



# Information Theory and Channel Coding

Prof. Rodrigo C. de Lamare  
CETUC, DEE, PUC-Rio, Brazil  
[delamare@puc-rio.br](mailto:delamare@puc-rio.br)

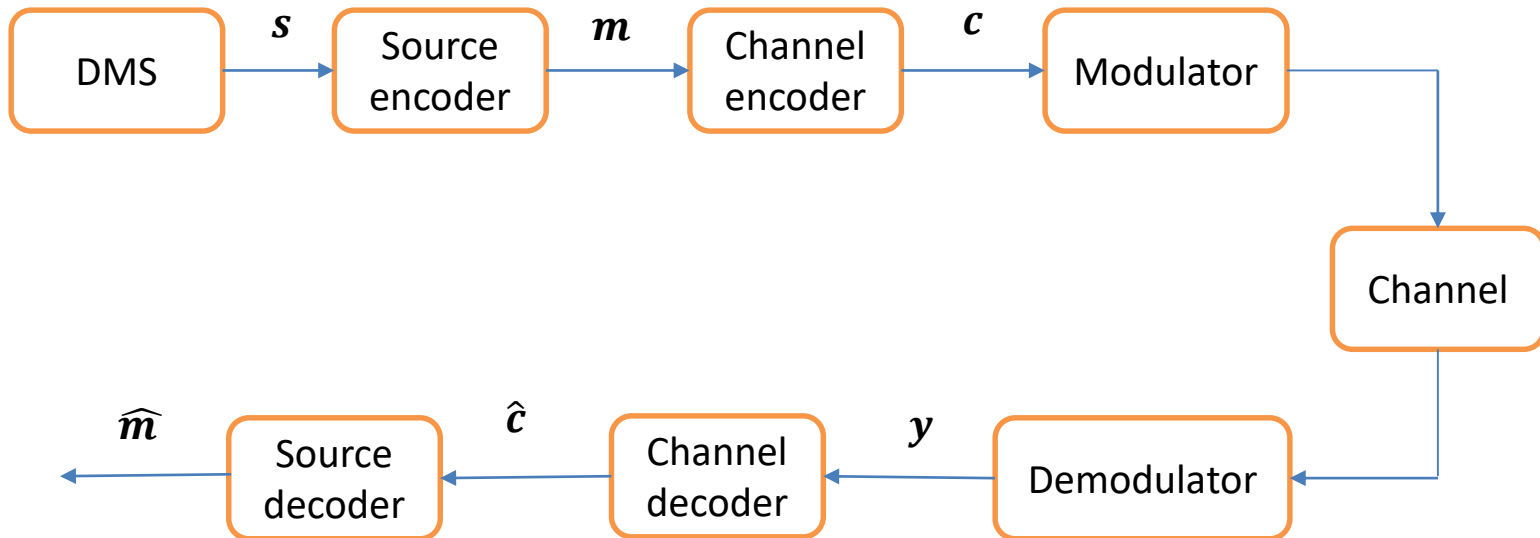


## IV. Channel coding

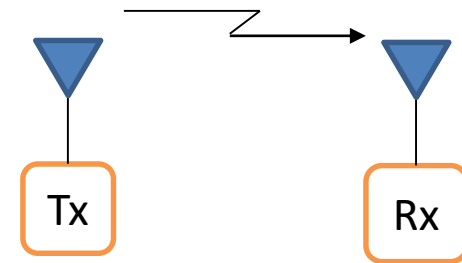
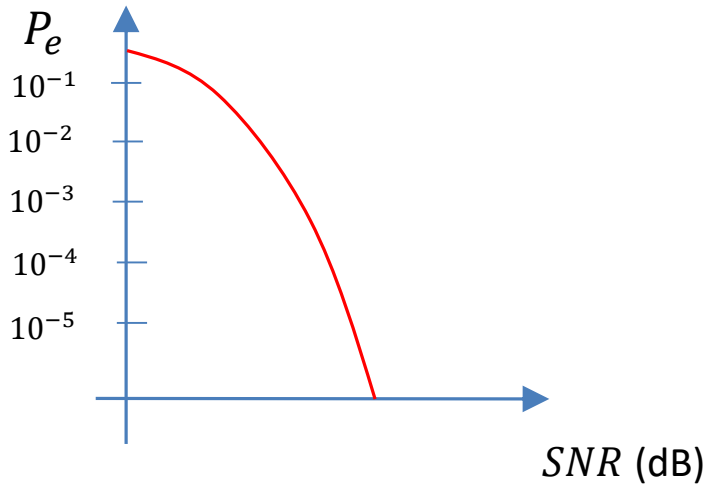
- In this chapter, we study the need for channel coding, derive the channel coding theorem and examine implications of the channel coding theorem.
- In particular, we examine the fundamental limit of how reliably information can be transmitted over a channel given some key parameters.
- We present a mathematical model of a digital communication system and how it can benefit from channel coding.
- We derive the channel coding theorem of Shannon using an approach based on the Markov inequality.
- We then examine implications of the channel coding theorem and how the probability of error of transmitted symbols can be made arbitrarily small.

# A. Digital communications model

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



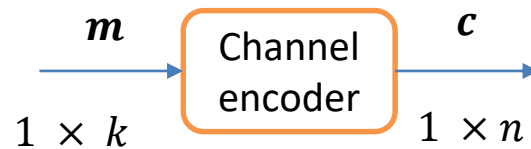
- Reliability is an important goal in digital communications that is often measured in terms of probability of symbol error  $P_e$ .



- In order to obtain reliable communication links and transmissions, we need to employ channel coding.



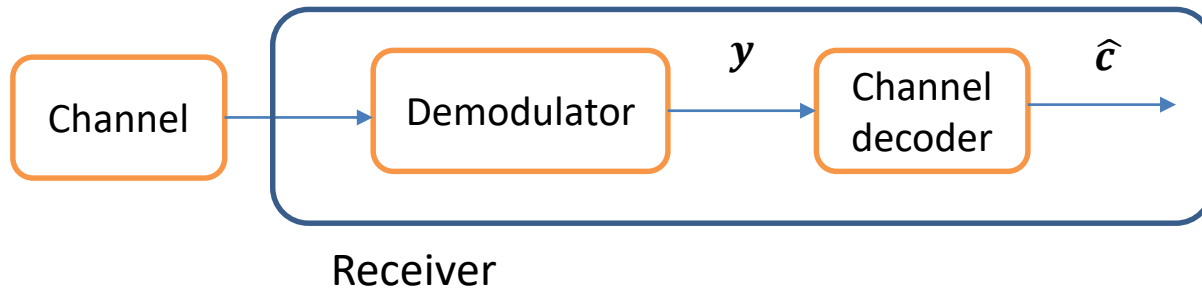
- Channel coding increases the resistance against channel errors in digital transmissions.
- The basic idea of channel coding is to introduce redundancy.



- A message  $m$  with  $k$  bits is mapped into a codeword  $c$  with  $n$  code bits, which is then transmitted.
- This redundancy translates into the code rate

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

- The receiver must deal with thermal noise often modelled as additive Gaussian noise and with the inverse mapping/decoding.



- Fundamental question:
  - Is there any channel coding scheme that allows transmission of messages with probability of error smaller than a small positive number  $\epsilon$  ?



## B. 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,

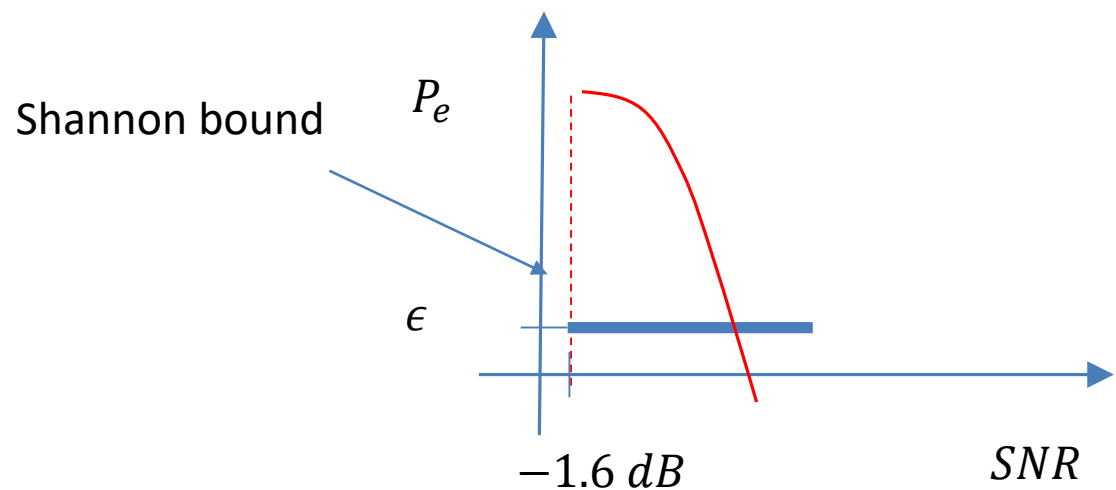
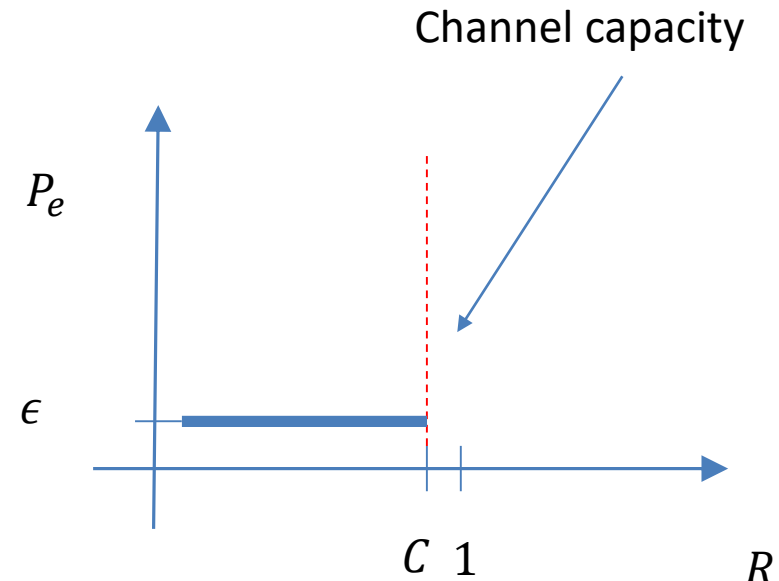
$$P_e \rightarrow \epsilon$$

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

Conversely, if  $R > C$  there is no coding scheme capable of delivering a  $P_e$  arbitrarily small. This is known as the converse theorem.

# Interpretation of the theorem

- For code rate  $R \leq C$  we can transmit information with arbitrarily low  $P_e$ .
- The theorem considers random codes but powerful channel codes could be designed close to capacity.
- Alternatively, a designer could use lower rates and approach the Shannon bound.







## C. Proof of the channel coding theorem

- Standard proof in textbooks:
  - Based on joint typicality  $\rightarrow$  generation of long sequences with certain properties.
  - Use of the asymptotic equipartition property (AEP): analog of the law of large numbers.
  - According to AEP, typical sets of sequences of random variables are generated with equally probable elements.
- Our approach: based on Markov's inequality

Yuval Lomnitz, Meir Feder, "A simpler derivation of the coding theorem",  
<https://arxiv.org/pdf/1205.1389.pdf>



- Consider the Markov inequality of a random variable  $x$  given by

$$P(x \geq t) \leq \frac{E[x]}{t}$$

Let us also consider the following assumptions:

- Channel codes are assumed random:  $x = [x_1 \dots x_n]$
- Entries of  $x$  are independent and identically distributed (i.i.d.) random variables, which yield the joint pdf

$$p_x(\mathbf{X}) = \prod_{i=1}^n p_{x_i}(X_i)$$



- System model:

$$\begin{aligned}\mathbf{y} &= \mathbf{x} + \mathbf{n} \\ &= [y_1 \dots y_n]\end{aligned}$$

- The noisy codeword  $\mathbf{y}$  at the output of the channel is random and its elements are i.i.d., i.e.,

$$p_{\mathbf{y}}(\mathbf{Y}) = \prod_{i=1}^n p_{y_i}(Y_i)$$

- We assume that maximum likelihood decoding is employed:

$$\hat{\mathbf{x}} = \arg \max p_{\mathbf{y}|\mathbf{x}}(\mathbf{Y}|\mathbf{X})$$



Probability of error over a set of codes:

- Let  $X_m, m = 1, 2, \dots, 2^{nR}$  be the independent codes of  $x, y$ .
- Consider the event  $E_m$  where  $X_m$  leads to the inequality on the a posteriori probability

$$P_{y|x}(Y|X_m) \geq P_{y|x}(Y|X)$$

- Therefore, we have

$$P(E_m|X, Y) = P(P_{y|x}(Y|X_m) \geq P_{y|x}(Y|X)|Y, X)$$



- Further developing the previous expression, we obtain

$$\begin{aligned} P(E_m | X, Y) &= P(P_{y|x}(Y | X_m) \geq P_{y'|x}(Y | X) | Y, X) \\ &\stackrel{\text{Markov}}{\leq} \frac{E[P_{y|x}(Y | X) | Y, X]}{P_{y|x}(Y | X)} \\ &= \sum_{X_m \in \mathcal{X}^n} \frac{P_{y|x}(Y | X) | Y, X}{P_{y|x}(Y | X)} \\ &= \sum_{X_m \in \mathcal{X}^n} \frac{P_{y|x}(Y | X) P_x(X_m)}{P_{y|x}(Y | X)} \\ &= \frac{P_y(Y)}{P_{y|x}(Y | X)} \end{aligned}$$



- Using the union bound, the probability of error conditioned on  $x, y$  is bounded by

$$\begin{aligned} P_{e|x,y} &\leq P \left\{ \bigcup_{i=1}^{2^{nR}} E_m | \mathbf{X}, \mathbf{Y} \right\} \\ &\leq 2^{nR} P(E_m | \mathbf{X}, \mathbf{Y}) \\ &\leq 2^{nR} \frac{P_y(\mathbf{Y})}{P_{y|x}(\mathbf{Y} | \mathbf{X})} \end{aligned}$$



- Then, we analyse the behaviour of  $P_{e|x,y}$  for the DMC channel.
- Using the law of large numbers, we have

$$\begin{aligned}\frac{1}{n} \log_2 P(E_m | \mathbf{X}, \mathbf{Y}) &= \frac{1}{n} \log_2 \frac{P_y(\mathbf{Y})}{P_{y|x}(\mathbf{Y} | \mathbf{X})} \\ &= \frac{1}{n} \sum_{i=1}^n \log_2 \frac{P_{y_i}(Y_i)}{P_{y_i|x_i}(Y_i | X_i)} \\ &\xrightarrow[n \rightarrow \infty]{\text{in prob.}} E \left[ \log_2 \frac{P_y(Y)}{P_{y|x}(Y | X)} \right] \triangleq -I(x, y), \text{ bits}\end{aligned}$$

where  $x, y$  are two random variables that are distributed according to  $p_{y|x}(Y | X)p_x(X)$ .



- From the law of large numbers, it follows that for any  $\epsilon, \delta > 0$  there is a sufficiently large  $n$  such that with probability  $(1 - \epsilon)$  we have

$$\frac{1}{n} \log_2 \frac{P_y(Y)}{P_{y|x}(Y|X)} \leq \delta - I(x, y)$$

- When the above expression is satisfied then we have

$$\begin{aligned} P_{e|x,y} &\leq 2^{nR} 2^{n(\delta - I(x,y))} \\ &= 2^{-nR(I(x,y) - \delta - R)} \end{aligned}$$





- The expression  $P_{e|x,y}$  can be averaged to obtain a bound on the probability of symbol error  $P_e$
- The probability of symbol error  $P_e$  is limited by the union bound which is given by

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

which can be made arbitrarily small if  $R < I(x,y)$  or equivalently  $R \leq C$  for small  $\epsilon$  and  $\delta$

- For  $n \rightarrow \infty$  with  $R$  fixed, we have

$$P_e \leq \epsilon$$



## D. Implications of the channel coding theorem

- Let us consider a repetition code used for digital transmission over a BSC with crossover or transition probability  $p = 10^{-2}$ .
- For such a BSC with probability  $p = 10^{-2}$  the capacity is given by

$$\begin{aligned} C &= 1 - p \log_2 p - (1 - p) \log_2(1 - p) \\ &= 0.9192 \end{aligned}$$

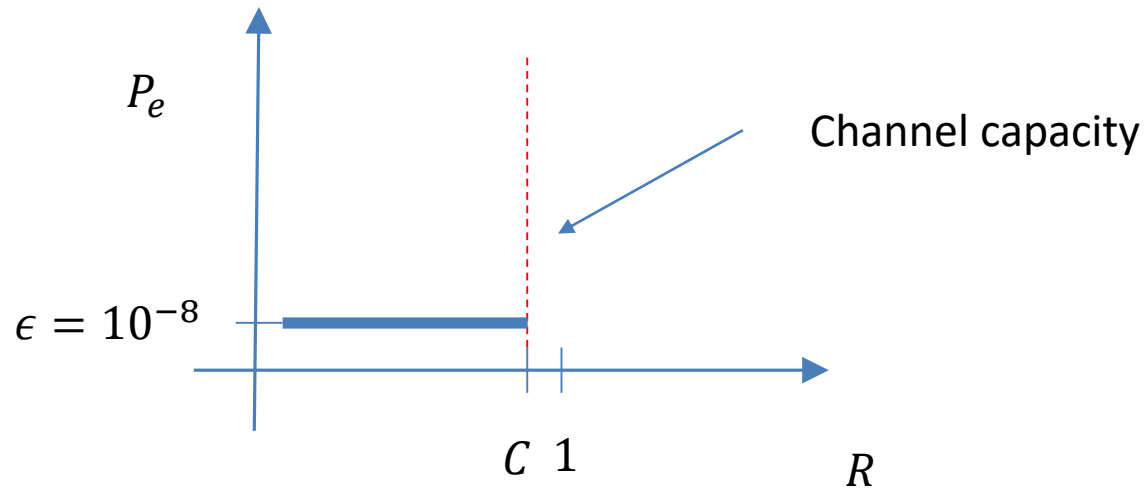
- Using the channel coding theorem, we know that for  $\epsilon > 0$  and  $R < 0.9192$  there exists a channel code with  $n$  sufficiently large, code rate  $R$  and a decoding algorithm that results in

$$P_e \leq \epsilon$$



# Illustration

- For  $\epsilon = 10^{-8}$ , we have





- Consider a repetition code that works as follows:
  - Each bit of the message  $m$  is repeated multiple times.
  - For each bit (0 or 1) we repeat it  $n$  times, where  $n = 2m + 1$  and  $n$  is an odd integer.
- The decoding of such code employs the majority logic decoding principle that works as follows:
  - If the number of 1s  $\geq$  the number of 0s  $\rightarrow$  the decoder decides for 1
  - If the number of 1s  $<$  the number of 0s  $\rightarrow$  the decoder decides for 0



- The probability of symbol error is given by

$$P_e = \sum_{i=m+1}^n \binom{n}{i} p^i (1-p)^{n-i},$$

where  $p$  is the transition probability of the BSC channel.

- The probability of error  $P_e$  is often used as a figure of merit and measured against the SNR or another useful quantity.

# Performance

- The performance of the repetition code can be illustrated by measuring the probability of error  $P_e$  against the code rate  $R$ .

Code rate (R)	Probability of symbol error $P_e$
1	$10^{-2}$
$\frac{1}{3}$	$3 \times 10^{-4}$
$\frac{1}{5}$	$10^{-6}$
$\frac{1}{7}$	$4 \times 10^{-7}$
$\frac{1}{9}$	$10^{-8}$
$\frac{1}{11}$	$5 \times 10^{-10}$

