Information Theory and Channel Coding

Prof. Rodrigo C. de Lamare
CETUC, DEE, PUC-Rio, Brazil
delamare@puc-rio.br

Textbooks and Assessment

- Textbooks:
  - Thomas Cover and Joy Thomas, Elements of Information Theory, 2nd Edition, 2006.
  - Jorge Moreira and Patrick Farrell, Essentials of Error-Control Coding, 2006, Wiley.

- Assessment:
  - 1 Exam papers (E) on Information Theory.
  - 1 Project (P) on a topic on Channel Coding chosen by the student in agreement with the lecturer.
  - 8 Lists of exercises (LE) on all topics.
  - Final grade (FG) = (E + P + LE)/3

Syllabus

Part II: Channel coding

I. Introduction
   - Fundamentals, system parameters, channels
   - Shannon limits and decoding principles

II. Linear block codes
   - Encoding
   - Syndrome decoding

III. Low-density parity-check (LDPC) codes
   - Encoding and design
   - Decoding with message passing

IV. Convolutional codes
   - Encoding
   - Decoding with the Viterbi algorithm

V. Turbo codes
   - Encoding
   - Iterative decoding

VI. Polar codes

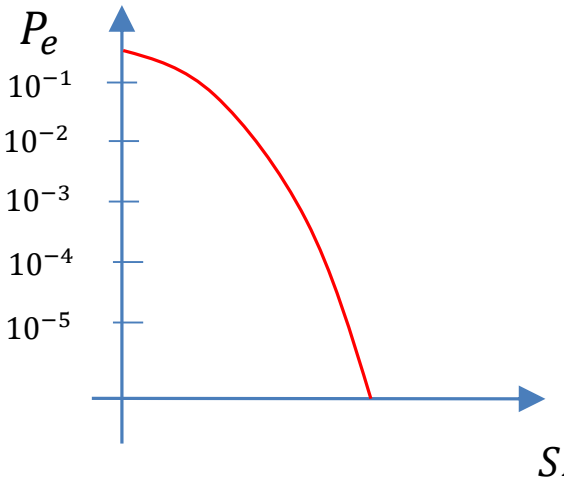
I. Introduction

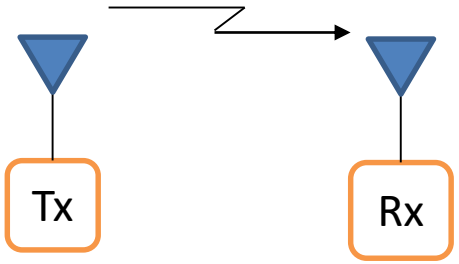
- Channel coding deals with mathematical mappings of messages with the purpose of transmitting data and protecting them from errors.
- The design of channel codes often involves mathematical mappings of messages using various strategies to produce codewords.
- Codewords are then transmitted over communication channels.
- At the receiver, a decoder is often used to perform decoding of a received signal and estimate the transmitted codeword.

A. Motivation

- The fundamental problem in communications is to reproduce a message that has been transmitted at the receiver.

- Reliable transmissions are an important goal in digital communications that is often measured in terms of probability of symbol error P_e.





- In order to obtain reliable transmissions, we need to employ channel coding techniques that are often designed for specific purposes.

- Why use channel codes?
  - Detection and correction of errors.
  - To increase the reliability of data transmission.
- Rationale:
  - To introduce redundancy.
  - To design codes with enough structure so that they can be easily decoded.
- System parameters:
  - Transmit power P_T
  - Waveforms
  - Bandwidth B

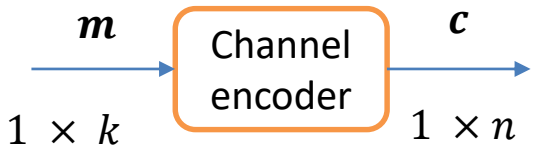
  } Under control

  - Noise
  - Fading

  } Out of control

- Channel coding increases the resistance against channel errors in digital transmissions.

- The basic idea of channel coding is to introduce redundancy.



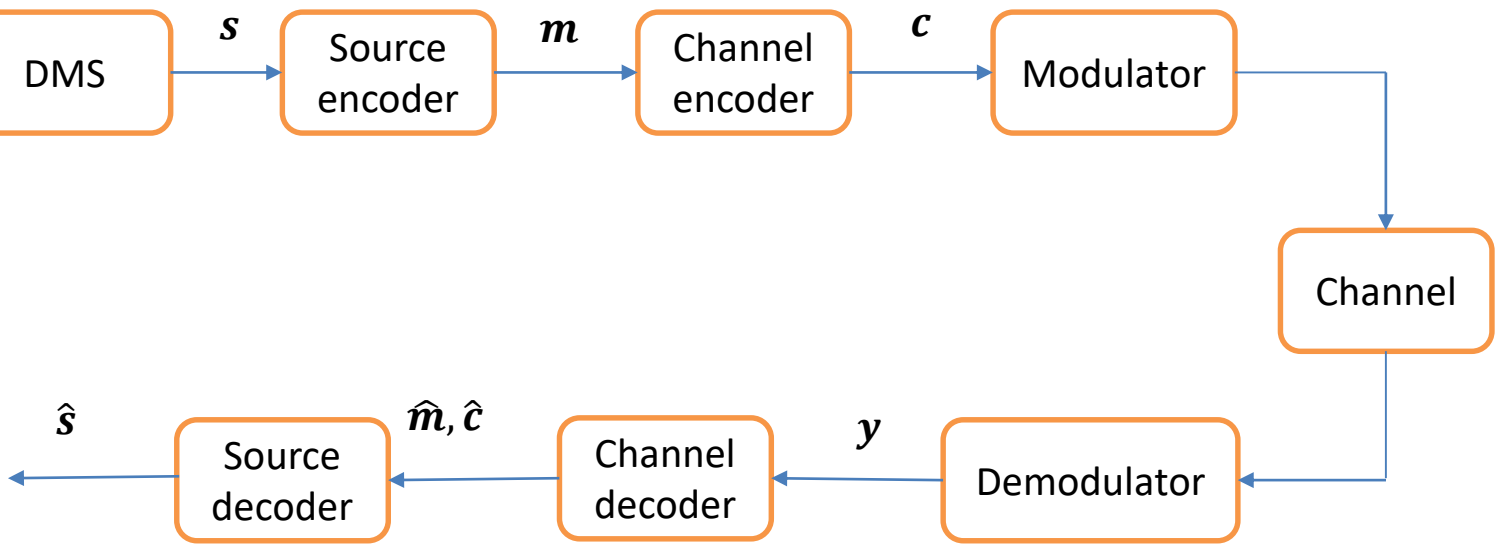
- A message $\boldsymbol{m}$ with k bits is mapped into a codeword $\boldsymbol{c}$ with n code bits, which is then transmitted.

- This redundancy translates into the code rate

$$R = \frac{k}{n}, \quad 0 < R < 1$$

B. Digital communication system

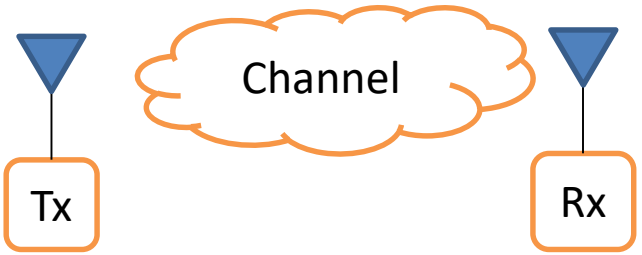
- Digital transmission over a channel with capacity C involves several operations such as source coding, channel coding, modulation and decoding.



- In what follows, we will detail all the quantities outlined in the block diagram.

- The message vector $\boldsymbol{m} = [m_0\, m_1\, \ldots m_{k-1}]$ contains k 0s and 1s .
- The data rate is given by $\frac{1}{T_b}$ bits / s
- The transmit power is P_T in Watts
- The codeword $\boldsymbol{c} = [c_0\, c_1\, \ldots c_{n-1}]$ contains n 0s and 1s .
- The code rate is $R = \frac{k}{n}$
- The noise power is $P_N = \sigma^2 = \frac{N_0}{2}$
- The received power is described by $P_R = L\, P_T$, where L is the propagation loss.
- The signal-to-noise ratio (SNR) is $SNR = \frac{P_R}{P_N}$

C. Communication channels



- Physical medium between the transmitter (Tx) and the receiver (Rx)
- Discrete versus continuous models.
- Memoryless channels versus channels with memory.

- Data transmission over communication channels is performed in blocks of (coded) bits, i.e., data packets. In this case, we have the input block

$$\boldsymbol{x} = [x_1\ x_2\ \dots x_n]$$

and the output of the channel

$$\boldsymbol{y} = [y_1\ y_2\ \dots y_n]$$

- A key property of discrete memoryless channels is given by

$$p_{\boldsymbol{y}|\boldsymbol{x}}(Y_1, Y_2, \dots, Y_n | X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} p_{y_i|x_i}(Y_j|X_i)$$

- The above property extends to probabilities and allow simpler decoding of data blocks due to the decoupling of the conditional pdfs and prob.

Example 1

Consider a continuous memoryless channel with bandwidth B and the following received signal:

$$y(t) = x(t) + n(t),$$

where $x(t)$ is the information transmitted and $n(t)$ is AWGN with zero mean and variance $\sigma^2 = \frac{N_0}{2}$.

a) Draw a block diagram and a simple graph diagram of this channel.

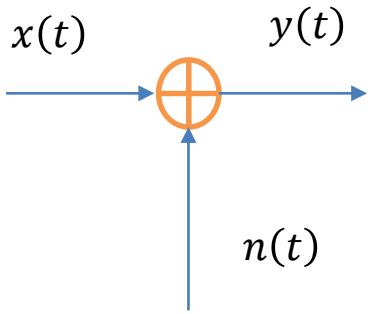
b) Describe the capacity of the channel.

c) Write down the probability of symbol error for binary signalling.

Solution:

a) This model for the received signal refers to a continuous channel model whose block diagram is given by

whereas a graph diagram for this channel is illustrated by



b) Since the channel model is continuous the information capacity is given by

$$C = B \log_2 \left(1 + \frac{P_R}{N_0 B}\right) \text{ bits/ s}$$

c) Since binary signalling is employed, the probability of symbol error for BPSK and PAM 2 is

$$P_e = Q\left(\sqrt{\frac{d^2}{2N_0}}\right)$$

Example 2

Consider a binary memoryless channel with the following received signal:

$$y_k = x_k + n_k, \quad k = 0,1,$$

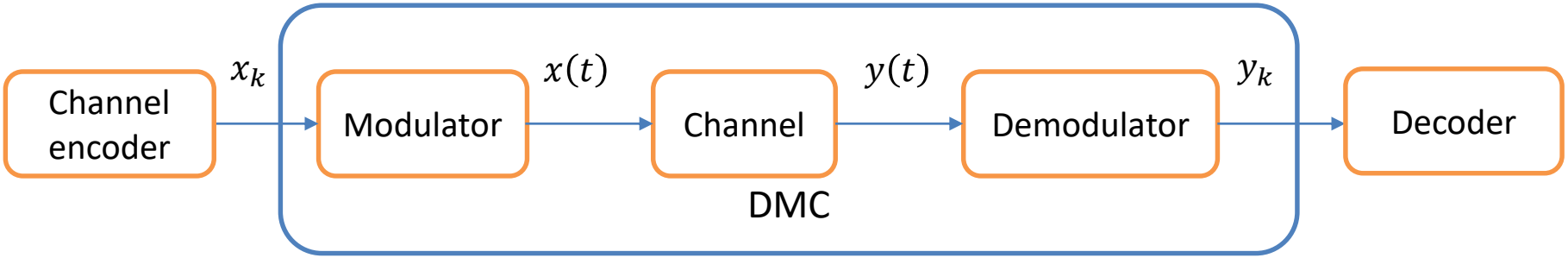
where $x_k = X_k$ with $X_0 = 0, X_1 = 1$ is the information transmitted and n_k is the noise that might change the transmitted bit.

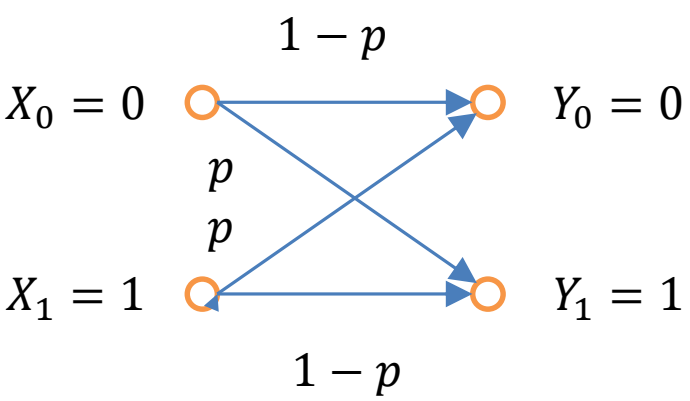
a) Draw a block diagram of this channel.

b) Compute the capacity of the channel.

Solution:

a) A communication system with a discrete memoryless channel can be represented by

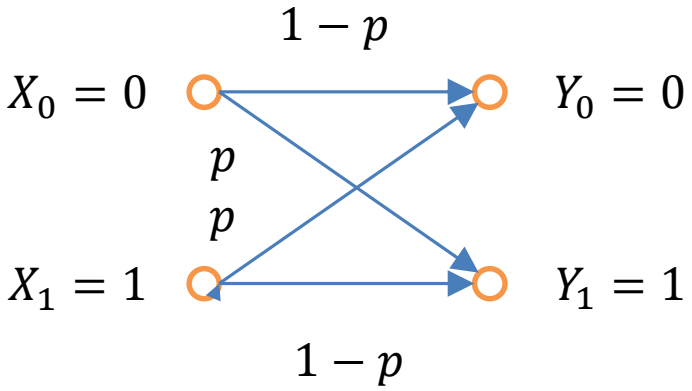




The input probabilities are described by

$p(X_0) = P(X_0 = 0)$

$p(X_1) = P(X_1 = 1)$



The transition probabilities are given by

$p(Y_0|X_0) = 1 - p$

$p(Y_1|X_1) = 1 - p$

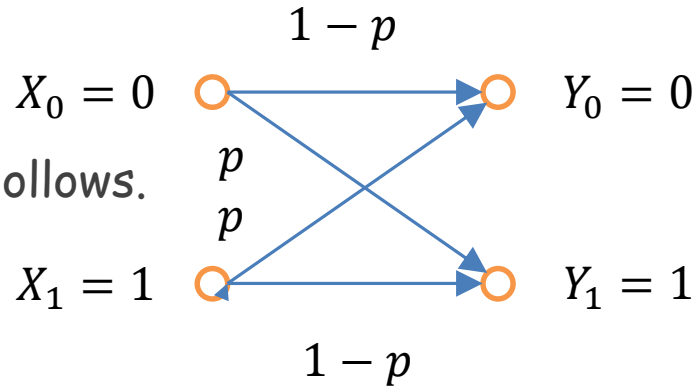
$p(Y_1|X_0) = p$

$p(Y_0|X_1) = p$

The output probabilities are described by

$$p(Y_0) = \sum_{j=0}^{J-1} p(Y_0|X_j)p(X_j) = p(Y_0|X_0)p(X_0) + p(Y_0|X_1)p(X_1) = (1-p)p(X_0) + pp(X_1)$$

$$p(Y_1) = \sum_{j=0}^{J-1} p(Y_1|X_j)p(X_j) = p(Y_1|X_0)p(X_0) + p(Y_1|X_1)p(X_1) = pp(X_0) + (1-p)p(X_1)$$



b) We compute the capacity of the BSC as follows.

We know that the entropy $H(x)$ is maximized when $p(X_0) = p(X_1) = \frac{1}{2}$, where x_0 and x_1 are 0 and 1, respectively.

The mutual information $I(x, y)$ is similarly maximized as described by

$$C = I(x, y) \text{ when } p(X_0) = p(X_1) = \frac{1}{2},$$

where

$p(Y_0|X_0) = 1 - p = p(Y_1|X_1)$

$p(Y_1|X_0) = p = p(Y_0|X_1)$

By substituting the transition probabilities in $I(x,y)$, we obtain

$$I(x,y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(Y_k, X_j) \log_2 \left[\frac{p(Y_k|X_j)}{p(Y_k)} \right]$$

With $J = K = 2$ and then setting $p(X_0) = p(X_1) = \frac{1}{2}$, we have

$$\begin{aligned}
C &= \max_{p(X_j)} \sum_{j=0}^{1} \sum_{k=0}^{1} p(Y_k, X_j) \log_2 \left[\frac{p(Y_k|X_j)}{p(Y_k)} \right] \\
&= p(Y_0, X_0) \log_2 \left[\frac{p(Y_0|X_0)}{p(Y_0)} \right] + p(Y_0, X_1) \log_2 \left[\frac{p(Y_0|X_1)}{p(Y_0)} \right] \\
&\quad + p(Y_1, X_0) \log_2 \left[\frac{p(Y_1|X_0)}{p(Y_1)} \right] + p(Y_1, X_1) \log_2 \left[\frac{p(Y_1|X_1)}{p(Y_1)} \right] \\
&= p(Y_0|X_0)\, p(X_0) \log_2 \left[\frac{p(Y_0|X_0)}{p(Y_0)} \right] + p(Y_0|X_1) p(X_1) \log_2 \left[\frac{p(Y_0|X_1)}{p(Y_0)} \right] \\
&\quad + p(Y_1|X_0) p(X_0) \log_2 \left[\frac{p(Y_1|X_0)}{p(Y_1)} \right] + p(Y_1|X_1) p(X_1) \log_2 \left[\frac{p(Y_1|X_1)}{p(Y_1)} \right] \\
&= \frac{1-p}{2} \log_2 [2(1-p)] + \frac{p}{2} \log_2 [2p] + \frac{p}{2} \log_2 [2p] + \frac{1-p}{2} \log_2 [2(1-p)] \\
&= 1 + p \log_2 p + (1-p) \log_2 (1-p)
\end{aligned}$$

D. Shannon's theorems

i) Source coding theorem:

Given a discrete memoryless source with entropy $H(\xi)$, the average codeword length for any lossless encoding scheme is bounded by

$$\bar{l} \geq H(\xi)$$

The entropy $H(\xi)$ is the fundamental limit of compression, i.e., the limit to the average number of bits per source symbol required to represent a discrete memoryless source.

In a source encoding scheme, when $l_{min} = H(\xi)$, the efficiency is given by

$$\eta = \frac{H(\xi)}{\bar{l}}$$

Shannon, Claude Elwood (July 1948). "A Mathematical Theory of Communication" (PDF). *Bell System Technical Journal*. **27** (3): 379–423.

ii) Channel capacity theorem:

The information capacity of a continuous channel bandlimited to B Hz perturbed by additive white Gaussian noise with power spectral density $\frac{N_0}{2}$ is given by

$$C = B \log_2 \left(1 + \frac{P}{N_0 B}\right), \quad \text{bits/ s}$$

where P is average transmit power.

This theorem shows that given P and B we can transmit information at a rate of C bits per second.

iii) Channel coding theorem:

For a discrete memoryless channel with capacity C that transmits information at a rate $R \leq C$ there exists a coding scheme in which the probability of error can be made arbitrarily small, that is,

$$P_e \leq \epsilon + 2^{-nR(I(x,y)-\delta-R)}$$

and

$$P_e \to \epsilon$$

when the block length $n \to \infty$. This is known as achievability.

In an ideal system, we transmit at a rate equal to $R_b = C$ bits /s.

If we take into account $P = E_b C$, where E_b is the transmit energy per bit, we have

$$\frac{C}{B} = \log_2\left(1 + \frac{P}{N_0 B}\right) = \log_2\left(1 + \frac{E_b C}{N_0 B}\right)$$

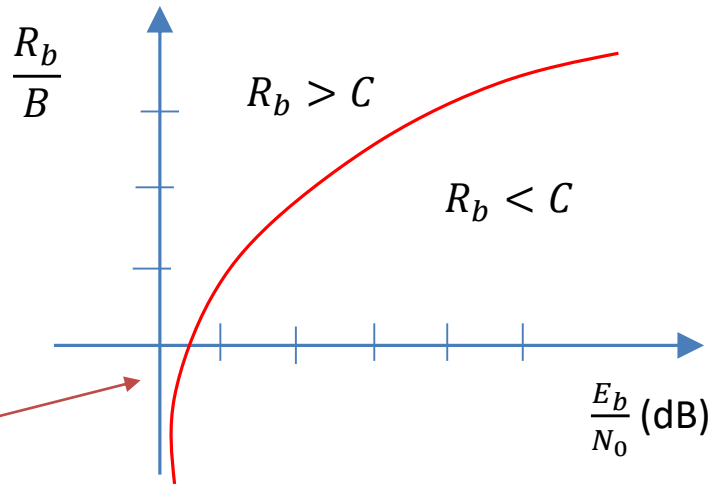
The spectral efficiency is the ratio of energy per bit by power spectral density is given by

$$\frac{E_b}{N_0} = \frac{2^{\frac{C}{B}} - 1}{\frac{C}{B}}$$

When $B \to \infty$ $\frac{E_b}{N_0}$ approaches

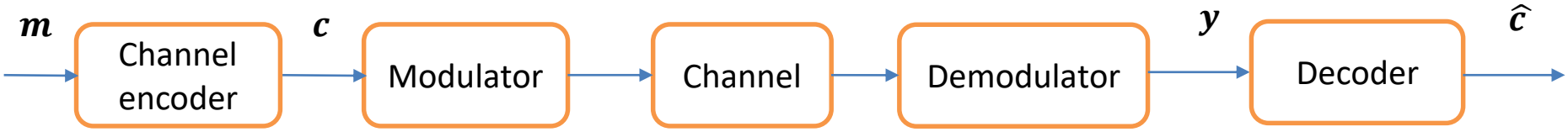
$$\left(\frac{E_b}{N_0}\right)_\infty = \lim_{B \to \infty}\left(\frac{E_b}{N_0}\right)$$

$$= \frac{1}{\log_2 e} = 0.693 \text{ or } -1.6 \text{ dB}$$



E. Decoding principles

- Let us consider a communication system illustrated by



- The task of the decoder is to observe $\mathbf{y} = \boldsymbol{Y}$ and produce an estimate $\hat{\boldsymbol{c}}$ of $\boldsymbol{c}$
- Because of the unique mapping of the channel encoder, we have

$$\hat{\boldsymbol{m}} = \boldsymbol{m} \text{ if } \hat{\boldsymbol{c}} = \boldsymbol{c}$$

- A general decoding rule then becomes

Compute $\hat{\boldsymbol{c}}$ for each $\boldsymbol{y}$ received

- The conditional error probability of the decoder is described by

$$P_{e|\boldsymbol{y}} = P(\hat{\boldsymbol{c}} \neq \boldsymbol{c}|\boldsymbol{y})$$

- The error probability can then be obtained by averaging over all received vectors processed at the receiver and is given by

$$P_e = \sum_{\boldsymbol{y}} P_{e|\boldsymbol{y}} P_{\boldsymbol{y}}$$

- Optimal decoding rules rely on the above quantities and their optimization to compute $\hat{\boldsymbol{c}}$.

MAP decoding

- In the Maximum a Posteriori (MAP) decoding strategy the decoder computes $\hat{\boldsymbol{c}}$ according to

$$\hat{\boldsymbol{c}} = \min_{\mathbf{c}} P_{e|\boldsymbol{y}} = P(\hat{\boldsymbol{c}} \neq \boldsymbol{c}|\boldsymbol{y})$$

- Since the minimization of $P_{e|\boldsymbol{Y}} = P(\hat{\boldsymbol{c}} \neq \boldsymbol{c}|\boldsymbol{y})$ is equivalent to the maximization of $P(\hat{\boldsymbol{c}} = \boldsymbol{c}|\boldsymbol{y})$, we can alternatively compute

$$\hat{\boldsymbol{c}} = \max_{\mathbf{C}} P(\hat{\boldsymbol{c}} = \boldsymbol{c}|\boldsymbol{y})$$

- Using Bayes's rule, we can then write the MAP decoding rule as

$$\hat{\boldsymbol{c}} = \max_{\boldsymbol{c}} P(\hat{\boldsymbol{c}} = \boldsymbol{c}|\boldsymbol{y})$$

$$= \max_{\boldsymbol{c}} \frac{p_{\boldsymbol{y}|\boldsymbol{c}}(\boldsymbol{Y}|\boldsymbol{C})P(\boldsymbol{c})}{p_{\boldsymbol{y}}(\boldsymbol{Y})}$$

ML decoding

- In the Maximum Likelihood (ML) decoding strategy the decoder assumes that all codewords are equiprobable and compute $\hat{\boldsymbol{c}}$ according to

$$\hat{\boldsymbol{c}} = \max_{\boldsymbol{c}} \frac{p_{\boldsymbol{y}|\boldsymbol{c}}(\boldsymbol{Y}|\boldsymbol{C})P(\boldsymbol{c})}{p_{\boldsymbol{y}}(\boldsymbol{Y})}$$
$$= \max_{\boldsymbol{c}} \; p_{\boldsymbol{y}|\boldsymbol{c}}(\boldsymbol{Y}|\boldsymbol{C})$$

- The ML decoder is often simpler than the MAP decoder in the computation and results in similar performance for many codes.

- For iterative strategies, extensive use of prior information changes the probabilities of the codewords, resulting in advantages for MAP decoding.

Example 3

Suppose that you would like to employ the ML decoder to decode a codeword with length n transmitted over a discrete memoryless channel (DMC).

a) Simplify the ML decoder by taking into account the DMS property.

b) Assume that the messages 00, 01, 10 and 11 have been encoded into the codewords 00100, 01110, 10001 and 11000, respectively, and that the received vector is $\boldsymbol{y} = [01111]$. Use the Hamming distance and the BSC with probabilities p and $1 - p$ to perform ML decoding.

Solution:

A DMC has the following property adapted to this problem:

$$p_{\boldsymbol{y}|\boldsymbol{c}}(Y_1, Y_2, \ldots, Y_n | C_1, C_2, \ldots, C_n) = p_{\boldsymbol{y}|\boldsymbol{c}}(\boldsymbol{Y}|\boldsymbol{C}) = \prod_{i=1}^{n} p_{y_j|c_i}(Y_j|C_i)$$

The ML decoder is then computed by

$$\hat{\boldsymbol{c}} = \max_{\boldsymbol{c}} \; p_{\boldsymbol{y}|\boldsymbol{c}}(\boldsymbol{Y}|\boldsymbol{C}) = \max_{\boldsymbol{c}} \; \prod_{i=1}^{n} p_{y_j|c_i}(Y_j|C_i)$$

Since $\log(x)$ is a monotonically increasing function of x, we have

$$\begin{aligned}
\hat{\boldsymbol{c}} &= \max_{\boldsymbol{c}} \; p_{\boldsymbol{y}|\boldsymbol{c}}(\boldsymbol{C}|\boldsymbol{C}) \\
&= \max_{\boldsymbol{c}} \; \prod_{i=1}^{n} p_{y_j|c_i}(Y_j|C_i) \\
&= \max_{\boldsymbol{c}} \; \log \prod_{i=1}^{n} p_{y_j|c_i}(Y_j|C_i) \\
&= \max_{\boldsymbol{c}} \; \sum_{i=1}^{n} \log \left(p_{y_j|c_i}(Y_j|C_i) \right)
\end{aligned}$$

The above form can greatly simplify computations in digital hardware.

b) The ML decoder computes

$$\hat{\boldsymbol{c}} = \max_{\mathbf{c}} \sum_{i=1}^{n} \log\left(p_{y_j|c_i}(Y_j|C_i)\right) = \max_{\mathbf{C}} \left(d \log p + (n-d)\log(1\ -p)\right)$$

where d is the Hamming distance, i.e., the number of positions that one vector differs from the other.

The Hamming distances of the codewords are

$$d([01111],[00100]) = 3$$
$$d([01111],[01110]) = 1$$
$$d([01111],[10001]) = 4$$
$$d([01111],[11000]) = 4$$

This means that the ML decoder will choose the codeword 01110 that corresponds to the message 01 and that the error occurred in the last bit.