Electronic Colloquium on Computational Complexity, Report No. 67 (2024)

Breaking Square-Root Loss Barriers via Min-Entropy

Mi-Ying Huang * Xinyu Mao * Guangxu Yang * Jiapeng Zhang *

April 10, 2024

Abstract

Information complexity is one of the most powerful tools to prove information-theoretical lower bounds, with broad applications in communication complexity and streaming algorithms. A core notion in information complexity analysis is the Shannon entropy. Though it has some convenient properties, such as chain rules, Shannon entropy still has inherent limitations. One of the most notable barriers is the square-root loss, which is reflected by the square-root gap between entropy gaps and statistical distances, e.g., Pinsker's inequality.

To break this barrier, we introduce a new method based on min-entropy analysis. Building on this new method, we prove the following three results.

- A tight $\Omega(n/k)$ randomized lower bounds of the k-party *Tree Pointer Jumping* problems, improving an $\Omega(n/k^2)$ lower bounds by Chakrabarti, Cormode, and McGregor (STOC 08).
- An $\Omega(n/k + \sqrt{n})$ lower bounds of the *Chained Index* problem for *oblivious communication protocols*, improving an $\Omega(n/k^2)$ lower bound by Cormode, Dark, and Konrad (ICALP 19). Here, oblivious means that the length of each message does not depend on the input.
- An $\Omega(n/k - k)$ lower bounds of the *Chained Index* problem for *non-oblivious protocols*. To the best of our knowledge, this is the first lower bound for non-oblivious protocols.

Since both problems served as hard problems for numerous applications for streaming problems, our new lower bounds improve these streaming lower bounds directly.

On the technical side, unlike Shannon entropy, min-entropy does not have nice properties such as chain rules. To address this issue, we adopt the structure-vs-pseudorandomness decomposition used by Göös, Pitassi, and Watson (FOCS 17) and Yang and Zhang (STOC 24), where both papers used this decomposition to prove communication lower bounds. In this paper, we extend this method into streaming settings, contributing a new toolkit for proving streaming lower bounds.

1 Introduction

Information complexity is one of the most powerful tools in proving communication complexity lower bounds [CSWY01, BYJKS04, BBCR10, DOR21, OR23] and streaming lower bounds [BYJKS04, CCM08, AMOP08, GO16, ACK19, BGW20, LZ23, BGL+24]

*Research supported by NSF CAREER award 2141536.
Thomas Lord Department of Computer Science, University of Southern California.
Email: {miying.huang,xinyumao, guangxuy, jiapengz}@usc.edu

ISSN 1433-8092

The idea of information complexity is to analyze the mutual information between the inputs held by the communication parties and the communication transcript. The definition of information complexity is similar to communication complexity, with information cost replacing communication cost. For a protocol Π, a popular notion of information cost is defined by

$$\mathrm{IC}(\Pi) := \mathrm{I}(\boldsymbol{X}; \Pi(\boldsymbol{X}, \boldsymbol{Y})|\boldsymbol{Y}) + \mathrm{I}(Y; \Pi(\boldsymbol{X}, \boldsymbol{Y})|\boldsymbol{Y}),$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the input distribution of Alice and Bob respectively and I is the mutual information. Intuitively, $\mathrm{IC}(\Pi)$ captures the mutual information of the inputs and the communication transcript, which is a lower bound of the communication cost. Besides this specific definition, there are many different variants which are smartly designed for diverse applications. However, they all share a similar idea: capture the information cost (usually by Shannon entropy) between the input distribution and the transcript.

Despite a vast number of applications successfully given by the information complexity-based approaches, this framework still has some inherent limitations. Indeed, some significant barriers are *not only* associated with some specific variants of information cost notions, but further deeply caused by the *entropy* itself. In this direction, one notable limitation is the square-root loss barrier.

Square-root loss barrier. We first use a simple example to illustrate this phenomenon. Let $\boldsymbol{I}$ be a random variable that outputs 1 with probability $1/2 + \varepsilon$ and 0 with probability $1/2 - \varepsilon$. This is a biased coin with a $\Theta(\varepsilon)$ statistical distance to the uniform distribution. However, on the other hand, the entropy gap between them has only $\Theta(\varepsilon^2)$. This square gap is not significant if ε is a constant. However, the loss would become very large when it becomes very small. Beyond this simple example, this is indeed a *general gap between entropy loss and statistical distance*. For example, any result that applies Pinsker's inequality has a good chance of creating this gap.

Lemma 1.1 (Pinsker's inequality)**.** *If P and Q are two distributions, then*

$$D_{TV}(P, Q) \leq \sqrt{\frac{1}{2} D_{KL}(P \| Q)}$$

Here $D_{TV}(P, Q)$ is the total variation distance of P and Q and $D_{KL}(P\|Q)$ is the KL-divergence of P and Q.

This quadratic gap makes it difficult to get good bounds via entropy-based analysis in many applications. For instance, proofs of *multiparty unique-set disjointness* [BYJKS04], *set disjointness under product distribution* [DOR21, OR23], the *chained index problem* [CDK18], *multi-party pointer jumping problem* [Cha07], *tree pointer jumping problem* [CCM08], *pointer chasing problem* [NW91], among others, all meet the square-root loss comparing the upper bounds.

Besides concrete examples, this square-root loss also appears in fundamental problems such as *direct-sum questions*. We quote a comment from a nice paper by Yehudayoff [Yeh20] here.

It appears in the parallel repetition theorem and is connected to the 'strong parallel repetition' conjecture, which is motivated by Khot's unique games conjecture [Kho02]. The 'strong parallel repetition' conjecture was falsified by Raz [Raz11]; showing this square-root loss is necessary for parallel repetition. This loss also appears in direct sums and products in communication complexity [BBCR10, BRWY13], which is related to the question of optimal compression of protocols. It is still unclear if the square-root loss is necessary for the direct sum question.

Towards breaking the square-root loss. Since this loss limits many applications, there is a good amount of work made progress to resolve this barrier [CKS03, Gro09, Jay09, BM13, Yeh20]. For example, Jayram [Jay09] proved tight lower bounds for the multiparty unique set-disjointness, resolving the square-root loss by [BYJKS04, CKS03]; Braverman and Moitra [BM13] proved tight lower bounds for the unique set-disjointness for all probabilities; and Yehudayoff [Yeh20] improved the pointer chasing problem, addressed the square-root loss by [NW91].

Despite avoiding the square-root loss barrier for some specific problems, these efforts are ad-hoc with some intelligent analysis on non-standard variants of Shannon entropy. Hence, it is hard to extend them for broader applications. A natural question arises: Could we use any measurement other than the Shannon entropy (or its close variants)?

Now, we revisit the example above. For a random variable $\boldsymbol{X}$ supported on $\{0,1\}^n$ with entropy $\mathrm{H}(\boldsymbol{X}) \geq n - \varepsilon$, we know the statistical distance between $\boldsymbol{X}$ and the uniform distribution is $\Theta(\sqrt{\varepsilon})$ by Pinsker's inequality. In further, it is hard to improve Pinsker's inequality as it is tight in general. However, on the other hand, for a random variable $\boldsymbol{X}$ with *min-entropy* $n - \varepsilon$, a simple calculation shows that the statistical distance between $\boldsymbol{X}$ and the uniform distribution is $\Theta(\varepsilon)$. In this paper, min-entropy is a good candidate for avoiding square root loss in general settings.

Analysis of min-entropy via structure-vs-pseudorandomness. Though the min-entropy itself does not meet the square-root loss, there are other challenges in analyzing it. One of the most significant challenges is that, unlike the Shannon entropy, there is *no chain rule* for min-entropy, where a chain rule is an essential tool in entropy-bases analysis.

In order to overcome this issue, we adopt the structure-vs-pseudorandomness decomposition to serve as the "chain rule" in min-entropy analysis. This approach has been successfully applied in sunflower lemmas [LSZ19, ALWZ20] and query-to-communication lifting theorems [GLM+16, GPW17, LMM+22, YZ24]. Though this approach has been successfully applied in several areas, it has not been studied in *streaming settings*. In this paper, we extend this approach to streaming problems. Beyond the two problems studied in this paper, we believe the min-entropy-based analysis could provide more applications to streaming problems.

1.1 Our Results and Their Applications to Streaming Problems

Building on min-entropy analysis, we improve the lower bounds for two communication problems: 1) *Tree Pointer Jumping problem* [CCM08] and 2) *Chained Index problem* [CDK18]. Combined with previous reductions, our new results also give many applications in streaming problems (see Section 1.1.3 for more details).

1.1.1 Tree Pointer Jumping Problem

The Tree Pointer Jumping problem is a communication problem introduced by Chakrabarti, Cormode, and McGregor [CCM08] with applications in streaming lower bounds. For $t, k \geq 2$, we consider a complete k-level t-ary tree T rooted at v_1. The k-party Tree Pointer Jumping problem, denoted by $TPJ_{k,t}(\phi)$, takes as an input a function $\phi : V(T) \to [t]$, where $V(T)$ is the set of nodes of T. For each input ϕ, we define the functions g_ϕ by,

$$g_\phi(v) = \begin{cases} \text{the } \phi(v)\text{-th child of } v & \text{if } v \text{ is not an internal node;} \\ \phi(v) & \text{if } v \text{ is a leaf.} \end{cases}$$

The output of $TPJ_{k,t}(\phi)$ is defined by $TPJ_{k,t}(\phi) := g_\phi(g_\phi(\cdots g_\phi(v_1)\cdots))$. In the communication setting, the input ϕ is distributed to k players. They play a communication game as follows:

- Player i receives the labels of the i-th level nodes, i.e., the first player receives $\phi(v_1)$,…, and the last player receives the labels of the leaves.
- In each round, players send messages in reverse order: *from the last player to the first player*. The cost of this round is the total number of bits sent by all players.
- Players could communicate $(k-1)$ rounds, and the *first player* outputs the answer.

The goal of the players is to compute $TPJ_{k,t}(\phi)$ while minimizing *the maximum cost of each round*. For any $(r-1)$-round protocol Π, we use $R_{\max}(\Pi)$ to denote the maximum communication cost in all rounds. In this direction, [CCM08] first proved the following lower bound.

Theorem 1.2 ([CCM08]). *Let μ_k denote the uniform distribution over all functions $\phi : V(T) \to [t]$. Then for any $(k-1)$-round protocol Π with*

$$\Pr_{\phi \leftarrow \mu_k}[\Pi(\phi) = TPJ_{k,t}(\phi)] \geq 2/3,$$

we have that $R_{\max}(\Pi) = \Omega(t/k^2)$.

Chakrabarti, Cormode , and McGregor [CCM08] first used Theorem 1.2 to improve multi-pass streaming lower bound for median finding. Later on, Chakrabarti and Wirth [CW16] used this theorem to show a pass/approximation trade-off for the SET-COVER in the semi-streaming setting. In this paper, we improve the lower bound from Theorem 1.2 based on min-entropy analysis.

Theorem 1.3. *Let μ_k denote the uniform distribution over all functions $\phi : V(T) \to [t]$. Then for any $(k-1)$-round protocol Π with*

$$\Pr_{\phi \leftarrow \mu_k}[\Pi(\phi) = TPJ_{k,t}(\phi)] \geq 2/3,$$

we have that $R_{\max}(\Pi) = \Omega(t/k)$.

As the corollaries, our improved lower bounds can be directly used to improve the downstream applications in [CCM08] and [CW16].

1.1.2 Chained Index Problem

The Chained Index problem, introduced by Cormode, Dark and, Konrad [CDK18], is another useful tool with many applications in streaming lower bounds [CDK18, FNFSZ20, FNFSZ22, BKO22, DDK23]. For this problem, we consider the following communication setting.

- There are k players. Each player i receives an input $z_i = (\sigma_i, x_i) \in [n] \times \{0,1\}^n$
- It is promised that $x_1(\sigma_2) = \cdots = x_{k-1}(\sigma_k)$. Here $x_i(\sigma_{i+1})$ is the σ_{i+1}-th coordinate of x_i.
- Their goal is to compute $x_i(\sigma_{i+1})$ through a one-way communication from the first player to the last player, where the last player should output the answer.

We say that a one-way protocol solves the Chained Index problem if for every input $(z_1, \ldots, z_k)$, the last player always outputs the correct answer with probability $2/3$. The communication cost of this protocol is the total communication bits of all players. Built on the information complexity, [CDK18] proved the following lower bounds for the Chained Index problem.

Theorem 1.4 ([CDK18]). *Any one-way communication protocol that solves the Chained Index problem has randomized communication complexity $\Omega(n/k^2)$.*

Since it has been introduced, many streaming lower bounds [CDK18, FNFSZ20, FNFSZ22, BKO22, DDK23] were built on Theorem 1.4. We list some of them in Section 1.1.3.

Theorem 1.5 ([CDK18]). *Any algorithm for the explicit vertex stream model that finds a c-approximation to $\alpha(G)$ with probability at least $2/3$ requires $\Omega(\frac{n^2}{c^7})$ space.*

For a restricted range of k, [CDK18] also announced the following result without proof.

Theorem 1.6 ([CDK18]). *For $k < (\frac{n}{\log n})^{1/4}$, any one-way communication protocol that solves the Chained Index problem has randomized communication complexity $\Omega(n/k)$.*

Theorem 1.4 was obtained by direct entropy-based analysis. However, we do not see a proof for 1.6 so far. In this paper, we prove the following lower bounds for the Chained Index problem without any range restriction on k.

Theorem 1.7. *Any one-way communication protocol that solves the Chained Index problem has randomized communication complexity $\Omega(n/k - k)$. For oblivious protocols, i.e., protocols in which each message has a predetermined length independent of the input, the lower bound can be improved to $\Omega(n/k + \sqrt{n})$.*

Remark 1.8. To the best of our knowledge, all previous results only apply to oblivious protocols, e.g., [Cha07, CDK18]; this is the first lower bound for non-oblivious protocols.

In Theorem 1.7, we consider a blackboard communication setting: any player can see all messages from previous players. This lower bound is indeed stronger than the lower bound in the streaming setting by [CDK18] where the i-th player can only see the $(i-1)$-th player's message. [CDK18] conjectured the tight bound should be $\Omega(n)$ (independent with k) in the streaming setting. However, in the blackboard setting, the upper bound is not clear. We suspect there is a chance to improve the upper bound into $O(n/k)$. We leave it as an interesting open problem.

It is worth noting that the Chained Index problem is a promised problem. We developed the structure-vs-pseudorandomness decomposition techniques to deal with the promise. Roughly, we prove that probabilities under promise can be bounded by the probabilities under uniform distribution during our main decomposition and sampling process.

1.1.3 Applications to Streaming Problems

Since many streaming lower bounds were built on the hardness of the Tree Pointer Jumping problem or the Chained Index problem, we automatically improve these steaming lower bounds. We list some applications below.

Corollary 1.9. *Any algorithm for the explicit vertex stream model that finds a c-approximation to $\alpha(G)$ with probability at least $2/3$ requires $\Omega(n^2/c^6)$ space.*

This corollary improves the previous lower of $\Omega(n^2/c^7)$ given by [CDK18].

Corollary 1.10. *For every $\epsilon > 0$, there is an integer $p_0 \geq 2$ such that the following holds for any randomized p-player protocol for Max-Card-k with $k = p \geq p_0$. If the protocol has an approximation guarantee of $1/2+\epsilon$, then one of the players sends a message of length at least $\Omega(N\epsilon/p^2)$.*

This corollary improves the previous lower of $\Omega(N\epsilon/p^3)$ given by [FNFSZ20].

Corollary 1.11. *A data stream algorithm for Submodular Maximization subject to k Matroid Constraints, whose only access to the matroids is via the common independence oracle, and with expected approximation ratio $k - \epsilon$ for some $\epsilon \in [0, k-1)$ must use $\Omega(\frac{\epsilon n}{k^3} \log k)$ memory.*

This corollary improves the previous lower of $\Omega(\frac{\epsilon n}{k^5} \log k)$ given by [FNFSZ22].

Corollary 1.12. *For any $t \geq 2$, any algorithm for the geometric maximum independent set that can distinguish between an independent set of size 1 and t and succeeds with probability at least $2/3$ on streams of 2-intervals must use at least $\Omega(n/t^2)$ bits of memory.*

This corollary improves the previous lower of $\Omega(n/t^3)$ given by [BKO22].

Corollary 1.13. *Any one-pass constant error randomized algorithm with approximation factor α for Interval Selection on weighted arbitrary-length intervals in insertion-only streams needs $\Omega(\frac{1}{\alpha} \cdot \min\{\Delta^{\frac{1}{2\alpha}}, \frac{n}{2^{2\alpha}}\})$ bits of space.*

This corollary improves the previous lower of $\Omega(\frac{1}{\alpha^2} \cdot \min\{\Delta^{\frac{1}{2\alpha}}, \frac{n}{2^{2\alpha}}\})$ given by [DDK23].

Corollary 1.14. *Let $c > 1$ be a constant. Let A be a p-pass streaming algorithm that approximates the optimum value of SET-COVER$_{n,m}$ over instances to a factor smaller than $n^{1/(p+1)}/(c(p+1)^2)$ with probability at least $2/3$. Then A must use $\Omega(n^c/p^2)$ bits of space. This space lower bound applies to instances with $m = \Theta(n^{cp})$.*

This corollary improves the previous $\Omega(n^c/p^3)$ space lower bound given by [CW16].

Beyond direct applications, we believe our novel method could provide more applications for streaming problems.

Paper organization. In Section 2, we give preliminaries. Section 3 shows an almost tight bound for the Tree Pointer Jumping problem. Here, we show the lower bound in Section 3.1 and the upper bound in Section 3.2. In Section 4, we prove an improved lower bounds for the Chained Index problem.

2 Preliminary

Notations. We use capital letters X to denote a set and bold symbols like $\boldsymbol{R}$ to denote random variables. For a set X, we use $\boldsymbol{X}$ to denote the random variable uniformly distributed over the set X. We introduce the formal definition of *min-entropy* and *min-entropy deficiency*.

Definition 2.1 (Min-entropy)**.** The min-entropy of a random variable $\boldsymbol{R}$ is defined by

$$\mathrm{H}_\infty(\boldsymbol{R}) \stackrel{\text{def}}{=} \min_{x \in \mathrm{supp}(\boldsymbol{R})} \log\left(\frac{1}{\mathbf{Pr}[\boldsymbol{R} = x]}\right)$$

Definition 2.2 (Min-entropy deficiency)**.** Let $\boldsymbol{R}$ be a random variable on a domain U. Its min-entropy deficiency is defined by $D_\infty(\boldsymbol{R}) \stackrel{\text{def}}{=} \log|U| - \mathrm{H}_\infty(\boldsymbol{R})$,

In structure-vs-pseudorandomness decomposition, one of the most important notions, which captures the pseudorandomness, is the block-wise density.

Definition 2.3 (Block-wise density [GLM+16])**.** For $\gamma > 0$. A random variable $\boldsymbol{X}$ supported on U^n is said to be *γ-dense* if for all nonempty $I \subseteq [n]$, we have that $\mathrm{H}_\infty(\boldsymbol{X}(I)) \geq \gamma \cdot |I| \cdot \log|U|$, here $\boldsymbol{X}(I)$ is the marginal distribution of $\boldsymbol{X}$ on the set I.

The following lemma tells us that a flat distribution could be decomposed by a combination of random variables with dense properties by fixing some positions:

Lemma 2.4 (Density-restoring partition)**.** *Let $\gamma \in (0,1)$. Let X be a subset of $[N]^M$ and $J \subseteq [M]$. Suppose that there exists an $\beta \in [N]^{\overline{J}}$ such that $\forall x \in X, x(\overline{J}) = \beta$. Then, there exists a partition $X = X^1 \cup X^2 \cup \cdots \cup X^r$ and every X^i is associated with a set $I_i \subseteq J$ and a value $\alpha_i \in \{0,1\}^{I_i}$ that satisfy the following properties.*

1. *$\forall x \in X^i, x(I_i) = \alpha_i$;*
2. *$\boldsymbol{X}^i(J \setminus I_i)$ is γ-dense;*
3. *$D_\infty\left(\boldsymbol{X}^i(J \setminus I_i)\right) \leq D_\infty\left(\boldsymbol{X}(J)\right) - (1-\gamma)|I_i| \log N + \delta_i$, where $\delta_i \stackrel{\text{def}}{=} \log(|X|/|\cup_{j \geq i} X^j|)$.*

Proposition 2.5. *Let $Z_1, \ldots, Z_T$ be a partition of set Z. Then*

$$\sum_{i=1}^{T} \frac{|Z_i|}{|Z|} \cdot \log|Z_i| \geq \log|Z| - \log T.$$

3 Tree Pointer Jumping

Recall that the Tree Pointer Jumping problem is associated with a k-level t-ary tree T and a function $\phi : V(T) \to [t]$. Here, the i-th player receives the labels of the i-th level nodes.

For each node v in the i-th level of T, we represent it by a number in $[t^{i-1}]$. Hence, the input space of the Tree Pointer Jumping could be written $[t]^{t^0} \times \cdots \times [t]^{t^{k-2}} \times [t]^{t^{k-1}}$. For each input $x_i = (x_i(1), \ldots, x_i(t^{i-1})) \in [t]^{t^{i-1}}$, it is indeed a function from $[t^{i-1}]$ to $[t]$, i.e., the i-th level labels. For an input $(x_1, \ldots, x_t)$, recall the associated function $g_\phi : V(T) \to V(T) \cup [t]$ is defined by

$$g_\phi(v) = \begin{cases} \text{the } \phi(v)\text{-th child of } v & \text{if } v \text{ is not a leaf;} \\ \phi(v) & \text{otherwise.} \end{cases}$$

Let v_1 be the root of the tree T. For an $\ell > 1$, we define $v_{\ell,\phi} = g_\phi(v_{\ell-1,\phi})$. When ϕ is clear in the context, we simply shorthand as $(v_1, \ldots, v_k)$. The goal is to compute $TPJ(\phi) = \phi(v_k) = x_k(v_k)$.

3.1 Proof of Lower Bounds

In communication protocols, it is known that the players' messages partition input space into many rectangles. Our proof's main idea is to decompose these into some nice rectangles further. Formally, we define the pseudorandom rectangles below.

Definition 3.1. For $\gamma > 0$ and $i \geq 1$, we say that a set $X_i \subseteq [t]^{t^{i-1}}$ is γ-structure if there is a set $I \subseteq [t^{i-1}]$ and $\alpha \in \{0,1\}^I$ such that,

1. $\forall x_i \in X_i, x_i(I) = \alpha$, here we call I the set of fixed coordinates
2. $X_i([t^{i-1}] \setminus I)$ is γ-dense.

We say that a rectangle $R := X_1 \times \cdots \times X_k$ is γ-structure if all sets $X_1, \ldots, X_k$ are γ-structure. Notice that for each γ-structure R, there is a list of sets of fixed coordinates $I_1, \ldots, I_k$.

Let μ_k be the uniform distribution on the input space. Recall that our main goal is to prove the, for any $(k-1)$-round communication protocol Π with

$$\Pr_{\phi=(x_1,\ldots,x_t)\leftarrow\mu_k}[v_{k,\phi} = \Pi(x_1, \ldots, x_k)] \geq 2/3,$$

we have that $R_{\max}(\Pi) = \Omega(t/k)$. To prove this result, we consider the following sampling process (Algorithm 1). In addition to sampling $(x_1, \ldots, x_k)$, we also sample some auxiliary random variables in this process to help the analysis.

Algorithm 1: The decomposition and sampling process

$R^{(0)} \leftarrow R^{\text{root}}$;
$J_i^{(0)} \leftarrow [t^{i-1}]$ for $i = 1, \cdots, k$;
for $\ell \leftarrow 1$ **to** k **do**
  $X_1 \times \cdots \times X_k \leftarrow R^{(\ell-1)}$;
  for $i \leftarrow k$ **to** $\ell + 1$ **do**
    let $c_{\ell,i}$ be the total communication bits of player i in the ℓ-th round;
    X_i is partitioned into $2^{c_{\ell,i}}$ rectangles $X_i^1, \ldots, X_i^{2^{c_{\ell,i}}}$ by the protocol message;
    sample a random element $\boldsymbol{j}$, which is equal to $j \in [2^{c_{\ell,i}}]$ w.p. $|X_j^j|/|X_i|$;
    $Y_i \leftarrow X_i^j$;
    decompose Y_i by Lemma 2.4 with $J = J_i^{(\ell-1)}$, get $Y^1, \cdots, Y^r, I_1, \cdots, I_r$;
    sample a random element $\boldsymbol{s}$, which is equal to $s \in [r]$ w.p. $|Y^s|/|Y_i|$;
    $Z_i \leftarrow Y^s$;
    $I_i^\ell \leftarrow I_s$ ▷ store the newly fixed indices;
    $J_i^{(\ell)} \leftarrow J_i^{(\ell-1)} \setminus I_i^{(\ell)}$ ▷ store the unfixed indices;
  uniformly sample an element $\boldsymbol{x}_\ell$ from X_ℓ;
  $R^{(\ell)} \leftarrow X_1 \times \cdots \times X_{\ell-1} \times \{\boldsymbol{x}_\ell\} \times Z_{\ell+1} \times \cdots \times Z_k$;
Output the only element $(x_1, \ldots, x_k)$ in $R^{(k)}$.

We observe that the output distribution by Algorithm 1 is exactly the uniform distribution μ_k. In addition to sampling μ_k, Algorithm 1 also samples random variables $\boldsymbol{R}^{(\ell)}, \boldsymbol{J}_i^{(\ell)}$ for $i > \ell \geq 1$. The crux is that in each round ℓ, we maintain a rectangle $R^{(\ell)}$ such that

- We fix the first ℓ players' inputs. Hence $(v_1, \ldots, v_{\ell+1})$ are determined;
- $R^{(\ell)}$ is a γ-structure, with alive coordinates $J_i^{(\ell)}$ for every $i > \ell$.

In order to prove the lower bound, our goal is to show that in the round ℓ, the players are hard to guess $(v_{\ell+1}, \ldots, v_k)$ even with the knowledge of $v_1, \ldots, v_\ell$. On the other hand, since $R^{(\ell)}$ is a γ-structure, it is sufficient to that with high probability, the event $v_i \in J_i^{(\ell)}$ will happen. We formalize the intuition by following lemmas.

Lemma 3.2. *Let $\gamma = 1 - \frac{1}{\log t}$. For each $\ell > 1$,*

$$\mathbf{Pr}\left[\exists i > \ell,\ v_\ell \in J_\ell^{(\ell-1)} \wedge\ v_i \in I_i^{(\ell)}\right] \leq \frac{2}{t}\,\mathbf{E}\left[\sum_{i>\ell}\left|\boldsymbol{I}_i^{(\ell)}\right|\right]$$

Proof. In order to prove this lemma, we consider a tuple $\Gamma \stackrel{\text{def}}{=} \left(R^{(\ell-1)}, I_{\ell+1}^{(\ell)}, \ldots, I_k^{(\ell)}, B\right)$, where

$B \stackrel{\text{def}}{=} \{v \text{ in the } (\ell+1)\text{-th level of the tree} : v \text{ is the ancestor of some nodes from } I_i^{(\ell)},\ \text{for some } i > \ell\}$

Now for any fixed tuple Γ, we have that

$$\mathbf{Pr}\left[\exists i > \ell,\ v_\ell \in J_\ell^{(\ell-1)} \wedge\ v_i \in I_i^{(\ell)} \,\middle|\, \Gamma\right] = \mathbf{Pr}\left[v_\ell \in J_\ell^{(\ell-1)} \wedge\ v_{\ell+1} \in B \,\middle|\, \Gamma\right]$$

Notice that if $v_\ell \in J_\ell^{(\ell-1)}$ happens, the random variable $\boldsymbol{v}_{\ell+1}$ is γ-dense. It implies that for any v,

$$\mathbf{Pr}\left[v_\ell \in J_\ell^{(\ell-1)}\ \wedge\ \boldsymbol{v}_{\ell+1} = v \,\middle|\, \Gamma\right] \leq t^{-\gamma} \leq \frac{2}{t}$$

Hence we have that,

$$\mathbf{Pr}\left[v_\ell \in J_\ell^{(\ell-1)} \wedge\ v_{\ell+1} \in B \,\middle|\, \Gamma\right] \leq \frac{2}{t} \cdot |B| \leq \frac{2}{t}\sum_{i>\ell}\left|I_i^{(\ell)}\right|$$

Now by taking the expectation on the choices of Γ, we then finish the proof. □

We then upper bound $\left|\boldsymbol{I}_i^{(\ell)}\right|$ in the following lemma.

Lemma 3.3. *Let Π be a $(k-1)$-round protocol such that players communicate at most c bits in each round, then we have that,*

$$\mathbf{E}\left[\sum_{\ell=1}^{k-1}\sum_{i>\ell}\left|\boldsymbol{I}_i^{(\ell)}\right|\right] \leq 2kc$$

Proof. Consider in the ℓ-th round of communication. For every player $i > \ell$, the rectangle X_i is partitioned into $X_i^1, \ldots, X_i^{2^{c_{\ell,i}}}$ and that $Y_i \leftarrow X_i^j$ with probability $\frac{|X_i^j|}{|X_i|}$. In this process, we have that,

$$\mathop{\mathbf{E}}_j\left[D_\infty\left(Y_i(J_i^{(\ell-1)})\right) - D_\infty\left(X_i(J_i^{(\ell-1)})\right)\right] = \sum_{j=1}^{2^{c_i}} \frac{\left|X_i^j(J_i^{(\ell-1)})\right|}{\left|X_i(J_i^{(\ell-1)})\right|} \log\left(\frac{\left|X_i(J_i^{(\ell-1)})\right|}{\left|X_i^j(J_i^{(\ell-1)})\right|}\right) \leq c_{\ell,i}$$

Then by the density restoring lemma (Lemma 2.4), the rectangle Y_i is partitioned into $Y^1, \ldots, Y^r$ together with the index sets $I_1, \cdots, I_r$. Recall that $\boldsymbol{Z}_i \leftarrow Y^s$ with probability $\frac{|Y^s|}{|Y_i|}$. By the third item in Lemma 2.4, we have that,

$$\mathop{\mathbf{E}}_s\left[D_\infty\left(\boldsymbol{Z}_i(\boldsymbol{J}_i^{(\ell)})\right) - D_\infty\left(Y_i(J_i^{(\ell-1)})\right)\right] \leq \mathop{\mathbf{E}}_s\left[-(1-\gamma)\cdot(\log t)\cdot\left|\boldsymbol{I}_i^{(\ell)}\right|\right] + \mathop{\mathbf{E}}_s[\delta_s] = \mathop{\mathbf{E}}_s\left[-\left|\boldsymbol{I}_i^{(\ell)}\right|\right] + \mathop{\mathbf{E}}_s[\delta_s]$$

Here the last equality follows from the fact that $\gamma = 1 - \frac{1}{\log t}$. Recall that $\delta_s = \log\left(|Y_i| / \left|\cup_{p \geq s} Y^p\right|\right)$, then we have that,

$$\mathbf{E}[\delta_s] = \sum_s \frac{|Y^s|}{|Y_i|} \log\left(\left|Y^i\right| / \left|\cup_{p \geq k} Y^p\right|\right) \leq \int_0^1 \log \frac{1}{1-x} dx = 1.$$

For the case where $c_{\ell,i} = 0$, the number of partitions is only 1 and we get $\mathbb{E}[\delta_s] = 0$. Hence, we deduce that $\mathbf{E}[\delta_s] \leq c_{\ell,i}$. Combining these inequalities, we get

$$\mathop{\mathbf{E}}_{j,s}\left[D_\infty\left(Z_i(J_i^{(\ell)})\right) - D_\infty\left(X_i(J_i^{(\ell-1)})\right)\right] \leq 2c_{\ell,i} - \mathbf{E}\left[\left|I_i^{(\ell)}\right|\right]$$

By taking a summation on all players $i > \ell$, we have that

$$\sum_{i>\ell} \mathop{\mathbf{E}}_{j,s}\left[D_\infty\left(Z_i(J_i^{(\ell)})\right) - D_\infty\left(X_i(J_i^{(\ell-1)})\right)\right] \leq \sum_{i>\ell}\left(2c_{\ell,i} - \mathbf{E}\left[\left|I_i^{(\ell)}\right|\right]\right) \leq 2c - \sum_{i>\ell} \mathbf{E}\left[\left|I_i^{(\ell)}\right|\right]$$

By the definition of Algorithm 1, it is clear that for every $i > \ell$, Z_i defined in the ℓ-th round is used as X_i in the $(\ell+1)$-th round. Hence, by summing all rounds, we have

$$\sum_\ell \sum_{i>\ell} \mathbf{E}\left[\left|I_i^{(\ell)}\right|\right] \leq 2kc - \left(\sum_\ell \sum_{i>\ell} \mathop{\mathbf{E}}_{j,s}\left[D_\infty\left(Z_i(J_i^{(\ell)})\right)\right] - D_\infty(R^{\text{root}})\right)$$

Here $D_\infty(R^{\text{root}})$ is the first round deficiency and $D_\infty\left(Z_i(J_i^{(\ell)})\right)$ is the last round deficiency. By the definition of deficiency, we have that

1. $D_\infty(R^{\text{root}}) = 0$ as R^{root} is equal to the input space;
2. For all i, ℓ, $D_\infty\left(Z_i(J_i^{(\ell)})\right) \geq 0$ as the deficiency is always non-negative.

Together, we conclude that $\sum_\ell \sum_{i>\ell} \mathbf{E}\left[\left|I_i^{(\ell)}\right|\right] \leq 2kc$. □

Combining these two lemmas, we can then prove the main theorem.

Theorem 3.4 (Restate of Theorem 1.3)**.** *Let μ_k denote the uniform distribution over the input space. Then for any $(k-1)$-round protocol Π with*

$$\mathop{\mathbf{Pr}}_{\phi=(x_1,\ldots,x_t)\leftarrow\mu_k}[v_k = \Pi(x_1,\ldots,x_k)] \geq 2/3,$$

we have that $R_{\max}(\Pi) = \Omega(t/k)$.

Proof. We consider an event E defined by $E := \left(v_2 \in J_2^{(1)}, \cdots, v_k \in J_k^{(k-1)}\right)$. Then we have that

$$\mathbf{Pr}[v_k = \Pi(x_1,\ldots,x_k)] = \mathbf{Pr}[v_k = \Pi(x_1,\ldots,x_k) \wedge E] + \mathbf{Pr}[v_k = \Pi(x_1,\ldots,x_k) \wedge \neg E]$$

By Lemma 3.2 and Lemma lemma 3.3,

$$\begin{aligned}
\mathbf{Pr}[\neg E] &= \mathbf{Pr}\left[v_{\ell+1} \notin J_{\ell+1}^{(\ell)} \text{ for some } \ell \in [k-1]\right] \\
&= \mathbf{Pr}\left[\bigcup_{\ell=1}^{k-1} \left\{v_i \in I_i^{(\ell)} \text{ for some } i > \ell \ \wedge \ v_\ell \in J_\ell^{(\ell-1)}\right\}\right] \\
&\leq \sum_{\ell=1}^{k-1} \mathbf{Pr}\left[v_i \in I_i^{(\ell)} \text{ for some } i > \ell \ \wedge \ v_\ell \in J_\ell^{(\ell-1)}\right] && \text{(by union bound)} \\
&\leq \sum_{\ell=1}^{k-1} \frac{2}{t} \mathbf{E}\left[\sum_{i>\ell} \left|I_i^{(\ell)}\right|\right] && \text{(by Lemma 3.2)} \\
&\leq O\left(\frac{kc}{t}\right) && \text{(by Lemma 3.3)}
\end{aligned}$$

On the other hand, if the event E happens, i.e., $v_2 \in J_2^{(1)}, \cdots, v_k \in J_k^{(k-1)}$, it implies that after the $(k-1)$-th round communication,

$$\mathrm{H}_\infty(Z_k(v_k)) \geq \left(1 - \frac{1}{\log t}\right) \log t = \log t - 1$$

Notice that $(k-1)$-th round communication is actually the last round of communication. It implies that

$$\mathbf{Pr}[v_k = \Pi(x_1, \ldots, x_k) \mid E] \leq 2^{-\log t + 1} = 2/t,$$

Combining the two observations, we see that the overall success probability of the protocol is

$$\mathbf{Pr}[v_k = \Pi(x_1, \ldots, x_k)] \leq O\left(\frac{kc}{t}\right)$$

Hence, the protocol can succeed with probability $2/3$ only if $c = \Omega\left(\frac{t}{k}\right)$. □

3.2 A Communication Protocol Matches the Lower Bounds

Now we present a protocol Π with $R_{\max}(\Pi) = \widetilde{O}(t/k)$.

Algorithm 2: Tree Pointer Jumping Protocol

Input: An input $\phi = (x_1, \ldots, x_k)$ where each player i receives the input x_i.
Output: The value $s \in [t]$

for $\ell \leftarrow 1$ **to** $k-1$ **do**
  if *the* $(\ell+1)$*-th player can not fully determine* $v_{\ell+1}$ **then**
    Player $(\ell+1)$ selects a random subset S_ℓ of size $|S_\ell| = 2t/k$ from the children of v_ℓ, and writes $\{(u, x_{\ell+1}(u))\}_{u \in S_\ell}$ to the blackboard ;
    Player ℓ writes $x_\ell(v_\ell)$ to the blackboard ▷ Note that $x_\ell(v_\ell)$ is $v_{\ell+1}$;
    All players compute whether $v_{\ell+1} \in S_\ell$. If so, they compute $v_{\ell+2} = x_{\ell+1}(v_{\ell+1})$;
  else
    Player $(\ell+1)$ writes $v_{\ell+2} = x_{\ell+1}(v_{\ell+1})$ to the blackboard;

Theorem 3.5. *There exists a protocol (Algorithm 2) that computes the Tree Pointer Jumping problem with $\widetilde{O}(t/k)$ communication bits per round.*

Proof. The communication cost of Algorithm 2 in each round is at most

$$\frac{2t}{k} \cdot \log t + \log t = \widetilde{O}(t/k).$$

We observe that if there is a round ℓ such that $v_{\ell+1} \in S_\ell$, the players could correctly compute the output. Hence

$$\mathbf{Pr}[v_k \neq \Pi(x_1, \ldots, x_k)] \leq \mathbf{Pr}\left[\forall \ell, v_{\ell+1} \notin S_\ell\right] \leq (1 - 2/k)^k \leq 0.3$$

The theorem then follows. □

4 Lower Bounds for Chained Index

Recall that in the Chained Index problem, the player i receives an input $z_i = (\sigma_i, x_i) \in [n] \times \{0,1\}^n$. The players aim to compute $x_{k-1}(\sigma_k)$ through a one-way communication. In this section, we show an improved lower bound for the Chained Index problem. In light of Yao's principle, we consider the following hard distribution.

The distribution χ_k

1. Uniformly sample $\sigma_1, ..., \sigma_k \in [n]$.
2. Sample $(x_1, \ldots, x_k) \leftarrow (\{0,1\}^n)^k$ conditioned on $x_1(\sigma_2) = \cdots = x_{k-1}(\sigma_k)$.
3. Output $z = (z_1, \ldots, z_k)$ where $z_i = (\sigma_i, x_i)$ for every $i \in [k]$.

For a subset $R \subseteq ([n] \times \{0,1\}^n)^k$, define the weight of R under χ_k as

$$\chi_k(R) \stackrel{\text{def}}{=} \mathop{\mathbf{Pr}}_{z \leftarrow \chi_k}[z \in R] = \frac{\#\{((\sigma_1, x_1), \ldots, (\sigma_k, x_k)) \in R : x_1(\sigma_2) = \cdots = x_{k-1}(\sigma_k)\}}{n^k \cdot 2^{(k-1)(n-1)+n+1}}.$$

We prove the following lower bound.

Theorem 4.1. *Let $\varepsilon \in (0, 1/4]$. Let Π be a protocol that has 2ε advantage, i.e.,*

$$\mathop{\mathbf{Pr}}_{z=(\sigma_1,x_1,\ldots,\sigma_k,x_k)\leftarrow\chi_k}[\Pi(z) = x_{k-1}(\sigma_k)] \geq \frac{1}{2} + 2\varepsilon.$$

Then we have that $\mathrm{CC}(\Pi) \geq \frac{\varepsilon^2}{8} \cdot n/k - k$.

Extension to the large k regime for oblivious protocols. Note that the above lower bound is vacuous when $k = \omega(\sqrt{n})$. If we are restricted to *oblivious protocols*, i.e., each message in a protocol has a predetermined length independent of the input, then we can strengthen the lower bound so as to work with large k's by a simple reduction. Formally, we have

Theorem 4.2. *Let $\varepsilon \in (0, 1/4]$. If Π is an oblivious protocol for the Chained Index problem that has 2ε advantage, then $\mathrm{CC}(\Pi) = \Omega(n/k + \sqrt{n})$.*

We use a decomposition and sampling process DS, as shown in Algorithm 3, in our analysis. DS takes as input a protocol Π, and samples a rectangle R that is contained in Π_v for some leaf node v. Our proof proceeds in two steps:

1. Section 4.1 shows that the accuracy of Π is captured by a quantity called *average fixed size*, which is a natural quantity that arises in the running of DS.
2. Section 4.2 proves that the average fixed size can be bounded from above by $O(k \cdot \mathrm{CC}(\Pi))$. Consequently, if Π enjoys high accuracy, we get a lower bound of $\mathrm{CC}(\Pi)$.

We first recall some basic definitions.

k-party one-way protocols. A deterministic k-party one-way communication protocol Π is specified by a rooted binary tree. For every internal vertex v,

- it has 2 children, denoted by $\Pi(v, 0)$ and $\Pi(v, 1)$;
- v is owned by some part — we denote the owner by $\mathsf{owner}(v) \in [k]$;
- every leaf node specifies an output.

Starting from the root, the owner of the current node cur partitions its input space into two parts X_0 and X_1, and sets the current node to $\Pi(\mathsf{cur}, b)$ if its input belongs to X_b.

The *communication complexity* of Π, denoted by $\mathrm{CC}(\Pi)$, is the depth of the tree. On a path from root to some leaf, each time the owner switches, we call it a new *round*; in a one-way protocol, the label of the owner is non-decreasing.

Fact 4.3. *The set of all inputs that leads to an internal vertex v is a rectangle, denoted by $\Pi_v = X_1 \times \cdots \times X_k$.*

Normalized Protocol. We normalized a protocol Π as follows so as to make it defined on all inputs including those not in $\mathrm{supp}(\chi_k)$. For the i-th party, given input $(\sigma_i, x_i) \in [n] \times \{0,1\}^n$ and previous transcripts trans, output 0 if the input is *invalid*, i.e., given trans, there is no input in $\mathrm{supp}(\chi_k)$ matches x_i. Otherwise the i-th party outputs 1 and proceed as Π. Clearly, by normalizing we communicate k more bits.

Lemma 4.4 (Loop Invariant)**.** *After each iteration in Algorithm 3,*

- *$R \subseteq \Pi_v$;*
- *for all $i \in [k]$, $\boldsymbol{X}_i(J_i)$ is γ-dense.*

Proof. The first item is true because every time v is updated, R is updated accordingly to a sub-rectangle of Π_v and updating R into its sub-rectangles does not violate this condition.

Since we applied density restoring partition at the end of each iteration, the second item is guaranteed by Lemma 2.4 and the way that X_i, J_i are updated. □

Algorithm 3: Decomposition and Sampling Process DS

Input: A protocol Π

Output: A rectangle $R = (\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_k\} \times X_k)$ and k sets $J_1, \ldots, J_k \subseteq [n]$.

for $i \in [k]$ **do**
  $X_i := \{0,1\}^n, J_i := [n]$. `// Initialization`
Sample $\sigma_1 \leftarrow [n]$.
$v :=$ root of $\Pi, R := (\{\sigma_1\} \times \{0,1\}^n) \times ([n] \times \{0,1\}^n)^{k-1}$, bad := FALSE.
while *v is not a leaf node* **do**
  $i := \mathsf{owner}(v), u_0 := \Pi(v,0), u_1 := \Pi(v,1)$.
  `//Loop invariant: (1)` $R \subseteq \Pi_v$`; (2)` $\boldsymbol{X}_i(J_i)$ `is` γ`-dense.`
  X_i is partitioned into $X_i = X^0 \cup X^1$ according to Π.
  Sample $\boldsymbol{b}$ such that $\mathbf{Pr}[\boldsymbol{b} = b] = \frac{\chi_k(R^b)}{\chi_k(R)}$ where
  $R^b = (\{\sigma_1\} \times X_1) \times \cdots (\{\sigma_i\} \times X^b) \times ([n] \times \{0,1\}^n)^{k-i}$ for $b \in \{0,1\}$.
  `//`R `is always a shorthand for` $(\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_i\} \times X_i) \times ([n] \times \{0,1\}^n)^{k-i}$
  Update $X_i := X^{\boldsymbol{b}}$.
  Let $X_i = X^1 \cup \cdots \cup X^m$ be the decomposition of X_i promised by Lemma 2.4 with associated sets $I_1, \ldots, I_m \subseteq J_i$. `// Invoking Lemma 2.4 with` $\gamma = 1 - \frac{2\varepsilon}{k}, J = J_i, N = 2$.
  Sample $\boldsymbol{j} \in [m]$ such that $\mathbf{Pr}[\boldsymbol{j}' = j] = \frac{\chi_k(R^j)}{\chi_k(R)}$ where
  $R^j = (\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_i\} \times X^j) \times ([n] \times \{0,1\}^n)^{k-i}$ for $j \in [m]$.
  Update $X_i := X^j, J_i := J_i \setminus I_j$.
  if $\mathsf{owner}(u_{\boldsymbol{b}}) \neq i$ **then**
    Sample $\sigma_{i+1} \in [n]$ such that $\mathbf{Pr}[\sigma_{i+1} = \rho] = \frac{\chi_k(R^\rho)}{\chi_k(R)}$ where
    $R^\rho = (\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_i\} \times X_i) \times (\{\rho\} \times \{0,1\}^n) \times ([n] \times \{0,1\}^n)^{k-i-1}$ for $\rho \in [n]$.
    if $\sigma_{i+1} \notin J_i$ **then** bad := TRUE;
  Output $R = (\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_k\} \times X_k)$.

4.1 Relating Accuracy and Average Fixed Size

Lemma 4.5 (Relating accuracy and avarage fixed size). *Assume that* $\gamma \geq \log\left[1 + \left(\frac{1-2\varepsilon}{1+2\varepsilon}\right)^{1/k}\right]$. *Then*

$$\Pr_{z=(\sigma_1,x_1,\ldots,\sigma_k,x_k)\leftarrow\chi_k}[\Pi(z) = x_{k-1}(\sigma_k)] \leq \frac{1}{2} + \varepsilon + \frac{4}{n} \cdot \mathop{\mathbf{E}}_{(R,J_1,\ldots,J_k)\leftarrow \mathsf{DS}(\Pi)}\left[\sum_{j\in[k]} |\overline{J_i}|\right].$$

Remark 4.6. $\gamma \stackrel{\text{def}}{=} 1 - \frac{2\varepsilon}{k}$ satisfies the condition. Indeed,

$$\log\left[1 + \left(\frac{1-2\varepsilon}{1+2\varepsilon}\right)^{1/k}\right] \leq \left(\frac{1-2\varepsilon}{1+2\varepsilon}\right)^{1/k} \leq 1 - \frac{1}{k} \cdot \frac{4\varepsilon}{1+2\varepsilon} \leq 1 - \frac{2\varepsilon}{k},$$

where the first inequality is by $\log(1+x) \leq x$, and the second is by $(1-x)^r \leq 1 - rx$ for $x \in (-1, 0)$ and $r \in (0, 1)$.

The proof of the lemma is obtained through the following two lemmas. The first lemma readily says that conditioned on the flag bad is not raised, Π has little advantage in the rectangle R output by $\mathsf{DS}(\Pi)$. The second lemma shows the probability that the flag is raised is bounded in terms of the average fixed size.

Lemma 4.7. *If* $\mathsf{DS}(\Pi)$ *outputs* $(R, J_1, \ldots, J_k)$ *and* $\mathsf{bad} = \textsc{False}$ *in the end, then*

$$\Pr_{z=(\rho_1,x_1,\ldots,\rho_k,x_k)\leftarrow R}[\Pi(z) = x_{k-1}(\rho_k) | x_1(\rho_2) = \cdots = x_{k-1}(\rho_k)] \leq \frac{1}{2} + \varepsilon.$$

Lemma 4.8. $\Pr_{\mathsf{DS}(\Pi)}[\mathsf{bad} = \textsc{True}] \leq \frac{4}{n} \cdot \mathbf{E}_{(R,J_1,\ldots,J_k)\leftarrow\mathsf{DS}(\Pi)}\left[\sum_{j\in[k]} |\overline{J_k}|\right]$.

Next, we first prove Lemma 4.5 using the above two lemma.

Proof of Lemma 4.5. Note that in the running of $\mathsf{DS}(\Pi)$, we first sample $\sigma_1, \ldots, \sigma_k \leftarrow [n]$ and then always update R to a randomly chosen rectangle; the probability of each rectangle being chosen is proportional to its weight under χ_k. Consequently,

$$\begin{aligned}
&\Pr_{z=(\sigma_1,x_1,\ldots,\sigma_k,x_k)\leftarrow\chi_k}[\Pi(z) = x_{k-1}(\sigma_k)] \\
&= \Pr_{\substack{(R,J_1,\ldots,J_k)\leftarrow\mathsf{DS}(\Pi)\\(\sigma_1,x_1,\ldots,\sigma_k,x_k)\leftarrow R}}[\Pi(z) = x_{k-1}(\sigma_k) \mid x_1(\sigma_2) = \cdots = x_{k-1}(\sigma_k)] \\
&\leq \Pr_{\mathsf{DS}(\Pi)}[\mathsf{bad} = \textsc{True}] + \Pr_{\substack{(R,J_1,\ldots,J_k)\leftarrow\mathsf{DS}(\Pi)\\z=(\sigma_1,x_1,\ldots,\sigma_k,x_k)\leftarrow R}}[\Pi(z) = x_{k-1}(\sigma_k) | x_1(\sigma_2) = \cdots = x_{k-1}(\sigma_k) \wedge \mathsf{bad} = \textsc{False}] \\
&\leq \frac{1}{2} + \varepsilon + \frac{4}{n} \cdot \mathop{\mathbf{E}}_{(R,J_1,\ldots,J_k)\leftarrow\mathsf{DS}(\Pi)}\left[\sum_{i=1}^{k} |\overline{J_i}|\right].
\end{aligned}$$

where the last step is by lemma 4.7 and lemma 4.8. □

It remains to prove the two lemmas.

Proof of lemma 4.7. Say $R = (\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_k\} \times X_k)$. Since bad = FALSE in the end, we have $\sigma_{i+1} \in J_i$ for all $i \in [k-1]$. By lemma 4.4, we have $\mathrm{H}_\infty(\boldsymbol{X}_i(\sigma_{i+1})) \geq \gamma$ for all i. Since R is contained in some leaf node of Π, Π output the same answer in R, say $b^* \in \{0,1\}$. Note that for $b \in \{0,1\}$,

$$\Pr_{z=(\rho_1,x_1,\ldots,\rho_k,x_k)\leftarrow R}[x_1(\rho_2) = \cdots = x_{k-1}(\rho_k) = b] = \prod_{i\in[k-1]} \Pr_{x_i\leftarrow X_i}[x_i(\sigma_{i+1}) = b],$$

since we must have $\rho_i = \sigma_i$. Write $p_i \stackrel{\text{def}}{=} \Pr_{x^i\leftarrow X^i}\left[x^i(\sigma_i) = b^*\right]$. Then we have

$$\begin{aligned}
&\Pr_{z=(\rho_1,x_1,\ldots,\rho_k,x_k)\leftarrow R}[\Pi(z) = x_{k-1}(\rho_k) | x_1(\rho_2) = \cdots = x_{k-1}(\rho_k)] \\
=& \Pr_{z=(\rho_1,x_1,\ldots,\rho_k,x_k)\leftarrow R}[x_1(\rho_2) = \cdots = x_{k-1}(\rho_k) = b^* | x_1(\rho_2) = \cdots = x_{k-1}(\rho_k)] \\
=& \Pr_{z=(\rho_1,x_1,\ldots,\rho_k,x_k)\leftarrow R}[x_1(\sigma_2) = \cdots = x_{k-1}(\sigma_k) = b^*] \Big/ \Pr_{z=(\rho_1,x_1,\ldots,\rho_k,x_k)\leftarrow R}[x_1(\sigma_2) = \cdots = x_{k-1}(\sigma_k)] \\
=& \frac{\prod_{i\in[k]} p_i}{\prod_{i\in[k]} p_i + \prod_{i\in[k]}(1-p_i)} = \frac{1}{1 + \prod_{i\in[k]}(1/p_i - 1)}.
\end{aligned}$$

Since $\mathrm{H}_\infty(\boldsymbol{X}_i(\sigma_{i+1})) \geq \gamma$ for all i, we have $p_i \in [1-2^{-\gamma}, 2^{-\gamma}]$, which implies

$$\frac{1}{1 + \prod_{i\in[k]}(1/p_i - 1)} \leq \frac{1}{1 + (2^\gamma - 1)^k}.$$

Since we assumed $\gamma \geq \log\left[1 + \left(\frac{1-2\varepsilon}{1+2\varepsilon}\right)^{1/k}\right]$, it holds that $\frac{1}{1+(2^\gamma-1)^k} \leq \frac{1}{2} + \varepsilon$, concluding the proof. □

Proof of lemma 4.8. Let $\mathcal{B}_t$ denote the event that the flag bad is raised when $i = t$ (i.e., when the i-th round ends) for the first time. Clearly, $\Pr[\text{bad} = \text{TRUE}] = \sum_{t=1}^{k-1} \Pr[\mathcal{B}_t]$. It suffices to bound each $\Pr[\mathcal{B}_t]$.

Fix $t \in [k-1]$ and the random coins coin used for the first $(t-1)$ rounds, i.e., until Line 17 is reached with $i = t-1$). Let $R_{t-1} = (\{\sigma_1\} \times X_1 \times \cdots \times (\{\sigma_t\} \times \{0,1\}^n) \times ([n] \times \{0,1\}^n)^{k-t}$ be the value of rectangle R when running $\mathsf{DS}(\Pi)$ using coin until the t-th round begins. The core of our proof is to compare the process with one that runs under uniform weight instead of the weight under χ_k; this is why we can deal with the promise.

- Let Real_t be the process that runs $\mathsf{DS}(\Pi)$ until the t-th round begins with coin, then run the t-th round with fresh random coins.
- Let $\mathcal{U}_k$ denote the uniform distribution over the input space $([n] \times \{0,1\}^n)^k$. Consider the following process, denoted by Unif_t: run $\mathsf{DS}(\Pi)$ until the t-th round begins with coin, then run the t-th round with χ_k replaced by $\mathcal{U}_k$.

Note that during the execution of Real_t and Unif_t, the partitions are the same, and the only difference is that when choosing $\boldsymbol{b}, \boldsymbol{j}, \sigma_{t+1}$, the probabilities are different. Let $\widehat{X}_t, \widehat{J}_t, \widehat{\sigma}_{t+1}$ be a possible value of X_t, J_t, σ_{t+1} at the end of the t-th round. In Real_t we update R according to χ_k, and thus the probability that $X_t = \widehat{X}_t, \sigma_{t+1} = \widehat{\sigma}_{t+1}$ in the end of Real_t equals

$$p(\widehat{X}_i, \widehat{\sigma}_{t+1}) = \frac{\chi_k((\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_i\} \times \widehat{X}_i) \times (\{\widehat{\sigma}_{t+1}\} \times \{0,1\}^n) \times ([n] \times \{0,1\}^n)^{k-t-1})}{\chi_k(R_{i-1})}.$$

Similarly, the probability that $X_i = \widehat{X}_i, \sigma_{t+1} = \widehat{\sigma}_{t+1}$ in the end of Unif_i equals

$$q(\widehat{X}_i, \widehat{\sigma}_{t+1}) = \frac{\mathcal{U}_k((\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_i\} \times \widehat{X}_i) \times (\{\widehat{\sigma}_{t+1}\} \times \{0,1\}^n) \times ([n] \times \{0,1\}^n)^{k-t-1})}{\mathcal{U}_k(R_{i-1})} = \frac{|\widehat{X}_i|}{n2^n}.$$

The next claim reveals a connection between the two probabilities, whose proof is by direct calculation and is deferred to the appendix.

Claim 4.9. *For all possible value $\widehat{X}_i, \widehat{\sigma}_{t+1}$, $p(\widehat{X}_i, \widehat{\sigma}_{t+1}) \le 2q(\widehat{X}_t, \widehat{\sigma}_{t+1})$.*

Since J_t is determined by the value of X_t and the event $\mathcal{B}_t$ is determined by X_t and σ_{t+1}, the above claim implies that $\mathbf{Pr}_{\mathsf{Real}_t}[\mathcal{B}_t] \le 2\,\mathbf{Pr}_{\mathsf{Unif}_t}[\mathcal{B}_t]$. Note that in Unif_t, σ_{t+1} is chosen uniformly at random, and thus

$$\Pr_{\mathsf{Unif}_t}[\mathcal{B}_t] \le \mathop{\mathbf{E}}_{\mathsf{Unif}_t}\left[|\overline{J_t}|\right]/n \le 2 \mathop{\mathbf{E}}_{\mathsf{Real}_t}\left[|\overline{J_t}|\right]/n.$$

Taking expectation over coin we get $\mathbf{Pr}[\mathcal{B}_t] \le \frac{4}{n} \cdot \mathbf{E}_{\mathsf{DS}(\Pi)}\left[|\overline{J_t}|\right]$, as desired. □

4.2 Average Fixed Size is Bounded by Communication

Now that the accuracy of a protocol Π is bounded from above by the average fixed size (i.e., $\mathbf{E}_{\mathsf{DS}(\Pi)}\left[|\overline{J_1}| + \cdots + |\overline{J_k}|\right]$), in what follows we show that the average fixed size is at most $O(k \cdot \mathrm{CC}(\Pi))$. Formally, we prove that

Lemma 4.10. *Assume that Π is a normalized protocol. Then*

$$\mathop{\mathbf{E}}_{(R,J_1,\ldots,J_k)\leftarrow\mathsf{DS}(\Pi)}\left[\sum_{j\in[k]} |\overline{J_k}|\right] \le \frac{4}{1-\gamma} \cdot \mathrm{CC}(\Pi).$$

Proof. The proof strategy is similar to the proof of lemma 4.8. Fix $t \in [k-1]$ and consider $\mathbf{E}_{\mathsf{DS}(\Pi)}\left[|\overline{J_t}|\right]$ Fix the random coins coin used for the first $(t-1)$ rounds, i.e., until Line 17 is reached with $i = t-1$). Let Real_t and Unif_t be defined as in the proof of lemma 4.8. Moreover, let c_t denote the number of bits sent by the t-th party, i.e., the number of iterations in the t-th round. By a standard density increment argument (prove later), we have

Claim 4.11. $\mathbf{E}_{\mathsf{Unif}_t}\left[|\overline{J_t}|\right] \le \frac{2}{1-\gamma}\mathbf{E}_{\mathsf{Unif}_t}[c_t]$.

Since the value of J_t is determined by the value of X_t, we get

$$\mathop{\mathbf{E}}_{\mathsf{DS}(\Pi)}\left[|\overline{J_t}|\right] = \mathop{\mathbf{E}}_{\mathsf{coin},\mathsf{Real}_t}\left[|\overline{J_t}|\right] \le 2 \mathop{\mathbf{E}}_{\mathsf{coin},\mathsf{Unif}_t}\left[|\overline{J_t}|\right] \le \frac{2}{1-\gamma} \mathop{\mathbf{E}}_{\mathsf{coin},\mathsf{Unif}_t}[c_t] \le \frac{4}{1-\gamma} \mathop{\mathbf{E}}_{\mathsf{coin},\mathsf{Real}_t}[c_t].$$

where the first inequality is by claim 4.9, the second is by claim 4.11, and the last inequality holds since Π is a normalized protocol — it only writes one bit on invalid inputs. Note that $\mathbf{E}_{\mathsf{coin},\mathsf{Real}_i}[c_i]$ is the expected number of bits written by the i-th party. Hence, by summing up all i's, we get the desired result. □

It remains to prove claim 4.11.

Proof of claim 4.11. We shall prove this lemma by density increment argument. That is, we study the change of the density function $D_\infty(\boldsymbol{X}_t(J_t))$. in each iteration. Let $\boldsymbol{\phi_t}$ be the value of $D_\infty(\boldsymbol{X}_t(J_t))$ at the end of the ℓ-th iteration.

We fix the random coins used for the first $(\ell-1)$ iterations and consider the updates in the current iteration.

1. First, X_t is partitioned into $X_t = X^0 \cup X^1$ according to Π. Then, X_t is updated to X^b with probability $\frac{|X^b|}{|X|}$. Consequently, $D_\infty(\boldsymbol{X_t}(J_t))$ will increase as $|X_t|$ shrinks, and in expectation (over the random choice of $\boldsymbol{b}$) the increment is

$$\sum_{b\in\{0,1\}} \frac{|X^b|}{|X_t|} \log\left(\frac{|X_t|}{|X^b|}\right) \le 1. \tag{1}$$

2. Next, we further partition X_t according to lemma 2.4. Say X is partitioned into $X_t = X^1 \cup \cdots \cup X^m$ and let $I_1, \ldots, I_m$ be the index sets promised by lemma 2.4; and for all $j \in [m]$ we have

$$D_\infty(\boldsymbol{X}^j(J_t \setminus I_j)) \le D_\infty(\boldsymbol{X}_t(J_t)) - (1-\gamma)|I_j| + \delta_j,$$

where $\delta_j = \log(|X_t|/\cup_{v\ge j} X^v)$. With probability $p_j \stackrel{\text{def}}{=} |X^j|/|X_t|$, we update $X_t := X^j$ and $J_t := J_t \setminus I_j$. Therefore, taking expectation over the random choice of $\boldsymbol{j}$, the density function will decrease by

$$D_\infty(\boldsymbol{X_t}(J_t)) - \mathop{\mathbf{E}}_{j\leftarrow \boldsymbol{j}}\left[D_\infty(\boldsymbol{X}^j_{\boldsymbol{t}}(J_t \setminus I_j))\right] \ge \mathop{\mathbf{E}}_{j\leftarrow \boldsymbol{j}}\left[(1-\gamma)\cdot|I_j| - \delta_j\right]. \tag{2}$$

Note that $\delta_j \stackrel{\text{def}}{=} \log\frac{1}{\sum_{v\ge j} p_v}$ and thus

$$\mathop{\mathbf{E}}_{j\leftarrow \boldsymbol{j}}\left[\delta_j\right] = \sum_{j=1}^{m} p_j \log\frac{1}{\sum_{v\ge j} p_j} \le \int_0^1 \frac{1}{1-x}\mathrm{d}x \le 1. \tag{3}$$

Let $\mathcal{F}_{\ell-1}$ be the σ-algebra generated by the random coins used for the first $(\ell-1)$ iterations. Let $\boldsymbol{\beta}_\ell$ be the increment of $|\overline{J_t}|$ in the ℓ-th iteration. Observe that $\boldsymbol{\beta}_\ell = |I_j|$ by definition. By eq. (2) and eq. (3), taking expectation over random choice of $\boldsymbol{j}$, $D_\infty(\boldsymbol{X}_t(J_t))$ decrease by at least $(1-\gamma)\cdot\mathbf{E}\left[\boldsymbol{\beta}_\ell \mid \mathcal{F}_{\ell-1}\right] - 1$ due to the density restoring partition. Then

$$\mathbf{E}\left[\boldsymbol{\phi}_\ell - \boldsymbol{\phi}_{\ell-1}\right] = \mathbf{E}\left[\mathbf{E}\left[\boldsymbol{\phi}_\ell - \boldsymbol{\phi}_{\ell-1} \mid \mathcal{F}_{\ell-1}\right]\right] \le \mathbf{E}\left[1 - ((1-\gamma)\cdot\boldsymbol{\beta}_\ell - 1)\right]. \tag{4}$$

In the beginning, $\boldsymbol{\phi}_0 = D_\infty(\{0,1\}^n) = 0$. Since the density function is always non-negative by definition, we have $\boldsymbol{\phi}_{\boldsymbol{c}_t} \ge 0$ and thus $\mathbf{E}\left[\boldsymbol{\phi}_{\boldsymbol{c}_t} - \boldsymbol{\phi}_0\right] \ge 0$. On the other hand, by telescoping,

$$\mathbf{E}\left[\boldsymbol{\phi}_{\boldsymbol{c}_t} - \boldsymbol{\phi}_0\right] = \mathbf{E}\left[\sum_{\ell=1}^{\boldsymbol{c}_t}(\boldsymbol{\phi}_\ell - \boldsymbol{\phi}_{\ell-1})\right] \le \mathbf{E}\left[\sum_{\ell=1}^{\boldsymbol{c}_t}(\boldsymbol{\beta}_\ell + 2)\right],$$

where the inequality follows from eq. (4). Observe that $\sum_{t=1}^{\boldsymbol{c}_t}\boldsymbol{\beta}_t = |\overline{J_t}|$ by definition. We conclude that

$$\mathbf{E}\left[|\overline{J_t}|\right] = \mathbf{E}\left[\sum_{\ell=1}^{\boldsymbol{c}_t}\boldsymbol{\beta}_\ell\right] \le \frac{2\,\mathbf{E}\left[\boldsymbol{c}_t\right]}{1-\gamma},$$

as desired. □

4.3 Putting it Together

Now we are prepared to prove theorem 4.1.

Proof of theorem 4.1. We first normalize Π so as to make it accepts all inputs in $([n] \times \{0,1\}^n)^k$. Denoted by Π' the normalized protocol, then we have $\mathrm{CC}(\Pi') \leq \mathrm{CC}(\Pi) + k$.

Set $\gamma = 1 - \frac{2\varepsilon}{k}$. One can check that γ satisfies the requirement in lemma 4.5. By lemma 4.5 and lemma 4.10, we have

$$\texttt{Accuracy}(\Pi') \stackrel{\text{def}}{=} \Pr_{z=(\sigma_1,x_1,\ldots,\sigma_k,x_k)\leftarrow\chi_k}[\Pi'(z) = x_{k-1}(\sigma_k)] \leq \frac{1}{2} + \varepsilon + \frac{4}{n} \cdot \frac{k}{2\varepsilon} \cdot 4\mathrm{CC}(\Pi'). \tag{5}$$

Since Π', Π have the same output on valid inputs and we assumed $\texttt{Accuracy}(\Pi) \geq \frac{1}{2} + 2\varepsilon$, we get $\texttt{Accuracy}(\Pi') \geq \frac{1}{2} + 2\varepsilon$. Combining with eq. (5) we conclude that $\mathrm{CC}(\Pi') \geq \frac{\varepsilon^2 n}{8k}$, meaning that $\mathrm{CC}(\Pi) \geq \frac{\varepsilon^2 n}{8k} - k$. □

References

[ACK19] Sepehr Assadi, Yu Chen, and Sanjeev Khanna. Polynomial pass lower bounds for graph streaming algorithms. In *Proceedings of the 51st Annual ACM SIGACT Symposium on theory of computing*, pages 265–276, 2019. 1

[ALWZ20] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 624–630, 2020. 3

[AMOP08] Alexandr Andoni, Andrew McGregor, Krzysztof Onak, and Rina Panigrahy. Better bounds for frequency moments in random-order streams. *arXiv preprint arXiv:0808.2222*, 2008. 1

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 67–76, 2010. 1, 2

[BGL+24] Mark Braverman, Sumegha Garg, Qian Li, Shuo Wang, David P Woodruff, and Jiapeng Zhang. A new information complexity measure for multi-pass streaming with applications. *arXiv preprint arXiv:2403.20283*, 2024. 1

[BGW20] Mark Braverman, Sumegha Garg, and David P Woodruff. The coin problem with applications to data streams. In *2020 ieee 61st annual symposium on foundations of computer science (focs)*, pages 318–329. IEEE, 2020. 1

[BKO22] Sujoy Bhore, Fabian Klute, and Jelle J Oostveen. On streaming algorithms for geometric independent set and clique. In *International Workshop on Approximation and Online Algorithms*, pages 211–224. Springer, 2022. 4, 5, 6

[BM13] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 161–170, 2013. 3

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 746–755. IEEE, 2013. 2

[BYJKS04] Ziv Bar-Yossef, Thathachar S Jayram, Ravi Kumar, and D Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004. 1, 2, 3

[CCM08] Amit Chakrabarti, Graham Cormode, and Andrew McGregor. Robust lower bounds for communication and stream computation. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 641–650, 2008. 1, 2, 3, 4

[CDK18] Graham Cormode, Jacques Dark, and Christian Konrad. Independent sets in vertex-arrival streams. *arXiv preprint arXiv:1807.08331*, 2018. 2, 3, 4, 5, 6

[Cha07] Amit Chakrabarti. Lower bounds for multi-player pointer jumping. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 33–45. IEEE, 2007. 2, 5

[CKS03] Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *18th IEEE Annual Conference on Computational Complexity, 2003. Proceedings.*, pages 107–117. IEEE, 2003. 3

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 270–278. IEEE, 2001. 1

[CW16] Amit Chakrabarti and Anthony Wirth. Incidence geometries and the pass complexity of semi-streaming set cover. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 1365–1373. SIAM, 2016. 4, 6

[DDK23] Jacques Dark, Adithya Diddapur, and Christian Konrad. Interval selection in data streams: Weighted intervals and the insertion-deletion setting. In *43rd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2023)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2023. 4, 5, 6

[DOR21] Nachum Dershowitz, Rotem Oshman, and Tal Roth. The communication complexity of multiparty set disjointness under product distributions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1194–1207, 2021. 1, 2

[FNFSZ20] Moran Feldman, Ashkan Norouzi-Fard, Ola Svensson, and Rico Zenklusen. The one-way communication complexity of submodular maximization with applications to streaming and robustness. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1363–1374, 2020. 4, 5, 6

[FNFSZ22] Moran Feldman, Ashkan Norouzi-Fard, Ola Svensson, and Rico Zenklusen. Submodular maximization subject to matroid intersection on the fly. In *30th Annual European Symposium on Algorithms (ESA 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022. 4, 5, 6

[GLM+16] Mika Goos, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM Journal on Computing*, 45(5):1835–1869, 2016. 3, 7

[GO16] Venkatesan Guruswami and Krzysztof Onak. Superlinear lower bounds for multipass graph processing. *Algorithmica*, 76:654–683, 2016. 1

[GPW17] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for bpp. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 132–143, 2017. 3

[Gro09] Andre Gronemeier. Asymptotically optimal lower bounds on the nih-multi-party information complexity of the and-function and disjointness. In *in Proc. of the 26th International Symposium on Theoretical Aspects of Computer Science, STACS*, pages 505–516, 2009. 3

[Jay09] TS Jayram. Hellinger strikes back: A note on the multi-party information complexity of and. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 562–573. Springer, 2009. 3

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, 2002. 2

[LMM+22] Shachar Lovett, Raghu Meka, Ian Mertz, Toniann Pitassi, and Jiapeng Zhang. Lifting with sunflowers. *Leibniz international proceedings in informatics*, 215, 2022. 3

[LSZ19] Shachar Lovett, Noam Solomon, and Jiapeng Zhang. From dnf compression to sunflower theorems via regularity. In *Proceedings of the 34th Computational Complexity Conference*, pages 1–14, 2019. 3

[LZ23] Shachar Lovett and Jiapeng Zhang. Streaming lower bounds and asymmetric set-disjointness. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 871–882. IEEE, 2023. 1

[NW91] Noam Nisan and Avi Widgerson. Rounds in communication complexity revisited. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 419–429, 1991. 2, 3

[OR23] Rotem Oshman and Tal Roth. The Communication Complexity of Set Intersection Under Product Distributions. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 95:1–95:20, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 1, 2

[Raz11] Ran Raz. A counterexample to strong parallel repetition. *SIAM Journal on Computing*, 40(3):771–777, 2011. 2

[Yeh20] Amir Yehudayoff. Pointer chasing via triangular discrimination. *Combinatorics, Probability and Computing*, 29(4):485–494, 2020. 2, 3

[YZ24] Guangxu Yang and Jiapeng Zhang. Communication lower bounds for collision problems via density increment arguments. *STOC 2024 (to appear)*, 2024. 3

A Proof of claim 4.9

Claim A.1 (claim 4.9 restated). *Let $t \le k$. Let $\sigma_1, \dots, \sigma_{t+1} \in [n], X_1, \dots, X_t \subseteq \{0,1\}^n$. For $\ell \in \{t, t+1\}$, define*

$$R_\ell \stackrel{\text{def}}{=} (\{\sigma_1\} \times X_1) \times \cdots \times (\{\sigma_{\ell-1}\} \times X_{\ell-1}) \times (\{\sigma_\ell\} \times \{0,1\}^n) \times ([n] \times \{0,1\}^n)^{k-\ell}.$$

Then

$$\frac{\chi_k(R_{t+1})}{\chi_k(R_t)} \le 2\frac{\mathcal{U}_k(R_{t+1})}{\mathcal{U}_k(R_t)}.$$

Proof of claim 4.9. To start with, observe that

$$\frac{\mathcal{U}_k(R_{t+1})}{\mathcal{U}_k(R_t)} = \frac{|R_{t+1}|}{|R_t|} = \frac{|X_t|}{n2^n}. \tag{6}$$

We claim that for $\ell \in \{t, t+1\}$,

$$\chi_k(R_\ell) = \frac{\#\{(x_1, \dots, x_{\ell-1}) \in X_1 \times \cdots \times X_{\ell-1} : x_1(\sigma_2) = \cdots = x_{\ell-1}(\sigma_\ell)\}}{n^\ell \cdot 2^{(n-1)\ell+1}}. \tag{7}$$

Then we have

$$\begin{aligned}
\chi_k(R_{t+1}) &= \frac{\#\{(x_1, \dots, x_t) \in X_1 \times \cdots \times X_t : x_1(\sigma_2) = \cdots = x_t(\sigma_{t+1})\}}{n^{t+1} \cdot 2^{(n-1)(t+1)+1}} \\
&\le \frac{\#\{(x_1, \dots, x_{t-1}) \in X_1 \times \cdots \times X_{t-1} : x_1(\sigma_2) = \cdots = x_{t-1}(\sigma_t)\} \cdot |X_t|}{n^{t+1} \cdot 2^{(n-1)(t+1)+1}} \\
&= \chi_k(R_t) \cdot \frac{|X_t|}{n2^{n-1}}.
\end{aligned}$$

where the first and the third equality is from eq. (7). Combining with eq. (6) we have the desired result.

It remains to show eq. (7). Suppose that $((\rho_1, x_1), \dots, (\rho_k, x_k)) \in R_\ell$ satisfies $x_1(\rho_2) = \cdots = x_{k-1}(\rho_k)$. Then we $\rho_1 = \sigma_1, \dots, \rho_\ell = \sigma_\ell$ and

$$x_1(\sigma_2) = \cdots = x_{\ell-1}(\sigma_\ell) = b \text{ for some } b \in \{0,1\}.$$

For every $\rho_{\ell+1}, \dots, \rho_k \in [n]$, there exists exactly $2^{(n-1)}$ possible values for each x_j with $\ell \le j \le k-1$ (with one bit fixed to be b) and 2^n possible values for x_k (which is not used at all). Therefore,

$$\begin{aligned}
\chi_k(R_\ell) &= \frac{\#\{((\rho_1, x_1), \dots, (\rho_k, x_k)) \in R_{t+1} : x_1(\rho_2) = \cdots = x_{k-1}(\rho_k)\}}{n^k \cdot 2^{(n-1)(k-1)+n+1}} \\
&= \frac{n^{k-\ell} \cdot 2^{(n-1)\cdot(k-1-\ell)+n} \cdot \#\{(x_1, \dots, x_{\ell-1}) \in X_1 \times \cdots \times X_{\ell-1} : x_1(\sigma_2) = \cdots = x_{\ell-1}(\sigma_\ell)\}}{n^k \cdot 2^{(n-1)(k-1)+n+1}} \\
&= \frac{\#\{(x_1, \dots, x_{\ell-1}) \in X_1 \times \cdots \times X_{\ell-1} : x_1(\sigma_2) = \cdots = x_{\ell-1}(\sigma_\ell)\}}{n^\ell \cdot 2^{(n-1)\ell+1}},
\end{aligned}$$

which is exactly what we wanted. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il