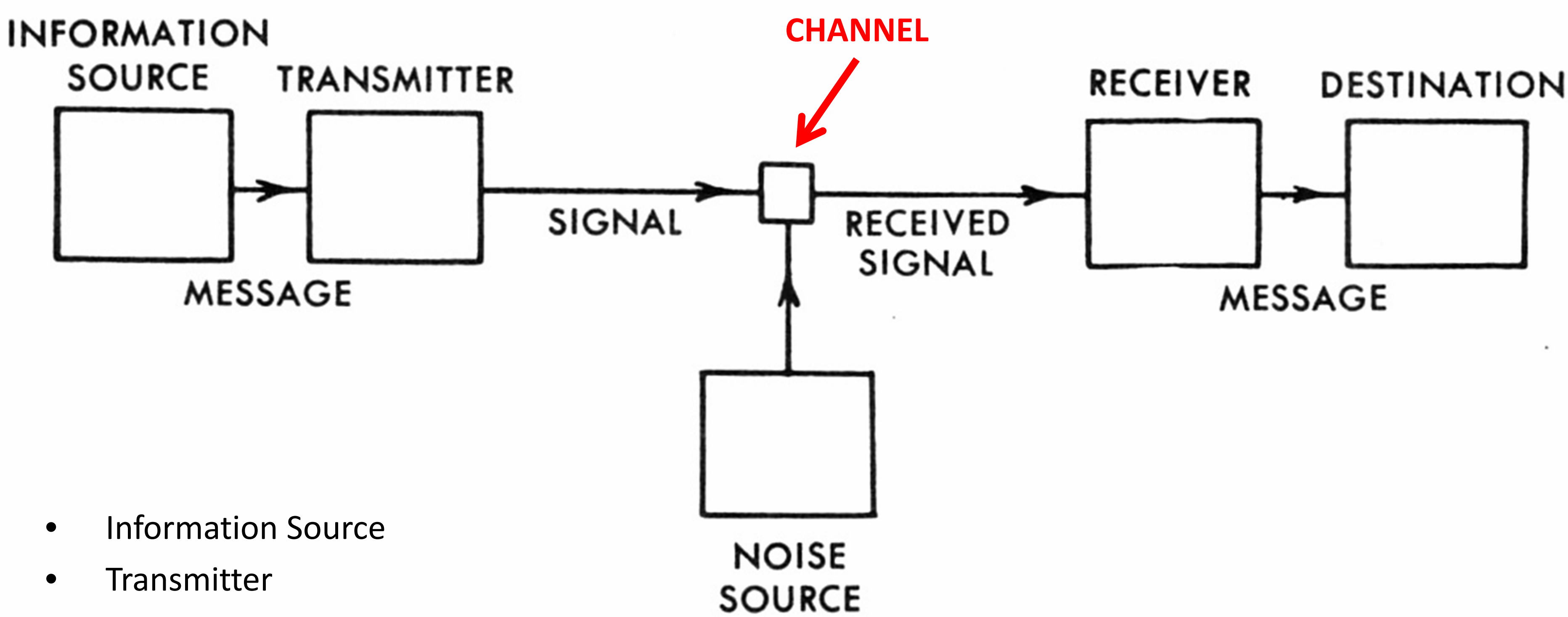
Introduction to Information Theory

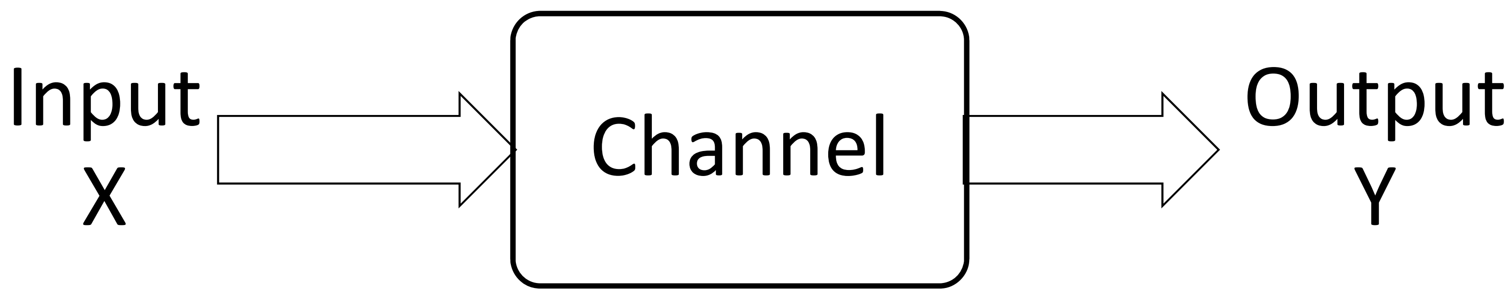
Part 4

A General Communication System

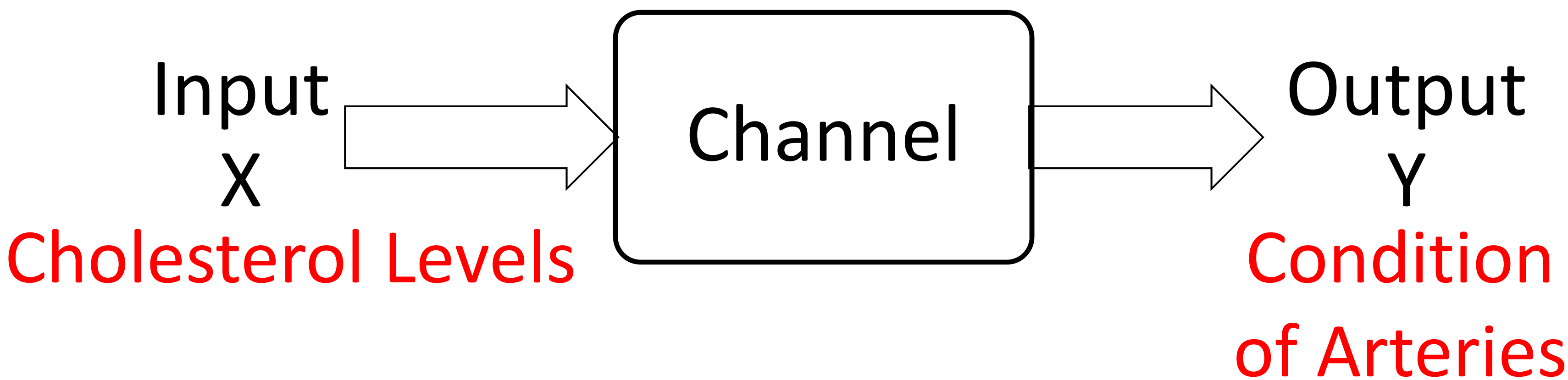


- Information Source
- Transmitter
- Channel
- Receiver
- Destination

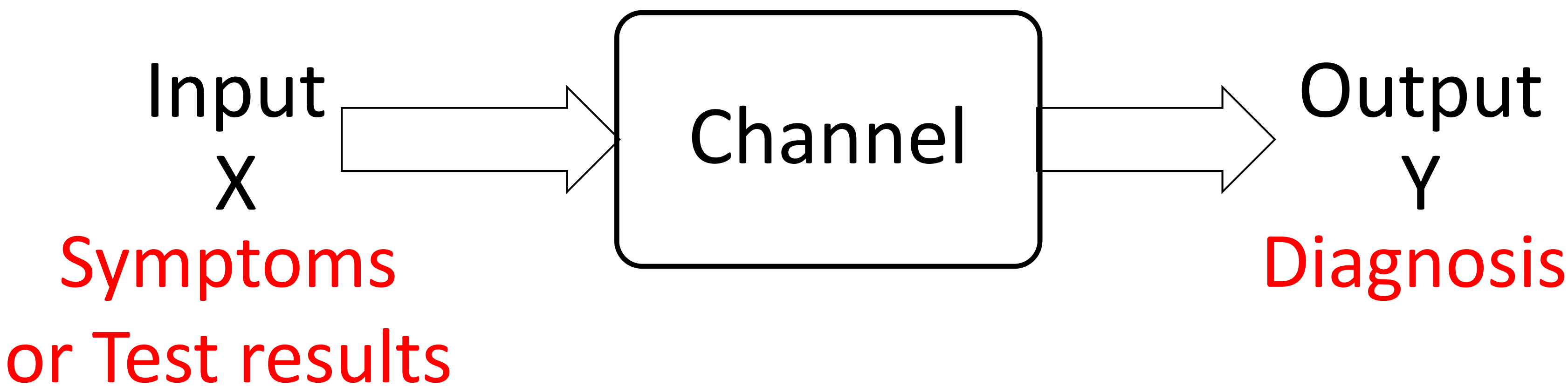
Information Channel



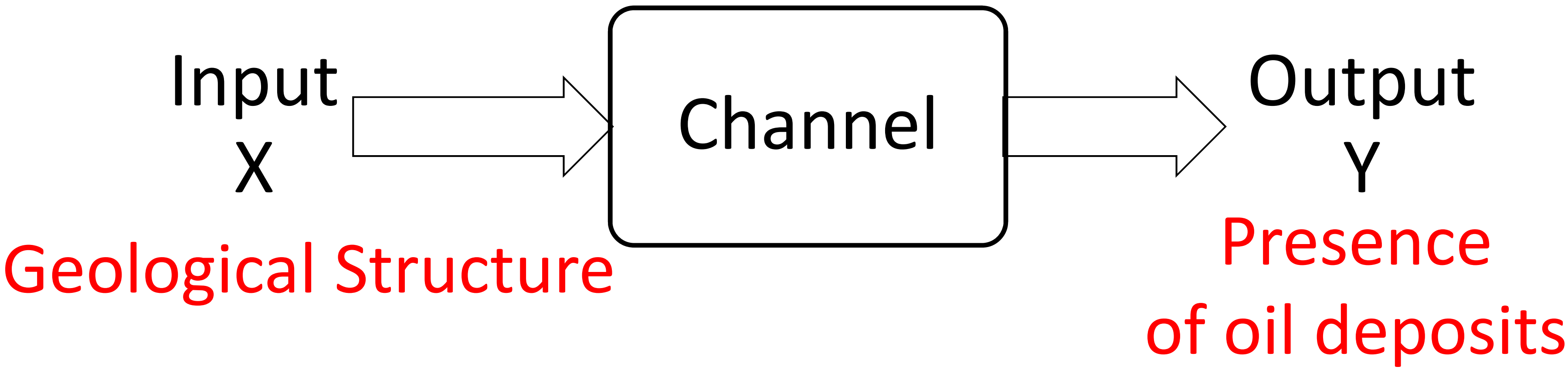
Information Channel



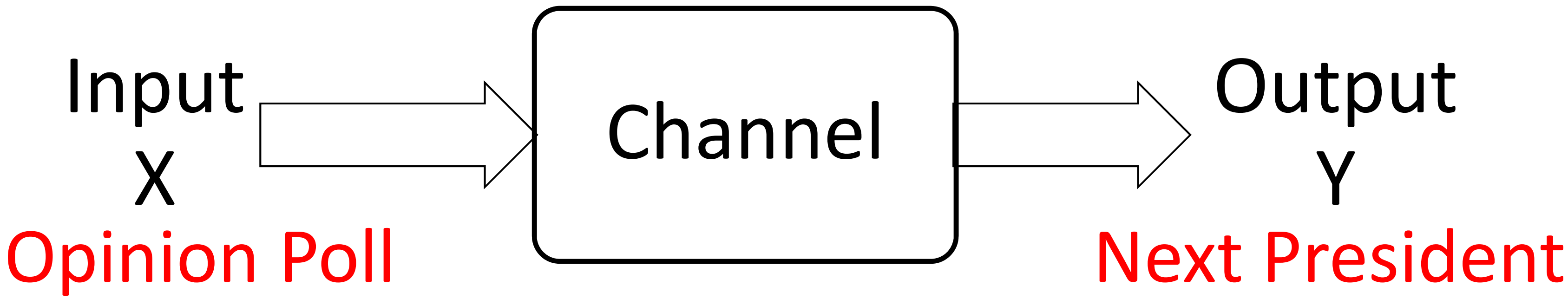
Information Channel



Information Channel

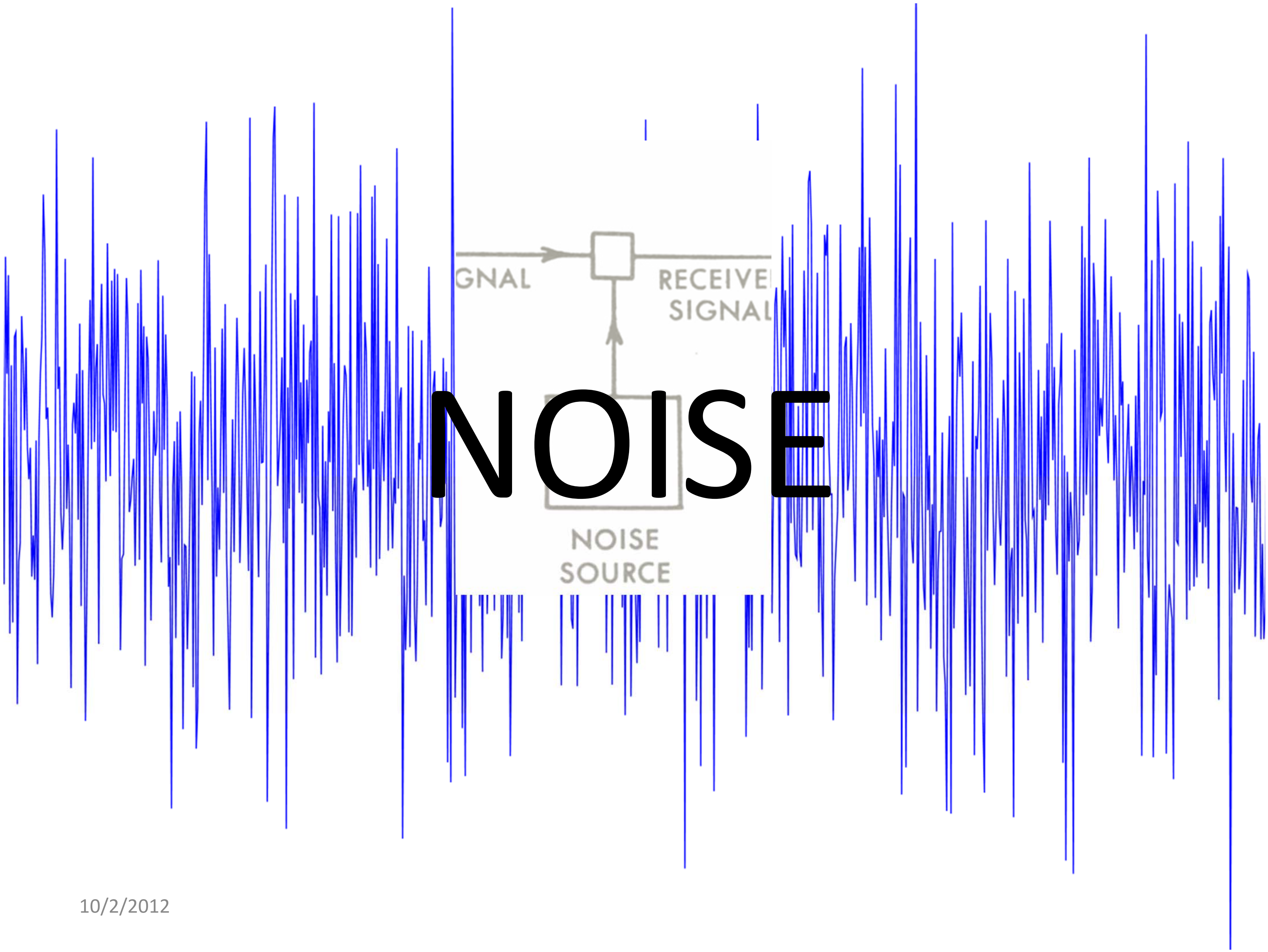


Information Channel



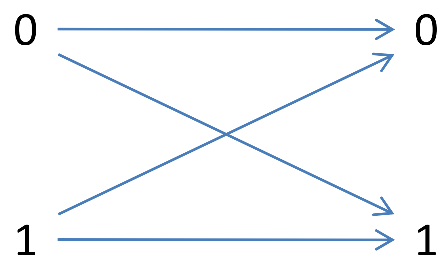
Perfect Communication (*Discrete Noiseless Channel*)

NOISE



Motivating Noise…

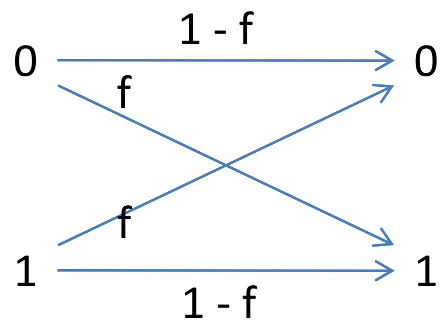
X
Transmitted
Symbol



Y
Received
Symbol

Motivating Noise…

f = 0.1, n = ~10,000



Motivating Noise…

Message: \$5213.75
Received: \$5293.75

1. Detect that an error has occurred.

2. Correct the error.

3. Watch out for the overhead.

Error Detection by Repetition

In the presence of 20% noise…

Message : $ 5 2 1 3 . 7 5

Transmission 1: $ 5 2 9 3 . 7 5
Transmission 2: $ 5 2 1 3 . 7 5
Transmission 3: $ 5 2 1 3 . 1 1
Transmission 4: $ 5 4 4 3 . 7 5
Transmission 5: $ 7 2 1 8 . 7 5

There is no way of knowing where the errors are.

Error Detection by Repetition

In the presence of 20% noise…

Message   : $ 5 2 1 3 . 7 5
Transmission 1: $ 5 2 9 3 . 7 5
Transmission 2: $ 5 2 1 3 . 7 5
Transmission 3: $ 5 2 1 3 . 1 1
Transmission 4: $ 5 4 4 3 . 7 5
Transmission 5: $ 7 2 1 8 . 7 5
Most common: $ 5 2 1 3 . 7 5

1. Guesswork is involved.
2. There is overhead.

Error Detection by Repetition

In the presence of 50% noise…

Message : $ 5 2 1 3 . 7 5

…

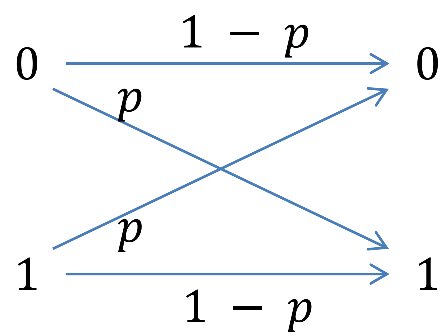
Repeat 1000 times!

1. Guesswork is involved.
   But it will almost never be wrong!

2. There is overhead.
   A LOT of it!

Binary Symmetric Channel (BSC) (Discrete Memoryless Channel)

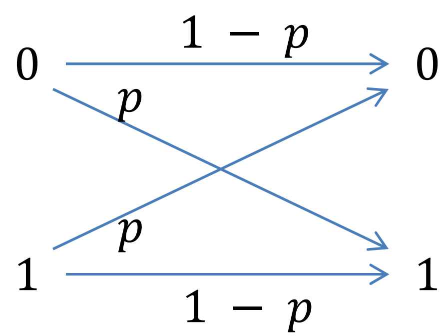
X
Transmitted
Symbol



Y
Received
Symbol

Binary Symmetric Channel (BSC)
(Discrete Memoryless Channel)

X
Transmitted
Symbol



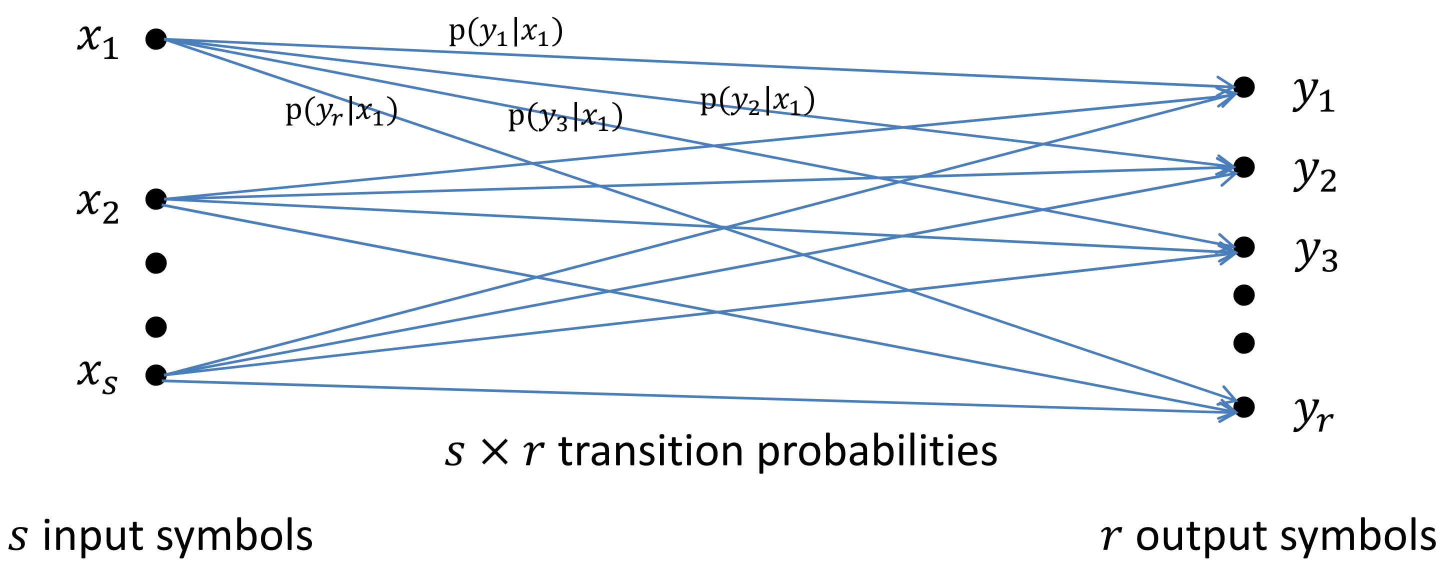
Y
Received
Symbol

Defined by a set of **conditional probabilities** (*aka* **transitional probabilities**)

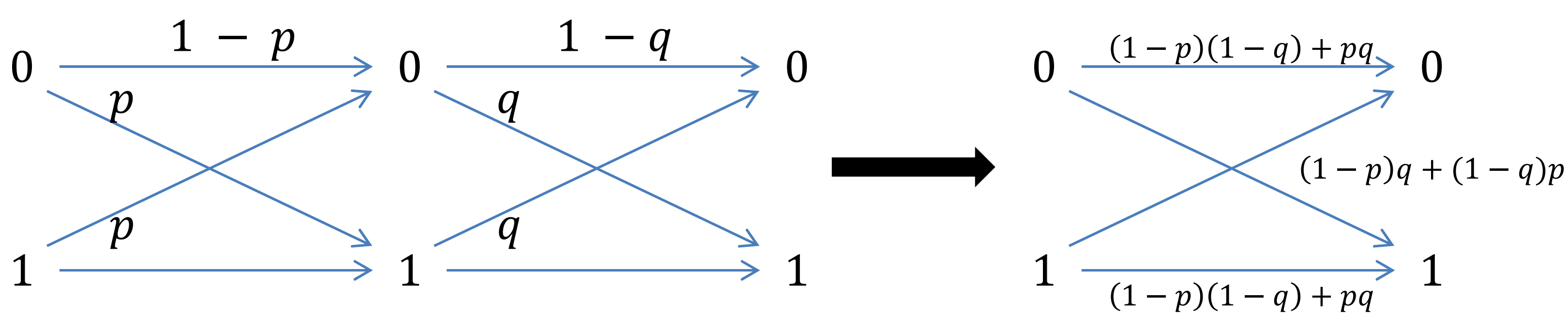
$$p(y|x) for\ all\ x \in X\ and\ y\ \in Y$$

The probability of y occurring at the output when x is the input to the channel.

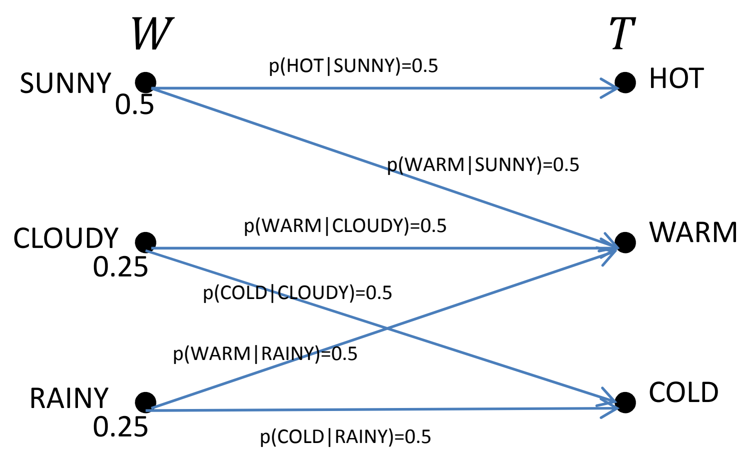
A General Discrete Channel



Channel With Internal Structure



The Weather Channel



Entropy

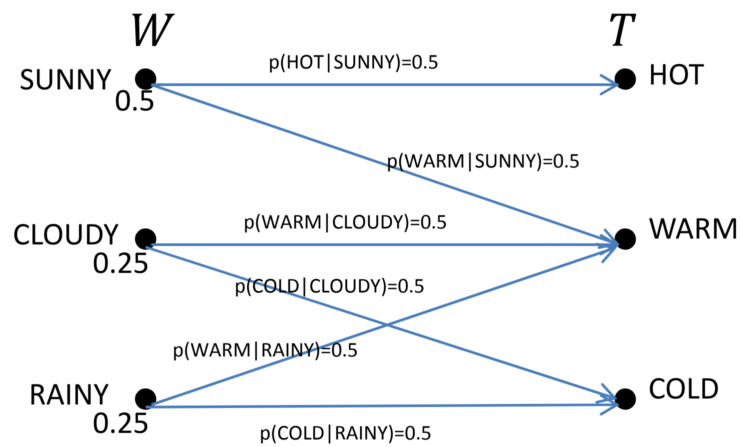
- X ,Y random variables with entropy $H(X)$ and $H(Y)$
- **Conditional Entropy:** Average entropy in Y, given knowledge of X.

$$\boldsymbol{H(Y|X) = \sum_{x_i \in X} \sum_{y_j \in Y} p(x_i, y_j)\, log \frac{1}{p(y_{j|x_i})}}$$

where $p\left(x_i, y_j\right) = p\left(y_j \middle| x_i\right) p(x_i)$

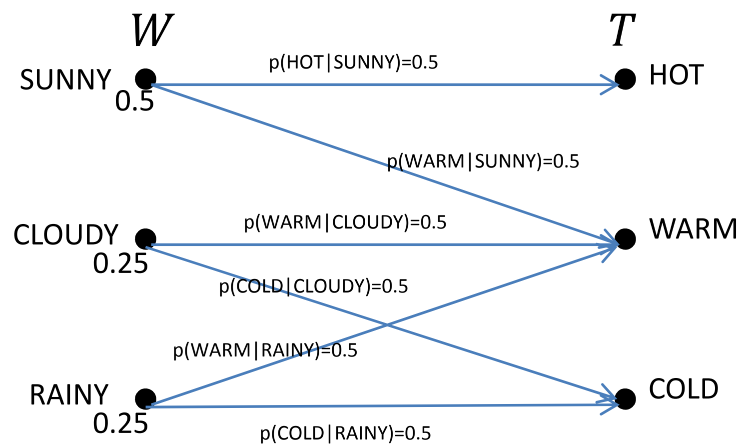
- **Joint Entropy:** $\boldsymbol{H(X,Y) = H(Y|X) + H(X)}$
  Entropy of the pair (X, Y)

The Weather Channel



Q. What is the Entropy, $H(W)$?

The Weather Channel



Q. What is the Entropy, $H(W)$?

$$
\begin{aligned}
H(W) &= 0.5\log(2) + 0.25\log(4) + 0.25\log(4) \\
&= 0.5 + 0.5 + 0.5 \\
&= 1.5 \text{ bits}
\end{aligned}
$$

Example

- $H(Y|x_i) = \sum_{y_j} p(y_j|x_i)\, log\, \frac{1}{p(y_j|x_i)}$

- Entropy of a toss of die, Y is
$$H(Y) = \log 6 = \ 2.59$$

- If outcome is HIGH (either 5 or 6):
$$H(Y|HIGH) = log2 = 1$$

- If outcome is LOW (either 1, 2, 3, or 4):
$$H(Y|LOW) = log4 = 2$$

- Conditional Entropy:
$$H(Y|X) = \frac{1}{3} log2 + \frac{2}{3} log4 = \frac{5}{3} = 1.67$$

Example

- $H(Y|x_i) = \sum_{y_j} p(y_j|x_i)\, log \frac{1}{p(y_j|x_i)}$

- Entropy of a toss of die, Y is

$$H(Y) = \log 6 = \ 2.59$$

- If outcome is HIGH (either 5 or 6):

$$H(Y|HIGH) = log2 = 1$$

- If outcome is LOW (either 1, 2, 3, or 4):

$$H(Y|LOW) = log4 = 2$$

- Conditional Entropy:

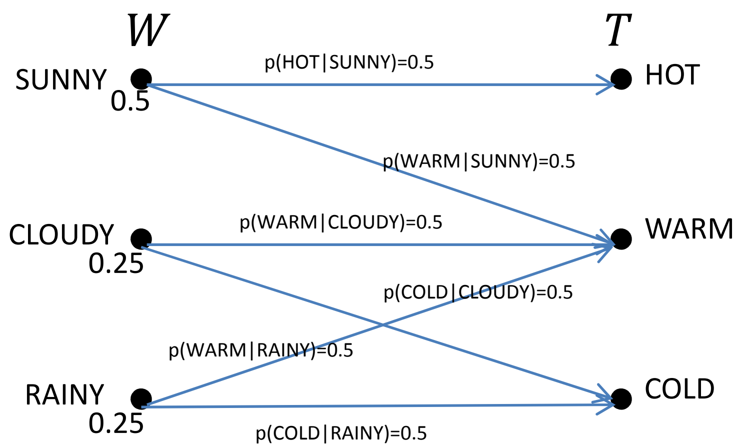
$$H(Y|X) = \frac{1}{3} log2 + \frac{2}{3} log4 = \frac{5}{3} = 1.67$$

Entropy Reduction:

$$\mathbf{0 <= \boldsymbol{H}(\boldsymbol{Y}|\boldsymbol{X}) <= \boldsymbol{H}(\boldsymbol{Y})}$$

Entropy of a variable Y is, on average, never increased by knowledge of another variable X.

The Weather Channel

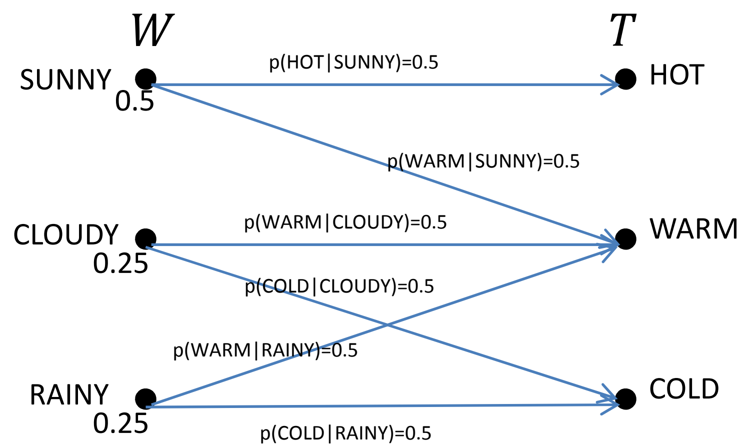


Q. What is the Entropy, $H(T|W)$?

$$H(T|SUNNY) = p(HOT|SUNNY)\log\left(\frac{1}{p(HOT|SUNNY)}\right) + p(WARM|SUNNY)\log\left(\frac{1}{p(WARM|SUNNY)}\right)$$
$$= 0.5\log(2) + 0.5\log(2)$$
$$= 1$$

$\therefore H(T|W) = 1$

The Weather Channel



Q. What is the Entropy, $H(T)$?

$$p(HOT) = 0.5 * 0.5 = 0.25$$
$$p(WARM) = 0.5 * 0.5 + 0.25 * 0.5 + 025 * 0.5 = 0.5$$
$$p(COLD) = 0.25 * 0.5 + 0.25 * 0.5 = 0.25$$

$$\therefore H(T) = 1.5$$

Mutual Information

- The mutual information of a random variable X given the random variable Y is

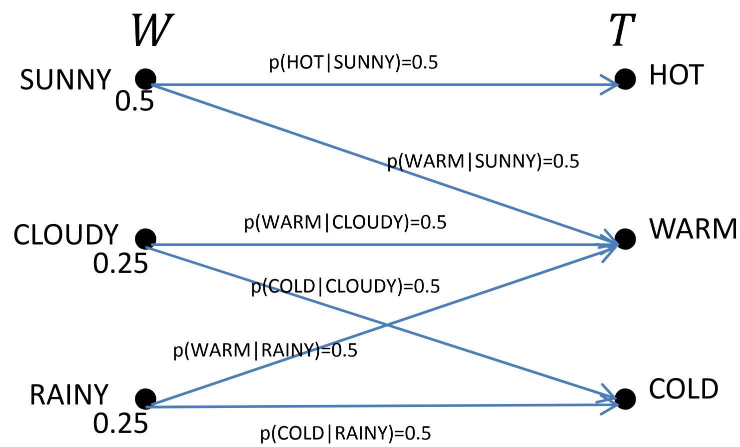
$$I(X;Y) = H(X) - H(X|Y)$$

It is the information about X transmitted by Y.

Mutual Information: Properties

- $I(X;Y) = H(X,Y) - H(X|Y) - H(Y|X)$
- $I(X;Y) = H(X) - H(X|Y)$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y)$ is symmetric in X and Y
- $I(X;Y) = \sum_x \sum_y p(x,y) log \frac{p(x,y)}{p(x)p(y)}$
- $I(X;Y) >= 0$
- $I(X;X) = H(X)$

The Weather Channel



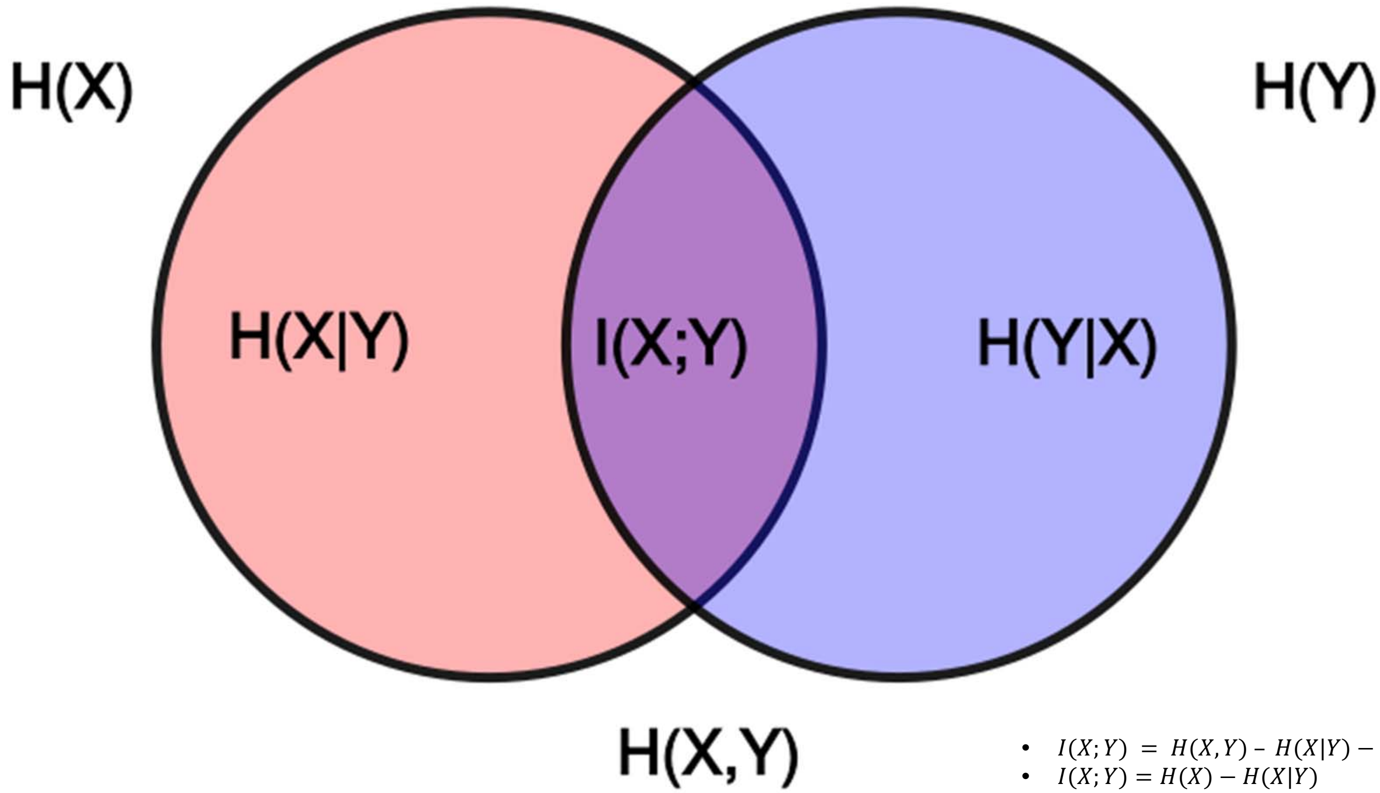
Q. What is the mutual information $I(W;T)$?

$$I(W;T) = H(T) - H(T|W) = 1.5 - 1.0 = 0.5$$

Also

$$I(T;W) = I(W;T) = 0.5$$

Entropy Concepts



- $I(X;Y) \; = \; H(X,Y) - H(X|Y) - H(Y|X)$
- $I(X;Y) = H(X) - H(X|Y)$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y)$ is symmetric in X and Y
- $I(X;Y) = \sum_x \sum_y p(x,y) log \frac{p(x,y)}{p(x)p(y)}$
- $I(X;Y) >= 0$
- $I(X;X) \; = \; H(X)$

Channel Capacity

- The capacity of a channel is the maximum possible mutual information that can be achieved between input and output by varying the probabilities of the input symbols.

  If X is the input channel and Y is the output, the capacity C is

$$\boldsymbol{c} = \max_{\boldsymbol{input\ probabilities}} \boldsymbol{I}(\boldsymbol{X};\boldsymbol{Y})$$

Channel Capacity

$$c = \max_{input\ probabilities} I(X;Y)$$

Mutual information about X given Y is the information transmitted by the channel and depends on the probability structure

- Input probabilities
- Transition probabilities
- Output probabilities

Channel Capacity

$$c = \max_{input\ probabilities} I(X;Y)$$

Mutual information about X given Y is the information transmitted by the channel and depends on the probability structure

– Input probabilities
– Transition probabilities: fixed by properties of channel
– Output probabilities: determined by input and transition probabilities

Channel Capacity

$$c = \max_{input\ probabilities} I(X;Y)$$

Mutual information about X given Y is the information transmitted by the channel and depends on the probability structure

- Input probabilities: can be adjusted by suitable coding
- Transition probabilities: fixed by properties of channel
- Output probabilities: determined by input and transition probabilities

Channel Capacity

$$\boldsymbol{c} = \max_{\boldsymbol{input\ probabilities}} \boldsymbol{I}(\boldsymbol{X};\boldsymbol{Y})$$

Mutual information about X given Y is the information transmitted by the channel and depends on the probability structure

- Input probabilities: can be adjusted by suitable coding
- Transition probabilities: fixed by properties of channel
- Output probabilities: determined by input and transition probabilities

That is, input probabilities determine mutual information and can be varied by coding. The maximum mutual information with respect to these input probabilities is the channel capacity.

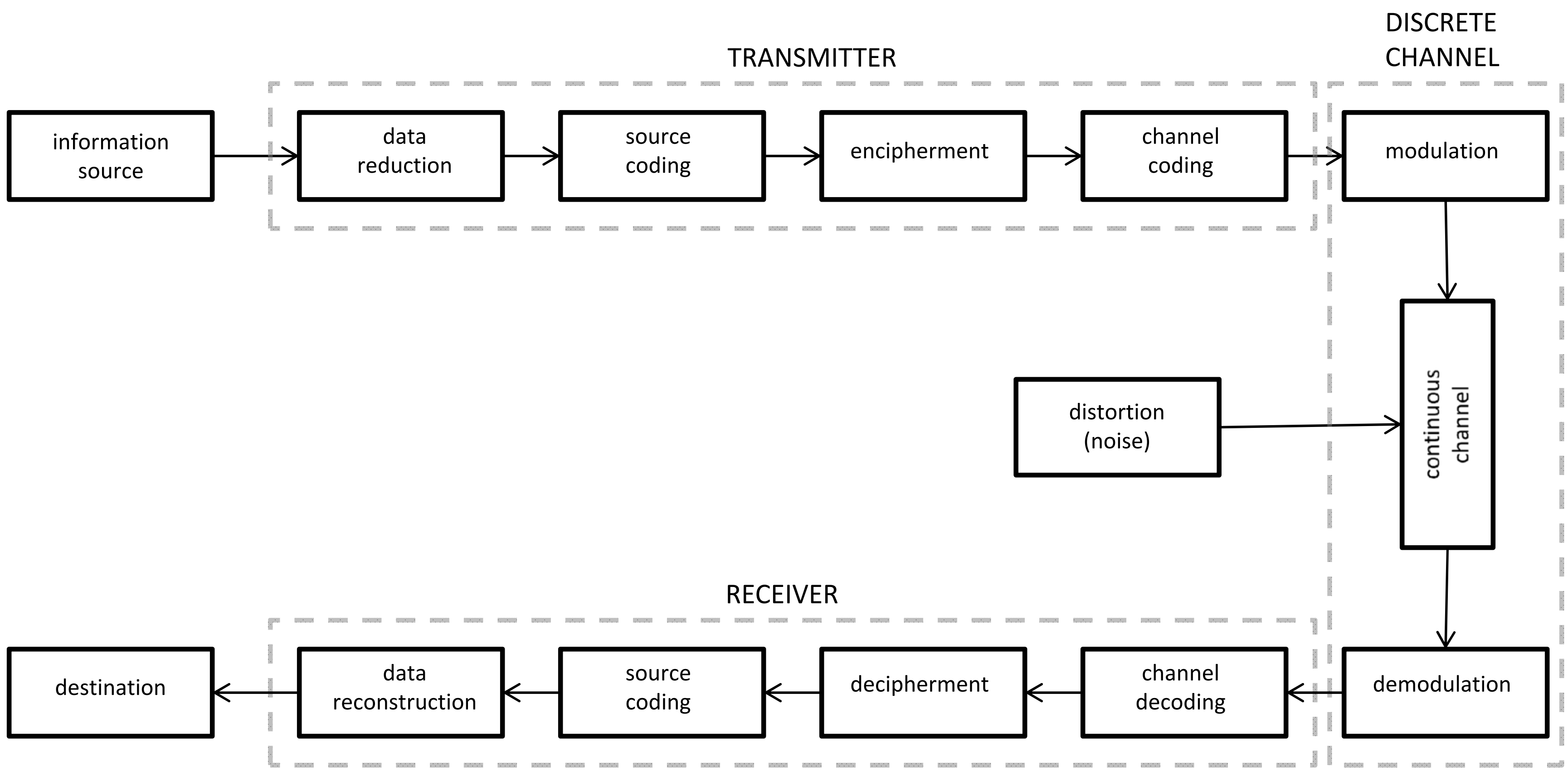
Shannon’s Second Theorem

- Suppose a discrete channel has capacity C and the source has entropy H

  If H < C there is a coding scheme such that the source can be transmitted over the channel with an arbitrarily small frequency of error.

  If H > C, it is not possible to achieve arbitrarily small error frequency.

Detailed Communication Model



Error Correcting Codes: Checksum

- ISBN: 0-691-12418-3

- 1*0+2*6+3*9+4*1+5*1+6*2+7*4+8*1+9*8
  = 168 mod 11 = 3

- This is a staircase checksum

Error Correcting Codes

- Hamming Codes (1950)
- Linear Codes
- Low Density Parity Codes (1960)
- Convolutional Codes
- Turbo Codes (1993)

MTC: Summary

- Information
- Entropy
- Source Coding Theorem
- Redundancy
- Compression
- Huffman Encoding
- Lempel-Ziv Coding
- Channel
- Conditional Entropy
- Joint Entropy
- Mutual Information
- Channel Capacity
- Shannon's Second Theorem
- Error Correction Codes

References

- Eugene Chiu, Jocelyn Lin, Brok Mcferron, Noshirwan Petigara, Satwiksai Seshasai: *Mathematical Theory of Claude Shannon: A study of the style and context of his work up to the genesis of information theory. MIT* 6.933J / STS.420J The Structure of Engineering Revolutions
- Luciano Floridi, 2010: *Information: A Very Short Introduction*, Oxford University Press, 2011.
- Luciano Floridi, 2011: *The Philosophy of Information*, Oxford University Press, 2011.
- James Gleick, 2011: *The Information: A History, A Theory, A Flood*, Pantheon Books, 2011.
- Zhandong Liu , Santosh S Venkatesh and Carlo C Maley, 2008: *Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples*, *BMC Genomics* 2008, **9:**509
- David Luenberger, 2006: *Information Science*, Princeton University Press, 2006.
- David J.C. MacKay, 2003: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- Claude Shannon & Warren Weaver, 1949: *The Mathematical Theory of Communication*, University of Illinois Press, 1949.
- W. N. Francis and H. Kucera: *Brown University Standard Corpus of Present-Day American English*, Brown University, 1967.
- Edward L. Glaeser: A Tale of Many Cities, New York Times, April 10, 2010. Available at: http://economix.blogs.nytimes.com/2010/04/20/a-tale-of-many-cities/
- Alan Rimm-Kaufman, The Long Tail of Search. Search Engine Land Website, September 18, 2007. Available at: http://searchengineland.com/the-long-tail-of-search-12198