Gap-Hamming-Distance: The Journey to an Optimal Lower Bound

Amit Chakrabarti

Dartmouth College

Main result joint with

Oded Regev, Tel Aviv University

Sublinear Algorithms Workshop at Bertinoro, May 2011

The Gap-Hamming-Distance Problem

Input: Alice gets $x \in \{0,1\}^n$, Bob gets $y \in \{0,1\}^n$.

Output:

- $\text{GHD}(x,y) = 1$ if $\Delta(x,y) > \frac{n}{2} + \sqrt{n}$
- $\text{GHD}(x,y) = 0$ if $\Delta(x,y) < \frac{n}{2} - \sqrt{n}$

Want: randomized, constant error protocol

Cost: Worst case number of bits communicated

$x =$	0	1	0	0	1	0	1	1	0	0	0	1
$y =$	0	0	0	0	0	0	1	1	1	0	0	1

$$n = 12; \quad \Delta(x,y) = 3 \in [6 - \sqrt{12}, 6 + \sqrt{12}]$$

Implications

Data stream lower bounds

- Distinct elements
- Frequency moments
- Norms
- Entropy
- General form of bound: $ps = \Omega(1/\varepsilon^2)$

Distributed functional monitoring lower bounds

Connections to differential privacy

The Reductions

E.g., Distinct Elements (Other problems: similar)

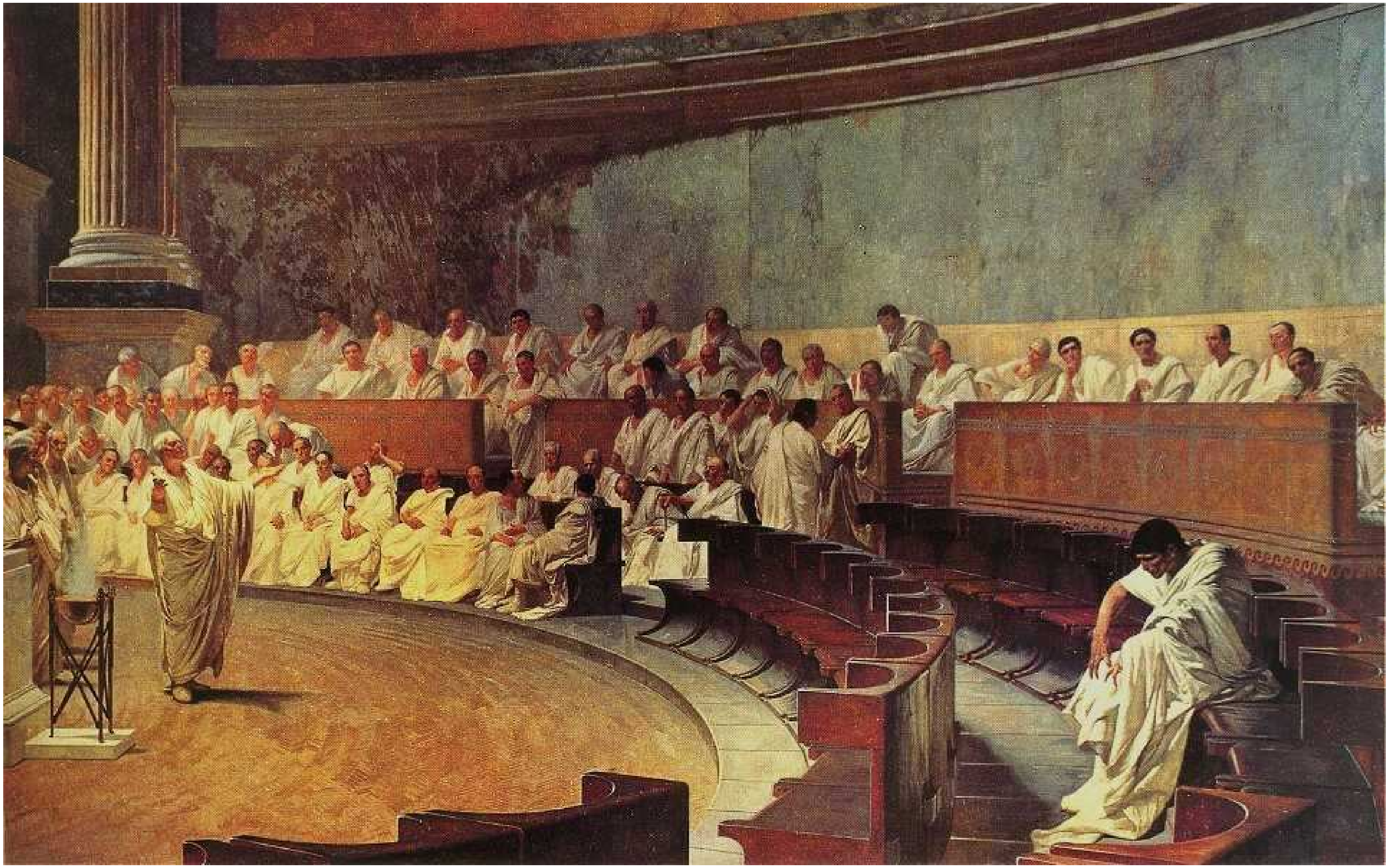
$x =$	**0**	**1**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**1**
σ :	(1,0)	(2,1)	(3,0)	(4,0)	(5,1)	(6,0)	(7,1)	(8,1)	(9,0)	(10,0)	(11,0)	(12,1)
$y =$	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**1**
τ :	(1,0)	(2,0)	(3,0)	(4,0)	(5,0)	(6,0)	(7,1)	(8,1)	(9,1)	(10,0)	(11,0)	(12,1)

Alice: $x \longmapsto \sigma = \langle (1, x_1), (2, x_2), \ldots, (n, x_n) \rangle$

Bob: $y \longmapsto \tau = \langle (1, y_1), (2, y_2), \ldots, (n, y_n) \rangle$

Notice: $F_0(\sigma \circ \tau) = n + \Delta(x, y) = \begin{cases} < \frac{3n}{2} - \sqrt{n}, & \text{or} \\ > \frac{3n}{2} + \sqrt{n}. \end{cases}$ Set $\varepsilon = \frac{1}{\sqrt{n}}$.

Ancient History



One-Pass Bounds

Indyk, Woodruff [FOCS 2003]

- Considered one-pass lower bound for DIST-ELEM
- Recognized relevance of GHD, difficulty of lower-bounding
- Defined "related" problem Π_{ℓ_2}, showed $\mathrm{R}^{\rightarrow}(\Pi_{\ell_2}) = \Omega(n)$
- Concluded $\Omega(\varepsilon^{-2})$ bound for $\mathrm{DIST\text{-}ELEM}_{m,\varepsilon}$ with $m = \widetilde{\Omega}(1/\varepsilon^9)$

One-Pass Bounds

Indyk, Woodruff [FOCS 2003]

- Considered one-pass lower bound for DIST-ELEM
- Recognized relevance of GHD, difficulty of lower-bounding
- Defined "related" problem Π_{ℓ_2}, showed $\mathrm{R}^{\rightarrow}(\Pi_{\ell_2}) = \Omega(n)$
- Concluded $\Omega(\varepsilon^{-2})$ bound for $\text{DIST-ELEM}_{m,\varepsilon}$ with $m = \widetilde{\Omega}(1/\varepsilon^9)$

Woodruff [SODA 2004]

- Worked with GHD itself, showed $\mathrm{R}^{\rightarrow}(\text{GHD}) = \Omega(n)$
- Very intricate combinatorial proof, with hairy probability estimations
- Conjectured $\mathrm{R}(\text{GHD}) = \Omega(n)$, implying multi-pass lower bounds

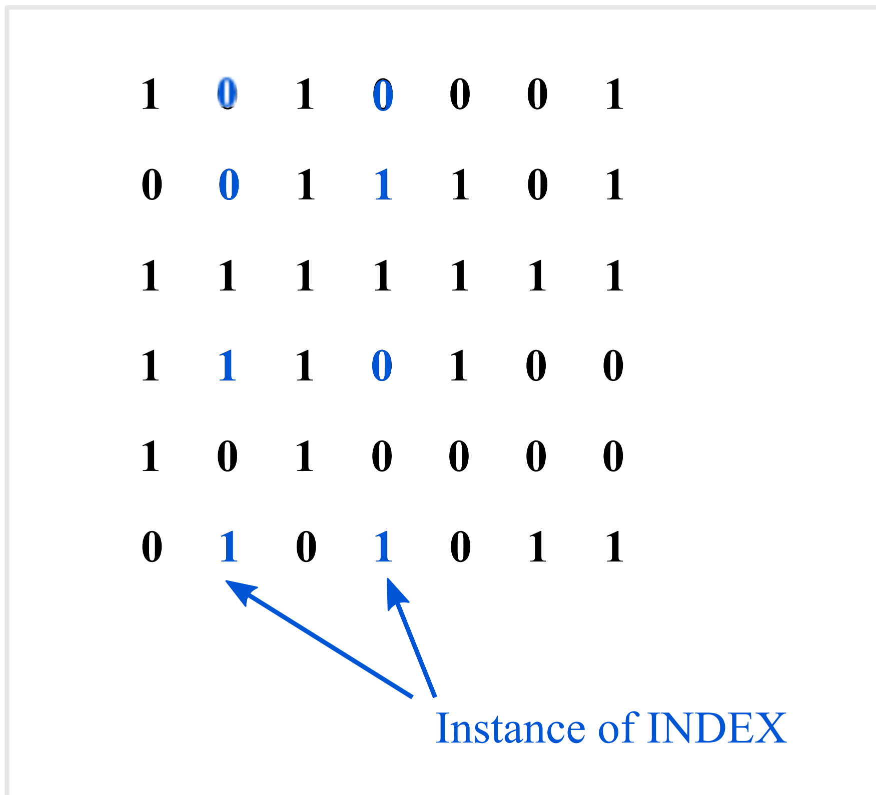
The VC-Dimension Technique

- Consider communication matrix of GHD as set system
- The system has $\Omega(n)$ VC-dimension

1	**0**	**1**	**0**	**0**	**0**	**1**
0	**0**	**1**	**1**	**1**	**0**	**1**
1	**1**	**1**	**1**	**1**	**1**	**1**
1	**1**	**1**	**0**	**1**	**0**	**0**
1	**0**	**1**	**0**	**0**	**0**	**0**
0	**1**	**0**	**1**	**0**	**1**	**1**

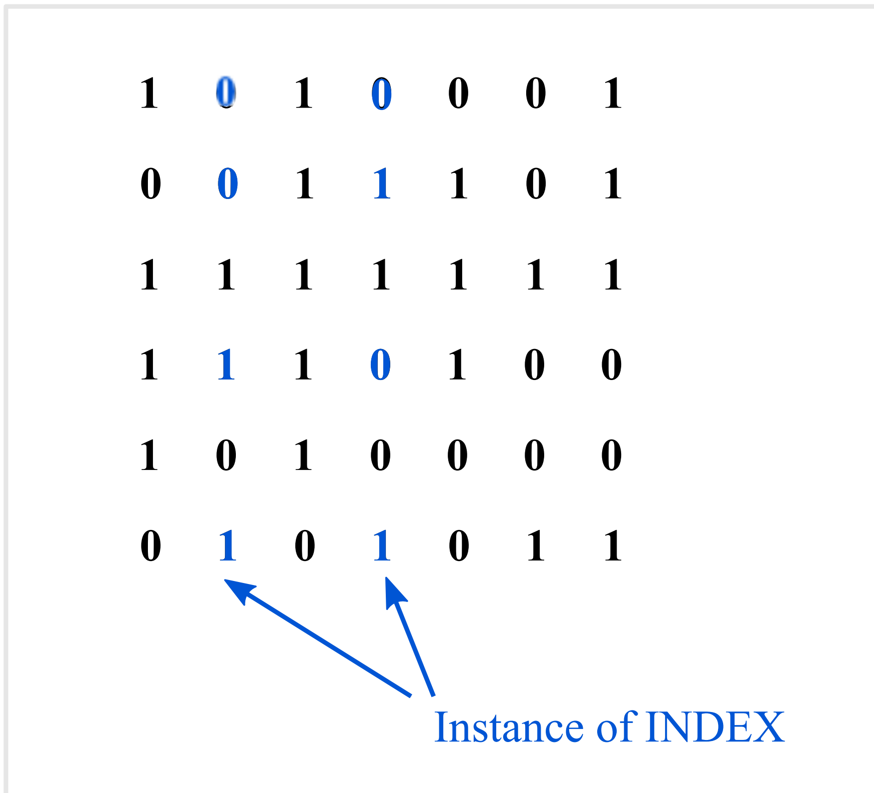
The VC-Dimension Technique

- Consider communication matrix of GHD as set system
- The system has $\Omega(n)$ VC-dimension



The VC-Dimension Technique

- Consider communication matrix of GHD as set system
- The system has $\Omega(n)$ VC-dimension



- Thus, $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$

The Middle Ages

A Nice Simplification

Jayram, Kumar, Sivakumar [circa 2005]

- Simpler proof of $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$
- *Much* simpler: direct reduction from INDEX
- Geometric intuition:

$$\text{Alice:} \quad x \in \{0,1\}^n \quad \longmapsto \quad \widetilde{x} \in \left\{ \frac{1}{\sqrt{n}}, -\frac{1}{\sqrt{n}} \right\}^n \in \mathbb{R}^n$$

$$\text{Bob:} \quad j \in [n] \quad \longmapsto \quad e_j = (0,\ldots,0,1,0,\ldots,0) \in \mathbb{R}^n$$

- Observe: $\langle \widetilde{x}, e_j \rangle \not\approx 0$, and x_j determined by $\mathrm{sgn}\langle \widetilde{x}, e_j \rangle$
- We've reduced INDEX to "gap-inner-product", or GIP

Inner Product $\leftrightarrow$ Hamming Distance

- Obviously, GHD $\to$ GIP:

$$\langle \widetilde{x}, \widetilde{y} \rangle = 1 - \frac{2\Delta(x,y)}{n}$$

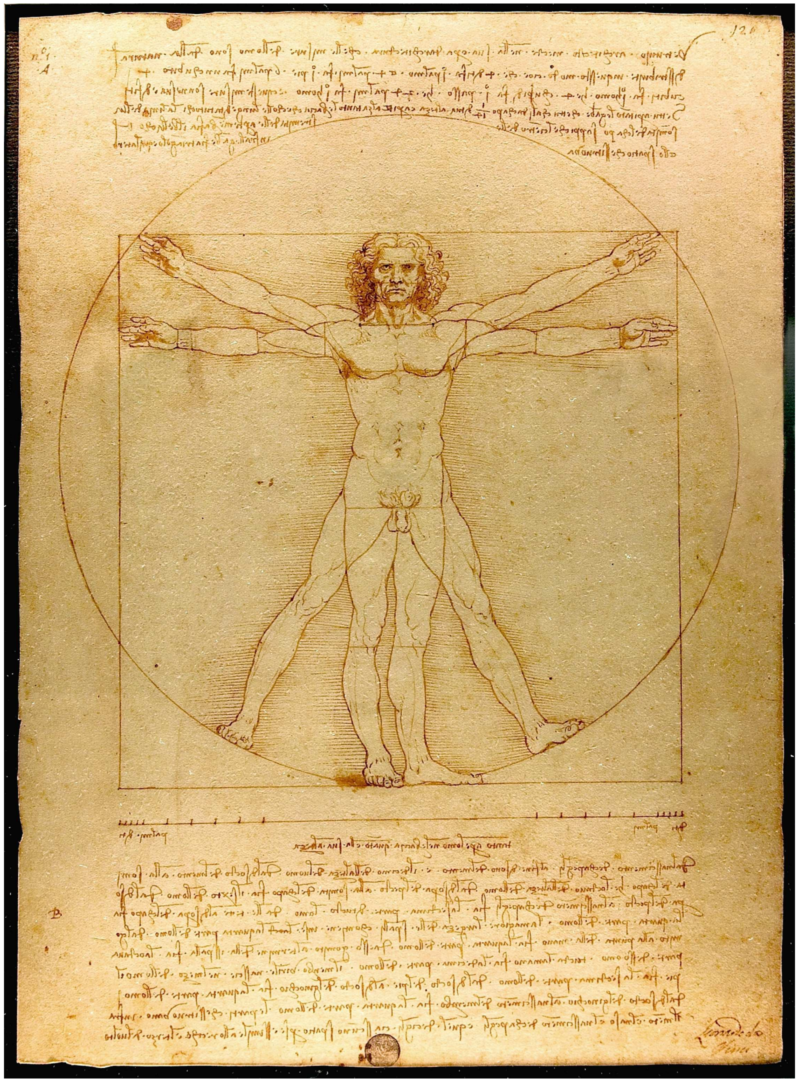
$$\langle \widetilde{x}, \widetilde{y} \rangle \gtrless \mp \frac{2}{\sqrt{n}} \Rightarrow \Delta(x,y) \lessgtr \frac{n}{2} \pm \sqrt{n}$$

- Also, GIP $\to$ GHD by "discretization transform":

  Pick random Gaussians $r_1, \ldots, r_N$, with $N = 10n$

$$\begin{array}{rlll} \text{Alice:} & \bar{x} \in \mathbb{R}^n & \longmapsto & x = (\operatorname{sgn}\langle \bar{x}, r_1 \rangle, \ldots, \operatorname{sgn}\langle \bar{x}, r_N \rangle) \in \{\pm 1\}^N \\ \text{Bob:} & \bar{y} \in \mathbb{R}^n & \longmapsto & y = (\operatorname{sgn}\langle \bar{y}, r_1 \rangle, \ldots, \operatorname{sgn}\langle \bar{y}, r_N \rangle) \in \{\pm 1\}^N \\ & \langle \bar{x}, \bar{y} \rangle \gtrless \mp \frac{1}{\sqrt{n}} & \overset{\text{whp}}{\Longrightarrow} & \Delta(x,y) \lessgtr \frac{N}{2} \pm O(\sqrt{N}) \end{array}$$

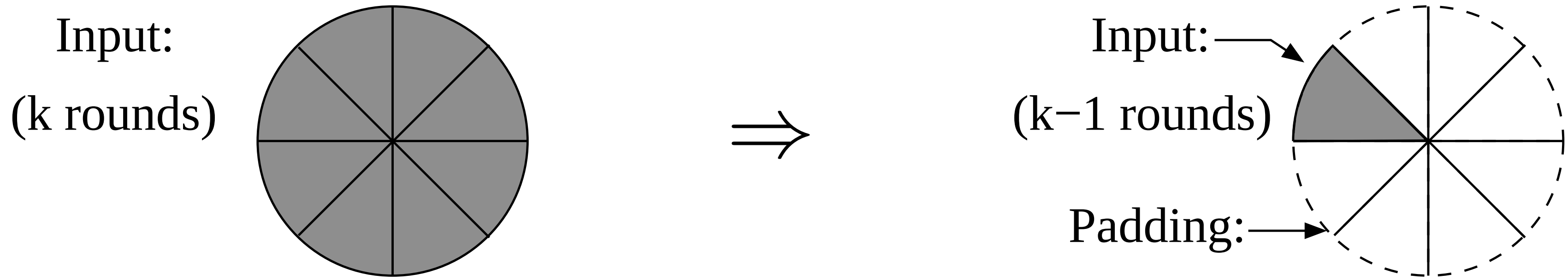
The Renaissance Era



Round Elimination

Brody, Chakrabarti [CCC 2009]

- Can we at least rule out a *two-pass* improvement for DIST-ELEM?
- A cheap first message makes little progress? Then rinse, repeat
- Tends to decimate problem [Miltersen-Nisan-Safra-Wigderson'98] [Sen'03]



Another VC-Dimension Argument: Subcube Lifting

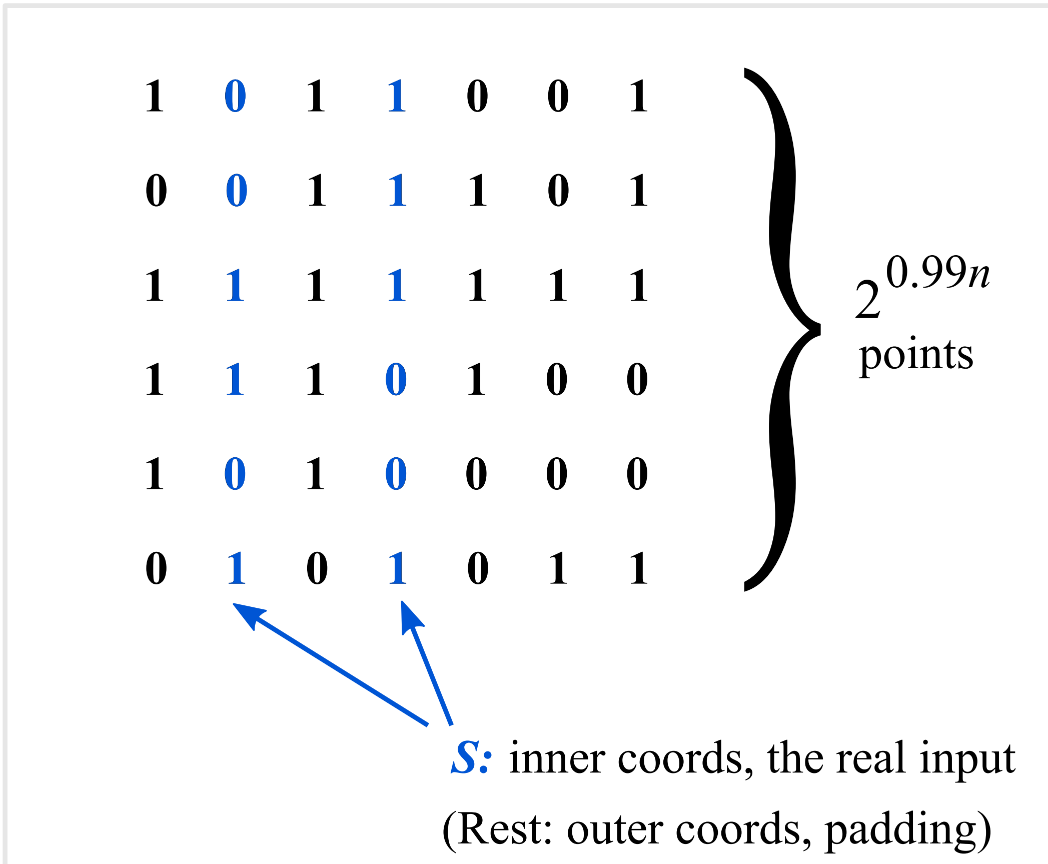
First message constant on large set:

1	**0**	**1**	**1**	**0**	**0**	**1**
0	**0**	**1**	**1**	**1**	**0**	**1**
1	**1**	**1**	**1**	**1**	**1**	**1**
1	**1**	**1**	**0**	**1**	**0**	**0**
1	**0**	**1**	**0**	**0**	**0**	**0**
0	**1**	**0**	**1**	**0**	**1**	**1**

} $2^{0.99n}$ points

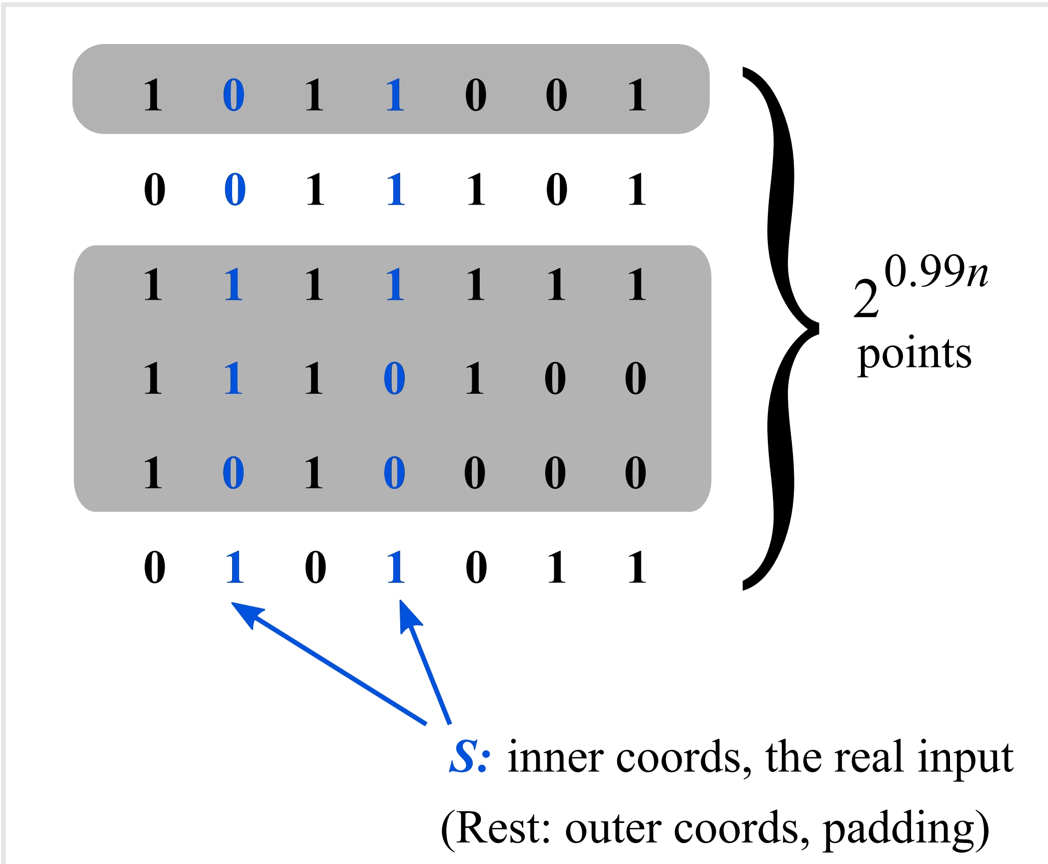
Another VC-Dimension Argument: Subcube Lifting

First message constant on large set:



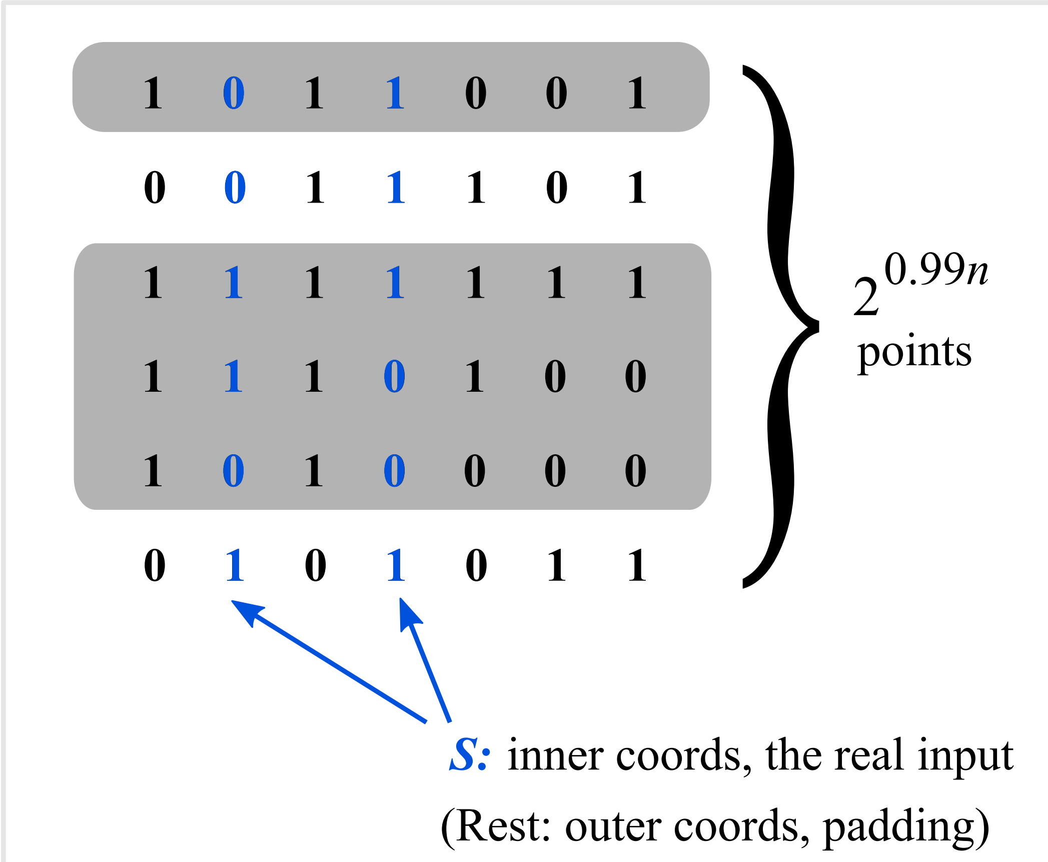
Another VC-Dimension Argument: Subcube Lifting

First message constant on large set:



Another VC-Dimension Argument: Subcube Lifting

First message constant on large set:



Alice, Bob lift their $(n/3)$-dim inputs from inner coords to full n-dim space

First message now redundant, so eliminate! [Brody-C.'09]

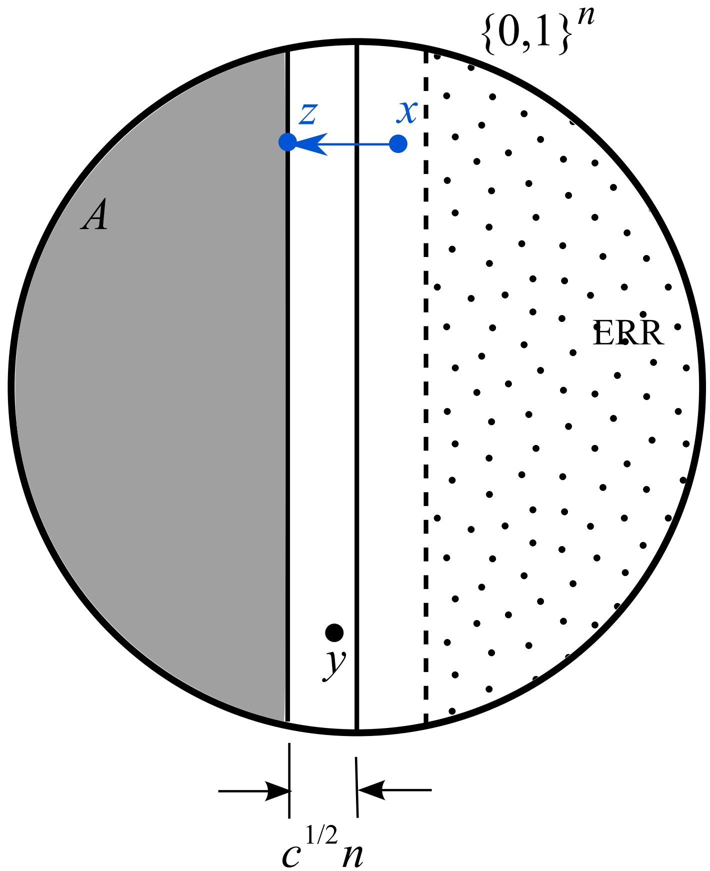
Better Round Elimination

Brody, Chakrabarti, Regev, Vidick, de Wolf [RANDOM 2010]

- Previous argument reduced dimension too rapidly
- Gives $R^k(\text{GHD}) = n/2^{O(k^2)}$
- Can improve to $R^k(\text{GHD}) = n/O(k^2)$

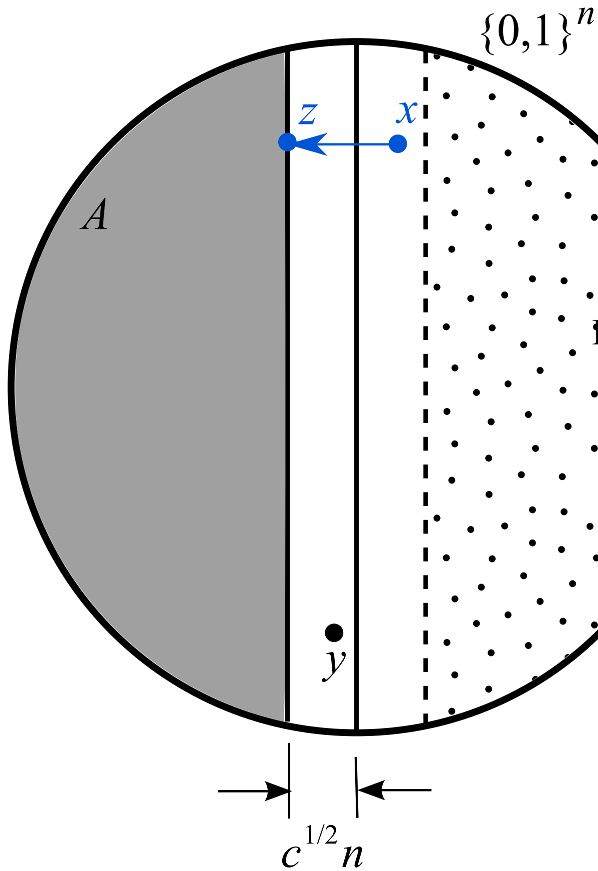
Round Elimination V2.0: Geometric Perturbation

First message constant over large set A



Round Elimination V2.0: Geometric Perturbation

First message constant over large set A



Alice: replace x with $z = \text{NearestNeighbour}(x, A)$

Modern History

Main Theorem

Chakrabarti, Regev [STOC 2011]

And now, we show:

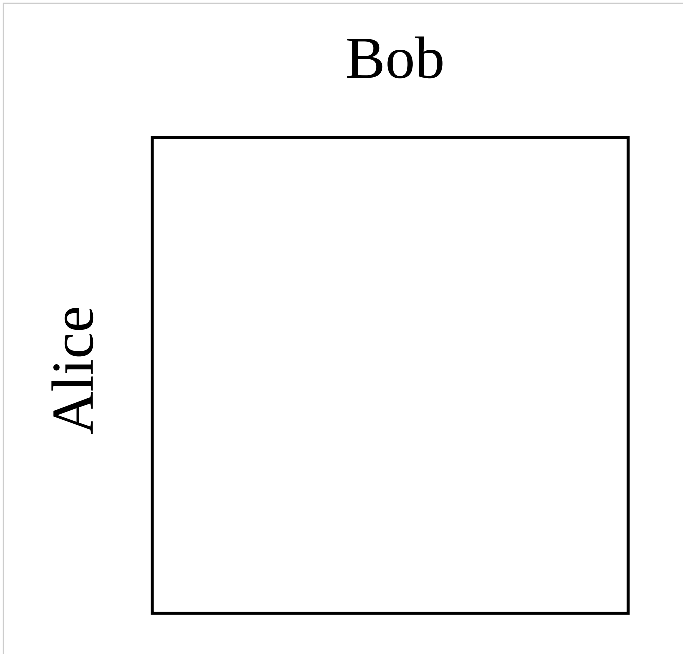
$$\mathrm{R}(\textsc{ghd}) = \Omega(n)$$

The Rectangle Property

Input universe $U = \{0,1\}^n \times \{0,1\}^n$

Deterministic protocol P, communicating $\leq c$ bits

partitions U into $\leq 2^c$ rectangles $A_i \times B_i$, where $A_i, B_i \subseteq \{0,1\}^n$

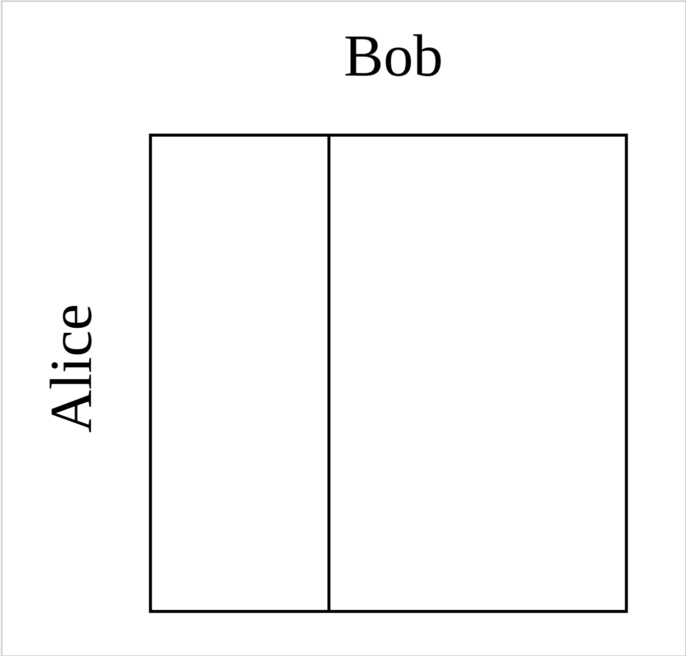


The Rectangle Property

Input universe $U = \{0,1\}^n \times \{0,1\}^n$

Deterministic protocol P, communicating $\leq c$ bits

partitions U into $\leq 2^c$ rectangles $A_i \times B_i$, where $A_i, B_i \subseteq \{0,1\}^n$

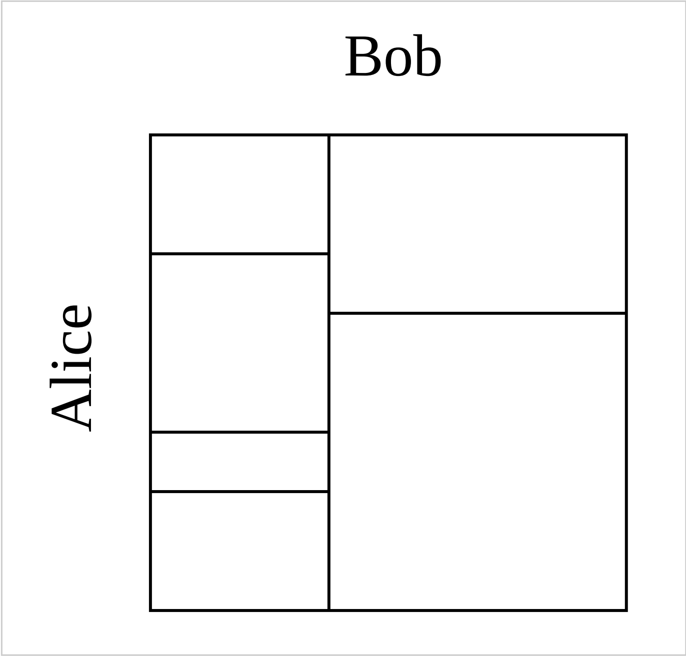


The Rectangle Property

Input universe $U = \{0,1\}^n \times \{0,1\}^n$

Deterministic protocol P, communicating $\leq c$ bits

partitions U into $\leq 2^c$ rectangles $A_i \times B_i$, where $A_i, B_i \subseteq \{0,1\}^n$

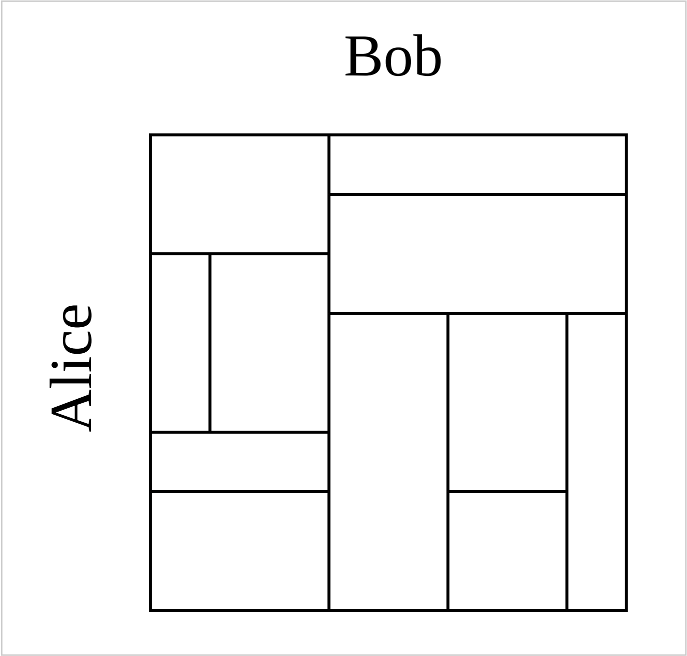


The Rectangle Property

Input universe $U = \{0,1\}^n \times \{0,1\}^n$

Deterministic protocol P, communicating $\leq c$ bits

partitions U into $\leq 2^c$ rectangles $A_i \times B_i$, where $A_i, B_i \subseteq \{0,1\}^n$

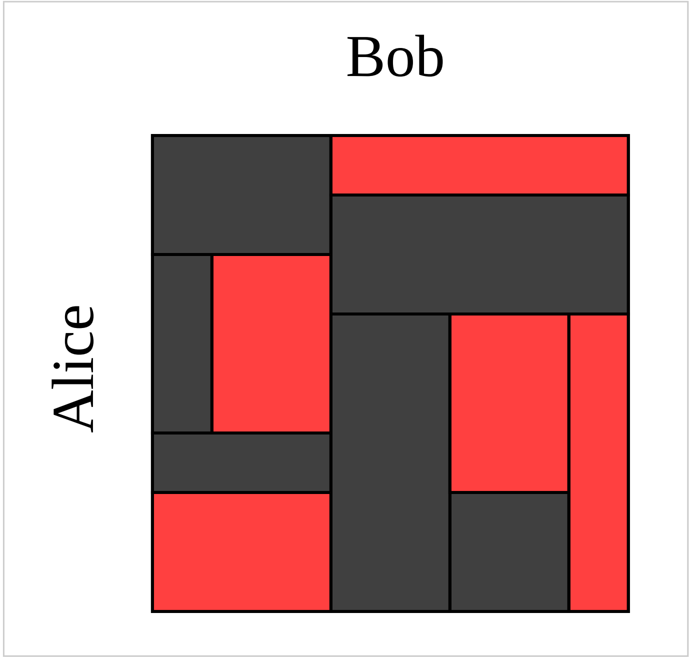


The Rectangle Property

Input universe $U = \{0,1\}^n \times \{0,1\}^n$

Deterministic protocol P, communicating $\leq c$ bits

partitions U into $\leq 2^c$ rectangles $A_i \times B_i$, where $A_i, B_i \subseteq \{0,1\}^n$

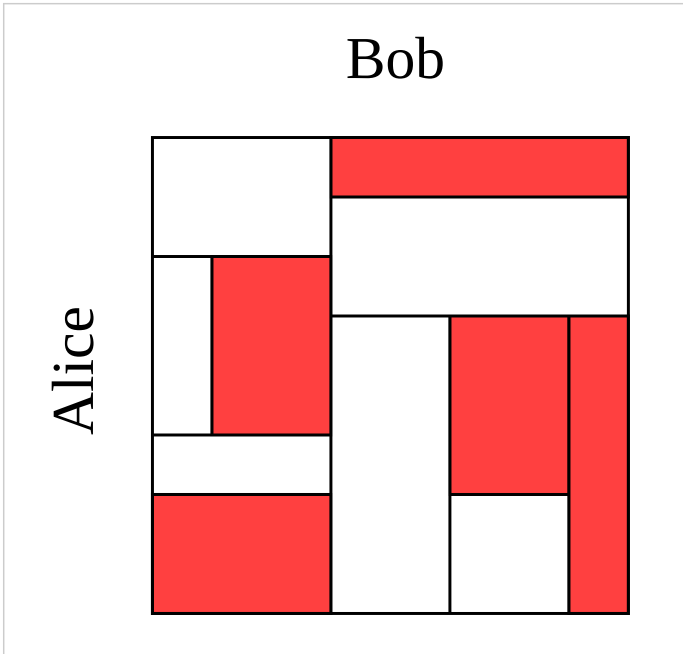


The Rectangle Property

Input universe $U = \{0,1\}^n \times \{0,1\}^n$

Deterministic protocol P, communicating $\leq c$ bits

partitions U into $\leq 2^c$ rectangles $A_i \times B_i$, where $A_i, B_i \subseteq \{0,1\}^n$



If P computes $f : U \to \{0,1\}$, then $f^{-1}(0) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

The Corruption Technique and a Twist

Deterministic: $f^{-1}(0) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

Randomized: $\{P \text{ outputs } 0\} = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

- Partition covers *most of* $f^{-1}(0)$
- Each R_i mostly *uncorrupted*: contains much fewer 1s than 0s.

The Corruption Technique and a Twist

Deterministic: $f^{-1}(0) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

Randomized: $\{P \text{ outputs } 0\} = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

- Partition covers *most of* $f^{-1}(0)$
- Each R_i mostly *uncorrupted*: contains much fewer 1s than 0s.

For lower bound:

- Show every large rectangle (size $\geq 2^{0.99n} \times 2^{0.99n}$) is *corrupted*

$$\mu_1(R) \;\geq\; \alpha\, \mu_0(R)$$

The Corruption Technique and a Twist

Deterministic: $f^{-1}(0) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

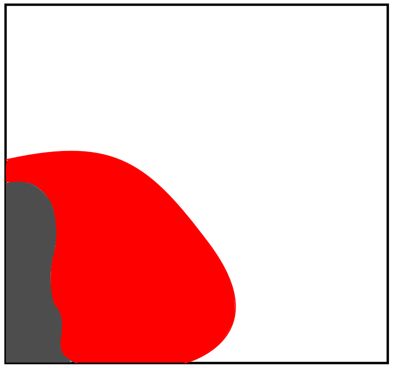
Randomized: $\{P \text{ outputs } 0\} = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

- Partition covers *most of* $f^{-1}(0)$
- Each R_i mostly *uncorrupted*: contains much fewer 1s than 0s.

For lower bound:

- Show every large rectangle (size $\geq 2^{0.99n} \times 2^{0.99n}$) is *corrupted*

$$\mu_1(R) \;\geq\; \alpha\,\mu_0(R)$$



The Corruption Technique and a Twist

Deterministic: $f^{-1}(0) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

Randomized: $\{P \text{ outputs } 0\} = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

- Partition covers *most of* $f^{-1}(0)$
- Each R_i mostly *uncorrupted*: contains much fewer 1s than 0s.

For lower bound:

- Show every large rectangle (size $\geq 2^{0.99n} \times 2^{0.99n}$) is *corrupted*

$$\mu_1(R) \;\geq\; \alpha\, \mu_0(R)$$

- Caveat: not true! E.g., $\{(x, y) : x_{1:100\sqrt{n}} = y_{1:100\sqrt{n}} = \vec{0}\}$

The Corruption Technique and a Twist

Deterministic: $f^{-1}(0) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

Randomized: $\{P \text{ outputs } 0\} = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

- Partition covers *most of* $f^{-1}(0)$
- Each R_i mostly *uncorrupted*: contains much fewer 1s than 0s.

For lower bound:

- Show every large rectangle (size $\geq 2^{0.99n} \times 2^{0.99n}$) is *corrupted*

$$\mu_1(R) \;\geq\; \alpha\, \mu_0(R)$$

- Caveat: not true! E.g., $\{(x,y) : x_{1:100\sqrt{n}} = y_{1:100\sqrt{n}} = \vec{0}\}$
- Show weaker inequality

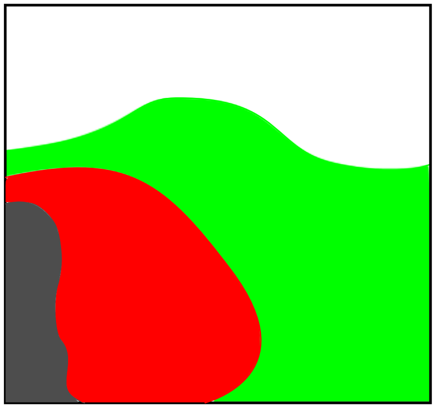
$$\mu_1(R) + \beta\, \mu_\star(R) \;\geq\; \alpha\, \mu_0(R) \qquad (\alpha > \beta)$$

Corruption with Jokers

Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$, and another distrib $\mu_\star$

Argue that for all large rectangles R, we have

$$\mu_1(R) + \beta\, \mu_\star(R) \;\geq\; \alpha\, \mu_0(R) \qquad\qquad (\alpha > \beta)$$



Corruption with Jokers

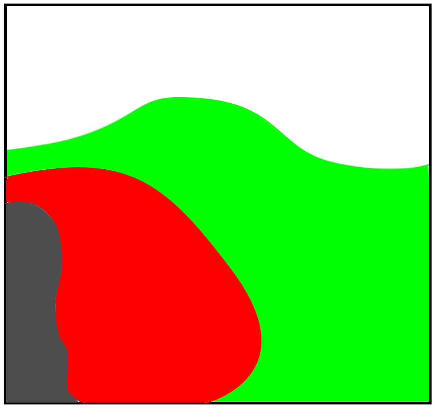
Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$, and another distrib $\mu_\star$

Argue that for all large rectangles R, we have

$$\mu_1(R) + \beta\,\mu_\star(R) \;\geq\; \alpha\,\mu_0(R) \qquad (\alpha > \beta)$$

Sum over partition $\{P \text{ outputs } 0\} = \bigcup_{i=1}^{2^c} R_i$:

$$\mu_1(P^{-1}(0)) + \beta\,\mu_\star(P^{-1}(0)) \;\geq\; \alpha\,\mu_0(P^{-1}(0))$$



Corruption with Jokers

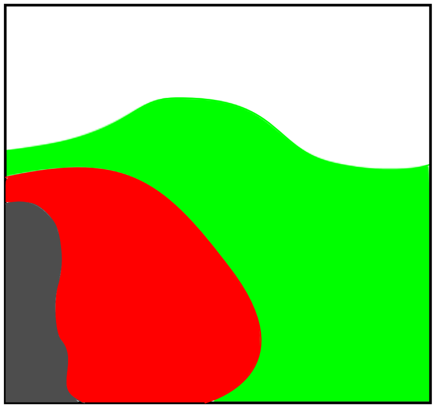
Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$, and another distrib $\mu_\star$

Argue that for all large rectangles R, we have

$$\mu_1(R) + \beta\,\mu_\star(R) \;\geq\; \alpha\,\mu_0(R) \qquad (\alpha > \beta)$$

Sum over partition $\{P \text{ outputs } 0\} = \bigcup_{i=1}^{2^c} R_i$:

$$\mu_1(P^{-1}(0)) + \beta\,\mu_\star(P^{-1}(0)) \;\geq\; \alpha\,\mu_0(P^{-1}(0)) \;\geq\; \alpha(1-\varepsilon)$$



Corruption with Jokers

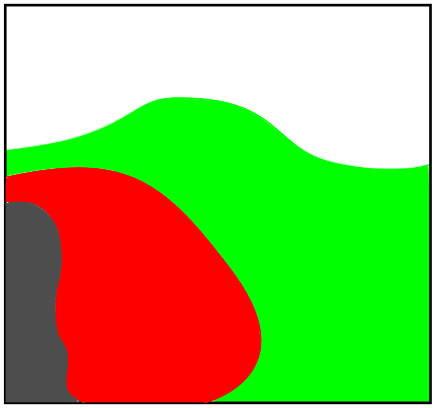
Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$, and another distrib $\mu_\star$

Argue that for all large rectangles R, we have

$$\mu_1(R) + \beta\,\mu_\star(R) \;\geq\; \alpha\,\mu_0(R) \qquad (\alpha > \beta)$$

Sum over partition $\{P \text{ outputs } 0\} = \bigcup_{i=1}^{2^c} R_i$:

$$\varepsilon + \beta \;\geq\; \mu_1(P^{-1}(0)) + \beta\,\mu_\star(P^{-1}(0)) \;\geq\; \alpha\,\mu_0(P^{-1}(0)) \;\geq\; \alpha(1-\varepsilon)$$



The Corruption Inequality and Its Proof

Let

$$\mu_0 = \text{Uniform on } \{(x,y) : \langle \widetilde{x}, \widetilde{y} \rangle = 0\}$$

$$\mu_1 = \text{Uniform on } \{(x,y) : \langle \widetilde{x}, \widetilde{y} \rangle = -10/\sqrt{n}\}$$

$$\mu_\star = \text{Uniform on } \{(x,y) : \langle \widetilde{x}, \widetilde{y} \rangle = 10/\sqrt{n}\}$$

The Key Inequality: For $|A|, |B| \geq 2^{0.99n}$

$$\frac{1}{2}(\mu_1(A \times B) + \mu_\star(A \times B)) \geq \frac{9}{10}\mu_0(A \times B)$$

"Inner product between large sets not too concentrated around zero"

The Corruption Inequality and Its Proof

Let

$$\mu_0 = \text{Uniform on } \{(x,y) : \langle \widetilde{x}, \widetilde{y} \rangle = 0\}$$

$$\mu_1 = \text{Uniform on } \{(x,y) : \langle \widetilde{x}, \widetilde{y} \rangle = -10/\sqrt{n}\}$$

$$\mu_\star = \text{Uniform on } \{(x,y) : \langle \widetilde{x}, \widetilde{y} \rangle = 10/\sqrt{n}\}$$

The Key Inequality: For $|A|, |B| \geq 2^{0.99n}$

$$\frac{1}{2}(\mu_1(A \times B) + \mu_\star(A \times B)) \geq \frac{9}{10}\,\mu_0(A \times B)$$

"Inner product between large sets not too concentrated around zero"

Proof Strategy: For $A, B \subseteq \mathbb{R}^n$ with $\gamma(A), \gamma(B) \geq 2^{-0.01n}$

distrib of $\langle \hat{x}, \hat{y} \rangle$ "spread out" like $N(0,1)$

where $\gamma = n$-dim Gaussian, $(\hat{x}, \hat{y}) \leftarrow A \times B$

Proof Details

Goal: For $A, B \subseteq \mathbb{R}^n$ with $\gamma(A), \gamma(B) \geq 2^{-0.01n}$

distrib of $\langle \hat{x}, \hat{y} \rangle$ "spread out" like $N(0,1)$

Think

$$A = \{\text{directions}\}$$

$$A_{\text{bad}} = \{\text{bad directions in } A\}$$

$$= \{\hat{x} \in A : \langle \hat{x}, \hat{y} \rangle \text{ not spread out, for } \hat{y} \leftarrow B\}$$

Proof Details

Goal: For $A, B \subseteq \mathbb{R}^n$ with $\gamma(A), \gamma(B) \geq 2^{-0.01n}$

distrib of $\langle \hat{x}, \hat{y} \rangle$ "spread out" like $N(0,1)$

Think

$$\begin{aligned} A &= \{\text{directions}\} \\ A_{\mathsf{bad}} &= \{\text{bad directions in } A\} \\ &= \{\hat{x} \in A : \langle \hat{x}, \hat{y} \rangle \text{ not spread out, for } \hat{y} \leftarrow B\} \end{aligned}$$

For a contradiction, suppose $\gamma(A_{\mathsf{bad}}) > 2^{-0.02n}$

Then (Raz's Lemma): A contains orthogonal bad dirs $\hat{x}_1, \ldots, \hat{x}_{n/2}$

Proof Details

Goal: For $A, B \subseteq \mathbb{R}^n$ with $\gamma(A), \gamma(B) \geq 2^{-0.01n}$

distrib of $\langle \hat{x}, \hat{y} \rangle$ "spread out" like $N(0,1)$

Think

$$\begin{aligned} A &= \{\text{directions}\} \\ A_{\text{bad}} &= \{\text{bad directions in } A\} \\ &= \{\hat{x} \in A : \langle \hat{x}, \hat{y} \rangle \text{ not spread out, for } \hat{y} \leftarrow B\} \end{aligned}$$

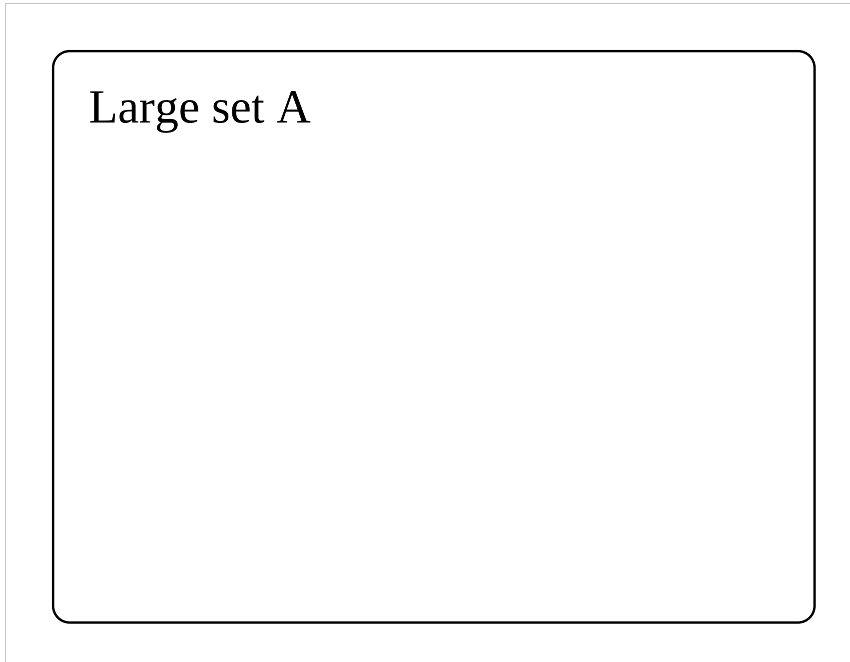
For a contradiction, suppose $\gamma(A_{\text{bad}}) > 2^{-0.02n}$

Then (Raz's Lemma): A contains orthogonal bad dirs $\hat{x}_1, \ldots, \hat{x}_{n/2}$

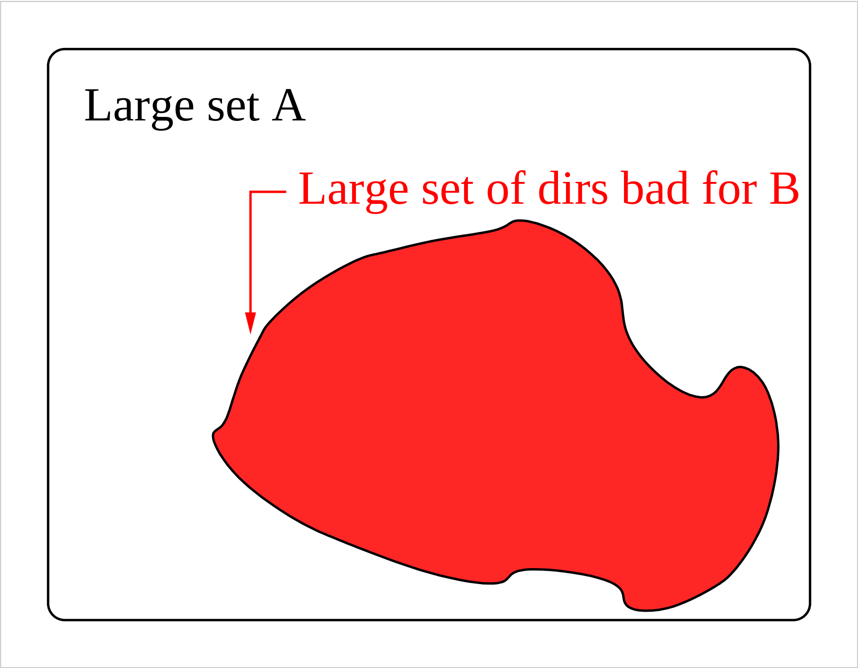
Therefore (Information Theory): $\hat{y} \leftarrow B$ can't have enough entropy

Contradicts $\gamma(B) \geq 2^{-0.01n}$

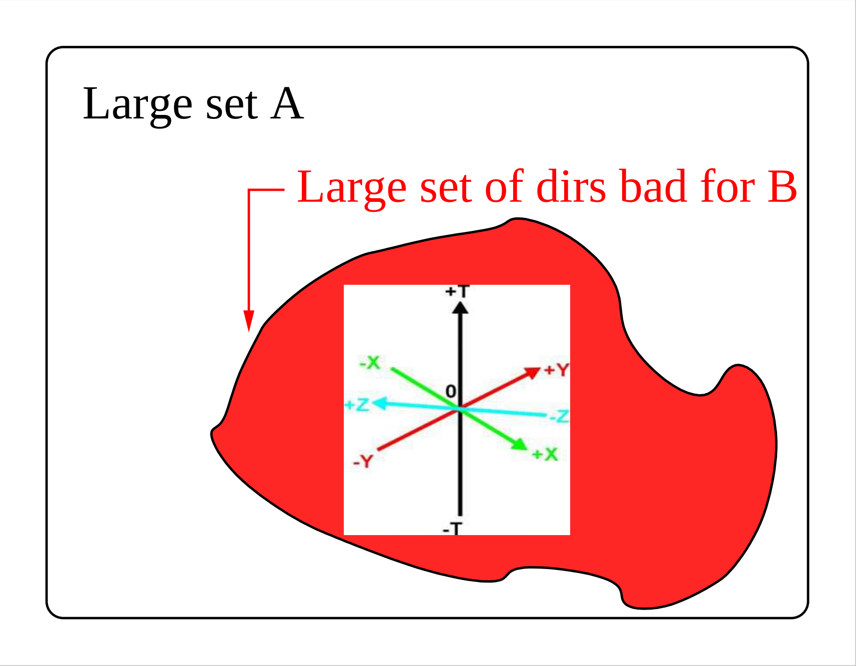
Geometric and Info Theoretic Intuition



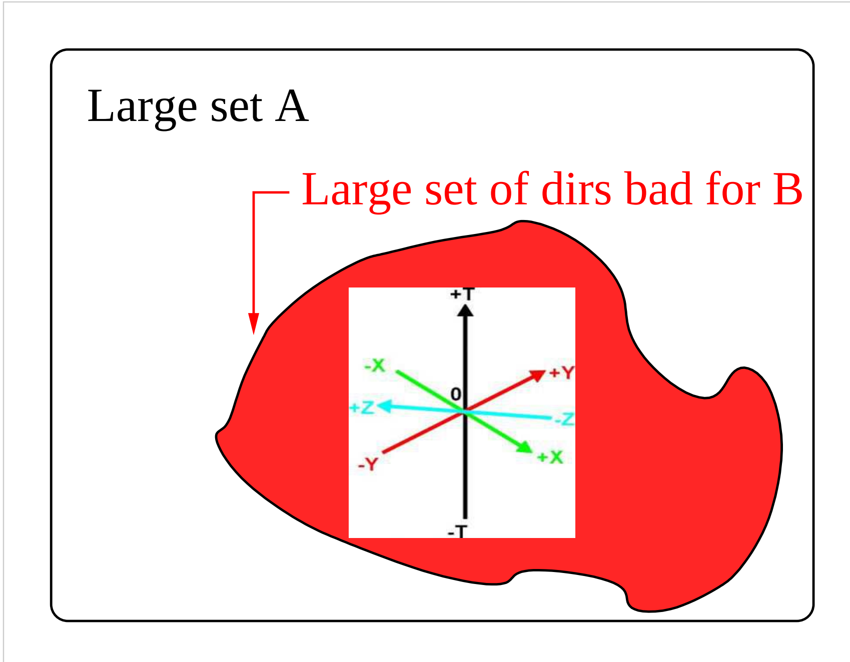
Geometric and Info Theoretic Intuition



Geometric and Info Theoretic Intuition



Geometric and Info Theoretic Intuition



$$
\begin{aligned}
0.99n \;\le\; \mathrm{H}(y) &\le \mathrm{H}(\langle y, x_1\rangle, \ldots, \langle y, x_n\rangle) \\
&= \textstyle\sum_{k=1}^{n/2} \mathrm{H}(\langle y, x_k\rangle \mid \langle y, x_1\rangle, \ldots, \langle y, x_{k-1}\rangle) \\
&\quad + \textstyle\sum_{k=n/2+1}^{n} \mathrm{H}(\langle y, x_k\rangle \mid \langle y, x_1\rangle, \ldots, \langle y, x_{k-1}\rangle) \\
&\le \textstyle\sum_{k=1}^{n/2} 0.7 + \sum_{k=n/2+1}^{n} 1 \;=\; 0.85n
\end{aligned}
$$

The Future

The Future

Two simplifications of our proof [not yet published]

- Vidick shows following anti-concentration inequality:

$$\mathbb{E}[\langle \widetilde{x}, \widetilde{y} \rangle^2] = \Omega(1/n)$$

Avoids "continuous information theory"; just concentration of measure

- Sherstov: anti-concetration gives corruption-based proof that

$$\mathrm{R}(\textsc{near-orthogonal}) = \Omega(n)$$

and reduces NEAR-ORTHOGONAL to GHD; thus avoids "jokers"

- Also, Sherstov proves anti-concentration using Talagrand's inequality

Conclusions

- Settled communication complexity of GHD, proving a long-conjectured $\Omega(n)$ bound
- As a result, understood multi-pass space complexity of a number of data stream problems

Conclusions

- Settled communication complexity of GHD, proving a long-conjectured $\Omega(n)$ bound
- As a result, understood multi-pass space complexity of a number of data stream problems

Open Problem

Prove that GHD is hard under the uniform distribution