Stanford University

ENGR 76
Information Science and Engineering

Winter 2026

Lecture 1

Introduction and Logistics

Siddharth Chandak

Course Staff

Instructors

Siddharth Chandak
Ph.D. Candidate, EE

Ayfer Ozgur
Professor, EE

TA

Andy Dong

Introduction

What is Information?

- Non-technical definitions -
  - *what is conveyed or represented by a particular arrangement of things.*
  - *facts and ideas, which can be represented (encoded) as various forms of data*
  - *content that can be stored, communicated, and used to understand or describe something.*
  - *and many others…*

What is Information?

- Non-technical definitions -
  - *what is conveyed or represented by a particular arrangement of things.*
  - *facts and ideas, which can be represented (encoded) as various forms of data*
  - *content that can be stored, communicated, and used to understand or describe something.*
  - *and many others…*

What is Information?

- Non-technical definitions -
  - *what is conveyed or represented by a particular arrangement of things.*
  - *facts and ideas, which can be represented (encoded) as various forms of data*
  - *content that can be stored, communicated, and used to understand or describe something.*
  - *and many others…*

What is Information?

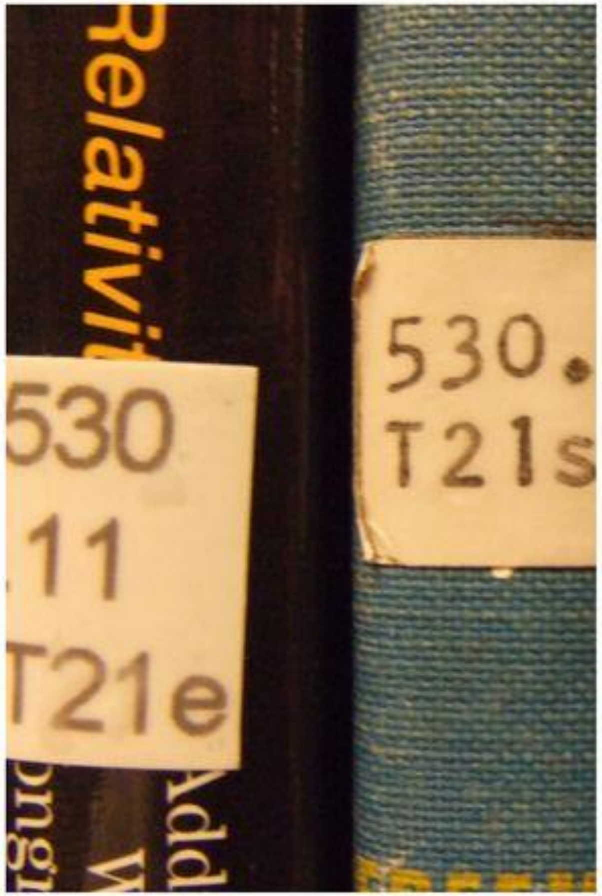
- Non-technical definitions -
  - *what is conveyed or represented by a particular arrangement of things.*
  - *facts and ideas, which can be represented (encoded) as various forms of data*
  - *content that can be stored, communicated, and used to understand or describe something.*
  - *and many others…*

Information Science and Engineering

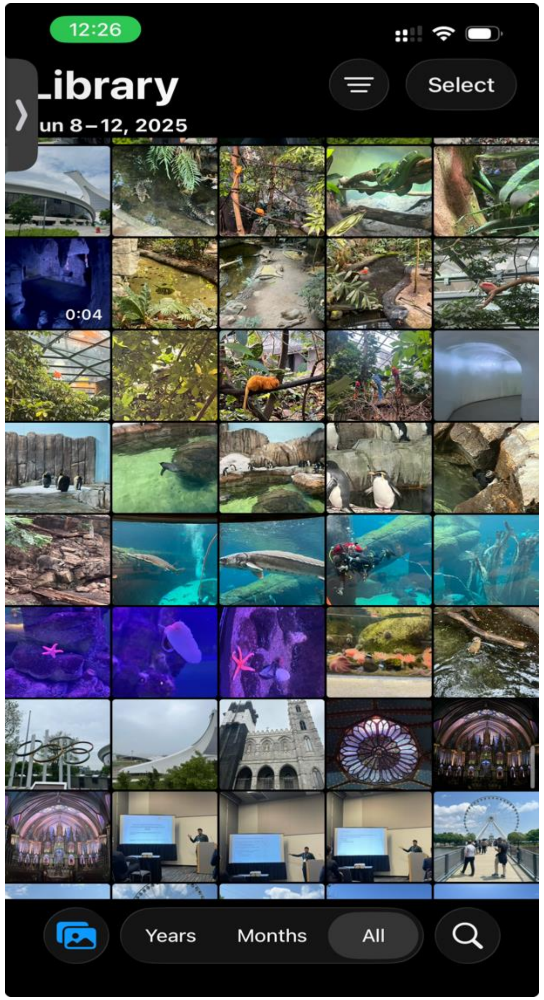
- What is information and how to quantify it?
- How to represent information?
- How to efficiently store information?
- How to communicate information?

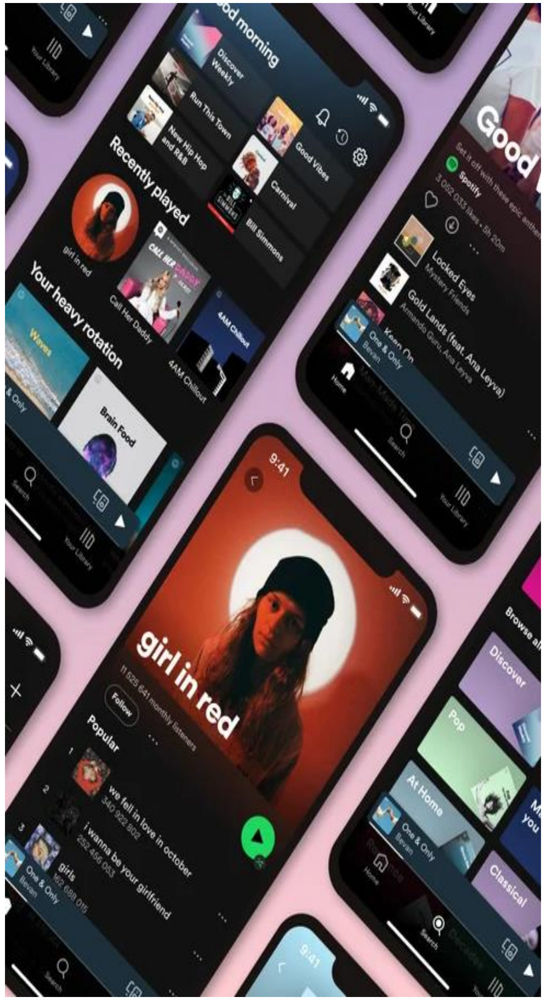
Representation and Storage

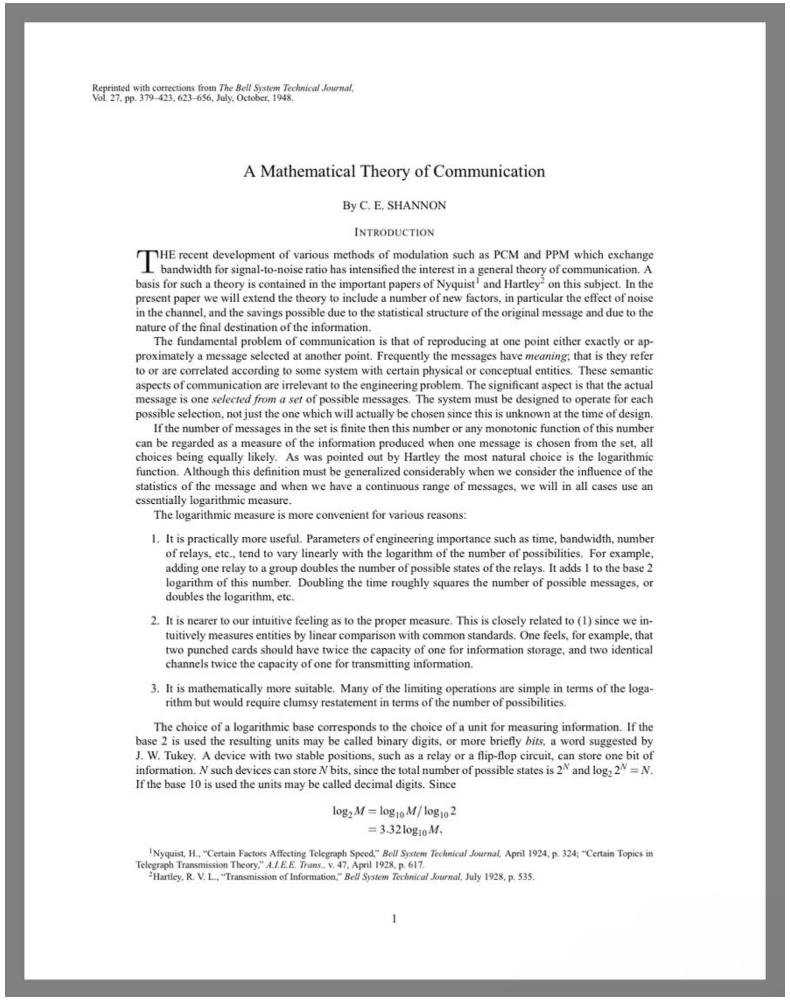
Storage Systems



Digital Storage Systems






Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number. Doubling the time roughly squares the number of possible messages, or doubles the logarithm, etc.

2. It is nearer to our intuitive feeling as to the proper measure. This is closely related to (1) since we intuitively measures entities by linear comparison with common standards. One feels, for example, that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information.

3. It is mathematically more suitable. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities.

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits*, a word suggested by J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2^N and $\log_2 2^N = N$. If the base 10 is used the units may be called decimal digits. Since

$$\log_2 M = \log_{10} M / \log_{10} 2$$
$$= 3.32 \log_{10} M,$$

[1]Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A.I.E.E. Trans.*, v. 47, April 1928, p. 617.
[2]Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

Digital Storage

- All data (e.g., text, images, audio, and video) is stored as sequences of **bits**.
  - A bit is the smallest unit of digital information and can take one of two values: **0 or 1**.
- Why digital?
  - Standardized binary representation allows for modular designs
  - Electronic circuits have become more digital
  - Better noise correction

Representation

How is data represented using bits?

Text represented using bits

- ASCII representation of characters
  - Each character is represented using a sequence of 8 bits (or 1 byte)

Symbol	Decimal	Binary
A	65	01000001
B	66	01000010
C	67	01000011
D	68	01000100
E	69	01000101
F	70	01000110
G	71	01000111
H	72	01001000
I	73	01001001
J	74	01001010
K	75	01001011
L	76	01001100
M	77	01001101
N	78	01001110
O	79	01001111
P	80	01010000
Q	81	01010001
R	82	01010010
S	83	01010011
T	84	01010100
U	85	01010101
V	86	01010110
W	87	01010111
X	88	01011000
Y	89	01011001
Z	90	01011010

Symbol	Decimal	Binary
a	97	01100001
b	98	01100010
c	99	01100011
d	100	01100100
e	101	01100101
f	102	01100110
g	103	01100111
h	104	01101000
i	105	01101001
j	106	01101010
k	107	01101011
l	108	01101100
m	109	01101101
n	110	01101110
o	111	01101111
p	112	01110000
q	113	01110001
r	114	01110010
s	115	01110011
t	116	01110100
u	117	01110101
v	118	01110110
w	119	01110111
x	120	01111000
y	121	01111001
z	122	01111010

Image represented using bits

- Raw representation of image
  - Image divided into pixels
  - Each pixel can take values from 0 to 255 (2^8 possible values)
    - Representing intensity
    - 8 bits

255	255	255	255	255	255	255	255	255
255	128	128	128	128	128	128	128	255
255	128	0	0	0	0	0	128	255
255	128	0	255	255	255	0	128	255
255	128	0	255	128	255	0	128	255
255	128	0	255	255	255	0	128	255
255	128	0	0	0	0	0	128	255
255	128	128	128	128	128	128	128	255
255	255	255	255	255	255	255	255	255

Raw representation takes too much space

- Consider a typical colored image (3 channels - RGB) of dimensions 4000 x 3000 pixels
  - Bits per pixel = 3 * 8 = 24 bits (3 bytes)
  - Total pixels = 4000 * 3000 = 12,000,000
  - Total size of image = **36,000,000 bytes (~36 MB)**
- 4000 images stored on phone would be **~144 GB**

More drastic for videos

- Video
  - Resolution: 1920 x 1080
  - Frame Rate: 30 fps
  - RGB, 8 bits per channel
- One hour of raw video: **~660 GB**

How much space does JPEG take?

- Raw representation of image: ~36MB
- JPEG image on our phone: 3-4 MB

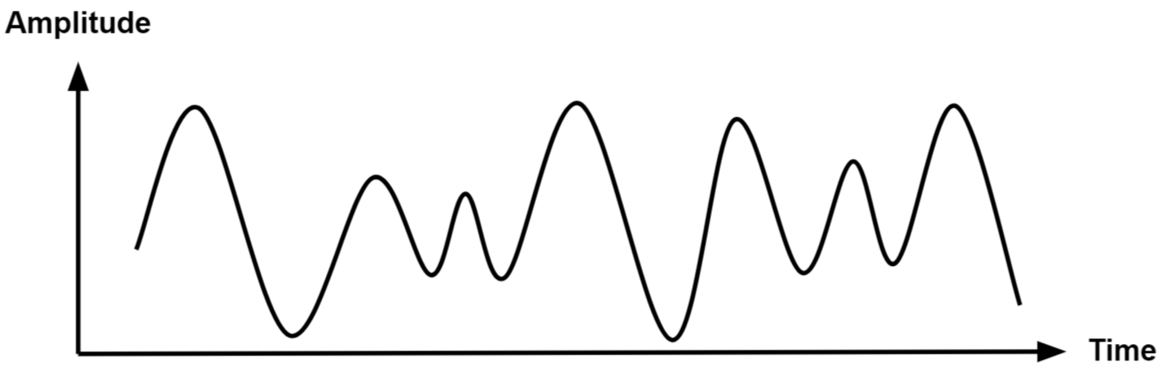
- **Project 1 – Image compressor**

Communication

Communication

- Sending information reliably over noisy channels

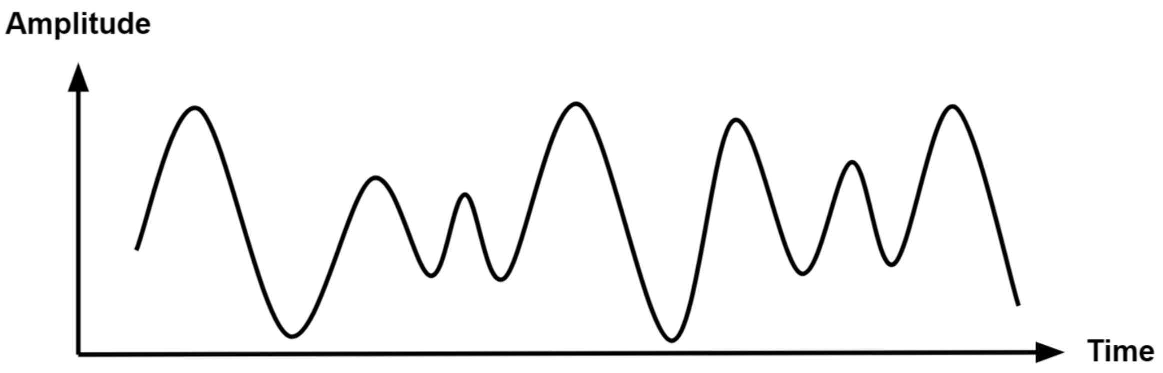
Analog vs Digital



Analog vs Digital Communication

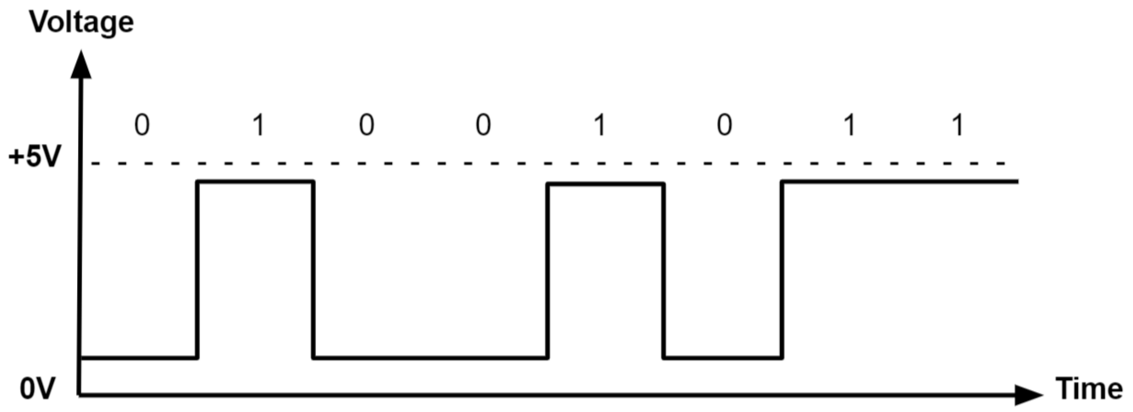
Analog Communication

- The message to be communicated is one of a continuum of possibilities.
    - Signal can take any value in a range at any given time



Digital Communication:

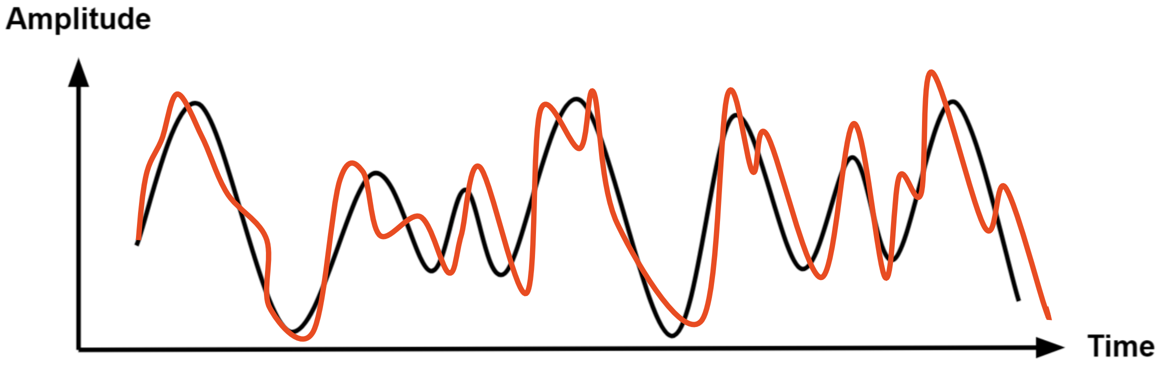
- The message to be communicated is one of a finite set of possible choices.



Analog vs Digital Communication

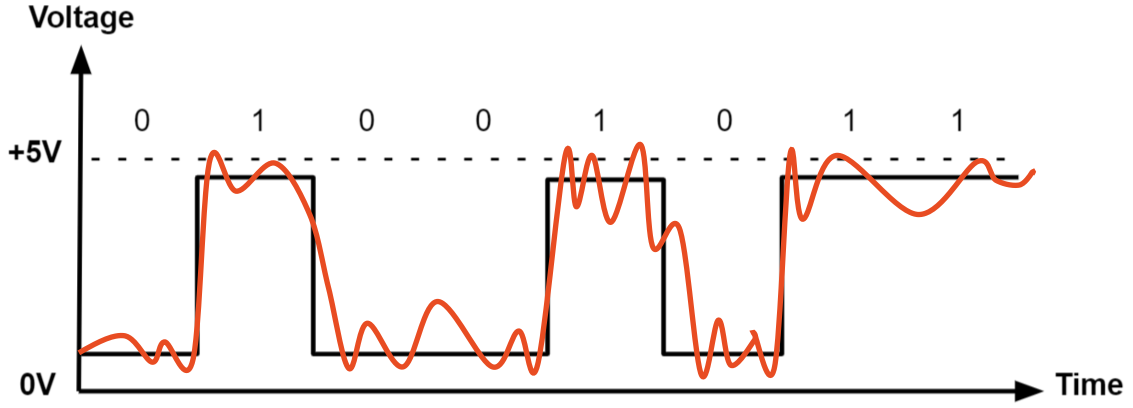
Analog Communication

- The message to be communicated is one of a continuum of possibilities.
- Can never fully remove the effects of noise.



Digital Communication:

- The message to be communicated is one of a finite set of possible choices.
- Can remove the effects of noise induced by the channel.

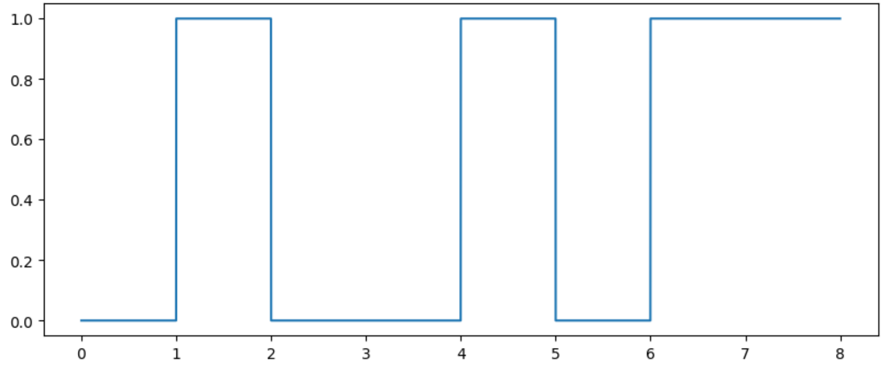


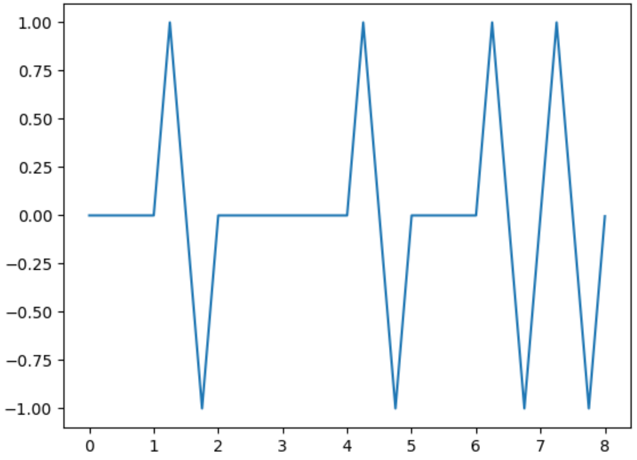
Converting bits to waveforms

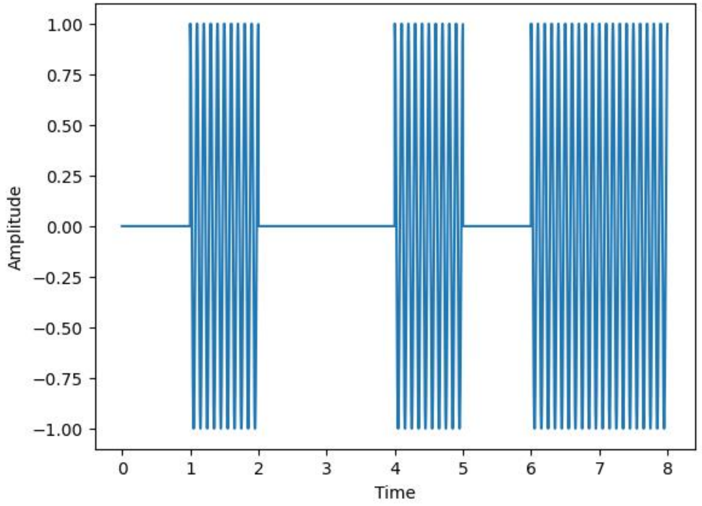
- How to communicate bits over a physical medium?
  - Cannot just send bits on an electric cable
  - Or cannot send bits ‘into the air’
- Need to convert bits into a waveform

Waveforms

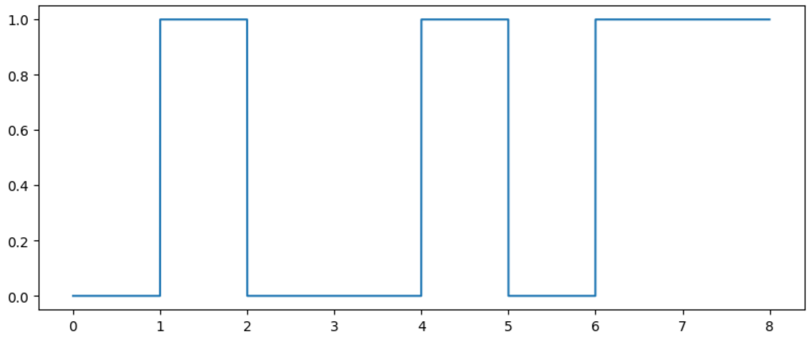
01001011 Modulation

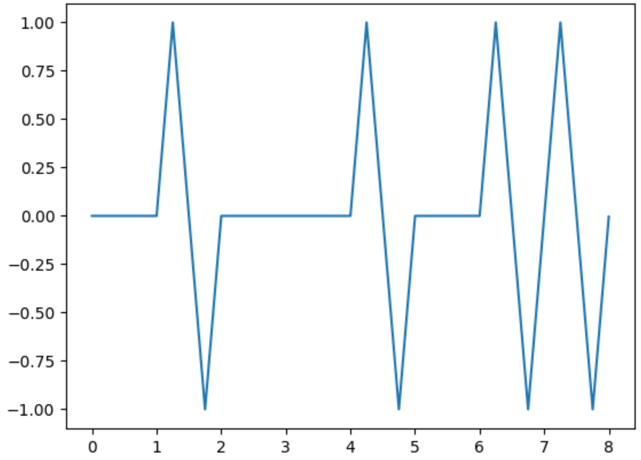


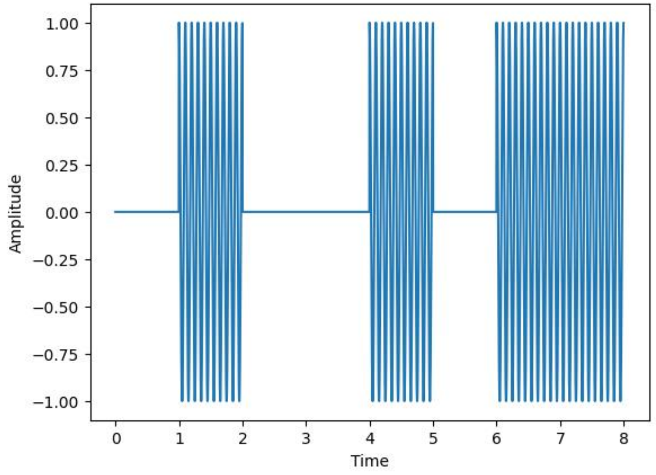




Waveforms







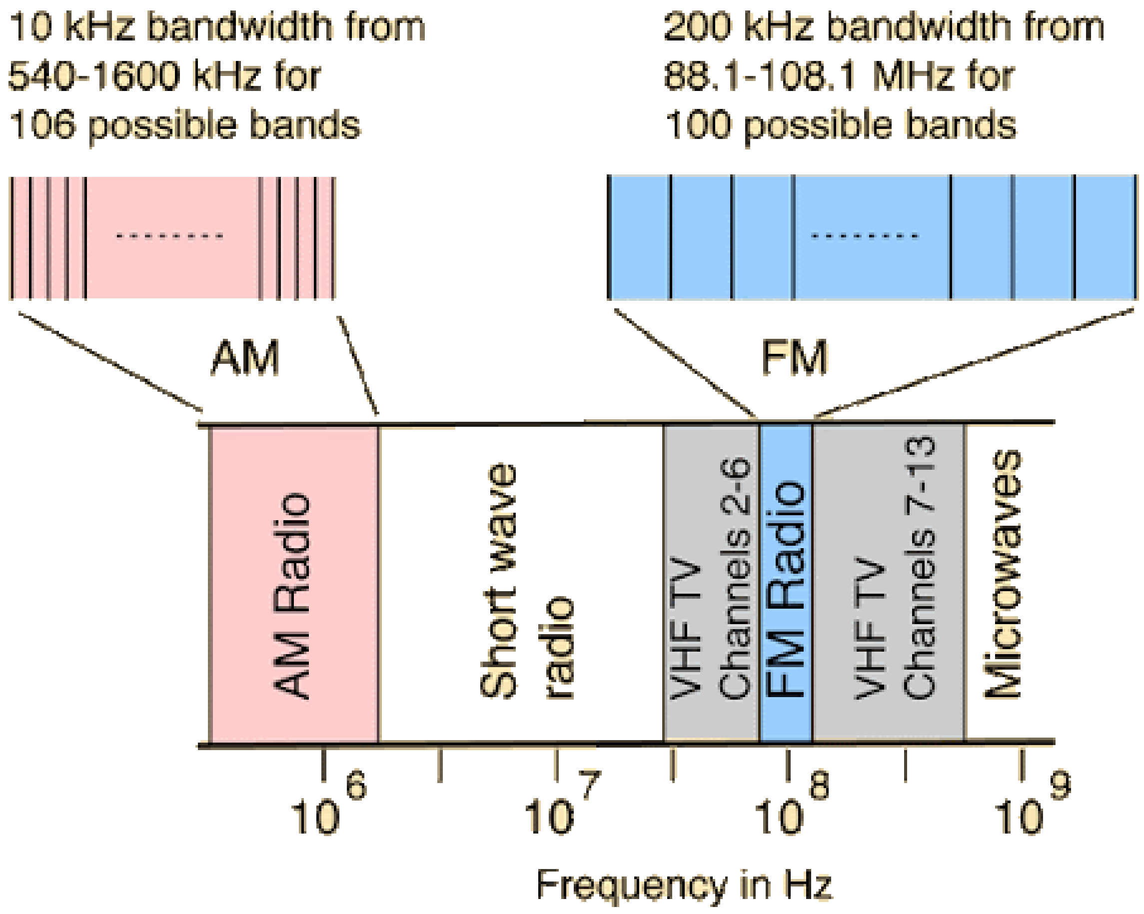
Demodulation → 01001011

Physical Medium

- We are working with a physical medium
  - Comes with its own challenges and limitations
- Example: Antennas
  - Specific **frequency** range for optimal performance
  - Our waveform/signal should be in that frequency range

Multiple messages

- How can multiple messages be transmitted at once (e.g., multiple radio channels)?
  - What does the 88.5 mean in 88.5 FM?
  - What are these bands? How is it possible that different channels operate in different bands?



Modulation

How do we ensure that the transmitted signal lies within its allocated band?

Noisy Channel

- ‘Noise’ everywhere
  - Parts of our transmitted signal might be lost (bits lost)
  - Or bits might get flipped
  - Our computer’s hard drive suffers errors in storage

How to achieve reliable communication in these noisy environments?

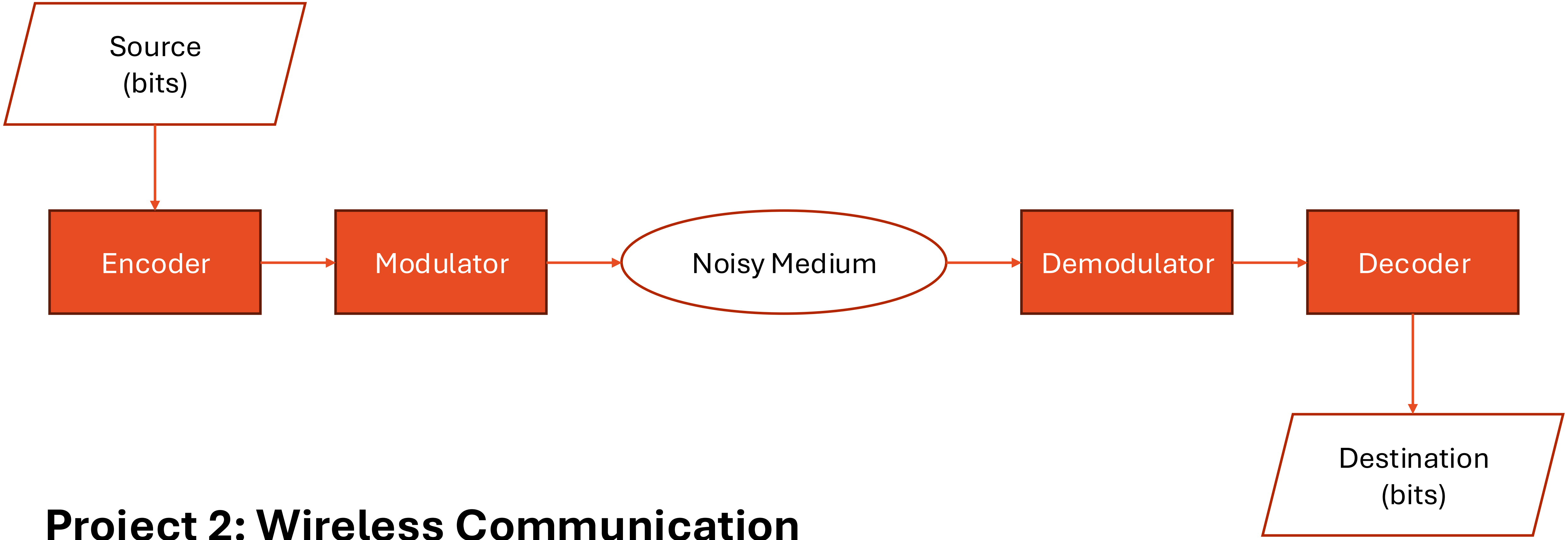
An example

An example

An example

An example

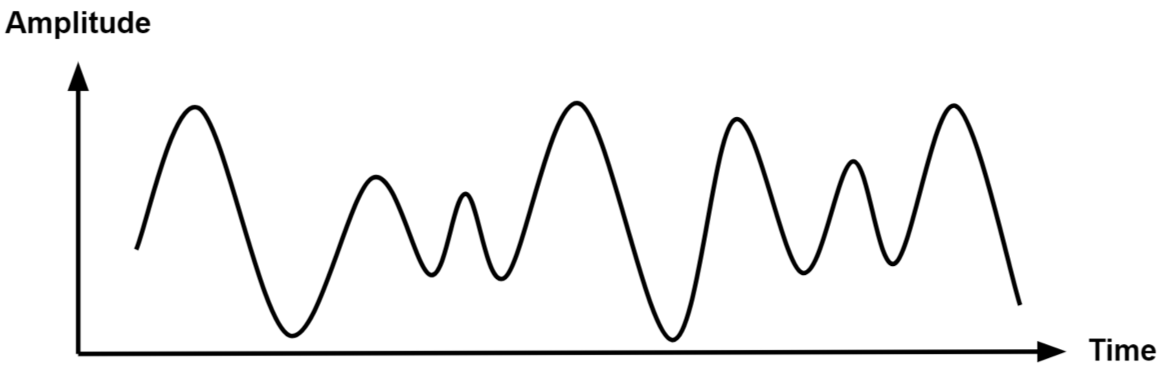
Communication System



Project 2: Wireless Communication

A ‘bit’ of history

Digital Systems



A Mathematical Theory of Communication

Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

Bits and Information

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits*, a word suggested by J. W. Tukey.

Introduced the term bits

2. THE DISCRETE SOURCE OF INFORMATION

We have seen that under very general conditions the logarithm of the number of possible signals in a discrete channel increases linearly with time. The capacity to transmit information can be specified by giving this rate of increase, the number of bits per second required to specify the particular signal used.

We now consider the information source. How is an information source to be described mathematically, and how much information in bits per second is produced in a given source? The main point at issue is the effect of statistical knowledge about the source in reducing the required capacity

A mathematical definition of information

Entropy and Compression

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant.

This theorem, and the assumptions required for its proof, are in no way necessary for the present theory. It is given chiefly to lend a certain plausibility to some of our later definitions. The real justification of these definitions, however, will reside in their implications.

Quantities of the form $H = -\Sigma\, p_i \log p_i$ (the constant K merely amounts to a choice of a unit of measure) play a central role in information theory as measures of information, choice and uncertainty. The form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics[8] where p_i is the probability of a system being in cell i of its phase space. H is then, for example, the H in Boltzmann's famous H theorem. We shall call $H = -\Sigma\, p_i \log p_i$ the entropy of the set of probabilities

Definition of entropy

&

How it acts as a lower limit for compression

(next week of classes)

Channels

Model of a noisy channel

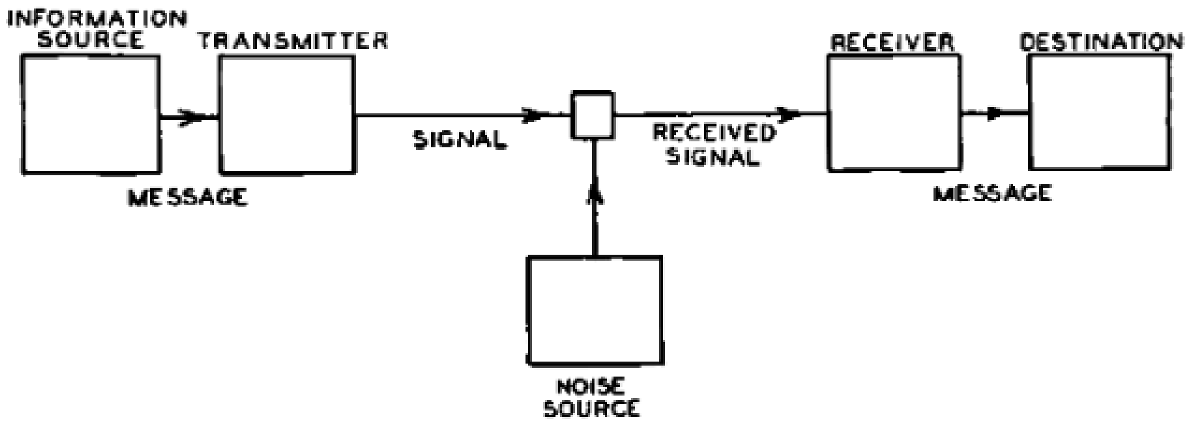


Fig. 1—Schematic diagram of a general communication system.

Source-Channel Separation Theorem:

For point-to-point communication, you can optimally compress data (source coding) and then reliably transmit it over a noisy channel (channel coding) by treating these tasks separately, without loss of performance compared to a joint system, as long as the source's rate is below the channel's capacity.

ENGR 76

ENGR 76

- What is information?
- How is it represented using bits?
- What makes compression possible?
- How are noisy environments and channels modeled?
- How do we manage reliable communication in a noisy environment?
- What makes multiplexing possible?
- What is the frequency domain representation and why is it important?

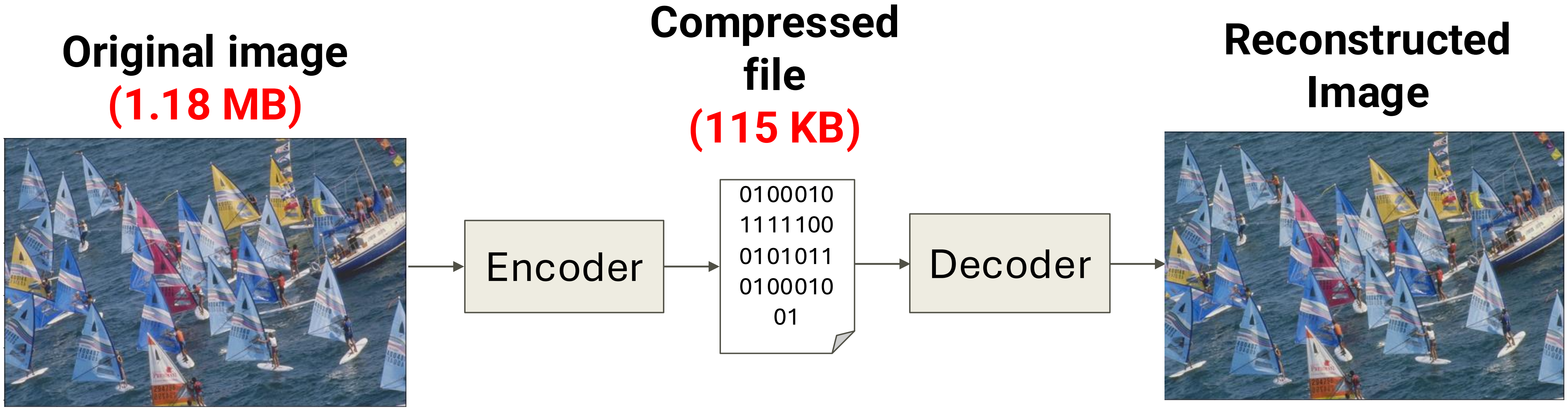
ENGR 76

- Explore some of the most important ideas and principles that enabled the “information age” we live in
- Develop the skills to bridge theory and practice - the art and craft of turning abstract ideas into real-world solutions
- Develop curiosity to go deeper into the working of these information systems
- Have fun!

Projects

- Explore hands-on how the tools learnt in class can be applied to different problems
  - Project 1: Image Compression
  - Project 2: Wireless Communication

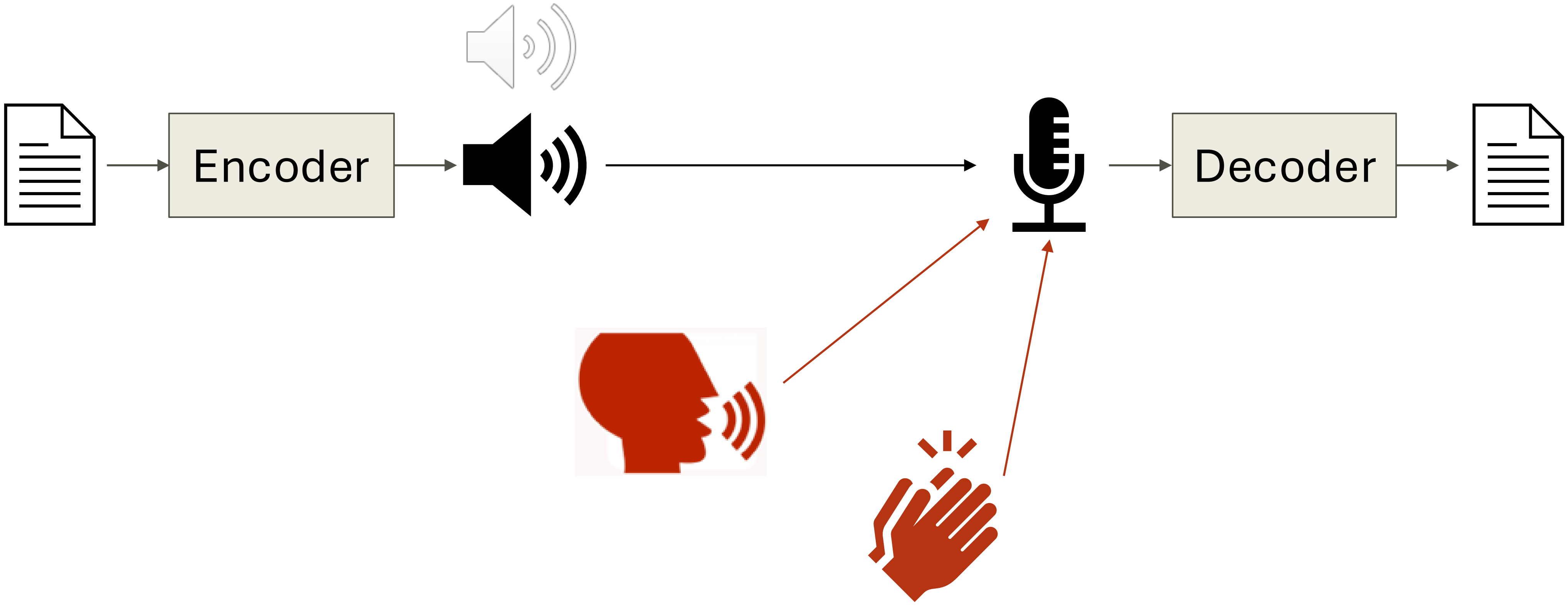
Project 1: Image Compression



Project 1: Image Compression

- Compress the same information into less space (LOSSLESS)
  - What is information? (Probability and Entropy)
  - Compression and fundamental limits (Huffman Coding)
- Remove information human eye can't see (LOSSY)
  - Frequency representation (Discrete Cosine Transform)
  - Quantization, Sampling and Reconstruction

Project 2: Wireless Communication



Project 2: Wireless Communication

- What is communication? (Modeling noisy channels)
- Converting bits to physical signals (Modulation)
- Bandwidth, spectrum shaping and sharing (Frequency-domain)
- Adding redundancy to correct errors (Error correcting codes)

Projects

- Logistics:
  - Project 0 – Python setup, tutorial, basic familiarization with images
  - Project 1a-1d, 2a-2c
  - Released every Friday, due next Friday 11:59pm
  - Solutions uploaded on Monday (to be used for the next assignment)
- Project Help Session
  - Uploaded to Canvas (video and slides)
  - Guidance on how to approach each project

Mini-PSets

- Questions on concepts covered in each week's class
- Released Thursday
- Due next Wednesday 11:59pm
- Discussion Sessions
  - Problems and solutions will be uploaded to Canvas

Exam

- In-class exam on Thursday of week 8 (February 26)
  - Both the theoretical concepts taught in lectures and the practical concepts covered in project assignments.
  - The questions corresponding to the theoretical concepts will be similar to the ones in mini-PSets and the associated discussion sections.

Grading

- Components:
  - Projects: 49%
  - Mini-PSets: 20%
  - Exam: 30%
  - Attendance (and participation): 1%

Late Policy

PLEASE READ CAREFULLY!

Late Policy

- **Mini-PSets**: We will drop your lowest mini-PSet score. No other extensions and exceptions will be provided.
- **Projects**: Weekly project assignments will be released each Friday, and they will be due the next Friday. The solutions will be uploaded on Monday. Each student will have a total of **6 late days for Project 0+1** (i.e., 0, 1a, ... , 1d) and **4 late days for Project 2** (i.e., 2a, ... , 2c). Once these late days are exhausted, any assignments turned in late will be penalized 10% per late day. However, no assignment will be accepted more than 3 days after its due date, at which point the solutions will be released automatically. Any fraction of a late day counts as one late day. These free late days are to be used only in cases of emergency and sickness. As assignments are released and due weekly, we strongly recommend using late days sparingly to avoid falling behind on subsequent deadlines. No exceptions to this policy will be granted to maintain structure and fairness in the class.

Feedback

- Non-anonymous feedback with each project
- Anonymous feedback form - https://forms.gle/PYYznnLyTBKdPa33A
  - Open throughout the course
  - Pinned on Ed
  - Please share any feedback at any time
  - Help improve the course both for you and others
  - Anonymous (requires Stanford login but email not shared)

Thank You!