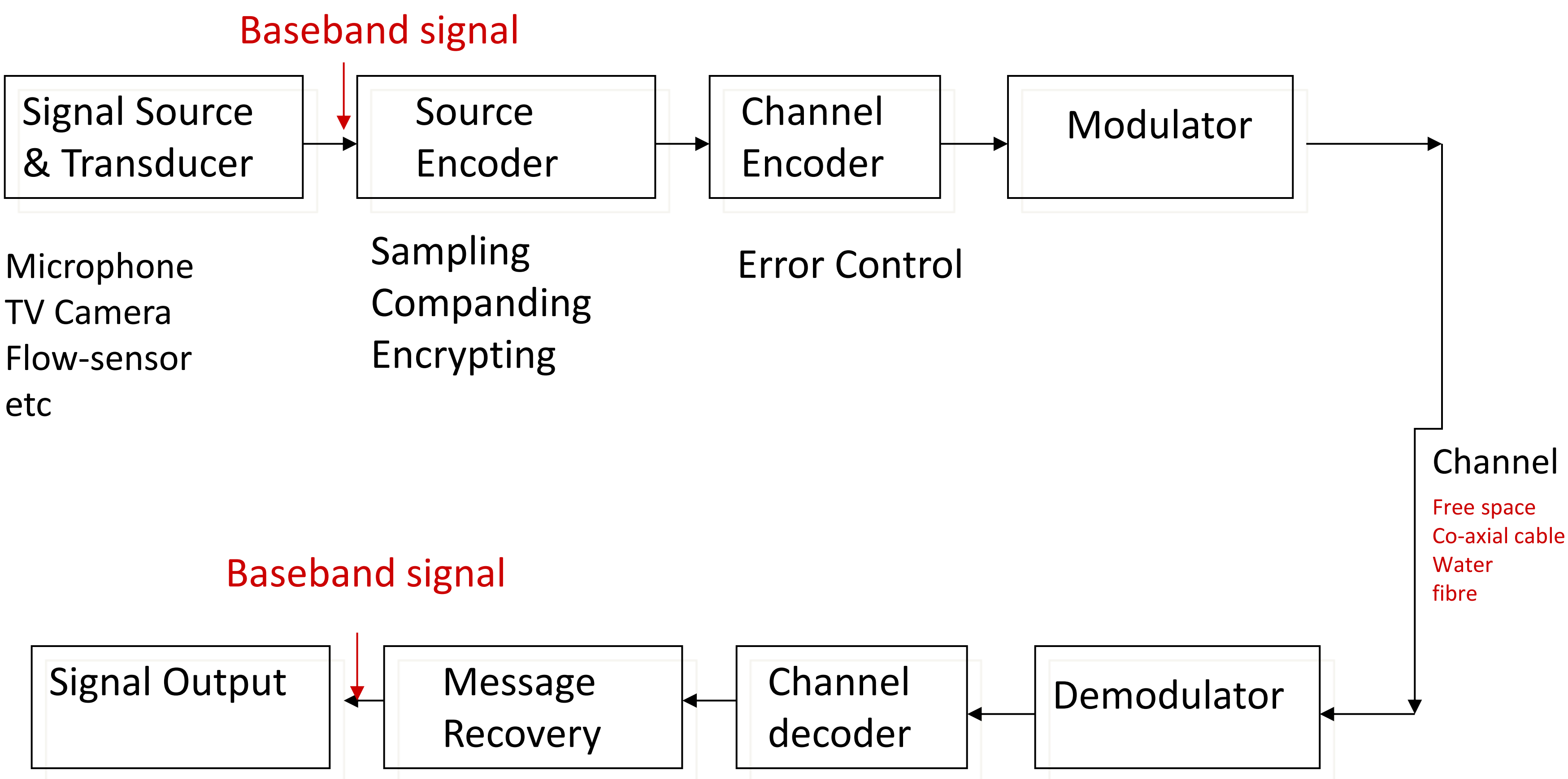
INFORMATION THEORY:
SOURCE CODING

1. Definition of Source Coding Terms
(a) Code Length
(b) Code efficiency
(c) Code redundancy
2. Source Coding Theory
3. Classification of Codes

EEEN 464 – Digital Communication

Thursday, 12 June 2025

DIGITAL COMMUNICATION



SOURCE CODING DEFINITION & OBJECTIVE

1. **Sources Coding is** the conversion of the output of a Discrete Memory-less Source (DMS) into a binary sequence.
2. The Objective of **Source Coding** is to minimize the average bit rate required to represent signal by reducing the redundancy of the information source.

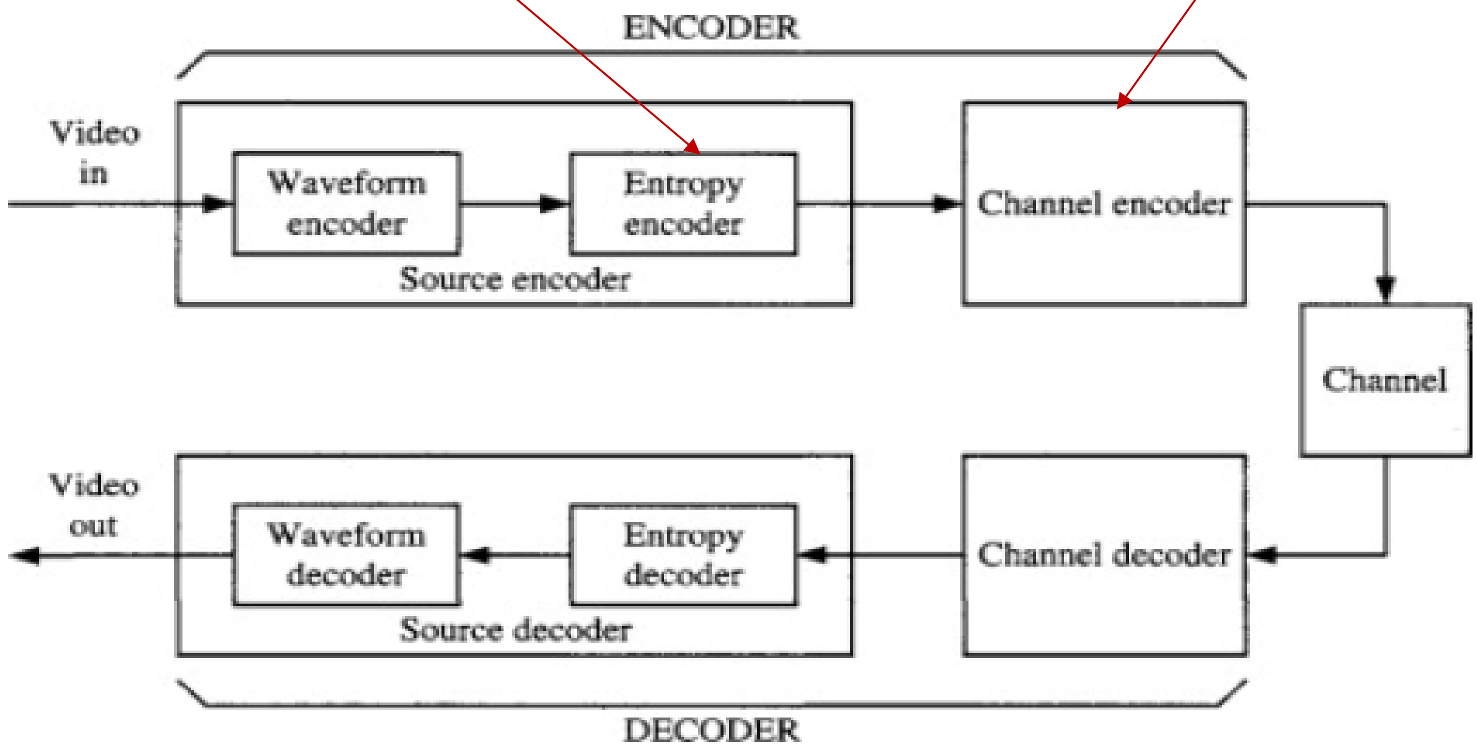
ENTROPY CODING DEFINITION & OBJECTIVE

1. **Entropy coding** is a loss-less compression technique that utilizes the statistical properties of data to assign shorter codes to more frequent symbols, effectively minimizing the amount of data needed to represent information;
2. **Entropy coding** has roots in information theory and is widely used in digital communication for image, audio, and video compression.

SOURCE & ENTROPY CODING

Entropy encoder takes data and compresses it by assigning variable-length codes based on the probability of each data symbol, essentially removing redundancy

Channel encoder conditions the compressed bitstream at the output of the source encoder to be suitable for transmission over the channel.



RECAP: CLASSIFICATION OF INFORMATION SOURCES

Information Sources fall into two categories:

a) **Memory sources** where the current symbol depends on the previous symbols,

b) **Memory-less sources** where current symbols are independent of previous symbols.

CODEWORD LENGTH DEFINITION

1. Assume X is a DMS with finite entropy H(X) and alphabet $\{x_1, x_2, .., x_n)$ and with corresponding probabilities $P(x_i)$ for i = 1,2,..,m.
2. If the binary code word assigned to a symbol x_i by the source encoder has length n, then the **codeword length** is the number of binary digits in the code word.

OTHER DEFINITIONS

1. The **average Codeword Length** is given by:

$$L = \sum_{i=1}^{m} P(x_i)n_i$$

Where n_i is the code length

2. **Code Efficiency** is defined as:

$$\eta = \frac{L_{\min}}{L}$$

where L_{min} is the minimum possible value of the average code length, L.

3. The code is said to be efficient when the code efficiency ,η tends to 1.

3. **Code redundancy** is defined as:

$$\gamma = 1 - \eta$$

SOURCE CODING THEOREM

1. The **Source coding theorem** states that a DMS X with entropy H(X), and **average code length L per symbol** is bound by

$$L \geq H(X)$$

2. Further, that L can be made closer to H(X) as desired through some suitable code.

CLASSIFICATION OF CODES

1. Fixed Length Code:

A code whose code length is fixed.

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

CLASSIFICATION OF CODES

2. Variable Length Code

A code whose length varies for different symbols.

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

CLASSIFICATION OF CODES - DISTINCT CODE

3. Distinct Code is a code in which each code word is distinguishable from other code words

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

CLASSIFICATION OF CODES – PREFIX FREE CODE

4. Prefix-free codes are codes in which no codeword can be formed by adding code symbols to another codeword

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

CLASSIFICATION OF CODES – UNIQUELY DECODABLE

5. Uniquely Decodable Codes

A code in which the original sequence can be reconstructed perfectly from the encoded binary sequence.

There's no ambiguity when decoding a received message

x_i	*Code 1*	*Code 2*	*Code 3*	*Code 4*	*Code 5*	*Code 6*
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

CLASSIFICATION OF CODES

6. Instantaneous Code (also referred to a prefix code) is a type of code where each codeword can be uniquely decoded as soon as it is received, meaning you don't need to wait for subsequent bits to identify a complete codeword.

- It is code which has **the end of any codeword is recognizable without examining subsequent code symbols.**
- Instantaneous Codes have the property that **no codeword is a prefix of another codeword.**

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

CLASSIFICATION OF CODES – OPTIMAL CODE

1. **Optimal code** is coding scheme that achieves the minimum possible average codeword length for a given set of data, essentially representing the best possible way to encode information with minimal redundancy.
2. **Optimal code** is the most efficient code for a particular data distribution, meaning it uses the least number of bits on average to represent each piece of data.

3. **A code is said to be optimal** if it is instantaneous and has a minimum average Length, L_{min}

WORKED EXAMPLE 1

1. A Discrete Memory-less Source X has alphabet $\{x_1, x_2\}$ and associated probabilities, $P(x_1)=0.9$, $P(x_2)=0.1$ where the symbols are encoded as:

x_i	$P(x_i)$	Code
x_1	0.9	0
x_2	0.1	1

Find the efficiency and redundancy of the code

EXAMPLE 1 SOLUTION

$$L = \sum_{i=1}^{2} P(x_i)n_i = (0.9)(1) + (0.1)(1) = 1\ \mathrm{b}$$

Entropy is:

$$\begin{aligned} H(X) &= -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i) \\ &= -0.9\log_2 0.9 - 0.1\log_2 0.1 \\ &= 0.469\ \mathrm{b/symbol} \end{aligned}$$

Code efficiency is

$$\eta = \frac{H(X)}{L} = 0.469 = 46.9\%$$

Code redundancy is:

$$\gamma = 1 - \eta = 0.531 = 53.1\%\ \textbf{Ans.}$$

EXAMPLE 2

2. A Discrete Memory-Less Source X has alphabet $\{x_1,x_2,x_3,x_4\}$ and a source coding as shown below.

x_i	$P(x_i)$	Code
X_1	0.81	0
X_2	0.09	10
X_3	0.09	110
X_4	0.01	111

Determine the efficiency and redundancy of the code.

EXAMPLE 2 - SOLUTION

- The average code word length, L is:

$$L = \sum_{i=1}^{4} P(a_i)n_i = 0.81(1) + 0.09(2) + 0.09(3) + 0.01(3) = 1.29 \text{ b/symbol}$$

- The entropy H(X) is given by:

$$\begin{aligned} H(X) &= -\sum_{i=1}^{4} P(a_i)\log_2 P(a_i) \\ &= -0.81 \log_2 0.81 - 0.09 \log_2 0.09 - 0.09 \log_2 0.09 - 0.01 \log_2 0.01 \end{aligned}$$

- Code Efficiency is therefore:

$$\eta = \frac{H(X)}{L} = \frac{0.938}{1.29} = 0.727$$

- Code redundancy is therefore:

$$\gamma = 1 - \eta = 0.273 = 27\%$$