Gadgetless Lifting Beats Round Elimination: Improved Lower Bounds for Pointer Chasing

Xinyu Mao * Guangxu Yang * Jiapeng Zhang *

April 10, 2024

Abstract

The notion of query-to-communication lifting theorems is a generic framework to convert query lower bounds into two-party communication lower bounds. Though this framework is very generic and beautiful, it has some inherent limitations such as it only applies to lifted functions. In order to address this issue, we propose gadgetless lifting, a new framework to prove communication lower bounds. The crux of the new approach is to lift lower bounds for a family of restricted protocols into lower bounds for general protocols.

To demonstrate the power of gadgetless lifting, we prove an $\Omega(n/k+k)$ communication lower bound on $(k-1)$-round distributional complexity of the $k$-step pointer chasing problem under the uniform input distribution, improving the $\Omega(n/k - k \log n)$ lower bound due to Yehudayoff (Combinatorics, Probability and Computing, 2020). Our lower bound almost matches the upper bound of $\tilde{O}(n/k + k)$ communication by Nisan and Wigderson (STOC 91).

A key step in gadgetless lifting is how to choose the definition of restricted protocols. We start with proving communication lower bounds for restricted protocols, and then lift it into general settings. In this paper, our definition of restricted protocols is inspired by the structure-vs-pseudorandomness decomposition by Göös, Pitassi, and Watson (FOCS 17) and Yang and Zhang (STOC 24).

Previously, round-communication trade-offs were mainly obtained by round elimination and information complexity. However, both methods have some obstacles in some settings. In general, we believe gadgetless lifting provides a new solution to address previous barriers by round elimination and information complexity.

1 Introduction

Pointer chasing is a well-known problem [RY20] that demonstrates the power of interaction in communication and has broad applications in different areas. It was used for proving monotone constant-depth hierarchy theorem [NW91, KPPY84], lower bounds on the time complexity of distributed computation [NDSP11], lower bounds on the space complexity of streaming algorithms [FKM*09, GO16, ACK19], adaptivity hierarchy theorem for property testing [CG18], exponential

\*Research supported by NSF CAREER award 2141536.
Thomas Lord Department of Computer Science, University of Southern California.
Email: \{xinyumao, guangxuy, jiapengz\}@usc.edu
separations in local differential privacy [JMR20], memory bounds for continual learning [CPP22] and limitations of the transformer architecture [PNP24]. It is a two-party function defined below.

**Definition 1.1 (k-step pointer chasing function).** For \( k \geq 1 \), the \( k \)-step pointer chasing function \( \text{PC}_k : [n]^n \times [n]^n \rightarrow \{0, 1\} \) is defined as follows. Given input \( f_A, f_B \in [n]^n \), for \( r = 0, 1, \ldots, k \) we recursively define pointers via

\[
\text{pt}_r(f_A, f_B) = \begin{cases} 
1 & \text{if } r = 0; \\
 f_A(\text{pt}_{r-1}(f_A, f_B)) & \text{if } r > 0 \text{ is odd}; \\
 f_B(\text{pt}_{r-1}(f_A, f_B)) & \text{if } r > 0 \text{ is even}.
\end{cases}
\]

The output of \( \text{PC}_k \) is the parity of the last pointer, namely, \( \text{PC}_k(f_A, f_B) \equiv \text{pt}_k(f_A, f_B) \mod 2 \).

**Upper bounds.** If Alice and Bob could communicate for \( k \) rounds, a simple protocol could be the following: Alice and Bob alternatively send \( f_A(\text{pt}_{r-1}(f_A, f_B)) \) or \( f_A(\text{pt}_{r-1}(f_A, f_B)) \). The total communication cost for this simple protocol is \( O(k \cdot \log n) \). However, if Alice and Bob can only communicate \( (k - 1) \) rounds, the upper bound then becomes non-trivial. In this direction, Nisan and Wigderson [NW91] proposed a randomized \( (k - 1) \)-round protocol with \( O((n/k + k) \log n) \) communication bits.

- In the beginning, Alice and Bob use public randomness to pick a set of coordinates \( I \subseteq [n] \) of size \( 10n/k \), and then send \( f_A(I) \) and \( f_B(I) \) to the other party.
- On the other hand, Alice and Bob also simulate \( (r \text{ rounds}) \) deterministic protocol but skip one round if one party finds that the pointer is located in \( I \).
- If the skip round never happens, Alice and Bob simply abort at the last round. A simple calculation shows the probability of this event is low.

This randomized protocol is indeed very simple. Alice and Bob only share coordinate-wise information. In fact, this is a restricted rectangle in our setting.

**Lower Bounds.** Consider \( (k - 1) \) round protocols where Alice speaks first. For deterministic protocols, Nisan and Wigderson [NW91] proved an \( \Omega(n - k \log n) \) communication lower bound. In the same paper, they also proved an \( \Omega(n/k^2 - k \log n) \) communication lower bound for protocols that achieve \( 2/3 \) accuracy under uniform input distribution.

Since then, lower bounds for pointer chasing and its close variants have been substantially studied by a good amount of papers [DGS84, DJ98, PRV01, Kla00, KNTSZ07, FKM09, GO16, ACK19]. Finally, Yehudayoff [Yeh20] proved an \( \Omega(n/k - k \log n) \) lower bound for protocols achieving constant advantage under uniform input distribution.

Now the main gap between the upper bound [NW91] and the lower bound [Yeh20] is the extra \( k \log n \) term. This gap becomes significant if \( k \geq \sqrt{n} \). In this paper, we further improve the lower bound and close the gap.
1.1 Our results

We prove that any protocol that achieves constant advantage under uniform input distribution must communicate $\Omega(n/k + k)$ bits.

**Theorem 1.2.** Let $\Pi$ be a $(k - 1)$-round protocol for $\text{PC}_k$ where Alice speaks first such that

$$\Pr_{f_A, f_B \leftarrow [n]^n}[\Pi(f_A, f_B) = \text{PC}_k(f_A, f_B)] \geq 2/3.$$

Then the communication complexity of $\Pi$ is $\Omega(n/k + k)$.

By Yao’s minimax principle, it implies a lower bound for the $(k - 1)$ round randomized communication complexity.

**Corollary 1.3.** Every $(k - 1)$-round randomized protocol for $\text{PC}_k$ with error at most $1/3$ (where Alice speaks first) has communication complexity $\Omega(n/k + k)$.

We observe there is still a $(\log n)$ gap between our lower bound and the protocol by [NW91]. We conjecture that our lower bound is tight and there is a chance to remove the $\log n$ factor in the upper bound side. A simple deterministic protocol with $(k - 1)$ rounds and $O(n)$ communication bits could be the following: Alice and Bob send the parity of $f_A(x)$ and $f_B(x)$ for all $x \in [n]$ in the beginning. Hence they can skip the last round as they already know the parity. This simple protocol shows that [NW91]’s protocol is not tight when $k = o(\log n)$. We believe similar ideas could be extended for large $k$.

**Applications.** Given the connections between $\text{PC}_k$ and diverse applications [FKM+09, NDSP11, CG18, JMR20, CPP22, PNP24], our improved lower bounds automatically lead to several applications. We list two applications below.

**Corollary 1.4** (Direct sum extension of pointer chasing). The $(k - 1)$-round randomized communication complexity of $\text{PC}_k$ with $d$ pairs of functions is $\Omega(d \cdot n/k^2 + d)$

This corollary improves the previous $\Omega(d \cdot n/k^3 - dk \log n - 2d)$ lower bound presented in [FKM+09], which has applications in BFS trees streaming lower bound.

**Corollary 1.5** (Exponential separations in local differential privacy). Let $A$ be a $(k - 1)$-round sequentially interactive $\epsilon$-locally private protocol solving $\text{PC}_k$ with error probability $\gamma \leq 1/3$. Then the sample complexity of $A$ is $\Omega\left(\frac{1}{\epsilon^2} \cdot (n/k + k)\right)$ and there is a $k$ round protocol with sample complexity $\widetilde{O}\left(\frac{k \log n}{\epsilon^2}\right)$.

This corollary improves the previous $\Omega\left(\frac{n}{\epsilon^2 k^2}\right)$ lower bound for $k < \sqrt{n/\log n}$ given by [JMR20].

1.2 Gadgetless Lifting: A New Framework to Prove Communication Lower Bounds

Query-to-communication lifting theorem [RM97, GPW15, GPW17, CFK+19, LMM+22] is a remarkable framework for proving communication complexity lower bounds with a wide variety of applications. Let $f : \mathbb{Z}^n \rightarrow \{0, 1\}$ be a boolean function and let $g : X \times Y \rightarrow Z$ be a two-party gadget function. The framework proceeds in two steps:
1. Given any communication protocol for $f \circ g^n$, we construct a decision tree for $f$ by simulating the communication protocol.

2. Prove a lower bound for the decision tree complexity of $f$.

Combined with these two steps, a communication lower bound for $f \circ g^n$ is then generically achieved. Though this framework is beautiful, however, it requires a gadget as $f$ is a one-party function. As a consequence, this framework only applies to lifted functions.

Many important problems, such as pointer chasing, do not have a lifted form. Therefore, lifting theorems do not apply to these questions. To address this limitation, we propose a new framework called gadgetless lifting. Unlike the query-to-communication lifting theorems that lift decision tree lower bounds into communication lower bounds, the gadgetless lifting lifts lower bounds for restricted protocols into lower bounds for general protocols. This new framework proceeds in two steps.

1. Identify a family of restricted protocols.

2. Simulate general protocols by restricted protocols and prove communication lower bounds for restricted protocols.

As the basic components in gadgetless lifting are the restricted protocols, the gadgets are no longer needed. A crux of gadgetless lifting is how to decide the restricted protocols. In this paper, we capture it as those protocols that “all shared useful information are local information”. For example, the protocol by [NW91] only share local information such as $f_A(x)$ or $f_B(x)$ for some $x \in [n]$. We formalize this idea by lemma 3.1, which is inspired by the structure-vs-pseudorandomness decomposition by Göös, Pitassi, and Watson [GPW17] and Yang and Zhang [YZ24].

In the study of lifted functions, it has been shown that query-to-communication lifting theorems bypassed some fundamental barriers from previous methods. Similarly, gadgetless lifting can also bypass obstacles from existing methods. We discuss two of them below.

**Avoiding the loss in round elimination method.** Previously, the only method to prove round-communication trade-offs is the round elimination method [NW91]. In these two papers, [NW91] and [Yeh20], the authors studied the pointer chasing problem via the round elimination method. Denote by $M_1, \ldots, M_t$ the messages sent in the first $t$ rounds, and let $Z_0, Z_1, \ldots, Z_t$ be the pointers. In round elimination-based proofs, [NW91, Yeh20] analyzed the random variable

$$R_t = (M_1, \ldots, M_t, Z_1, \ldots, Z_{t-1})$$

for $t \leq k$.

Particularly, they proved that $H(R_k) \geq \Omega(n/k)$. Together with the fact that $H(Z_1, \ldots, Z_k) = k \log n$, it implies that $H(M) \geq \Omega(n/k - k \log n)$. The $(k \log n)$ loss is associated with many round elimination-based analysis [NW91, Kla00, KNTSZ07, FKM+09, Yeh20]. In this paper, we avoid the $k \log n$ loss via the gadgetless lifting.

**Breaking square-root loss barrier in information complexity.** Another popular method in proving communication lower bounds is the information complexity. However, as mentioned by Yehudayoff [Yeh20], entropy-based analyses are likely to meet a square-root loss barrier. This barrier usually comes from applying Pinsker’s inequality (or its variant) to bound statistical distance from a small entropy gap. As a consequence, many results such as [NW91] can only prove an $\Omega(n/k^2 - k \log n)$ lower bound.
As mentioned in [Yeh20], the square-root loss also appears in many works when using the entropy-based method to prove lower bounds. For example, it appears in the parallel repetition theorem and is related to the ‘strong parallel repetition’ conjecture which is motivated by Khot’s unique games conjecture [Kho02]. This loss also appears in direct-sum theorems [BBCR10] and direct-product theorems [BRWY13] in communication complexity.

[Yeh20] overcomes this square-root loss barrier via a non-standard measurement called triangular discrimination. By contrast, our approach overcomes the barrier more naturally without using entropy.

Potential applications. We noticed that our method can also be naturally extended to multiparty settings such as the numbers in hand model and numbers on forehead model. Moreover, some important open problems, such as round-communication tradeoff of bipartite matching problem [DNO14, ANRW15, BVDBE+22] and set pointer chasing problem [FKM*09, GO16], are difficult to solve using the round elimination method due to its inherent limitations. Our method offers the potential to solve these challenging problems.

2 Preliminaries

Notations. We use capital letters $X$ to denote a set and use bold symbols like $\mathbf{R}$ to denote random variables. Particularly, for a set $X$, we use $\mathbf{X}$ to denote the random variable uniformly distributed over the set $X$. We use $\leftarrow$ to denote sampling from a distribution or choosing an element from a set uniformly at random.

2.1 Density-Restoring Partition

Min-entropy and dense distribution. For a random variable $X$, we use $\text{supp}(X)$ to denote the support of $X$.

Definition 2.1 (Min-entropy and deficiency). The min-entropy of a random variable $X$ is defined by

$$H_\infty(X) := \min_{x \in \text{supp}(X)} \log \left( \frac{1}{\Pr[X = x]} \right).$$

Suppose that $X$ is supported on $[n]^J$. We define the deficiency of $X$ as

$$D_\infty(X) := |J| \log n - H_\infty(X).$$

For $I \subseteq J$, $x \in [n]^J$, let $x(I) \overset{\text{def}}{=} (x(i))_{i \in I} \in [n]^I$ be the projection of $x$ on coordinates in $I$.

Definition 2.2 (Dense distribution). Let $\gamma \in (0, 1)$. A random variable $X$ supported on $[n]^J$ is said to be $\gamma$-dense if for all nonempty $I \subseteq J$, $H_\infty(x(I)) \geq \gamma |I| \log n$.

The following lemma is the crux of the structure-vs-pseudorandomness method by [GPW17]. It essentially says that a flat random variable could be decomposed into a convex combination of flat random variables with disjoint support and dense properties.

Lemma 2.3 (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 $\beta \in [n]^J$ such that $\forall x \in X$, $x(J) = \beta$. 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 $\alpha_i \in [n]^{\beta_i}$ that satisfy the following properties.
1. \( \forall x \in X^i, x(I_i) = \alpha_i \);
2. \( X^i(J \setminus I_i) \) is \( \gamma \)-dense;
3. \( D_{\infty}(X^i(J \setminus I_i)) \leq D_{\infty}(X(J)) - (1 - \gamma) \log n \cdot |I_i| + \delta_i \), where \( \delta_i \) is defined as \( \log(|X|/\bigcup_{j \geq i} X^j) \).

The proof of this lemma, simple and elegant, is included in the appendix for completeness.

2.2 Communication Protocols

We recall basic definitions and facts about communication protocols.

Protocol Tree. Let \( X \) and \( Y \) be the input space of Alice and Bob respectively. A deterministic communication protocol \( \Pi \) 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 either Alice or Bob — we denote the owner by \( \text{owner}(v) \);
- every leaf node specifies an output.

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

Fact 2.4. The set of all inputs that leads to an internal vertex \( v \) is a rectangle, denoted by \( \Pi_v = X_v \times Y_v \subseteq X \times Y \).

The communication complexity of \( \Pi \), denoted by \( \text{CC}(\Pi) \), is the depth of the tree. The round complexity of \( \Pi \), is the minimum number \( k \) such that in every path from the root to some leaf, the owner switches at most \( k - 1 \) times. Clearly, if a protocol has \( k \) round, then its communication complexity is at least \( k \). We can safely make the following assumptions for any protocol \( \Pi \):

- \( \Pi \) has \( k \) round on every input; and
- \( \Pi \) communicates \( \text{CC}(\Pi) \) bits on every input.

Indeed, for any protocol, we can add empty messages and rounds in the end, which boosts the communication complexity by a factor of 2.

3 Proof of Main Theorem

We use a decomposition and sampling process \( \text{DS} \), as shown in Algorithm 1, in our analysis. \( \text{DS} \) takes as input a protocol \( \Pi \), and samples a rectangle \( R \) that is contained in \( \Pi_v \) for some leaf node \( v \). Our proof proceeds in three steps:

1. First, section 3.1 analyzes crucial invariants during the running of \( \text{DS} \).
2. Next, section 3.2 shows that the accuracy of \( \Pi \) is captured by a quantity called average fixed size, which is a natural quantity that arises in the running of \( \text{DS} \).
3. Finally, section 3.3 proves that the average fixed size can be bounded from above by \( O(\text{CC}(\Pi)) \). Consequently, if \( \Pi \) enjoys high accuracy, we get a lower bound of \( \text{CC}(\Pi) \).
3.1 The Decomposition and Sampling Process

During the sampling process, we maintain a useful structure of $R$ mainly by a partitioning-then-sampling mechanism: At the beginning, $R$ is set to be the set of all inputs. Walking down the protocol tree, we decompose the rectangle into structured sub-rectangles; then we sample a decomposed rectangle with respect to its size. In the end, we arrive at a leaf node $v$ and a subrectangle of $\Pi_v$.

**Lemma 3.1** (Loop invariant). Set $\gamma \text{ def } 1 - \frac{0.1}{\log n}$. Then in the running of DS($\Pi$), we have the following loop invariants: After each iteration,$$
\begin{align*}
(\phi) \quad & X \times Y \subseteq \Pi_v; \\
(\bullet) \quad & X(J_A), Y(J_B) \text{ are } \gamma\text{-dense}; \\
(\forall) \quad & \text{there exists some } \alpha_A \in [n]^{\overline{J_A}}, \alpha_B \in [n]^{\overline{J_B}} \text{ such that } x(\overline{J_A}) = \alpha_A, y(\overline{J_B}) = \alpha_B \text{ for all } x \in X, y \in Y; \\
(\bullet) \quad & \text{there exists some } z_r \in [n] \text{ such that } pt_r(f_A, f_B) = z_r \text{ for all } (f_A, f_B) \in X \times Y.
\end{align*}
$$

**Proof.** Item $(\phi)$ is true because every time $v$ is updated, $X \times Y$ is updated accordingly to a sub-rectangle of $\Pi_v$ and updating $X \times Y$ into its sub-rectangles does not violate this condition.

Since we applied density restoring partition at the end of each iteration, Item $(\bullet)$ and $(\forall)$ is guaranteed by lemma 2.3 and the way that $X, Y, J_A, J_B$ are updated.

We prove the last item $(\bullet)$ by induction. Assume that the statement holds after the first $(t-1)$ iterations. WLOG, assume that at the beginning of the $t$-th iteration, $v$ is owned by Alice. Consider the following two cases.

- **Case 1.** Not a new round: Line 13 is not executed in the $t$-th iteration. Since $r$ remains unchanged and we only update $R$ to be a sub-rectangle of itself, the statement still holds.

- **Case 2.** A new round begins: Line 13 is executed and $r$ is increased by 1. Let $\rho$ denote the value of $r$ before Line 13, then after this iteration, we have $r = \rho + 1$. The induction hypothesis guarantees that there exists some $z_{\rho-1} \in [n]$ such that

$$pt_{\rho-1}(f_A, f_B) = z_{\rho-1} \text{ for all } (f_A, f_B) \in X \times Y.$$ 

Due to the partition and the update in Line 11 and Line 12, $|\text{supp}(X(z_{\rho-1}))| \leq n/2$. Hence, $X(z_{\rho-1})$ cannot be $\gamma$-dense as we set $\gamma = 1 - \frac{0.1}{\log n}$. Observe that after the update in Line 17, $X(J_A)$ is $\gamma$-dense. Consequently, we must have $z_{\rho-1} \in J_A$, and by item $(\forall)$, there exists some $z_\rho \in [n]$ such that $f_A(z_{\rho-1}) = z_\rho \forall f_A \in X$. By definition, for all $(f_A, f_B) \in X \times Y,$

$$pt_\rho(f_A, f_B) = f_A(pt_{\rho-1}(f_A, f_B)) = f_A(z_{\rho-1}) = z_\rho.$$ 

This is exactly the same statement after the $t$-th iteration (as we have $r = \rho + 1$).

\[\square\]

The restricted rectangles in this loop invariant are inspired by the protocols of Nisan and Wigderson [NW91]. This lemma aims to capture the fact that Alice and Bob cannot get any additional useful information other than coordinate-wise information during their communication.
Algorithm 1: The decomposition and sampling process DS

Input: A protocol $\Pi$ for the problem $PC_k$.
Output: A rectangle $R = X \times Y$, and $J_A, J_B \subseteq [n]$.

1. Initialize $v := \text{root of } \Pi, r := 1, X := Y := [n]^n, J_A := J_B := [n], \text{bad} := \text{false}.$
2. while $v$ is not a leaf node do
   // Invariant: (1) $X \times Y \subseteq \Pi_0$; (2) there exists some $z_{r-1} \in [n]$ such that $pt_{r-1}(f_A, f_B) = z_{r-1} \forall (f_A, f_B) \in X \times Y$.
   3. Let $u_0 := \Pi(v, 0), u_1 := \Pi(v, 1)$ be the two children of $v$.
   4. if $\text{owner}(v) = \text{Alice}$ then
      Partition $X$ into $X = X^0 \cup X^1$ such that $X^b \times Y \subseteq \Pi_{u_b}$ for $b \in \{0, 1\}$.
      Sample $b \in \{0, 1\}$ such that $\Pr[b = b] = |X^b|/|X|$ for $b \in \{0, 1\}$.
      Update $X := X^b, v := u_b$.
      9. if $\text{owner}(u_b) = \text{Bob}$ then
         // A new round.
         Further Partition $X$ into $X = X^0 \cup X^1$ where $X^b := \{f_A \in X : f_A(z_{r-1}) \mod 2 = b\}$.
         Sample $b' \in \{0, 1\}$ such that $\Pr[b' = b] = |X^b|/|X|$ for $b \in \{0, 1\}$.
         Update $X := X^{b'}, r := r + 1$.
   14. Let $X = X^1 \cup \cdots \cup X^m$ be the decomposition of $X$ promised by lemma 2.3 with associated sets $I_1, \ldots, I_m \subseteq J_A$.
   // Invoking lemma 2.3 with $J = J_A, M = n, \gamma = 1 - \frac{0.1}{\log n}$.
   16. Sample a random element $j \in [m]$ such that $\Pr[j = j] = |X^j|/|X|$ for $j \in [m]$.
   17. Update $X := X^j, J_A := J_A \setminus I_j$.
   19. if $\text{owner}(u_b) = \text{Bob} \land z_{r-1} \notin J_B$ then
      $\text{bad} := \text{true}$.
   20. if $\text{owner}(v) = \text{Bob}$ then
      Partition $Y$ into $Y = Y^0 \cup Y^1$ such that $X \times Y^b \subseteq \Pi_{u_b}$ for $b \in \{0, 1\}$.
      Sample $b \in \{0, 1\}$ such that $\Pr[b = b] = |Y^b|/|Y|$ for $b \in \{0, 1\}$.
      Update $Y := Y^b, v := u_b$.
      24. if $\text{owner}(u_b) = \text{Alice}$ then
         Further Partition $Y$ into $Y = Y^0 \cup Y^1$ where $Y^b := \{f_B \in Y : f_B(z_{r-1}) \mod 2 = b\}$.
         Sample $b' \in \{0, 1\}$ such that $\Pr[b' = b] = |Y^b|/|Y|$ for $b \in \{0, 1\}$.
         Update $Y := Y^{b'}, r := r + 1$.
   28. Let $Y = Y^1 \cup \cdots \cup Y^m$ be the decomposition of $Y$ promised by lemma 2.3 with associated sets $I_1, \ldots, I_m \subseteq J_B$.
   Sample a random element $j \in [m]$ such that $\Pr[j = j] = |Y^j|/|Y|$ for $j \in [m]$.
   30. Update $Y := Y^j, J_B := J_B \setminus I_j$.
   31. if $\text{owner}(u_b) = \text{Alice} \land z_{r-1} \notin J_A$ then
      $\text{bad} := \text{true}$.
   32.
3.2 Relating Accuracy and Average Fixed Size

From lemma 3.1 we know that the coordinates in \(J_A\) and \(J_B\) are fixed if we only look at the inputs in \(X \times Y\). Intuitively, the advantage of the protocol comes from such fixed coordinates, since the ‘alive’ coordinates \(J_A, J_B\) are dense in the sense that \(X(J_A), Y(J_B)\) is \(\gamma\)-dense. This intuition is formalized in the following lemma.

**Lemma 3.2** (Relating accuracy and average fixed size). Let \(\Pi\) be a \((k - 1)\)-round deterministic protocol where Alice speaks first. Then

\[
\Pr_{f_A,f_B \leftarrow [n]} [\Pi(f_A, f_B) = PC_k(f_A, f_B)] \leq \frac{n^{1-\gamma}}{2} + n^{-\gamma} \cdot (k - 1) \cdot \mathbf{E}_{(R,J_A,J_B) \leftarrow \text{DS}(\Pi)} [|J_A| + |J_B|].
\]

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

**Claim 3.3.** If DS(\(\Pi\)) outputs \((R = X \times Y, J_A, J_B)\) and bad = FALSE in the end, then

\[
\Pr_{(f_A,f_B) \leftarrow R} [\Pi(f_A, f_B) = PC_k(f_A, f_B)] \leq \frac{n^{1-\gamma}}{2}.
\]

**Claim 3.4.** \(\Pr_{\text{DS}(\Pi)} \left[ \text{bad} = \text{TRUE} \right] \leq n^{-\gamma} \cdot (k - 1) \cdot \mathbf{E}_{(R,J_A,J_B) \leftarrow \text{DS}(\Pi)} [|J_A| + |J_B|].\)

Next, we first prove lemma 3.2 using the above two claims, and the proof of the claims is followed.

**Proof of lemma 3.2.** Note that in the running of DS(\(\Pi\)), we always update \(R\) to a randomly chosen rectangle and the probability of each rectangle being chosen is proportional to its size. Consequently,

\[
\Pr_{f_A,f_B \leftarrow [n]} [\Pi(f_A, f_B) = PC_k(f_A, f_B)]
= \Pr_{(R,J_A,J_B) \leftarrow \text{DS}(\Pi),(f_A,f_B) \leftarrow R} [\Pi(f_A, f_B) = PC_k(f_A, f_B)]
\leq \Pr_{\text{DS}(\Pi)} \left[ \text{bad} = \text{TRUE} \right] + \Pr_{(R,J_A,J_B) \leftarrow \text{DS}(\Pi),(f_A,f_B) \leftarrow R} [\Pi(f_A, f_B) = PC_k(f_A, f_B) \land \text{bad} = \text{FALSE}]
\leq \frac{n^{1-\gamma}}{2} + n^{-\gamma} \cdot (k - 1) \cdot \mathbf{E}_{(R,J_A,J_B) \leftarrow \text{DS}(\Pi)} [|J_A| + |J_B|].
\]

where the last step is by claim 3.3 and claim 3.4. \(\square\)

It remains to prove the two claims.

**Proof of claim 3.3.** WLOG, assume \(k - 1\) is odd and the protocol always has \(k\) round. Let \(z_{k-1}\) be the pointer guaranteed by the loop invariant (lemma 3.1), i.e., \(pt_{k-1}(f_A, f_B) = z_{k-1}\) for all \((f_A, f_B) \in R\). Since bad = FALSE, we have \(z_{k-1} \in J_A\). Again by the loop invariant, \(H_{\infty}(X(z_{k-1})) \geq \gamma\). Moreover,
since $R$ is contained in some leaf node of $Π$, $Π$ output the same answer in $R$, say $b^* \in \{0, 1\}$. Consequently,

$$\Pr_{(f_A, f_B) \sim R} \left[ Π(f_A, f_B) = \text{PC}_k(f_A, f_B) \right] = \Pr_{f_A \sim X} \left[ f_A(z_{k-1}) \mod 2 = b^* \right] \leq \sum_{\sigma \in [n] : \sigma \mod 2 = b^*} \Pr_{f_A \sim X} \left[ f_A(z_{k-1}) = \sigma \right] \leq \frac{n}{2} \cdot n^{-\gamma}.$$ 


Proof of claim 3.4. Let $E_t$ denote the event that the flag bad is raised when $r = t + 1$ (i.e., when the $t$-th round ends) for the first time. Clearly, $\Pr \left[ \text{bad} = \text{TRUE} \right] = \sum_{t=1}^{\#_t} \Pr \left[ E_t \right]$. It suffices to bound each $\Pr \left[ E_t \right]$.

Assume $t$ is odd, meaning that Alice speaks in the $t$-th round. Let coin denote the randomness used for the first $(t - 1)$ rounds. Let $X^{(t-1)}$, $J_A^{(t-1)}$, $J_B^{(t-1)}$ be the sets $X$, $J_A$, $J_B$ when executing $DS(Π)$ using coin until the $t$-th round begins. Let $z_{t-1}$ be the pointer promised by the invariant. For $E_t$ to happen, we must have $\text{bad} = \text{FALSE}$ until the $t$-th round begins, meaning that $z_{t-1} \in J_A^{(t-1)}$.

Note that the random variable $z_t$ exactly has the same distribution as $X^{(t-1)}(z_{t-1})$. This is because, in the $t$-th round (i.e., until $r$ steps to $t + 1$), we decompose $X^{(t-1)}$ into finer sets and update $X$ to be one of them with probability proportional to their size. Therefore,

$$\Pr_{\text{coin'}} \left[ E_t \right] = \Pr_{\text{coin'}} \left[ z_t \not\in J_B^{(t-1)} \right] = \Pr_{f_A \sim X^{(t-1)}} \left[ f_A(z_{t-1}) \not\in J_B^{(t-1)} \right] = \sum_{\sigma \in J_B^{(t-1)}} \Pr_{f_A \sim X^{(t-1)}} \left[ f_A(z_{t-1}) = \sigma \right] \leq \left| J_B^{(t-1)} \right| \cdot n^{-\gamma},$$

where we fix coin and the probability runs over coin', the randomness used afterward; the last inequality holds because $z_{t-1} \in J_A^{(t-1)}$ and $X^{(t-1)}(J_A^{(t-1)})$ is $\gamma$-dense (by Item (i) in lemma 3.1). Averaging over coin, we get

$$\Pr_{DS(Π)} \left[ E_t \right] \leq \mathbb{E}_{\text{coin}} \left[ \left| J_B^{(t-1)} \right| \right] \cdot n^{-\gamma} \leq \mathbb{E}_{(R, J_A, J_B) \sim DS(Π)} \left[ \left| J_B \right| \right] \cdot n^{-\gamma},$$

where the second inequality holds because $J_B$ becomes smaller and smaller during the execution.

For even $t$'s, we analogously have $\Pr \left[ E_t \right] \leq \mathbb{E} \left[ |J_A| \right] \cdot n^{-\gamma}$, and hence the claim follows from union bound. 

\section{3.3 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., $\mathbb{E}_{(R, J_A, J_B) \sim DS(Π)} \left[ |J_A| + |J_B| \right]$), in what follows we show that the average fixed size is at most $O(\text{CC}(Π))$. Formally, we prove that

\textbf{Lemma 3.5.} Let $Π$ be a $(k - 1)$-round deterministic protocol where Alice speaks first. Then

$$\mathbb{E}_{(R, J_A, J_B) \sim DS(Π)} \left[ |J_A| + |J_B| \right] \leq \frac{3 \text{CC}(Π)}{(1 - \gamma) \log n}.$$
Remark 3.6. We shall set \( \gamma := 1 - \frac{0.1}{\log n} \) and hence the right-handed side equals 30CC(\( \Pi \)).

Proof. We shall prove this lemma by density increment argument. That is, we study the change of the density function
\[
\mathcal{D}_\infty(R) \overset{\text{def}}{=} \mathcal{D}_\infty(X(J_A)) + \mathcal{D}_\infty(Y(J_B)).
\]
in each iteration. Let \( \phi_t \) be the value of \( \mathcal{D}_\infty(R) \) at the end of the \( t \)-th iteration. Assume without loss of generality Alice speaks (i.e., owner(\( v \)) = Alice) in the \( t \)-th iteration.

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

1. First, \( X \) is partitioned into \( X = X^0 \cup X^1 \) according to \( \Pi \). Then, \( X \) is updated to \( X^b \) with probability \( \frac{|X^b|}{|X|} \). Consequently, \( \mathcal{D}_\infty(X(J_A)) \) will increase as \(|X|\) shrinks, and in expectation (over the random choice of \( b \)) the increment is
\[
\sum_{b \in \{0,1\}} \frac{|X^b|}{|X|} \log \left( \frac{|X|}{|X^b|} \right) \leq 1.
\]

2. Next, suppose that updating \( v \) leads to the switch of the owner, i.e., Line 13 is triggered. Since we also partition \( X \) into two parts and update \( X \) with probability proportional to the size of each part, the same argument applies. That is, taking expectation over the random choice of \( b' \), \( \mathcal{D}_\infty(X(J_A)) \) increases by at most 1 in expectation.

3. Finally, we further partition \( X \) according to lemma 2.3. Say \( X \) is partitioned into \( X = X^1 \cup \cdots \cup X^m \) and let \( I_1, \ldots, I_m \) be the index sets promised by lemma 2.3; and for all \( j \in [m] \) we have
\[
\mathcal{D}_\infty(X^j(J_A \setminus I_j)) \leq \mathcal{D}_\infty(X(J_A)) \leq (1 - \gamma) \log n|I_j| + \delta_j,
\]
where \( \delta_j = \log(|X|/\cup_{v \in X^b} X^v) \). With probability \( p_j \overset{\text{def}}{=} |X^j|/|X| \), we update \( X := X^j \) and \( J_A := J_A \setminus I_j \). Therefore, taking expectation over the random choice of \( j \), the density function will decrease by
\[
\mathcal{D}_\infty(X(J_A)) - \mathbb{E}_{j \sim J} \left[ \mathcal{D}_\infty(X^j(J_A \setminus I_j)) \right] \geq \mathbb{E}_{j \sim J} \left[ (1 - \gamma) \log n \cdot |I_j| - \delta_j \right]. \tag{3}
\]
Note that \( \delta_j \overset{\text{def}}{=} \log \frac{1}{\sum_{v \in J} p_v} \) and thus
\[
\mathbb{E}_{j \sim J} [\delta_j] = \sum_{j=1}^{m} p_j \log \frac{1}{\sum_{v \in J} p_v} \leq \int_{0}^{1} \frac{1}{1-x} \, dx \leq 1. \tag{4}
\]

Let \( \mathcal{F}_{t-1} \) be the \( \sigma \)-algebra generated by the random coins used for the first \((t - 1)\) iterations. Let \( \beta_t \) be the increment of \(|J_A|\) and \(|J_B|\) in the \( t \)-th iteration. Observe that \( \beta_t = |I_j| \) by definition. By eq. (3) and eq. (4), taking expectation over random choice of \( j \), \( \mathcal{D}_\infty(X(J_A)) \) decrease by at least \((1 - \gamma) \log n \cdot \mathbb{E} [\beta_t | \mathcal{F}_{t-1}] - 1\) due to the density restoring partition. Then
\[
\mathbb{E} [\phi_t - \phi_{t-1}] = \mathbb{E} [\mathbb{E} [\phi_t - \phi_{t-1} | \mathcal{F}_{t-1}]] \leq \mathbb{E} [1 + \eta_t - ((1 - \gamma) \log n \cdot \beta_t - 1)], \tag{5}
\]
where \( \eta_t \overset{\text{def}}{=} \mathbb{1}[\text{owner switches in the } t\text{-th iteration}]. \]
Write $c \overset{\text{def}}{=} \text{CC}(\Pi)$ and assume we always have $c$ iterations. \footnote{Namely, $\Pi$ communicates $c$ bits on all inputs.} In the beginning, $\varphi_0 = D_{\infty}([n]^n \times [n]^n) = 0$. Since the density function is always non-negative by definition, we have $\varphi_c \geq 0$ and thus $E[\varphi_c - \varphi_0] \geq 0$. On the other hand, by telescoping,

$$E[\varphi_c - \varphi_0] = \sum_{t=1}^{c} E[\varphi_t - \varphi_{t-1}] \leq 2c + \sum_{t=1}^{c} E[\eta_t - (1 - \gamma) \log n \cdot \beta_t],$$

where the inequality follows from eq. (5). Observe that $\sum_{t=1}^{c} \eta_t$ is at most $k$ and $\sum_{t=1}^{c} \beta_t = |\mathcal{J}_A| + |\mathcal{J}_B|$ by definition. We conclude that

$$E[|\mathcal{J}_A| + |\mathcal{J}_B|] = E\left[\sum_{t=1}^{c} \beta_t\right] \leq \frac{2c + k}{(1 - \gamma) \log n} \leq \frac{3c}{(1 - \gamma) \log n},$$

as desired. \hfill $\square$

**Proving the main theorem.** Now our main theorem easily follows from the two lemmas.

**Proof of theorem 1.2.** Set $\gamma \overset{\text{def}}{=} 1 - \frac{0.1}{\log n}$. By lemma 3.5 and lemma 3.2, we get

$$\text{Accuracy}(\Pi) \overset{\text{def}}{=} \Pr_{f_A, f_B \sim [n]^n} [\Pi(f_A, f_B) = \text{PC}_k(f_A, f_B)] \leq \frac{n^{1 - \gamma}}{2} + n^{-\gamma} \cdot (k - 1) \cdot \frac{3\text{CC}(\Pi)}{(1 - \gamma) \log n} \leq 0.54 + \frac{1.08(k - 1)}{n} \cdot 30\text{CC}(\Pi),$$

where we use $\frac{n^{1 - \gamma}}{2} \leq 0.54, n^{-\gamma} \leq \frac{1.08}{n}$. Since we assumed $\text{Accuracy}(\Pi) \geq 2/3$, we conclude that

$$\text{CC}(\Pi) \geq \frac{2/3 - 0.54}{1.08 \cdot 30} \cdot \frac{n}{k - 1} > 0.0039 \cdot \frac{n}{k - 1} = \Omega(n/k).$$

We also trivially have $\text{CC}(\Pi) \geq k - 1$ as $\Pi$ has $(k - 1)$ rounds; putting it together we conclude that $\text{CC}(\Pi) = \Omega(n/k + k)$. \hfill $\square$

**Acknowledgements**

We thank Sepehr Assadi for his helpful comments and suggestions.

**References**


Appendix

The following lemma and proof are from Lemma 5 in [GPW17].

Lemma 3.7 (lemma 2.3 restated). Let $\gamma \in (0, 1)$. Let $X$ be a subset of $[n]^M$ and $J \subseteq [M]$. Suppose that there exists an $\beta \in [n]^J$ 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 $\alpha_i \in \{0, 1\}^{|J|}$ that satisfy the following properties.

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

Proof. We prove it by a greedy algorithm as follows.

Algorithm 2: Greedy Algorithm

\begin{algorithm}
\begin{align*}
\textbf{Input:} & \ X \subseteq [n]^M \\
\textbf{Output:} & \text{A partition } X = X^1 \cup X^2 \cup \cdots \cup X^m \\
1 & \text{Initialize } i := 1. \\
2 & \text{while } X \neq \emptyset \text{ do} \\
3 & \quad \text{Let } I \subseteq J \text{ be a maximal subset (possibly } I = \emptyset) \text{ such that } H_\infty(X(I)) < \gamma |I| \log n \text{ and let } \\
& \quad \alpha_i \overset{\text{def}}{=} [n]^J \text{ be a witness of this fact, i.e., } \Pr[X(I) = \alpha_i] > n^{-\gamma |I|}. \\
4 & \quad X^i := \{x \in X : x(I) = \alpha_i\} \text{ and } I_i := I. \\
5 & \quad \text{Update } X := X \setminus X^i, J := J \setminus I_i, \text{ and } i := i + 1.
\end{align*}
\end{algorithm}

Item 1 is guaranteed by the construction of $X^i$ and $I_i$.

We prove Item 2 by contradiction. Assume towards contradiction that $X^i(J \setminus I_i)$ is not $\gamma$-dense for some $i$. By definition, there is a nonempty set $K \subseteq J \setminus I_i$ and $\beta \in [n]^K$ violating the min-entropy condition, namely, $\Pr[X(K) = \beta] > n^{-\gamma |K|}$. Write $X^\geq i \overset{\text{def}}{=} \cup_{j \geq i} X^j$. Then

$$\Pr[X^\geq i(I_i \cup K) = (\alpha_i, \beta)] = \Pr[X^\geq i(I_i) = \alpha_i] \cdot \Pr[X^i(K) = \beta] > n^{-\gamma |I_i|} \cdot n^{-\gamma |K|} = n^{-\gamma |I_i \cup K|},$$

where the first equality holds as $(X^\geq i)X^\geq i(I_i) = \alpha_i = X^i$. However, this means at moment that $I_i$ is chosen, the set $I_i \cup K \subseteq J$ also violates the min-entropy condition (witnessed by $(\alpha_i, \beta)$), contradicting the maximality of $I_i$.

Finally, Item 3 is proved by straightforward calculation:

$$D_\infty(X^i(J \setminus I_i)) = |J \setminus I_i| \log n - \log |X^i|$$
$$\leq (|J| \log n - |I_i| \log n) - \log \left( |X^\geq i| \cdot n^{-\gamma |I_i|} \right)$$
$$= (|J| \log n - \log |X|) - (1 - \gamma)|I_i| \cdot \log n + \log \left( \frac{|X|}{|X^\geq i|} \right)$$
$$= D_\infty(X(J)) - (1 - \gamma)|I_i| \log n + \delta_i.$$

$\square$