# Multi-pass Data Stream Lower Bounds via Round Elimination

Amit Chakrabarti

Dartmouth College

WAPMDS, IIT Kanpur, Dec 2009

# Lower Bounds Paradigms

**Algorithm design:**

**Lower bounds:**

# Lower Bounds Paradigms

**Algorithm design:**

divide & conquer, greedy, dynamic programming, LP relaxation, . . .

**Lower bounds:** ? ? ?

## Lower Bounds Paradigms

**Algorithm design:**

divide & conquer, greedy, dynamic programming, LP relaxation, . . .

**Lower bounds:** ? ? ?

- Information complexity paradigm [C.-Shi-Wirth-Yao'01]
- Round elimination paradigm [Miltersen-Nisan-Safra-Wigderson'95]

# Multi-Pass Lower Bounds

Data streams: two broad application scenarios

- **Networks:** Busy router, packets whizzing by
  - Web traffic statistics
  - Intrusion detection
- **Databases:** Huge DB, linear scan cheaper than random access
  - Query optimisation: join size estimation
  - Log analysis

# Multi-Pass Lower Bounds

Data streams: two broad application scenarios

- **Networks:** Busy router, packets whizzing by
  - Web traffic statistics
  - Intrusion detection
- **Databases:** Huge DB, linear scan cheaper than random access
  - Query optimisation: join size estimation
  - Log analysis
- DB setting: Multiple passes meaningful

**This talk:** Pass/space tradeoffs for some basic stream problems

## Data Stream Model

- Formally: input stream $= n$ tokens, each token $\in [m]$
  - Assume $\log m = \Theta(\log n)$
- Compute some function of stream, using
  - Small space, $s \ll m, n$ ... ideally, $s = O(\log n)$
  - Small number of passes, $p$

# Problems of Interest

**Class A:**

- Median

**Class B:**

- Distinct elements
- Frequency moments
- Empirical entropy

# Problems of Interest

**Class A:**

- Median

**Class B:**

- Distinct elements , $F_0$
- Frequency moments , $F_k = \sum_{i=1}^{m} \mathrm{freq}(i)^k$
- Empirical entropy , $H = \sum_{i=1}^{m} (\mathrm{freq}(i)/m) \cdot \log(m/\mathrm{freq}(i))$

# Problems of Interest

**Class A:**

- Median
- **Key question:** Want $s = O(\log n)$; then $p = ??$
  - Dates back to first "data streams" paper [Munro-Paterson'78]

**Class B:**

- Distinct elements , $F_0$
- Frequency moments , $F_k = \sum_{i=1}^{m} \mathrm{freq}(i)^k$
- Empirical entropy , $H = \sum_{i=1}^{m} (\mathrm{freq}(i)/m) \cdot \log(m/\mathrm{freq}(i))$

# Problems of Interest

**Class A:**

- Median
- **Key question:** Want $s = O(\log n)$; then $p = ??$
  - Dates back to first "data streams" paper [Munro-Paterson'78]

**Class B:**

- Distinct elements , $F_0$
- Frequency moments , $F_k = \sum_{i=1}^{m} \mathrm{freq}(i)^k$
- Empirical entropy , $H = \sum_{i=1}^{m} (\mathrm{freq}(i)/m) \cdot \log(m/\mathrm{freq}(i))$
- **Key question:** Want $\varepsilon$-approx; then $s = ??$
  - One-pass: $\widetilde{O}(\varepsilon^{-2})$, $\widetilde{\Omega}(\varepsilon^{-2})$ [BarYossef-J.-K.-S.-T.'02]; [Woodruff'04]
  - Dependence of $s$ on $n$: [A-M-S'96]; [C.-Khot-Sun'03]; [Gronemeier'09]

# Our Results (Answering the Key Questions)

**Class A: Median** [C.-Cormode-McGregor'08]

- Achieving $s = O(\log n)$ requires $p = \Omega(\log n)$
- If tokens randomly ordered, requires $p = \Omega(\log \log n)$

- Above lower bounds are tight [Guha-McGregor'07]

# Our Results (Answering the Key Questions)

**Class A: Median** [C.-Cormode-McGregor'08]

- Achieving $s = O(\log n)$ requires $p = \Omega(\log n)$
- If tokens randomly ordered, requires $p = \Omega(\log \log n)$
  - Specifically: $s \approx \Omega(n^{1/p})$ $\left[\Omega(n^{2^{-p}})\right]$ for adversarial [random] order
- Above lower bounds are tight [Guha-McGregor'07]

# Our Results (Answering the Key Questions)

**Class A: Median** [C.-Cormode-McGregor'08]

- Achieving $s = O(\log n)$ requires $p = \Omega(\log n)$
- If tokens randomly ordered, requires $p = \Omega(\log \log n)$
  - Specifically: $s \approx \Omega(n^{1/p})$ $\left[\Omega(n^{2^{-p}})\right]$ for adversarial [random] order
- Above lower bounds are tight [Guha-McGregor'07]

---

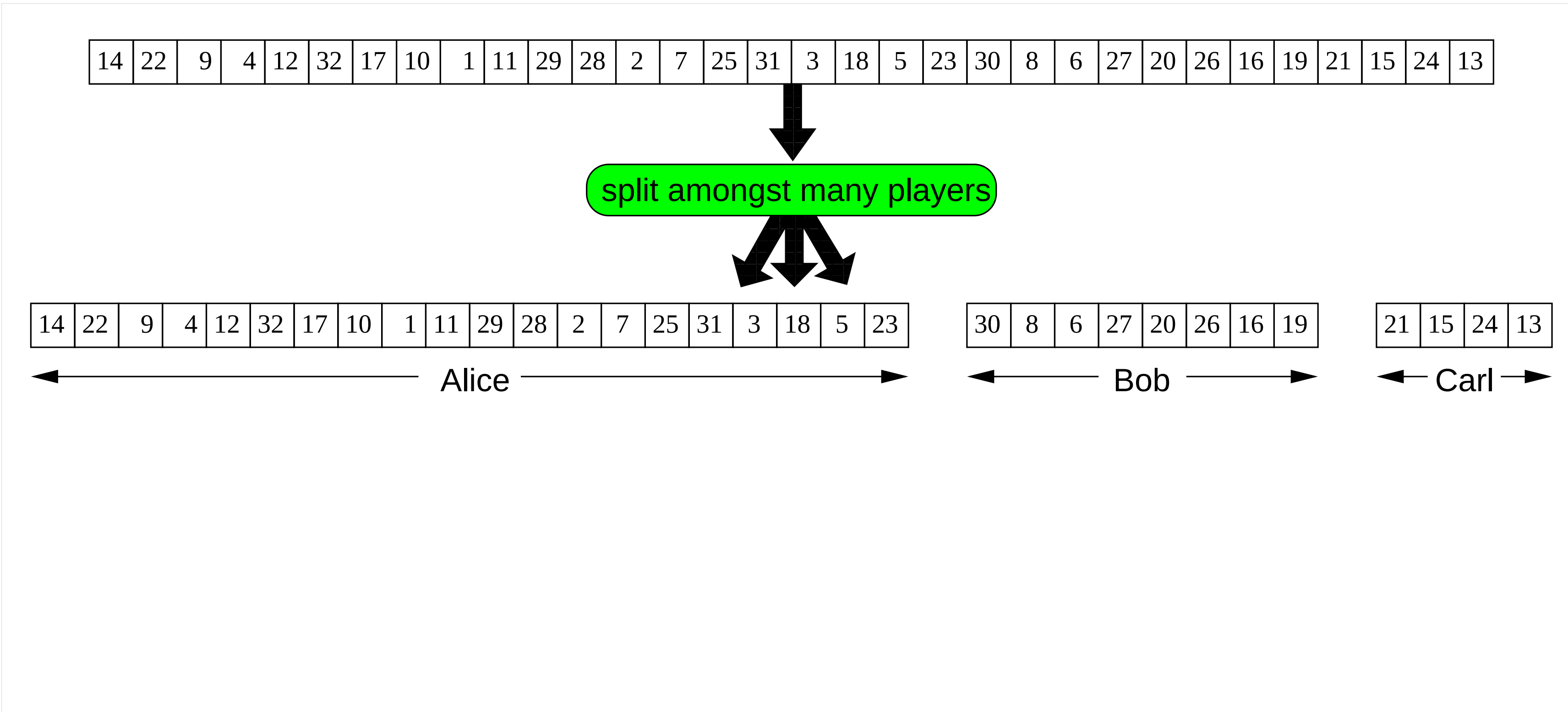
**Class B: Distinct elements** [Brody-C.'09]

- Need $s = \Omega(1/\varepsilon^2)$ space for any $p = O(1)$
  - Specifically: $s = \widetilde{\Omega}(1/(\varepsilon^2 p^2))$ [Brody-C.-Regev-Vidick-deWolf'10]
- Holds under random order, and even random data
- Matching upper bound, even with one pass and adversarial data

# Method: Reduce from Communication Complexity

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 22 | 9 | 4 | 12 | 32 | 17 | 10 | 1 | 11 | 29 | 28 | 2 | 7 | 25 | 31 | 3 | 18 | 5 | 23 | 30 | 8 | 6 | 27 | 20 | 26 | 16 | 19 | 21 | 15 | 24 | 13 |

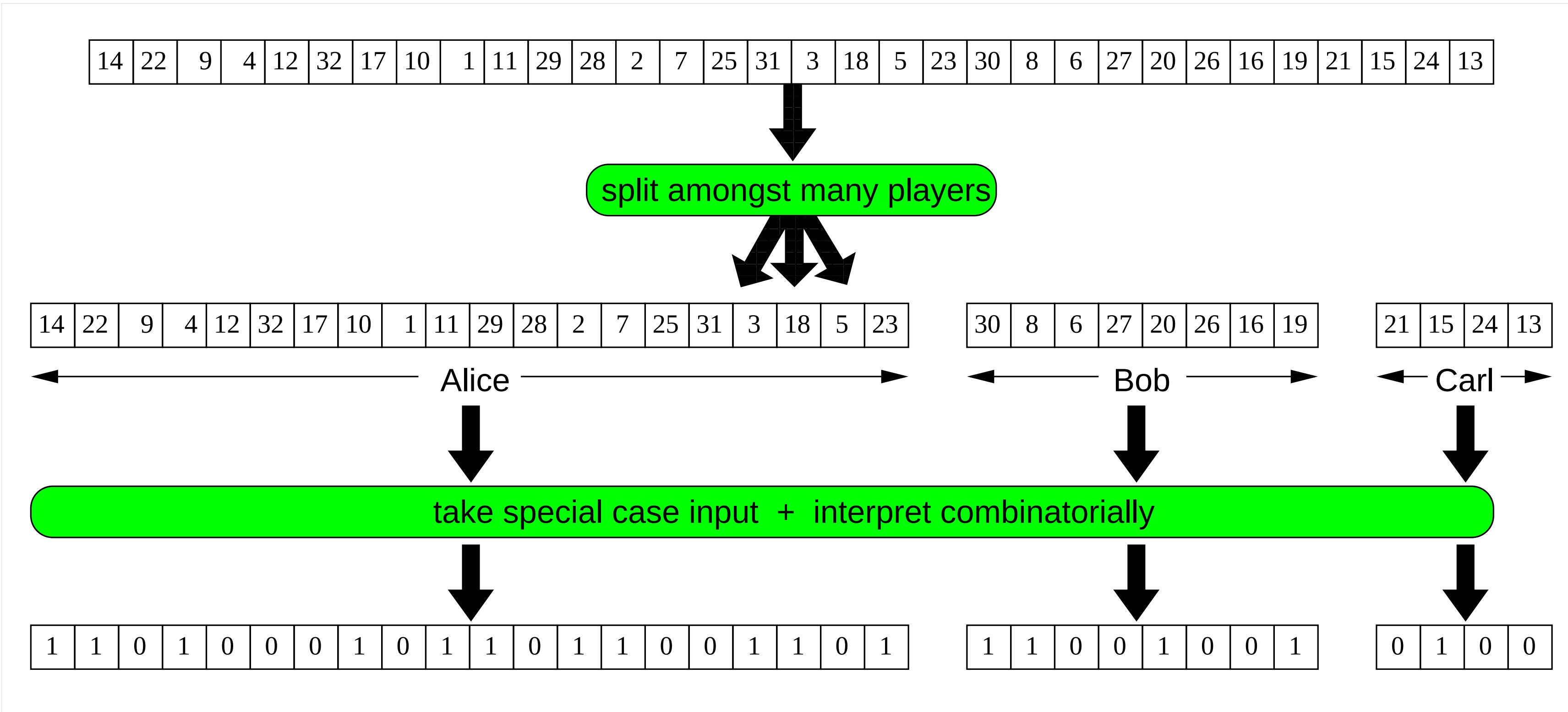
# Communication vs Data Stream




$p$-pass streaming algorithm $\Longrightarrow$ $\Theta(p)$-round communication protocol

messages $=$ memory contents of streaming algorithm

# Communication vs Data Stream




$p$-pass streaming algorithm $\Longrightarrow$ $\Theta(p)$-round communication protocol

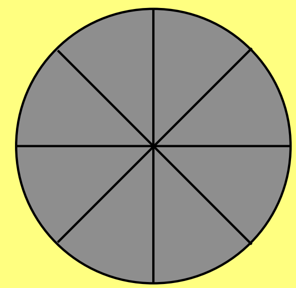
messages $=$ memory contents of streaming algorithm

# The Round Elimination Paradigm

If there exists...

| | | | | |
|---|---|---|---|---|
| Round 1: | **A** **msg1** | **B** **msg1** | **C** **msg1** | **D** **msg1** |
| Round 2: | **A** **msg2** | **B** **msg2** | **C** **msg2** | **D** **msg2** |
| Round 3: | **A** **msg3** | **B** **msg3** | **C** **msg3** | **D** **msg3** |

Input:

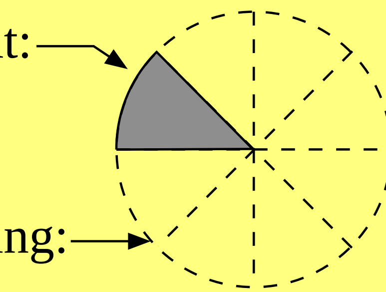


with short messages, then there exists...

| | | | | |
|---|---|---|---|---|
| Round 2: | **A** **msg2** | **B** **msg2** | **C** **msg2** | **D** **msg2** |
| Round 3: | **A** **msg3** | **B** **msg3** | **C** **msg3** | **D** **msg3** |

Input:

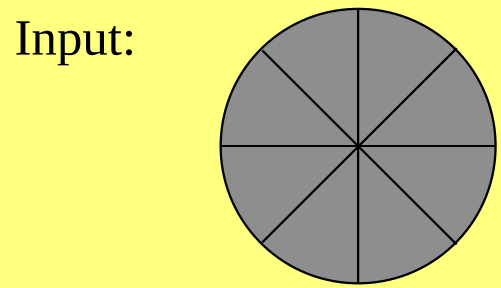
Padding:



# The Round Elimination Paradigm

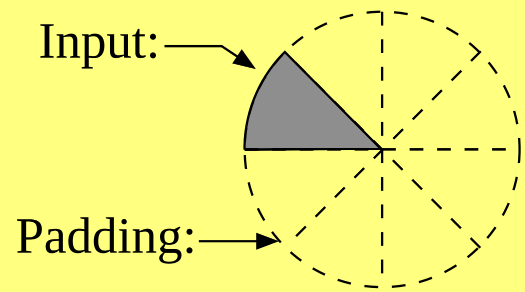
If there exists...

| | | | | |
|---|---|---|---|---|
| Round 1: | **A** msg1 | **B** msg1 | **C** msg1 | **D** msg1 |
| Round 2: | **A** msg2 | **B** msg2 | **C** msg2 | **D** msg2 |
| Round 3: | **A** msg3 | **B** msg3 | **C** msg3 | **D** msg3 |




with short messages, then there exists...

| | | | | |
|---|---|---|---|---|
| Round 2: | **A** msg2 | **B** msg2 | **C** msg2 | **D** msg2 |
| Round 3: | **A** msg3 | **B** msg3 | **C** msg3 | **D** msg3 |




Eventually, if original protocol too short,

then $0$-round protocol for a nontrivial problem $\implies$ Contradiction

# Class A: Median

## Tree Pointer Jumping

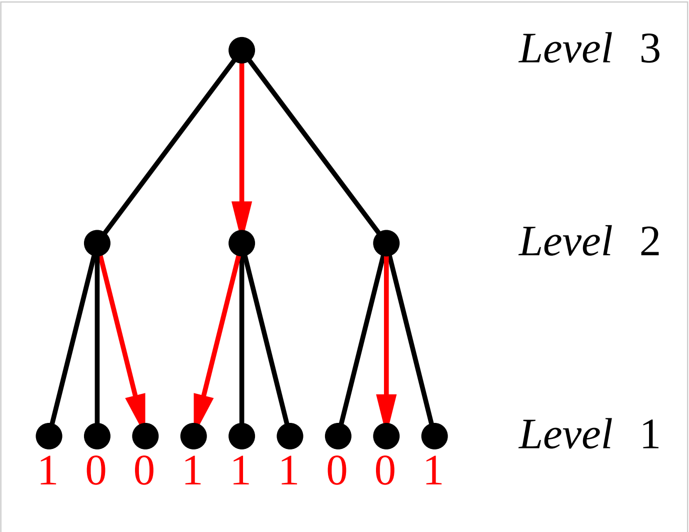
Complete $k$-level $t$-ary tree $T$

Input $\phi : V(T) \to [t]$ with $\phi(\text{leaf}) \in \{0,1\}$

Player $i$ knows $\phi$ at level $i$

$$g_\phi(v) \;:=\; \begin{cases} \phi(v)\text{-th child of } v, & \text{if } v \text{ internal} \\ \phi(v), & \text{if } v \text{ leaf} \end{cases}$$

Desired output $= g_\phi(g_\phi(\cdots g_\phi(\text{root})\cdots))$




Model: $k-1$ rounds of communication

Each round: (Plr 1, Plr 2, ..., Plr $k$)

Call this $\text{TPJ}_{k,t}$

# Weight-Based TPJ

**Theorem:** For uniform random input, $\frac{1}{3}$-error, $\mathrm{CC}^p(\textsc{tpj}_{p+1,t}) = \Omega(t/p^2)$

Contrast: $\mathrm{D}^p(\textsc{tpj}_{p+1,t}) = O(t)$ and $\mathrm{D}^{p+1}(\textsc{tpj}_{p+1,t}) = O(p \log t)$

# Weight-Based TPJ

**Theorem:** For uniform random input, $\frac{1}{3}$-error, $\mathrm{CC}^p(\textsc{tpj}_{p+1,t}) = \Omega(t/p^2)$

Contrast: $\mathrm{D}^p(\textsc{tpj}_{p+1,t}) = O(t)$ and $\mathrm{D}^{p+1}(\textsc{tpj}_{p+1,t}) = O(p \log t)$

Actually, use a variant $\textsc{w-tpj}$ (weight-based):

- Input specifies $x_v \in \{0,1\}^{\ell_v}$ with $\phi(v) = \frac{t}{2} + \mathrm{bias}(|x_v|)$
- Lengths $\ell_v = t^{\mathrm{level}(v)-1}$

Median lower bound: reduction from $\textsc{w-tpj}$ (next slide)

# Weight-Based TPJ

**Theorem:** For uniform random input, $\frac{1}{3}$-error, $\mathrm{CC}^p(\textsc{tpj}_{p+1,t}) = \Omega(t/p^2)$

Contrast: $\mathrm{D}^p(\textsc{tpj}_{p+1,t}) = O(t)$ and $\mathrm{D}^{p+1}(\textsc{tpj}_{p+1,t}) = O(p \log t)$

Actually, use a variant $\textsc{w-tpj}$ (weight-based):

- Input specifies $x_v \in \{0,1\}^{\ell_v}$ with $\phi(v) = \frac{t}{2} + \mathrm{bias}(|x_v|)$
- Lengths $\ell_v = t^{\mathrm{level}(v)-1}$

Median lower bound: reduction from $\textsc{w-tpj}$ (next slide)

**Robust communication complexity:** Above CC lower bound still holds when input bits allocated amongst players at random.

Relevant theory developed in [C.-Cormode-McGregor'08]

# Weight-Based TPJ

**Theorem:** For uniform random input, $\frac{1}{3}$-error, $\mathrm{CC}^p(\textsc{tpj}_{p+1,t}) = \Omega(t/p^2)$

Contrast: $\mathrm{D}^p(\textsc{tpj}_{p+1,t}) = O(t)$ and $\mathrm{D}^{p+1}(\textsc{tpj}_{p+1,t}) = O(p \log t)$

Actually, use a variant $\textsc{w-tpj}$ (weight-based):

- Input specifies $x_v \in \{0,1\}^{\ell_v}$ with $\phi(v) = \frac{t}{2} + \mathrm{bias}(|x_v|)$
- Lengths $\ell_v = t^{\mathrm{level}(v)-1}$
- For random order, $\ell_v \approx t^{2^{\mathrm{level}(v)-1}}$ (hence, smaller lower bound)

Median lower bound: reduction from $\textsc{w-tpj}$ (next slide)

**Robust communication complexity:** Above CC lower bound still holds when input bits allocated amongst players at random.

Relevant theory developed in [C.-Cormode-McGregor'08]

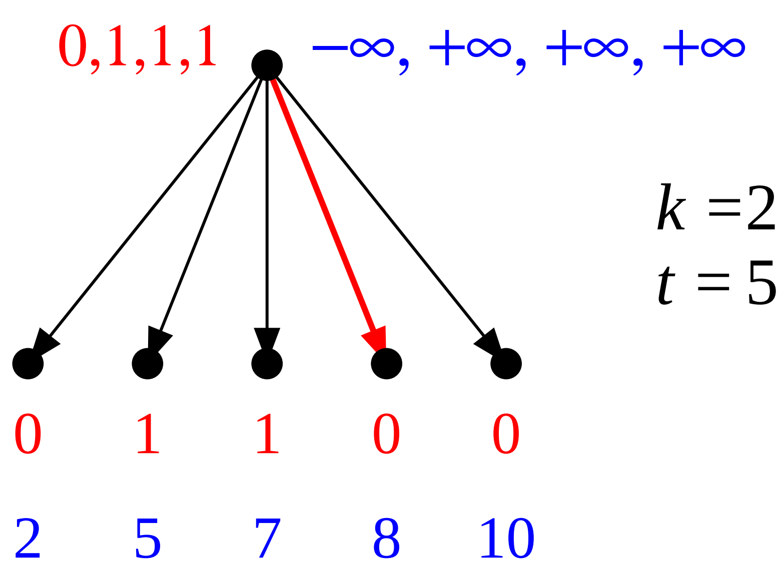
## From TPJ to Median

Map each input bit to an integer: $x \longmapsto$ multiset $S_x$, s.t.

$$\text{W-TPJ}(x) = \text{LSB}(\text{median}(S_x))$$

Basic idea, for $k = 2$ levels:

- At level 2, $0 \mapsto -\infty$ (min value) and $1 \mapsto +\infty$ (max value)
- At level 1, $x_i \mapsto 2i + x_i$ (for $i$th leaf)



# Class B: Distinct Elements

# The Gap-Hamming-Distance Problem

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

Output:

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

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

# The Reductions

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

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x =$ | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| $\sigma$ : | (1,0) | (2,1) | (3,0) | (4,0) | (5,1) | (6,0) | (7,0) | (8,0) | (9,0) | (10,0) | (11,0) | (12,1) |
| $y =$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| $\tau$ : | (1,0) | (2,0) | (3,0) | (4,0) | (5,0) | (6,0) | (7,0) | (8,0) | (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 $\varepsilon = \frac{1}{\sqrt{n}}$.

## State of Play, Jan. 2009

Using one round = one message...

Previous results [Indyk-Woodruff'03], [Woodruff'04], [C.-Cormode-McGregor'07]:

- For one-round protocols, $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$
- Implies the $\widetilde{\Omega}(\varepsilon^{-2})$ streaming lower bounds

## State of Play, Jan. 2009

Using one round = one message...

Previous results [Indyk-Woodruff'03], [Woodruff'04], [C.-Cormode-McGregor'07]:

- For one-round protocols, $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$
- Implies the $\widetilde{\Omega}(\varepsilon^{-2})$ streaming lower bounds

Key open questions:

- What is the two-way randomized complexity $\mathrm{R}(\textsc{ghd})$?
- Better algorithm for Distinct Elements (or $F_k$, or $H$) using two passes?

# State of Play, Jan. 2009

Using one round = one message...

Previous results [Indyk-Woodruff'03], [Woodruff'04], [C.-Cormode-McGregor'07]:

- For one-round protocols, $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$
- Implies the $\widetilde{\Omega}(\varepsilon^{-2})$ streaming lower bounds

Key open questions:

- What is the two-way randomized complexity $\mathrm{R}(\textsc{ghd})$?
- Better algorithm for Distinct Elements (or $F_k$, or $H$) using two passes?

# New Results

**Summer Thm:** $\mathrm{R}^{O(1)}(\textsc{ghd}) = \Omega(n)$; i.e., $O(1)$ rounds/passes no better

## State of Play, Jan. 2009

Using one round = one message...

Previous results [Indyk-Woodruff'03], [Woodruff'04], [C.-Cormode-McGregor'07]:

- For one-round protocols, $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$
- Implies the $\widetilde{\Omega}(\varepsilon^{-2})$ streaming lower bounds

Key open questions:

- What is the two-way randomized complexity $\mathrm{R}(\textsc{ghd})$?
- Better algorithm for Distinct Elements (or $F_k$, or $H$) using two passes?

## New Results

**Summer Thm:** $\mathrm{R}^{O(1)}(\textsc{ghd}) = \Omega(n)$; i.e., $O(1)$ rounds/passes no better
**Winter Thm:** $\mathrm{R}^{p}(\textsc{ghd}) = \widetilde{\Omega}(n/p^2)$; previously was $\widetilde{\Omega}(n/2^{O(p^2)})$

Remark: These hold under uniform input distribution

# A Simplification

Will prove distributional lower bound under uniform dist

In this setting, may as well work with threshold version, $\textsc{thd}$

- $\textsc{thd}(x, y) = 1$ if $\Delta(x, y) \geq \frac{n}{2}$
- $\textsc{thd}(x, y) = 0$ if $\Delta(x, y) < \frac{n}{2}$

# Round Elimination V1.0: Subcube Lifting

First message constant on large set:

$$\left.\begin{array}{ccccccc} \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \end{array}\right\} \begin{array}{l} 2^{0.99n} \\ \text{points} \end{array}$$

# Round Elimination V1.0: Subcube Lifting

First message constant on large set:

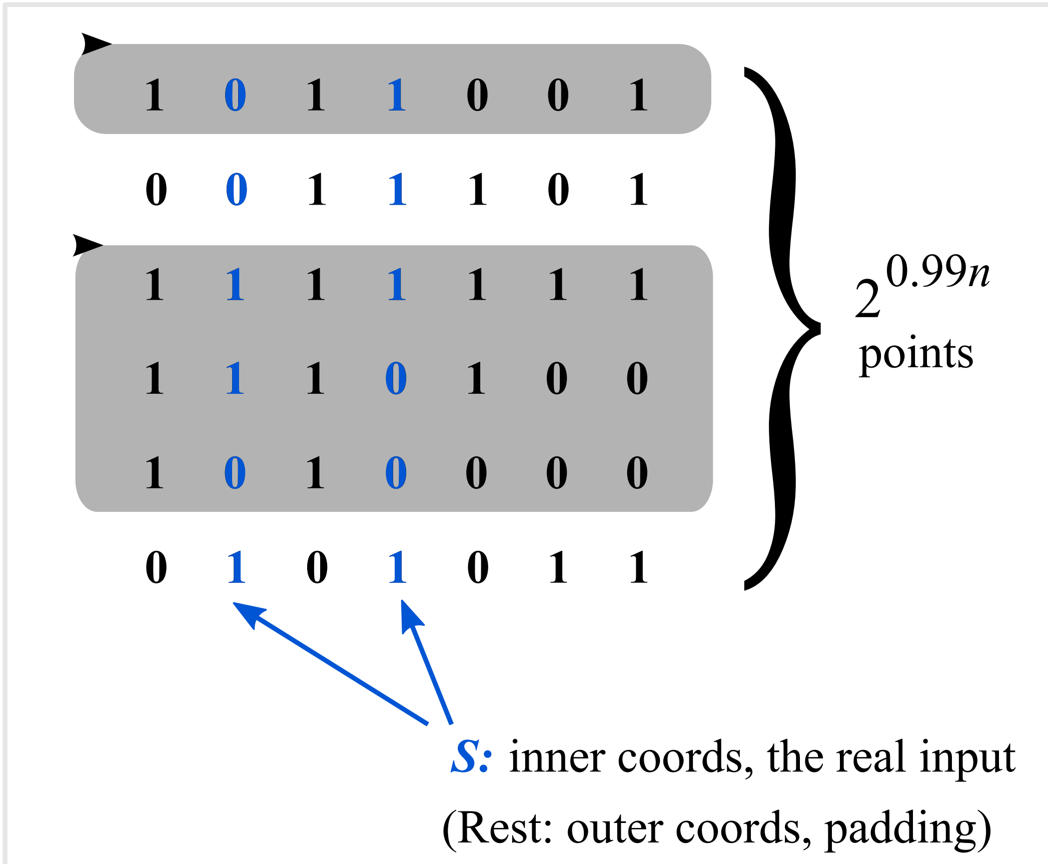
$$\left.\begin{matrix}
1 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 1
\end{matrix}\right\} 2^{0.99n} \text{ points}$$

***S*:** inner coords, the real input

(Rest: outer coords, padding)

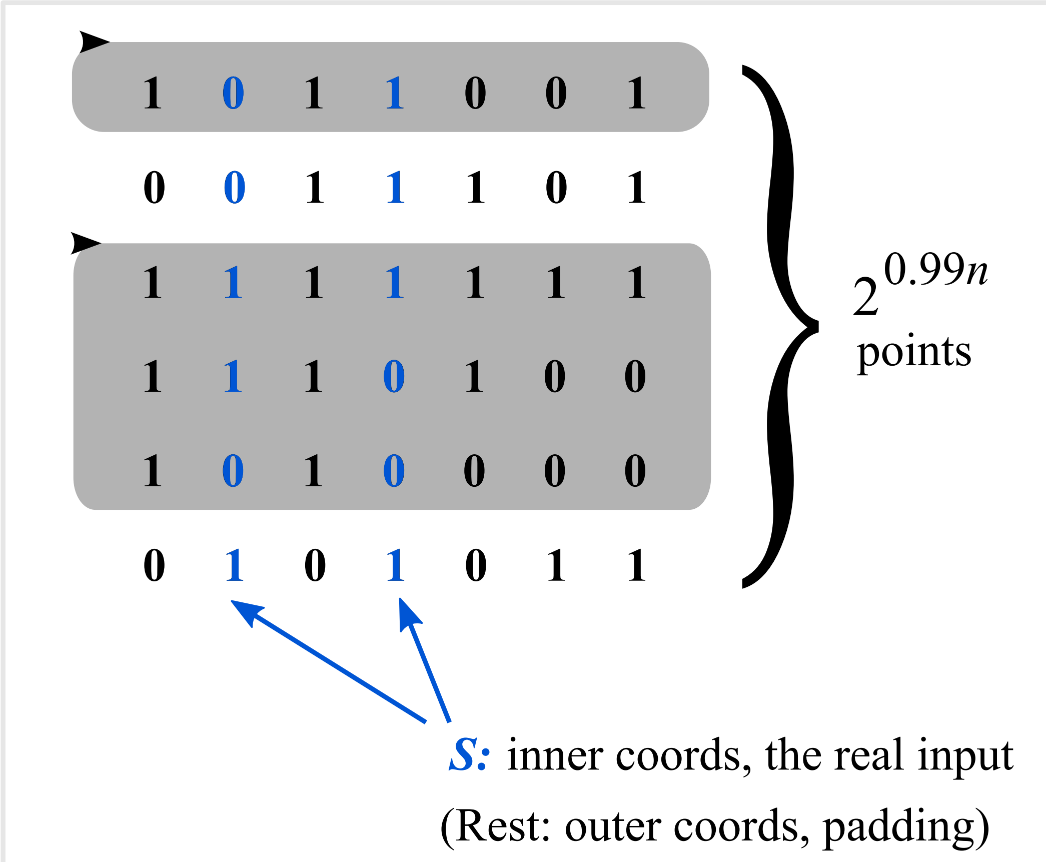
# Round Elimination V1.0: Subcube Lifting

First message constant on large set:



# Round Elimination V1.0: 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]

## Subcube Lifting: Wasteful?

- Each step: dimension $n \longrightarrow n/3$
- Inherently, can eliminate at most $O(\log n)$ rounds

  In fact, get $\mathrm{R}^p(\textsc{ghd}) = n/2^{O(p^2)}$
- Solved long-standing open problem (IITK 2006 list)... happy?

## Subcube Lifting: Wasteful?

- Each step: dimension $n \longrightarrow n/3$
- Inherently, can eliminate at most $O(\log n)$ rounds

  In fact, get $\mathrm{R}^p(\textsc{ghd}) = n/2^{O(p^2)}$
- Solved long-standing open problem (IITK 2006 list)... happy?

## Rethinking Round Elimination

- Crux: delete first round, solve simpler instance
- Simpler need not mean smaller!

## Subcube Lifting: Wasteful?

- Each step: dimension $n \longrightarrow n/3$
- Inherently, can eliminate at most $O(\log n)$ rounds

  In fact, get $\mathrm{R}^p(\textsc{ghd}) = n/2^{O(p^2)}$
- Solved long-standing open problem (IITK 2006 list)... happy?

## Rethinking Round Elimination

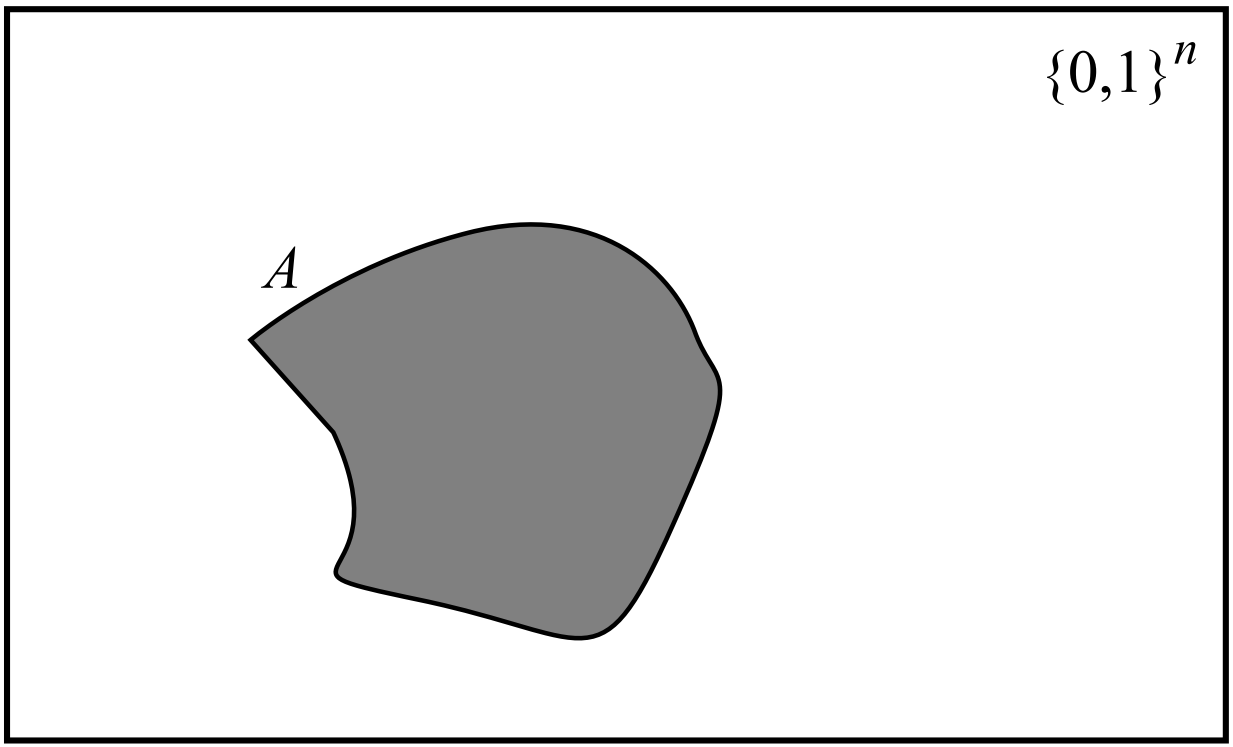
- Crux: delete first round, solve simpler instance
- Simpler need not mean smaller!

  E.g., could mean increased error prob.

# Round Elimination V2.0: Geometric Perturbation

Max message size = $cn$

First message constant over set $A$ of size $2^{n-cn}$

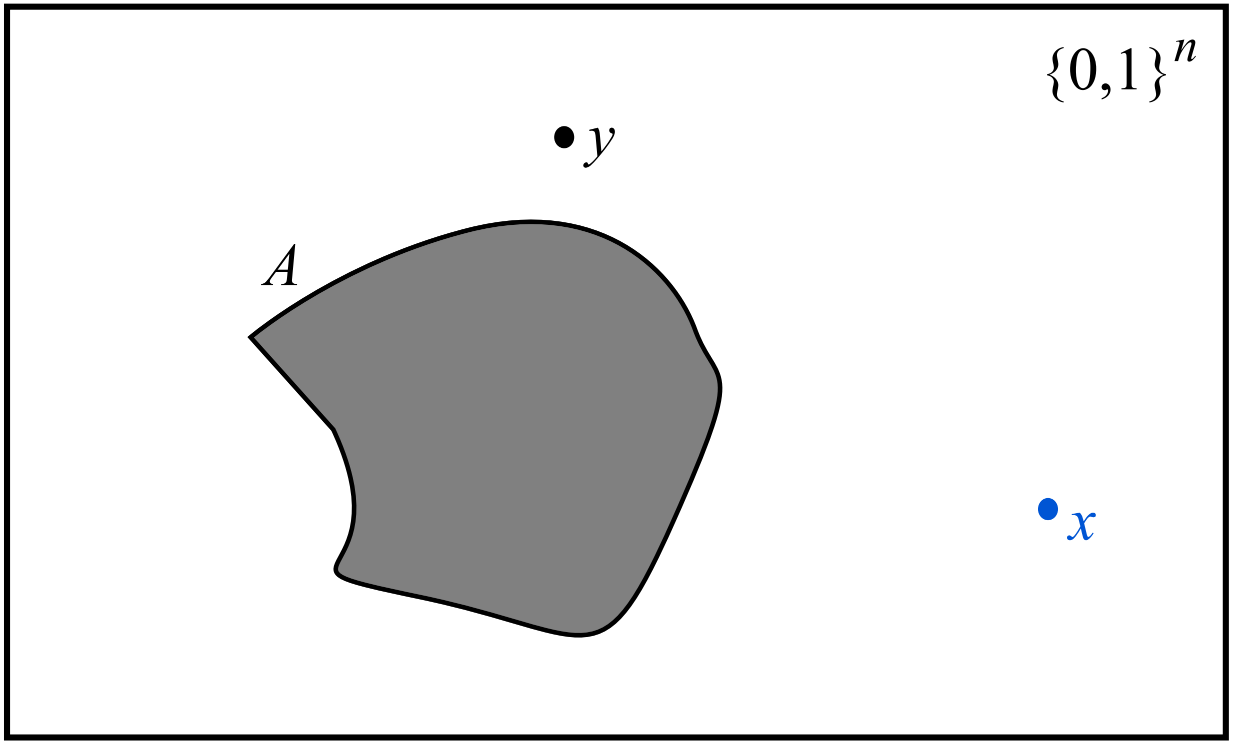



Alice: replace $x$ with $z = \mathsf{NearestNeighbour}(x, A)$

## Round Elimination V2.0: Geometric Perturbation

Max message size $= cn$

First message constant over set $A$ of size $2^{n-cn}$

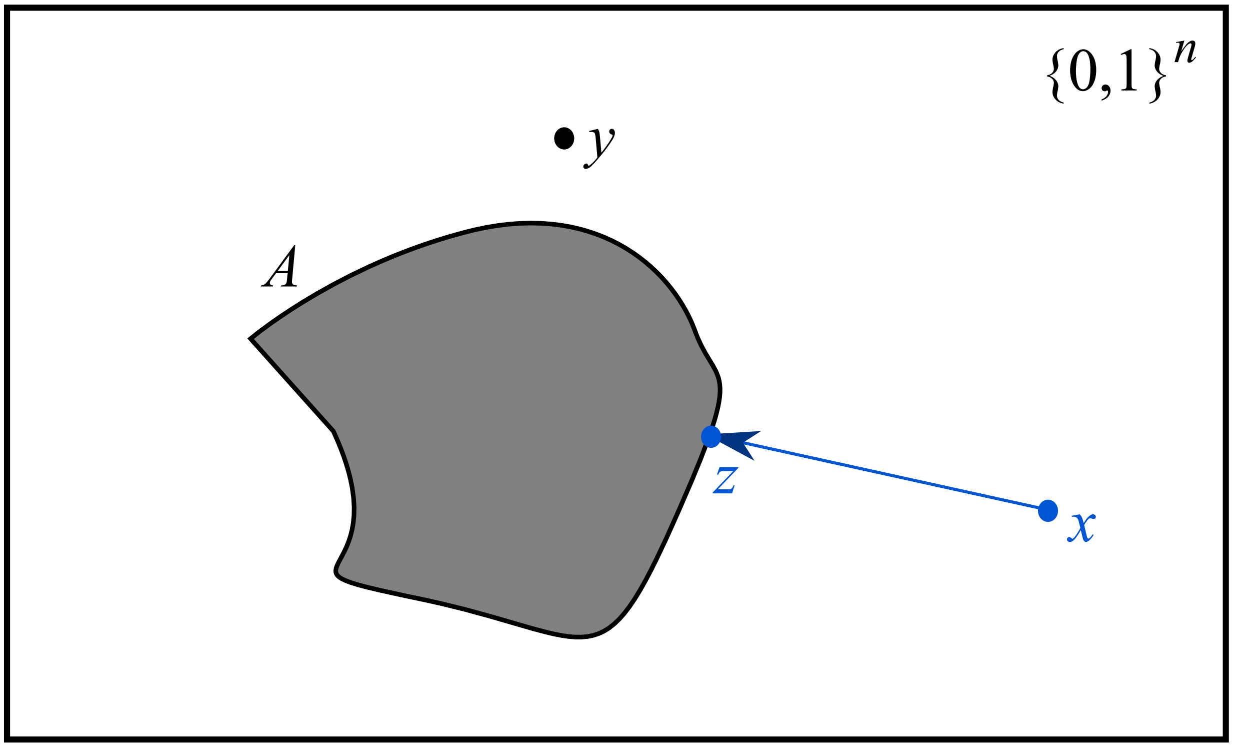



Alice: replace $x$ with $z = \text{NearestNeighbour}(x, A)$

# Round Elimination V2.0: Geometric Perturbation

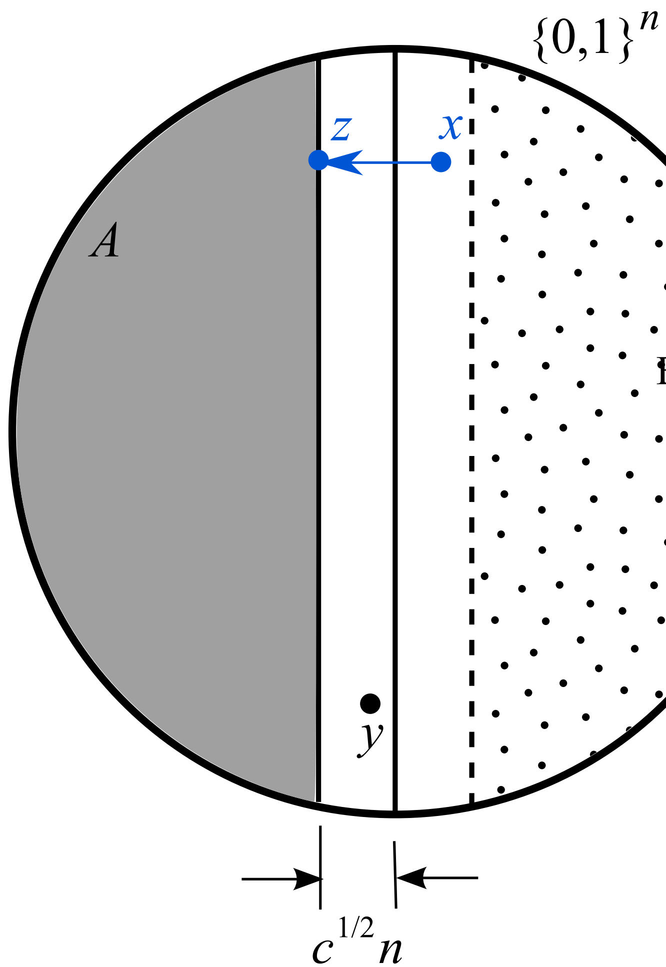
Max message size $= cn$

First message constant over set $A$ of size $2^{n-cn}$



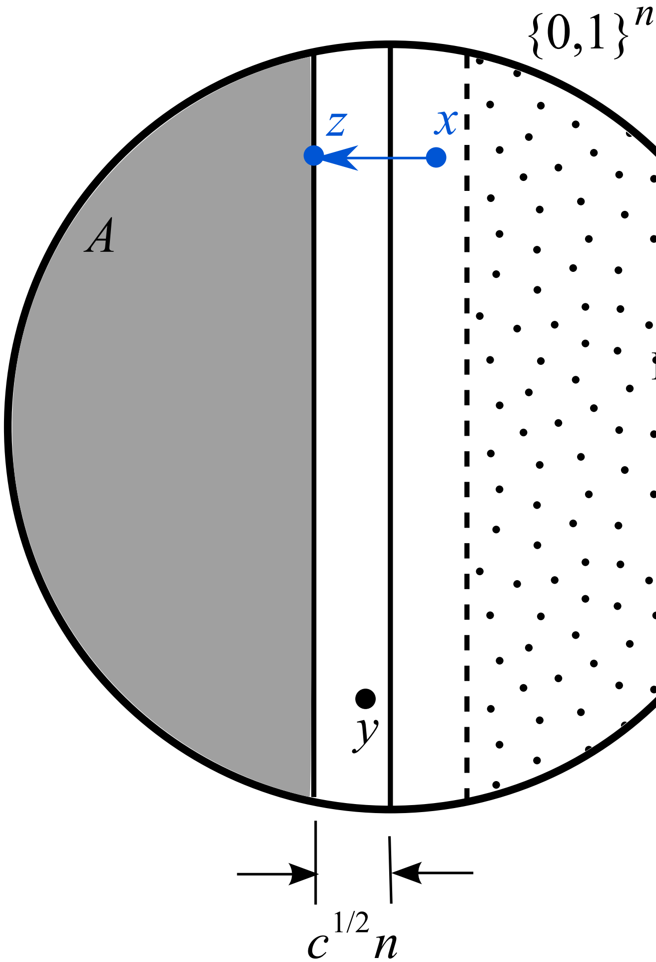

Alice: replace $x$ with $z = \mathsf{NearestNeighbour}(x, A)$

# Geometric Perturbation: A Better Picture




$\Pr[A] = 2^{-cn}$ ...... thus, w.h.p., $\Delta(x, z) \leq (\sqrt{cn}$ std devs$) = \sqrt{c} \cdot n$

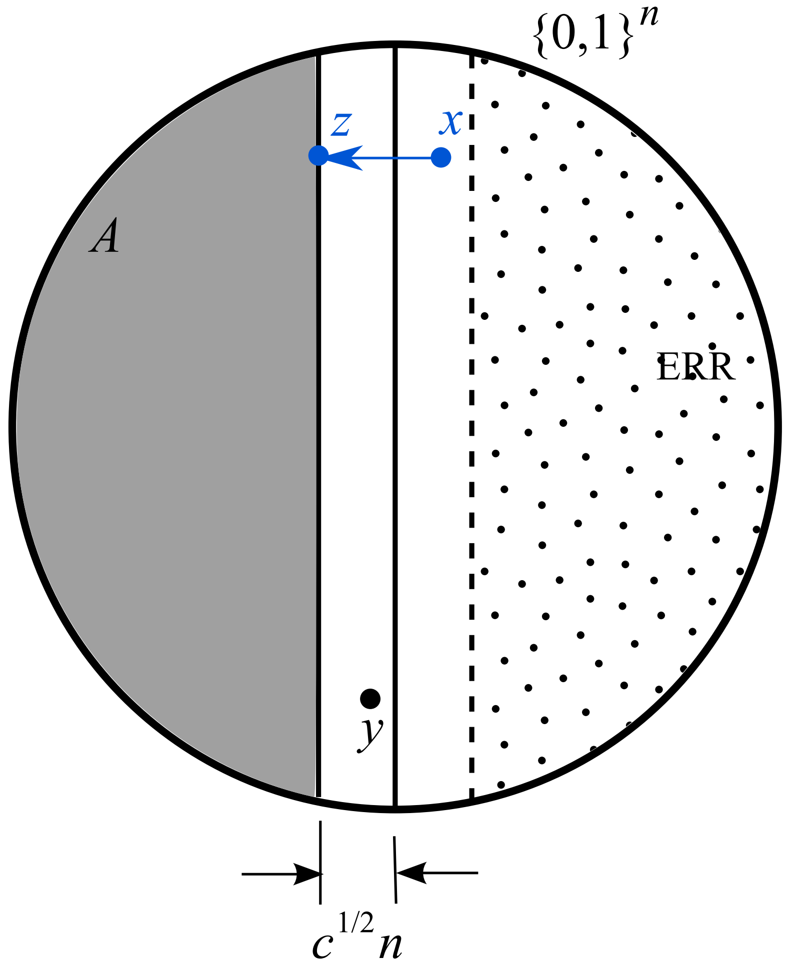
# Geometric Perturbation: A Better Picture




$\Pr[A] = 2^{-cn}$ . . . . . . thus, w.h.p., $\Delta(x,z) \leq (\sqrt{cn}$ std devs$) = \sqrt{c} \cdot n$

Assumed $A$ is Hamming ball . . . . . .

# Geometric Perturbation: A Better Picture




$\Pr[A] = 2^{-cn}$ . . . . . . thus, w.h.p., $\Delta(x, z) \le (\sqrt{cn}$ std devs$) = \sqrt{c} \cdot n$

Assumed $A$ is Hamming ball . . . . . . *that's indeed the worst case* [Harper'66]

# Round Elimination: Analysis

Alice: $x \in_R \{0,1\}^n \longmapsto z \sim \;??$; Bob: $y \in_R \{0,1\}^n$

Why does the shorter protocol work?

# Round Elimination: Analysis

Alice: $x \in_R \{0,1\}^n \longmapsto z \sim\ ??$; Bob: $y \in_R \{0,1\}^n$

Why does the shorter protocol work?

How can it fail? Two ways:

- $\mathcal{E}_1$: $\Delta(x, y)$ too close to $n/2$
- $\mathcal{E}_2$: Not near threshold, but $\textsc{thd}(x, y) \neq \textsc{thd}(z, y)$

# Round Elimination: Analysis

Alice: $x \in_R \{0,1\}^n \longmapsto z \sim\ ??$; Bob: $y \in_R \{0,1\}^n$

Why does the shorter protocol work?

How can it fail? Two ways:

- $\mathcal{E}_1$: $\Delta(x,y)$ too close to $n/2$
- $\mathcal{E}_2$: Not near threshold, but $\textsc{thd}(x,y) \neq \textsc{thd}(z,y)$

Estimating the probabilities:

- $\mathcal{E}_1$: "anticoncentration" of Binomial dist

## Round Elimination: Analysis

Alice: $x \in_R \{0,1\}^n \longmapsto z \sim\ ??$; Bob: $y \in_R \{0,1\}^n$

Why does the shorter protocol work?

How can it fail? Two ways:

- $\mathcal{E}_1$: $\Delta(x,y)$ too close to $n/2$
- $\mathcal{E}_2$: Not near threshold, but $\text{THD}(x,y) \neq \text{THD}(z,y)$

Estimating the probabilities:

- $\mathcal{E}_1$: "anticoncentration" of Binomial dist

$$\Pr\left[\ |\Delta(x,y) - n/2| < \delta\sqrt{n}\ \right] \ \leq\ \delta$$

# Round Elimination: Analysis

Alice: $x \in_R \{0,1\}^n \longmapsto z \sim\ ??$; Bob: $y \in_R \{0,1\}^n$

Why does the shorter protocol work?

How can it fail? Two ways:

- $\mathcal{E}_1$: $\Delta(x,y)$ too close to $n/2$
- $\mathcal{E}_2$: Not near threshold, but $\textsc{thd}(x,y) \neq \textsc{thd}(z,y)$

Estimating the probabilities:

- $\mathcal{E}_1$: "anticoncentration" of Binomial dist

  $$\Pr\left[\ |\Delta(x,y) - n/2| < \delta\sqrt{n}\ \right] \ \le\ \delta$$

- $\mathcal{E}_2$: shift to assume $x = \vec{0}$

  $$\Pr\left[\ |y| < n/2 - \delta\sqrt{n}\ \wedge\ |y \oplus z| > n/2\ \right] \ \le\ ??$$

## Round Elimination: Analysis

Alice: $x \in_R \{0,1\}^n \longmapsto z \sim\ ??$; Bob: $y \in_R \{0,1\}^n$

Why does the shorter protocol work?

How can it fail? Two ways:

- $\mathcal{E}_1$: $\Delta(x,y)$ too close to $n/2$
- $\mathcal{E}_2$: Not near threshold, but $\text{THD}(x,y) \neq \text{THD}(z,y)$

Estimating the probabilities:

- $\mathcal{E}_1$: "anticoncentration" of Binomial dist

  $$\Pr\left[\ |\Delta(x,y) - n/2| < \delta\sqrt{n}\ \right] \ \le\ \delta$$

- $\mathcal{E}_2$: shift to assume $x = \vec{0}$

  $$\Pr\left[\ |y| < n/2 - \delta\sqrt{n}\ \wedge\ |y \oplus z| > n/2\ \right] \ \le\ ??$$

  Recall: $|z| = \Delta(x,z) \le \sqrt{c} \cdot n$, w.h.p.

## Switcheroo

Fixed $y \in \{0,1\}^n$, with $|y| < n/2 - \delta\sqrt{n}$

Random $z \in_R \{0,1\}^n$, with $|z| \le \sqrt{c} \cdot n$

Recall: first message length $= cn$

$$\Pr\left[\ |y \oplus z| > n/2\ \right] \ \le\ ??$$

## Switcheroo

Fixed $y \in \{0,1\}^n$, with $|y| < n/2 - \delta\sqrt{n}$

Random $z \in_R \{0,1\}^n$, with $|z| \le \sqrt{c} \cdot n$

Recall: first message length $= cn$

$\Pr\left[\ |y \oplus z| > n/2\ \right] \ \le\ ??$

Random coordinate flipping: $y \longmapsto y \oplus z$

## Switcheroo

Fixed $y \in \{0,1\}^n$, with $|y| < n/2 - \delta\sqrt{n}$

Random $z \in_R \{0,1\}^n$, with $|z| \le \sqrt{c} \cdot n$

Recall: first message length $= cn$

$\Pr\left[\ |y \oplus z| > n/2\ \right]\ \le\ ??$

Random coordinate flipping: $y \longmapsto y \oplus z$

Expect $|y|$ to change by about $\sqrt{\sqrt{c} \cdot n}$

## Switcheroo

Fixed $y \in \{0,1\}^n$, with $|y| < n/2 - \delta\sqrt{n}$

Random $z \in_R \{0,1\}^n$, with $|z| \le \sqrt{c} \cdot n$

Recall: first message length $= cn$

$\Pr\left[\ |y \oplus z| > n/2\ \right] \ \le\ ??$

Random coordinate flipping: $y \longmapsto y \oplus z$

Expect $|y|$ to change by about $\sqrt{\sqrt{c} \cdot n}$

W.h.p., change is no more than $c^{1/4}\sqrt{n \log p}$ [Hoeffding'63]

We're good if this $= \delta\sqrt{n}$, i.e., if $\delta = c^{1/4} \log^{1/2} p$

# Switcheroo

Fixed $y \in \{0,1\}^n$, with $|y| < n/2 - \delta\sqrt{n}$

Random $z \in_R \{0,1\}^n$, with $|z| \leq \sqrt{c} \cdot n$

Recall: first message length $= cn$

$\Pr\left[\ |y \oplus z| > n/2\ \right] \ \leq\ ??$

Random coordinate flipping: $y \longmapsto y \oplus z$

Expect $|y|$ to change by about $\sqrt{\sqrt{c} \cdot n}$

W.h.p., change is no more than $c^{1/4}\sqrt{n \log p}$ [Hoeffding'63]

We're good if this $= \delta\sqrt{n}$, i.e., if $\delta = c^{1/4} \log^{1/2} p$

Overall error $= \delta$+(tiny) $\approx c^{1/4} \log^{1/2} p$

# Round Elimination: Wrap-Up

- Killed a message of length $cn$, adding $c^{1/4}\log^{1/2} p$ to error
- Have to do this $p$ times
- Final error must be $\Omega(1)$, else contradiction

$$\implies pc^{1/4}\log^{1/2} p = \Omega(1)$$

$$\implies \text{(max comm)} = \Omega(n/p^4 \log^2 p)$$

[Brody-C.-Regev-Vidick-deWolf'10]

## Round Elimination: Wrap-Up

- Killed a message of length $cn$, adding $c^{1/4}\log^{1/2} p$ to error
- Have to do this $p$ times
- Final error must be $\Omega(1)$, else contradiction

$$\implies pc^{1/4}\log^{1/2} p = \Omega(1)$$

$$\implies \text{(max comm)} = \Omega(n/p^4 \log^2 p)$$

- Work on sphere, not Hamming cube: $\mathrm{R}^p(\textsc{ghd}) = \Omega(n/p^2 \log p)$

$$x \in \{0,1\}^n \quad \longmapsto \quad \widetilde{x} \in \left\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right\}^n$$

$$\textsc{ghd} \quad \longmapsto \quad \text{Gap-Inner-Product}$$

[Brody-C.-Regev-Vidick-deWolf'10]

## Why Did This Take So Long?

Multi-pass lower bounds for Distinct Elements and $F_k$ has been an important open question since at least 2003. Why did it remain open for so long?

# Why Did This Take So Long?

Multi-pass lower bounds for Distinct Elements and $F_k$ has been an important open question since at least 2003. Why did it remain open for so long?

Underlying communication problem thorny!

## Why Did This Take So Long?

Multi-pass lower bounds for Distinct Elements and $F_k$ has been an important open question since at least 2003. Why did it remain open for so long?

Underlying communication problem thorny! Resists the "usual" attacks:

- Rectangle-based methods (discrepancy/corruption)
- Approximate polynomial degree
- Pattern matrix, Factorization norms [Sherstov'08], [Linial-Shraibman'07]
- Information complexity [C.-Shi-Wirth-Yao'01], [BarYossef-J.-K.-S.'02]

# Why Did This Take So Long?

Multi-pass lower bounds for Distinct Elements and $F_k$ has been an important open question since at least 2003. Why did it remain open for so long?

Underlying communication problem thorny! Resists the "usual" attacks:

- Rectangle-based methods (discrepancy/corruption)

  Matrix has large near-monochromatic rectangles

- Approximate polynomial degree

- Pattern matrix, Factorization norms [Sherstov'08], [Linial-Shraibman'07]

- Information complexity [C.-Shi-Wirth-Yao'01], [BarYossef-J.-K.-S.'02]

# Why Did This Take So Long?

Multi-pass lower bounds for Distinct Elements and $F_k$ has been an important open question since at least 2003. Why did it remain open for so long?

Underlying communication problem thorny! Resists the "usual" attacks:

- Rectangle-based methods (discrepancy/corruption)

  Matrix has large near-monochromatic rectangles

- Approximate polynomial degree

  Underlying predicate has approx degree $\widetilde{O}(\sqrt{n})$

- Pattern matrix, Factorization norms [Sherstov'08], [Linial-Shraibman'07]

- Information complexity [C.-Shi-Wirth-Yao'01], [BarYossef-J.-K.-S.'02]

## Why Did This Take So Long?

Multi-pass lower bounds for Distinct Elements and $F_k$ has been an important open question since at least 2003. Why did it remain open for so long?

Underlying communication problem thorny! Resists the "usual" attacks:

- Rectangle-based methods (discrepancy/corruption)

  Matrix has large near-monochromatic rectangles

- Approximate polynomial degree

  Underlying predicate has approx degree $\widetilde{O}(\sqrt{n})$

- Pattern matrix, Factorization norms [Sherstov'08], [Linial-Shraibman'07]

  Quantum communication upper bound $O(\sqrt{n}\log n)$

- Information complexity [C.-Shi-Wirth-Yao'01], [BarYossef-J.-K.-S.'02]

## Why Did This Take So Long?

Multi-pass lower bounds for Distinct Elements and $F_k$ has been an important open question since at least 2003. Why did it remain open for so long?

Underlying communication problem thorny! Resists the "usual" attacks:

- Rectangle-based methods (discrepancy/corruption)

  Matrix has large near-monochromatic rectangles

- Approximate polynomial degree

  Underlying predicate has approx degree $\widetilde{O}(\sqrt{n})$

- Pattern matrix, Factorization norms [Sherstov'08], [Linial-Shraibman'07]

  Quantum communication upper bound $O(\sqrt{n}\log n)$

- Information complexity [C.-Shi-Wirth-Yao'01], [BarYossef-J.-K.-S.'02]

  Hmm! Can't see a concrete obstacle

## Final Remarks

Summary:

1. Round elimination is a great paradigm for proving lower bounds (especially when you don't over-define it).
2. Gives clean proofs
3. Cases in point: Multi-player Pointer Jumping, Gap-Hamming-Distance
4. Data stream consequences

# Final Remarks

Summary:

1. Round elimination is a great paradigm for proving lower bounds (especially when you don't over-define it).
2. Gives clean proofs
3. Cases in point: Multi-player Pointer Jumping, Gap-Hamming-Distance
4. Data stream consequences

Open "problems":

1. Understand communication complexity of "gap problems" better... get further streaming results.
2. Apply round elimination to *your* favourite problem.

## Breaking News

Very recently, Oded Regev proved a remarkable new "correlation inequality" for Gaussian distributions.

This, plus a new generalization of the rectangle method, implies that $\mathrm{R}(\textsc{ghd}) = \Omega(n)$.