Communication Lower Bounds of Collision Problems via Density Increment Arguments

Guangxu Yang Jiapeng Zhang

USC University of Southern California

Outline

- Collision problems and motivations
- Motivations and previous results
- Our contribution and applications
- Proof Outline

Collision Problems

Collision Problems(query version)

On input $z = (z_1, \dots, z_M) \in [N]^M$ $(M = 2N)$ the goal is to output a collision, that is, a pair of distinct indices $i, j \in [M]$ such that $z_i = z_j$.

It has been studied exhaustively in quantum query complexity and cryptography [BHT98,BSMP91,Aar02]

We note that since $M = 2N$, there must exist a collision.

Natural two-party communication version :

Alice holds the first half of the bits of each z_i and Bob holds the second half of each z_i .

Collision Problems

Let $M = 2N$,Alice holds $x = (x_1, \dots, x_M) \in [\sqrt{N}]^M$ and Bob holds $y = (y_1, \dots, y_M) \in [\sqrt{N}]^M$ .The goal is find a collision, that is, distinct $i, j \in [M]$ such that $x_i y_i = x_j y_j$.

Motivations

Motivation from Cryptography: Multi-set double-intersection problem

Bauer, Farshim and Mazaheri (CRYPTO 2018)	
Collision-resistance of combiners for backdoored random oracles.	Information complexity

Motivation from Proof Complexity: Bit-pigeonhole principle problem (BPHP)

Hrubes and Pudlak (FOCS 2017), Göös and Jain (RANDOM 2022) and Itsykson and Riazanov (CCC 2021)	
Some proof system requires exponential size to refute BPHP can be exponential size.	Lifting theorems

A Simple Protocol

Collision Problems (uniform distribution)

Let $M = 2N$, Alice holds $\mathrm{x} = (x_1, \dots, x_M) \in [\sqrt{N}]^M$ and Bob holds $\mathrm{y} = (y_1, \dots, y_M) \in [\sqrt{N}]^M$. the goal is find a collision, that is, distinct $i, j \in [M]$ such that $x_i y_i = x_j y_j$.

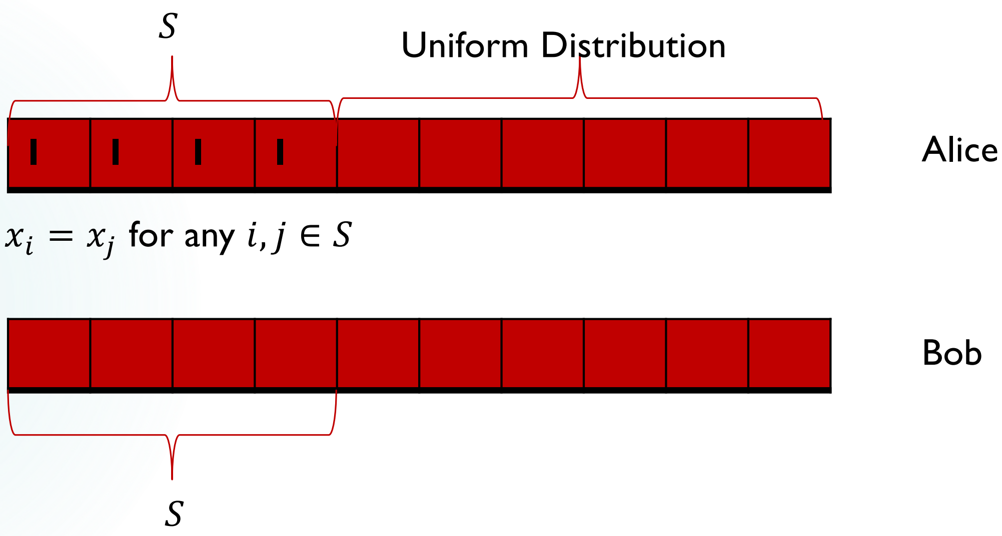
A Simple Protocol:

1. Alice randomly chooses $N^{1/4}$ coordinates with the same value and sends to Bob.
2. Bob can find a collision with high probability by the birthday paradox argument.

Proof by Birthday Paradox

Collision Problems (uniform distribution)

Let $M > N$, Alice holds $\mathrm{x} = (x_1, \ldots, x_M) \in [\sqrt{N}]^M$ and Bob holds $\mathrm{y} = (y_1, \ldots, y_M) \in [\sqrt{N}]^M$. the goal is find a collision, that is, distinct $i, j \in [M]$ such that $x_i y_i = x_j y_j$.



The value of each coordinate uniformly sampled from $[\sqrt{N}]$.

If $|\mathrm{S}| = N^{1/4}, \Pr[\exists\, i, j \in S, y_i = y_j] = \Omega(1)$ birthday paradox argument.

Previous Lower bounds:

Conjecture [BFM18,IR21,GJ22]

The communication lower bound of collision problem is $\Omega(N^{1/4})$

Göös and Jain (RANDOM 2022)

The communication lower bound of collision problem is $\Omega(N^{1/12})$

Their approach(Lifting theorem):

- Proving $\Omega(N^{1/3})$ communication lower bound of $Col \circ Ver$ via degree to rank lifting.
- Builds on lower bound of Col ∘ Ver , [GJ22] proves an $\Omega(N^{1/12})$ lower bound for BPHP via reductions.

Since there is a loss in the reduction [GJ22], the limitation of their framework is an $\Omega(N^{1/8})$ lower bound.

Our Contribution

Main Theorem

The communication lower bound of the collision problem is $\Omega(N^{1/4})$.

The protocol based on birthday paradox is almost optimal.

Technical Contribution:

1、 Using density-restoring partition [GLM+16,GPW17] for non-lifted functions. $f(g(x_1, y_1), \ldots, g(x_n, y_n))$

2、 Bypass the barriers by using the information complexity and lifting techniques.

Applications in cryptography and proof complexity

Bauer, Farshim and Mazaheri (CRYPTO 2018)

Collision-resistance of combiners for backdoored random oracles

Hrubes and Pudlak (FOCS 2017)

Every tree-like cutting planes of the weak bit pigeon hole principle BPHP , $M > N$, has size$2^{\Omega(N^{1/4})}$.

Göös and Jain (RANDOM 2022) and Itsykson and Riazanov (CCC 2021)

Any proof system that can be efficiently simulated by randomized protocols (most notably, tree-like Res($\oplus$)) requires exponential size to refute bit-pigeonhole formulas featuring M pigeons and N holes for arbitrary $M > N$.

Proof Outline

Baby version: One way communication

Send a messgae C

$\mathrm{x} = (x_1, \dots, x_M) \in [\sqrt{N}]^M$

$\mathrm{y} = (y_1, \dots, y_M) \in [\sqrt{N}]^M$

Bob outputs the collision $i, j \in [M]$ such that $x_i y_i = x_j y_j$

Theorem 1

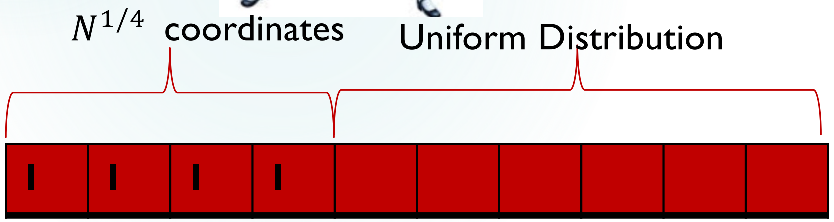
The one-way communication lower bound of the collision problem is $\Omega(N^{1/4})$.

Review: A Simple Protocol by Birthday Paradox

Collision Problems (uniform distribution)

Let $M > N$, Alice holds $\mathrm{x} = (x_1, \dots, x_M) \in [\sqrt{N}]^M$ and Bob holds $\mathrm{y} = (y_1, \dots, y_M) \in [\sqrt{N}]^M$. the goal is find a collision, that is, distinct $i, j \in [M]$ such that $x_i y_i = x_j y_j$.

The set of $N^{1/4}$ coordinates with same values $\longrightarrow$

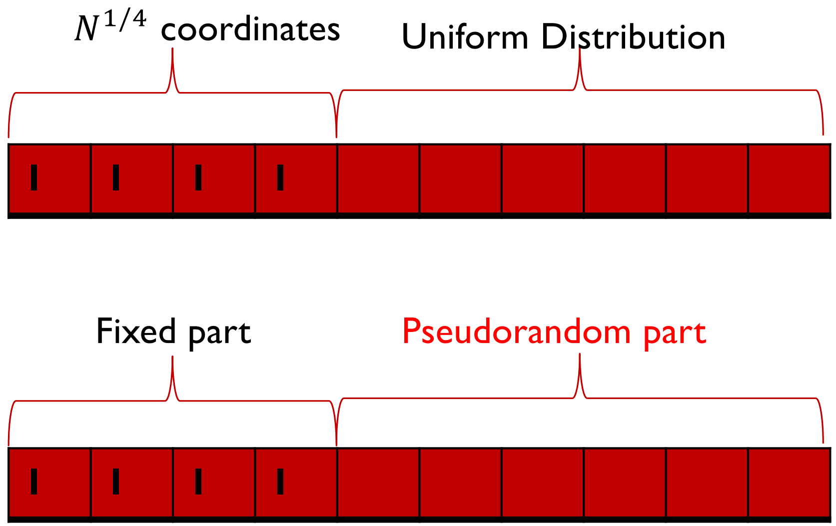


We want to prove this protocol is optimal !

Intuition: Normalize any protocol.

If Alice sends $N^{1/4}$ coordinates with same values:

If Alice sends a message M:



Structured sets

Let X be the random variable on inputs of Alice condition on message C.

There is a partition $X = \cup_i X^i$ such that for each i, X^i is a structured set and the expected size of fixed part is $O(|C|)$.

Intuition: One way communication

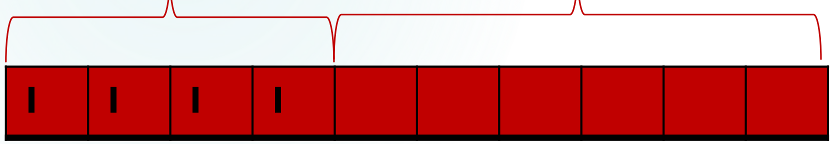
Send C with $|C| = o(N^{1/4})$

Let X be the random variable on inputs of Alice condition on message C.

There is a partition $X = \cup_i X^i$ such that for each i, X^i is a structured set and the expected size of fixed part is $o(N^{1/4})$.

Fixed part

Pseudorandom part

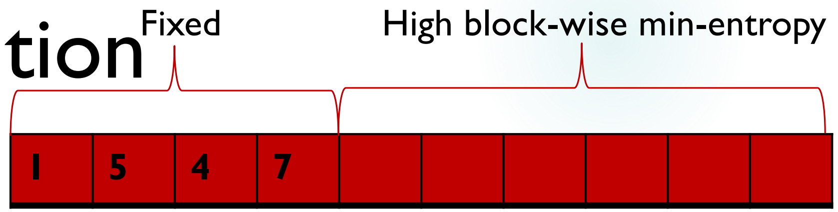


Structured sets

By birthday paradox, Bob can't find the collision with high probability for each X^i .

How to achieve such partition ?

Density-Restoring partition



Dense distribution [GLM+16]

Let $\boldsymbol{D}$ be a random variable on $[\sqrt{N}]^M$. We say that $\boldsymbol{D}$ is dense on J if for every subset $I \subseteq J$ it holds that

$$\mathrm{H}_\infty(\boldsymbol{D}_I) \geq \gamma \cdot |I| \cdot \log\sqrt{N}$$

and there is a $\alpha \in [\sqrt{N}]^{[M]\setminus J}$ such that $\Pr[\boldsymbol{D}_{[M]\setminus J} = \alpha] = 1$.

Set $\gamma = 1 - \frac{1}{\log\sqrt{N}}$

Lemma 1 [GLM+16,GPW17]

Density function of $X \subseteq [\sqrt{N}]^J$: $D_\infty(X) = |J| \cdot \log\sqrt{N} - H_\infty(X)$

For any $X \subseteq [\sqrt{N}]^J$, there is a partition $X = \cup_i X^i$ such that for each i,

- $X^i_{I_i} = \alpha_i$ and X^i is dense on J $\setminus$ I_i.
- $D_\infty(X^i_{J\setminus I_i}) \leq D_\infty(X) - |I_i| + \delta_i$ where $\delta_i = \log\frac{|X|}{|\cup_{j\geq i} X^j|}$

Density-Restoring Partition

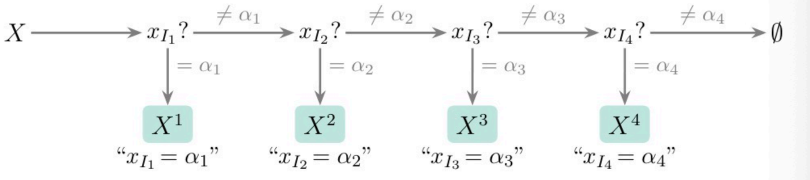
Lemma 1 [GLM+16,GPW17]

For any $X \subseteq [\sqrt{N}]^J$, there is a partition $X = \cup_i X^i$ such that for each i,

- There is a $I_i \subseteq J$ and $\alpha_i \in [\sqrt{N}]^{I_i}$ such that $X^i_{I_i} = \alpha_i$ and X^i is dense on $J \setminus I_i$.
- $D_\infty(X^i_{J\setminus I_i}) \leq D_\infty(X) - |I_i| + \delta_i$ where $\delta_i = \log\frac{|X|}{|\cup_{j\geq i} X^j|}$

Density Restoring Partition:

1. While $X \neq \emptyset$:
2. Initialize $i = 1$.
3. Let $I \subseteq J$ be a maximal subset (possibly $I = \emptyset$) such that X_I has min-entropy rate γ and let $\alpha_i \in [\sqrt{N}]^I$ be an outcome witnessing this: $\Pr[X_I = \alpha_i] > \sqrt{N}^{-\gamma|I|}$.
4. Output $X^i := \{ x \in X : x_I = \alpha_i \}$ and $I_i = I$.
5. Update $X \leftarrow X \setminus X^i$ and $J = J \setminus I$.



Proof of Lemma 1

Lemma 1 [GLM+16,GPW17]

For any $X \subseteq [\sqrt{N}]^J$, there is a partition $X = \bigcup_i X^i$ such that for each i,

- There is a $I_i \subseteq J$ and $\alpha_i \in [\sqrt{N}]^{I_i}$ such that $X^i_{I_i} = \alpha_i$ and X^i is dense on $J \setminus I_i$.
- $D_\infty(X^i_{J\setminus I_i}) \leq D_\infty(X) - |I_i| + \delta_i$ where $\delta_i = \log \frac{|X|}{|\bigcup_{j\geq i} X^j|}$

Proof outline: **The first part is proved by contradiction.**
If X^i is not dense on $J \setminus I_i$, then there is a non-empty set $K \subseteq J \setminus I_i$ and an outcome $\beta \in [\sqrt{N}]$^K violating the min-entropy condition.

Thus, the set $I_i \cup K \subseteq J$ and (α_i, β) violating the min-entropy condition this contradicts the maximality of I_i .

Proof of Lemma 1

Lemma 1 [GLM+16,GPW17]

Density function of $X \subseteq [\sqrt{N}]^J$: $D_\infty(X) = |J| \cdot \log\sqrt{N} - H_\infty(X)$

For any $X \subseteq [\sqrt{N}]^J$, there is a partition $X = \bigcup_i X^i$ such that for each i,

- There is a $I_i \subseteq J$ and $\alpha_i \in [\sqrt{N}]^{I_i}$ such that $X^i_{I_i} = \alpha_i$ and X^i is dense on $J \setminus I_i$.
- $D_\infty(X^i_{J\setminus I_i}) \leq D_\infty(X) - |I_i| + \delta_i$ where $\delta_i = \log\frac{|X|}{|\bigcup_{j\geq i} X^j|}$

Proof outline: **The second part is proved by straightforward calculation:**

$$
\begin{aligned}
D_\infty(X^i_{J\setminus I_i}) &= |J \setminus I_i| \log\sqrt{N} - \log|X^i| \\
&\leq (|J| \log\sqrt{N} - |I_i| \log\sqrt{N}) - \log(|\textstyle\bigcup_{j\geq i} X^j| \cdot \sqrt{N}^{-\gamma|I_i|}) \\
&= (|J| \log\sqrt{N} - \log|X|) - (1-\gamma)|I_i| \cdot \log\sqrt{N} + \log(|X|/|\textstyle\bigcup_{j\geq i} X^j|) \quad \left(\gamma = 1 - \frac{1}{\log\sqrt{N}}\right) \\
&\leq D_\infty(X) - |I_i| + \delta_i
\end{aligned}
$$

Proof of one way communication lower bound

Lemma 1 [GLM+16,GPW17]

Density function of $X \subseteq [\sqrt{N}]^J$: $D_\infty(X) = |J| \cdot \log\sqrt{N} - H_\infty(X)$

For any $X \subseteq [\sqrt{N}]^J$, there is a partition $X = \bigcup_i X^i$ such that for each i,

- There is a $I_i \subseteq J$ and $\alpha_i \in [\sqrt{N}]^{I_i}$ such that $X^i_{I_i} = \alpha_i$ and X^i is dense on $J \setminus I_i$.
- $D_\infty(X^i_{J\setminus I_i}) \le D_\infty(X) - |I_i| + \delta_i$ where $\delta_i = \log\frac{|X|}{|\bigcup_{j\ge i} X^j|}$

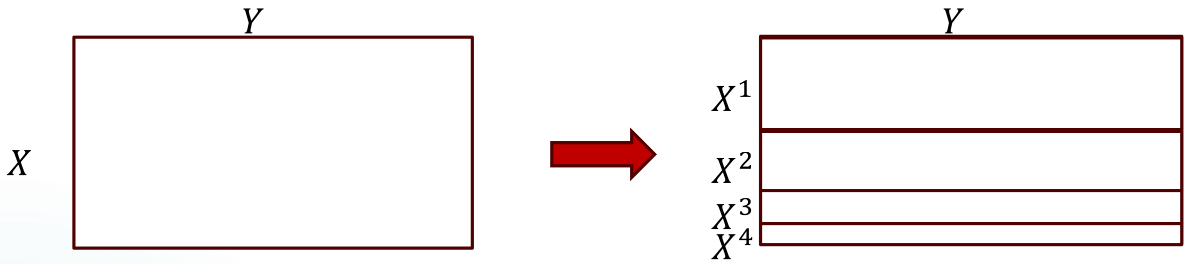
Moreover, let $J = [M]$ and $p_i = \frac{|X^i|}{|X|}$ denote the probability of set X^i,

$$E[\delta_i] = \sum_i p_i \cdot \log\frac{|X|}{|\bigcup_{j\ge i} X^j|} = \sum_i p_i \cdot \log\frac{1}{\sum_{j\ge i} p_j} \le \int_0^1 \log\frac{1}{x}\,dx = 1$$

$$E[|I_i|] \le D_\infty(X) - E[D_\infty(X^i_{J\setminus I_i})] + E[\delta_i] \le D_\infty(X) + E[\delta_i] \le M \cdot \log\sqrt{N} - H_\infty(X) + 1$$

$E[D_\infty(X^i_{J\setminus I_i})] \ge 0$ $\qquad$ $E[\delta_i] \le 1$

Proof of one way communication lower bound



X^i is dense on $[\mathrm{M}] \setminus I_i$

By Lemma 1, $E[|I_i|] \leq M \cdot \log\sqrt{N} - H_\infty(X) + 1 = M \cdot \log\sqrt{N} - \log|X| + 1$

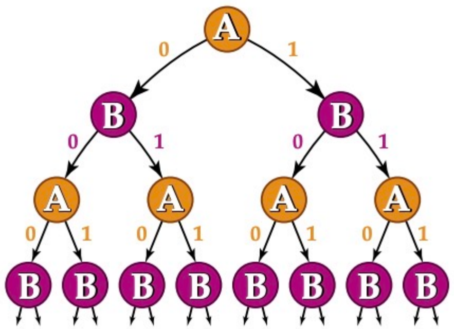
By birthday paradox, if Bob can find collision with high probability in X, then $E[I_i] = \Omega(N^{1/4})$.

$$|X| \leq \frac{\left(\sqrt{N}\right)^M}{2^{\Omega(N^{1/4})}}$$

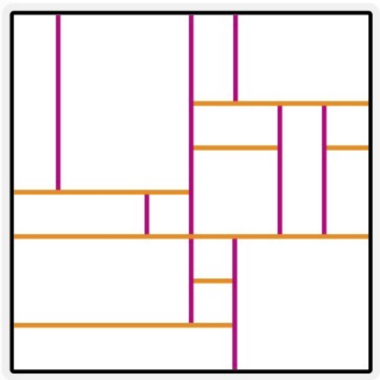
Two way communication lower bound ?

Two way communication lower bound

Protocol tree of Π



Partition of Truth Table



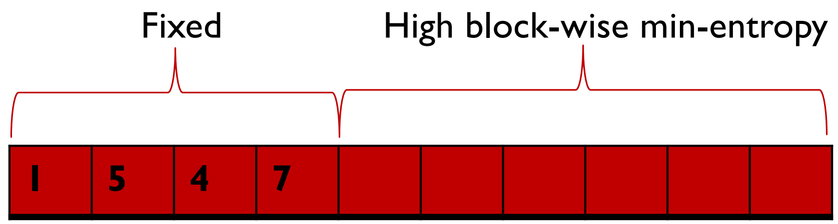
Π induces a partition of the truth table into at most $2^{|\Pi|}$ leaf rectangles.
The leaf rectangles are in 1-to-1 correspondence with the leaves of the protocol tree.

Decomposition Algorithm

Our decomposition algorithm has two process in each communication round:

- Density-Restoring partition:

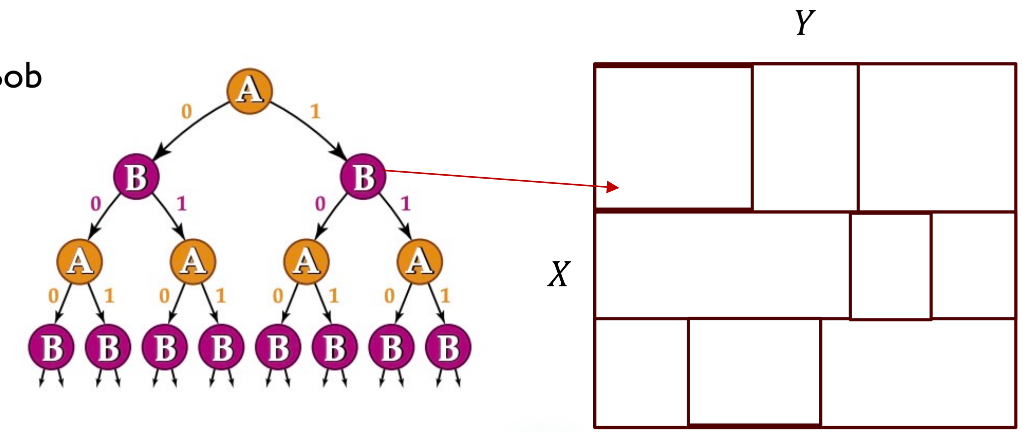
We further decompose the rectangle of each node into dense rectangles (X is dense on J_1 and Y is dense on J_2).



- Labeling process:

Labeling the inputs in each dense rectangle that Alice or Bob can find the collision.

Claim: The total probability of labeled inputs should be $\Omega(1)$ if Alice or Bob can find the collision with high probability.



Density-Restoring Partition (Alice speaks)

X_0 is dense on J_1 and Y is dense on J_2

Alice speaks, $X_0 = X \cup X^*$

Y is dense on J_2

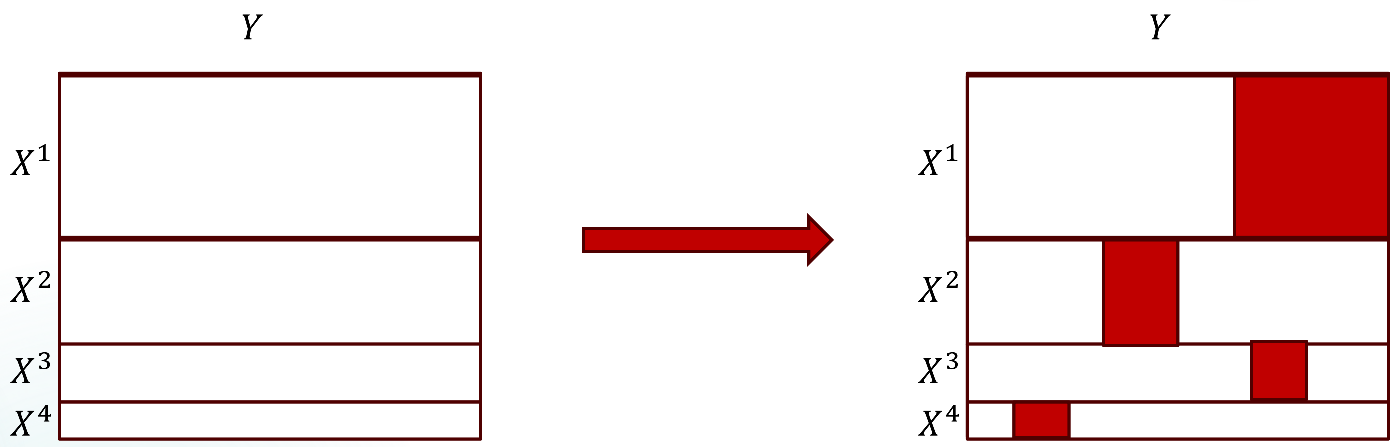
Y is dense on J_2 and X^i is dense on $J_1 \setminus I_i$

Lemma 1 [GLM+16,GPW17]

For any $X \subseteq \left[\sqrt{N}\right]^{J_1}$, there is a partition $X = \cup_i X^i$ such that for each i,

- $X^i_{I_i} = \alpha_i$ and X^i is dense on $J_1 \setminus I_i$.
- $D_\infty\left(X^i_{J_1 \setminus I_i}\right) \leq D_\infty(X) - |I_i| + \delta_i$ where $\delta_i = \log \frac{|X|}{|\cup_{j \geq i} X^j|}$

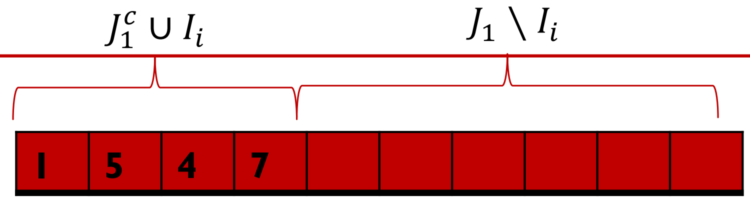
Labeling Process(Alice speaks)



Y is dense on J_2 and X^i is dense on $J \setminus I_i$

Labeling process:

For each i, s^i be the value of fixed part $I_i \cup J^c$ in X^i.



We define $Y^i = \{y \in Y : y_i = y_j \ for\ some\ i,j \in I_i \cup J_1^c\ with\ s_i^i = s_j^i\ \}$

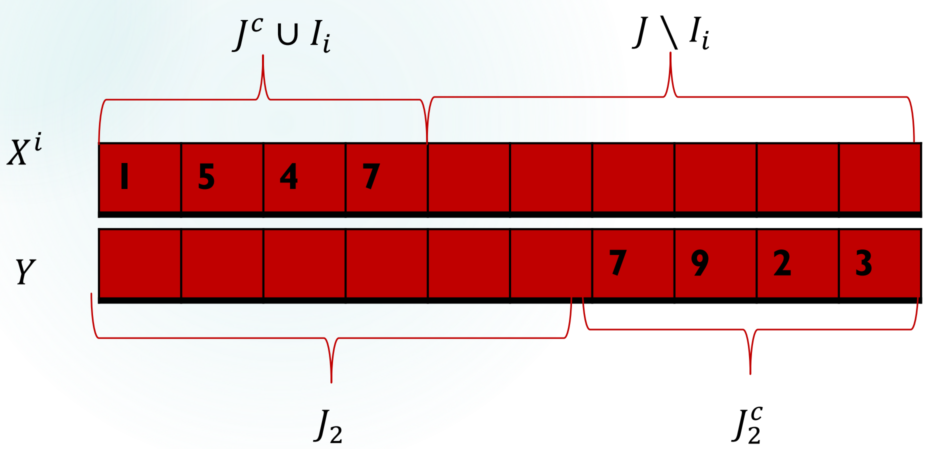
The inputs of Bob that can find the collision in fixed part of Alice.

Labeling the inputs in the rectangle $X^i \times Y^i$ if it don't be labeled in previous rounds.

Proof of Lemma 2

Lemma 2

In the labeling process, for each i, the probability of labeled inputs in $X^i \times Y$ is at most $\frac{2 \cdot (|I_i \cup J^c|^2 - |J^c|^2)}{\sqrt{N}}$



We only consider the case: $J^c \cup I_i \subseteq J_2$

If $J^c \cup I_i \cap J_2^c \neq \emptyset$,

- **Either there is no collision in** $J^c \cup I_i \cap J_2^c$
- **Or if there is a collision in** $J^c \cup I_i \cap J_2^c$ **, there inputs must be labeled in previous rounds.**

Proof of Lemma 2

Lemma 2

In the labeling process, for each i, the probability of labeled inputs in $X^i \times Y$ is at most $\frac{2 \cdot (|I_i \cup J^c|^2 - |J^c|^2)}{\sqrt{N}}$

Proof outline:

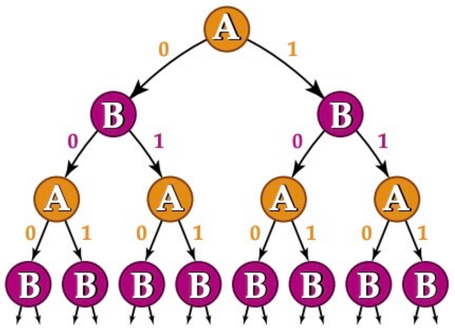
For any $(i, j) \in I_i \cup J_1^c \; with \; s_i^i = s_j^i$, since $i, j \in J_2$ and Y is dense on J_2

$$\Pr[y_i = y_j] \le \frac{2}{\sqrt{N}}$$

The lemma 2 holds by union bound.

Decomposition Algorithm

Protocol tree of Π



Assume Alice speaks,

In each communication iteration, for each dense rectangle $X \times Y$ is decomposed into $X_0 \times Y$ and $X_1 \times Y$.

Doing density-restoring partition on X_0 and X_1 to further decompose $X_0 \times Y$ and $X_1 \times Y$ into dense rectangles.

Labeling the inputs in dense rectangles.

- Lemma 3: The expected size of fixed coordinates in leaf rectangles is at most $\mathrm{O}(\mathrm{CC}(\Pi))$.
- Lemma 4: If the expected size of fixed coordinates in leaf rectangles is $\mathrm{o}\left(N^{\frac{1}{4}}\right)$, Alice or Bob can find the collision with $o(1)$ probability.

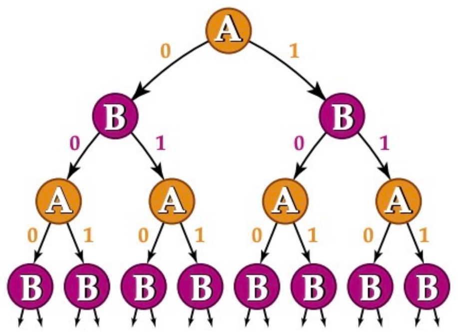
Proof of Lemma 3

Lemma 3: The expected size of fixed coordinates in leaf rectangles is at most $O(CC(\Pi))$.

Proof outline via density increment argument:

Density function is the average of density function of current all dense rectangles.

Protocol tree of Π



Density function is 0 at the beginning of protocol tree.

In each communication round, the density function increase at most 1.

By Lemma 1, in the density-restoring partition, for each i

$$E[D_\infty(X^i_{J\setminus I_i})] \leq E[D_\infty(X)] - E[|I_i|\} + E[\delta_i]$$

the density function decreases at least $E[|I_i|] - E[\delta_i] \leq E_i[|I_i|] - 1$.

Since the density function is always non-negative, the expected size of fixed coordinates = $\sum E[|I_i|] \leq 2 \cdot CC(\Pi)$

Proof of Lemma 4

Lemma 4: If the expected size of fixed coordinates in leaf rectangles is $o\left(N^{\frac{1}{4}}\right)$, then Alice or Bob can find the collision with $o(1)$ probability.

Proof outline:

Claim: the total probability of labeled inputs should be $\Omega(1)$ if Alice or Bob can find the collision with high probability.

By Lemma 2, in each labeling process,

the probability of labeled inputs increase at most $\mathrm{E}[\frac{2\cdot(|I_i \cup J^c|^2 - |J^c|^2)}{\sqrt{N}}]$. $\mathrm{E}\left[min\{ 1, \frac{2\cdot(|I_i \cup J^c|^2 - |J^c|^2)}{\sqrt{N}}\}\right]$.

Thus, let J_1 and J_2 be the random variables on fixed coordinates of Alice and Bob's sides . Taking summation in all communication rounds,

the total probability of labeled inputs is at most $\mathrm{E}[\frac{2\cdot(|J_1^c|^2 + |J_2^c|^2)}{\sqrt{N}}]$.

$$\mathrm{E}\left[\frac{2\cdot(|J_1^c|^2 + |J_2^c|^2)}{\sqrt{N}}\right] = \Omega(1). \qquad \mathrm{E}\left[min\{ 2, \frac{2\cdot(|J_1^c|^2 + |J_2^c|^2)}{\sqrt{N}}\}\right] = \Omega(1) .$$

Summary and Proof Outline

Main Theorem
The communication lower bound of the collision problem is $\Omega(N^{1/4})$.

The proof outline is as follows:

Decomposition Algorithm: In each communication iteration, do density-restoring partition and labeling process for each dense rectangle.

Lemma 3: The expected size of fixed coordinates in leaf rectangles is at most $\mathrm{O}(\mathrm{CC}(\Pi))$

Proved by density increment arguments.

Lemma 4: The expected size of fixed coordinates in leaf rectangles is at least $\Omega\left(N^{\frac{1}{4}}\right)$.

Proved by birthday paradox argument.

Other Applications and Open Problems

Main Theorem

The communication lower bound of the collision problem is $\Omega(N^{1/4})$.

Numbers on forehead model?

This result will have important applications in proof complexity.

Thank you for listening ☺