Introduction to Information Theory (§1)

← MacKay

① How to **measure** information? How to ask the **most informative** questions?

"bit" ... but: 

→ "entropy"

"guess a number" game → data science, ML

② How to **compress** a data source? lossless: FLAC, ZIP, GIF, ... lossy: JPG, MP3, MP4, ...

③ How to **reliably** send information over unreliable **channels**? LTE, Blu-ray, QR-codes, ...

1948: Shannon, "A Mathematical Theory of Information" solved ①–③ "in theory"

origins: telecommunication + physics

Morse (1830s): ●E ●●● S (1830s)

1920s Bell labs

thermodynamics (1870+) Boltzmann, Gibbs, ...

info ~ log(# voltage levels) ~ log(# possible signals)

Nyquist — Hartley — abstraction!

today: engineering + theory (efficient codes, beyond i.i.d.) + (quantum)

☆ LOGISTICS

Compression

Suppose we want to compress a message in $\{A,B,C,D\} = \mathcal{A}$:

source message $x \in \mathcal{A}$ — [Compressor] → code word $c \in \{0,1\}^\ell$ — [decompr.] → decompressed message $\hat{x} \in \mathcal{A}$

WANT: $x = \hat{x}$

4 possible messages $(2^2 = 4)$ → need $\ell = 2$

x	c
A	00
B	01
C	10
D	11

why not 0, 1, 00, 01 → prefix

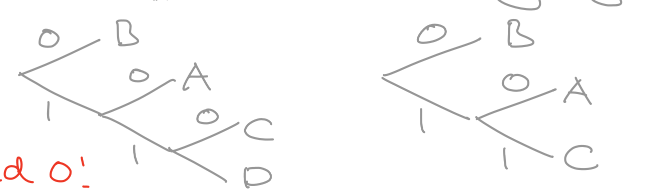
In general: $2^\ell \geq \#\mathcal{A} \Rightarrow \ell \geq \log_2(\#\mathcal{A})$

Can we do better? Imagine some messages are more frequent than others...

			code I	code II
A	sunshine	44%	10	10
B	rain	55%	0	0
C	snow	0.99%	110	11
D	hurricane	0.01%	111	0

longer (code I) — reused 0! (code II)

both can be decoded nicely! e.g.



Code I: lossless, average length = 1.46 $\ll 2$?!
Code II: lossy! $P_{error} = 0.01\%$, average length ≈ 1.45

How to do even better? Look at blocks of messages!

↳ **SHANNON:** Optimal rate of compression is $\approx 1.06 \frac{\text{bits}}{\text{message}}$ ← entropy of source (but...)

BREAK

Communicating over Noisy Channels

Examples of noisy channels & how to avoid:

* Scratch on Bluray disk	Don't do it!
* Loud party	Tell people not to shout!
* Mail arrives crumpled	Pay your postman more!
* Bad signal	Build more cell phone towers!
* Bit flip on hard disk	Shield better

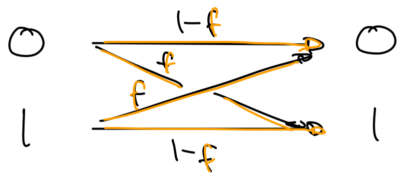
} ∈ or infeasible

↖ SATA mandates $P_{read\ error} \ll 10^{-14}$ ⇝ Reed-Solomon, LDPC codes

Mathematical model:

input → Channel → output $\qquad p(\text{output} \mid \text{input})$

e.g. binary symmetric channel:

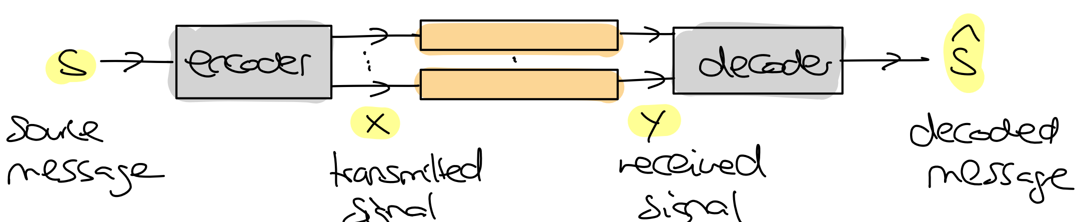


$p(1|0) = p(0|1) = f$
$p(0|0) = p(1|1) = 1-f$

f = probability of bit flip

assume we know f !!!

How to reduce error? Introduce redundancy by encoding message!



WANT: $S = \hat{S}$ with high probability!

Repetition Code R_3:

* encoder:

S	$X = X_1X_2X_3$
0	000
1	111

* decoder: majority vote

$Y = Y_1Y_2Y_3$	$\hat{S}$
000	0
001 / 010 / 100	0
011 / 101 / 110	1
111	1

* analysis: Can deal with ≤ 1 bit flip

$\Rightarrow$ $P_{error} = Pr(2 \text{ or } 3 \text{ bit flips}) = 3 \cdot f^2(1-f) + f^3 \approx 3f^2$ if f small

← $< f$ as long as $f < \frac{1}{2}$ [Ex.]

e.g. $f = 10\% = 0.1$: $P_{error} = 0.028 \approx 0.03 = 3\%$ 😊

* $\text{rate} = \frac{\# \text{ source msg bits}}{\# \text{ channel uses}} = \frac{1}{3}$

[Ex.] Is this decoder optimal? Yes, if $f \leq 50\%$. What if $f = 50\%$? No information! ☆

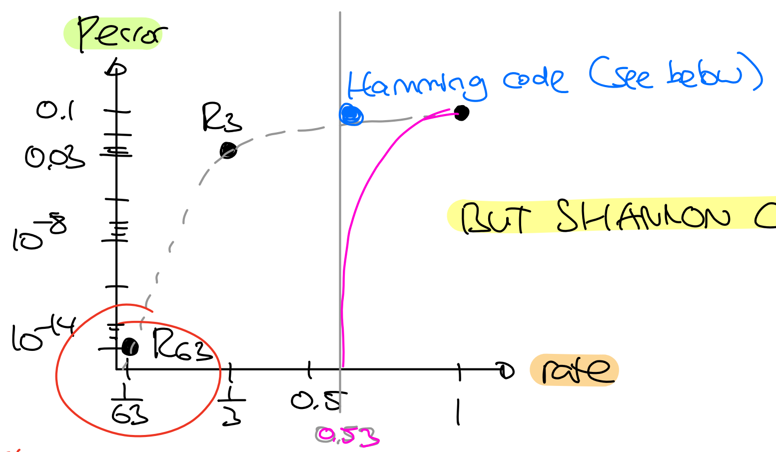
What if we repeat $N > 3$ times?

$$P_{error} = Pr(\geq \tfrac{N}{2} \text{ bit flips}) = \sum_{k \geq \frac{N}{2}} \binom{N}{k} f^k (1-f)^{N-k} \sim 2^N f^{N/2} (1-f)^{N/2}$$

(= : Thursday; ~ : Later)

at $\text{rate} = \frac{1}{N}$

e.g. $f = 10\%$: $P_{error} \sim 0.6^N$



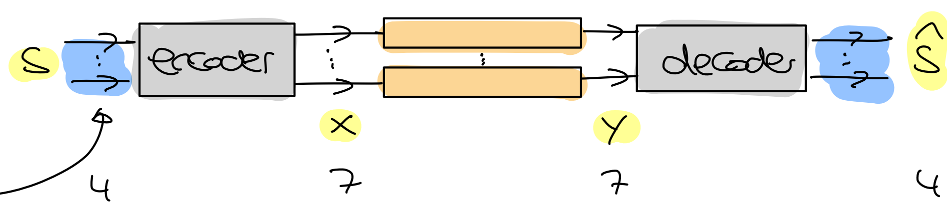
BUT SHANNON CAN DO BETTER! (see below)

it seems like rate → 0 if perror → 0

How can we find more & better codes?

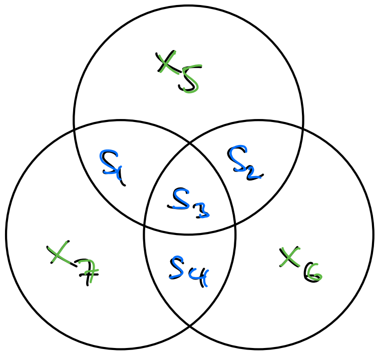
Block codes:

Encode more than one symbol at a time



We did not cover the following in class: but the TAs discussed it in the tutorials

(7,4)-Hamming code:



$x_1 = s_1 \ \dots \ x_4 = s_4$

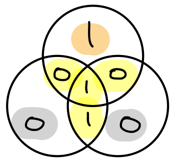
$x_5, \dots, x_7$ chosen such that sum in each circle even

("parity bits")

$x_1 \cdots x_4$ ‖

$S = S_1 \cdots S_4$	$x_5 x_6 x_7$
0000	000
0001	011
0010	111
0011	100
...	

e.g.



Any two codewords differ by 3 or more bits!

↳ can correct single bit flips

How to decode?

① Compute parities in all three circles: $z_1 = y_1 \oplus y_2 \oplus y_3 \oplus y_5 \pmod 2$

$\vdots$

z_3

② If at least one $z_i \neq 0$:

Flip unique bit that is only in circles with $z_i \neq 0$

$z = z_1 z_2 z_3$	000	001	010	100	011	101	110	111
flipped bit	—	y_7	y_6	y_5	y_4	y_1	y_2	y_3

$\Rightarrow$ $P_{\text{block error}} \leq \Pr(\geq 2 \text{ bit flips}) \sim \binom{7}{2} f^2 (1-f)^5 = 21 f^2$

$P_{\text{bit error}} = \frac{1}{4} \sum_{k=1}^{4} \Pr(\hat{S}_k \neq S_k) \sim 9 f^2$

exercise class

$\text{rate} = \frac{4}{7}$

SHANNON: For $f = 10\%$, can reliably send at optimal rate ≈ 0.53 !?

$\frac{\text{source bits}}{\text{channel uses}}$

(but...)

Wednesday: Probability theory recap + entropy (towards compression)