Mathematical Theories of Communication:
Old and New

Madhu Sudan
Harvard University
Communication = What?

- **Today: Digital Communication**
  - i.e., Communicating bits/bytes/data ...
  - As opposed to “radio waves for sound”

- **Challenge? Communication can be ...**
  - ... expensive: (e.g., satellite with bounded energy to communicate)
  - ... noisy: (bits flipped, DVD scratched)
  - ... interactive: (complicates above further)
  - ... contextual (assumes you are running specific OS, IPvX)
Theory = Why?

- Why build theory and not just use whatever works?
  - Ad-hoc solutions work today, but will they work tomorrow?
  - Formulating problem allows comparison of solutions!
    - And some creative solutions might be surprisingly better than naïve!
  - Better understanding of limits.
  - What can not be achieved ...
Old? New?

- Why new theories? Were old ones not good enough?
  - Quite the opposite: Old ones were too good.
  - They provided right framework and took us from ground level to “orbit”!
  - And now we can explore all of “space” ... but new possibilities leads to new challenges.
Problem from the 1940s: Advent of digital age.

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

- Can repeat (every letter of) message to improve reliability:
  
  WWW EEE AAA RRR EEE NNN OOO WWW ...
  
  ↓
  
  WXW EEA ARA SSR EEE NMN OOP WWW ...

- Elementary Reasoning:
  
  \[\begin{align*}
  \uparrow \text{repetitions} & \Rightarrow \downarrow \text{Prob. decoding error; but still +ve} \\
  \uparrow \text{length of transmission} & \Rightarrow \uparrow \text{expected # errors.}
  \end{align*}\]

- Combining above: Rate of repetition coding \(\rightarrow 0\) as length of transmission increases.

- Belief (pre1940):
  
  \[\begin{align*}
  \text{Rate of any scheme} & \rightarrow 0 \text{ as length} \rightarrow \infty
  \end{align*}\]
Shannon’s Theory [1948]

- Sender “Encodes” before transmitting
- Receiver “Decodes” after receiving

Encoder/Decoder arbitrary functions.

\[ E : \{0,1\}^k \rightarrow \{0,1\}^n \]
\[ D : \{0,1\}^n \rightarrow \{0,1\}^k \]

- Rate \( = \frac{k}{n} \);
- Requirement: \( m = D(E(m) + \text{error}) \) w. high prob.
- What are the best \( E, D \) (with highest Rate)?
Shannon’s Theorem

- If every bit is flipped with probability $p$:
  - Rate $\rightarrow 1 - H(p)$ can be achieved.
  - $H(p) \triangleq p \log_2 \frac{1}{p} + (1 - p) \log_2 \frac{1}{1-p}$
  - This is best possible.
- Examples:
  - $p = 0 \Rightarrow Rate = 1$
  - $p = \frac{1}{2} \Rightarrow Rate = 0$
  - Monotone decreasing for $p \in (0, \frac{1}{2})$
  - Positive rate for $p = 0.4999$; even if $k \rightarrow \infty$
Shannon’s contributions

- Far-reaching architecture:

- Profound analysis:
  - First (?) use of probabilistic method.

- Deep Mathematical Discoveries:
  - Entropy, Information, Bit?
Challenges post-Shannon

- Encoding/Decoding functions not “constructive”.
  - Shannon picked $E$ at random, $D$ brute force.
  - Consequence:
    - $D$ takes time $\sim 2^k$ to compute (on a computer).
    - $E$ takes time $2^{2^k}$ to find!

- Algorithmic challenge:
  - Find $E, D$ more explicitly.
  - Both should take time $\sim k, k^2, k^3$ ... to compute
  - Solutions: 1948-2017
Chomsky: Theory of “Language”

- Formalized syntax in languages mathematically (with caveats).
  - Initially goal was to understand inter-human communication!
    - Structure behind languages
    - Implication on acquisition.
  - Major side-effect: Use of “context-free grammars” to specify programming languages!
    - Automated parsing, compiling, interpretation!
Modern Theories:

- Communication is interactive
  - E.g., "buying airline ticket online"
- Involve large inputs from either side:
  - Travel agent = vast portfolio of airlines+flights
  - Traveler = complex collection of constraints and preferences.
- Best protocol ≠ Travel agent sends brochure.
  ≠ Traveller sends entire list of constraints.
- How to model? What happens if errors happen? How well should context be shared?
The model (with shared randomness)

\[ f : (x, y) \mapsto \Sigma \]

\[ R = $$$ \]

\[ CC(f) = \# \text{ bits exchanged by best protocol} \]

Usually studied for lower bounds. This talk: CC as +ve model.
Some short protocols!

- **Problem 1:** Alice \( x_1, \ldots, x_n \); Bob \( y_1, \ldots, y_n \);
- **Want to know:** \( (x_1 - y_1) + \cdots + (x_n - y_n) \);
- **Solution:** Alice \( \rightarrow \) Bob: \( S \equiv x_1 + \cdots + x_n \)
  
  Bob \( \rightarrow \) Alice: \( S - (y_1 + \cdots + y_n) \);

- **Problem 2:** Alice \( x_1, \ldots, x_n \); Bob \( y_1, \ldots, y_n \);
- **Want to know:** \( (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 \);
- **Solution?**
  - Deterministically: Needs \( n \) bits of communication.
  - Randomized: Say Alice+Bob \( \leftarrow r_1, r_2, \ldots, r_n \in \{-1, +1\} \) random.
  
  Alice \( \rightarrow \) Bob: \( S_1 = x_1^2 + \cdots + x_n^2; \quad S_2 = r_1 x_1 + \cdots r_n x_n \)
  
  Bob \( \rightarrow \) Alice: \( T_1 = y + \cdots + y; \quad T_2 = r_1 y_1 + \cdots r_n y_n \)
  
  **Thm:** \( \text{Exp} \ [S_1 + T_1 - 2S_2 T_2] = (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 \)
Application to Buying Air Tickets

- If we can express every flight and every user’s preference as \( n \) numbers
  - (commonly done in Machine Learning)
  - Then \#bits communicated \( \approx 2 \). description of final itinerary.
  - Only two rounds of communication!

- Challenge: Express user preferences as numbers!
  - Not yet there ... but soon your cellphones will do it!
Interaction + Errors: Schulman

- Consider distributed update of shared document.
- What if there are errors in interaction?
  - Error must be detected immediately?
    - Or else all future communication wasted.
  - But too early detection might lead to false alarms!

Typical interaction:
Server → User: Current state of document
User → Server: Update
Interactive Coding Schemes

- If bits are flipped with probability $p$ what is the rate of communication?
- Limits still not precisely determined! But linear for $p \leq \frac{1}{8}$ (scales more like $1 - \sqrt{H(p)}$)
- Non-trivial mathematics! Some still not fully constructive
- ... surprising even given Shannon theory!
Sales Pitch + Intro

- Most of communication theory [a la Shannon, Hamming]:
  - Built around sender and receiver perfectly synchronized.
  - So large context (codes, protocols, priors) ignored.

- Most Human communication (also device-device)
  - ... does not assume perfect synchronization.
  - So context is relevant:
    - Qualitatively (receiver takes wrong action)
    - and Quantitatively
Aside: Contextual Proofs & Uncertainty

- Mathematical proofs assume large context.
  - “By some estimates a proof that 2+2=4 in ZFC would require about 20000 steps ... so we will use a huge set of axioms to shorten our proofs – namely, everything from high-school mathematics” [Lehman, Leighton, Meyer]

- Context shortens proofs. But context is uncertain!
  - What is “high school mathematics”
    - Is it a fixed set of axioms?
    - Or a set from which others can be derived?
      - Is the latter amenable to efficient reasoning?
      - What is efficiency with large context?
Sales Pitch + Intro

- Most of communication theory [a la Shannon, Hamming]:
  - Built around sender and receiver perfectly synchronized.
  - So large context (codes, protocols, priors) ignored.

- Most Human communication (also device-device)
  - ... does not assume perfect synchronization.
  - So context is relevant:
    - Qualitatively (receiver takes wrong action)
    - and Quantitatively (inputs are long!!)

- Theory? What are the problems?
  - Starting point = Shannon? Yao?
Communication Complexity

The model (with shared randomness)

\[ f : (x, y) \mapsto \Sigma \]

\[ R = $$ \]

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

\[ CC(f) = \# \text{ bits exchanged} \]

by best protocol

\[ f(x, y) \text{ w.p. } 2/3 \]
Aside: Easy CC Problems [Ghazi,Kamath,S’15]

Exist problems with large inputs and small communication?

- Equality testing:
  \[ EQ(x, y) = 1 \iff x = y; \quad O(1) \quad \text{ Protocol: } \]

- Hamming distance:
  \[ H_k(x, y) = 1 \iff \Delta(x, y) \leq k; \quad poly(k) \quad \text{ Protocol: } \]

- Small set intersection:
  \[ \cap_k (x, y) = 1 \iff \text{wt}(x), \text{wt}(y) \leq k; \quad O(k) \quad \text{ Protocol: } \]

Unstated philosophical contribution of CC a la Yao:
Communication with a **focus** ("only need to determine \( f(x, y) \")
**can be more effective** (shorter than \( |x|, H(x), H(y), I(x; y) \ldots \))
Many possibilities. Ongoing effort.

Alice+Bob may have estimates of $x$ and $y$

- More generally: $x, y$ close (in some sense).
- Knowledge of $f$ – function Bob wants to compute
  - may not be exactly known to Alice!
- Shared randomness
  - Alice + Bob may not have identical copies.
1. Compression

- **Classical compression**: Alice $\leftarrow P, m \sim P$; Bob $\leftarrow P$;
  - Alice $\rightarrow$ Bob: $y = E_P(m)$; Bob $\hat{m} = D_P(y) \triangleq m$;
  - [Shannon]: $\hat{m} = m$; w. $\mathbb{E}_{m \sim P}[|E_P(m)|] \leq H(P) + 1$  
    $H(P) \triangleq \mathbb{E}_{m \sim P}[-\log P(m)]$

- **Uncertain compression** [Juba, Kalai, Khanna, S.]
  - Alice $\leftarrow P, m \sim P$; Bob $\leftarrow Q$;
  - Alice $\rightarrow$ Bob: $y = E_P(m)$; Bob $\hat{m} = D_Q(y) \triangleq m$;
  - $P, Q \Delta$-close: $\forall m |\log P(m) - \log Q(m)| \leq \Delta$
  - Can we get $\mathbb{E}_{m \sim P}[|E_P(m)|] \leq O(H(P) + \Delta)$?
  - [JKKS] – Yes – with shared randomness.
  - [Haramaty+S.] – Deterministically $O(H(P) + \log \log |\Omega|)$
Deterministic Compression: Challenge

- Say Alice and Bob have rankings of $N$ players.
  - 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?

With Elad Haramaty: $\tilde{O}(\log^* n)$
Compression as a proxy for language

- Information theoretic study of language?
- Goal of language: **Effective means of expressing information/action.**
- Implicit objective of language: **Make frequent messages short. Compression!**
- **Frequency = Known globally? Learned locally?**
  - If latter – every one can’t possibly agree on it;
  - Yet need to agree on language (mostly)!
  - Similar to problem of Uncertain Compression.
- **Studied formally in**
  - [Ghazi, Haramaty, Kamath, S. ITCS 17]
2. Imperfectly Shared Randomness

- Recall: Communication becomes more effective with randomness.
  - Identity, Hamming Distance, Small Set Intersection, Inner Product.

- How does performance degrade if players only share correlated variables:
  - E.g. Alice $\leftarrow r$; Bob $\leftarrow s$. $(r, s) = (r_i, t_i)_i$ i.i.d.
    $r_i, s_i \in \{-1, 1\}$; $\mathbb{E}[r_i] = \mathbb{E}[s_i] = 0$; $\mathbb{E}[r_is_i] = \rho \in (0, 1)$;

- [CGMS ’16]:
  - Comm. With perfect randomness = $k$
    $\Rightarrow$ Comm. With imperfect randomness = $O_\rho(2^k)$
Imperfectly Shared Randomness

- **Easy (Complete) Problem:**
  - **Gap Inner Product:** $x, y \in \mathbb{R}^n$
  - $GIP_{c,s}(x, y) = 1$ if $\langle x, y \rangle \geq \epsilon \cdot |x|_2 \cdot |y|_2$; 
    $= 0$ if $\langle x, y \rangle \leq 0$
  - Decidable with $O(\rho \left( \frac{1}{\epsilon^2} \right))$ (o.w.) communication

- **Hard Problem:**
  - **Sparse Gap Inner Product:** $GIP$ on sparse $x$
    - $x \in \{0,1\}^n, y \in \{-1,1\}^n;$ $|x|_1 = 2\epsilon n$
  - Classical communication = $O \left( \log \frac{1}{\epsilon} \right)$ [uses sparsity]
  - No way to use sparsity with imperfect randomness.
3. Functional Uncertainty

- [Ghazi, Komargodski, Kothari, S. ‘16]
- Recall positive message of Yao’s model:
  - Communication can be brief, if Alice knows what function $f(x, y)$ Bob wants to compute.
- What if Alice only knows $f$ approximately?
  - Can communication still be short?
The Model

- Recall Distributional Complexity:
  - \( (x, y) \sim \mu; \) \( \text{error}_\mu(\Pi) \overset{\text{def}}{=} \Pr_{x,y \sim \mu} [f(x, y) \neq \Pi(x, y)] \)
  - Complexity: \( cc_{\mu, \epsilon}(f) \overset{\text{def}}{=} \min_{\Pi: \text{error}_\mu(\Pi) \leq \epsilon} \left\{ \max_{x,y} |\Pi(x, y)| \right\} \)

- Functional Uncertainty Model - I:
  - Adversary picks \( f, g \). Nature picks \( (x, y) \sim \mu \)
  - Alice \( \leftarrow (f, x) \); Bob \( \leftarrow (g, y) \); Compute \( g(x, y) \)
  - Promise: \( \delta_\mu(f, g) \overset{\text{def}}{=} \Pr_{\mu} [f(x, y) \neq g(x, y)] \leq \delta_0 \)
  - Goal: Compute (any) \( \Pi(x, y) \) with \( \delta_\mu(g, \Pi) \leq \epsilon_1 \)
    - (just want \( \epsilon_1 \to 0 \) as \( \delta_0 \to 0 \))
  - If \( (f, g) \) part of input; this is complexity of what?
Modelling Uncertainty

- Modelled by graph $\mathcal{G}$ of possible inputs
- Protocols know $\mathcal{G}$ but not $(f, g)$
- \[
    cc_{\mu, \epsilon}(\mathcal{G}) \overset{\text{def}}{=} \max_{(f, g) \in \mathcal{G}} \{cc_{\mu, \epsilon}(g)\}
\]
- \[
    \delta_{\mu}(\mathcal{G}) \overset{\text{def}}{=} \max_{(f, g) \in \mathcal{G}} \{\delta_{\mu}(f, g)\}
\]
- Uncertain error:
  \[
  \text{error}_{\mu, \mathcal{G}}(\Pi) \overset{\text{def}}{=} \max_{(f, g) \in \mathcal{G}} \{\Pr[\mu(g(x, y) \neq \Pi(f, g, x, y))]\}
  \]
- Uncertain complexity:
  \[
  Ucc_{\mu, \epsilon}(\mathcal{G}) \overset{\text{def}}{=} \min_{\Pi: \text{error}(\Pi) \leq \epsilon} \{\max_{f, g, x, y} \{||\Pi(f, g, x, y)||\}\}
  \]
- Compare $cc_{\mu, \epsilon_0}(\mathcal{G})$ vs. $Ucc_{\mu, \epsilon_1}(\mathcal{G})$
  want $\epsilon_1 \to 0$ as $\epsilon_0, \delta_{\mu}(\mathcal{G}) \to 0$
Main Results

- **Thm 1**: (-ve) \( \exists \mathcal{G}, \mu \) s.t. \( \delta_\mu(\mathcal{G}) = o(1); \text{cc}_{\mu, o(1)}(\mathcal{G}) = 1; \)

but \( U\text{cc}_{\mu, 1}(\mathcal{G}) = \Omega(\sqrt{n}) \); ( \( n = |x| = |y| \) )

- **Thm 2**: (+ve) \( \forall \mathcal{G}, \mu, \) product \( U\text{cc}_{\mu, \epsilon_1}(\mathcal{G}) = O\left(\text{cc}_{\mu, \epsilon_0}^{\text{oneway}}(\mathcal{G})\right) \)

where \( \epsilon_1 \to 0 \) as \( \epsilon_0, \delta_\mu(\mathcal{G}) \to 0 \)

- **Thm 2'**: (+ve) \( \forall \mathcal{G}, \mu, \)

\( U\text{cc}_{\mu, \epsilon_1}(\mathcal{G}) = O\left(\text{cc}_{\mu, \epsilon_0}^{\text{oneway}}(\mathcal{G}) \cdot (1 + I(x; y))\right) \)

where \( \epsilon_1 \to 0 \) as \( \epsilon_0, \delta_\mu(\mathcal{G}) \to 0 \)

and \( I(x; y) = \text{Mutual Information between } x; y \)

Protocols are not continuous wrt the function being computed.
Details of Negative Result

- $\mu: x \sim U(\{0,1\}^n); y = Noisy(x) ; \Pr[x_i \neq y_i] = 1/\sqrt{n}$;
- $G = \{((\oplus_S (x \oplus y), \oplus_T (x \oplus y))) | |S \oplus T| = o(\sqrt{n})\}$
  - $\oplus_S (z) = \oplus_{i \in S} z_i$
- Certain Comm: Alice $\rightarrow$ Bob: $\oplus_T (x)$
- $\delta_\mu (G) = \max_{S,T} \{\Pr[\oplus_S (x \oplus y) \neq \oplus_T (x \oplus y)]\}$
  
  $$= \max_{U: |U| = o(\sqrt{n})} \left\{ \Pr_{z \sim \text{Bernoulli}(\frac{1}{\sqrt{n}})^n} [\oplus_U (z) = 1] \right\} = o(1)$$
- Uncertain Lower bound:
  - Standard $cc_{\mu,\varepsilon} (F)$ where $F((S, x); (T, y)) = \oplus_T (x \oplus y)$;
  - Lower bound obtain by picking $(S, T)$ randomly:
    - $S$ uniform; $T$ noisy copy of $S$
Positive result (Thm. 2)

- Consider comm. Matrix Protocol for \( g \) partitions matrix in \( 2^k \) blocks.
- Bob wants to know which block?
- Common randomness:
  \[ y_1, \ldots, y_m \]
- Alice → Bob:
  \[ f(x, y_1) \ldots f(x, y_m) \]
- Bob (whp) recognizes block and uses it.
- \( m = O(k) \) suffices.
Analysis Details

1. W.p. $1 - \sqrt{\epsilon}$, $\exists j$ s.t. $\delta \left( g(x_j; \cdot), g(x; \cdot) \right) \leq \sqrt{\epsilon}$
   - Main idea: If $\Pi g(x) = \Pi g(x_j)$ then w.h.p. $\delta \left( g(x_j; \cdot), g(x; \cdot) \right) \leq \sqrt{\epsilon}$

2. If $j \in [K]$ s.t. $\delta \left( g(x_j; \cdot), g(x; \cdot) \right) \geq 2\sqrt{\epsilon}$ then $\Pr[j$ is selected$] = \exp(-m)$.
   - But Step 2. works only if $y_i \sim \mu_x$
Thm 2’: Main Idea

- Now can not sample \( y_1, \ldots, y_m \) independent of \( x \)
- Instead use [HJMR’07] to sample \( y_i \sim \mu_x \)
  - Each sample costs \( I(x; y) \)
- Analysis goes through ...
4. Contextual Proofs and Uncertainty?

- Scenario: Alice + Bob start with axioms $A$: subset of clauses on $X_1, \ldots, X_n$
- Alice wishes to prove $A \Rightarrow C$ for some clause $C$
- But proof $\Pi : A \Rightarrow C$ may be long ($\sim 2^{\sqrt{n}}$)
- Context to rescue: Maybe Alice + Bob share context $D \leftrightarrow A$; and contextual proof $\Pi' : D \Rightarrow C$ short ($\text{poly}(n)$)
- Uncertainty: Alice’s Context $D_A \neq D_B$ (Bob’s context)
  - Alice writes proof $\Pi' : D_A \Rightarrow C$
  - When can Bob verify $\Pi'$ given $D_B$?

- Scenario: Alice + Bob start with axioms $A$: subset of clauses on $X_1, \ldots, X_n$
- Alice wishes to prove $A \Rightarrow C$ for some clause $C$
  - Alice writes proof $\Pi': D_A \Rightarrow C$
  - When can Bob verify $\Pi'$ given $D_B$?
  - Surely if $D_A \subseteq D_B$
  - What if $D_A \setminus D_B = \{C'\}$ and $\Pi'': D_B \Rightarrow C'$ is one step long?
  - Can Bob still verify $\Pi': D_A \Rightarrow C$ in $\text{poly}(n)$ time?
  - Need feasible data structure that allows this!
    - None known to exist. Might be harder than Partial Match Retrieval ...
Summarizing

- Perturbing “common information” assumptions in Shannon/Yao theory, lead to many changes.
  - Some debilitating
  - Some not; but require careful protocol choice.
- In general: Communication protocols are not continuous functions of the common information.
- Uncertain model ($\mathcal{G}$) needs more exploration!
- Some open questions from our work:
  - Tighten the gap: $cc(f) \cdot I$ vs. $cc(f) + \sqrt{I}$
  - Multi-round setting? Two rounds?
  - What if randomness & function imperfectly shared? [Prelim. Results in [Ghazi+S’17]]
Contextual Communication & Uncertainty
Communication Complexity: Yao

The model \((\text{with shared randomness})\)

\[ x \quad f : (x, y) \mapsto \Sigma \]

\[ R = $$$ \]

\[ CC(f) = \# \text{ bits exchanged by best protocol} \]

\[ f(x, y) \quad \text{w.p. 2/3} \]
Boole’s “modest” ambition

“The design of the following treatise is to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolical language of a Calculus, and upon this foundation to establish the science of Logic and construct its method; to make that method itself the basis of a general method for the application of the mathematical doctrine of Probabilities; and, finally, to collect from the various elements of truth brought to view in the course of these inquiries some probable intimations concerning the nature and constitution of the human mind.” [G.Boole, “On the laws of thought ...” p.1]
Sales Pitch + Intro

- Most of communication theory [a la Shannon, Hamming]:
  - Built around sender and receiver perfectly synchronized.
  - So large context (codes, protocols, priors) ignored.

- Most Human communication (also device-device)
  - ... does not assume perfect synchronization.
  - So context is relevant:
    - Qualitatively (receiver takes wrong action)
    - and Quantitatively
Aside: Contextual Proofs & Uncertainty

- Mathematical proofs assume large context.
  - “By some estimates a proof that 2+2=4 in ZFC would require about 20000 steps ... so we will use a huge set of axioms to shorten our proofs – namely, everything from high-school mathematics” [Lehman, Leighton, Meyer]

- Context shortens proofs. But context is uncertain!
  - What is “high school mathematics”
    - Is it a fixed set of axioms?
    - Or a set from which others can be derived?
      - Is the latter amenable to efficient reasoning?
      - What is efficiency with large context?
**Sales Pitch + Intro**

- Most of communication theory [a la Shannon, Hamming]:
  - Built around sender and receiver perfectly synchronized.
  - So large context (codes, protocols, priors) ignored.

- Most Human communication (also device-device)
  - ... does not assume perfect synchronization.
  - So context is relevant:
    - Qualitatively (receiver takes wrong action)
    - and Quantitatively (inputs are long!!)

- Theory? What are the problems?
  - Starting point = Shannon? Yao?
The model (with shared randomness)

\[ f : (x, y) \mapsto \Sigma \]

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

\[ CC(f) = \# \text{ bits exchanged by best protocol} \]

\[ f(x, y) \text{ w.p. } 2/3 \]
Aside: Easy CC Problems [Ghazi, Kamath, S’15]

Exist Problems with large inputs and small communication?

- Equality testing:
  \[ EQ(x, y) = 1 \iff x = y; \quad \mathcal{C}(EQ) = O(1) \]

- Hamming distance:
  \[ H_k(x, y) = 1 \iff \Delta(x, y) \leq k; \quad \mathcal{C}(H_k) = O(k \log k) \] [Huang et al.]

- Small set intersection:
  \[ \cap_k(x, y) = 1 \iff \text{wt}(x), \text{wt}(y) \leq k \& \exists i: \forall s \cdot R \cdot x_i = y_i = 1; \quad \mathcal{C}(\cap_k) = O(k) \] [Håstad, Wigderson]

- Gap (Real) Inner Product:
  \[ x, y \in \mathbb{R}^n; \quad x^2, y^2 = 1; \quad \mathcal{G}(x, y) = 1 \iff x, y \geq c; \quad = 0 \iff x, y \leq s; \quad \mathcal{C}(\mathcal{G}) = O(1/c - s) \] [Alon, Matias, Szegedy]

Protocol:
Fix ECC \( E : \{0,1\}^n \rightarrow \{0,1\}^N \)

Use common randomness to hash \([n] \rightarrow \mathcal{P}(k)\)

\[ x = (x_1, \ldots, x_n), \quad y = (y_1, \ldots, y_n) \]

\[ \langle x, y \rangle \triangleq \sum_i x_i y_i \]

Unstated philosophical contribution of CC a la Yao:

Communication with a focus ("only need to determine \( f(x, y) \)) can be more effective (shorter than \(|x|, H(x), H(y), I(x; y) \ldots \))
Modelling Shared Context + Imperfection

- Many possibilities. Ongoing effort.
- Alice+Bob may have estimates of $x$ and $y$.
  - More generally: $x, y$ close (in some sense).
- Knowledge of $f$ – function Bob wants to compute
  - may not be exactly known to Alice!
- Shared randomness
  - Alice + Bob may not have identical copies.
1. Compression

- **Classical compression**: Alice $\leftarrow P, m \sim P$; Bob $\leftarrow P$;
  - Alice $\rightarrow$ Bob: $y = E_P(m)$; Bob $\hat{m} = D_P(y) \approx m$;
  - [Shannon]: $\hat{m} = m$; w. $\mathbb{E}_{m \sim P}[|E_P(m)|] \leq H(P) + 1$
    \[ H(P) \equiv \mathbb{E}_{m \sim P}[-\log P(m)] \]

- **Uncertain compression** [Juba,Kalai,Khanna,S.]
  - Alice $\leftarrow P, m \sim P$; Bob $\leftarrow Q$;
  - Alice $\rightarrow$ Bob: $y = E_P(m)$; Bob $\hat{m} = D_Q(y) \approx m$;
  - $P, Q \Delta$-close: $\forall m |\log P(m) - \log Q(m)| \leq \Delta$
  - Can we get $\mathbb{E}_{m \sim P}[|E_P(m)|] \leq O(H(P) + \Delta)$?
  - [JKKS] – Yes – with shared randomness. Universe
  - [Haramaty+S.] – Deterministically $O(H(P) + \log \log |\Omega|)$
Deterministic Compression: Challenge

- Say Alice and Bob have rankings of $N$ players.
  - Rankings = bijections $\pi, \sigma : [N] \to [N]$
  - $\pi(i)$ = 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?
    - With Elad Haramaty: $\tilde{O}(\log^* n)$
Compression as a proxy for language

- Information theoretic study of language?
- Goal of language: Effective means of expressing information/action.
- Implicit objective of language: Make frequent messages short. Compression!
- Frequency = Known globally? Learned locally?
  - If latter – every one can’t possibly agree on it;
  - Yet need to agree on language (mostly)!
  - Similar to problem of Uncertain Compression.
  - Studied formally in
    [Ghazi, Haramaty, Kamath, S. ITCS 17]
2. Imperfectly Shared Randomness

- Recall: Communication becomes more effective with randomness.
  - Identity, Hamming Distance, Small Set Intersection, Inner Product.
- How does performance degrade if players only share correlated variables:
  - E.g. Alice $\leftarrow r$; Bob $\leftarrow s$. $(r, s) = (r_i, t_i)_i$ i.i.d.
    - $r_i, s_i \in \{-1, 1\}$; $\mathbb{E}[r_i] = \mathbb{E}[s_i] = 0$; $\mathbb{E}[r_is_i] = \rho \in (0,1)$;
  - [CGMS ’16]:
    - Comm. With perfect randomness $= k$
    - $\Rightarrow$ Comm. With imperfect randomness $= O_\rho(2^k)$
Imperfectly Shared Randomness

**Easy (Complete) Problem:**
- **Gap Inner Product:** \( x, y \in \mathbb{R}^n \)
- \( GIP_{c,s}(x, y) = 1 \text{ if } \langle x, y \rangle \geq \epsilon \cdot |x|_2 \cdot |y|_2; \)
  \( = 0 \text{ if } \langle x, y \rangle \leq 0 \)
- Decidable with \( O_\rho \left( \frac{1}{\epsilon^2} \right) \) (o.w.) communication

**Hard Problem:**
- **Sparse Gap Inner Product:** \( GIP \) on sparse \( x \)
  - \( x \in \{0,1\}^n, y \in \{-1,1\}^n; \ |x|_1 = 2\epsilon n \)
- Classical communication = \( O \left( \log \frac{1}{\epsilon} \right) \) [uses sparsity]
- No way to use sparsity with imperfect randomness.
3. Functional Uncertainty

- [Ghazi, Komargodski, Kothari, S. ‘16]
- Recall positive message of Yao’s model:
  - Communication can be brief, if Alice knows what function $f(x, y)$ Bob wants to compute.
- What if Alice only knows $f$ approximately?
  - Can communication still be short?
The Model

- Recall Distributional Complexity:
  - $(x, y) \sim \mu$; \( \text{error}_\mu(\Pi) \stackrel{\text{def}}{=} \Pr_{x,y \sim \mu} [f(x, y) \neq \Pi(x, y)] \)
  - Complexity: \( cc_{\mu, \epsilon}(f) \stackrel{\text{def}}{=} \min_{\Pi : \text{error}_\mu(\Pi) \leq \epsilon} \{ \max_{x,y} \{|\Pi(x, y)|\} \} \)

- Functional Uncertainty Model - I:
  - Adversary picks \( f, g \). Nature picks \( (x, y) \sim \mu \)
  - Alice \( \leftarrow (f, x) \); Bob \( \leftarrow (g, y) \); Compute \( g(x, y) \)
  - Promise: \( \delta_\mu(f, g) \stackrel{\text{def}}{=} \Pr_{\mu} [f(x, y) \neq g(x, y)] \leq \delta_0 \)
  - Goal: Compute (any) \( \Pi(x, y) \) with \( \delta_\mu(g, \Pi) \leq \epsilon_1 \)
    - (just want \( \epsilon_1 \to 0 \) as \( \delta_0 \to 0 \))
  - If \( (f, g) \) part of input; this is complexity of what?
Modelling Uncertainty

- Modelled by graph $\mathcal{G}$ of possible inputs
- Protocols know $\mathcal{G}$ but not $(f, g)$

\[
cc_{\mu, \epsilon}(\mathcal{G}) \overset{\text{def}}{=} \max_{(f, g) \in \mathcal{G}} \{cc_{\mu, \epsilon}(g)\}
\]

\[
\delta_{\mu}(\mathcal{G}) \overset{\text{def}}{=} \max_{(f, g) \in \mathcal{G}} \{\delta_{\mu}(f, g)\}
\]

Uncertain error:

\[
error_{\mu, \mathcal{G}}(\Pi) \overset{\text{def}}{=} \max_{(f, g) \in \mathcal{G}} \{\Pr[\mu(g(x, y) \neq \Pi(f, g, x, y))]\}
\]

Uncertain complexity:

\[
Ucc_{\mu, \epsilon}(\mathcal{G}) \overset{\text{def}}{=} \min_{\Pi: error(\Pi) \leq \epsilon} \{\max_{f, g, x, y} \{|\Pi(f, g, x, y)|\}\}
\]

Compare $cc_{\mu, \epsilon_0}(\mathcal{G})$ vs. $Ucc_{\mu, \epsilon_1}(\mathcal{G})$

Want $\epsilon_1 \to 0$ as $\epsilon_0, \delta_{\mu}(\mathcal{G}) \to 0$
Main Results

- **Thm 1**: (-ve) \( \exists \mathcal{G}, \mu \) s.t. \( \delta_\mu (\mathcal{G}) = o(1); cc_{\mu,o(1)}(\mathcal{G}) = 1 \); but \( Ucc_{\mu,1}(\mathcal{G}) = \Omega(\sqrt{n}) \); ( \( n = |x| = |y| \))

- **Thm 2**: (+ve) \( \forall \mathcal{G} \), product \( \mu \)

\[
Ucc_{\mu,\varepsilon_1}(\mathcal{G}) = O \left( cc_{\mu,\varepsilon_0}^{oneway}(\mathcal{G}) \right)
\]

where \( \varepsilon_1 \to 0 \) as \( \varepsilon_0, \delta_\mu(\mathcal{G}) \to 0 \)

- **Thm 2’**: (+ve) \( \forall \mathcal{G}, \mu \)

\[
Ucc_{\mu,\varepsilon_1}^{oneway}(\mathcal{G}) = O \left( cc_{\mu,\varepsilon_0}^{oneway}(\mathcal{G}) \cdot (1 + I(x;y)) \right)
\]

where \( \varepsilon_1 \to 0 \) as \( \varepsilon_0, \delta_\mu(\mathcal{G}) \to 0 \)

and \( I(x;y) = \) Mutual Information between \( x; y \)

Protocols are not continuous wrt the function being computed
Details of Negative Result

- $\mu: x \sim U(\{0,1\}^n); y = Noisy(x) ; \Pr[x_i \neq y_i] = 1/\sqrt{n} ;$
- $G = \{(\bigoplus_S (x \bigoplus y), \bigoplus_T (x \bigoplus y))| |S \bigoplus T| = o(\sqrt{n})\}$
  - $\bigoplus_S (z) = \bigoplus_{i \in S} z_i$
- Certain Comm: Alice $\rightarrow$ Bob: $\bigoplus_T (x)$
- $\delta_\mu(G) = \max_{S,T} \{\Pr[x \bigoplus_S (x \bigoplus y) \neq \bigoplus_T (x \bigoplus y)]\}$
  $$= \max_{U: |U| = o(\sqrt{n})} \left\{ \Pr_{z \sim \text{Bernoulli}(1/\sqrt{n})^n} \left[ \bigoplus_U (z) = 1 \right] \right\} = o(1)$$
- Uncertain Lower bound:
  - Standard $cc_{\mu,\varepsilon}(F)$ where $F((S, x); (T, y)) = \bigoplus_T (x \bigoplus y)$;
  - Lower bound obtain by picking $(S, T)$ randomly:
    - $S$ uniform; $T$ noisy copy of $S$
Positive result (Thm. 2)

- Consider comm. Matrix Protocol for $g$ partitions matrix in $2^k$ blocks
- Bob wants to know which block?
- Common randomness: $y_1, \ldots, y_m$
- Alice $\rightarrow$ Bob: $f(x, y_1) \ldots f(x, y_m)$
- Bob (whp) recognizes block and uses it.
- $m = O(k)$ suffices.

“CC preserved under uncertainty for one-way communication if $x \perp y$”
Analysis Details

1. W.p. $1 - \sqrt{\epsilon}$, $\exists j$ s.t. $\delta \left( g(x_j, \cdot), g(x, \cdot) \right) \leq \sqrt{\epsilon}$
   
   Main idea: If $\Pi g(x) = \Pi g(x_j)$ then w.h.p.
   
   $\delta \left( g(x_j, \cdot), g(x, \cdot) \right) \leq \sqrt{\epsilon}$

2. If $j \in [K]$ s.t. $\delta \left( g(x_j, \cdot), g(x, \cdot) \right) \geq 2\sqrt{\epsilon}$ then
   
   $\Pr[j \text{ is selected}] = \exp(-m)$.

   But Step 2. works only if $y_i \sim \mu_x$
Thm 2’: Main Idea

- Now cannot sample \(y_1, \ldots, y_m\) independent of \(x\)
- Instead use [HJMR’07] to sample \(y_i \sim \mu_x\)
  - Each sample costs \(I(x; y)\)

- Analysis goes through …
4. Contextual Proofs and Uncertainty?

- Scenario: Alice + Bob start with axioms $A$: subset of clauses on $X_1, \ldots, X_n$
- Alice wishes to prove $A \Rightarrow C$ for some clause $C$
- But proof $\forall: A \Rightarrow C$ may be long ($\sim 2\sqrt{n}$)
- Context to rescue: Maybe Alice + Bob share context $D \iff A$; and contextual proof $\forall': D \Rightarrow C$ short ($\text{poly}(n)$)
- Uncertainty: Alice’s Context $D_A \neq D_B$ (Bob’s context)
  - Alice writes proof $\forall': D_A \Rightarrow C$
  - When can Bob verify $\forall'$ given $D_B$?

- Scenario: Alice + Bob start with axioms $A$: subset of clauses on $X_1, \ldots, X_n$
- Alice wishes to prove $A \Rightarrow C$ for some clause $C$
  - Alice writes proof $\Pi': D_A \Rightarrow C$
  - When can Bob verify $\Pi'$ given $D_B$?
    - Surely if $D_A \subseteq D_B$
    - What if $D_A \setminus D_B = \{C'\}$ and $\Pi'': D_B \Rightarrow C'$ is one step long?
    - Can Bob still verify $\Pi': D_A \Rightarrow C$ in $\text{poly}(n)$ time?
  - Need feasible data structure that allows this!
    - None known to exist. Might be harder than Partial Match Retrieval ...
Summarizing

- Perturbing “common information” assumptions in Shannon/Yao theory, lead to many changes.
  - Some debilitating
  - Some not; but require careful protocol choice.
- In general: Communication protocols are not continuous functions of the common information.
- Uncertain model ($\mathcal{G}$) needs more exploration!
- Some open questions from our work:
  - Tighten the gap: $cc(f) \cdot I$ vs. $cc(f) + \sqrt{I}$
  - Multi-round setting? Two rounds?
  - What if randomness & function imprecfectly shared? [Prelim. Results in [Ghazi+S’17]]
Thank You!