Lecture 5: 1/26

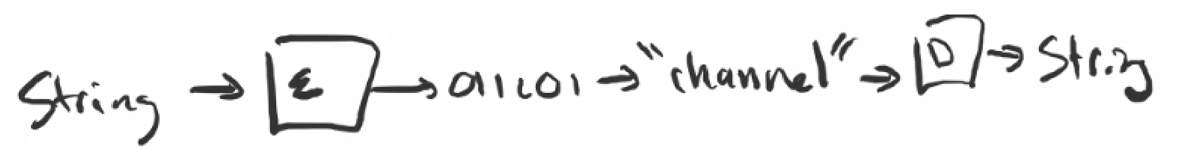
Paper

- Deep dive into area of interest in the course topics
  - Implement a state-of-the-art algorithm relevant to a problem that you're interested in
  - Coming up with a new analysis
  - Feb 15—proposal 1: Form groups of at most 3, set up zoom
  - Marh 5—progress report
  - End of qtr—6 pg LaTeX report

Class notes:

- First class started with Shannon "channel" model

  Figure 5.1

  - Represents communication across a wire, memory, a database, or interprocess communication
  - Result: if channel is noiseless, the maximum rate at which you can send data is governed by $H(x)$.
- What if the channel is noisy?
  - Why would the channel be noisy?
    - Real-world storage are nonideal—they have bit flips
    - Digital communication
    - Adversarial noise

- Today's class: **What can we do in a noisy channel?**

- To understand noisy channels, we split the problem into two parts

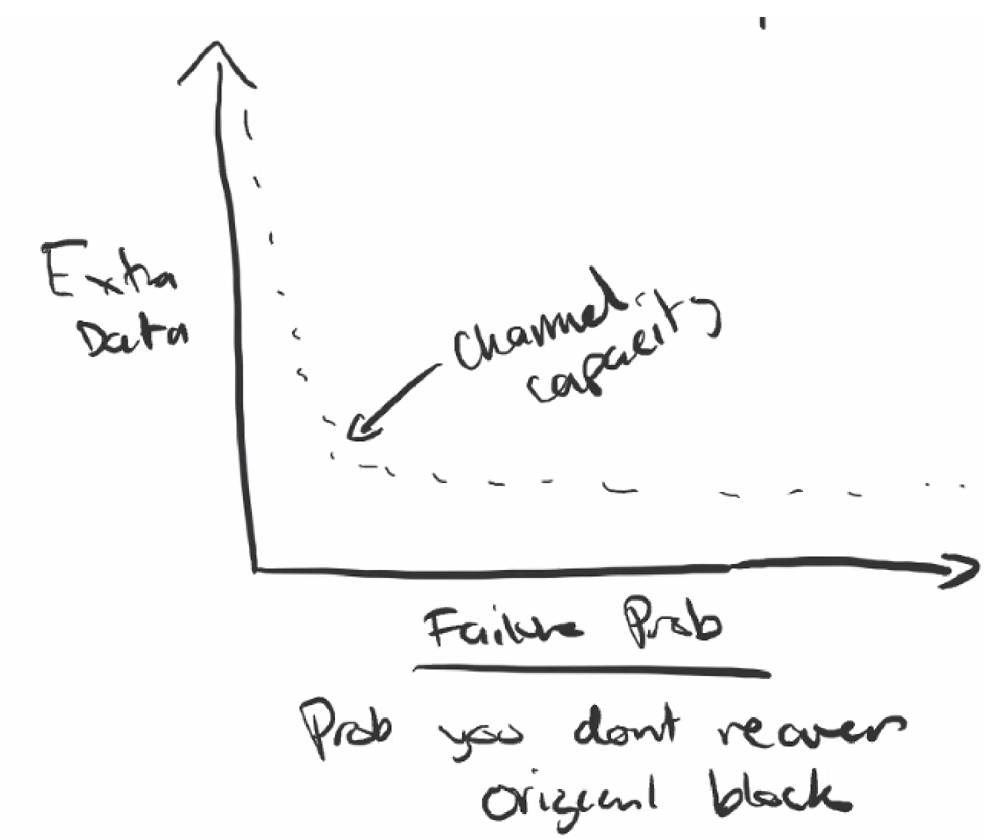
  Figure 5.2

  - On one side, sequences of zeros and ones coming into a channel
  - On the other side, sequences of zeros and ones coming out of a channel
- Week 1: source encoding → Week 3: **channel** encoding
  - Assume that the noise in the channel is of the form: for each bit, randomly flip it.
  - Assume a sequence: 100110100101
    1. Break the sequence into "blocks" of size B
    2. Add redundancy to each block
    3. Send over channel
  - Version 1: replicate each block by 2x
    - Since we can't discriminate between the two halves, this is useless
  - Version 2: replicate each block by 3x (or some odd number) and have them vote
    - How many failures (bit flips) can this scheme tolerate?
      - 1: if one bit flips in two blocks, then they will overrule the correct version of the block
      - For a more general n replications, the answer is $\lfloor \frac{n}{2} \rfloor$

- This is a lot of work! Paying 3x data to tolerate 1 bit failure
- Version 3: given a block B, add a **parity bit** (1 if the number of ones is odd, 0 if the number of ones is even)
  - How many failures (bit flips) can this code tolerate?
    - None: we can detect one error, but we don't know where it occurred
  - Replication factor: $\frac{B+1}{B}$
  - Can we have something similar that can actually tolerate a failure?

Channel capacity: how much you have to pay for reliable transmission

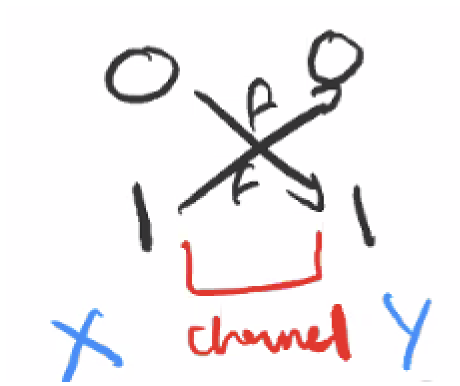
Figure 5.3



- What is the best we can do?

1. Binary symmetric channel: every bit gets independently flipped with probability "f"

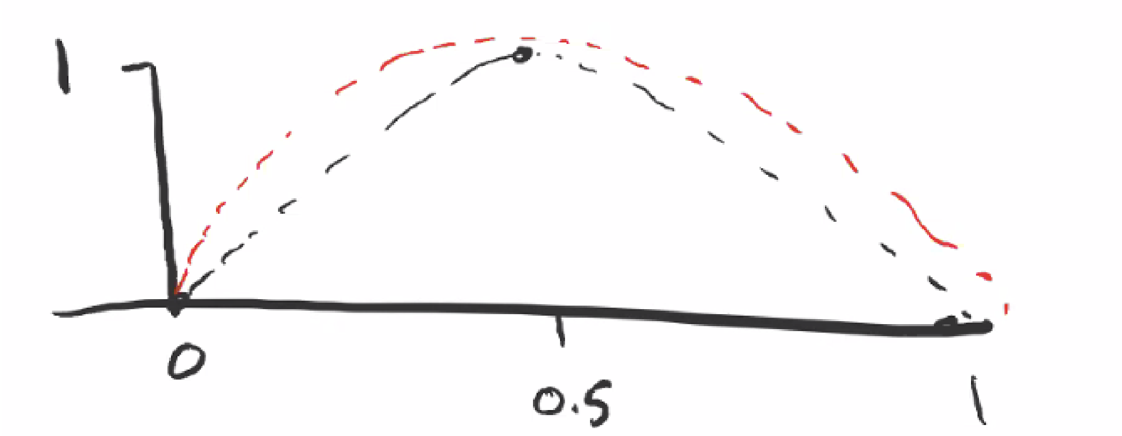
   Figure 5.4



2. Have a side channel so that X knows that its data was not properly transmitted, and so it can try again

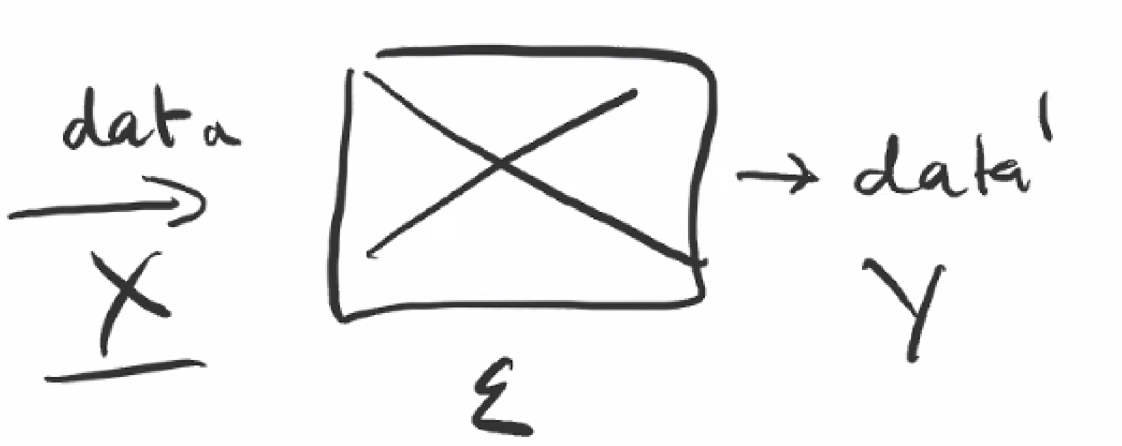
   - X and Y are both Bernoulli random variables
   - What is the expected number of retries per bit?
   - $E(retries) = 1 + f * E(retries)$
     - $E(retries) = \frac{1}{1-f}$
     - However, this isn't quite correct, since once we get above $f > 0.5$, we can just flip the bits to get a better # of retries (assume the data is wrong)
     - $E(retries) = \frac{1}{max(f,1-f)}$
   - $g(f) = \frac{1}{max(f,1-f)} - 1$ (extra data)

     Figure 5.5



- The red ($H(f)$ = entropy of the channel) is the upper bound of the graph

Figure 5.6



- To tolerate errors, you have to "transmit" E (transmit E worth of information)
  - $H(y|x, E) = 0$ (if we knew x, E, we would be able to determine y)
  - $H(x|y, E) = 0$
  - $H(E|x, y) = 0$
    - The above three imply that $H(x, y) = H(x, E)$
  - Noiseless case: we have to transmit $H(x)$ worth of data
  - Noisy case: we have to transmit $H(x, E)$ worth of data

- **Shannon's Noisy Channel Theorem**
  1. $I(x, y)$: mutual information

     $I(x, y) = H(x) + H(y) - H(x, y)$, or the difference of information assuming independence and the actual information

     - Measure of dependence
  2. $C =$ max over source dists $I(x, y)$
     - Assume the source x is as random as possible, i.e. 1 with prob .5 and 0 with prob .5

$$\begin{aligned} I(x, y) = C &= H(x) + H(y) - H(x, y) \\ &= H(x) + H(y) - H(x, E) \\ &= H(x) + H(y) - (H(x) + H(E)) \\ &= H(y) - H(E) \end{aligned}$$

     Line 2: amount of information minus amount of noise info you have to send)

     Line 4: entropy of data at the destination minus entropy of noise
  3. C is the **channel capacity**
     - $R(P_b) \leq \frac{C}{1-H(P_b)}$ where P_b is the desired error probability. (The right hand side is the red line in figure 5.5)
- Intuition: When we have a noisy channel, we tolerate errors by encoding the noise. By replicating something, it's like getting repeated observations of the pattern of bit flips.

On Thursday, we're going to talk about better channel codes

- Reed-Solomon codes