Lower bounds on information complexity via zero-communication protocols

Iordanis Kerenidis

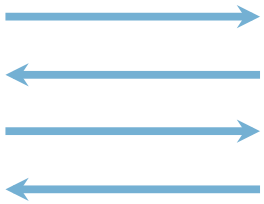
CNRS, LIAFA-Univ Paris Diderot 7

Joint work: Sophie Laplante, Virginie Lerays, Jérémie Roland, David Xiao

To appear: FOCS 2012

Communication complexity

$x \in \{0,1\}^n$



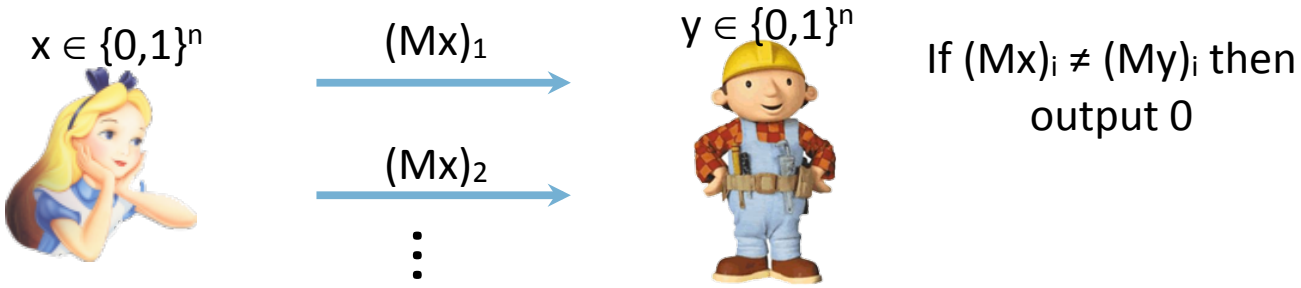
$y \in \{0,1\}^n$

- Alice and Bob want to compute f(x, y)
- Minimal communication needed?
  - Trivial protocol with communication n
- Initiated by [Yao '79]
- Wide application to circuit complexity, VLSI, streaming, distributed computing, data structures, etc.

Measuring complexity of communication

- Basic measure: number of bits transmitted
- But bits may be "useless"
- Example: EQUALITY

M = random invertible n x n matrix



- Many bits communicated, Ω(n) (on avg)
- BUT, Bob learns O(1) bits about x
- Question: how to measure amount of information transmitted?

Outline

1. Introducing Information Complexity

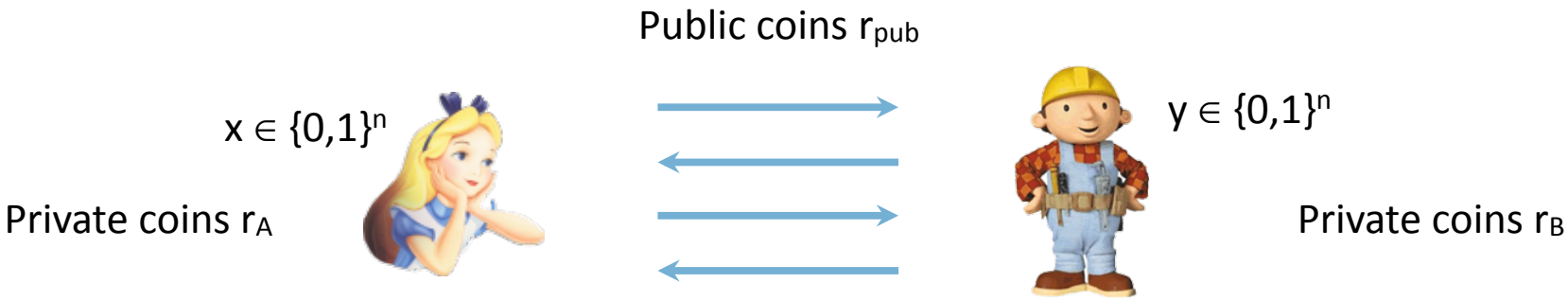
2. Relating Information to Communication Complexity

3. Zero-communication protocols and Information Complexity

Entropy

- Entropy: $H(X) \stackrel{\text{def}}{=} \sum_{x \in supp(X)} Pr[X = x] \log (1/Pr[X = x])$
  - "Uncertainty about X"
- Conditional entropy: $H(X \mid Y) \stackrel{\text{def}}{=} \sum_{y \in supp(Y)} Pr[Y = y] H(X \mid Y = y)$
  - "Uncertainty about X knowing Y"
- Mutual information: $\mathbf{I}(X ; Y) \stackrel{\text{def}}{=} H(X) - H(X \mid Y)$
  - "How much information does Y reveal about X"
- Conditional Mutual Information: $\mathbf{I}(X ; Y \mid Z) \stackrel{\text{def}}{=} H(X \mid Z) - H(X \mid YZ)$
  - "Already knowing Z, how much additional information does Y reveal about X"

Information complexity



- Fix protocol π, a distribution μ over inputs
- (X, Y, Π) : sample $(X, Y) \leftarrow \mu$, let Π be transcript of $\pi(X, Y)$

- $CC(\pi) = \max_{x,\ y,\ \text{random coins}} |\ \pi(x, y)\ |$
- $IC_\mu(\pi) = \mathbf{I}(X\ ;\ \Pi\ |\ Y) + \mathbf{I}(Y\ ;\ \Pi\ |\ X)$

"Information Alice learns about Bob's input from transcript + vice versa"

- π is (μ,ε)-good for f if $\Pr_{\pi,(X,Y)\leftarrow\mu}[\pi(X,Y) = f(X,Y)] \geq 1 - \varepsilon$
- $R_{\mu,\varepsilon}(f) = \inf_{\pi\ (\mu,\varepsilon)\text{-good for f}} CC(\pi)$
- $IC_{\mu,\varepsilon}(f) = \inf_{\pi\ (\mu,\varepsilon)\text{-good for f}} IC_\mu(\pi)$

History of IC

- Introduced in context of privacy by [BarYehuda-Chor-Kushilevitz-Orlitsky'92, Klauck'02]
- Introduced in context of communication complexity by [Chakrabarti-Shi-Wirth-Yao'01]
- Lower bound for disjointness [BarYossef-Jayram-Kumar-Sivakumar'04], Tribes [Jayram-Kumar-Sivakumar'03]
- Used to study direct sum questions [Chakrabarti-Shi-Wirth-Yao'01, Jain-Radhakrishnan-Sen'03, JRS'05, Harsha-Jain-McAllester-Radhakrishnan'07, Barak-Braverman-Chen-Rao'10, Braverman-Rao'11, Braverman'12]

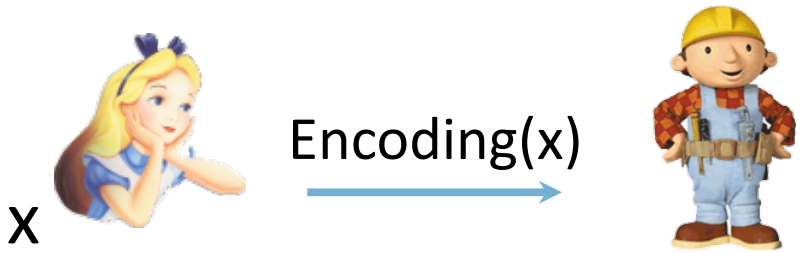
Plan

1. Introducing Information Complexity
2. Relating Information to Communication Complexity
3. Zero-communication protocols and Information Complexity

Compression

- Source Coding Theorem [Shannon'49]:
  Given a source X with entropy H(X), can encode X using H(X)+1 bits on avg
  - Huffman coding.
  - Asymptotically, can use H(X) bits
  - Non-interactive

- One-way communication: Information C = Communication C



What about interaction?

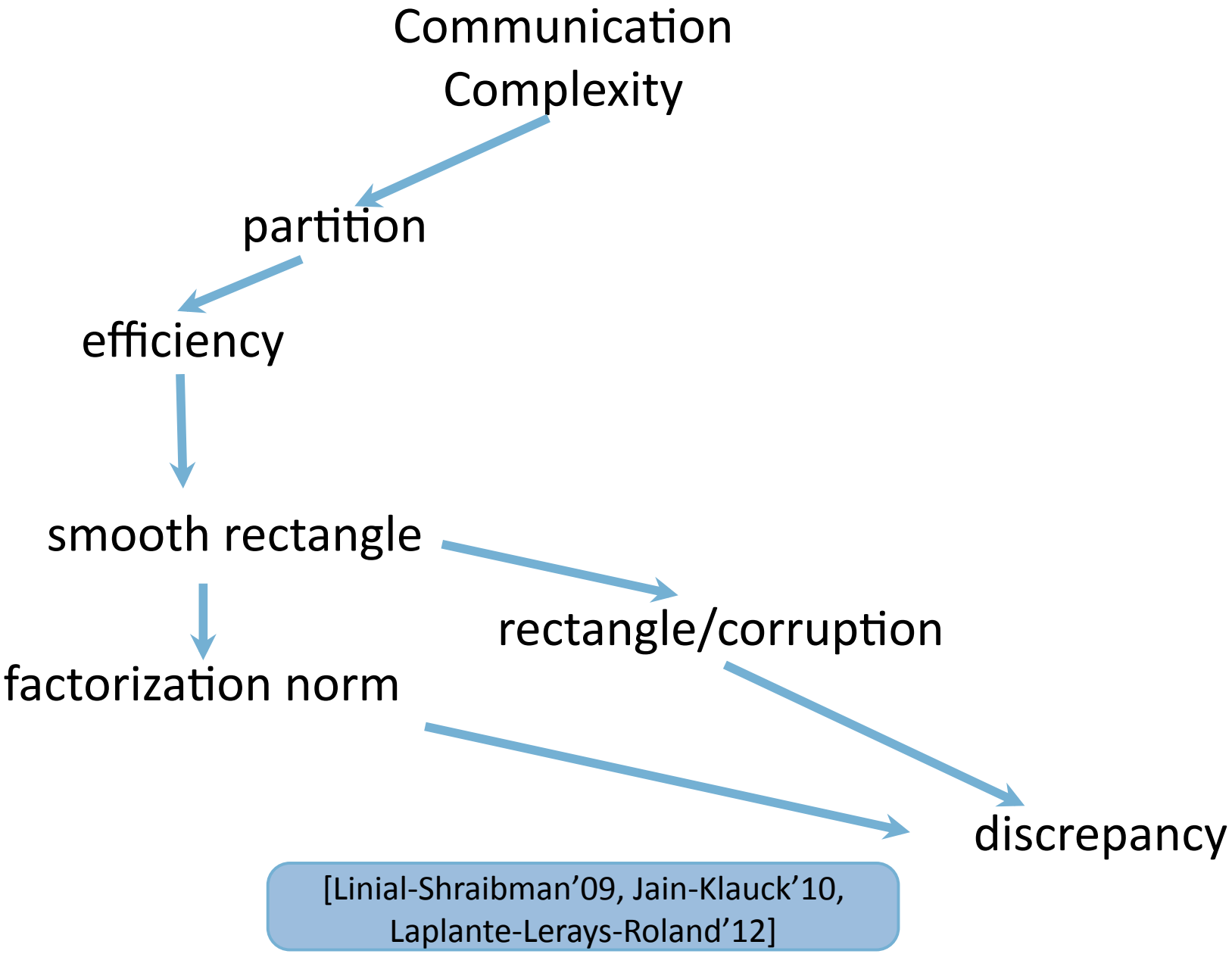
Information vs. Communication Complexity

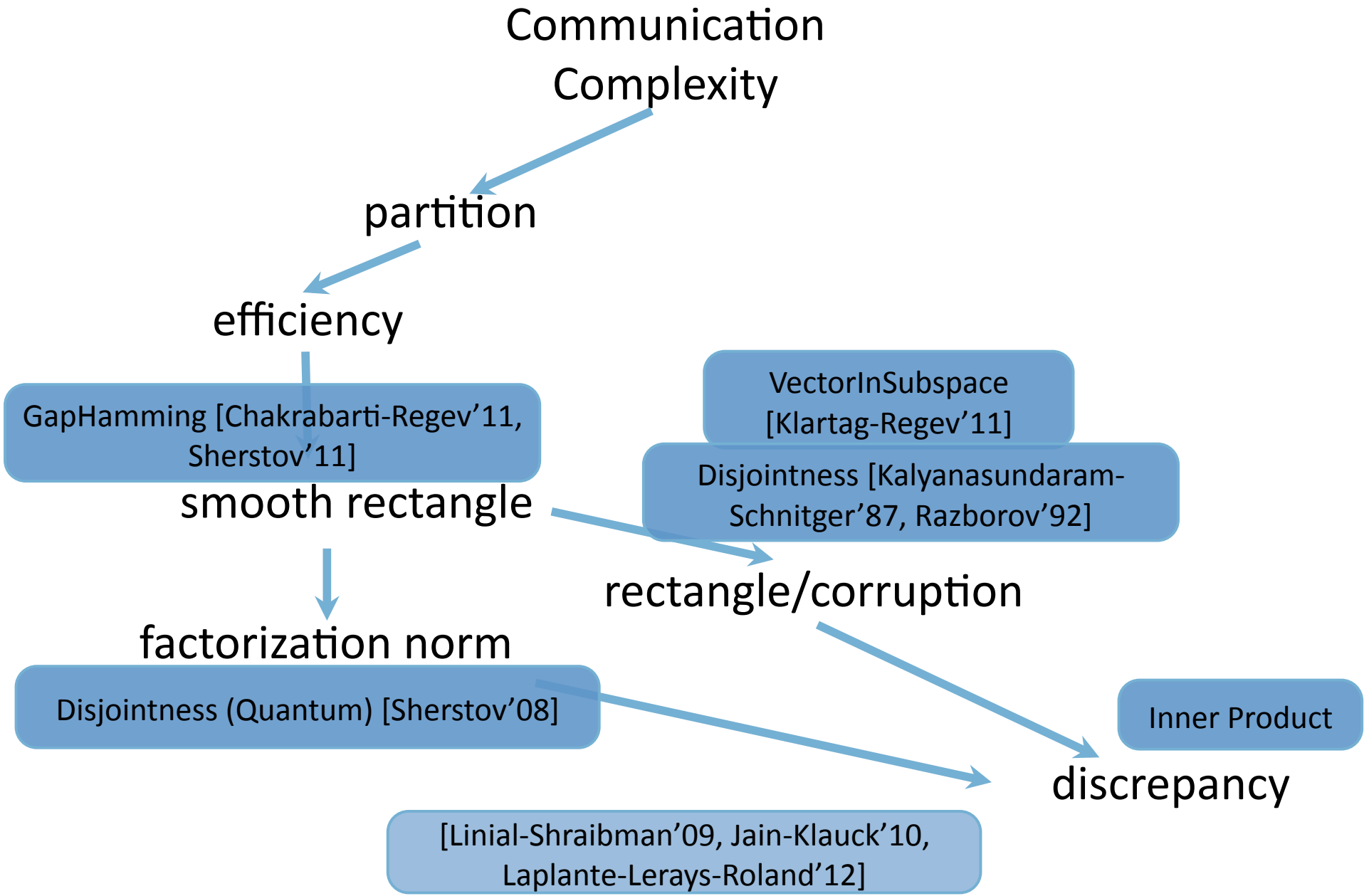
- How close are IC and CC? Does it hold that IC = CC?
- Conjecture 1: for any ε, μ, f, it holds that $R_{\mu,\varepsilon}(f) = \boldsymbol{O}(IC_{\mu,\varepsilon}(f))$

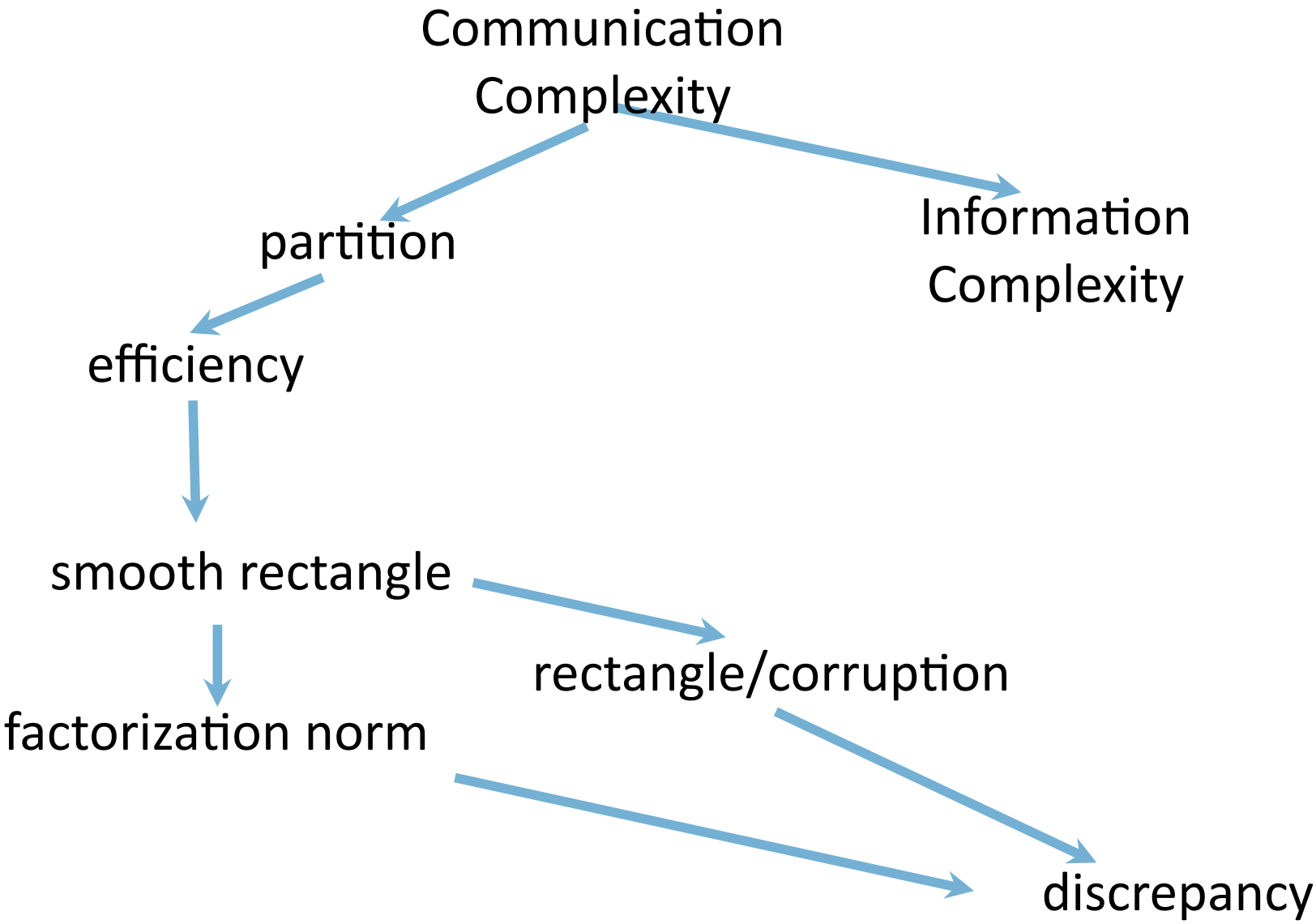
- Main application: assuming Conjecture 1, can prove direct sum for CC
  - We know that $R_{\mu^k,\varepsilon}(f^k) \geq IC_{\mu^k,\varepsilon}(f^k)$ and $IC_{\mu^k,\varepsilon}(f^k) = k * IC_{\mu,\varepsilon}(f)$ [Braverman11]
  - Therefore $R_{\mu^k,\varepsilon}(f^k) \geq k * IC_{\mu,\varepsilon}(f) \geq \Omega(k * R_{\mu,\varepsilon}(f))$

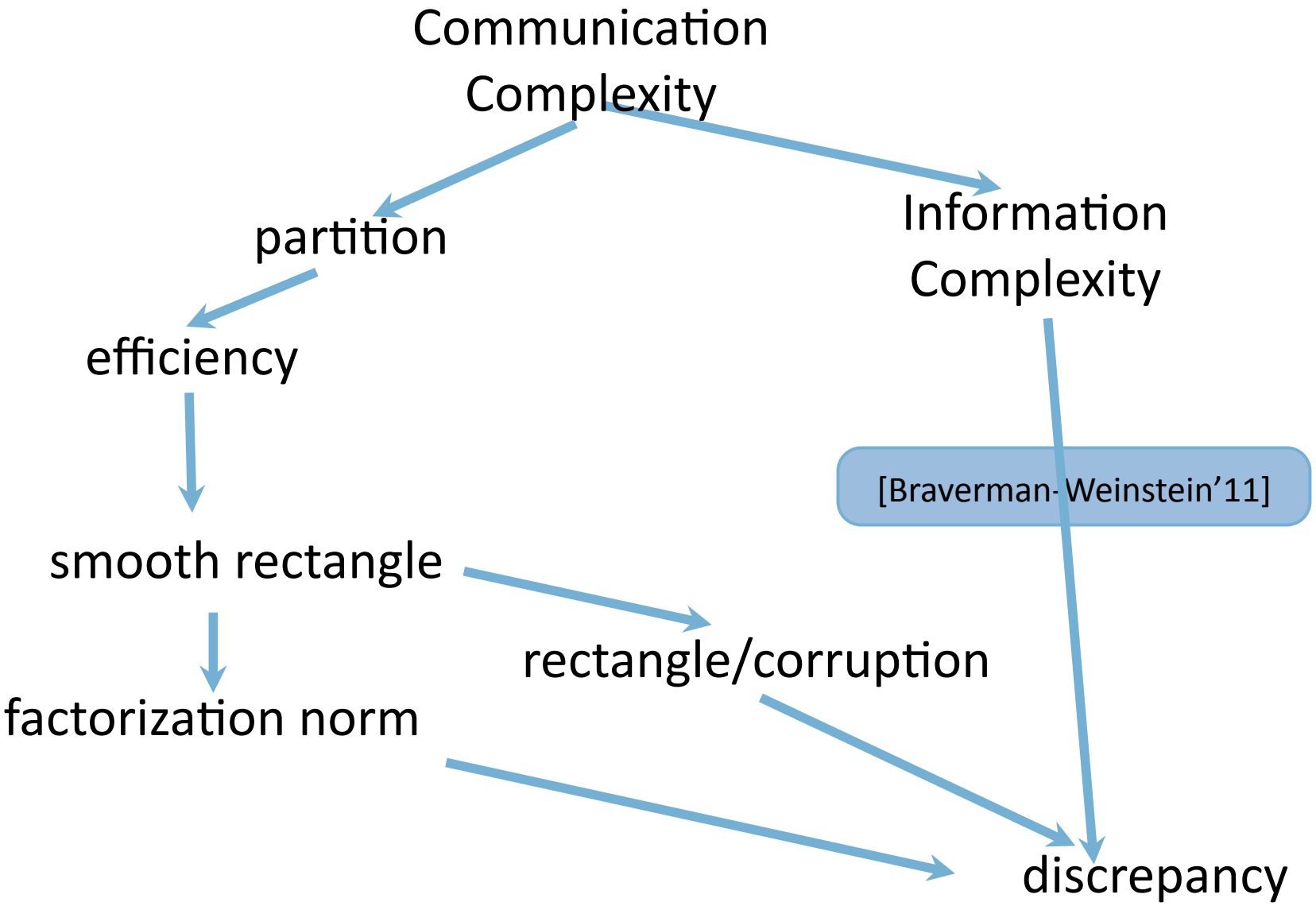
- Theorem [Braverman-Rao'11]: $IC(f) = \lim_{k\to\infty} R(f^k) / k$

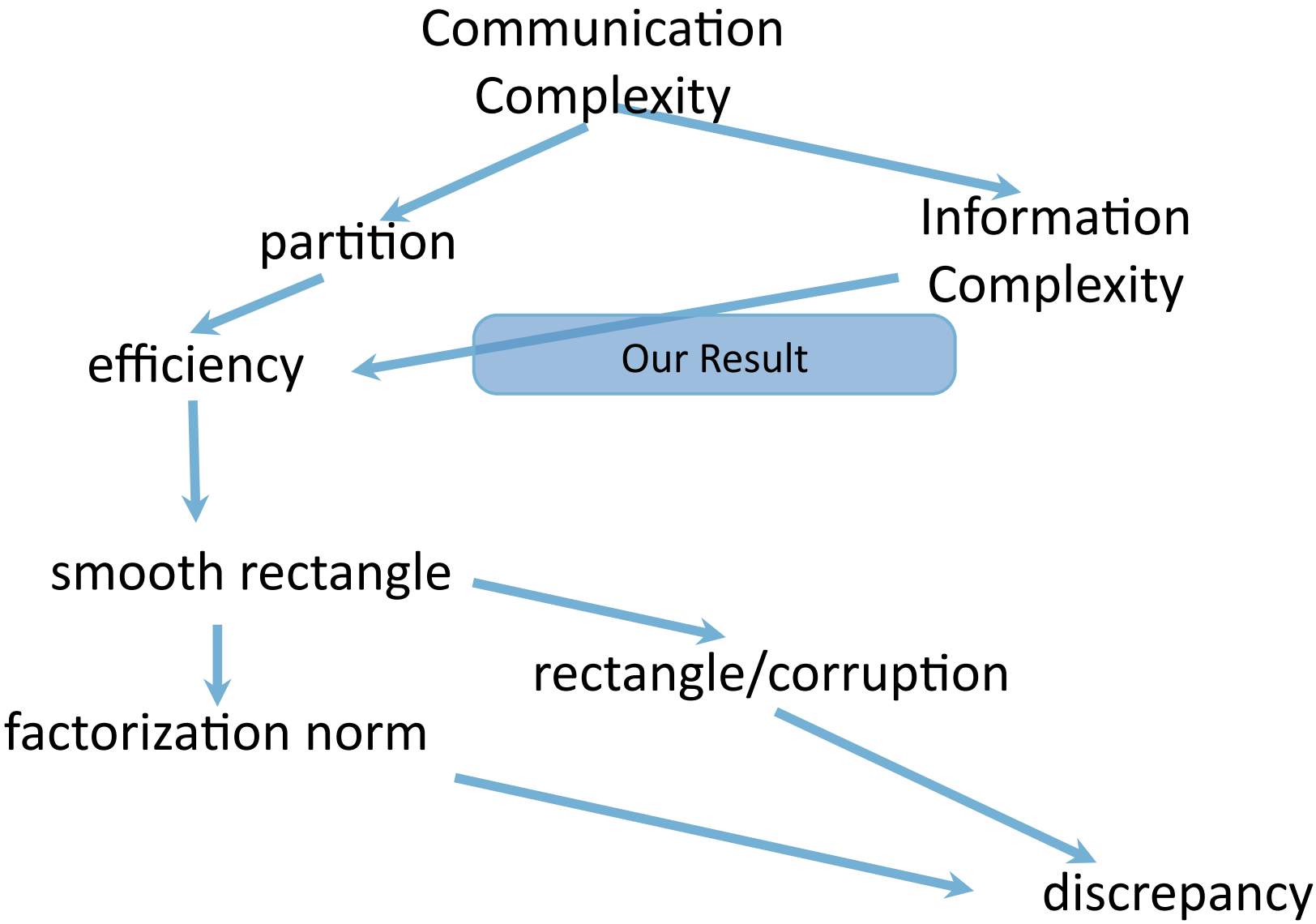
- Theorem [Braverman'12]: for any π, exists τ with $CC(\tau) \leq \boldsymbol{2}^{\boldsymbol{O}(IC(\pi))}$


Communication
Complexity
partition
efficiency
smooth rectangle
factorization norm
rectangle/corruption
discrepancy
[Linial-Shraibman'09, Jain-Klauck'10,
Laplante-Lerays-Roland'12]


Communication Complexity
partition
efficiency
GapHamming [Chakrabarti-Regev'11, Sherstov'11]
smooth rectangle
VectorInSubspace [Klartag-Regev'11]
Disjointness [Kalyanasundaram-Schnitger'87, Razborov'92]
rectangle/corruption
factorization norm
Disjointness (Quantum) [Sherstov'08]
Inner Product
discrepancy
[Linial-Shraibman'09, Jain-Klauck'10, Laplante-Lerays-Roland'12]


Communication
Complexity
partition
Information
Complexity
efficiency
smooth rectangle
rectangle/corruption
factorization norm
discrepancy


Communication Complexity
partition
efficiency
smooth rectangle
factorization norm
Information Complexity
[Braverman- Weinstein'11]
rectangle/corruption
discrepancy


Communication Complexity
partition
Information Complexity
efficiency
Our Result
smooth rectangle
rectangle/corruption
factorization norm
discrepancy

Main Theorem

Theorem: for all ε, μ, f, it holds that $IC_{\mu,\varepsilon}(f) \geq \Omega(\log 1/\eta_{\mu,\varepsilon}(f))$, i.e.
IC subsumes efficiency

Corollaries

1. All known CC lower bounds imply same lower bound on IC

2. All known problems with tight LBs satisfy Direct Sum Theorems
   $R_{\mu^k,\varepsilon}(f^k) \geq IC_{\mu^k,\varepsilon}(f^k) \geq k * IC_{\mu,\varepsilon}(f) = k * R_{\mu,\varepsilon}(f)$

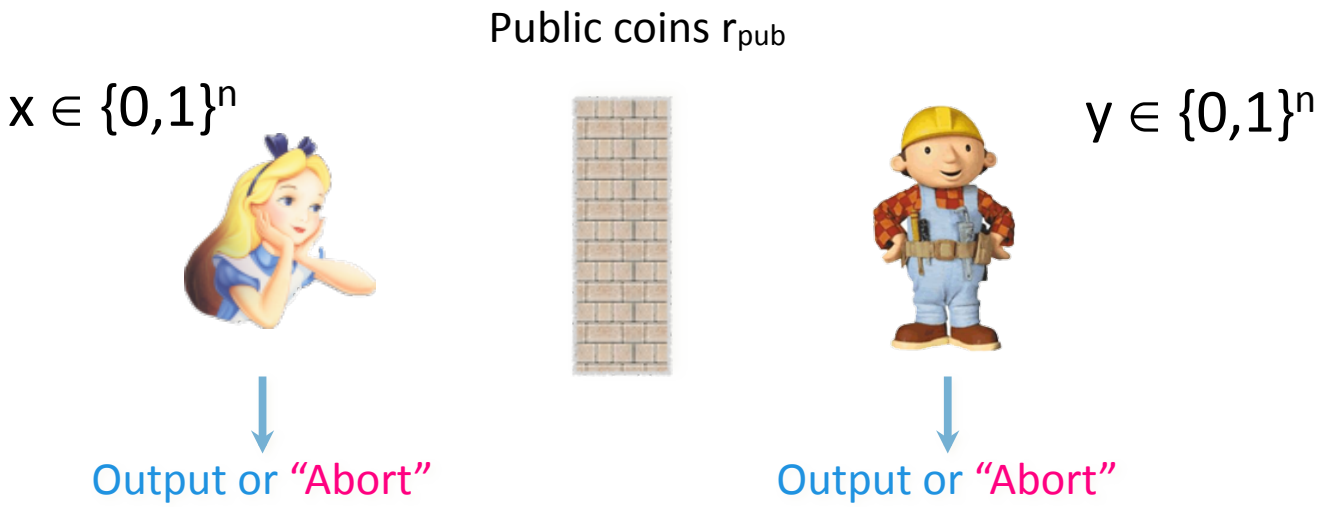
3. Quantum 1-way communication exponentially smaller than IC
   - VectorInSubspace problem [Klartag-Regev'11]: $\boldsymbol{O}(\log n)$ vs. $\Omega(n^{1/3})$

4. $IC_{\mu,\varepsilon}(\text{GapHamming}) = \Omega(n)$ and Direct Sum [CR'11, Sherstov'11])

Plan

1. Introducing Information Complexity
2. Relating Information to Communication Complexity
3. Zero-communication protocols and Information Complexity

Zero-communication protocols



Motivated by "detection loophole" in Bell experiments

- Same as previously, except:
  - No communication
  - Players may output "Abort"
  - Output of protocol: z if both players output z, otherwise "Abort"
  - π is (μ,ε)-good for f if $\Pr_{(X,Y)\leftarrow\mu}[\pi(X,Y) = f(X,Y) \mid \pi(X,Y) \neq \text{"Abort"}] \geq 1 - \varepsilon$
- Efficiency of protocol π defined as $\eta(\pi) = \Pr[\pi \text{ does not output "Abort"}]$
  - Define $\eta_{\mu,\varepsilon}(f) = \sup_{\pi \text{ } (\mu,\varepsilon)\text{-good for } f} \eta(\pi)$

Compressing CC to Zero communication

r_{pub} (= transcript t)

$x \in \{0,1\}^n$

$y \in \{0,1\}^n$

- Theorem: if π (μ,ε)-good for f using communication, then can build ZCP τ (μ,ε)-good for f with $\eta(\tau) \geq 2^{-CC(\pi)}$
  - Proof: τ uses r_{pub} to guess random transcript t
  - Alice (resp. Bob) sees if t is consistent with x (resp. y)
  - If so, output what π outputs, otherwise output "Abort"
  - Pr[t not abort] $\geq 2^{-CC(\pi)}$

$$CC_{\mu,\varepsilon}(f) \geq \Omega(\log 1/\eta_{\mu,\varepsilon}(f))$$

Compressing IC to Zero communication

$r_{pub} = (t_1, t_2, t_3, \ldots, t_K)$, hash h

$x \in \{0,1\}^n$

$y \in \{0,1\}^n$

1. Create set of individually accepted transcripts
   - Using r_{pub}, generate candidate transcripts $t_1, \ldots, t_K$ ($K \approx 2^{CC(\pi) + O(IC(\pi))}$)
   - Alice & Bob each decide whether to accept each t_i
   - Conditioned on t_i accepted, distribution of t_i same as $\pi(x,y)$
   - roughly $2^{O(IC(\pi))}$ accepted transcripts by each one

2. Find common accepted transcript
   - Using r_{pub}, pick a hash function that gives 0 w.p. $2^{-O(IC(\pi))}$
   - Alice and Bob: If there exists accepted transcript that hashes to 0, then output according to that, else Abort.
   - Pr[not abort] $\geq 2^{-O(IC(\pi))}$

Compressing IC to Zero communication

$r_{pub} = (t_1, t_2, t_3, \ldots, t_K)$, hash h

$x \in \{0,1\}^n$

$y \in \{0,1\}^n$

Theorem: for all ε, μ, f, it holds that $IC_{\mu,\varepsilon}(f) \geq \Omega(\log 1/\eta_{\mu,\varepsilon}(f))$

Conclusions and open questions

- Does IC = CC?
  - Yes for one-way
  - Yes, for all known functions. No counterexample candidate
  - Separating IC from CC will need new lower bound techniques
- How does IC compare to partition bound?
- What about zero error?
  - There exists a gap between IC and CC. Better definition?

Thank you

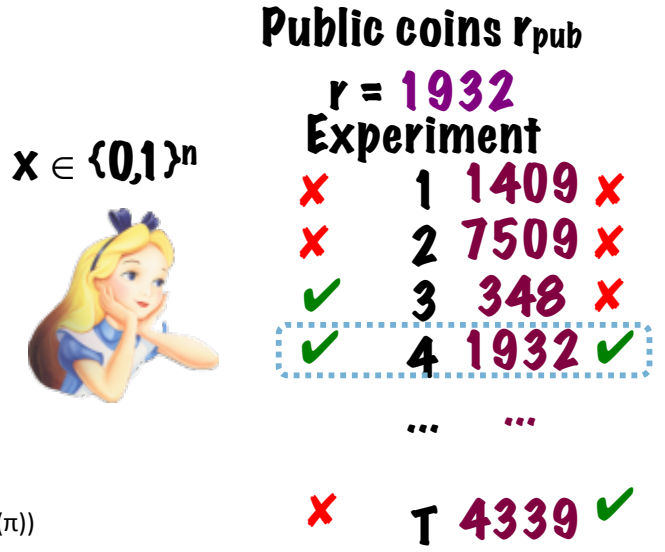
PhD and Postdoc positions available

Single sampling experiment

- Same as [Braverman-Weinstein'12]
- Define:
  - For transcript t, let t_A (resp. t_B) denote Alice's (resp. Bob's) messages in t
  - Define $p_x(t) = \Pr[\text{On } x, \text{ Alice outputs } t_A \text{ when Bob responds } t_B]$
  - Define $p_{y|x}(t) = \Pr_{(X,Y)\leftarrow\mu \mid X=x}[\text{On } y, \text{ Bob outputs } t_B \text{ when Alice responds } t_A]$
  - Define $q_y(t)$, $q_{x|y}(t)$ similarly
  - $p_x\, p_{y|x}$ gives distribution of $\pi(X, Y) \mid X = x$
  - $q_y\, q_{x|y}$ gives distribution of $\pi(X, Y) \mid Y = y$
  - $p_x\, q_y$ gives distribution of $\pi(x, y)$
  - Alice knows p_x and $p_{y|x}$ and Bob knows q_y and $q_{x|y}$
- Single sample protocol:
  1. Using r_{pub} select t uniformly, $\alpha, \beta \leftarrow [0, 2^{O(IC(\pi))}]$
  2. Alice accepts if $\alpha \le p_x(t)$ and $\beta \le 2^{O(IC(\pi))}\, p_{y|x}(t)$
  3. Bob accepts if $\alpha \le 2^{O(IC(\pi))}\, q_{x|y}(t)$ and $\beta \le q_y(t)$

- **Small IC implies for almost all t, $p_{y|x}(t) = q_y(t)$ and $q_{x|y}(t) = p_x(t)$**
- **Conditioned on t accepted, distribution of t same as $\pi(x,y)$**

Finding consistent experiment



- Set $K = 2^{O(IC(\pi))}$
  1. Using r_{pub}, pick random integer $r \leftarrow [K]$
  2. Using r_{pub}, for i'th experiment pick random integer $h_i \leftarrow [K]$
  3. Let i_0 be Alice's first accepting transcript, j_0 Bob's first accepting transcript
  4. Alice checks if $h_{i_0} = r$, if so give same output as i_0'th transcript else abort
  5. Same for Bob

- Set T so that $\Pr[i_0 = j_0] \geq 2^{-O(IC(\pi))}$
- $\Pr[h(i_0) = r] = 2^{-O(IC(\pi))}$
- Also have to deal with collisions...
- Overall: $\Pr[\text{not abort}] \geq 2^{-O(IC(\pi))}$
- Output distribution same as single experiment output distribution