

# Coding Theory

CSE 445/545

February 1, 2023

# Make sure to check out the syllabus!

CSE 4/545

Syllabus

Piazza

Schedule

Homeworks

Mini Project

Autolab

Book

## CSE 445/545 (Coding Theory) Syllabus

Spring 2023

Mondays, Wednesdays and Fridays, 4:00-4:50pm, Cooke  121.

### Please note

It is **your responsibility** to make sure you read and understand the contents of this syllabus. If you have any questions, please contact the instructor.

## Academic Integrity

### Penalty for academic integrity violation

In accordance with the current departmental policy on academic integrity violations, we will follow this procedure in CSE 4/545:

1. If the violation is the student's second academic violation, then it will result in an automatic **F** letter grade in the course.
2. If the violation is the first ever academic violation, then it will result in a **minimum of a letter grade reduction** in the grade for the course **and zero in the relevant assignment**. If the violation is serious enough, then it can result in an **F in the course**. While it gives me no pleasure in failing students, I will do so since I

# Let me know if you're not on piazza

**piazza** CSE 4/545 ▾ [Q & A](#) [Resources](#) [Statistics](#) ▾ [Manage Class](#)  Atri Rudra

LIVE Q&A Drafts project exam logistics other feedback hw0 proof-reading lectures hw1 hw2 hw3 hw4 hw5 hw6 hw7 hw8 hw9 hw10

Unread Updated Unresolved Following  

[New Post](#)  

Show Actions

PINNED

- Instr Poll on Ben's OH Times** 1/27/23  
In the poll below, please select all of the potential times for office hours that work for you, even if you can only att 
- Instr Background feedback** 1/22/23  
For me to get a better sense of your background, please fill in these piazza polls:  
Linear Algebra: @7Abstract Algebra: 
- Search for Teammates!** 1/11/23  
• 2 Open Teammate Searches 

YESTERDAY

- Instr Lecture 1 stuff** 10:12 PM  
I won't be doing this for all the followup lectures but wanted to give y'all a headsup that lecture 1 material i 
- Instr My OH is on** 11:36 AM  
Just to clarify that I do have my OH from 11:30am–12:20pm today even though we have not had our first lecture yet. Feel 

LAST WEEK

- Instr Welcome to Piazza!** Sun  
Students,Welcome to Piazza! We'll be conducting all class-related discussion here this term. The quicker you begin a 
- Instr Why this course?** Sun  
Why are you taking this course? 

**Class at a Glance** Updated 0 seconds ago. [Reload](#) [Go to Live Q&A](#)

 **no unread posts**  **no unanswered questions**  **no unanswered followups**

**license status** active instructor license  
16 total posts  
267 total contributions  
0 instructors' responses  
0 students' responses  
**1 sec** avg. response time

Student Enrollment  60 enrolled out of 75 (estimated) [Edit](#)

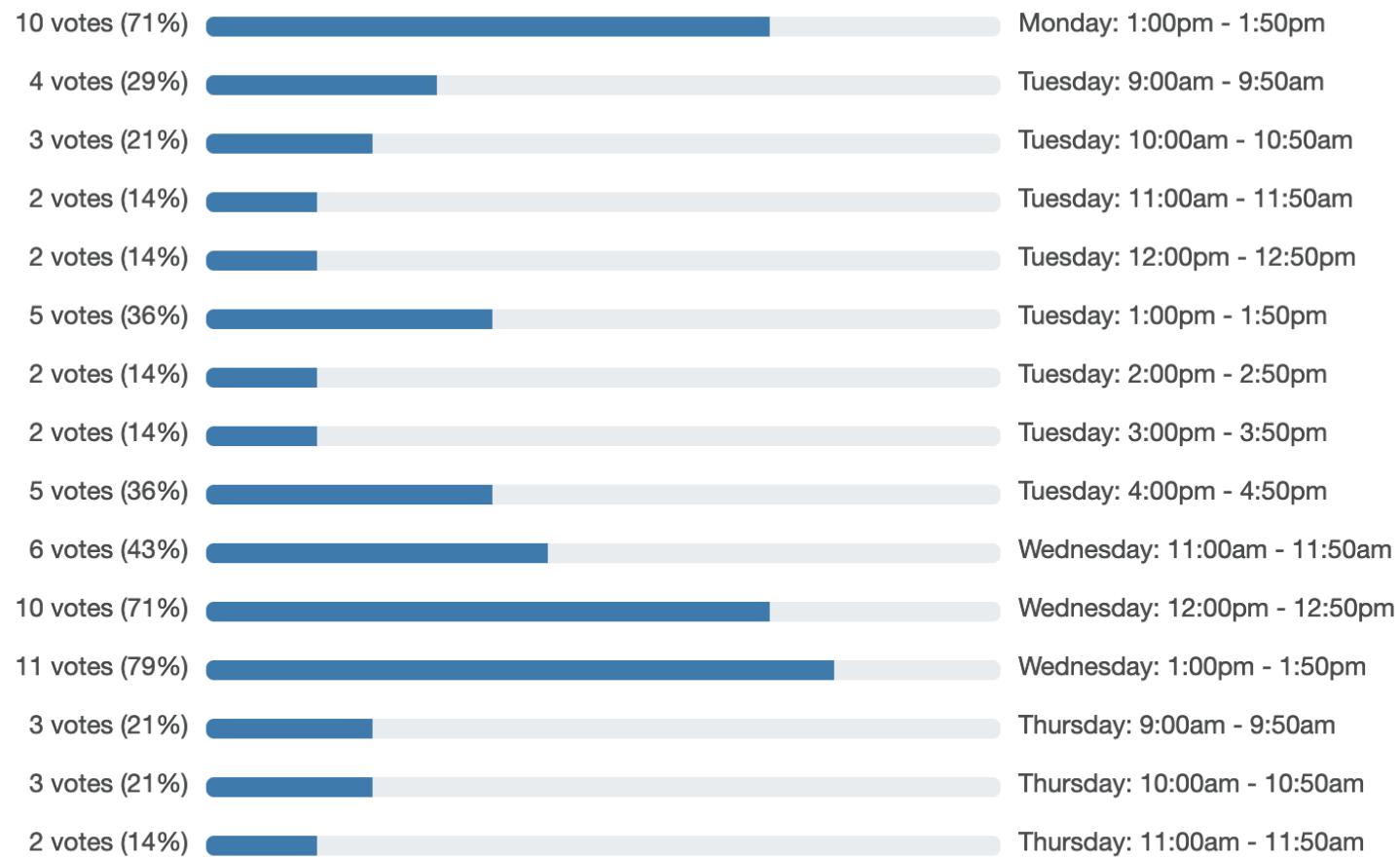
Download us in the app store:  

**Share Your Class**  
Professors appreciate Piazza best when they see how it is being used.  
Allow colleagues to view your class through a demo link - a restricted, read only version of your class where all students' names are anonymized and all student information hidden.  
[https://piazza.com/demo\\_login?nid=lcs7u45n33p2ar&auth=8404ff0](https://piazza.com/demo_login?nid=lcs7u45n33p2ar&auth=8404ff0)  
Opening this link in the same browser will log you out as atri@buffalo.edu

# By tonight fill in poll for Ben's office hours

## Poll on Ben's OH Times closes in 2 day(s)

A total of 14 voter(s) in 101 hours



Make sure you see 4/545 on Autolab

## Courses

### Current

---

CSE440/441/540:  
ML and  
Society (s23)

Homework

White Supremacy and Buffalo

[COURSE PAGE](#)

CSE445/545:  
Coding  
Theory (s23)

[COURSE PAGE](#) [GRADE](#)  
[SECTION](#)

# While applications are numerous...

This course (lectures and HWs) will focus ONLY on proofs

where in the above " $|E(\mathbf{m})|$ " is short for "being conditioned on  $E(\mathbf{m})$  being transmitted" and the inequality follows from the union bound (Proposition 3.1.5) and the fact that  $D$  is MLD.

Noting that  $\Delta(E(\mathbf{m}'), \mathbf{y}) \leq \Delta(E(\mathbf{m}), \mathbf{y}) \leq (p + \varepsilon')n$  (see Figure 6.6), by (6.9) we have

$$\begin{aligned} \mathbb{E}_E [\mathbb{1}_{D(\mathbf{y}) \neq \mathbf{m}}] &\leq \sum_{\mathbf{m}' \neq \mathbf{m}} \Pr [E(\mathbf{m}') \in B(\mathbf{y}, (p + \varepsilon')n) | E(\mathbf{m})] \\ &= \sum_{\mathbf{m}' \neq \mathbf{m}} \frac{|B(\mathbf{y}, (p + \varepsilon')n)|}{2^n} \end{aligned} \quad (6.10)$$

$$\begin{aligned} &\leq \sum_{\mathbf{m}' \neq \mathbf{m}} \frac{2^{H(p + \varepsilon')n}}{2^n} \\ &< 2^k \cdot 2^{-n(1 - H(p + \varepsilon'))} \end{aligned} \quad (6.11)$$

$$\begin{aligned} &\leq 2^{n(1 - H(p + \varepsilon)) - n(1 - H(p + \varepsilon'))} \\ &= 2^{-n(H(p + \varepsilon) - H(p + \varepsilon'))}. \end{aligned} \quad (6.12) \quad (6.13)$$

In the above, (6.10) follows from the fact that the choice for  $E(\mathbf{m}')$  is independent of  $E(\mathbf{m})$ . (6.11) follows from the upper bound on the volume of a Hamming ball (Proposition 3.3.3), while (6.12) follows from our choice of  $k$ .

Using (6.13) in (6.8), we get

$$\begin{aligned} \mathbb{E}_E \left[ \Pr_{\mathbf{e} \sim \text{BSC}_p} [D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \right] &\leq e^{-(\varepsilon')^2 n/2} + 2^{-n(H(p + \varepsilon) - H(p + \varepsilon'))} \sum_{\mathbf{y} \in B(E(\mathbf{m}), (p + \varepsilon')n)} \Pr[\mathbf{y} | E(\mathbf{m})] \\ &\leq e^{-(\varepsilon')^2 n/2} + 2^{-n(H(p + \varepsilon) - H(p + \varepsilon'))} \leq 2^{-\delta' n}, \end{aligned} \quad (6.14)$$

where the second inequality follows from the fact that

$$\sum_{\mathbf{y} \in B(E(\mathbf{m}), (p + \varepsilon')n)} \Pr[\mathbf{y} | E(\mathbf{m})] \leq \sum_{\mathbf{y} \in \{0, 1\}^n} \Pr[\mathbf{y} | E(\mathbf{m})] = 1$$

and the last inequality follows for large enough  $n$ , say  $\varepsilon' = \varepsilon/2$  and by picking  $\delta' > 0$  to be small enough. (See Exercise 6.3.)

Thus, we have shown that for any arbitrary  $\mathbf{m}$  the average (over the choices of  $E$ ) decoding error probability is small. However, we still need to show that the decoding error probability is exponentially small for *all* messages *simultaneously*. Towards this end, as the bound holds for each  $\mathbf{m}$ , we have

$$\mathbb{E}_{\mathbf{m}} \left[ \mathbb{E}_E \left[ \Pr_{\mathbf{e} \sim \text{BSC}_p} [D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \right] \right] \leq 2^{-\delta' n}.$$

# Questions/Comments?



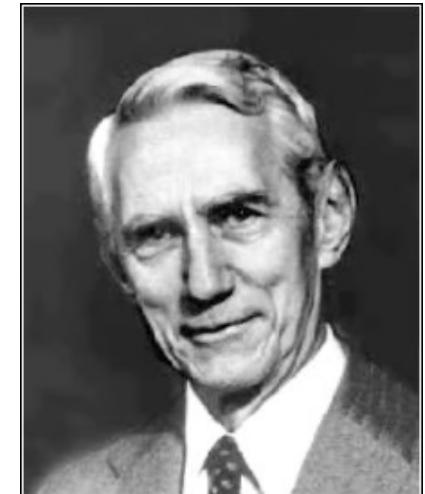
# The birth of coding theory

Claude E. Shannon

“A Mathematical Theory of Communication”

1948

Gave birth to Information theory



Richard W. Hamming

“Error Detecting and Error Correcting Codes”

1950



# Structure of the course

## Part I: Combinatorics

What can and cannot be done with codes

## Part II: Algorithms

How to use codes efficiently

## Part III: Applications

Applications in (theoretical) Computer Science

# Redundancy vs. Error-correction

**Repetition code:** Repeat every bit say 100 times

Good error correcting properties

Too much redundancy

**Parity code:** Add a parity bit

Minimum amount of redundancy

Bad error correcting properties

Two errors go completely undetected

1	1	1	0	0	1
---	---	---	---	---	---

Neither of these codes are satisfactory

1	0	0	0	0	1
---	---	---	---	---	---

# Two main challenges in coding theory

## Problem with parity example

Messages mapped to codewords which do not differ in many places

Need to pick a lot of codewords that differ a lot from each other

## Efficient decoding

Naive algorithm: check received word with all codewords

# The fundamental tradeoff

Correct as **many errors** as possible with as **little redundancy** as possible

Can one achieve the “optimal” tradeoff with  
*efficient* encoding and decoding ?

# Rest (of the semester) on the board...

