01

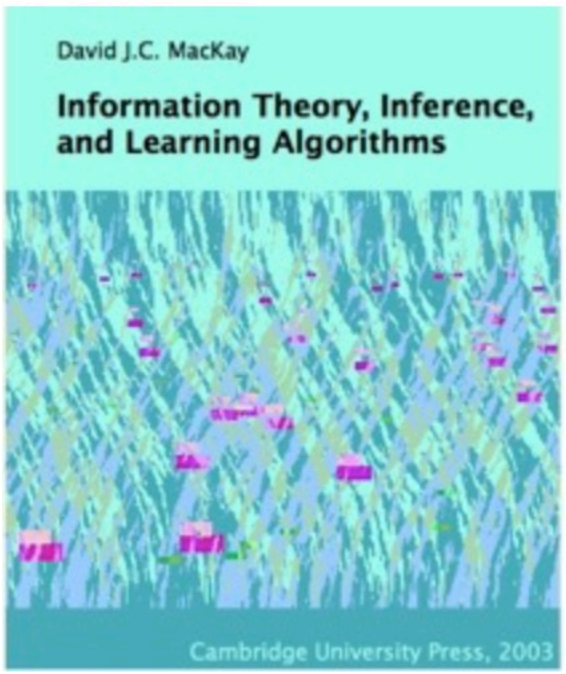
Course Overview

Notice

- **Author**
  - **João Moura Pires (jmp@fct.unl.pt)**

Bibliography

- **Many examples are extracted and adapted from:**



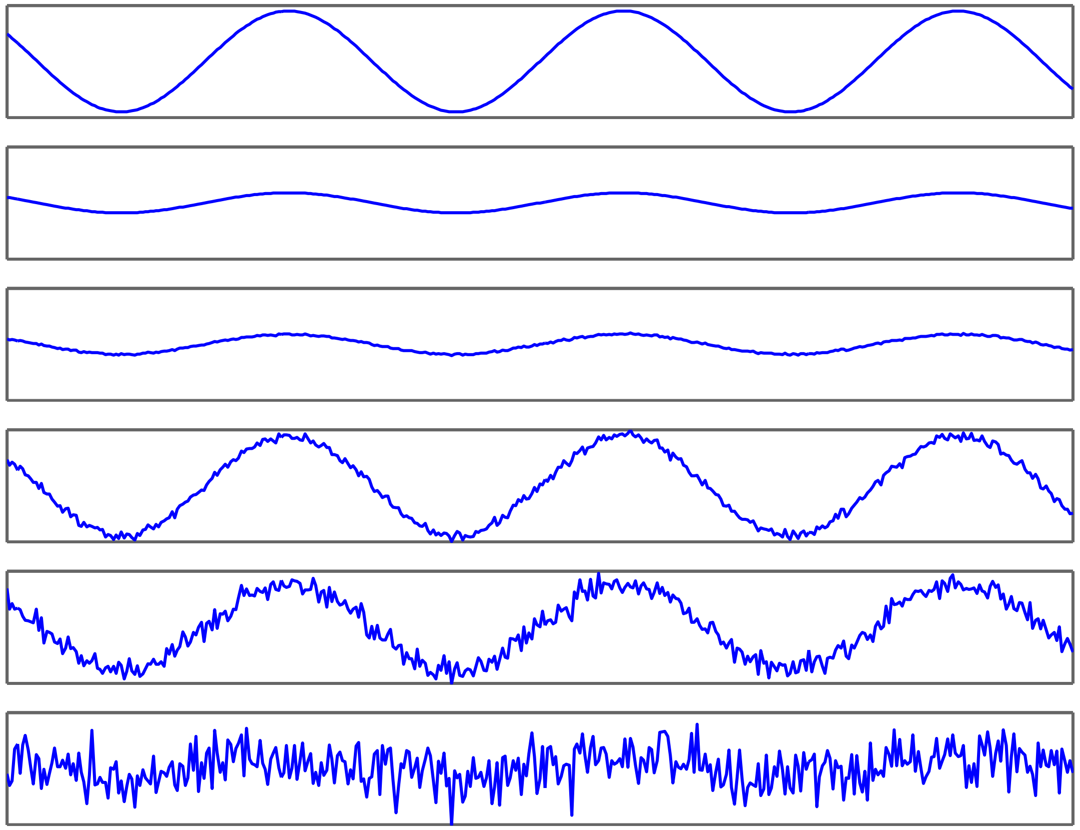
Information Theory, Inference, and Learning Algorithms
David J.C. MacKay
2005, Version 7.2

- **And some slides were based on Iain Murray course**
  - **http://www.inf.ed.ac.uk/teaching/courses/it/2014/**

Table of Contents

Context and Motivation

Analog versus Digital Communication



Signal

Attenuate

Add noise

Boost

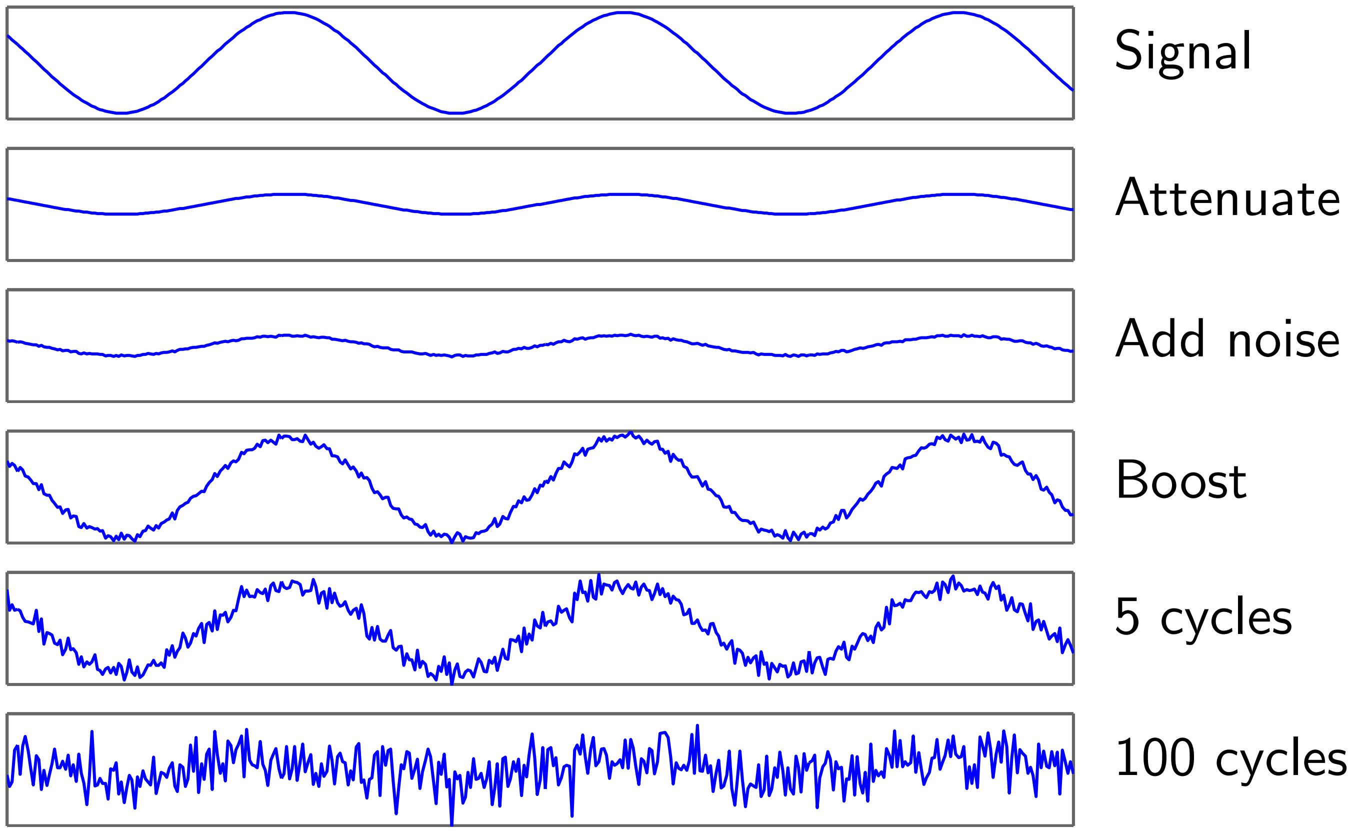
5 cycles

100 cycles

Econding by amplitude modulation

Analog

Analog versus Digital Communication

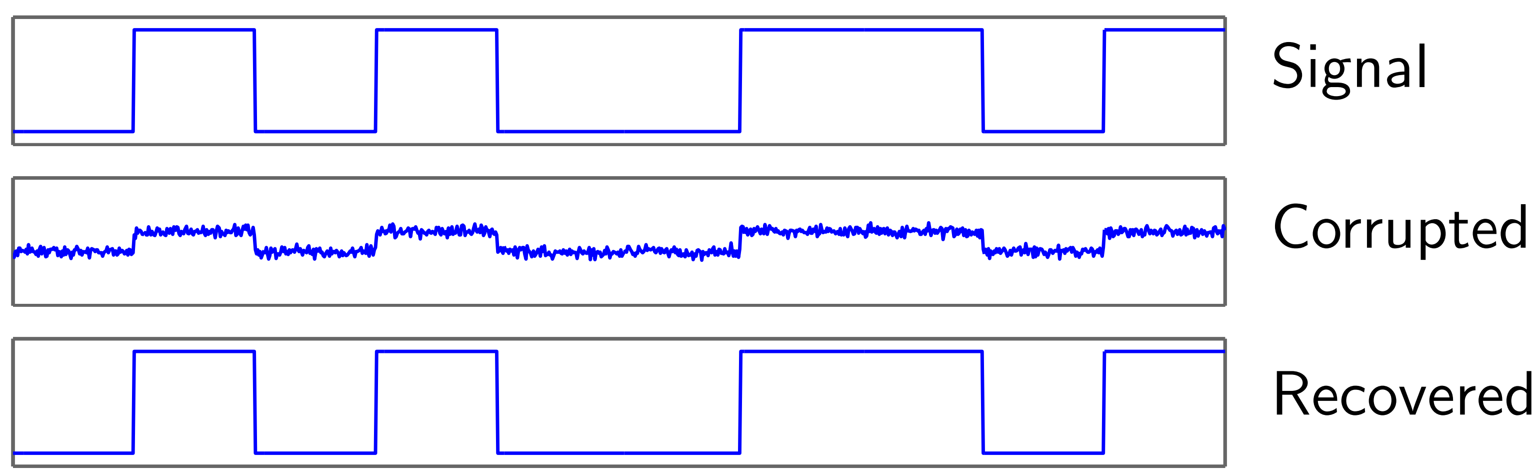


Econding by amplitude modulation

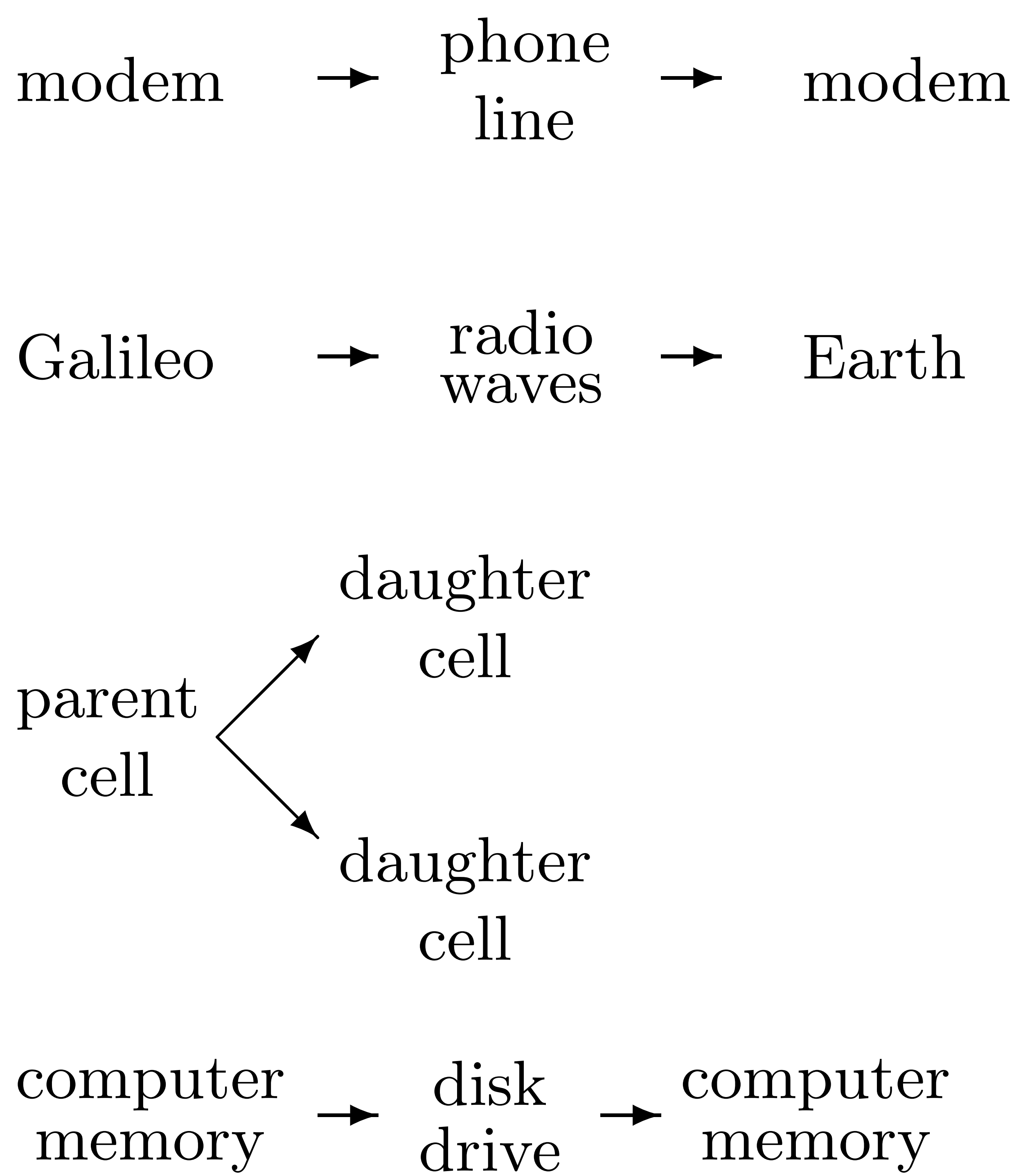
Analog

Digital

digital encoding

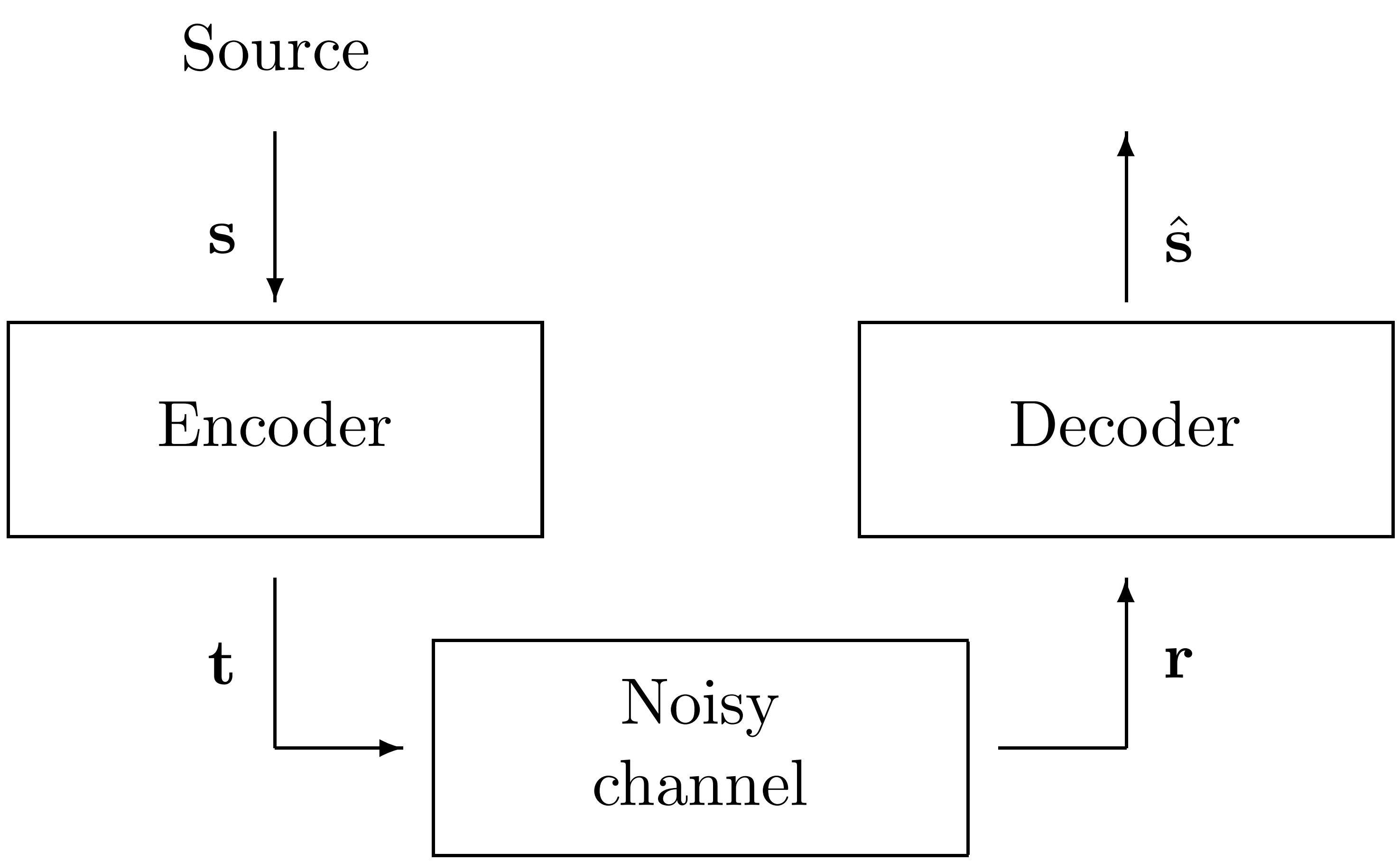


Examples of noisy communication channels



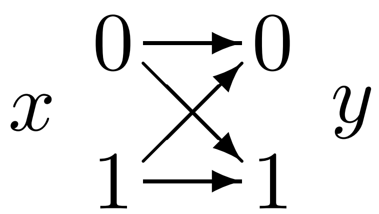
General digital communication system

- The role of the **encoder is to introduce systematically redundancy** to make possible to the decoder (which know the encoding process) to discover the sent message even if some bits were flipped by the noise channel



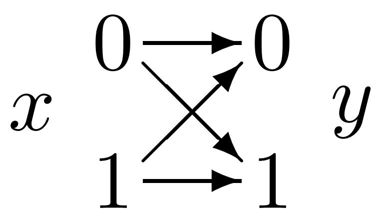
FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Symmetric Binary Channel

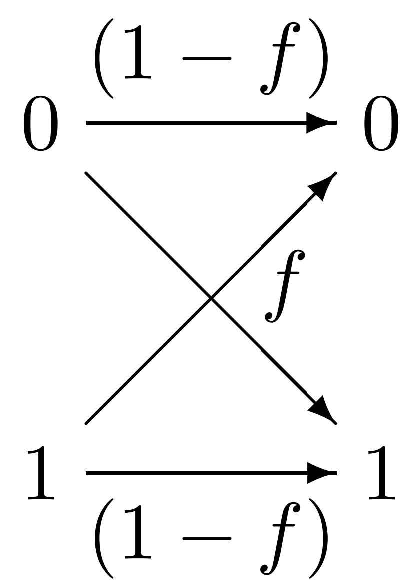


$$
\begin{aligned}
P(y=0 \mid x=0) &= 1-f; & P(y=0 \mid x=1) &= f; \\
P(y=1 \mid x=0) &= f; & P(y=1 \mid x=1) &= 1-f.
\end{aligned}
$$

Symmetric Binary Channel

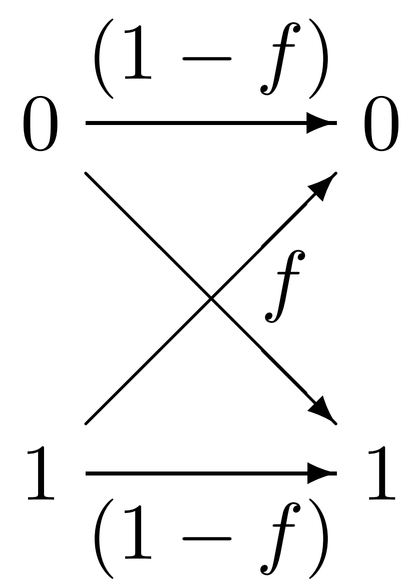


$$
\begin{aligned}
P(y=0\,|\,x=0) &= 1-f; & P(y=0\,|\,x=1) &= f;\\
P(y=1\,|\,x=0) &= f; & P(y=1\,|\,x=1) &= 1-f.
\end{aligned}
$$



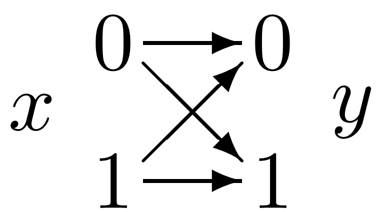
Symmetric Binary Channel

$$
x \;\begin{array}{ccc} 0 & \longrightarrow & 0 \\ & \times & \\ 1 & \longrightarrow & 1 \end{array}\; y
\qquad
\begin{array}{lcllcl}
P(y=0\,|\,x=0) & = & 1-f; & P(y=0\,|\,x=1) & = & f; \\
P(y=1\,|\,x=0) & = & f; & P(y=1\,|\,x=1) & = & 1-f.
\end{array}
$$

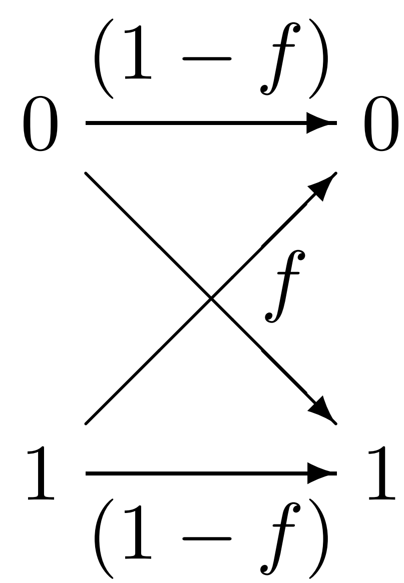


$$f = 0.1$$

Symmetric Binary Channel



$$
\begin{aligned}
P(y=0 \mid x=0) &= 1-f; & P(y=0 \mid x=1) &= f; \\
P(y=1 \mid x=0) &= f; & P(y=1 \mid x=1) &= 1-f.
\end{aligned}
$$



$f = 0.1$

Perfect communication over an noisy communication channel?

- A useful disk drive would flip no bits at all in its entire lifetime. If we expect to read and write a gigabyte per day for ten years, we **require a bit error probability of the order of 10^{-15}, or smaller**.

Perfect communication over an noisy communication channel?

- A useful disk drive would flip no bits at all in its entire lifetime. If we expect to read and write a gigabyte per day for ten years, we **require a bit error probability of the order of 10^{-15}, or smaller**.

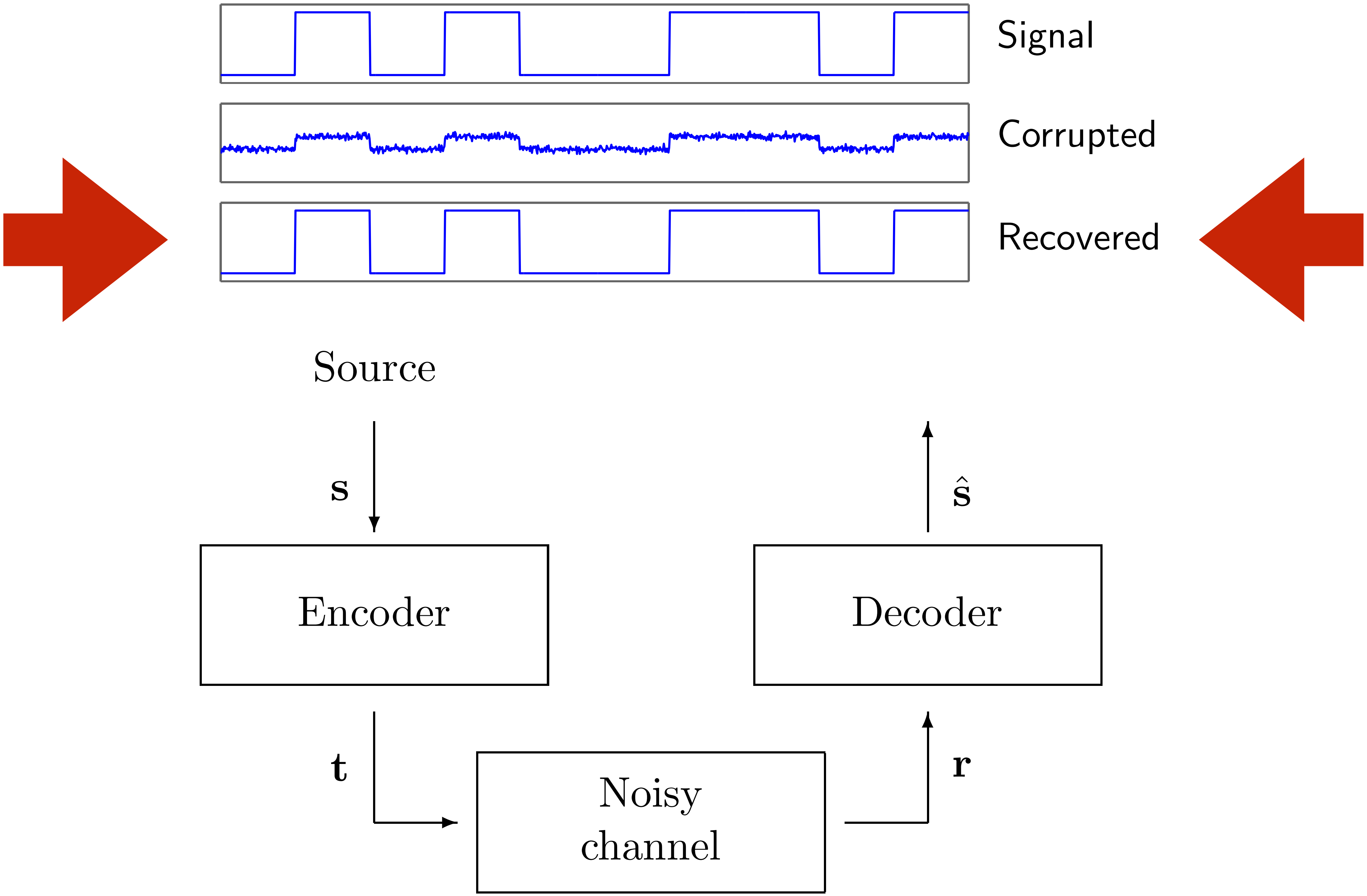
- **Physical solutions**
  - Incremental improvements
  - Increasing costs

Perfect communication over an noisy communication channel?

- A useful disk drive would flip no bits at all in its entire lifetime. If we expect to read and write a gigabyte per day for ten years, we **require a bit error probability of the order of 10^{-15}, or smaller**.

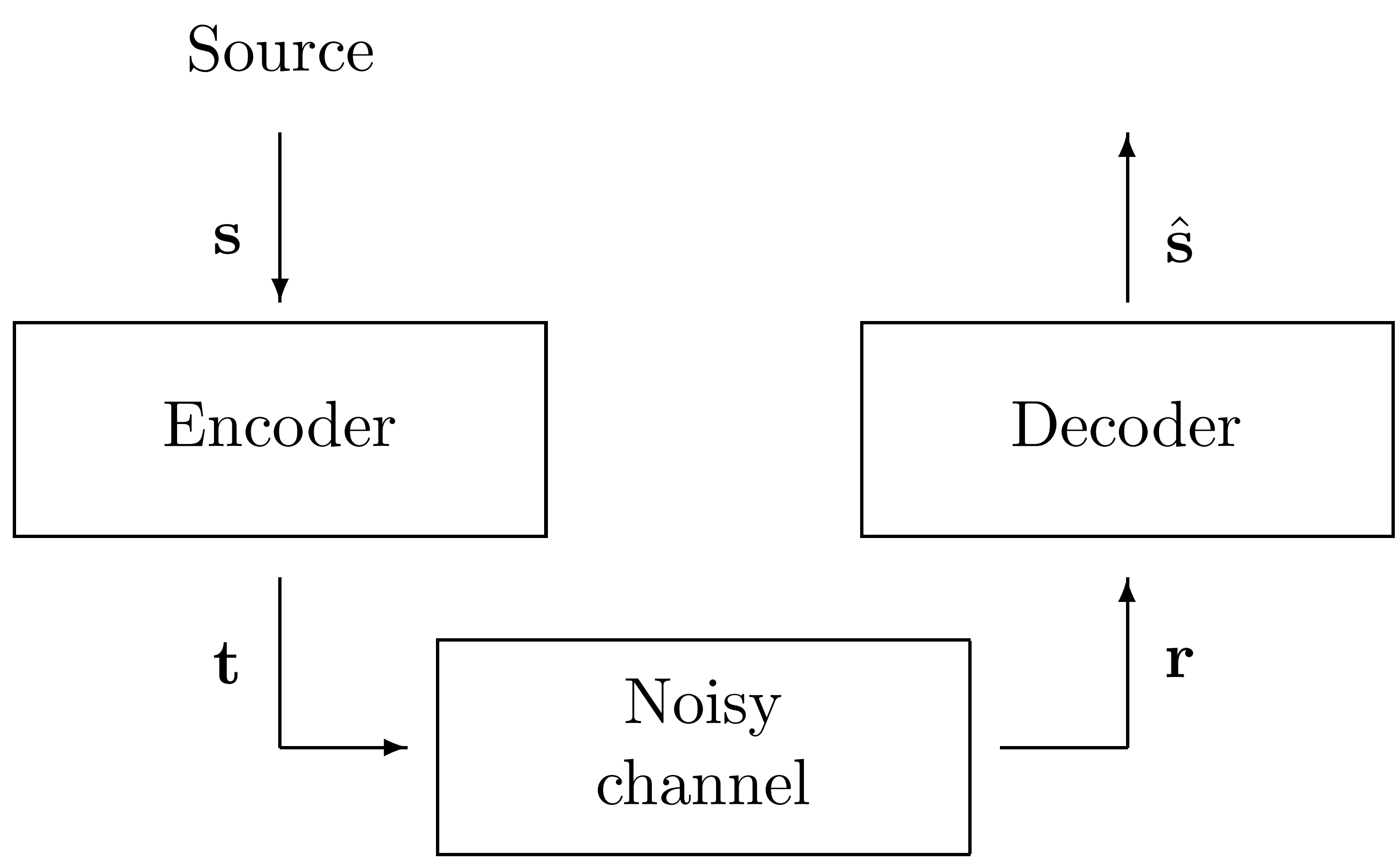
- **Physical solutions**
  - Incremental improvements
  - Increasing costs
- **System Solutions**
  - Can turn noisy channels into reliable communication channels (with the only cost being a computational requirement at the encoder and decoder)

General digital communication system



Information Theory and Coding Theory

- The **role of the encoder** is to introduce **systematically redundancy** to make possible to the decoder (which know the encoding process) to discover the sent message even if some bits were flipped by the noise channel.



FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Information Theory and Coding Theory

- The **role of the encoder** is to introduce **systematically redundancy** to make possible to the decoder (which know the encoding process) to discover the sent message even if some bits were flipped by the noise channel.

- **Information Theory** is concerned with the theoretical limitations and potentials of such systems. '**What is the best error-correcting performance we could achieve?**'

- **Coding Theory** is concerned with the creation of **practical encoding and decoding systems**

Course Organization and Overview

Syllabus

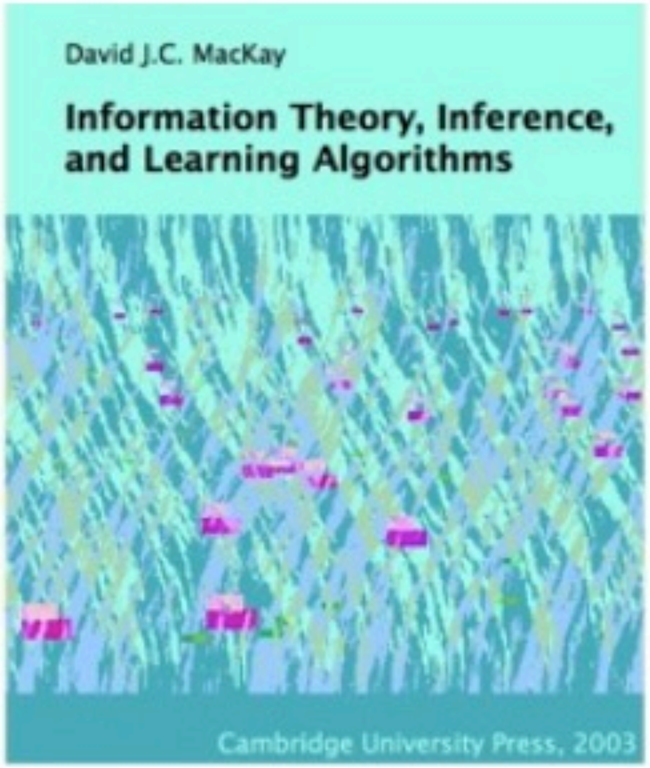
1 - Introduction
- Main problems addressed by Information Theory and its creation.
- Relations with other bodies of knowledge.
- Information Theory Overview.

2 - Foundational Concepts
- Entropy for discrete variables.
- Channel capacity for noiseless channels.
- Source Coding Shannon Theorem
- Data Compression
- Kolmogorov Complexity
- Joint distributions
- Mutual Information
- Condicional Entropy
- Noise and cross comunicativos
- Error correcting codes
- Channel capacity for noisy-channels.
- The Noisy-Channel Coding Theorem
- Extension for the continuum domain.
- Information Theory on other knowledge domains

3 - Probability and Inference

4 - Neural Networks

Bibliography



Information Theory, Inference, and Learning Algorithms
David J.C. MacKay
2005, Version 7.2

http://www.inference.org.uk/mackay/

http://www.inference.org.uk/mackay/itila/

Weekly routine

- **Lectures - 1 x 2 h**

- **The lab sessions - 1 x 2 h**
  - **Training problem solving**
  - **Project developing**

- **The recommended readings**

- **The recommended actions**

- **Meetings for student support if required**

Evaluation rules

- **The students performance evaluation includes two individual written tests and a small project.**
  - **Final Grade = 35% Test1 + 35% Test2 + 30% Project**
- **To successful conclude the following constraints are applied:**
  - Project >= 10;
  - Test1 >= 8; Test2 >= 8;
  - Average of Test1 and Test2 >= 10;
  - Final Grade >= 10.
- **The students that get a Project >= 10 and do satisfy the constraints on the tests, may have an exam which grade will replace the tests in the final grade calculation**

Web Site: http://ti.ssdi.di.fct.unl.pt

TI 20/21
Information Theory

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

News / Home /

- Home
- News
- Information
- Resources
- Summaries
- Training

Information Theory (IT) is a 6 ECTS unit and belongs to the free block B (year 4 of FCT profile). This unit is offered to students from 4th and 5th years of Mestrado Integrado em Engenharia Informática (MIEI) and to Pós-Graduação em Criptografia e Informação (PGCI).
This unit presents the foundations of Information Theory and shows its applications to Computer Science, Statistical Inference and Machine Learning.

This course is provided by Departamento de Informática (DI) da Faculdade de Ciências e Tecnologia (FCT) da Universidade Nova de Lisboa (UNL).

See the News! (last update: Setember14, 2020) - IMPORTANT INFORMATION

If it is your **first time visit** on this site, I suggest you to take a look on:

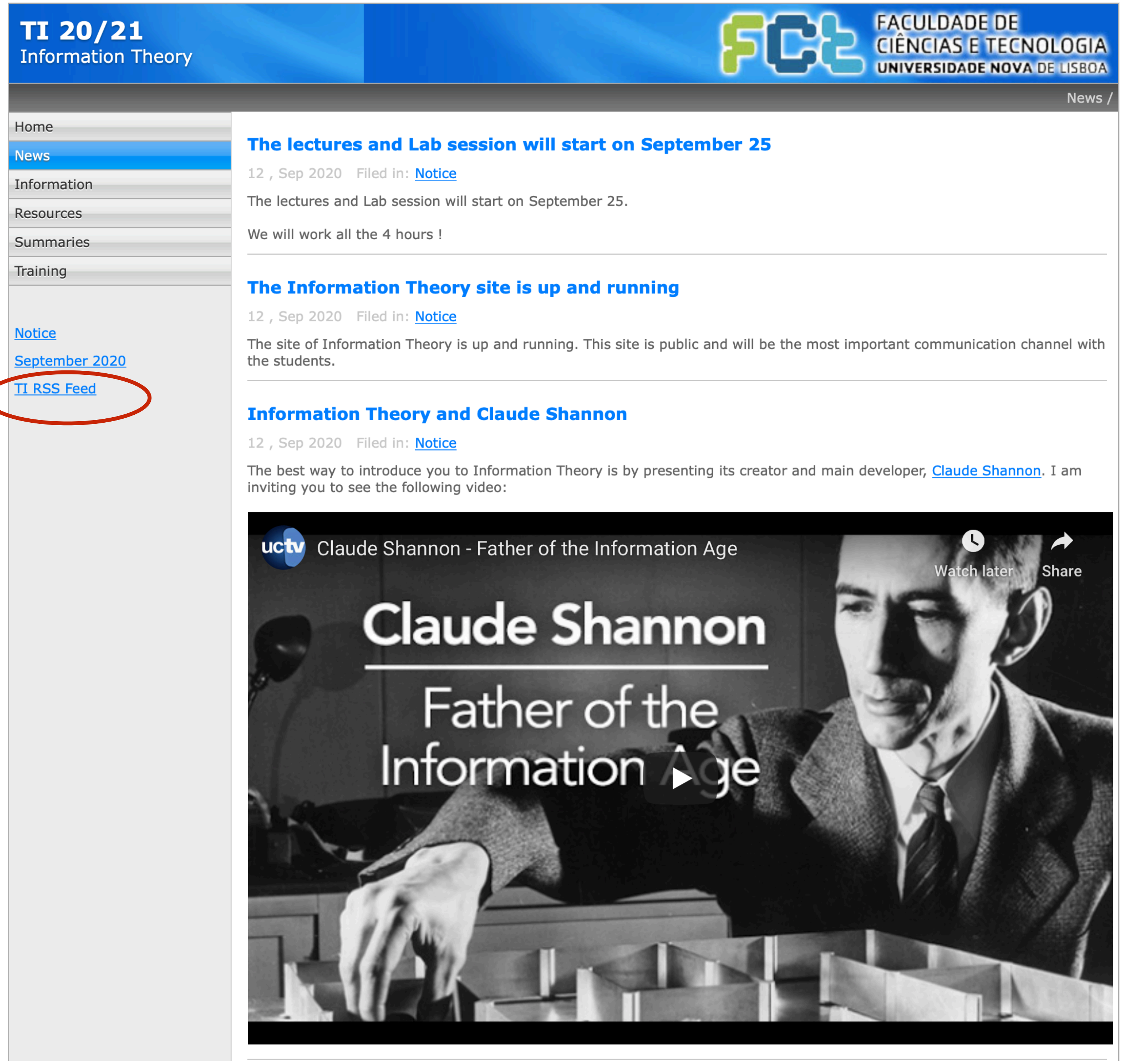
- News
- Information
- Information / Evaluation Rules

I hope we will have a nice semester !

NOVA LINCS STAResearch.NET João Moura Pires

TI - Teoria de Informação

Web Site: News



Web Site: Information

TI 20/21
Information Theory

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

News / Information /

- Home
- News
- Information
  - Bibliography
  - Sylabus
  - Evaluation Rules
  - Schedule
- Resources
- Summaries
- Training

Information Theory (IT) is a 6 ECTS unit and belongs to the free block B (year 4 of FCT profile). This unit is offered to students from 4th and 5th years of Mestrado Integrado em Engenharia Informática (MIEI) and to Pós-Graduação em Criptografia e Informação (PGCI).
This unit presents the foundations of Information Theory and shows its applications to Computer Science, Statistical Inference and Machine Learning.

This course is provided by Departamento de Informática (DI) da Faculdade de Ciências e Tecnologia (FCT) da Universidade Nova de Lisboa (UNL).

Objectives:
Knowledge:

- The main Information Theory concepts, including Entropy, Information, Condicional Entropy, Channel Capacity.
- Identify these concepts in different contexts of communication systems, Storage, Data Processing and Inference.
- The main Information Theory's Theorem, the source coding theorem (with and without noise), its role, impact and application areas.
- The core aspects of Compression and error correcting codes.
- General principles and approaches to cryptography.
- Information Theory's application examples to different knowledge areas.

Application:

- Apply Entropy and Information concepts to Computer Science and Machine Learning.
- Develop the main components of data compression algorithms or error correcting codes.

Soft-Skills:

- - Improve you ability to read and understand papers with a significant formal component. Be able to provide examples that illustrate the concepts and techniques discussed.
- - Improve your team-work skills.
- - Improve your communication skills (oral and written) on formal subjects.
- - Propose and develop simple but formal notation.

Prerequisites:
Basic knowledge of Probability and Statistics.
Basic knowledge and Practice of computer programming.

Teacher
Prof. João Moura Pires (jmp@fct.unl.pt) at office P3/2 and Tel: 10746.

Schedule (see at Schedule that will be updated)
Lectures:

- **English** (if required) spoken lectures

Office hours:

- **TBD**
- Other time slots if you get previously an appointment

NOVA LINCS STAResearch.NET João Moura Pires

- Home
- News
- Information
  - Bibliography
  - Sylabus
  - Evaluation Rules
  - Schedule
- Resources
- Summaries
- Training

Information Theory (IT) is a 6 ECTS unit and belongs to the free block B (year 4 of FCT profile). This unit is offered to students from 4th and 5th years of Mestrado Integrado em Engenharia Informática (MIEI) and to Pós-Graduação em Criptografia e Informação (PGCI).
This unit presents the foundations of Information Theory and shows its applications to Computer Science, Statistical Inference and Machine Learning.

This course is provided by Departamento de Informática (DI) da Faculdade de Ciências e Tecnologia (FCT) da Universidade Nova de Lisboa (UNL).

Objectives:
Knowledge:

- The main Information Theory concepts, including Entropy, Information, Condicional Entropy, Channel Capacity.
- Identify these concepts in different contexts of communication systems, Storage, Data Processing and Inference.
- The main Information Theory's Theorem, the source coding theorem (with and without noise), its role, impact and application areas.
- The core aspects of Compression and error correcting codes.
- General principles and approaches to cryptography.
- Information Theory's application examples to different knowledge areas.

Application:

- Apply Entropy and Information concepts to Computer Science and Machine Learning.
- Develop the main components of data compression algorithms or error correcting codes.

Soft-Skills:

- - Improve you ability to read and understand papers with a significant formal component. Be able to provide examples that illustrate the concepts and techniques discussed.
- - Improve your team-work skills.
- - Improve your communication skills (oral and written) on formal subjects.
- - Propose and develop simple but formal notation.

Prerequisites:
Basic knowledge of Probability and Statistics.
Basic knowledge and Practice of computer programming.

Teacher
Prof. João Moura Pires (jmp@fct.unl.pt) at office P3/2 and Tel: 10746.

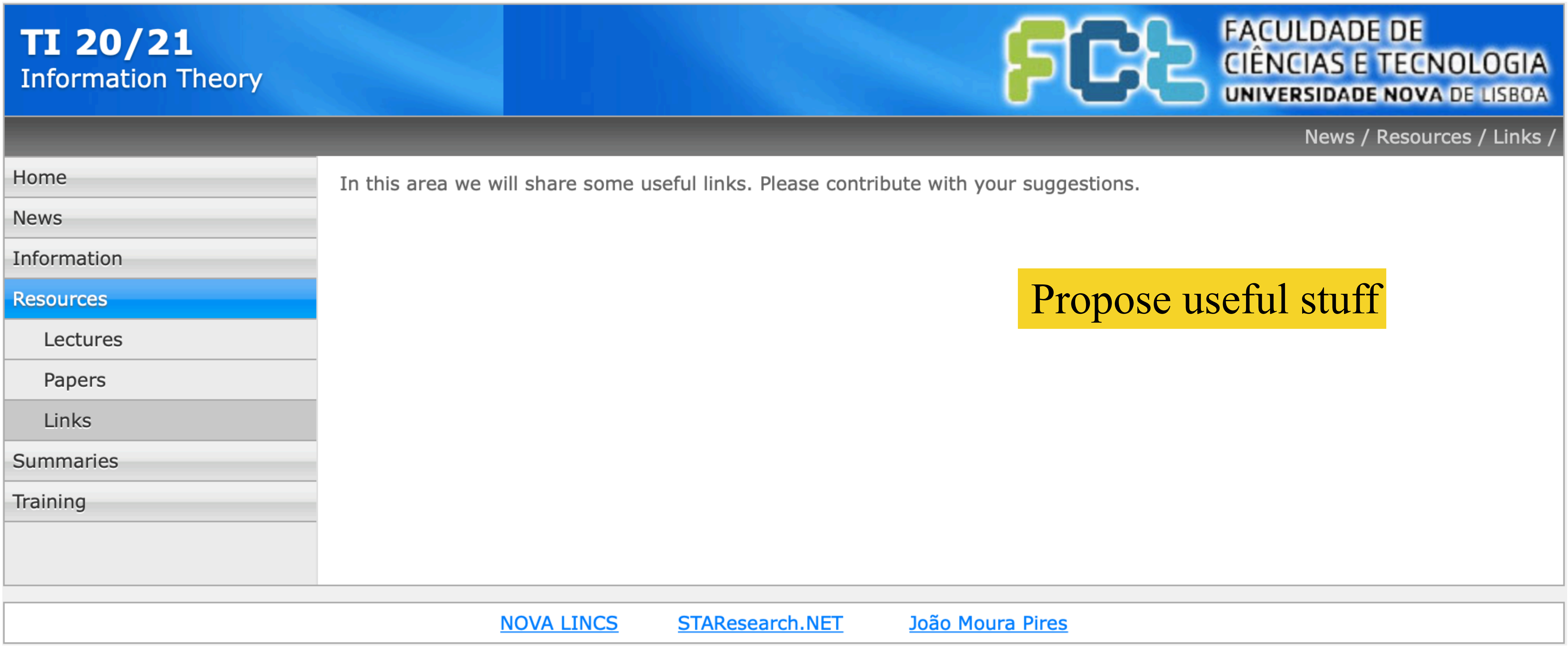
Schedule (see at Schedule that will be updated)
Lectures:

- **English** (if required) spoken lectures

Office hours:

- **TBD**
- Other time slots if you get previously an appointment

Web Site: Resources



Web Site: Summaries

TI 20/21
Information Theory

FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

News / Summaries /

- Home
- News
- Information
- Resources
- Summaries
- Training

Lectures
Labs

Summaries - RSS Feed

[P01]: Repeating and Block codes

25 Sep 2020, 02:10 PM Filed in: Labs

The Binomial distribution
Approximating x! and
Error Probability of RN
Decoding some message with H(7,4)
Probability of block error PB in H(7,4)
Noise vectors that give the all-zero syndrome in H(7,4)
Design an error-correcting code (*)
H(14, 8) code can correct any two errors?

[T01]: Course overview

25 Sep 2020, 10:10 AM Filed in: Lectures

What we mean by "Information Theory"? What for? Why Information Theory is important? Why is important to study Information Theory?

Course Organization and Overview: Syllabus; Bibliography; Evaluation rules; important dates, etc..

Recommended Readings: (i) Information Theory, Inference, and Learning Algorithms from David MacKay, 2015, pages 1 - 16; **(ii)** The introduction of "A Mathematical Theory of Communication, Claude Shannon, 1948", pages 1-2.

Recommended Activities: (i) see the following video: "Claude Shannon - Father of the Information Age". **(ii)** Visit the various sections of this site.

To Know:

- Why is important the ideia of Digital communications?
- What was the main question that Shannon try to address with Information Theory?
- What is on of the most important result of Shannon's work?
- Understand the General Digital Communication system; What is the role of the Encoder (and the corresponding decoder).
- Binary Symmetric Channel; what is f?
- What is Pb, PB and R (rate)?
- Understand the Repetition codes, (RN).
- Understand the Block codes, the Linear Block codes, the Hamming code H(7, 4)

Home
News
Information
Resources
Summaries
Training

Lectures
Labs

Summaries - RSS Feed

[P01]: Repeating and Block codes

25 Sep 2020, 02:10 PM Filed in: Labs

The Binomial distribution
Approximating x! and
Error Probability of RN
Decoding some message with H(7,4)
Probability of block error PB in H(7,4)
Noise vectors that give the all-zero syndrome in H(7,4)
Design an error-correcting code (*)
H(14, 8) code can correct any two errors?

[T01]: Course overview

25 Sep 2020, 10:10 AM Filed in: Lectures

What we mean by "Information Theory"? What for? Why Information Theory is important? Why is important to study Information Theory?

Course Organization and Overview: Syllabus; Bibliography; Evaluation rules; important dates, etc..

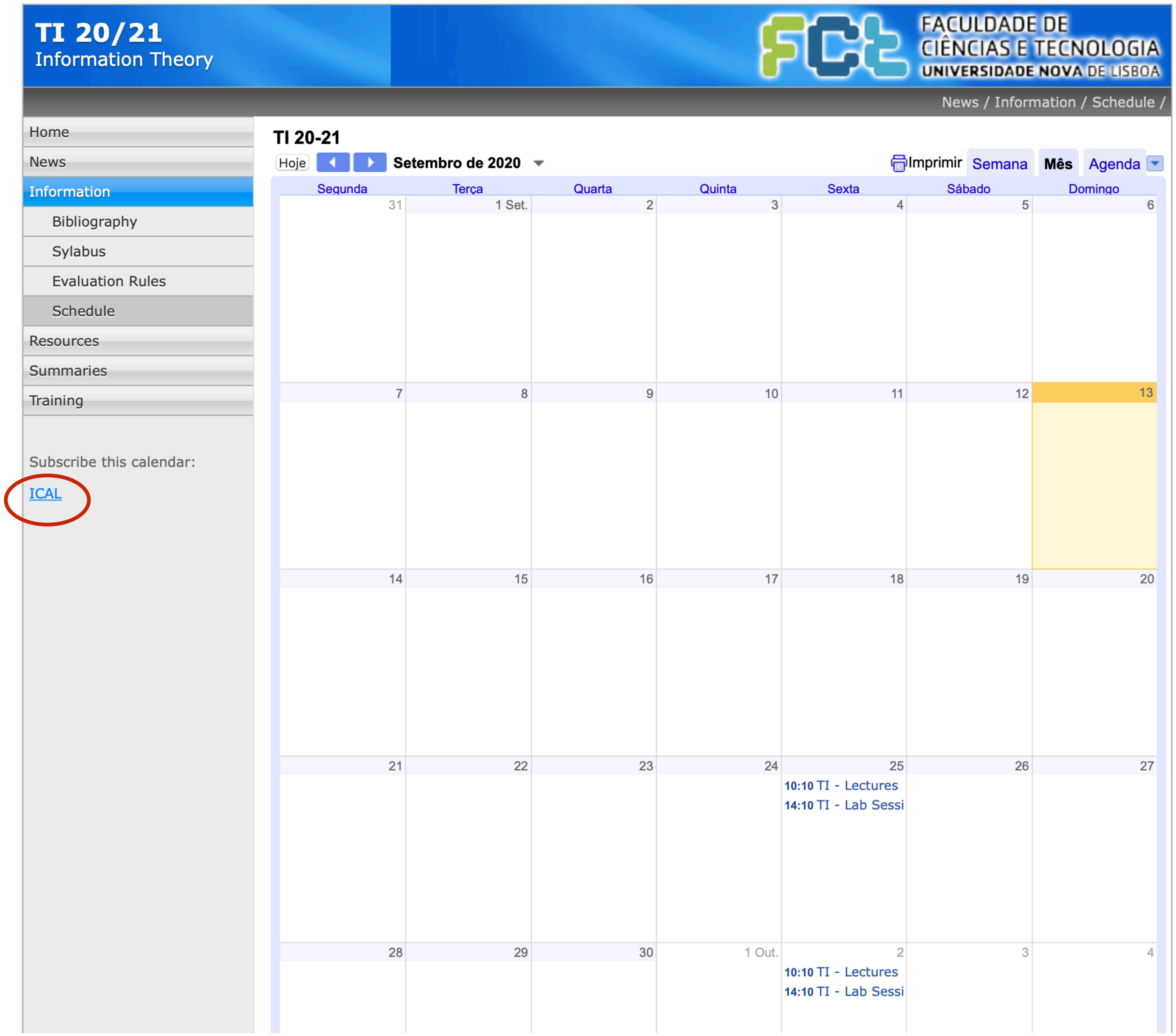
Recommended Readings: (i) Information Theory, Inference, and Learning Algorithms from David MacKay, 2015, pages 1 - 16; **(ii)** The introduction of "A Mathematical Theory of Communication, Claude Shannon, 1948", pages 1-2.

Recommended Activities: (i) see the following video: "Claude Shannon - Father of the Information Age" **(ii)** Visit the various sections of this site.

To Know:

- Why is important the ideia of Digital communications?
- What was the main question that Shannon try to address with Information Theory?
- What is on of the most important result of Shannon's work?
- Understand the General Digital Communication system; What is the role of the Encoder (and the corresponding decoder).
- Binary Symmetric Channel; what is f?
- What is Pb, PB and R (rate)?
- Understand the Repetition codes, (RN).
- Understand the Block codes, the Linear Block codes, the Hamming code H(7, 4)

Web Site: Information / Schedule



Important Dates

- **Test 1: Week starting at November 9th**

- **Test 2: Week starting at January 4th**

- **Project Specification: Up to 13 November**

- **Project delivery: Up to 19 December**

- **Project Oral discussion: 21, 22 December**

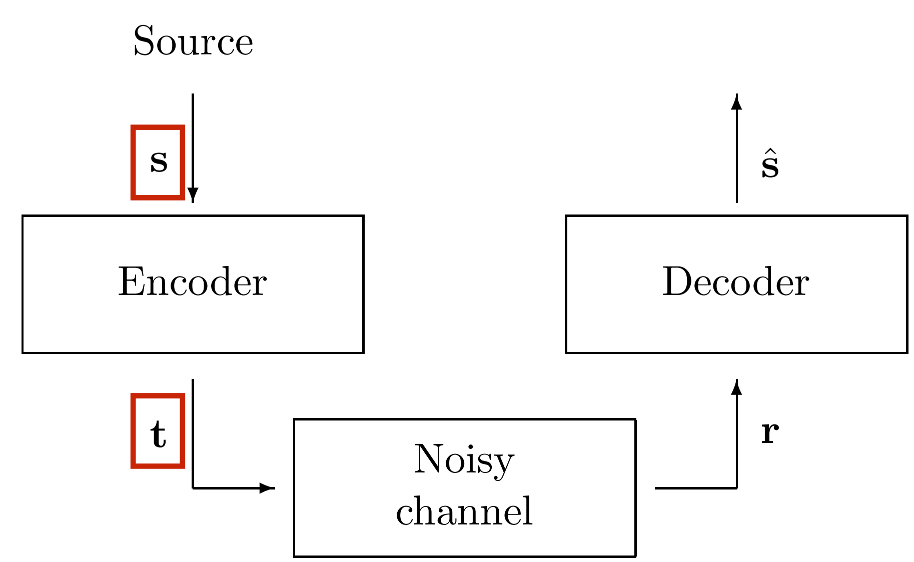
Let’s code. Repeating Codes

General digital communication system

- A straightforward idea is to repeat every bit of the message a prearranged number of times.

Source sequence $\mathbf{s}$	Transmitted sequence $\mathbf{t}$
0	000
1	111

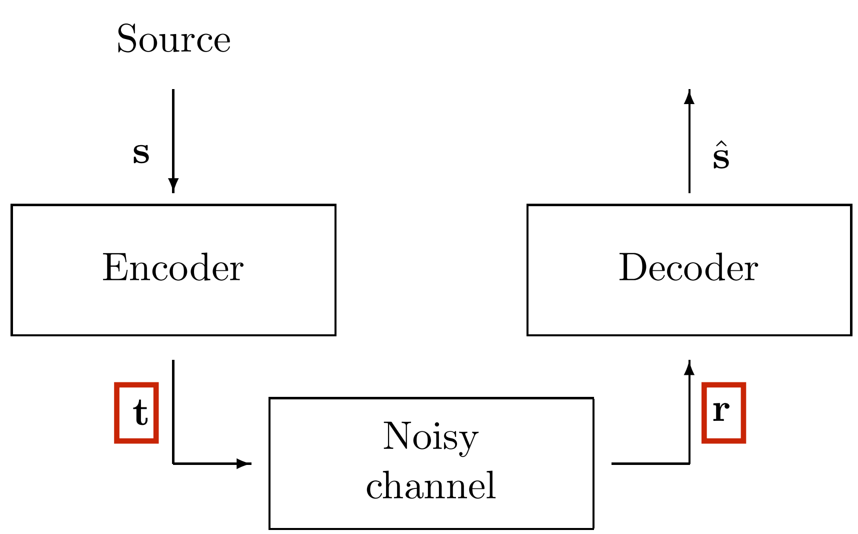
The repetition code R_3



Transmit R3 messages over a BSC

- Transmit R3 messages over a Binary Symmetric Channel with *f* = 0.1

- We can describe the channel as 'adding' a sparse noise vector ***n*** to the transmitted vector ***t*** (adding in modulo 2 arithmetic):
  - A **zero** on ***n*** does not change the transmitted bit
  - A **one** on ***n*** does change the transmitted bit $0 \to 1; \quad 1 \to 0$

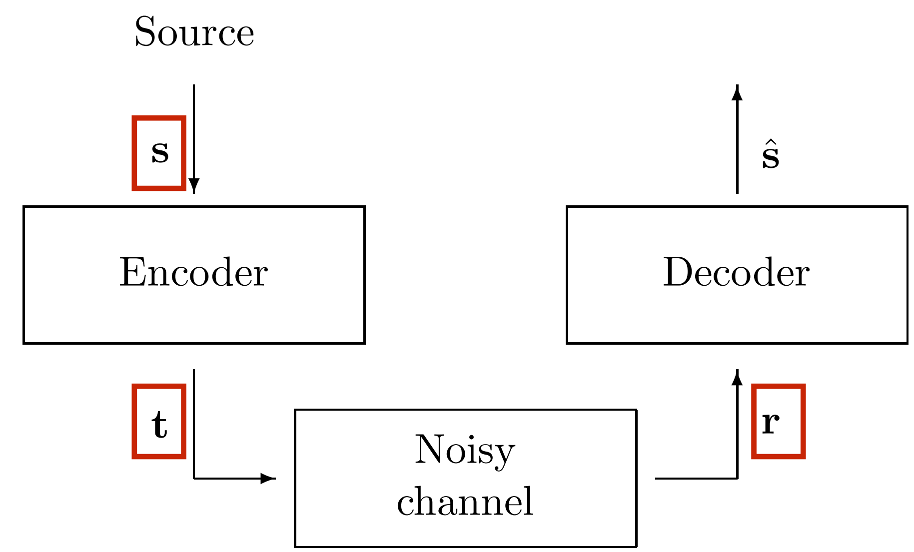


$$r = t \oplus n$$

Transmit R3 messages over a BSC

- Lets send a message with few bits

$$\mathbf{s} = 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 1\ \ 0$$



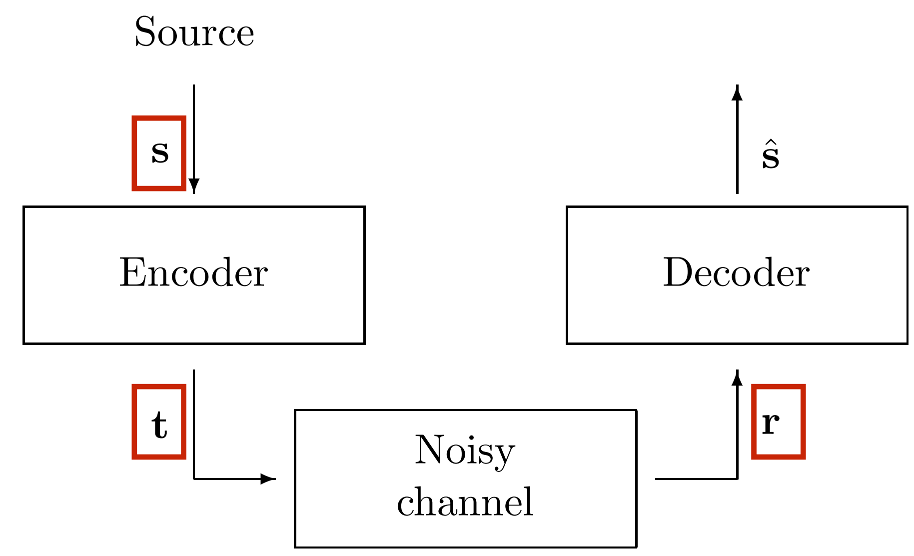
- The transmitted message according to different noise vectors

$\mathbf{s}$	0	0	1	0	1	1	0

Transmit R3 messages over a BSC

- Lets send a message with few bits

$$\mathbf{s} = 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 1\ \ 0$$



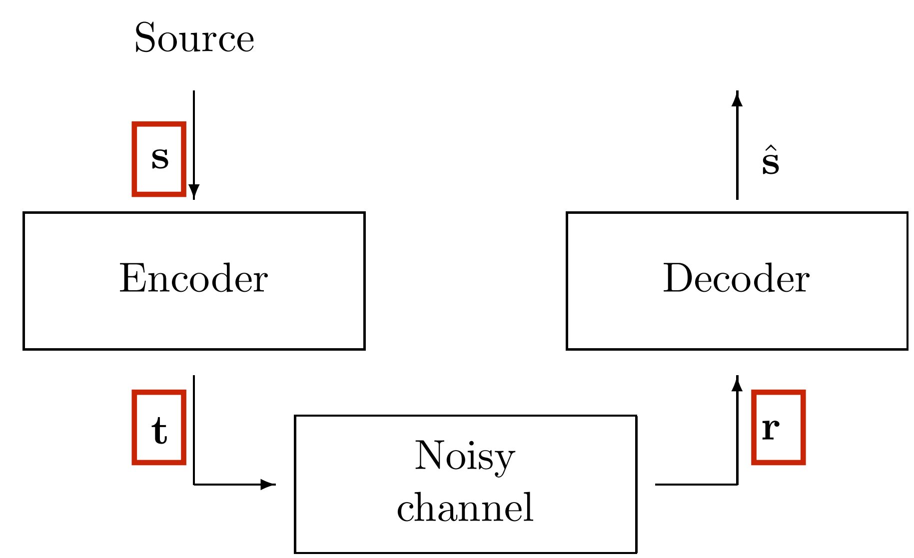
- The transmitted message according to different noise vectors

$$\begin{array}{cccccccc} \mathbf{s} & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ \mathbf{t} & \overbrace{0\,0\,0} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} \end{array}$$

Transmit R3 messages over a BSC

- Lets send a message with few bits

$$\mathbf{s} = 0 \;\; 0 \;\; 1 \;\; 0 \;\; 1 \;\; 1 \;\; 0$$



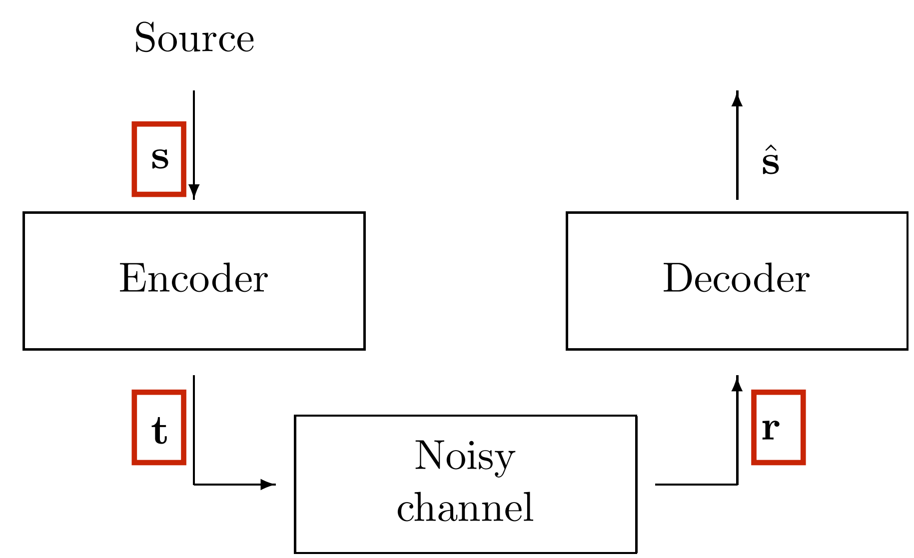
- The transmitted message according to different noise vectors

$$\begin{array}{cccccccc}
\mathbf{s} & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
\mathbf{t} & \overbrace{0\,0\,0} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} \\
\mathbf{n} & 0\,0\,0 & 0\,0\,1 & 0\,0\,0 & 0\,0\,0 & 1\,0\,1 & 0\,0\,0 & 0\,0\,0
\end{array}$$

Transmit R3 messages over a BSC

- Lets send a message with few bits

$$\mathbf{s} = 0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0$$



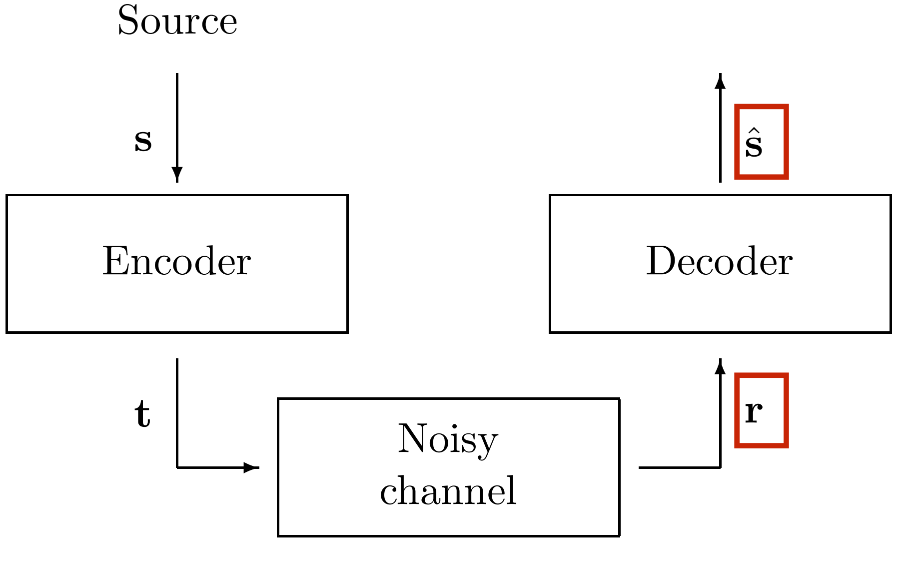
- The transmitted message according to different noise vectors

$$\begin{array}{lccccccc}
\mathbf{s} & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
\mathbf{t} & \overbrace{0\,0\,0} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} \\
\mathbf{n} & 0\,0\,0 & 0\,0\,1 & 0\,0\,0 & 0\,0\,0 & 1\,0\,1 & 0\,0\,0 & 0\,0\,0 \\
\cline{2-8}
\mathbf{r} & 0\,0\,0 & 0\,0\,1 & 1\,1\,1 & 0\,0\,0 & 0\,1\,0 & 1\,1\,1 & 0\,0\,0
\end{array}$$

How should we decode this received vector?

- The optimal algorithm looks at the received bits three at a time and takes a **majority vote.**
  - More 0, take a 0
  - More 1s, take a 1

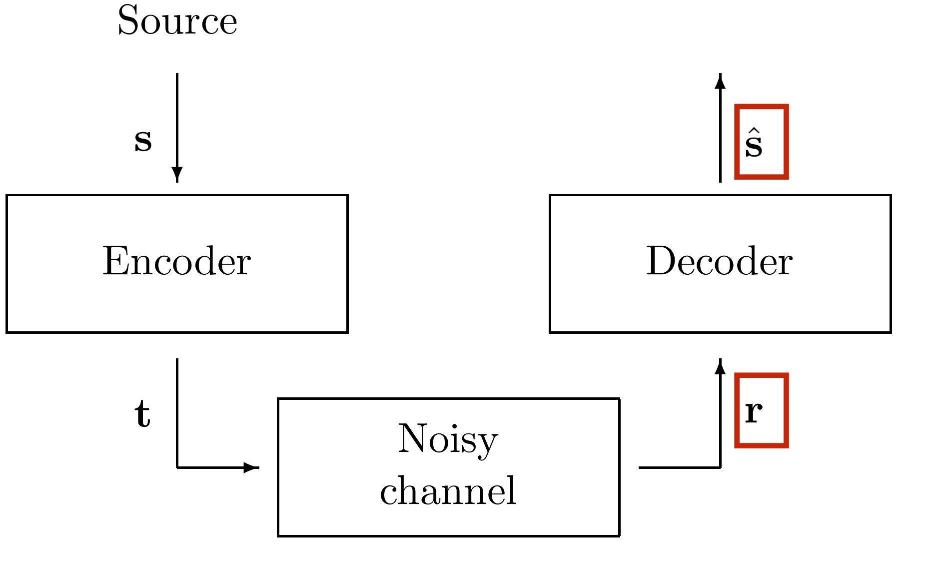
Received sequence $\mathbf{r}$
000
001
010
100
101
110
011
111



How should we decode this received vector?

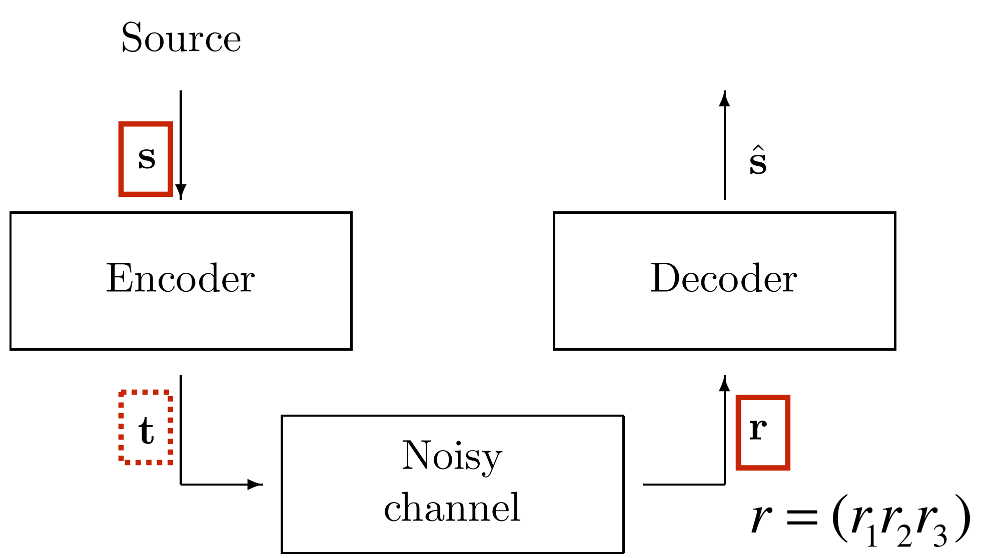
- The optimal algorithm looks at the received bits three at a time and takes a **majority vote.**
  - More 0, take a 0
  - More 1s, take a 1

Received sequence $\mathbf{r}$	Decoded sequence $\hat{\mathbf{s}}$
000	0
001	0
010	0
100	0
101	1
110	1
011	1
111	1



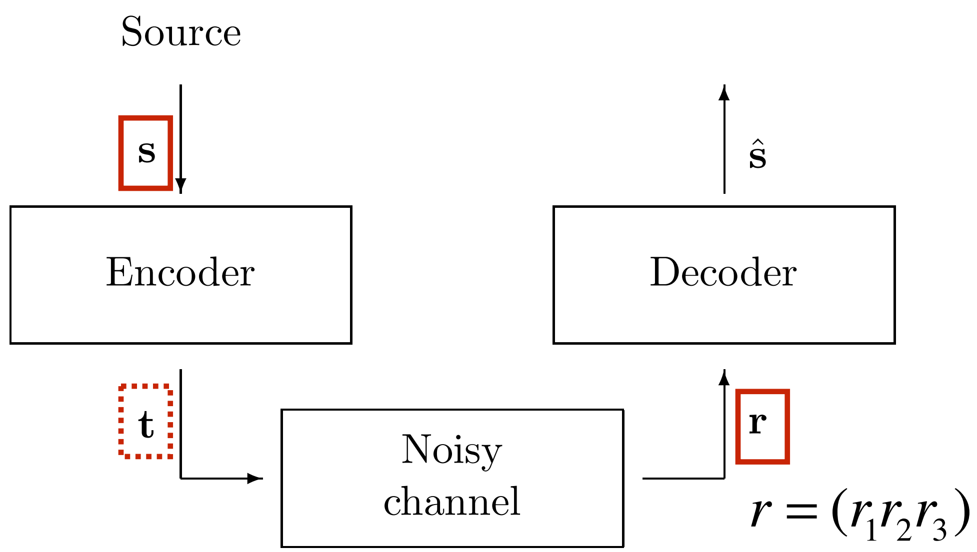
How should we decode this received vector?

- The optimal algorithm looks at the received bits three at a time and takes a **majority vote.**
- The optimal decoding decision (optimal in the sense of having the smallest probability of being wrong) is to find **which value of *s* is most probable, given *r*.**



How should we decode this received vector?

- The optimal algorithm looks at the received bits three at a time and takes a **majority vote.**
- The optimal decoding decision (optimal in the sense of having the smallest probability of being wrong) is to find **which value of *s* is most probable, given *r*.**



Posteriori probability of s

$$P(s \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s) P(s)}{P(r_1 r_2 r_3)}$$

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=1) P(s=1)}{P(r_1 r_2 r_3)}$$

$$P(s=0 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=0) P(s=0)}{P(r_1 r_2 r_3)}$$

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=1)P(s=1)}{P(r_1 r_2 r_3)} \qquad P(s=0 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=0)P(s=0)}{P(r_1 r_2 r_3)}$$

- Depends on
  - *prior probability* $P(s)$
  - the data-dependent term $P(r_1 r_2 r_3 \mid s)$ - the *likelihood* of s

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=1) P(s=1)}{P(r_1 r_2 r_3)} \qquad P(s=0 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=0) P(s=0)}{P(r_1 r_2 r_3)}$$

- Depends on
  - *prior probability* $P(s)$
  - the data-dependent term $P(r_1 r_2 r_3 \mid s)$ - the *likelihood* of s

- The normalizing constant $P(r_1 r_2 r_3)$ does not needs to be calculated to find the most probable s given the received sequence

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1r_2r_3) = \frac{P(r_1r_2r_3 \mid s=1)P(s=1)}{P(r_1r_2r_3)}$$

$$P(s=0 \mid r_1r_2r_3) = \frac{P(r_1r_2r_3 \mid s=0)P(s=0)}{P(r_1r_2r_3)}$$

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=1) P(s=1)}{P(r_1 r_2 r_3)} \qquad P(s=0 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=0) P(s=0)}{P(r_1 r_2 r_3)}$$

- Assumptions:

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1r_2r_3) = \frac{P(r_1r_2r_3 \mid s=1)P(s=1)}{P(r_1r_2r_3)} \qquad P(s=0 \mid r_1r_2r_3) = \frac{P(r_1r_2r_3 \mid s=0)P(s=0)}{P(r_1r_2r_3)}$$

- Assumptions:
  - We assume that the *prior probability are equal*: $P(s = 0) = P(s = 1) = 0.5$

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=1) P(s=1)}{P(r_1 r_2 r_3)} \qquad P(s=0 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=0) P(s=0)}{P(r_1 r_2 r_3)}$$

- Assumptions:
  - We assume that the *prior probability are equal*: $P(s = 0) = P(s = 1) = 0.5$
  - We assume that the channel is a Binary Symmetric Channel with noise level $f < 0.5$

How should we decode this received vector?

- **Which value of *s* is most probable, given *r*.**

$$P(s=1 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=1)P(s=1)}{P(r_1 r_2 r_3)} \qquad P(s=0 \mid r_1 r_2 r_3) = \frac{P(r_1 r_2 r_3 \mid s=0)P(s=0)}{P(r_1 r_2 r_3)}$$

- Assumptions:
  - We assume that the *prior probability are equal*: $P(s = 0) = P(s = 1) = 0.5$
  - We assume that the channel is a Binary Symmetric Channel with noise level $f < 0.5$

- So maximizing $P(s \mid r_1 r_2 r_3)$ just requires to maximize the the *likelihood* $P(\mathbf{r} \mid s)$

$$P(\mathbf{r} \mid s) = P(\mathbf{r} \mid \mathbf{t}(s)) = \prod_{n=1}^{N} P(r_n \mid t_n(s))$$

N = 3 is the number of bits in the block

The BSC has no memory !

How should we decode this received vector?

- So maximizing $P(s \mid r_1 r_2 r_3)$ just requires to maximize the the *likelihood* $P(\boldsymbol{r} \mid s)$

$$P(\boldsymbol{r} \mid s) = P(\boldsymbol{r} \mid \boldsymbol{t}(s)) = \prod_{n=1}^{N} P(r_n \mid t_n(s))$$

N = 3 is the number of bits in the block

The BSC has no memory !

- Where $P(r_n \mid t_n)$,

$$P(r_n \mid t_n) = \begin{cases} (1-f) & \text{if} \quad r_n = t_n \\ f & \text{if} \quad r_n \neq t_n \end{cases}$$

How should we decode this received vector?

- So maximizing $P(s \mid r_1 r_2 r_3)$ just requires to maximize the the *likelihood* $P(\boldsymbol{r} \mid s)$

$$P(\boldsymbol{r} \mid s) = P(\boldsymbol{r} \mid \boldsymbol{t}(s)) = \prod_{n=1}^{N} P(r_n \mid t_n(s))$$

N = 3 is the number of bits in the block

The BSC has no memory !

- Where $P(r_n \mid t_n)$,

$$P(r_n \mid t_n) = \begin{cases} (1-f) & \text{if} \quad r_n = t_n \\ f & \text{if} \quad r_n \neq t_n \end{cases}$$

- and the *likelihood ratio* for the to possible hypotheses is,

$$\frac{P(\boldsymbol{r} \mid s=1)}{P(\boldsymbol{r} \mid s=0)} = \prod_{n=1}^{N} \frac{P(r_n \mid t_n(1))}{P(r_n \mid t_n(0))}$$

$$\frac{P(r_n \mid t_n(1))}{P(r_n \mid t_n(0))} = \frac{1-f}{f} \quad \text{if} \quad r_n = 1$$

$$\frac{P(r_n \mid t_n(1))}{P(r_n \mid t_n(0))} = \frac{f}{1-f} \quad \text{if} \quad r_n = 0$$

How should we decode this received vector?

- The *likelihood ratio* for the two possible hypotheses is,

$$\frac{P(\boldsymbol{r} \mid s=1)}{P(\boldsymbol{r} \mid s=0)} = \prod_{n=1}^{N} \frac{P(r_n \mid t_n(1))}{P(r_n \mid t_n(0))}$$

$$\frac{P(r_n|t_n(1))}{P(r_n|t_n(0))} = \frac{1-f}{f} \quad if \quad r_n = 1$$

$$\frac{P(r_n|t_n(1))}{P(r_n|t_n(0))} = \frac{f}{1-f} \quad if \quad r_n = 0$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

How should we decode this received vector?

- The *likelihood ratio* for the two possible hypotheses is,

$$\frac{P(\boldsymbol{r} \mid s=1)}{P(\boldsymbol{r} \mid s=0)} = \prod_{n=1}^{N} \frac{P(r_n \mid t_n(1))}{P(r_n \mid t_n(0))}$$

$$\frac{P(r_n|t_n(1))}{P(r_n|t_n(0))} = \frac{1-f}{f} \quad if \quad r_n = 1$$

$$\frac{P(r_n|t_n(1))}{P(r_n|t_n(0))} = \frac{f}{1-f} \quad if \quad r_n = 0$$

- Since f < 0.5, $\gamma = \frac{1-f}{f} > 1$ and so,
  - The winning hypothesis is the one with the most ‘votes’.
  - Each vote counting for a factor of γ in the *likelihood ratio*.

How should we decode this received vector?

- The majority vote to decode R3.

Received sequence $\mathbf{r}$	Likelihood ratio $\frac{P(\mathbf{r} \mid s = 1)}{P(\mathbf{r} \mid s = 0)}$	Decoded sequence $\hat{\mathbf{s}}$
000	γ^{-3}	0
001	γ^{-1}	0
010	γ^{-1}	0
100	γ^{-1}	0
101	γ^{1}	1
110	γ^{1}	1
011	γ^{1}	1
111	γ^{3}	1

Decoding a message

- The optimal algorithm looks at the received bits three at a time and takes a **majority vote.**

$$\mathbf{s} = 0\ 0\ 1\ 0\ 1\ 1\ 0$$

$\mathbf{s}$	0	0	1	0	1	1	0
$\mathbf{t}$	000	000	111	000	111	111	000
$\mathbf{n}$	000	001	000	000	101	000	000
$\mathbf{r}$	000	001	111	000	010	111	000
$\hat{\mathbf{s}}$	0	0	1	0	0	1	0
corrected errors		$\star$					
undetected errors					$\star$		

- no errors: the message is correctly decoded
- one error: the original is recovered
- two or three errors: the message is incorrectly decoded.

Probability of error of R3 coding?

Probability of error of R3 coding?

- An error is made by R3 if two or more bits are flipped in a block of three.

Probability of error of R3 coding?

- An error is made by R3 if two or more bits are flipped in a block of three.
- The error probability of R3 is a sum of two terms:
  - the probability that all three bits are flipped = f^3;
  - the probability that exactly two bits are flipped, $3f^2(1 - f)$.

Probability of error of R3 coding?

- An error is made by R3 if two or more bits are flipped in a block of three.
- The error probability of R3 is a sum of two terms:
  - the probability that all three bits are flipped = f^3;
  - the probability that exactly two bits are flipped, $3f^2(1 - f)$.

$$\boxed{P_b = 3f^2(1-f) + f^3}$$

Probability of error of R3 coding?

- An error is made by R3 if two or more bits are flipped in a block of three.
- The error probability of R3 is a sum of two terms:
  - the probability that all three bits are flipped = f^3;
  - the probability that exactly two bits are flipped, $3f^2(1 - f)$.

$$\boxed{P_b = 3f^2(1-f)+f^3}$$

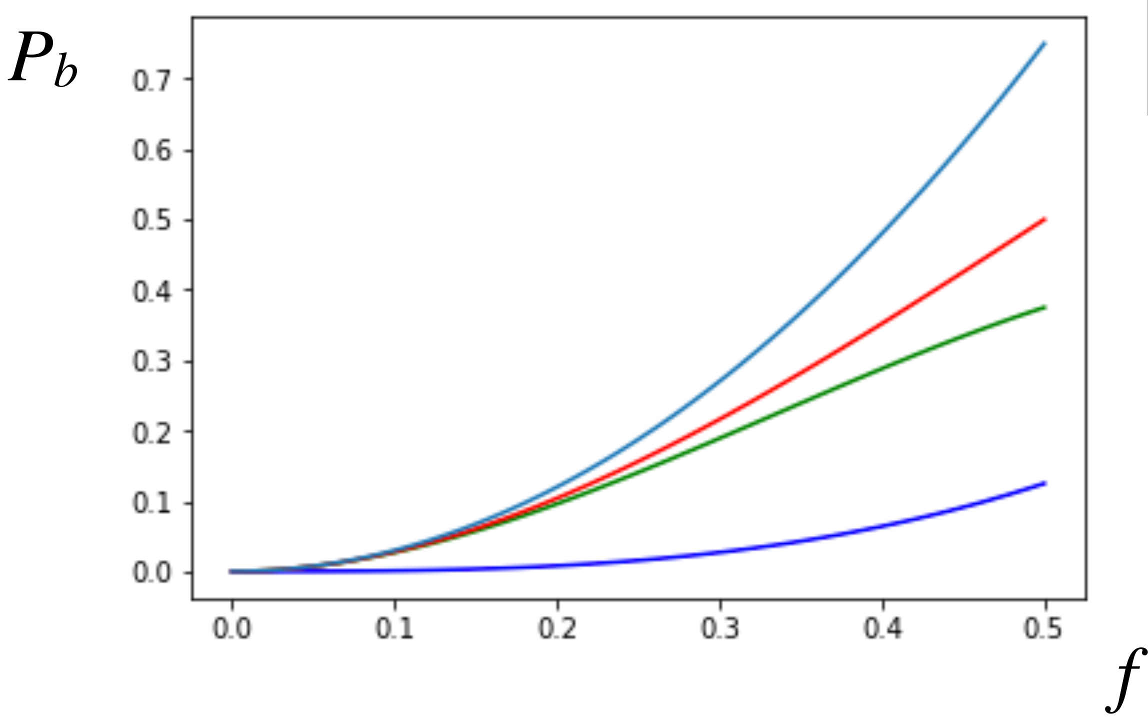
$$\boxed{P_b = 3f^2 - 2f^3}$$

Probability of error of R3 coding?

- An error is made by R3 if two or more bits are flipped in a block of three.
- The error probability of R3 is a sum of two terms:
  - the probability that all three bits are flipped $= f^3$;
  - the probability that exactly two bits are flipped, $3f^2(1 - f)$.

$$\boxed{P_b = 3f^2(1-f) + f^3}$$

$$\boxed{P_b = 3f^2 - 2f^3}$$



$\boxed{3f^2}$

any error

2 bits errors

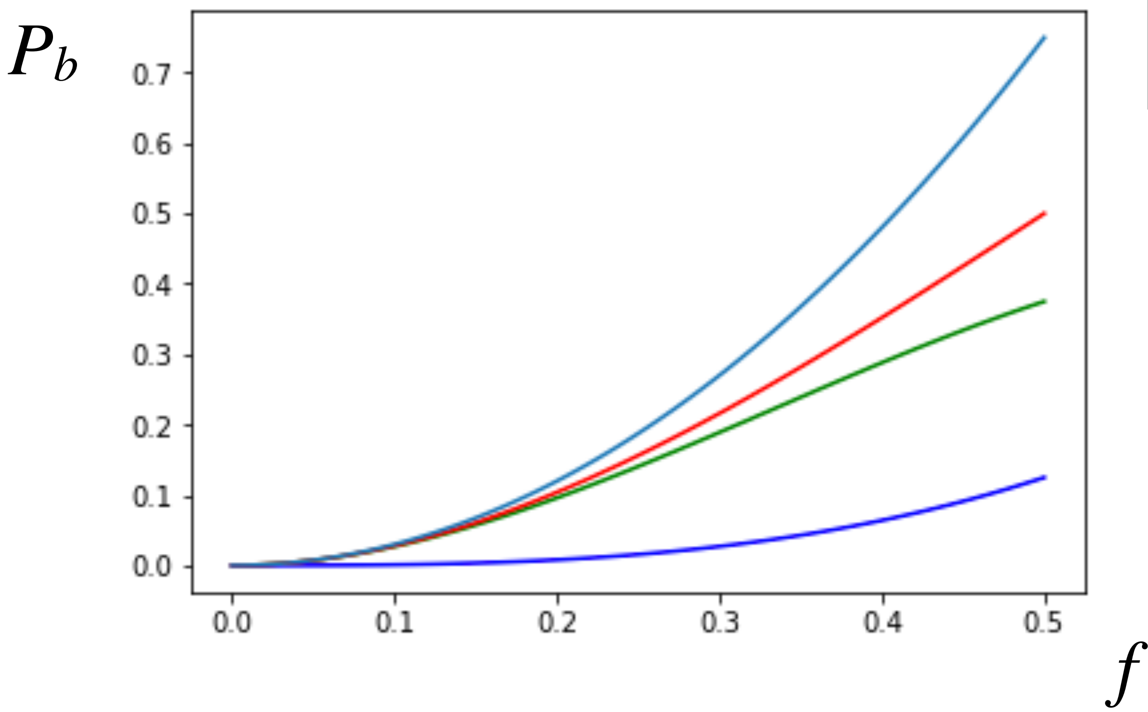
3 bits errors

Probability of error of R3 coding?

- An error is made by R3 if two or more bits are flipped in a block of three.
- The error probability of R3 is a sum of two terms:
  - the probability that all three bits are flipped = f^3;
  - the probability that exactly two bits are flipped, $3f^2(1-f)$.

$$\boxed{P_b = 3f^2(1-f) + f^3}$$

$$\boxed{P_b = 3f^2 - 2f^3}$$



$\boxed{3f^2}$

any error

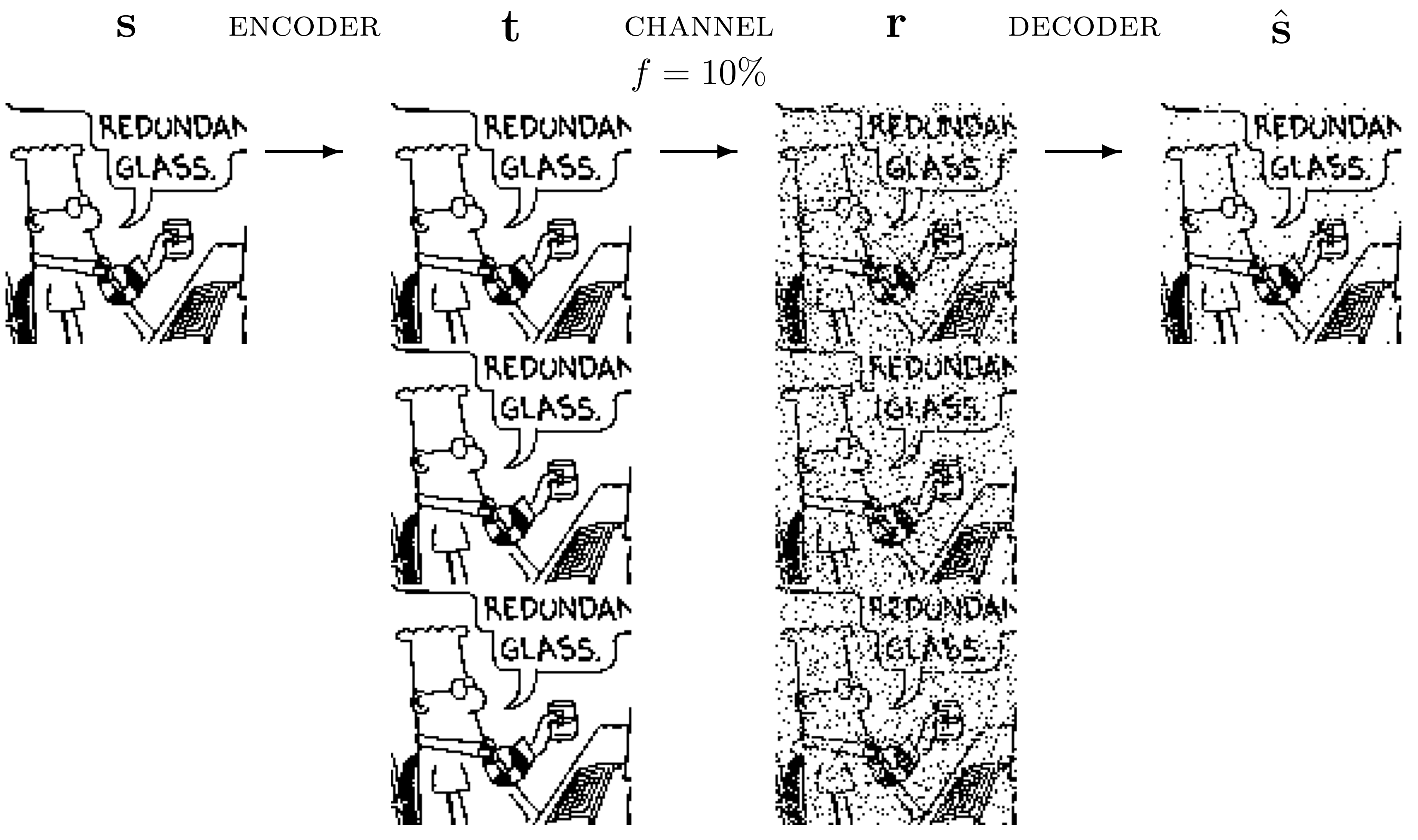
2 bits errors

3 bits errors

$$\boxed{f = 0.1} \longrightarrow \boxed{P_b = 0.03}$$

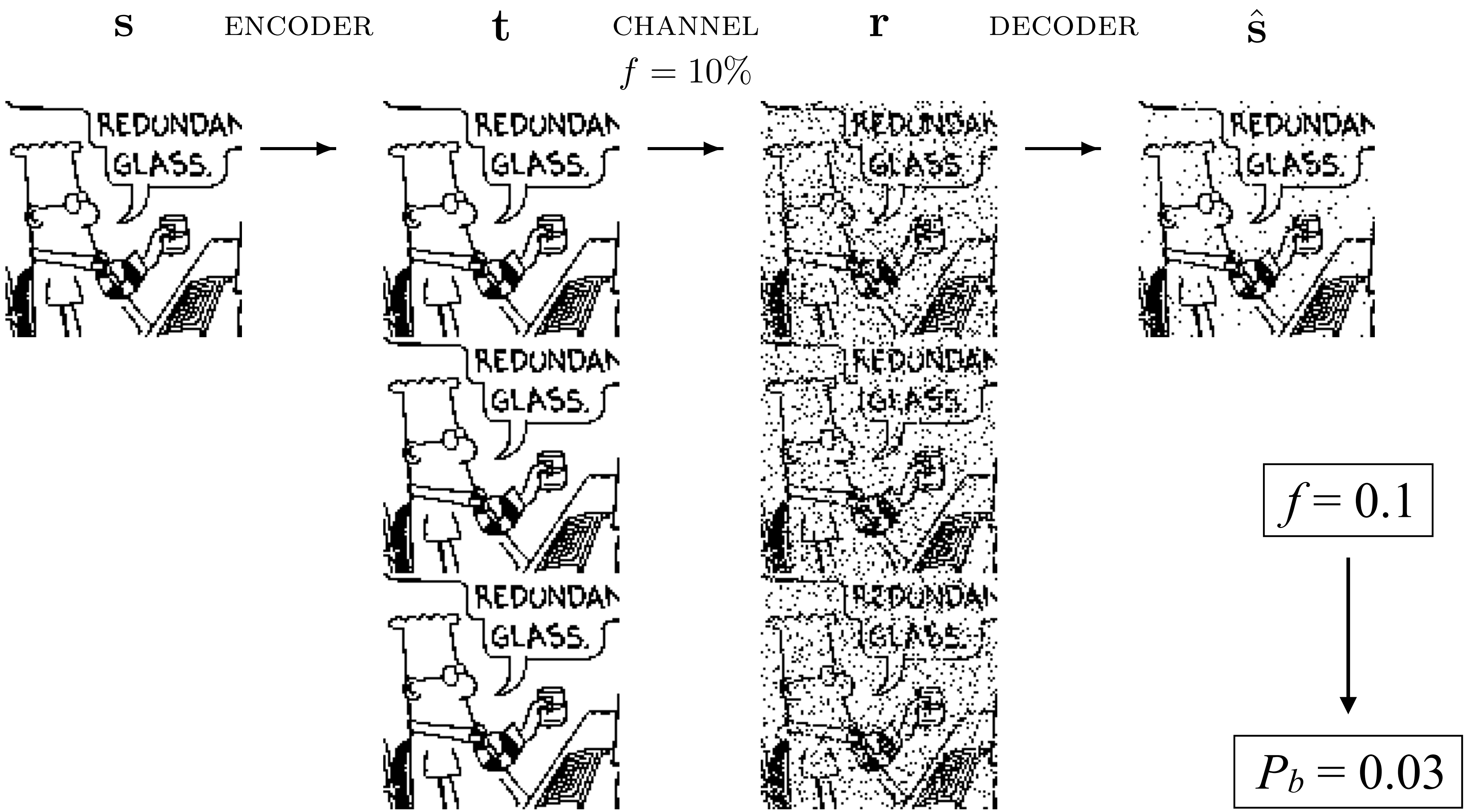
The improvement on Pb with R3

- Assuming a BSC with $f = 0.1$



The improvement on Pb with R3

- Assuming a BSC with f = 0.1



The improvement on P_b with R_3: What is the cost?

- Assuming a BSC with $f = 0.1$, with R_3 we reduce the error probability from 0.1 to 0.03, but.

The improvement on P_b with R_3: What is the cost?

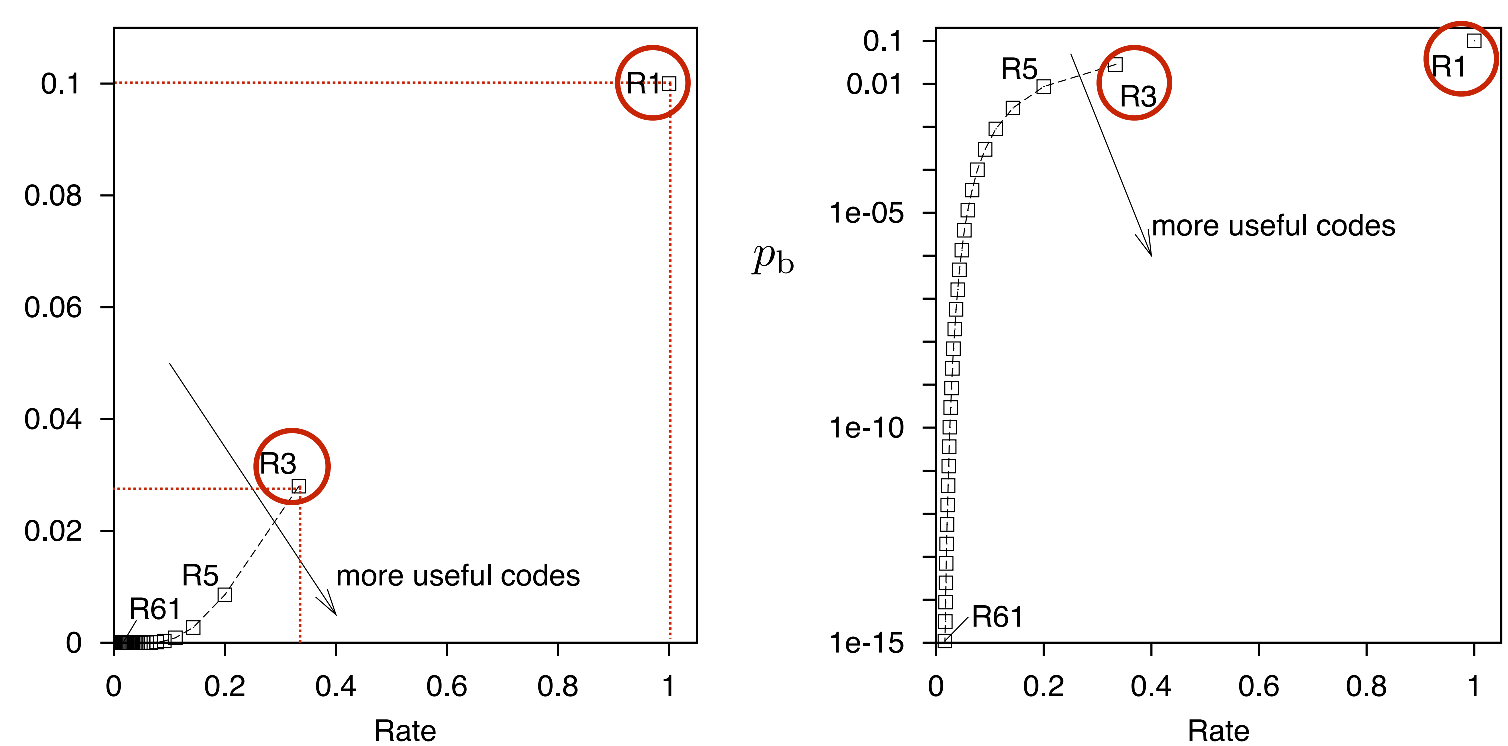
- Assuming a BSC with f = 0.1, with R_3 we reduce the error probability from 0.1 to 0.03, but.
- The rate of information transfer has fallen by a factor of three

The improvement on P_b with R_3: What is the cost?

- Assuming a BSC with f = 0.1, with R_3 we reduce the error probability from 0.1 to 0.03, but.
- The rate of information transfer has fallen by a factor of three
  - We would need **three** of the original noisy gigabyte disk drives in order to create a one-gigabyte disk drive with pb = 0.03 !

The improvement on P_b with R_3: What is the cost?

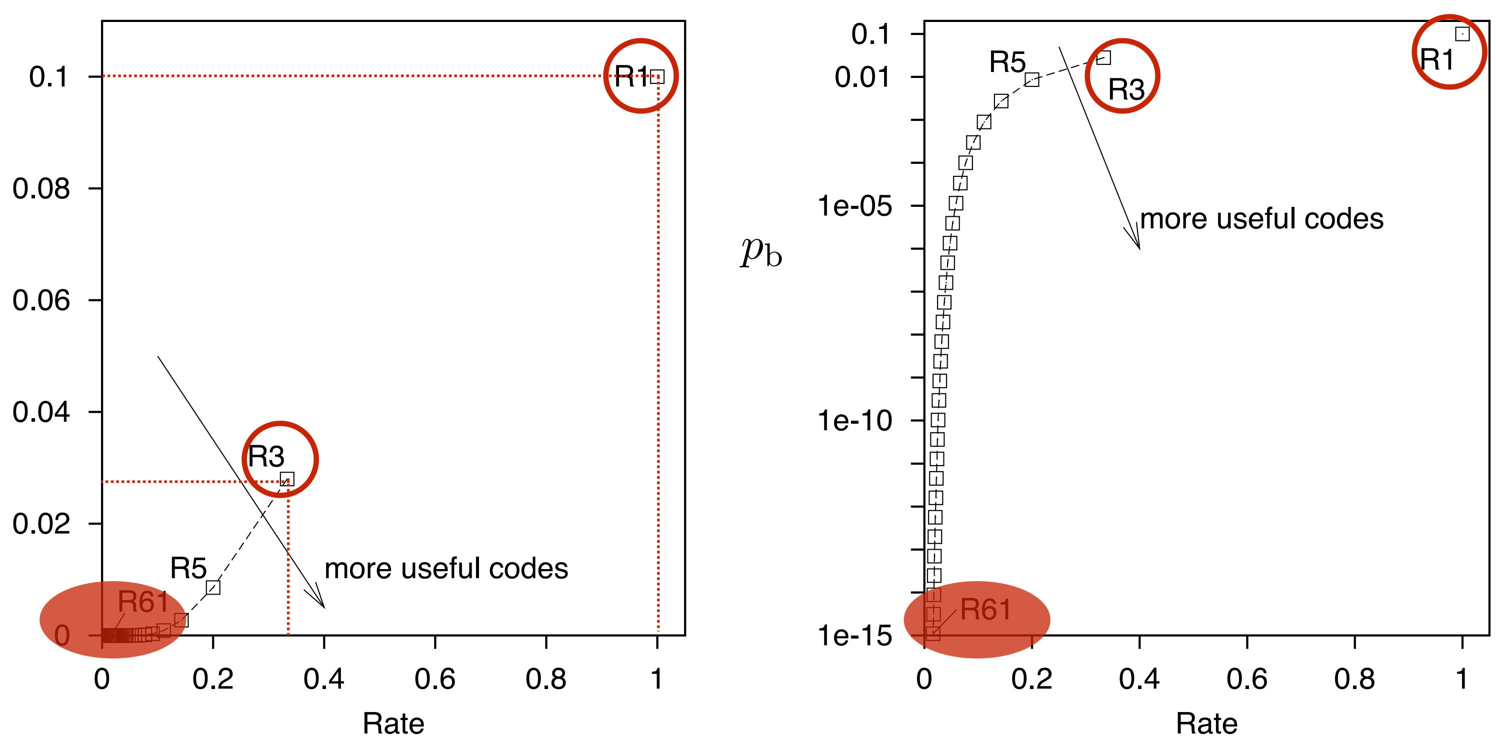
- Assuming a BSC with f = 0.1, with R_3 we reduce the error probability from 0.1 to 0.03, but.
- The rate of information transfer has fallen by a factor of three
  - We would need **three** of the original noisy gigabyte disk drives in order to create a one-gigabyte disk drive with pb = 0.03 !



Log scale

What improvements could we expect? At What rate?

- Can we push the error probability lower, to the values required for a sellable disk drive (e.g. 10^{-15}) ?
- So to build a single gigabyte disk drive with the required reliability from noisy gigabyte drives with f = 0.1, we would need **60** of the noisy disk drives



Log scale

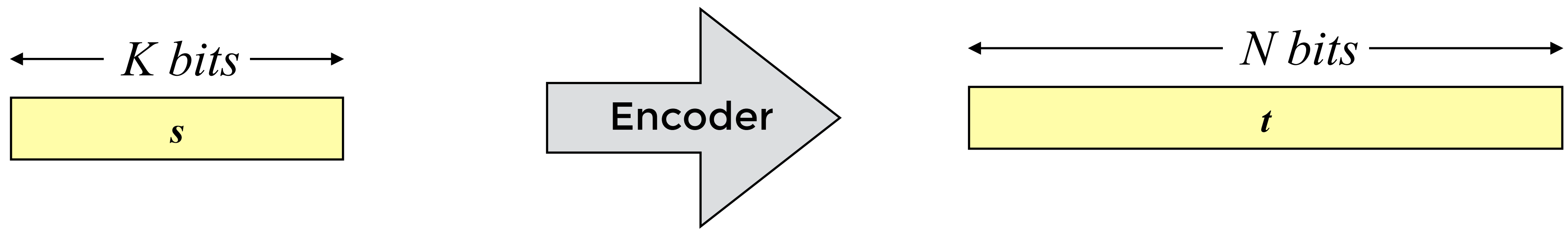
Block codes – the (7, 4) Hamming code

Block Codes

- Add redundancy to **blocks of data** instead of encoding one bit at a time

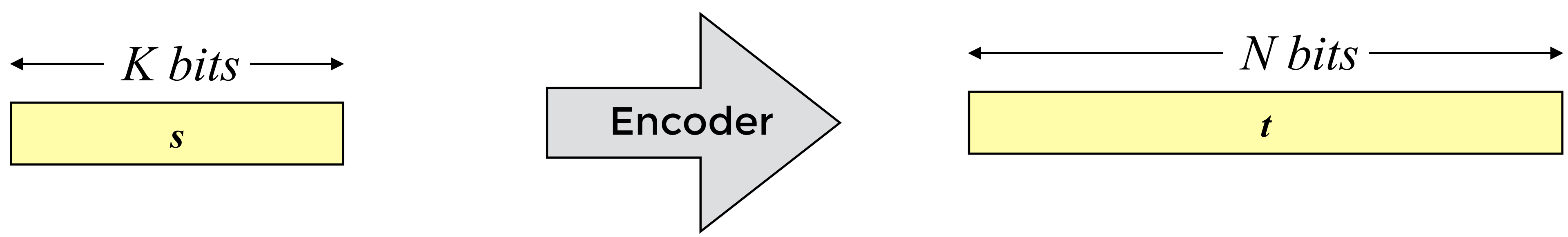
Block Codes

- Add redundancy to **blocks of data** instead of encoding one bit at a time
- A **block code** is a rule for converting a **sequence of source bits** s, of length ***K***, say, into a transmitted sequence t of length N bits.
  - To add redundancy, $N > K$

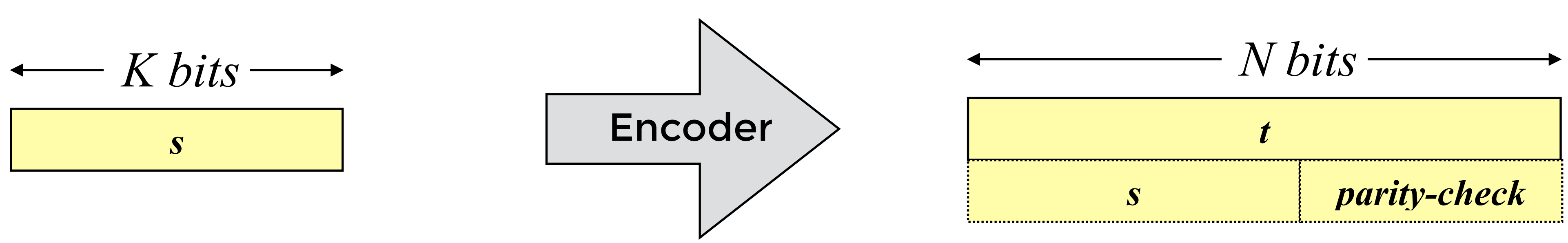


Block Codes

- Add redundancy to **blocks of data** instead of encoding one bit at a time
- A **block code** is a rule for converting a **sequence of source bits** s, of length K, say, into a transmitted sequence t of length N bits.
  - To add redundancy, $N > K$

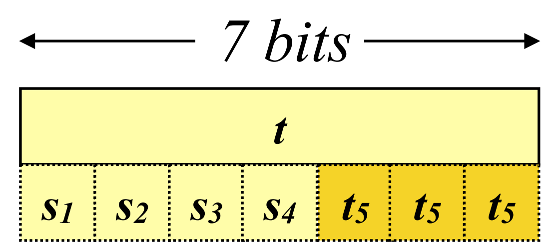


- In a **linear block code**, the extra $N - K$ bits are **linear functions** of the original K bits



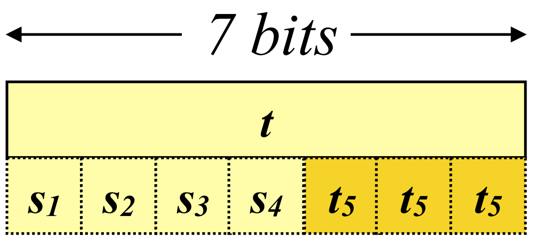
Hamming Code (7, 4)

- Linear block code with $N = 7$; $K = 4$
  - 4 source bits
  - 3 parity check bits



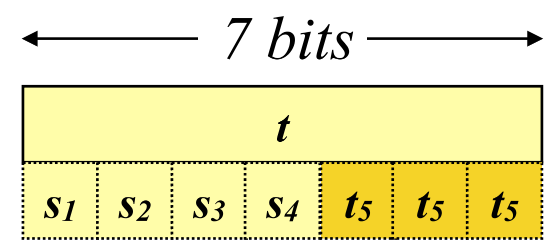
Hamming Code (7, 4)

- Linear block code with $N = 7$; $K = 4$
  - 4 source bits
  - 3 parity check bits
- The 3 parity check bits are linear combinations of the message bits



Hamming Code (7, 4)

- Linear block code with $N = 7$; $K = 4$
  - 4 source bits
  - 3 parity check bits



- The 3 parity check bits are linear combinations of the message bits

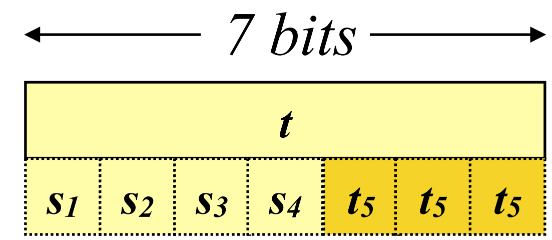
$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

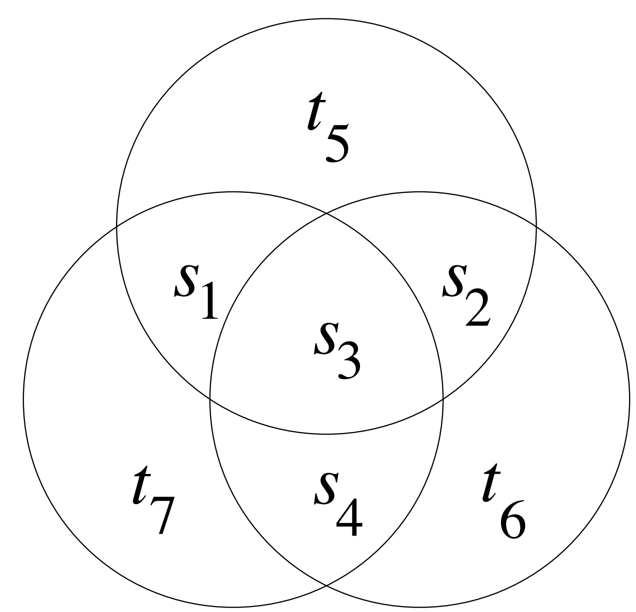
Hamming Code (7, 4)

- Linear block code with *N* = 7; *K* = 4
  - 4 source bits
  - 3 parity check bits



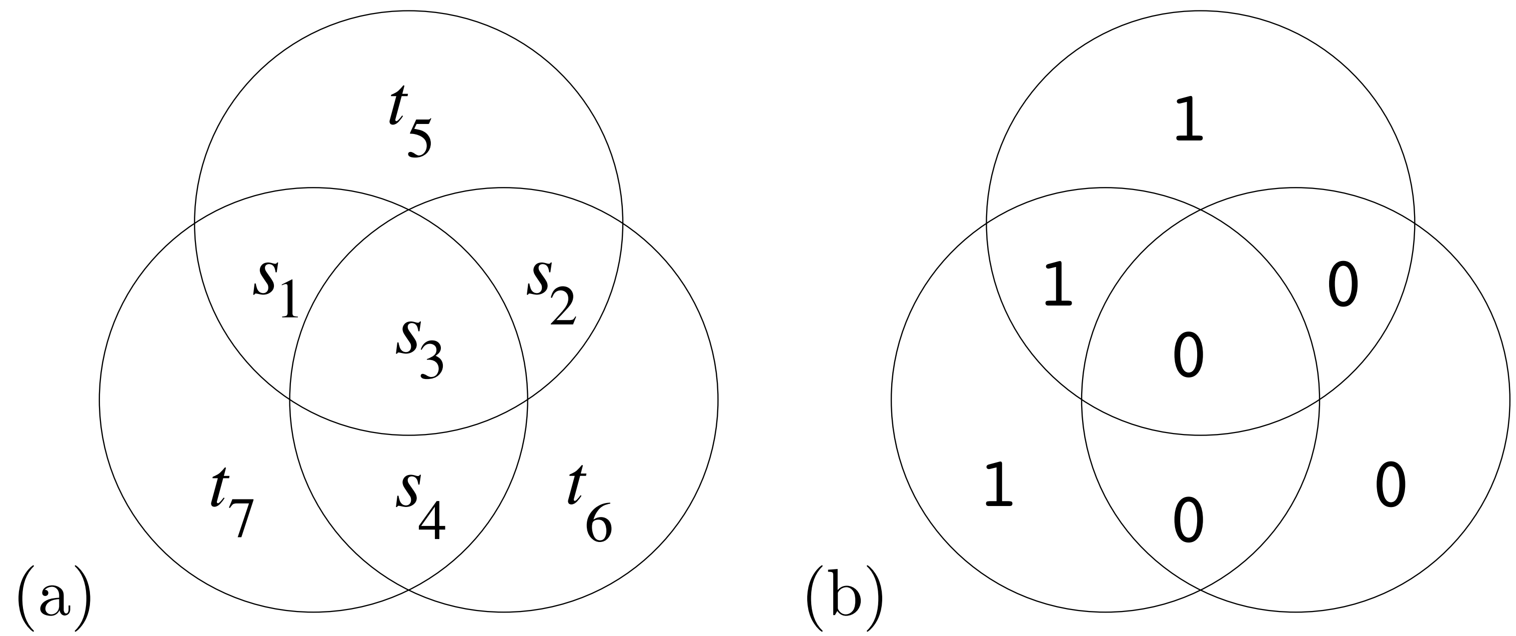
- The 3 parity check bits are linear combinations of the message bits

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$



Hamming Code (7, 4)

- Transmitting s = **1000**



$$t_5 = s_1 \oplus s_2 \oplus s_3 \qquad t_5 = 1 \oplus 0 \oplus 0$$

$$t_6 = s_2 \oplus s_3 \oplus s_4 \qquad t_6 = 0 \oplus 0 \oplus 0$$

$$t_7 = s_1 \oplus s_3 \oplus s_4 \qquad t_7 = 1 \oplus 0 \oplus 0$$

Hamming Code (7, 4): codes for the 2^4 messages

s	t	s	t	s	t	s	t
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

- In H(7, 4) any pair of codewords differ from each other in at **least three bits**
  - What this suggest in terms of its capabilities of detecting errors?
  - Or even in terms of its capabilities of correcting errors?

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Hamming Code (7, 4): Matricial form

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

s and **t** as column vectors

$$\boldsymbol{t} = G^T \boldsymbol{s}$$

$$G^T = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{array} \begin{array}{c} \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right] \end{array}$$

s and **t** as row vectors

$$\boldsymbol{t} = \boldsymbol{s}G$$

$$G = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{ccccccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \end{array} \\ \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right] \end{array}$$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)

- We assume a Binary Symmetric Channel and that all source vectors are equiprobable.

Decoding the Hamming Code (7, 4)

- We assume a Binary Symmetric Channel and that all source vectors are equiprobable.
- The **optimal decoder** identifies the source vector $\boldsymbol{s}$ whose encoding $\boldsymbol{t}(\boldsymbol{s})$ **differs from the received vector $\boldsymbol{r}$ in the fewest bits**. This corresponds to find the closest codeword of $\boldsymbol{r}$.

Decoding the Hamming Code (7, 4)

- We assume a Binary Symmetric Channel and that all source vectors are equiprobable.
- The **optimal decoder** identifies the source vector $\mathbf{s}$ whose encoding $\mathbf{t}(\mathbf{s})$ **differs from the received vector $\mathbf{r}$ in the fewest bits**. This corresponds to find the closest codeword of $\mathbf{r}$.
- Since *any pair of codewords differ from each other in at least three bits*, the H(7, 4) **will detect and correct any error on a single bit**. It will be misleading with errors on two bits.

Decoding the Hamming Code (7, 4)

- We assume a Binary Symmetric Channel and that all source vectors are equiprobable.
- The **optimal decoder** identifies the source vector s whose encoding $t(s)$ **differs from the received vector r in the fewest bits**. This corresponds to find the closest codeword of r.
- Since *any pair of codewords differ from each other in at least three bits*, the H(7, 4) **will detect and correct any error on a single bit**. It will be misleading with errors on two bits.
- Each error on one bit is associated to a ***syndrome***.

Decoding the Hamming Code (7, 4)

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

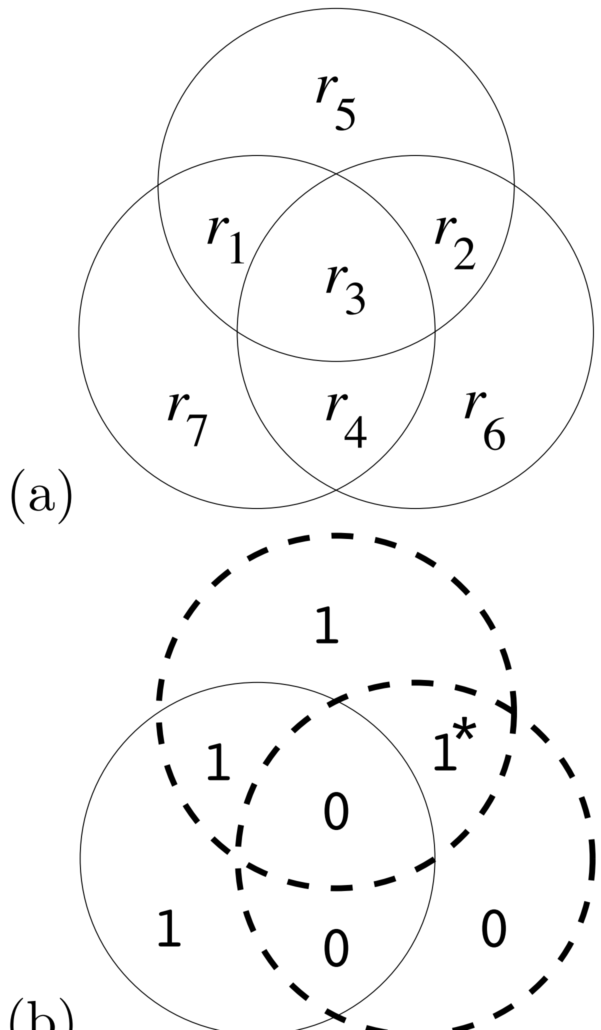
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



$\boldsymbol{s}$ = 1000

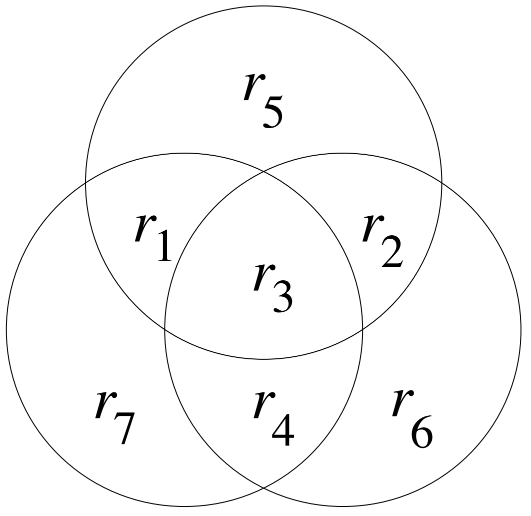
$\boldsymbol{t}$ = 1000**101**

$\boldsymbol{n}$ = 0**1**00000

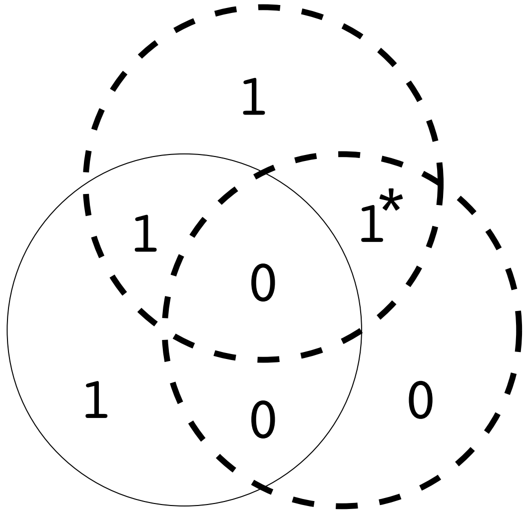
$\boldsymbol{r}$ = 1100101

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(b)

Which circles violate the parity check?

$\boldsymbol{s} = 1000$

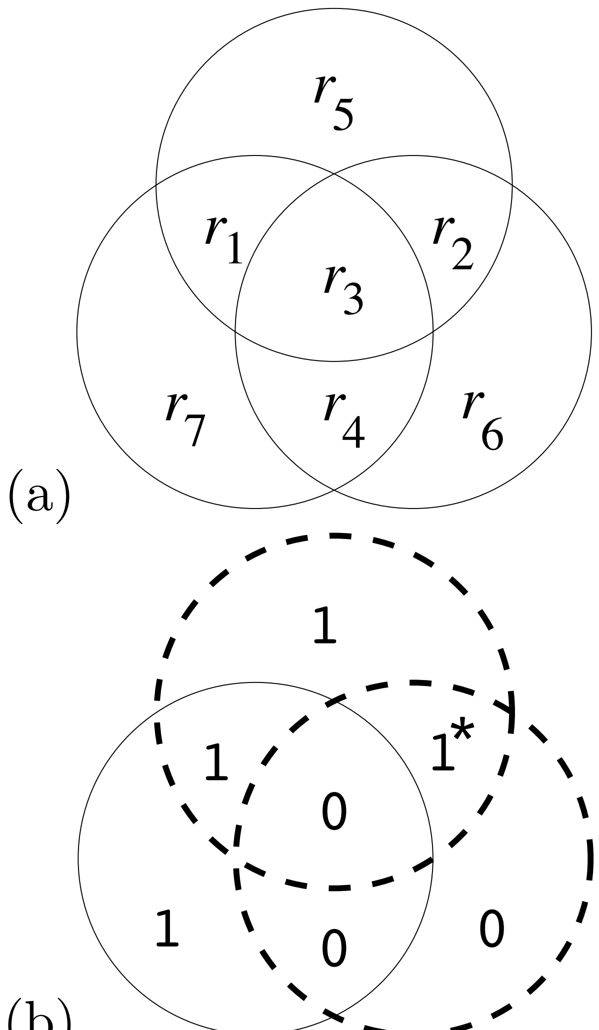
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

$\boldsymbol{s} = 1000$

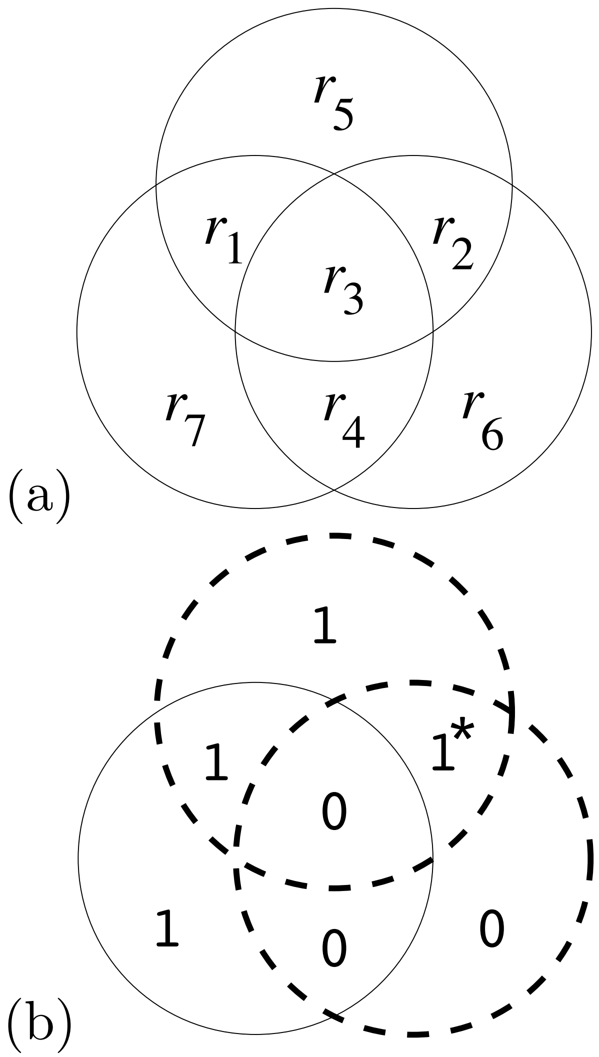
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

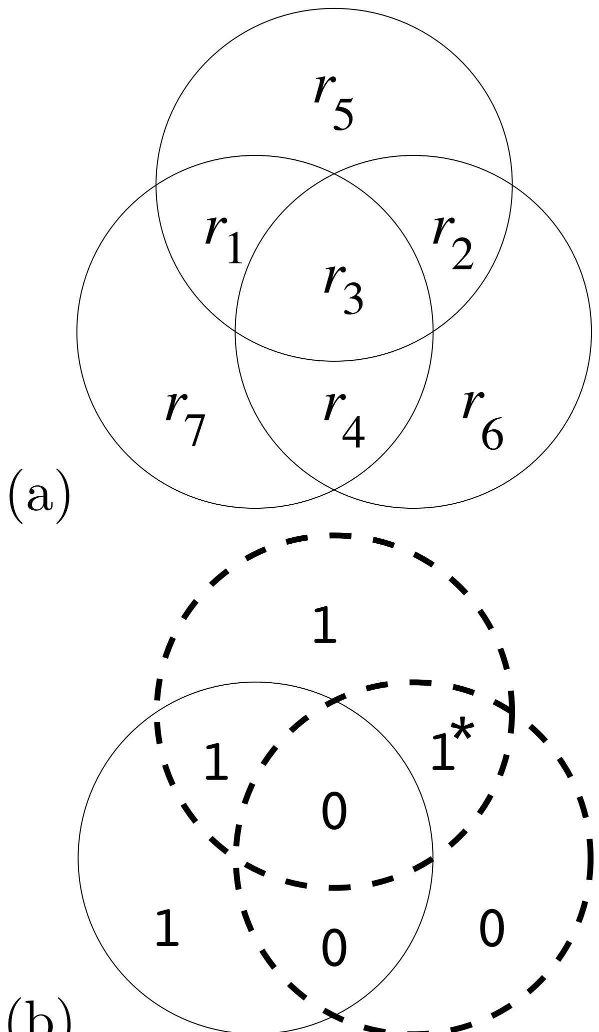
$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

only r_2 ! (the flipped bit)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

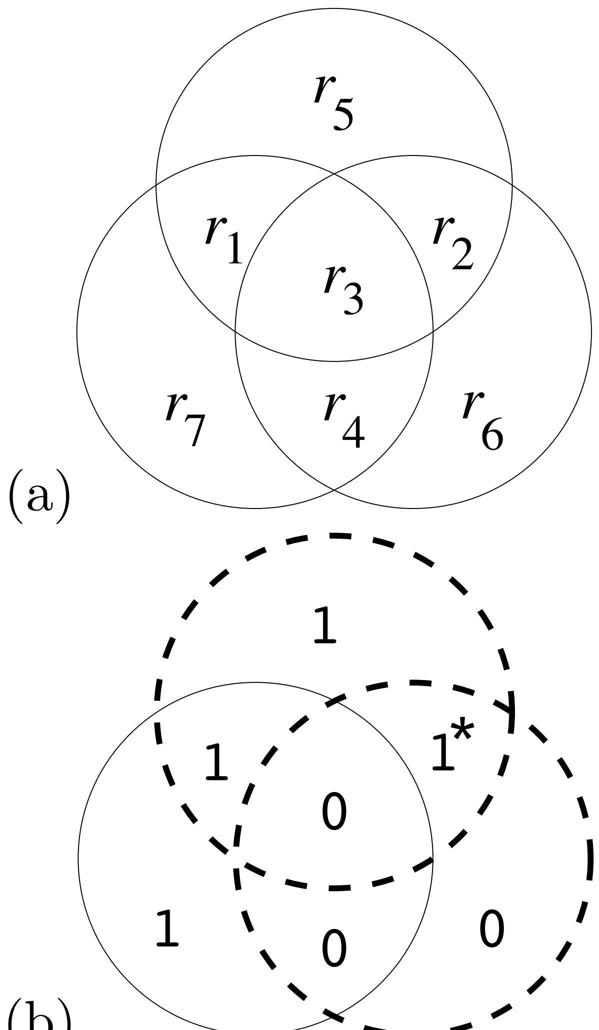
$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

only r_2 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

$\boldsymbol{s} = 1000$

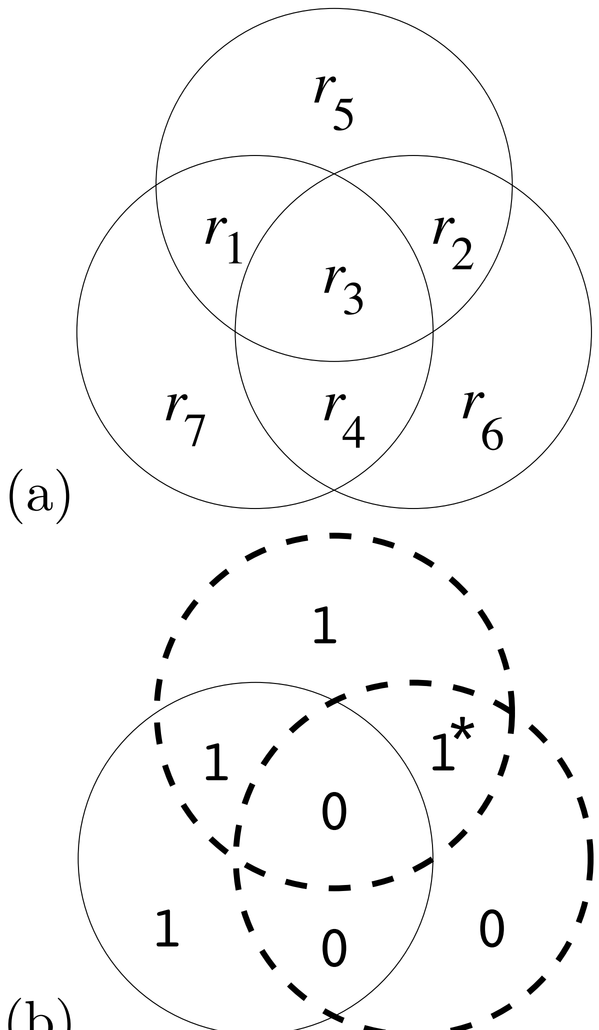
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

only r_2 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

z = (110)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)

0

Decoding the Hamming Code (7, 4)

$$s = 1000$$

0

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

0

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

0

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

0

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Decoding the Hamming Code (7, 4)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

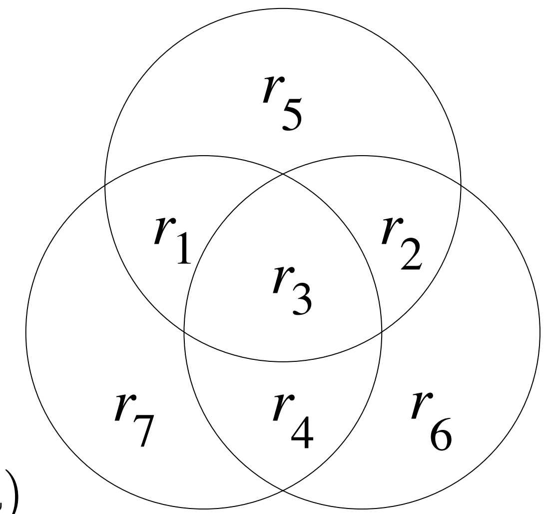
$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

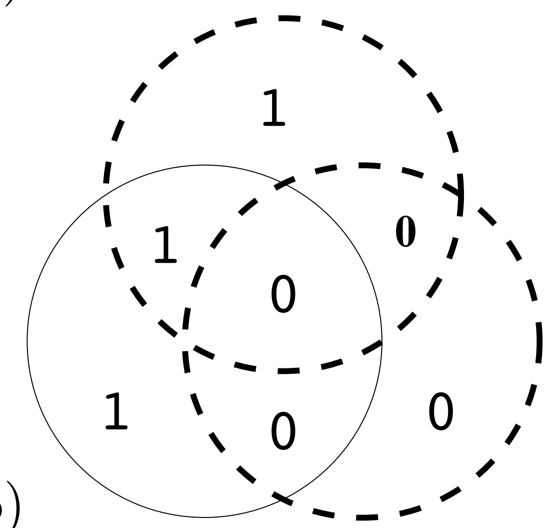
0

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(b)

$\boldsymbol{s}$ = 1000

$\boldsymbol{t}$ = 1000**101**

$\boldsymbol{n}$ = 0**1**00000

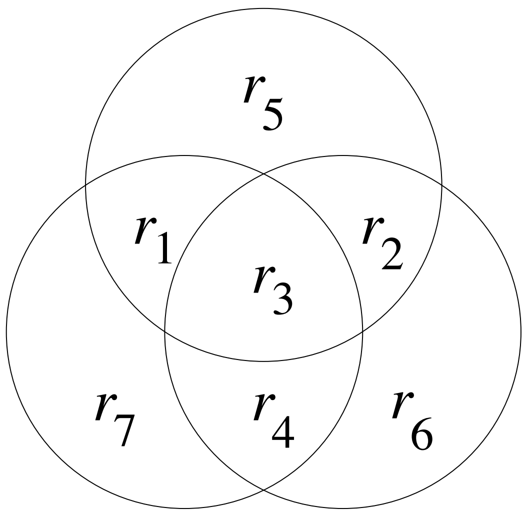
$\boldsymbol{r}$ = 1100101

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

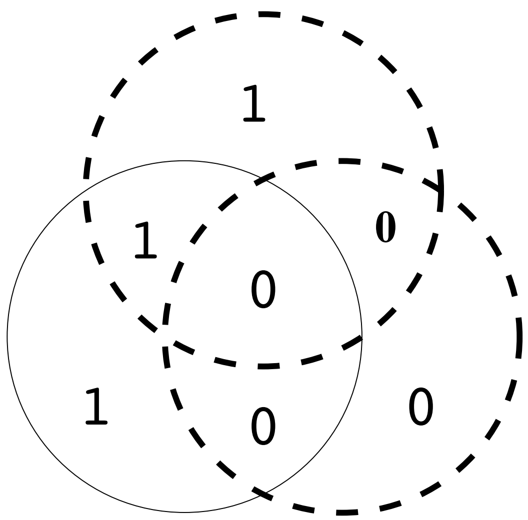
$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(b)

Which circles violate the parity check?

$\boldsymbol{s} = 1000$

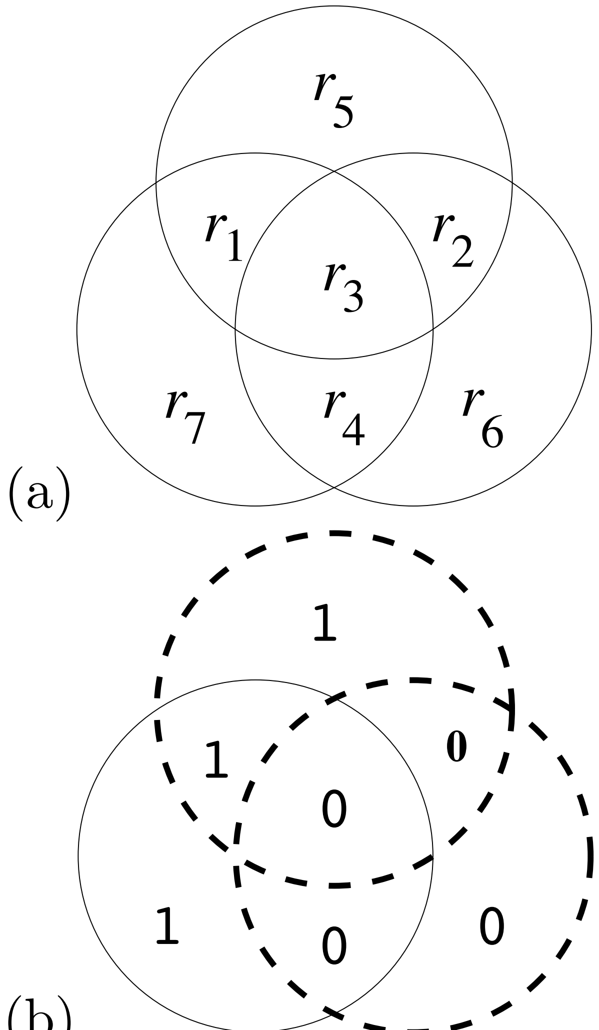
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

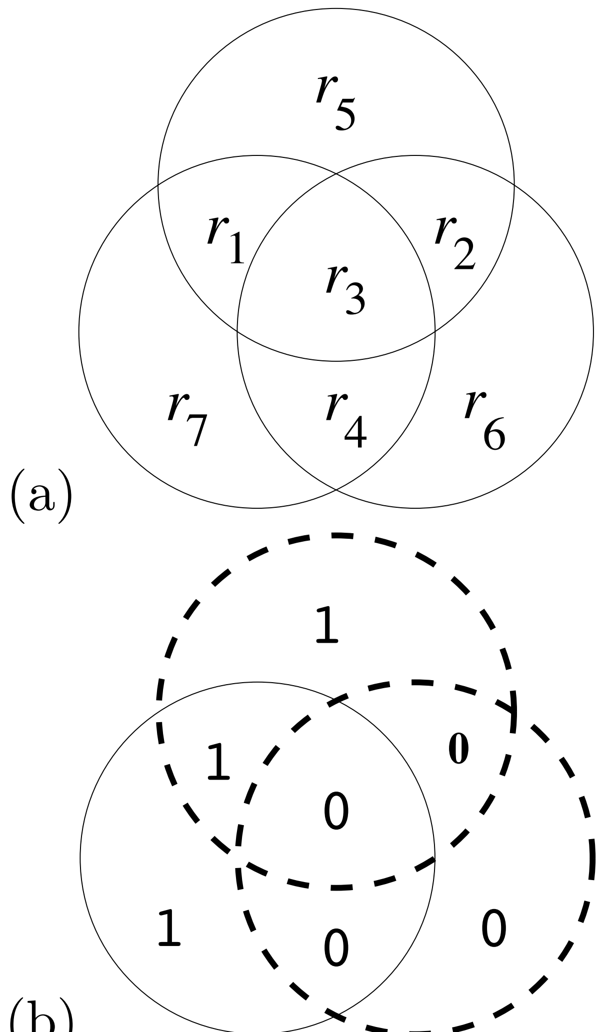
$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

$\boldsymbol{s} = 1000$

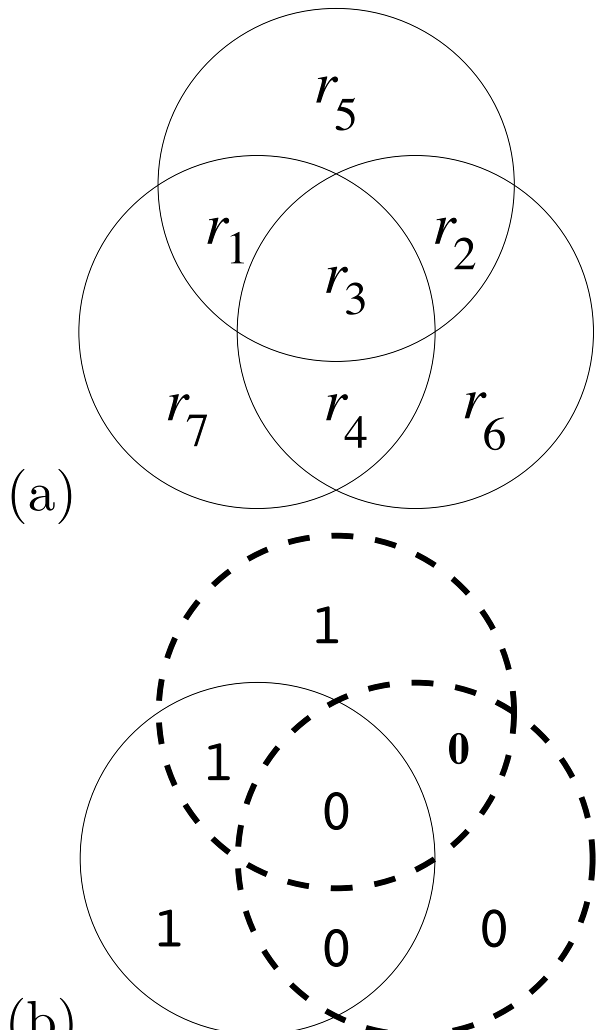
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

only r_2 ! (the flipped bit)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

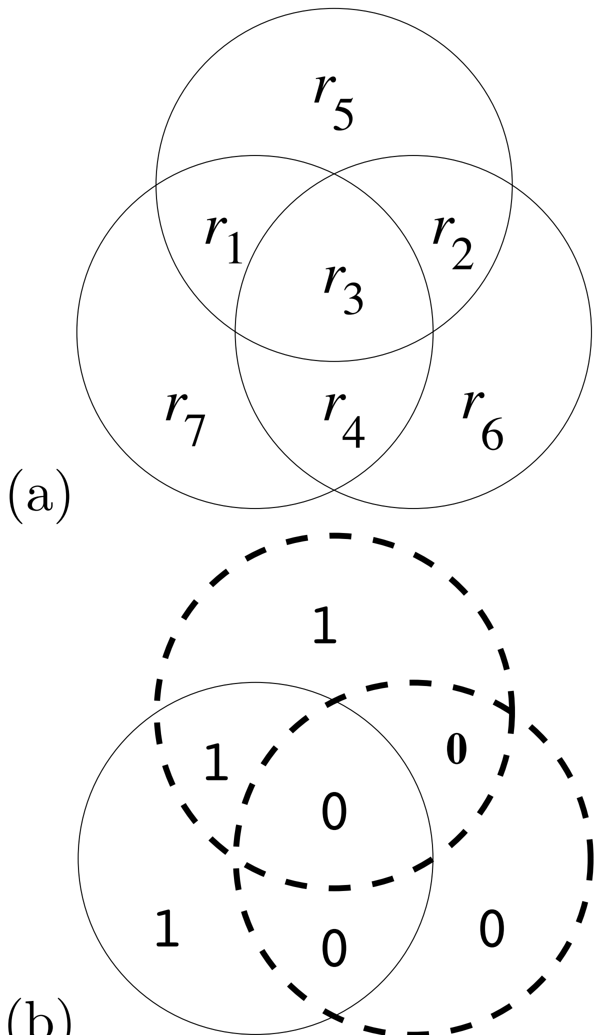
$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

only r_2 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

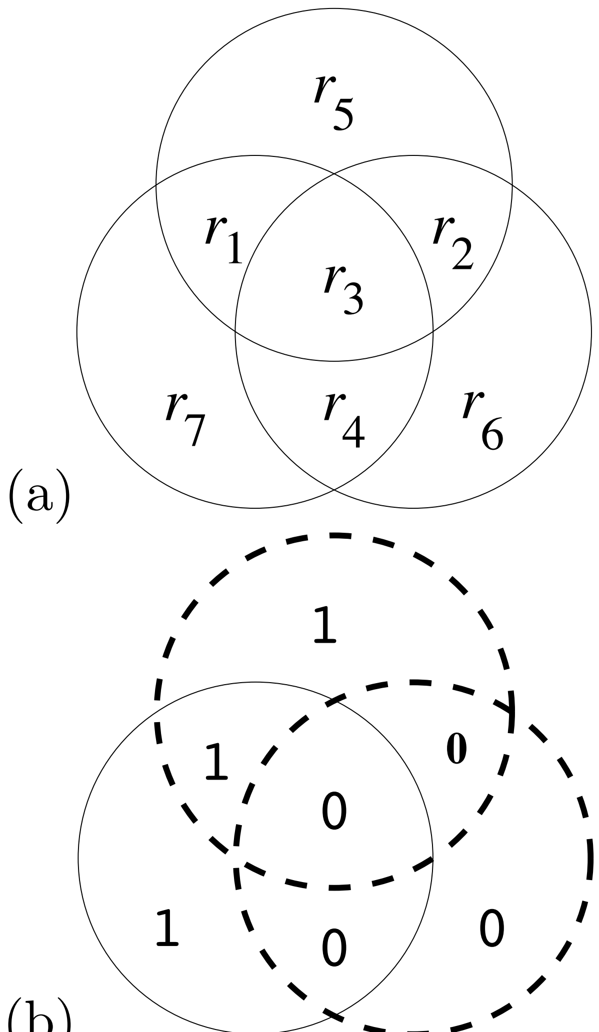
$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

The circles associated to t_5 and t_6

Which bits are involved in all circles with a violation?

only r_2 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

z = (110)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0\mathbf{1}00000$

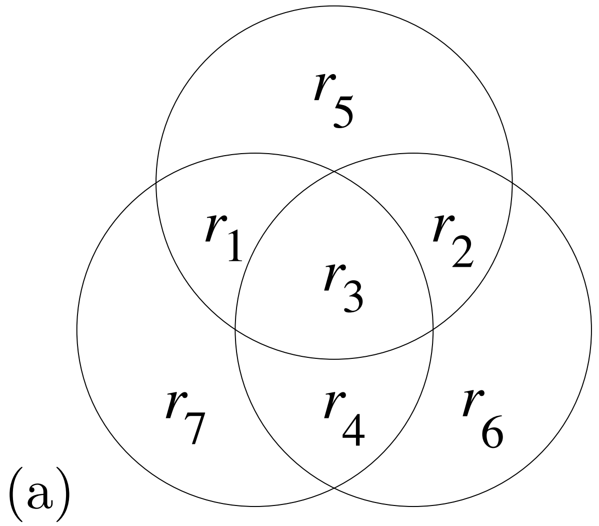
$\boldsymbol{r} = 1100101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



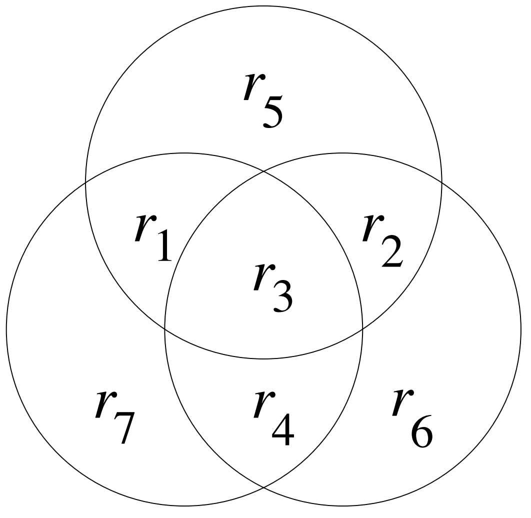
$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0000\mathbf{1}00$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)

$\boldsymbol{s}$ = 1000

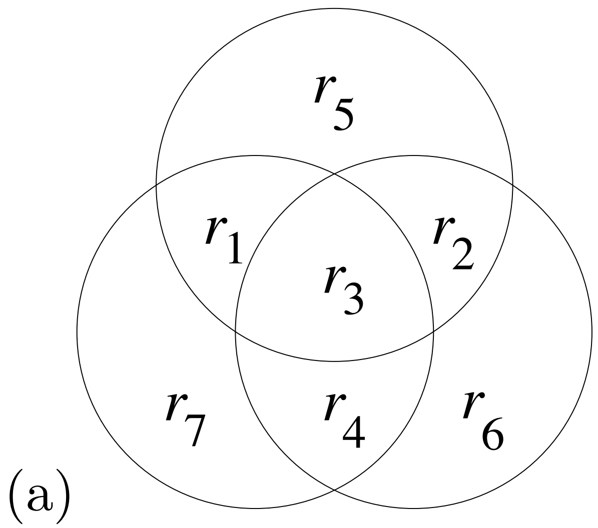
$\boldsymbol{t}$ = 1000**101**

$\boldsymbol{n}$ = 0000**1**00

$\boldsymbol{r}$ = 1000001

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



Which circles violate the parity check?

$\boldsymbol{s} = 1000$

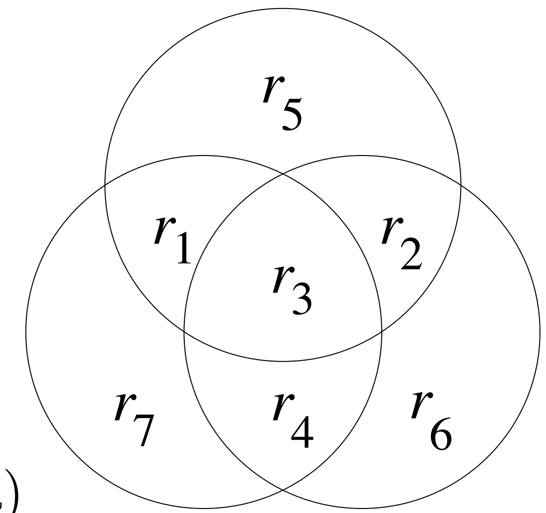
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0000\mathbf{1}00$

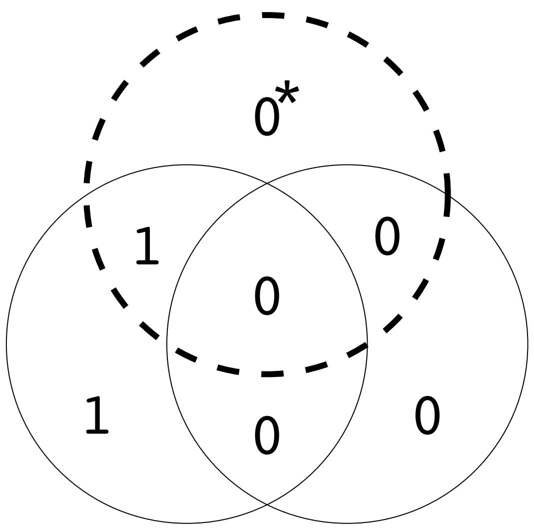
$\boldsymbol{r} = 1000001$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(c)

Which circles violate the parity check?

$\boldsymbol{s} = 1000$

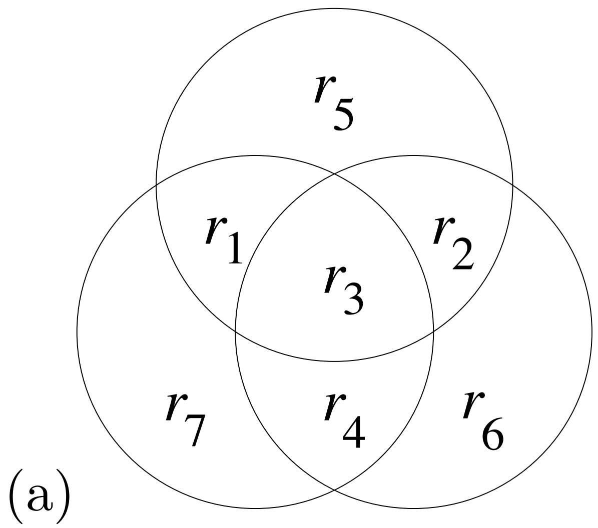
$\boldsymbol{t} = 1000\mathbf{101}$

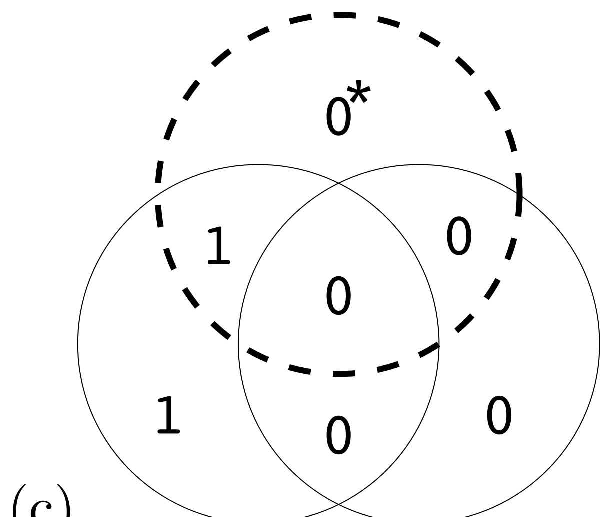
$\boldsymbol{n} = 0000\mathbf{1}00$

$\boldsymbol{r} = 1000001$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)





Which circles violate the parity check?

The circles associated to t_5

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0000\mathbf{1}00$

$\boldsymbol{r} = 1000001$

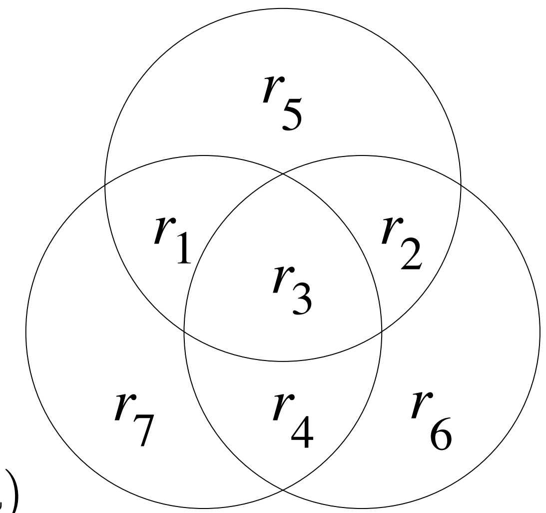
$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

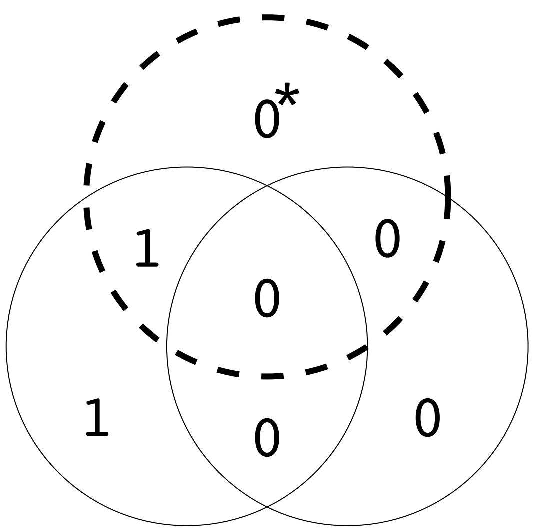
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Decoding the Hamming Code (7, 4)



(a)



(c)

Which circles violate the parity check?

The circles associated to t_5

Which bits are involved in all circles with a violation?

$\boldsymbol{s} = 1000$

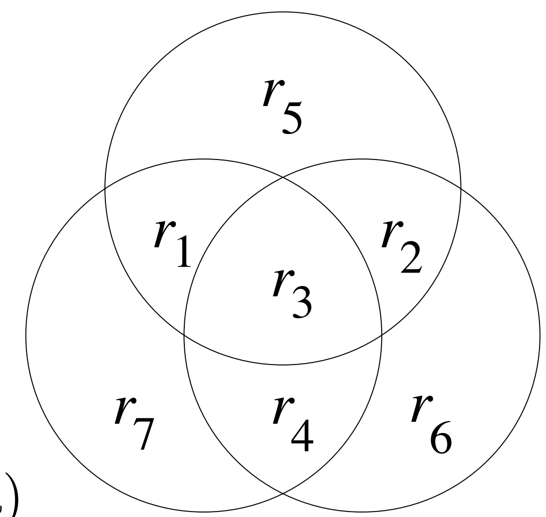
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0000\mathbf{1}00$

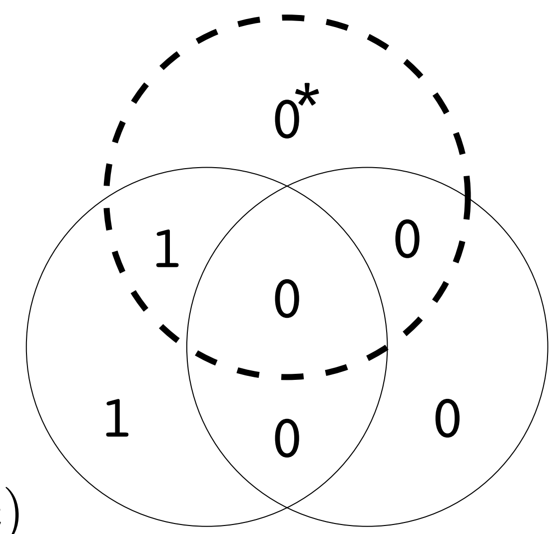
$\boldsymbol{r} = 1000001$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(c)

Which circles violate the parity check?

The circles associated to t_5

Which bits are involved in all circles with a violation?

only r_5 ! (the flipped bit)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0000\mathbf{1}00$

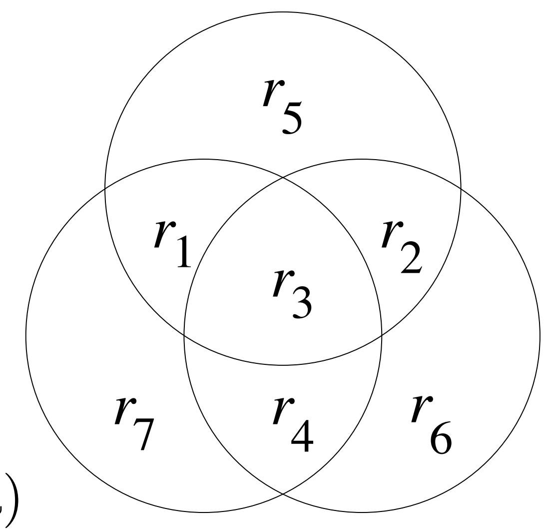
$\boldsymbol{r} = 1000001$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

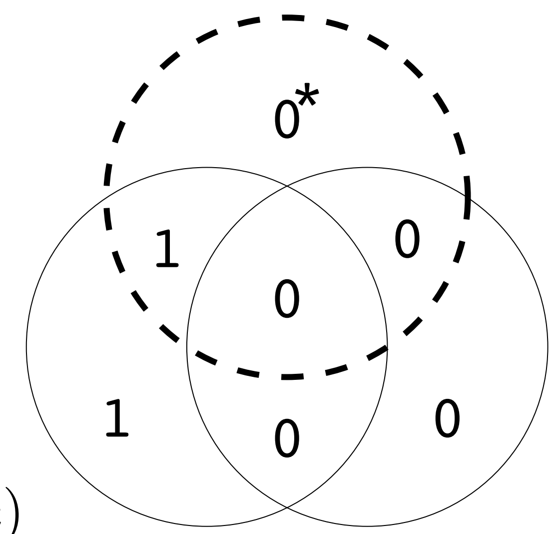
$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(c)

Which circles violate the parity check?

The circles associated to t_5

Which bits are involved in all circles with a violation?

only r_5 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

$\boldsymbol{s} = 1000$

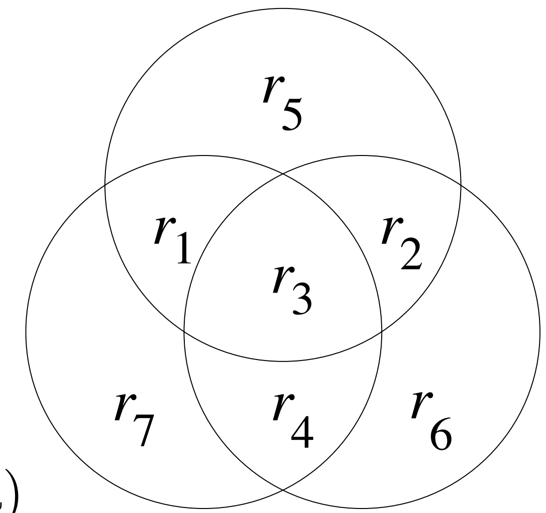
$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0000\mathbf{1}00$

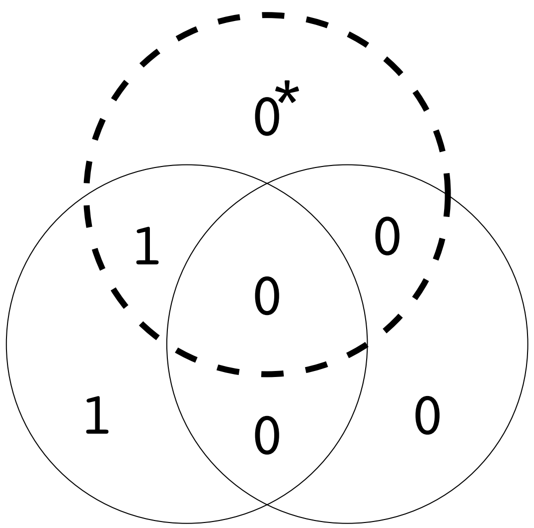
$\boldsymbol{r} = 1000001$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(c)

Which circles violate the parity check?

The circles associated to t_5

Which bits are involved in all circles with a violation?

only r_5 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

z = (100)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0000\mathbf{1}00$

$\boldsymbol{r} = 1000001$

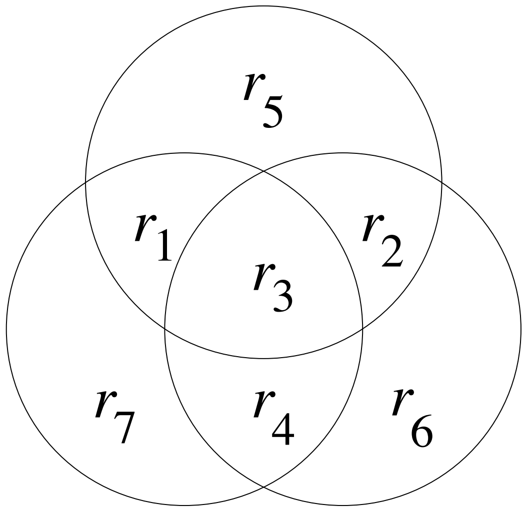
$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

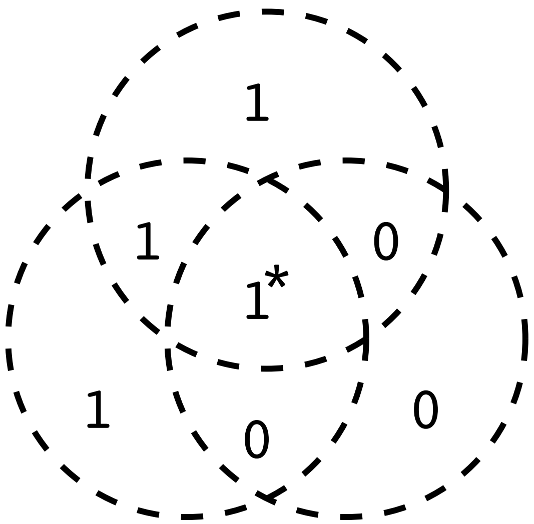
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Decoding the Hamming Code (7, 4)



(a)



(d)

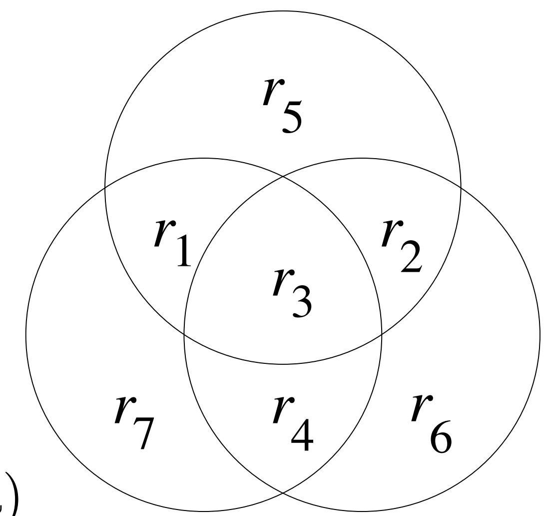
$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

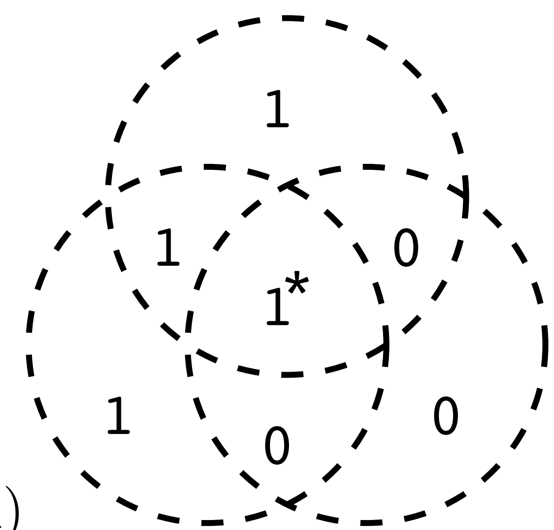
$\boldsymbol{n} = 00\mathbf{1}0000$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$
$$t_6 = s_2 \oplus s_3 \oplus s_4$$
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(d)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 00\mathbf{1}0000$

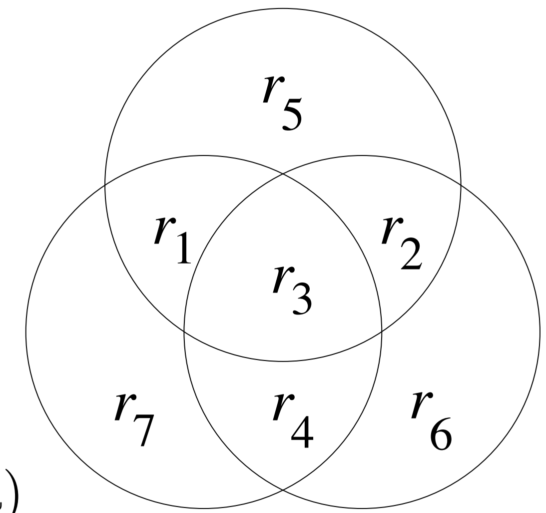
$\boldsymbol{r} = 1010101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

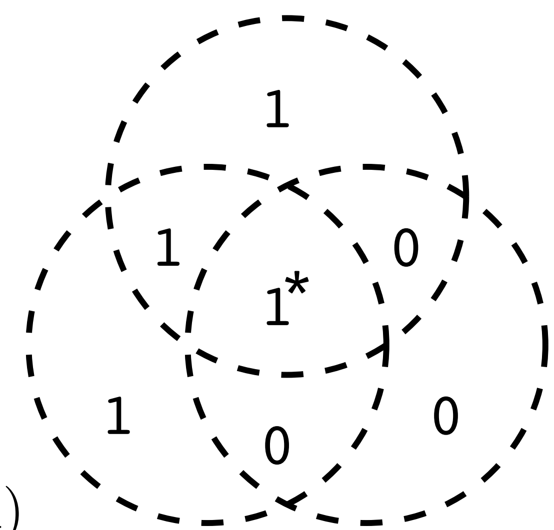
$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(d)

Which circles violate the parity check?

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 00\mathbf{1}0000$

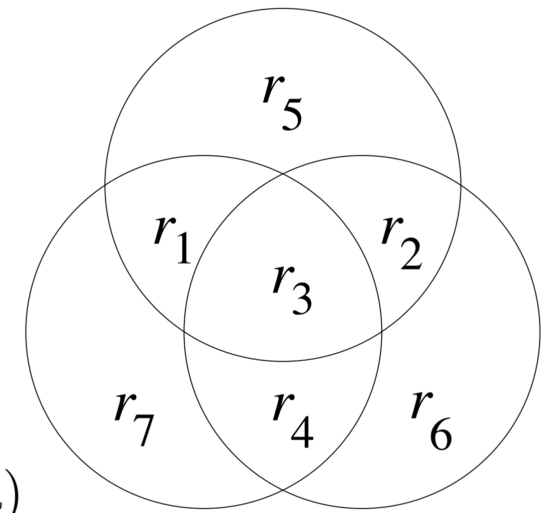
$\boldsymbol{r} = 1010101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

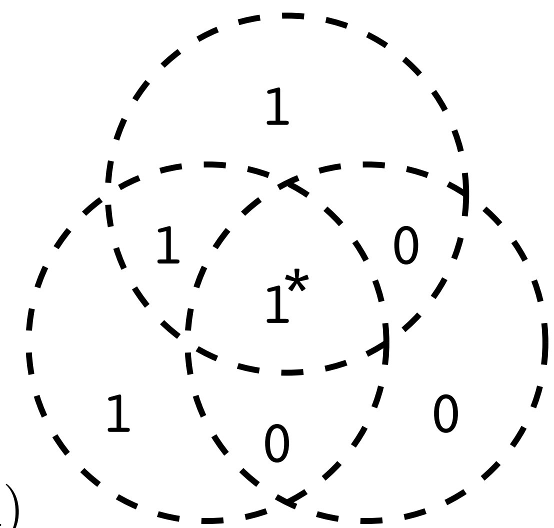
$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(d)

Which circles violate the parity check?

The circles associated to t_5, t_6, t_7.

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 0010000$

$\boldsymbol{r} = 1010101$

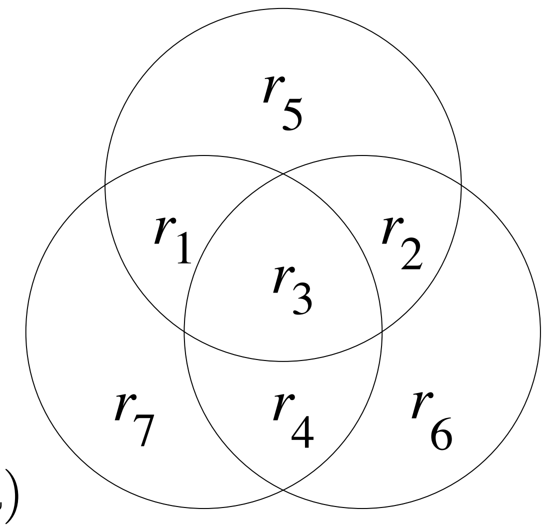
$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

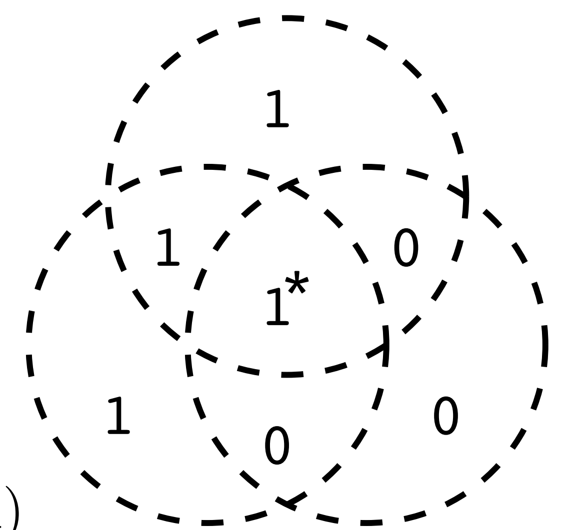
$$t_7 = s_1 \oplus s_3 \oplus s_4$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Decoding the Hamming Code (7, 4)



(a)



(d)

Which circles violate the parity check?

The circles associated to t_5, t_6, t_7.

Which bits are involved in all circles with a violation?

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 00\mathbf{1}0000$

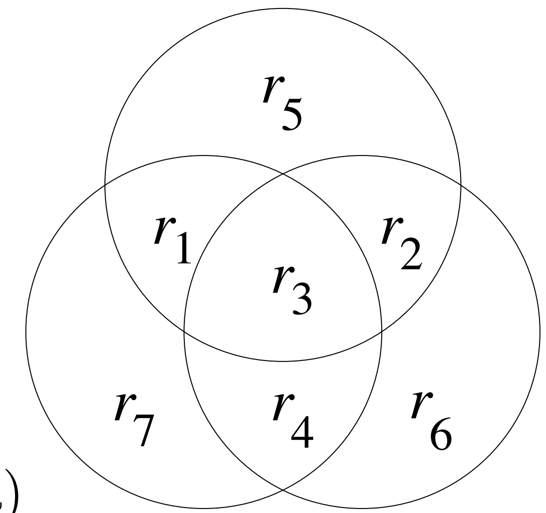
$\boldsymbol{r} = 1010101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

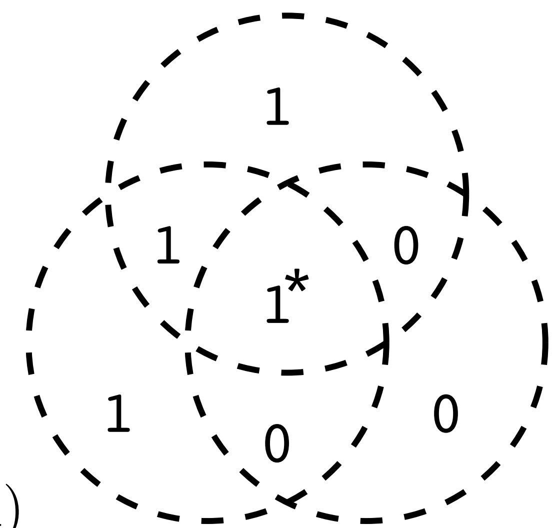
$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(d)

Which circles violate the parity check?

The circles associated to t_5, t_6, t_7.

Which bits are involved in all circles with a violation?

only r_3 ! (the flipped bit)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 00\mathbf{1}0000$

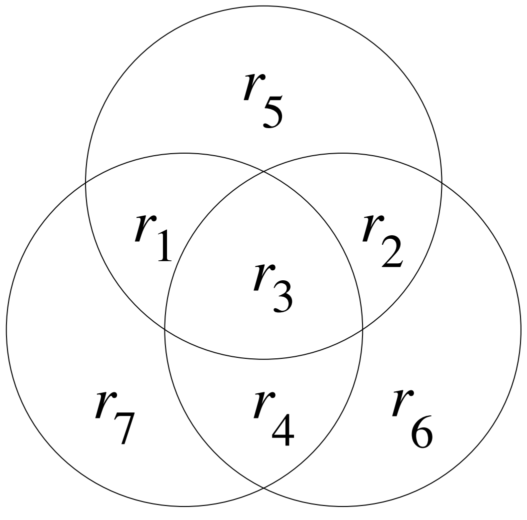
$\boldsymbol{r} = 1010101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

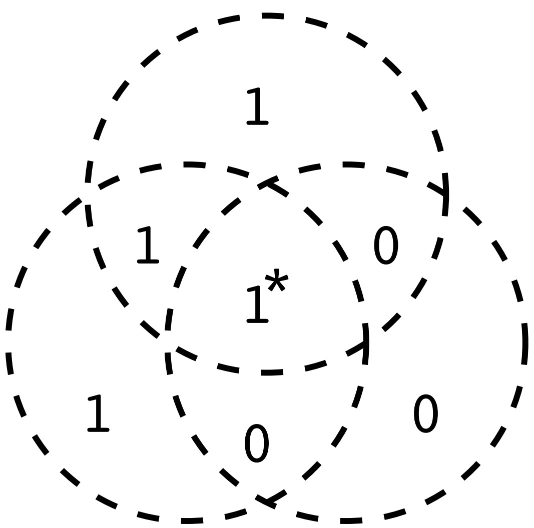
$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)



(a)



(d)

Which circles violate the parity check?

The circles associated to t_5, t_6, t_7.

Which bits are involved in all circles with a violation?

only r_3 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

$\boldsymbol{s}$ = 1000

$\boldsymbol{t}$ = 1000**101**

$\boldsymbol{n}$ = 00**1**0000

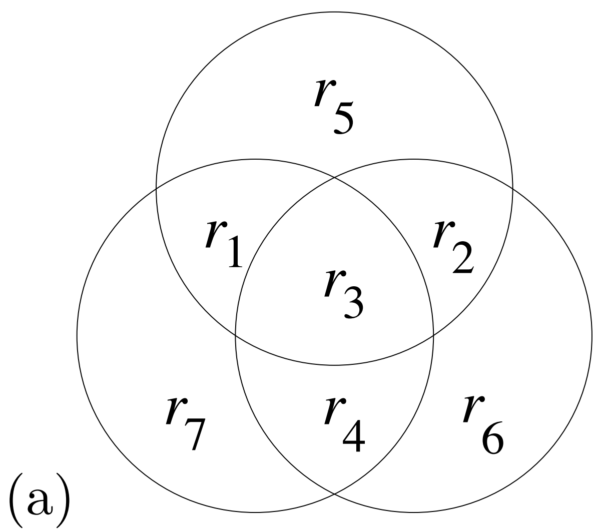
$\boldsymbol{r}$ = 1010101

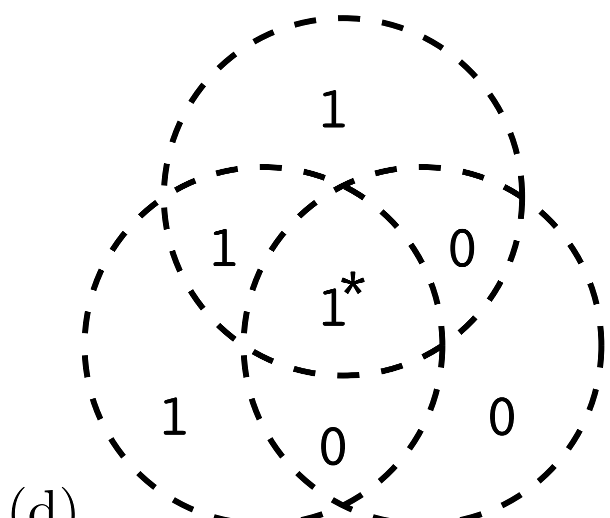
$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4)





Which circles violate the parity check?

The circles associated to t_5, t_6, t_7.

Which bits are involved in all circles with a violation?

only r_3 ! (the flipped bit)

The syndrome to this error is based on the parity of the circles

z = (111)

$\boldsymbol{s} = 1000$

$\boldsymbol{t} = 1000\mathbf{101}$

$\boldsymbol{n} = 00\mathbf{1}0000$

$\boldsymbol{r} = 1010101$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Decoding the Hamming Code (7, 4)

- Flipping any one of the seven bits, we get a different **syndrome** is obtained in each case.
  - Seven non-zero syndromes, one for each bit.
  - The all-zero syndrome

Syndrome $\mathbf{z}$	000	001	010	011	100	101	110	111
Unflip this bit	*none*	r_7	r_6	r_4	r_5	r_1	r_2	r_3

- The optimal decoder **unflips at most one bit**, depending on the syndrome.
- Any **two-bit error pattern** will lead to a decoded seven-bit **vector that contains three errors.**

Decoding the Hamming Code (7, 4): Matricial view

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

s and **t** as column vectors

$$\mathbf{t} = G^T \mathbf{s}$$

$$G^T = \begin{array}{c|cccc} & s_1 & s_2 & s_3 & s_4 \\ \hline t_1 & 1 & 0 & 0 & 0 \\ t_2 & 0 & 1 & 0 & 0 \\ t_3 & 0 & 0 & 1 & 0 \\ t_4 & 0 & 0 & 0 & 1 \\ t_5 & 1 & 1 & 1 & 0 \\ t_6 & 0 & 1 & 1 & 1 \\ t_7 & 1 & 0 & 1 & 1 \end{array}$$

s and **t** as row vectors

$$\mathbf{t} = \mathbf{s}G$$

$$G = \begin{array}{c|ccccccc} & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \\ \hline s_1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ s_2 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ s_3 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ s_4 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}$$

$$t_5 = s_1 \oplus s_2 \oplus s_3$$

$$t_6 = s_2 \oplus s_3 \oplus s_4$$

$$t_7 = s_1 \oplus s_3 \oplus s_4$$

Decoding the Hamming Code (7, 4): Matricial view

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

  s and **t** as column vectors

$$\boldsymbol{t} = G^T \boldsymbol{s}$$

$$G^T = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{array} \begin{array}{c} \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right] \end{array}$$

Decoding the Hamming Code (7, 4): Matricial view

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

  - **s** and **t** as column vectors

$$\boldsymbol{t} = G^T \boldsymbol{s}$$

$$G^T = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{array} \begin{array}{c} s_1 \quad s_2 \quad s_3 \quad s_4 \\ \left[\begin{array}{c} \\ \mathrm{I}_4 \\ \\ \hline \\ \mathrm{P} \\ \\ \end{array} \right] \end{array}$$

Decoding the Hamming Code (7, 4): Matricial view

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

  s and **t** as column vectors

$$\boldsymbol{t} = G^T \boldsymbol{s}$$

$$G^T = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{array} \begin{array}{c} s_1 \quad s_2 \quad s_3 \quad s_4 \\ \left[\begin{array}{c} \\ \mathrm{I}_4 \\ \\ \hline \\ \mathrm{P} \\ \\ \end{array} \right] \end{array}$$

$$H = \begin{bmatrix} P & I_3 \end{bmatrix} = \begin{bmatrix} \mathrm{P} & \mathrm{I}_3 \end{bmatrix}$$

Decoding the Hamming Code (7, 4): Matricial view

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

s and **t** as column vectors

$$\boldsymbol{t} = G^T \boldsymbol{s}$$

$$G^T = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{array} \begin{array}{c} \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} \\ \left[\begin{array}{c} \mathrm{I}_4 \\ \hline \mathrm{P} \end{array} \right] \end{array}$$

$$H = \begin{bmatrix} P & I_3 \end{bmatrix} = \begin{bmatrix} \mathrm{P} & \mathrm{I}_3 \end{bmatrix}$$

The syndrome **z** is calculated with the received sequence and H

$$\boldsymbol{z} = H\boldsymbol{r}$$

Decoding the Hamming Code (7, 4): Matricial view

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

s and **t** as column vectors

$$\boldsymbol{t} = G^T \boldsymbol{s}$$

$$G^T = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{array} \begin{array}{c} s_1 \quad s_2 \quad s_3 \quad s_4 \\ \left[\begin{array}{c} \\ \mathrm{I}_4 \\ \\ \\ \hline \\ \mathrm{P} \\ \\ \end{array} \right] \end{array}$$

$$H = \left[\begin{array}{cc} P & I_3 \end{array} \right] = \left[\begin{array}{c|c} \mathrm{P} & \mathrm{I}_3 \end{array} \right]$$

The syndrome **z** is calculated with the received sequence and H

$$\boldsymbol{z} = H\boldsymbol{r}$$

All the codewords t of the code satisfy

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = H\boldsymbol{t}$$

Decoding the Hamming Code (7, 4): Matricial view

- Because the Hamming code is a linear code, it can be written compactly in terms of matrices

  s and **t** as column vectors

$$\boldsymbol{t} = G^T \boldsymbol{s}$$

$$G^T = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{array} \begin{array}{c} \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right] \end{array}$$

$$H = \begin{bmatrix} P & I_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

The syndrome **z** is calculated with the received sequence and H

$$\boldsymbol{z} = H\boldsymbol{r}$$

All the codewords t of the code satisfy

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = H\boldsymbol{t}$$

Error Probability of H(7, 4)

- A **decoding error** will occur whenever the noise has **flipped more than one bit** in a block of seven.

Error Probability of H(7, 4)

- A **decoding error** will occur whenever the noise has **flipped more than one bit** in a block of seven.
- The **probability of block error** is thus the **probability that two or more bits are flipped** in a block

$$P_B = \sum_{r=2}^{7} \binom{7}{r} f^r (1-f)^{7-r}$$

Error Probability of H(7, 4)

- A **decoding error** will occur whenever the noise has **flipped more than one bit** in a block of seven.
- The **probability of block error** is thus the **probability that two or more bits are flipped** in a block

$$P_B = \sum_{r=2}^{7} \binom{7}{r} f^r (1-f)^{7-r}$$

- The **probability of bit error** (for the source bits) is simply three sevenths of the probability of block error.

$$P_b = \frac{3}{7} P_B$$

Error Probability of H(7, 4)

- A **decoding error** will occur whenever the noise has **flipped more than one bit** in a block of seven.
- The **probability of block error** is thus the **probability that two or more bits are flipped** in a block

$$P_B = \sum_{r=2}^{7} \binom{7}{r} f^r (1-f)^{7-r}$$

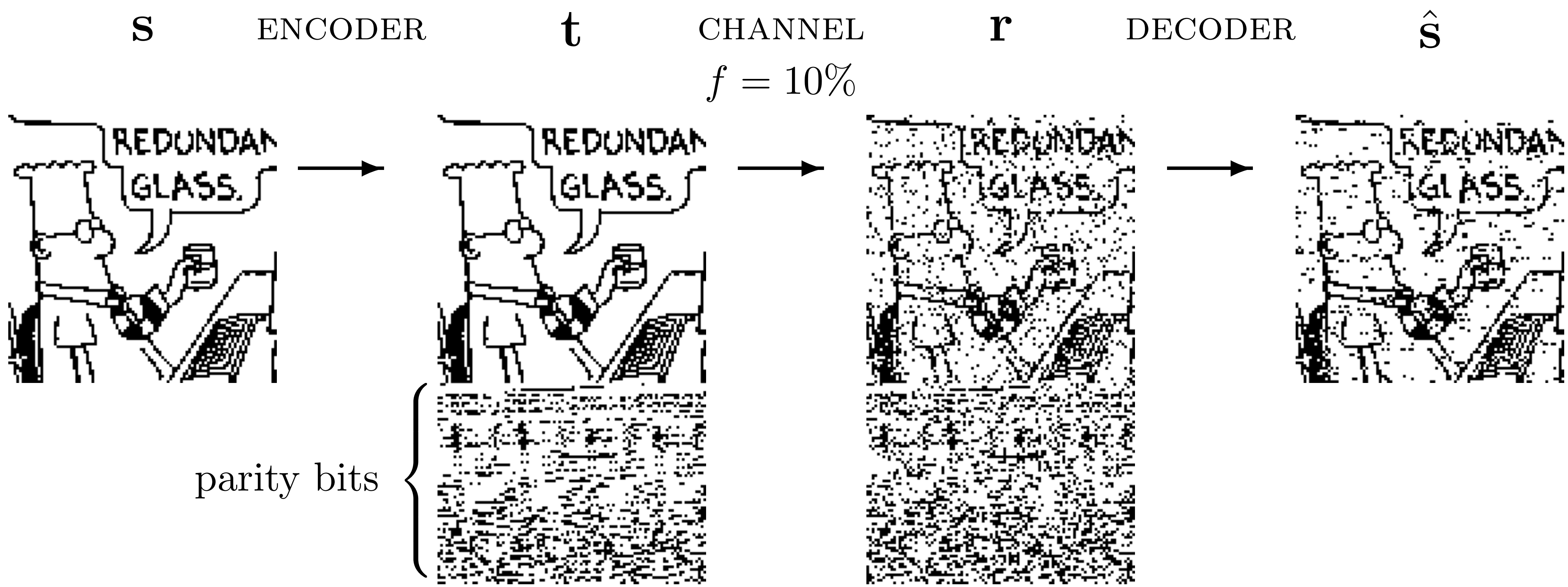
- The **probability of bit error** (for the source bits) is simply three sevenths of the probability of block error.

$$P_b = \frac{3}{7} P_B$$

- The Hamming code communicates at a **rate**, R = 4/7.

Decoding the Hamming Code (7, 4)

- The probability of decoded bit error is about 7% (*)



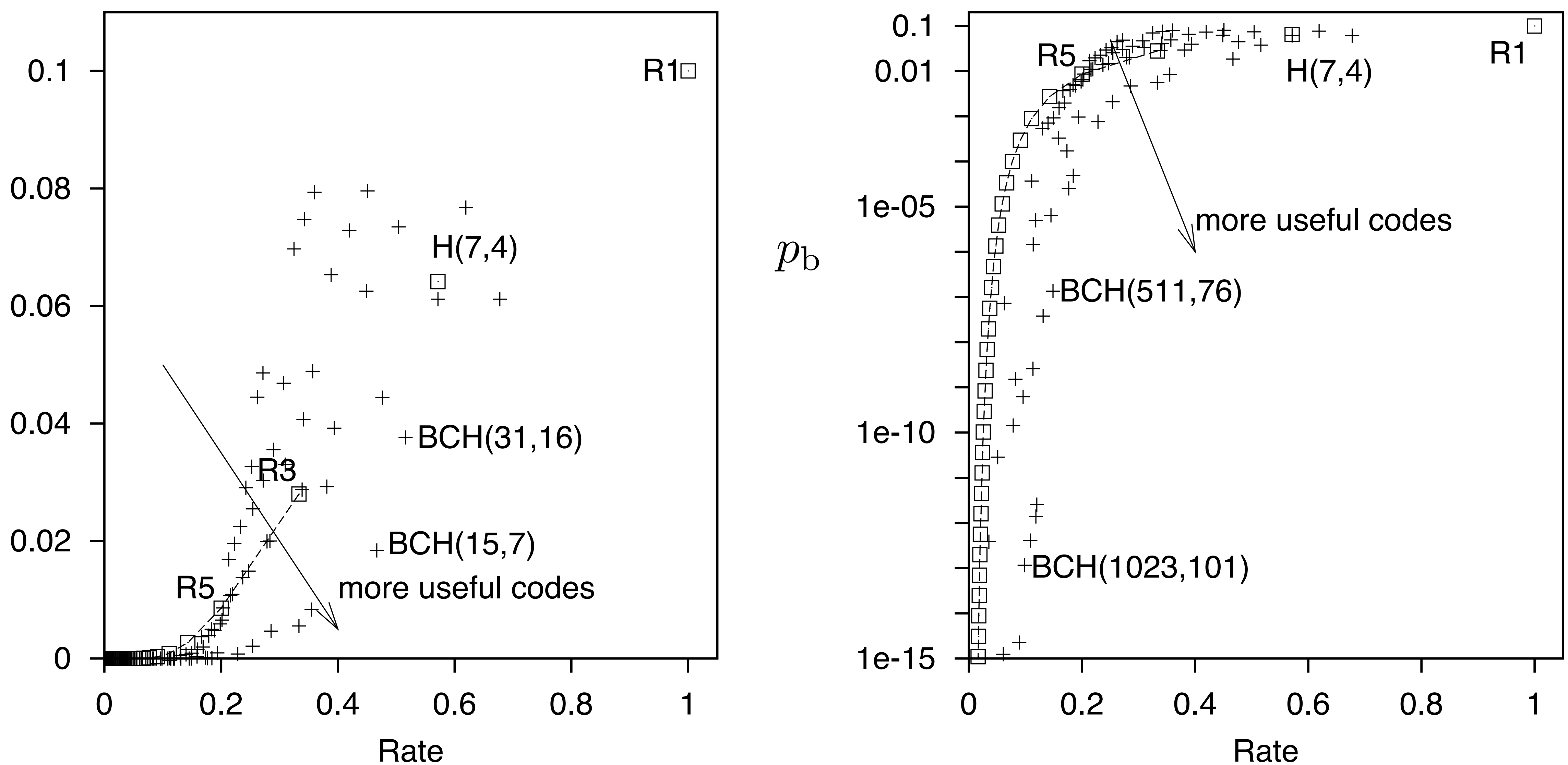
(*) Transmitting 10 000 source bits over a binary symmetric channel with f = 10% using a (7, 4) Hamming code. The probability of decoded bit error is about 7%.

What performance can the best codes achieve?

Code's performances

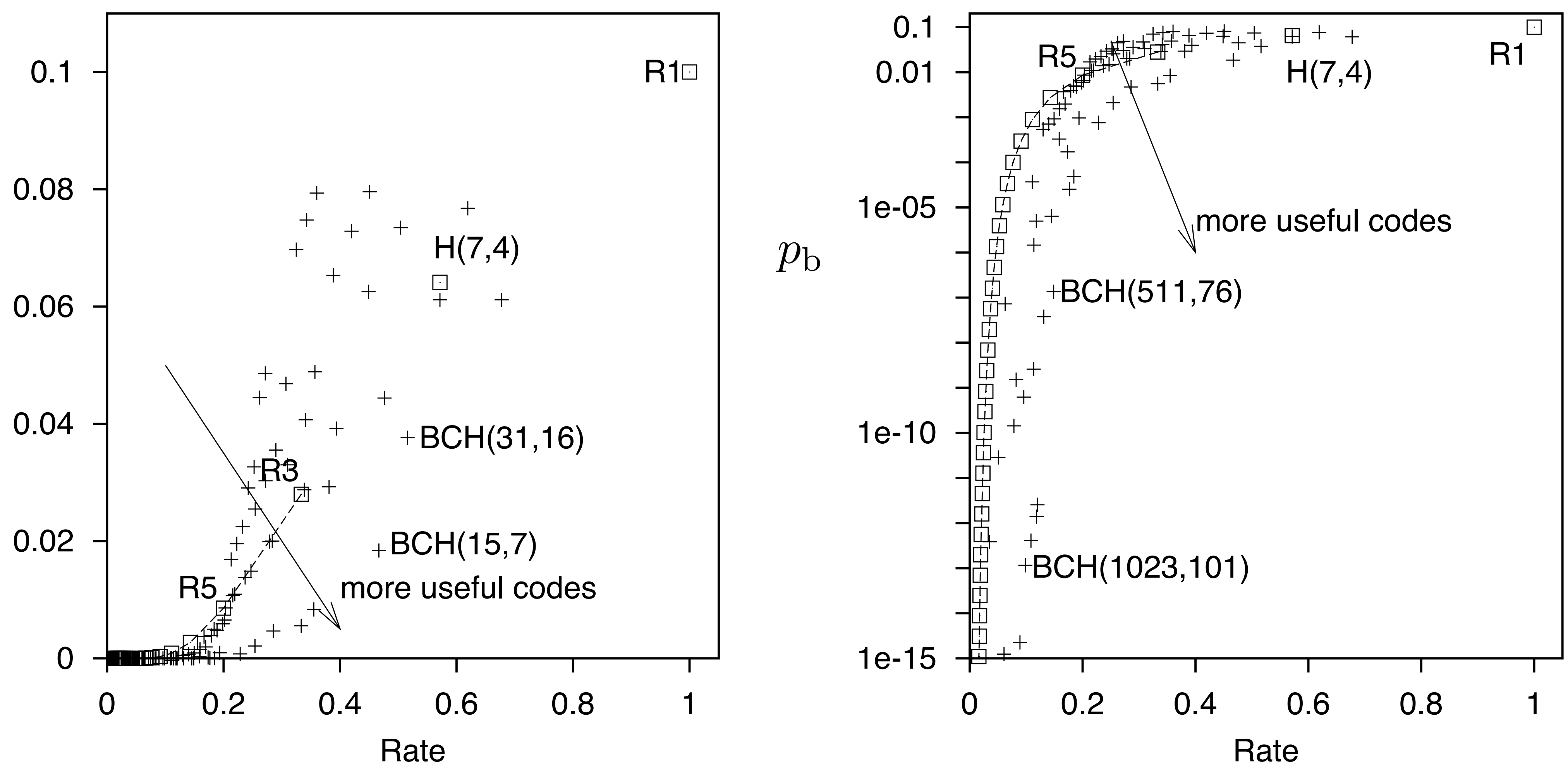
- Error probability P_b versus rate R for repetition codes, H(7,4) and BCH codes (generalization of Hamming codes)

Over a binary symmetric channel with $f = 0.1$



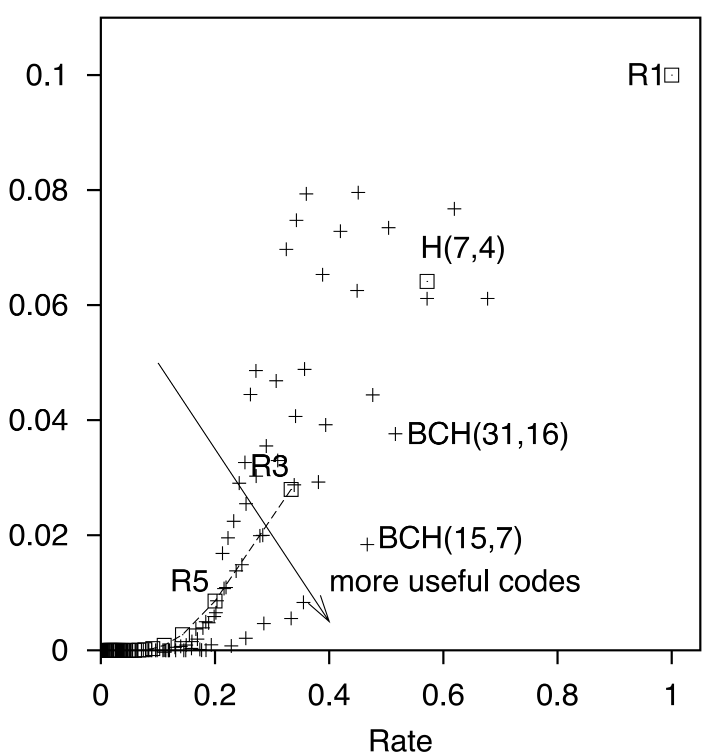
Code's performances

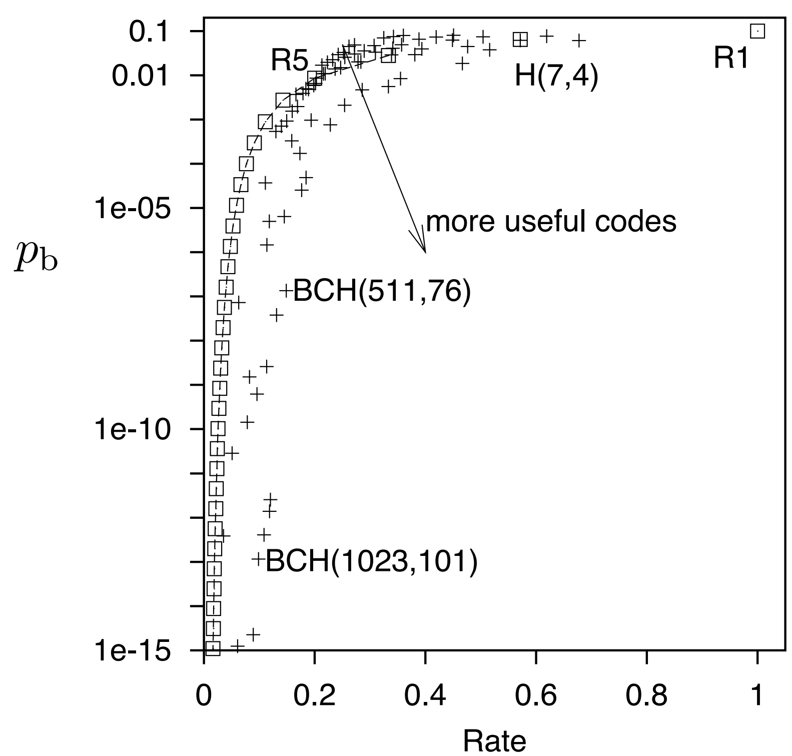
- Linear codes look better than repetition codes
- But in all cases it **looks that we need a rate near to zero to get a very small error probability**



What performance can the best codes achieve?

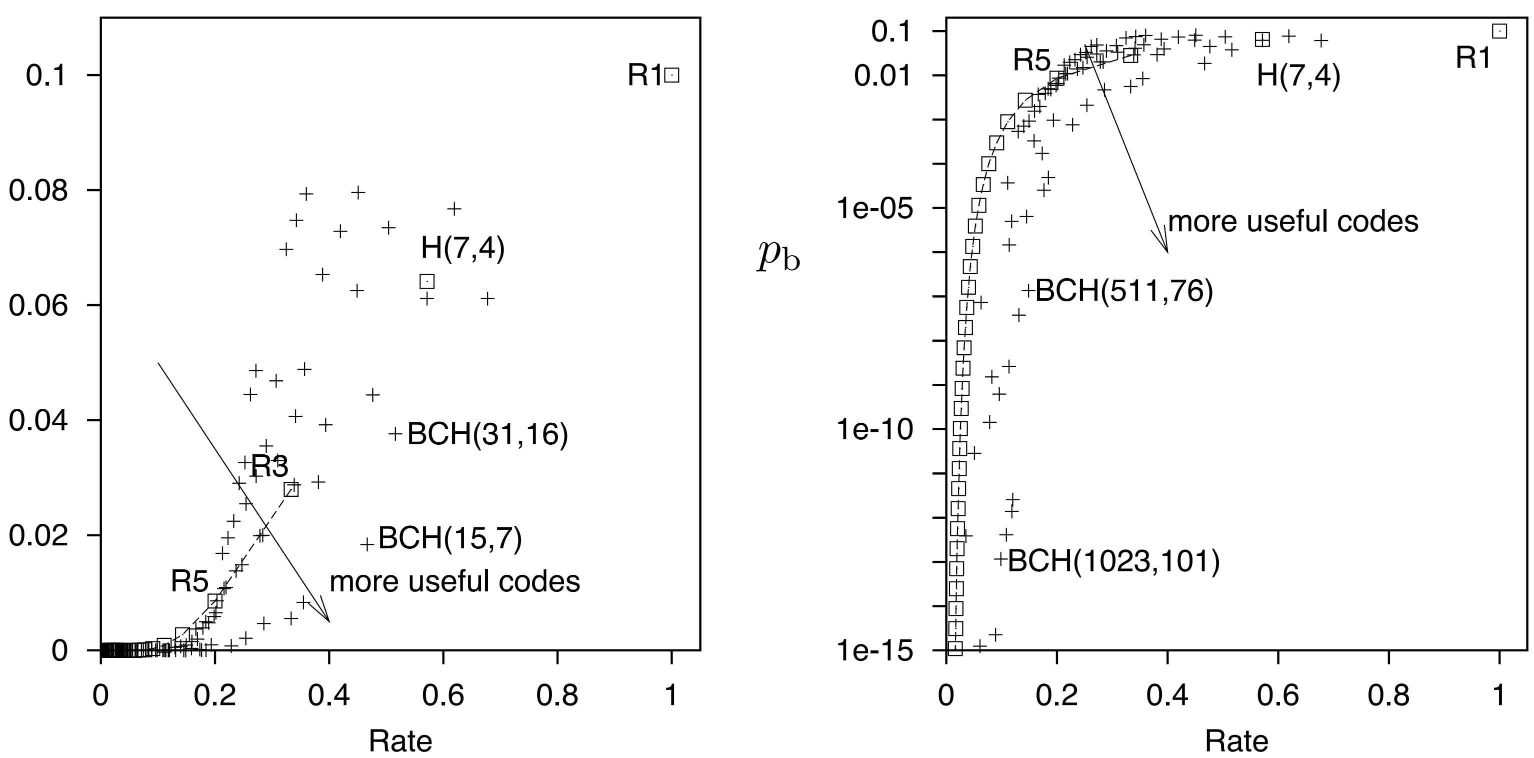
- Goals:
  - Reduce the decoded bit-error probability P_b
  - We would like to keep the rate R large.
- What points in the (R, P_b) plane are **achievable**?





What performance can the best codes achieve?

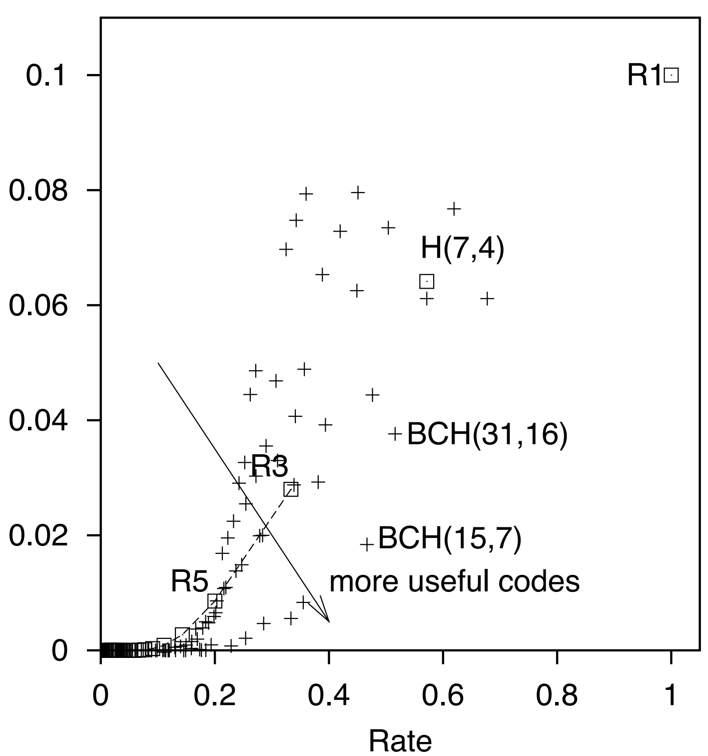
- Goals:
  - Reduce the decoded bit-error probability P_b
  - We would like to keep the rate R large.
- What points in the (R, P_b) plane are **achievable**?

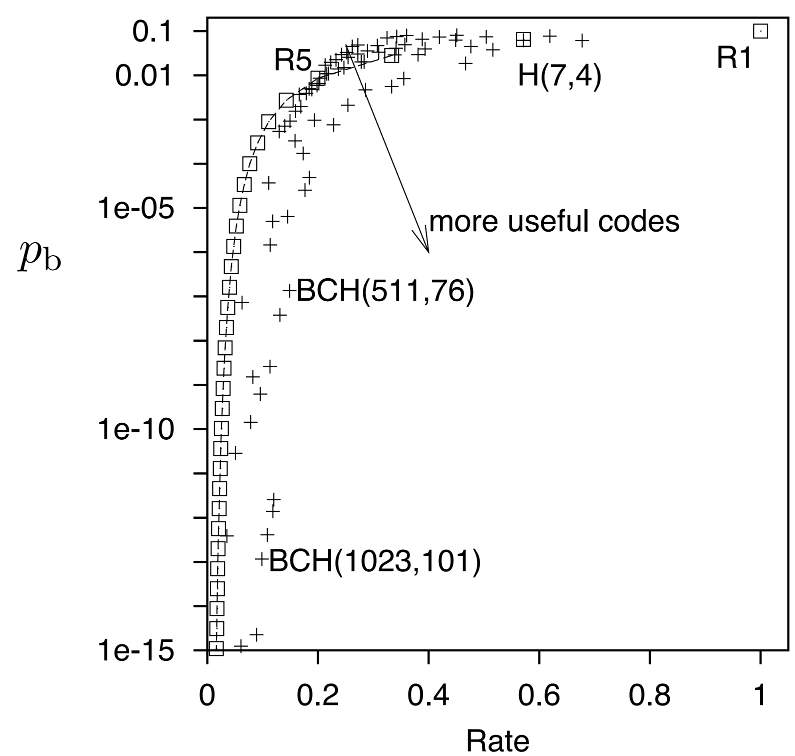


A Mathematical Theory of Communication, **Claude Shannon**, 1948

What performance can the best codes achieve?

- Goals:
  - Reduce the decoded bit-error probability P_b
  - We would like to keep the rate R large.
- What points in the (R, P_b) plane are **achievable**?



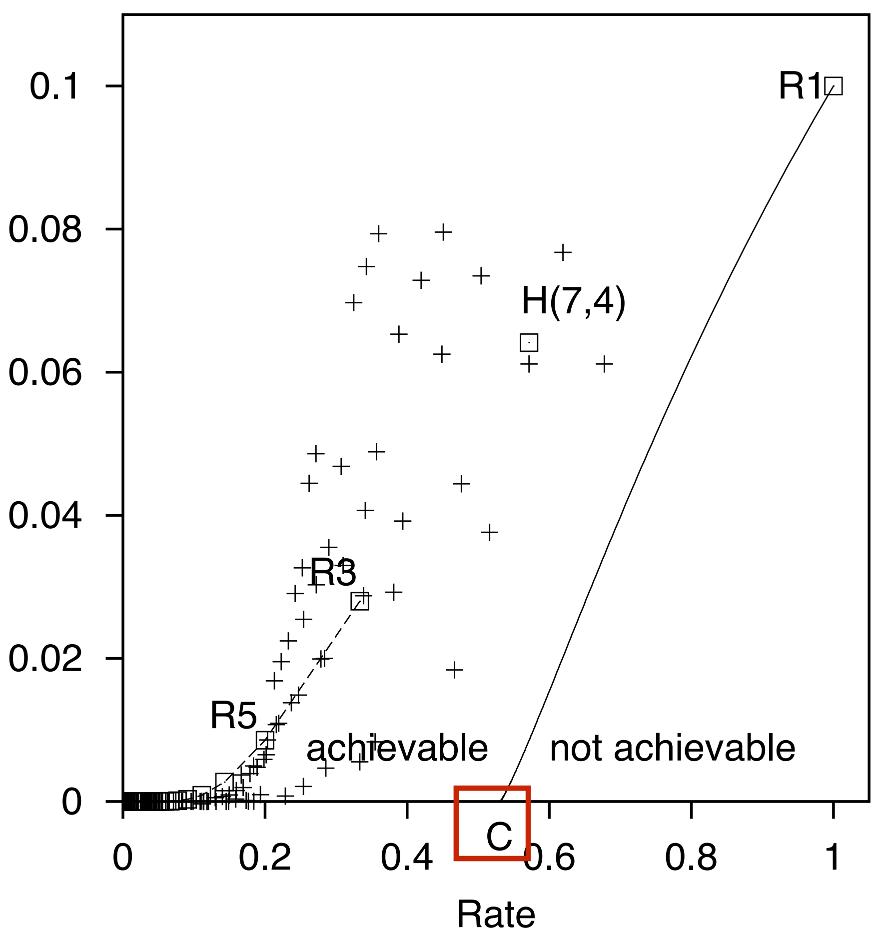


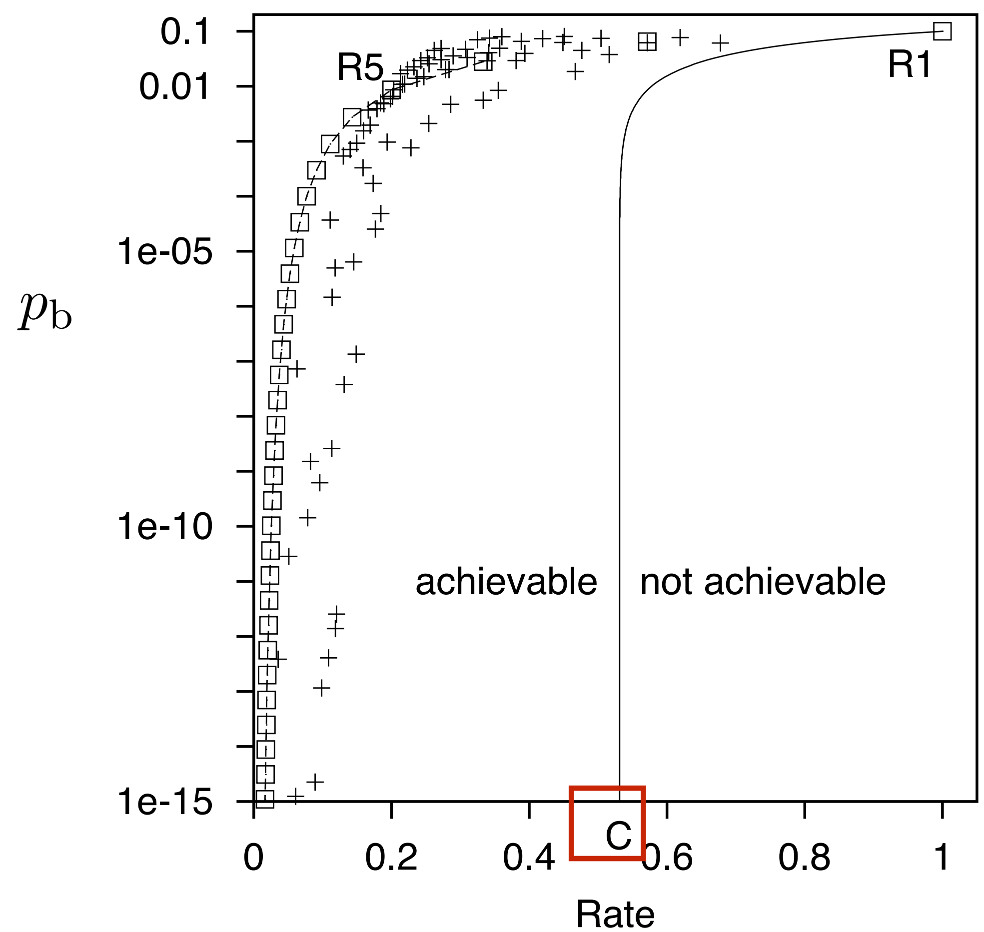
A Mathematical Theory of Communication, **Claude Shannon**, 1948

Created the field of Information Theory and solved most of its fundamental problems.

What performance can the best codes achieve?

- The widespread belief that the boundary between achievable and nonachievable points in the (R, P_b) plane was **a curve passing through the origin $(R, P_b) = (0,0)$**
- Shannon proved that the boundary between achievable and nonachievable points **meets the R axis at a non-zero value $R = C$**



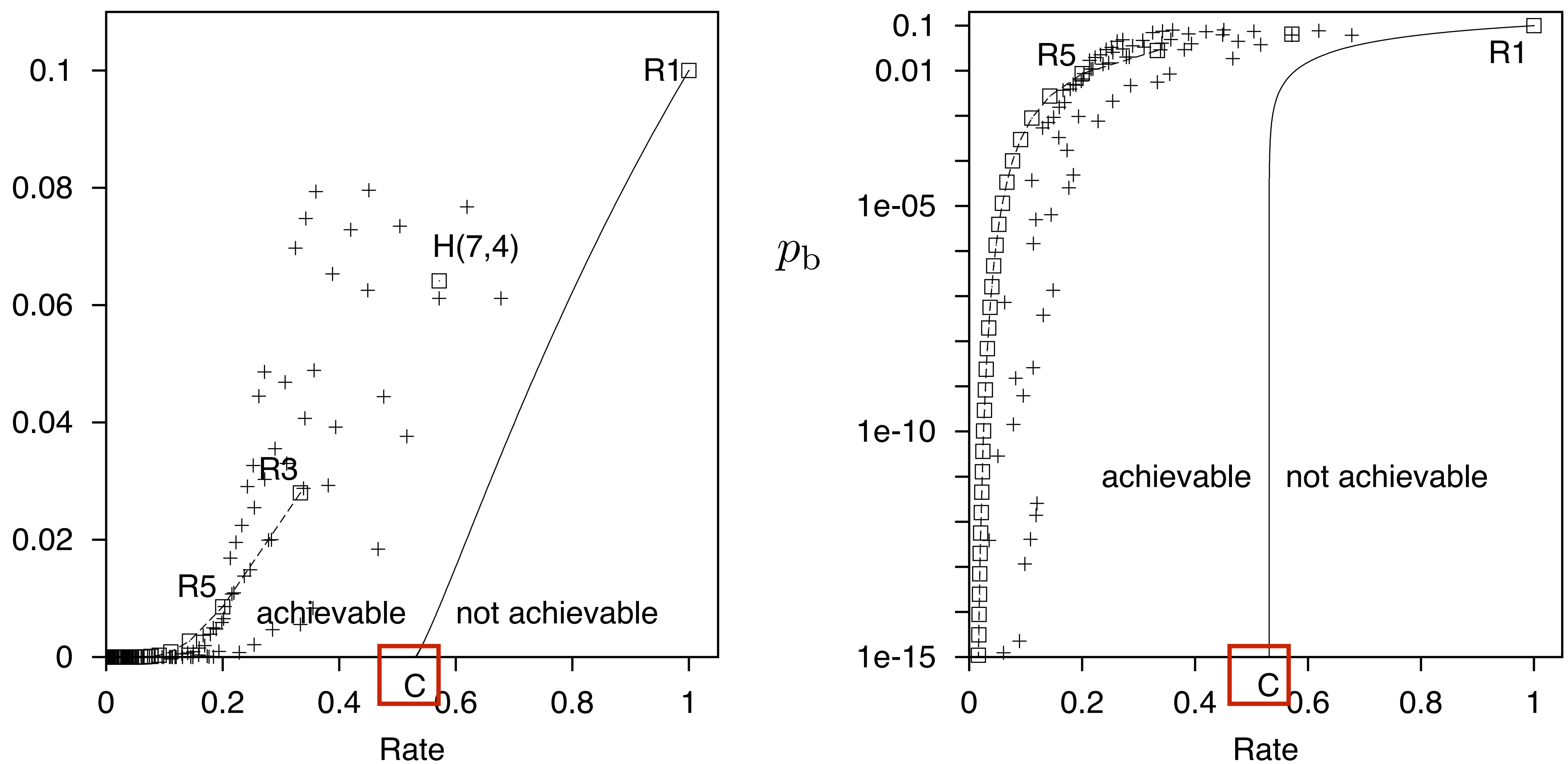


For BSC with $f = 0.1$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

What performance can the best codes achieve?

- Shannon proved that the boundary between achievable and nonachievable points **meets the *R* axis at a non-zero value *R* = *C***
- For any channel, **there exist codes** that make it possible to communicate with **arbitrarily small probability of error P_b at non-zero rates**.



For BSC with $f = 0.1$

What performance can the best codes achieve?

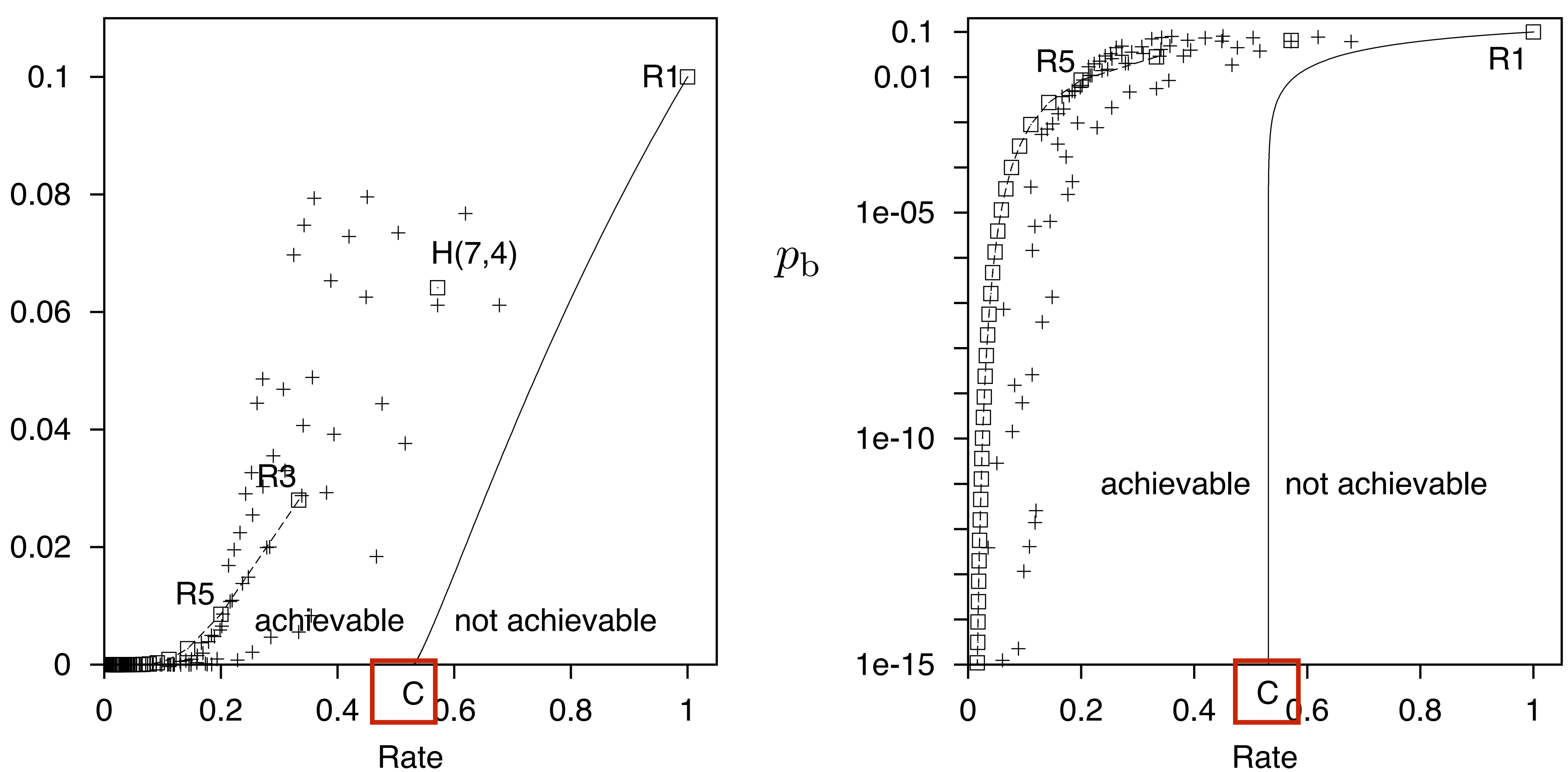
- C is the channel ***capacity***

$$C(f) = 1 - H_2(f) = 1 - \left[f \log_2 \tfrac{1}{f} + (1-f)\log_2 \tfrac{1}{1-f} \right]$$

- and the curve separating the regions

$$R = C / (1 - H_2(p_b))$$

For BSC with $f = 0.1$ $C = 0.53$



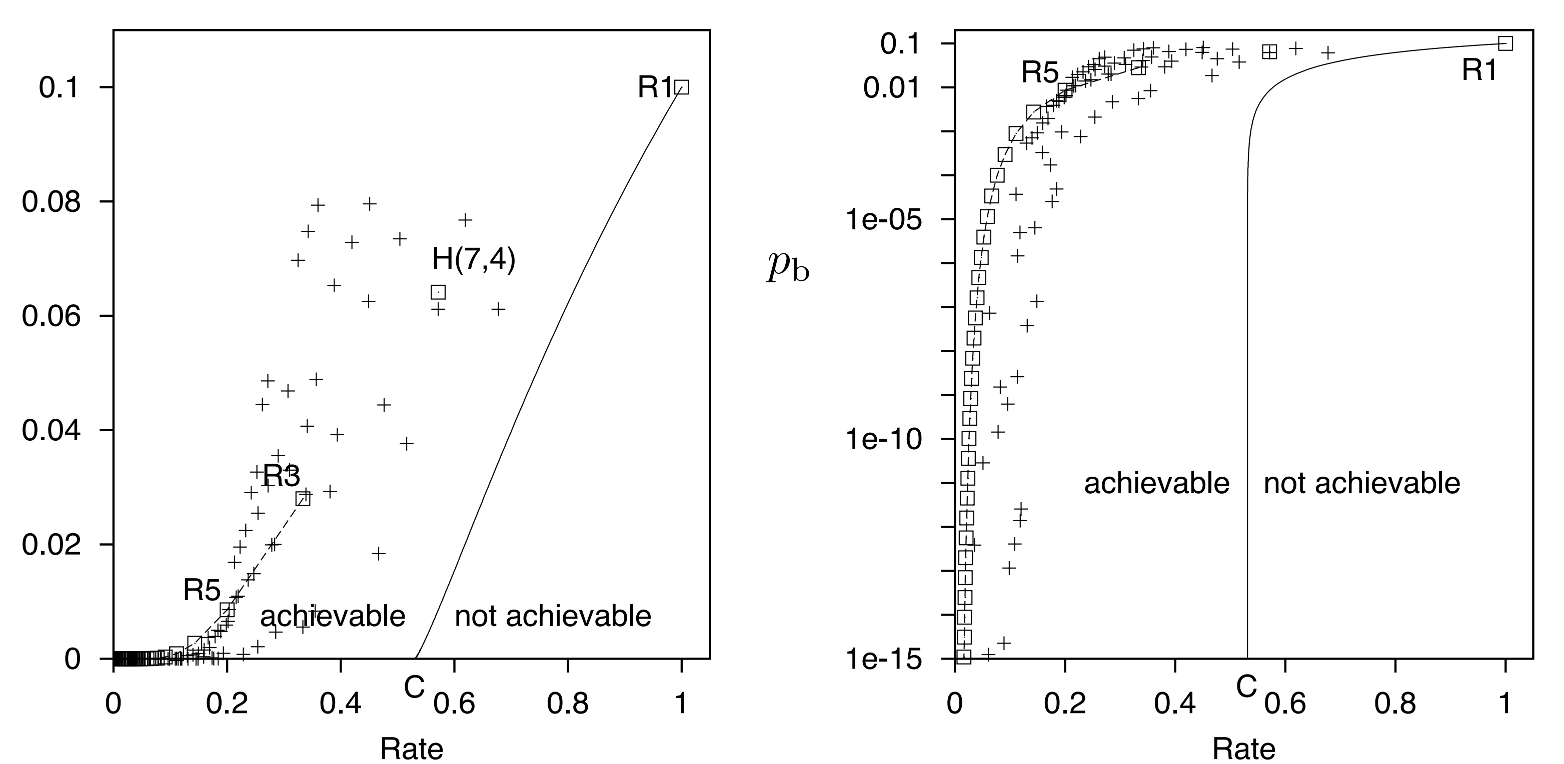
What is the impact?

- Consider a (noisy) disk drive with $f = 0.1$
  - The code R_3 could communicate over this channel with $P_b = 0.03$ at a rate $R = 1/3$
  - The code R_3 could communicate over this channel with $P_b \approx 10^{-15}$ at a rate $R = 1/60$

What is the impact?

- Consider a (noisy) disk drive with $f = 0.1$
  - The code R_3 could communicate over this channel with $P_b = 0.03$ at a rate $R = 1/3$
  - The code R_3 could communicate over this channel with $P_b \approx 10^{-15}$ at a rate $R = 1/60$
- According to Shannon you don't need 60 disks to get a performance of $P_b \approx 10^{-15}$.

  You can get that performance with just 2 disks ! (0.5 < 0.53)



FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Further Reading and Summary

Q&A

Further Reading

- **Recommend Readings**
  - Information Theory, Inference, and Learning Algorithms from David MacKay, 2015, pages 1 - 16.
- **Supplemental readings:**
  - The introduction of “A Mathematical Theory of Communication, **Claude Shannon**, 1948”, pages 1-2.
  - See the **movie**: “Claude Shannon - Father of the Information Age”

What you should know

- Why is important the ideia of Digital communications?
- What was the main question that Shannon try to address with Information Theory?
- What is on of the most important result of Shannon's work?

- Concepts:
  - General Digital Communication system
  - What is the role of the Encoder (and the corresponding decoder)
  - BSC; what is f
  - What is P_b, P_B and R (rate)?
  - Understand the Repetition codes, (R_N)
  - Understand the Block codes, the Linear Block codes, the Hamming code H(7, 4)

Further Reading and Summary

Q&A