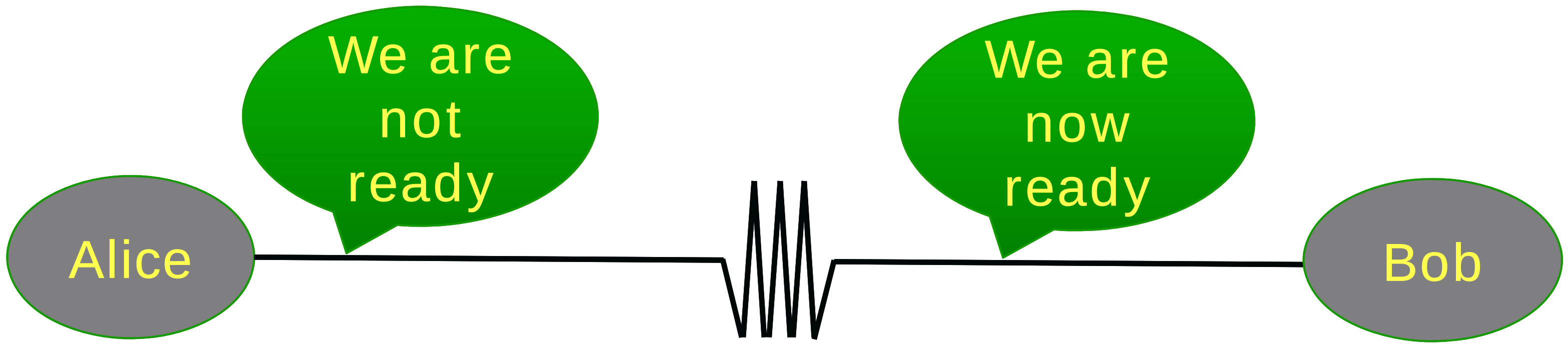
Reliable Meaningful Communication

Madhu Sudan
Microsoft, Cambridge, USA

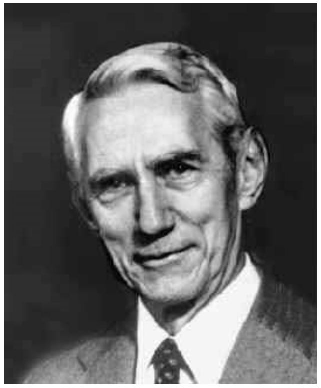
Reliable Communication?

- Problem from the 1940s: Advent of digital age.



- Communication media are always noisy
  - But digital information less tolerant to noise!

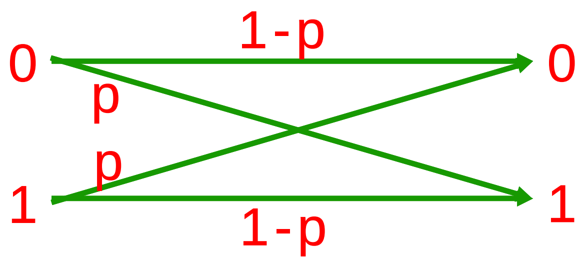
Theory of Communication



- [Shannon, 1948]
  - Model for noisy communication channels
  - Architecture for reliable communication
  - Analysis of some communication schemes
  - Limits to any communication scheme

Modelling Noisy Channels

- Channel = Probabilistic Map from Input to Output
  - Example: Binary Symmetric Channel (BSC(p))



Some limiting values

- p=0
  - Channel is perfectly reliable.
  - No need to do anything to get 100% utilization

    (1 bit of information received/bit sent)

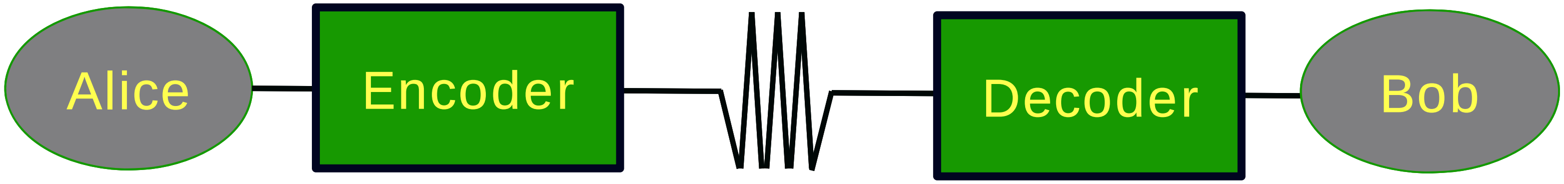
- p=½
  - Channel output independent of sender's signal.
  - No way to get any information through.

    (0 bits of information received/bit sent)

Lessons from Repetition

- Can repeat (retransmit) message bits many times
  - E.g., 0100 $\rightarrow$ 000 111 000 000
  - Decoding: take majority
    - E.g., 010 110 011 100 $\rightarrow$ 0110
  - Utilization rate = 1/3
  - More we repeat, more reliable the transmission.
  - More information we have to transmit, less reliable is the transmission.
- Tradeoff inherent in all schemes?
- What do other schemes look like?

Shannon's Architecture



- Sender "Encodes" before transmitting
- Receiver "Decodes" after receiving
- Encoder/Decoder arbitrary functions.

$$E: \{0,1\}^k \to \{0,1\}^n$$
$$D: \{0,1\}^n \to \{0,1\}^k$$

  - Rate $= \frac{k}{n}$;
  - Hope: Usually $m = D(E(m) + \text{error})$

Shannon's Analysis

- Coding Theorem:
  - For every p, there exists Encoder/Decoder that corrects p fraction errors with high probability with Rate $\rightarrow 1 - H(p)$
  - $H(p)$: Binary entropy function:
    - $H(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$
    - $H(0) = 0; H\left(\frac{1}{2}\right) = 1$; $H(p)$ monotone $0 < p < \frac{1}{2}$
  - So if $p = .499$; Channel still has utility!
  - Note on probability: Goes to 1 as $k \rightarrow \infty$

Limit theorems

- Converse Coding Theorem:
  - If Encoder/Decoder have Rate $> 1 - H(p)$ then decoder output wrong with prob. $1 - \exp(-n)$.

- Entropy is right measure of loss due to error.
- Entropy = ?
  - Measures uncertainty of random variable.
  - (In our case: Noise).

An aside: Data Compression

- Noisy encoding + Decoding $\Rightarrow$ Message + Error
  - (Receiver knows both).
  - Total length of transmission $= n$
  - Message length $= n - H(p).n$
  - So is error-length $= H(p).n$?
- Shannon's Noiseless Coding Theorem:
  - Information (modelled as random variable) can be compressed to its entropy … with some restrictions
  - General version due to Huffman

1948-2013?

- [Hamming 1950]: Error-correcting codes
  - More "constructive" look at encoding/decoding functions.
- Many new codes/encoding functions:
  - Based on Algebra, Graph-Theory, Probability.
- Many novel algorithms:
  - Make encoding/decoding efficient.
  - Result:
    - Most channels can be exploited.
    - Even if error is not probabilistic.
    - Profound influence on practice.

Modern Challenges

New Kind of Uncertainty

- Uncertainty always has been a central problem:
  - But usually focusses on uncertainty introduced by the channel
  - Rest of the talk: Uncertainty at the endpoints (Alice/Bob)
- Modern complication:
  - Alice+Bob communicating using computers
  - Both know how to program.
  - May end up changing encoder/decoder (unintentionally/unilaterally).
- Alice: How should I "explain" to Bob?
- Bob: What did Alice mean to say?

New Era, New Challenges:

- Interacting entities not jointly designed.
  - Can’t design encoder+decoder jointly.
  - Can they be build independently?
  - Can we have a theory about such?
    - Where we prove that they will work?

    - Hopefully:
      - YES
      - And the world of practice will adopt principles.

Example Problem

- Archiving data
  - Physical libraries have survived for 100s of years.
  - Digital books have survived for five years.
  - Can we be sure they will survive for the next five hundred?

- Problem: Uncertainty of the future.
  - What formats/systems will prevail?
  - Why aren't software systems ever constant?

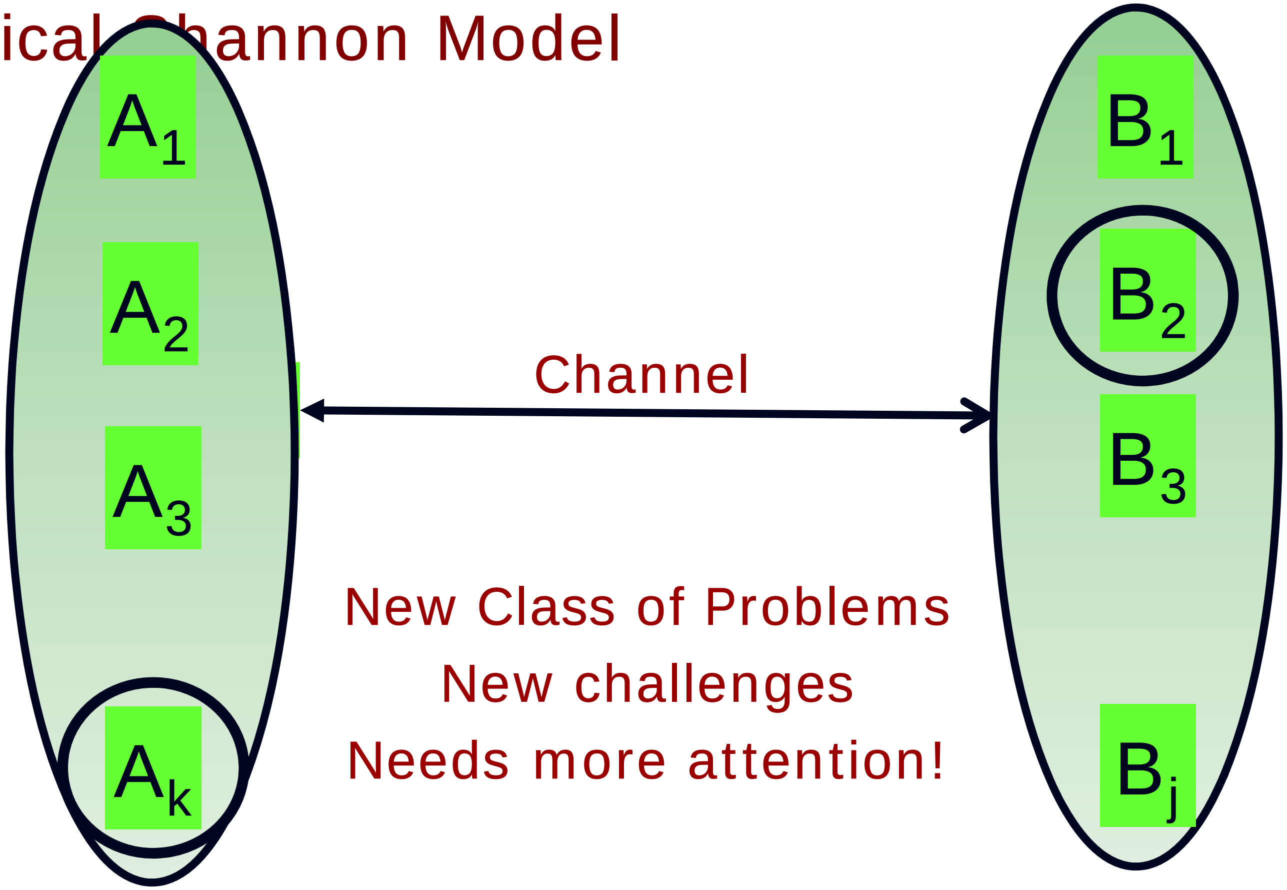
Challenge:

- If Decoder does not know the Encoder, how should it try to guess what it meant?
- Similar example:
  - Learning to speak a foreign language
    - Humans do … (?)
      - Can we understand how/why?
      - Will we be restricted to talking to humans only?
      - Can we learn to talk to “aliens”? Whales? ☺
- Claim:
  - Questions can be formulated mathematically.
  - Solutions still being explored.

Modelling uncertainty

Uncertain Communication Model

Classical Shannon Model



New Class of Problems

New challenges

Needs more attention!

Language as compression

- Why are dictionaries so redundant+ambiguous?
  - Dictionary = map from words to meaning
  - For many words, multiple meanings
  - For every meaning, multiple words/phrases
  - Why?
- Explanation: "Context"
  - Dictionary:
    - Encoder: Context1 × Meaning → Word
    - Decoder: Context2 × Word → Meaning
    - Tries to compress length of word
    - Should works even if Context1 ≠ Context2
- [Juba,Kalai,Khanna,S'11],[Haramaty,S'13]: Can design encoders/decoders that work with uncertain context.

A challenging special case

- Say Alice and Bob have rankings of N movies.
  - Rankings = bijections $\pi, \sigma : [N] \rightarrow [N]$
  - $\pi(i)$ = rank of i^{th} player in Alice's ranking.
- Further suppose they know rankings are close.
  - $\forall\, i \in [N]$: $|\pi(i) - \sigma(i)| \leq 2$.
- Bob wants to know: Is $\pi^{-1}(1) = \sigma^{-1}(1)$
- How many bits does Alice need to send (non-interactively).
  - With shared randomness – $O(1)$
  - Deterministically?
    - $O(1)$? $O(\log N)$? $O(\log\log\log N)$?

Meaning of Meaning?

- Difference between meaning and words
  - Exemplified in
    - Turing machine vs. universal encoding
    - Algorithm vs. computer program
  - Can we learn to communicate former?
    - Many universal TMs, programming languages
- [Juba,S.'08], [Goldreich,Juba,S.'12]:
  - Not generically …
  - Must have a goal: what will we get from the bits?
  - Must be able to sense progress towards goal.
  - Can use sensing to detect errors in understanding, and to learn correct meaning.
- [Leshno,S'13]:
  - Game theoretic interpretation

Communication as Coordination Game [Leshno,S.'13]

- Two players playing series of coordination games
  - Coordination?
    - Two players simultaneously choose 0/1 actions.
    - "Win" if both agree:
      - Alice's payoff: not less if they agree
      - Bob's payoff: strictly higher if they agree.
    - How should Bob play?
      - Doesn't know what Alice will do. But can hope to learn.
      - Can he hope to eventually learn her behavior and (after finite # of miscoordinations) always coordinate?
- Theorem:
  - Not Deterministically (under mild "general" assumptions)
  - Yes with randomness (under mild restrictions)

Summary

- Understanding how to communicate meaning is challenging:
  - Randomness remains key resource!
  - Much still to be explored.
  - Needs to incorporate ideas from many facets
    - Information theory
    - Computability/Complexity
    - Game theory
    - Learning, Evolution ...
- But Mathematics has no boundaries ...

Thank You!