I427 Coding Theory
Chapter 1: Error Correcting Codes

Brian Kurkoski

Japan Advanced Institute of Science and Technology

2023 October

Contents

1.1 Error Correcting Code Concepts

Success of Coding Theory

Coding theory is:

- the design of error-correcting codes for
- reliable communications over unreliable channels

If you use media on your smartphone, then you used an error-correcting code:

- error-correcting codes for WiFi and mobile data
- error-correcting codes for reliable flash storage

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

Unreliable Communications Channels and Applications

Examples of unreliable communications channels

- Wireless communication channels are unreliable due to noise and interference
- Flash memories are unreliable due to device failure and endurance problems.

Unreliable Communications Channels and Applications

Examples of unreliable communications channels

- Wireless communication channels are unreliable due to noise and interference
- Flash memories are unreliable due to device failure and endurance problems.

Communications applications in **space**:

Mobile data and WiFi LDPC codes, polar codes, convolutional codes

Digital Television LDPC codes, BCH codes

ADSL/fiber optic LDPC codes, convolutional codes

Unreliable Communications Channels and Applications

Examples of unreliable communications channels

- Wireless communication channels are unreliable due to noise and interference
- Flash memories are unreliable due to device failure and endurance problems.

Communications applications in **space**:

Mobile data and WiFi LDPC codes, polar codes, convolutional codes

Digital Television LDPC codes, BCH codes

ADSL/fiber optic LDPC codes, convolutional codes

Communications applications in **time**, i.e. data storage:

Flash memories LDPC codes

Distributed Storage Reed-Solomon codes

Hard disk drives, DVDs and CDs Reed-Solomon codes

Information Theory vs. Coding Theory

The two areas of **information theory**[1] and **coding theory** deal with reliable communications over unreliable channels.

Information Theory:

- Deals with fundamental limits of communications
- Block length is infinite
- Does not tell us about practical code design

Coding Theory:

- Design of practical, finite-length error-correcting codes
- Goal is low rate of decoding errors for communication systems
- Complexity of the decoder should be low

[1] Taught at JAIST as I232 Information Theory

Two Types of Error Correcting Codes

Finite-field error-correcting codes

- Defined on a finite field
- Most are binary error-correcting codes
- Widely used in practice

Two Types of Error Correcting Codes

Finite-field error-correcting codes

- Defined on a finite field
- Most are binary error-correcting codes
- Widely used in practice

Lattices

- Defined on the real numbers
- Uses the same real-valued algebra as the wireless channel
- Lattices for wireless networks: recent research results assuming infinite-length lattices
- Many opportunities for research in finite-length lattices and their applications

Unreliable Channels

SEND␣$100

SEND␣$500

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

Adding Redundancy

- Errors can be reduced by using redundancy[2]
- Language naturally has redundancy

Suppose you received the following message, where some symbols have been erased:

`DO YO_ HA_E ANY QUES_IO_S?`

Because of the redundancy of language, you can see the original message is:

[2] redundancy: inclusion of extra components which are not strictly necessary to functioning, in case of failure in other components

Adding Redundancy

- Errors can be reduced by using redundancy[2]
- Language naturally has redundancy

Suppose you received the following message, where some symbols have been erased:

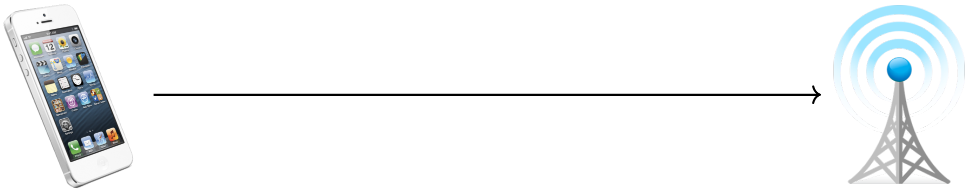
```
DO YO_ HA_E ANY QUES_IO_S?
```

Because of the redundancy of language, you can see the original message is:

```
DO YOU HAVE ANY QUESTIONS?
```

[2] redundancy: inclusion of extra components which are not strictly necessary to functioning, in case of failure in other components

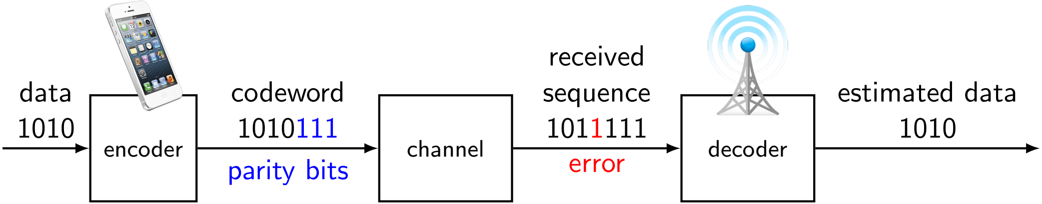
Communications Uses Binary Data



1010 → 1011

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

Error-Correcting Codes: Reliable Communications over Unreliable Channels

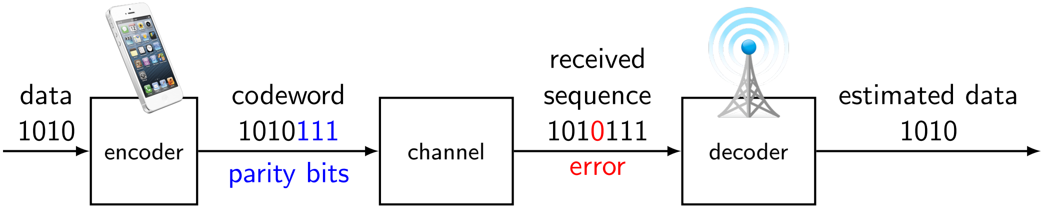


This problem is solved using channel coding:

- An **encoder** adds parity bits (or redundancy) — the message is now longer
- The channel may be a probabilistic model of the errors
- The **decoder** recovers the original message — if there are not too many errors.

Coding theory is the design of the encoder and decoder

Channel Coding: Reliable Communications over Unreliable Channels



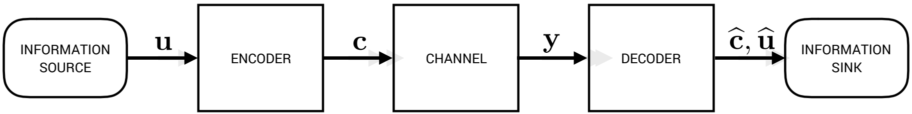
The code rate R is an important part of an error-correcting code

$$\text{Code rate } R = \frac{\#\text{ of message bits}}{\#\text{ of codeword bits}} = \frac{4}{7} \approx 0.571$$

High code rate R carries more information. Low rate R is more reliable.

1.2 Communication System Model

Model of a communication system with encoder, channel and decoder.



- Information source: Produces information sequence $\mathbf{u}$
- Encoder: Maps $\mathbf{u}$ to codeword $\mathbf{c}$
- Channel: Output $\mathbf{y}$ is noise version of $\mathbf{c}$
- Encoder: Outputs estimate $\widehat{\mathbf{c}}$ and $\widehat{\mathbf{u}}$ which is "similar to" $\mathbf{y}$
- Information sink: If $\widehat{\mathbf{u}} = \mathbf{u}$, then success! Otherwise, errors occurred.

Formal Definition

Definition

An *error-correcting code* $\mathcal{C}$ consists of M codewords:

$$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\}$$

where each codeword (or constellation point) is an n-tuple:

$$\mathbf{c} = [c_1, c_2, \ldots, c_n].$$

Each c_i is from a specified alphabet, often the binary alphabet.

Definition

The *code rate* R of a code is:

$$R = \frac{1}{n} \log_2 M$$

★3

Example: Repeat-by-7 Code

Consider the repeat-by-seven code. The information source produces $u = 0$ or $u = 1$, and encodes to $\mathbf{c}$ according to:

u	$\mathbf{c}$
0	0000000
1	1111111

★1

Example: Hamming Code

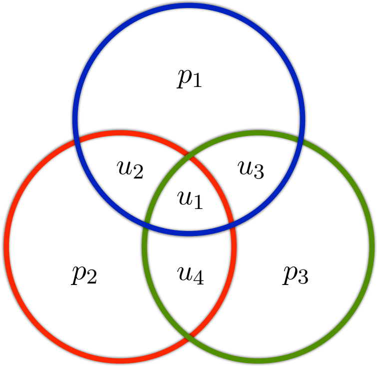
Graphical representation of a Hamming code.

- Information bits are u_1, u_2, u_3, u_4
- Parity bits are p_1, p_2, p_3.
- Codeword is
  $\mathbf{c} = [u_1, u_2, u_3, u_4, p_1, p_2, p_3]$

The seven code bits must satisfy the following condition:

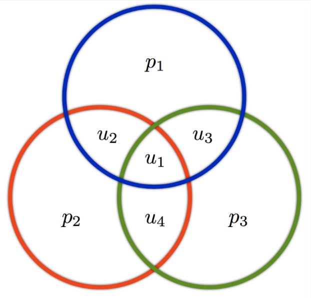
The number of 1's inside each circle must be even.

★2



Self-Study Question (SSQ) on LMS

SSQ (Self-Study Quiz): Decoding Hamming (7,4) code



A (7,4) Hamming codeword of the form $\mathbf{c} = [u_1, u_2, u_3, u_4, p_1, p_2, p_3]$ is transmitted. Decode each sequence $\mathbf{y}$ to the corresponding codeword $\hat{\mathbf{c}}$:

(a) $\mathbf{y} = [0, 0, 0, 0, 0, 0, 1]$. $\hat{\mathbf{c}} =$

Which is better?

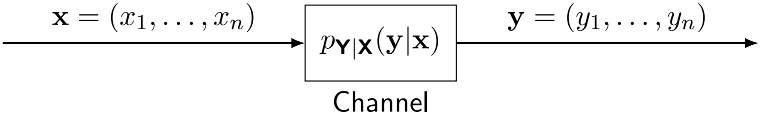
Which of the following two codes is better?

- Repeat-7 code, or
- Hamming code?

Both use the channel 7 times.

1.3 Channel Models

Memoryless Channels



- 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:

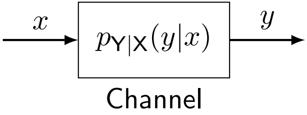
$$\mathbf{y} = (y_1, y_2, \ldots, y_n)$$

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

$$\begin{aligned} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= p_{\mathsf{Y}|\mathsf{X}}(y_1|x_1)p_{\mathsf{Y}|\mathsf{X}}(y_2|x_2)\cdots p_{\mathsf{Y}|\mathsf{X}}(y_n|x_n) \\ &= \prod_{i=1}^{n} p_{\mathsf{Y}|\mathsf{X}}(y_i|x_i). \end{aligned}$$

1.3.1 Discrete Memoryless Channel (DMC)

Channel model:



In a general DMC the output y is **discrete**

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

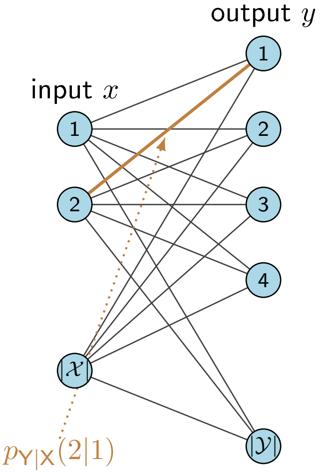
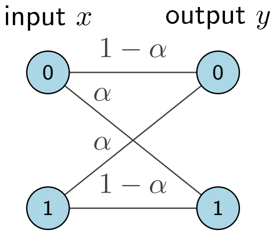


Figure 1: Transition diagram for DMC.

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:

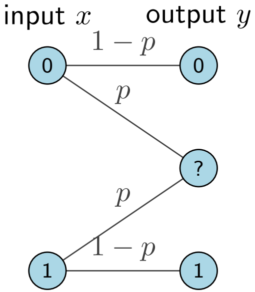
$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix},$$

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 "?"

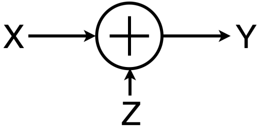


For a parameter $0 \leq p \leq 1$, the probability transition matrix $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ is:

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

1.3.2 Gaussian Channel Model

The additive-white Gaussian noise (AWGN) channel model is:



X, Y and Z are continuous random variables.

- **Input:** $\mathsf{X} \sim p_{\mathsf{X}}(x)$, Transmit power constraint: $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$
- **Noise:** $\mathsf{Z} \sim \mathcal{N}(0, N)$, Gaussian with noise power N
- **Output:** $\mathsf{Y} = \mathsf{X} + \mathsf{Z}$

Transmit Power Constraint P

The power of a signal is the square of its value.

Transmit Power Constraint P

The power of a signal is the square of its value.

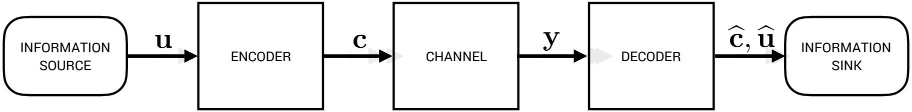
Maximum signal power P is the channel input power constraint:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$$

1.4 Channel Capacity

- Question: Given a channel $p_{\mathsf{Y}|\mathsf{X}}(y|x)$, what is the highest possible communications rate with 0 errors?
- Answer: The channel capacity C.

Channel Capacity



- Decoding error probability: word error rate (WER), bit-error rate (BER):

$$\text{WER} = \Pr(\mathbf{u} \neq \widehat{\mathbf{u}} | \mathbf{u} \text{ was transmitted}, \mathbf{y} \text{ was received}).$$
$$\text{BER} = \Pr(u_i \neq \widehat{u}_i | u_i \text{ was transmitted}, \mathbf{y} \text{ was received})$$

- Channel capacity is maximum possible communications rate
- Must let block length n go to infinity.

Mutual Information

Definition

Consider random variables X and Y with a joint probability distribution function $p_{\mathsf{X},\mathsf{Y}}(x, y)$ and marginal distributions $p_{\mathsf{X}}(x)$ and $p_Y(y)$. Then $I(\mathsf{X};\mathsf{Y})$ is given by:

$$I(\mathsf{X};\mathsf{Y}) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{\mathsf{X},\mathsf{Y}}(x, y) \log \frac{p_{\mathsf{X},\mathsf{Y}}(x, y)}{p_{\mathsf{X}}(x) p_{\mathsf{Y}}(y)}.$$

High mutual information means that X tells you a lot about Y.

Definition of Channel Capacity

Definition

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

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

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

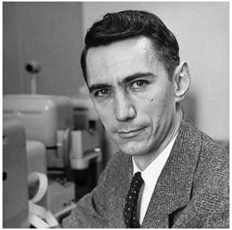
$$p^*_{\mathsf{X}}(x) = \arg\max_{p_{\mathsf{X}}(x)} I(\mathsf{X};\mathsf{Y}).$$

Shannon's Channel Coding Theorem

Proposition

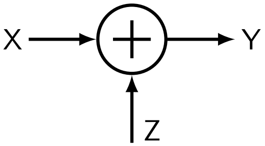
Channel Coding Theorem For every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with probability of decoding error going to 0 as $n \to \infty$. Conversely, any sequence of $(2^{nR}, n)$ codes with probability of decoding error going to 0 must have $R \leq C$.

- A channel has a capacity C
- This is the upper limit for the code rate R
- This limit can only be achieved for $n \to \infty$
- Finite-length codes cannot achieve capacity



Claude E. Shannon

Gaussian Channel Capacity



Proposition

The capacity of AWGN channel with power constraint P and noise variance σ^2 is:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2}\right) \text{ bits per transmission}$$

The capacity-achieving input distribution is $p_X^*(x)$ is a zero-mean Gaussian with variance P.

Capacity of the Binary-Input AWGN Channel

For the special case when the input alphabet is limited to $\mathcal{X} = \{-1, +1\}$, we have the binary-input AWGN channel. The capacity of this channel is given by:

$$C = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(y-1)^2/2\sigma^2} \log_2 \frac{2}{1 + e^{-2y/\sigma^2}} dy$$

and is achieved by choosing $p_{\mathsf{X}}(0) = p_{\mathsf{X}}(1) = \frac{1}{2}$.

Capacity of AWGN and BI-AWGN Channels

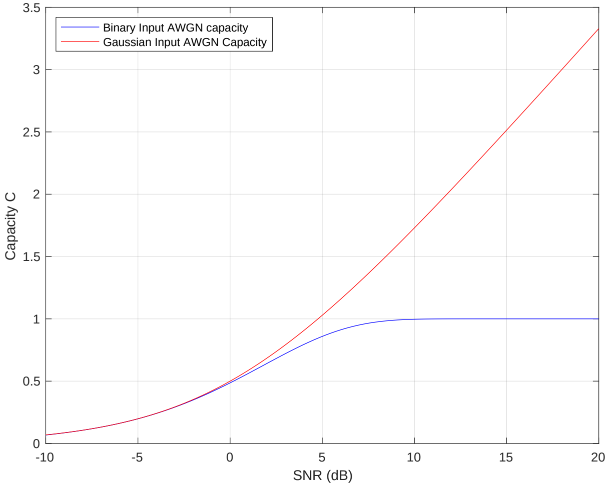


Figure 2: Capacity of the AWGN channel, with Gaussian inputs and binary inputs (BI-AWGN). Left side: power-limited domain. Right side: bandwidth-limited domain.

The Challenge: How to Achieve Channel Capacity

- The channel coding theorem assumes random codes, which are not practical to use.

How to design *practical* codes for reliable communications?

- For any finite n, there will be a nonzero probability of error
- As n gets larger the probability of error may decrease
- Decoding complexity is a concern.

Challenge: How to Achieve Capacity?

Given code rate $R = 1/2$, what is the worse channel we can use?

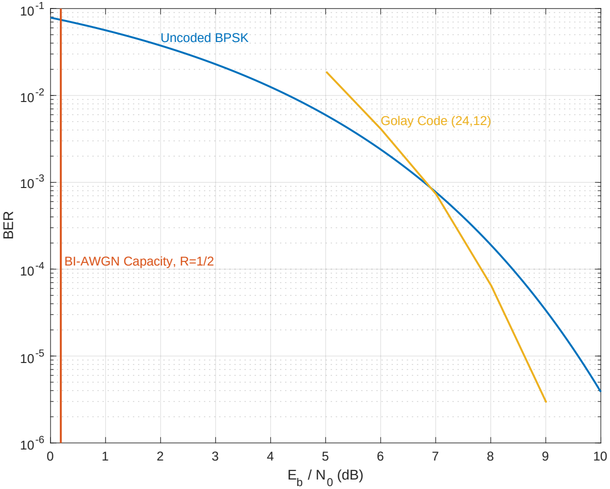


Figure 3: Bit-error rate for the uncoded BPSK, the extended Golay code (24,12), and the

Homework

Homework 1 is due October 18 at 18:00. See LMS for details.