# EC5.102: Information and Communication

(Lec-2)

# Source coding-2

(5-March-2025)

**Arti D. Yardi**

Email address: arti.yardi@iiit.ac.in

Office: A2-204, SPCRC, Vindhya A2, 1st floor

# Summary of the last class

# Recap

- Introduction to source coding:
  - Definition
  - Examples
  - Expected length of code $L(C)$
  - Encoding
  - Decoding
  - Decoding error
- (Informal statement of Source coding theorem): For zero-decoding error (or for "lossless" data compression), we need $L(C) \geq H(X)$.

- Types of source codes
  - Singular and nonsingular source code
  - Unique decodability of a source code
  - Prefix or instantaneous source code

# Singular and nonsingular source code

- A code is said to be nonsingular if every element of $\mathcal{X}$ maps into a different string in $\mathcal{D}^*$, i.e.,

$$x \neq x' \Rightarrow C(x) \neq C(x') \text{ for any } x, x' \in \mathcal{X}$$

- Consider the source code for $\mathcal{X} = \{a\ b\ c\ d\}$ defined as follows.

$$C(a) = 0\ 0 \quad C(b) = 0\ 1 \quad C(c) = 1\ 0 \quad C(d) = 1\ 1$$

  This code is nonsingular.

- Consider the source code for $\mathcal{X} = \{a\ b\ c\ d\}$ defined as follows.

$$C(a) = 0 \quad C(b) = 0 \quad C(c) = 0\ 1 \quad C(d) = 1\ 1$$

  This code is singular.

- Why is it desirable to have nonsingular code?
- Is it desirable to have nonsingular code? **Key idea in SCT!**

# Unique decodability of a source code

- Consider the following nonsingular source code for $\mathcal{X} = \{a\ b\ c\ d\}$ defined as follows.

$$C(a) = 0 \quad C(b) = 1 \quad C(c) = 0\ 1 \quad C(d) = 1\ 1$$

- File: a c b a a a d a a b b a a b c d
- Source coded file is given by

0, 0 1, 1, 0, 0, 0, 1 1, 0, 0, 1, 1, 0, 0, 1, 0 1, 1 1

- We have used "comma" to separate codewords and this ensures unique decodability.
- Without using comma, source coded file is given by

0 0 1 1 0 0 0 1 1 0 0 1 1 0 0 1 0 1 1 1 Wrong

0 0 1 1 0 0 0 1 1 0 0 1 1 0 0 1 0 1 1 1 Correct

- Can we get rid of this "comma" and still ensure unique decodability? Yes!

# Uniquely decodable source code: Example

- Consider the following nonsingular source code for $\mathcal{X} = \{a\ b\ c\ d\}$ defined as follows.

$$C(a) = 0 \quad C(b) = 1\ 0 \quad C(c) = 1\ 1\ 0 \quad C(d) = 1\ 1\ 1$$

- File: a c b a a a d a a b b a a b c d
- Source coded file is given by

$$0\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 1\ 1$$

- We can decode this file uniquely without using comma!
- We shall next define this formally.

# Definition: Uniquely decodable source code

- Concatenation $C(x_1)C(x_2)$ of two codewords $C(x_1) = 0\ 0$ and $C(x_2) = 1\ 1$ is

$$C(x_1)C(x_2) = 0\ 0\ 1\ 1$$

- The extension $C^*$ of a code $C$ is the mapping from finite-length strings of $\mathcal{X}$ to to finite-length strings of $\mathcal{D}$, defined by

$$C(x_1, x_2, \ldots, x_n) = C(x_1)C(x_2)\ldots C(x_n)$$

- A code is called uniquely decodable if its extension is nonsingular.
- Any encoded string in a uniquely decodable code has only one possible source string producing it.
- However, one may have to look at the entire string to determine even the first symbol in the corresponding source string.

# Unique decodability: Example

- Which of the following code is uniquely decodable?

1. $\{0, 10, 11\}$
2. $\{0, 01, 11\}$
3. $\{0, 01, 10\}$
4. $\{0, 01\}$
5. $\{0, 01, 10, 11\}$
6. $\{110, 11, 10\}$
7. $\{110, 11, 100, 00, 10\}$

1. Yes
2. Yes
3. No
4. Yes
5. No
6. Yes
7. Yes

# Prefix code or instantaneous source code

- A code is called a prefix code or instantaneous code if no codeword is a prefix of any other codeword.
- For example, the following code is a prefix code.

$$C(a) = 0 \quad C(b) = 1\ 0 \quad C(c) = 1\ 1\ 0 \quad C(d) = 1\ 1\ 1$$

- Example:

0 1 0 1 1 1 1 1 0 1 0

0, 1 0 1 1 1 1 1 0 1 0

0, 1 0, 1 1 1 1 1 0 1 0

0, 1 0, 1 1 1, 1 1 0 1 0

0, 1 0, 1 1 1, 1 1 0, 1 0

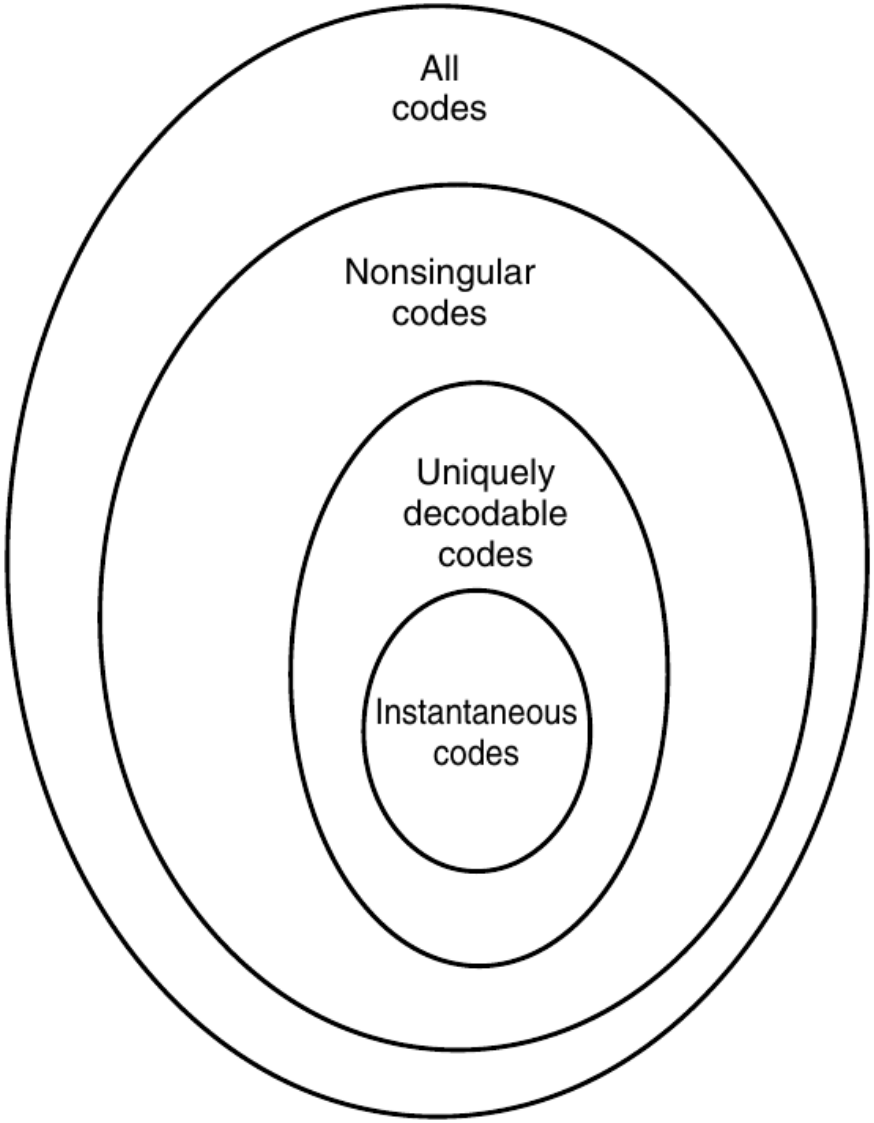
- Instantaneous code: We can look down the sequence of code symbols and add commas to separate the codewords without looking at later symbols.
- It is a self-punctuating code!

# Classes of source codes

| $X$ | Singular | Nonsingular, But Not Uniquely Decodable | Uniquely Decodable, But Not Instantaneous | Instantaneous |
|---|---|---|---|---|
| 1 | 0 | 0 | 10 | 0 |
| 2 | 0 | 010 | 00 | 10 |
| 3 | 0 | 01 | 11 | 110 |
| 4 | 0 | 10 | 110 | 111 |

- Verify each property: Homework

# Classes of source codes



# Criteria for Optimal Code Design

- Code must be uniquely decodable.
- Code should be a prefix or instantaneous.
  - Prefix Condition: A code is said to satisfy prefix condition if no code word is prefix to another code word.
  - Prefix condition is necessary and sufficient condition for a code to be uniquely decodable and instantaneous.
- Need smaller average codeword length among all prefix codes.
- Huffman codes are uniquely decodable, instantaneous codes with minimum average codeword length. In this sense, they are optimal.

# Summary

- Introduction to source coding
  - Definition of source code
  - Optimal source code
  - Types of source code:
    - Singular and nonsingular source code
    - Unique decodability of a source code
    - Prefix or instantaneous source code
  - Criteria for optimal code design