I232E Information Theory (2023)
Course Introduction

Brian Kurkoski

2023 April

Information Theory

Information theory is the mathematical fundamentals of:

- information compression and
- information transmission

Information Compression

Reduce the amount of space required for storing data:

- Text — e.g. ZIP files
- Video — MP4
- Images — JPEG
- Audio — MP3

Information theory studies the mathematical fundamentals of compression.

Information Transmission

Reliable transmission in space:

- Mobile data and 4G
- WiFi
- fiber optic

Reliable transmission in time:

- flash memories, SSDs
- hard drives

Information theory studies the mathematical fundamentals of transmission.

Success of Information Theory

If you use media on your smartphone, then information theory has benefited you:

- error-correcting codes for WiFi and mobile data
- error-correcting codes for reliable flash storage
- compression of audio, video and pictures

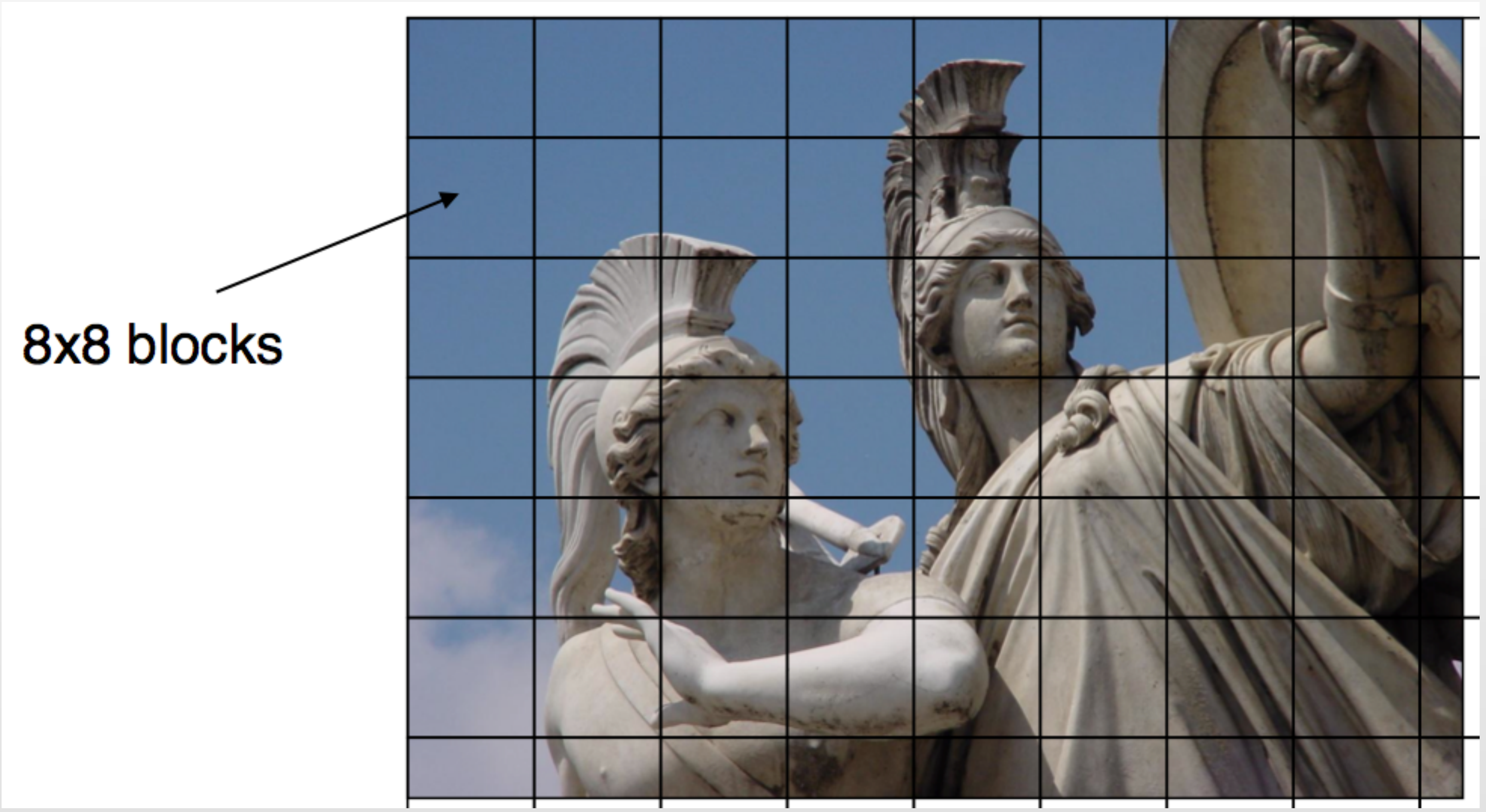
https://jp.techcrunch.com/2022/03/15/iphone13-greeen/

Questions Addressed in This Class

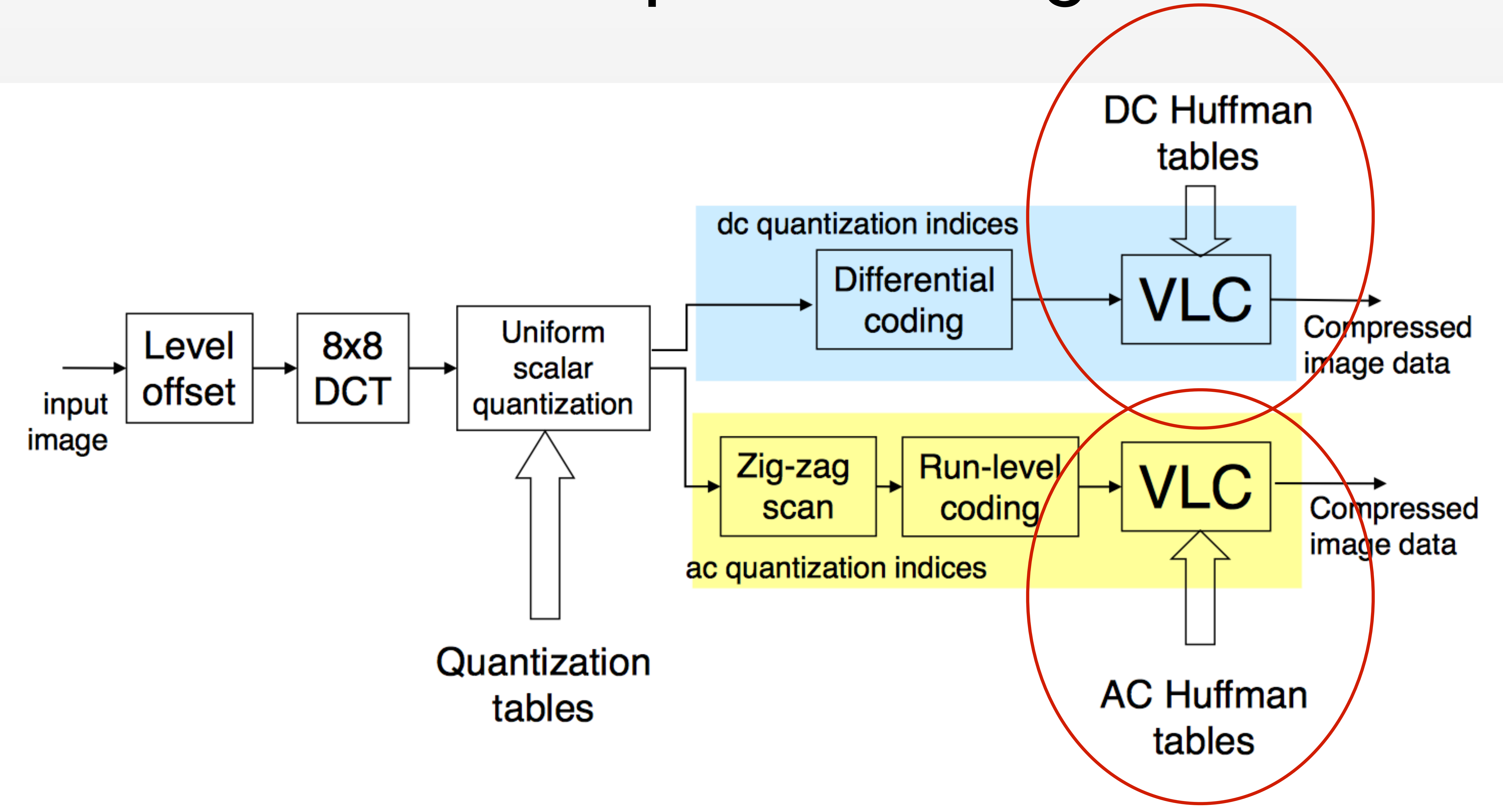
- How to measure information?
- How to compress information losslessly?
- What are the limits on compression of information?
- What are the limits on reliable communication?
- What are the limits on lossy compression of information?
- What are the limits when two users transmit simultaneously?

Huffman Codes Compress Information Losslessly

source image



JPEG compression algorithm



JPEG compression uses Huffman codes for compression

We do not study signal processing aspects of image compression

http://www.stanford.edu/class/ee398a/handouts/lectures/08-JPEG.pdf


PNG format, File size 736 kB


JPEG compression 100%, File Size 124 kB


JPEG compression 1%, File Size 5 kB

Mathematical Statement of This Tradeoff

Rate R: number of bits needed to represent the source [file size]

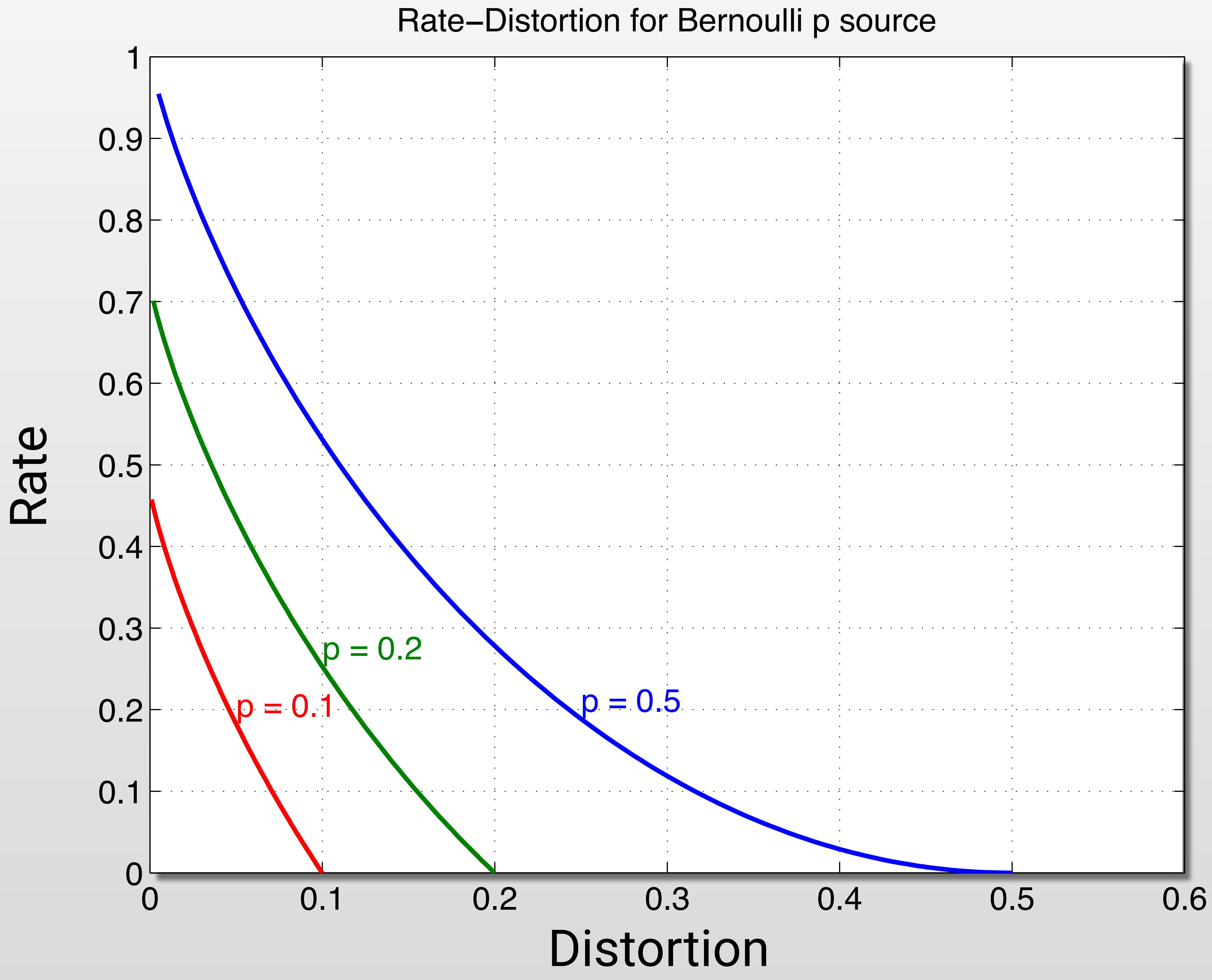
Distortion D: How close the reconstruction is to the original [quality]

Instead of images, we study binary sequences. The Rate R is:

$$R = h(p) - h(D)$$

where p is probability of a "1"

$$h(p) = -p \log p - (1-p) \log p$$



Unreliable Channels

SEND_$100 → SEND_$500

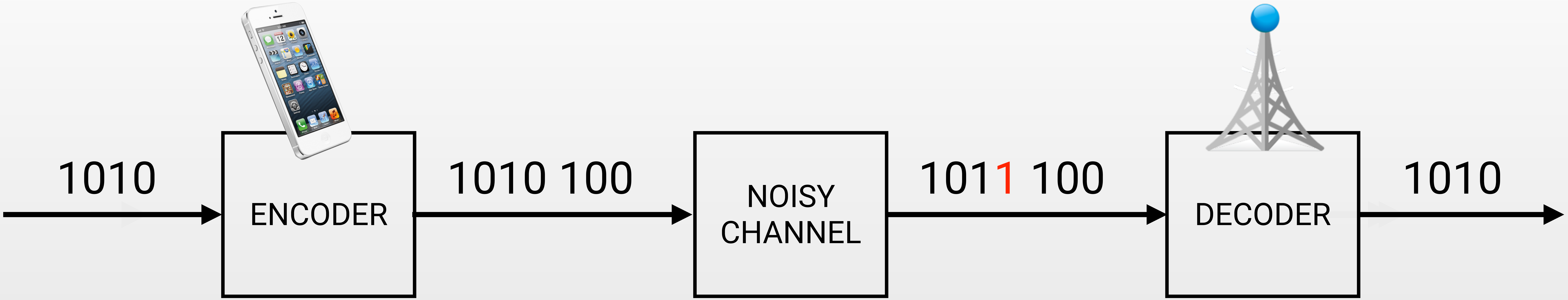
- Every day we send important information
- One error can completely change the meaning

Unreliable Channels

1010 → 1011

- Errors can occur in any physical communication medium
- Represent the message using numbers, usually binary bits.
- Errors: Transmitter sends zero, but it is received as one (or vice versa)

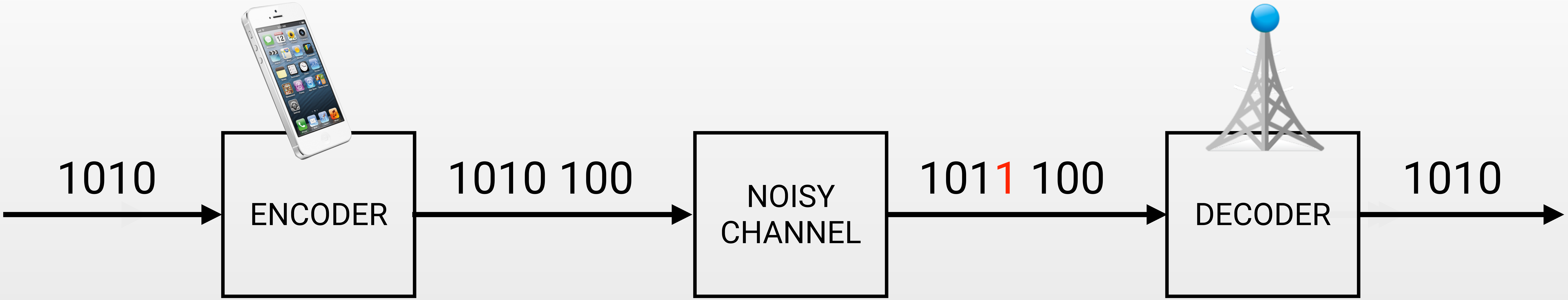
Reliable Communications over Unreliable Channels



This problem is solved using **channel coding**

- An **encoder** adds parity bits — the message is now longer
- The **channel** is a probability model of the errors
- The **decoder** recovers the original message — if there are not too many errors.

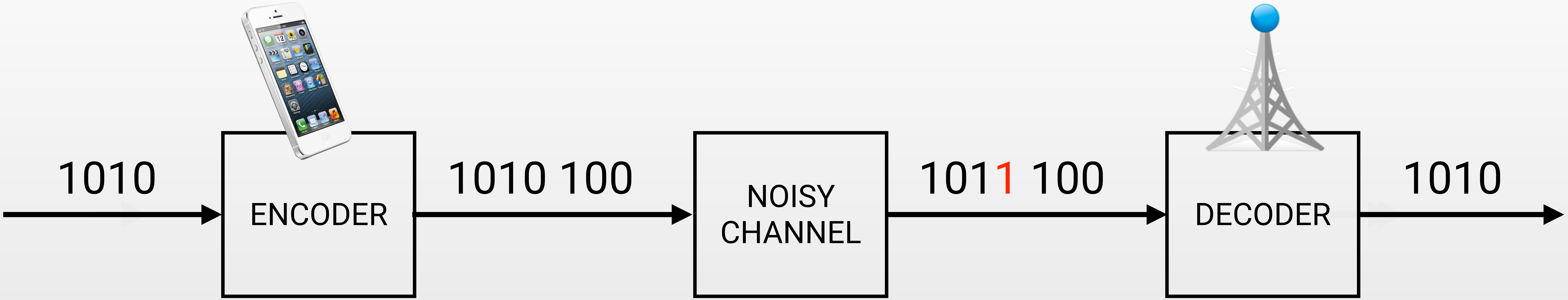
Reliable Communications over Unreliable Channels



Two aspects of channel coding:

- How to design specific codes that work well — covered in I437 Coding Theory course at JAIST
- What are the fundamental limits of communications — covered in this course.

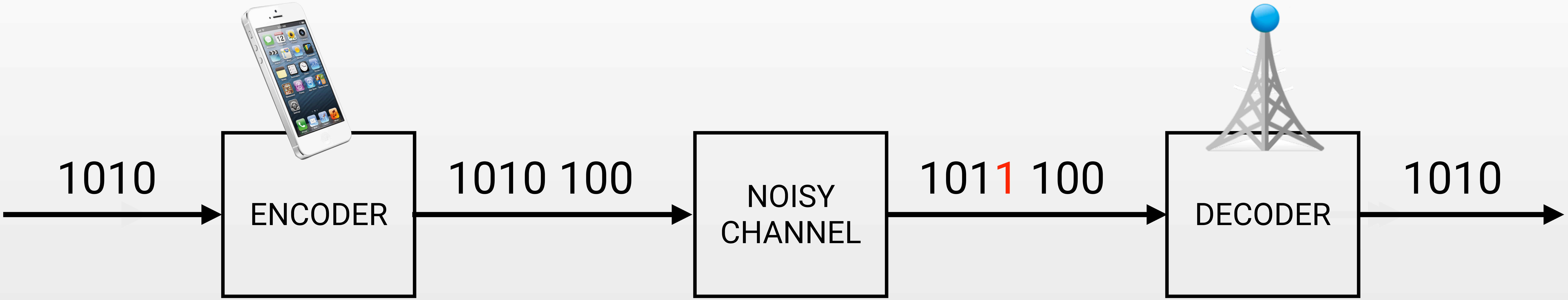
Fundamental Limits of Communications: Code Rate



$$\text{Code Rate} = \frac{\#\text{ message bits}}{\#\text{ codeword bits}} = \frac{4}{7} = 0.571$$

- Higher rates carry more information. Lower rates are more reliable.
- **Central question**: What is the highest possible rate with 0 errors?

Fundamental Limits: Channel Capacity



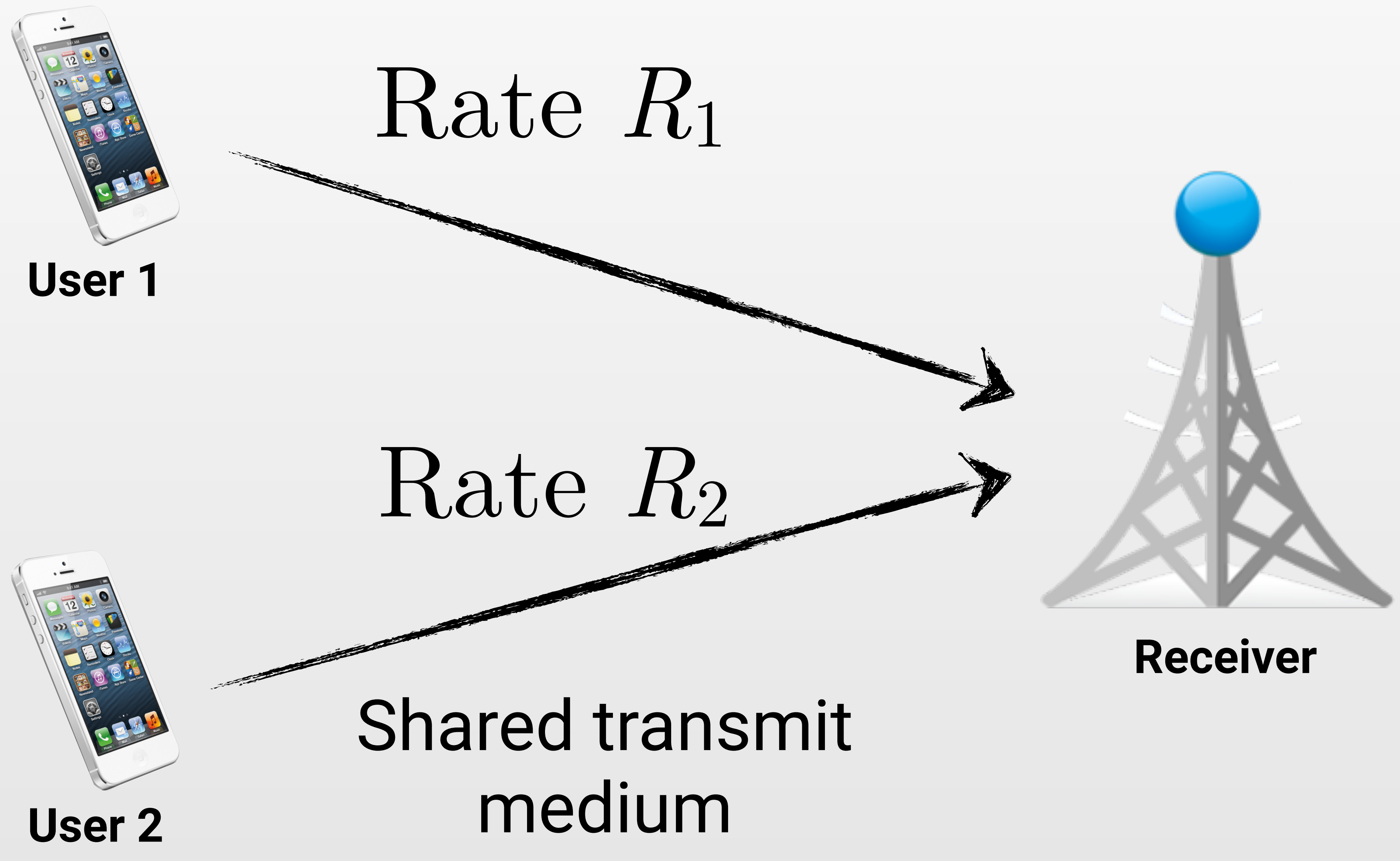
- The **channel capacity** is a fundamental property of a channel
- Channel Coding Theorem: Error-free communications if and only if:

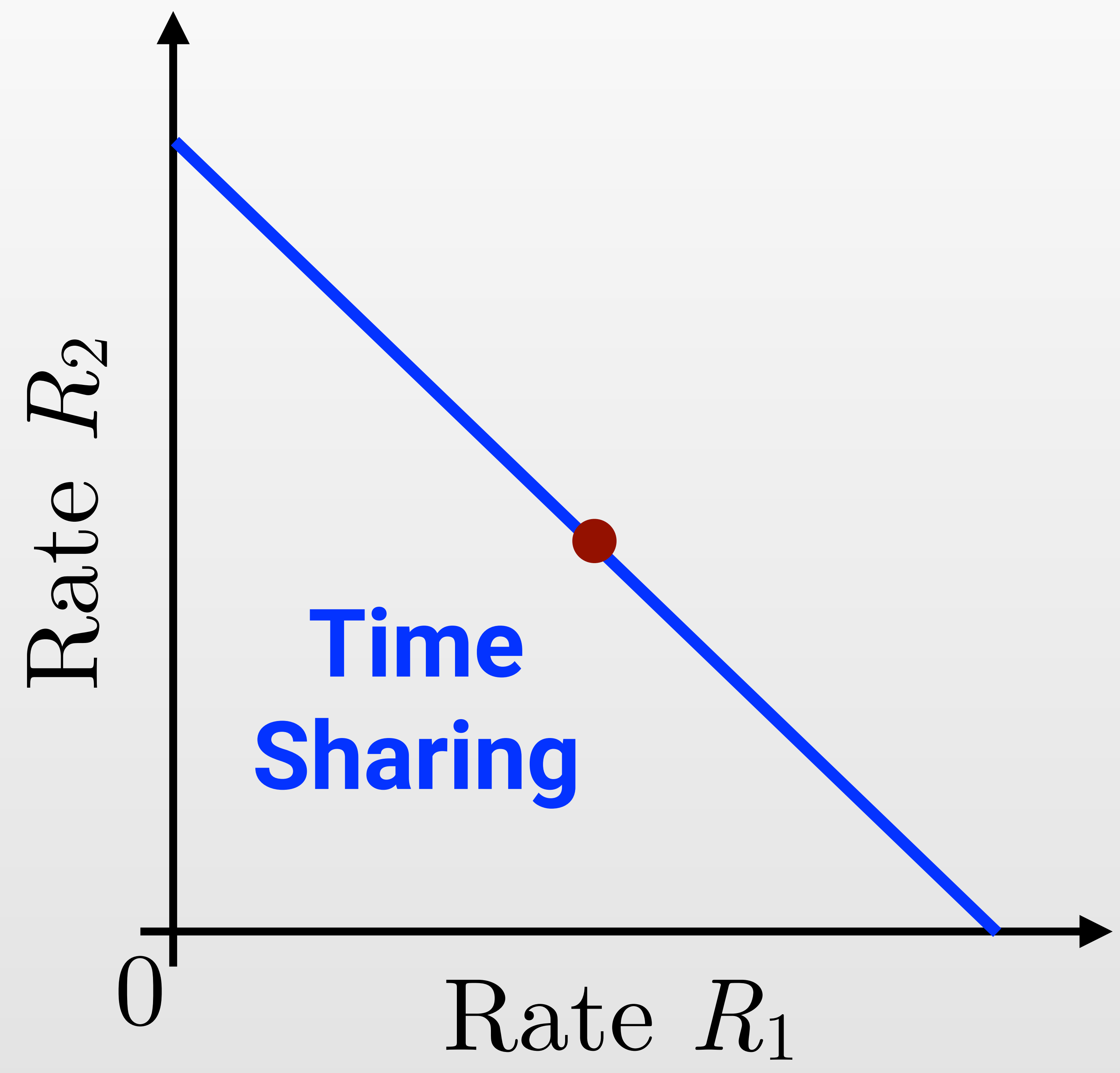
$$\text{Code rate} < \text{Channel Capacity}$$

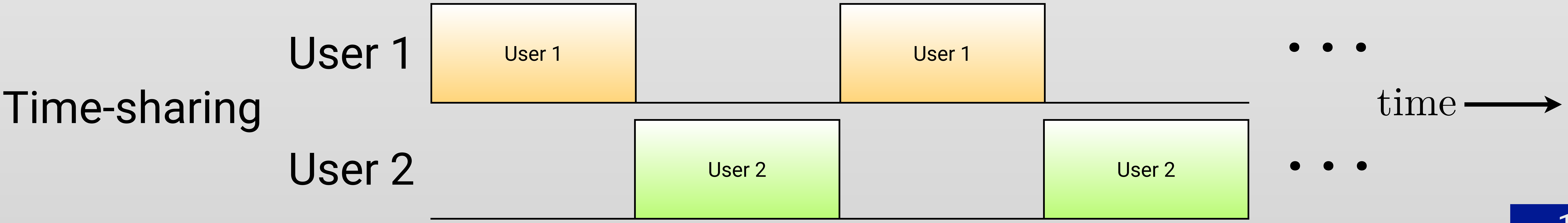
- For an important case of the Gaussian channel:

$$R < C = \frac{1}{2}\log\left(1 + \mathrm{SNR}\right)$$

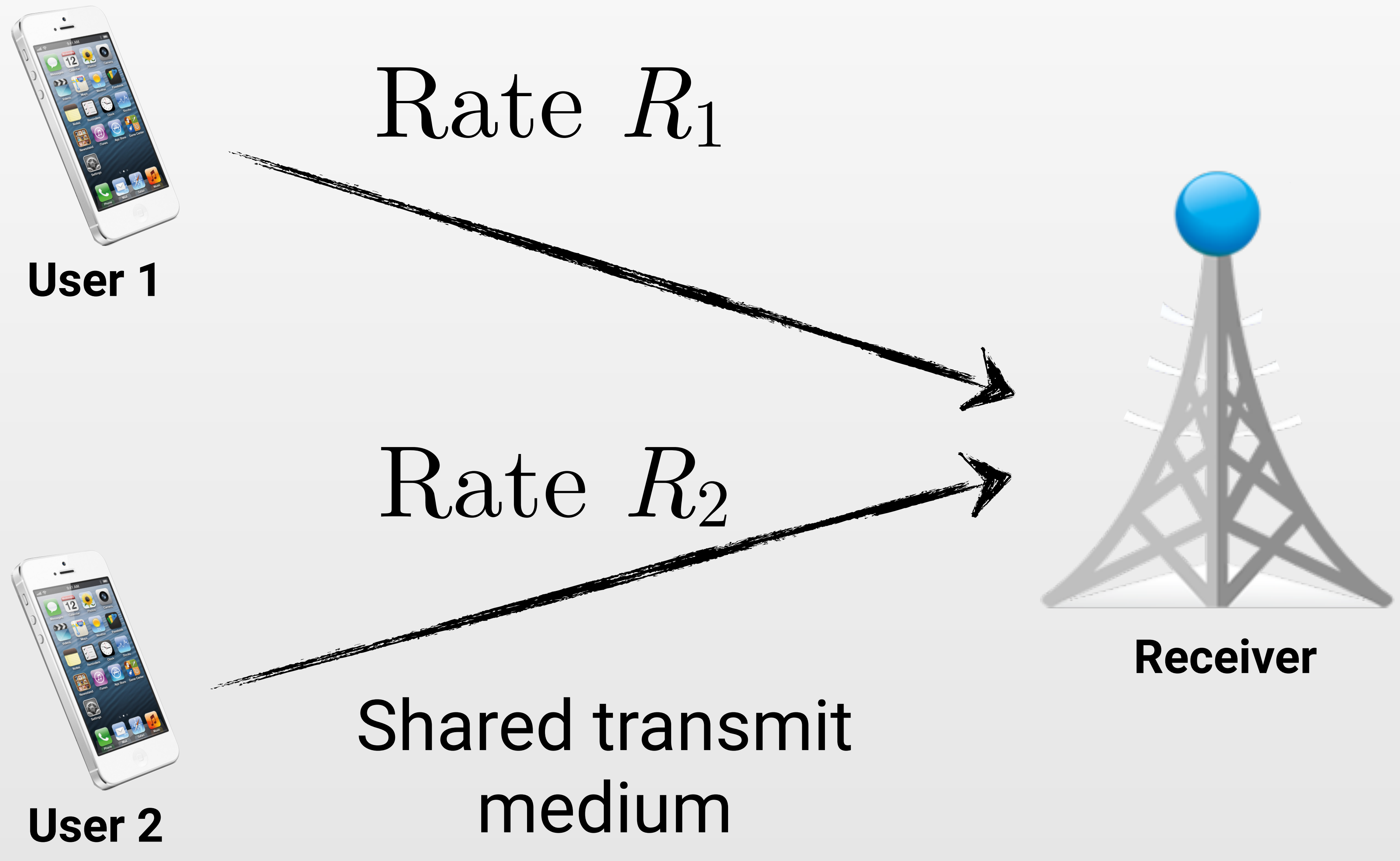
Two Users Transmitting at the Same Time

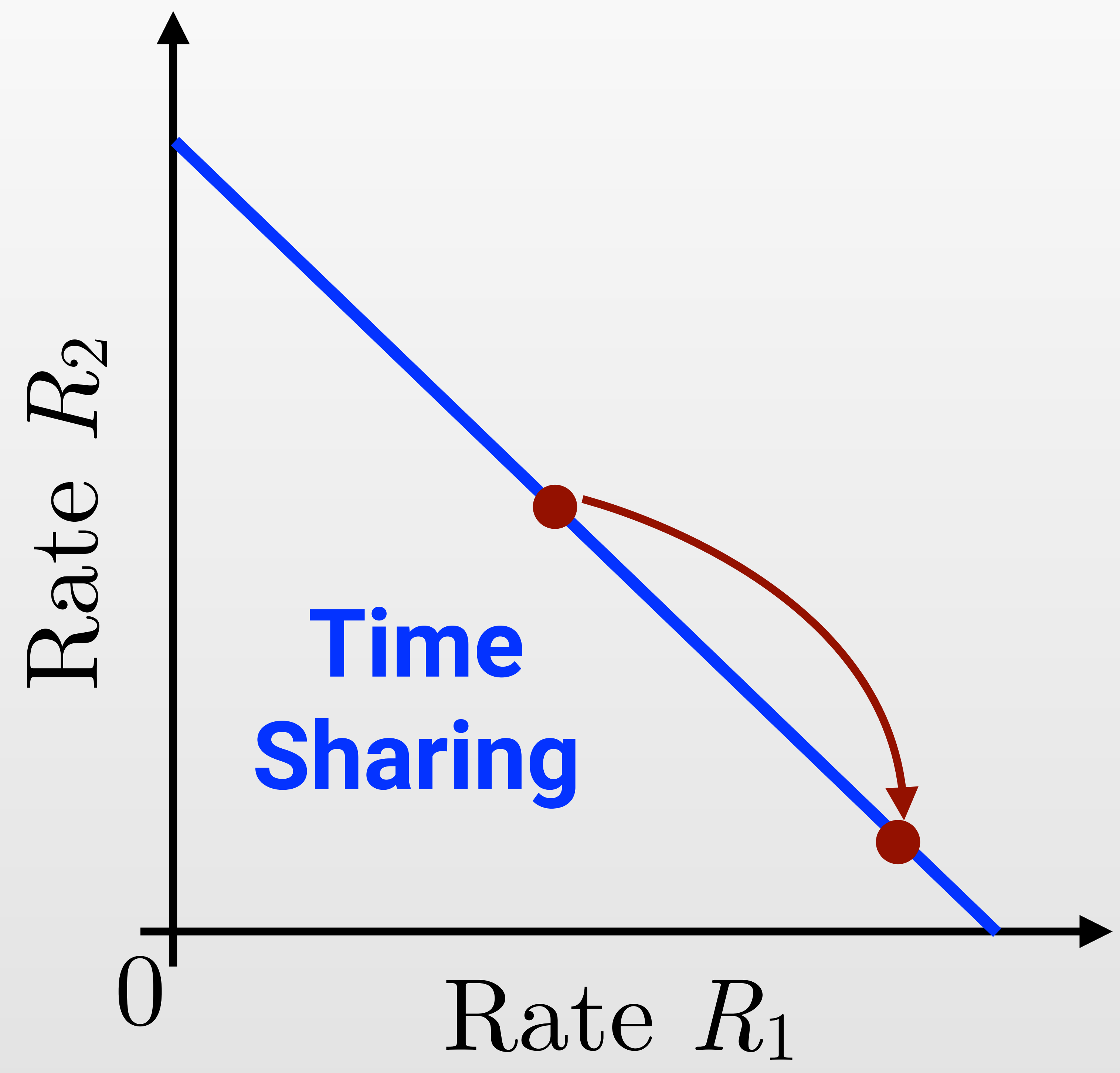


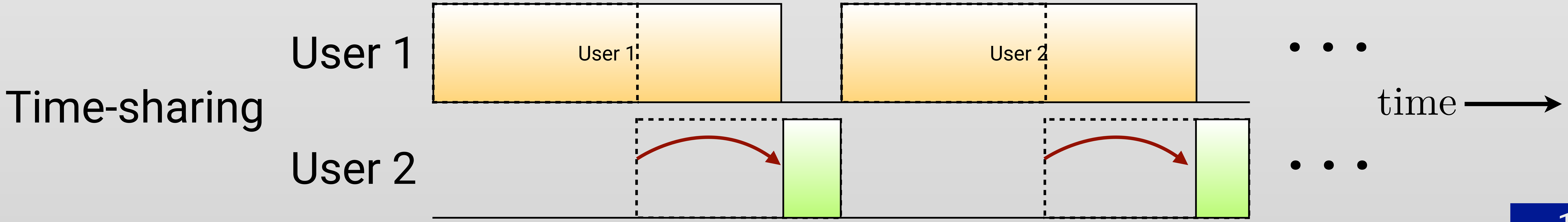




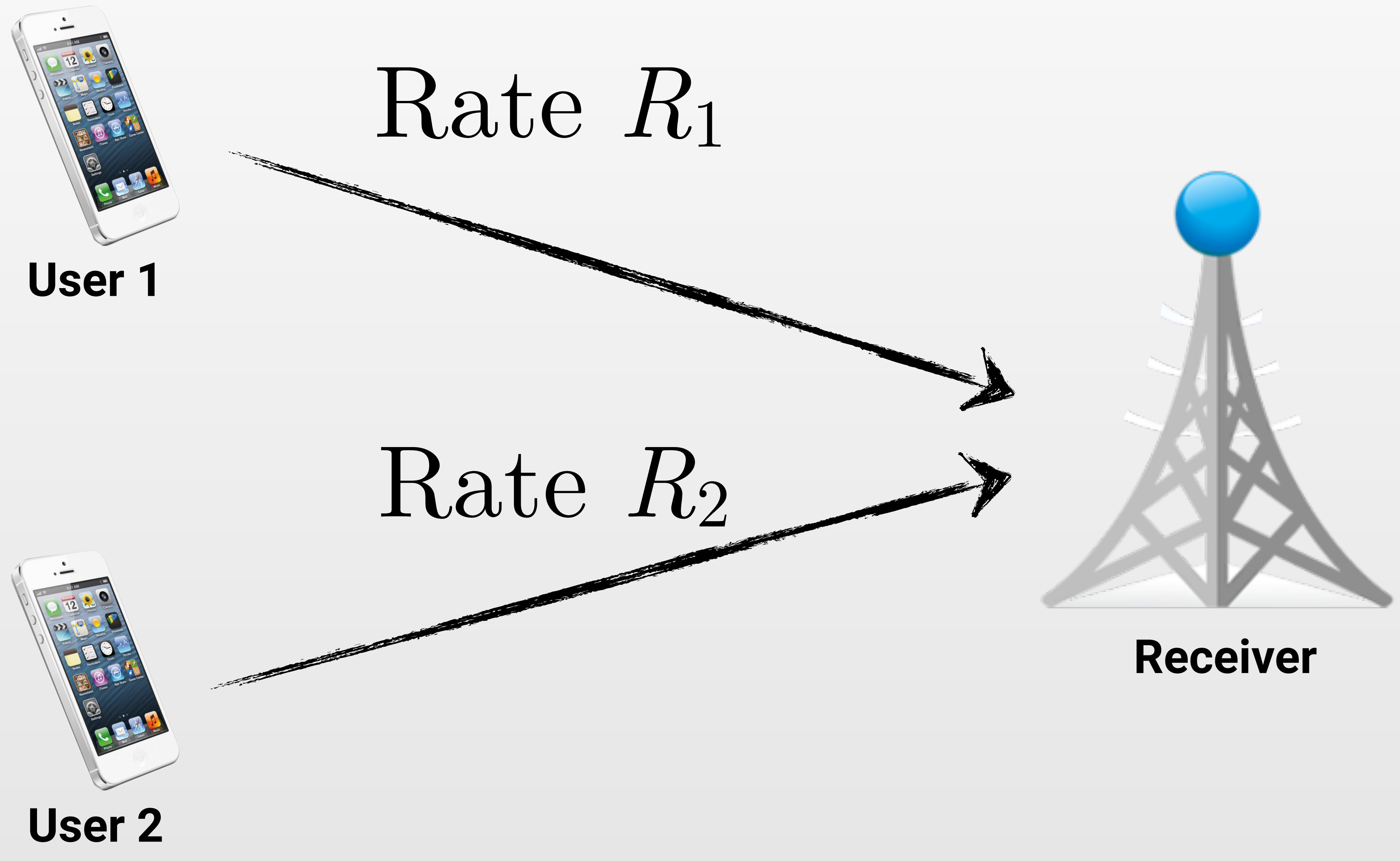
Two Users Transmitting at the Same Time

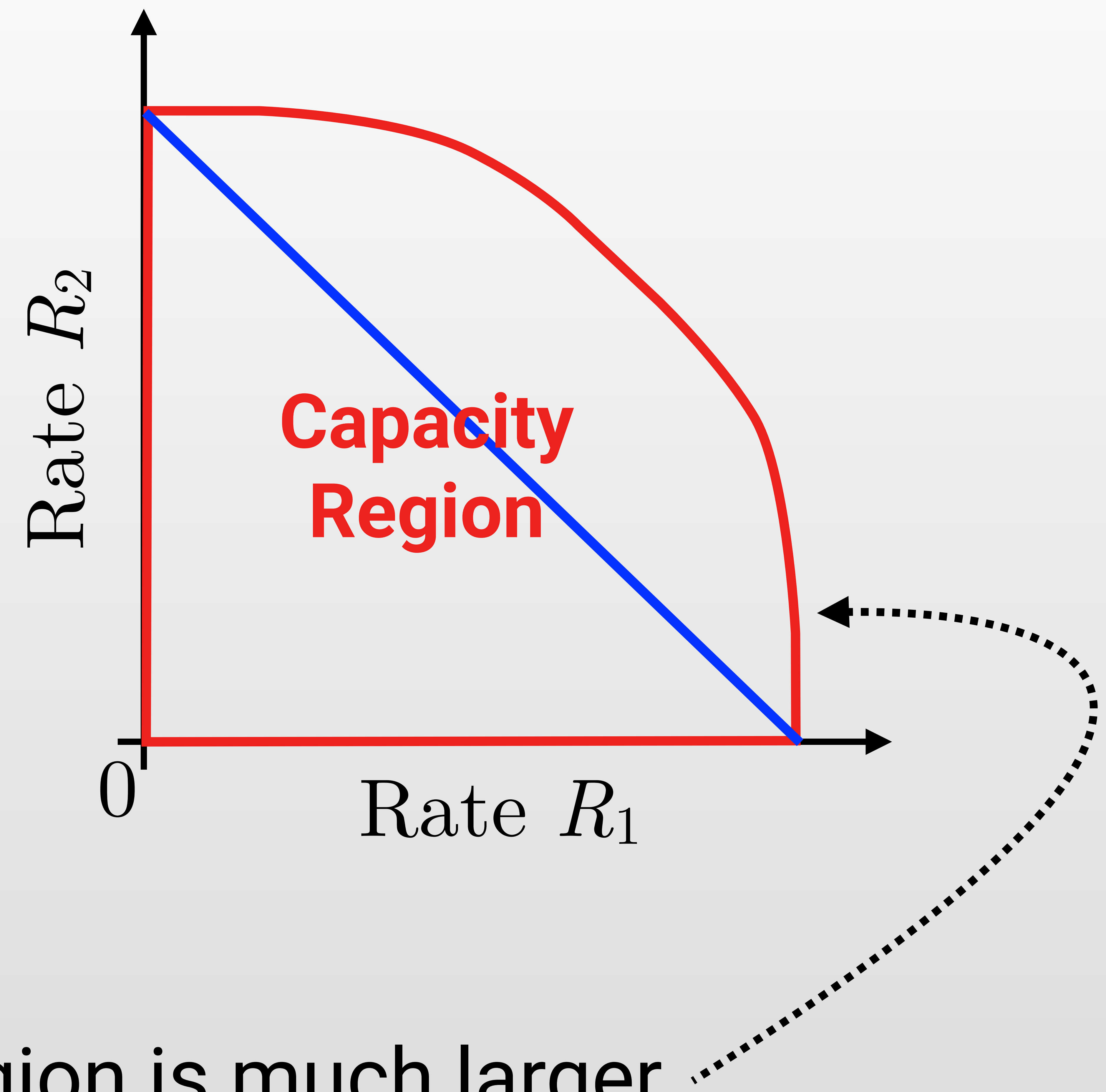






Two Users Transmitting at the Same Time





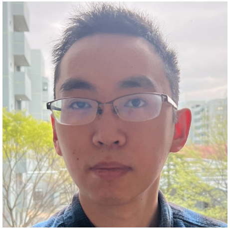
We will show that the capacity region is much larger

Both users can transmit at rates higher than time sharing

Course Staff

Brian Kurkoski
Instructor

Shunqi Huang
Teaching Assistant



Jiajie Xue
Teaching Assistant

Calendar and Grading

Instructor: Brian Kurkoski

Lectures		Tutorial Hour
Monday 9:00	Wednesday 10:50	Monday 13:30
	4/12	—
4/17	4/19	4/17
4/24	4/26	4/24
5/1	—	5/1
5/8	5/10	5/8
5/15	5/17	5/15 midterm exam
5/22	5/24	5/22
5/29	5/31	5/29
6/5 final exam		

Grade is based on:

- Midterm Exam 20%
- Final Exam 50%
- Pop Quiz 5%
- Homework 25%

Lecture Topics

Lecture 1 (today): Introduction and Entropy

Lecture 2: Tour of Probability Theory

Lecture 3: Mutual Information

Lecture 4: Source Coding 1

Lecture 5: Source Coding 2

Lecture 6: Source Coding 3

Lecture 7: Source Coding for Markov Chains

Lecture Topics

Lecture 7: Channel Coding 1

Lecture 8: Channel Coding 2

Lecture 9: Differential Entropy and the Gaussian Chanel

Lecture 10: Rate-Distortion Theory

Lecture 11: Network Information Theory 1

Lecture 12: Network Information Theory 2

Lecture 13: Optimization in Information Theory 1

Lecture 14: Optimization in Information Theory 2

Prerequisites: Before Taking This Course

Student should have studied the following topics:

1. **Probability theory**, for example:
   - Bayes' rule
   - Binomial distribution
   - Law of large numbers
2. **Basic calculus**

Other tools will be introduced as the course progresses.

Some "mathematical maturity" is needed — understand the problem and its mathematical justifications. This course includes various theorems and their proofs.

Study Materials

Textbook:

- *Information Theory Lecture Notes* by Brian Kurkoski. Provided on the course website.
- *Elements of Information Theory* by Cover and Thomas. Several copies available in the library. Japanese translation is available.

Study Materials

Textbook:

- *Information Theory Lecture Notes* by Brian Kurkoski. Provided on the course website.
- *Elements of Information Theory* by Cover and Thomas. Several copies available in the library. Japanese translation is available.

Video:

- We recorded and edited lectures in 2013
- Access from course website
- Some students find these useful

Study Materials: Self-Study Questions

Self-Study questions (SSQ) are simple questions to check your own understanding:

- available on the LMS
- unlimited number of attempts
- correct answers are shown after attempt
- does not count for your grade

Lecture notes

SSQ 1.1. Let X be a ternary random variable with $\mathsf{X} = \{1, 2, 3\}$ with probability distribution $p_{\mathsf{X}}(x)$:

$$p_{\mathsf{X}}(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{4} & \text{if } x = 2 \\ \frac{1}{4} & \text{if } x = 3 \end{cases} \tag{1.7}$$

Calculate $H(\mathsf{X})$. *Go to the course website for solutions and more SSQs.*

Course Web Site

Let X be a ternary random variable with $\mathcal{X} = \{1, 2, 3\}$ with probability distribution $p_X(x)$:

$$p_X(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{4} & \text{if } x = 2 \\ \frac{1}{4} & \text{if } x = 3 \end{cases}$$

Calculate $H(X)$.

Answer:

Pop Quiz: Begins Each Lecture

Each lecture begins with a short 5-minute quiz, called "Pop Quiz":

- Before each lecture, read the first section of the lecture notes
- Preparation for Pop Quiz is listed on the LMS:

  Lecture 2 Pop Quiz Preparation

  Preparation for the **online** Pop Quiz at the beginning of Lecture 2.

  To prepare for the pop quiz (see Pop Quiz 2), you should:

  - Read Section 2.1 of Information Theory Lecture Notes. You should be able to solve problems like Example 2.1 to Example 2.4.

- The pop quiz is online in the LMS, and must be completed within 5 minutes

In this way, you should study before you attend the lecture.

Homework

Homework is issued for each lecture

- First, attempt to solve the homework by yourself.

Homework

Homework is issued for each lecture

- First, attempt to solve the homework by yourself.
- If you cannot solve by yourself, bring your questions to tutorial hours
- Homework is due 18:00 same day as tutorial hours

	Tutorial Hours	Deadline	TA and Grader
Homework 1–2	4/24	18:00	Shunqi Huang
Homework 3–4	5/1	18:00	Jiajie Xue
Homework 5–6	5/8	18:00	Shunqi Huang
Homework 7–8		No submission	(midterm exam is 5/15)
Homework 9–10	5/22	18:00	Jiajie Xue
Homework 11	5/29	18:00	Shunqi Huang

Homework: Programming

Some homework assignments require programming:

- Simple numerical calculations
- Confirm your understanding of mathematical principles
- Lecture notes contain examples (in Matlab)
- You can use your favorite programming language

Homework: Collaboration Policy

You are encouraged to discuss and collaborate with other students on the homework. But:

- You must express your own understanding of the problem
- You must write or type by yourself
- Exams will test your understanding of information theory, use the homeworks to learn

Discussion Forum

The LMS has a Discussion Forum:

- Pleases use the Discussion Forum to ask questions
- Avoid sending instructors email for homework questions
- Used to send announcements

Summary — Information Theory Is:

Information theory is:

- an elegant mathematical approach
- . . . to solving important engineering questions
- . . . of information communication, compression and storage.

We start with practical problems, and make abstract mathematical models

This is a challenging topic, but the results are beautiful