I232 Information Theory
Chapter 7: Channel Coding and Channel Capacity

Brian Kurkoski

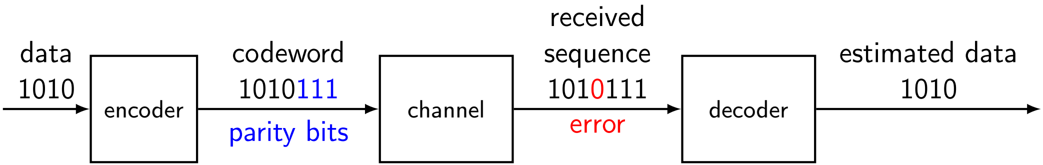
Japan Advanced Institute of Science and Technology

2023 May

Channel Coding Motivation

- Wireless Communicaitons
- Data storage — SSDs, flash memory, hard drives
- Optical communications
- Distributed data storage
- Blockchain

Reliable Communications Over Unreliable Channels



Code Rate R

$$R = \frac{\#\ \text{messasge bits}}{\#\ \text{codeword symbols}}$$

For example:

$$= \frac{\text{length } 1010}{\text{length } 1010111} = \frac{4}{7}$$

Central Question

Under what condition is
reliable communication possible?

$$R < C$$

where $C = \max I(\mathsf{X};\mathsf{Y})$ is capacity.

Fundamental Question for Channel Coding

Given an unreliable communications channel, what is the greatest rate at which reliable communications is possible?

Outline

7.1 Communication System Model

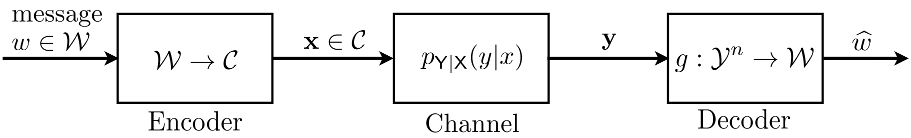


Figure 1: A model of a communications system.

- Encoder: encodes message w to codeword $\mathbf{x}$
- Channel: model of how noise occurs in transmission
- Decoder: decodes $\mathbf{y}$ to estimated message $\widehat{w}$

Goal: decoder output $\widehat{w}$ should be equal to w. Otherwise, an error has occurred.

7.1.1 Encoder

The encoder maps messages to codewords. The terms message, code, encoder and rate are defined as follows.

Definition

A *message* W is random variable representing one of M information symbols:

$$\mathcal{W} = \{1, 2, \ldots, M\}.$$

W is uniformly distributed.

Encoder — (M, n) code

Definition

An (M, n) *code* having a *codebook* $\mathcal{C}$ consists of M vectors:

$$\mathcal{C} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_M \end{bmatrix}$$

where each *codeword* $\mathbf{x}_i$ consists of n symbols:

$$\mathbf{x} = (x_1, x_2, \ldots, x_n),$$

with $x_i \in \mathcal{X}$.

The codebook alphabet is $\mathcal{X}$. For a binary code, $\mathcal{X} = \{0, 1\}$.

Encoder — Rate

Definition

The *rate* R of an (M, n) code is:

$$R = \frac{1}{n} \log M.$$

If we take log base 2, then the units of R is bits per transmission.

The rate R measures how much information a code can carry for each channel use:

- For a code with n symbols, the channel is used n times.
- The code carries $\log M$ bits of information — in other words, we need $\log M$ bits to select one of the codewords.
- Then, $\frac{1}{n} \log M$ is the number of bits transmitted per channel use.

Encoder

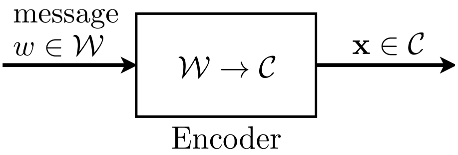


Figure 2: An encoder maps information to codewords.

Definition

An *encoder* is a mapping from the M messages of $\mathcal{W}$ to the M codewords of $\mathcal{C}$:

$$\mathcal{W} \to \mathcal{C}$$

Example of a Ternary code

Example

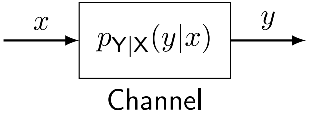
The table below gives an encoding mapping for a ternary code with $\mathcal{X} = \{0, 1, 2\}$:

message w	codeword $\mathbf{x}$
1	2 1 0 2 1 0
2	2 0 0 0 1 1
3	2 2 0 2 2 1
4	1 0 2 0 2 1

What are M, n and R?

7.1.2 Discrete Memoryless Channel (DMC)

The channel model is a DMC:



A *discrete memoryless channel* (DMC) consists of:

- an input alphabet $\mathcal{X}$,
- an output alphabet $\mathcal{Y}$ and
- a conditional probability distribution $p_{Y|X}(y|x)$.

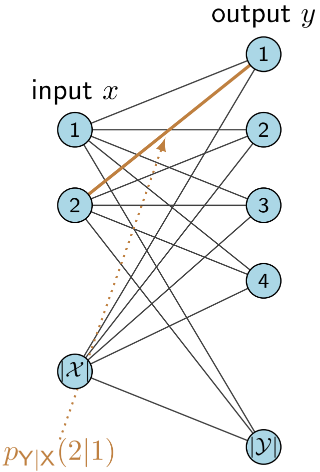


Figure 3: Transition diagram for DMC.

Discrete Memoryless Channel

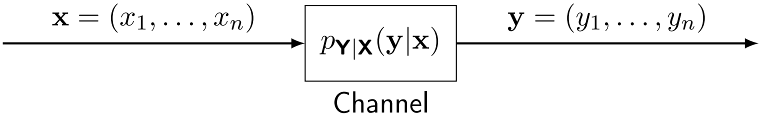


Figure 4: Use the channel n times.

- The codeword $\mathbf{x}$ is the input to the channel:

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

- The sequence $\mathbf{y}$ is the output of the channel:

$$\mathbf{y} = (y_1, y_2, \ldots, y_n)$$

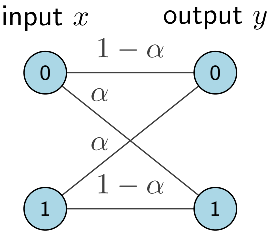
- For a memoryless channel, the joint conditional distribution is:

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p_{Y|X}(y_1|x_1) p_{Y|X}(y_2|x_2) \cdots p_{Y|X}(y_n|x_n)$$

Binary Symmetric Channel (BSC)

In the binary symmetric channel (BSC), an error occurs with probability α.

It has binary inputs and binary outputs.



The probability transition matrix $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ is:

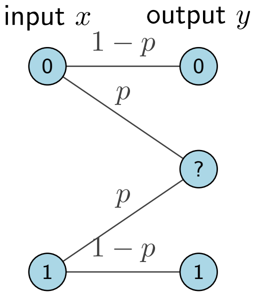
$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix},$$

where $0 \leq \alpha \leq 1$. There is no error with probability $1-\alpha$.

Binary Erasure Channel (BEC)

In the binary erasure channel (BEC), an erasure occurs with probability p.

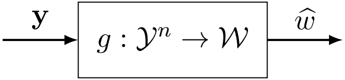
It has binary inputs, and three outputs: 0, 1 and an erasure symbol "?"



For a parameter $0 \leq p \leq 1$, the probability transition matrix $p_{\mathsf{Y|X}}(y|x)$ is:

$$p_{\mathsf{Y|X}}(y|x) = \begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}.$$

7.1.3 Decoder



Decoder

The *decoding function* g maps channel output $\mathbf{y}$ to estimated message $\widehat{w}$:

$$\widehat{w} = g(\mathbf{y})$$

If $\widehat{w} = w$ then there is no error. If $\widehat{w} \neq w$, then an error occurred.

Probabilities of error:

$$\begin{aligned} \text{conditional probability of error} \qquad & \lambda_w = \Pr\left(\widehat{\mathsf{W}} \neq w | \mathsf{W} = w\right) \\ \text{average probability of error} \qquad & P_e = \frac{1}{M} \sum_{w \in \mathcal{W}} \lambda_w \end{aligned}$$

7.2 Example Using Repeat Code

Repeat code:

- Repeats the message n times
- Simple, low code rate, can correct many errors.

Encoder Message set is $\mathcal{W} = \{0, 1\}$. With $n = 5$, codewords are $\mathbf{x}(0) = 00000$ and $\mathbf{x}(1) = 11111$.

Channel Binary symmetric channel (BSC) with error probability $\alpha = 0.1$ ★1

Decoder is "majority vote"

Majority Vote Decoder for Repeat Code

Majority vote decoding rule (n odd):

- If channel output $\mathbf{y}$ has $(n-1)/2$ or more zeros, then estimated message is $\widehat{w} = 0$.
- If channel output $\mathbf{y}$ has $(n-1)/2$ or more ones, then estimated message is $\widehat{w} = 1$.

The symbol with the most "votes" wins.

Example of decoding rule when $n = 5$:

$\mathbf{y}$ has ...	example $\mathbf{y}$	estimated codeword $\widehat{\mathbf{x}}$	estimated message $\widehat{w}$
0 ones	00000	00000	0
1 one	00010	00000	0
2 ones	10010		

Majority Vote Decoder for Repeat Code

Majority vote decoding rule (n odd):

- If channel output $\mathbf{y}$ has $(n-1)/2$ or more zeros, then estimated message is $\widehat{w} = 0$.
- If channel output $\mathbf{y}$ has $(n-1)/2$ or more ones, then estimated message is $\widehat{w} = 1$.

The symbol with the most "votes" wins.

Example of decoding rule when $n = 5$:

$\mathbf{y}$ has ...	example $\mathbf{y}$	estimated codeword $\widehat{\mathbf{x}}$	estimated message $\widehat{w}$
0 ones	00000	00000	0
1 one	00010	00000	0
2 ones	10010	00000	0
3 ones	01110		

Majority Vote Decoder for Repeat Code

Majority vote decoding rule (n odd):

- If channel output $\mathbf{y}$ has $(n-1)/2$ or more zeros, then estimated message is $\widehat{w} = 0$.
- If channel output $\mathbf{y}$ has $(n-1)/2$ or more ones, then estimated message is $\widehat{w} = 1$.

The symbol with the most "votes" wins.

Example of decoding rule when $n = 5$:

$\mathbf{y}$ has ...	example $\mathbf{y}$	estimated codeword $\widehat{\mathbf{x}}$	estimated message $\widehat{w}$
0 ones	00000	00000	0
1 one	00010	00000	0
2 ones	10010	00000	0
3 ones	01110	11111	1
4 ones	10111	11111	1
5 ones	11111	11111	1

★poll

Example: Repeat Code

Questions:

1. What is the code rate?

For BSC with $\alpha = 0.1$:

2. What is the probability of error λ_0 and λ_1?
3. What is the average probability of error n?

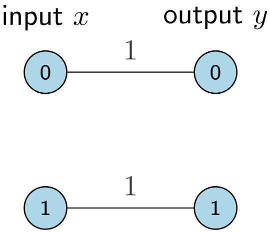
★2

7.3 Channel Capacity

7.3.1 Motivating Examples

Motivate channel capacity by considering how many bits simple channels can carry.

Consider the zero-error channel (error-free channel):



This channel adds no errors, for example:

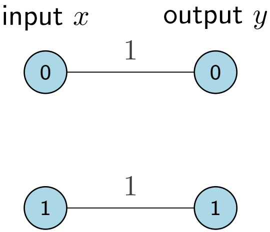
$$\text{input } \mathbf{x} = 1010100111$$
$$\text{output } \mathbf{y} = 1010100111$$

How many bits can be transmitted for each channel use? Answer:

7.3.1 Motivating Examples

Motivate channel capacity by considering how many bits simple channels can carry.

Consider the zero-error channel (error-free channel):



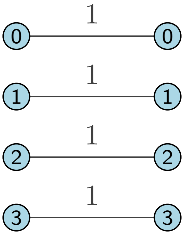
This channel adds no errors, for example:

$$\text{input } \mathbf{x} = 1010100111$$
$$\text{output } \mathbf{y} = 1010100111$$

How many bits can be transmitted for each channel use? Answer: **1 bit/channel use**

Channel Capacity: 4-input Example

Consider this four input, four output channel with no errors:

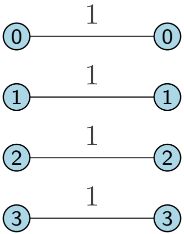


Note: You can choose one of the four inputs as the message to transmit.

How many bits can be transmitted for each channel use? Answer:

Channel Capacity: 4-input Example

Consider this four input, four output channel with no errors:

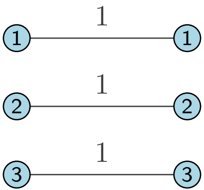


Note: You can choose one of the four inputs as the message to transmit.

How many bits can be transmitted for each channel use? Answer: **2 bit/channel use**

Channel Capacity: 3-input Example

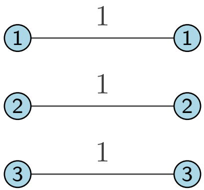
Consider this three input, three output channel with no errors:



How many bits can be transmitted for each channel use? Answer:

Channel Capacity: 3-input Example

Consider this three input, three output channel with no errors:



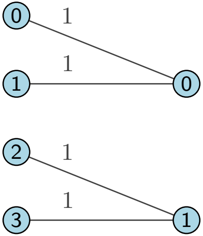
How many bits can be transmitted for each channel use? Answer:

$$\log 3 \approx 1.585 \textbf{ bits/channel use}$$

Deal with non-integer number of bits by averaging over many channel uses.

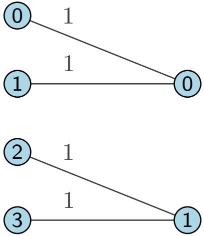
Channel Capacity: 4-Input, 2-Output Example

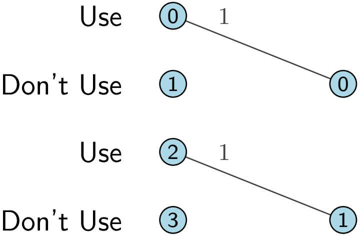
How many bits can be transmitted for each channel use?



Channel Capacity: 4-Input, 2-Output Example

How many bits can be transmitted for each channel use?





Answer: Only use two of the four inputs. **1 bit/channel use**

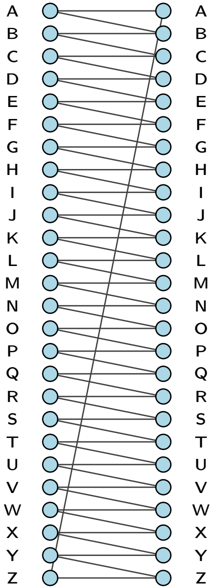
Key point: the choice of inputs is important.

Noisy Keyboard Channel

Suppose we have a noisy keyboard:

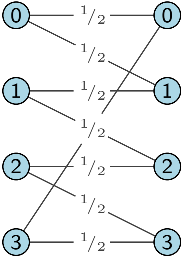
- If you press "A", the keyboard will output "A" or "B" with probability 0.5 each. Et cetera.
- While we can consider the capacity of this channel, consider a simpler channel on the next slide.

image credit: Wikipedia/Michael Maggs/CC BY-SA



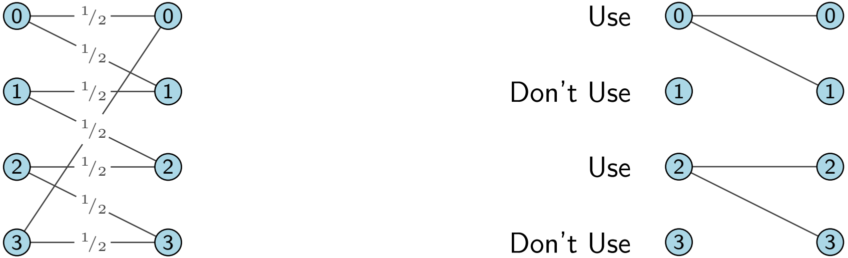
Channel Capacity: Simplified Keyboard Channel

How many bits can be transmitted for each channel use?



Channel Capacity: Simplified Keyboard Channel

How many bits can be transmitted for each channel use?



Answer: Only use two of the four inputs. **1 bit/channel use**

Key points: A channel has a maximum number of bits it can carry. Choosing how to use the inputs is important.

★poll

7.3.2 Definition of Channel Capacity

Definition

For a discrete memoryless channel $p_{\mathsf{Y}|\mathsf{X}}(y|x)$, the *"information" capacity* C of a discrete memoryless channel is:

$$C = \max_{p_{\mathsf{X}}(x)} I(\mathsf{X};\mathsf{Y}).$$

Definition

An optimal $p^*_{\mathsf{X}}(x)$ is called the *capacity-achieving input distribution*:

$$p^*_{\mathsf{X}}(x) = \arg\max_{p_{\mathsf{X}}(x)} I(\mathsf{X};\mathsf{Y}).$$

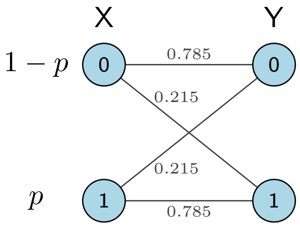
Five Properties of Channel Capacity

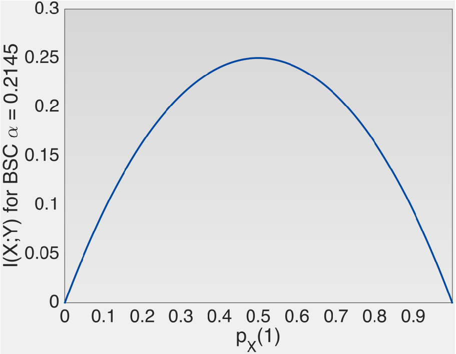
Five properties related to channel capacity are given:

1. $C \geq 0$.
2. $C \leq \log |\mathcal{X}|$,
3. $C \leq \log |\mathcal{Y}|$.
4. $I(\mathsf{X};\mathsf{Y})$ is a continuous function of $p_{\mathsf{X}}(x)$.
5. $I(\mathsf{X};\mathsf{Y})$ is a concave function of $p_{\mathsf{X}}(x)$.

Example: Capacity of BSC with $\alpha = 0.215$

For a BSC with $\alpha = 0.215$, the input distribution is $p_{\mathsf{X}}(0) = 1 - p$ and $p_{\mathsf{X}}(1) = p$.



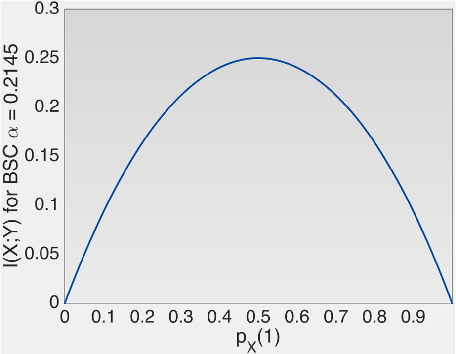


$I(\mathsf{X};\mathsf{Y})$ is a continuous function of p.
$I(\mathsf{X};\mathsf{Y})$ is a concave function of p.

Example: Capacity of BSC with $\alpha = 0.215$

For a BSC with $\alpha = 0.215$, the input distribution is $p_{\mathsf{X}}(0) = 1 - p$ and $p_{\mathsf{X}}(1) = p$.
Recall the definition of capacity:

$$C = \max_{p_{\mathsf{X}}(x)} I(\mathsf{X};\mathsf{Y})$$



What is the capacity?

$$C =$$

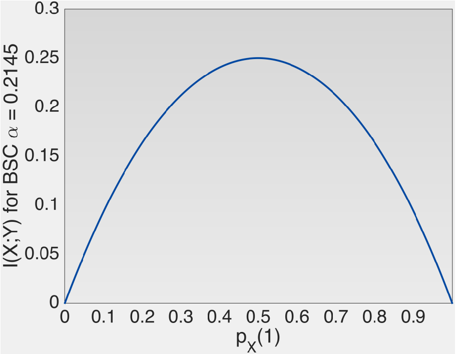
What is the capacity-achieving distribution?

$$p^* =$$

Example: Capacity of BSC with $\alpha = 0.215$

For a BSC with $\alpha = 0.215$, the input distribution is $p_{\mathsf{X}}(0) = 1 - p$ and $p_{\mathsf{X}}(1) = p$.
Recall the definition of capacity:

$$C = \max_{p_{\mathsf{X}}(x)} I(\mathsf{X};\mathsf{Y})$$



What is the capacity?

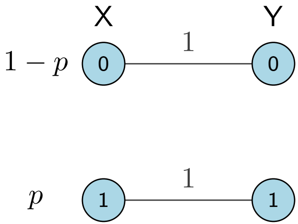
$$C = 0.25$$

What is the capacity-achieving distribution?

$$p^* = 0.5$$

7.3.3 Capacity of the Zero-Error Channel

Find the capacity of the zero-error channel having input distribution $p_X(x) = [1 - p, p]$:



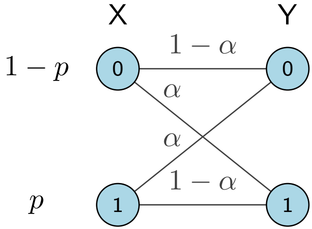
Clearly, the capacity is $C = 1$. Let's verify that analytically.

Recall the binary entropy function $h(p) = -p \log p - (1 - p) \log(1 - p)$.

★3

7.3.4 Capacity of the Binary Symmetric Channel (BSC)

Consider the general BSC with error probability $0 \leq \alpha \leq 1$ having input distribution $p_{\mathsf{X}}(x) = [1-p, p]$:



Proposition

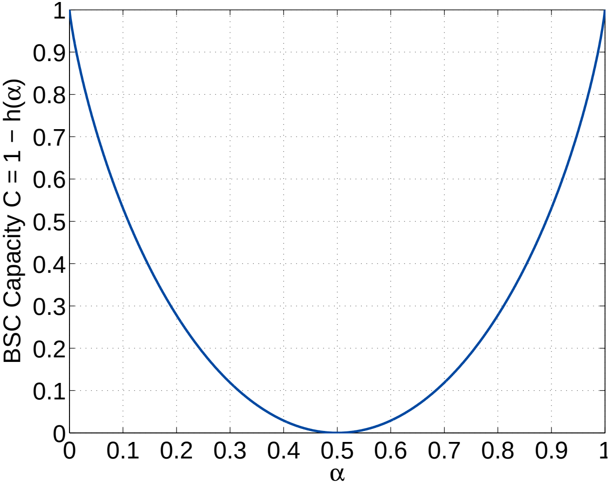
The capacity of the binary symmetric channel (BSC) with error probability α is:

$$C = 1 - h(\alpha)$$

with capacity-achieving input distribution $p^*_{\mathsf{X}}(x) = [\frac{1}{2}, \frac{1}{2}]$. ★4

Capacity of the BSC

The capacity of the BSC is $C = 1 - h(\alpha)$.



Class Info

- Tutorial Hours: Monday, May 8 at 13:30. Ask questions about homework.
- Homework 5 and 6 on LMS. Deadline Monday, May 8 at 18:00.
- Next lecture: Wednesday, May 10. Channel Coding Theorem. There will be a pop quiz on Fano's inequality — understand the proof of Fano's inequality.
- Midterm exam on May 15 at 13:30.
- Homework 7 on LMS (soon)

Midterm Exam

The exam is closed book. You may use:

- One page of notes, A4-sized paper, double-sided OK.
- Blank scratch paper

You may not use anything else: No printed materials, including books, lecture notes, and slides. No notes (except as above). No internet-connected devices. No calculators. You may need to perform a 2×2 matrix inverse.

Exam Content

- Covers Chapters 1–6
- Study Homework 1–6. Solutions to Homework 1–6 are provided.
- No programming questions.

Practice problems will be provided.