Settling Nondeterministic Communication Complexity of Graph Connectivity

Zhijun Zhang*
INSAIT, Sofia University "St. Kliment Ohridski"

Abstract

Determining the communication complexity of graph connectivity has been one of the longest standing open problems in communication complexity since the seminal work of Babai, Frankl, and Simon [FOCS'86]. A simple protocol transmits a spanning forest using $\mathcal{O}(N \log N)$ communication on N-vertex graphs. While this was shown optimal for deterministic protocols by Hajnal, Maass, and Turán [STOC'88], the optimal nondeterministic and randomized complexities have remained unresolved for decades.

We settle the nondeterministic communication complexity of graph connectivity by proving an optimal $\Omega(N \log N)$ lower bound. The best prior bound was $\Omega(N \log \log N)$ due to Raz and Spieker [Combinatorica'95]. Our bound is shown by upper bounding the maximum size of 1-monochromatic rectangles. Generalizing our techniques to weighted, almost 1-monochromatic rectangles yields an $\Omega(N \cdot \log(\min(1/\varepsilon, N)))$ lower bound for randomized protocols with error probability ε. In particular, this bound is optimal for any polynomially small ε.

For constant error probability, we also prove that any r-round protocol requires $\Omega(N/r \cdot \log N)$ communication, which is optimal for every constant r. Previously, optimal $\Omega(N \log N)$ lower bounds for graph connectivity were known only for 1-round [Sun, Woodruff; APPROX/RANDOM'15] and 2-round protocols [Radhakrishnan, Reddy, Venkat; ITCS'26]. Our proof introduces a new communication problem which can be viewed as a composition of inner product and pointer chasing problems.

Our lower bounds can be extended to other fundamental graph properties, and also imply optimality of known query and streaming algorithms for related problems.

*zhijun.zhang@insait.ai. This research was partially funded by the Ministry of Education and Science of Bulgaria (support for INSAIT, part of the Bulgarian National Roadmap for Research Infrastructure).

Contents

1 Introduction **1**
- 1.1 Our Results . . . 1
- 1.2 Related Work . . . 3

2 Technical Overview **4**
- 2.1 Nondeterministic Lower Bound . . . 4
- 2.2 Randomized Lower Bound with Constant Error . . . 6

3 Preliminaries **7**

4 Nondeterministic Lower Bound for Connectivity **8**
- 4.1 Proof of Lemma 4.3 . . . 10

5 Lower Bound for Connectivity with Small Error **14**
- 5.1 Proof of Lemma 5.2 . . . 18

6 Lower Bound for Connectivity with Constant Error **21**
- 6.1 Proof of Theorem 6.1 . . . 22
- 6.2 Lower Bounds for Other Graph Properties . . . 25

7 Lower Bound for IP∘PC **26**
- 7.1 A Weaker Lower Bound . . . 26
- 7.2 Proof of Theorem 7.1 . . . 33

AI Methodology **37**

A Basic Tools from Information Theory **44**

B Round-Efficient Protocols for PC-Perm **45**

C Lower Bound for IP∘PC with More Rounds **49**

1 Introduction

Graph problems have been extensively studied in communication complexity since the early works of [BFS86, HMT88]. Examples include connectivity [RS95, FKM+08, IKL+12, SW15, RRV26], matching [GKK12, BvdBE+22, CJSY26], vertex/edge coloring [FM25, CMNS25], minimum cuts [ASW25, JNS26], vertex covers [DGR25], planarity [ĎP89], and many more.

One notable problem is connectivity, denoted Conn_N. In the standard two-party communication model [Yao79], Alice and Bob receive disjoint edge sets $E_\mathsf{A}, E_\mathsf{B}$ of an N-vertex undirected graph[1] $G = (V, E_\mathsf{A} \cup E_\mathsf{B})$. The goal is to jointly decide if G is connected.

A simple one-way deterministic protocol solves Conn_N in $\mathcal{O}(N \log N)$ communication: Alice sends a spanning forest of her edges E_A to Bob, who then updates the spanning forest using his edges E_B and then decides connectivity of G. [HMT88] proved a matching lower bound, showing that no deterministic protocol with $o(N \log N)$ communication exists, regardless of the number of rounds. For randomized protocols, the seminal work of [BFS86] gave a reduction from set disjointness to graph connectivity. When combined with the $\Omega(N)$ lower bound for set disjointness [KS92, Raz92, BJKS04], it implies that any randomized protocol for graph connectivity requires $\Omega(N)$ communication. Closing the gap between lower and upper bounds remains a major open problem for nondeterministic and randomized protocols, which has been highlighted for decades; see for example [BFS86, RS95, IKL+12, AEG+22, RRV26].

Prior to this work, the strongest nondeterministic lower bound was $\Omega(N \log \log N)$ due to [RS95]. Their lower bound actually holds for the more restricted bipartite version bConn_N, where G is promised to be a balanced bipartite graph with $N/2$ vertices on each side, and both E_A and E_B are guaranteed to be perfect matchings across the bipartition[2]. Equivalently, $E_\mathsf{A}, E_\mathsf{B}$ can be viewed as permutations a, b over $[N/2]$, and G is connected if and only if $b^{-1}a$ is a single $(N/2)$-cycle.

No superlinear $\omega(N)$ lower bound is known for randomized protocols in general. When the number of rounds is bounded, the first progress was an optimal $\Omega(N \log N)$ lower bound for 1-round protocols due to [SW15]. Recently, the same lower bound was shown for 2-round protocols by [RRV26]. For every constant $r \geq 3$, [RRV26] also gave an $\Omega(N \log^{(2r+1)} N)$ lower bound for r-round protocols.

1.1 Our Results

Our first main contribution is settling the nondeterministic communication complexity of graph connectivity. [RS95] explicitly conjectured that $\Omega(N \log N)$ communication is required for nondeterministic protocols. We resolve the conjecture in the affirmative.

Theorem 1.1 (Informal version of Theorem 4.1)**.** *The nondeterministic communication complexity of N-vertex graph connectivity is $\Theta(N \log N)$.*

[1] To avoid confusion, we use N to denote the number of vertices in the input graph throughout this paper. n is reserved for the size parameter of auxiliary communication problems used for reduction.

[2] Technically, bConn_N may induce multigraphs with parallel edges. This can be easily fixed by splitting every vertex into a length-1 path, resulting in simple bipartite graphs with $\Theta(N)$ vertices.

Like many nondeterministic communication lower bounds, the proof of Theorem 1.1 is via upper bounds on the maximum size of 1-monochromatic rectangles. In Section 2.1, we sketch our new ideas and give a brief comparison with the approach of [RS95].

For randomized protocols, transcripts do not necessarily correspond to monochromatic rectangles anymore. Nevertheless, our techniques can be generalized to show that bConn_N does not have large weighted rectangles that are close to being 1-monochromatic. This yields an error-dependent randomized lower bound.

Theorem 1.2 (Informal version of Theorem 5.1)**.** *The randomized communication complexity of N-vertex graph connectivity with error probability $\varepsilon = \varepsilon(N)$ is $\Omega(N \cdot \log(\min(1/\varepsilon, N)))$.*

It is worth noting that Theorem 1.2 implies an optimal $\Omega(N \log N)$ lower bound for any polynomially small error probability $\varepsilon = 1/\mathsf{poly}(N)$. For any $\varepsilon = o(1)$, it gives a superlinear $\omega(N)$ lower bound, improving the $\Omega(N)$ bound of [BFS86]. We also remark that any improvement to the $\Omega(\log(1/\varepsilon))$ dependency on ε would immediately imply $\omega(N)$ lower bounds even for protocols with constant error. This is because $\mathcal{O}(\log(1/\varepsilon))$ repetitions suffice to decrease the error probability from constant to ε.

For randomized protocols with constant error, we also prove tradeoffs between the number of rounds and the total communication.

Theorem 1.3 (Informal version of Theorem 6.1)**.** *The r-round randomized communication complexity of N-vertex graph connectivity with constant error probability is $\Omega(N/r \cdot \log N)$.*

In particular, for every constant r, Theorem 1.3 implies optimal $\Omega(N \log N)$ lower bounds for any r-round protocols, improving [SW15, RRV26]. Also note that our lower bound is always superlinear in the case of $r = o(\log N)$, while an $\Omega(N/r \cdot \log N)$ lower bound for $r = \Omega(\log N)$ already follows from the $\Omega(N)$ bound of [BFS86], which holds even for protocols with unbounded rounds.

We emphasize that the proof of Theorem 1.3 is not a generalization of previous approaches. Indeed, the lower bounds in [RRV26] crucially rely on a specific variant of pointer chasing. Appendix B presents a round-efficient protocol that can jump over quadratically many layers for their specific variant, which tends to be an unfavorable feature for lower bound proofs[3]. Section 2.2 includes a more detailed discussion.

Motivated by this, our proof of Theorem 1.3 introduces a new two-party communication problem $\mathsf{IP}\circ\mathsf{PC}_{n,r}$, which is essentially a composition of two well-studied communication problems, namely inner product and pointer chasing, as the name suggests. Concretely, the input of $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ consists of two bit vectors $X, Y \in \{0,1\}^n$ and $2r$ permutations $A_1, \ldots, A_r, B_1, \ldots, B_r$ over $[n]$. Alice holds all A_i, B_i for odd i, while Bob holds X, Y as well as all A_i, B_i for even i. The goal is to output the binary inner product $\langle A_r \cdots A_1 X, B_r \cdots B_1 Y \rangle$, where $A_1, \ldots, A_r, B_1, \ldots, B_r$ are identified with their corresponding permutation matrices. At a high level, before outputting the final inner product, protocols have to simultaneously "chase" each of the $2n$ input bits through r layers of permutations (i.e., "pointers").

[3] With current techniques, it is unclear how to prove substantially stronger lower bounds for these problems when the number of rounds is significantly smaller than the number of layers.

We prove the following strong lower bound[4] for $\mathsf{IP}\circ\mathsf{PC}_{n,r}$. Unlike most of the known lower bounds for pointer chasing problems (see Section 1.2), Theorem 1.4 neither loses any $\mathsf{poly}(r, \log n)$ factor nor has iterated-logarithm-type dependency on r. We believe that $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ can be a powerful alternative to pointer chasing and may be of independent interest.

Theorem 1.4 (Informal version of Theorems 7.1 and C.1)**.** *The r-round randomized communication complexity of $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ with constant error probability is $\Omega(n \log n)$, for $r \ll n$.*

We remark that our lower bounds for graph connectivity can be extended to other fundamental graph properties. In Theorem 6.2, we give lower bounds similar to Theorem 1.3 for deciding if the input graph is bipartite, cycle-free, Eulerian, Hamiltonian, or planar.

Our lower bounds also immediately imply optimality of known query and streaming algorithms. In particular, Theorem 1.1 means that the zero-error cut query algorithm for connectivity in [AEG+22, Theorem 1.2] has optimal expected query complexity $\Theta(N)$. Their algorithm actually outputs a spanning forest, whereas our lower bound holds even for algorithms that only decide connectivity; see Corollary 4.2 for details.

Another important corollary is that any constant-pass streaming algorithm for global minimum cuts in unweighted graphs, also known as edge connectivity, requires $\Omega(N \log N)$ space. It implies that the two-pass algorithm in [AD21] is indeed optimal, even in logarithmic factors. This is formalized in Corollary 6.3.

1.2 Related Work

Pointer chasing. Initially introduced by [PS84], the problem of pointer chasing $\mathsf{PC}_{n,k}$ has played a central role in communication complexity. Specifically, the problem instance consists of k functions $f_1, \ldots, f_k : [n] \to [n]$, together with some fixed starting vertex v_0. Alice owns f_i for odd i and Bob owns f_i for even i. For $i \in [k]$, recursively define the i-th target vertex $v_i = f_i(v_{i-1})$. The goal is for both parties to output v_k. For the bit version $\mathsf{bPC}_{n,k}$, protocols only need to decide the parity of v_k.

In the original version of [PS84], Alice holds a single function f_{A}, and $f_1 = f_3 = \cdots = f_{\mathsf{A}}$. Similarly for Bob, $f_2 = f_4 = \cdots = f_{\mathsf{B}}$. A straightforward protocol uses only $\mathcal{O}(k \log n)$ communication in k rounds, by exchanging target vertices v_i. A long line of work has thus focused on protocols with $r \le k-1$ rounds, or k rounds but starting with Bob. [NW93] gave a $(k-1)$-round randomized protocol using $\mathcal{O}((n/k+k) \cdot \log n)$ communication. On the lower bound side, a series of results [NW93, Kla00, Yeh20, MYZ25] leads to an $\Omega(n/k + k)$ randomized lower bound even for the bit version. Additionally, for constant k, [DJS98] showed a k-round deterministic protocol starting with Bob that uses $\mathcal{O}(n \log^{(k-1)} n)$ communication, which is shown to be tight in [PRV01] via a matching $\Omega(n \log^{(k-1)} n)$ randomized lower bound.

For reductions to connectivity and other related graph problems, it is also useful to study the variants of pointer chasing where each function f_i is promised to be a permutation; see for example [AN21, RRV26]. We denote these variants by $\mathsf{PC\text{-}Perm}_{n,k}, \mathsf{bPC\text{-}Perm}_{n,k}$.

[4] The communication lower bound in Theorem 1.4 is independent of the number of rounds, so long as r is not too large; see Theorems 7.1 and C.1 for the precise range of r.

Beyond constant error. The standard randomized communication complexity concerns protocols with some fixed constant error probability, typically 1/3. In many cases, however, it is also important to understand tradeoffs between communication and error probability. That is, we want to obtain error-dependent communication bounds. Examples[5] include error probabilities close to 1 (i.e., small success probabilities) [BRWY13, BW15, JPY16], close to 1/2 (i.e., small advantages) [BM13, Yu22], close to 0 [BGPW13, HPZZ21, LMdW23], and asymmetric error probabilities [GW16].

Other computation models. Communication complexity has proven to be an extremely useful tool in a wide variety of computation models. For example, via standard connections, strong results are known for streaming [FKM+08, GKK12, SW15, GO16, AD21, AKZ24, AKNS24, ABK+26, EKSZ26], sketching [NY19, AKO20, Yu21, AKZ25], and query algorithms [AEG+22, BvdBE+22, ASW25, JNS26].

As a fundamental graph property, connectivity in particular has been studied in many different settings other than two-party communication. The most closely related one is multi-party communication [PVZ16, WZ17]. Besides, efficient algorithms and strong lower bounds for graph connectivity have also been established for streaming [FKM+08, SW15], sketching [AGM12, NY19, Yu21], and query algorithms [ACK21, AEG+22].

Quantum communication complexity. With the rapid development of quantum computation, quantum communication complexity has attracted increasing attention as well. Strong quantum bounds are shown for many classical communication problems, such as graph connectivity [IKL+12, EKNP14], pointer chasing [JRS02a, JRS02b, KNTZ07], inner product [Kre95, NS06, CvDNT13], and set disjointness [BCW98, HdW02, Raz03, JRS03, AA05, KSdW07, BGK+18].

2 Technical Overview

We provide a high-level overview of our proofs in this section. Section 2.1 is dedicated to the nondeterministic lower bound and compares our approach with [RS95]. Towards the end, we also briefly discuss how to generalize our techniques to prove error-dependent lower bounds for randomized protocols. Section 2.2 first motivates our new two-party communication problem $\mathsf{IP}\circ\mathsf{PC}$ by a comparison with the pointer-chasing-based approach of [RRV26], and then sketches the intuition for the communication lower bound on $\mathsf{IP}\circ\mathsf{PC}$.

2.1 Nondeterministic Lower Bound

Nondeterministic communication complexity is by definition closely related to monochromatic rectangles. In particular, upper bounds on the maximum size of 1-monochromatic rectangles imply nondeterministic communication lower bounds.

Let $n = N/2$. The approach of [RS95] focused on a special collection $\mathcal{S}$ of $n! = 2^{n\log n - \mathcal{O}(n)}$ permutation pairs $(a, b) \in \mathsf{bConn}_N^{-1}(1)$ such that any 1-monochromatic rectangle $\mathcal{A} \times \mathcal{B}$

[5] The first two cases are also closely related to direct product theorems and XOR lemmas.

contains at most $(n/\mathsf{poly}\log(n))^n = 2^{n\log n-\Omega(n\log\log n)}$ elements of $\mathcal{S}$. This yields their $\Omega(n\log\log n)$ nondeterministic lower bound. A closer look reveals that all permutation pairs $(a,b)\in\mathcal{S}$ actually have distinct a (and distinct b). Thus, the argument of [RS95] upper bounds the projection of $(\mathcal{A}\times\mathcal{B})\cap\mathcal{S}$ onto either coordinate. If we additionally assume that every row (column) of $\mathcal{A}\times\mathcal{B}$ contains some element of $\mathcal{S}$, [RS95] essentially bounds the maximum height (width) of certain 1-monochromatic rectangles, i.e.,

$$\max_{\mathcal{A}\times\mathcal{B}\subseteq\mathsf{bConn}_N^{-1}(1)}|\mathcal{A}| = \max_{\mathcal{A}\times\mathcal{B}\subseteq\mathsf{bConn}_N^{-1}(1)}|\mathcal{B}|,$$

subject to the assumption on $\mathcal{A}\times\mathcal{B}$. Instead, we directly upper bound the (logarithmic) size of all 1-monochromatic rectangles:

$$S_n = \max_{\mathcal{A}\times\mathcal{B}\subseteq\mathsf{bConn}_N^{-1}(1)}\log|\mathcal{A}\times\mathcal{B}|.$$

Consider two independent random variables $A\sim\mathcal{A}$ and $B\sim\mathcal{B}$, where $\mathcal{A}\times\mathcal{B}$ is the maximizer in the above definition, so $\mathbb{H}(A,B)=S_n$. In the bipartite graph view, A,B induce a Hamiltonian cycle C of length $N=2n$ among left vertices $[n]_{\mathsf{L}}$ and right vertices $[n]_{\mathsf{R}}$. Moreover, the edges of C are provided by A,B alternately.

The starting point of our proof is the following observation (Lemma 4.4). Suppose we know one edge $(u_{\mathsf{L}},v_{\mathsf{R}})$ supplied by A, so $A(u)=v$. Its two adjacent edges[6] $(u_{\mathsf{L}},B(u)_{\mathsf{R}})$ and $(B^{-1}(v)_{\mathsf{L}},v_{\mathsf{R}})$ must be supplied by B. Imagine "shortcutting" these three edges by a new edge $(B^{-1}(v)_{\mathsf{L}},B(u)_{\mathsf{R}})$. Then, we actually get another cycle of length $2(n-1)=N-2$ on all vertices except $u_{\mathsf{L}},v_{\mathsf{R}}$. In other words, up to renaming of the remaining vertices, we have reduced the case to 1-monochromatic rectangles for a smaller input size.

Let $A_{(u,v)}:[n]\backslash\{u\}\to[n]\backslash\{v\}$ denote the restriction of A on $[n]\backslash\{u\}$. Also let $B_{(u,v)}:[n]\backslash\{u\}\to[n]\setminus\{v\}$ be the restriction of B on $[n]\setminus\{u\}$ except redefining $B_{(u,v)}(B^{-1}(v))=B(u)$. Note that specifying $A(u)=v$ reveals at most $\log n$ bits of information on average while $\mathbb{H}(A_{(u,v)},B_{(u,v)})$ is at most S_{n-1} by the argument above. As A is fully determined given $A(u)=v$ and $A_{(u,v)}$, it now boils down to upper bounding the remaining randomness in B.

Equivalently, we want to show that there cannot be too many $b'\neq b\in\mathcal{B}$ satisfying $b'_{(u,v)}=b_{(u,v)}$ for any fixed $b\in\mathcal{B}$. A trivial upper bound of $n-1$ immediately follows from the observation that $b_{(u,v)}$ specifies $n-1$ edges for b', out of which only one is newly introduced by "shortcutting". However, an increase of $\log n+\log(n-1)$ for each recursive step leads to an upper bound of $\log(n!)+\log((n-1)!)$ on S_n. Meanwhile, it can be verified that $|\mathsf{bConn}_N^{-1}(1)|=n!(n-1)!$. No meaningful result can be derived from such a bound.

The key to a much tighter bound is realizing that $\mathcal{A}\times\mathcal{B}$ being 1-monochromatic actually imposes more restrictions on b', as $\mathsf{bConn}_N(a,b')=1$ must also hold for all $a\in\mathcal{A}$. A careful analysis (Lemma 4.5) establishes two necessary conditions for $b'_{(u,v)}=b_{(u,v)}$:

1. It must be that[7] $b^{-1}b'=(u\ w\ b^{-1}(v))$ for some $w\in[n]\setminus\{u,b^{-1}(v)\}$;
2. Moreover, $u,b^{-1}(v),w$ must follow the same cyclic order $(u\ b^{-1}(v)\ w)$ along the cycle $b^{-1}a$ for every $a\in\mathcal{A}$.

[6] These are all distinct edges as $B(u)\neq v$. Otherwise, the graph is not Hamiltonian for $n>1$.

[7] See Section 3 for the cycle decomposition representation of permutations and also Figure 1 for a detailed illustration of the first condition.

The first condition essentially rephrases the requirement in the trivial upper bound above. Intuitively, the more important second condition is useful because the cyclic order of many triples of vertices must be fixed if we want to have a lot of w satisfying the condition. On the other hand, a large $\mathcal{A}$ means high randomness in the cycle $b^{-1}a$, which in turn implies that most triples of vertices should not have a fixed cyclic order. The conflict between these two requirements suggests a better upper bound.

Formalizing this intuition involves a delicate edge coloring argument (Lemma 4.6). After all, we manage to show that the remaining randomness in B conditioned on $B_{(u,v)}$ is in fact bounded away from $\log n$ by at least[8] $\Omega(\mathbb{H}(A)/n) = \Omega(S_n/n)$, up to lower order terms, when averaged over (u, v). Putting everything together, we have that

$$S_n \lesssim \log n + S_{n-1} + \log n - \Omega\left(\frac{S_n}{n}\right).$$

Solving the recurrence then yields

$$S_n \lesssim (2 - \Omega(1)) \cdot n \log n,$$

concluding our proof.

As to randomized protocols, we generalize our techniques to upper bound the size of any weighted, almost 1-monochromatic rectangle. Specifically, suppose

$$\Pr(\mathsf{bConn}_N(A, B) = 0) \leq \delta,$$

for independent random permutations A, B. We prove an upper bound (Lemma 5.2) on $\mathbb{H}(A, B)$, similar to the nondeterministic one, with an additional $\mathcal{O}(\sqrt{\delta} \cdot \mathsf{poly}(n))$ loss. Indeed, a typical $A_{(u,v)} \times B_{(u,v)}$ is also δ-close to 1-monochromatic on average. Thus, the recursive approach for the nondeterministic setting still applies. The only source of the additional $\mathcal{O}(\sqrt{\delta} \cdot \mathsf{poly}(n))$ loss comes from upper bounding the remaining randomness in B (Lemma 5.3). Overall, as typical transcripts on input from $\mathsf{bConn}_N^{-1}(1)$ correspond to weighted, almost 1-monochromatic rectangles, optimal $\Omega(n \log n)$ lower bounds can be derived for polynomially small error probabilities. The full lower bound for the entire scale of error probabilities then follows by repetition.

2.2 Randomized Lower Bound with Constant Error

The 2-round lower bound of [RRV26] for randomized protocols solving connectivity crucially relies on the bit version $\mathsf{bPC\text{-}Perm}_{n,k}$ of pointer chasing with permutations. Ordinarily, the literature around pointer chasing primarily concerns the separation between $(k-1)$-round and k-round protocols. Different variants are typically irrelevant to the communication complexity, up to logarithmic factors, in such a case. Upper bounds mentioned in Section 1.2 all apply to $\mathsf{PC}_{n,k}$, while many lower bounds can be adapted to hold even for $\mathsf{bPC\text{-}Perm}_{n,k}$.

However, the situation changes dramatically if we do care about the logarithmic gap. [RRV26] gave a protocol for solving[9] $\mathsf{PC}_{n,k}$ with $\mathcal{O}(nk \log \log n)$ communication in roughly

[8] Here, we actually assume $|\mathcal{A}| \geq |\mathcal{B}|$. This is without loss of generality by symmetry.

[9] Their protocols are stated for $\mathsf{PC\text{-}Perm}_{n,k}$, but also work for $\mathsf{PC}_{n,k}$ in general.

$k/3$ rounds. The round complexity can be reduced to about $k/4$ for the bit version. No protocols with significantly fewer rounds are known. In sharp contrast, we present in Appendix B round-efficient protocols for solving $\mathsf{PC\text{-}Perm}_{n,k}$ and $\mathsf{bPC\text{-}Perm}_{n,k}$ with $\mathcal{O}(nk \log\log n)$ communication in only $\mathcal{O}(\sqrt{k})$ rounds.

The 2-round lower bound of [RRV26] for connectivity follows from an optimal $\Omega(n \log n)$ communication lower bound for 2-round[10] protocols solving $\mathsf{bPC\text{-}Perm}_{n,7}$. Both their protocol and our protocol can solve $\mathsf{bPC\text{-}Perm}_{n,6}$ with $\mathcal{O}(n \log\log n)$ communication in 2 rounds. With 3 rounds and $\mathcal{O}(n \log\log n)$ communication, their protocol can only solve $\mathsf{bPC\text{-}Perm}_{n,10}$, whereas our protocol can solve $\mathsf{bPC\text{-}Perm}_{n,12}$. The smallest hard instance that has not been ruled out is thus $\mathsf{bPC\text{-}Perm}_{n,13}$. For more rounds and general k, to the best of our knowledge, it is unknown if $\Omega(n \log n)$ communication lower bounds can be proved for any variant of pointer chasing, even for protocols with far fewer than $\sqrt{k}$ rounds.

The protocol in Appendix B essentially preprocesses disjoint intervals of layers in parallel. For each interval, almost balanced partitions into $\Theta(\log n)$ parts are constructed for each layer such that they are consistent with the permutations, at the communication cost of $\mathcal{O}(n \log\log n)$ per layer. Then, revealing the starting vertex allows the protocol to jump over each interval while only requiring the exact pointers for vertices within the same part. In a nutshell, the protocol is only forced to send full information about a $\Theta(1/\log n)$ fraction of the pointers in each layer, once the starting vertex is fixed.

This motivates our new two-party communication problem $\mathsf{IP}\circ\mathsf{PC}_{n,r}$. Recall that protocols for $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ have to output $\langle A_r \cdots A_1 X, B_r \cdots B_1 Y \rangle$, where X, Y are bit vectors and $A_1, \ldots, A_r, B_1, \ldots, B_r$ are permutations. Intuitively, permutations $A_1, \ldots, A_r$ are expected to send a random X to a random bit vector for the final inner product. Similarly for $B_1, \ldots, B_r$ and Y. Therefore, if some protocol only chased a small number of coordinates, it should not be able to predict the inner product very well.

The proof of the communication lower bound for $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ involves delicate information-theoretic arguments. The high level plan is straightforward: we keep track of the joint distribution of $A_i \cdots A_1 X$ and $B_i \cdots B_1 Y$ in round i during execution of the protocol, by bounding its deviation from uniform. Nevertheless, proving the tight bound, in particular ensuring no $\mathsf{poly}(r, \log n)$ factor loss, turns out to be quite nontrivial and requires extensive use of Rényi entropies $\mathbb{H}_\alpha(\cdot)$ (specifically, for $\alpha = 2, \infty$, namely collision entropy and min-entropy). Plenty of further intuitions and more technical details are provided throughout the full proof in Section 7.

Once lower bounds for $\mathsf{IP}\circ\mathsf{PC}$ are shown, our reduction to graph connectivity uses a layered construction similar to those in [AN21, RRV26].

3 Preliminaries

Notation. For $n \geq 1$, we write $[n]$ as a shorthand for $\{1, \ldots, n\}$. For any tuple $x = (x_1, x_2, \ldots)$ and $i \geq 1$, define $x_{<i}$ to be $(x_1, \ldots, x_{i-1})$. Similarly define $x_{\leq i}$, $x_{\geq i}$, etc. For

[10] To better align with our convention for graph connectivity protocols, we assume throughout this paper that only the receiver of the last message has to output. The results in [RRV26] were stated in the standard context of pointer chasing, where both parties need to output. This changes the number of rounds by 1. Both their protocol and our protocol start with Bob.

two bit vectors x, y, we write $\langle x, y \rangle$ for their binary inner product. $\circ$ denotes composition of functions. Logarithms are always base 2. Iterated logarithm of order k is denoted $\log^{(k)}(\cdot)$.

Permutations. For $n \geq 1$, an n-permutation σ is a bijection from $[n]$ to $[n]$. Let $\mathcal{P}_n$ be the collection of all n-permutations, so $|\mathcal{P}_n| = n!$. We also use the standard cycle decomposition representation for permutations. For example, $\sigma = (1\ 4)(2\ 10\ 5)$ means that $\sigma(1) = 4$, $\sigma(4) = 1$, $\sigma(2) = 10$, etc., and all other elements (of omitted 1-cycles) map to themselves. For uniqueness, the smallest element of each cycle is written first, and the cycles are arranged in increasing order of their smallest elements. The number of cycles in the decomposition of σ is denoted by $\mathsf{cyc}(\sigma)$. Let $\mathcal{C}_n \subseteq \mathcal{P}_n$ be the collection of all n-cycles, which are exactly n-permutations σ with $\mathsf{cyc}(\sigma) = 1$, so $|\mathcal{C}_n| = (n-1)!$. Additionally, with slight abuse of notation, an n-permutation σ can be equivalently represented by an $n \times n$ binary matrix where $\sigma_{j,i} = \mathbb{1}(j = \sigma(i))$ (i.e., mapping the i-th standard basis vector to the $\sigma(i)$-th). So $\sigma^{-1} = \sigma^\top$. The specific representation we use will be clear from context.

Information theory. For any distribution Q, the support of Q is denoted by $\mathsf{supp}(Q)$. We use $x \sim Q$ to mean an outcome drawn according to Q. For any set S, we also write $x \sim S$ when x is drawn from S uniformly at random. With slight abuse of notation, we may write a random variable X interchangeably with its distribution. For $\alpha \in [0, \infty]$, the Rényi entropy of order α of a random variable X is denoted by $\mathbb{H}_\alpha(X)$. The cases of $\alpha = 1, 2, \infty$ are also known as Shannon entropy, collision entropy, and min-entropy, respectively. We use $\mathbb{D}_\alpha(X\|Y)$ to denote the Rényi divergence of order α of X from Y. The case of $\alpha = 1$ is also known as Kullback-Leibler (KL) divergence. When Y follows the uniform distribution over a universe $\mathcal{U} \supseteq \mathsf{supp}(X)$, we simply write $\mathbb{D}_\alpha(X) = \mathbb{D}_\alpha(X\|Y) = \log|\mathcal{U}| - \mathbb{H}_\alpha(X)$. When there is no confusion, we use the convention $\mathbb{H}(\cdot) = \mathbb{H}_1(\cdot)$ and $\mathbb{D}(\cdot\|\cdot) = \mathbb{D}_{\mathsf{KL}}(\cdot\|\cdot) = \mathbb{D}_1(\cdot\|\cdot)$. The mutual information between two random variables X and Y is denoted by $\mathbb{I}(X;Y)$. Appendix A provides a summary of the necessary background and tools we use in information theory. Throughout, random variables are denoted by uppercase letters, and we use the corresponding lowercase letters for their outcomes.

Communication complexity. We use the standard two-party communication model in this paper. Deterministic communication complexity is denoted by $\mathbf{D}(\cdot)$ while nondeterministic communication complexity is denoted by $\mathbf{N}(\cdot)$. For $r \geq 1$ and $\varepsilon \in [0, 1]$, we use $\mathbf{R}^{(r)}_\varepsilon(\cdot)$ for the r-round randomized communication complexity with error probability ε. When the number of rounds is unbounded, we simply write $\mathbf{R}_\varepsilon(\cdot)$. For any constant error probability $\varepsilon \in (0, 1/2)$, the subscript is omitted, as it can always be assumed without loss of generality to be $1/3$ by repetition. We use the convention that Alice sends to Bob in the first round and that output is only required for the receiver of the last message, unless explicitly stated otherwise. Throughout, communication is always measured in the worst case.

4 Nondeterministic Lower Bound for Connectivity

In this section, we prove Theorem 4.1. The upper bound follows from a simple one-way deterministic protocol that communicates to Bob a spanning forest of Alice's edges.

Theorem 4.1. *For even $N > 1$, the nondeterministic communication complexity of bConn_N is $\mathbf{N}(\mathsf{bConn}_N) = \Theta(N \log N)$, regardless of the number of rounds.*

As an immediate corollary, we obtain an optimal query lower bound for graph connectivity in the zero-error setting. It implies that the cut query algorithm in [AEG+22, Theorem 1.2] is optimal, even if only connectivity has to be decided (without outputting a spanning forest).

Corollary 4.2. *For $N \geq 1$, any zero-error randomized cut query algorithm that decides connectivity in N-vertex graphs, requires at least $\Omega(N)$ queries in expectation.*

Proof. By Markov's inequality, any zero-error protocol with $o(N)$ queries in expectation can be transformed into another one with $o(N)$ queries in the worst case, that has one-sided constant error only in the connected case. As the total number of possible inputs is no more than $2^{\mathcal{O}(N^2)}$, the use of randomness can be reduced to $\mathcal{O}(\log N)$ by Newman's theorem [New91], while preserving one-sided constant error. In the meantime, each cut query can be simulated using $\mathcal{O}(\log N)$ communication. Altogether, it implies a randomized protocol with $o(N \log N)$ communication and one-sided constant error, contradicting Theorem 4.1, since nondeterministic protocols are at least as easy. □

We prove Theorem 4.1 in the rest of this section. Fix sufficiently large $n = N/2$. The total number of input combinations $a, b \in \mathcal{P}_n$ with $\mathsf{bConn}_N(a, b) = 1$ is $T = n!(n-1)!$. To see this, pick an arbitrary permutation $b \in \mathcal{P}_n$ and an arbitrary cycle $c \in \mathcal{C}_n$. Then, $\mathsf{bConn}_N(bc, b) = 1$ as $b^{-1}(bc) = c$. Furthermore, (bc, b) enumerates exactly all possibilities in $\mathsf{bConn}_N^{-1}(1)$, and there are $|\mathcal{P}_n| = n!$ choices for b and $|\mathcal{C}_n| = (n-1)!$ choices for c.

Let $\mathcal{A} \times \mathcal{B} \subseteq \mathcal{P}_n \times \mathcal{P}_n$ be any 1-monochromatic rectangle for bConn_N. We want to show that the size of $\mathcal{A} \times \mathcal{B}$ must be substantially smaller than T. In particular, any bound of the form $\log|\mathcal{A} \times \mathcal{B}| \leq (2 - \Omega(1)) \cdot n \log n$ would suffice to prove Theorem 4.1. This is because $\log T = (2 - o(1)) \cdot n \log n$, implying that at least $2^{\Omega(n \log n)}$ 1-monochromatic rectangles are required to cover $\mathsf{bConn}_N^{-1}(1)$.

To this end, define

$$S_n = \max_{\mathcal{A} \times \mathcal{B} \subseteq \mathsf{bConn}_N^{-1}(1)} \log|\mathcal{A} \times \mathcal{B}|.$$

Lemma 4.3 shows a recurrence for S_n that is strong enough to obtain the desired bound.

Lemma 4.3. *For $n > 1$, it holds that*

$$\left(1 + \frac{1}{4n}\right) \cdot S_n \leq S_{n-1} + 2 \log n + \mathcal{O}(\log \log n).$$

We first prove Theorem 4.1, assuming Lemma 4.3, whose proof is deferred to Section 4.1.

Proof of Theorem 4.1. For some sufficiently large constant K, we solve the recurrence in Lemma 4.3 as follows:

$$\begin{aligned} S_n &\leq \frac{4n}{4n+1} \cdot S_{n-1} + \frac{4n}{4n+1} \cdot (2 \log n + K \log \log n) \\ &\leq \frac{4n(4n-4)}{(4n+1)(4n-3)} \cdot S_{n-2} + \left(\frac{4n}{4n+1} + \frac{4n(4n-4)}{(4n+1)(4n-3)}\right) \cdot (2 \log n + K \log \log n) \end{aligned}$$

$$\vdots$$

$$\leq \left(\prod_{i\in[2,n]} \frac{4i}{4i+1}\right) \cdot S_1 + \left(\sum_{i\in[2,n]} \prod_{j\in[i,n]} \frac{4j}{4j+1}\right) \cdot (2\log n + K\log\log n)$$

$$\leq \left(\sum_{i\in[n]} \prod_{j\in[i,n]} \frac{4j}{4j+1}\right) \cdot (2\log n + K\log\log n), \tag{1}$$

as $S_1 = \log 1 = 0$. Define

$$f_n = \sum_{i\in[n]} \prod_{j\in[i,n]} \frac{4j}{4j+1}.$$

Then, we have that $f_1 = 4/5$ and that $f_n = 4n/(4n+1) \cdot (f_{n-1}+1)$. By induction on n, it can be verified that $f_n = 4n/5$. Plugging it back into Equation (1) yields

$$S_n \leq \frac{8}{5} n\log n + \mathcal{O}(n\log\log n).$$

Observe that at least $T/2^{S_n}$ 1-monochromatic rectangles are required to cover $\mathsf{bConn}_N^{-1}(1)$. Consequently, the nondeterministic communication complexity of bConn_N is at least

$$\log\frac{T}{2^{S_n}} = \log T - S_n \geq (2-o(1))\cdot n\log n - \frac{8}{5}n\log n - \mathcal{O}(n\log\log n) = \Omega(N\log N),$$

by Stirling's approximation, as claimed. □

4.1 Proof of Lemma 4.3

We now prove the recurrence in Lemma 4.3. Fix $\mathcal{A}\times\mathcal{B}$ to be the maximizer in the definition of S_n. Without loss of generality, assume $|\mathcal{A}| \geq |\mathcal{B}|$. Consider two independent and uniformly random variables $A \sim \mathcal{A}$ and $B \sim \mathcal{B}$, so we have that

$$\mathbb{H}(A) + \mathbb{H}(B) = \mathbb{H}(A,B) = S_n \qquad \text{and} \qquad \mathbb{H}(A) \geq \frac{S_n}{2}.$$

Fix $u, v \in [n]$. Suppose $a(u) = v$ for some $a \in \mathcal{A}$. Then, $b(u) \neq v$ must hold for all $b \in \mathcal{B}$, as otherwise $b^{-1}a$ contains a singleton cycle, contradicting $(a,b) \in \mathcal{A}\times\mathcal{B} \subseteq \mathsf{bConn}_N^{-1}(1)$. Define $a_{(u,v)} : [n]\setminus\{u\} \to [n]\setminus\{v\}$ to be the restriction of a on $[n]\setminus\{u\}$. Also define $b_{(u,v)} : [n]\setminus\{u\} \to [n]\setminus\{v\}$ to be the same as b except that $b_{(u,v)}(b^{-1}(v)) = b(u)$. That is, $b_{(u,v)}$ redirects the preimage of v, namely $b^{-1}(v) \neq u$, to the image of u, namely $b(u) \neq v$. For any $u, v \in [n]$ such that $a(u) = v$ for some $a \in \mathcal{A}$, define

$$\mathcal{A}_{(u,v)} = \{a_{(u,v)} \mid a \in \mathcal{A}, a(u) = v\}, \qquad \text{and} \qquad \mathcal{B}_{(u,v)} = \{b_{(u,v)} \mid b \in \mathcal{B}\}.$$

Lemma 4.4 shows that $\mathcal{A}_{(u,v)} \times \mathcal{B}_{(u,v)}$ also constitutes a 1-monochromatic rectangle for $n-1$, after renaming both $[n]\setminus\{u\}$ and $[n]\setminus\{v\}$ to $[n-1]$.

Lemma 4.4. *For any $u, v \in [n]$ such that $a(u) = v$ for some $a \in \mathcal{A}$, it must be that $\widehat{b}^{-1}\widehat{a}$ consists of a single $(n-1)$-cycle on $[n]\setminus\{u\}$, for all $(\widehat{a},\widehat{b}) \in \mathcal{A}_{(u,v)} \times \mathcal{B}_{(u,v)}$.*

Proof. Fix any $a \in \mathcal{A}$ such that $a(u) = v$, and $b \in \mathcal{B}$. Consider a bipartite graph with vertex set $[n]$ on each side. We use subscripts L, R to distinguish vertices on two sides. Note that each of a and b corresponds to a perfect matching, and their union forms a single $2n$-cycle, due to $(a, b) \in \mathsf{bConn}_N^{-1}(1)$. Furthermore, $b(u)_\mathsf{R}, u_\mathsf{L}, v_\mathsf{R}, b^{-1}(v)_\mathsf{L}$ actually constitute a length-3 subpath on the cycle. The net effect of $a_{(u,v)}, b_{(u,v)}$ is simply replacing this subpath with an edge between $b(u)_\mathsf{R}$ and $b^{-1}(v)_\mathsf{L}$. As a result, the edges corresponding to $a_{(u,v)}, b_{(u,v)}$ form a single $2(n-1)$-cycle on vertices excluding $u_\mathsf{L}, v_\mathsf{R}$, concluding the proof. □

After conditioning on $A(u) = v$, observe that $A_{(u,v)}$ is now a uniformly random element of $\mathcal{A}_{(u,v)}$. However, $B_{(u,v)}$ is not necessarily uniform over $\mathcal{B}_{(u,v)}$ as each $b_{(u,v)}$ may arise from a different number of $b \in \mathcal{B}$. Nevertheless, Lemma 4.4 implies that

$$\log|\mathcal{A}_{(u,v)} \times \mathcal{B}_{(u,v)}| \le S_{n-1}.$$

As will be seen later, this is exactly the source of the S_{n-1} term in the recurrence in Lemma 4.3. In addition, since $A(u)$ has at most n possibilities, its uncertainty contributes one $\log n$ term. What remains is to bound the contribution coming from multiplicity of $b_{(u,v)}$.

To this end, for $x \neq y \in [n]$, let $\mathcal{W}_{(x,y)}(b)$ be the collection of $w \in [n] \setminus \{x, y\}$ such that $(w\ x\ y)$ is compatible with the cycle $b^{-1}a$ for every $a \in \mathcal{A}$. That is, w, x, y follow the cyclic order $(w\ x\ y)$, instead of the only other possibility $(w\ y\ x)$, along $b^{-1}a$.

Lemma 4.5 upper bounds the multiplicity of $b_{(u,v)}$.

Lemma 4.5. *For any $u, v \in [n]$ such that $a(u) = v$ for some $a \in \mathcal{A}$, and $b \in \mathcal{B}$, it holds that*

$$|\{b' \in \mathcal{B} \mid b'_{(u,v)} = b_{(u,v)}\}| \le |\mathcal{W}_{(u,b^{-1}(v))}(b)| + 1.$$

Proof. Let $\widehat{b} = b_{(u,v)}$ and $z = b^{-1}(v) \neq u$ for notational convenience. We use the same bipartite graph view as in the proof of Lemma 4.4. The net effect from $b^{-1}a$ to $b_{(u,v)}^{-1}a_{(u,v)}$ is replacing the length-3 subpath $(b(u)_\mathsf{R}, u_\mathsf{L}, v_\mathsf{R}, z_\mathsf{L})$ by the edge between $b(u)_\mathsf{R}$ and z_L, so $\widehat{b}(z) = b(u)$. Similarly, for any $b' \in \mathcal{B}$ such that $b'_{(u,v)} = b_{(u,v)} = \widehat{b}$, the net effect from $(b')^{-1}a$ to $(b'_{(u,v)})^{-1}a_{(u,v)}$ is replacing the length-3 subpath $(b'(u)_\mathsf{R}, u_\mathsf{L}, v_\mathsf{R}, w_\mathsf{L})$ by the edge between $b'(u)_\mathsf{R}$ and w_L, where $w = (b')^{-1}(v) \neq u$, so $\widehat{b}(w) = b'(u)$.

Observe that $b, b', \widehat{b}$ all share the same image on $[n] \setminus \{u, w, z\}$. If $w = z$, then we additionally have $b'(u) = \widehat{b}(w) = \widehat{b}(z) = b(u)$ and $b'(w) = v = b(z)$ by definition. As a result, it must be that $b' = b$.

Now suppose $w \neq z$. b provides edges $(z_\mathsf{L}, v_\mathsf{R})$ and $(u_\mathsf{L}, b(u)_\mathsf{R})$, while b' provides edges $(w_\mathsf{L}, v_\mathsf{R})$ and $(u_\mathsf{L}, b'(u)_\mathsf{R})$. In the meantime, due to $b_{(u,v)} = \widehat{b} = b'_{(u,v)}$, we know that $b'(z) = b'_{(u,v)}(z) = \widehat{b}(z) = b(u)$, so b' also provides the edge $(z_\mathsf{L}, b(u)_\mathsf{R})$. Similarly, we have that $b(w) = b_{(u,v)}(w) = \widehat{b}(w) = b'(u)$, so b also provides the edge $(w_\mathsf{L}, b'(u)_\mathsf{R})$. Altogether, we obtain a 6-cycle $(u_\mathsf{L}, b'(u)_\mathsf{R}, w_\mathsf{L}, v_\mathsf{R}, z_\mathsf{L}, b(u)_\mathsf{R})$ using alternating edges provided by b', b. As all other edges corresponding to $\widehat{b}$ are common to b, b', it actually implies that $b^{-1}b' = (u\ w\ z)$. See Figure 1 for an illustration of the 6-cycle.

Furthermore, not all $w \in [n] \setminus \{u, z\}$ are valid. In particular, it must be that

$$(b')^{-1}a = (u\ w\ z)^{-1}b^{-1}a = (u\ z\ w)b^{-1}a \in \mathcal{C}_n,$$

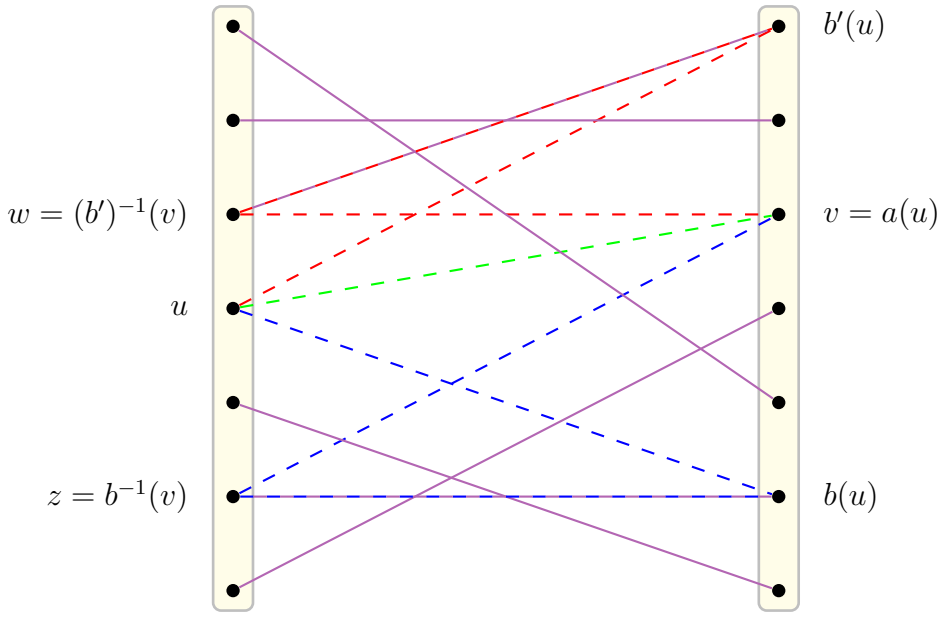


Figure 1: An illustration of the 6-cycle in the proof of Lemma 4.5, with $n = 7$. Red and blue edges correspond to b' and b, respectively. The green edge corresponds to a. Solid purple (=red+blue) edges correspond to $\widehat{b} = b_{(u,v)} = b'_{(u,v)}$. Dashed edges are not included in $\widehat{b}$.

for every $a \in \mathcal{A}$, due to $(a, b') \in \mathsf{bConn}_N^{-1}(1)$. Suppose u, w, z follow the cyclic order $(u\ w\ z)$ along $b^{-1}a$. Equivalently, as $b^{-1}a \in \mathcal{C}_n$, we have that $b^{-1}a = (u\ \widehat{u}\ w\ \widehat{w}\ z\ \widehat{z})$ for some potentially empty sequences $\widehat{u}, \widehat{w}, \widehat{z}$ such that $\widehat{u}, \widehat{w}, \widehat{z}$ constitute a disjoint partition of $[n] \setminus \{u, w, z\}$. Then, we can get

$$(b')^{-1}a = (u\ \widehat{u})(w\ \widehat{w})(z\ \widehat{z}),$$

consisting of three disjoint cycles, a contradiction.

On the other hand, if u, w, z follow the cyclic order $(u\ z\ w)$ along $b^{-1}a$ instead, then we have that $b^{-1}a = (u\ \widehat{u}\ z\ \widehat{z}\ w\ \widehat{w})$. It implies that

$$(b')^{-1}a = (u\ \widehat{u}\ w\ \widehat{w}\ z\ \widehat{z})$$

still consists of a single cycle. Since the above analysis holds for every $a \in \mathcal{A}$, the maximum number of such w, or equivalently b', is exactly $|\mathcal{W}_{(u,z)}(b)|$ by definition.

Together with the possibility of $b' = b$, this concludes the proof. □

Intuitively, if the cycle $b^{-1}a$ is close to random, $\mathcal{W}_{(u,b^{-1}(v))}(b)$ should be relatively small, as otherwise the cyclic order of many triples $(w, u, b^{-1}(v))$ is already fixed. Lemma 4.6 formalizes this intuition, providing the last major piece to the proof of Lemma 4.3.

Lemma 4.6. *For any $b \in \mathcal{B}$, it holds that*

$$\mathop{\mathbb{E}}_{u\sim[n]} \mathop{\mathbb{E}}_{a\sim A} \log\left(|\mathcal{W}_{(u,b^{-1}(a(u)))}(b)| + 1\right) \le \log n - \frac{\mathbb{H}(A)}{2n} + \mathcal{O}(\log\log n).$$

Proof. Let $C = b^{-1}A$. Note that $C \in \mathcal{C}_n$ because of $(A, b) \in \mathsf{bConn}_N^{-1}(1)$. For $u \in [n]$, we say that vertex u has an outgoing edge to $C(u)$ and an incoming edge from $C^{-1}(u)$. In the following, we will color all $n(n-1)$ potential directed edges such that the color ϕ_u of u's outgoing edge and the color ψ_u of u's incoming edge, for all vertices u, uniquely determine the cycle C.

Partition all potential edges (u, v) into $g = \mathcal{O}(\log n)$ groups such that the i-th group consists of all edges (u, v) satisfying $2^{i-1} \le |\mathcal{W}_{(u,v)}(b)| + 1 < 2^i$. This is well-defined because $0 \le |\mathcal{W}_{(u,v)}(b)| \le n-2$ by definition. For each group i, independently sample $w_1^{(i)}, \ldots, w_{t_i}^{(i)} \sim [n]$, where $t_i = \Theta(n/2^i \cdot \log n)$. For any edge (u, v) in group i, by Chernoff bound, there exists $j \in [t_i]$ such that $w_j^{(i)} \in \mathcal{W}_{(u,v)}(b) \cup \{u\}$, with probability at least $1 - 1/\mathsf{poly}(n)$. Define the color of edge (u, v) to be (i, j). By union bound over all $n(n-1)$ potential edges, such $w_1^{(i)}, \ldots, w_{t_i}^{(i)}$ for every group i and thus the overall coloring can be fixed.

Consider $U \sim [n]$ and $V = C(U)$. Let (I, J) denote the color of (U, V). It holds that

$$\begin{aligned}
\mathbb{H}(I, J) &= \mathbb{H}(I) + \mathbb{H}(J \mid I) && \text{(by chain rule of entropy (Fact A.1))} \\
&\le \log g + \sum_{i \in [g]} \Pr(I = i) \cdot \log t_i \\
&\le \mathcal{O}(\log \log n) + \sum_{u,v \in [n]} \Pr((U, V) = (u, v)) \cdot \log \frac{\mathcal{O}(n \log n)}{|\mathcal{W}_{(u,v)}(b)| + 1} \\
&\le \log n + \mathcal{O}(\log \log n) - \mathop{\mathbb{E}}_{u \in [n]} \mathop{\mathbb{E}}_{c \sim C} \log\big(|\mathcal{W}_{(u,c(u))}(b)| + 1\big) \\
&= \log n + \mathcal{O}(\log \log n) - \mathop{\mathbb{E}}_{u \in [n]} \mathop{\mathbb{E}}_{a \sim A} \log\big(|\mathcal{W}_{(u,b^{-1}(a(u)))}(b)| + 1\big).
\end{aligned}$$

Note that (U, V) is a uniformly random edge of C. Thus, for outgoing edges, we can get

$$\begin{aligned}
\mathop{\mathbb{E}}_{u \sim [n]} \mathbb{H}(\phi_u) &= \mathop{\mathbb{E}}_{u \sim [n]} \mathbb{H}(I, J \mid U = u) \\
&= \mathbb{H}(I, J \mid U) \\
&\le \mathbb{H}(I, J) \\
&\le \log n + \mathcal{O}(\log \log n) - \mathop{\mathbb{E}}_{u \in [n]} \mathop{\mathbb{E}}_{a \sim A} \log\big(|\mathcal{W}_{(u,b^{-1}(a(u)))}(b)| + 1\big).
\end{aligned}$$

A similar argument applies to incoming edges.

It now remains to prove that C is indeed uniquely determined by ϕ_u, ψ_u for all $u \in [n]$. For every color (i, j), we show that all edges in C with color (i, j) are determined by

$$\mathcal{U}_{(i,j)} = \{u \in [n] \mid \phi_u = (i, j)\} \qquad \text{and} \qquad \mathcal{V}_{(i,j)} = \{v \in [n] \mid \psi_v = (i, j)\}.$$

That is, we need to recover a perfect matching between $\mathcal{U}$ and $\mathcal{V}$ such that every matching edge (u, v) has color (i, j) and satisfies $v = C(u)$. For any $u \in \mathcal{U}$ and $v \in \mathcal{V}$, the edge (u, v) having color (i, j) means $w_j^{(i)} \in \mathcal{W}_{(u,v)}(b) \cup \{u\}$ by construction of the coloring. By definition of $\mathcal{W}_{(u,v)}(b)$, it further implies that either $w_j^{(i)} = u$, or $(w_j^{(i)}\ u\ v)$ is compatible with every $b^{-1}a$ with $a \in \mathcal{A}$, including C. In other words, if we break the cyclic order of C right before $w_j^{(i)}$ into a linear order, v must come after u.

The perfect matching between $\mathcal{U}, \mathcal{V}$ induced by C satisfies that $v = C(u)$ comes immediately after u in the linear order, so the difference in their rankings is exactly 1. Observe that the sum of these differences over all matching edges is fixed, regardless of the perfect matching. As a result, any other perfect matching would have some matching edge (u, v) with negative difference in their rankings, meaning that v comes before u, a contradiction to the color of edge (u, v) being (i, j). Consequently, we can construct a bipartite graph between U and V, and add an edge (u, v) if and only if it has color (i, j). The unique perfect matching then recovers all edges in C with color (i, j), as claimed.

This concludes the proof by rearranging the terms, as

$$\mathbb{H}(A) = \mathbb{H}(C) \leq \sum_{u \in [n]} \mathbb{H}(\phi_u) + \sum_{u \in [n]} \mathbb{H}(\psi_u).$$

□

Now, we are ready to prove Lemma 4.3.

Proof of Lemma 4.3. For any $u \in [n]$, (A, B) can be equivalently represented as

$$(A(u), A_{(u,A(u))}, B_{(u,A(u))}, \beta_{(u,A(u))}(B)),$$

where $\beta_{(u,A(u))}(B) \in [|\mathcal{W}_{(u,B^{-1}(A(u)))}(B)| + 1]$ is a unique label among all $b' \in \mathcal{B}$ such that $b'_{(u,A(u))} = B_{(u,A(u))}$. This is well-defined because of Lemma 4.5. Now observe that

$$\begin{aligned}
\mathbb{H}(A, B) &= \mathbb{H}\big(A(u), A_{(u,A(u))}, B_{(u,A(u))}, \beta_{(u,A(u))}(B)\big) \\
&= \mathbb{H}(A(u)) + \mathbb{H}\big(A_{(u,A(u))}, B_{(u,A(u))} \mid A(u)\big) + \mathbb{H}\big(\beta_{(u,A(u))}(B) \mid A, B_{(u,A(u))}\big) \\
&\qquad \text{(by chain rule of entropy (Fact A.1))} \\
&\leq \log n + S_{n-1} + \mathbb{H}(\beta_{(u,A(u))}(B) \mid A, B_{(u,A(u))}) \\
&\qquad \text{(by Lemma 4.4 and averaging over } v = A(u)) \\
&\leq \log n + S_{n-1} + \mathop{\mathbb{E}}_{(a,b)\sim(A,B)} \log\big(|\mathcal{W}_{(u,b^{-1}(a(u)))}(b)| + 1\big).
\end{aligned}$$

By Lemma 4.6 and averaging over $b \sim \mathcal{B}$ and $u \sim [n]$, it implies that

$$\mathbb{H}(A, B) \leq \log n + S_{n-1} + \log n - \frac{\mathbb{H}(A)}{2n} + \mathcal{O}(\log \log n).$$

The proof is completed by rearranging the terms, since $\mathbb{H}(A) \geq S_n/2$ by assumption. □

5 Lower Bound for Connectivity with Small Error

This section is dedicated to the proof of error-dependent randomized lower bounds for connectivity, as shown in Theorem 5.1.

Theorem 5.1. *For even $N > 1$, and $\varepsilon = \varepsilon(N) \in (0, 1/3)$, any randomized protocol for bConn_N with error probability ε requires at least $\mathbf{R}_\varepsilon(\mathsf{bConn}_N) = \Omega(N \cdot \log(\min(1/\varepsilon, N)))$ communication, regardless of the number of rounds.*

Fix sufficiently large $n = N/2$, and $\varepsilon = \varepsilon(N) \in (0, 1/3)$. The proof of Theorem 5.1 can be viewed as a generalization of the analysis in Section 4. In a nutshell, the key idea is to show that bConn_N does not even admit large almost 1-monochromatic rectangles, including weighted ones. Concretely, the distance of a product distribution[11] Q on $\mathcal{P}_n \times \mathcal{P}_n$ from being 1-monochromatic is defined to be

$$\Delta(Q) = \Pr_{(a,b)\sim Q}(\mathsf{bConn}_N(a, b) = 0).$$

For $\delta \in [0, 1]$, define

$$S_{n,\delta} = \max_{Q:\Delta(Q)\le\delta} \mathbb{H}(Q),$$

where Q ranges over product distributions only.

Lemma 5.2 is the analog of Lemma 4.3. In fact, the special case of $\delta = 0$ recovers the nondeterministic setting.

Lemma 5.2. *For $n \ge 1$ and $\delta \in [0, 1]$, it holds that*

$$S_{n,\delta} \le S'_{n,\delta},$$

where $S'_{n,\delta}$ is defined recursively by

$$\left(1 + \frac{1}{4n}\right) \cdot S'_{n,\delta} = S'_{n-1,\delta} + 2\log n + K \cdot \left(\log\log n + \sqrt{\delta} n^3 \log n\right),$$

for $n > 1$, starting with $S'_{1,\delta} = 0$, for some sufficiently large constant K.

Intuitively, on input $(A, B) \in \mathsf{bConn}_N^{-1}(1)$, the protocol should mostly produce transcripts that lead to output 1 and that correspond to almost 1-monochromatic rectangles. We now prove Theorem 5.1, assuming Lemma 5.2. We will prove Lemma 5.2 in Section 5.1.

Proof of Theorem 5.1. With $S_{1,\delta} = S'_{1,\delta} = 0$, a similar argument to the analysis in the proof of Theorem 4.1 solves the recurrence in Lemma 5.2 to

$$S_{n,\delta} \le S'_{n,\delta} \le \frac{8}{5} n \log n + \mathcal{O}\left(n \log\log n + \sqrt{\delta} n^4 \log n\right). \tag{2}$$

First suppose for the sake of contradiction that there exists a randomized protocol with error probability $\varepsilon \le n^{-10}$ and communication $L = o(n \log n)$. We will later see that the case of $\varepsilon > n^{-10}$ follows by repetition. It suffices to consider independent and uniformly random input $A, B \sim \mathcal{P}_n$. We emphasize, however, that we do not derandomize the protocol and insist that it has error probability ε on every possible input.

Let $\Gamma = (M, R)$ be the transcript of the protocol, where M, R denote the message and public randomness, respectively. Since A and B are still independent conditioned on Γ, with slight abuse of notation, we write $\Delta(\Gamma) = \Delta(A, B \mid \Gamma)$. We also assume the output $O = O(\Gamma)$ is a function of Γ, as otherwise the receiver of the last message can send its output using one extra bit of communication. Replacing L with $L+1$ does not change its asymptotic behavior.

[11] That is, Q is exactly the product of its two marginals.

In addition, Γ is said to be typical if $\Delta(\Gamma)$ is no more than $\delta = n^{-8}$. Let $T = T(\Gamma)$ denote the indicator of Γ being typical. For notational convenience, also let $Z = \mathsf{bConn}_N(A, B)$, so $\Pr(Z = 0 \mid \gamma) \le \delta$ for any γ such that $T(\gamma) = 1$.

Note that (A, B) conditioned on $Z = 1$ is just the uniform distribution[12] on $\mathsf{bConn}_N^{-1}(1)$, which has size $n!(n-1)!$ as shown in Section 4. Therefore, we can get

$$\mathbb{H}(A, B \mid Z = 1) = 2n \log n - \mathcal{O}(n), \tag{3}$$

by Stirling's approximation. In the meantime, by definition, observe that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\gamma|Z=1} O \cdot \Delta(\gamma) &= \sum_{(a,b)\in\mathsf{bConn}_N^{-1}(1)} \Pr((A, B) = (a, b) \mid Z = 1) \cdot \mathop{\mathbb{E}}_{\gamma|(a,b)} O \cdot \Delta(\gamma) \\
&= n \cdot \sum_{(a,b)\in\mathsf{bConn}_N^{-1}(1)} \Pr((A, B) = (a, b)) \cdot \mathop{\mathbb{E}}_{\gamma|(a,b)} O \cdot \Delta(\gamma) \\
&\le n \cdot \sum_{(a,b)\in\mathsf{bConn}_N^{-1}(1)} \Pr((A, B) = (a, b)) \cdot \mathop{\mathbb{E}}_{\gamma|(a,b)} \Pr_{(a',b')|\gamma} (\mathsf{bConn}_N(a', b') \ne O) \\
&\le n \cdot \mathop{\mathbb{E}}_{a,b\sim\mathcal{P}_n} \mathop{\mathbb{E}}_{\gamma|(a,b)} \Pr_{(a',b')|\gamma} (\mathsf{bConn}_N(a', b') \ne O) \\
&= n \cdot \mathop{\mathbb{E}}_{\gamma} \mathop{\mathbb{E}}_{(a,b)|\gamma} \Pr_{(a',b')|\gamma} (\mathsf{bConn}_N(a', b') \ne O) \\
&= n \cdot \mathop{\mathbb{E}}_{\gamma} \Pr_{(a',b')|\gamma} (\mathsf{bConn}_N(a', b') \ne O) \\
&= n \cdot \Pr_{a',b'\sim\mathcal{P}_n} \mathop{\mathbb{E}}_{\gamma|(a',b')} (\mathsf{bConn}_N(a', b') \ne O) \\
&\le n\varepsilon.
\end{aligned}$$

By Markov's inequality, it implies that

$$\Pr(O = 1, T = 0 \mid Z = 1) = \Pr_{\gamma|Z=1} (O \cdot \Delta(\gamma) > \delta) \le \frac{\mathbb{E}_{\gamma|Z=1} O \cdot \Delta(\gamma)}{\delta} \le \frac{n\varepsilon}{\delta}.$$

Since both O and T are functions of Γ only, A and B remain independent conditioned on any fixed γ such that $O(\gamma) = 1$ and $T(\gamma) = 1$. By averaging over all such γ, Equation (2) implies that

$$\mathbb{H}(A, B \mid \Gamma, O = 1, T = 1) \le \frac{8}{5} n \log n + \mathcal{O}(n \log \log n),$$

as $\delta = n^{-8}$. Furthermore, it must be that

$$\begin{aligned}
&\mathbb{H}(A, B \mid \Gamma, O = 1, T = 1) \\
&\quad = \mathbb{H}(A, B, Z \mid \Gamma, O = 1, T = 1) \\
&\quad = \mathbb{H}(A, B \mid \Gamma, Z, O = 1, T = 1) + \mathbb{H}(Z \mid \Gamma, O = 1, T = 1) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(by chain rule of entropy (Fact A.1))} \\
&\quad \ge \Pr(Z = 1 \mid O = 1, T = 1) \cdot \mathbb{H}(A, B \mid \Gamma, Z = 1, O = 1, T = 1) \\
&\quad \ge (1 - \delta) \cdot \mathbb{H}(A, B \mid \Gamma, Z = 1, O = 1, T = 1),
\end{aligned}$$

[12] But it is not a product distribution.

implying that

$$\begin{aligned}\mathbb{H}(A,B \mid \Gamma, Z=1, O=1, T=1) &\le \frac{1}{1-\delta}\cdot \mathbb{H}(A,B \mid \Gamma, O=1, T=1)\\ &\le \frac{8}{5}n\log n + \mathcal{O}(n\log\log n),\end{aligned}$$

due to $\delta = n^{-8}$ as well. Altogether, we can get

$$\begin{aligned}&\mathbb{H}(A,B \mid \Gamma, Z=1)\\ &\quad\le \mathbb{H}(A,B,O,T \mid \Gamma, Z=1)\\ &\quad= \mathbb{H}(O,T \mid \Gamma, Z=1) + \mathbb{H}(A,B \mid \Gamma, O, T, Z=1)\\ &\qquad\qquad\qquad

5.1 Proof of Lemma 5.2

This section constitutes the proof of Lemma 5.2. We take a similar approach to the proof of Lemma 4.3. For any $u, v \in [n]$, we extend the definition of $b_{(u,v)}$ such that it is simply the restriction of b on $[n] \setminus \{u\}$ if $b(u) = v$.

The following Lemma 5.3 is the analog of Lemmas 4.5 and 4.6 in the randomized setting.

Lemma 5.3. *For $\delta \in [0,1]$, and any product distribution Q on $\mathcal{P}_n \times \mathcal{P}_n$ that is δ-close to 1-monochromatic, it holds that*

$$\mathop{\mathbb{E}}_{u\sim[n]} \mathbb{H}(B \mid A(u), B_{(u,A(u))}) \le \log n - \frac{\mathbb{H}(Q_{\mathsf{A}})}{2n} + \mathcal{O}\Big(\log\log n + \sqrt{\delta} n^3 \log n\Big),$$

where $(A, B) \sim Q$, and Q_{A} is the marginal of A.

Proof. For any fixed $u, v \in [n]$ and $b \in \mathcal{P}_n$, it holds that

$$|\{b' \in \mathcal{P}_n \mid b'_{(u,v)} = b_{(u,v)}\}| \le n. \tag{5}$$

Indeed, this can be shown by straightforward case analysis similar to the proof of Lemma 4.5. Also note that $\mathbb{H}(Q_{\mathsf{A}})/(2n) \le \mathcal{O}(\log n)$. Consequently, we can assume without loss of generality that $\sqrt{\delta} n^3 < 1$, as otherwise the $\mathcal{O}(\sqrt{\delta} n^3 \log n)$ term already dominates the upper bound and is trivially larger than the expectation on LHS.

Let Q_{B} be the marginal of B, and $\mathcal{B}^*$ the collection of $b \in \mathcal{B} = \mathsf{supp}(Q_{\mathsf{B}})$ such that

$$\Delta(Q_{\mathsf{A}} \times \{b\}) \le \sqrt{\delta}.$$

Since Q is δ-close to 1-monochromatic, by Markov's inequality, we know that

$$\Pr_{b\sim Q_{\mathsf{B}}}(b \in \mathcal{B}^*) \ge 1 - \sqrt{\delta}.$$

Let I denote the indicator of $B \in \mathcal{B}^*$, so

$$\Pr(I = 0) \le \sqrt{\delta} < 1.$$

As a result, we can get

$$\begin{aligned}
\mathbb{H}(B \mid A(u), B_{(u,A(u))}) &\le \mathbb{H}(B, I \mid A(u), B_{(u,A(u))}) \\
&= \mathbb{H}(I \mid A(u), B_{(u,A(u))}) + \mathbb{H}(B \mid A(u), B_{(u,A(u))}, I) \\
&\qquad\qquad \text{(by chain rule of entropy (Fact A.1))} \\
&\le 1 + \Pr(I = 0) \cdot \mathbb{H}(B \mid A(u), B_{(u,A(u))}, I = 0) \\
&\qquad + \Pr(I = 1) \cdot \mathbb{H}(B \mid A(u), B_{(u,A(u))}, I = 1) \\
&\le 1 + \sqrt{\delta} \cdot \log n + \mathbb{H}(B \mid A(u), B_{(u,A(u))}, I = 1),
\end{aligned} \tag{6}$$

due to Equation (5).

Let Q^*_{B} be the distribution Q_{B} conditioned on $I = 1$, so it suffices to upper bound

$$\mathop{\mathbb{E}}_{u\sim[n]} \mathbb{H}(B^* \mid A(u), B^*_{(u,A(u))}),$$

where $(A, B^*) \sim (Q_{\mathsf{A}}, Q^*_{\mathsf{B}})$. Define the neighborhood of b inside $\mathcal{B}^*$ to be

$$\mathcal{B}^*(b) = \{b' \in \mathcal{B}^* \mid \exists u, v \in [n], b'_{(u,v)} = b_{(u,v)}\}.$$

Note that $\mathcal{B}^*(b)$ always contains b itself and that its size is upper bounded by $n \cdot n^2 = n^3$ because of Equation (5). Also define the neighborhood of b inside $\mathcal{A} = \mathsf{supp}(Q_{\mathsf{A}})$ to be

$$\mathcal{A}^*(b) = \{a \in \mathcal{A} \mid \forall b' \in \mathcal{B}^*(b), (a, b') \in \mathsf{bConn}_N^{-1}(1)\}.$$

Let J denote the indicator of $A \in \mathcal{A}^*(B^*)$. By union bound, we have that

$$\Pr(J = 0 \mid B^* = b) \le \sqrt{\delta} \cdot |\mathcal{B}^*(b)| \le \sqrt{\delta} n^3 < 1.$$

Then, it must hold that

$$\begin{aligned}
\mathbb{H}(B^* \mid A(u), B^*_{(u,A(u))}) &\le \mathbb{H}(B^*, J \mid A(u), B^*_{(u,A(u))}) \\
&= \mathbb{H}(J \mid A(u), B^*_{(u,A(u))}) + \mathbb{H}(B^* \mid A(u), B^*_{(u,A(u))}, J) \\
&\qquad \text{(by chain rule of entropy (Fact A.1))} \\
&\le 1 + \Pr(J = 0) \cdot \mathbb{H}(B^* \mid A(u), B^*_{(u,A(u))}, J = 0) \\
&\quad + \Pr(J = 1) \cdot \mathbb{H}(B^* \mid A(u), B^*_{(u,A(u))}, J = 1) \\
&\le 1 + \sqrt{\delta} n^3 \cdot \log n + \Pr(J = 1) \cdot \mathbb{H}(B^* \mid A(u), B^*_{(u,A(u))}, J = 1), \quad (7)
\end{aligned}$$

again due to Equation (5).

Let $Q^*_{\mathsf{A}}(b)$ be the distribution Q_{A} conditioned on $B^* = b$ and $J = 1$, so we can get

$$\begin{aligned}
\mathbb{H}(Q_{\mathsf{A}}) &= \mathbb{H}(Q_{\mathsf{A}} \mid B^*) \\
&\le \mathbb{H}(Q_{\mathsf{A}}, J \mid B^*) \\
&= \mathbb{H}(J \mid B^*) + \mathbb{H}(Q_{\mathsf{A}} \mid B^*, J) \qquad \text{(by chain rule of entropy (Fact A.1))} \\
&\le 1 + \Pr(J = 0) \cdot \mathbb{H}(Q_{\mathsf{A}} \mid B^*, J = 0) + \Pr(J = 1) \cdot \mathbb{H}(Q_{\mathsf{A}} \mid B^*, J = 1) \\
&\le 1 + \sqrt{\delta} n^3 \cdot n \log n + \Pr(J = 1) \cdot \mathbb{H}(Q^*_{\mathsf{A}}(B^*) \mid B^*, J = 1). \quad (8)
\end{aligned}$$

By definition, $\mathcal{A}^*(b) \times \mathcal{B}^*(b)$ is a 1-monochromatic rectangle, meaning that Lemmas 4.5 and 4.6 actually can be applied[13]. Therefore, we also have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{u \in [n]} \mathbb{H}(B^* \mid A(u), B^*_{(u,A(u))}, J = 1) \\
&\quad \le \mathop{\mathbb{E}}_{u \in [n]} \mathop{\mathbb{E}}_{(a,b) \sim (Q_{\mathsf{A}} \times Q^*_{\mathsf{B}}) \mid J=1} \log |\{b' \in \mathcal{B}^* \mid b'_{(u,a(u))} = b_{(u,a(u))}\}| \\
&\quad = \mathop{\mathbb{E}}_{b \sim Q^*_{\mathsf{B}} \mid J=1} \mathop{\mathbb{E}}_{u \in [n]} \mathop{\mathbb{E}}_{a \sim Q^*_{\mathsf{A}}(b)} \log |\{b' \in \mathcal{B}^*(b) \mid b'_{(u,a(u))} = b_{(u,a(u))}\}| \\
&\quad \le \mathop{\mathbb{E}}_{b \sim Q^*_{\mathsf{B}} \mid J=1} \left(\log n - \frac{\mathbb{H}(Q^*_{\mathsf{A}}(b))}{2n} + \mathcal{O}(\log \log n) \right) \qquad \text{(by Lemmas 4.5 and 4.6)} \\
&\quad = \log n - \frac{\mathbb{H}(Q^*_{\mathsf{A}}(B^*) \mid B^*, J = 1)}{2n} + \mathcal{O}(\log \log n). \quad (9)
\end{aligned}$$

[13] Technically, Lemma 4.6 is stated for A uniformly distributed on its support $\mathcal{A}$. Nevertheless, it actually holds for any distribution of A. The same proof applies verbatim.

Plugging Equation (9) into Equations (6) and (7) and averaging over $u \sim [n]$, we have that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{u\sim[n]} \mathbb{H}(B \mid A(u), B_{(u,A(u))}) \\
&\quad \le \log n - \frac{\Pr(J=1)\cdot \mathbb{H}(Q^*_{\mathsf{A}}(B^*) \mid B^*, J=1)}{2n} + \mathcal{O}\Big(\log\log n + \sqrt{\delta} n^3 \log n\Big) \\
&\quad \le \log n - \frac{\mathbb{H}(Q_{\mathsf{A}})}{2n} + \mathcal{O}\Big(\log\log n + \sqrt{\delta} n^3 \log n\Big),
\end{aligned}$$

by Equation (8), concluding the proof. □

Now, we are ready to prove Lemma 5.2.

Proof of Lemma 5.2. The proof is by induction on n. The base case of $n = 1$ holds because

$$S_{1,\delta} = S'_{1,\delta} = 0.$$

For $n > 1$, fix Q to be the maximizer in the definition of $S_{n,\delta}$. Consider two independent random variables $(A, B) \sim Q$, so

$$\mathbb{H}(A) + \mathbb{H}(B) = \mathbb{H}(A, B) = S_{n,\delta},$$

as Q is a product distribution. Without loss of generality, assume

$$\mathbb{H}(A) \ge \frac{S_{n,\delta}}{2}.$$

For any $u \in [n]$, (A, B) can be equivalently represented as

$$(A(u), A_{(u,A(u))}, B_{(u,A(u))}, \beta_{(u,A(u))}(B)),$$

where $\beta_{(u,A(u))}(B) \in [|\{b' \in \mathcal{B} \mid b'_{(u,A(u))} = B_{(u,A(u))}\}|]$ is a unique label among all $b' \in \mathcal{B}$ such that $b'_{(u,A(u))} = B_{(u,A(u))}$. Also define[14]

$$\delta_{(u,v)} = \Pr\big(\mathsf{bConn}_{N-2}(A_{(u,v)}, B_{(u,v)}) = 0 \mid A(u) = v\big).$$

It must hold that

$$\mathop{\mathbb{E}}_{v\sim A(u)} \delta_{(u,v)} \le \delta,$$

as taking restriction always preserves connectivity. Also observe that $A_{(u,v)}$ and $B_{(u,v)}$ remain independent conditioned on any $v = A(u)$. Consequently, we can get

$$\begin{aligned}
\mathbb{H}(A, B) &= \mathbb{H}\big(A(u), A_{(u,A(u))}, B_{(u,A(u))}, \beta_{(u,A(u))}(B)\big) \\
&= \mathbb{H}(A(u)) + \mathbb{H}\big(A_{(u,A(u))}, B_{(u,A(u))} \mid A(u)\big) + \mathbb{H}\big(\beta_{(u,A(u))}(B) \mid A, B_{(u,A(u))}\big) \\
&\qquad\qquad \text{(by chain rule of entropy (Fact A.1))} \\
&\le \mathbb{H}(A(u)) + \mathop{\mathbb{E}}_{v\sim A(u)} \mathbb{H}\big(A_{(u,v)}, B_{(u,v)} \mid A(u) = v\big) + \mathbb{H}\big(B \mid A(u), B_{(u,A(u))}\big)
\end{aligned}$$

[14] Technically, $A_{(u,v)}, B_{(u,v)}$ map from $[n]\setminus\{u\}$ to $[n]\setminus\{v\}$. Renaming both domain and codomain to $[n-1]$ makes them valid input to bConn_{N-2}.

$$
\begin{aligned}
&\le \mathbb{H}(A(u)) + \mathop{\mathbb{E}}_{v\sim A(u)} S_{n-1,\delta_{(u,v)}} + \mathbb{H}\big(B \mid A(u), B_{(u,A(u))}\big) \\
&\le \mathbb{H}(A(u)) + \mathop{\mathbb{E}}_{v\sim A(u)} S'_{n-1,\delta_{(u,v)}} + \mathbb{H}\big(B \mid A(u), B_{(u,A(u))}\big) \quad \text{(by inductive hypothesis)} \\
&\le \log n + S'_{n-1,\delta} + \mathbb{H}(B \mid A(u), B_{(u,A(u))}), \qquad \text{(by Jensen's inequality)}
\end{aligned}
$$

where we use the fact that $S'_{n-1,\delta}$ is both concave[15] and nondecreasing in δ. By Lemma 5.3 and averaging over $u \sim [n]$, it implies that

$$
\mathbb{H}(A,B) \le \log n + S'_{n-1,\delta} + \log n - \frac{\mathbb{H}(A)}{2n} + \mathcal{O}\Big(\log\log n + \sqrt{\delta} n^3 \log n\Big).
$$

Since $\mathbb{H}(A) \ge S_{n,\delta}/2$ by assumption, rearranging the terms yields

$$
\left(1 + \frac{1}{4n}\right) \cdot S_{n,\delta} \le S'_{n-1,\delta} + 2\log n + \mathcal{O}\Big(\log\log n + \sqrt{\delta} n^3 \log n\Big),
$$

concluding the proof by picking a sufficiently large constant K. □

6 Lower Bound for Connectivity with Constant Error

We prove round-dependent randomized lower bounds for connectivity with constant error in this section. As will be seen in Theorem 6.2, the reduction can be easily adapted to prove similar lower bounds for deciding other fundamental graph properties.

Theorem 6.1. *For even $N > 1$, and $r = o(\log N)$, any r-round randomized protocol for bConn_N with constant error probability less than $1/2$ requires at least $\mathbf{R}^{(r)}(\mathsf{bConn}_N) = \Omega(N/r \cdot \log N)$ communication.*

Theorem 6.2. *For $N \ge 1$, and $r = o(\log N)$, any r-round randomized protocol requires at least $\Omega(N/r \cdot \log N)$ communication to decide with constant error probability less than $1/2$ if an N-vertex graph is (1) bipartite; (2) cycle-free; (3) Eulerian; (4) Hamiltonian; (5) planar.*

Another direct consequence of Theorem 6.1 is a multi-pass streaming lower bound for connectivity, as shown in Corollary 6.3. Since global minimum cuts can only be harder, it also fully resolves the optimality of the two-pass algorithm in [AD21], even if any constant number of passes is allowed.

Corollary 6.3. *For $N \ge 1$, and $p = o(\log N)$, any p-pass randomized streaming algorithm that decides connectivity in N-vertex graphs with constant error probability less than $1/2$, requires at least $\Omega(N/p^2 \cdot \log N)$ space.*

Proof. By standard reductions, any p-pass S-space streaming algorithm can be simulated by a $(2p-1)$-round communication protocol with $O(pS)$ communication. Consequently, the corollary follows from Theorem 6.1. □

[15] This is exactly the reason for introducing $S'_{n,\delta}$. Concavity does not necessarily hold for $S_{n,\delta}$ as defined.

Remark 6.4. *We remark that the 2-round $\Omega(N \log N)$ communication lower bound for connectivity in [RRV26] alone does not automatically imply optimality of [AD21], as the standard simulation of 2-pass streaming algorithms uses $2 \cdot 2 - 1 = 3$ communication rounds.*

Proofs of Theorems 6.1 and 6.2 utilize the following communication lower bound for $\mathsf{IP} \circ \mathsf{PC}$, which will be proved in Section 7.

Theorem 7.1. *For $n \geq 1$, and $r = o(\log n)$, any r-round protocol for $\mathsf{IP} \circ \mathsf{PC}_{n,r}$ with constant error probability less than $1/2$ requires at least $\mathbf{R}^{(r)}(\mathsf{IP} \circ \mathsf{PC}_{n,r}) = \Omega(n \log n)$ communication.*

In Section 6.1, we present the reduction from $\mathsf{IP} \circ \mathsf{PC}$. Theorem 6.2 is shown in Section 6.2.

6.1 Proof of Theorem 6.1

This section constitutes the proof of Theorem 6.1. Fix sufficiently large even N, and $r = o(\log N)$. We first prove the lower bound for Conn_N.

At a high level, the input permutations A_i, B_i of $\mathsf{IP} \circ \mathsf{PC}_{n,r}$ are encoded by perfect matchings between consecutive layers in the constructed graph instance. The two bit vectors X, Y decide the structure of interleaved special layers so that the number of cycles is determined by the output of $\mathsf{IP} \circ \mathsf{PC}_{n,r}$. Similar approaches have been taken to provide reductions from other pointer-chasing-type communication problems; see for example [AN21, RRV26].

The graph instance $G = (V, E)$ is constructed as follows.

Vertices. V consists of $4\ell + 1$ layers $V_0, \ldots, V_{4\ell}$, each with $4n$ vertices, where $\ell = 2r + 1$. The total number of vertices is $N = (4\ell + 1) \cdot 4n = \Theta(rn)$. Each vertex is indexed by $(i, j, k) \in [0, 4\ell] \times [n] \times [4]$, so vertices in the same layer are partitioned into groups of size 4.

Edges. E consists of different types of edges. All edges are between two consecutive layers, or $V_{4\ell}$ and V_0. The type of an edge is determined solely by the layers that it connects. All edges connecting the same two layers are owned by the same party.

First consider *permutation edges*. For $i \in [r]$ and $j \in [n]$, add the identity matching (i.e., vertices with the same last index are connected) between each of the following groups:

1. $(i, j, [4])$ and $(i + 1, A_i(j), [4])$;
2. $(r + i, j, [4])$ and $(r + i + 1, B_{r-i+1}^{-1}(j), [4])$;
3. $(\ell + i, j, [4])$ and $(\ell + i + 1, B_i(j), [4])$;
4. $(\ell + r + i, j, [4])$ and $(\ell + r + i + 1, A_{r-i+1}^{-1}(j), [4])$;
5. $(2\ell + i, j, [4])$ and $(2\ell + i + 1, A_i(j), [4])$;
6. $(2\ell + r + i, j, [4])$ and $(2\ell + r + i + 1, B_{r-i+1}^{-1}(j), [4])$;
7. $(3\ell + i, j, [4])$ and $(3\ell + i + 1, B_i(j), [4])$;
8. $(3\ell + r + i, j, [4])$ and $(3\ell + r + i + 1, A_{r-i+1}^{-1}(j), [4])$.

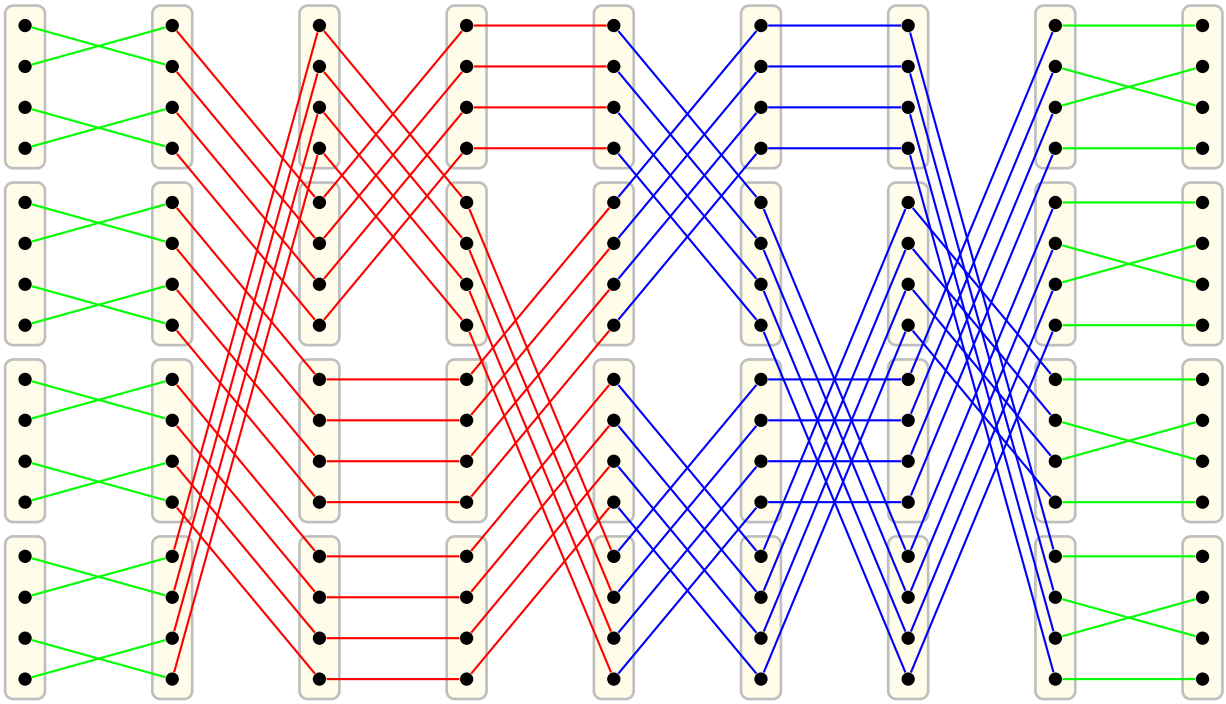


Figure 2: An illustration of the first $\ell+1$ layers of permutation and product edges in G, with $n = 4, r = 3$ and thus $\ell = 7$. Each yellow group represents a group of 4 vertices in the same layer. Red edges correspond to permutations $A_1 = (1\ 2\ 3\ 4)$, $A_2 = (1\ 2)$, $A_3 = (2\ 4\ 3)$. Blue edges correspond to inverses of permutations $B_1 = (1\ 3\ 2\ 4)$, $B_2 = (2\ 4)$, $B_3 = (1\ 2)(3\ 4)$ in reverse order. Green edges are product edges.

Essentially, the 4ℓ layers of edges are equally partitioned into 4 blocks. In each block, all but the first layer of edges are permutation edges, with each layer corresponding to one of the $2r$ permutations. The first and the third blocks encode the permutations in the order of $A_1, \ldots, A_r, B_r^{-1}, \ldots, B_1^{-1}$, while the second and the fourth blocks encode the inverse permutations in the reverse order, namely $B_1, \ldots, B_r, A_r^{-1}, \ldots, A_1^{-1}$. Therefore, while preserving the last index, the net effect of permutation edges on the middle index is simply permutations C, C^{-1}, C, C^{-1} in the four blocks, respectively, where

$$C = (B_r \cdots B_1)^{-1} A_r \cdots A_1. \tag{10}$$

The first layer in each block is reserved for *product edges*. Specifically, let $\sigma_{\mathsf{X}} = (1\ 2)(3\ 4)$ and $\sigma_{\mathsf{Y}} = (2\ 3)$ be 4-permutations. For $j \in [n]$, add the following product edges:

1. Edges corresponding to $(\sigma_{\mathsf{X}})^{X_j}$ between $(0, j, [4])$ and $(1, j, [4])$;
2. Edges corresponding to $(\sigma_{\mathsf{Y}})^{Y_j}$ between $(\ell, j, [4])$ and $(\ell + 1, j, [4])$;
3. Edges corresponding to $(\sigma_{\mathsf{X}})^{-X_j} = (\sigma_{\mathsf{X}})^{X_j}$ between $(2\ell, j, [4])$ and $(2\ell + 1, j, [4])$;
4. Edges corresponding to $(\sigma_{\mathsf{Y}})^{-Y_j} = (\sigma_{\mathsf{Y}})^{Y_j}$ between $(3\ell, j, [4])$ and $(3\ell + 1, j, [4])$.

Observe that product edges always preserve the middle index. More important, for any bits $x, y \in \{0, 1\}$, the net effect of product edges on the last index is exactly

$$(\sigma_{\mathsf{Y}})^{-y} (\sigma_{\mathsf{X}})^{-x} (\sigma_{\mathsf{Y}})^{y} (\sigma_{\mathsf{X}})^{x} = \tau^{x \cdot y}, \tag{11}$$

where $\tau = (1\ 4)(2\ 3)$. Indeed, if either x or y is 0, then it trivially evaluates to the identity permutation. Otherwise, $\sigma_{\mathsf{Y}} \sigma_{\mathsf{X}} \sigma_{\mathsf{Y}} \sigma_{\mathsf{X}} = \tau$. Note that τ^2 is simply the identity permutation.

Lastly, there are also *fixed edges*. For now, the fixed edges only connect $V_{4\ell}$ and V_0. We will insert a few more layers of fixed edges for bConn_N. Concretely, for $j \in [n-1]$, add the identity matching between $(4\ell, j, [4])$ and $(0, j+1, [4])$. In the meantime, add the edges corresponding to the cyclic 4-permutation $\pi = (1\ 2\ 3\ 4)$ between $(4\ell, n, [4])$ and $(0, 1, [4])$.

This completes the construction of G. See Figure 2 for an illustration of (part of) permutation and product edges. Note that it only depicts the first block of permutation edges and two adjacent layers of product edges. It does not contain the fixed edges.

Lemma 6.5 provides an exact characterization of the structure of G.

Lemma 6.5. *G consists of a single N-cycle if $\langle A_r \cdots A_1 X, B_r \cdots B_1 Y \rangle = 0$, and of three disjoint cycles of length $N/2, N/4, N/4$ if $\langle A_r \cdots A_1 X, B_r \cdots B_1 Y \rangle = 1$.*

Proof. The observations above, especially Equations (10) and (11), establish that every vertex $(0, j, k) \in V_0$ is connected via a unique and disjoint path to

1. $(\ell, C(j), (\sigma_{\mathsf{X}})^{X_j}(k)) \in V_\ell$ by the first block;
2. $(2\ell, j, (\sigma_{\mathsf{Y}})^{Y_{C(j)}} (\sigma_{\mathsf{X}})^{X_j}(k)) \in V_{2\ell}$ by the second block;
3. $(3\ell, C(j), (\sigma_{\mathsf{X}})^{-X_j} (\sigma_{\mathsf{Y}})^{Y_{C(j)}} (\sigma_{\mathsf{X}})^{X_j}(k)) \in V_{3\ell}$ by the third block;
4. $(4\ell, j, (\sigma_{\mathsf{Y}})^{-Y_{C(j)}} (\sigma_{\mathsf{X}})^{-X_j} (\sigma_{\mathsf{Y}})^{Y_{C(j)}} (\sigma_{\mathsf{X}})^{X_j}(k)) \in V_{4\ell}$ by the fourth block.

By Equation (11), the net effect is exactly $\tau^{X_j \cdot Y_{C(j)}}$ on the last index.

Furthermore, observe that the identity matchings added as fixed edges essentially compose the net effect for every $j \in [n]$ in sequence. In more detail, excluding the fixed edges corresponding to π, G is the union of four disjoint paths, each with $N/4$ vertices, from $(0, 1, k)$ to $(4\ell, n, \tau^O(k))$ for $k \in [4]$, where

$$O = \sum_{j \in [n]} X_j \cdot Y_{C(j)} = \langle A_r \cdots A_1 X, B_r \cdots B_1 Y \rangle \pmod 2.$$

To see the last step, note that for $j' \in [n]$, the j'-th bit of $A_r \cdots A_1 X$ is the $(A_r \cdots A_1)^{-1}(j')$-th bit of X, while the j'-th bit of $B_r \cdots B_1 Y$ is the $(B_r \cdots B_1)^{-1}(j')$-th bit of Y. Plugging in $j = (A_r \cdots A_1)^{-1}(j')$ yields the desired equality by Equation (10).

Finally, the proof is completed by observing $\pi\tau = (2\ 4)$. □

Now, we are ready to prove Theorem 6.1.

Proof of Theorem 6.1. Given any $\mathsf{IP} \circ \mathsf{PC}_{n,r}$ input $(X, Y, A_{\le r}, B_{\le r})$, all edges of the graph instance G can be constructed by either Alice or Bob. Indeed, permutation edges corresponding to A_i, B_i (or A_i^{-1}, B_i^{-1}) are owned by Alice for odd i and are owned by Bob for even i. Bob also owns both product edges and fixed edges. Therefore, Lemma 6.5 implies that any protocol for Conn_N also solves $\mathsf{IP} \circ \mathsf{PC}_{n,r}$ with the same number of rounds, error probability, and communication. The lower bound for Conn_N then follows from the lower bound for $\mathsf{IP} \circ \mathsf{PC}_{n,r}$ (Theorem 7.1) as $n = \Theta(N/r)$ and thus $r = o(\log n)$.

For bConn_N, we additionally insert a constant number of layers of fixed edges so that the graph instance G is bipartite and that Alice and Bob own alternating layers of edges. Concretely, the current construction already has alternating ownership except that:

1. Layers of edges corresponding to A_r, B_r^{-1} (or B_r, A_r^{-1}) are owned by the same party. To fix this, simply insert an additional layer of fixed edges in between, corresponding to the identity permutation, that is owned by the opposite party.

2. Bob owns both edges between V_0, V_1 and the ones between $V_{4\ell}, V_0$. This can also be fixed by inserting an additional layer of fixed edges after V_0, corresponding to the identity permutation, that is owned by Alice.

Then, the modified graph has a balanced bipartition such that both Alice and Bob own a perfect matching across the bipartition. This concludes the proof[16] for bConn_N since the number of layers increases only by a constant (exactly $4 + 1 = 5$, to be precise). □

6.2 Lower Bounds for Other Graph Properties

Theorem 6.2 can be shown via standard modifications to the reduction in Section 6.1.

[16] Technically, we only construct G for N of the form $(4\ell + 1 + 5) \cdot 4n$. For any other even N, simply let $n = \lfloor N/(16\ell + 24) \rfloor$. Splitting the fixed edge $((4\ell, n, 1), (0, 1, 2))$ into a path of length $N - (16\ell + 24) \cdot n + 1$ ensures exactly N vertices. Assign the subdivided edges to the two parties alternately. No analysis is affected.

Proof of Theorem 6.2. The proof of Theorem 6.1 actually constructs a bipartite graph G consisting of either a single even cycle or three disjoint even cycles such that the two fixed edges $((4\ell, n, 4), (0, 1, 1))$, $((4\ell, n, 2), (0, 1, 3))$ are contained in different cycles.

The lower bound for bipartiteness follows by subdividing each edge above into a path of length 2. Indeed, if there is a single even cycle, G remains bipartite, as the length of the cycle increases by 2. Otherwise, G contains two odd cycles and is not bipartite anymore.

For cycle-freeness, remove the above edges. In the case of a single cycle, removing the above edges breaks the cycle. Otherwise, G still contains one cycle.

A graph is Eulerian if and only if it is connected and every vertex has an even degree. Since every vertex has degree exactly 2 in G, it is Eulerian if and only if it is connected.

Similarly, G is Hamiltonian if and only if it is connected.

As to planarity, add three new vertices u_1, u_2, u_3, and construct a 5-clique among u_1, u_2, u_3 and $(0, 1, 1), (0, 1, 3)$, except for the edge between $(0, 1, 1)$ and $(0, 1, 3)$. If G is a single cycle, then the cycle completes a K_5 minor and thus the resulting graph is not planar. Otherwise, since $(0, 1, 1)$ and $(0, 1, 3)$ are in two disconnected cycles, the resulting graph remains planar.

This concludes the proof as the lower bound for each of the above graph properties follows from a similar reduction from $\mathsf{IP}\circ\mathsf{PC}_{n,r}$. □

7 Lower Bound for IP∘PC

In this section, we prove the communication lower bound for $\mathsf{IP}\circ\mathsf{PC}$, as promised in Section 6.

Theorem 7.1. *For $n \geq 1$, and $r = o(\log n)$, any r-round protocol for $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ with constant error probability less than $1/2$ requires at least $\mathbf{R}^{(r)}(\mathsf{IP}\circ\mathsf{PC}_{n,r}) = \Omega(n \log n)$ communication.*

Remark 7.2. *The requirement of $r = o(\log n)$ can actually be relaxed to many more rounds. We remark that Theorem 7.1 already suffices to prove the lower bound for connectivity in Section 6, and larger r does not help obtain a stronger lower bound. Nevertheless, we extend the lower bound in Appendix C, as it may be of independent interest.*

The rest of this section constitutes the proof of Theorem 7.1. We first prove a weaker version of Theorem 7.1 in Section 7.1, showing that $\mathbf{R}^{(r)}(\mathsf{IP}\circ\mathsf{PC}_{n,r}) = \Omega(n/r \cdot \log n)$. It already demonstrates many of the crucial ideas for the full proof. As to graph connectivity, by the same reduction as in Section 6, the weaker version is sufficient for proving the optimal lower bound of $\mathbf{R}^{(r)}(\mathsf{Conn}_N) = \Omega(N \log N)$ for constant rounds. However, it yields a worse $1/r^2$ dependency on r for superconstant rounds, as compared to Theorem 6.1.

In Section 7.2, we give a tighter analysis of how much randomness in the input permutations can actually be lost during the protocol. This completes the full proof of Theorem 7.1.

7.1 A Weaker Lower Bound

Fix sufficiently large n, and $r = o(\log n)$. Suppose that there exists an r-round protocol with constant error probability $\varepsilon < 1/2$ and communication $L = o(n/r \cdot \log n)$. We can actually assume $\varepsilon \leq 1/3$ by repetition. We further assume that the protocol is deterministic and has error probability ε on uniformly random input.

The high level plan is to show that the pair of initially independent and uniformly random bit vectors (X, Y) retains almost full entropy through the protocol, and eventually, after taking inner product, leads to one random output bit with hardly any bias. More specifically, let $(W_0, Z_0) = (X, Y)$ and recursively define $(W_i, Z_i) = (A_i W_{i-1}, B_i Z_{i-1})$ for $i \in [r]$. In round i, the i-th pair of permutations (A_i, B_i) should still remain close to random because the total communication is limited. Then, intuitively, (W_i, Z_i) should also have enough randomness, provided that (W_{i-1}, Z_{i-1}) is sufficiently random to start with.

Lemma 7.3 formalizes the above intuition in the ideal case where (W_{i-1}, Z_{i-1}) were uniformly random. With Q being the current distribution of (A_i, B_i), it essentially upper bounds the "distance" of (W_i, Z_i) from being uniformly random, in terms of the "distance" of (A_i, B_i) from being uniformly random (although the "distances" are measured by Rényi divergence of different orders).

Lemma 7.3. *For any distribution Q on $\mathcal{P}_n \times \mathcal{P}_n$, it holds that*

$$\mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} 2^{\mathbb{D}_2(Q^{(w,z)})} \le 2^{\frac{\mathbb{D}_\infty(Q)+\mathcal{O}(n)}{\log n}},$$

where $Q^{(w,z)}$ is the distribution of $(aw, bz) \in \{0,1\}^{2n}$ when $(a, b) \sim Q$.

The proof of Lemma 7.3 uses the following technical lemma.

Claim 7.4. *For $m \ge 1$, permutations $\sigma_1, \ldots, \sigma_m \in \mathcal{P}_n$, and any distribution Q on $(\mathcal{P}_n)^m$, it holds that*

$$\mathop{\mathbb{E}}_{(\sigma'_1,\ldots,\sigma'_m)\sim Q} 2^{\sum_{i\in[m]} \mathsf{cyc}(\sigma_i^{-1}\sigma'_i)} \le 2^{\frac{\mathbb{D}_\infty(Q)+\mathcal{O}(mn)}{\log n}}.$$

Proof. Observe that σ'_i can always be decomposed as $\sigma'_i = \sigma_i \sigma''_i$ for a unique $\sigma''_i \in \mathcal{P}_n$. It then holds that $\sigma_i^{-1}\sigma'_i = \sigma''_i$. Moreover, σ''_i jointly follow the same distribution Q, up to renaming induced by σ_i. Thus, we can assume without loss of generality that all σ_i are simply the identity permutation.

Let $t = (\mathbb{D}_\infty(Q) + Kmn)/\log n$, for some sufficiently large constant K. The contribution to the expectation is at most 2^t, for any $\sigma'_1, \ldots, \sigma'_m \in \mathcal{P}_n$ such that $\sum_{i\in[m]} \mathsf{cyc}(\sigma'_i) \le t$. In the meantime, we claim that the total number of possible combinations of $\sigma'_1, \ldots, \sigma'_m \in \mathcal{P}_n$ such that $\sum_{i\in[m]} \mathsf{cyc}(\sigma'_i) > t$ is at most $2^{mn} n^{mn-t}$. Assume the claim for now. Note that $\sum_{i\in[m]} \mathsf{cyc}(\sigma'_i) \le mn$ always holds by definition. As a result, by definition of min-entropy, the total contribution in this case is at most

$$2^{mn} \cdot 2^{mn} n^{mn-t} \cdot 2^{-\mathbb{H}_\infty(Q)} = 2^{2mn+mn\log n - t\log n - m\log(n!) + \mathbb{D}_\infty(Q)} = o(1),$$

where we use Stirling's approximation. The proof concludes once we show the claim.

To see the claim, consider constructing the unique cycle decomposition representation of σ'_i by writing its elements in increasing order. If the current element $j \in [n]$ will be the smallest one in the cycle containing it, then it can only be appended to the end of the current partial representation, as a 1-cycle for now. Otherwise, it can be inserted into existing cycles by being appended to one of the $j-1$ previous elements. Overall, there are at most $mn - \sum_{i\in[m]} \mathsf{cyc}(\sigma'_i) < mn - t$ such elements in total across all σ'_i, each with at most n options, whereas all other elements have only a single option. This concludes the proof as there are no more than 2^{mn} ways to pick such elements. □

Now, we are ready to prove Lemma 7.3.

Proof of Lemma 7.3. By definition of collision entropy,

$$\begin{aligned}\mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} 2^{\mathbb{D}_2(Q^{(w,z)})} &= 2^{2n}\cdot \mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} 2^{-\mathbb{H}_2(Q^{(w,z)})} \\ &= 2^{2n}\cdot \mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n}\ \mathop{\mathbb{E}}_{(u,v),(u',v')\sim Q^{(w,z)}} \mathbb{1}((u,v)=(u',v')) \\ &= 2^{2n}\cdot \mathop{\mathbb{E}}_{(a,b),(a',b')\sim Q}\ \mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} \mathbb{1}((aw,bz)=(a'w,b'z)).\end{aligned}$$

Observe that the condition $aw = a'w$ is equivalent to $a^{-1}a'w = w$. Since $a^{-1}a'$ is also a permutation, this happens if and only if w is constant when restricted to every cycle in the decomposition of $a^{-1}a'$. For uniformly random w, such an event occurs with probability exactly $2^{\mathsf{cyc}(a^{-1}a')-n}$. A similar argument applies to $bz = b'z$. As the two parts are independent for fixed $(a,b),(a',b')$, we can get

$$\mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} \mathbb{1}((aw,bz)=(a'w,b'z)) = 2^{\mathsf{cyc}(a^{-1}a')+\mathsf{cyc}(b^{-1}b')-2n}.$$

Continuing, we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} 2^{\mathbb{D}_2(Q^{(w,z)})} &= \mathop{\mathbb{E}}_{(a,b),(a',b')\sim Q} 2^{\mathsf{cyc}(a^{-1}a')+\mathsf{cyc}(b^{-1}b')} \\ &\le \max_{a,b\in\mathcal{P}_n} \mathop{\mathbb{E}}_{(a',b')\sim Q} 2^{\mathsf{cyc}(a^{-1}a')+\mathsf{cyc}(b^{-1}b')}.\end{aligned}$$

The lemma then follows from Claim 7.4 for $m = 2$. □

In reality, we can only expect (W_{i-1}, Z_{i-1}) to be close to random. To this end, for $i \in [r]$, let M_i denote the i-th message. Define[17] $L_i = \sum_{j\in[i]}|M_j|$ and $\Phi_i = \sum_{j\in[i]}(L_j + Kn)/\log n$, for some sufficiently large constant K. Also let $\Gamma_i = (X, Y, A_{<i}, B_{<i}, M_{\le i})$ denote the partial transcript up to round i, together with the input variables that we are going to additionally reveal. We emphasize that Γ_i is only for analysis purposes and is not revealed to the parties during execution of the protocol. Also note that it includes the i-th message, but not the i-th pair of permutations. In particular, (w_{i-1}, z_{i-1}) is fixed once $\Gamma_i = \gamma_i$ is revealed, whereas (W_i, Z_i) is not. For convenience, define[18] $L_0 = \Phi_0 = 0$ and $\Gamma_0 = \varnothing$.

Given[19] $(\Gamma_{i-1}, W_{i-1}, Z_{i-1}, M_i) = (\gamma_{i-1}, w_{i-1}, z_{i-1}, m_i)$, first define $\mathcal{E}_i$ to be the event that $\mathbb{D}_2(W_i, Z_i \mid \gamma_{i-1}, w_{i-1}, z_{i-1}, m_i) > \Phi_i$. For convenience, define $\mathcal{E}_0 = \varnothing$ to be a never happening event. We want to show that $\mathcal{E}_i$ rarely happens. That is, $(W_i, Z_i) = (A_i w_{i-1}, B_i z_{i-1})$ should still be sufficiently random.

In order to prove this, also define $\mathcal{E}'_i$ to be the event that $\mathbb{D}_\infty(A_i, B_i \mid \gamma_{i-1}, m_i) > L_i + n$. We first show in Lemma 7.5 that with extremely high probability, a typical partial transcript of total length L_i should not cause $\mathcal{E}'_i$ to happen.

[17] Assume each message is of fixed length. The full proof in Section 7.2 does not assume this. Specifically, any prefix-free message of varying length is allowed so long as the total message length is bounded.

[18] Γ_0 excludes (X, Y) due to the definition of $\mathcal{E}_1, \mathcal{E}'_1$ below. This will be useful in later proofs.

[19] Note that Γ_{i-1} excludes A_{i-1}, B_{i-1} (or X, Y for $i = 1$), so it does not fix (W_{i-1}, Z_{i-1}).

Lemma 7.5. *For $i \in [r]$, it holds that*

$$\Pr(\mathcal{E}'_i) \leq 2^{-n}.$$

Proof. We prove the lemma in the case of odd i. The case of even i is similar.

At the beginning of the protocol, all permutations are independent and uniformly random. In particular, revealing $A_{<(i-1)}, B_{<(i-1)}$ (and X, Y for $i > 1$) does not change the fact that Alice's remaining input permutations, namely $A_{i'}, B_{i'}$ for odd $i' \geq i$, are still independent and uniformly random. Let $t = \lceil r/2 \rceil - \lceil (i-1)/2 \rceil$ be the number of remaining pairs of input permutations for Alice. Because of the rectangle property of deterministic communication protocols, each message M_j for $j \in [i]$ either belongs to Alice, subdividing Alice's possible input into no more than $2^{|M_j|}$ parts, or belongs to Bob, keeping Alice's possible input the same. Overall, the average support size of $\{(A_{i'}, B_{i'})\}_{\text{odd } i' \geq i}$ after conditioning is at least $S = (n!)^{2t} 2^{-L_i}$. Moreover, the support size is smaller than $S \cdot 2^{-n}$, with probability at most 2^{-n} over $(\gamma_{i-1}, m_i) \sim (\Gamma_{i-1}, M_i)$. To see this, the total support size of $\{(A_{i'}, B_{i'})\}_{\text{odd } i' \geq i}$ for all such (γ_{i-1}, m_i) is at most

$$S \cdot 2^{-n} \cdot 2^{L_i} = (n!)^{2t} \cdot 2^{-n}.$$

By the rectangle property, this leads to the 2^{-n} probability bound because Alice always holds a uniform distribution on the remaining support of $\{(A_{i'}, B_{i'})\}_{\text{odd } i' \geq i}$.

Fix any $(\Gamma_{i-1}, M_i) = (\gamma_{i-1}, m_i)$ such that

$$|\mathsf{supp}(\{(A_{i'}, B_{i'})\}_{\text{odd } i' \geq i} \mid \gamma_{i-1}, m_i)| \geq (n!)^{2t} 2^{-L_i - n}.$$

For any fixed $(A_i, B_i) = (a_i, b_i)$, the total number of compatible $\{(A_{i'}, B_{i'})\}_{\text{odd } i' \geq i}$ is at most $(n!)^{2(t-1)}$. Due to uniformity, we can upper bound its marginal probability as follows:

$$\Pr((A_i, B_i) = (a_i, b_i) \mid \gamma_{i-1}, m_i) \leq \frac{(n!)^{2(t-1)}}{(n!)^{2t} 2^{-L_i - n}} = \frac{2^{L_i + n}}{(n!)^2}.$$

Consequently, by definition of min-entropy,

$$\mathbb{D}_\infty(A_i, B_i \mid \gamma_{i-1}, m_i) = 2\log(n!) - \mathbb{H}_\infty(A_i, B_i \mid \gamma_{i-1}, m_i) \leq 2\log(n!) + \log \frac{2^{L_i + n}}{(n!)^2} = L_i + n,$$

or equivalently, $\mathcal{E}'_i$ does not happen. This concludes the proof as such (γ_{i-1}, m_i) occurs with probability at least $1 - 2^{-n}$, as shown above. □

Next, we upper bound the probability of event $\mathcal{E}_i$ in Lemma 7.6.

Lemma 7.6. *For $i \in [r]$, it holds that*

$$\Pr(\mathcal{E}_i \mid \overline{\mathcal{E}}_{i-1}, \overline{\mathcal{E}}'_i) \leq 2^{-\frac{n}{\log n}},$$

where $\overline{\mathcal{E}}_{i-1}, \overline{\mathcal{E}}'_i$ are the complement events of $\mathcal{E}_{i-1}, \mathcal{E}'_i$, respectively.

Proof. First consider $i = 1$. $\overline{\mathcal{E}}_0$ trivially happens by definition. The rectangle property of deterministic communication protocols implies that $(W_0, Z_0) = (X, Y)$ is uniformly random even conditioned on any $M_1 = m_1$. This is because M_1 depends only on Alice's input while X, Y are owned by Bob. Therefore, for any m_1 satisfying $\overline{\mathcal{E}}'_1$ (and $\gamma_0 = \varnothing$), Lemma 7.3 implies that

$$\mathop{\mathbb{E}}_{(w_0,z_0)|m_1} 2^{\mathbb{D}_2(W_1,Z_1|w_0,z_0,m_1)} = \mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} 2^{\mathbb{D}_2((A_1,B_1)^{(w,z)}|m_1)} \le 2^{\frac{L_1+Kn/2}{\log n}},$$

for sufficiently large constant K. By Markov's inequality, we can get

$$\Pr_{(w_0,z_0)|m_1}(\mathcal{E}_1) \le 2^{\frac{L_1+Kn/2}{\log n}} \cdot 2^{-\Phi_1} = 2^{-\frac{Kn/2}{\log n}} \le 2^{-\frac{n}{\log n}}.$$

Averaging over m_1 satisfying $\overline{\mathcal{E}}'_1$ yields the desired lemma statement.

For $i > 1$, $\overline{\mathcal{E}}'_i$ implies that

$$\mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} 2^{\mathbb{D}_2((A_i,B_i)^{(w,z)}|\gamma_{i-1},m_i)} \le 2^{\frac{L_i+Kn/2}{\log n}},$$

for sufficiently large constant K. By Markov's inequality, the total number of possible combinations of $w, z \in \{0,1\}^n$ such that $\mathbb{D}_2((A_i, B_i)^{(w,z)} \mid \gamma_{i-1}, m_i) > \Phi_i$ is at most

$$2^{2n} \cdot 2^{\frac{L_i+Kn/2}{\log n}} \cdot 2^{-\Phi_i} = 2^{2n-\Phi_{i-1}-\frac{Kn/2}{\log n}}.$$

At the same time, $\overline{\mathcal{E}}_{i-1}$ means that

$$\mathbb{D}_2(W_{i-1}, Z_{i-1} \mid \gamma_{i-2}, w_{i-2}, z_{i-2}, m_{i-1}) \le \Phi_{i-1},$$

or equivalently,

$$\mathbb{H}_2(W_{i-1}, Z_{i-1} \mid \gamma_{i-2}, w_{i-2}, z_{i-2}, m_{i-1}) \ge 2n - \Phi_{i-1}.$$

In fact, it further implies that

$$\mathbb{H}_2(W_{i-1}, Z_{i-1} \mid \gamma_{i-1}, m_i) \ge 2n - \Phi_{i-1}.$$

To see this, for even i, after revealing $(\Gamma_{i-2}, W_{i-2}, Z_{i-2}, M_{i-1})$, both A_{i-2}, B_{i-2} (or X, Y for $i = 2$) and M_i are functions of Bob's input only, while (W_{i-1}, Z_{i-1}) depends only on Alice's input. As a result, additional conditioning does not change its distribution, due to the rectangle property of deterministic communication protocols. A similar argument applies to odd i. Furthermore, the distribution of (A_i, B_i) is conditionally independent of (W_{i-1}, Z_{i-1}), again because of the rectangle property. Altogether, for any (γ_{i-1}, m_i) satisfying both $\overline{\mathcal{E}}_{i-1}$ and $\overline{\mathcal{E}}'_i$, we have that

$$\begin{aligned}
&\Pr_{(w_{i-1},z_{i-1})|(\gamma_{i-1},m_i)}(\mathcal{E}_i) \\
&\quad= \mathop{\mathbb{E}}_{(w_{i-1},z_{i-1})|(\gamma_{i-1},m_i)} \mathbb{1}\left(\mathbb{D}_2\left((A_i, B_i)^{(w_{i-1},z_{i-1})} \mid \gamma_{i-1}, m_i\right) > \Phi_i\right) \\
&\quad= \sum_{w,z\in\{0,1\}^n} \Pr((W_{i-1}, Z_{i-1}) = (w, z) \mid \gamma_{i-1}, m_i) \cdot \mathbb{1}\left(\mathbb{D}_2\left((A_i, B_i)^{(w,z)} \mid \gamma_{i-1}, m_i\right) > \Phi_i\right)
\end{aligned}$$

$$\leq \left(2^{-\mathbb{H}_2(W_{i-1},Z_{i-1}|\gamma_{i-1},m_i)} \cdot 2^{2n-\Phi_{i-1}-\frac{Kn/2}{\log n}}\right)^{\frac{1}{2}}$$
$$\leq 2^{-\frac{Kn/4}{\log n}}$$
$$\leq 2^{-\frac{n}{\log n}},$$

by Cauchy-Schwarz inequality and definition of collision entropy. The lemma then follows by averaging over (γ_{i-1}, m_i) satisfying both $\overline{\mathcal{E}}_{i-1}$ and $\overline{\mathcal{E}}'_i$. □

In the last round of the protocol, after revealing (Γ_{r-1}, M_r), which contains all messages and input permutations except (A_{r-1}, B_{r-1}) and (A_r, B_r), we can expect (W_r, Z_r) to be still close to random. In fact, the sender of M_r owns (A_r, B_r), uniformly distributed on its remaining support. Meanwhile, the receiver of M_r owns (A_{r-1}, B_{r-1}), also uniformly distributed on its remaining support. More importantly, (A_{r-1}, B_{r-1}) and (A_r, B_r) are conditionally independent, due to the rectangle property of deterministic communication protocols. Since the receiver has to output based solely on (A_{r-1}, B_{r-1}), it cannot predict $\langle W_r, Z_r\rangle$ well.

This is formalized in Lemma 7.7, where O denotes the output of the protocol.

Lemma 7.7. *It holds that*

$$\Pr(O = \langle W_r, Z_r\rangle \mid \overline{\mathcal{E}}_{r-1}, \overline{\mathcal{E}}'_r) \leq \frac{1}{2} + 2^{-\Omega(n)}.$$

Proof. First note that it suffices to upper bound the advantage

$$\Pr(O = \langle W_r, Z_r\rangle \mid \overline{\mathcal{E}}_{r-1}, \overline{\mathcal{E}}'_r) - \Pr(O \neq \langle W_r, Z_r\rangle \mid \overline{\mathcal{E}}_{r-1}, \overline{\mathcal{E}}'_r)$$

by $2^{-\Omega(n)}$. Fix any $(\Gamma_{r-1}, M_r) = (\gamma_{r-1}, m_r)$. We know that[20] (A_{r-1}, B_{r-1}) and (A_r, B_r) are independent due to the rectangle property, as discussed above. Since the receiver of $M_r = m_r$ has to output O based only on its own input and the full transcript, we also know that $O = O(A_{r-1}, B_{r-1})$ is now a function of (A_{r-1}, B_{r-1}) only and is independent of (A_r, B_r), while (W_r, Z_r) is fully determined by (A_r, B_r) and is independent of (A_{r-1}, B_{r-1}) conditioned on (W_{r-1}, Z_{r-1}). Thus, we can get

$$\begin{aligned}
&\Pr(O = \langle W_r, Z_r\rangle \mid \gamma_{r-1}, m_r) - \Pr(O \neq \langle W_r, Z_r\rangle \mid \gamma_{r-1}, m_r) \\
&\quad = \mathop{\mathbb{E}}_{(w_r,z_r,o)|(\gamma_{r-1},m_r)} (-1)^{o+\langle w_r, z_r\rangle} \\
&\quad = \sum_{w,z\in\{0,1\}^n} f(w,z)g(w,z) \\
&\quad \leq \left(\sum_{w,z\in\{0,1\}^n} f(w,z)^2\right)^{\frac{1}{2}} \left(\sum_{w,z\in\{0,1\}^n} g(w,z)^2\right)^{\frac{1}{2}}, \qquad (12)
\end{aligned}$$

by Cauchy-Schwarz inequality, where

$$f(w,z) = \mathop{\mathbb{E}}_{(a_{r-1},b_{r-1})|\gamma_{r-1},m_r} \mathbb{1}((W_{r-1}, Z_{r-1}) = (w,z)) \cdot (-1)^{O(a_{r-1},b_{r-1})},$$

[20] More precisely, (X, Y) should be used in place of (A_0, B_0) throughout the proof if $r = 1$.

and

$$g(w,z) = \mathop{\mathbb{E}}_{(w_r,z_r)\sim(A_r,B_r)^{(w,z)}|(\gamma_{r-1},m_r)} (-1)^{\langle w_r,z_r\rangle}.$$

Observe that[21]

$$\begin{aligned}|f(w,z)| &\le \Pr((W_{r-1},Z_{r-1}) = (w,z) \mid \gamma_{r-1}, m_r)\\ &= \Pr((W_{r-1},Z_{r-1}) = (w,z) \mid \gamma_{r-2}, w_{r-2}, z_{r-2}, m_{r-1}),\end{aligned}$$

because (W_{r-1}, Z_{r-1}) is conditionally independent of (A_{r-2}, B_{r-2}) and M_r, again by the rectangle property. Consequently, by definition of collision entropy, we have that

$$\begin{aligned}\sum_{w,z\in\{0,1\}^n} f(w,z)^2 &\le 2^{-\mathbb{H}_2(W_{r-1},Z_{r-1}|\gamma_{r-2},w_{r-2},z_{r-2},m_{r-1})}\\ &= 2^{-2n+\mathbb{D}_2(W_{r-1},Z_{r-1}|\gamma_{r-2},w_{r-2},z_{r-2},m_{r-1})}\\ &\le 2^{-2n+\Phi_{r-1}},\end{aligned} \tag{13}$$

for any $(\gamma_{r-2}, w_{r-2}, z_{r-2}, m_{r-1})$ satisfying $\overline{\mathcal{E}}_{r-1}$.

Furthermore, let Q denote the distribution of (A_r, B_r) conditioned on $(\Gamma_{r-1}, M_r) = (\gamma_{r-1}, m_r)$, for notational convenience. Define Q' to be the distribution of $a^{-1}b$ for $(a,b) \sim Q$. By definition of min-entropy, we have that

$$2^{\mathbb{D}_\infty(Q')} = n! \cdot 2^{-\mathbb{H}_\infty(Q')} \le n! \cdot n! \cdot 2^{-\mathbb{H}_\infty(Q)} = 2^{\mathbb{D}_\infty(Q)},$$

because at most $n!$ pairs of (a,b) can map to the same $a^{-1}b$. Equivalently, it must be that $\mathbb{D}_\infty(Q') \le \mathbb{D}_\infty(Q)$. Then, we can get

$$\begin{aligned}\sum_{w,z\in\{0,1\}^n} g(w,z)^2 &= \sum_{w,z\in\{0,1\}^n} \left(\mathop{\mathbb{E}}_{(a,b)\sim Q}(-1)^{\langle aw,bz\rangle}\right)\left(\mathop{\mathbb{E}}_{(a',b')\sim Q}(-1)^{\langle a'w,b'z\rangle}\right)\\ &= \sum_{w,z\in\{0,1\}^n} \mathop{\mathbb{E}}_{(a,b),(a',b')\sim Q}(-1)^{\langle aw,bz\rangle+\langle a'w,b'z\rangle}\\ &= \sum_{w,z\in\{0,1\}^n} \mathop{\mathbb{E}}_{(a,b),(a',b')\sim Q}(-1)^{\langle w,a^{-1}bz\rangle-\langle w,(a')^{-1}b'z\rangle}\\ &= \mathop{\mathbb{E}}_{c,c'\sim Q'} \sum_{w,z\in\{0,1\}^n} (-1)^{\langle w,(c-c')z\rangle}\\ &= \mathop{\mathbb{E}}_{c,c'\sim Q'} \sum_{z\in\{0,1\}^n} 2^n \cdot \mathbb{1}((c-c')z = \mathbf{0})\\ &= 2^n \cdot \mathop{\mathbb{E}}_{c,c'\sim Q'} \sum_{z\in\{0,1\}^n} \mathbb{1}\left(c^{-1}c'z = z\right)\\ &= 2^n \cdot \mathop{\mathbb{E}}_{c,c'\sim Q'} 2^{\mathsf{cyc}(c^{-1}c')},\end{aligned} \tag{14}$$

where in the last step, we use an argument similar to the one in the proof of Lemma 7.3.

[21] For convenience, we use the convention $m_0 = \varnothing$. Similarly for any variable with negative subscript.

Plugging Equations (13) and (14) into Equation (12), and by Claim 7.4 for $m = 1$, we have that

$$\begin{aligned}
&\Pr(O = \langle W_r, Z_r\rangle \mid \gamma_{r-1}, m_r) - \Pr(O \neq \langle W_r, Z_r\rangle \mid \gamma_{r-1}, m_r) \\
&\quad\leq \left(2^{-2n+\Phi_{r-1}} \cdot 2^n \cdot 2^{\frac{\mathbb{D}_\infty(A_r, B_r \mid \gamma_{r-1}, m_r)+\mathcal{O}(n)}{\log n}}\right)^{\frac{1}{2}} \\
&\quad\leq \left(2^{-n+\Phi_{r-1}} \cdot 2^{\frac{L_r+Kn}{\log n}}\right)^{\frac{1}{2}} \\
&\quad= \left(2^{-n+\Phi_r}\right)^{\frac{1}{2}},
\end{aligned}$$

for any (γ_{r-1}, m_r) satisfying both $\overline{\mathcal{E}}_{r-1}$ and $\overline{\mathcal{E}}'_r$, and sufficiently large constant K. This concludes the proof by averaging over (γ_{r-1}, m_r), as

$$\Phi_r \leq \frac{rL + Krn}{\log n} = o(n),$$

for $L = o(n/r \cdot \log n)$ and $r = o(\log n)$. □

Combining Lemmas 7.5 to 7.7 yields the desired $\Omega(n/r \cdot \log n)$ lower bound. Indeed, we have that

$$\begin{aligned}
&\Pr(O = \langle W_r, Z_r\rangle) \\
&\quad\leq \Pr(O = \langle W_r, Z_r\rangle \mid \overline{\mathcal{E}}_{r-1}, \overline{\mathcal{E}}'_r) + \Pr(\mathcal{E}'_r) + \Pr(\mathcal{E}_{r-1}) \\
&\quad\leq \frac{1}{2} + 2^{-\Omega(n)} + 2^{-n} + \Pr(\mathcal{E}_{r-1}) && \text{(by Lemmas 7.5 and 7.7)} \\
&\quad\leq \frac{1}{2} + 2^{-\Omega(n)} + 2^{-n} + \Pr(\mathcal{E}_{r-1} \mid \overline{\mathcal{E}}_{r-2}, \overline{\mathcal{E}}'_{r-1}) + \Pr(\mathcal{E}'_{r-1}) + \Pr(\mathcal{E}_{r-2}) \\
&\quad\leq \frac{1}{2} + 2^{-\Omega(n)} + 2^{-n} + 2^{-\frac{n}{\log n}} + 2^{-n} + \Pr(\mathcal{E}_{r-2}) && \text{(by Lemmas 7.5 and 7.6)} \\
&\quad\leq \frac{1}{2} + 2^{-\Omega(n)} + 2^{-n} + 2^{-\frac{n}{\log n}} + 2^{-n} + \Pr(\mathcal{E}_{r-2} \mid \overline{\mathcal{E}}_{r-3}, \overline{\mathcal{E}}'_{r-2}) + \Pr(\mathcal{E}'_{r-2}) + \Pr(\mathcal{E}_{r-3}) \\
&\qquad\vdots \\
&\quad\leq \frac{1}{2} + 2^{-\Omega(n)} + 2^{-n} + (r-1) \cdot \left(2^{-\frac{n}{\log n}} + 2^{-n}\right) + \Pr(\mathcal{E}_0) \\
&\quad= \frac{1}{2} + o(1),
\end{aligned}$$

contradicting the error probability $\varepsilon \leq 1/3$.

7.2 Proof of Theorem 7.1

We now consider any protocol with communication $L = n/1000 \cdot \log n$. The major bottleneck of the analysis in Section 7.1 is that the min-entropy loss in (A_i, B_i) is upper bounded only in terms of the total length L_i of all previous messages, which could be as large as L. Then, a total loss of $rL/\log n$ in Φ_r would fail the advantage bound in Lemma 7.7.

Intuitively, however, at most L_i bits of information in total have been revealed across all pairs of permutations (A_j, B_j) for $j \in [i]$. In other words, (A_i, B_i) should only be charged for the actual amount of revealed information, instead of the maximum possible in the worst case. After a closer look at the proof in Section 7.1, we can realize that the full version of Theorem 7.1 would follow if the upper bound in Lemma 7.3 were in terms of the standard Shannon entropy loss $\mathbb{D}_1(Q)$ rather than min-entropy loss $\mathbb{D}_\infty(Q)$. Indeed, although $\mathbb{D}_1(A_i, B_i \mid \gamma_{i-1}, m_i)$ could be as large as L individually for each i, its sum over all i should be at most L in expectation, by chain rule of Shannon entropy (Fact A.1). This would yield a total loss of only $L/\log n$ in Φ_r.

Unfortunately, the difference between Shannon entropy and min-entropy can be unbounded in general, and may be at least $\Omega(\log(n!)) = \Omega(n \log n)$ in our case. Charging it in every round would immediately destroy all the analysis.

Nevertheless, it turns out we can actually guarantee that Shannon entropy and min-entropy are almost the same, by associating each round with a "virtual message" $\widehat{M}_i$. In order to properly define $\widehat{M}_i$, we need to change the definition of Γ_i at the same time. Concretely, starting with $\Gamma_0 = \varnothing$, recursively define

$$\widehat{M}_i = \lfloor -\log \Pr(A_i, B_i \mid \gamma_{i-1}, m_i) \rfloor \qquad \text{and} \qquad \Gamma_i = (X, Y, A_{<i}, B_{<i}, M'_{\le i}),$$

where $M'_i = (M_i, \widehat{M}_i)$ is the new "combined message". Again, we emphasize that $\widehat{M}_i$ is only for analysis purposes and is not communicated by the parties throughout the protocol. At the same time, it is also worth noting that the owner of (A_i, B_i) can actually compute $\widehat{M}_i$ based only on its own input and the partial transcript, including virtual messages, so far. This is because the rectangle property of deterministic communication protocols ensures that the inputs of the two parties are always independent even conditioned on any partial transcript. Consequently, $\widehat{M}_i$ can be treated just like the actual message M_i in the analysis, and in particular preserves the rectangle property itself. Also, no assumption on fixed message lengths is required.

The following Lemma 7.8 shows that this virtual message $\widehat{M}_i$ has negligible impact on the communication complexity.

Lemma 7.8. *For $i \in [r]$, it holds that*

$$\log(\widehat{M}_i + 1) = \mathcal{O}(\log n).$$

Proof. We prove the lemma in the case of odd i. The case of even i is similar.

Due to the rectangle property of deterministic communication protocols, Alice always holds a uniform distribution on $\mathsf{supp}(\{(A_{i'}, B_{i'})\}_{\text{odd } i' \ge i} \mid \gamma_{i-1}, m_i)$ after conditioning on any $(\Gamma_{i-1}, M_i) = (\gamma_{i-1}, m_i)$. As a result, any possible (A_i, B_i) has marginal probability at least $(n!)^{-2r}$, as the universe is of size at most $(n!)^{2r}$. This concludes the proof as

$$\log \log(n!)^{2r} = \mathcal{O}(\log(rn \log n)) = \mathcal{O}(\log n).$$

□

By Lemma 7.8, the extra length caused by $\widehat{M}_i$ is at most $\mathcal{O}(\log(\widehat{M}_i + 1)) = \mathcal{O}(\log n)$. More importantly, revealing $\widehat{M}_i$ partitions $\mathsf{supp}(A_i, B_i \mid \gamma_{i-1}, m_i)$ into different scales, and all

probabilities within the same scale differ by at most a factor of 2. Specifically, we have that

$$\mathbb{D}_\infty(A_i, B_i \mid \gamma_{i-1}, m_i') \le \mathbb{D}_1(A_i, B_i \mid \gamma_{i-1}, m_i') + 1. \tag{15}$$

The full proof of Theorem 7.1 essentially follows by replacing M_i with the combined message M_i' in the analysis in Section 7.1. Indeed, change the definition of Φ_i to be

$$\Phi_i = \sum_{j \in [i]} \frac{\mathbb{D}_1(A_j, B_j \mid \gamma_{j-1}, m_j') + Kn}{\log n},$$

which is now a random variable depending on (γ_{i-1}, m_i'). The event $\mathcal{E}_i$ is now defined with respect to the updated threshold Φ_i. In the meantime, we no longer use the events $\mathcal{E}_i'$. Instead, define a single new event $\mathcal{E}'$ to be

$$\sum_{i \in [r]} \mathbb{D}_1(A_i, B_i \mid \gamma_{i-1}, m_i') > 20L.$$

With the updated definitions, analogs to Lemmas 7.5 to 7.7 can be shown using quite similar arguments. We do not repeat all the proofs and only sketch a few required adaptations below. The new statements are summarized in Lemma 7.9.

Lemma 7.9. *It holds that*

1. $$\Pr(\mathcal{E}') \le \frac{1}{10}.$$

2. *For $i \in [r]$,*
$$\Pr(\mathcal{E}_i \mid \overline{\mathcal{E}}_{i-1}) \le 2^{-\frac{n}{\log n}}.$$

3. $$\Pr(O = \langle W_r, Z_r \rangle \mid \overline{\mathcal{E}}_{r-1}, \overline{\mathcal{E}}') \le \frac{1}{2} + 2^{-\Omega(n)}.$$

Proof. Since the input is uniformly random at the beginning of the protocol, we have that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y,a_{\le r}, b_{\le r}, m'_{\le r}} \sum_{i \in [r]} \mathbb{D}_1(A_i, B_i \mid \gamma_{i-1}, m_i') \\
&\quad = \sum_{i \in [r]} \mathbb{I}(A_i, B_i; \Gamma_{i-1}, M_i') \\
&\quad = \mathbb{I}(A_1, B_1; M_1') \\
&\qquad + \sum_{i \in [2,r]} \left(\mathbb{I}(A_i, B_i; X, Y, A_{<(i-1)}, B_{<(i-1)}) + \mathbb{I}(A_i, B_i; M'_{\le i} \mid X, Y, A_{<(i-1)}, B_{<(i-1)}) \right) \\
&\qquad\qquad \text{(by chain rule of mutual information (Fact A.2))} \\
&\quad = \mathbb{I}(A_1, B_1; M_1') + \sum_{i \in [2,r]} \mathbb{I}(A_i, B_i; M'_{\le i} \mid X, Y, A_{<(i-1)}, B_{<(i-1)})
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{I}(A_1, B_1; M'_1) + \sum_{i\in[2,r]} \mathbb{I}(A_i, B_i; M'_{\leq i} \mid X, Y, A_{<i}, B_{<i}) && \text{(by Fact A.3)}\\
&\leq \mathbb{I}(X, Y, A_1, B_1; M'_{\leq r}) + \sum_{i\in[2,r]} \mathbb{I}(A_i, B_i; M'_{\leq r} \mid X, Y, A_{<i}, B_{<i})\\
&= \mathbb{I}(X, Y, A_{\leq r}, B_{\leq r}; M'_{\leq r}) && \text{(by chain rule of mutual information (Fact A.2))}\\
&\leq \mathbb{H}_1(M'_{\leq r})\\
&\leq \mathbb{H}_1(M_{\leq r}) + \mathbb{H}_1(\widehat{M}_{\leq r})\\
&\leq L + \mathcal{O}(r \log n) && \text{(by Lemma 7.8)}\\
&\leq 2L.
\end{aligned}
$$

Item 1 then follows from Markov's inequality.

For Item 2, almost the same proof as for Lemma 7.6 applies, except that the upper bound on $\mathbb{D}_2((A_i, B_i)^{(w,z)} \mid \gamma_{i-1}, m'_i)$ now comes from Lemma 7.3 and Equation (15), instead of being guaranteed by $\overline{\mathcal{E}}'_i$.

Item 3 is shown by a similar argument to the one for Lemma 7.7, where $\overline{\mathcal{E}}'$ now ensures

$$\Phi_r \leq \frac{20L + Krn}{\log n} \leq \frac{n}{10},$$

for $r = o(\log n)$. □

Finally, we are ready to prove the full version of Theorem 7.1.

Proof of Theorem 7.1. By Lemma 7.9, we have that

$$
\begin{aligned}
\Pr(O = \langle W_r, Z_r\rangle) &\leq \Pr(O = \langle W_r, Z_r\rangle \mid \overline{\mathcal{E}}_{r-1}, \overline{\mathcal{E}}') + \Pr(\mathcal{E}') + \Pr(\mathcal{E}_{r-1})\\
&\leq \frac{1}{2} + 2^{-\Omega(n)} + \frac{1}{10} + \Pr(\mathcal{E}_{r-1}) && \text{(by Items 1 and 3)}\\
&\leq \frac{3}{5} + 2^{-\Omega(n)} + \Pr(\mathcal{E}_{r-1} \mid \overline{\mathcal{E}}_{r-2}) + \Pr(\mathcal{E}_{r-2})\\
&\leq \frac{3}{5} + 2^{-\Omega(n)} + 2^{-\frac{n}{\log n}} + \Pr(\mathcal{E}_{r-2}) && \text{(by Item 2)}\\
&\leq \frac{3}{5} + 2^{-\Omega(n)} + 2^{-\frac{n}{\log n}} + \Pr(\mathcal{E}_{r-2} \mid \overline{\mathcal{E}}_{r-3}) + \Pr(\mathcal{E}_{r-3})\\
&\;\;\vdots\\
&\leq \frac{3}{5} + 2^{-\Omega(n)} + (r-1)\cdot 2^{-\frac{n}{\log n}} + \Pr(\mathcal{E}_0)\\
&= \frac{3}{5} + o(1),
\end{aligned}
$$

a contradiction to the assumption $\epsilon \leq 1/3$. This concludes the proof. □

AI Methodology

Extensive interaction with ChatGPT was used to assist with exploring, formulating, and improving proof ideas for several lemmas in this paper. The tool materially affected Lemmas 4.6, 5.3 and C.2. In particular,

1. ChatGPT discovered the edge coloring argument for proving Lemma 4.6;

2. ChatGPT produced a preliminary version of the error-dependent randomized lower bound with worse $\Omega(\log\log(1/\varepsilon))$ dependency on ε. The approach was slightly different but to some extent could be viewed as a weaker version of Lemma 5.3 where the additional loss term was exponential in n. It alone was useful only for exponentially small error probability. Direct sum arguments were used to extend it to the full scale of error probabilities. The proof has been significantly simplified and improved by human authors to derive the current $\Omega(\log(1/\varepsilon))$ dependency.

3. ChatGPT suggested the use of Hölder's inequality to optimize the lower order term in proving Lemma C.2. It relaxed the round requirement of the lower bound for $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ but did not improve the communication bound for $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ or graph connectivity.

ChatGPT was also used for proof-reading and minor copy-editing. Human authors take full responsibility for correctness and exposition of all content in the final manuscript.

References

[AA05] Scott Aaronson and Andris Ambainis. Quantum search of spatial regions. *Theory Comput.*, 1(4):47–79, 2005. 4

[ABK+26] Sepehr Assadi, Soheil Behnezhad, Christian Konrad, Kheeran K. Naidu, and Janani Sundaresan. Settling the pass complexity of approximate matchings in dynamic graph streams. *J. ACM*, 73(4):25:1–25:83, 2026. 4

[ACK21] Sepehr Assadi, Deeparnab Chakrabarty, and Sanjeev Khanna. Graph connectivity and single element recovery via linear and OR queries. In *29th Annual European Symposium on Algorithms*, volume 204 of *LIPIcs*, pages 7:1–7:19, 2021. 4

[AD21] Sepehr Assadi and Aditi Dudeja. A simple semi-streaming algorithm for global minimum cuts. In *4th Symposium on Simplicity in Algorithms*, pages 172–180, 2021. 3, 4, 21, 22

[AEG+22] Simon Apers, Yuval Efron, Pawel Gawrychowski, Troy Lee, Sagnik Mukhopadhyay, and Danupon Nanongkai. Cut query algorithms with star contraction. In *63rd IEEE Annual Symposium on Foundations of Computer Science*, pages 507–518, 2022. 1, 3, 4, 9

[AGM12] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Analyzing graph structure via linear measurements. In *Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 459–467, 2012. 4

[AKNS24] Sepehr Assadi, Christian Konrad, Kheeran K. Naidu, and Janani Sundaresan. O(log log n) passes is optimal for semi-streaming maximal independent set. In *56th Annual ACM Symposium on Theory of Computing*, pages 847–858, 2024. 4

[AKO20] Sepehr Assadi, Gillat Kol, and Rotem Oshman. Lower bounds for distributed sketching of maximal matchings and maximal independent sets. In *ACM Symposium on Principles of Distributed Computing*, pages 79–88, 2020. 4

[AKZ24] Sepehr Assadi, Gillat Kol, and Zhijun Zhang. Optimal multi-pass lower bounds for MST in dynamic streams. In *56th Annual ACM Symposium on Theory of Computing*, pages 835–846, 2024. 4

[AKZ25] Sepehr Assadi, Gillat Kol, and Zhijun Zhang. Rounds vs. communication tradeoffs for maximal independent sets. *SIAM J. Comput.*, 54(5):FOCS22-20–FOCS22-59, 2025. 4

[AN21] Sepehr Assadi and Vishvajeet N. Graph streaming lower bounds for parameter estimation and property testing via a streaming XOR lemma. In *53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 612–625, 2021. 3, 7, 22

[ASW25] Aditya Anand, Thatchaphol Saranurak, and Yunfan Wang. Deterministic edge connectivity and max flow using subquadratic cut queries. In *2025 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 124–142, 2025. 1, 4

[BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Thirtieth Annual ACM Symposium on the Theory of Computing*, pages 63–68, 1998. 4

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, 1986. 1, 2

[BGK+18] Mark Braverman, Ankit Garg, Young Kun Ko, Jieming Mao, and Dave Touchette. Near-optimal bounds on the bounded-round quantum communication complexity of disjointness. *SIAM J. Comput.*, 47(6):2277–2314, 2018. 4

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Symposium on Theory of Computing Conference*, pages 151–160, 2013. 4

[BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. 1

[BM13] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Symposium on Theory of Computing Conference*, pages 161–170, 2013. 4

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *54th Annual IEEE Symposium on Foundations of Computer Science*, pages 746–755, 2013. 4

[BvdBE+22] Joakim Blikstad, Jan van den Brand, Yuval Efron, Sagnik Mukhopadhyay, and Danupon Nanongkai. Nearly optimal communication and query complexity of bipartite matching. In *63rd IEEE Annual Symposium on Foundations of Computer Science*, pages 1174–1185, 2022. 1, 4

[BW15] Mark Braverman and Omri Weinstein. An interactive information odometer and applications. In *Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 341–350, 2015. 4

[CJSY26] Yu Cheng, Tianle Jiang, Pachara Sawettamalya, and Huacheng Yu. On the communication complexity of maximum matching and negative-weight shortest paths. In *34th Annual European Symposium on Algorithms*, volume 388 of *LIPIcs*, pages 72:1–72:22, 2026. 1

[CMNS25] Yi-Jun Chang, Gopinath Mishra, Hung Thuan Nguyen, and Farrel D. Salim. Round and communication efficient graph coloring. In *ACM Symposium on Principles of Distributed Computing*, pages 360–371, 2025. 1

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006. 44

[CvDNT13] Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp. Quantum entanglement and the communication complexity of the inner product function. *Theor. Comput. Sci.*, 486:11–19, 2013. 4

[DGR25] Mahsa Derakhshan, Andisheh Ghasemi, and Rajmohan Rajaraman. One-way communication complexity of minimum vertex cover in general graphs. In *52nd International Colloquium on Automata, Languages, and Programming*, volume 334 of *LIPIcs*, pages 66:1–66:19, 2025. 1

[DJS98] Carsten Damm, Stasys Jukna, and Jirí Sgall. Some bounds on multiparty communication complexity of pointer jumping. *Comput. Complex.*, 7(2):109–127, 1998. 3

[ĎP89] Pavol Ďuriš and Pavel Pudlák. On the communication complexity of planarity. In *Fundamentals of Computation Theory*, volume 380 of *Lecture Notes in Computer Science*, pages 145–147, 1989. 1

[EKNP14] Michael Elkin, Hartmut Klauck, Danupon Nanongkai, and Gopal Pandurangan. Can quantum communication speed up distributed computation? In *ACM Symposium on Principles of Distributed Computing*, pages 166–175, 2014. 4

[EKSZ26] Klim Efremenko, Gillat Kol, Raghuvansh R. Saxena, and Zhijun Zhang. Universally optimal streaming algorithm for random walks in dense graphs. In *17th Innovations in Theoretical Computer Science Conference*, volume 362 of *LIPIcs*, pages 55:1–55:20, 2026. 4

[FKM+08] Joan Feigenbaum, Sampath Kannan, Andrew McGregor, Siddharth Suri, and Jian Zhang. Graph distances in the data-stream model. *SIAM J. Comput.*, 38(5):1709–1727, 2008. 1, 4

[FM25] Maxime Flin and Parth Mittal. $(\Delta + 1)$ vertex coloring in O(n) communication. *Distributed Comput.*, 38(1):19–29, 2025. 1

[GKK12] Ashish Goel, Michael Kapralov, and Sanjeev Khanna. On the communication and streaming complexity of maximum bipartite matching. In *Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 468–485, 2012. 1, 4

[GO16] Venkatesan Guruswami and Krzysztof Onak. Superlinear lower bounds for multipass graph processing. *Algorithmica*, 76(3):654–683, 2016. 4

[GW16] Mika Göös and Thomas Watson. Communication complexity of set-disjointness for all probabilities. *Theory Comput.*, 12(9):1–23, 2016. 4

[HdW02] Peter Høyer and Ronald de Wolf. Improved quantum communication complexity bounds for disjointness and equality. In *19th Annual Symposium on Theoretical Aspects of Computer Science*, volume 2285 of *Lecture Notes in Computer Science*, pages 299–310, 2002. 4

[HMT88] András Hajnal, Wolfgang Maass, and György Turán. On the communication complexity of graph properties. In *20th Annual ACM Symposium on Theory of Computing*, pages 186–191, 1988. 1

[HPZZ21] Dawei Huang, Seth Pettie, Yixiang Zhang, and Zhijun Zhang. The communication complexity of set intersection and multiple equality testing. *SIAM J. Comput.*, 50(2):674–717, 2021. 4

[IKL+12] Gábor Ivanyos, Hartmut Klauck, Troy Lee, Miklos Santha, and Ronald de Wolf. New bounds on the classical and quantum communication complexity of some graph properties. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 18 of *LIPIcs*, pages 148–159, 2012. 1, 4

[JNS26] Yonggang Jiang, Danupon Nanongkai, and Pachara Sawettamalya. Minimum s-t cuts with fewer cut queries. In *2026 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 258–296, 2026. 1, 4

[JPY16] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for two-party bounded-round public-coin communication complexity. *Algorithmica*, 76(3):720–748, 2016. 4

[JRS02a] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *43rd Symposium on Foundations of Computer Science*, pages 429–438, 2002. 4

[JRS02b] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. The quantum communication complexity of the pointer chasing problem: The bit version. In *Foundations of Software Technology and Theoretical Computer Science*, volume 2556 of *Lecture Notes in Computer Science*, pages 218–229, 2002. 4

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of set disjointness. In *44th Symposium on Foundations of Computer Science*, pages 220–229, 2003. 4

[Kla00] Hartmut Klauck. On quantum and probabilistic communication: Las Vegas and one-way protocols. In *Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 644–651, 2000. 3

[KNTZ07] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication. *IEEE Trans. Inf. Theory*, 53(6):1970–1982, 2007. 4

[Kre95] Ilan Kremer. Quantum communication. Master's thesis, Hebrew University of Jerusalem, 1995. 4

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discret. Math.*, 5(4):545–557, 1992. 1

[KSdW07] Hartmut Klauck, Robert Spalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007. 4

[LMdW23] Olivier Lalonde, Nikhil S. Mande, and Ronald de Wolf. Tight bounds for the randomized and quantum communication complexities of equality with small error. In *43rd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 284 of *LIPIcs*, pages 32:1–32:18, 2023. 4

[MYZ25] Xinyu Mao, Guangxu Yang, and Jiapeng Zhang. Gadgetless lifting beats round elimination: Improved lower bounds for pointer chasing. In *16th Innovations in Theoretical Computer Science Conference*, volume 325 of *LIPIcs*, pages 75:1–75:14, 2025. 3

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39(2):67–71, 1991. 9

[NS06] Ashwin Nayak and Julia Salzman. Limits on the ability of quantum states to convey classical messages. *J. ACM*, 53(1):184–206, 2006. 4

[NW93] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM J. Comput.*, 22(1):211–219, 1993. 3

[NY19] Jelani Nelson and Huacheng Yu. Optimal lower bounds for distributed and streaming spanning forest computation. In *Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1844–1860, 2019. 4

[PRV01] Stephen Ponzio, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. The communication complexity of pointer chasing. *J. Comput. Syst. Sci.*, 62(2):323–355, 2001. 3

[PS84] Christos H. Papadimitriou and Michael Sipser. Communication complexity. *J. Comput. Syst. Sci.*, 28(2):260–269, 1984. 3

[PVZ16] Jeff M. Phillips, Elad Verbin, and Qin Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. *SIAM J. Comput.*, 45(1):174–196, 2016. 4

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992. 1

[Raz03] Alexander A Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003. 4

[RRV26] Jaikumar Radhakrishnan, Chaitanya Reddy, and Rakesh Venkat. Optimal two-round communication lower bound for graph connectivity via pointer chasing. In *17th Innovations in Theoretical Computer Science Conference*, volume 362 of *LIPIcs*, pages 110:1–110:20, 2026. 1, 2, 3, 4, 6, 7, 22

[RS95] Ran Raz and Boris Spieker. On the "log rank"-conjecture in communication complexity. *Comb.*, 15(4):567–588, 1995. 1, 2, 4, 5

[SW15] Xiaoming Sun and David P. Woodruff. Tight bounds for graph problems in insertion streams. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 40 of *LIPIcs*, pages 435–448, 2015. 1, 2, 4

[WZ17] David P. Woodruff and Qin Zhang. When distributed computation is communication expensive. *Distributed Comput.*, 30(5):309–323, 2017. 4

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *11th Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979. 1

[Yeh20] Amir Yehudayoff. Pointer chasing via triangular discrimination. *Comb. Probab. Comput.*, 29(4):485–494, 2020. 3

[Yu21] Huacheng Yu. Tight distributed sketching lower bound for connectivity. In *2021 ACM-SIAM Symposium on Discrete Algorithms*, pages 1856–1873, 2021. 4

[Yu22] Huacheng Yu. Strong XOR lemma for communication with bounded rounds: (extended abstract). In *63rd IEEE Annual Symposium on Foundations of Computer Science*, pages 1186–1192, 2022. 4

A Basic Tools from Information Theory

We introduce some definitions from information theory that we use in this paper. Interested readers are referred to the excellent textbook by Cover and Thomas [CT06] for an introduction to information theory. For notational convenience, we may identify a random variable with its distribution.

The *Shannon entropy* of a random variable X, with distribution μ supported on universe $\mathcal{U}$, is defined to be

$$\mathbb{H}(X) = \sum_{x \in \mathcal{U}} \mu(x) \cdot \log \frac{1}{\mu(x)}.$$

The *conditional entropy* of X conditioned on Y is defined as

$$\mathbb{H}(X \mid Y) = \mathop{\mathbb{E}}_{y \sim Y} \mathbb{H}(X \mid Y = y).$$

For two distributions μ, ν supported on the same universe $\mathcal{U}$, the *Kullback-Leibler (KL) divergence* of μ from ν is defined to be

$$\mathbb{D}_{\mathsf{KL}}(\mu \| \nu) = \sum_{x \in \mathcal{U}} \mu(x) \cdot \log \frac{\mu(x)}{\nu(x)}.$$

For $\alpha \in [0, 1) \cup (1, \infty)$, the *Rényi entropy* of order α of a random variable X, with distribution μ supported on universe $\mathcal{U}$, is defined to be

$$\mathbb{H}_\alpha(X) = \frac{1}{1-\alpha} \cdot \log \sum_{x \in \mathcal{U}} \mu(x)^\alpha.$$

For $\alpha = 1$, taking the limit recovers Shannon entropy:

$$\mathbb{H}_1(X) = \lim_{\alpha \to 1} \mathbb{H}_\alpha(X) = \mathbb{H}(X).$$

The *min-entropy* of X is defined to be

$$\mathbb{H}_\infty(X) = \lim_{\alpha \to \infty} \mathbb{H}_\alpha(X) = -\log \max_{x \in \mathcal{U}} \mu(x).$$

The case of $\alpha = 2$ is also known as *collision entropy*:

$$\mathbb{H}_2(X) = -\log \sum_{x \in \mathcal{U}} \mu(x)^2,$$

as the sum is exactly the probability of collision for two independently drawn samples.

For two distributions μ, ν supported on the same universe $\mathcal{U}$, the *Rényi divergence* of order α of μ from ν is defined to be

$$\mathbb{D}_\alpha(\mu \| \nu) = \frac{1}{\alpha - 1} \cdot \log \sum_{x \in \mathcal{U}} \mu(x)^\alpha \nu(x)^{1-\alpha},$$

for $\alpha \in (0,1) \cup (1,\infty)$, and extended to $\alpha = 1, \infty$ by limits. $\mathbb{D}_1(\cdot\|\cdot)$ is exactly KL divergence. If ν is the uniform distribution on $\mathcal{U}$, we simply write

$$\mathbb{D}_\alpha(\mu) = \mathbb{D}_\alpha(\mu\|\nu) = \log|\mathcal{U}| - \mathbb{H}_\alpha(\mu).$$

The *mutual information* between two random variables X, Y is defined to be

$$\mathbb{I}(X;Y) = \mathbb{H}(X) - \mathbb{H}(X \mid Y) = \mathbb{H}(Y) - \mathbb{H}(Y \mid X),$$

and the *conditional mutual information* between X, Y conditioned on Z is defined as

$$\mathbb{I}(X;Y \mid Z) = \mathop{\mathbb{E}}_{z\sim Z} \mathbb{I}(X;Y \mid Z = z).$$

Throughout, we use the following basic properties.

Fact A.1 (Chain rule of Shannon entropy)**.** *For random variables A, B, C, it holds that*

$$\mathbb{H}(A, B \mid C) = \mathbb{H}(A \mid B, C) + \mathbb{H}(B \mid C).$$

Fact A.2 (Chain rule of mutual information)**.** *For random variables A, B, C, D, it holds that*

$$\mathbb{I}(A, B; C \mid D) = \mathbb{I}(A; C \mid B, D) + \mathbb{I}(B; C \mid D).$$

Fact A.3. *For random variables A, B, C, D such that $A \perp D \mid C$, it holds that*

$$\mathbb{I}(A; B \mid C) \le \mathbb{I}(A; B \mid C, D).$$

Fact A.4. *For random variables A, B, C, D such that $A \perp D \mid B, C$, it holds that*

$$\mathbb{I}(A; B \mid C) \ge \mathbb{I}(A; B \mid C, D).$$

B Round-Efficient Protocols for PC-Perm

We present round-efficient protocols for PC-Perm and bPC-Perm in this section. For simplicity, only the receiver of the last message is required to output for protocols in this section. It only changes the round complexity by 1.

Theorem B.1. *For $n, r \ge 1$, and $k \le r^2$, there exists an r-round deterministic protocol, starting with Bob, for $\mathsf{PC\text{-}Perm}_{n,k}$ with $\mathcal{O}(nr^2 \log\log n)$ communication.*

Moreover, for $k \le r^2 + r$, there exists an r-round deterministic protocol, starting with Bob, for $\mathsf{bPC\text{-}Perm}_{n,k}$ with $\mathcal{O}(nr^2 \log\log n)$ communication.

The protocol for proving Theorem B.1 utilizes Algorithm 1 as a subroutine, which solves $\mathsf{PC\text{-}Perm}_{n,2\tau-1}$ in τ rounds with $\mathcal{O}(n\tau \log\log n)$ communication, assuming the proper party starts communication. More crucially, it has the "delayed starting vertex" property, as formalized in Lemma B.2.

Algorithm 1 A τ-round protocol for $\mathsf{PC\text{-}Perm}_{n,2\tau-1}$.

Set $p \leftarrow 1$ for Alice and $p \leftarrow 0$ for Bob.
Initialize $L = \lceil \log n \rceil$, and $\chi_i(v) \leftarrow 0$ for $i \in [\tau]$ and $v \in [n]$.
if $p \equiv \tau \pmod 2$ **then**
    Set $\chi_\tau(v) \leftarrow v \bmod L$ for $v \in [n]$.
    Set $\chi_{\tau-1}(v) \leftarrow \chi_\tau(f_\tau(v))$ for $v \in [n]$.
end if
for $i = 1, \ldots, \tau - 1$ **do**
    if $p \equiv \tau - i + 1 \pmod 2$ **then**
        Send $\chi_{\tau-i}, \chi_{\tau+i-1}$.
    else
        Receive $\chi_{\tau-i}, \chi_{\tau+i-1}$.
        Set $\chi_{\tau-i-1}(v) \leftarrow \chi_{\tau-i}(f_{\tau-i}(v))$ for $v \in [n]$.
        Set $\chi_{\tau+i}(v) \leftarrow \chi_{\tau+i-1}(f^{-1}_{\tau+i}(v))$ for $v \in [n]$.
    end if
end for
if $p = 1$ **then**
    Get input v_0.
    Send v_0, and $\chi_0, \chi_{2\tau-1}$.
    Send $f_i(v)$ for odd $i \in [2\tau - 1]$ and all $v \in [n]$ with $\chi_{i-1}(v) = \chi_0(v_0)$.
else
    Receive v_0, and $\chi_0, \chi_{2\tau-1}$.
    Receive $f_i(v)$ for odd $i \in [2\tau - 1]$ and all $v \in [n]$ with $\chi_{i-1}(v) = \chi_0(v_0)$.
    Output $f_{2\tau-1} \circ \cdots \circ f_1(v_0)$.
end if

Lemma B.2. *For $n, \tau \geq 1$, there exists a τ-round deterministic protocol for* $\mathsf{PC\text{-}Perm}_{n,2\tau-1}$ *with $\mathcal{O}(n\tau \log\log n)$ communication such that*[22]

1. *The first message is sent by Alice for odd τ and Bob for even τ.*
2. *The last message is received by Bob.*
3. *The starting vertex v_0 is given to Alice only at the beginning of the last round.*

Proof. Fix sufficiently large n, and $\tau \geq 1$. We prove the lemma using Algorithm 1. Items 1 and 2 follow directly from the initialization in Line 1 and the conditions in Lines 8 and 16. They also ensure an alternating speaking order for Alice and Bob. Line 17 implies Item 3.

The correctness of the protocol is shown by induction on round i. In the base case of $i = 0$, the party with $p \equiv \tau \pmod 2$ holds f_τ as input. Due to the condition in Line 3, it (almost) equally partitions each of layers τ and $\tau - 1$ into $L = \Theta(\log n)$ color classes in Lines 4 and 5. Furthermore, the color of each vertex v in layer $\tau - 1$ is the same as the color of $f_\tau(v)$ in layer τ. In other words, the coloring is consistent with the pointers for layer τ.

In round $i \in [\tau-1]$, by induction, the sender knows the colorings of layers $\tau-i, \ldots, \tau+i-1$, while the receiver knows the same colorings except the two outer layers $\tau - i$ and $\tau + i - 1$. Once the receiver gets the colorings of layers $\tau - i, \tau + i - 1$ from the sender, it further extends them to layers $\tau - i - 1, \tau + i$ in Lines 12 and 13. This is well-defined because it holds $f_{\tau-i}, f_{\tau+i}$ as input.

In the last round, Alice sends the newly obtained starting vertex v_0, along with the colorings of layers $0, 2\tau - 1$. Upon receiving the message, Bob now knows the colorings of all layers. More importantly, for every $i \in [2\tau - 1]$, the i-th target vertex v_i must have the same color as v_0 because the colorings are consistent with all pointers. Consequently, it suffices for Alice to additionally send her pointers only for vertices with color $\chi_0(v_0)$. This ensures that Bob has enough knowledge to output correctly.

As to the total amount of communication, only $L = \Theta(\log n)$ colors are used, so sending each coloring costs at most $\mathcal{O}(n \log L) = \mathcal{O}(n \log\log n)$ bits in Lines 9 and 18. In the meantime, each layer has $\Theta(n/L)$ vertices of any specific color. Sending pointers for these vertices costs $\mathcal{O}(n/L \cdot \log n) = \mathcal{O}(n)$ per layer owned by Alice. In total, the protocol communicates $\mathcal{O}(n\tau \log\log n)$ bits. This concludes the proof. □

The "delayed starting vertex" property in Item 3 of Lemma B.2 enables us to run r parallel instances of Algorithm 1 on disjoint layers and gradually chase through more and more layers at once. Concretely, Algorithm 2 utilizes this intuition to quadratically improve the round complexity.

Now, we are ready to prove Theorem B.1.

Proof of Theorem B.1. Fix sufficiently large n, and $r \geq 1$. Assume without loss of generality that $k = r^2$. Since the case of $r = 1$ holds trivially, assume $r \geq 2$ as well. We first prove by induction on i that Algorithm 2 solves $\mathsf{PC\text{-}Perm}_{n,k}$. By Item 1 of Lemma B.2, Algorithm 1 starts with Alice for odd τ and Bob for even τ. Thus, swapping the roles of Alice and Bob for odd i ensures that all instances of Algorithm 1 start with Bob.

[22] In fact, Algorithm 1 can also solve $\mathsf{PC\text{-}Perm}_{n,2\tau}$ as Bob owns $f_{2\tau}$ and thus can output correctly in Line 23. Nevertheless, due to parity constraints, this extra layer does not help improve Algorithm 2.

Algorithm 2 An r-round protocol for $\mathsf{PC\text{-}Perm}_{n,r^2}$.

Set $p \leftarrow 1$ for Alice and $p \leftarrow 0$ for Bob.
for $i = 2, \ldots, r$ in parallel **do**
    Run in parallel an instance of Algorithm 1 with $\tau = i$ on $f_{(i-1)^2}, \ldots, f_{i^2-1}$, swapping Alice and Bob for odd i.
    Give $v_{(i-1)^2-1}$ to Alice (or Bob for odd i) at the beginning of round i.
    Get v_{i^2-1} from Bob (or Alice for odd i) at the end of round i.
end for
if $p \equiv r^2 \pmod 2$ **then**
    Output $f_{r^2}(v_{r^2-1})$.
end if

Algorithm 3 An r-round protocol for $\mathsf{bPC\text{-}Perm}_{n,r^2+r}$.

Set $p \leftarrow 1$ for Alice and $p \leftarrow 0$ for Bob.
if $p \equiv r^2 + r \pmod 2$ **then**
    Initialize $b_{r^2+r-1}(v) \leftarrow f_{r^2+r}(v) \bmod 2$ for $v \in [n]$.
end if
Run in parallel an instance of Algorithm 2 on layers $1, \ldots, r^2$.
for $i = 1, \ldots, r-1$ **do**
    if $p \equiv r^2 + r - i + 1 \pmod 2$ **then**
        Send b_{r^2+r-i}.
    else
        Receive b_{r^2+r-i}.
        Set $b_{r^2+r-i-1}(v) \leftarrow b_{r^2+r-i}(f_{r^2+r-i}(v))$ for $v \in [n]$.
    end if
end for
if $p \equiv r^2 + 1 \pmod 2$ **then**
    Send b_{r^2}.
else
    Receive b_{r^2}.
    Get v_{r^2} from Algorithm 2.
    Output $b_{r^2}(v_{r^2})$.
end if

For the base case of $i = 1$, Alice knows v_0. For $i \in [2, r]$, layer $(i-1)^2$ is held by Alice for even i and Bob for odd i. Also verify that $i^2 - 1 - (i-1)^2 + 1 = 2i - 1 = 2\tau - 1$. Therefore, the i-th instance of Algorithm 1 can be run on layers $(i-1)^2, \ldots, i^2 - 1$, with the roles of Alice and Bob swapped for odd i. In round i, the $(i-1)$-th instance has already completed, so $v_{(i-1)^2-1}$ is held by Alice for even i and Bob for odd i by inductive hypothesis (or v_0 is already known to Alice for $i = 2$). It then implies that Lines 28 and 29 are well-defined, due to Items 2 and 3 of Lemma B.2. After all r rounds, both v_{r^2-1} and f_{r^2} are known to Alice for odd r and Bob for even r, so Line 32 outputs correctly. The total communication is

$$\sum_{\tau=1}^{r} \mathcal{O}(n\tau \log\log n) = \mathcal{O}(nr^2 \log\log n).$$

For the bit version, simply propagate the bits backwards from the last layer, in parallel to Algorithm 2. This helps to chase through r extra layers. The algorithm is formalized in Algorithm 3. Observe that $r^2 + r = r(r+1)$ is always even, so the condition in Line 40 ensures that backward propagation starts with Bob as well. The additional communication is only $\mathcal{O}(n)$ per layer in Lines 41 and 48. As to correctness, the receiver of round i owns f_{r^2+r-i} and thus can derive $b_{r^2+r-i-1}$ by induction. After $r-1$ rounds of backward propagation, b_{r^2} is known to Alice for even r and Bob for odd r. Round r communicates it to the other party and thus ensures correct output, concluding the proof. □

C Lower Bound for IP∘PC with More Rounds

In this section, we prove the following strengthening of Theorem 7.1.

Theorem C.1. *For $n \geq 1$, $\eta \in (0, 1)$, and $r = \mathcal{O}(n^\eta)$, any r-round protocol for $\mathsf{IP}\circ\mathsf{PC}_{n,r}$ with constant error probability less than $1/2$ requires at least $\mathbf{R}^{(r)}(\mathsf{IP}\circ\mathsf{PC}_{n,r}) = \Omega_\eta(n \log n)$ communication.*

The only constraint of the analysis in Section 7 that limits $r = o(\log n)$ is the $\mathcal{O}(rn/\log n)$ loss in Φ_r, which can be traced back to Lemma 7.3. Therefore, the key to proving Theorem C.1 is a stronger alternative to Lemma 7.3, as formalized in Lemma C.2.

Lemma C.2. *For $\zeta \in (0, 1)$ and any distribution Q on $\mathcal{P}_n \times \mathcal{P}_n$, it holds that*

$$\mathop{\mathbb{E}}_{w,z\sim\{0,1\}^n} 2^{\mathbb{D}_2(Q^{(w,z)})} \leq 2^{\frac{\mathbb{D}_\infty(Q)}{\zeta \log n} + \mathcal{O}_\zeta(n^\zeta)},$$

where $Q^{(w,z)}$ is the distribution of $(aw, bz) \in \{0,1\}^{2n}$ when $(a, b) \sim Q$.

The proof of Lemma C.2 uses Hölder's inequality, which can be viewed as a generalization of Cauchy-Schwarz inequality.

Lemma C.3 (Hölder's inequality)**.** *For $p, q > 1$ such that $1/p + 1/q = 1$, and any two real-valued random variables X, Y, it holds that*

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} (\mathbb{E}|Y|^q)^{\frac{1}{q}}.$$

Proof of Lemma C.2. Arguments similar to the proofs of Lemma 7.3 and Claim 7.4 show that it suffices to upper bound

$$\mathop{\mathbb{E}}_{(a,b)\sim Q} 2^{\mathsf{cyc}(a)+\mathsf{cyc}(b)}.$$

To this end, for $p, q > 1$ such that $1/p + 1/q = 1$, we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{(a,b)\sim Q} 2^{\mathsf{cyc}(a)+\mathsf{cyc}(b)} &= \mathop{\mathbb{E}}_{a,b\sim\mathcal{P}_n} (n!)^2 \cdot Q(a,b) \cdot 2^{\mathsf{cyc}(a)+\mathsf{cyc}(b)} \\ &\le \left(\mathop{\mathbb{E}}_{a,b\sim\mathcal{P}_n} (n!)^{2p} \cdot Q(a,b)^p\right)^{\frac{1}{p}} \left(\mathop{\mathbb{E}}_{a,b\sim\mathcal{P}_n} 2^{q\cdot(\mathsf{cyc}(a)+\mathsf{cyc}(b))}\right)^{\frac{1}{q}}, \end{aligned} \tag{16}$$

by Hölder's inequality (Lemma C.3). Also observe that

$$\mathop{\mathbb{E}}_{a,b\sim\mathcal{P}_n} (n!)^2 \cdot Q(a,b) = (n!)^2 \cdot (n!)^{-2} = 1.$$

Therefore, we can get

$$\begin{aligned}\mathop{\mathbb{E}}_{a,b\sim\mathcal{P}_n} (n!)^{2p} \cdot Q(a,b)^p &\le \left(\max_{a,b\in\mathcal{P}_n} (n!)^2 \cdot Q(a,b)\right)^{p-1} \\ &\le \left(2^{2\log(n!)} \cdot 2^{-\mathbb{H}_\infty(Q)}\right)^{p-1} \\ &= 2^{(p-1)\cdot\mathbb{D}_\infty(Q)}. \end{aligned} \tag{17}$$

In the meantime, using the cycle decomposition representation of permutations, we also have

$$\begin{aligned}\mathop{\mathbb{E}}_{a,b\sim\mathcal{P}_n} 2^{q\cdot(\mathsf{cyc}(a)+\mathsf{cyc}(b))} &= \left(\mathop{\mathbb{E}}_{a\sim\mathcal{P}_n} 2^{q\cdot\mathsf{cyc}(a)}\right)^2 \\ &= \prod_{i\in[n]} \left(\frac{1}{i}\cdot 2^q + \left(1-\frac{1}{i}\right)\cdot 1\right)^2 \\ &= \prod_{i\in[n]} \left(1+\frac{2^q-1}{i}\right)^2 \\ &\le \exp\left(2\cdot\sum_{i\in[n]}\frac{2^q-1}{i}\right) \\ &= \exp(\mathcal{O}(2^q\cdot\ln n)). \end{aligned} \tag{18}$$

To conclude the proof, simply plug Equations (17) and (18), as well as $(p-1)/p = 1/q$ and $q = \zeta\log n$, into Equation (16). □

Now, we are ready to prove Theorem C.1.

Proof of Theorem C.1. Fix $\zeta = (1-\eta)/2 > 0$, so that $rn^\zeta = o(n)$, and sufficiently large n. Change the definition of Φ_i to be

$$\Phi_i = \sum_{j\in[i]} \left(\frac{\mathbb{D}_1(A_j, B_j \mid \gamma_{j-1}, m'_j)}{\zeta\log n} + Kn^\zeta\right),$$

for some sufficiently large constant K, and adjust the event $\mathcal{E}_i$ to be defined with respect to the updated threshold Φ_i. Then, conditioned on $\overline{\mathcal{E}}'$, we can get

$$\Phi_r \leq \frac{20L}{\zeta \log n} + Krn^{\zeta} \leq \frac{n}{10},$$

for $L = \zeta n/1000 \cdot \log n$. Therefore, it can be verified that Lemma 7.9 still holds, except that $2^{-n/\log n}$ becomes $2^{-n^{\zeta}/\log n}$ in Item 2. The same argument then concludes the proof. □