Lower Bounds for Gap-Hamming-Distance and Consequences for Data Stream Algorithms

Amit Chakrabarti

(Joint work with Joshua Brody)

Dartmouth College

DIMACS/DyDAn Workshop, March 2009
Status of Certain Streaming Problems, Jan 2009

Problems:

- Distinct elements
- Frequency moments
- Empirical entropy

One-pass, randomized, $\varepsilon$-approximate:

- Space upper bound: $\tilde{O}(\varepsilon^{-2})$
- Space lower bound: $\tilde{\Omega}(\varepsilon^{-2})$

Do multiple passes help?
Problems:

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

One-pass, randomized, $\varepsilon$-approximate: $\left| \frac{\text{output}}{\text{answer}} - 1 \right| \leq \varepsilon$

- Space upper bound: $\tilde{O}(\varepsilon^{-2})$
- Space lower bound: $\tilde{\Omega}(\varepsilon^{-2})$

Do multiple passes help?
Gap-Hamming Lower Bound

Status of Certain Streaming Problems, Jan 2009

Problems:

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

One-pass, randomized, $\varepsilon$-approximate: $|\frac{\text{output}}{\text{answer}} - 1| \leq \varepsilon$

- Space upper bound: $\tilde{O}(\varepsilon^{-2})$
- Space lower bound: $\tilde{\Omega}(\varepsilon^{-2})$

Do multiple passes help? If not, why not?

Amit Chakrabarti
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 $\Delta(x, y) < \frac{n}{2} - \sqrt{n}$

Problem: Design randomized, constant error protocol to solve this

Cost: Worst case number of bits communicated

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

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

\[
x = \begin{bmatrix}
0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
\sigma: \langle (1,0), (2,1), (3,0), (4,0), (5,1), (6,0), (9,0), (8,0), (9,0), (10,0), (11,0), (12,1) \rangle
\]

\[
y = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1
\end{bmatrix}
\]

\[
\tau: \langle (1,0), (2,0), (3,0), (4,0), (5,0), (6,0), (9,0), (8,0), (9,1), (10,0), (11,0), (12,1) \rangle
\]

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

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

Notice: \( F_0(\sigma \circ \tau) = n + \Delta(x,y) = \left\{ \begin{array}{ll}
< \frac{3n}{2} - \sqrt{n}, & \text{or} \\
\geq \frac{3n}{2} + \sqrt{n}.
\end{array} \right. \) Set \( \varepsilon = \frac{1}{\sqrt{n}} \).
Communication to Streaming

\[ p \text{-pass streaming algorithm} \implies (2p - 1)\text{-round communication protocol} \]

messages = memory contents of streaming algorithm

And Thus

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

- For one-round protocols, \( R \rightarrow (GHD) = \Omega(n) \)
- Implies the \( \tilde{\Omega}(\varepsilon^{-2}) \) streaming lower bounds
Communication to Streaming

$p$-pass streaming algorithm $\implies (2p - 1)$-round communication protocol

messages = memory contents of streaming algorithm

And Thus

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

- For one-round protocols, $R(\overrightarrow{GHD}) = \Omega(n)$
- Implies the $\tilde{\Omega}(\varepsilon^{-2})$ streaming lower bounds

Key open questions:

- What is the unrestricted randomized complexity $R(\text{GHD})$?
- Better algorithm for Distinct Elements (or $F_k$, or $H$) using two passes?
Our Results

Previous Results (Communication):

• One-round (one-way) lower bound: \( R \rightarrow (GHD) = \Omega(n) \) \[\text{Woodruff’04}\]

• Simplification, clever reduction from INDEX \[\text{Jayram-Kumar-Sivakumar}\]

• Multi-round case: \( R(GHD) = \Omega(\sqrt{n}) \) \[\text{Folklore}\]
Previous Results (Communication):

- One-round (one-way) lower bound: $R \rightarrow (\text{GHD}) = \Omega(n)$ [Woodruff’04]
- Simplification, clever reduction from INDEX [Jayram-Kumar-Sivakumar]
  Hard distribution “contrived,” non-uniform
- Multi-round case: $R(\text{GHD}) = \Omega(\sqrt{n})$ [Folklore]
Our Results

Previous Results (Communication):

- One-round (one-way) lower bound: $R \rightarrow (GHD) = \Omega(n)$ \ [Woodruff’04]
- Simplification, clever reduction from INDEX \ [Jayram-Kumar-Sivakumar]
  Hard distribution “contrived,” non-uniform
- Multi-round case: $R(GHD) = \Omega(\sqrt{n})$ \ [Folklore]
  Reduction from DISJOINTNESS using “repetition code”
  Hard distribution again far from uniform
Our Results

Previous Results (Communication):

- One-round (one-way) lower bound: $R \rightarrow (\text{GHD}) = \Omega(n)$ [Woodruff’04]
- Simplification, clever reduction from INDEX [Jayram-Kumar-Sivakumar]
  Hard distribution “contrived,” non-uniform
- Multi-round case: $R(\text{GHD}) = \Omega(\sqrt{n})$ [Folklore]
  Reduction from DISJOINTNESS using “repetition code”
  Hard distribution again far from uniform

What we show:

- Theorem 1: $\Omega(n)$ lower bound for any $O(1)$-round protocol
  Holds under uniform distribution
Our Results

Previous Results (Communication):

- One-round (one-way) lower bound: $R \rightarrow (GHD) = \Omega(n)$ [Woodruff’04]
- Simplification, clever reduction from INDEX [Jayram-Kumar-Sivakumar]
  Hard distribution “contrived,” non-uniform
- Multi-round case: $R(GHD) = \Omega(\sqrt{n})$ [Folklore]
  Reduction from DISJOINTNESS using “repetition code”
  Hard distribution again far from uniform

What we show:

- Theorem 1: $\Omega(n)$ lower bound for any $O(1)$-round protocol
  Holds under uniform distribution
- Theorem 2: one-round, deterministic: $D \rightarrow (GHD) = n - \Theta(\sqrt{n \log n})$
- Theorem 3: $R \rightarrow (GHD) = \Omega(n)$ (simpler proof, uniform distrib)
Technique: Round Elimination

Base Case Lemma: There is no “nice” 0-round GHD protocol.

Round Elimination Lemma: If there is a “nice” $k$-round GHD protocol, then there is a “nice” $(k - 1)$-round GHD protocol.
Technique: Round Elimination

**Base Case Lemma:** There is no 0-round GHD protocol with error $< \frac{1}{2}$.

**Round Elimination Lemma:** If there is a “nice” $k$-round GHD protocol, then there is a “nice” $(k - 1)$-round GHD’ protocol.
Technique: Round Elimination

Base Case Lemma: There is no 0-round GHD protocol with error $< \frac{1}{2}$.

Round Elimination Lemma: If there is a “nice” $k$-round GHD protocol, then there is a “nice” $(k - 1)$-round GHD$'$ protocol.

- The $(k - 1)$-round protocol will be solving a “simpler” problem
- Parameters degrade with each round elimination step
The problem:

\[
\text{GHD}_{c,n}(x, y) = \begin{cases} 
1, & \text{if } \Delta(x, y) \geq n/2 + c\sqrt{n}, \\
0, & \text{if } \Delta(x, y) \leq n/2 - c\sqrt{n}, \\
\ast, & \text{otherwise.}
\end{cases}
\]
**Parametrized Gap-Hamming-Distance Problem**

The problem:

\[
GHD_{c,n}(x, y) = \begin{cases} 
1, & \text{if } \Delta(x, y) \geq n/2 + c\sqrt{n}, \\
0, & \text{if } \Delta(x, y) \leq n/2 - c\sqrt{n}, \\
\star, & \text{otherwise.}
\end{cases}
\]

Hard input distribution:

\[\mu_{c,n} : \text{uniform over } (x, y) \text{ such that } |\Delta(x, y) - n/2| \geq c\sqrt{n}\]
Parametrized Gap-Hamming-Distance Problem

The problem:

\[
\text{GHD}_{c,n}(x, y) = \begin{cases} 
1, & \text{if } \Delta(x, y) \geq n/2 + c\sqrt{n}, \\
0, & \text{if } \Delta(x, y) \leq n/2 - c\sqrt{n}, \\
\ast, & \text{otherwise.}
\end{cases}
\]

Hard input distribution:

\[\mu_{c,n} : \text{uniform over } (x, y) \text{ such that } |\Delta(x, y) - n/2| \geq c\sqrt{n}\]

Protocol assumptions (eventually, will lead to contradiction):

- Deterministic \(k\)-round protocol for \(\text{GHD}_{c,n}\)
- Each message is \(s \ll n\) bits
- Error probability \(\leq \varepsilon\), under distribution \(\mu_{c,n}\)
Round Elimination

Main Construction: Given $k$-round protocol $P$ for $\text{GHD}_{c,n}$, construct $(k - 1)$-round protocol $Q$ for $\text{GHD}_{c',n'}$
Main Construction: Given $k$-round protocol $\mathcal{P}$ for $\text{GHD}_{c,n}$, construct $(k-1)$-round protocol $\mathcal{Q}$ for $\text{GHD}_{c',n'}$

First Attempt:

• Fix Alice’s first message $m$ in $\mathcal{P}$, suitably
**Main Construction:** Given $k$-round protocol $\mathcal{P}$ for $\text{GHD}_{c,n}$, construct $(k-1)$-round protocol $\mathcal{Q}$ for $\text{GHD}_{c',n'}$

First Attempt:

- Fix Alice’s first message $m$ in $\mathcal{P}$, suitably

- Protocol $\mathcal{Q}_1$:
  - Input: $x', y' \in \{0, 1\}^A$ where $A \subseteq [n]$, $|A| = n'$
  - Extend $x' \rightarrow x$ s.t. Alice sends $m$ on input $x$
  - Extend $y' \rightarrow y$ uniformly at random
  - Output $\mathcal{P}(x, y)$; Note: first message unnecessary
**Round Elimination**

**Main Construction:** Given $k$-round protocol $\mathcal{P}$ for $\text{GHD}_{c,n}$, construct $(k-1)$-round protocol $\mathcal{Q}$ for $\text{GHD}_{c',n'}$

First Attempt:

- Fix Alice’s first message $m$ in $\mathcal{P}$, suitably
- Protocol $\mathcal{Q}_1$:
  - Input: $x', y' \in \{0, 1\}^A$ where $A \subseteq [n]$, $|A| = n'$
  - Extend $x' \rightarrow x$ s.t. Alice sends $m$ on input $x$
  - Extend $y' \rightarrow y$ uniformly at random
  - Output $\mathcal{P}(x, y)$; Note: first message unnecessary
- Errors: $\mathcal{Q}_1$ correct, unless
  - $BAD_1$: $\text{GHD}_{c',n'}(x', y') \neq \text{GHD}_{c,n}(x, y)$.
  - $BAD_2$: $\text{GHD}_{c,n}(x, y) \neq \mathcal{P}(x, y)$.
Main Construction: Given \( k \)-round protocol \( P \) for \( \text{GHD}_{c,n} \), construct \((k - 1)\)-round protocol \( Q \) for \( \text{GHD}_{c',n'} \)

First Attempt:

- Fix Alice’s first message \( m \) in \( P \), suitably

- Protocol \( Q_1 \):
  - Input: \( x', y' \in \{0, 1\}^A \) where \( A \subseteq [n], \ |A| = n' \)
  - Extend \( x' \to x \) s.t. Alice sends \( m \) on input \( x \) (why possible?)
  - Extend \( y' \to y \) uniformly at random
  - Output \( P(x, y) \); Note: first message unnecessary

- Errors: \( Q_1 \) correct, unless
  - \( BAD_1 \): \( \text{GHD}_{c',n'}(x', y') \neq \text{GHD}_{c,n}(x, y) \).
  - \( BAD_2 \): \( \text{GHD}_{c,n}(x, y) \neq P(x, y) \).
Fixing Alice’s first message:

• Call $x$ good if $\Pr_y[P(x, y) \neq \text{GHD}_{c,n}(x, y)] \leq 2\varepsilon$

  Then $\#\{\text{good } x\} \geq 2^{n-1}$ (Markov)

• Let $M = M_m = \{\text{good } x : \text{Alice sends } m \text{ on input } x\}$.

• Fix $m$ to maximize $|M|$; then $|M| \geq 2^{n-1-s}$.
Fixing Alice’s first message:

- Call $x$ good if $\Pr_y[\mathcal{P}(x, y) \neq \text{GHD}_{c,n}(x, y)] \leq 2\varepsilon$
  
  Then $\#\{\text{good } x\} \geq 2^{n-1}$ (Markov)

- Let $M = M_m = \{\text{good } x : \text{Alice sends } m \text{ on input } x\}$.

- Fix $m$ to maximize $|M|$; then $|M| \geq 2^{n-1-s}$.

Shattering:

- Say $S \subseteq \{0, 1\}^n$ shatters $A \subseteq [n]$ if $\#\{x|_A : x \in S\} = 2^{|A|}$

- $\text{VCD}(S) := \text{size of largest } A \text{ shattered by } S$

**Sauer’s Lemma:** If $\text{VCD}(S) < \alpha n$ then $|S| < 2^{nH(\alpha)}$. 
Fixing Alice’s first message:

- Call $x$ good if $\Pr_y[\mathcal{P}(x, y) \neq \text{GHD}_{c,n}(x, y)] \leq 2\varepsilon$
  
  Then $\#\{\text{good } x\} \geq 2^{n-1}$ (Markov)

- Let $M = M_m = \{\text{good } x : \text{Alice sends } m \text{ on input } x\}$.

- Fix $m$ to maximize $|M|$; then $|M| \geq 2^{n-1-s}$.

Shattering:

- Say $S \subseteq \{0, 1\}^n$ shatters $A \subseteq [n]$ if $\#\{x|_A : x \in S\} = 2^{|A|}$

- $\text{VCD}(S) := \text{size of largest } A \text{ shattered by } S$

**Sauer’s Lemma:** If $\text{VCD}(S) < \alpha n$ then $|S| < 2^{nH(\alpha)}$.

**Corollary:** $\text{VCD}(M) \geq n' := n/3 \quad (\text{Because } s \ll n)$
Fixing Alice’s first message:

- Call \( x \) good if \( \Pr_y[\mathcal{P}(x, y) \neq \text{GHD}_{c, n}(x, y)] \leq 2\varepsilon \)

  Then \( \#\{\text{good } x\} \geq 2^{n-1} \) (Markov)

- Let \( M = M_m = \{\text{good } x : \text{Alice sends } m \text{ on input } x\} \).

- Fix \( m \) to maximize \( |M| \); then \( |M| \geq 2^{n-1-s} \).

Shattering:

- Say \( S \subseteq \{0, 1\}^n \) shatters \( A \subseteq [n] \) if \( \#\{x|_A : x \in S\} = 2^{|A|} \)

- \( \text{VCD}(S) := \text{size of largest } A \text{ shattered by } S \)

**Sauer’s Lemma:** If \( \text{VCD}(S) < \alpha n \) then \( |S| < 2^{nH(\alpha)} \).

**Corollary:** \( \text{VCD}(M) \geq n' := n/3 \) (Because \( s \ll n \))

Extend \( x' \rightarrow x \): pick \( x \in M \) such that \( x' = x|_A \)
Recall $BAD_1$: $\text{GHD}_{c',n'}(x',y') \neq \text{GHD}_{c,n}(x,y)$.

Notation: $x = x' \circ x''$, $y = y' \circ y''$, $n = n' + n''$. 
Recall $BAD_1$: $\text{GHD}_{c',n'}(x',y') \neq \text{GHD}_{c,n}(x,y)$.

Notation: $x = x' \circ x''$, $y = y' \circ y''$, $n = n' + n''$.

Definition: $x''$, $y''$ nearly orthogonal if $|\Delta(x'',y'') - n''/2| < 2\sqrt{n''}$. 

The First Bad Event
Recall $BAD_1$: $\text{GHD}_{c',n'}(x',y') \neq \text{GHD}_{c,n}(x,y)$.

Notation: $x = x' \circ x''$, $y = y' \circ y''$, $n = n' + n''$.

Definition: $x''$, $y''$ nearly orthogonal if $|\Delta(x'',y'') - n''/2| < 2\sqrt{n''}$.

Lemma: $\Pr_{y''}[x'',y'' \text{ nearly orthogonal}] > 7/8$. (Binom distrib tail)
The First Bad Event

Recall $BAD_1$: $\text{GHD}_{c',n'}(x', y') \neq \text{GHD}_{c,n}(x, y)$.

Notation: $x = x' \circ x''$, $y = y' \circ y''$, $n = n' + n''$.

Definition: $x''$, $y''$ nearly orthogonal if $|\Delta(x'', y'') - n''/2| < 2\sqrt{n''}$.

**Lemma:** $\Pr_{y''}[x'', y'' \text{ nearly orthogonal}] > 7/8$. (Binom distrib tail)

**Lemma:** If $x''$, $y''$ nearly orthogonal and $c' \geq 2c$, then

- $\text{GHD}_{c',n'}(x', y') = 1 \implies \text{GHD}_{c,n}(x, y) = 1$
- $\text{GHD}_{c',n'}(x', y') = 0 \implies \text{GHD}_{c,n}(x, y) = 0$
Recall $BAD_1$: $\text{GHD}_{c',n'}(x',y') \neq \text{GHD}_{c,n}(x,y)$.

Notation: $x = x' \circ x'', y = y' \circ y'', n = n' + n''$.

Definition: $x'', y''$ nearly orthogonal if $|\Delta(x'', y'') - n''/2| < 2\sqrt{n''}$.

Lemma: $\Pr_{y''}[x'', y'' \text{ nearly orthogonal}] > 7/8$. (Binom distrib tail)

Lemma: If $x'', y''$ nearly orthogonal and $c' \geq 2c$, then

- $\text{GHD}_{c',n'}(x',y') = 1 \implies \text{GHD}_{c,n}(x,y) = 1$
- $\text{GHD}_{c',n'}(x',y') = 0 \implies \text{GHD}_{c,n}(x,y) = 0$

Corollary: $\Pr[BAD_1] < 1/8$. 

Amit Chakrabarti
Recall $BAD_2$: $\text{GHD}_{c,n}(x, y) \neq \mathcal{P}(x, y)$.

Bounding $\Pr[BAD_2]$ is subtle:

- $x$ is good, so $\Pr[\mathcal{P} \text{ errs} \mid x] \leq 2\varepsilon$
  - But this requires $(x, y) \sim \mu_{c,n}$

- Random extension $(x', y') \rightarrow (x, y)$ is not $\sim \mu_{c,n}$.
The Second Bad Event

Recall $BAD_2$: $\text{GHD}_{c,n}(x, y) \neq \mathcal{P}(x, y)$.

Bounding $\Pr[BAD_2]$ is subtle:

- $x$ is good, so $\Pr[\mathcal{P} \text{ errs} \mid x] \leq 2\varepsilon$  
  - But this requires $(x, y) \sim \mu_{c,n}$

- Random extension $(x', y') \rightarrow (x, y)$ is not $\sim \mu_{c,n}$.

- Actual distrib (fixed $x$, random $y$):
  - $(x, y) \sim (\mu_{c',n'} \mid x) \otimes \text{Unif}_{n''}$
  - $y$ uniform over a subset of $\{0,1\}^n$, just like in $\mu_{c,n}$
Recall $BAD_2$: $GHD_{c,n}(x, y) \neq \mathcal{P}(x, y)$.

Bounding $\Pr[BAD_2]$ is subtle:

- $x$ is good, so $\Pr[\mathcal{P} \text{ err} \mid x] \leq 2\varepsilon$
  - But this requires $(x, y) \sim \mu_{c,n}$

- Random extension $(x', y') \rightarrow (x, y)$ is not $\sim \mu_{c,n}$.

- Actual distrib (fixed $x$, random $y$):
  - $(x, y) \sim (\mu_{c',n'} \mid x) \otimes \text{Unif}_{n''}$
  - $y$ uniform over a subset of $\{0, 1\}^n$, just like in $\mu_{c,n}$

**Lemma:** $\Pr[BAD_2] = O(\varepsilon)$. 
Round Elimination, First Attempt (Recap)

Putting it together:

- $\mathcal{P}$ is $k$-round $\varepsilon$-error protocol for $\text{GHD}_{c,n}$

- $Q_1$ is $(k - 1)$-round $\varepsilon'$-error protocol for $\text{GHD}_{c',n'}$ with
  - $c' = 2c$, $n' = n/3$
  - $\varepsilon' = 1/8 + O(\varepsilon)$
Putting it together:

- $\mathcal{P}$ is $k$-round $\varepsilon$-error protocol for $\text{GHD}_{c,n}$
- $\mathcal{Q}_1$ is $(k-1)$-round $\varepsilon'$-error protocol for $\text{GHD}_{c',n'}$ with
  - $c' = 2c$, $n' = n/3$
  - $\varepsilon' \leq 1/8 + 16\varepsilon$ ← Can’t repeat this argument!
Round Elimination, Second Attempt

Putting it together:

- **P** is $k$-round $\varepsilon$-error protocol for $\text{GHD}_{c,n}$
- **$Q_1$** is $(k - 1)$-round $\varepsilon'$-error protocol for $\text{GHD}_{c',n'}$ with
  - $c' = 2c$, $n' = n/3$
  - $\varepsilon' \leq 1/8 + 16\varepsilon$ \text{ ← Can’t repeat this argument!}

Second attempt: protocol **$Q$**:

- Repeat $Q_1$ $2^{O(k)}$ times in parallel, take majority
- Blows up communication by $2^{O(k)}$
- Error is now $\varepsilon' = O(\varepsilon)$
  - Analysis even more subtle: not just a Chernoff bound
Eventual Round Elimination Lemma

Lemma: If there is a $k$-round, $\varepsilon$-error protocol for $\text{GHD}_{c,n}$ in which each player sends $s \ll n$ bits, then there is a $(k - 1)$-round, $O(\varepsilon)$-error protocol for $\text{GHD}_{2c,n/3}$ in which each player sends $2^{O(k)} s$ bits.

Recall Base Case Lemma: There is no zero-round protocol with error $< 1/2$. 
**Eventual Round Elimination Lemma**

**Lemma:** If there is a $k$-round, $\varepsilon$-error protocol for $\text{GHD}_{c,n}$ in which each player sends $s \ll n$ bits, then there is a $(k - 1)$-round, $O(\varepsilon)$-error protocol for $\text{GHD}_{2c,n/3}$ in which each player sends $2^{O(k)} s$ bits.

Recall Base Case Lemma: There is no zero-round protocol with error $< 1/2$.

**Consequence: Main Theorem**

**Theorem:** There is no $o(n)$-bit, $\frac{1}{3}$-error, $O(1)$-round randomized protocol for $\text{GHD}_{c,n}$. In other words, $R^{O(1)}(\text{GHD}) = \Omega(n)$. 
Eventual Round Elimination Lemma

**Lemma:** If there is a $k$-round, $\varepsilon$-error protocol for $\text{GHD}_{c,n}$ in which each player sends $s \ll n$ bits, then there is a $(k - 1)$-round, $O(\varepsilon)$-error protocol for $\text{GHD}_{2c,n/3}$ in which each player sends $2^{O(k)}s$ bits.

Recall Base Case Lemma: There is no zero-round protocol with error $< 1/2$.

Consequence: Main Theorem

**Theorem:** There is no $o(n)$-bit, $\frac{1}{3}$-error, $O(1)$-round randomized protocol for $\text{GHD}_{c,n}$. In other words, $R^{O(1)}(\text{GHD}) = \Omega(n)$.

More Specific: $R^k(\text{GHD}) = n/2^{O(k^2)}$. 

Amit Chakrabarti
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 $\tilde{O} (\sqrt{n})$

- 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 $\tilde{O}(\sqrt{n})$

- 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 $\tilde{O}(\sqrt{n})$

- 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
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 $\tilde{O}(\sqrt{n})$

- 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
  I’m biased (I helped invent it, so it’s my pet technique)
Open Problems

1. The key problem here: Settle $R(\text{GHD})$.


3. This should help with other streaming problems, e.g., longest increasing subsequence.