EE653 - Coding Theory

Lecture 1: Introduction & Overview

Dr. Duy Nguyen

San Diego State University
Leadership Starts Here

Outline

Administration

- Hours and Location
  - Lectures: MW 4:00pm – 5:15pm
  - Location: P-148
  - Office hours: MW 2:00pm – 3:00pm or by email appointments

- Course webpage:
  http://engineering.sdsu.edu/~nguyen/EE653/index.html

- Instructor:
  - Name: Dr. Duy Nguyen
  - Office: E-408
  - Phone: 619-594-2430
  - Email: duy.nguyen@sdsu.edu
  - Webpage: http://engineering.sdsu.edu/~nguyen

- Teaching Assistant: N/A

Syllabus

- Prerequisite
  - EE 558 - Digital Communications
  - Knowledge of MATLAB programming

- References

1. Shu Lin and Daniel J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, 2nd Ed., Prentice Hall, 2004.
2. B. Sklar, *Digital Communications: Fundamentals and Applications*, 2nd Ed., Prentice Hall, 2001.
3. J. Proakis, *Digital Communications*, 4th Ed., McGraw-Hill, 2000.

Assessments

- Assessments: 20% Homework, 15% Quiz, 15% Midterm Exam, 20% Project, and 30% Final Exam (Open-Book)
- Homework assignments: Bi-weekly, Total: 5. Late submission: maximum 1 day, 20% score deducted
- Research Project: In-depth study or original research topic
  - Project Proposal: 1 page (%5)
  - Project Report: 5-7 pages (double-column) (%10)
  - Presentation: 15 minutes - End of semester (%5)
- Midterm: Monday, Mar 06
- Final: Monday, May 08 at 15:30 – 17:30
- Grades:

90–100	A/−
75–89	B/±
60–74	C/±
50–59	D/+

Schedule

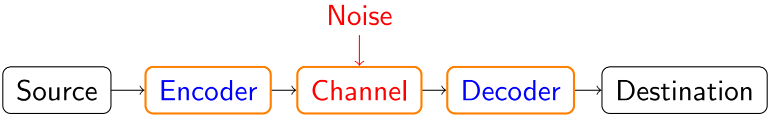
Week	Day	Task	Week	Day	Task
1	M		9	M	
Jan 16	W	First day of class	Mar 13	W	
2	M		10	M	HW4 out, HW3 due
Jan 23	W		Mar 20	W	
3	M	HW1 out	BREAK	M	Spring break
Jan 30	W		Mar 27	W	Spring break
4	M		11	M	Quiz 2
Feb 6	W		Apr 3	W	
5	M	HW2 out, HW1 due	12	M	HW5 out, HW4 due
Feb 13	W		Apr 10	W	
6	M	Quiz 1	13	M	Quiz 3
Feb 20	W		Apr 17	W	
7	M	HW3 out, HW2 due	14	M	HW5 due
Feb 27	W		Apr 24	W	
8	M	Midterm Exam	15	M	Project presentation
Mar 6	W	Project proposal due	May 1	W	Final Report due

Topics to Cover

- Mathematical background
  - Related background on Abstract Algebra
- Linear block codes
  - Hamming codes
  - Reed-Muller codes
- Cyclic codes
  - Cyclic codes
  - BCH codes
  - Reed-Solomon codes
- Convolutional codes
- Advanced Topics: Turbo codes, Low-Density Parity Check (LDPC) codes, trellis coded modulation (TCM), bit-interleaved coded modulation (BICM)

Outline

What is Coding for?



- Source Coding
  - The process of compressing the data using fewer bits to remove redundancy
  - Shannon's source coding theorem establishes the limits to possible data compression: entropy
- Channel Coding or Error Control Coding
  - The process of adding redundancy to information data to better withstand the effects of channel impairments
  - Shannon-Hartley's capacity theorem establishes the limits for data transmission with an arbitrary small error probability

What is Source Coding?

- Forming efficient descriptions of information sources
- Reduction in memory to store or bandwidth resources to transport sample realizations of the source data
- Discrete sources: entropy to define the average self-information for the symbols in an alphabet

$$\mathrm{H}(X) = -\sum_{j=1}^{N} p_j \log_2(p_j)$$

- Maximum entropy with equal probability $1/N$ for all symbols

$$0 \leq \mathrm{H}(X) \leq \log_2(N)$$

- Compress source signals to the entropy limit
- Examples: entropy of binary sources

What is Error Control Coding?

- Coding for reliable digital storage and transmission

  "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." (Claude Shannon 1948)

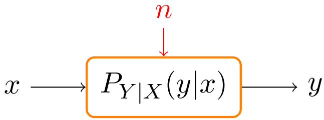
- Proper encoding can reduce errors to any desired level as long as the information rate is less than the capacity of the channel

- What is Error Control Coding?
  - Adding redundancy for error detection and/or correction
  - Automatic Repeat reQuest (ARQ): error detection only - easy and fast with parity check bits. If there is an error, retransmission is necessary (ACK vs NAK)
  - Forward ECC: both error detection and correction - more complicated encoding and decoding techniques

- Focus of this course: channel encoding and decoding!

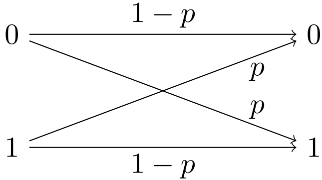
Communication Channel

- Physical medium: used to send the signal from TX to RX
- Describe the transition probability from input to output



- *Noiseless binary channel:* input is reproduced exactly at output

- *Binary symmetric channel:* cross probability p



Channel Capacity

- Example of AWGN channel: $y = x + n$, $n \sim \mathcal{N}(0, N)$, $\mathbb{E}\left[|x|^2\right] = S$
  - Mutual information

$$\mathrm{I}(x;y) = \mathrm{H}(y) - \mathrm{H}(y|x)$$

  - Capacity of a channel

$$C = \max_{p(x_i)} \mathrm{I}(x;y)$$

  - Gaussian distribution has the highest entropy

$$\mathrm{H}(y|x) = \mathrm{H}(n) = \frac{1}{2}\log\left[2\pi e N\right]$$

  - $\mathrm{H}(y)$ is maximum if y is Gaussian $\rightarrow$ x is also Gaussian

$$\mathrm{H}(y) = \frac{1}{2}\log\left[2\pi e(S+N)\right]$$

  - Shannon-Hartley theorem on channel capacity with Gaussian input

$$C = \frac{1}{2}\log\left(1 + \frac{S}{N}\right) \quad \text{nats/s/Hz}$$

Outline

Example 1: Repetition Code

- Repetition code: Repeat each bit $(n-1)$ times
- Code rate $1/n$, denoted as R_n
- Encoding rule for R_5 code:
  - $0 \rightarrow 00000$
  - $1 \rightarrow 11111$
- Decoding rule:
  - Majority decoding rule: choose bit that occurs more frequently

- Example with R_5 code: We have information bits 10. After encoding, we have 1111100000. If 0110111000 is received (some bits are in error):
  - We first decode 01101 to 1
  - We then decode 11000 to 0
  - Decoded bits: 10

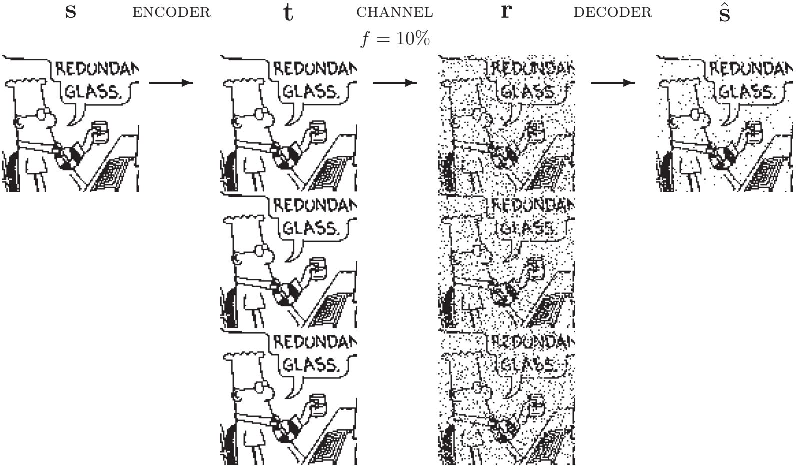
How Good Is Repetition Code?

- Without *repetition code*, assume the probability of error is p
- With R_n code, the probability of error is:

$$P_E = \sum_{i=(n+1)/2}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

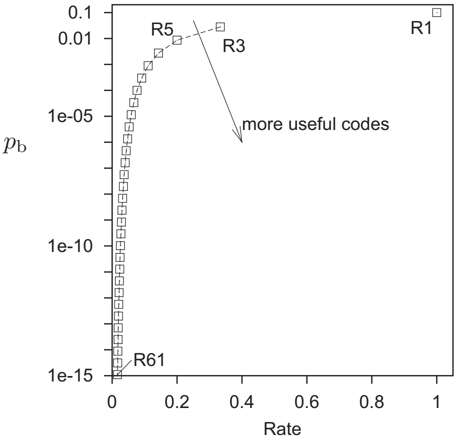
- Repetition is the simplest code: Is it a good code?
- With $p = 10^{-1}$ and R_3 code, overall error P_E is 2×10^{-2}
- Not good if n is small. If n is large: Overhead burden

How Good is Repetition Code?



Source: David J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*.

How Good is Repetition Code?



Source: David J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*.

Example 2: Cyclic Redundancy Check (CRC)

- Check values are added to information. If the check values do not match, re-transmission is requested
- CRC: Used for error detection, not correction
- Simple to implement in binary hardware, easy to analyze mathematically, and particularly good at detecting common errors
- Commonly used in digital networks and storage devices; Ethernet and many other standards
- CRC is a special case of Cyclic Codes
- In this course, most of the time, the focus is on Forward Error Correction (FEC): a one-way system employing error-correcting codes that automatically correct errors detected at the receiver

What is a "Good" Code?

- For a bandwidth W, power P, Gaussian noise power spectral density N_0, there exists a coding scheme that drives the probability of error arbitrarily close to 0, as long as the transmission rate R is smaller than the Shannon capacity limit C:

$$\boxed{C = W \log_2 \left(1 + \frac{P}{W N_0}\right) \quad \text{(bits/s)}}$$

- Consider the normalized channel capacity (spectral efficiency) $\eta = C/W\,\text{(bits/s/Hz)}$ with $P = C E_b$, where E_b: energy per bit:

$$\eta = \frac{C}{W} = \log_2 \left(1 + \frac{C}{W} \frac{E_b}{N_0}\right)$$

- Then we have

$$\boxed{\frac{E_b}{N_0} = \frac{2^\eta - 1}{\eta}}$$

[1] Claude E. Shannon, *A Mathematical Theory of Communication.* Bell System Technical Journal, 27, 379–423 & 623–656, 1948.

Capacity Approaching Coding Schemes

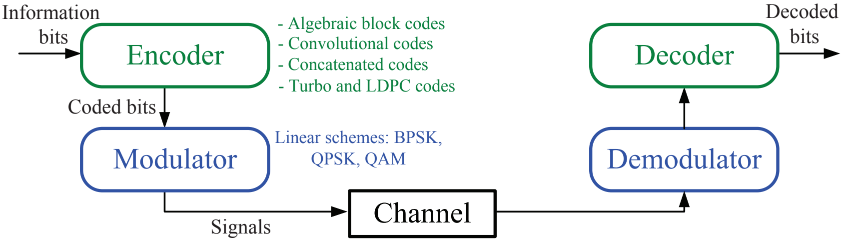
- If $R > C$: no way for a reliable transmission
- If $R \leq C$: the results of the theorem were based on the idea of random coding
  - The theorem was proved using random coding bound
  - Block length must goes to infinity
- No explicit/practical coding scheme was provided
- A holy grail for communication engineers and coding theorist
  - Finding a scheme with performance close to what was promised by Shannon: Capacity-approaching schemes
  - Complexity in implementation of those schemes
- High performing coding schemes only found very recently!

A Brief History of Error Control Coding

- Linear block codes: Hamming code (1950), Reed-Muller code (1954)
- Cyclic codes: BCH code (1960), Reed-Solomon (1960)
- LDPC, 1963
- TCM, 1976 & 1982
- Turbo codes, 1993
- BICM, 1996
- The rediscovery of LDPC, 1996
- Fountain codes: LT code (2003), Raptor code (2006)
- Polar code, 2009

Outline

Digital Communication System



Digital Communication System

- Information $\mathbf{u} = 1001$; Using repetition code R_3, we have coded bits $\mathbf{v} = 111000000111$
- Now we can use BPSK modulation scheme:

$$\text{Bit 0:} \; -\sqrt{E_s}\sqrt{\frac{2}{T_b}}\cos\left(2\pi f_c t\right)$$
$$\text{Bit 1:} \; +\sqrt{E_s}\sqrt{\frac{2}{T_b}}\cos\left(2\pi f_c t\right)$$

- Baseband model $r[m] = x[m] + w[m]$, with $x[m] = \pm\sqrt{E_s}; w[m] \sim \mathcal{N}(0, N_0/2)$: AWGN
- What can we do with $r[m]$? Hard-decision decoding and soft-decision decoding

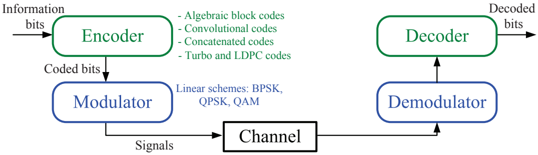
Digital Communication System

- If hard-decision decoding, the uncoded bit error probability is $p = \mathcal{Q}\left(\sqrt{2E_s/N_0}\right)$. We will then have a binary symmetric channel (BSC) with transition probability p
- Here, $\mathcal{Q}(x)$ is the complementary error function, defined as

$$\mathcal{Q}(x) = \frac{1}{2\pi}\int_x^{\infty} \mathrm{e}^{-y^2/2}\mathrm{d}y$$

- Given p, we should be able to calculate the bit error probability of our information sequence
- Soft-decision decoding: offers significant performance. We will talk later on about it

Maximum Likelihood Decoding



- Information $\mathbf{u}$; coded information or coded bits $\mathbf{v}$
- After modulation, we have transmitted signals $\mathbf{x}$. For the moment, let's assume we use BPSK so that length of $\mathbf{v}$ and $\mathbf{x}$ are the same
- At the receiver, we receive $\mathbf{r}$. From $\mathbf{r}$, the decoder needs to produce an estimate $\hat{\mathbf{u}}$
- Equivalently, since there is one-to-one correspondence between information sequence $\mathbf{u}$ and coded sequence $\mathbf{v}$, the decoder can produce an estimate $\hat{\mathbf{v}}$

Maximum Likelihood Decoding

- Clearly, $\hat{\mathbf{u}} = \mathbf{u}$ if and only if $\hat{\mathbf{v}} = \mathbf{v}$.
- A decoding rule is a strategy for choosing an estimated of $\hat{\mathbf{v}}$ for each possible received sequence $\mathbf{r}$, e.g., the hard decision decoding rule.
- Given that $\mathbf{r}$ is received, the *conditional error probability* of the decoder is defined as:

$$P(E|\mathbf{r}) \triangleq P\left(\hat{\mathbf{v}} \neq \mathbf{v}|\mathbf{r}\right)$$

- The error probability of the decoder is then given by:

$$P(E) = \sum_{\mathbf{r}} P(E|\mathbf{r})P(\mathbf{r})$$

Maximum Likelihood Decoding

- $P(\mathbf{r})$ is independent of decoding rule, since $\mathbf{r}$ is produced prior to decoding. Hence, an optimal decoding rule, that is, one that minimize $P(E)$ must minimize $P(E|\mathbf{r}) = P\left(\hat{\mathbf{v}} \neq \mathbf{v}|\mathbf{r}\right)$ for all $\mathbf{r}$.
- Now, note that minimizing $P\left(\hat{\mathbf{v}} \neq \mathbf{v}|\mathbf{r}\right)$ is equivalent to maximizing $P\left(\hat{\mathbf{v}} = \mathbf{v}|\mathbf{r}\right)$. Therefore, an optimal decoding rule is to choose a codeword $\mathbf{v}$ that maximizes

$$P(\mathbf{v}|\mathbf{r}) = \frac{P(\mathbf{r}|\mathbf{v})P(\mathbf{v})}{P(\mathbf{r})}$$

- If we assume all information sequences $\mathbf{u}$ are equally likely, it would be the same for all coded sequences $\mathbf{v}$. As such, $P(\mathbf{v}$ are the same for all $\mathbf{v}$. It means that for an optimal decoding rule, we need to find a codeword $\mathbf{v}$ to maximize $P(\mathbf{r}|\mathbf{v})$: Maximum Likelihood Decoding (MLD) rule.

MLD with DMC and BSC

- If we assume channel is discrete and memoryless channel (DMC), i.e., each received symbol r_i depends only on the corresponding transmitted symbol x_i (or v_i), we have $P(\mathbf{r}|\mathbf{v}) = \prod_i P(r_i|v_i)$
- So MLD is equivalent to maximize the log-likelihood function:

$$\log P(\mathbf{r}|\mathbf{v}) = \sum_i \log P(r_i|v_i)$$

  i.e, we need to choose $\mathbf{v}$ to maximize the above sum
- Now, we consider a special case of BSC channel, i.e., $\mathbf{r}$ is a binary sequence that may differ from transmitted sequence $\mathbf{v}$ in some positions owning to the channel noise. For this BSC, assume when $r_i \neq v_i$, $P(r_i|v_i) = p$. Of course, when $r_i = v_i$, $P(r_i|v_i) = 1 - p$

MLD with DMC and BSC

- Now, let $d(\mathbf{r}, \mathbf{v})$ be the distance between $\mathbf{r}$ and $\mathbf{v}$, that is, the number of positions in which $\mathbf{r}$ and $\mathbf{v}$ differ. Since they are binary sequences, this distance is called *Hamming distance*.
- Assume a block length of n, we then have:

$$\begin{aligned}\sum_i \log P(r_i|v_i) &= d(\mathbf{r}, \mathbf{v}) \log p + [n - d(\mathbf{r}, \mathbf{v})] \log(1-p) \\ &= d(\mathbf{r}, \mathbf{v}) \log \frac{p}{1-p} + n \log(1-p)\end{aligned}$$

- So what is MLD rule now?