The Communication Complexity of Multiparty Set Disjointness under Product Distributions

Nachum Dershowitz
Tel Aviv University
Tel Aviv, Israel
nachumd@tau.ac.il

Rotem Oshman
Tel Aviv University
Tel Aviv, Israel
roshman@tau.ac.il

Tal Roth
Tel Aviv University
Tel Aviv, Israel
roth1@mail.tau.ac.il

ABSTRACT

In the multiparty number-in-hand set disjointness problem, we have k players, with private inputs $X_1, \ldots, X_k \subseteq [n]$. The players' goal is to check whether $\bigcap_{\ell=1}^{k} X_\ell = \emptyset$. It is known that in the shared blackboard model of communication, set disjointness requires $\Omega(n \log k + k)$ bits of communication, and in the coordinator model, it requires $\Omega(kn)$ bits. However, these two lower bounds require that the players' inputs can be highly correlated.

We study the communication complexity of multiparty set disjointness under product distributions, and ask whether the problem becomes significantly easier, as it is known to become in the two-party case. Our main result is a nearly-tight bound of $\tilde{\Theta}(n^{1-1/k} + k)$ for both the shared blackboard model and the coordinator model. This shows that in the shared blackboard model, as the number of players grows, having independent inputs helps less and less; but in the coordinator model, when k is very large, having independent inputs makes the problem much easier. Both our upper and our lower bounds use new ideas, as the original techniques developed for the two-party case do not scale to more than two players.

CCS CONCEPTS

• **Theory of computation → Communication complexity**.

KEYWORDS

communication complexity, set disjointness, product distributions

ACM Reference Format:
Nachum Dershowitz, Rotem Oshman, and Tal Roth. 2021. The Communication Complexity of Multiparty Set Disjointness under Product Distributions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21), June 21–25, 2021, Virtual, Italy*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3406325.3451106

1 INTRODUCTION

The set disjointness problem is a central problem in communication complexity, and lower bounds on the communication complexity of set disjointness have wide-ranging applications in circuit complexity, streaming algorithms, data structures, distributed computing, and other areas (the many variants of the problem and its applications have inspired several surveys, e.g., [11, 20]). Moreover, the search for lower bounds for set disjointness in various settings and models has led to the development of powerful combinatorial and information-theoretic techniques, which are now ubiquitous in communication complexity.

In its simplest form, the set disjointness problem asks two players, Alice and Bob, to determine whether their inputs sets, $X, Y \subseteq \{1, \ldots, n\}$ (resp.) intersect. The celebrated lower bound of [18, 19] shows that $\Omega(n)$ bits must be exchanged between the players, even using randomness and allowing for a constant error probability. However, before the linear lower bound was proven, [3] showed that under *product distributions* — that is, if we require that the players' inputs be independent of one another — the communication complexity of disjointness is only $\tilde{\Theta}(\sqrt{n})$ bits (with constant distributional error over the input distribution). In other words, set disjointness is significantly easier under product distributions than it is under arbitrary input distributions.

In recent years, the study of set disjointness has been extended to the multiparty setting, where we have k players with inputs $X^1, \ldots, X^k \subseteq [n]$, and our goal is to determine whether $\bigcap_{\ell \in [k]} X^k = \emptyset$. Here and throughout the paper, we study the *number-in-hand* model, where each input X_i is known only to player i (rather than the *number-on-forehead* model, where each input X_i is known to all the players *except* player i). A promise version of disjointness has important applications in streaming (see, e.g., [1, 4, 13]), and connections and applications in distributed computing and auction theory have led to the development of further lower bounds [6–9, 22]. In particular, it is known that in the *shared blackboard* model, where the players communicate by writing messages on a "shared blackboard" that all players can see, the communication complexity of k-party set disjointness is $\Theta(n \log k + k)$ [7]. On the other hand, in the *coordinator* model, where players can only interact by sending and receiving messages to a special party called the coordinator, the communication complexity rises to $\Theta(kn)$ [6]. These lower bounds imply communication lower bounds in the message-passing model, where a large number of servers compute on an input that is partitioned between them (see [2, 12, 14, 21, 22] and many others for examples of upper and lower bounds in this setting).

Our results. In this paper we study multiparty set disjointness under *product distributions*, where the players' inputs are independent of one another, and ask whether and by how much restricting to product distributions makes the problem easier. Recall that for unrestricted set disjointness, the shared blackboard model and the coordinator model display a gap of $\tilde{\Theta}(k)$ (in the shared blackboard the complexity is $\Theta(n \log k + k)$, but in the coordinator it is $\Theta(kn)$). Curiously, we show that under product distributions, as the number

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
STOC '21, June 21–25, 2021, Virtual, Italy
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8053-9/21/06…$15.00
https://doi.org/10.1145/3406325.3451106

of players increases, disjointness converges to the same complexity in both models: the communication complexity is $\tilde{\Theta}(n^{1-1/k}+k)$ for both. This means that in the shared blackboard, the more players we have, "the less useful" it is to restrict to product distributions — the problem becomes harder and harder as k increases, until for $k = \Omega(\log n)$ players it becomes as hard as it is for arbitrary distributions, up to polylogarithmic factors. On the other hand, in the coordinator model, the more players we have, the *more* useful it is to restrict to product distributions (assuming $k = \Omega(\log n)$): since the unrestricted complexity is $\Theta(kn)$ [6], the gap between the restricted and the unrestricted complexity grows with the number of players.

The formal statement of our results is as follows. Let $Disj^{\mu,\epsilon}_{n,k}$ denote the task of solving k-player disjointness over n elements, with distributional error at most ϵ over the input distribution μ.

Theorem 1. *For any constant $\epsilon \in (0,1)$, any $n, k \in \mathbb{N}$, and any product distribution μ over $\{0,1\}^{n\times k}$,*

(1) If $k < \log n$, then the expected communication complexity of $Disj^{\mu,\epsilon}_{n,k}$ is
- *$O(k + n^{1-1/k}\log n\lceil \log\log n/\log k\rceil)$ in the shared blackboard model, and*
- *$O(kn^{1-1/k}\log n\lceil \log\log n/\log k\rceil)$ in the coordinator model.*

(2) If $k \geq \log n$, then in both the shared blackboard and coordinator models, the expected communication complexity of $Disj^{\mu,\epsilon}_{n,k}$ is $O(k + n\log^2 n)$.

Our lower bound is proven for the shared black board model, but it also applies to the coordinator model, which the shared blackboard can simulate at no additional cost:

Theorem 2. *For a sufficiently small constant error $\epsilon \in (0,1)$, there exists a product distribution μ such that the expected communication complexity of $Disj^{\mu,\epsilon}_{n,k}$ is*

(1) $\Omega(k + n^{1-\frac{1}{k}}/k^2)$, if $k \leq \log n/6$; and
(2) $\Omega(k + n/\log^2 n)$, if $k > \log n/6$.

Applications. Beyond its intrinsic interest, our lower bound of $\tilde{\Omega}(n^{1-1/k})$ implies lower bounds for the communication complexity of various statistical and graph problems, when the input is partitioned between k servers, and each server's input is *independent* of the others'. Set disjointness reduces to many such problems, so lower bounds carry over. For example, using the reduction from [22],[1] we get a communication lower bound of $\tilde{\Omega}(n^{1-1/k})$ on graph connectivity with k servers, even in a graph where the presence or absence of each edge is independent of all the other edges (but the edges are not identically distributed). The ultimate conclusion is that this problem, and others like it, do not become trivial when the servers' inputs are independent.

Our techniques. Interestingly, it turns out that neither the upper bound nor the lower bound technique of [3] readily generalize to $k > 2$ players.[2] The *symmetrization* lower bound technique [16] also does not apply here. Therefore, our upper bound is based on different ideas than [3], and whereas [3] used a combinatorial lower bound argument (the corruption bound), our lower bound is information-theoretic.

Our lower bound also does not use the typical direct sum argument [10] that is often used in information-theoretic disjointness lower bounds (e.g., in [4–8, 13, 15]).

Next, we sketch the usual approach to information-theoretic disjointness lower bounds, and why it does not quite work for our setting.

Information-theoretic lower bounds for disjointness. Information-theoretic lower bounds in communication complexity measure the amount of information that a communication protocol must reveal about the inputs of the players. Since this information is always bounded by the length of the protocol's transcript, a lower bound on the information complexity of a function implies a communication lower bound as well. Working with information can be more convenient because of properties such as the chain rule — essentially, information is *additive*, and allows us to formalize statements such as "the information revealed about X, Y together is the sum of the information about X and the information about Y".

Many information-theoretic lower bounds for disjointness work only for protocols with small *worst-case error*: even though the lower bound works with a hard input distribution, we require the protocol to solve *every* input with low error, including inputs that are not in the support of the hard distribution. This approach is unsuitable for us, because we are interested in *distributional* error: we are given a product input distribution μ, and the protocol only needs to have low error probability over the average input drawn from μ. The textbook [17] gives a distributional version of the two-party lower bound, which forms the basis of our lower bound.

It is convenient to view the inputs X, Y to the players as the characteristic vectors of their sets. The lower bound of [17] works with the following input distribution μ:[3]

- We choose a random coordinate $i \in [n]$, and sample (X_i, Y_i) uniformly from $\{0,1\}^2$.
- For each remaining coordinate $j \neq i$, we sample (X_j, Y_j) uniformly from $\{(0,0),(1,0),(0,1)\}$.

Note that under μ we have $Disj(X,Y) = \neg(X_i \wedge Y_i)$, because no coordinate other than i can be in the intersection. The proof then shows that any protocol that sends $o(n)$ bits can typically only reveal $o(n)/n = o(1)$ bits about X_i, Y_i, and that $o(1)$ bits do not suffice to discover whether $X_i \wedge Y_i = 1$. Therefore, any protocol with communication $o(n)$ must have high error.[4]

The distribution μ given above is not a product distribution. When we work with a product distribution, we can no longer have the answer to disjointness depend only on a single coordinate which we as external observers know, but the protocol does not — this implies dependence between the inputs. Instead, a hard product distribution for disjointness is one where the answer is "spread

[1] In [22] the reduction is from a different problem, which [22] defined and analyzed, as [22] preceded the disjointness lower bound of [6]. However, the reductions of [22] are easily modified to work with disjointness instead.

[2] Nor do the techniques of [5], which interpolated between the $\Theta(n)$ unrestricted complexity and the $\Theta(\sqrt{n})$ complexity for product distributions, by showing that when the players' inputs have mutual information k between them, the communication complexity is $\Theta(\sqrt{n(k+1)})$. The upper bound in [5] is a clever modification of [3], and the lower bound is an adaptation of Razborov's lower bound [18].

[3] As does Razborov's original lower bound [18], using different constants.

[4] This is a highly informal description of the lower bound, and it glosses over many crucial details. We refer the interested reader to the excellent presentation in [17].

out" over all the coordinates: let μ' be the distribution where all the input bits $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are iid Bernoulli variables with probability $1/\sqrt{n}$ of being 1.[5] Now, each $i \in [n]$ has probability $1/n$ of being in the intersection, independent of the other coordinates. Together, we get a constant probability that there is an intersection.

The main source of technical difficulty in our lower bound is that under μ', it is not enough to argue that the protocol cannot reveal much information about a typical *single* coordinate $i \in [n]$. A single coordinate has probability only $1/n$ of being in the intersection! Instead, we must argue that even after observing the transcript of the protocol, there is a large *set* of coordinates that we have learned very little about, and which remain nearly independent of one another. We then carefully "add up" the tiny uncertainty that the protocol has about each individual coordinate, and prove that all together the protocol cannot distinguish the case where the input is disjoint from the case where it is intersecting.

Organization. The remainder of the paper is organized as follows. In Section 2 we introduce our notation and review some basic notions from information theory that are used in our lower bound proof. Next, we give our protocol for product distributions — actually, two separate protocols, depending on the number of players: in Section 3 we give a protocol for the case where $k \geq \log n$, and in Section 4 a protocol for the case where $k < \log n$. Finally, in Section 5, we give the proof of our $\tilde{\Omega}(n^{1-1/k})$ lower bound for disjointness under a product distribution.

2 PRELIMINARIES

Notation. We use boldface to denote random variables. Abusing notation, we sometimes conflate a random variable with the distribution from which it is drawn. If $\boldsymbol{A}$ is a random variable and $\mathcal{E}$ is an event, then $\boldsymbol{A}|_{\mathcal{E}}$ denotes the distribution of $\boldsymbol{A}$ conditioned on $\mathcal{E}$.

The input to the k players is denoted $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k \in \{0,1\}^n$, and we use $\boldsymbol{X}_i^\ell$ to denote the i-th coordinate of player ℓ's input. We also use $\boldsymbol{X}_{<i}^\ell$ to denote the vector consisting of coordinates $1, \ldots, i-1$ of player ℓ's input, and similarly for superscripts (e.g., $\boldsymbol{X}_i^{<\ell}$ denotes coordinate i in the inputs of players $1, \ldots, \ell-1$). We denote by $\boldsymbol{X}_i^{-\ell}$ the i-th coordinate of all players *except* for player ℓ, i.e. $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{\ell-1}, \boldsymbol{X}_i^{\ell+1}, \ldots, \boldsymbol{X}_i^k$. Finally, $\boldsymbol{X}_J$ denotes the coordinates $j \in J$ of all players' inputs.

For a vector $\boldsymbol{V}$, we denote by "$\boldsymbol{V} = \overline{1}$" the event that all of $\boldsymbol{V}$'s coordinates are 1.

The shared blackboard model. In this classical model of multiparty communication, we have k players, with private inputs $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k$. The players communicate by writing on a *shared blackboard* that all players can see. At any point in the protocol, the identity of the next player to write on the board is determined by the current contents of the board. When the protocol ends, its output is the last bit written on the board. We refer to the contents of the board as the *transcript* of the protocol, and use the random variable $\boldsymbol{M}$ to denote it.

[5] This is very nearly the distribution used in [3], except that there the inputs were two uniformly distributed sets of size $\sqrt{n}$. For our purposes it is nicer to avoid the dependencies between coordinates.

Set disjointness. The k-player set disjointness problem is formally defined as

$$\mathrm{Disj}_{n,k}(X^1, \ldots, X^k) := \neg \bigvee_{i=1}^{n} \bigwedge_{\ell=1}^{k} X_i^\ell.$$

However, since some of the literature takes the complement, i.e. $\bigvee_{i=1}^n \bigwedge_{\ell=1}^k X_i^\ell$, as the definition of disjointness, we avoid confusion by using "intersecting" ($\neg\emptyset$) and "non-intersecting" ($\emptyset$) to refer to the answer, instead of 0 and 1.

Background on information theory. To measure the amount of information a protocol reveals about its inputs, we use *mutual information*:

Definition 1 (Mutual information and conditional mutual information)**.** *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be random variables. The* mutual information *between $\boldsymbol{A}$ and $\boldsymbol{B}$ is given by*

$$\mathsf{I}(\boldsymbol{A}\,;\,\boldsymbol{B}) := H(\boldsymbol{A}) - H(\boldsymbol{A} \mid \boldsymbol{B}).$$

For random variables $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$, the conditional mutual information *between $\boldsymbol{A}$ and $\boldsymbol{B}$ given $\boldsymbol{C}$ is:*

$$\mathsf{I}(\boldsymbol{A}\,;\,\boldsymbol{B} \mid \boldsymbol{C}) := H(\boldsymbol{A} \mid \boldsymbol{C}) - H(\boldsymbol{A} \mid \boldsymbol{B}, \boldsymbol{C}).$$

Here, $H(\boldsymbol{X})$ is the entropy of $\boldsymbol{X}$, and $H(\boldsymbol{X} \mid \boldsymbol{Y})$ denotes the conditional entropy of $\boldsymbol{X}$ given $\boldsymbol{Y}$.

For an event $\mathcal{E}$, we sometimes abuse notation and denote $\mathsf{I}(\boldsymbol{A}; \boldsymbol{B} \mid \mathcal{E}) := \mathsf{I}(\boldsymbol{A}|_{\mathcal{E}}\,;\,\boldsymbol{B}|_{\mathcal{E}})$.

To measure the difference between two distributions, we use KL divergence:

Definition 2 (KL divergence)**.** *For two distributions μ,μ' supported over a set χ, the* KL divergence *of μ from μ' is:*

$$\mathrm{D}\left(\mu \,\middle\|\, \mu'\right) := \sum_{x \in \chi} \mu(x) \log \frac{\mu(x)}{\mu'(x)}.$$

We sometimes use $\mathrm{D}(p \,||\, p')$ as short-hand notation for the divergence between the Bernoulli distributions with probability p and p' (resp.) of being 1. Our lower bound also uses *Pinsker's inequality*, which asserts that for any $p, p' \in (0,1)$ we have

$$|p - p'| \leq \sqrt{\mathrm{D}(p \,||\, p') \ln 2/2}.$$

The mutual information between two variables $\boldsymbol{A}, \boldsymbol{B}$ is the *expected divergence* of $\boldsymbol{A}$'s posterior distribution given $\boldsymbol{B}$, from $\boldsymbol{A}$'s prior distribution (or vice-versa):

Property 1. *For any random variables $\boldsymbol{A}, \boldsymbol{B}$ we have* $\mathsf{I}(\boldsymbol{A}\,;\,\boldsymbol{B}) = \mathbb{E}_{b \sim \boldsymbol{B}}\left[\mathrm{D}\left(\boldsymbol{A}|_{\boldsymbol{B}=b} \,||\, \boldsymbol{A}\right)\right]$.

The following technical lemmas will be useful in our lower bound. The first bounds the "difference" between two Bernoulli random variables, in terms of their KL divergence:

Lemma 1. *Let $p \in (0, 1/3)$, $p' \in (0,1)$ and $\alpha \in (0, 1/2)$. If $\mathrm{D}(p \,||\, p') \leq p\alpha^2/40 \ln 2$, then $p'/p \in (1-\alpha, 1+\alpha)$.*

The second lemma follows from the first. It bounds the effect that conditioning on an event $\boldsymbol{B} = b$ can have on the distribution of a Bernoulli variable $\boldsymbol{A}$, in terms of the mutual information $\mathsf{I}(\boldsymbol{A}; \boldsymbol{B})$ and the probabilities $\Pr(\boldsymbol{A} = 1), \Pr(\boldsymbol{B} = b)$.

Lemma 2. *Fix $\alpha \in (0, 1/2)$. Let $\boldsymbol{A} \sim \mathrm{B}(p)$, where $p \in (0, 1/3)$, and let $\boldsymbol{B}$ be a random variable and $b \in \mathit{support}(\boldsymbol{B})$. Finally, let $p' \in (0, 1)$ such that $\boldsymbol{A}|_{\boldsymbol{B}=b} \sim \mathrm{B}(p')$. If $\mathrm{I}(\boldsymbol{A}\,;\,\boldsymbol{B}) \le \Pr(\boldsymbol{B} = b) \cdot p\alpha^2 40 \ln 2$, then $p'/p \in (1-\alpha, 1+\alpha)$.*

This last observation is useful in our lower bound:

Lemma 3. *Let $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$ be ranodm variables, then $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$ are independent iff for any $i \in [m]$, $\boldsymbol{R}_i$ is independent of $\boldsymbol{R}_{-i}$.*

3 PROTOCOL FOR $k \ge \log n$ PLAYERS

We prove that for any product distribution μ, the constant-error distributional communication complexity of disjointness under μ is $\tilde{O}(n^{1-1/k} + k)$. We give two separate protocols: in the current section we handle the case where $k \ge \log n$, and in the next section, the case $k < \log n$.

The "large k" case, $k \ge \log n$, is much easier than the "small k" case, because in this case the upper bound we are aiming for is $\tilde{O}(n^{1-1/\Omega(\log n)} + k) = \tilde{O}(n + k)$, i.e., slightly super-linear in the number of elements. Given $i \in [n]$, let $\boldsymbol{N}_i = \left|\left\{\ell \in [k] \;\middle|\; \boldsymbol{X}_i^\ell = 0\right\}\right|$ denote the number of players that have 0 in coordinate i. Also, let us say that coordinate i is *negligible* if $\Pr\left(i \in \bigcap_\ell \boldsymbol{X}^\ell\right) < \epsilon/n$. Our protocol for large k relies on the following simple observation, which holds for any product input distribution:[6]

Lemma 4. *If i is a non-negligible coordinate, then $\mathbb{E}[\boldsymbol{N}_i] \le \ln(n/\epsilon)$.*

Our protocol simply asks each player ℓ to announce the set of non-negligible coordinates i such that $\boldsymbol{X}_i^\ell = 0$. We then announce "intersecting" iff there is some non-negligible coordinate where no player announced it has zero. By Lemma 4, the expected communication cost of the protocol is $O(n \cdot \log^2 n)$, since the expected number of players that announce a given non-negligible coordinate is $O(\log n)$. The error of the protocol is bounded by ϵ, since we only err if there is some negligible coordinate in the intersection, and the probability that this occurs is at most $n \cdot (\epsilon/n) = \epsilon$.

4 UPPER BOUND FOR SMALL k

In this section we present a protocol for the case where $k < \log n$. We begin by showing how to handle input distributions that have a constant (or "small enough") expected intersection size, and then give a general protocol that can handle any product distribution.

4.1 Handling Distributions with a Small Expected Intersection

Recall that we are trying to show a protocol for $Disj_{n,k}$ with communication cost $\widetilde{O}(k + n^{1-1/k})$ in the shared blackboard model, and $\widetilde{O}(kn^{1-1/k})$ in the coordinator model. In this section we will show such a protocol for the simple case where the *expected intersection size* of the product distribution is $O(1)$. We refer to this protocol as our *base protocol*. (The protocol can actually handle a larger class of distributions, as we will see, but for the sake of intuition, let us think of distributions with a constant expected intersection size for now.)

[6] Specifically, we rely on the fact that for each $i \in [n]$, the bits $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^k$ are independent. This is true whenever the full inputs $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^\ell$ are independent.

Beyond just solving disjointness, the protocol computes the pointwise-AND of the inputs, and produces a *witness*, in the form of a string $W \in ([k] \cup \{\top\})^n$, such that

- If $\bigwedge_{\ell=1}^k \boldsymbol{X}_i^\ell = 0$, then W_i is the index of a player $\ell \in [k]$ such that $\boldsymbol{X}_i^\ell = 0$ (if there is more than one such player, one is chosen according to a deterministic rule described in the next section).
- If $\bigwedge_{\ell=1}^k \boldsymbol{X}_i^\ell = 1$, then $W_i = \top$.

The witness will be important for the general protocol we give in Section 4.2.

The base protocol is based on the following observation: if the expected intersection size is small, then for most elements $i \in [n]$, there is at least one player that is "not too likely" to have i in its input. This is because if all players are likely to have i in their input, then i is likely to be in the intersection, but we assumed that the expected intersection size is small. The base protocol partitions the elements $[n]$ into sets $I^1, \ldots, I^k$, such that in total, for all players $\ell \in [k]$, the expected sizes of $\boldsymbol{X}^\ell \cap I^\ell$ sum up to $O(n^{1-1/k})$. This partition is fixed in advance (before the inputs are seen).

We now describe the base protocol in the shared blackboard model; the protocol for the coordinator model is similar, and will be defined formally in the next section.

When the protocol begins, each player ℓ announces $\boldsymbol{X}^\ell \cap I^\ell$, and any element in $I^\ell \setminus \boldsymbol{X}^\ell$ (that is, any element of I^ℓ that is missing from player ℓ's input) is immediately ruled out, as we know that it cannot be in the intersection. For the remaining elements,

$$\boldsymbol{T} \coloneqq \bigcup_{\ell \in [k]} \boldsymbol{X}^\ell \cap I^\ell,$$

we go over the players in order; each player ℓ announces $\boldsymbol{T} \setminus \boldsymbol{X}^\ell$; we then remove these elements from $\boldsymbol{T}$, setting $\boldsymbol{T} \leftarrow \boldsymbol{T} \cap \boldsymbol{X}^\ell$. After going through all the players, if $\boldsymbol{T} \neq \emptyset$, we announce that the inputs are not disjoint, and otherwise we announce that they are disjoint.

Details of the protocol. For each $i \in [n]$, let $\boldsymbol{Z}_i$ be an indicator for an intersection in coordinate i:

$$\boldsymbol{Z}_i = \bigwedge_{\ell=1}^k \boldsymbol{X}_i^\ell.$$

Also, let S denote the expected intersection size:

$$S \coloneqq \mathbb{E}\left[\sum_{i=1}^n \boldsymbol{Z}_i\right] = \mathbb{E}\left[\left|\bigcap_{\ell=1}^k \boldsymbol{X}^\ell\right|\right].$$

We prove that there exists a partition of the elements to the players, such that in expectation, the players' actual inputs *do not contain* most of the elements assigned to them. This allows us to quickly rule out many elements, and focus on a small set of remaining candidates that might still be in the intersection.

Lemma 5. *There exists a partition $I^1, \ldots, I^k$ of $[n]$ such that*

$$\mathbb{E}\left[\sum_{\ell=1}^k \left|\boldsymbol{X}^\ell \cap I^\ell\right|\right] \le \sum_{i=1}^n \mathbb{E}[\boldsymbol{Z}_i]^{1/k} \le S^{1/k} n^{1-1/k}.$$

Proof. Because the inputs $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k$ are independent, the bits $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^k$ of each coordinate $i \in [n]$ are also independent, and

therefore

$$\mathbb{E}[Z_i] = \mathbb{E}\left[\bigwedge_{\ell=1}^{k} X_i^\ell\right] = \prod_{\ell=1}^{k} \mathbb{E}\left[X_i^\ell\right].$$

It follows that for each $i \in [n]$, there exists some $\ell \in [k]$ such that

$$\mathbb{E}\left[X_i^\ell\right] \le \mathbb{E}[Z_i]^{1/k}. \tag{1}$$

We use this observation to construct the partition: define, for each $\ell \in \{1, \ldots, k\}$,

$$I^\ell := \left\{ i \in [n] \;\middle|\; \mathbb{E}\left[X_i^\ell\right] \le \mathbb{E}[Z_i]^{1/k} \right\} \setminus \bigcup_{j=1}^{\ell} I^j.$$

This is indeed a partition of $[n]$, because of (1). Furthermore,

$$\begin{aligned}
\mathbb{E}\left[\sum_{\ell=1}^{k} \left|X^\ell \cap I^\ell\right|\right] &= \mathbb{E}\left[\sum_{\ell=1}^{k} \sum_{i \in I^\ell} X_i^\ell\right] = \sum_{\ell=1}^{k} \sum_{i \in I^\ell} \mathbb{E}\left[X_i^\ell\right] \\
&\le \sum_{\ell=1}^{k} \sum_{i \in I^\ell} \mathbb{E}[Z_i]^{1/k} \\
&= \sum_{i=1}^{n} \mathbb{E}[Z_i]^{1/k} \qquad (I^1, \ldots, I^k \text{ is a partition}) \\
&\le \left(\sum_{i=1}^{n} \mathbb{E}[Z_i]\right)^{1/k} \left(\sum_{i=1}^{n} 1\right)^{1-1/k} \\
&\qquad \text{(Hölder's inequality)} \\
&= S^{1/k} n^{1-1/k}. \qquad \square
\end{aligned}$$

We are now ready to describe the base protocol. As we said, in addition to solving set disjointness, the protocol produces a *witness*: a string $W \in ([k] \cup \{\top\})^n$ that indicates, for each coordinate that is not in the intersection, the index of some player that has 0 in this coordinate. The witness is a deterministic function of the transcript of the protocol.

In the shared blackboard model, the protocol proceeds as follows.

(1) Each player $\ell \in [k]$ announces $X^\ell \cap I^\ell$. Let
$$T^0 := \bigcup_{\ell \in [k]} \left(X^\ell \cap I^\ell\right)$$
be the set written on the board. Following this step, only elements in T^0 remain candidates for being in the intersection.
(2) We go over the players in order, $\ell = 1, \ldots, k$: player ℓ announces $T^{\ell-1} \setminus X^\ell$, and all players update $T^\ell := T^{\ell-1} \cap X^\ell$.
(3) We announce that the intersection is empty iff $T^k = \emptyset$.

The witness W is defined as follows: for each $i \in [n]$,

- If $i \notin T^0$, then we set W_i to the index ℓ such that $i \in I^\ell$.
- If $i \in T^0$, then since $T^k \subseteq T^{k-1} \subseteq \ldots \subseteq T^0$, there are two cases:
  - If $i \in T^k$ then we set $W_i = \top$,
  - If $i \notin T^k$ then there is exactly one index $\ell \in [k]$ such that $i \in T^{\ell-1} \setminus T^\ell$, and we set W_i to this index.

In the coordinator model, the protocol proceeds as follows.

(1) Each player $\ell \in [k]$ sends $X^\ell \cap I^\ell$ to the coordinator.
(2) The coordinator sends
$$T := \bigcup_{\ell \in [k]} \left(X^\ell \cap I^\ell\right)$$
to all players. Following this step, only elements in T remain candidates for being in the intersection.
(3) Each player $\ell \in [k]$ sends $X^\ell \cap T$ to the coordinator.
(4) The coordinator computes a witness $W = \{W_i\}_{i \in [n]}$ as follows, and sends it to all the players: for each $i \in [n]$, denote by ℓ the index such that $i \in I^\ell$. Then:
   - If $i \notin T$, then we set $W_i = \ell$.
   - If $i \in T$, then there are two cases:
     - If for all $\ell' \in [k] \setminus \{\ell\}$, we have $i \in X^{\ell'}$, then we set $W_i = \top$,
     - Otherwise, i.e., if there exists $\ell' \in [k] \setminus \{\ell\}$ such that $i \notin X^{\ell'}$, then we set W_i to be the minimal such index ℓ'.
(5) The players announce that there is an intersection iff W contains a $\top$.

Lemma 6. *The base protocol always solves disjointness correctly and produces a proper witness. Its expected communication cost is*

$$O\left(k + \left(\sum_{i=1}^{n} \mathbb{E}[Z_i]^{1/k}\right) \log n\right) \quad = \quad O(k \;+\; S^{1/k} n^{1-1/k} \log n)$$

in the shared blackboard model, and in the coordinator model the expected cost is

$$O\left(\left(\sum_{i=1}^{n} \mathbb{E}[Z_i]^{1/k}\right) k(\log n + \log k)\right) = O(S^{1/k} n^{1-1/k} k(\log n + \log k)).$$

Proof. We prove the claim for the shared blackboard; the analysis for the coordinator model is similar.

Correctness. The protocol outputs "intersecting" iff some coordinate of W is $\top$. Thus, it is sufficient to prove that the witness W is a proper witness, that is, W_i is the index of some player ℓ with $X_i^\ell = 0$ if there is such a player, and $\top$ otherwise. The fact that the witness is proper is evident from the protocol: for each coordinate i, if $i \notin T^0$ and $i \in I^\ell$, then we have $i \notin X^\ell$, so setting $W_i = \ell$ is proper. Otherwise, if $i \in T^{\ell-1} \setminus T^\ell$, then $i \notin X^\ell$, because $T^\ell = T^{\ell-1} \cap X^\ell$. And finally, if $i \in T^k$, then we have $i \in X^\ell$ for all $\ell \in [k]$, and accordingly we set $W_i = \top$.

Communication cost. In the first step of the protocol, the set written on the board is $\bigcup_{\ell \in [k]} X^\ell \cap I^\ell$, and its expected size is $O\left(\sum_{i=1}^{n} \mathbb{E}[Z_i]^{1/k}\right) = O(S^{1/k} n^{1-1/k})$ by Lemma 5. Therefore, the expected number of bits on the board in this step is

$$O\left(k + \left(\sum_{i=1}^{n} \mathbb{E}[Z_i]^{1/k}\right) \log n\right) = O(k + S^{1/k} n^{1-1/k} \log n).$$

In the second step, each coordinate in T^0 is written at most once, so the expected cost is again

$$O\left(k + \left(\sum_{i=1}^{n} \mathbb{E}[Z_i]^{1/k}\right) \log n\right) = O(k + S^{1/k} n^{1-1/k} \log n). \qquad \square$$

4.2 The General Protocol

The base protocol handles distributions where the expected intersection size is small; now suppose we have an input distribution where the intersection is large. If the probability that the inputs

intersect is close to 1, we can simply guess that the inputs do intersect — and risk erring, but only with small probability. Thus, assume that

$$\Pr\left[\bigcap_{\ell\in[k]} \boldsymbol{X}^\ell \neq \emptyset\right] \leq 1-\epsilon$$

for some $\epsilon \in (0,1)$. Together, the fact that the expected intersection size is large, while the probability of an intersection is bounded away from 1, imply that the indicators $\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_n$ of an intersection in the individual coordinates must be *correlated*. We would like to exploit this correlation to reduce the general case to the base case (where we have a constant-sized intersection).

The reduction takes a recursive form: in each step, we find a maximal set $I \subseteq [n]$ of "negatively-correlated" coordinates (not in the usual sense of negative correlation, but rather in a sense we define below). We would naïvely like to have the following properties:

Property 2. *The expected intersection size* inside I *is* constant*:*

$$\mathbb{E}\left[\sum_{i\in I} \boldsymbol{Z}_i\right] = O(1).$$

Intuitively, this property should hold because the coordinates in I are negatively correlated with one another, so if one of them is in the intersection, the others tend not to be. Therefore, we can use the base protocol to check whether there is an intersection inside I, and if there is, we halt.

Property 3. *The remaining coordinates, $[n]\setminus I$, are "positively correlated" with the coordinates in I (otherwise we would add them to I, as I is a* maximal *set of negatively-correlated coordinates).*

This means that conditioned on the event that there is no intersection in I, the expected intersection size in $[n]\setminus I$ is much smaller than the prior. We recurse on the set $[n]\setminus I$.

As it turns out, the above plan yields a protocol with $\approx \log n$ iterations, each with an expected communication cost of $O(k + n^{1-1/k}\log n)$. We would like to reduce the number of iterations to $\approx \log\log n/\log k$ (without increasing the expected communication cost per iteration), as the resulting protocol will have both better round complexity as well as better overall communication cost. For this purpose, we *weaken* our first requirement to:

Property 4. *The set I satisfies:*

$$\sum_{i=1}^{n} \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k} = O\left(n^{1-1/k}\right).$$

Note that by Hölder's inequality this property is indeed weaker than Property 2, so intuitively, it should hold for the same reasons. Moreover, by Lemma 6, the weaker requirement still guarantees an expected communication cost of $O(k + n^{1-1/k}\log n)$ per iteration. Weakning Property 2 will allow us to add more indices to the set I at every iteration, so the protocol will require fewer iterations to complete; in the next sections will show that $\log\log n/\log k$ iterations are enough.

Our protocol works only for product distributions; in order to recurse on the set $[n]\setminus I$, we must ensure that the players' inputs remain independent conditioned on what they have seen so far. For the shared blackboard, this is easy — all players see the full transcript of the protocol on the shared blackboard, and it is well-known that conditioning on the transcript of a protocol does not create dependence between the inputs. In the coordinator model, however, the players do not see the entire transcript — only the coordinator does; each player sees only the messages the coordinator sent it, and these messages can create dependencies. This is where the *witness* produced by the base protocol comes in: the coordinator sends to all players the witness $\boldsymbol{W}$ that it computed from their messages, and we prove that conditioned on the witness, the players' remaining inputs are independent.

We note that while the base protocol is described in Section 4.1 as operating on the universe $[n]$, this is merely for the sake of convenience. In the sequel, when we call the base protocol, we let $I \subseteq [n]$ be the set of coordinates on which we want to solve disjointness using the base protocol.

4.2.1 Negatively-Correlated Coordinates. Recall that a pair of real-valued random variables $\boldsymbol{A},\boldsymbol{B}$ are said to be *negatively correlated* if

$$\mathrm{Cov}(A,B) = \mathbb{E}\left[A\cdot B\right] - \mathbb{E}\left[A\right]\mathbb{E}\left[B\right] \leq 0.$$

This definition is easily extended to a larger number of random variables, $\boldsymbol{R}_1,\ldots,\boldsymbol{R}_m$, by requiring that

$$\mathbb{E}\left[\prod_{i=1}^{m} \boldsymbol{R}_i\right] \leq \prod_{i=1}^{m} \mathbb{E}\left[\boldsymbol{R}_i\right].$$

We will generalize this notion further, by using a *weighted* version of the last inequality. For the sake of concreteness, we restrict attention to Bernoulli random variables, but the definition is easily stated for real-valued variables as well.

Definition 3 (φ-negatively-correlated indicators)**.** *Let $\varphi : [0,1] \to [0,1]$ be a function. The Bernoulli random variables $\boldsymbol{B}_1,\ldots,\boldsymbol{B}_m$ are said to be* φ-negatively correlated *if:*

$$\mathbb{E}\left[\prod_{i=1}^{m} (1-\boldsymbol{B}_i)\right] \leq \prod_{i=1}^{m} \left(1-\varphi\left(\mathbb{E}\left[\boldsymbol{B}_i\right]\right)\right).$$

Note that in the special case where $m=2$ and φ is the identity function, the new definition coincides with the standard definition of negative correlation for two variables $1-\boldsymbol{B}_1, 1-\boldsymbol{B}_2$. The reason we take the complements $(1-\boldsymbol{B}_i)$ instead of the indicators themselves $(\boldsymbol{B}_i)$ is that we are actually interested in the event of *not* having an intersection in a given coordinate, and the indicator for this event is $1-\boldsymbol{Z}_i$ (for coordinate i).[7]

The following two properties of φ-negatively-correlated indicators are key to our protocol. First, we can relate the expectations of these variables to the probability that none of them take the value 1, as follows:

Lemma 7. *If $\boldsymbol{B}_1,\ldots,\boldsymbol{B}_m$ are φ-negatively-correlated, then*

$$\Pr\left(\bigwedge_{i=1}^{m} (\boldsymbol{B}_i = 0)\right) \leq e^{-\sum_{i=1}^{m}\varphi(\mathbb{E}[\boldsymbol{B}_i])}.$$

[7] For two random variables $\boldsymbol{B}_1,\boldsymbol{B}_2$, we have $\mathrm{Cov}(\boldsymbol{B}_1,\boldsymbol{B}_2) = \mathrm{Cov}(1-\boldsymbol{B}_1, 1-\boldsymbol{B}_2)$, and hence $\boldsymbol{B}_1,\boldsymbol{B}_2$ are negatively correlated iff $1-\boldsymbol{B}_1, 1-\boldsymbol{B}_2$ are negatively correlated

Proof. We can write

$$\Pr\left(\bigwedge_{i=1}^{m}(\boldsymbol{B}_i=0)\right)=\Pr\left(\prod_{i=1}^{m}(1-\boldsymbol{B}_i)=1\right)=\mathbb{E}\left[\prod_{i=1}^{m}(1-\boldsymbol{B}_i)\right].$$

Since $\boldsymbol{B}_1,\ldots,\boldsymbol{B}_m$ are φ-negatively-correlated, and using the fact that $1-x\le e^{-x}$ for all $x\ge 0$, we have

$$\mathbb{E}\left[\prod_{i=1}^{m}(1-\boldsymbol{B}_i)\right]\le\prod_{i=1}^{m}\left(1-\varphi\left(\mathbb{E}\left[\boldsymbol{B}_i\right]\right)\right)\le\prod_{i=1}^{m}e^{-\varphi(\mathbb{E}[\boldsymbol{B}_i])}$$
$$=e^{-\sum_{i=1}^{m}\varphi(\mathbb{E}[\boldsymbol{B}_i])}.$$ □

The next property asserts that if we have a maximal subset I of φ-negatively-correlated indicators out of some larger set of indicators, then conditioned on all indicators in I taking the value zero, we can bound the expected sum of the remaining indicators:

Lemma 8. *Let $\boldsymbol{B}_1,\ldots,\boldsymbol{B}_m$ be Bernoulli random variables, and let $I\subseteq[m]$ be a maximal subset such that $\{\boldsymbol{B}_i\}_{i\in I}$ are φ-negatively-correlated. Let $J:=[m]\setminus I$. If $\Pr\left(\boldsymbol{B}_I=\overline{0}\right)>0$, then*

$$\mathbb{E}\left[\sum_{j\in J}\boldsymbol{B}_j\,\middle|\,\boldsymbol{B}_I=\overline{0}\right]\le\sum_{j\in J}\varphi(\mathbb{E}\left[\boldsymbol{B}_j\right]).$$

Proof. By linearity of expectation, it suffices to show that for each $j\in J$,

$$\mathbb{E}\left[\boldsymbol{B}_j\,\middle|\,\boldsymbol{B}_I=\overline{0}\right]\le\varphi(\mathbb{E}\left[\boldsymbol{B}_j\right]).$$

To that end, let $j\in J$. Since $I\subseteq[m]$ is maximal and $j\notin I$, the indicators $\{\boldsymbol{B}_i\}_{i\in I}\cup\{\boldsymbol{B}_j\}$ are *not* φ-negatively-correlated, so

$$\mathbb{E}\left[\left(1-\boldsymbol{B}_j\right)\cdot\prod_{i\in I}(1-\boldsymbol{B}_i)\right]>\left(1-\varphi(\mathbb{E}\left[\boldsymbol{B}_j\right])\right)\cdot\prod_{i\in I}\left(1-\varphi(\mathbb{E}\left[\boldsymbol{B}_i\right])\right). \quad (2)$$

For the left-hand side, we can write

$$\begin{aligned}
&\mathbb{E}\left[\left(1-\boldsymbol{B}_j\right)\cdot\prod_{i\in I}(1-\boldsymbol{B}_i)\right]\\
&=\Pr\left[\left(\boldsymbol{B}_j=0\right)\wedge\prod_{i\in I}(1-\boldsymbol{B}_i)=1\right]\\
&=\Pr\left(\boldsymbol{B}_j=0\,\middle|\,\boldsymbol{B}_I=\overline{0}\right)\Pr\left(\prod_{i\in I}(1-\boldsymbol{B}_i)=1\right)\\
&=\Pr\left(\boldsymbol{B}_j=0\,\middle|\,\boldsymbol{B}_I=\overline{0}\right)\mathbb{E}\left[\prod_{i\in I}(1-\boldsymbol{B}_i)\right]\\
&\le\Pr\left(\boldsymbol{B}_j=0\,\middle|\,\boldsymbol{B}_I=\overline{0}\right)\prod_{i\in I}\left(1-\varphi(\mathbb{E}[\boldsymbol{B}_i])\right),
\end{aligned}$$

where the last step used the fact that $\{\boldsymbol{B}_i\}_{i\in I}$ are φ-negatively-correlated. Together with (2), we obtain

$$\left(1-\varphi(\mathbb{E}\left[\boldsymbol{B}_j\right])\right)\cdot\prod_{i\in I}\left(1-\varphi(\mathbb{E}\left[\boldsymbol{B}_i\right])\right)>\Pr\left(\boldsymbol{B}_j=0\,\middle|\,\boldsymbol{B}_I=\overline{0}\right)\cdot\prod_{i\in I}\left(1-\varphi(\mathbb{E}[\boldsymbol{B}_i])\right). \quad (3)$$

Since the range of φ is $[0,1]$, the term $\prod_{i\in I}\left(1-\varphi(\mathbb{E}\left[\boldsymbol{B}_i\right])\right)$ is non-negative, and in fact it must be positive (otherwise (3) cannot hold). Dividing both sides of (3) by this term yields

$$1-\varphi(\mathbb{E}\left[\boldsymbol{B}_j\right])>\Pr\left(\boldsymbol{B}_j=0\,\middle|\,\boldsymbol{B}_I=\overline{0}\right)=1-\mathbb{E}\left[\boldsymbol{B}_j\,\middle|\,\boldsymbol{B}_I=\overline{0}\right],$$

and the claim follows. □

4.2.2 Partitioning the Coordinates. We define a concrete weight function $\varphi:[0,1]\to[0,1]$ and a partition $[n]=I\cup J$ of the coordinates, as follows:

$$\varphi(x):=\frac{x^{1/k}}{n^{1-1/k}}, \quad (4)$$

and let $I\subseteq[n]$ be a maximal set of indices such that $\{\boldsymbol{Z}_i\}_{i\in I}$ is φ-negatively-correlated. As above, let $J:=[n]\setminus I$.

Based on the properties we showed above for φ-negatively-correlated indicators, we obtain the following properties of the partition $[n]=I\cup J$.

Lemma 9. *For all $\epsilon\in(0,1)$, if $\Pr\left(\bigcap_{\ell=1}^{k}\boldsymbol{X}_I^{\ell}=\emptyset\right)>\epsilon$, then*

$$\sum_{i\in I}\mathbb{E}[\boldsymbol{Z}_i]^{1/k}\le\ln\left(\frac{1}{\epsilon}\right)n^{1-1/k}.$$

Proof. By Lemma 7 and our definition of φ,

$$e^{-\sum_{i\in I}\mathbb{E}[\boldsymbol{Z}_i]^{1/k}/n^{1-1/k}}=e^{-\sum_{i\in I}\varphi(\mathbb{E}[\boldsymbol{Z}_i])}\ge\Pr\left(\bigcap_{\ell=1}^{k}\boldsymbol{X}_I^{\ell}=\emptyset\right)>\epsilon.$$

Taking the natural logarithm and re-arranging yields the claim. □

Lemma 10. *Conditioned on having no intersection in I, the expected intersection size in J is bounded by*

$$\mathbb{E}\left[\sum_{j\in J}\boldsymbol{Z}_j\,\middle|\,\boldsymbol{Z}_I=\overline{0}\right]\le\left(\mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{Z}_i\right]\right)^{1/k}.$$

Proof. Using Lemma 8, we obtain

$$\begin{aligned}
\mathbb{E}\left[\sum_{j\in J}\boldsymbol{Z}_j\,\middle|\,\boldsymbol{Z}_I=\overline{0}\right]&\le\sum_{j\in J}\varphi(\mathbb{E}\left[\boldsymbol{Z}_j\right])\le\sum_{i=1}^{n}\frac{\mathbb{E}\left[\boldsymbol{Z}_j\right]^{1/k}}{n^{1-1/k}}\\
&\le\frac{1}{n^{1-1/k}}\left(\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i\right]\right)^{1/k}\left(\sum_{i=1}^{n}1\right)^{1-1/k}\quad\text{(by Hölder's inequality)}\\
&=\frac{1}{n^{1-1/k}}\left(\mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{Z}_i\right]\right)^{1/k}n^{1-1/k}=\left(\mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{Z}_i\right]\right)^{1/k}.
\end{aligned}$$ □

4.2.3 Preserving Independence. Let $I=\{i_1,\ldots,i_m\}\subseteq[n]$ be the set of coordinates on which we call the base protocol, let $I_1,\ldots,I_k\subseteq I$ be the partition computed by the base protocol, and let $\boldsymbol{W}_I\in([k]\cup\{\top\})^m$ be the witness returned (as defined in Section 4.1). Finally, let $J=[n]\setminus I$ be the set on which we recurse if we do not find an intersection inside I. We prove that conditioned on the witness $\boldsymbol{W}_I$, the players' remaining inputs are independent:

Lemma 11. *For each concrete witness $w\in\mathit{support}(\boldsymbol{W}_I)$, the random variables $\boldsymbol{X}_J^1,\ldots,\boldsymbol{X}_J^{\ell}$ are independent conditioned on the event $\boldsymbol{W}_I=w$.*

Proof. For this proof it is convenient to use the language of mutual information. To prove that the inputs are independent, by Lemma 3, it suffices to show that for each $\ell \in [k]$ we have

$$\mathsf{I}\left(\boldsymbol{X}_J^\ell \,;\, \boldsymbol{X}_J^{-\ell} \,\middle|\, \boldsymbol{W}_I = w\right) = 0.$$

For every $w \in ([k] \cup \{\top\})^m$, the event $\{\boldsymbol{W}_I = w\}$ is equivalent to some partial assignment to the random variables $\boldsymbol{X}_I^\ell$ and $\boldsymbol{X}_I^{-\ell}$. Let us denote by $\boldsymbol{Y}_I^\ell, \boldsymbol{Y}_I^{-\ell}$ the bits of $\boldsymbol{X}_I^\ell, \boldsymbol{X}_I^{-\ell}$ that are fixed under the event $\{\boldsymbol{W}_I = w\}$, and denote by $\overline{a}, \overline{b}$ their respective assignments, i.e. under this notation, the event $\{\boldsymbol{W}_I = w\}$ is equivalent to the event $\left\{\boldsymbol{Y}_I^\ell = \overline{a} \wedge \boldsymbol{Y}_I^{-\ell} = \overline{b}\right\}$. Hence it suffices to show that:

$$\mathsf{I}\left(\boldsymbol{X}_J^\ell \,;\, \boldsymbol{X}_J^{-\ell} \,\middle|\, \boldsymbol{Y}_I^\ell = \overline{a} \wedge \boldsymbol{Y}_I^{-\ell} = \overline{b}\right) = 0.$$

In fact, since mutual information is non-negative, and since by definition $\mathsf{I}(\boldsymbol{A}; \boldsymbol{B} | \boldsymbol{C}) = \mathbb{E}_{c \sim \boldsymbol{C}} \mathsf{I}(\boldsymbol{A}; \boldsymbol{B} | \boldsymbol{C} = c)$, it suffices to show that:

$$\mathsf{I}\left(\boldsymbol{X}_J^\ell \,;\, \boldsymbol{X}_J^{-\ell} \,\middle|\, \boldsymbol{Y}_I^\ell, \boldsymbol{Y}_I^{-\ell}\right) = 0,$$

but observe that this holds, since

$$\begin{aligned}\mathsf{I}\left(\boldsymbol{X}_J^\ell \,;\, \boldsymbol{X}_J^{-\ell} \,\middle|\, \boldsymbol{Y}_I^\ell, \boldsymbol{Y}_I^{-\ell}\right) &\le \mathsf{I}\left(\boldsymbol{X}_J^\ell, \boldsymbol{Y}_I^\ell \,;\, \boldsymbol{X}_J^{-\ell}, \boldsymbol{Y}_I^{-\ell}\right) \\ &\le \mathsf{I}\left(\boldsymbol{X}^\ell \,;\, \boldsymbol{X}^{-\ell}\right) \\ &= 0.\end{aligned}$$

□

When we described the base protocol for the coordinator model, we said that the coordinator first sends the set $\boldsymbol{T} = \bigcup_\ell \left(\boldsymbol{X}^\ell \cap I^\ell\right)$ to all players, and upon receiving their response, the coordinator sends the witness $\boldsymbol{W}$. However, for convenience, in the sequel we represent the trascript seen by an individual player by simply the witness $\boldsymbol{W}$, omitting the set $\boldsymbol{T}$, except when we discuss the communication cost of the protocol. For all purposes other than the communication cost, the set $\boldsymbol{T}$ may be omitted, because it can be computed from $\boldsymbol{W}$: for each $i \in [n]$ we have $i \in \boldsymbol{T}$ iff $\boldsymbol{W}_i$ is *not* the index of the player $\ell \in [k]$ such that $i \in I^\ell$. The same holds for the shared blackboard model.

4.2.4 The Protocol. We are now ready to describe our full protocol. Throughout the protocol, all players keep track of the following:

- The set $U \subseteq [n]$ of coordinates that have not been ruled out as being in the intersection; initially, $U = [n]$.
- A witness $W \in ([k] \cup \{\top\})^{[n] \setminus U}$ for the coordinates we have already handled. We represent the witness as the concatenation of the individual witnesses returned by calls to the base protocol. The witness is initially empty.

Our protocol executes in several iterations, calling the base protocol at most once at each iteration, and may stop at the end of an iteration under certain conditions. Let μ be the input distribution (which is assumed to be known in advance to all participants). Assuming the protocol reaches the r-th iteration, we denote by $\boldsymbol{W}_r$ the witness returned by the base protocol at that iteration, and otherwise we define $\boldsymbol{W}_r$ to be the empty string. We further denote by $\boldsymbol{W}_{\le r}$ the concatenation of the witnesses $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_r$, and by $\mu|W$ the distribution of the inputs conditioned on the event $\boldsymbol{W}_{\le r} = W$. Finally, if $W = W_1 \circ \ldots \circ W_r$ is the concatenation of r witnesses, we denote $|W| \coloneqq r$.

The protocol executes as follows:

(1) Repeat for $N \coloneqq \left\lceil \frac{\log \log n}{\log k} \right\rceil$ iterations:
  (1a) The players compute (without communication):
    (i) For each remaining coordinate $i \in U$, the value
    $$\varphi_i \coloneqq \varphi\left(\mathop{\mathbb{E}}_{\mu|W}[\boldsymbol{Z}_i]\right).$$
    (ii) A maximal subset $I \subseteq U$ such that (conditioned on the event $\boldsymbol{W}_{\le r} = W$) $\{\boldsymbol{Z}_i\}_{i \in I}$ are φ-negatively-correlated. (If there is more than one such subset, the players choose one using some predetermined tie-breaking mechanism.)
    Let $J \coloneqq U \setminus I$.
  (1b) If $\Pr_{\mu|W}\left(\bigcap_{\ell=1}^k \boldsymbol{X}_I^\ell = \emptyset\right) \le \epsilon$, the players output "intersecting" and halt the protocol.
  (1c) Otherwise, the players run the base protocol on the coordinates in I, that is, on $X_I^1, \ldots, X_I^k$.
  (1d) The players examine the witness w returned by the base protocol: if it indicates that there is an intersection, they announce "intersecting" and halt. Otherwise, the players update the universe and the current witness by setting $U \leftarrow J$, and $W \leftarrow W \circ w$ (where $\circ$ stands for concatenation).
(2) Finally, the players run the base protocol on $X_U^1, \ldots, X_U^k$ and output its answer.

We point out that at the end of each iteration, when the players update their witness W, this also has the effect of updating the distribution they use to compute the various expectations and probabilities during the protocol, because the players use $\mu|W$ for this purpose.

4.2.5 Analysis. Next, we analyze the expected communication cost and the error of the protocol. In the sequel, we typically use the subscript $r \in [N]$ to indicate values associated with iteration r in Step 1 of the protocol. For convenience, we sometimes refer to step 2 as iteration $N + 1$.

Expected communication cost. We analyze the cost of the protocol in the shared blackboard model; the analysis for the coordinator model is similar.

We define the following random variables:

- For each $r = 1, \ldots, N$, let $\boldsymbol{C}_r$ denote the number of bits sent during the r-th iteration in Step 1 of the protocol, or 0 if we do not reach the r-th iteration.
- Let $\boldsymbol{C}_{N+1}$ be the number of bits sent in Step 2 of the protocol, or 0 if we do not reach Step 2.
- Let $\boldsymbol{I}_r \subseteq [n]$ be the set of coordinates on which the base protocol is called in iteration r, or the empty set if we do not execute iteration r.
- Let $\boldsymbol{U}_r \subseteq [n]$ be the set of coordinates that have not yet ruled out as being in the intersection at the beginning of iteration r (i.e., the value of the variable U of the protocol at the beginning of iteration r).
- Let $\boldsymbol{J}_r \coloneqq \boldsymbol{U}_r \setminus \boldsymbol{I}_r$.
- Let $\boldsymbol{R}$ be the number of iterations completed in Step 1 of the protocol before halting, or $N + 1$ if the protocol reached Step 2.

- Finally, it will also be convenient sometimes to use the notation $\boldsymbol{W}_{<r}$ to denote $\boldsymbol{W}_{\leq r-1}$.

Since our protocol is deterministic, the witness $\boldsymbol{W}_{\leq r}$ for a given iteration r is a deterministic function of the inputs (and the input distribution). We can also "read off" the global variables $\boldsymbol{U}_r, \boldsymbol{I}_r, \boldsymbol{W}_r$ from $\boldsymbol{W}_{\leq r}$, and determine exactly when the protocol halted. Let $\mathcal{W}_r$ be the set of witnesses $W_{<r}$ of length $r-1$, which cause the protocol to reach the r-th iteration, and also satisfy

$$\Pr_{\mu|W_{<r}}\left(\bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell}_{\boldsymbol{I}_r} = \emptyset\right) > \epsilon.$$

In other words, $\mathcal{W}_r$ is the set of witnesses under which the protocol reaches iteration r and calls the base protocol in this iteration.

We begin by bounding the expected communication cost of the individual iterations in Step 1:

Lemma 12. *In the shared blackboard model, for each $r = 1, \ldots, N$ we have*

$$\mathop{\mathbb{E}}_{\mu}[\boldsymbol{C}_r] = O\left(k + n^{1-1/k}\log n\right).$$

Proof. The only communication in a given iteration results from calling the base protocol on the sets $\boldsymbol{X}^1_{\boldsymbol{I}_r}, \ldots, \boldsymbol{X}^k_{\boldsymbol{I}_r}$, and this only occurs if we reach iteration r and do not halt in Step 1b (where, if the intersection probability is too high, we halt and guess that the inputs are intersecting). Thus, we need only consider the witnesses in $\mathcal{W}_r$. For each such $W_{<r} \in \mathcal{W}_r$, Lemma 9 asserts that

$$\sum_{i \in \boldsymbol{I}_r} \mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{Z}_i]^{1/k} \leq \ln\left(\frac{1}{\epsilon}\right) n^{1-1/k}.$$

Plugging this bound into Lemma 6, we obtain

$$\mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{C}_r] = O\left(k + n^{1-1/k}\log n\right).$$

Since this holds point-wise for any witness $W_{<r} \in \mathcal{W}_r$, and since $\boldsymbol{C}_r = 0$ whenever $\boldsymbol{W}_{<r} \notin \mathcal{W}_r$, all together we have

$$\begin{aligned}\mathop{\mathbb{E}}_{\mu}[\boldsymbol{C}_r] &= \mathop{\mathbb{E}}_{\mu}[\boldsymbol{C}_r \mid \mathcal{W}_r]\Pr_{\mu}(\mathcal{W}_r) + 0 \cdot \Pr_{\mu}(\neg\mathcal{W}_r)\\ &= O\left(k + n^{1-1/k}\log n\right).\end{aligned}$$

□

Next, we show that in each iteration of Step 1, if we do not halt, then the expected intersection size decreases by the k-th root compared to the previous iteration, until it becomes constant. Let

$$\boldsymbol{S}_r := \left|\bigcap_{\ell \in [k]} \boldsymbol{X}^{\ell}_{\boldsymbol{U}_r}\right| = \sum_{i \in \boldsymbol{U}_r} \boldsymbol{Z}_j$$

denote the intersection size at the beginning of the r-th iterations of Step 1, or 0 if the protocol has already halted prior to iteration r.

Lemma 13. *For each $1 \leq r \leq N$,*

$$\mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_{r+1}] \leq \mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_r]^{1/k}.$$

Proof. Consider again the set $\mathcal{W}_r$ of witnesses of length $r-1$ under which the protocol reaches Step 1c in the r-th iteration, and invokes the base protocol. Lemma 10 implies that for every $W_{<r} \in \mathcal{W}_r$:

$$\mathop{\mathbb{E}}_{\mu|W_{<r}}\left[\boldsymbol{S}_{r+1} \,\middle|\, \boldsymbol{Z}_{\boldsymbol{I}_r} = \overline{0}\right] \leq \mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{S}_r]^{1/k}.$$

The same holds even without conditioning on the event $\boldsymbol{Z}_{\boldsymbol{I}_r} = \overline{0}$: if $\boldsymbol{Z}_{\boldsymbol{I}_r} \neq \overline{0}$, then the protocol finds an intersection in iteration r, causing it to halt before reaching iteration $r+1$. In this case we have $\boldsymbol{S}_{r+1} = 0$ (by definition). Thus,

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{S}_{r+1}]\\ &= \Pr_{\mu|W_{<r}}\left(\boldsymbol{Z}_{\boldsymbol{I}_r} = \overline{0}\right)\mathop{\mathbb{E}}_{\mu|W_{<r}}\left[\boldsymbol{S}_{r+1} \,\middle|\, \boldsymbol{Z}_{\boldsymbol{I}_r} = \overline{0}\right] + \Pr_{\mu|W_{<r}}\left(\boldsymbol{Z}_{\boldsymbol{I}_r} \neq \overline{0}\right)\cdot 0\\ &\leq \mathop{\mathbb{E}}_{\mu|W_{<r}}\left[\boldsymbol{S}_{r+1} \,\middle|\, \boldsymbol{Z}_{\boldsymbol{I}_r} = \overline{0}\right]\\ &\leq \mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{S}_r]^{1/k}.\end{aligned}$$

Since this holds point-wise for any witness $W_{<r} \in \mathcal{W}_r$, and since $\boldsymbol{S}_{r+1} = 0$ whenever $\boldsymbol{W}_{<r} \notin \mathcal{W}_r$, all together we have:

$$\begin{aligned}\mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_{r+1}] &= \sum_{W_{<r} \in support(\boldsymbol{W}_{<r})} \Pr_{\mu}(\boldsymbol{W}_{<r} = W_{<r}) \mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{S}_{r+1}]\\ &= \sum_{W_{<r} \in \mathcal{W}_r} \Pr_{\mu}(\boldsymbol{W}_{<r} = W_{<r}) \mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{S}_{r+1}]\\ &\leq \sum_{W_{<r} \in \mathcal{W}_r} \Pr_{\mu}(\boldsymbol{W}_{<r} = W_{<r}) \mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{S}_r]^{1/k}\\ &\leq \sum_{W_{<r} \in support(\boldsymbol{W}_{<r})} \Pr_{\mu}(\boldsymbol{W}_{<r} = W_{<r}) \mathop{\mathbb{E}}_{\mu|W_{<r}}[\boldsymbol{S}_r]^{1/k}\\ &\leq \mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_r]^{1/k}. \qquad \text{(Jensen's inequality)}\end{aligned}$$

This completes the proof. □

Corollary 1. *We have $\mathbb{E}_{\mu}[\boldsymbol{S}_{N+1}] \leq 2$.*

Proof. Applying Lemma 13 N times, we get that

$$\mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_{N+1}] \leq \left(\mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_1]\right)^{1/k^N}.$$

Since $\mathbb{E}_{\mu}[\boldsymbol{S}_1] \leq n$ and $k^N = k^{\lceil \log\log n/\log k\rceil} \geq k^{\log_k \log n} = \log n$,

$$\left(\mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_1]\right)^{1/k^N} \leq n^{1/\log n} = 2,$$

which proves the claim. □

Corollary 2. *We have $\mathbb{E}_{\mu}[\boldsymbol{C}_{N+1}] = O(k + n^{1-1/k}\log n)$.*

Proof. Whenever $\boldsymbol{R} \leq N$, we do not reach Step 2 of the protocol, and both $\boldsymbol{C}_{N+1} = 0$ and $\boldsymbol{S}_{N+1} = 0$ by definition. Therefore,

$$\mathop{\mathbb{E}}_{\mu}[\boldsymbol{C}_{N+1}] = \Pr(\boldsymbol{R} = N+1)\mathop{\mathbb{E}}_{\mu}[\boldsymbol{C}_{N+1} \mid \boldsymbol{R} = N+1],$$

and similarly,

$$\mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_{N+1}] = \Pr(\boldsymbol{R} = N+1)\mathop{\mathbb{E}}_{\mu}[\boldsymbol{S}_{N+1} \mid \boldsymbol{R} = N+1].$$

When $\boldsymbol{R} = N+1$, we do call the base protocol, and by Lemma 6 the expected communication cost is:

$$\mathop{\mathbb{E}}_{\mu}[C_{N+1} \mid \boldsymbol{R} = N+1] = O(k + \mathop{\mathbb{E}}_{\mu}[S_{N+1} \mid \boldsymbol{R} = N+1]^{1/k}\, n^{1-1/k} \log n).$$

All together we have:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mu}[\boldsymbol{C}_{N+1}] = \Pr(\boldsymbol{R} = N+1) \mathop{\mathbb{E}}_{\mu}[\boldsymbol{C}_{N+1} \mid \boldsymbol{R} = N+1] \\
&= \Pr(\boldsymbol{R} = N+1)\, O(k + \mathop{\mathbb{E}}_{\mu}[S_{N+1} \mid \boldsymbol{R} = N+1]^{1/k}\, n^{1-1/k} \log n) \\
&= O(k + \mathop{\mathbb{E}}_{\mu}[S_{N+1}]^{1/k}\, n^{1-1/k} \log n) \\
&= O(k + n^{1-1/k} \log n),
\end{aligned}$$

where the last equality follows from Corollary 1. □

Putting everything together, we see that the expected communication cost of the protocol is given by

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mu}\left[\sum_{r=1}^{N} \boldsymbol{C}_r + \boldsymbol{C}_{N+1}\right] &\le (N+1) O(k + n^{1-1/k} \log n) \\
&= O(k + \lceil \log\log n / \log k \rceil n^{1-1/k} \log n).
\end{aligned}$$

Error probability. For every $r \in [N]$, let $\mathcal{W}_{<\epsilon,r}$ denote the set of witnesses $W_{<r}$ of length $|W_{<r}| = r-1$ under which the protocol reaches the r-th iteration and we have:

$$\Pr_{\mu|W_{<r}}\left(\bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell}_{\boldsymbol{I}_r} = \emptyset\right) \le \epsilon.$$

If $\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r}$, then the protocol halts in Step 1b of the r-th iteration, and declares that the player's inputs are intersecting. The protocol errs only if there is some iteration r where $\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r}$, but the players' inputs are in fact disjoint. In addition, the events $\{\boldsymbol{W}_{<1} \in \mathcal{W}_{<\epsilon,1}\}, \ldots, \{\boldsymbol{W}_{<N} \in \mathcal{W}_{<\epsilon,N}\}$ are disjoint, as the event $\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r}$ implies that the protocol halts in the r-th iteration and does not reach the next iteration. All together, we can bound the error probability of the protocol as follows:

$$\begin{aligned}
&\Pr_{\mu}(\text{ The protocol errs }) \\
&= \sum_{r\in[R]} \Pr_{\mu}\left(\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r} \wedge \bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell} = \emptyset\right) \\
&= \sum_{r\in[R]} \sum_{W_{<r}\in\mathcal{W}_{<\epsilon,r}} \Pr_{\mu}(\boldsymbol{W}_{<r} = W_{<r}) \Pr_{\mu|W_{<r}}\left(\bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell} = \emptyset\right) \\
&\le \sum_{r\in[R]} \sum_{W_{<r}\in\mathcal{W}_{<\epsilon,r}} \Pr_{\mu}(\boldsymbol{W}_{<r} = W_{<r}) \Pr_{\mu|W_{<r}}\left(\bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell}_{\boldsymbol{I}_r} = \emptyset\right) \\
&\le \sum_{r\in[R]} \sum_{W_{<r}\in\mathcal{W}_{<\epsilon,r}} \Pr_{\mu}(\boldsymbol{W}_{<r} = W_{<r}) \cdot \epsilon \\
&\qquad\qquad \text{(by the definition of } \mathcal{W}_{<\epsilon,r}) \\
&= \epsilon \cdot \sum_{r\in[R]} \Pr_{\mu}(\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r}) \le \epsilon. \qquad \text{(disjoint events)}
\end{aligned}$$

This concludes our analysis of the protocol's communication cost and error probability.

A note on round complexity. The protocol we presented here runs in $O(\log\log n/\log k)$ iterations, where each iteration requires the players (and the coordinator, in the coordinator model) to speak a constant number of times. However, in the shared-blackboard version of the base protocol as currently written, the players need to speak *one after the other*. It is not hard to see that this can be improved to $O(\log n)$ rounds where the players speak *simultaneously* (e.g., using the protocol from [8]), yielding a protocol with poly-logarithmic round complexity. It is interesting to ask whether under product distributions, set disjointness can be solved in *constant* rounds, while still achieving optimal communication complexity. (This is ruled out for general distributions in [8].)

5 LOWER BOUND

In this section we prove a lower bound of $\Omega(n^{1-1/k}/k^2)$ on set disjointness with a specific product distribution μ, assuming that $k = O(\log n)$. For $k = \omega(\log n)$, this trivially implies a lower bound of $\Omega(n/\log^2 n)$: simply take $O(\log n)$ players whose inputs are drawn from μ, and pad up to k by adding $k - O(\log n)$ more players with a fixed input of $[n]$.

5.1 High Level Overview of the Proof

Our lower bound uses the product distribution μ where each coordinate $\boldsymbol{X}^{\ell}_{i}$ of every player $\ell \in [k]$ has iid probability $1/n^{1/k}$ of being 1. Under this distribution, the prior probability that a given element $i \in [n]$ will be in the intersection is:

$$\Pr\left(i \in \bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell}\right) = \left(\frac{1}{n^{1/k}}\right)^{k} = 1/n.$$

Thus, there is a constant probability that the intersection is empty:

$$\Pr\left(\bigcap_{\ell} \boldsymbol{X}^{\ell} = \emptyset\right) = \left(1 - \frac{1}{n}\right)^{n} \in \left(\frac{1}{2e}, \frac{1}{e}\right),$$

assuming $n \ge 2$.

Let Π be a protocol with error at most ϵ on μ, where ϵ is a sufficiently small constant. Let $M_{\emptyset}$ be the set of transcripts of Π where the output is "non-intersecting" ($\emptyset$), and let $\boldsymbol{M}$ be a random variable representing the transcript of the protocol.

Since Π has low error, for a typical transcript $m \in M_{\emptyset}$, there should be a low *posterior* probability given $\boldsymbol{M} = m$ that the input is intersecting:

$$\begin{aligned}
\epsilon &\ge \Pr\left(\Pi \text{ outputs "}\emptyset\text{" but } \bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell} \ne \emptyset\right) \\
&= \sum_{m\in M_{\emptyset}} \Pr(\boldsymbol{M} = m) \Pr\left(\bigcap_{\ell=1}^{k} \boldsymbol{X}^{\ell} \ne \emptyset \,\middle|\, \boldsymbol{M} = m\right). \qquad (5)
\end{aligned}$$

Our lower bound shows that if Π sends fewer than $\Omega(n^{1-1/k}/k^2)$ bits in expectation, then for a typical transcript $m \in M_{\emptyset}$ we have fairly *high* posterior probability of an intersection, violating (5). To show this, we prove that there is a large subset $J = J(m) \subseteq [n]$ of coordinates, of size $|J| = \Omega(n)$, such that

(1) For each $i \in J$, the posterior probability $\Pr\left(i \in \bigcap_{\ell} \boldsymbol{X}^{\ell} \,\middle|\, \boldsymbol{M} = m\right)$ remains close to $1/n$, and

(2) The coordinates in J remain "almost independent" of one another given $\boldsymbol{M} = m$.

Together, (1) and (2) allow us to carefully "collect" the intersection probabilities of the coordinates in J, and show that there is a high overall intersection probability given $\boldsymbol{M} = m$. This part of the proof is delicate, because we are working with very low-probability events: informally, we need to "add up" $\Omega(n)$ events of the form $i \in \bigcap_\ell \boldsymbol{X}^\ell$, where each event has probability $\approx 1/n$. A protocol that sends roughly $n^{1-1/k}$ bits can create dependencies on the order of $n^{1-1/k}/n = 1/n^{1/k}$ between the $\Theta(n)$ coordinates in J (measured in mutual information), and this is enough to overwhelm the $\approx 1/n$ probability of the event we are interested in. To deal with this, we exploit the independence of the players' inputs, and show that even though there can be dependence of $\approx 1/n^{1/k}$ between two coordinates $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^k$ and $\boldsymbol{X}_j^1, \ldots, \boldsymbol{X}_j^k$ (where $i, j \in J$), the dependence between the events $\bigwedge_\ell \boldsymbol{X}_i^\ell = 1$ and $\bigwedge_\ell \boldsymbol{X}_j^\ell = 1$ (or in other words, the events $i \in \bigcap_\ell \boldsymbol{X}^\ell$ and $j \in \bigcap_\ell \boldsymbol{X}^\ell$) is only $\approx 1/n$.

5.2 Proof of the Lower Bound

Let $\epsilon, C \in (0,1)$ be sufficiently small constants. Fix a protocol Π that communicates fewer than $Cn^{1-1/k}/k^2$ bits in expectation, when the inputs are drawn from μ. Let $\boldsymbol{M}$ be a random variable denoting the transcript of Π.

Good transcripts. A typical transcript of Π conveys $O(n^{1-1/k}/k^2)$ bits of information about the players' input, in the following sense:

$$\mathop{\mathbb{E}}_{m\sim\boldsymbol{M}} \mathsf{D}\left(\boldsymbol{X}|_{\boldsymbol{M}=m} \,||\, \boldsymbol{X}\right) = \mathsf{I}(\boldsymbol{M};\boldsymbol{X}) \le Cn^{1-1/k}/k^2.$$

This allows us to bound the amount by which conditioning on the event $\boldsymbol{M} = m$, for a typical transcript m, distorts the input distribution. We define the *good transcripts* of Π to be the transcripts that output "non-intersecting" ($\emptyset$) and reveal $O(n^{1-1/k}/k^2)$ information about the input, which means that they do not distort the input distribution too much or cause too many dependencies between different coordinates.

Definition 4. *We say that a transcript m of Π is* good *if it satisfies:*

(1) The output of m is "non-intersecting" ($\emptyset$).

(2) The information m gives about the input is "not much higher than average": for some small constant $\alpha_1 \in (0,1)$,

(a) Conditioning on $\boldsymbol{M} = m$ does not distort the distribution of the typical coordinate by too much:

$$\sum_{i=1}^n \mathsf{D}\left(\boldsymbol{X}_i|_{\boldsymbol{M}=m} \,||\, \boldsymbol{X}_i\right) \le \alpha_1 n^{1-1/k}/k^2,$$

and

(b) Conditioning on $\boldsymbol{M} = m$ does not create too much dependence between individual coordinates:

$$\sum_{i=1}^n \mathsf{I}\left(\boldsymbol{X}_i; \boldsymbol{X}_{<i} \mid \boldsymbol{M} = m\right) \le \alpha_1 n^{1-1/k}/k^2.$$

We prove that there is constant probability of getting a good transcript:

Lemma 14. *There is a constant $\gamma \in (0,1)$ such that*

$$\Pr\left(\boldsymbol{M} \text{ is good}\right) \ge \gamma.$$

Proof. We go through the conditions of Definition 4, and bound the probability that a transcript $m \sim \boldsymbol{M}$ fails to satisfy each condition.

Condition (1) holds with good probability because there is constant probability that the *correct* answer is "$\emptyset$", and Π has low error:

$$\begin{aligned}\Pr\left(\boldsymbol{M} \text{ outputs "}\emptyset\text{"}\right) &\ge \Pr\left(\bigcap_\ell \boldsymbol{X}^\ell = \emptyset \text{ and } \boldsymbol{M} \text{ outputs "}\emptyset\text{"}\right)\\ &\ge \Pr\left(\bigcap_\ell \boldsymbol{X}^\ell = \emptyset\right) - \Pr\left(\boldsymbol{M} \text{ errs}\right)\\ &\ge \frac{1}{2e} - \epsilon.\end{aligned}$$

Next, let us bound the probability that condition (2a) or (2b) does not hold, by bounding the probability that the *sum* of the quantities in those conditions is too large. This follows from the fact that

$$\mathsf{I}(\boldsymbol{M};\boldsymbol{X}) \le \mathsf{H}(\boldsymbol{M}) \le Cn^{1-1/k}/k^2.$$

By the chain rule, we can also write

$$\begin{aligned}\mathsf{I}(\boldsymbol{M};\boldsymbol{X}) &= \sum_{i=1}^n \left[\mathsf{I}(\boldsymbol{X}_i;\boldsymbol{M}|\boldsymbol{X}_{<i})\right]\\ &= \sum_{i=1}^n \left[\mathsf{I}(\boldsymbol{X}_i;\boldsymbol{M}|\boldsymbol{X}_{<i}) + \mathsf{I}(\boldsymbol{X}_i;\boldsymbol{X}_{<i})\right] && (\mathsf{I}(\boldsymbol{X}_i;\boldsymbol{X}_{<i}) = 0)\\ &= \sum_{i=1}^n \left[\mathsf{I}(\boldsymbol{X}_i;\boldsymbol{X}_{<i},\boldsymbol{M})\right]\\ &= \sum_{i=1}^n \left[\mathsf{I}(\boldsymbol{X}_i;\boldsymbol{X}_{<i}|\boldsymbol{M}) + \mathsf{I}(\boldsymbol{X}_i;\boldsymbol{M})\right]\\ &= \sum_{i=1}^n \left[\mathop{\mathbb{E}}_{m\sim\boldsymbol{M}} \mathsf{I}(\boldsymbol{X}_i;\boldsymbol{X}_{<i}|\boldsymbol{M}=m) + \mathop{\mathbb{E}}_{m\sim\boldsymbol{M}} \mathsf{D}\left(\boldsymbol{X}_i|_{\boldsymbol{M}=m} \,||\, \boldsymbol{X}_i\right)\right]\\ &= \mathop{\mathbb{E}}_{m\sim\boldsymbol{M}}\left[\sum_{i=1}^n \left[\mathsf{I}(\boldsymbol{X}_i;\boldsymbol{X}_{<i}|\boldsymbol{M}=m) + \mathsf{D}\left(\boldsymbol{X}_i|_{\boldsymbol{M}=m} \,||\, \boldsymbol{X}_i\right)\right]\right].\end{aligned}$$

Since $\mathsf{I}(\boldsymbol{M};\boldsymbol{X}) \le Cn^{1-1/k}/k^2$, and since mutual information and divergence are non-negative, we obtain by Markov that the probability that either condition (2a) or condition (2b) is violated is at most C/α_1.

Taking ϵ, C sufficiently small, there is constant probability that neither condition (1) nor conditions (2a), (2b) are violated, yielding a good transcript. □

A good set of indices. We show that for any good transcript m, there is a large subset $J \subseteq [n]$ of indices that m "does not give a lot of information about", and which remain nearly-independent of one another:

Lemma 15. *Let α_2 be a sufficiently small constant. For any good transcript m, there exists a set $J = J(m) \subseteq [n]$ of size $|J| \ge n/2$, such that for all $i \in J$, $\ell \in [k]$:*

(1) $\mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) \le \alpha_2/(n^{1/k}k^2)$, and

(2) $\mathsf{I}\left(\boldsymbol{X}_i^\ell\,;\, \boldsymbol{X}_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M}=m\right) \le \alpha_2/(n^{1/k}k^2)$.

We write J instead of $J(m)$ when the transcript m is clear from the context.

The existence of the set $J(m)$ follows fairly immediately from the definition of a good transcript, since condition (2) of Definition 4 essentially requires that an average coordinate $i \in [n]$ satisfy the conditions of the lemma. In fact, condition (2) is stronger, since it applies to *all players together*, and here we apply the bound to each individual player. It is possible that the lower bound could be improved by a factor of k using a more careful averaging argument to argue about the *typical* player rather than all players.

For each $i \in J$, the divergence $\mathrm{D}\left(X_i^\ell|_{M=m} \,\middle\|\, X_i^\ell\right)$ is very small, and this means the posterior probability that $X_i^\ell = 1$ given $M = m$ is very close to the prior:

Lemma 16. *For each $i \in J$ and $\ell \in [k]$ we have:*

$$\Pr\left(X_i^\ell = 1 \,\middle|\, M = m\right) \in \left(\frac{1 - 1/(4k)}{n^{1/k}}, \frac{1 + 1/(4k)}{n^{1/k}}\right).$$

This follows from Lemma 1, taking α_2 sufficiently small.

We remark that if only the different coordinates $i \in [n]$ were *independent* given $M = m$, we would now be finished with the proof: since Π is a communication protocol, conditioning on $M = m$ does not create dependencies between the players' inputs, so for each $i \in J$,

$$\Pr\left(i \in \bigcap_\ell X^\ell \,\middle|\, M = m\right) = \prod_\ell \Pr\left(X_i^\ell = 1 \,\middle|\, M = m\right) \geq \left(\frac{1 - 1/(4k)}{n^{1/k}}\right)^k \approx 1/n.$$

Therefore,

$$\mathbb{E}\left[\left|\bigcap_\ell X^\ell\right| \,\middle|\, M = m\right] \geq \sum_{i \in J} \Pr\left(\bigwedge_\ell X_i^\ell = 1 \,\middle|\, M = m\right) = \Omega(1).$$

If the coordinates were independent, the fact that the expected intersection size is $\Omega(1)$ would imply constant probability of a non-empty intersection, even given $M = m$, meaning the protocol errs w.h.p.; unfortunately, conditioning on $M = m$ does create dependencies between the coordinates, so we must proceed more carefully.

Adding up the intersection probabilities. Our goal now is to prove that if m is a good transcript, then

$$\Pr\left(\bigcap_\ell X^\ell \neq \emptyset \,\middle|\, M = m\right) \geq 0.01.$$

We "collect" the coordinates in J one-by-one, handle the dependencies between them, and show that the probability of an intersection roughly "adds up" over the coordinates.

Let $\mathcal{D}$ be the event that $\bigcap_\ell X^\ell = \emptyset$, and given a transcript m, let $\mathcal{D}_{<i}(m)$ be the event that there is no intersection at any coordinate of $J(m)$ that is smaller than i:

$$\mathcal{D}_{<i}(m) = \bigwedge_{j \in J(m) : j < i} \left(\bigwedge_\ell X_j^\ell = 0\right).$$

As usual, we write $\mathcal{D}_{<i}$ instead of $\mathcal{D}_{<i}(m)$ when m is clear from the context. Note that

$$\Pr\left(\mathcal{D} \mid M = m\right) \leq \Pr\left(\forall i \in J : \bigwedge_\ell X_i^\ell = 0 \,\middle|\, M = m\right) = \prod_{i \in J} \Pr\left(\bigwedge_\ell X_i^\ell = 0 \,\middle|\, M = m, \mathcal{D}_{<i}\right). \tag{6}$$

The key lemma that allows us to "collect" the intersection probabilities using (6) is the following:

Lemma 17. *Let m be a good transcript such that $\Pr\left(\mathcal{D} \mid M = m\right) > 0.6$. Then for each $i \in J$ we have:*

$$\Pr\left(\bigwedge_\ell X_i^\ell = 1 \,\middle|\, M = m, \mathcal{D}_{<i}\right) \geq \frac{1}{2n}.$$

Since $|J| = \Omega(n)$, plugging the bound from Lemma 17 into (6) shows that there is constant, non-zero probability that the intersection is non-empty given $M = m$. But m outputs "$\emptyset$", so this contributes to the error of the protocol.

In the proof of Lemma 17 we will need to handle the dependencies between the coordinates. Technically, we will need to show that even when we condition on $M = m$ and on $\mathcal{D}_{<i}$ (i.e., no intersection up to coordinate i, exclusive), the distribution of coordinate i itself is not distorted by much. To do so we will rely on the following properties of the event $\mathcal{D}_{<i}$:

Lemma 18. *For every good transcript m, coordinate $i \in [n]$ and player $\ell \in [k]$,*

$$\mathrm{I}\left(X_i^\ell \,;\, \mathbf{1}_{\mathcal{D}_{<i}} \,\middle|\, M = m, X_i^{<\ell} = \overline{1}\right) \leq \mathrm{I}\left(X_i^\ell \,;\, X_{J_{<i}}^\ell \,\middle|\, M = m\right).$$

Lemma 19. *Let m be a good transcript such that $\Pr\left(\mathcal{D} \mid M = m\right) > 0.6$. Then for each $i \in J$ and $\ell \in [k]$,*

$$\Pr\left(\mathcal{D}_{<i} \,\middle|\, M = m, X_i^{<\ell} = \overline{1}\right) \geq 1/2.$$

Here, $\mathbf{1}_{\mathcal{D}_{<i}}$ is an indicator for the event $\mathcal{D}_{<i}$.

Before proving Lemmas 18 and 19, let us show how they are used to prove Lemma 17.

Proof of Lemma 17. We go through the players $\ell \in [k]$, and prove that

$$\Pr\left(X_i^\ell = 1 \,\middle|\, M = m, \mathcal{D}_{<i}, X_i^{<\ell} = \overline{1}\right) \geq \frac{(1 - 1/(4k))^2}{n^{1/k}}. \tag{7}$$

This immediately implies the lemma, as

$$\Pr\left(\bigwedge_\ell X_i^\ell = 1 \,\middle|\, M = m, \mathcal{D}_{<i}\right) = \prod_{\ell=1}^k \Pr\left(X_i^\ell = 1 \,\middle|\, M = m, \mathcal{D}_{<i}, X_i^{<\ell} = \overline{1}\right) \geq \left(\frac{(1 - 1/(4k))^2}{n^{1/k}}\right)^k \geq \frac{1}{2n}. \qquad \text{(by (7))}$$

In the last step, we used the fact that $(1 - 1/(4k))^{2k} \geq 1/2$ for all $k \geq 1$.

Thus, let us prove (7). By Lemma 16, we already know that

$$\Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m\right) \geq \frac{1 - 1/(4k)}{n^{1/k}}. \tag{8}$$

To prove (7), we need to show that

$$\Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, \mathcal{D}_{<i}, X_i^{<\ell} = \overline{1}\right) \geq \left(1 - \frac{1}{4k}\right) \Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m\right).$$

In other words, we need to bound the effect of conditioning on $\mathcal{D}_{<i}$ and on $X_i^{<\ell} = \overline{1}$.

Introducing the conditioning on the event $X_i^\ell = \overline{1}$ is easy: since Π is a communication protocol, the inputs of the players remain independent of one another, even given $\boldsymbol{M} = m$. Thus,

$$\Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) = \Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m\right). \tag{9}$$

Next, we carefully introduce the conditioning on $\mathcal{D}_{<i}$, using Lemma 2 to bound the effect. Note that the event $X_i^\ell = 1$ has quite small probability, roughly $1/n^{1/k}$; to argue that conditioning on $\mathcal{D}_{<i}$ does not "wipe out" the event $X_i^\ell = 1$, we need to argue that the dependence between $\mathcal{D}_{<i}$ and X_i^ℓ is also on the order of $1/n^{1/k}$. And indeed this is the case: since $\mathcal{D}_{<i}$ is an event that depends only on the coordinates in $J_{<i}$,

$$\begin{aligned}
&\mathsf{I}\left(X_i^\ell \,;\, \mathbf{1}_{\mathcal{D}_{<i}} \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) \\
&\leq \mathsf{I}\left(X_i^\ell \,;\, X_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M} = m\right) && \text{(Lemma 18)} \\
&\leq \frac{\alpha_2}{n^{1/k} k^2} && \text{(by definition of } J\text{)} \\
&\leq \Pr\left(\mathcal{D}_{<i} \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) \\
&\qquad \cdot \frac{\Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) \cdot \left(\frac{1}{4k}\right)^2}{40 \ln 2},
\end{aligned}$$

where the last step used (8), (9), Lemma 19, and also assumed that α_2 is small enough. Now, taking $\alpha = 1/(4k)$, Lemma 2 yields

$$\begin{aligned}
&\Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}, \mathcal{D}_{<i}\right) \\
&\geq (1 - \alpha) \cdot \Pr\left(X_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) \\
&\geq \left(1 - \frac{1}{4k}\right)^2 \cdot \frac{1}{n^{1/k}}. && \text{(by (8), (9))}
\end{aligned}$$

This completes the proof. □

We now return to Lemmas 18 and 19.

Proof of Lemma 18. Recall that $\mathcal{D}_{<i}$ is an event that depends only on the bits $X_{J_{<i}}$. Thus, by the data processing inequality and the chain rule,

$$\begin{aligned}
&\mathsf{I}\left(X_i^\ell \,;\, \mathbf{1}_{\mathcal{D}_{<i}} \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) \\
&\leq \mathsf{I}\left(X_i^\ell \,;\, X_{J_{<i}}^\ell X_{J_{<i}}^{-\ell} \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) \\
&= \mathsf{I}\left(X_i^\ell \,;\, X_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) \\
&\qquad + \mathsf{I}\left(X_i^\ell \,;\, X_{J_{<i}}^{-\ell} \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}, X_{J_{<i}}^\ell\right) \\
&= \mathsf{I}\left(X_i^\ell \,;\, X_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M} = m\right).
\end{aligned}$$

The last equality follows from the fact that Π is a communication protocol, and the inputs of different players are independent a-priori; given $\boldsymbol{M} = m$, the inputs of different players remain independent, even conditioned on parts of the input. □

Proof of Lemma 19. Let m be a good transcript with

$$\Pr\left(\mathcal{D} \mid \boldsymbol{M} = m\right) > 0.6.$$

Then, since $\mathcal{D}_{<i}$ is implied by $\mathcal{D}$, we also have

$$\Pr\left(\mathcal{D}_{<i} \mid \boldsymbol{M} = m\right) > 0.6.$$

We introduce the conditioning on $X_i^{<\ell} = \overline{1}$ step-by-step, each time conditioning on one additional bit being 1: for each $t \in [\ell]$, let

$$p_t = \Pr\left(\mathcal{D}_{<i} \,\middle|\, \boldsymbol{M} = m, X_i^{<t} = \overline{1}\right).$$

We will show that $|p_t - p_{t-1}| \leq 1/(10k)$ for each $t \in [\ell]$, and conclude that $|p_\ell - p_0| \leq 1/10$. Therefore:

$$\begin{aligned}
\Pr\left(\mathcal{D}_{<i} \,\middle|\, \boldsymbol{M} = m, X_i^{<\ell} = \overline{1}\right) &\geq \Pr\left(\mathcal{D}_{<i} \mid \boldsymbol{M} = m\right) - 1/10 \\
&> 0.6 - 0.1 > 1/2,
\end{aligned}$$

as desired.

Fix $t \in [\ell]$, and let us bound the effect of conditioning on $X_i^t = 1$ (i.e., the difference between p_t and p_{t-1}). By Lemma 18, and by definition of the set J,

$$\begin{aligned}
\mathsf{I}\left(\mathbf{1}_{\mathcal{D}_{<i}}; X_i^t \,\middle|\, \boldsymbol{M} = m, X_i^{<t} = \overline{1}\right) &\leq \mathsf{I}\left(X_{J_{<i}}^t; X_i^t \,\middle|\, \boldsymbol{M} = m\right) \\
&\leq \alpha_2/(n^{1/k} k^2).
\end{aligned} \tag{10}$$

By Property 1 and the definition of p_{t-1} and p_t, we have

$$\begin{aligned}
&\mathsf{I}\left(\mathbf{1}_{\mathcal{D}_{<i}}; X_i^t \,\middle|\, \boldsymbol{M} = m, X_i^{<t} = \overline{1}\right) \\
&\geq \Pr\left(X_i^t = 1 \,\middle|\, \boldsymbol{M} = m, X_i^{<t} = \overline{1}\right) \mathsf{D}\left(p_t \,||\, p_{t-1}\right),
\end{aligned}$$

and together with (10) we obtain

$$\mathsf{D}\left(p_t \,||\, p_{t-1}\right) \leq \frac{\alpha_2}{\Pr\left(X_i^t = 1 \,\middle|\, \boldsymbol{M} = m, X_i^{<t} = \overline{1}\right) \cdot n^{1/k} k^2}. \tag{11}$$

While $X_i^t = 1$ is an unlikely event, it is not *too* unlikely – from Lemma 16 and the independence of the inputs we get that $\Pr\left(X_i^t = 1 \,\middle|\, \boldsymbol{M} = m, X_i^{<t} = \overline{1}\right) \geq 1/(2n^{1/k})$, so (11) yields:

$$\mathsf{D}\left(p_t \,||\, p_{t-1}\right) \leq 2\alpha_2/k^2.$$

By Pinsker's inequality, we have:

$$|p_t - p_{t-1}| \leq \sqrt{\mathsf{D}\left(p_t \,||\, p_{t-1}\right) \ln 2/2} \leq 1/(10k),$$

assuming $\alpha_2 \leq 1/(100 \ln 2)$. This concludes the proof. □

Putting everything together. From Lemma 17 we deduce that every good transcript has noticeable probability of erring:

Corollary 3. *If m is a good transcript, then*

$$\Pr(\mathcal{D} \mid \boldsymbol{M} = m) \le e^{-1/4}.$$

Proof. If $\Pr(\mathcal{D} \mid \boldsymbol{M} = m) \le 0.6$ then we are done, because $0.6 < e^{-1/4}$. Thus, suppose $\Pr(\mathcal{D} \mid \boldsymbol{M} = m) > 0.6$. Applying Lemma 17, we get

$$\begin{aligned}
\Pr(\mathcal{D} \mid \boldsymbol{M} = m) &\le \Pr\left(\forall i \in J : \bigwedge_\ell X_i^\ell = 0 \;\middle|\; \boldsymbol{M} = m\right) \\
&= \prod_{i \in J} \Pr\left(\bigwedge_\ell X_i^\ell = 0 \;\middle|\; \boldsymbol{M} = m, \mathcal{D}_{<i}\right) \\
&\le \left(1 - \frac{1}{2n}\right)^{|J|} && \text{(Lemma 17)} \\
&\le \left(1 - \frac{1}{2n}\right)^{n/2} \le e^{-1/4}. && \square
\end{aligned}$$

Since every good transcript errs with reasonable probability, and since there is reasonable probability that the protocol's transcript is good, we can bound the protocol's error probability from below:

Corollary 4. *The protocol errs with probability at least $\gamma(1 - e^{-1/4})$.*

Proof. Recall that the output of every good transcript is "non-intersecting" ($\emptyset$). Therefore,

$$\begin{aligned}
\Pr(\boldsymbol{M} \text{ errs}) &\ge \sum_{\text{good } m} \Pr(\boldsymbol{M} = m) \Pr\left(\bigcap_\ell X^\ell \ne \emptyset \;\middle|\; \boldsymbol{M} = m\right) \\
&\ge \sum_{\text{good } m} \Pr(\boldsymbol{M} = m)\,(1 - e^{-1/4}) && \text{(Cor. 3)} \\
&= \Pr(\boldsymbol{M} \text{ is good}) \cdot (1 - e^{-1/4}) \\
&\ge \gamma(1 - e^{-1/4}). && \text{(Lemma 14)}
\end{aligned}$$

$\square$

Setting C small enough, we can set the relevant constants such that we get $\epsilon < \gamma(1 - e^{-1/4})$, and obtain a contradiction.

ACKNOWLEDGMENTS

The authors wish to thank Toniann Pitassi and Orr Fischer for fruitful discussions.

This work is supported by the Len Blavatnik and the Blavatnik Family foundation and by the Israeli Science Foundation under Grant No. 2801/20.

REFERENCES

[1] Noga Alon, Yossi Matias, and Mario Szegedy. 1999. The Space Complexity of Approximating the Frequency Moments. *J. Comput. System Sci.* 58, 1 (1999), 137–147.

[2] Pranjal Awasthi, Ainesh Bakshi, Maria-Florina Balcan, Colin White, and David P. Woodruff. 2019. Robust Communication-Optimal Distributed Clustering Algorithms. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, Vol. 132. 18:1–18:16.

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

[4] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. 2002. An Information Statistics Approach to Data Stream and Communication Complexity. In *43rd Symposium on Foundations of Computer Science (FOCS 2002)*. 209–218.

[5] Ralph Bottesch, Dmitry Gavinsky, and Hartmut Klauck. 2015. Correlation in Hard Distributions in Communication Complexity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM (LIPIcs, Vol. 40)*. 544–572.

[6] Mark Braverman, Faith Ellen, Rotem Oshman, Toniann Pitassi, and Vinod Vaikuntanathan. 2013. A Tight Bound for Set Disjointness in the Message-Passing Model. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*. 668–677.

[7] Mark Braverman and Rotem Oshman. 2015. On Information Complexity in the Broadcast Model. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing, PODC 2015*. 355–364.

[8] Mark Braverman and Rotem Oshman. 2017. A Rounds vs. Communication Tradeoff for Multi-Party Set Disjointness. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017*. 144–155.

[9] Joshua Brody, Amit Chakrabarti, Ranganath Kondapally, David P. Woodruff, and Grigory Yaroslavtsev. 2014. Beyond Set Disjointness: The Communication Complexity of Finding the Intersection. In *Proceedings of the 2014 ACM Symposium on Principles of Distributed Computing (PODC '14)*. 106–113.

[10] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. 2001. Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001*. 270–278.

[11] Arkadev Chattopadhyay and Toniann Pitassi. 2010. The story of set disjointness. *ACM SIGACT News* 41, 3 (2010), 59–85.

[12] Jiecao Chen, He Sun, David Woodruff, and Qin Zhang. 2016. Communication-Optimal Distributed Clustering. In *Advances in Neural Information Processing Systems*, Vol. 29. 3727–3735.

[13] André Gronemeier. 2009. Asymptotically Optimal Lower Bounds on the NIH-Multi-Party Information Complexity of the AND-Function and Disjointness. In *26th International Symposium on Theoretical Aspects of Computer Science, STACS 2009 (LIPIcs, Vol. 3)*. 505–516.

[14] Zengfeng Huang, Bozidar Radunovic, Milan Vojnovic, and Qin Zhang. 2020. Communication complexity of approximate maximum matching in the message-passing model. *Distributed Computing* 33, 6 (2020), 515–531.

[15] T. S. Jayram. 2009. Hellinger Strikes Back: A Note on the Multi-party Information Complexity of AND. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009 (Lecture Notes in Computer Science, Vol. 5687)*. 562–573.

[16] Jeff M. Phillips, Elad Verbin, and Qin Zhang. 2012. Lower bounds for number-in-hand multiparty communication complexity, made easy. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012*. 486–501.

[17] Anup Rao and Amir Yehudayoff. 2020. *Communication Complexity: and Applications.*

[18] Alexander A Razborov. 1990. On the distributional complexity of disjointness. In *International Colloquium on Automata, Languages, and Programming*. 249–253.

[19] Georg Schnitger and Bala Kalyanasundaram. 1987. The probabilistic communication complexity of set intersection. In *Proceedings of the Second Annual Conference on Structure in Complexity Theory 1987*.

[20] Alexander A Sherstov. 2014. Communication complexity theory: Thirty-five years of set disjointness. In *International Symposium on Mathematical Foundations of Computer Science*. 24–43.

[21] David P. Woodruff and Qin Zhang. 2012. Tight Bounds for Distributed Functional Monitoring. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing (STOC '12)*. 941–960.

[22] David P. Woodruff and Qin Zhang. 2013. When Distributed Computation Is Communication Expensive. In *Distributed Computing: 27th International Symposium, DISC 2013*. 16–30.