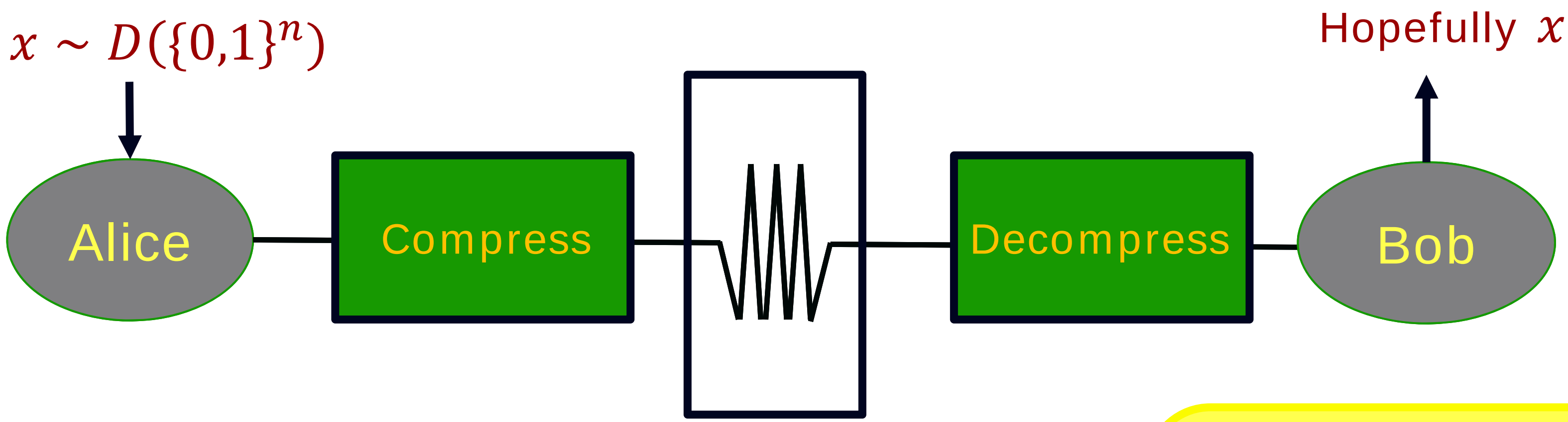
# Imperfectly Shared Randomness in Communication

**Madhu Sudan**
Microsoft Research

Joint work with Clément Canonne (Columbia), Venkatesan Guruswami (CMU) and Raghu Meka (UCLA).

# Communication (Complexity)

- Recall Shannon (Noiseless setting)




- What will Bob do with $x$?
  - Often knowledge of $x$ is overkill.
  - [Yao]'s model:
    - Bob has private information $y$.
    - Wants to know $f(x,y) \in \{0,1\}$.
    - Can we get away with much less communication?

In general, model allows interaction. For this talk, only one way comm.

# Example:

- Parity:
  - $x = x_1 x_2 \dots x_n; y = y_1 y_2 \dots y_n$;
  - $f(x, y) = \sum_i (x_i + y_i) \pmod 2 \triangleq \bigoplus_i (x_i \oplus y_i)$
- Solution:
  - Alice sends $a = \bigoplus_i x_i$ to Bob.
  - Bob computes $b = \bigoplus_i y_i$. Outputs $a \oplus b$.
- 1 bit of communication!
- (No distributional assumption on $x$!)

# Randomness in Communication

- As in many aspects of CS, randomness often helps find (more efficient) solutions.
- Two “Probabilistic Communication” Models:
  - Private randomness:
    - Alice tosses random coins and uses that to determine what to send to Bob.
  - Shared randomness:
    - Alice and Bob share random string $r \in \{0,1\}^*$
    - Alice’s message depends on $r$
    - Bob’s use of message depends on $r$.
- Det. CC $\geq$ Private. CC $\geq$ Shared. CC

# Example: Equality Testing

- $f(x,y) = 1$ if $x = y$ and $0$ o.w.
  - Deterministically: Communicate $\Omega(n)$ bits
  - With private randomness:
    - Alice encodes $x \mapsto E(x)$; $(E: \{0,1\}^n \to \{0,1\}^N)$
    - Picks $i \leftarrow_U [N]$; sends $(i, E(x)_i)$ to Bob.
    - Bob receives $(i, b)$ and outputs 1 if $E(y)_i = b$
    - Priv. CC $= O(\log n)$ bits
  - With shared randomness:
    - Alice and Bob share $i$.
    - Alice sends $E(x)_i$.
    - Shared CC $= O(1)$ bits.

# This talk: Imperfect Sharing

- Generic motivation:
  - Where does the shared randomness come from?
    - Nature/Collective experience $\Rightarrow$ Noisy
  - Do parties have to agree on their shares perfectly?
    - Can they get away with imperfection?
    - Is their a price?

# Model: Imperfectly Shared Randomness

- Alice $\leftarrow r$; and Bob $\leftarrow s$ where $(r, s) =$ i.i.d. sequence of correlated pairs $(r_i, s_i)_i$; $r_i, s_i \in \{-1, +1\}; \mathbb{E}[r_i] = \mathbb{E}[s_i] = 0; \mathbb{E}[r_i s_i] = \rho \geq 0$ .
- Notation:
  - $isr_\rho(f)$ = cc of $f$ with $\rho$-correlated bits.
  - $psr(f)$: Perfectly Shared Randomness cc.
  - $priv(f)$: cc with PRIVate randomness
- Starting point: for Boolean functions $f$
  - $psr(f) \leq isr_\rho(f) \leq priv(f) \leq psr(f) + \log n$
  - What if $psr(f) \ll \log n$? E.g. $psr(f) = O(1)$

# Results

- Model first studied by [Bavarian et al.'14] ("Independently and earlier").
  - They show $isr(\text{Equality}) = O(1)$

- Our Results:
  - Generally: $psr(f) \leq k \Rightarrow isr(f) \leq 2^k$
  - Converse: $\exists f$ with $psr(f) \leq k \,\&\, isr(f) \geq 2^k$

## Equality Testing (our proof)

$$X = (X_1, \dots, X_N)$$
$$Y = (Y_1, \dots, Y_N)$$
$$\langle X, Y \rangle \triangleq \sum_i X_i Y_i$$

- Key idea: Think inner products.
  - Encode $x \mapsto X = E(x); y \mapsto Y = E(y); X, Y \in \{-1, +1\}^N$
    - $x = y \Rightarrow \langle X, Y \rangle = N$
    - $x \neq y \Rightarrow \langle X, Y \rangle \leq N/2$
- Estimating inner products:
  - Using ideas from low-distortion embeddings …
  - Alice: Picks Gaussian $G \in \mathbb{R}^N$, sends $\langle G, X \rangle$
  - Bob: has $G' \sim_\rho G$; compares $\langle G, X \rangle$ with $\langle G', Y \rangle$
  - (mod details): $O_\rho(1)$ bits suffice if $G \approx_\rho G'$
  - [Bavarian et al.] Alternate protocol.

# General Communication

- Idea: All communication $\leq$ Inner Products
  - For each random string $R$
    - Alice's message = $i_R \in [2^k]$
    - Bob's output = $f_R(i_R)$ where $f_R : [2^k] \rightarrow \{0,1\}$
    - W.p. $\geq \frac{2}{3}$ over $R$, $f_R(i_R)$ is the right answer.

# General Communication

- For each random string $R$
  - Alice's message = $i_R \in [2^k]$
  - Bob's output = $f_R(i_R)$ where $f_R: [2^k] \to \{0,1\}$
  - W.p. $\geq \frac{2}{3}$, $f_R(i_R)$ is the right answer.
- Vector representation:
  - $i_R \mapsto x_R \in \{0,1\}^{2^k}$ (unit coordinate vector)
  - $f_R \mapsto y_R \in \{0,1\}^{2^k}$ (truth table of $f_R$).
  - $f_R(i_R) = \langle x_R, y_R \rangle$; Acc. Prob. $\propto \langle X, Y \rangle; X = (x_R)_R; Y = (y_R)_R$
  - Gaussian protocol estimates inner products of unit vectors to within $\pm\epsilon$ with $O\left(\frac{1}{\epsilon^2}\right)$ communication.

# Main Technical Result: Matching lower bound

- There exists (promise) problem $f$ s.t.
  - $psr(f) \leq k$
  - $isr_\rho(f) \geq \exp(k)$

- The Problem:
  - Gap Sparse Inner Product (G-Sparse-IP).
  - Alice gets sparse $x \in \{0,1\}^n$; $\mathrm{wt}(x) \approx 2^{-k} \cdot n$
  - Bob gets $y \in \{0,1\}^n$
  - Promise: $\langle x, y \rangle \geq (.9)2^{-k} \cdot n$ or $\langle x, y \rangle \leq (.6)2^{-k} \cdot n$.
  - Decide which.

# $psr$ Protocol for G-Sparse-IP

- Idea: $x_i \neq 0 \ \Rightarrow \ \ y_i$ correlated with answer.
- Use (perfectly) shared randomness to find random index $i$ s.t. $x_i \neq 0$ .
- Shared randomness: $i_1, i_2, i_3, \ldots$ uniform over $[n]$
- Alice → Bob: smallest index $j$ s.t. $x_{i_j} \neq 0$.
- Bob: Accept if $y_{i_j} = 1$
- Expect $j \approx 2^k; psr \leq k$.

# ISR lower bounds

- Challenge: Usual CC lower bounds give a distribution and prove lower bound against it.
- G-Sparse-IP has a low-complexity protocol for every input, with shared randomness.
- Thus for every distribution, there exists a deterministic low-complexity protocol!
- So usual method can't work ...

- Need to fix strategy first and then "tailor-make" a hard distribution for the strategy ...

# ISR lower bound for GSIP: Overview

- Strategies: Alice $f_r(x) \in [\ell]$; Bob $g_s(y) \in \{0,1\}^\ell$;
- Two possibilities:
  - Case 1: Alice's strategy and Bob's strategy have common highly "influential coordinate":
    - ($i$ s.t. flipping $x_i$ changes Alice's message etc.)
    - Leads to protocol for "agreement distillation" [We prove this is impossible.]
  - Case 2: Strategies have no common influential variable:
    - Invariance Principle ⇒ Solves some Gaussian problem
    - High complexity lower bound for Gaussian problem. (Details shortly)

## Case 1: Agreement Distillation

- Problem: Charlie $\leftarrow r$; Dana $\leftarrow s$; $(r, s)$ $\rho$-correlated
- Goal: Charlie outputs $u$; Dana outputs $v$;

$$H_\infty(u), H_\infty(v) \geq t; \qquad \Pr[u = v] \geq \gamma$$

- Lemma: With zero communication $\gamma = 2^{-\Omega(t)}$;
- Proof: "Small-set expansion of noisy hypercube"
  - Well-known by now … application of Bonami's lemma.
  - See, e.g., [Analysis of Boolean functions, O'Donnell]

- Corollary: For $c$ bits of communication,

$$c \geq \epsilon \cdot t + \log \gamma$$

# Completing Case 1

- Bad $\triangleq$ $\{i \mid \Pr_r[\mathrm{Inf}_i(f_r) \geq \text{high}] \geq \text{large}\}$
  $\cup \{i \mid \Pr_s[\mathrm{Inf}_i(g_s) \geq \text{high}] \geq \text{large}\}$
- Fact: (for our defn of influence) any function has bounded number of high influence variables.
- (By Fact + Markov) Can assume $|\mathrm{Bad}| \leq \epsilon \cdot n$.
- Distributions on Yes and No instances:
  - No: $x$ random sparse $\in \{0,1\}^n$ ; $y \leftarrow_U \{0,1\}^n$
  - Yes: Same as No on Bad coordinates.
    - On rest, $y_i$ is more likely to be 1 if $x_i = 1$.

# Completing Case 1 (contd.)

- Agreement strategy for Charlie + Dana:
  - Charlie: $i \in [n] \setminus \mathrm{Bad}$ s.t. $\mathrm{Inf}_i(f_r)$ high.
  - Dana: $j \in [n] \setminus \mathrm{Bad}$ s.t. $\mathrm{Inf}_j(g_s)$ high.
- Analysis:
  - $H_\infty(i), H_\infty(j)$ large since $i, j \notin \mathrm{Bad}$.
  - $i = j$?: Case 1 assumption.

- Combined with lower bound for agreement distillation, implies Case 1 can't occur

# Case 2: No common influential variable

- Key Lemma: Fix $r, s$; let $f = f_r$ and $g = g_s$.
  If $\ell$ small ($\approx 2^{2^k}$) and $f, g$ distinguish Yes/No then $f, g$ have common influential variable.
- Idea: Use "Invariance Principle":
  - Remarkable theorem: Mossel, O'Donnell, Oleskiewicz; Mossel++;
  - Informal form: f,g low-degree polynomials with no common influential variable $\Rightarrow$ $\mathrm{Exp}_{x,y}[f(x)g(y)] \approx \mathrm{Exp}_{X,Y}[f(X)g(Y)]$
    - where $x, y$ Boolean $n$-wise product dist.
    - and $X, Y$ Gaussian $n$-wise product dist.

## The Gaussian-IP Problem

- Suppose we can get the "perfect" invariance theorem for us …

- Would transform:

  Sol'n for G-Sparse-IP → Sol'n for G-Gaussian-IP
  - Alice, Bob get Gaussian unit vectors $X, Y \in \mathbb{R}^n$
  - Yes: $\langle X, Y \rangle \geq 2^{-k}$; No: $\langle X, Y \rangle \leq 0$

- Theorem: Non-sparse $X \Rightarrow CC \geq 2^k$ bits
  - Formally [Bar Yossef et al.]: Can reduce "indexing" to G-Gaussian-IP.

# Invariance Principle + Challenges

- Informal Invariance Principle: $f, g$ low-degree polynomials with no common influential variable $\Rightarrow \mathrm{Exp}_{x,y}[f(x)g(y)] \approx \mathrm{Exp}_{X,Y}[f(X)g(Y)]$
  - where $x, y$ Boolean $n$-wise product dist.
  - and $X, Y$ Gaussian $n$-wise product dist
- Challenges [+ Solutions]:
  - Our functions not low-degree [Smoothening]
  - Our functions not real-valued
    - $g: \{0,1\}^n \rightarrow \{0,1\}^\ell$: [Truncate range to $[0,1]^\ell$]
    - $f: \{0,1\}^n \rightarrow [\ell]$: [???, [work with $\Delta(\ell)$]]

# Invariance Principle + Challenges

- Informal Invariance Principle: $f, g$ low-degree polynomials with no common influential variable
  $\Rightarrow \mathrm{Exp}_{x,y}[f(x)g(y)] \approx \mathrm{Exp}_{X,Y}[\boldsymbol{f}(X)\boldsymbol{g}(Y)]$ (caveat $f \approx \boldsymbol{f}; g \approx \boldsymbol{g}$)
- Challenges
  - Our functions not low-degree [Smoothening]
  - Our functions not real-valued [Truncate]
  - Quantity of interest is not $f(x) \cdot g(y)$ ...
    - [Can express quantity of interest as inner product. ]
  - ... (lots of grunge work ...)
- Get a relevant invariance principle (next)

# Invariance Principle for CC

- Thm: For every convex $K_1, K_2 \subseteq [-1,1]^\ell$
  $\exists$ transformations $T_1, T_2$ s.t.
  if $f: \{0,1\}^n \to K_1$ and $g: \{0,1\}^n \to K_2$
  have no common influential variable, then
  $F = T_1 f: \mathbb{R}^n \to K_1$ and $G = T_2 g: \mathbb{R}^n \to K_2$ satisfy
  $\mathrm{Exp}_{x,y}[\langle f(x), g(y)\rangle] \approx \mathrm{Exp}_{X,Y}[\langle F(X), G(Y)\rangle]$

  - Main differences: $f, g$ vector-valued.
  - Functions are transformed: $f \mapsto F; g \mapsto G$
  - Range preserved exactly ($K_1 = \Delta(\ell); K_2 = [0,1]^\ell$)!
    - So $F, G$ are still communication strategies!

## Summarizing

- $k$ bits of comm. with perfect sharing
  $\rightarrow 2^k$ bits with imperfect sharing.
- This is tight (for one-way communication)
  - Invariance principle for communication
  - Agreement distillation
  - Low-influence strategies

## Conclusions

- Imperfect agreement of context important.
  - Dealing with new layer of uncertainty.
  - Notion of scale (context LARGE)
- Many open directions+questions:
  - Imperfectly shared randomness:
    - One-sided error?
    - Does interaction ever help?
    - How much randomness?
    - More general forms of correlation?

# Thank You!