Communication amid Uncertainty

Madhu Sudan
Microsoft, Cambridge, USA

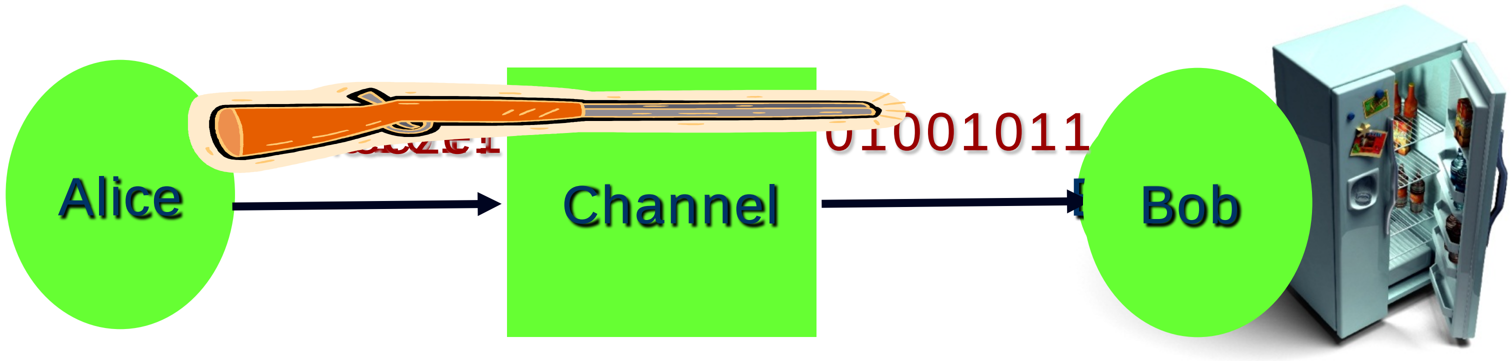
Based on:

- Universal Semantic Communication – Juba & S. (STOC 2008)
- Goal-Oriented Communication – Goldreich, Juba & S. (JACM 2012)
- Compression without a common prior … – Kalai, Khanna, Juba & S. (ICS 2011)

Overview

- I. Motivations/Ramblings
- II. Example: Compression
- III. General “Goal-oriented communication”

The Meaning of Bits



- Is this perfect communication?

- What if Alice is trying to send instructions?
  - In other words … an algorithm
  - Does Bob understand the correct algorithm?
  - What if Alice and Bob speak in different (programming) languages?

Shannon on Semantics

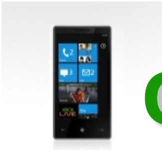
"*The semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages*"

Shannon '48

Claim: Today, the semantics are becoming important to engineering.

Importance of semantics

- Why is semantics (relatively) important today?
  - Factor 1: Success of the Shannon program:
    - Reliability, in syntactic sense, has been achieved.
  - Factor 2: Communication vs. Computing.

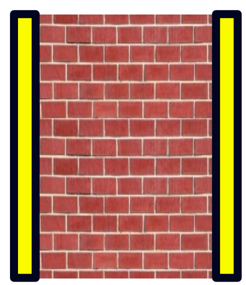


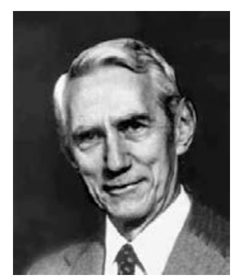
Communication vs. Computation

- Interdependent technologies: Neither can exist without other
- Technologies/Products/Commerce developed (mostly) independently.
  - Early products based on clean abstractions of the other.
  - Later versions added other capability as afterthought.
  - Today products … deeply integrated.
- Deep theories:

Well separated … and have stayed that way

Turing '36



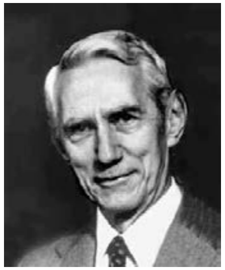


Shannon '48

Time for the theoretical wall to come down?

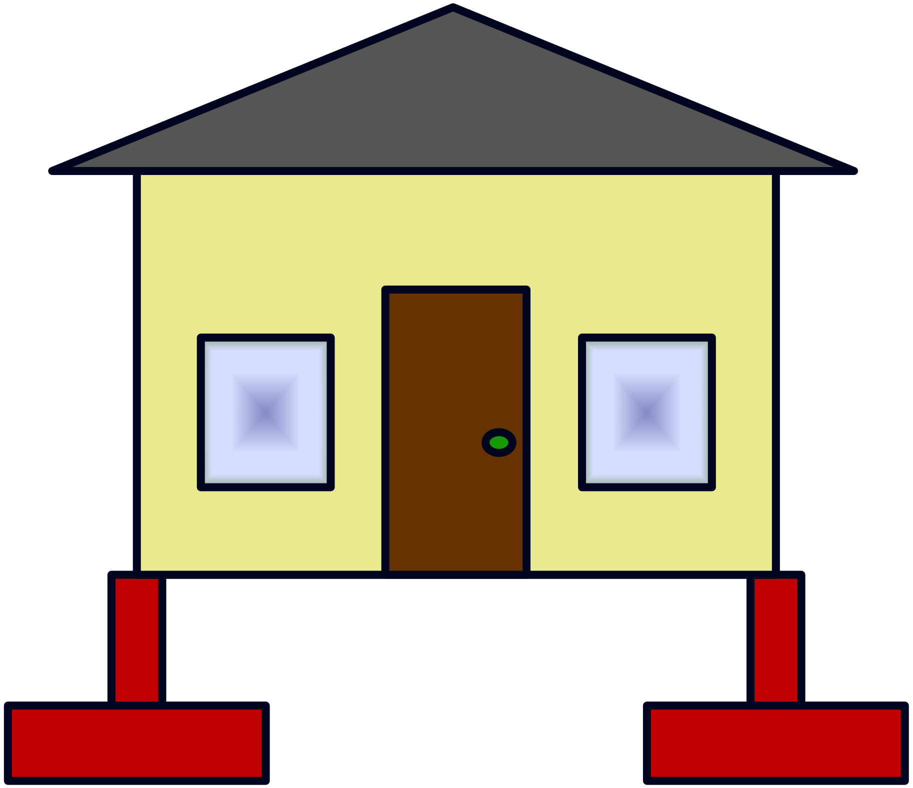
Consequences of the wall

- Computing theory:
  - Fundamental principle = Universality
  - You can program your computer to do whatever you want.
- Communication principle:
  - Centralized design (Encoder, Decoder, Compression, IPv4, TCP/IP).
  - You can NOT program your device!
- Contradiction! But does it matter?
  - Yes! As in semantics



Role of theory?

- Ideally: Foundations of practice!

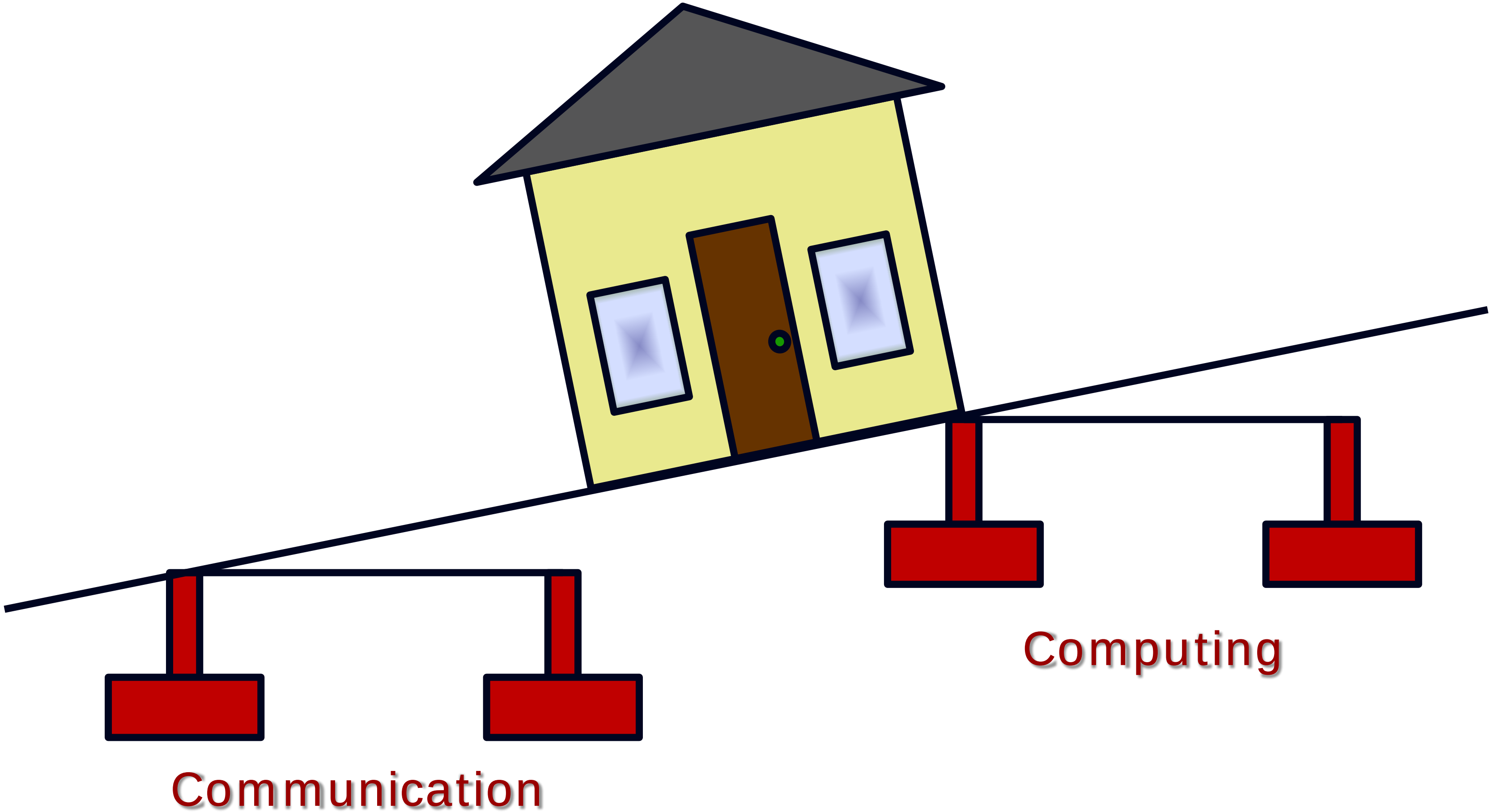


Application

Theory layer

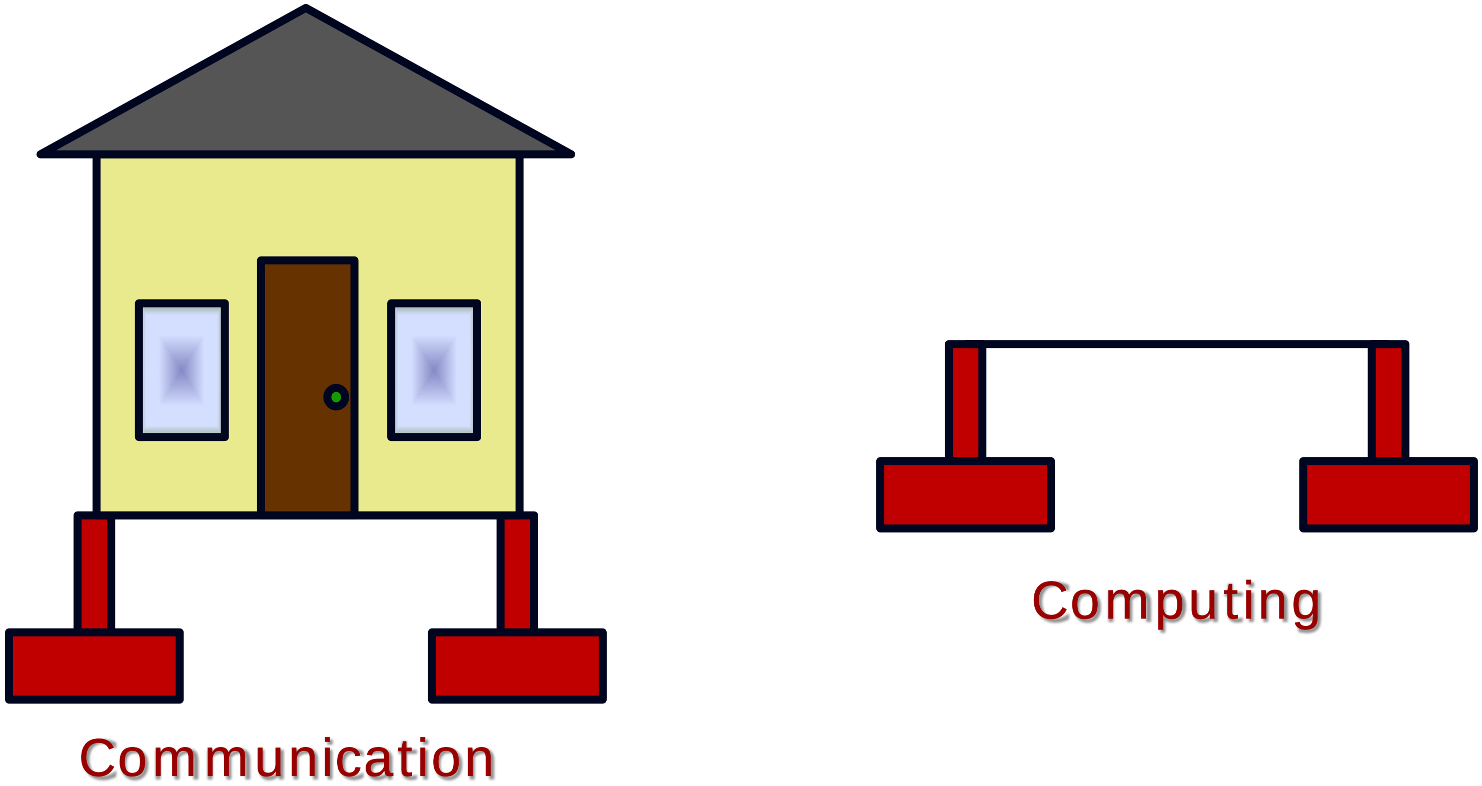
Communication vs. Computing

- Option 1



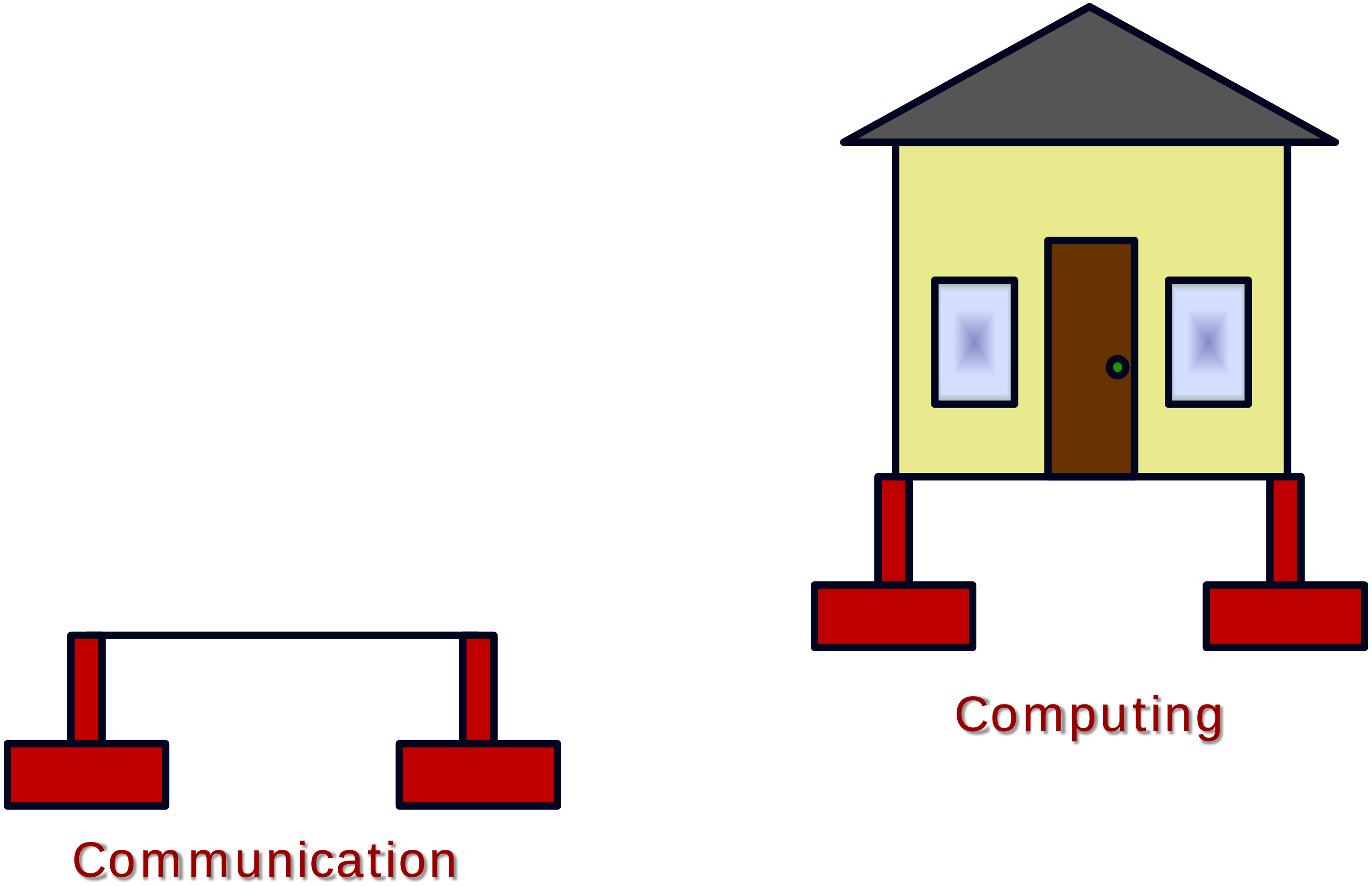
Communication vs. Computing

- Option 2



Communication vs. Computing

- Option 3



Good News/ Bad News

- Good: We are mostly practicing option 2 or 3!
- Bad:
  - Lost opportunities.
  - Vulnerabilities.
  - Inefficiency.
  - Incompatibilities.

Sample problems:

- Universal printing:
  - You are visiting a new university. Can your machine not learn how to print on the local printer, without requiring installation?
- Projecting from your laptop:
  - Machines that learn to communicate, and learn to understand each other.
- Digital libraries:
  - Data that lives forever (communication across time), while devices change.

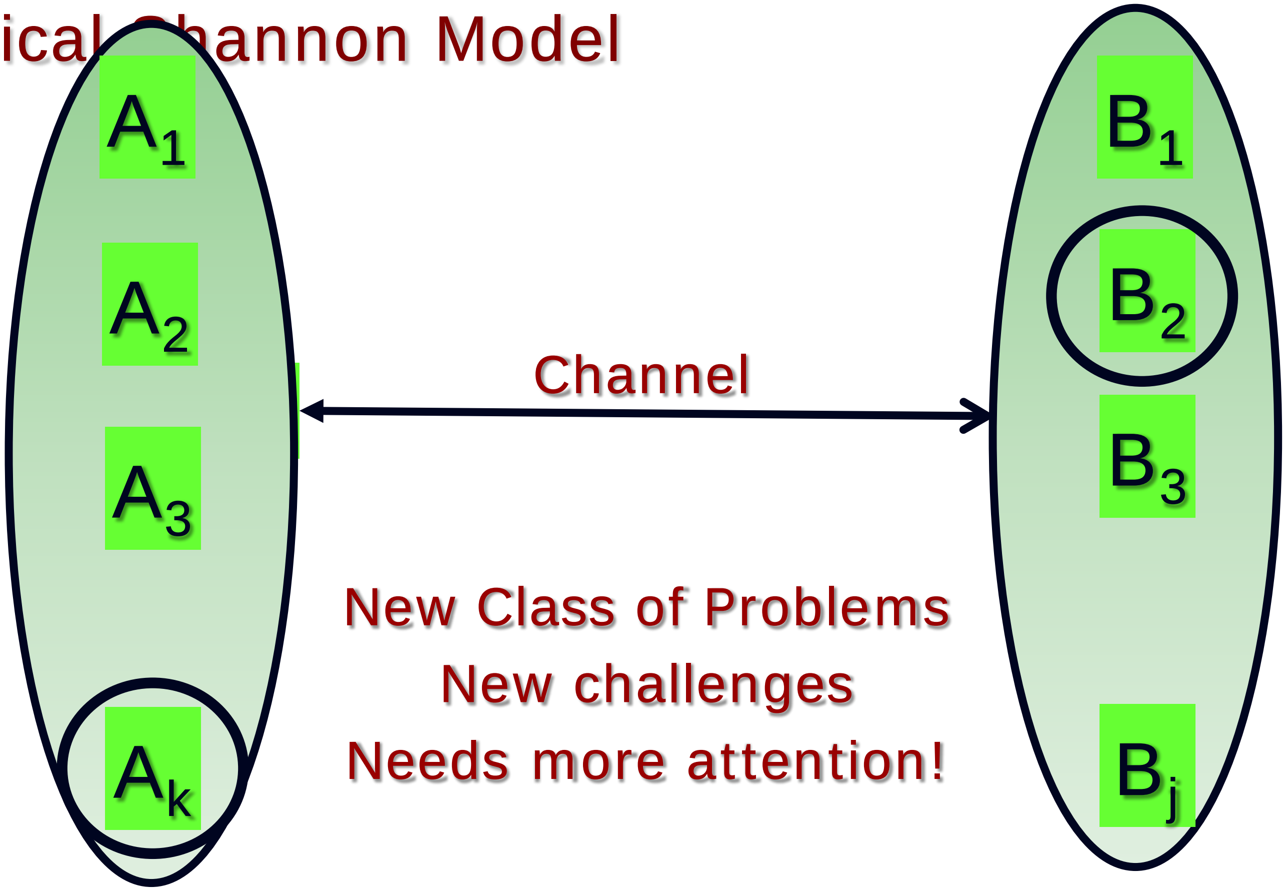
Essence of "semantics": Uncertainty

- Recall Shannon:
  - *"The significant aspect is that the actual message is one selected from a set of possible messages"*
- Essence of unreliability today:
  - Sender and receiver in disagreement about set of possible messages (or about their meaning).

Modelling uncertainty

Semantic Communication Model

Classical Shannon Model



New Class of Problems

New challenges

Needs more attention!

[Kalai,Khanna,J.,S. – ICS 2011] Compression in this setting: Leads to ambiguous, redundant compression

II: Non-interactive communication: Compression

Motivation: Human Communication

- Human communication (dictated by languages, grammars) very different.
  - Grammar: Rules, often violated.
  - Dictionary: Often multiple meanings to a word.
  - Redundant: But not as in any predefined way (not an error-correcting code).
  - Our thesis: Emerges from uncertainty:
    - Sender of message uncertain about receiver's background/context/prior.
    - Will try to explain in the context of Redundancy

Behavioral aspects of natural communication

- (Vast) Implicit context.
- Sender sends increasingly long messages to receiver till receiver “gets” (the meaning of) the message.
- Sender may use feedback from receiver if available; or estimates receiver’s knowledge if not.
- Language provides sequence of (increasingly) long ways to represent a message.
- Question: What is the benefit of choosing short/long messages?

Model:

- Reason to choose short messages: Compression.
  - Channel is still a scarce resource; still want to use optimally.

- Reason to choose long messages (when short ones are available): Reducing ambiguity.
  - Sender unsure of receiver's prior (context).

    ("uncertainty")
  - Sender wishes to ensure receiver gets the message, no matter what its prior (within reason).

Model

- Wish to design encoding/decoding schemes (E/D) to be used as follows:
  - Sender has distribution P on M = {1,2,...,N}
  - Receiver has distribution Q on M = {1,2,...,N}
  - Sender gets $X \in M$
  - Sends E(P,X) to receiver.
  - Receiver receives Y = E(P,X)
  - Decodes to $\hat{X}$ = D(Q,Y)

  - Want: X = $\hat{X}$ (provided P,Q close),
    - While minimizing $Exp_{X \leftarrow P} |E(P,X)|$

Contrast with some previous models

- Universal compression?
  - Doesn't apply: P,Q are not finitely specified.
  - Don't have a sequence of samples from P; just one!
- K-L divergence?
  - Measures inefficiency of compressing for Q if real distribution is P.
  - But assumes encoding/decoding according to same distribution Q.
- Semantic Communication:
  - Uncertainty of sender/receiver; but no special goal.

Closeness of distributions:

- P is α-close to Q if for all $X \in M$,

$$\frac{1}{\alpha} \le \frac{P(X)}{Q(X)} \le \alpha$$

- P α-close to Q $\Rightarrow$ $D(P||Q), D(Q||P) \le \log\alpha$.

Dictionary = Shared Randomness?

- Modelling the dictionary: What should it be?

- Simplifying assumption – it is shared randomness, so …

- Assume sender and receiver have some shared randomness R and X is independent of R.
  - Y = E(P,X,R)
  - $\hat{X}$ = D(Q,Y,R)

- Want $\forall X,\ \Pr_R\left[\hat{X} = X\right] \geq 1 - \epsilon$

Solution (variant of Arith. Coding)

- Use R to define sequences
  - $R_1\,[1], R_1\,[2], R_1\,[3], \ldots$
  - $R_2\,[1], R_2\,[2], R_2\,[3], \ldots$
  - …
  - $R_N\,[1], R_N\,[2], R_N\,[3], \ldots .$
- $E_\alpha(P, x, R) = R_x[1 \ldots L]$, where L chosen s.t. $\forall z \neq x$

  Either $R_z[1 \ldots L] \neq R_x[1 \ldots L]$

  Or $P(z) < \frac{P(x)}{\alpha^2}$
- $D_\alpha(Q, y, R) = \text{argmax}_{\hat{x}}\ \{Q(\hat{x})\}$ among $\hat{x} \in \{\, z \mid R_z[1 \ldots L] = y\}$

Performance

- Obviously decoding always correct.
- Easy exercise:
  - $\text{Exp}_X\ [E(P,X)] = H(P) + 2\log\alpha$
- Limits:
  - No scheme can achieve $(1-\epsilon)\cdot[H(P)+\log\alpha]$
  - Can reduce randomness needed.

Implications

- Reflects the tension between ambiguity resolution and compression.
  - Larger the α ((estimated) gap in context), larger the encoding length.
- Coding scheme reflects the nature of human process (extend messages till they feel unambiguous).
- The "shared randomness" is a convenient starting point for discussion
  - Dictionaries do have more structure.
  - But have plenty of entropy too.
  - Still … should try to do without it.

III: Interactive Setting:
Goal-Oriented Communication

Back to meaning

- What if sender is sending instructions?
  - Can we ensure receiver decodes the right instructions?
      - Freeze?

- Translation of bits to instructions?
  - Well studied in language/computer science.
  - (Many) “Complete” languages/codebooks exist.
    - Each translates bits to meaning.
    - All equivalent (upto “Kolmogorov constant”)
    - But not same.

Goal of communication

- Easy negative result:
  - (Due to plethora of languages/codebooks): In finite time, **can't** guarantee "receiver understands instructions."
  - Is this bad?
    - If receiver can not distinguish correct instructions from incorrect ones, why should it try to do so?
- Goals of communication:
  - Communication is not an end in itself, it a means to achieving some end.
  - Semantic communication:
    - Help communication achieve its goal.
    - Use progress towards goal to understand meaning.

Computation as a goal [Juba+S., 2008]

- The lens of computational complexity:
  - To prove some resource is useful:
    - Step 1: Identify hardest problems one can solve without the resource.
    - Step 2: Show presence of resource can help solve even harder problems.
- Classical resources:
  - CPU speed, Memory, Non-determinism, Randomness ...
- In our case:
  - Communication in presence of understanding.
  - Communication w/o understanding.

Computation as a goal

- Model: Simple user talking to powerful server.
  - Class of problems user can solve on its own:

    ~ probabilistic polynomial time (P).
  - Class of problems user can solve with perfect understanding of server:

    ~ Any problem. (Even uncomputable!)
  - Class of problems user can solve without understanding of server:

    ~ Polynomial space.
- Roughly: If you are solving problems and can verify solutions, then this helps. If you have a solution, you are done. If not, you've found some error in communication.
- Moral: Communication helps, even with misunderstanding, but misunderstanding introduces limits.

Summarizing results of [GJS 2012]

- But not all goals are computational.
  - We use communication mostly for (remote) control.
  - Intellectual/informational goals are rare(r).
- Modelling general goals, in the presence of misunderstanding:
  - Non-trivial, but can be done.
  - Results extend those from computational setting:
    - Goals can be achieved if user can sense progress towards goal, servers are "forgiving" and "helpful"

Useful lessons

- User/Server can be designed separately.
- Each should attempt to model its “uncertainty” about the other.
- Each should plan for uncertainty:
  - Server: By assuming some short “interrupt” sequence.
  - User: By always checking its progress.

Future goals

- Broadly:
  - Information-theoretic study of human communication, with uncertainty as an ingredient.
    - Should exploit natural restrictions of humans:
      - Limited ability to learn/infer/decode.
      - Limited bandwidth.
  - Conversely, use human interactions to create alternate paradigms for “designed communications.
    - Place semantics on solid foundations.

Thank You!