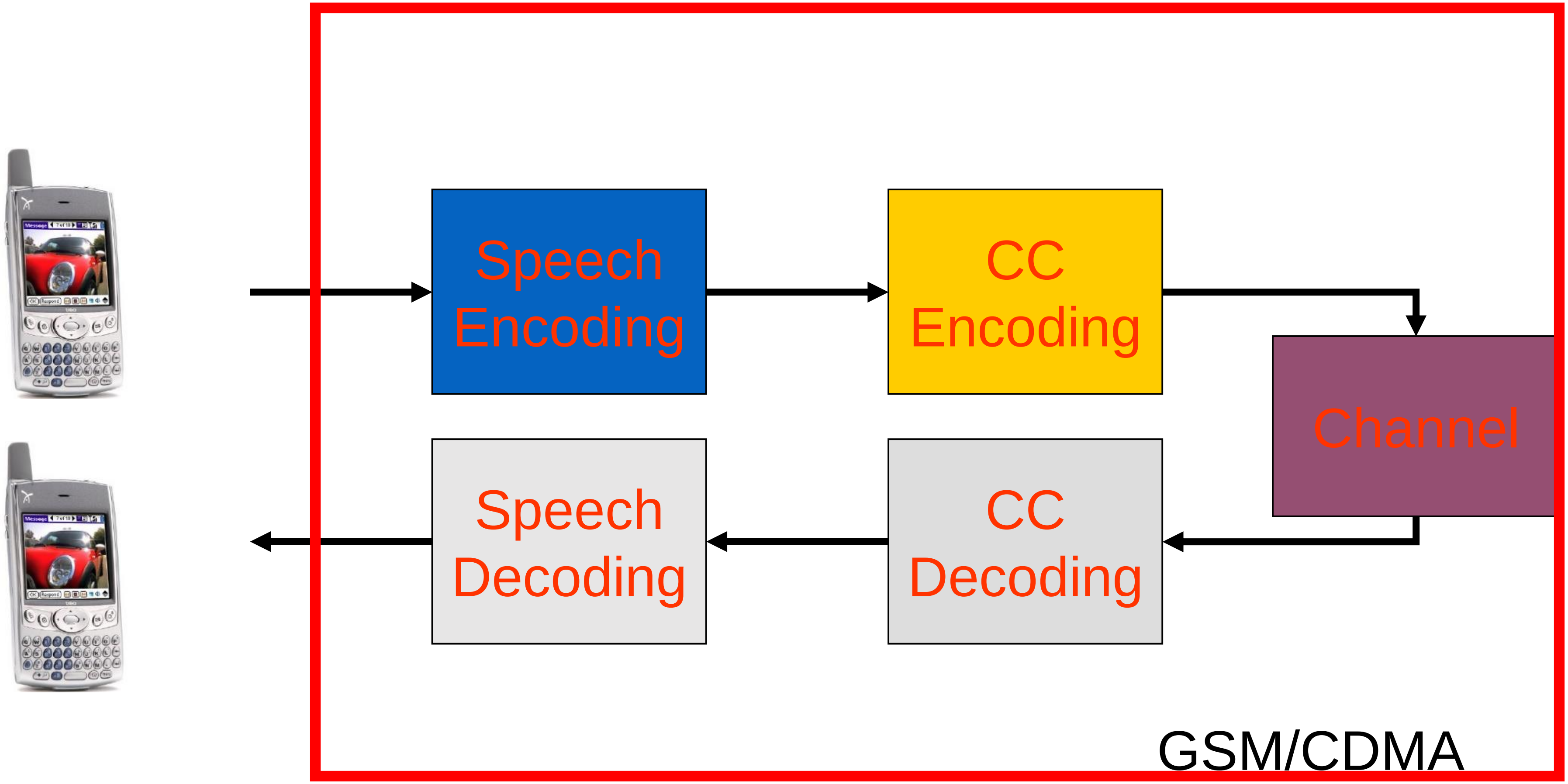
Zip	=	Source Encoding	Data Compression
Unzip	=	Source Decoding	Data Decompression
Add CRC	=	Channel Encoding	Error Protection
Verify CRC	=	Channel Decoding	Error Correction

Example: VCD and DVD



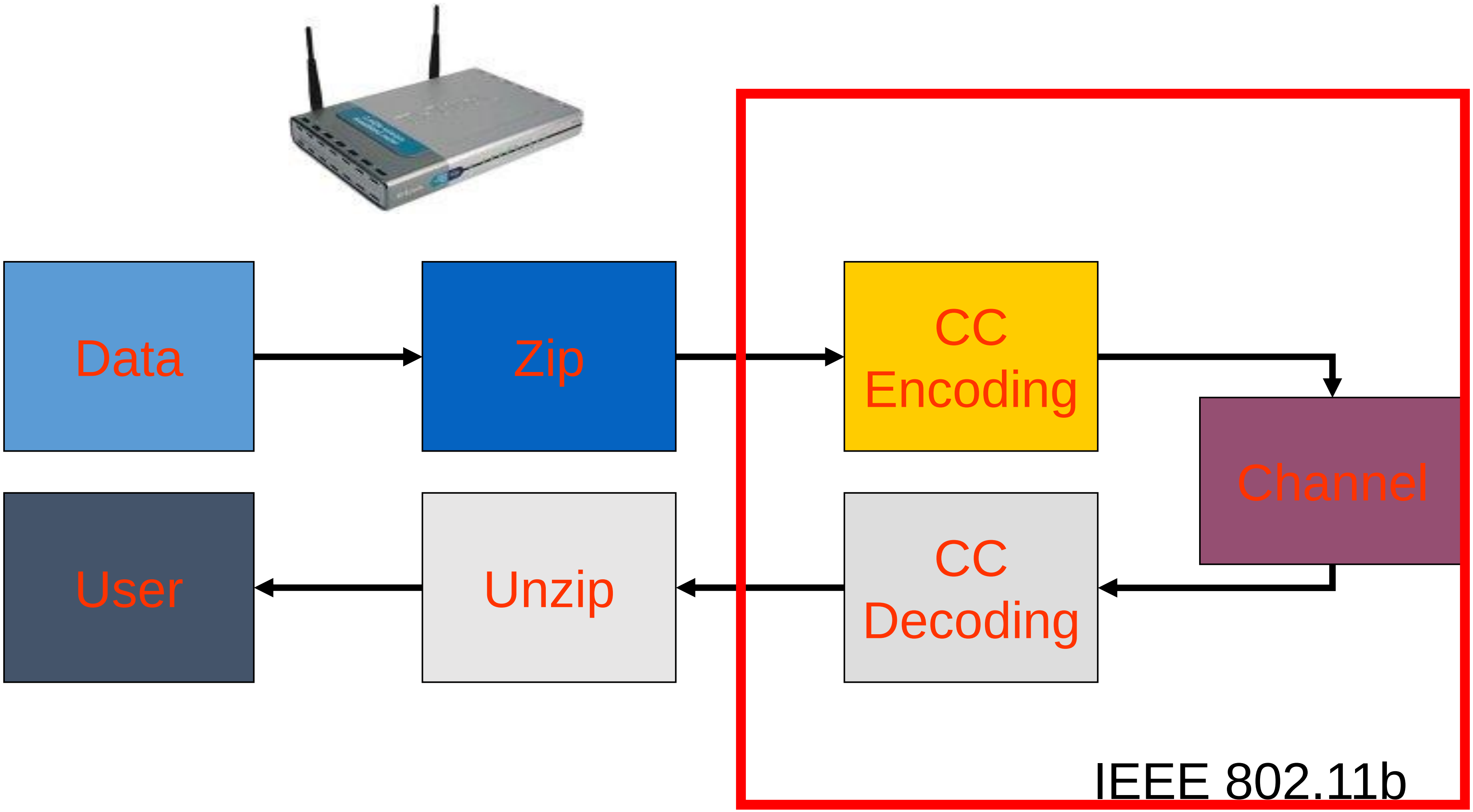
RS stands for Reed-Solomon Code.

Example: Cellular Phone

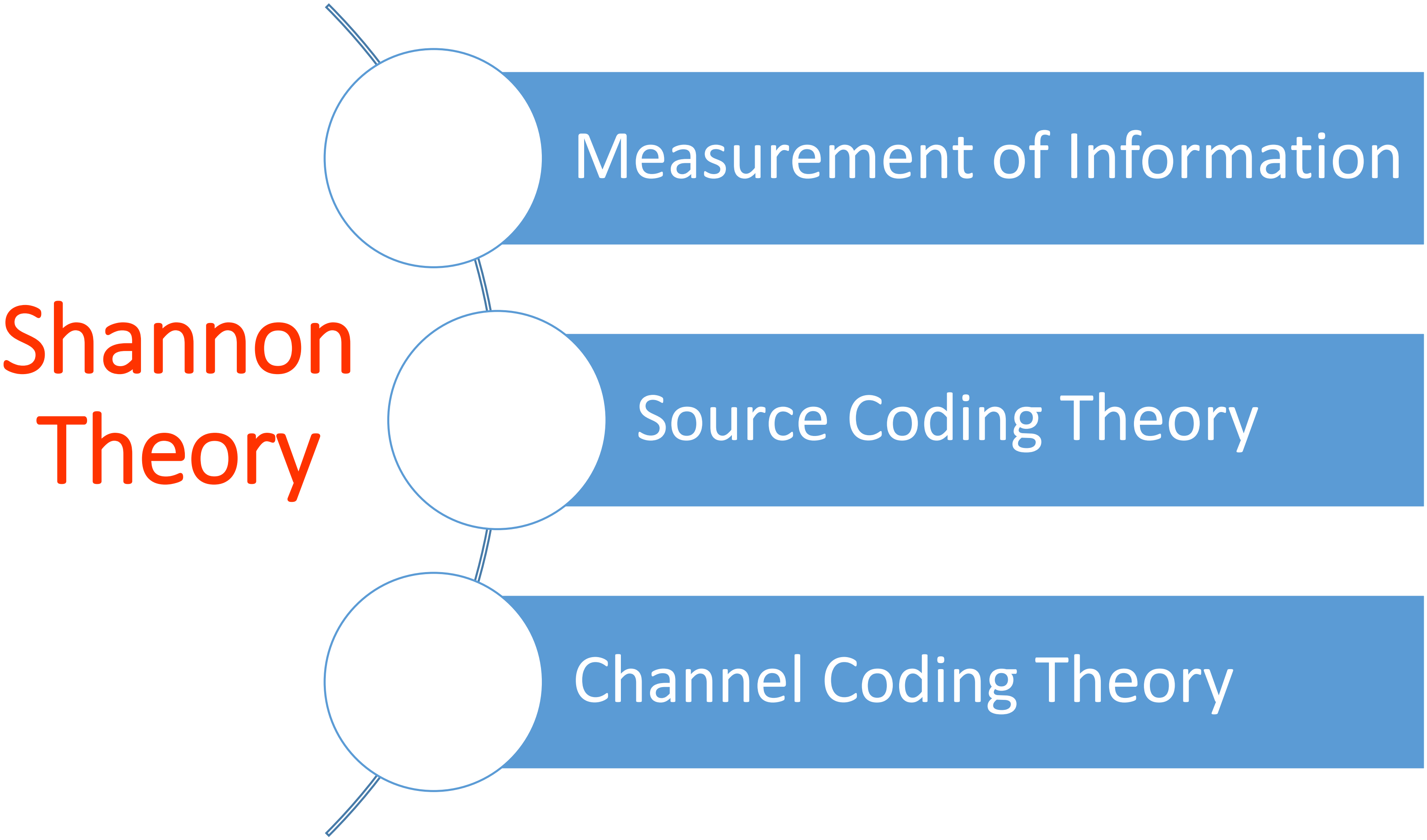


CC stands for Convolutional Code.

Example: WLAN IEEE 802.11b



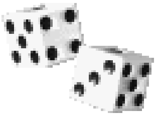
CC stands for Convolutional Code.


Shannon Theory
Measurement of Information
Source Coding Theory
Channel Coding Theory

Measurement of Information

Shannon's first question is

"How to measure information in terms of bits?"



= ? bits

= ? bits

= ? bits

= ? bits

All events are probabilistic!

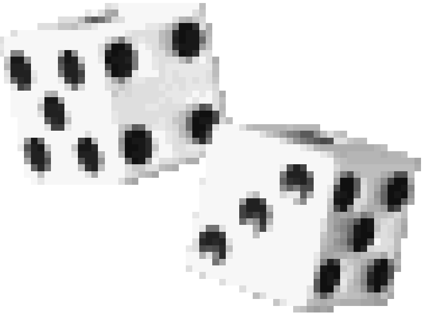
- Using Probability Theory, Shannon showed that there is only one way to measure information in terms of number of bits:

$$H(X) = -\sum_{x} p(x) \log_2 p(x)$$

called the entropy function

For example

- Tossing a dice:
  - Outcomes are 1,2,3,4,5,6
  - Each occurs at probability 1/6
  - Information provided by tossing a dice is



$$H = -\sum_{i=1}^{6} p(i)\log_2 p(i) = -\sum_{i=1}^{6} p(i)\log_2 p(i)$$

$$= -\sum_{i=1}^{6} \frac{1}{6}\log_2 \frac{1}{6} = \log_2 6 = 2.585 \text{ bits}$$

Wait!
It is nonsense!

The number 2.585-bits is not an integer!!

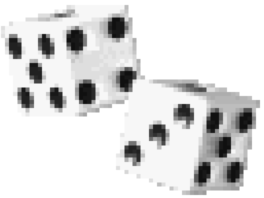
What does you mean?

Shannon's Source Coding Theorem

Shannon showed:

"To reliably store the information generated by some random source ***X***, you need no more/less than, on the average, ***H(X)*** bits for each outcome."

Meaning:



- If I toss a dice 1,000,000 times and record values from each trial

  1,3,4,6,2,5,2,4,5,2,4,5,6,1,....

- In principle, I need 3 bits for storing each outcome as 3 bits covers 1-8. So I need 3,000,000 bits for storing the information.
- Using ASCII representation, computer needs 8 bits=1 byte for storing each outcome
- The resulting file has size 8,000,000 bits

But Shannon said:

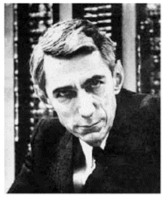
- You only need 2.585 bits for storing each outcome.
- So, the file can be compressed to yield size

2,585,000 bits

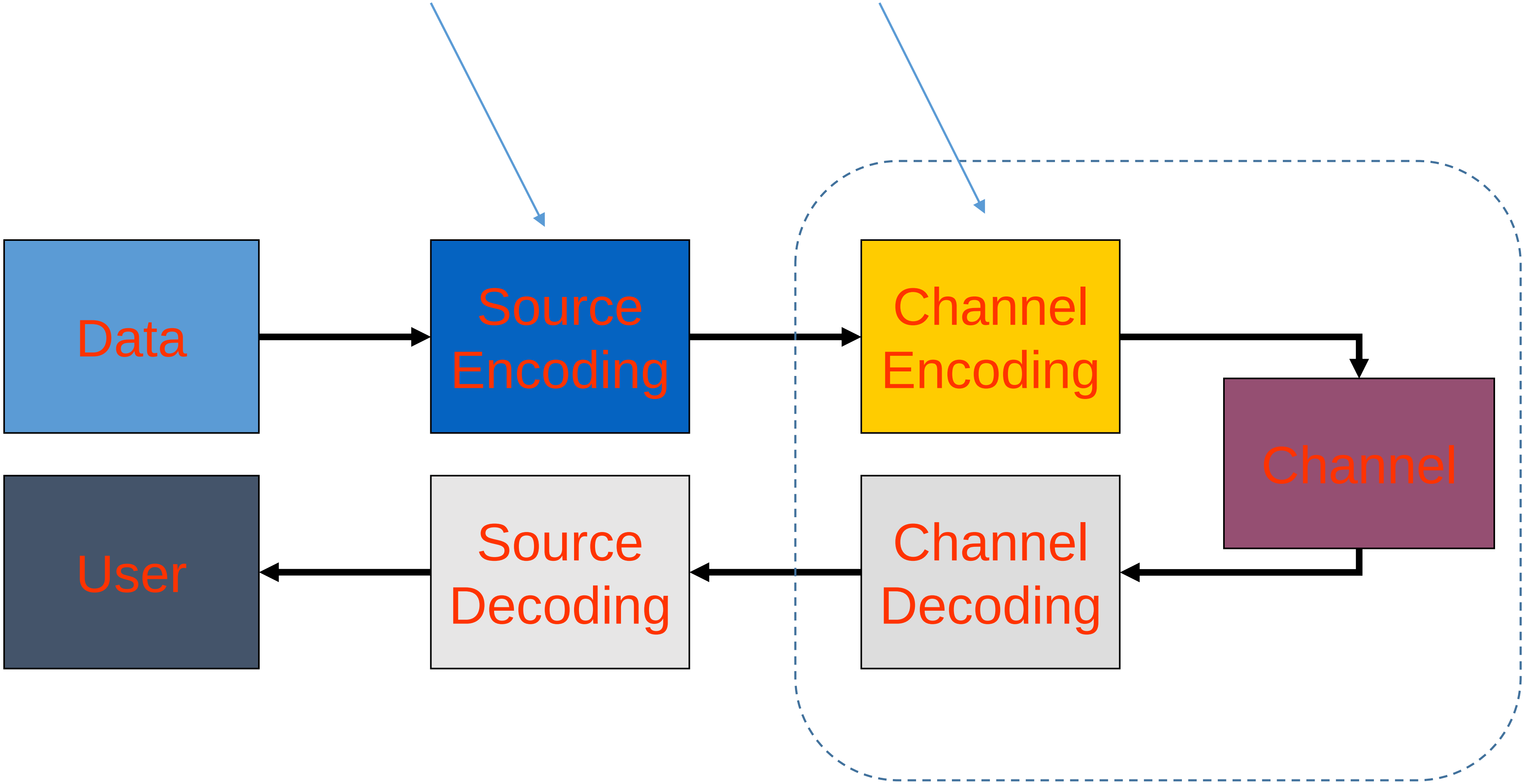
Unfortunately…

I do not know exactly HOW?

Let’s Do Some Test!

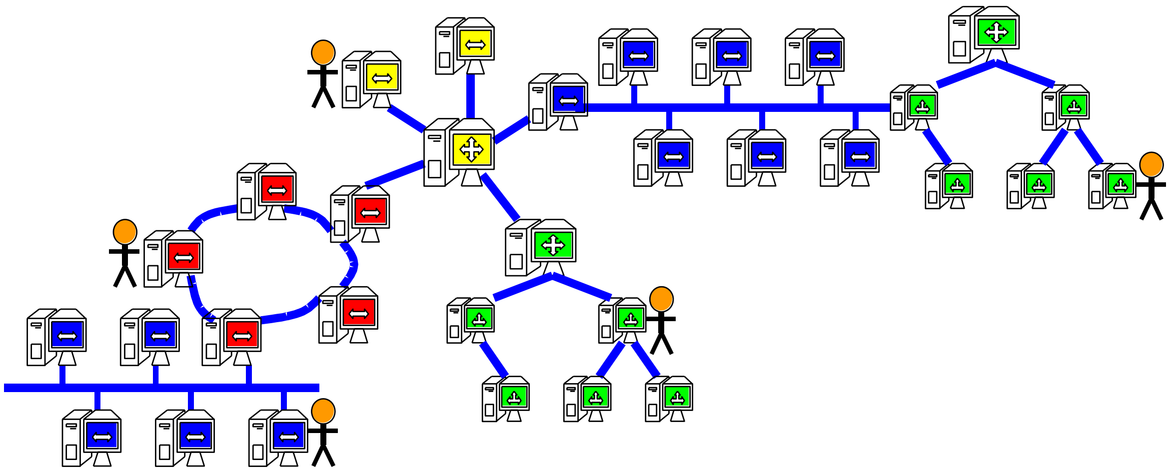
	File Size	Compression Ratio
No Compression	8,000,000 bits	100%
Winzip	2,930,736 bits	36.63%
WinRAR	2,859,336 bits	35.74%
Shannon 	2,585,000 bits	32.31%

So far… but how about?



The Simplest Case: Computer Network

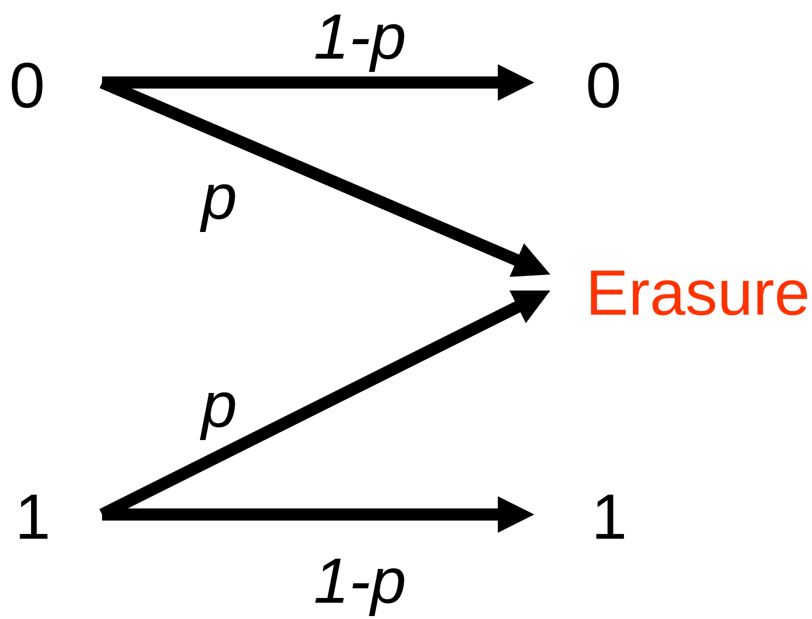
Communications over computer network,
ex. Internet



The major channel impairment herein is
Packet Loss

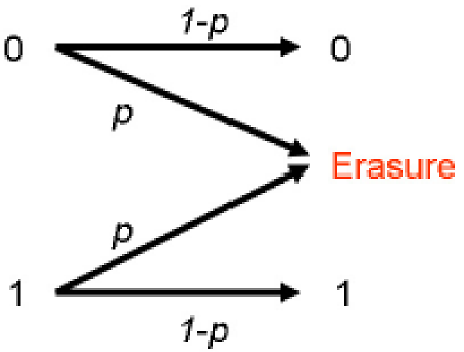
Binary Erasure Channel

Impairment like "packet loss" can be viewed as Erasures. Data that are erased mean they are lost during transmission…



p is the packet loss rate in this network

Once a binary symbol is erased,
it can not be recovered…
Ex:



- Say, Alice sends 0,1,0,1,0,0 to Bob

But the network was so poor that Bob only received 0,?,0,?,0,0

- So, Bob asked Alice to send again

Only this time he received 0,?,?,1,0,0

and Bob goes CRAZY!

- What can Alice do?

What if Alice sends
0000,1111,0000,1111,0000,0000
Repeating each transmission four times!

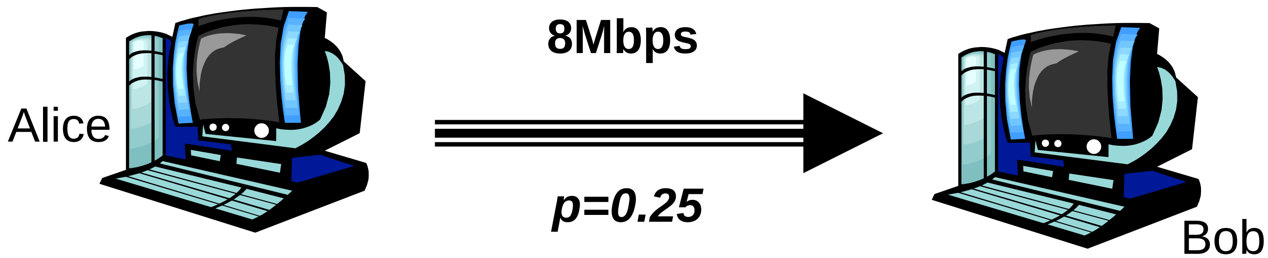
What Good Can This Serve?

- Now Alice sends 0000,1111,0000,1111,0000,0000
- The only cases Bob can not read Alice are for example

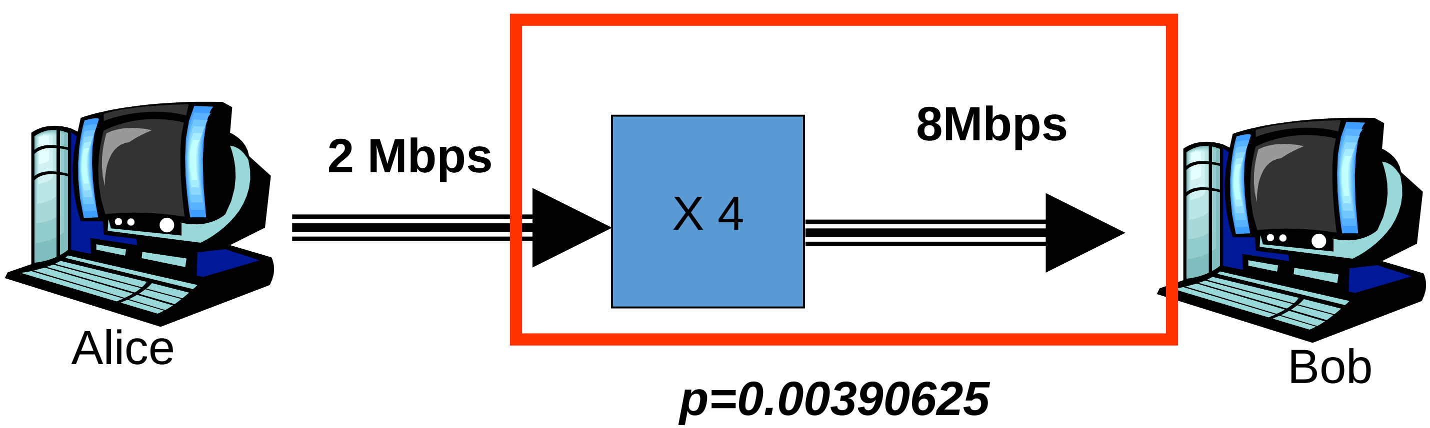
  ????,1111,0000,1111,0000,0000

  all the four symbols are erased.
- But this happens at probability p^4

- Thus if the original network has packet loss rate *p=0.25*, by repeating each symbol 4 times, the resulting system has packet loss rate $p^4=0.00390625$
- *But if the data rate in the original network is 8M bits per second*



With repetition, Alice can only transmit at 2 M bps



Shannon challenged:

Is repetition the best Alice can do?

Shannon's Channel Coding Theorem

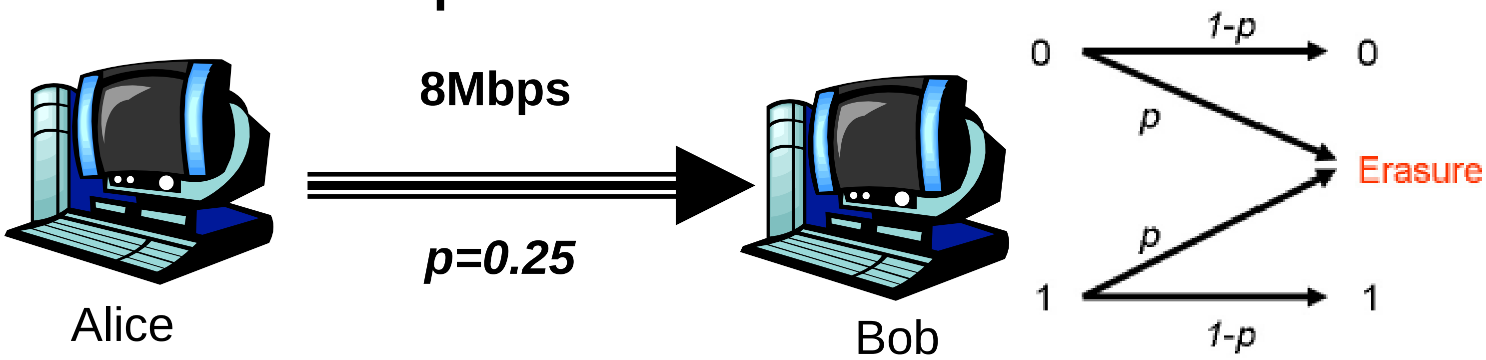
- Shannon answered:

  "Give me a channel and I can compute a quantity called capacity, *C* for that channel. Then reliable communication is possible only if your data rate stays below *C*."

What does Shannon mean?

Shannon means

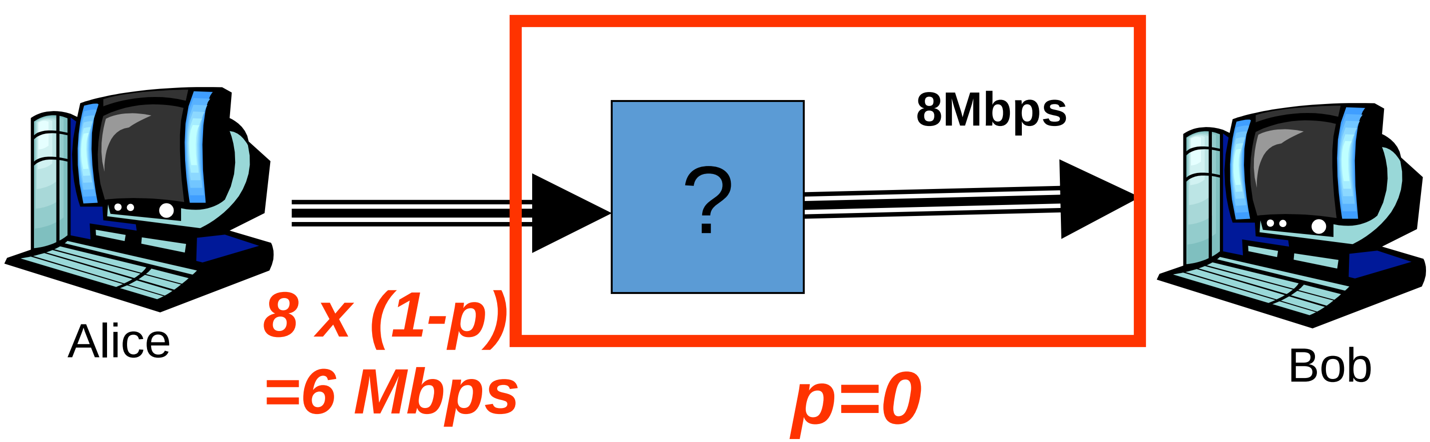
In this example:



He calculated the channel capacity

$C = 1 - p = 0.75$

And there exists coding scheme such that:



Unfortunately…

I do not know exactly HOW?

But With 50 Years of Hard Work

- **We have discovered a lot of good codes:**
  - Hamming codes
  - Convolutional codes,
  - Concatenated codes,
  - Low density parity check (LDPC) codes
  - Reed-Muller codes
  - Reed-Solomon codes,
  - BCH codes,
  - Finite Geometry codes,
  - Cyclic codes,
  - Golay codes,
  - Goppa codes
  - Algebraic Geometry codes,
  - Turbo codes
  - Zig-Zag codes,
  - Accumulate codes and Product-accumulate codes,
  - …

We now come very close to the dream Shannon had 50 years ago! ☺

Nowadays…

Source Coding Theorem has applied to

- JPEG 2000 — Image Compression
- MPEG — Audio/Video Compression
- Data Compression
- MP3 — Audio Compression

Channel Coding Theorem has applied to

- VCD/DVD – Reed-Solomon Codes
- Wireless Communication – Convolutional Codes
- Optical Communication – Reed-Solomon Codes
- Computer Network – LT codes, Raptor Codes
- Space Communication