The Communication Complexity of Multiparty Set Disjointness Under Product Distributions

Thesis submitted in partial fulfillment of the requirements for the M.Sc. degree in the School of Computer Science, Tel-Aviv University

by

Tal Roth

The research for this thesis has been carried out at Tel-Aviv University under the supervision of

Professor Rotem Oshman and **Professor Nachum Dershowitz**

December 2020

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisors, Professor Rotem Oshman and Professor Nachum Dershowitz, for their endless patience, willingness to always advise on both technical and academic issues, and for giving me both room to take risks as well as a safety net when attempts fail. Most of all, I'm thankful that they made this not only a tremendous learning experience, but also a lot of fun...

I would like to thank my friends Tal Yankovitz, Orr Fischer, Roni Con and Dror Chawin, for lending me their ears countless times, many hours of technical discussions and for their encouragement.

To my wife Tammy, and my kids Shiri, Lia and Eitan, this is all for you.

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.

Contents

Chapter 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., [CP10, She14]). 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 [SK87, Raz90] 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, [BFS86] 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 thesis, 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., [AMS99, BJKS02, Gro09]), and connections and applications in distributed computing and auction theory have led to the development of further lower bounds [WZ13, BCK+14, BEO+13, BO15, BO17]. 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 cost of k-party set disjointness is $\Theta(n \log k + k)$ [BO15]. 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 cost rises to $\Theta(kn)$ [BEO+13]. 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 [WZ12, WZ13, CSWZ16, ABB+19, HRVZ20] and many

others for examples of upper and lower bounds in this setting).

Our results. In this thesis we study multiparty set disjointness under *product distributions*, 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 of players increases, disjointness converges to the same cost 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 cost is $\Theta(kn)$ [BEO+13], the gap between the restricted and the unrestricted costs grows with the number of players.

The formal statement of our results is as follows. Let $Disj_{n,k}^{\mu,\epsilon}$ 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 $Disj_{n,k}^{\mu,\epsilon}$ 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 cost of $Disj_{n,k}^{\mu,\epsilon}$ is $O(k + n \log^2 n)$.*

Our lower bound is proven for the shared blackboard 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 cost of $Disj_{n,k}^{\mu,\epsilon}$ 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 cost 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 [WZ13],[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 [BFS86] readily generalize to $k > 2$ players.[2] Therefore, we came up with a new upper bound based on different ideas than [BFS86], and whereas [BFS86] used a combinatorial lower bound argument (the corruption bound), our lower bound is information theoretic. In Chapter 6, we sketch the upper and lower bounds of [BFS86], and explain why they break down when there are more than two players.

Our lower bound also does not use the typical direct sum argument [CSWY01] that is often used in information-theoretic disjointness lower bounds (e.g., in [BJKS02, Gro09, Jay09, BEO+13, BO15, BGK15, BO17]). We believe that our approach may have applications in other settings that are not amenable to the standard direct sum, such as proving information-theoretic lower bounds for the number-on-forehead model.

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 cost 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 [RY20] 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 [RY20] works with the following input distribution μ:[3]

[1] In [WZ13] the reduction is from a different problem, which [WZ13] defined and analyzed, as [WZ13] preceded the disjointness lower bound of [BEO+13]. However, the reductions of [WZ13] are easily modified to work with disjointness instead.

[2] Nor do the techniques of [BGK15], which interpolated between the $\Theta(n)$ unrestricted cost and the $\Theta(\sqrt{n})$ cost 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 [BGK15] is a clever modification of [BFS86], and the lower bound is an adaptation of Razborov's lower bound [Raz90].

[3] As does Razborov's original lower bound [Raz90], using different constants.

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

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 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 thesis is organized as follows. In Chapter 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 in Chapters 3 and 4. Next, in Chapter 5, we prove our $\tilde{\Omega}(n^{1-1/k})$ lower bound for disjointness under a product distribution. Finally, in Chapter 6, we discuss in detail the limitations of the upper and lower bound techniques that were introduced in [BFS86], which motivated the development of our new upper bound and lower bound techniques.

[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 [RY20].

[5] This is very nearly the distribution used in [BFS86], 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.

Chapter 2

Preliminaries

The shared blackboard model. We have k players with private inputs $X^1, \ldots, X^k$ who wish to cooperate in order to compute some function $f(X^1, \ldots, X^k)$ of their inputs. At each point in time, one (and only one) player is allowed to write a message on the shared blackboard, visible to all players. The identity of the player whose turn it is to speak, is a deterministic function of everything written on the shared blackboard so far. At the end of the execution, the last player writes the value of $f(X^1, \ldots, X^k)$ on the shared blackboard. A *transcript* of a certain execution is everything that was written on the shared blackboard during that execution.

The coordinator model. In this model, the players communicate over private channels. There is an extra player without any input – the *coordinator*, and the players may only communicate with the coordinator. At the end of the protocol, the coordinator computes the value $f(X^1, \ldots, X^k)$ and sends it to the first player. The transcript of an execution of a protocol, is everything transmitted by the players and coordinator during the execution.

In both shared blackboard and coordinator models, the players (and coordinator) are allowed to use public random bits.

Notation. We use boldface to denote random variables. We will denote by $\boldsymbol{X}_i^k$ the i-th coordinate of player k. Consider a set of random variables $\{\boldsymbol{X}_i^\ell\}_{i\in[n],\ell\in[k]}$. Throughout this thesis, we will use the following notations:

$$\begin{aligned}
\boldsymbol{X}^\ell &:= \boldsymbol{X}_1^\ell, \ldots, \boldsymbol{X}_n^\ell \\
\boldsymbol{X}^{-\ell} &:= \boldsymbol{X}^1, \ldots, \boldsymbol{X}^{\ell-1}, \boldsymbol{X}^{\ell+1}, \ldots, \boldsymbol{X}^k \\
\boldsymbol{X}^{<\ell} &:= \boldsymbol{X}^1, \ldots, \boldsymbol{X}^{\ell-1} \\
\boldsymbol{X}_i &:= \boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^k
\end{aligned}$$

Let $J \subseteq [n]$, $i \in [n]$. Then:

$$\begin{aligned}
\boldsymbol{X}_J^\ell &:= \{\boldsymbol{X}_j^\ell\}_{j\in J} \\
J_{<i} &:= \{j \in J \mid j < i\} \\
\boldsymbol{X}_{J_{<i}}^\ell &:= \{\boldsymbol{X}_j^\ell\}_{j\in J_{<i}}
\end{aligned}$$

Sometimes for $a \in \{0,1\}$, we will denote by $\boldsymbol{X}^{\ell}_{J_{<i}} = \overline{a}$ the event:

$$\left\{\boldsymbol{X}^{\ell}_{J_{<i}} = \overline{a}\right\} := \bigwedge_{j \in J_{<i}} \left(\boldsymbol{X}^{\ell}_{j} = a\right),$$

and similarly, denote by $\boldsymbol{X}_i = \overline{a}$ the event:

$$\{\boldsymbol{X}_i = \overline{a}\} := \bigwedge_{\ell \in [k]} \left(\boldsymbol{X}^{\ell}_{i} = a\right).$$

For some random variables $\boldsymbol{A} \sim \mu$ and $\boldsymbol{B}$, and b in the support of $\boldsymbol{B}$, we denote by $\boldsymbol{A}|_{\boldsymbol{B}=b}$ a random variable distributed according to the distribution $\mu|_{\boldsymbol{B}=b}$. Finally, we sometimes use $\mathsf{D}\left(p \,||\, p'\right)$ as short-hand notation for the KL divergence between two Bernoulli random variables with probabilities p, p' (resp.) of being 1.

Problem statement. In this thesis, we study the *Disjointness* problem, defined as follows: for $X^1, \ldots, X^k \in \{0,1\}^n$:

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

Background on information theory. Our lower bound is based on *information theory*, we therefore require the following notions:

Definition 1 (Entropy and conditional entropy)**.** *Let $\boldsymbol{X} \sim \mu$ be a random variable with support χ. Then the* entropy *of $\boldsymbol{X}$ is:*

$$H(\boldsymbol{X}) := \sum_{x \in \chi} \Pr(\boldsymbol{X} = x) \log \frac{1}{\Pr(\boldsymbol{X} = x)}.$$

For two jointly distributed random variables $\boldsymbol{X}$ and $\boldsymbol{Y} \sim \mu_Y$ the conditional entropy *of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is:*

$$H(\boldsymbol{X} \mid \boldsymbol{Y}) := \mathop{\mathbb{E}}_{y \sim \mu_Y} [H(\boldsymbol{X} \mid \boldsymbol{Y} = y)].$$

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

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

Definition 3 (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:*

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

For an event $\mathcal{E}$, we sometimes denote:

$$\mathsf{I}(\boldsymbol{A}\,;\,\boldsymbol{B} \mid \mathcal{E}) := \mathsf{I}(\boldsymbol{A}|_{\mathcal{E}}\,;\,\boldsymbol{B}|_{\mathcal{E}}).$$

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

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

We require the following properties of mutual information, and technical Lemmas:

Property 1. *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$ *be RVs, then:*

$$\mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B}\right) = \mathop{\mathbb{E}}_{\boldsymbol{B}}\left[\mathsf{D}\left(\boldsymbol{A}|_{\boldsymbol{B}}\,||\,\boldsymbol{A}\right)\right].$$

Property 2 (Data processing inequality)**.** *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$ *be RVs, and let* f *be a function defined over the support of* $\boldsymbol{A}$*, then we have that:*

$$\mathsf{I}\left(f\left(\boldsymbol{A}\right)\,;\,\boldsymbol{B}\right) \leq \mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B}\right).$$

Property 3 (Monotonicity of mutual information)**.** *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$*,* $\boldsymbol{B'}$*,* $\boldsymbol{C}$ *be RVs, then:*

$$\mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B}\mid\boldsymbol{C}\right) \leq \mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B},\boldsymbol{B'}\,\middle|\,\boldsymbol{C}\right),$$

Property 4 (Public V.S. private information)**.** *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$*,* $\boldsymbol{B'}$*,* $\boldsymbol{C}$ *be RVs, then:*

$$\mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B}\,\middle|\,\boldsymbol{B'},\boldsymbol{C}\right) \leq \mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B},\boldsymbol{B'}\,\middle|\,\boldsymbol{C}\right).$$

The following is a convenient criterion for testing independence of RVs:

Lemma 1. *Let* $\boldsymbol{R}_1,\ldots,\boldsymbol{R}_n$ *be RVs.* $\forall i\in[n]$ *denote:*

$$\boldsymbol{R}_{-i} := \boldsymbol{R}_1,\ldots,\boldsymbol{R}_{i-1},\boldsymbol{R}_{i+1},\ldots,\boldsymbol{R}_n,$$

Then:

$$\boldsymbol{R}_1,\ldots,\boldsymbol{R}_n \text{ are independent} \iff \forall i\in[n]:\mathsf{I}\left(\boldsymbol{R}_i\,;\,\boldsymbol{R}_{-i}\right)=0.$$

Proof. "$\implies$": Assume that $\boldsymbol{R}_1,\ldots,\boldsymbol{R}_n$ are independent. By definition, $\forall r_1,\ldots,r_n$ such that $\forall j\in[n]$, $r_j\in \mathit{support}(\boldsymbol{R}_j)$, we have that:

$$\Pr\left(\bigwedge_{i=1}^{n}\boldsymbol{R}_i=r_i\right)=\prod_{i=1}^{n}\Pr\left(\boldsymbol{R}_i=r_i\right). \tag{2.1}$$

Now let $i\in[n]$. We will prove that $\forall r_1,\ldots,r_n$ such that $\forall j\in[n]$, $r_j\in \mathit{support}(\boldsymbol{R}_j)$, denote by $\{\boldsymbol{R}_{-i}=r_{-i}\}$ the event:

$$\{\boldsymbol{R}_{-i}=r_{-i}\} := \left\{\bigwedge_{j\in[n]\setminus\{i\}}\boldsymbol{R}_j=r_j\right\},$$

Then we have that:

$$\Pr\left(\boldsymbol{R}_i=r_i\wedge\boldsymbol{R}_{-i}=r_{-i}\right)=\Pr\left(\boldsymbol{R}_i=r_i\right)\cdot\Pr\left(\boldsymbol{R}_{-i}=r_{-i}\right). \tag{2.2}$$

before proving (2.2), note that it implies the claim, as by the properties of KL-divergence, it

will imply that:

$$\mathsf{D}\left(\boldsymbol{R}_i\boldsymbol{R}_{-i} \,||\, \boldsymbol{R}_i \times \boldsymbol{R}_{-i}\right) = 0,$$

Hence we will get that:

$$\begin{aligned}\mathsf{I}\left(\boldsymbol{R}_i\,;\,\boldsymbol{R}_{-i}\right) &= \mathsf{D}\left(\boldsymbol{R}_i\boldsymbol{R}_{-i} \,||\, \boldsymbol{R}_i \times \boldsymbol{R}_{-i}\right)\\ &= 0,\end{aligned}$$

As required. Let us now proceed to prove (2.2). Observe that by (2.1), it is enough to show that:

$$\Pr\left(\boldsymbol{R}_{-i} = r_{-i}\right) = \prod_{j\in[n]\setminus\{i\}} \Pr\left(\boldsymbol{R}_j = r_j\right).$$

Now observe that:

$$\begin{aligned}\Pr\left(\boldsymbol{R}_{-i} = r_{-i}\right) &= \sum_{r_i\in support(\boldsymbol{R}_i)} \Pr\left(\boldsymbol{R}_i = r_i \wedge \boldsymbol{R}_{-i} = r_{-i}\right) && \text{(law of total probability)}\\ &= \sum_{r_i\in support(\boldsymbol{R}_i)} \left(\Pr\left(\boldsymbol{R}_i = r_i\right) \prod_{j\in[n]\setminus\{i\}} \Pr\left(\boldsymbol{R}_j = r_j\right)\right) && \text{(by (2.1))}\\ &= \left(\prod_{j\in[n]\setminus\{i\}} \Pr\left(\boldsymbol{R}_j = r_j\right)\right)\left(\sum_{r_i\in support(\boldsymbol{R}_i)} \Pr\left(\boldsymbol{R}_i = r_i\right)\right)\\ &= \prod_{j\in[n]\setminus\{i\}} \Pr\left(\boldsymbol{R}_j = r_j\right),\end{aligned}$$

Which proves (2.2), as required.

" $\Longleftarrow$ " Now assume that $\forall i \in [n] : \mathsf{I}\left(\boldsymbol{R}_i\,;\,\boldsymbol{R}_{-i}\right) = 0$. $\forall i \in [n]$, denote:

$$\boldsymbol{R}_{<i} := \boldsymbol{R}_1, \ldots, \boldsymbol{R}_{i-1}.$$

Observe that $\forall i \in [n]$:

$$\begin{aligned}\mathsf{I}\left(\boldsymbol{R}_i\,;\,\boldsymbol{R}_{<i}\right) &\leq \mathsf{I}\left(\boldsymbol{R}_i\,;\,\boldsymbol{R}_{-i}\right) && \text{(by the monotonicity of mutual information [3])}\\ &= 0, && \text{(by assumption)}\end{aligned}$$

Hence:

$$\begin{aligned}\mathsf{D}\left(\boldsymbol{R}_i\boldsymbol{R}_{<i} \,||\, \boldsymbol{R}_i \times \boldsymbol{R}_{<i}\right) &= \mathsf{I}\left(\boldsymbol{R}_i\,;\,\boldsymbol{R}_{<i}\right)\\ &= 0.\end{aligned}$$

By the properties of KL divergence, this implies that $\forall r_1, \ldots, r_i$ such that $\forall j \in [i]$, $r_j \in support(\boldsymbol{R}_j)$, we have that:

$$\Pr\left(\boldsymbol{R}_i = r_i \wedge \left(\bigwedge_{j\in[i-1]} \boldsymbol{R}_j = r_j\right)\right) = \Pr\left(\boldsymbol{R}_i = r_i\right) \cdot \Pr\left(\bigwedge_{j\in[i-1]} \boldsymbol{R}_j = r_j\right).$$

By an easy induction on $i \in [n]$, this implies that $\forall r_1, \ldots, r_n$ such that $\forall j \in [n]$, $r_j \in \mathit{support}(\boldsymbol{R}_j)$, we have that:

$$\Pr\left(\bigwedge_{i=1}^{n} \boldsymbol{R}_i = r_i\right) = \prod_{i=1}^{n} \Pr\left(\boldsymbol{R}_i = r_i\right),$$

Which implies that $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ are independent, as required. □

Corollary 1. *Let $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k$ be independent RVs such that $\forall \ell \in [k]$:*

$$\boldsymbol{X}^\ell := \boldsymbol{X}^\ell_1, \ldots, \boldsymbol{X}^\ell_n,$$

For RVs $\boldsymbol{X}^\ell_1, \ldots, \boldsymbol{X}^\ell_n$, and let $I_1, \ldots, I_k \subseteq [n]$, then $\boldsymbol{X}^1_{I_1}, \ldots, \boldsymbol{X}^k_{I_k}$ are independent RVs.

Proof. Observe that $\forall \ell \in [k]$:

$$\begin{aligned}
\mathsf{I}\left(\boldsymbol{X}^\ell_{I_\ell} ; \boldsymbol{X}^1_{I_1}, \ldots, \boldsymbol{X}^{\ell-1}_{I_{\ell-1}}, \boldsymbol{X}^{\ell+1}_{I_{\ell+1}}, \ldots, \boldsymbol{X}^k_{I_k}\right) &\leq \mathsf{I}\left(\boldsymbol{X}^\ell ; \boldsymbol{X}^1, \ldots, \boldsymbol{X}^{\ell-1}, \boldsymbol{X}^{\ell+1}, \ldots, \boldsymbol{X}^k\right) \\
&\qquad \text{(by the monotonicity of mutual information [3])} \\
&= \mathsf{I}\left(\boldsymbol{X}^\ell ; \boldsymbol{X}^{-\ell}\right) \\
&= 0, \qquad \text{(by Lemma [1])}
\end{aligned}$$

Hence by Lemma [1]: $\boldsymbol{X}^1_{I_1}, \ldots, \boldsymbol{X}^k_{I_k}$ are independent RVs. □

Chapter 3

Upper Bound for Small k

In this chapter 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.

3.1 Useful Lemmas

We begin by stating a few technical lemmas that will be useful later.

Lemma 2 (Hölder's inequality)**.** *Let $p, q > 0$ be such that $\frac{1}{p} + \frac{1}{q} = 1$, $n \in \mathbb{N}$, and $a_1, \ldots, a_n$, $b_1 \ldots, b_n \in \mathbb{R}_{\geq 0}$. Then:*

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}},$$

And in particular:

$$\sum_{i=1}^{n} a_i^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} a_i\right)^{\frac{1}{p}} n^{1-\frac{1}{p}}.$$

3.2 Handling Distributions with a Small Expected Intersection

Overview. Recall that we are trying to show a protocol for $Disj_{n,k}$ with communication complexity $\widetilde{O}(k + n^{1-1/k})$ 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 $O(1)$ (note that in general it may be as large as n). 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 $W_i = \top$.

We will refer to this protocol as our *base protocol*.

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 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,

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

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. Fix $n, k \in \mathbb{N}$, and let $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k \subseteq \{0,1\}^n$ be independent RVs representing the players' inputs. For $i \in [n]$, we 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:

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

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

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

Proof. For each $i \in [n]$, by Corollary 1 we have that $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^k$ are independent, hence we have:

$$\mathbb{E}\left[\boldsymbol{Z}_i\right] = \mathbb{E}\left[\bigwedge_{\ell=1}^{k} \boldsymbol{X}_i^\ell\right] = \prod_{\ell=1}^{k} \mathbb{E}\left[\boldsymbol{X}_i^\ell\right].$$

Thus, there exists some $\ell \in [k]$ such that

$$\mathbb{E}\left[\boldsymbol{X}_i^\ell\right] \leq \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k}.$$

We use this to construct the partition: define, for each $\ell \in [k]$,

$$I^0 := \emptyset,$$

$$I^\ell := \left\{ i \in [n] \;\middle|\; \mathbb{E}\left[\boldsymbol{X}_i^\ell\right] \leq \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k} \right\} \setminus I^{\ell-1}.$$

Now we have:

$$\begin{aligned}
\mathbb{E}\left[\sum_{\ell=1}^{k} \left|\boldsymbol{X}^\ell \cap I^\ell\right|\right] &= \mathbb{E}\left[\sum_{\ell=1}^{k}\sum_{i\in I^\ell} \boldsymbol{X}_i^\ell\right] = \sum_{\ell=1}^{k}\sum_{i\in I^\ell} \mathbb{E}\left[\boldsymbol{X}_i^\ell\right] \leq \sum_{\ell=1}^{k}\sum_{i\in I^\ell} \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k} \\
&= \sum_{i=1}^{n} \mathbb{E}[\boldsymbol{Z}_i]^{1/k} && (I^1, \ldots, I^k \text{ is a partition}) \\
&\leq \left(\sum_{i=1}^{n} \mathbb{E}[\boldsymbol{Z}_i]\right)^{1/k} \left(\sum_{i=1}^{n} 1\right)^{1-1/k} && (\text{Hölder's inequality (2)}) \\
&= S^{1/k} n^{1-1/k}. && \square
\end{aligned}$$

We are now ready to describe the base protocol. As we mentioned above, 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 $\boldsymbol{X}^\ell \cap I^\ell$. Let

$$\boldsymbol{T}^0 := \bigcup_{\ell \in [k]} \boldsymbol{X}^\ell \cap I^\ell$$

be the set written on the board. Following this step, only elements in $\boldsymbol{T}^0$ remain candidates for being in the intersection.

(2) We go over the players in order, $\ell = 1, \ldots, k$: player ℓ announces $\boldsymbol{T}^{\ell-1} \setminus \boldsymbol{X}^\ell$, and all players update $\boldsymbol{T}^\ell := \boldsymbol{T}^{\ell-1} \cap \boldsymbol{X}^\ell$.

(3) We announce that the intersection is empty iff $\boldsymbol{T}^k = \emptyset$.

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

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

In the coordinator model, the protocol proceeds as follows.

(1) Each player $\ell \in [k]$ sends $\boldsymbol{X}^\ell \cap I^\ell$ to the coordinator.

(2) The coordinator sends

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

to all players. Following this step, only elements in $\boldsymbol{T}$ remain candidates for being in the intersection.

(3) Each player $\ell \in [k]$ sends $\boldsymbol{X}^\ell \cap \boldsymbol{T}$ to the coordinator.

(4) The coordinator sends the witness $\boldsymbol{W} = \{\boldsymbol{W}_i\}_{i \in \boldsymbol{T}}$ to all players. The witness is defined as follows: for each $i \in [n]$, denote by ℓ the index such that $i \in I^\ell$. Then:

- If $i \notin \boldsymbol{T}$, then we set $\boldsymbol{W}_i = \ell$.
- If $i \in \boldsymbol{T}$, then there are two cases:
  - If for all $\ell' \in [k] \setminus \{\ell\}$, we have that $i \in \boldsymbol{X}^{\ell'}$ then we set $\boldsymbol{W}_i = \top$,
  - If exists $\ell' \in [k] \setminus \{\ell\}$ such that $i \notin \boldsymbol{X}^{\ell'}$ then we set $\boldsymbol{W}_i$ to be the minimal such index ℓ'.

(5) We announce that there is an intersection iff $\boldsymbol{W}$ contains a $\top$.

Lemma 4. *The base protocol always solves disjointness correctly and produces a proper witness. Its expected bit complexity is* $O\left(k + \left(\sum_{i=1}^n \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k}\right) \log n\right) = O(k + S^{1/k} n^{1-1/k} \log n)$ *in the shared blackboard model, and* $O\left(\left(\sum_{i=1}^n \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k}\right) k(\log n + \log k)\right) = O(S^{1/k} n^{1-1/k} k(\log n + \log k))$ *in the coordinator model.*

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

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

Bit complexity. In the first step of the protocol, the set written on the board is $\bigcup_{\ell \in [k]} \boldsymbol{X}^\ell \cap I^\ell$, which has expected size $O\left(\sum_{i=1}^n \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k}\right) = O(S^{1/k} n^{1-1/k})$ by Lemma 3. Therefore, the expected number of bits on the board in this step is $O\left(k + \left(\sum_{i=1}^n \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k}\right) \log n\right) = O(k + S^{1/k} n^{1-1/k} \log n)$. In the second step, each coordinate in $\boldsymbol{T}^0$ is written at most once, so the expected cost is again $O\left(k + \left(\sum_{i=1}^n \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k}\right) \log n\right) = O(k + S^{1/k} n^{1-1/k} \log n)$.

Note that in both protocols (i.e. for the shared blackboard and coordinator models), each player talks at most twice (including the coordinator). □

3.3 The General Protocol

3.3.1 High-Level Plan: Exploiting Correlations

The base protocol handles distributions where the expected intersection size is constant; 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 5. *The expected intersection size* inside I, $\mathbb{E}\left[\sum_{i\in I} \boldsymbol{Z}_i\right]$, *is* constant.

Intuitively, this property holds 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 6. *The remaining coordinates, $[n]\setminus I$, are "positively correlated" with the coordinates in I (otherwise we would add them to I).*

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 recur 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 7. *The set I satisfies that* $\sum_{i=1}^{n} \mathbb{E}\left[\boldsymbol{Z}_i\right]^{1/k} = O\left(n^{1-1/k}\right)$.

Note that by Hölder's inequality this property is indeed weaker than property 5, hence intuitively it should hold for the same reasons. Moreover, observe that by Lemma 4, using property 7, we have that the expected communication cost per iteration is still $O(k + n^{1-1/k} \log n)$, as with property 5. Weakning property 5 will allow us to add more indices to the set I (at each iteration), so the protocol will require less iterations to complete; in the next sections will show formally that $\log\log n / \log k$ iterations are enough.

Our protocol works only for product distributions; in order to recur 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 board, 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. To break any such dependencies, 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 (remaining) players' inputs remain independent.

We note that while the base protocol is described in Section 3.2 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.

3.3.2 Negatively-Correlated Coordinates

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

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

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

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

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 4 (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:*

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

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).[1]

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

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

[1] For two random variables $\boldsymbol{B}_1$,$\boldsymbol{B}_2$, we have that $\mathrm{Cov}(\boldsymbol{B}_1, \boldsymbol{B}_2) = \mathrm{Cov}(1 - \boldsymbol{B}_1, 1 - \boldsymbol{B}_2)$ 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

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

This proves the claim. □

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 6. *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*

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

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

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

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

$$\mathbb{E}\left[(1-\boldsymbol{B}_j)\cdot\prod_{i\in I}(1-\boldsymbol{B}_i)\right] > (1-\varphi(\mathbb{E}[\boldsymbol{B}_j]))\cdot\prod_{i\in I}(1-\varphi(\mathbb{E}[\boldsymbol{B}_i])). \tag{3.1}$$

For the left-hand side, we can write

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

where the last step used the fact that $\{\boldsymbol{B}_i\}_{i\in I}$ are φ-negatively-correlated. Together with (3.1),

we obtain

$$(1-\varphi(\mathbb{E}[\boldsymbol{B}_j]))\cdot\prod_{i\in I}(1-\varphi(\mathbb{E}[\boldsymbol{B}_i]))>\Pr\left(\boldsymbol{B}_j=0\;\middle|\;\boldsymbol{B}_I=\overline{0}\right)\cdot\prod_{i\in I}(1-\varphi(\mathbb{E}[\boldsymbol{B}_i]))\,. \tag{3.2}$$

Since φ's range is $[0,1]$, the term $\prod_{i\in I}(1-\varphi(\mathbb{E}[\boldsymbol{B}_i]))$ is non-negative, and in fact it must be positive (otherwise (3.2) cannot hold). Dividing both sides of (3.2) by this term yields

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

and the claim follows. □

3.3.3 Partitioning the Coordinates

Let us define a concrete $\varphi:[0,1]\to[0,1]$ and a partition $[n]=I\cup J$ of the coordinates, as follows:

$$\varphi(x):=\frac{x^{1/k}}{n^{1-1/k}}, \tag{3.3}$$

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

Lemma 7. *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 5 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 8. *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 6, we obtain

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

3.3.4 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 3.2). Finally, let $J = [n] \setminus I$ be the set on which we will recur if we do not find an intersection inside I.

We prove that conditioned on the witness $\boldsymbol{W}_I$, the players' (remaining) inputs remain independent:

Lemma 9. *For each concrete witness $w \in$ 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 1 it suffices to show that for each $\ell \in [k]$ we have

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

Now observe that the 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 random variables of $\boldsymbol{X}_I^\ell$,$\boldsymbol{X}_I^{-\ell}$ that get assigned under the event $\{\boldsymbol{W}_I = w\}$, and denote by $\overline{a}, \overline{b}$ their respective assignments, i.e. under these notations, 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, 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:

$$\begin{aligned}
\mathsf{I}\left(\boldsymbol{X}_J^\ell \,;\, \boldsymbol{X}_J^{-\ell} \,\middle|\, \boldsymbol{Y}_I^\ell, \boldsymbol{Y}_I^{-\ell}\right) &\le \mathsf{I}\left(\boldsymbol{X}_J^\ell, \boldsymbol{Y}_I^\ell \,;\, \boldsymbol{X}_J^{-\ell}, \boldsymbol{Y}_I^{-\ell}\right) && \text{(property 4)} \\
&\le \mathsf{I}\left(\boldsymbol{X}^\ell \,;\, \boldsymbol{X}^{-\ell}\right) && \text{(property 3)} \\
&= 0 && \text{(Lemma 1+assumption)}
\end{aligned}$$

□

3.3.5 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. For convenience, 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 will execute in several iterations, calling the base protocol once at each iteration, and may stop at the end of an iteration under certain conditions. Assuming the protocol reaches the r-th iteration, we will denote by $\boldsymbol{W}_r$ the witness returned by the base protocol at that iteration, and otherwise we have that $\boldsymbol{W}_r$ equals the empty string. We further denote by $\boldsymbol{W}_{\leq r}$ the concatenation of the witnesses $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_r$.

Notation 1. *Let $W = W_1 \circ \cdots \circ W_r$ be a concatenation of r witnesses, we define $|W| := r$, and $\mu|W$ to be to be the distribution of players' inputs after seeing the witness W, i.e. conditioned on the event $\boldsymbol{W}_{\leq r} = W$.*

The protocol executes as follows:

1. Repeat for $N := \left\lceil \dfrac{\log \log n}{\log k} \right\rceil$ iterations:
   - (1a) The players compute (without communication):
     - i. For each remaining coordinate $i \in U$, the value
     $$\varphi_i := \varphi\left(\mathop{\mathbb{E}}_{\mu|W}[\boldsymbol{Z}_i]\right).$$
     - ii. A maximal subset $I \subseteq U$ such that $\{\boldsymbol{Z}_i\}_{i \in I}$ are φ-negatively-correlated. (If there is more than one such subset, the players choose one using some predetermined mechanism, e.g., they choose the lexicographically-smallest one.) Let $J := U \setminus I$.
   - (1b) If $\Pr_{\mu|W}\left(\bigcap_{\ell=1}^{k} \boldsymbol{X}_I^\ell = \emptyset\right) \leq \epsilon$, the players output "intersecting" and halt the protocol.
   - (1c) Otherwise, the players run the base protocol 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 distributions as follows: they set $U \leftarrow J$, 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.

3.3.6 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.

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, and 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.

Note that at each iteration $r \in [N]$, the base protocol determintes if there is an intersection for some set of coordinates, We will denote this set of coordinates as $\boldsymbol{I}_r \subseteq [n]$ (or the empty set if we do not reach the r-th iteration). Similarly, we define $\boldsymbol{U}_r \subseteq [n]$ to be the set of coordinates that have not been ruled out as being in the intersection in the beginning of the r-th iteration. For simplicity, we define $\boldsymbol{J}_r := \boldsymbol{U}_r \setminus \boldsymbol{I}_r$.

We define $\boldsymbol{R}$ to 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 $\boldsymbol{W}_{\leq r-1}$.

Note that since our protocol is deterministic, for all $r \in [N]$ we have that the witness $\boldsymbol{W}_{\leq r}$ 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.

We will use the following notation:

Notation 2. *for each $r = 1, \ldots, N+1$, Let $\mathcal{W}_r$ be the set of witnesses $w_{<r}$ of length $|w_{<r}| = r-1$ that imply that the protocol reached the r-th iteration, and such that*

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

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

Lemma 10. *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 7 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 4, 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

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

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

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

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 11. *For each* $1 \le r \le N$

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

Proof. Let us consider again the set of witnesses $\mathcal{W}_r$ of length $r-1$ that imply that the protocol reaches Step 1c in the r-th iteration, and invokes the base protocol. First, observe that Lemma 8 implies that for every $w_{<r} \in \mathcal{W}_r$:

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

Let us obtain a similar equation without the conditioning on the event $\boldsymbol{Z}_{\boldsymbol{I}_r} = \overline{0}$; observe that for every $w_{<r} \in \mathcal{W}_r$, if $\boldsymbol{Z}_{\boldsymbol{I}_r} \neq \overline{0}$ then the protocol halts before the $r+1$-th iteration, hence by definition $\boldsymbol{S}_{r+1} \equiv 0$. Hence for every $w_{<r} \in \mathcal{W}_r$:

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

Since this holds point-wise for any witness $w_{<r} \in \mathcal{W}_r$, and since $\boldsymbol{S}_{r+1} \equiv 0$ whenever $\boldsymbol{W}_{<r} \notin \mathcal{W}_r$,

all together we have:

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

Which completes the proof. □

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

Proof. Applying Lemma 11 N times, we get that

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

Since $\mathbb{E}_{\mu}[\boldsymbol{S}_1] \le n$ and $k^N = k^{\lceil \log\log n/\log k\rceil} \ge k^{\log_k \log n} = \log n$, we get that

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

which completes the proof. □

Corollary 3. *We have* $\mathbb{E}_{\mu}[\boldsymbol{C}_{N+1}] = O(k + n^{1-1/k}\log n)$.

Proof. Whenever $\boldsymbol{R} \le N$, we do not reach Step 2 of the protocol, and both $\boldsymbol{C}_{N+1} \equiv 0$ and $\boldsymbol{S}_{N+1} \equiv 0$ by definition, hence we have that:

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

and similarly we have that:

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

When $\boldsymbol{R} = N+1$, we do call the base protocol, and by Lemma 4 the expected communication cost is:

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

All together we have:

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

Where the last equality follows from Corollary 2. □

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

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

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

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

Recall that if $\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r}$, then the protocol will halt in Step 1b of the r-th iteration, and declare that the player's inputs are intersecting.

Observe that the protocol may only err if in some iteration r, we have that $\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r}$ but the players' inputs are disjoint. In addition, observe that the events $\{\boldsymbol{W}_{<1} \in \mathcal{W}_{<\epsilon,1}\}, \ldots,$ $\{\boldsymbol{W}_{<N} \in \mathcal{W}_{<\epsilon,N}\}$ are disjoint events, as the event $\boldsymbol{W}_{<r} \in \mathcal{W}_{<\epsilon,r}$ implies that the protocol halts in the r-th iteration. All together this implies that:

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

Chapter 4

Upper Bound for Large k

When $k = \Omega(\log n)$, it is no longer worthwhile to use our protocol from Chapter 3, as $n^{1-1/k} = \Theta(n)$ in this case. Instead we give a much simpler protocol that exploits the fact that when working with a product distribution, any coordinate that has $\Omega(\log n)$ expected zeroes (across all players) also has negligible probability of being in the intersection.

The following lemma is a simple special case of Chernoff, which we prove for the sake of completeness:

Lemma 12. *Fix $\alpha \in (0, a)$, and let $\boldsymbol{A} = \boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$, where $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ are independent Bernoulli random variables satisfying*

$$\mathbb{E}\left[\sum_i^m \boldsymbol{A}_i\right] > \ln(1/\alpha). \tag{4.1}$$

Then

$$\Pr\left(\boldsymbol{A} = \overline{0}\right) < \alpha.$$

Proof. We can write

$$\begin{aligned}\Pr\left(\boldsymbol{A} = \overline{0}\right) = \prod_{i=1}^m \left(1 - \Pr\left(\boldsymbol{A}_i = 1\right)\right) &\leq e^{-\sum_{i=1}^m \Pr(\boldsymbol{A}_i = 1)} \\ &= e^{-\sum_{i=1}^m \mathbb{E}[\boldsymbol{A}_i]} < e^{-\ln(1/\alpha)} = \alpha.\end{aligned}$$

□

Corollary 4. *Let $i \in [n]$ be a coordinate such that $\mathbb{E}\left[\left|\left\{\ell \in [k] : \boldsymbol{X}_i^\ell = 0\right\}\right|\right] > \ln(n/\epsilon)$. Then $\Pr\left(i \in \bigcap_{\ell \in [k]} \boldsymbol{X}^\ell\right) < \epsilon/n$.*

Proof. Follows from the lemma, by taking $\boldsymbol{A} = (1 - \boldsymbol{X}_i^1, \ldots, 1 - \boldsymbol{X}_i^k)$. Observe that the random variables $1 - \boldsymbol{X}_i^1, \ldots, 1 - \boldsymbol{X}_i^k$ are indeed independent, since for every $\ell \in [k]$, we have that:

$$\begin{aligned}\mathsf{I}\left(1 - \boldsymbol{X}_i^\ell \,;\, 1 - \boldsymbol{X}_i^1, \ldots, 1 - \boldsymbol{X}_i^{\ell-1}, 1 - \boldsymbol{X}_i^{\ell+1}, \ldots, 1 - \boldsymbol{X}_i^k\right) &\leq \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_i^{-\ell}\right) && \text{(property 2)} \\ &= 0 && \text{(Lemma 1)}\end{aligned}$$

Hence by Lemma 1, the RVs $1 - \boldsymbol{X}_i^1, \ldots, 1 - \boldsymbol{X}_i^k$ are independent. □

The protocol. Corollary 4 implies a simple deterministic simultaneous protocol: "ignore" any coordinate where the number of expected zeroes exceeds n/ϵ, and send all the zeroes in the remaining coordinates. Let $I \subseteq [n]$ be the set of coordinates i such that $\Pr\left(i \in \bigcap_{t\in[k]} \boldsymbol{X}^t\right) \geq \epsilon/n$. Each player $\ell \in [k]$ executes the following:

1. Let $S_\ell = \left\{i \in I \;\middle|\; \boldsymbol{X}_i^\ell = 0\right\}$.
2. Announce S_ℓ, by writing it on the board (for the shared blackboard model) or sending it to the coordinator (in the coordinator model).
3. Announce "intersecting" iff $\bigcap_{t\in[k]} S_t \neq \emptyset$. (This is evaluated by each player in the shared blackboard model, or by the coordinator in the coordinator model.)

Lemma 13. *The protocol errs with probability at most ϵ and communicates $O\left(k + n\log(n/\epsilon)\log n\right)$ bits in expectation.*

Proof. By Corollary 4, for each coordinate $i \in I$ we have $\mathbb{E}\left[\left|\left\{\ell \in [k] : \boldsymbol{X}_i^\ell = 0\right\}\right|\right] \leq \ln(n/\epsilon)$. Thus, the expected number of zeroes in $\boldsymbol{X}_I^1, \ldots, \boldsymbol{X}_I^\ell$ is bounded by $|I|\cdot\ln(n/\epsilon) = O\left(n\left(\log n + \log(1/\epsilon)\right)\right)$. Hence the expected communication complexity of the protocol is $O\left(k + n\left(\log n + \log(1/\epsilon)\right)\log n\right)$.

As for the error, observe that the protocol errs iff there is an intersection outside I. However, I consists only of coordinates $i \in [n]$ such that $\Pr\left(i \in \bigcap_{\ell\in[k]} \boldsymbol{X}^\ell\right) < \epsilon/n$. By union bound,

$$\Pr\left(\exists i \notin I : i \in \bigcap_{\ell\in[k]} \boldsymbol{X}^\ell\right) \leq \sum_{i\notin I} \Pr\left(i \in \bigcap_{\ell\in[k]} \boldsymbol{X}^\ell\right) < n\cdot(\epsilon/n) = \epsilon. \qquad \square$$

Chapter 5

Lower Bound

In this chapter we prove our lower bound of $\Omega(n^{1-1/k}/k^2)$ for Disjointness under product distributions, 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 $[n]$.

The lower bound is information theoretic. We begin by introducing the notation that will be used throughout the proof, reviewing the relevant definitions, and stating some technical lemmas.

5.1 Preliminaries

5.1.1 Useful Inequalities

The following technical facts will be used in the proof.

Fact 1. *For each $m \in \mathbb{R}^+$ we have $\left(1 - \frac{1}{m}\right)^m \leq 1/e$.*

Fact 2 (Bernoulli's Inequality)**.** *For each $t \geq 1$ and $x \in [0, 1]$,*

$$(1 - x)^t \geq 1 - xt.$$

5.1.2 Information Theory

We state several technical lemmas that will be used to bound the effect of conditioning on various random variables and events in the proof, as well as to move between mutual information, KL divergence, and probabilities of Bernoulli random variables.

Lemma 14. *Let $p \in (0, 1/3)$, $p' \in (0, 1)$ and $\alpha \in (0, 1/2)$. If $\mathsf{D}(p \,||\, p') \leq p\alpha^2/40 \ln 2$, then $1 - \alpha \leq p'/p \leq 1 + \alpha$.*

Lemma 15. *Fix $\alpha \in (0, 1/2)$. Let $\boldsymbol{A} \sim \mathsf{B}(p)$, 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 $\boldsymbol{A}|_{\boldsymbol{B}=b} \sim \mathsf{B}(p')$. If*

$$\mathsf{I}(\boldsymbol{A}\,;\,\boldsymbol{B}) \leq \Pr(\boldsymbol{B} = b) \cdot \frac{p\alpha^2}{40 \ln 2},$$

then

$$\frac{p'}{p} \in (1-\alpha, 1+\alpha)\,.$$

Lemma 16. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ be random variables such that $\boldsymbol{A}$ is independent of $\boldsymbol{C}$. Then*

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B}) + \mathsf{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{B}) = \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}),$$

and in particular:

1. *(Conditioning on an independent variable does not decrease information):*

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B}) \leq \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}),$$

2. *(Reversal lemma):*

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{B}) \leq \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}).$$

Claim 1 (Removing obstructions)**.** *Let $\boldsymbol{X}, \boldsymbol{X'}, \boldsymbol{Y}, \boldsymbol{Y'}$ be RVs, such that the pair $(\boldsymbol{X}, \boldsymbol{X'})$ is independent of the pair $(\boldsymbol{Y}, \boldsymbol{Y'})$, and let $y \in$ support$(\boldsymbol{Y})$. Then:*

$$\mathsf{I}\left(\boldsymbol{X}\,;\,\boldsymbol{Y'X'} \,\middle|\, \boldsymbol{Y}=y\right) = \mathsf{I}\left(\boldsymbol{X}\,;\,\boldsymbol{X'} \,\middle|\, \boldsymbol{Y}=y\right) = \mathsf{I}\left(\boldsymbol{X}\,;\,\boldsymbol{X'}\right).$$

5.1.3 Properties of Communication Protocols

Lemma 17 (Protocols do not create dependencies between their inputs)**.** *Let $\boldsymbol{M}$ denote the transcript of a deterministic protocol Π over the inputs $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k$, and let $\ell \in [k]$. Then*

$$\mathsf{I}(\boldsymbol{X}^\ell; \boldsymbol{X}^{-\ell} \mid \boldsymbol{M}) \leq \mathsf{I}(\boldsymbol{X}^\ell; \boldsymbol{X}^{-\ell}).$$

5.2 Setup: Constants and Distributions

Let:

$$\begin{aligned}\epsilon_1 &:= \frac{1}{800e}\\ \alpha_M &:= \frac{1}{8e}\\ \alpha_J &:= 1/2.\end{aligned}$$

Let $C > 0$ be a constant that satisfies all the following requirements:

$$\frac{1}{\alpha_J \alpha_M C} \leq \frac{1}{2560 \ln 2} \leq \frac{1}{1280 \ln 2} < \frac{1}{100 \ln 2}. \tag{5.1}$$

Note that it is enough to take:

$$C := \frac{2560 \ln 2}{\alpha_J \alpha_M}. \tag{5.2}$$

The number of players and the input size. Let $N \in \mathbb{N}$ be such that for every $n > N$,

$$\left(1 - \frac{1}{n}\right)^n \geq \frac{1}{2e}. \tag{5.3}$$

We consider only input sizes $n > N$. As for the number of players k, we require

$$2 \leq k \leq \frac{\log n}{\log 6}.$$

Note this implies that:

$$\frac{1}{n^{1/k}} \leq \frac{1}{6}. \tag{5.4}$$

The input distribution. Fix $n > N$. The input distribution for our lower bound is given by

$$\mu = \mu_n^k := \mu_{X^1} \times \ldots \times \mu_{X^k},$$

where

$$\mu_{X^1} = \ldots = \mu_{X^k} = \mathrm{Ber}\left(\frac{1}{n^{1/k}}\right)^n.$$

That is, all the input bits are iid Bernoulli random variables with probability $1/n^{1/k}$ of being 1.

Let $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k$ be the players' respective inputs. Let $\mathcal{E}_\emptyset$ denote the event that $\bigcap_{\ell \in [k]} \boldsymbol{X}^\ell = \emptyset$.

Property 8. *Under μ we have $\Pr(\mathcal{E}_\emptyset) \geq 1/(2e)$.*

Proof. Since the players' inputs are independent, and the coordinates of each input are also independent,

$$\Pr\left(\bigcap_{\ell \in [k]} \boldsymbol{X}^\ell = \emptyset\right) = \prod_{i=1}^{n} \left(1 - \Pr\left(\bigwedge_{\ell \in [k]} \boldsymbol{X}_i^\ell = 1\right)\right) = \left(1 - \left(\frac{1}{n^{1/k}}\right)^k\right)^n = \left(1 - \frac{1}{n}\right)^n \geq \frac{1}{2e},$$

where the last inequality holds by our choice of $n > N$. □

5.3 Proof of the Lower Bound

The formal statement of our lower bound is as follows:

Theorem 3. *Every deterministic protocol Π for $Disj_{n,k}$ in the shared blackboard model with transcript length at most $n^{1-\frac{1}{k}}/(Ck^2)$ errs with probability at least ϵ_1 on the the input distribution μ.*

In order to prove the theorem, fix a deterministic protocol Π for $Disj_{n,k}$ with transcript length at most $n^{1-\frac{1}{k}}/(Ck^2)$. We will show that Π's error must be at least ϵ_1. We assume that Π's error is at most $1/(8e)$, as otherwise Π's error trivially exceeds $\epsilon_1 < 1/(8e)$. Let the random variable $\boldsymbol{M} = \Pi\left(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k\right)$ denote the transcript of Π on inputs $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k \in \{0,1\}^n$.

5.3.1 High Level Overview of the Proof

We will show that if the protocol's transcript is short, then the protocol errs with some constant probability, implying that protocols that ensure low (enough) error require long transcripts.

We will start by defining a set of "good" transcripts G. Observe that under our input distribution μ, a low-error protocol must output "non-intersecting" with constant probability, because μ has constant probability that the inputs will not intersect. Moreover, if the protocol's transcript is short, we expect it to give little information about the typical coordinate $\boldsymbol{X}_i^\ell$ and its relation to other coordinates. We therefore define G to be the set of transcripts satisfying these two requirements, and note that the protocol's transcript is in G with constant probability. We will show that every transcript in G has a constant error probability, which will complete the proof.

Next, given a transcript $m \in G$, we continue by defining a *good set of coordinates* $J(m)$. By definition of G, the transcript m does not tell us much about most coordinates $i \in [n]$, so we let $J(m)$ be a set of $\Omega(n)$ coordinates that m does not convey much information about, and which remain "nearly independent" given m.

Next, for all and indices $i \in J(m)$, we denote by $\mathcal{E}_{\emptyset,<i}$ the event that an intersection did not occur at any index of $J(m)$ lower than i. We continue by showing that for all players $\ell \in [k]$, we have that:

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, \mathcal{E}_{\emptyset,<i}, \boldsymbol{X}_i^{<\ell} = \overline{1}\right) \approx \Pr\left(\boldsymbol{X}_i^\ell = 1\right), \tag{5.5}$$

as proving (5.5) will imply that every $i \in J(m)$ has intersection probability roughly $1/n$, hence bounding away from 1 the probability that the inputs are indeed disjoint, implying that the transcript errs with constant probability.

Now, observe that as a first step towards proving (5.5), by the definition of G we have that for all transcripts $m \in G$, indices $i \in J(m)$, and players $\ell \in [k]$:

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m\right) \approx \Pr\left(\boldsymbol{X}_i^\ell = 1\right).$$

Secondly, since the random variables $\boldsymbol{X}_i^\ell$ and $\boldsymbol{X}_i^{<\ell}$ are independent (even conditioned on $\boldsymbol{M} = m$), we have that:

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}\right) = \Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m\right),$$

So it suffices to show that:

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}, \mathcal{E}_{\emptyset,<i}\right) \approx \Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}\right). \tag{5.6}$$

In order to show that (5.6) holds, we first show that for all coordinates $i \in [n]$ and players $l \in [k]$ we have that:

$$\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \mathbf{1}_{\mathcal{E}_{\emptyset,<i}} \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}\right) \leq \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M} = m\right). \tag{5.7}$$

Note that (5.7), together with the definition of $J(m)$ tells us that the value of $\boldsymbol{X}_i^\ell$ does not give a lot of information about whether the event $\mathcal{E}_{\emptyset,<i}$ occurred or not (and vice versa), even con-

ditioned on the event $\{\boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}\}$.[1] Intuitively this should imply (5.6), but technically, since mutual information is in fact an average over events, we still need to show that the event $\mathcal{E}_{\emptyset,<i}|_{\boldsymbol{M}=m,\boldsymbol{X}_i^{<\ell}=\overline{1}}$ is likely (say, occurs with constant probability).

In order to show that the event $\mathcal{E}_{\emptyset,<i}|_{\boldsymbol{M}=m,\boldsymbol{X}_i^{<\ell}=\overline{1}}$ is likely, first observe that the event $\mathcal{E}_{\emptyset,<i}|_{\boldsymbol{M}=m}$ must be likely, as it upper bounds the probability that the inputs are disjoint (and hence that the transcript answers correctly). So it is only left to verify that the supplementary condition $\boldsymbol{X}_i^{<\ell} = \overline{1}$ does not reduce the probability of $\mathcal{E}_{\emptyset,<i}$ too much. We now consider a sequence of events: $\{\boldsymbol{X}_i^{<2} = \overline{1}\}, \{\boldsymbol{X}_i^{<3} = \overline{1}\}, \ldots, \{\boldsymbol{X}_i^{<\ell} = \overline{1}\}$, and aim to show that for all $1 < \ell' < \ell$:

$$\delta_{\mathrm{SD}}\left(\mathbf{1}_{\mathcal{E}_{\emptyset,<i}|_{\boldsymbol{M}=m,\boldsymbol{X}_i^{<\ell'}=\overline{1}}}, \mathbf{1}_{\mathcal{E}_{\emptyset,<i}|_{\boldsymbol{M}=m,\boldsymbol{X}_i^{<\ell'+1}=\overline{1}}}\right) = O\left(\frac{1}{k}\right) \tag{5.8}$$

(for a suitably small constant), as this together with the triangle inequality will imply that:

$$\Pr\left(\mathcal{E}_{\emptyset,<i} \;\middle|\; \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}\right) = \Omega(1).$$

Finally, we show that eq. (5.8) holds by appealing to (5.7) again, and applying Pinsker's inequality.[2]

5.3.2 Good Transcripts

Denote by $G \subseteq \mathit{support}(\boldsymbol{M})$ the set of transcripts m that satisfy all the following requirements:

1. The output of the protocol upon producing transcript m is "non-intersecting".
2. The amount of information that the transcript conveys about the input, and similarly the amount of dependencies it creates between the coordinates of any individual input, is "not much higher than average":

$$\sum_{i=1}^{n}\sum_{\ell\in[k]}\left(\mathsf{D}\left(\boldsymbol{X}_i^{\ell}|_{\boldsymbol{M}=m} \;\middle\|\; \boldsymbol{X}_i^{\ell}\right) + \mathsf{I}(\boldsymbol{X}_i^{\ell}\,;\, \boldsymbol{X}_{<i}^{\ell} \mid \boldsymbol{M} = m)\right) \leq \frac{1}{\alpha_M}\cdot\frac{n^{1-\frac{1}{k}}}{Ck^2}. \tag{5.9}$$

To show that there is a constant probability of getting a good transcript, we show that there is constant probability that the first condition holds, and a very high probability that the second condition holds. For convenience, let us denote by G_i the set of transcripts that satisfy requirement $i \in \{1, 2\}$.

Claim 2. *We have that:*

$$\Pr\left(\boldsymbol{M} \in G_1\right) \geq 1/(4e).$$

Proof. Recall that by assumption, it holds that:

$$\Pr\left(\boldsymbol{M} \text{ errs}\right) \leq \frac{1}{8e}.$$

[1] This roughly follows from the fact that $\mathcal{E}_{\emptyset,<i}$ is a function of $\boldsymbol{X}_{J_{<i}}$, which by the definition of J does not give a lot of information about $\boldsymbol{X}_i^{\ell}$.

[2] In order to move from mutual information to KL-divergence (which is needed for Pinsker's inequality), we use the fact that the event $\{\boldsymbol{X}_i^{\ell'+1} = 1\}$ is not "too unlikely".

Now, for the probability that $\boldsymbol{M} \in G_1$, that is, that the transcript announces "non-intersecting", we have

$$\Pr\left(\boldsymbol{M} \in G_1\right) \geq \Pr\left(\mathcal{E}_\emptyset \wedge \boldsymbol{M} \text{ does not err}\right) \geq \Pr\left(\mathcal{E}_\emptyset\right) - \Pr\left(\boldsymbol{M} \text{ errs}\right) \geq \frac{1}{2e} - \frac{1}{8e} > \frac{1}{4e}. \qquad \square$$

Claim 3. *We have* $\Pr\left(\boldsymbol{M} \in G_2\right) \geq 1 - \alpha_M$.

Proof. First note that

$$\mathsf{I}\left(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k \,;\, \boldsymbol{M}\right) \leq H\left(\boldsymbol{M}\right) \leq |M| \leq \frac{n^{1-\frac{1}{k}}}{Ck^2}. \tag{5.10}$$

On the other hand,

$$\begin{aligned}
\mathsf{I}\left(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^k \,;\, \boldsymbol{M}\right) &= \sum_{\ell \in [k]} \mathsf{I}\left(\boldsymbol{X}^\ell \,;\, \boldsymbol{M} \,\middle|\, \boldsymbol{X}^{<\ell}\right) \\
&\geq \sum_{\ell \in [k]} \mathsf{I}\left(\boldsymbol{X}^\ell \,;\, \boldsymbol{M}\right) \qquad (\boldsymbol{X}^\ell, \boldsymbol{X}^{<\ell} \text{ independent + Lemma 16}) \\
&= \sum_{\ell \in [k]} \sum_{i=1}^{n} \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{M} \,\middle|\, \boldsymbol{X}_{<i}^\ell\right).
\end{aligned}$$

By the chain rule, for each $\ell \in [k]$ and $i \in [n]$,

$$\begin{aligned}
\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{M} \,\middle|\, \boldsymbol{X}_{<i}^\ell\right) &= \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{M}\right) + \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M}\right) \qquad \text{(Lemma 16)} \\
&= \mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right)\right] + \mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M} = m\right)\right]
\end{aligned}$$

Together with (5.10), we have

$$\begin{aligned}
\frac{n^{1-\frac{1}{k}}}{Ck^2} &\geq \sum_{\ell \in [k]} \sum_{i=1}^{n} \mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right)\right] + \mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M} = m\right)\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\sum_{\ell \in [k]} \sum_{i=1}^{n} \mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) + \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M} = m\right)\right] \\
&\qquad \text{(linearity of expectation)} \\
&= \mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\sum_{i=1}^{n} \sum_{\ell \in [k]} \mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) + \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M} = m\right)\right].
\end{aligned}$$

Finally, since mutual information and KL-divergence are non-negative, Markov's inequality yields:

$$\Pr_{\boldsymbol{M}}\left(\sum_{i=1}^{n} \sum_{\ell \in [k]} \mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) + \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M} = m\right) \geq \frac{1}{\alpha_M} \cdot \frac{n^{1-\frac{1}{k}}}{Ck^2}\right) \leq \alpha_M. \qquad \square$$

All together, we see that there is decent probability of getting a good transcript:

Lemma 18. *We have* $\Pr(\boldsymbol{M} \in G) \geq 1/(8e)$.

Proof. We have shown that $\Pr(\boldsymbol{M} \in G_1) \geq 1/(4e)$ and $\Pr(\boldsymbol{M} \notin G_2) \leq \alpha_M = 1/(8e)$. It follows that

$$\Pr(\boldsymbol{M} \in G) = \Pr(\boldsymbol{M} \in G_1 \wedge \boldsymbol{M} \in G_2) \geq \Pr(\boldsymbol{M} \in G_1) - \Pr(\boldsymbol{M} \notin G_2) \geq 1/(8e). \qquad \square$$

5.3.3 Good Subset of Indices and Its Properties

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

Lemma 19. *For each $m \in G$, there exists a set $J = J(m) \subseteq [n]$ of size $|J| \geq (1 - \alpha_J)n$, such that for all $i \in J$, $\ell \in [k]$:*

1. $\mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) \leq 1/(\alpha_J \alpha_M C n^{1/k} k^2)$, *and*
2. $\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M} = m\right) \leq 1/(\alpha_J \alpha_M C n^{1/k} k^2)$.

Proof. Denote by $J \subseteq [n]$ the set of indices that satisfy:

$$\sum_{\ell \in [k]} \left(\mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) + \mathsf{I}(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \mid \boldsymbol{M} = m) \right) \leq \frac{1}{\alpha_J} \cdot \frac{1}{\alpha_M} \cdot \frac{1}{C n^{1/k} k^2}.$$

By the definition of G (5.3.2) and Markov's inequality, $|J| \geq (1-\alpha_J)n$. Since mutual information and KL divergence are non-negative, we have, for each $i \in J$ and $\ell \in [k]$,

1. $\mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) \leq 1/(\alpha_J \alpha_M C n^{1/k} k^2)$, and
2. $\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M} = m\right) \leq 1/(\alpha_J \alpha_M C n^{1/k} k^2)$.

Finally, observe that by the monotonicity of mutual information (property 3):

$$\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M} = m\right) \leq \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{<i}^\ell \,\middle|\, \boldsymbol{M} = m\right),$$

and this completes the proof.

$\square$

We use J as short-hand notation for $J(m)$, when m is clear from the context. We also refer to the indices $i \in J$ as "good indices" (or "good coordinates"), but we note that this is slight abuse of the definition, because it is the set J that is good (not the individual indices in it).

One key property of the coordinates in J is that since $\boldsymbol{M}$ does not give much information about the indices in $J(\boldsymbol{M})$, their posterior probabilities given $\boldsymbol{M}$ are close to the prior:

Lemma 20. *For each $m \in G$, $i \in J(m)$ and $\ell \in [k]$,*

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

Proof. By definition of J, for each $i \in J$ and $\ell \in [k]$,

$$\mathsf{D}\left(\boldsymbol{X}_i^\ell|_{\boldsymbol{M}=m} \,\middle\|\, \boldsymbol{X}_i^\ell\right) \leq \frac{1}{\alpha_J \alpha_M C n^{1/k} k^2} \leq \frac{1}{2 \cdot 40 \ln 2 \cdot n^{1/k} 16 k^2},$$

where the last inequality follows from the definition of C (eq. (5.1)). The claim now follows from Lemma 14, taking $p = 1/n^{1/k} \in (0, 1/3)$ and $\alpha = 1/(4k) \in (0, 1/2)$. □

5.3.4 Adding Up the Intersection Probabilities of the Good Coordinates

Fix a good transcript, $m \in G$. Our ultimate goal is to show the following:

Lemma 21. *For all $m \in G$,*

$$\Pr\left(\neg\mathcal{E}_\emptyset \mid \boldsymbol{M} = m\right) \geq 1 - e^{-1/4} = \Omega(1).$$

Since all good transcripts $m \in G$ output "non-intersecting", and $\Pr(\boldsymbol{M} \in G)$ is fairly high, this means the protocol's error is large.

Conditioned on $\boldsymbol{M} = m$ where $m \in G$, the bits $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^k$ at any good coordinate $i \in J(m)$ are close to their prior distribution (Lemma 20). Moreover, since we are working with a communication protocol, the bits $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^k$ remain independent conditioned on $\boldsymbol{M} = m$. Therefore the probability of an intersection, $\boldsymbol{X}_i = \overline{1}$, is close to its prior of $1/n$. Since there are $|J(m)| = \Theta(n)$ good coordinates, the expected intersection size in $J(m)$ is $\Theta(1)$. If the coordinates in $J(m)$ were still independent of one another, we could now conclude that there is a constant probability of getting an intersection in $J(m)$, but unfortunately, conditioned on $\boldsymbol{M} = m$, these coordinates are *not* independent – they are only "close" to independent (by definition of the set of good coordinates). Thus, to prove that there is a constant probability of having an intersection in $J(m)$, we "collect" the coordinates one-by-one, handle the dependencies between them, and show that the probability of an intersection roughly "adds up" over the coordinates.

Let $\mathcal{E}_{\emptyset,<i}$ be the event that there is no intersection at any coordinate inside $J(m)$ that is smaller than i. The key lemma that allows us to "collect" the intersection probabilities is the following:

Lemma 22. *For each $m \in G$ and $i \in J(m)$, if $\Pr(\mathcal{E}_\emptyset \mid \boldsymbol{M} = m) > 0.6$, then it holds that*

$$\Pr\left(\boldsymbol{X}_i = \overline{1} \,\middle|\, \boldsymbol{M} = m, \mathcal{E}_{\emptyset,<i}\right) \geq \frac{1}{2n}.$$

Lemma 22 asserts that for any good transcript m and coordinate $i \in J(m)$, conditioned on $\boldsymbol{M} = m$ and on having no intersection in the coordinates of J below i (the event $\mathcal{E}_{\emptyset,<i}$), the probability of having an intersection in coordinate i is at least $1/(2n)$. To prove the lemma, we iterate over the players $\ell \in [k]$, and prove that even conditioned on $\boldsymbol{X}_i^{<\ell} = \overline{1}$ (as well as $\boldsymbol{M} = m, \mathcal{E}_{\emptyset,<i}$), we still have good probability that $\boldsymbol{X}_i^\ell = 1$. All together, this implies that with good probability, $\boldsymbol{X}_i = \overline{1}$. Before proving Lemma 22 formally, we state two lemmas that will be used in the proof.

Since we start out with a product distribution, any dependencies between $\boldsymbol{X}_i^\ell$ and $\boldsymbol{X}_i^{<\ell}$ must arise from conditioning on $\boldsymbol{M} = m, \mathcal{E}_{\emptyset,<i}$. However, conditioning on the transcript of a protocol

does not create any dependencies between the players' inputs, so it is really the event $\mathcal{E}_{\emptyset,<i}$ that is "problematic". Thus, a key step in the proof is to show that conditioning on $\mathcal{E}_{\emptyset,<i}$ does not create strong dependencies. We start by showing that the event $\mathcal{E}_{\emptyset,<i}$ does not give a lot of information about the i-th coordinate:

Lemma 23. *For every coordinate $i \in [n]$ and player $\ell \in [k]$ we have that:*

$$\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \mathbf{1}_{\mathcal{E}_{\emptyset,<i}} \,\middle|\, \boldsymbol{M}=m, \boldsymbol{X}_i^{<\ell}=\overline{1}\right) \le \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M}=m\right).$$

Proof. By the data processing inequality,

$$\begin{aligned}\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \mathbf{1}_{\mathcal{E}_{\emptyset,<i}} \,\middle|\, \boldsymbol{M}=m, \boldsymbol{X}_i^{<\ell}=\overline{1}\right) &\le \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{J_{<i}}^\ell \boldsymbol{X}_{J_{<i}}^{-\ell} \,\middle|\, \boldsymbol{M}=m, \boldsymbol{X}_i^{<\ell}=\overline{1}\right) \qquad \text{(Property 2)}\\ &= \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M}=m\right),\end{aligned}$$

Where the last equality follows from Claim 1, by taking $X := \boldsymbol{X}_i^\ell$, $X' := \boldsymbol{X}_J^\ell$, $Y' := \boldsymbol{X}_J^{-\ell}$, $(Y=y) := (\boldsymbol{X}_i^{<\ell}=\overline{1})$. □

We then use Lemma 23 to show that conditioning on some (or all) of the input in coordinate i does not reduce the probability of the event $\mathcal{E}_{\emptyset,<i}$ by much. Specifically, if $\mathcal{E}_\emptyset$ has high probability given $\boldsymbol{M}=m$ (and therefore so does $\mathcal{E}_{\emptyset,<i}$, which is implied by $\mathcal{E}_\emptyset$), then $\mathcal{E}_{\emptyset,<i}$ retains high probability even after conditioning on some bits in coordinate i being 1:

Lemma 24. *If $\Pr\left(\mathcal{E}_\emptyset \mid \boldsymbol{M}=m\right) > 0.6$, then for each $m \in G$, $i \in J(m)$ and $\ell \in [k]$*

$$\Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M}=m, \boldsymbol{X}_i^{<\ell}=\overline{1}\right) > 0.5.$$

Before proving Lemma 24, let us show how it is used to prove Lemma 22.

Proof of Lemma 22. As we said above, we would like to show that for all $i \in J$, $\ell \in [k]$,

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M}=m, \mathcal{E}_{\emptyset,<i}, \boldsymbol{X}_i^{<\ell}=\overline{1}\right) \ge \frac{\left(1-\frac{1}{4k}\right)^2}{n^{1/k}}, \tag{5.11}$$

as this will easily imply Lemma 22.

First, since conditioning on the transcript of a protocol does not create dependence between the inputs, we have

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M}=m, \boldsymbol{X}_i^{<\ell}=\overline{1}\right) = \Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M}=m\right),$$

and by Lemma 20,

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M}=m\right) \in \left(\frac{1-\frac{1}{4k}}{n^{1/k}}, \frac{1+\frac{1}{4k}}{n^{1/k}}\right). \tag{5.12}$$

In particular, then,

$$\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M}=m, \boldsymbol{X}_i^{<\ell}=\overline{1}\right) \ge \frac{1}{2n^{1/k}}. \tag{5.13}$$

Next, we carefully introduce the conditioning on $\mathcal{E}_{\emptyset,<i}$, using Lemma 15 to bound the effect. By choice of J, we know that there is not much dependence between the coordinates $J_{<i}$ and coordinate i given $\boldsymbol{M} = m$, and since $\mathcal{E}_{\emptyset,<i}$ is an event that depends only on the coordinates in $J_{<i}$, we have

$$\begin{aligned}
\mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \mathbf{1}_{\mathcal{E}_{\emptyset,<i}} \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}\right) &\le \mathsf{I}\left(\boldsymbol{X}_i^\ell \,;\, \boldsymbol{X}_{J_{<i}}^\ell \,\middle|\, \boldsymbol{M} = m\right) && \text{(Lemma 23)} \\
&\le \frac{1}{\alpha_J \alpha_M C n^{1/k} k^2} && \text{(definition of } J \text{ (19))} \\
&\le \Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell}\right) \cdot \frac{\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m\right) \cdot (1/(16k^2))}{40 \ln 2}. \\
& && \text{(by (5.13), Lemma 24, and the definitions of } \alpha_J, \alpha_M, C \text{ (5.1))}
\end{aligned}$$

Therefore, taking $\alpha = 1/(4k)$, Lemma 15 yields

$$\begin{aligned}
\Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}, \mathcal{E}_{\emptyset,<i}\right) &\ge (1-\alpha) \cdot \Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m\right) \\
&\ge \left(1 - \frac{1}{4k}\right)^2 \cdot \frac{1}{n^{1/k}}. && \text{(by (5.12))}
\end{aligned}$$

To complete the proof, it remains only to observe that

$$\Pr\left(\boldsymbol{X}_i = \overline{1} \,\middle|\, \boldsymbol{M} = m, \mathcal{E}_{\emptyset,<i}\right) = \prod_{\ell=1}^{k} \Pr\left(\boldsymbol{X}_i^\ell = 1 \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}, \mathcal{E}_{\emptyset,<i}\right) \ge \left(\frac{\left(1 - \frac{1}{4k}\right)^2}{n^{1/k}}\right)^k \ge \frac{1}{2n}.$$

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

Finally, let us prove Lemma 24, and conclude the proof of the lower bound.

Proof of Lemma 24. Recall that by assumption it holds that $\Pr\left(\mathcal{E}_\emptyset \mid \boldsymbol{M} = m\right) > 0.6$. The event $\mathcal{E}_{\emptyset,<i}$ is implied by $\mathcal{E}_\emptyset$, so in particular, $\Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m\right) > 0.6$. In order to prove the lemma, we need to show that conditioning on $\boldsymbol{X}_i^{<\ell} = \overline{1}$ does not reduce the probability of $\mathcal{E}_{\emptyset,<i}$ by much; this is delicate, because $\boldsymbol{X}_i^{<\ell} = \overline{1}$ is a highly unlikely event.

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

$$p_t = \Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<t}\right).$$

We will show that the difference $|p_t - p_{t-1}|$ is small for each $t \in [\ell]$, and conclude that $|p_\ell - p_0|$ is small; that is, $\Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell}\right)$ is close to $\Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m\right)$, which we know is high.

To that end, fix $t \in [\ell]$, let us study the effect of adding conditioning on $\boldsymbol{X}_i^t = 1$. The dependence between the events $\mathcal{E}_{\emptyset,<i}$ and $\boldsymbol{X}_i^t = 1$ is small, because

$$\begin{aligned}
\mathsf{I}\left(\mathbf{1}_{\mathcal{E}_{\emptyset,<i}} \,;\, \boldsymbol{X}_i^t \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<t} = \overline{1}\right) &\le \mathsf{I}\left(\boldsymbol{X}_{J_{<i}}^t \,;\, \boldsymbol{X}_i^t \,\middle|\, \boldsymbol{M} = m\right) && \text{(by Lemma 23)} \\
&\le \frac{1}{\alpha_J \alpha_M C n^{1/k} k^2}. && (5.14)
\end{aligned}$$

In the last step, we used the fact that $i \in J$, so there is not much dependence between it and

the preceding bits in J given $\boldsymbol{M} = m$.

While $\boldsymbol{X}_i^t = 1$ is an unlikely event, it is not *too* unlikely: The players' inputs are independent given $\boldsymbol{M} = m$ (by Lemma 17), hence also are the i-th coordinates of the players' inputs (by Corollary 1). Together with Lemma 20 we have that:

$$\Pr\left(\boldsymbol{X}_i^t = 1 \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<t} = \overline{1}\right) = \Pr\left(\boldsymbol{X}_i^t = 1 \,\middle|\, \boldsymbol{M} = m\right) \geq \frac{1 - 1/(4k)}{n^{1/k}} \geq \frac{1}{2n^{1/k}}. \tag{5.15}$$

We therefore have, by definition of p_t, p_{t-1} and Property 1:

$$\begin{aligned}\mathsf{I}\left(\mathbf{1}_{\mathcal{E}_{\emptyset,<i}} \,;\, \boldsymbol{X}_i^t \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<t} = \overline{1}\right) &\geq \Pr\left(\boldsymbol{X}_i^t = 1 \,\middle|\, \boldsymbol{M} = 1, \boldsymbol{X}_i^{<t} = \overline{1}\right) \cdot \mathsf{D}\left(p_t \,||\, p_{t-1}\right) \\ &\geq \frac{1}{2n^{1/k}} \cdot \mathsf{D}\left(p_t \,||\, p_{t-1}\right). \qquad \text{(by (5.15))}\end{aligned}$$

Together with (5.14), we obtain

$$\mathsf{D}\left(p_t \,||\, p_{t-1}\right) \leq 2n^{1/k} \cdot \frac{1}{\alpha_J \alpha_M C n^{1/k} k^2} \leq \frac{1}{50 \ln 2 k^2}. \qquad \text{(by definition of } C \text{ (5.1))}$$

Finally, by Pinsker,

$$|p_t - p_{t-1}| \leq \sqrt{\frac{\ln 2}{2} \mathsf{D}\left(p_t \,||\, p_{t-1}\right)} \leq \sqrt{\frac{\ln 2}{2} \cdot \frac{1}{50 \ln 2 k^2}} = \frac{1}{10k}.$$

We have now shown that

$$\left|\Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell}\right) - \Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m\right)\right| = |p_\ell - p_0| \leq \sum_{t=1}^{\ell} |p_t - p_{t-1}| \leq k \cdot \frac{1}{10k} = \frac{1}{10}.$$

Thus,

$$\Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m, \boldsymbol{X}_i^{<\ell} = \overline{1}\right) \geq \Pr\left(\mathcal{E}_{\emptyset,<i} \,\middle|\, \boldsymbol{M} = m\right) - \frac{1}{10} > 0.6 - \frac{1}{10} = 0.5. \qquad \square$$

We now conclude the lower bound, by first proving Lemma 21, i.e. showing that each *good* transcript leads to error with high probability, and then that the protocol as a whole errs with probability at least ϵ_1.

Proof of Lemma 21. First, note that if $\Pr\left(\mathcal{E}_\emptyset \mid \boldsymbol{M} = m\right) \leq 0.6$, then the lemma holds trivially (as $0.6 < 1/e^{1/4}$). Therefore it only remains to consider the case where $\Pr\left(\mathcal{E}_\emptyset \mid \boldsymbol{M} = m\right) > 0.6$.

Recall that we defined $\mathcal{E}_{\emptyset,<i}$ to be the event that there is no intersection at any coordinate of J up to (but excluding) i. Therefore,

$$\begin{aligned}\Pr\left(\left(\bigcap_{\ell \in [k]} \boldsymbol{X}_J^\ell\right) = \emptyset \,\middle|\, \boldsymbol{M} = m\right) &= \Pr\left(\bigwedge_{i \in J} \left(\boldsymbol{X}_i \neq \overline{1}\right) \,\middle|\, \boldsymbol{M} = m\right) \\ &= \prod_{j \in J} \Pr\left(\boldsymbol{X}_i \neq \overline{1} \,\middle|\, \boldsymbol{M} = m, \mathcal{E}_{\emptyset,<i}\right) \\ &\leq \left(1 - \frac{1}{2n}\right)^{|J|}. \qquad \text{(Lemma 22)}\end{aligned}$$

Since $|J| \geq (1-\alpha_J)n$,

$$\left(1-\frac{1}{2n}\right)^{|J|} \leq \left(1-\frac{1}{2n}\right)^{(1-\alpha_J)n} \leq e^{-(1-\alpha_J)/2} = e^{-1/4},$$

where the second inequality holds due to Fact 1. □

Corollary 5. *The protocol errs with probability at least* ϵ_1.

Proof. Consider some specific $m \in G$. Since m is a good transcript, it outputs "non-intersecting". By Lemma 21, given $\boldsymbol{M} = m$, there is good probability that the inputs *do* intersect; whenever this occurs, the protocol errs.

More formally, we can write

$$\begin{aligned}
\Pr(M \text{ errs}) &\geq \sum_{m \in G} \Pr(M \text{ errs} \mid \boldsymbol{M} = m) \Pr(\boldsymbol{M} = m) \\
&\geq \sum_{m \in G} \Pr(\neg \mathcal{E}_\emptyset \mid \boldsymbol{M} = m) \Pr(\boldsymbol{M} = m) \\
&\geq \sum_{m \in G} \left(1 - e^{-1/4}\right) \cdot \Pr(\boldsymbol{M} = m) && \text{(Lemma 21)} \\
&= \left(1 - e^{-1/4}\right) \Pr(\boldsymbol{M} \in G) \geq \left(1 - e^{-1/4}\right) \cdot \frac{1}{8e}. && \text{(Lemma 18)}
\end{aligned}$$

We see that the protocol errs with probability at least $(1-e^{-1/4})/(8e) > 1/(800e) = \epsilon_1$. □

5.4 Proofs of the Technical Lemmas

Lemma 25 ("Technical Lemma 25", page 23, in [BO17])**.** *Let* $p \in (0, 1/3)$, $\alpha \in (-1, 1/2)$, *then:*

$$\mathsf{D}((1+\alpha)p \,||\, p) \geq \frac{1}{4 \ln 2} \cdot p\alpha^2.$$

Lemma 26 ("Technical Lemma 26", page 23, in [BO17])**.** *Let* $p \in (0, 1/2)$, $\alpha \geq 1/2$ *such that* $(1+\alpha)p \leq 1$, *then:*

$$\mathsf{D}((1+\alpha)p \,||\, p) \geq \frac{1}{10} \cdot p(1+\alpha).$$

Lemma (Restating Lemma 14)**.** *Let* $\mu \sim Ber(p)$, $p \in (0, 1/3)$, $\alpha \in (0, 1/2)$ *and Let* $\mu' \sim Ber(p')$ *such that* $\mathsf{D}(\mu' \,||\, \mu) \leq \frac{1}{40 \ln 2} \cdot p\alpha^2$, *then:*

$$\frac{p'}{p} \in (1-\alpha, 1+\alpha).$$

Proof. Assume towards showing a contradiction that:

$$\frac{p'}{p} \notin (1-\alpha, 1+\alpha),$$

and divide to cases:

Case 1.

$$\frac{p'}{p} = 0.$$

Since $p > 0$, this implies that $p' = 0$. Then:

$$\mathsf{D}\left(\mu' \,\middle|\middle|\, \mu\right) = \mathsf{D}\left(p' \,\middle|\middle|\, p\right) = \mathsf{D}\left(0 \,||\, p\right) = 1 \cdot \log\left(\frac{1}{1-p}\right) = -\log(1-p) = -\frac{\ln(1-p)}{\ln 2} \geq \frac{p}{\ln 2}$$
$$> \frac{p\alpha^2}{40 \ln 2},$$

contradiction.

Case 2.

$$\frac{p'}{p} \in (0, 1-\alpha).$$

Denote:

$$\alpha' := 1 - \frac{p'}{p} \in (\alpha, 1),$$

I.e.:

$$p' = (1-\alpha')p.$$

Since we have that $-\alpha' \in (-1, -\alpha) \subseteq (-1, 1/2)$:

$$\begin{aligned}
\mathsf{D}\left(\mu' \,\middle|\middle|\, \mu\right) = \mathsf{D}\left(p' \,\middle|\middle|\, p\right) &= \mathsf{D}\left((1-\alpha')p \,\middle|\middle|\, p\right) \\
&\geq \frac{1}{4 \ln 2} \cdot p\alpha'^2 && \text{(by Lemma 25)} \\
&> \frac{1}{4 \ln 2} \cdot p\alpha^2 && (\alpha' > \alpha) \\
&> \frac{1}{40 \ln 2} \cdot p\alpha^2,
\end{aligned}$$

contradiction.

Case 3.

$$\frac{p'}{p} > 1 + \alpha.$$

Denote:

$$\beta := \frac{p'}{p} - 1 > \alpha,$$

which implies that:

$$p' = (1+\beta)p.$$

If $\beta < 1/2$, then:

$$\begin{aligned}
\mathsf{D}\left(\mu' \,||\, \mu\right) &= \mathsf{D}\left(p' \,||\, p\right) = \mathsf{D}\left((1+\beta)p \,||\, p\right) \\
&\geq \frac{1}{4\ln 2} \cdot p\beta^2 && \text{(by Lemma 25)} \\
&> \frac{1}{4\ln 2} \cdot p\alpha^2 && (\beta > \alpha) \\
&> \frac{1}{40\ln 2} \cdot p\alpha^2,
\end{aligned}$$

contradiction. Otherwise $\beta \geq 1/2$, hence:

$$\begin{aligned}
\mathsf{D}\left(\mu' \,||\, \mu\right) &= \mathsf{D}\left(p' \,||\, p\right) = \mathsf{D}\left((1+\beta)p \,||\, p\right) \\
&\geq \frac{p(1+\beta)}{10} && \text{(by Lemma 26)} \\
&\geq \frac{1.5p}{10} > \frac{0.5^2 p}{40\ln 2} \\
&> \frac{p\alpha^2}{40\ln 2}, && (\alpha < 1/2)
\end{aligned}$$

contradiction.

□

Lemma (Restating Lemma 15)**.** *Let $\boldsymbol{A} \sim Ber(p)$, $p \in (0, 1/3)$, $\alpha \in (0, 1/2)$, $\boldsymbol{B}$ an RV with $b \in support(\boldsymbol{B})$, such that:*

$$\mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B}\right) \leq \Pr\left(\boldsymbol{B}=b\right) \cdot \frac{p\alpha^2}{40\ln 2},$$

and denote: $\boldsymbol{A}|_{\boldsymbol{B}=b} \sim Ber(p')$. Then:

$$\frac{p''}{p} \in (1-\alpha, 1+\alpha)\,.$$

Proof. By Lemma 14, it is enough to show that:

$$\mathsf{D}\left(p \,||\, p'\right) \leq p\alpha^2/40\ln 2.$$

Observe that by Property 1 and by the non-negativity of KL-divergence, it follows that:

$$\mathsf{D}\left(\boldsymbol{A} \,||\, \boldsymbol{A}|_{\boldsymbol{B}=b}\right) \leq \frac{1}{\Pr(\boldsymbol{B}=b)}\mathsf{I}\left(\boldsymbol{A}\,;\,\boldsymbol{B}\right),$$

which completes the proof. □

Lemma (Restating Lemma 16)**.** *Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ be RVs, such that $\boldsymbol{A}$ is independent of $\boldsymbol{C}$, then:*

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B}) + \mathsf{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{B}) = \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}),$$

and in particular:

1. *(conditioning does not decrease information):*

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B}) \leq \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}).$$

2. (reversal Lemma)

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{B}) \leq \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}).$$

Proof. Using the chain rule:

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B},\boldsymbol{C}) = \mathsf{I}(\boldsymbol{A};\boldsymbol{B}) + \mathsf{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{B}), \tag{5.16}$$

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B},\boldsymbol{C}) = \mathsf{I}(\boldsymbol{A};\boldsymbol{C}) + \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}) = \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}). \tag{5.17}$$

Hence:

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B}) + \mathsf{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{B}) = \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}).$$

From the non-negativity of information, it follows that:

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{B}) \leq \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}),$$

and

$$\mathsf{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{B}) \leq \mathsf{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}),$$

which proves both desired properties. □

Claim (Restating Claim 1)**.** *Let $\boldsymbol{X},\boldsymbol{X'},\boldsymbol{Y},\boldsymbol{Y'}$ be RVs, such that the pair $(\boldsymbol{X},\boldsymbol{X'})$ is independent of the pair $(\boldsymbol{Y},\boldsymbol{Y'})$, and let $y \in support(\boldsymbol{Y})$. Then:*

$$\mathsf{I}\left(\boldsymbol{X}\,;\,\boldsymbol{Y'X'} \,\middle|\, \boldsymbol{Y}=y\right) = \mathsf{I}\left(\boldsymbol{X}\,;\,\boldsymbol{X'} \,\middle|\, \boldsymbol{Y}=y\right) = \mathsf{I}\left(\boldsymbol{X}\,;\,\boldsymbol{X'}\right).$$

Proof. Let's prove each equality separately:

Lemma 27.

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y'X'} \mid \boldsymbol{Y}=y) = \mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{X'} \mid \boldsymbol{Y}=y).$$

Proof. By the chain rule:

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y'X'} \mid \boldsymbol{Y}=y) = \mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{X'} \mid \boldsymbol{Y}=y) + \mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y'} \mid \boldsymbol{Y}=y,\boldsymbol{X'}),$$

so it suffices to show that:

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y'} \mid \boldsymbol{Y}=y,\boldsymbol{X'}) = 0.$$

Since information is non-negative, it suffices to show that:

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y'} \mid \boldsymbol{Y},\boldsymbol{X'}) = 0.$$

But this is true, as by property 4 of mutual information:

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y'}\mid\boldsymbol{Y},\boldsymbol{X'})\leq\mathsf{I}(\boldsymbol{X}\boldsymbol{X'}\,;\,\boldsymbol{Y}\boldsymbol{Y'})=0.\qquad\square$$

Lemma 28.

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{X'}\mid\boldsymbol{Y}=y)=\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{X'}).$$

Proof. Note that by the definition of mutual information:

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{X'}):=H(\boldsymbol{X})-H(\boldsymbol{X}\mid\boldsymbol{X'}),$$

and:

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{X'}\mid\boldsymbol{Y}=y)=H(\boldsymbol{X}\mid\boldsymbol{Y}=y)-H(\boldsymbol{X}\mid\boldsymbol{Y}=y,\boldsymbol{X'}),$$

So it suffices to show that:

$$H(\boldsymbol{X})=H(\boldsymbol{X}\mid\boldsymbol{Y}=y),\tag{5.18}$$

and

$$H(\boldsymbol{X}\mid\boldsymbol{X'})=H(\boldsymbol{X}\mid\boldsymbol{Y}=y,\boldsymbol{X'}).\tag{5.19}$$

Now note that by monotonicity of mutual information (property 3):

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y})\leq\mathsf{I}(\boldsymbol{X}\boldsymbol{X'}\,;\,\boldsymbol{Y}\boldsymbol{Y'})=0,$$

hence $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, and particularly for all $x\in\mathit{support}(\boldsymbol{X})$ we have that:

$$\Pr\left(\boldsymbol{X}=x\right)=\Pr\left(\boldsymbol{X}=x\mid\boldsymbol{Y}=y\right),$$

and hence $\boldsymbol{X}|_{\boldsymbol{Y}=y}$ and $\boldsymbol{X}$ are identically distributed, and in particular have the same Shannon entropy, i.e.:

$$H(\boldsymbol{X})=H(\boldsymbol{X}\mid\boldsymbol{Y}=y),$$

proving eq. (5.18). Now, note that by the monotonicity of mutual information (property 3):

$$\mathsf{I}(\boldsymbol{X'}\,;\,\boldsymbol{Y})\leq\mathsf{I}(\boldsymbol{X}\boldsymbol{X'}\,;\,\boldsymbol{Y}\boldsymbol{Y'})=0,$$

hence for all $\tilde{x}\in\mathit{support}\left(\boldsymbol{X'}\right)$ we have that:

$$\Pr\left(\boldsymbol{X'}=\tilde{x}\,\middle|\,\boldsymbol{Y}=y\right)=\Pr\left(\boldsymbol{X'}=\tilde{x}\right).\tag{5.20}$$

Now note that:

$$\begin{aligned}\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y}\mid\boldsymbol{X'})&\leq\mathsf{I}(\boldsymbol{X}\boldsymbol{X'}\,;\,\boldsymbol{Y})&&\text{(property 4 of mutual information)}\\&\leq\mathsf{I}(\boldsymbol{X}\boldsymbol{X'}\,;\,\boldsymbol{Y}\boldsymbol{Y'})&&\text{(property 3 of mutual information)}\\&=0.&&\text{(by assumption)}\end{aligned}$$

Since mutual information is non-negative, this implies that for all $\tilde{x}\in\mathit{support}\left(\boldsymbol{X'}\right)$, we have

that:

$$\mathsf{I}(\boldsymbol{X}\,;\,\boldsymbol{Y}\mid\boldsymbol{X'}=\tilde{x})=0.$$

and hence $\boldsymbol{X}|_{\boldsymbol{X'}=\tilde{x},\boldsymbol{Y}=y}$ and $\boldsymbol{X}|_{\boldsymbol{X'}=\tilde{x}}$ are identically distributed, and in particular for all $\tilde{x}\in$ *support* $(\boldsymbol{X'})$, we have that:

$$H(\boldsymbol{X}|\boldsymbol{X'}=\tilde{x},\boldsymbol{Y}=y)=H(\boldsymbol{X}|\boldsymbol{X'}=\tilde{x}). \tag{5.21}$$

Now this implies that:

$$\begin{aligned}
H(\boldsymbol{X}|\boldsymbol{X'},\boldsymbol{Y}=y) &:= \mathop{\mathbb{E}}_{\boldsymbol{X'}}[H(\boldsymbol{X}|\boldsymbol{X'},\boldsymbol{Y}=y)]\\
&= \sum_{\tilde{x}\in support(\boldsymbol{X'})} \Pr\left(\boldsymbol{X'}=\tilde{x}\;\middle|\;\boldsymbol{Y}=y\right) H(\boldsymbol{X}|\boldsymbol{X'}=\tilde{x},\boldsymbol{Y}=y)\\
&= \sum_{\tilde{x}\in support(\boldsymbol{X'})} \Pr\left(\boldsymbol{X'}=\tilde{x}\right) H(\boldsymbol{X}|\boldsymbol{X'}=\tilde{x},\boldsymbol{Y}=y) && \text{(eq. (5.20))}\\
&= \sum_{\tilde{x}\in support(\boldsymbol{X'})} \Pr\left(\boldsymbol{X'}=\tilde{x}\right) H(\boldsymbol{X}|\boldsymbol{X'}=\tilde{x}) && \text{(eq. (5.21))}\\
&= \mathop{\mathbb{E}}_{\boldsymbol{X'}}[H(\boldsymbol{X}|\boldsymbol{X'})]\\
&=: H(\boldsymbol{X}|\boldsymbol{X'}),
\end{aligned}$$

proving (5.19). □

□

Lemma (Restating Lemma 17)**.** *Let* $\boldsymbol{M}$ *denote the transcript of a deterministic protocol* Π *over the inputs* $\boldsymbol{X}^1,\ldots,\boldsymbol{X}^k$*, and let* $\ell\in[k]$*. Then:*

$$\mathsf{I}(\boldsymbol{X}^\ell;\boldsymbol{X}^{-\ell}\mid\boldsymbol{M})\le\mathsf{I}(\boldsymbol{X}^\ell;\boldsymbol{X}^{-\ell}).$$

Proof. Instead of re-proving the claim for 2-players, consider the following: Look at a deterministic **2**-player protocol Π', obtained from Π by re-labeling the vertices owned by player ℓ as owned by Alice, and the other vertices as owned by Bob. Now denote: $\boldsymbol{X'}:=\boldsymbol{X}^\ell$, $\boldsymbol{Y'}:=\boldsymbol{X}^{-\ell}$ and observe that Π' is a deterministic 2-player protocol over $\boldsymbol{X'},\boldsymbol{Y'}$. Denote this protocol's transcript by $\boldsymbol{M'}$, and observe that: $\boldsymbol{M'}=\boldsymbol{M}$. Then, Since Π' is a deterministic 2-player protocol over $\boldsymbol{X'},\boldsymbol{Y'}$:

$$\mathsf{I}(\boldsymbol{X'};\boldsymbol{Y'}\mid\boldsymbol{M'})\le\mathsf{I}(\boldsymbol{X'};\boldsymbol{Y'}).$$

Substituting the assigned values, we get the desired result. □

Chapter 6

Limitations of Prior Work

In this chapter, we will discuss certain limitations of the related prior work. These limitations have motivated our development of new techniques, as described in the previous chapters.

6.1 Limitations of the Lower Bound of Babai, Frankl and Simon

Babai, Frankl and Simon showed in [BFS86] a lower bound of $\Omega(\sqrt{n})$ bits on the communication complexity of the 2-players Disjointness problem for some specific product distribution. We will give a high level overview of their proof, and then explain why certain simple attempts to extend their proof technique to 3 players fail. Note that it is possible that some more elborate modifications to their proof will yield the desired lower bound for 3 players (or more), but we were not able to find such a proof, hence our information-theoretic lower bound, which uses an entirely different technique. Note that it is reasonable to start with 3 players, which is the simplest scenario to analyze (besides the 2 players scenario).

6.1.1 Overview of the Lower Bound of BFS

Babai, Frankl and Simon ([BFS86]) use the following hard distribution for their lower bound: the inputs of Alice and Bob are sets of size $\sqrt{n}$, chosen uniformly and independently from a universe of n elements. Formally, we denote:

$$U := \binom{[n]}{\sqrt{n}},$$

and let μ_X and μ_Y be the marginal distributions of Alice's and Bob's inputs (respectively). We then define

$$\mu_X = \mu_Y := Unif(U).$$

We will denote by $\boldsymbol{X} \sim \mu_X, \boldsymbol{Y} \sim \mu_Y$ the random variables corresponding to Alice's and Bob's inputs (respectively). The input distribution is given by $\mu = \mu_X \times \mu_Y$. For convenience, we sometimes abuse notation by letting $\mu(X)$ or $\mu(Y)$ denote the marginal probability of a set $X \subseteq U$ or $Y \subseteq U$ (respectively).

The lower bound of [BFS86] uses the *corruption technique*: to derive the $\sqrt{n}$ lower bound, [BFS86] shows that there exists some positive constant $c > 0$, such that for every *combinatorial*

rectangle $X \times Y \subseteq U^2$ with

$$\Pr_{(\boldsymbol{X},\boldsymbol{Y})\sim\mu}(\boldsymbol{X} \cap \boldsymbol{Y} \neq \emptyset \mid \boldsymbol{X} \in X, \boldsymbol{Y} \in Y) \leq \epsilon, \tag{6.1}$$

either X or Y must be "small": formally, either $\mu(X)$ or $\mu(Y)$ must be upper-bounded by $2^{-c\sqrt{n}}$. A rectangle satisfying (6.1) is often called an "almost monochromatic 1-rectangle" (ϵ-AM1R).

The proof continues by fixing a combinatorial rectangle $X \times Y$ such that (6.1) holds, and such that either $\mu(X)$ or $\mu(Y)$ is at least $2^{-c\sqrt{n}}$ (where c is determined later). Without loss of generality, we assume that $\mu(X) \geq 2^{-c\sqrt{n}}$.

Next, the proof shows that since X is "large", it can be "represented" by a small collection $X' \subseteq U$, which covers a constant fraction of the n elements in the universe, and such that the typical $y \in Y$ only intersects few of the sets in X'. This essentially completes the proof, since it implies that the average $y \in Y$ must be disjoint from a large portion of the elements in the universe, and therefore there cannot be too many such $y \in Y$. More formally, [BFS86] shows:

Lemma. *There exists a collection $X' \subseteq X$ such that:*

1. $|X'| = \Theta(\sqrt{n})$.
2. $\left|\bigcup_{x\in X'} x\right| \geq \frac{n}{6}$.
3. *It holds that:*

$$\Pr_{(\boldsymbol{X},\boldsymbol{Y})\sim\mu}\left(\boldsymbol{X} \cap \boldsymbol{Y} \neq \emptyset \;\middle|\; \boldsymbol{X} \in X', \boldsymbol{Y} \in Y\right) \leq 2\epsilon.$$

This, together with an averaging argument, implies that most $y \in Y$ also intersect only a small fraction of $x \in X'$; that is, there exists a collection $Y' \subseteq Y$ of cardinality at least $|Y|/2$, such that for every $y \in Y'$ we have that:

$$\Pr_{\boldsymbol{X}\sim\mu_X}\left(\boldsymbol{X} \cap y \neq \emptyset \;\middle|\; \boldsymbol{X} \in X'\right) \leq 4\epsilon.$$

Observe that lower-bounding the size of the collection Y' also lower bounds the size of Y, and indeed we can see that each $y \in Y'$ is determined by:

1. The sets $x \in X'$ which y intersects (at most $4\epsilon \cdot |X'|$ of those).
2. The choice of y's $\sqrt{n}$ elements from those elements that are not "forbidden", that is, those elements that are not in any set $x \in X'$ which is disjoint from y.

We see that for every $y \in Y'$, there are at least $n/6 - 4\epsilon \cdot |X'| \cdot \sqrt{n}$ "forbidden" elements for y, hence there are at most $5n/6 + 4\epsilon \cdot |X'| \cdot \sqrt{n}$ possible elements, which, for small enough ϵ, is at most $8n/9$. It follows that $|Y'|$ is upper-bounded by $\binom{|X'|}{4\epsilon|X'|}\binom{8n/9}{\sqrt{n}} < |U| \cdot 2^{-c\sqrt{n}}$ (for suitably chosen constants ϵ and c).

6.1.2 Limitations on Generalizing the BFS Lower Bound

Let us now try to generalize the BFS lower bound proof to 3 players. We will present three attempts and show where they fail, indicating – to some extent – the limitations of the combinatorial proof used by BFS.

For 3 players, our goal is to show a lower bound of $\Omega\left(n^{2/3}\right)$, matching the lower bound we proved using information-theoretic techniques in Theorem 3.

Let us start by re-defining the hard product distribution to work for 3 players. In the 2-player case, Alice and Bob each got a set of size $\sqrt{n}$. It is easy to see that in the 3-player case, the size of each player's inputs must be $\tilde{\Omega}(n^{2/3})$: if one player, say Alice, has a set of size $o\left(\frac{n^{2/3}}{\log n}\right)$, then she can transmit her set to Bob. Bob can then intersect his set with Alice's set, and transmit the resulting set to Charlie, who can now determine if the sets intersect or not. The overall communication complexity will be $o(n^{2/3})$. Let us therefore re-define:

$$U := \binom{[n]}{n^{2/3}},$$

and let $\mu_X = \mu_Y = \mu_Z = Unif(U)$ be the marginal distributions of the players' inputs. We denote by $\boldsymbol{X} \sim \mu_X, \boldsymbol{Y} \sim \mu_Y, \boldsymbol{Z} \sim \mu_Z$ the random variables indicating the inputs of Alice, Bob and Charlie (respectively).

Let us now try to follow the footsteps of the BFS proof: we start by fixing a combinatorial rectangle $X \times Y \times Z \subseteq U^3$, such that

$$\Pr_{(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})\sim\mu}(\boldsymbol{X} \cap \boldsymbol{Y} \cap \boldsymbol{Z} \neq \emptyset \mid \boldsymbol{X} \in X, \boldsymbol{Y} \in Y, \boldsymbol{Z} \in Z) \leq \epsilon. \tag{6.2}$$

Recall that the BFS lower bound proof has two steps:

(1) Assuming that X is large, we "represent" X by a small collection $X' \subseteq X$ of sets that together cover $\Theta(n)$ elements and roughly preserve the intersection probability with Y;

(2) "Represent" Y by a collection $Y' \subseteq Y$ that includes a large fraction of the elements in Y, and contains only sets that *each* intersect only a small fraction of the sets in X'.

The BFS proof then shows that that $|Y'|$ cannot be too large, which also implies that $|Y|$ is not too large.

We consider three attempts to generalize the BFS proof.

Attempt I: Applying step (1) to X and step (2) to $Y \times Z$. Let us assume that we have a collection $X' \subseteq X$ which is "small", covers $\Theta(n)$ elements, and satisfies for every $x \in X'$:

$$\Pr_{(\boldsymbol{Y},\boldsymbol{Z})\sim\mu_Y\times\mu_Z}(x \cap \boldsymbol{Y} \cap \boldsymbol{Z} \neq \emptyset \mid \boldsymbol{Y} \in Y, \boldsymbol{Z} \in Z) \leq 2\epsilon.$$

The BFS proof shows that there exists such a set whenever $\mu(X)$ is sufficiently large; in our case, let us simply assume that it exists (we are interested in showing that the proof cannot go through even in this case).

In Step (2), the BFS proof uses an averaging argument to deduce the existence of a collection $Y' \subseteq Y$ which includes most of the elements in Y, while having roughly the same probability of intersection with an element of X' as the overall set Y. The proof is then completed by showing that this upper-bounds the cardinality of Y' (and hence of Y).

In the 3-player case, following the same reasoning, we can deduce the existence of a collection $S' \subseteq Y \times Z$ with cardinality at least half of that of $Y \times Z$, such that for every $(y, z) \in S$,

$$\Pr_{\boldsymbol{X} \sim \mu_X} \left(\boldsymbol{X} \cap y \cap z \neq \emptyset \;\middle|\; \boldsymbol{X} \in X' \right) \leq 4\epsilon. \tag{6.3}$$

note that S' is *not necessarily a combinatorial rectangle.*

To complete the proof, we would need to show that the cardinality of S is upper-bounded by $\left|U^2\right| \cdot 2^{-\Omega(n^{2/3})}$ (for some constant c'). This would yield an upper bound of $2^{-\Omega(n^{2/3})}$ on $\mu(X \times Y \times Z)$, as desired. Unfortunately, even under the assumptions above, the size of S is not bounded by $2^{-\Omega(n^{2/3})}$: there exists a large collection $S \subseteq Y \times Z$, of cardinality at least $|U|^{\,2} \cdot 2^{-\Theta(n^{1/3})}$, such that for any non-empty $X' \subseteq X$, and for every $(y, z) \in S$, we have

$$\Pr_{\boldsymbol{X} \sim \mu_X} \left(\boldsymbol{X} \cap y \cap z \neq \emptyset \;\middle|\; \boldsymbol{X} \in X' \right) = 0.$$

In other words, even though (6.3) is satisfied, the size of S' exceeds our desired bound.

The collection S is defined by taking all disjoint pairs $y, z \in U$, i.e.:

$$S := \left\{ (y, z) \in U^2 \;\middle|\; y \cap z = \emptyset \right\}.$$

A simple combinatorial calculation shows that

$$\Pr_{(\boldsymbol{Y}, \boldsymbol{Z}) \sim \mu_Y \times \mu_Z} (\boldsymbol{Y} \cap \boldsymbol{Z} = \emptyset) \geq 2^{-\Theta(n^{1/3})}.$$

Attempt II: applying step (1) to $X \times Y$, then step (2) to Z. We now consider a different strategy: "representing" $X \times Y$ by a small collection $S \subseteq X \times Y$ of sets that "cover" most of the universe and roughly preserve the intersection probability, and then arguing that the remaining dimension of the rectangle, Z, must be small.

Fix a combinatorial rectangle $X \times Y \times Z \subseteq U^3$, with intersection probability at most ϵ inside it. Assume that $\mu(X \times Y) \geq 2^{-cn^{2/3}}$, otherwise the rectangle is small and we need not consider it. We would like to find a "representation of $X \times Y$", a collection of pairs $S \subseteq X \times Y$ with the following properties:

- $|S| \leq n^{1/3}$.
- The intersection probability in $S \times Z$ is at most 2ϵ (the exact constant 2 is of course not important here, but for simplicity we use the same constant as BFS).
- S "covers" $\Theta(n)$ elements of the universe. Here, the correct interpretation of "covers" that would allow the proof to go through is the following: define

$$C(S) = \bigcup_{(x,y) \in S} (x \cap y).$$

Then we would like to have $|C(S)| \geq \alpha n$, where $\alpha \in (0, 1]$ is a sufficiently large constant.

If we can find such a set, we could apply step (2) to find a collection $Z' \subseteq Z$ of size $|Z'| \geq |Z|/2$, such that for each $z \in Z'$, the intersection probability inside $S \times \{z\}$ is at most 4ϵ. A

simple counting argument, similar to the one used in the 2-player proof, would then show that $|Z| \leq 2|Z'| \leq 2^{-\Omega(n^{2/3})} \cdot |U|$ and complete the proof. Unfortunately, we will see that attempting to prove that such a collection S exists is likely to fail. First, observe that the properties of S imply that most pairs (x, y) in S should have a large intersection, i.e. $|x \cap y| = \Theta(n^{2/3})$ for most (x, y) in S. This would be critical later on in understanding why this proof attempt fails, as we will show that it is unreasonable to expect that the intersection size of the typical (x, y) would be much larger than $n^{1/3}$.

Secondly, recall that the BFS argument for showing a small collection X' that "represents" X starts by applying Markov's inequality to deduce that some large collection $X'' \subseteq X$ exists of cardinality at least $|X|/2$ and such that for any $x \in X''$, the combinatorial rectangle $\{x\} \times Y$ has intersection probability at most 2ϵ. This step is necessary, as we would like to claim that since X'' is large, there must exist a small sub–collection $X' \subseteq X''$ that covers a constant fraction of the elements in the world, and such that $X' \times Y$ has intersection probability at most 2ϵ. Hence for the BFS proof it was necessary to apply the Markov argument to obtain the intermediate collection X'', as otherwise there would be no guarantee that the intersection probability of $X' \times Y$ is small.

Attempting to apply the same Markov argument in the 3-player scenario, we obtain a collection of pairs $S' \subseteq X \times Y$ containing most of the pairs in $X \times Y$, and such that for any pair $(x, y) \in S'$, the combinatorial rectangle $\{x\} \times \{y\} \times Z$ has intersection probability at most 2ϵ. Unfortunately, S' itself does not have to be a combinatorial rectangle, and in fact, all we can assume about S' is the assumptions about its cardinality and intersection probability mentioned above.

At this point, we can see why this proof attempt fails; $S \subseteq S'$ is required to satisfy that for most elements $(x, y) \in S$ the intersection of x and y has cardinality at least $\Omega(n^{2/3})$, but there exists a large collections S' such that *every* pair $(x, y) \in S'$ has intersection size roughly $n^{1/3}$. To see this, recall that the intersection *size* of a random set of fixed size, with another fixed set follows the hypergeometric distribution. The Chernoff-like tail bounds for this distribution imply that with high probability, two random sets of size $n^{2/3}$ will have intersection size at most $2n^{1/3}$. Consequently, if we define S' to be the collection of pairs of intersection size at most $n^{2/3}$, then S' has large cardinality (in fact, $\mu(S') = \Omega(1)$), but one cannot extract the desirable sub-collection S from S', as S is required to have mostly elements of intersection size $n^{2/3} \gg 2n^{1/3}$.

Attempt III: applying step (1) to X and to Y, then step (2) to Z. We now consider a different strategy: instead of using the collection X' to "represent" the collection X, and then arguing about $Y \times Z$, let us consider what happens if we use a small collection X' to "represent" X and a small collection Y' to "represent" Y, and then argue about Z. One may hope that if there is a large set of elements in the universe covered by both X' and Y', then we can employ an argument similar to the one that BFS used for 2 players. Unfortunately, we will see that this type of argument fails for some combinatorial rectangles.

For 3 players, we start with a combinatorial rectangle $X \times Y \times Z \subseteq U^3$ with intersection probability at most ϵ. If $\mu(X) \geq 2^{-cn^{2/3}}$ for some suitable constant $c > 0$, then, following the

proof of BFS, we assume again that there exists a collection $X' \subseteq X$, of cardinality $\Theta(n^{1/3})$, which covers $\Theta(n)$ elements in the universe, and moreover, the combinatorial rectangle $X' \times Y \times Z$ has intersection probability at most 2ϵ inside it. If $\mu(Y) \geq 2^{-cn^{2/3}}$ as well, then a similar argument would show that there exists a similar collection $Y' \subseteq Y$ that also covers $\Theta(n)$ elements in the universe, and such that the intersection probability in the combinatorial rectangle $X' \times Y' \times Z$ is at most 4ϵ. Markov's inequality then implies that there exists a collection $Z' \subseteq Z$ of cardinality at least $|Z|/2$, such that for every $z \in Z'$, the combinatorial rectangle $X' \times Y' \times \{z\}$ has intersection probability at most 8ϵ. Observe that this implies that for every set $z \in Z'$, there are at most $8\epsilon|X'| \cdot |Y'|$ pairs $(x, y) \in X' \times Y'$, such that z intersects $x \cap y$. We would like to use this property to bound the size of $|Z'|$, and hence also of $|Z|$.

As in the 2-player argument, we can describe each $z \in Z'$ by

- The collection of pairs $S \subseteq X' \times Y'$ that z is "allowed" to intersect – there are at most $8\epsilon|X'| \cdot |Y'|$ of these;
- The choice of z's $n^{2/3}$ elements from those elements that are not "forbidden", that is, those elements that are not in any $x \cap y$ for $(x, y) \notin S$.

Note that, once again, S is not necessarily a combinatorial rectangle.

As in the 2 players argument, for a collection of pairs $S \subset X' \times Y'$ of cardinality $8\epsilon|X'| \cdot |Y'|$, we denote by $U(S)$ the set of allowed elements for a set guarunteed not to intersect any set $s = x \cap y$ such that $(x, y) \in (X' \times Y') \setminus S$. Observe that any set $z \in Z'$ may be described by the collection of pairs $S \subseteq X' \times Y'$ it is allowed to intersect, and the choice of $|z| = n^{2/3}$ elements of the universe $U(S)$. It follows that:

$$\mu(Z') \leq \sum_{\substack{S \subseteq X' \times Y' \\ s.t.\ |S| = 8\epsilon|X'| \cdot |Y'|}} \frac{\binom{|U(S)|}{n^{2/3}}}{|U|}. \tag{6.4}$$

Naïvely, one could hope to show that, as in the 2 players proof, there exists some constant $0 < \alpha < 1$ such that for every such S, $|U(S)| \leq \alpha n$. Then, a simple counting argument (similar to the one used for 2 players) would imply that the right-hand-side expression in (6.4) is at most $2^{-\Theta(n^{2/3})}$, hence completing the proof for 3 players. Unfortunately, there exists collections X' and Y' that cover all n elements in the universe, and such that a fraction of at least $2^{-\Theta(n^{1/3})}$ of the S's satisfy that $\binom{|U(S)|}{n^{2/3}} = |U|$. Hence for these collections X', Y', the right-hand-side expression in (6.4) is at least $2^{-\Theta(n^{1/3})} \gg 2^{-\Theta(n^{2/3})}$, and hence this argument cannot imply the desired lower bound.

Let us now describe the collections X' and Y': let X' be any collection of $n^{1/3}$ pairwise-disjoint sets in U, and set $Y' = X'$. Observe that both X' and Y' cover all the n elements in the universe. Now consider the collection of pairs $G = \{(x, x) \mid x \in X'\}$. Observe that since the sets in X' are pairwise-disjoint, for every pair $(x, y) \in (X' \times Y') \setminus G$ it holds that $x \cap y = \emptyset$. It follows that for every collection $S \subseteq X' \times Y'$ that contains G, the set of possible elements $U(S)$ contains *all* the n elements of the universe, and hence $\binom{|U(S)|}{n^{2/3}} = |U|$.

It remains to show that the fraction of collections $S \subseteq X' \times Y'$ of cardinality $8\epsilon|X'| \cdot |Y'|$ that contain G is large. Observe that the cardinality of X' and Y' is $n^{1/3}$. The total number

of sub-collections of cardinality $8\epsilon|X'| \cdot |Y'|$ is $\binom{n^{2/3}}{8\epsilon n^{2/3}}$. Since $|G| = |X'|$, the number of such collections S that also contain G is $\binom{n^{2/3}-n^{1/3}}{8\epsilon n^{2/3}-n^{1/3}}$. Hence the fraction of collections S that also contain G is:

$$\frac{\binom{n^{2/3}-n^{1/3}}{8\epsilon n^{2/3}-n^{1/3}}}{\binom{n^{2/3}}{8\epsilon n^{2/3}}} = \frac{\binom{8\epsilon n^{2/3}}{n^{1/3}}}{\binom{n^{2/3}}{n^{1/3}}} \geq (4\epsilon)^{n^{1/3}} = 2^{-\Theta(n^{1/3})}.$$

6.2 Limitations of the Upper Bound of Babai, Frankl and Simon

In addition to their celebrated lower bound, Babai, Frankl and Simon showed in [BFS86] an *upper bound* of $O(\sqrt{n}\log n)$ bits on the communication complexity of the 2-players Disjointness problem for any product distribution.

Unfortunately, it seems like the protocol used in [BFS86] cannot be easily adapted to more than 2 players, which motivated our upper bound algorithms described in chapters 3 and 4. We will give a high level overview of the protocol used in [BFS86], and then explain its limitations when trying to adapt it to 3 or more players.

6.2.1 Overview of the Upper Bound of BFS

The protocol described in [BFS86] works in iterations, where Alice and Bob both maintain a current "universe" $U \subseteq [n]$, where initially $U = [n]$, and each iteration decreases the size of the universe U by (at least) $\sqrt{n}$ elements, until Alice's input restricted to U (i.e. $X \cap U$) has cardinality at most $\sqrt{n}$, where Alice can simply send it to Bob (using at most $n\log n$ bits of communication).

At each iteration, Alice first tells Bob whether her restricted input $X \cap U$ is at most $\sqrt{n}$ (in which case she also sends her restricted input to Bob who can determine if an intersection occured), or whether it is at least $\sqrt{n}$. In the latter case, Bob first checks whether the probability that a random set intersects his input $Y \cap U$ is at least ϵ, where the set is sampled randomly from Alice's distribution, conditioned on the current universe U and the fact that the set has cardinality at least $\sqrt{n}$. If this intersection probability is less than ϵ, then Bob delares "X and Y are intersecting". Otherwise, Bob publicly samples (an infinite number) of such random sets independently, and sends Alice the index of the first set that is disjoint from $Y \cap U$. Alice and Bob then remove all the elements of this random set from U, and continue to the next iteration.

Now observe that after at most $\sqrt{n}$ iterations the protocol must terminate, and that at each iteration where Alice's input is still large, Alice uses 1 bit of communication, and Bob uses on expectation at most $O(\log(1/\epsilon))$ bits to communicate the index of the disjoint random set to Alice. Assuming that Bob never declared "intersecting", then at some point Alice's input will be at most $\sqrt{n}$ bits, and then Alice uses another $O(\sqrt{n}\log n)$ bits to transmit her set to Bob.

6.2.2 Limitations on Generalizing the BFS upper bound

When considering an adaptation of the [BFS86] protocol to 3 players: Alice, Bob and Charlie, we can naturally consider two variants: one where Bob and Charlie try to jointly select a random set from Alice's distribution, and one where Charlie selects a pair of random sets: one from Alice's distribution, and one from Bob's distribution. In both cases, all 3 players maintain a

universe U and try to decrease the cardinality of one of the input sets to be at most $n^{2/3}$ when restricted to U.

Attempt I: Bob and Charlie jointly select a random set from Alice's distribution. In this variant, in the beginning of each iteration, Alice tells Bob and Charlie whether her restricted input $X \cap U$ has cardinality at most $n^{2/3}$ (in which case she sends it to Bob, who itersects it with his set and sends the result to Charlie). If Alice's restricted set has cardinality at least $n^{2/3}$, then Bob and Charlie would like to jointly sample a random set drawn from Alice's input conditioned on U and set cardinality, who is disjoint from the *intersection* of Bob and Charlie's inputs: $Y \cap Z \cap U$.

Unfortunately, it seems unlikely that Bob and Charlie can find such a random set without first finding the exact intersection $Y \cap Z \cap U$, which is a harder task than to tell whether their sets intersect or not, and generally no better upper bound than n bits of communication is known for this task.

Attempt II: Charlie selects a random pair of sets from Alice and Bob's distribution. In this variant, in the beginning of each iteration, Alice and Bob tell Charlie whether their respective inputs have cardinality at most $n^{2/3}$ when restricted to U. As in the previous variant, if this condition holds for Alice and/or Bob, they send their small set to Charlie who can use it to deterministically detect an intersection using an additional $n^{2/3}\log(n)$ bits of communication.

If Both the input of Alice and Bob are small, then Charlie would like to publicly sample (an infinite sequence of) pairs of random sets: one from Alice's input distribution, and one from Bob's input distribution, conditioned on U and set cardinalities being at least $n^{2/3}$. Note that in this case, Charlie can indeed correctly sample such a pair, and if the typical pair is disjoint from Charlie's input $Z \cap U$, then such a pair could be identified with a small index. Unfortunately, in this case we have no guarantee on the *size* of the intersection of the pair of random sets (it is possible that the intersection of the random sets is empty), hence we have no guarantee that the protocol will terminate after $O(n^{2/3})$ iterations (or at all, for that matter).

Chapter 7

Conclusions and Open Problems

In this thesis we prove a bound of $\widetilde{\Theta}\left(n^{1-1/k}+k\right)$ on the communication complexity of the k-player Set Disjointness problem under product distributions, for the number-in-hand *shared blackboard* and *coordinator* communication models. In order to prove this bound, we introduce new techniques for both the upper and lower bounds, and explain why it seems hard to extend techniques used in previous work to the scenario discussed in this thesis.

We conclude this thesis with a list of open problems:

Problem 1. *Can the information-theoretic lower bound technique used in this thesis be extended to show a lower bound for Set Disjointness in the NOF communication model?*

Note that while our protocol for $k \geq \log n$ requires 1 simultaneous round (in the average case), our protocol for $k < \log n$ requires $O(\log\log n/\log k)$ simultaneous rounds, which we do not know to be tight.

Problem 2. *Can the number of rounds for $k < \log n$ be improved to $O(1)$?*

[BGK15] showed a smooth interpolation between the communication complexity bound for product distributions and the bound for general distributions, for 2 players.

Problem 3. *Is it possible to show a similar interpolation for k-player Disjointness in the coordinator model? I.e. show smooth interpolation between the $\widetilde{\Theta}\left(n^{1-1/k}+k\right)$ bound for product distributions, and the $\Theta(kn)$ bound for general distributions?*

Bibliography

[ABB+19] Pranjal Awasthi, Ainesh Bakshi, Maria-Florina Balcan, Colin White, and David P. Woodruff. Robust Communication-Optimal Distributed Clustering Algorithms. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132, pages 18:1–18:16, 2019.

[AMS99] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

[BCK+14] Joshua Brody, Amit Chakrabarti, Ranganath Kondapally, David P. Woodruff, and Grigory Yaroslavtsev. Beyond set disjointness: The communication complexity of finding the intersection. In *Proceedings of the 2014 ACM Symposium on Principles of Distributed Computing*, PODC '14, pages 106–113, 2014.

[BEO+13] Mark Braverman, Faith Ellen, Rotem Oshman, Toniann Pitassi, and Vinod Vaikuntanathan. A tight bound for set disjointness in the message-passing model. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 668–677, 2013.

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

[BGK15] Ralph Bottesch, Dmitry Gavinsky, and Hartmut Klauck. Correlation in hard distributions in communication complexity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA*, volume 40 of *LIPIcs*, pages 544–572, 2015.

[BJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, pages 209–218, 2002.

[BO15] Mark Braverman and Rotem Oshman. On information complexity in the broadcast model. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing, PODC 2015, Donostia-San Sebastián, Spain, July 21 - 23, 2015*, pages 355–364, 2015.

[BO17] Mark Braverman and Rotem Oshman. A rounds vs. communication tradeoff for multi-party set disjointness. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 144–155, 2017.

[CP10] Arkadev Chattopadhyay and Toniann Pitassi. The story of set disjointness. *ACM SIGACT News*, 41(3):59–85, 2010.

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

[CSWZ16] Jiecao Chen, He Sun, David Woodruff, and Qin Zhang. Communication-optimal distributed clustering. In *Advances in Neural Information Processing Systems*, volume 29, pages 3727–3735, 2016.

[Gro09] André Gronemeier. 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, February 26-28, 2009, Freiburg, Germany, Proceedings*, volume 3 of *LIPIcs*, pages 505–516, 2009.

[HRVZ20] Zengfeng Huang, Bozidar Radunovic, Milan Vojnovic, and Qin Zhang. Communication complexity of approximate maximum matching in the message-passing model. *Distributed Computing*, 33(6):515–531, 2020.

[Jay09] T. S. Jayram. 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, Berkeley, CA, USA, August 21-23, 2009. Proceedings*, volume 5687 of *Lecture Notes in Computer Science*, pages 562–573, 2009.

[Raz90] Alexander A Razborov. On the distributional complexity of disjointness. In *International Colloquium on Automata, Languages, and Programming*, pages 249–253, 1990.

[RY20] Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. 2020.

[She14] Alexander A Sherstov. Communication complexity theory: Thirty-five years of set disjointness. In *International Symposium on Mathematical Foundations of Computer Science*, pages 24–43, 2014.

[SK87] Georg Schnitger and Bala Kalyanasundaram. The probabilistic communication complexity of set intersection. In *Proceedings of the Second Annual Conference on*

Structure in Complexity Theory, Cornell University, Ithaca, New York, USA, June 16-19, 1987, 1987.

[WZ12] David P. Woodruff and Qin Zhang. Tight bounds for distributed functional monitoring. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 941–960, 2012.

[WZ13] David P. Woodruff and Qin Zhang. When distributed computation is communication expensive. In *Distributed Computing: 27th International Symposium, DISC 2013, Jerusalem, Israel, October 14-18, 2013. Proceedings*, pages 16–30, 2013.

תקציר

בבעיית זרות הקבוצות עם מספר-ביד מרובת-המשתתפים, ישנם k שחקנים עם קלטים פרטיים $X_1, \ldots, X_k \subseteq [n]$. מטרתם של השחקנים היא לבדוק האם $\bigcap_{\ell=1}^{k} X_\ell = \emptyset$. ידוע כי במודל התקשורת ה"לוח המשותף" , בעיית זרות הקבוצות דורשת $\Omega(n \log k + k)$ סיביות של תקשורת, ובמודל ה"מתאם", הבעיה דורשת $\Omega(kn)$ סיביות. עם זאת, שני חסמים תחתונים אלה דורשים שקלטי השחקנים יוכלו להיות מתואמים מאוד.

בעבודה זו אנו חוקרים את סיבוכיות התקשורת רבת-המשתתפים של בעיית הזרות של קבוצות תחת התפלגות מכפלה, ושואלים האם הבעיה הופכת להיות קלה משמעותית, כפי שידוע שמתקיים במקרה של שני שחקנים. התוצאה המרכזית שלנו היא חסם כמעט הדוק של $\widetilde{\Theta}\left(n^{1-\frac{1}{k}} + k\right)$ עבור מודל הלוח המשותף ועבור מודל המתאם. תוצאה זו מראה כי במודל הלוח המשותף, ככל שמספר השחקנים גדל, כך העובדה שלשחקנים יש קלטים בלתי תלויים עוזרת פחות ופחות, אך במודל המתאם, כאשר k הוא גדול, העובדה שלשחקנים יש קלטים בלתי תלויים הופכת את הבעיה לקלה משמעותית. גם החסם העליון שלנו וגם החסם התחתון משתמשים ברעיונות חדשים, כיוון שהשיטות המקוריות שפותחו עבור המקרה של שני שחקנים לא ניתנות להרחבה פשוטה ליותר משני שחקנים.

הפקולטה למדעים
מדויקים ע"ש ריימונד
ובברלי סאקלר
אוניברסיטת תל אביב

סיבוכיות התקשורת רבת-המשתתפים של בעיית הזרות של קבוצות תחת התפלגות מכפלה

חיבור זה הוגש כחלק מהדרישות לקבלת תואר "מוסמך במדעי המחשב" (M.Sc.)

מאת

טל רוט

העבודה נכתבה בבית הספר למדעי המחשב באוניברסיטת תל אביב

בהנחיית

פרופ' רותם אושמן ופרופ' נחום דרשוביץ

דצמבר 2020