Communication Amid Uncertainty

Madhu Sudan
Harvard

Based on

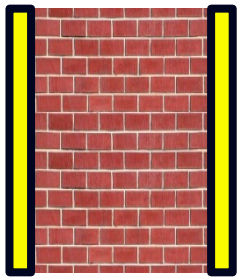
- Juba, S. (STOC 2008, ITCS 2011)
- Goldreich, Juba, S. (JACM 2011)
- Juba, Kalai, Khanna, S. (ITCS 2011)
- Haramaty, S. (ITCS 2014)
- Canonne, Guruswami, Meka, S. (ITCS 2015)
- Ghazi, Kamath, S. (SODA 2016)
- Ghazi, Komargodski, Kothari, S. (SODA 2016)
- Leshno, S. (manuscript)

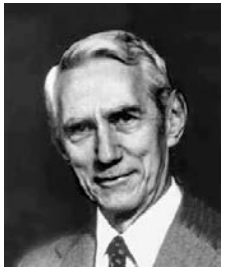
Communication vs. Computation

- 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



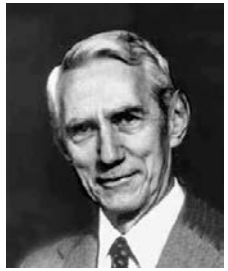


Shannon '48

Consequences of the wall

- Computing theory:
  - Fundamental principle = Universality
  - You can program your computer to do whatever you want.
  - ⇒ Heterogeneity of devices

- Communication theory:
  - Centralized design (Encoder, Decoder, Compression, IPv4, TCP/IP).
  - You can NOT program your device!
  - ⇐ Homogeneity of devices



- Contradiction! But does it matter?
  - Yes!

Sample problems:

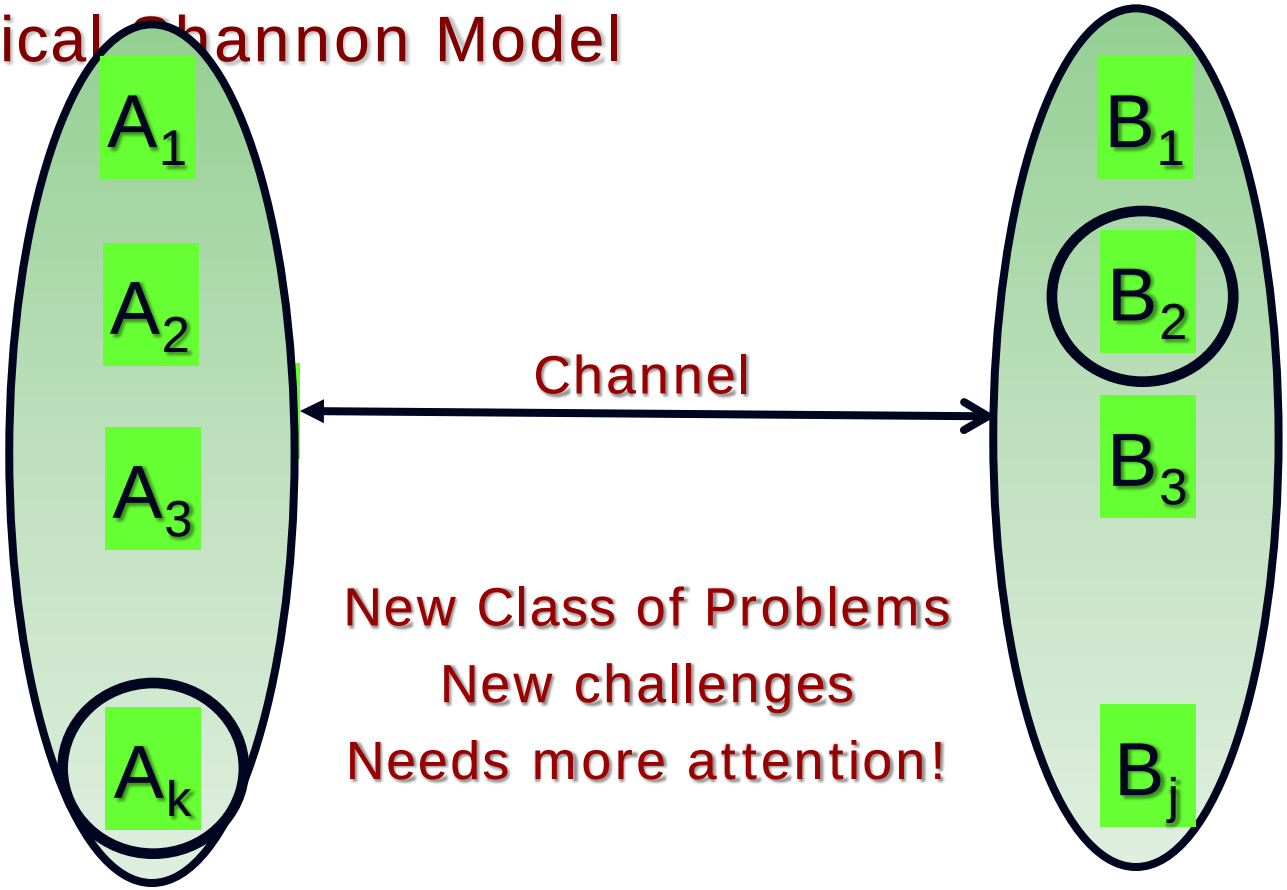
- Universal printing:
  - You are visiting a friend. You can use their Wifi network, but not their printer. Why?
- Projecting from your laptop:
  - Machines that learn to communicate, and learn to understand each other.
- Digital libraries:
  - Data that lives forever (communication across time), while devices change.

Essence of "semantics": Uncertainty

- Shannon:
  - *"The significant aspect is that the actual message is one selected from a set of possible messages"*
- Essence of unreliability today:
  - Context: Determines set of possible messages.
    - dictionary, grammar, general knowledge
    - coding scheme, prior distribution, communication protocols …
  - Context is HUGE; and not shared perfectly;

Modelling uncertainty

Uncertain Communication Model
Classical Shannon Model



New Class of Problems
New challenges
Needs more attention!

Hope

- Better understanding of existing mechanisms
  - In natural communication
  - In “ad-hoc” (but “creative”) designs
- What problems are they solving?
- Better solutions?
  - Or at least understand how to measure the quality of a solution.

II: Uncertain Compression

Human-Human Communication

$$M_1 = w_{11}, w_{12}, \ldots$$
$$M_2 = w_{21}, w_{22}, \ldots$$
$$M_3 = w_{31}, w_{32}, \ldots$$
$$M_4 = w_{41}, w_{42}, \ldots$$
$$\ldots$$

- Role of dictionary = ?
  - [Juba, Kalai, Khanna, S. 11]
- Dictionary: list of words representing message
  - words appear against multiple messages
  - multiple words per message.
- How to decide which word to use? Context!
  - Encoding: Given message, use shortest unambiguous word in current context.
  - Decoding: Given word, use most likely message in current context, (among plausible messages)
- Context = ???. Prob. distribution on messages

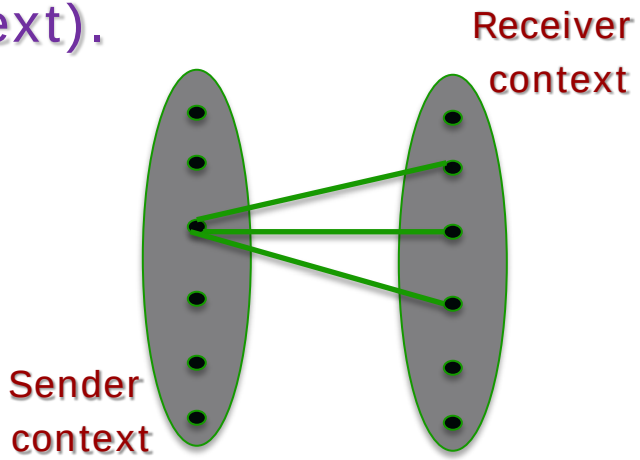
$$P_i = \text{Prob}\,[\text{message} = M_i]$$

Human Communication - 2

$M_1 = w_{11}, w_{12}, \ldots$
$M_2 = w_{21}, w_{22}, \ldots$
$M_3 = w_{31}, w_{32}, \ldots$
$M_4 = w_{41}, w_{42}, \ldots$
$\ldots$

- Good (Ideal?) dictionary
  - Should compress messages to entropy of context: $H(P = \langle P_1, \ldots, P_N \rangle)$.
- Even better dictionary?
  - Should not assume context of sender/receiver identical!
  - Compression should work even if sender **uncertain** about receiver (or receivers' context).

Theorem [JKKS]: If dictionary is "random" then compression achieves message length $H(P) + \Delta$, if sender and receiver distributions are "Δ-close".



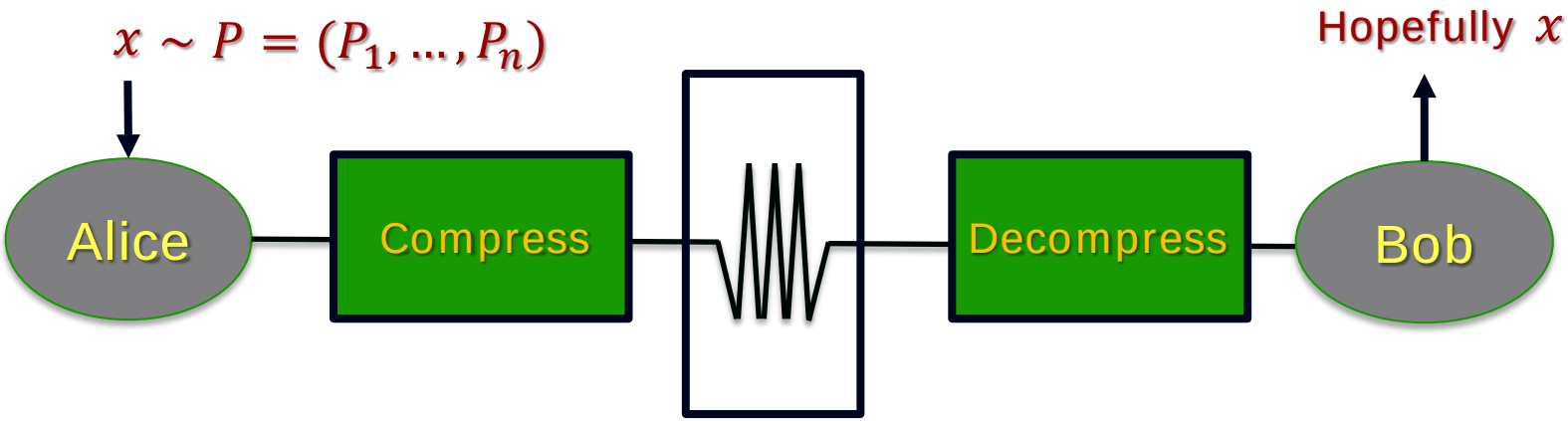
Implications

- Reflects tension between ambiguity resolution and compression.
  - Larger the gap in context (Δ), larger the encoding length.
- Coding scheme reflects human communication?
- "Shared randomness" debatable assumption:
  - Dictionaries do have more structure.
  - Deterministic communication? [Haramaty+S,14]
  - Randomness imperfectly shared? Next ...

III: Imperfectly Shared Randomness

Communication (Complexity)

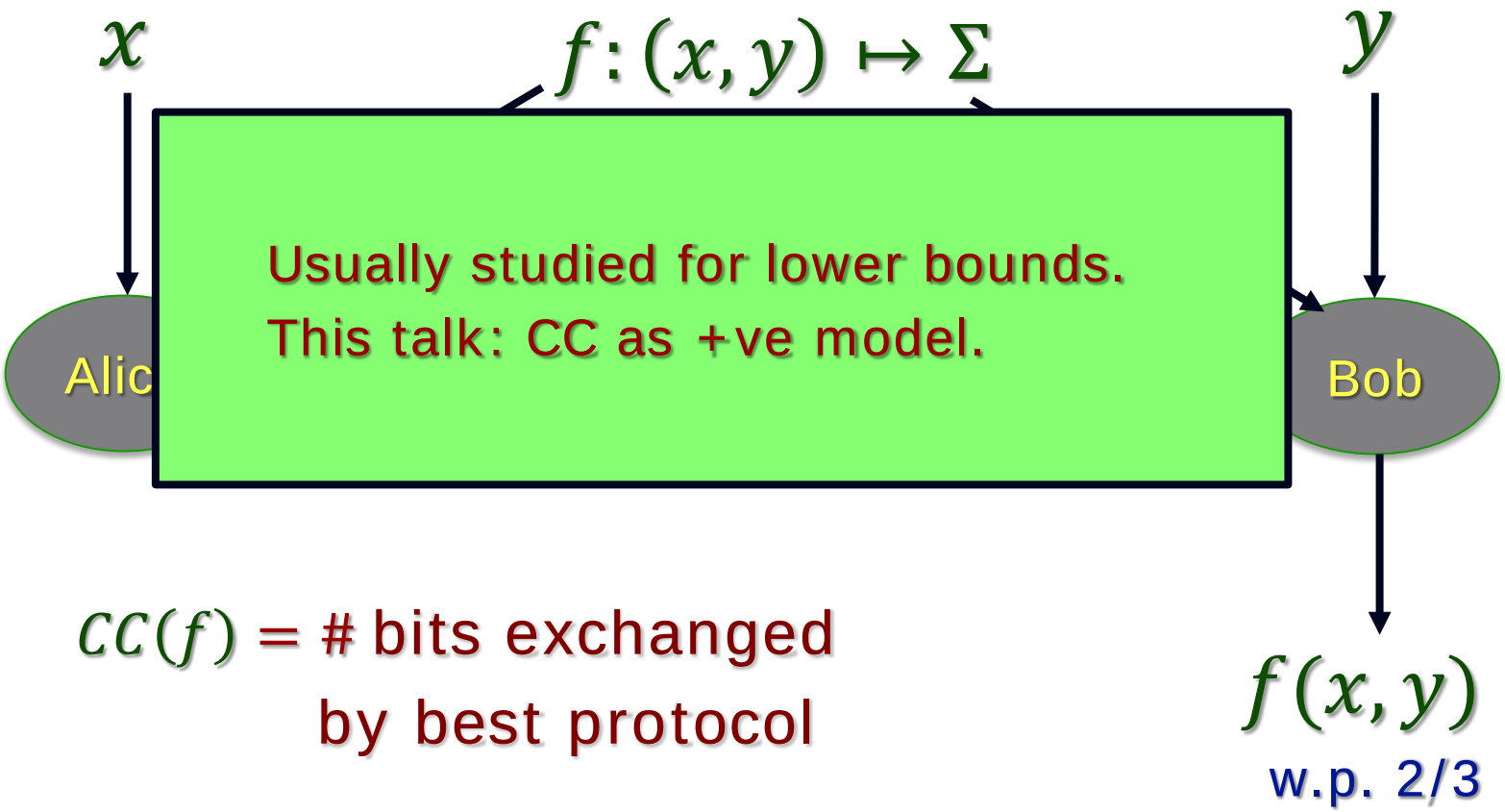
- Compression (Shannon, Noiseless Channel)



- What will Bob do with x?
  - Often knowledge of x is overkill.

(Recall) Communication Complexity

The model (with shared randomness)

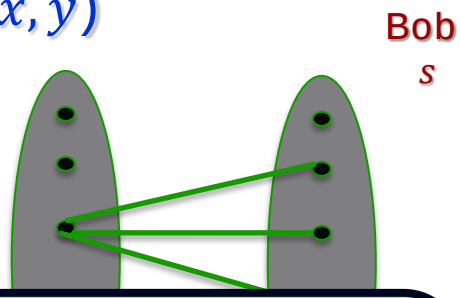


Usually studied for lower bounds.
This talk: CC as +ve model.

$CC(f) =$ # bits exchanged by best protocol

Brief history

- ∃ problems where Alice can get away with much fewer bits of communication.
  - Example: $\oplus(x,y) \triangleq \oplus_i (x_i \oplus y_i)$
  - But very few such deterministically.
- Enter Randomness:
  - Alice & Bob share random string r (ind. of x, y)
  - Many more problems; Example:
    - $\mathrm{Eq}(x,y) = 1$ if $x = y$ and 0 otherwise
      - Deterministically: $\Theta(n)$



[Ghazi, Kamath, S., SODA 2016]: Taxonomy of simple problems;
Many interesting problems and protocols!
can be more <u>effective</u> (shorter than $|x|, H(x), H(y), I(x;y) \ldots$)

Results

- [Newman '90s]:

  CC without sharing $\leq CC$ with sharing $+ \log n$
- But additive cost of $\log n$ may be too much.
  - Compression! Equality!!
- Model recently studied by [Bavarian et al.'14]
  - Equality: $O(1)$ bit protocol w. imperfect sharing
- Our Results: [Canonne, Guruswami, Meka, S.'15]
  - Compression: $O(H(P) + \Delta)$
  - Generally: k bits with shared randomness

    $\Rightarrow O(2^k)$ bits with imperfect sharing.
  - $k \to 2^k$ loss is necessary.

Some General Lessons

- Compression Protocol:
  - Adds “error-correction” to [JKKS] protocol.
    - Send shortest word that is far from words of other high probability messages.
    - Another natural protocol.
- General Protocol:
  - Much more “statistical”
    - Classical protocol for Equality:
      - Alice sends random coordinate of ECC(x)
    - New Protocol
      - ~ Alice send # 1’s in random subset of coordinates.

IV: Focussed Communication

Model

- Bob wishes to compute $f(x,y)$; Alice knows $g \approx f$;
- Alice, Bob given g, f explicitly. (New input size $\sim 2^n$)
- Modelling Questions:
  - What is $\approx$?
  - Is it reasonable to expect to compute $f(x,y)$?
    - E.g., $f(x,y) = f'(x)$? Can't compute $f(x,y)$ without communicating x
- Our Choices:
  - Assume x, y come from a distribution μ
  - $f \approx g$ if $f(x,y)$ usually equals $g(x,y)$
  - Suffices to compute $h(x,y)$ for $h \approx f$

Results

- Thm [Ghazi,Komargodski,Kothari,S]: $\exists f \approx g$, with $CC(f), CC(g) = 1$, but uncertain complexity $\approx \sqrt{n}$
- Thm [GKKS]: But not if x independent of y (in 1-way setting).
  - 2-way setting, even 2-round, open!
- Main Idea:
  - Canonical 1-way protocol for f:
    - Alice + Bob share random $y_1, \ldots y_m \in \{0,1\}^n$.
    - Alice sends $g(x, y_1), \ldots, g(x, y_m)$ to Bob.
    - Protocol used previously … but not as "canonical".
  - Canonical protocol robust when $f \approx g$.

Conclusions

- Positive view of communication complexity: Communication with a focus can be effective!
- Context Important:
  - (Elephant in the room: Huge, unmentionable)
  - New layer of uncertainty.
  - New notion of scale (context LARGE)
    - Importance of $o(\log n)$ additive factors.
- Many "uncertain" problems can be solved without resolving the uncertainty (which is a good thing)
- Many open directions+questions

Thank You!