

The Virtues of Redundancy

The Virtues of Redundancy

An Introduction to Error-Correcting Codes

Paul H. Siegel

Director, CMRR

University of California, San Diego

Outline

- **Information Storage and Transmission**
- **Repetition Codes**
- **Hamming Code (with a **live** demo!)**
- **Shannon's Coding Theorem**
- **Compact Audio Disk Code (another **live** demo!)**
- **Summary**

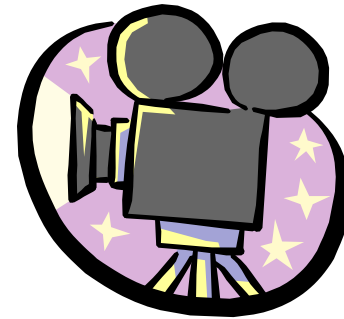
Information

- We all generate and use information:

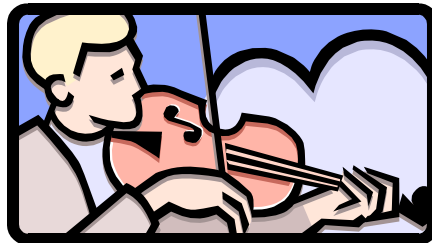
Text



Images



Music



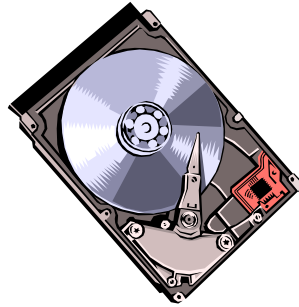
**Scientific
Measurements**



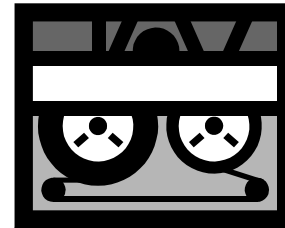
Information Storage

- We often **store** the information for later use:

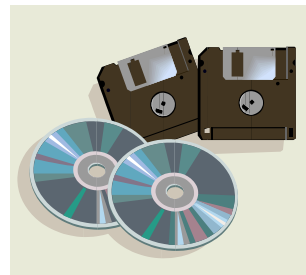
- Disk drive



- Tape drive



- Zip/diskette drive



- CD-R/W

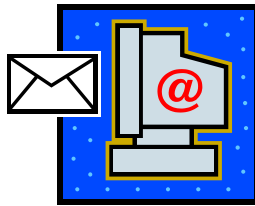
Information Storage

- **And we want our storage devices to be:**
 - **Reliable** (no errors!)
 - **Compact** (lots of storage capacity in a small space!)
 - **Inexpensive** (almost free!)

Information Transmission

- We also **send** our information to others:

- Email



- Fax



- Wireless telephony



- Satellite broadcast



Information Transmission

- **And we want the transmission system to be:**
 - **Reliable (no errors!)**
 - **Fast (instant downloads!)**
 - **Inexpensive (almost free!)**

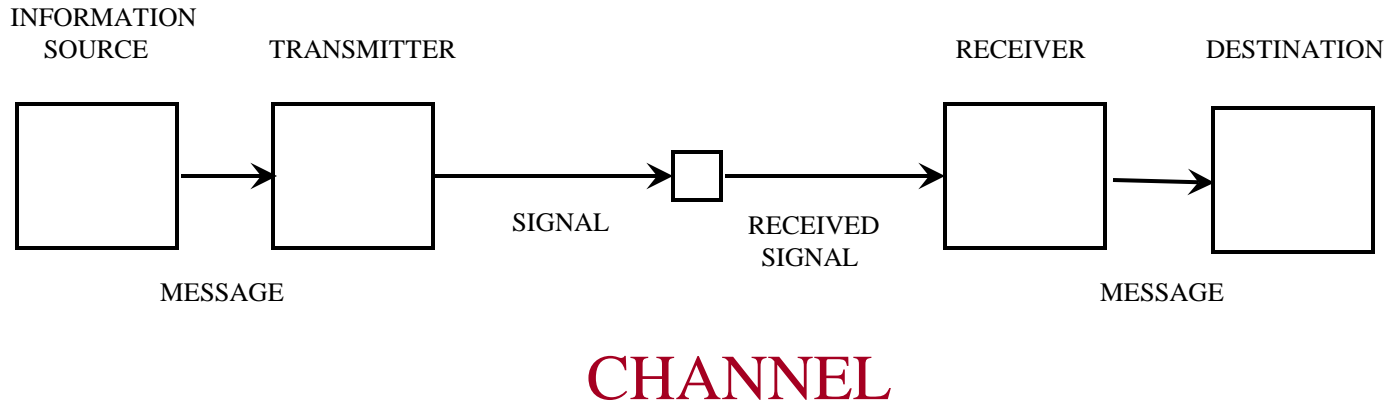
Data Storage and Transmission

- **Data storage and data transmission are two sides of the same coin – communication systems.**

A **data storage** system communicates information through **time**, i.e., “from now to then.”

A **data transmission** system communicates information through **space**, i.e., “from here to there.”

Communication System



Digitized Information

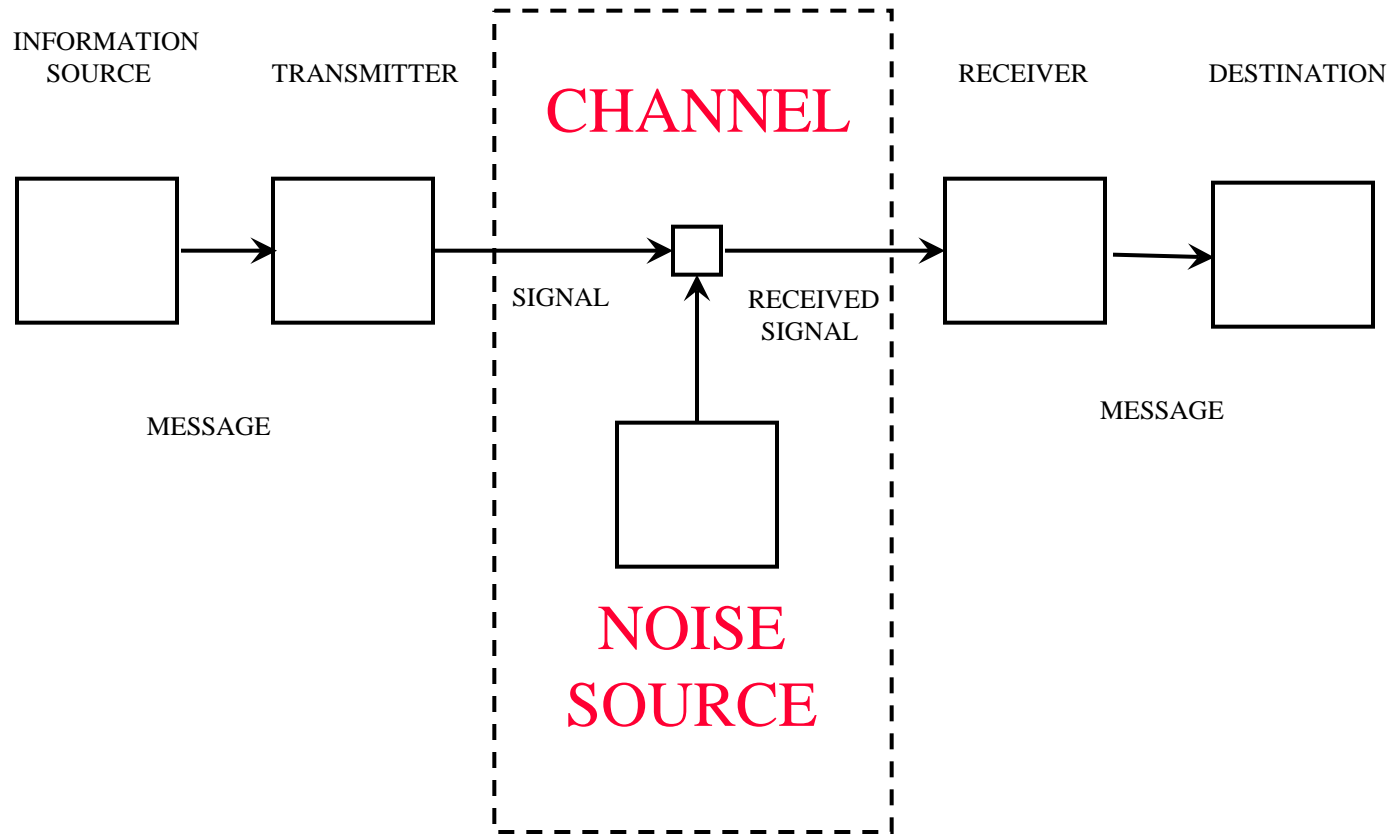
- We often represent the information **digitally**, as a sequence of numbers.
- Each number is converted to a unique string of bits (**0**'s and **1**'s).
- The **0**'s and **1**'s are converted to an electrical signal that is used to store or transmit the information.

“Noisy” Channels

- When we retrieve or receive the signal, it gets converted back to a sequence of 0’s and 1’s.
- But an evil force is at work in the channel...**NOISE!**
- Electrical and magnetic devices can distort the electrical signal, and the recovered sequence of bits may have errors:

1111111111 → 1111101111
sent received

A Noisy Communication System

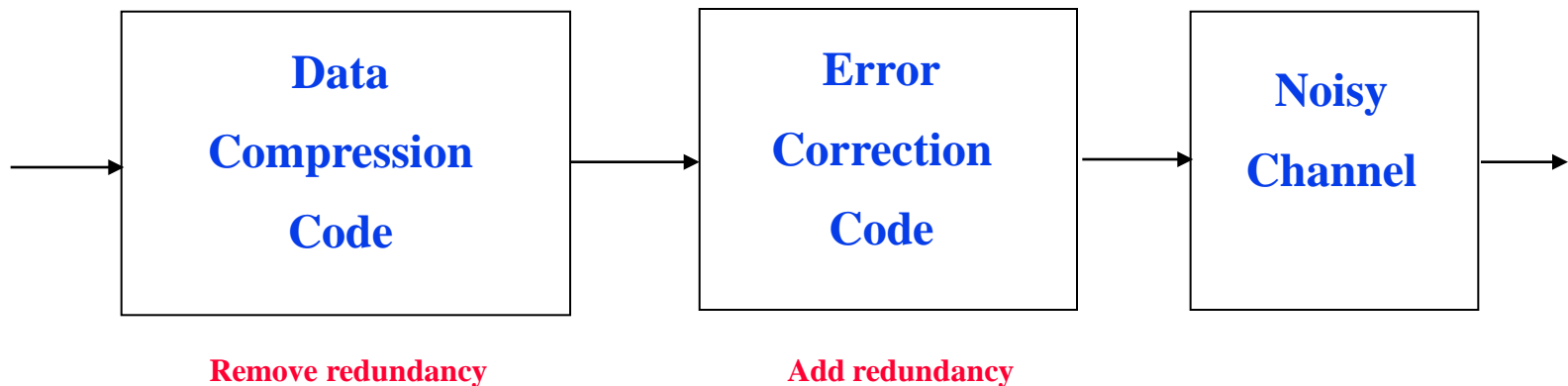


Coding to the Rescue

- To combat the effects of noise, we use an

Error Correction Code (ECC).

- The ECC **adds redundancy** in a special way that allows us to **correct** the bit errors caused by the noisy channel.



Repetition

The simplest form of redundancy is: **Repetition!**

Sender

“0”

I said “0”

One more time: “0”

Receiver

Did she say “1” ?

Sounded like “0”

Sounded like “0” again

She was sending “0”

3-Repetition Code

- **Encoding rule: Repeat each bit 3 times**

Example:

1 . 0 . 1 . \rightarrow 111 . 000 . 111 .

- **Decoding rule: Majority vote!**

Examples of received codewords:

110 . 000 . 111 . \rightarrow 1 . 0 . 1 . **Error-free!**

111 . 000 . 010 . \rightarrow 1 . 0 . 0 . **Error!**

How good is this 3-repetition code?

- The code can correct 1 bit error per 3-bit codeword.
- The price we pay in redundancy is measured by the **efficiency or rate** of the code, denoted by **R**:

$$\mathbf{R} = \text{\#information bits} / \text{\# bits in codeword}$$

- For the 3-repetition code: **R=33%**

How good is this 3-repetition code?

Suppose that, on average, the noisy channel flips

1 code bit in 100

Then, on average, the 3-repetition code makes

only 1 information bit error in 3333 bits!

Can we do better?

How about repeat 5 times?

On average, only 1 bit error in 100,000 bits.

How about repeat 7 times?

On average, only 1 bit error in 2,857,142 bits

If we let the number of repetitions grow and grow, we can approach **perfect reliability** !

What's the catch????

The catch is:

As the number of repetitions grows to infinity, the transmission rate shrinks to zero!!!

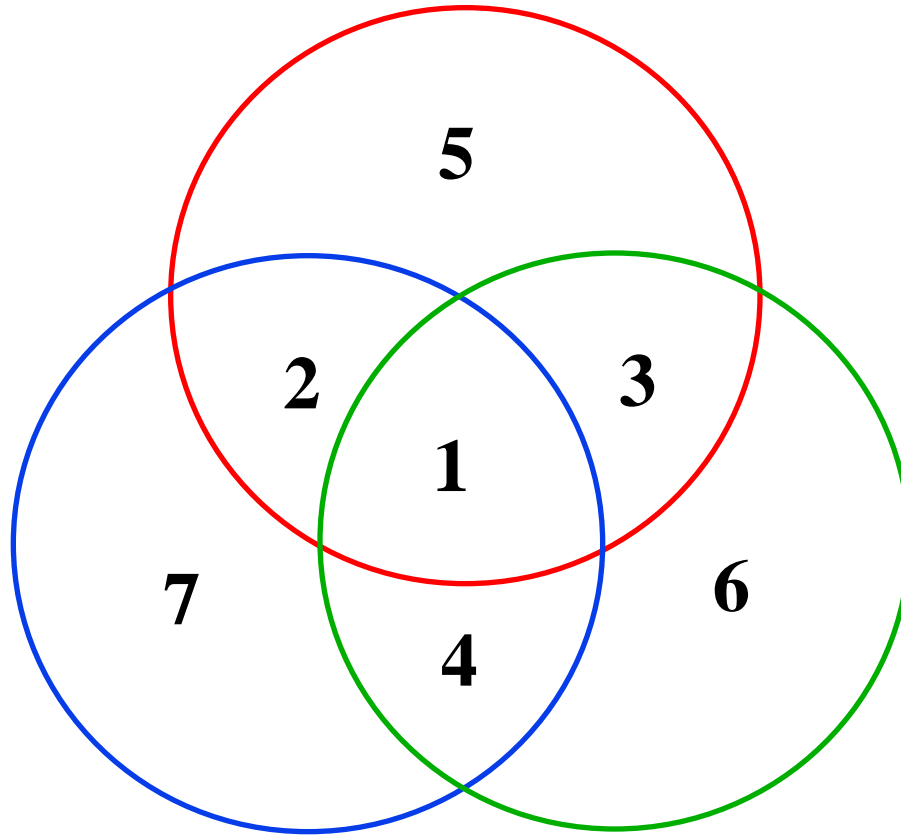
This means: slow data transmission / low storage density.

Is there a better (more efficient) error correcting code?

Hamming Code

- **Invented by Richard W. Hamming in 1948.**
- **7-bit codewords: 4 information bits, 3 redundant bits.**
- **Corrects 1 bit error per codeword**
(like 3-repetition code)
- **Code efficiency: $R = 4/7 \approx 57\%$**
(compared to rate $R \approx 33\%$ for 3-repetition code)

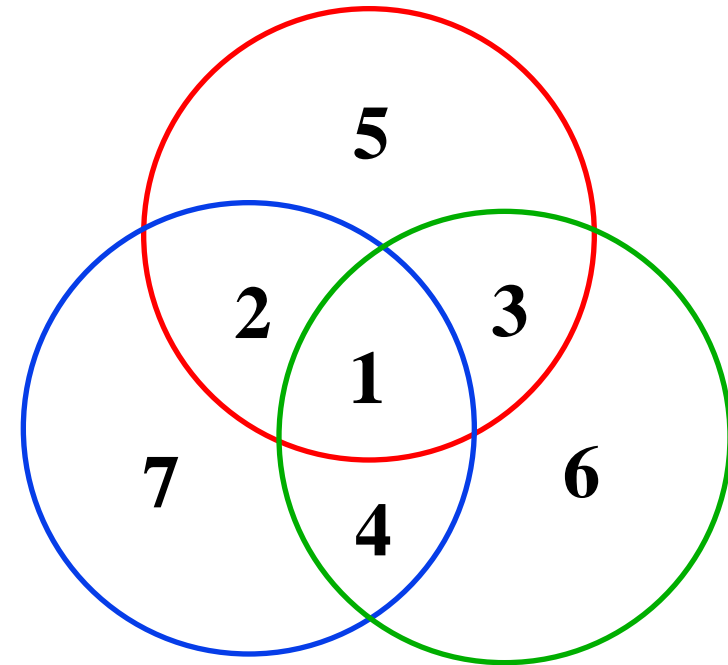
Hamming Code



Hamming Code: Encoder

- **Simple encoding rule:**

1. Insert data bits in 1,2,3,4.
2. Insert “parity” bits in 5,6,7 to ensure an even number of 1’s in each circle.



Hamming Code

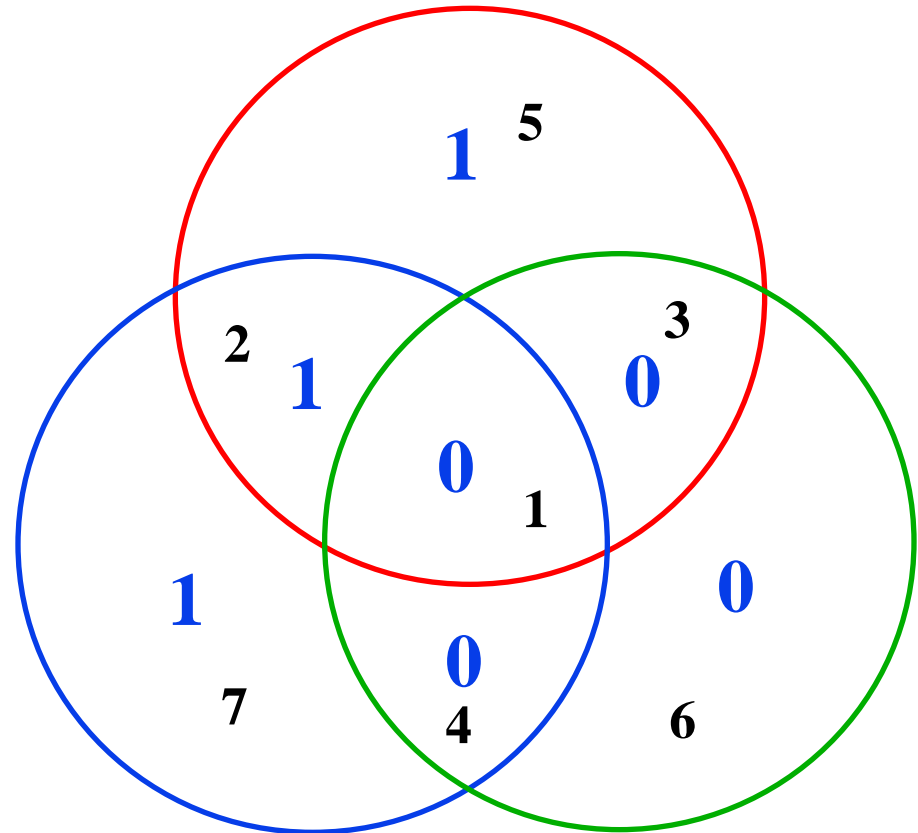
Information: **0 1 0 0**

Parity bit 5: **1**

Parity bit 6: **0**

Parity bit 7: **1**

Codeword: **0 1 0 0 1 0 1**



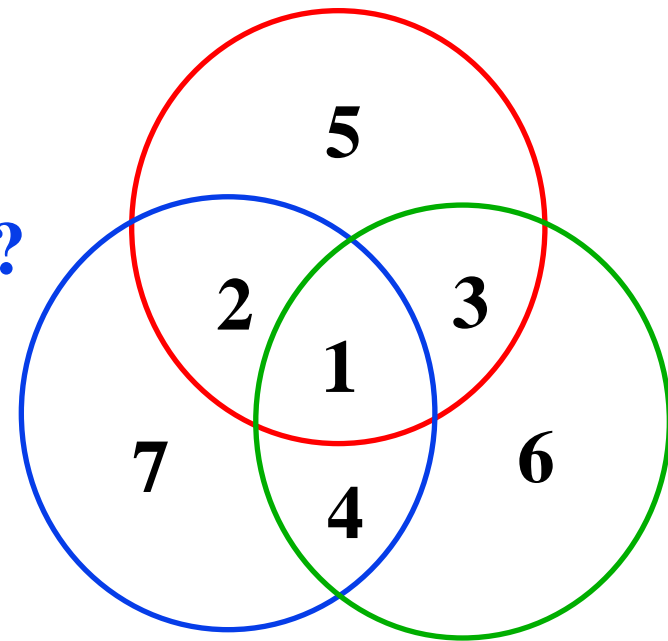
Hamming Code: Decoder

- **Simple decoding rule:**

1. **Check the parity of each circle – is the number of 1's even or odd ?**

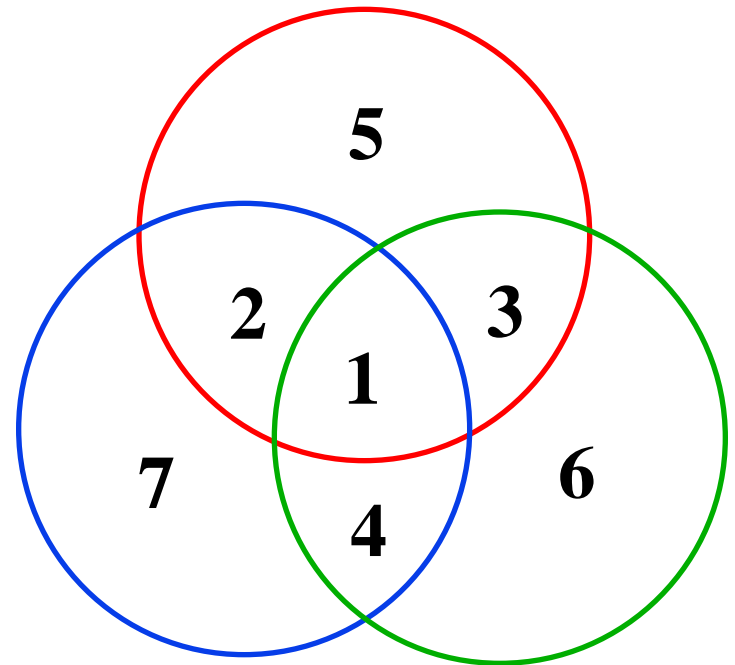
The pattern of parities is called the syndrome .

2. **Look up the syndrome in the decoding table.**



Hamming Code: Decoding Table

Syndrome			Flipped bit
even	even	even	None!
odd	odd	odd	1
odd	even	odd	2
odd	odd	even	3
even	odd	odd	4
odd	even	even	5
even	odd	even	6
even	even	odd	7



How good is the Hamming code?

Suppose that, on average, the noisy channel flips

1 bit in 100

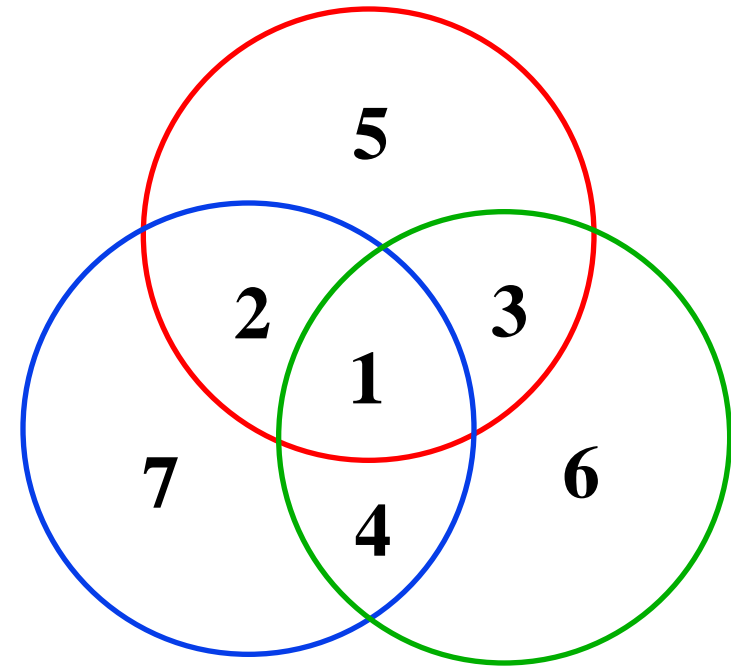
Then, on average, the Hamming code makes

On average, only 1 error in 1111 !!

(Not as good as 3-repetition, but efficiency is 57% compared to 33%)

Matrix description of Hamming Code

Bit #	1	2	3	4	5	6	7
Red circle	1	1	1	0	1	0	0
Green circle	1	0	1	1	0	1	0
Blue circle	1	1	0	1	0	0	1



**This is the “parity-check matrix” for the code.
Encoding and decoding can be described algebraically .**

Practical ECC

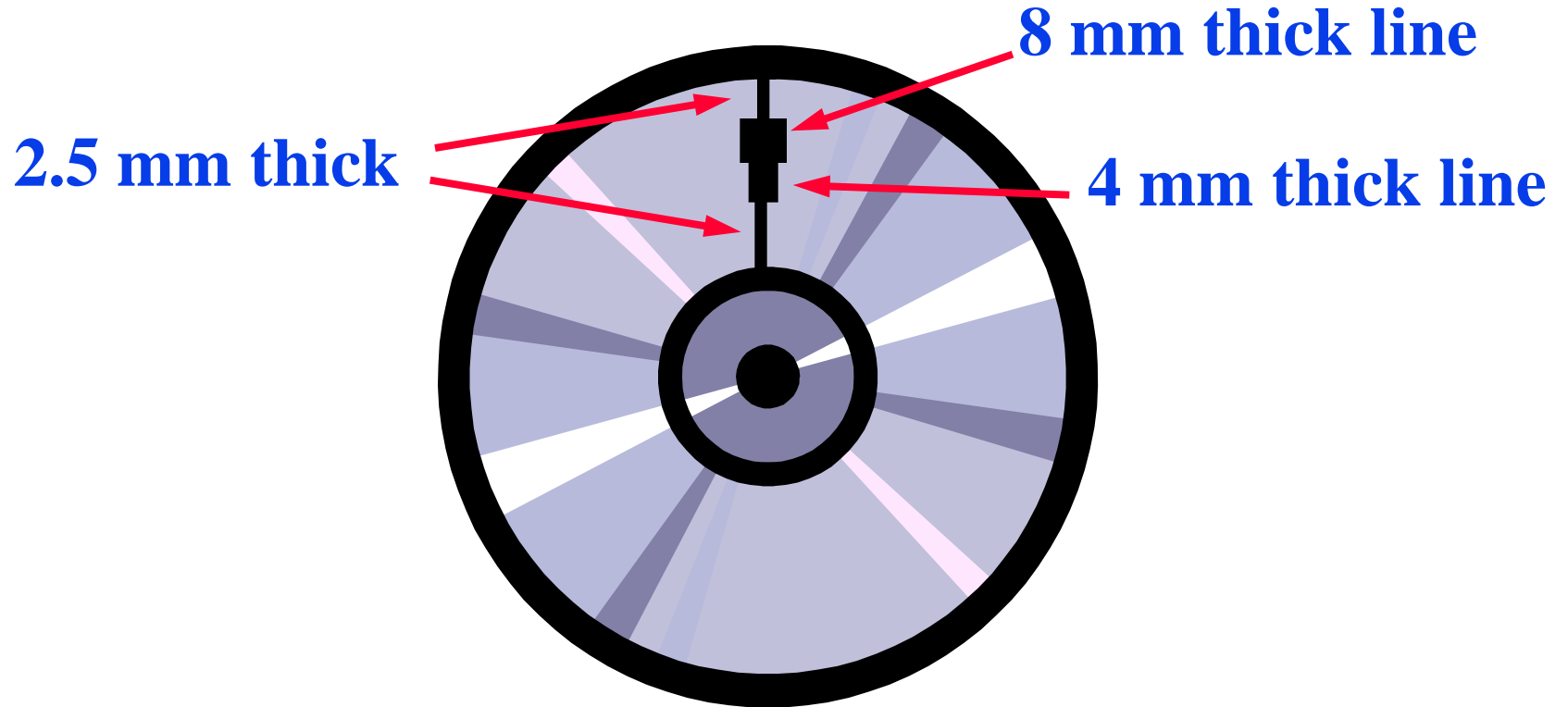
- **Lots of “algebraic” codes can be designed by choosing different parity-check matrices.**
- **Many useful and powerful codes have been found and are used in many data transmission and data storage applications.**

Compact Disc ECC

- The compact disk uses a powerful code with **75%** efficiency
- It can correct a burst of about **4000 bits**, which corresponds to a scratch of length about **1-tenth** of an inch (2.5 mm).
- Using the redundancy in musical signals, the CD player can “fix” a burst of **12,300 bits** - a scratch of length about **3-tenths** of an inch (7.5 mm)!

CD Demonstration

(Warning: do not attempt this at home!)



How “good” can a code be?

- What is the best tradeoff between efficiency (transmission speed / storage density) and error-correcting capability (reliability)?
- Does the efficiency have to go to zero in order for the average number of decoder errors to approach zero (as in the repetition codes)?

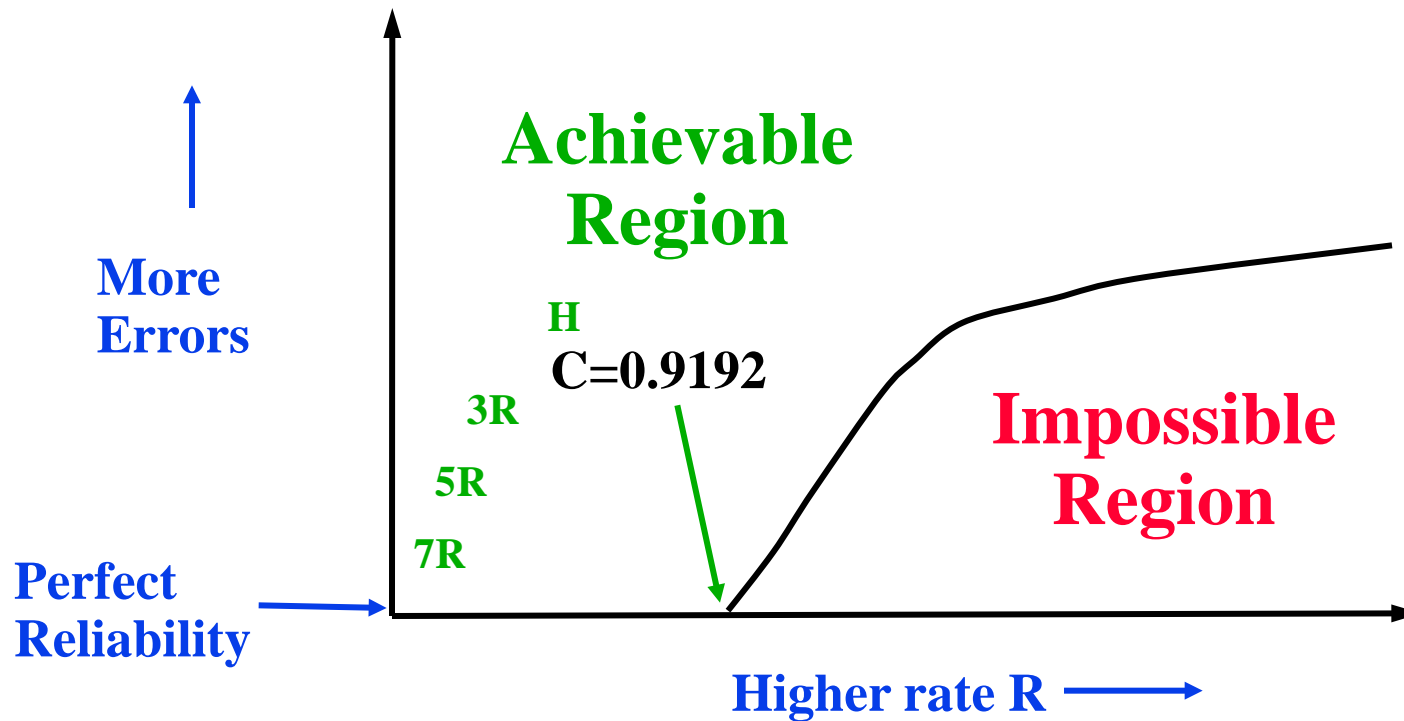
Amazingly, the answer is NO !

Shannon's Coding Theorem

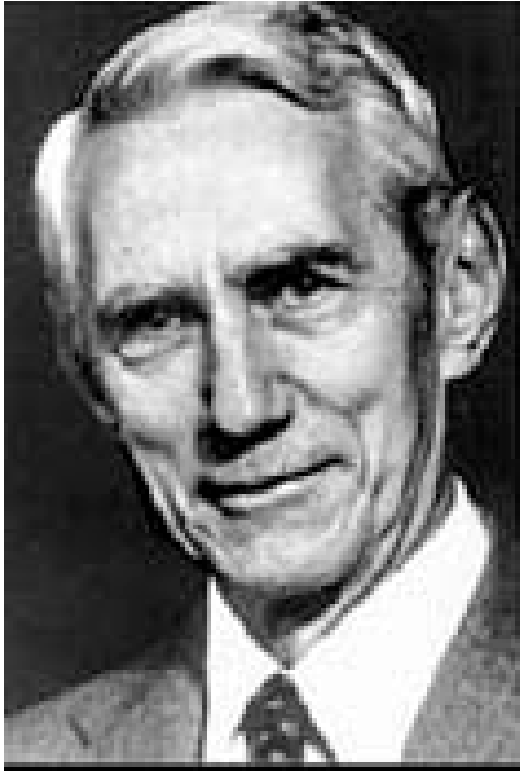
- A mathematical result proved by Claude E. Shannon (Bell Labs) - in 1948 (coincidentally!)
- For any noisy channel, there is a maximum achievable efficiency (larger than 0), called the **channel capacity**, and denoted by **C**.
- For any rate **R smaller than C**, there exist codes that can approach **perfect reliability!**
- For any rate **R greater than C**, perfect reliability is **IMPOSSIBLE !**

Shannon's Coding Theorem

For the channel that flips 1 bit in 100 bits,
the capacity is $C = 91.92\%$!!!



Claude E. Shannon



Claude Elwood Shannon
1916 - 2001



(CMRR Lobby)

The Inscription

CLAUDE ELWOOD SHANNON

1916 – 2001

FATHER OF INFORMATION THEORY

**HIS FORMULATION OF THE MATHEMATICAL
THEORY OF COMMUNICATION PROVIDED
THE FOUNDATION FOR THE DEVELOPMENT OF
DATA STORAGE AND TRANSMISSION SYSTEMS
THAT LAUNCHED THE INFORMATION AGE.**

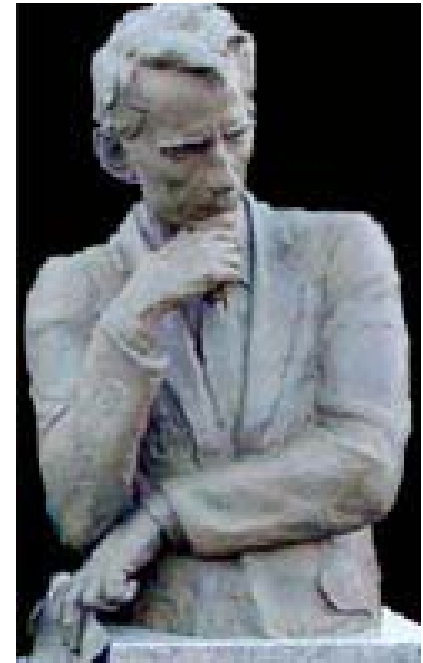
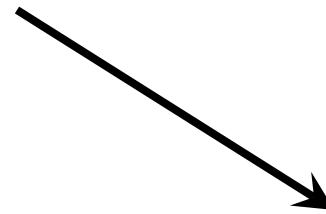
DEDICATED OCTOBER 16, 2001

EUGENE DAUB, SCULPTOR

The Formula on the “Paper”

**The paper shows the formula
for the capacity of a discrete
noisy channel [Shannon, 1948]**

$$C = \text{Max} (H(x) - H_y(x))$$



Capacity-approaching codes

- Shannon proved that lots of such “good” codes exist... **but** he did **not** show how to construct the encoders and decoders!
- Finding such practical good codes has been the “holy grail” of coding theorists since 1948.
- Turbo codes (1993) and Low-Density Parity-Check codes (2001) finally have come close to achieving Shannon’s theoretical limits !!! Hooray!!
- Are coding theorists out of work? **No way....**

Summary

- **Coding theory is fun and useful!**
- **You've been a nice audience – Thanks!**