Electronic Colloquium on Computational Complexity, Report No. 2 (2015)

The Communication Complexity of Number-In-Hand Set Disjointness with No Promise

Mark Braverman[*] and Rotem Oshman[†]

January 2, 2015

Abstract

Set disjointness is one of the most fundamental problems in communication complexity. In the multi-party number-in-hand version of set disjointness, k players receive private inputs $X_1, \ldots, X_k \subseteq \{1, \ldots, n\}$, and their goal is to determine whether or not $\bigcap_{i=1}^{k} X_i = \emptyset$. In this paper we prove a tight lower bound on the randomized communication complexity of multi-party number-in-hand set disjointness in the shared blackboard model. Our main tool is information complexity. Intuitively, in order to "become convinced" that their sets are disjoint, the players must discover, for each element $j \in [n]$, some player i such that $j \notin X_i$; this information is worth $n \log k$ bits. We are able to formalize this information and show that the players must learn a total of $\Omega(n \log k)$ bits of information about each other's inputs, and this implies a communication lower bound of $\Omega(n \log k)$ as well. Overall, we obtain the tight bound $\Theta(n \log k + k)$ on the problem, and give a simple matching deterministic upper bound.

1 Introduction

Set disjointness is one of the most fundamental problems in communication complexity; it has been studied in the classical two-player model [KS92, Raz92], in the number-on-forehead model (e.g., [She13]), and in many other settings (see [CP10]); it has applications in various areas ranging from streaming [AMS99] to data structures [Pat11], distributed computing [SHK$^+$12], circuit lower bounds [CFL83] and many other examples.

In this note we study the multiparty number-in-hand communication complexity of set disjointness, with communication over a shared blackboard. In this classical setup, there are k players, each with a private input $X_i \in \{0,1\}^n$, and the players wish to compute a joint function $f(X_1, \ldots, X_k)$ of their inputs. Specifically, in set disjointness, we interpret each X_i as a subset of $[n]$, and the players need to determine whether $\bigcap_{i=1}^{n} X_i = \emptyset$; formally, we shall denote

$$\textsc{Disj}_{n,k}(X_1, \ldots, X_k) = \neg \bigvee_{j=1}^{n} \bigwedge_{i=1}^{k} X_i^j,$$

where X_i^j is coordinate j of player i's input.

[*] Department of Computer Science, Princeton University, email: mbraverm@cs.princeton.edu. Research supported in part by an NSF CAREER award (CCF-1149888), a Turing Centenary Fellowship, a Packard Fellowship in Science and Engineering, and the Simons Collaboration on Algorithms and Geometry.

[†] Department of Computer Science, Tel Aviv University, email: roshman@tau.ac.il. Research supported in part by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation, Grant No. 4/11.

ISSN 1433-8092

A trivial reduction from the two-player lower bound of [KS92, Raz92] shows that $\Omega(n)$ bits are also required to solve set disjointness with k players, and an easy argument shows that $\Omega(k)$ is also a lower bound. As for upper bounds, there is a simple protocol with communication complexity $O(n \log n + k)$: the players go in order, with each player i writing on the board the coordinates j where $X_i^j = 0$, unless they already appear on the board. A player that has no new zero coordinates to contribute writes a single bit to indicate this. After all players have taken their turn, if there is some coordinate that does not appear on the board, then this coordinate is in the intersection.

A-priori, it is not obvious whether the "right answer" is $\Theta(n + k)$, $\Theta(n \log n + k)$, or somewhere in between. After all, in some cases where one might naively expect a logarithmic factor to arise, it does not: an example is the randomized protocol of Håstad and Wigderson [HW07], which solves two-player set disjointness under the promise that $|X| = |Y| = s$ in $O(s)$ bits, instead of $O(s \log n)$. Moreover, two players can even compute *the exact intersection* of their sets X, Y using $O(s)$ bits when $|X| = |Y| = s$: this result is proven in [BGPW13], which uses information complexity to compute the exact (up to lower-order terms) number of bits required for any s, and in [BCK$^+$14] it is shown that only $\log^* s$ rounds are needed to achieve a communication complexity of $O(s)$.

In this paper we show that the randomized communication complexity of set disjointness with k players is actually $\Theta(n \log k + k)$. Intuitively, the factor $\log k$ arises because the protocol must "find", for each $j \in [n]$, some player i with $X_i^j = 0$, before it can declare that $\bigcap_{i=1}^k X_i = 0$. The index of such a player is "worth" $\log k$ bits of information. For sufficiently large n this also translates to $O(\log k)$ bits of *communication* (for small n, the term $\Theta(k)$ dominates the communication complexity).

We remark that a *promise* version of set disjointness has received significant attention in the number-in-hand model, due to its connections to streaming lower bounds [AMS99]. In this problem the players are promised that either the sets intersect at exactly one element, or the sets are pairwise disjoint. The communication complexity of promise set disjointness is $\Theta(n/k)$ [Gro09] (note again the absence of the logarithmic factor), and in this paper we use some of the techniques developed in [BYJKS04] to prove lower bounds on promise set disjointness — specifically, the notion of *conditional information cost*, and the technique by which one decomposes the problem into many smaller problems (*direct sum*). (In [BYJKS04] the lower bound proved is $\Omega(n/k^2)$, and this was gradually improved until [Gro09] gave the tight lower bound of $\Omega(n/k)$.)

1.1 Information Complexity

Our main tool in this paper is *information complexity* [CSWY01, BYJKS04]: in order to prove a lower bound on the communication complexity of set disjointness we will show that any protocol that solves it must reveal a lot of information about the players' inputs, in the information-theoretic sense — the entropy of the inputs is significantly reduced once the transcript of the protocol is observed. Since the amount of information revealed by the transcript is bounded by the transcript's length, a lower bound on the information revealed also implies the same lower bound on communication.

The precise notion of information cost that we use in this paper was introduced in [BYJKS04], and is called *conditional information cost*. We will work with a distribution on inputs $(\boldsymbol{X}, \boldsymbol{D}) \sim \mu$ which has an "auxiliary variable" $\boldsymbol{D}$, conditioned on which the inputs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are independent. Conditional information cost measures the information an external observer learns about the inputs, given $\boldsymbol{D}$.

One of the advantages of this approach is that it allows us to decompose the problem into many small problems, prove that each small problem is "a little bit hard", and obtain a lower bound for the overall problem that is the sum of the bounds on the smaller problems. This is called *direct sum* [BYJKS04]. In our case (as in [BYJKS04]), since $\textsc{Disj}_{n,k} = \bigvee_{j=1}^n \bigwedge_{i=1}^k X_i^j$, we will break the problem up into n instances of one-bit AND, prove a lower bound of $\Omega(\log k)$ on AND, and obtain the lower bound of $\Omega(n \log k)$ for disjointness.

In order to show that computing AND must reveal $\Omega(\log k)$ bits of information about the input, we argue that whenever the output of the protocol is 0, the protocol's transcript must "point" to some player i whose input is zero. This is formalized by analyzing the posterior probability that $X_i = 0$

given the transcript. Roughly speaking, under our input distribution, the prior probability that $X_i = 0$ is only $O(1/k)$, but we will show that for "most" transcripts, there is a player i with a *constant* posterior probability of $X_i = 0$ given the transcript. The divergence between the prior and the posterior then is $\Omega(\log k)$ (see below for the formal definition of divergence), which corresponds to $\Omega(\log k)$ bits of information revealed.

1.2 Organization

We begin in Section 2 with a review of some basic notions from information theory. In Section 3 we prove the two lower bounds on set disjointness, of $\Omega(n \log k)$ and $\Omega(k)$ respectively. Section 3.1 gives the formal definition of conditional information cost and the direct sum statement (we include a proof for completeness). We then prove in Section 3.2 that computing a single-bit AND with k players has a conditional information cost of $\Omega(\log k)$; together with the direct sum lemma, this yields a lower bound of $\Omega(n \log k)$ on the conditional information cost and communication complexity of $\textsc{Disj}_{n,k}$.

In Section 3.3 we give an easy proof showing that computing a single-bit AND with k players requires $\Omega(k)$ bits of communication (even using randomness), which serves both to demonstrate a gap of $\Omega(k/\log k)$ between communication and information cost in the shared blackboard model, and also implies a lower bound of $\Omega(k)$ on the communication complexity of $\textsc{Disj}_{n,k}$. Finally, in Section 4 we give a deterministic protocol for $\textsc{Disj}_{n,k}$ with communication complexity $O(n \log k + k)$, which is optimal in light of our lower bounds in Section 3.

2 Background on Information Theory

Our main lower bound is based on *information complexity* [CSWY01] and follows the framework introduced in [BYJKS04]. We shall use the following basic notions. In general, for variables $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_\ell$ with joint distribution μ, we let $\mu(\boldsymbol{A}_i)$ denote the marginal distribution of $\boldsymbol{A}_i$, and $\mu(\boldsymbol{A}_i \mid \boldsymbol{A}_j = a_j)$ denote the distribution of $\boldsymbol{A}_i$ conditioned on $\boldsymbol{A}_j = a_j$ (and similarly for more variables).

Definition 1 (Entropy and conditional entropy)**.** *The* entropy *of a random variable $\boldsymbol{X} \sim \mu$ with support $\mathcal{X}$ is given by*

$$H(\boldsymbol{X}) = \sum_{x \in \mathcal{X}} \Pr_{\mu}[\boldsymbol{X} = x] \log \frac{1}{\Pr_\mu[\boldsymbol{X} = x]}.$$

For two random variables $\boldsymbol{X}, \boldsymbol{Y}$ with joint distribution μ, the conditional entropy of $\boldsymbol{X}$ given $\boldsymbol{Y}$ *is*

$$H(\boldsymbol{X} \mid \boldsymbol{Y}) = \mathop{\mathbb{E}}_{\boldsymbol{y} \sim \mu(\boldsymbol{Y})} \sum_{x \in \mathcal{X}} \Pr_{\mu(\boldsymbol{X} \mid \boldsymbol{Y} = \boldsymbol{y})}[\boldsymbol{X} = x] \log \frac{1}{\Pr_{\mu(\boldsymbol{X} \mid \boldsymbol{Y} = \boldsymbol{y})}[\boldsymbol{X} = x]}.$$

Definition 2 (KL divergence)**.** *Given two distributions μ_1, μ_2 with support $\mathcal{X}$, the* KL divergence *of μ_1 from μ_2 is*

$$\mathsf{D}\left(\frac{\mu_1}{\mu_2}\right) = \sum_{x \in \mathcal{X}} \mu_1(x) \log \frac{\mu_1(x)}{\mu_2(x)}.$$

Definition 3 (Mutual information and conditional mutual information)**.** *The* mutual information *between two random variables $\boldsymbol{X}, \boldsymbol{Y}$ is*

$$\mathrm{I}(\boldsymbol{X}; \boldsymbol{Y}) = H(\boldsymbol{X}) - H(\boldsymbol{X} \mid \boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y} \mid \boldsymbol{X}).$$

The conditional mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}$ given $\boldsymbol{Z}$ *is*

$$\mathrm{I}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z}) = H(\boldsymbol{X} \mid \boldsymbol{Z}) - H(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{Z}) = H(\boldsymbol{Y}, \boldsymbol{Z}) - H(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}).$$

Mutual information and KL divergence are related as follows:

$$\mathrm{I}(\boldsymbol{X};\boldsymbol{Y}) = \mathsf{D}\left(\frac{\mu(\boldsymbol{X},\boldsymbol{Y})}{\mu(\boldsymbol{X})\mu(\boldsymbol{Y})}\right) = \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\mu(\boldsymbol{Y})} \mathsf{D}\left(\frac{\mu(\boldsymbol{X} \mid \boldsymbol{Y}=\boldsymbol{y})}{\mu(\boldsymbol{X})}\right) = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu(\boldsymbol{X})} \mathsf{D}\left(\frac{\mu(\boldsymbol{Y} \mid \boldsymbol{X}=\boldsymbol{x})}{\mu(\boldsymbol{Y})}\right),$$

and similarly for conditional mutual information.

We will require the following useful lemma, from, e.g., [BR11]:

Lemma 1. *If $\boldsymbol{A}$ and $\boldsymbol{C}$ are independent given $\boldsymbol{D}$, then for any variable $\boldsymbol{B}$,*

$$\mathrm{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C},\boldsymbol{D}) \geq \mathrm{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{D}).$$

Proof.

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

□

Another easy fact is the following (also noted in, e.g., [BR11]):

Lemma 2. *If $\boldsymbol{A}$ and $\boldsymbol{C}$ are independent given $\boldsymbol{D}$, then for any $\boldsymbol{B}$ we have*

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

Proof. By the chain rule,

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

Since $\boldsymbol{A}$ and $\boldsymbol{C}$ are independent given $\boldsymbol{D}$ we have $\mathrm{I}(\boldsymbol{A};\boldsymbol{C} \mid \boldsymbol{D}) = 0$, and the claim follows. □

3 Lower Bounds

3.1 $\Omega(n \log k)$ Lower Bound

Conditional information cost and direct sum. To prove a lower bound of $\Omega(n \log k)$ on the communication complexity of $\textsc{Disj}_{n,k}$, we use the notion of *conditional information cost*, introduced in [BYJKS04].

Definition 4. *Let Π be a randomized protocol, and let μ be a distribution on $(\{0,1\}^n)^k \times \mathcal{D}$ for some domain $\mathcal{D}$. The* conditional information cost *of Π with respect to μ is given by*

$$\mathop{\mathrm{CIC}}_{\mu}(\Pi) = \mathop{\mathrm{I}}_{(\boldsymbol{X},\boldsymbol{D})\sim\mu}(\boldsymbol{\Pi};\boldsymbol{X} \mid \boldsymbol{D}).$$

For a problem P and an error parameter $\epsilon \in (0,1)$, we define

$$\mathop{\mathrm{CIC}}_{\mu,\epsilon}(P) = \inf_{\Pi} \mathop{\mathrm{CIC}}_{\mu}(\Pi),$$

where the infimum is taken over all protocols that solve P with worst-case error ϵ.

Note that, since $\mathrm{I}(\boldsymbol{A};\boldsymbol{B} \mid \boldsymbol{C}) \leq H(\boldsymbol{A}) \leq |\boldsymbol{A}|$ always, a lower bound on the conditional information cost of a problem P implies a corresponding lower bound on the communication complexity of P.

Following [BYJKS04], we decompose the disjointness problem into n copies of k-player $\textsc{And}$ on a single bit, and argue that the cost adds up linearly:

Lemma 3. *Let μ be a distribution on $\{0,1\}^k \times \mathcal{D}$, with the following properties:*

1. *For any (X, D) in the support of μ, $\bigwedge_{i=1}^{k} X_i = 0$, and*
2. *For any $D \in \mathcal{D}$, when we draw $(\boldsymbol{X}, \boldsymbol{D}) \sim \mu$ conditioned on $\boldsymbol{D} = D$, the variables $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are independent.*

Then

$$\operatorname*{CIC}_{\mu^n,\epsilon}(\textsc{Disj}_{n,k}) \geq n \cdot \operatorname*{CIC}_{\mu,\epsilon}(\textsc{And}_k).$$

Proof. We show that for any protocol Π that solves $\textsc{Disj}_{n,k}$ with error ϵ, there is a protocol Θ that solves $\textsc{And}_k$ with error ϵ and has

$$\operatorname*{I}_{(\boldsymbol{U},\boldsymbol{F})\sim\mu}(\boldsymbol{U}; \Theta \mid \boldsymbol{F}) \leq \frac{\operatorname{I}_{(\boldsymbol{X},\boldsymbol{D})\sim\mu^n}(\boldsymbol{X}; \Pi' \mid \boldsymbol{D})}{n}.$$

The claim follows.

We construct Θ as follows: on input U, the players generate an input X for Π by agreeing on a uniform coordinate $i \in [n]$ and on a vector $\boldsymbol{H} \sim \mu(\boldsymbol{F})^{n-1}$ using public randomness, setting $X^i = U$, and then each player p privately samples each coordinate X_p^j for $j \neq i$ from its distribution given H^j. Then they call Π with the resulting input. Since under μ we have $\bigwedge X^j = 0$ for each $j \neq i$, the output to disjointness on X is 1 iff $\bigwedge U = 1$, so Θ has the same error probability as Π.

On input $(\boldsymbol{U}, \boldsymbol{F}) \sim \mu$, let $(\boldsymbol{X}, \boldsymbol{D})$ be the input to Π generated as above. The distribution of $(\boldsymbol{X}, \boldsymbol{D})$ is μ^n. We have

$$\begin{aligned}
\operatorname*{I}_{(\boldsymbol{U},\boldsymbol{F})\sim\mu}(\boldsymbol{U}; \Theta \mid \boldsymbol{F}) &= \operatorname*{I}_{(\boldsymbol{X},\boldsymbol{D})\sim\mu^n}\left(\boldsymbol{X}^{\boldsymbol{i}}; \Pi, \boldsymbol{i}, \boldsymbol{D}^{-\boldsymbol{i}} \mid \boldsymbol{D}^{\boldsymbol{i}}\right) \\
&= \operatorname*{I}_{(\boldsymbol{X},\boldsymbol{D})\sim\mu^n}\left(\boldsymbol{X}^{\boldsymbol{i}}; \Pi \mid \boldsymbol{i}, \boldsymbol{D}\right) && \text{(by Lemma 2)} \\
&\leq \operatorname*{I}_{(\boldsymbol{X},\boldsymbol{D})\sim\mu^n}\left(\boldsymbol{X}^{\boldsymbol{i}}; \Pi \mid \boldsymbol{i}, \boldsymbol{X}^{<\boldsymbol{i}}, \boldsymbol{D}\right) && \text{(by Lemma 1)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \operatorname*{I}_{(\boldsymbol{X},\boldsymbol{D})\sim\mu^n}\left(\boldsymbol{X}^{i}; \Pi \mid \boldsymbol{X}^{<i}, \boldsymbol{D}\right) = \frac{\operatorname{I}_{(\boldsymbol{X},\boldsymbol{D})\sim\mu^n}(\boldsymbol{X}; \Pi' \mid \boldsymbol{D})}{n}.
\end{aligned}$$

□

In the next section we will give a distribution μ satisfying the conditions of the lemma above, under which $\mathrm{CIC}_{\mu,\delta}(\textsc{And}_k) \geq \Omega(\log k)$ for some small constant error probability δ. Therefore, by the lemma, $\mathrm{CIC}_{\mu,\delta}(\textsc{Disj}_{n,k}) \geq \Omega(n \log k)$.

3.2 Lower Bound for $\textsc{And}_k$

We use the following distribution μ:

(1) Select one player $\boldsymbol{Z} \in [k]$; this player receives zero.

(2) Every other player's input is an iid Bernoulli RV with probability $1/k$ of being 0.

That is, for each $i \neq z$ we have

$$\Pr\left[\boldsymbol{X}_i = b \mid \boldsymbol{Z} = z\right] = \mathsf{Ber}_{1/k}(b).$$

Our analysis will be conditional in $\boldsymbol{Z}$. We will also assume that the error δ of the protocol is bounded by some small constant we can choose.

The distribution is motivated as follows: we must always assign zero to some player, in order to satisfy the conditions of Lemma 3. Intuitively, we wish to have as few zeroes as possible in the input, to make it hard for the protocol to find one of them and become "convinced" that $\bigwedge X = 0$. However, this concern is balanced by the fact that we require a residual entropy of at least $\Omega(\log k)$ even when we condition on the identity of the special zero player (to get the lower bound on the conditional information cost); so we cannot assign 1 to all players except the special player. We would ideally choose a small constant number of non-special players and assign them zero, but this would not be a product distribution; we emulate this idea by assigning each player 0 w.p. $1/k$.

Outline of the lower bound. Under the distribution μ described above, it is somewhat likely (constant probability) that besides the special player $\boldsymbol{Z}$, exactly one other player receives zero; we will condition on this event, which makes the distribution "symmetric" in a sense—a uniformly random pair of players receive zero, and the others receive one. Notice that conditioned on exactly two players receiving zero and on $\boldsymbol{Z}$, the identity of the other player that received zero is uniform in $[k] \setminus \{\boldsymbol{Z}\}$, so it is worth roughly $\log k$ bits of information. Our proof will argue that the protocol must *find* some player that received zero, and since, due to the symmetry, it "cannot tell" which is the special player $\boldsymbol{Z}$, it finds the *other* player that got zero with probability $1/2$. This will show that the conditional information cost of the protocol is $\Omega(\log k)$.

Posterior probabilities. What does it mean for the protocol to "find a zero"? We formalize this in terms of the *posterior probability distribution* of the input $\boldsymbol{X}$ given the transcript and $\boldsymbol{Z}$. "Finding a zero" will mean that for some $i \in [k]$, the posterior probability of $\boldsymbol{X}_i = 0$ given the transcript is *constant* (whereas the prior was only $1/k$).

Recall that the conditional information cost can be re-phrased in terms of the KL-divergence between the posterior and prior of $\boldsymbol{X}$:

$$\mathrm{I}(\boldsymbol{\Pi}; \boldsymbol{X} \mid \boldsymbol{Z}) = \mathop{\mathbb{E}}_{\boldsymbol{\ell},\boldsymbol{z}} \mathsf{D}\left(\frac{\mu(\boldsymbol{X} \mid \boldsymbol{\Pi} = \boldsymbol{\ell}, \boldsymbol{Z} = \boldsymbol{z})}{\mu(\boldsymbol{X} \mid \boldsymbol{Z} = \boldsymbol{z})} \right).$$

It is more convenient to work with the distributions of the individual inputs $\boldsymbol{X}_i$ rather than the distribution of the entire input $\boldsymbol{X}$. And this will be sufficient to prove the lower bound:

Lemma 4.

$$\mathrm{I}(\boldsymbol{\Pi}; \boldsymbol{X} \mid \boldsymbol{Z}) \geq \sum_{i=1}^{k} \mathop{\mathbb{E}}_{\boldsymbol{\ell},\boldsymbol{z}} \mathsf{D}\left(\frac{\mu(\boldsymbol{X}_i \mid \boldsymbol{\Pi} = \boldsymbol{\ell}, \boldsymbol{Z} = \boldsymbol{z})}{\mu(\boldsymbol{X}_i \mid \boldsymbol{Z} = \boldsymbol{z})} \right)$$

Proof. By the chain rule,

$$\mathrm{I}(\boldsymbol{\Pi}; \boldsymbol{X} \mid \boldsymbol{Z}) = \sum_{i=1}^{k} \mathrm{I}(\boldsymbol{\Pi}; \boldsymbol{X}_i \mid \boldsymbol{X}_{<i}, \boldsymbol{Z}).$$

Because $\boldsymbol{X}_{<i}$ is independent of $\boldsymbol{X}_i$ given $\boldsymbol{Z}$, Lemma 1 implies that

$$\sum_{i=1}^{k} \mathrm{I}(\boldsymbol{\Pi}; \boldsymbol{X}_i \mid \boldsymbol{X}_{<i}, \boldsymbol{Z}) \geq \sum_{i=1}^{k} \mathrm{I}(\boldsymbol{\Pi}; \boldsymbol{X}_i \mid \boldsymbol{Z}) = \sum_{i=1}^{k} \mathop{\mathbb{E}}_{\boldsymbol{\ell},\boldsymbol{z}} \mathsf{D}\left(\frac{\mu(\boldsymbol{X}_i \mid \boldsymbol{\Pi} = \boldsymbol{\ell}, \boldsymbol{Z} = \boldsymbol{z})}{\mu(\boldsymbol{X}_i \mid \boldsymbol{Z} = \boldsymbol{z})} \right),$$

and we are done. □

When $\boldsymbol{z} = i$, the divergence is zero: we know in advance that $\boldsymbol{X}_i = 0$, so the posterior and the prior are the same. So in fact we are interested in bounding

$$\mathrm{I}(\boldsymbol{\Pi}; \boldsymbol{X} \mid \boldsymbol{Z}) \geq \frac{1}{k} \sum_{i=1}^{k} \sum_{z \neq i} \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \pi \mid \boldsymbol{Z} = z} \mathsf{D}\left(\frac{\mu(\boldsymbol{X}_i \mid \boldsymbol{\Pi} = \boldsymbol{\ell}, \boldsymbol{Z} = z)}{\mu(\boldsymbol{X}_i \mid \boldsymbol{Z} = z)} \right). \tag{1}$$

Our goal is to show that the average transcript ℓ "points" to at least one player that received zero and is not the special player—that is, for some $i \neq z$, the posterior probability of $\boldsymbol{X}_i = 0$ given $\boldsymbol{\Pi} = \ell, \boldsymbol{Z} = z$ is constant. Since the prior is only $1/k$, this gives us a divergence of $\Omega(\log k)$: specifically, if the posterior probability of 0 is p, then

$$\begin{aligned}
&\mathsf{D}\left(\frac{\mu(\boldsymbol{X}_i \mid \boldsymbol{\Pi} = \ell, \boldsymbol{Z} = z)}{\mu(\boldsymbol{X}_i \mid \boldsymbol{Z} = z)}\right) = \sum_{b=0,1} \Pr\left[\boldsymbol{X}_i = b \mid \boldsymbol{\Pi} = \ell, \boldsymbol{Z} = z\right] \log \frac{\Pr\left[\boldsymbol{X}_i = b \mid \boldsymbol{\Pi} = \ell, \boldsymbol{Z} = z\right]}{\Pr\left[\boldsymbol{X}_i = b \mid \boldsymbol{Z} = z\right]} \\
&= p \log \frac{p}{\frac{1}{k}} + (1-p) \log \frac{1-p}{1-\frac{1}{k}} \\
&= p \log p + p \log k + (1-p)\log(1-p) + (1-p) \log \frac{1}{1-\frac{1}{k}} \\
&\geq p \log k - H(p) \geq p \log(k) - 1. \qquad (2)
\end{aligned}$$

(As usual, we assume the convention that $0 \log 0 = 0$.) Plugging (2) into (1), with constant p, will yield the lower bound. But first we must show that indeed the average transcript "points" to a player that received zero.

Finding zeroes. In order to analyze the posterior probabilities, we examine the structure of the protocol. As usual, we will use the fact that for any particular transcript, the probability of getting this transcript can be broken up into the product of functions that each depend only on the input to a single player:

Lemma 5. *For any leaf (transcript) ℓ, there exist functions $\left\{q^\ell_{i,b}\right\}_{i \in [k], b \in \{0,1\}}$ such that*

$$\Pr\left[\boldsymbol{\Pi}(X) = \ell\right] = \prod_{i=1}^{k} q^\ell_{i,X_i}.$$

Proof. For convenience, we assume that all transcripts of the protocol have the same length. We prove the claim for any prefix of a transcript as well as for complete transcripts. Let $\boldsymbol{\Pi}_s$ denote the first s symbols of the transcript. The proof is by induction on the prefix length; the base case is the empty prefix, for which the claim is immediate.

Suppose we have already proven the claim for a prefix ℓ of length s, and let $\ell' = \ell a$ for some symbol a. Let i be the identity of the player whose turn it is to speak after seeing transcript ℓ. Then the probability that the transcript of length $s+1$ on input X will be ℓa is given by

$$\Pr\left[\boldsymbol{\Pi}_{s+1}(X) = \ell a\right] = \Pr\left[\boldsymbol{\Pi}_s(X) = \ell\right] \cdot \Pr\left[\text{player } i \text{ says } a \text{ on input } X_i, \text{ transcript } \ell\right].$$

By the induction hypothesis we can write

$$\Pr\left[\boldsymbol{\Pi}_s(X) = \ell\right] = \prod_{i=1}^{k} q^\ell_{i,X_i}.$$

To prove the claim for $\ell' = \ell a$ we set $q^{\ell'}_{j,b} = q^\ell_{j,b}$ for each $j \neq i$, and for player i we set $q^{\ell'}_{i,b} = q^\ell_{i,b} \cdot \Pr\left[\text{player } i \text{ says } a \text{ on input } b, \text{ transcript } \ell\right]$.

□

As we said above, we want to argue that in many leafs (transcripts), some player "shows his hand" and reveals that its input was zero; indeed, we will show that for some player $i \neq \boldsymbol{Z}$, the probability of getting this transcript given $\boldsymbol{X}_i = 0$ is $\Omega(k)$ times larger than given $\boldsymbol{X}_i = 1$. More formally, let us denote

$$\alpha_i^\ell := q_{i,0}^\ell / q_{i,1}^\ell. \tag{3}$$

(The case where $q_{i,1}^\ell = 0$ is trivial, as in this case the posterior probability of zero is 1, so we will generally assume $q_{i,1}^\ell > 0$.) We will show that "many" leafs ℓ have $\max_i \alpha_i^\ell = \Omega(k)$. To bring the argument back to the posterior probability of $\boldsymbol{X}_i = 0$ given $\mathbf{\Pi} = \ell$ we will use Bayes' rule.

To apply Bayes' rule, let us calculate the probabilities involved. For $i \neq z$ and $b \in \{0, 1\}$,

$$\begin{aligned}
&\Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X}_i = b, \boldsymbol{Z} = z\right] = \sum_X \Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X} = X, \boldsymbol{Z} = z\right] \Pr\left[\boldsymbol{X} = X \mid \boldsymbol{X}_i = b, \boldsymbol{Z} = z\right] \\
&= \sum_{X : X_i = b, X_z = 0} \left(\prod_{j \neq i,z}^{k} \mathsf{Ber}_{1/k}(X_i) \cdot \prod_{j=1}^{k} q_{j,X_j}^\ell \right) \\
&= q_{i,b}^\ell \cdot q_{z,0}^\ell \sum_{X_{-i,z}} \left(\prod_{j \neq i,z}^{k} q_{j,X_j}^\ell \, \mathsf{Ber}_{1/k}(X_i) \right).
\end{aligned}$$

We also have

$$\begin{aligned}
\Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{Z} = z\right] &= \sum_{b=0,1} \Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X}_i = b, \boldsymbol{Z} = z\right] \Pr\left[\boldsymbol{X}_i = b \mid \boldsymbol{Z} = z\right] \\
&= q_{z,0}^\ell \sum_{X_{-i,z}} \left(\prod_{j \neq i,z}^{k} q_{j,X_j}^\ell \, \mathsf{Ber}_{1/k}(X_i) \right) \cdot \left[\frac{1}{k} q_{i,0}^\ell + \left(1 - \frac{1}{k}\right) q_{i,1}^\ell \right].
\end{aligned}$$

Therefore the posterior probability of $\boldsymbol{X}_i = b$ at leaf ℓ, given $\boldsymbol{Z} = z$ for $z \neq i$, is

$$\begin{aligned}
\Pr\left[\boldsymbol{X}_i = b \mid \mathbf{\Pi} = \ell, \boldsymbol{Z} = z\right] &= \frac{\Pr\left[\boldsymbol{X}_i = b \mid \boldsymbol{Z} = z\right] \Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X}_i = b, \boldsymbol{Z} = z\right]}{\Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{Z} = z\right]} \\
&= \frac{\mathsf{Ber}_{1/k}(b) q_{i,b}^\ell}{\frac{1}{k} q_{i,0}^\ell + \left(1 - \frac{1}{k}\right) q_{i,1}^\ell}.
\end{aligned}$$

Since this does not depend on the specific value of $\boldsymbol{Z}$, we also have

$$\begin{aligned}
\Pr\left[\boldsymbol{X}_i = b \mid \mathbf{\Pi} = \ell, \boldsymbol{Z} \neq i\right] &= \sum_{z \neq i} \Pr\left[\boldsymbol{X}_i = b \wedge \boldsymbol{Z} = z \mid \mathbf{\Pi} = \ell, \boldsymbol{Z} \neq i\right] \\
&= \sum_{z \neq i} \Pr\left[\boldsymbol{Z} = z \mid \mathbf{\Pi} = \ell, \boldsymbol{Z} \neq i\right] \Pr\left[\boldsymbol{X}_i = b \mid \mathbf{\Pi} = \ell, \boldsymbol{Z} = z\right] \\
&= \sum_{z \neq i} \Pr\left[\boldsymbol{Z} = z \mid \mathbf{\Pi} = \ell, \boldsymbol{Z} \neq i\right] \frac{\mathsf{Ber}_{1/k}(b) q_{i,b}^\ell}{\frac{1}{k} q_{i,0}^\ell + \left(1 - \frac{1}{k}\right) q_{i,1}^\ell} \\
&= \frac{\mathsf{Ber}_{1/k}(b) q_{i,b}^\ell}{\frac{1}{k} q_{i,0}^\ell + \left(1 - \frac{1}{k}\right) q_{i,1}^\ell} \cdot \sum_{z \neq i} \Pr\left[\boldsymbol{Z} = z \mid \mathbf{\Pi} = \ell, \boldsymbol{Z} \neq i\right] \\
&= \frac{\mathsf{Ber}_{1/k}(b) q_{i,b}^\ell}{\frac{1}{k} q_{i,0}^\ell + \left(1 - \frac{1}{k}\right) q_{i,1}^\ell}.
\end{aligned}$$

In particular, for a transcript ℓ, the posterior probability of $\boldsymbol{X}_i = 0$ given $\boldsymbol{Z} \neq i$ and $\mathbf{\Pi} = \ell$ is

$$\frac{q_{i,0}^\ell}{q_{i,0}^\ell + (k-1) q_{i,1}} = \frac{\alpha_i^\ell}{\alpha_i^\ell + k - 1} \geq \frac{\alpha_i^\ell}{\alpha_i^\ell + k}, \tag{4}$$

unless $q_{i,1} = 0$, in which case the posterior is of course 1. If $\alpha_i^\ell = \Omega(k)$ then this becomes constant.

"Good leafs". Our task now is to show that for "many" leafs ℓ we indeed have, for some $i \in [k]$, $\alpha_i^\ell = \Omega(k)$; in other words, $\max_i \alpha_i^\ell = \Omega(k)$.

We focus our attention on inputs that have exactly two zeroes, and our "good leafs" will also be chosen with respect to their behavior on these inputs. Let $\mathcal{X}_c$ denote the set of inputs with c zeroes for $c \in [k]$. Note that the correct answer for AND_k is 1 on inputs in $\mathcal{X}_0$ (there is only one such input, 1^k) and 0 on inputs in $\mathcal{X}_c$ for $c \geq 1$.

Let π_2 be the distribution of leafs conditioned on the input being in $\mathcal{X}_2$:

$$\pi_2(\ell) = \sum_{X \in \mathcal{X}_2} \left[\mu(X|_{\mathcal{X}_2}) \prod_{i=1}^{k} q_{i,X_i}^\ell \right].$$

Let L be the set of leafs satisfying the following constraints: for each $\ell \in L$, the output of the protocol is 0, and also

$$\pi_2(\ell) \geq C \cdot \prod_{i=1}^{k} q_{i,1}^\ell,$$

where C is some large constant whose value will be chosen later. In other words, L is the set of leafs with output 0 which "strongly prefer" inputs with two zeroes over 1^k. (In particular, these leafs do not contribute much to the error of the protocol.)

To show that L has large mass under π_2, let us partition the complement of L_2 into two sets B_0, B_1 based on the output value on each leaf. Neither set can be large:

- The leafs in B_1 cannot have large mass under π_2 because they yield the wrong output (1) on all inputs in $\mathcal{X}_2$:

$$\begin{aligned} \pi_2(B_1) &= \sum_{\ell \in B_1} \sum_{X \in \mathcal{X}_2} \mu(X|_{\mathcal{X}_2}) \Pr[\mathbf{\Pi} = \ell \mid \boldsymbol{X} = X] \\ &= \frac{1}{\mu(\mathcal{X}_2)} \sum_{X \in \mathcal{X}_2} \mu(X) \sum_{\ell \in B_1} \Pr[\mathbf{\Pi} = \ell \mid \boldsymbol{X} = X] \\ &\leq \frac{1}{\mu(\mathcal{X}_2)} \sum_{X \in \mathcal{X}_{\geq 1}} \mu(X) \sum_{\ell \in B_1} \Pr[\mathbf{\Pi} = \ell \mid \boldsymbol{X} = X] \\ &\leq \frac{1}{\mu(\mathcal{X}_2)} \Pr_\mu[\Pi \text{ outputs } 1] \leq \frac{\delta}{\mu(\mathcal{X}_2)}. \end{aligned}$$

- B_0 contains leafs on which the output is 0, but

$$\pi_2(\ell) < C \cdot \prod_{i=1}^{k} q_{i,1}^\ell.$$

These leafs contribute to the error of the protocol on 1^k at least in proportion to their mass under π_2, and therefore they also cannot have large mass:

$$\begin{aligned} \Pr\left[\mathbf{\Pi}(1^k) \text{ outputs } 0\right] &\geq \sum_{\ell \in B_0} \Pr\left[\mathbf{\Pi}(1^k) = \ell\right] \\ &= \sum_{\ell \in B_0} \prod_{i=1}^{k} q_{i,1}^\ell \\ &> \sum_{\ell \in B_0} \pi_2(\ell)/C = \pi_2(B_0)/C, \end{aligned}$$

and therefore $\pi_2(B_0) < C \cdot \delta$.

Together we have that assuming $C \cdot \delta < 1/200$ and $\delta/\mu(\mathcal{X}_2) < 1/200$ (both constant requirements), $\pi_2(B_0 \cup B_1) < 1/100$. Note that we can choose C arbitrarily large, and compensate by assuming a smaller error probability δ.

For any leaf $\ell \in L$, we have

$$\pi_2(\ell) = \sum_{X \in \mathcal{X}_2} \mu(X|_{\mathcal{X}_2}) \prod_{i=1}^{k} q^{\ell}_{i,X_i} \geq C \prod_{i=1}^{k} q^{\ell}_{i,1}.$$

Given membership in $\mathcal{X}_2$, all two-zero inputs are equally likely, so $\mu(X|_{\mathcal{X}_2}) = \frac{1}{\binom{k}{2}}$ for any $X \in \mathcal{X}_2$. Note that for each $X \in \mathcal{X}_2$ we can write $\prod_{i=1}^{k} q^{\ell}_{i,X_i} = q^{\ell}_{j_1,0} q^{\ell}_{j_2,0} \cdot \prod_{i \neq j_1,j_2} q^{\ell}_{i,1}$, where $j_1 \neq j_2$ are the two indices where X has zero. Dividing both sides by $\prod_{i=1}^{k} q^{\ell}_{i,1}$ (and renaming the indices for convenience), we obtain

$$\frac{1}{\binom{k}{2}} \sum_{i<j} \alpha^{\ell}_i \alpha^{\ell}_j \geq C.$$

Because $\sum_{i<j} a_i a_i \leq (\sum_i a_i)^2$ for any sequence $a_1, \ldots, a_N$, we get that

$$\sum_i \alpha^{\ell}_i \geq \sqrt{\frac{k(k-1)}{2} \cdot C} \geq \sqrt{\frac{k^2}{4} \cdot C} = \frac{\sqrt{C}}{2} k. \tag{5}$$

In other words, the *sum* of the coefficients is linear. However, our goal is to show that for many leafs the *maximum* coefficient is linear, and this does not necessarily hold for all of L. We focus our attention on the subset $L' \subseteq L$ of leafs ℓ satisfying

$$\Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X} \in \mathcal{X}_2\right] \geq \frac{1}{2} \Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X} \in \mathcal{X}_3\right],$$

that is, leafs that "like" inputs with two zeroes not much less than inputs with three zeroes. We have $\pi_2(L') \geq \pi_2(L) - 1/2$, because

$$\pi_2(L \setminus L') = \sum_{\ell \in L \setminus L'} \Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X} \in \mathcal{X}_2\right] \leq \frac{1}{2} \sum_{\ell \in L \setminus L'} \Pr\left[\mathbf{\Pi} = \ell \mid \boldsymbol{X} \in \mathcal{X}_3\right] \leq \frac{1}{2}.$$

Now fix $\ell \in L'$, and let us show that for some $i \in [k]$, the posterior probability of $X_i = 0$ given ℓ is $\Omega(k)$ times the posterior probability of 1. If there is some i for which $q^{\ell}_{i,1} = 0$, then we are done, so assume that this is not the case. Because $\ell \in L'$,

$$\sum_{X \in \mathcal{X}_2} \mu(X|_{\mathcal{X}_2}) \prod_{i=1}^{k} q^{\ell}_{i,X_i} \geq \frac{1}{2} \sum_{X \in \mathcal{X}_3} \mu(X|_{\mathcal{X}_3}) \prod_{i=1}^{k} q^{\ell}_{i,X_i},$$

that is,

$$\frac{1}{\binom{k}{2}} \sum_{i<j} \alpha^{\ell}_i \alpha^{\ell}_j \geq \frac{1}{2\binom{k}{3}} \sum_{i<j<m} \alpha^{\ell}_i \alpha^{\ell}_j \alpha^{\ell}_m. \tag{6}$$

Now assume for the sake of contradiction that for a constant C' whose value will be fixed later, we have $\alpha^{\ell}_i < C'k$ for each $i \in [k]$. Then

$$\begin{aligned}
6 \sum_{i<j<m} \alpha^{\ell}_i \alpha^{\ell}_j \alpha^{\ell}_m &= \left(\sum_i \alpha^{\ell}_i\right)^3 - 3 \sum_{i \neq j} \left(\alpha^{\ell}_i\right)^2 \alpha^{\ell}_j - \sum_i \alpha_i^3 \\
&> \left(\sum_i \alpha^{\ell}_i\right)^3 - 3C'k \sum_{i \neq j} \alpha^{\ell}_i \alpha^{\ell}_j - (C')^2 k^2 \sum_i \alpha_i \\
&\geq \left(\sum_i \alpha^{\ell}_i\right)^3 - 3C'k \left(\sum_i \alpha^{\ell}_i\right)^2 - (C')^2 k^2 \sum_i \alpha_i.
\end{aligned}$$

From (5) we know that $k \leq 2\sum_i \alpha_i^\ell/(\sqrt{C})$, so

$$6\sum_{i<j<m} \alpha_i^\ell \alpha_j^\ell \alpha_m^\ell \geq \left(\sum_i \alpha_i^\ell\right)^3 - 6\frac{C'}{\sqrt{C}}\left(\sum_i \alpha_i^\ell\right)^3 - 4\frac{(C')^2}{C}\left(\sum_i \alpha_i\right)^3.$$

If we choose, e.g., $C' < \sqrt{C}/100$, we get that

$$6\sum_{i<j<m} \alpha_i^\ell \alpha_j^\ell \alpha_m^\ell \geq \frac{1}{2}\left(\sum_i \alpha_i^\ell\right)^3 \overset{(5)}{\geq} \frac{1}{4}\sqrt{C}k\left(\sum_i \alpha_i^\ell\right)^2,$$

and therefore

$$\frac{1}{2\binom{k}{3}}\sum_{i<j<m} \alpha_i^\ell \alpha_j^\ell \alpha_m^\ell \geq \frac{\sqrt{C}}{16k^2}\left(\sum_i \alpha_i^\ell\right)^2.$$

This gives us a lower bound on the right-hand side in (6). The left-hand side is bounded from above by

$$\frac{1}{\binom{k}{2}}\sum_{i<j} \alpha_i^\ell \alpha_j^\ell \leq \frac{4}{k^2}\left(\sum_i \alpha_i^\ell\right)^2.$$

If we choose C sufficiently large we obtain a contradiction to (6). Note that the value of C' is constrained only by the value of C, so by increasing C (which requires only assuming a smaller error probability δ for the protocol) we can obtain an arbitrarily large lower bound $\max_i \alpha_i^\ell \geq C'k$.

Wrap-up. So far we have shown the following: fix a constant probability $p \in (0,1)$, and let $c > 0$ be a constant such that $c/(c+1) \geq p$. (We can find such c because $\lim_{c\to\infty} c/(c+1) = 1$.) Then there exist constants $\delta, p_2 \in (0,1)$ such that any protocol that solves AND with error probability at most δ has a set L' of leafs such that $\pi_2(L') \geq p_2$, and for each $\ell \in L'$ there is a player $i = i(\ell)$ with $\alpha_i^\ell \geq ck$. For this player we have

$$\Pr\left[\boldsymbol{X}_{i(\ell)} = 0 \mid \boldsymbol{\Pi} = \ell, \boldsymbol{Z} \neq i\right] \overset{(4)}{\geq} \frac{\alpha_{i(\ell)}^\ell}{\alpha_{i(\ell)}^\ell + k} \overset{(*)}{\geq} \frac{ck}{ck+k} \geq p. \tag{7}$$

$(*)$: The function $x/(x+k)$ is increasing in x for any $k > 0$, as

$$\frac{d}{dx}\left(\frac{x}{x+k}\right) = \frac{1}{x+k} - \frac{x}{(x+k)^2} = \frac{k}{(x+k)^2} > 0.$$

It follows from (2) that for any $z \neq i(\ell)$, assuming $k \geq 2^{2/p}$,

$$\mathsf{D}\left(\frac{\mu(\boldsymbol{X}_{i(\ell)} \mid \boldsymbol{\Pi} = \ell, \boldsymbol{Z} = z)}{\mu(\boldsymbol{X}_{i(\ell)} \mid \boldsymbol{Z} = z)}\right) \geq p\log k - 1 \geq (p/2)\log k, \tag{8}$$

and hence

$$
\begin{aligned}
\mathrm{I}(\boldsymbol{\Pi};\boldsymbol{X}\mid\boldsymbol{Z}) &\overset{(1)}{\geq} \frac{1}{k}\sum_{i=1}^{k}\sum_{z\neq i}\mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\pi|\boldsymbol{Z}=z}\mathsf{D}\left(\frac{\mu(\boldsymbol{X}_i\mid\boldsymbol{\Pi}=\boldsymbol{\ell},\boldsymbol{Z}=z)}{\mu(\boldsymbol{X}_i\mid\boldsymbol{Z}=z)}\right)\\
&=\frac{1}{k}\sum_{i=1}^{k}\sum_{z\neq i}\sum_{\ell}\Pr\left[\boldsymbol{\Pi}=\ell\mid\boldsymbol{Z}=z\right]\mathsf{D}\left(\frac{\mu(\boldsymbol{X}_i\mid\boldsymbol{\Pi}=\ell,\boldsymbol{Z}=z)}{\mu(\boldsymbol{X}_i\mid\boldsymbol{Z}=z)}\right)\\
&\geq\frac{1}{k}\sum_{i=1}^{k}\sum_{z\neq i}\sum_{\ell\in L'}\sum_{X\in\mathcal{X}_2}\Pr[\boldsymbol{X}=X\mid\boldsymbol{Z}=z]\Pr\left[\boldsymbol{\Pi}(X)=\ell\right]\mathsf{D}\left(\frac{\mu(\boldsymbol{X}_i\mid\boldsymbol{\Pi}=\ell,\boldsymbol{Z}=z)}{\mu(\boldsymbol{X}_i\mid\boldsymbol{Z}=z)}\right)\\
&\geq\frac{1}{k}\sum_{\ell\in L'}\sum_{z\neq i(\ell)}\sum_{X\in\mathcal{X}_2}\Pr[\boldsymbol{X}=X\mid\boldsymbol{Z}=z]\Pr\left[\boldsymbol{\Pi}(X)=\ell\right]\mathsf{D}\left(\frac{\mu(\boldsymbol{X}_{i(\ell)}\mid\boldsymbol{\Pi}=\ell,\boldsymbol{Z}=z)}{\mu(\boldsymbol{X}_{i(\ell)}\mid\boldsymbol{Z}=z)}\right)\\
&\geq\frac{p\log k}{2}\sum_{\ell\in L'}\sum_{X\in\mathcal{X}_2}\Pr\left[\boldsymbol{X}=X\mid\boldsymbol{Z}\neq i(\ell)\right]\Pr\left[\boldsymbol{\Pi}(X)=\ell\right]. \qquad (9)
\end{aligned}
$$

Observe that for any $X\in\mathcal{X}_2$ and $i\in[k]$ we have $\Pr\left[\boldsymbol{X}=X\mid\boldsymbol{Z}\neq i\right]\geq\Pr\left[\boldsymbol{X}=X\right]/2$, because if $X_i=1$ the probability of X is increased, and if $X_i=0$ we are ruling out one of two symmetric ways to obtain X from μ. Also, recall that $\pi_2(L')\geq p_2$, that is,

$$
p_2\leq\Pr_{\boldsymbol{X}\sim\mu|_{\mathcal{X}_2}}\left[\boldsymbol{\Pi}(\boldsymbol{X})\in L'\right]=\frac{1}{\mu(\mathcal{X}_2)}\sum_{X\in\mathcal{X}_2}\mu(X)\left[\boldsymbol{\Pi}(\boldsymbol{X})\in L'\right]. \qquad (10)
$$

Therefore,

$$
\begin{aligned}
\mathrm{I}(\boldsymbol{\Pi};\boldsymbol{X}\mid\boldsymbol{Z}) &\overset{(9)}{\geq}\frac{p\log k}{2}\sum_{\ell\in L'}\sum_{X\in\mathcal{X}_2}\Pr\left[\boldsymbol{X}=X\mid\boldsymbol{Z}\neq i(\ell)\right]\Pr\left[\boldsymbol{\Pi}(X)=\ell\right] && (11)\\
&\geq\frac{p\log k}{4}\sum_{\ell\in L'}\sum_{X\in\mathcal{X}_2}\mu(X)\Pr\left[\boldsymbol{\Pi}(X)=\ell\right]\\
&=\frac{p\log k}{4}\sum_{X\in\mathcal{X}_2}\mu(X)\Pr\left[\boldsymbol{\Pi}(X)\in L'\right]\overset{(10)}{\geq}\frac{p\log k}{4}\cdot p_2\cdot\mu(\mathcal{X}_2). && (12)
\end{aligned}
$$

Finally, we have

$$
\mu(\mathcal{X}_2)=\mathsf{Bin}_{k-1,1/k}(1)=(k-1)\cdot\frac{1}{k}\left(1-\frac{1}{k}\right)^{k-2}\geq\frac{1}{2}\cdot\left(e^{-1/k}\right)^{k-2}\geq\frac{1}{2e^{(k-2)/k}}\geq\frac{1}{2e}.
$$

We used the fact that for $x\in(0,1/2)$ we have $1-x\geq e^{-x}$. Together with (12), we see that

$$
|\Pi|\geq\mathrm{I}(\boldsymbol{\Pi};\boldsymbol{X}\mid\boldsymbol{Z})\geq\Omega(\log k).
$$

This completes the proof.

Our lower bound of $\Omega(n\log k)$ is tight for the external information complexity of $\textsc{Disj}_{n,k}$: if players go over the coordinates one by one, and write their inputs on the board until they encounter a zero, then they learn at most $O(n\log k)$ bits of information (for each coordinate, the index of some player that got zero). Below we show that for communication complexity there is also an additive factor of k.

3.3 Lower Bound of $\Omega(k)$ on the Communication Complexity of $\textsc{And}_k$

We now show that computing $\textsc{And}_k$ requires $\Omega(k)$ bits of communication, which trivially implies a lower bound of $\Omega(k)$ on $\textsc{Disj}_{n,k}$. Note that, since $\textsc{And}_k$ requires only $O(\log k)$ bits of information to solve (by simply having the players write their inputs in order until we reach some player that received 0), this implies a gap of $\Omega(k/\log(k))$ between communication and information complexity.

Lemma 6. *The randomized communication complexity of* $\textsc{And}_k$ *is* $\Omega(k)$.

Proof. As usual, we prove the lower bound for deterministic protocols under random inputs, and obtain the lower bound for randomized protocols with worst-case input. Let $\epsilon < 1/3$ be the error bound, and fix $\epsilon' > \epsilon$ such that $\epsilon/(1-\epsilon') < 1/2$.

Consider the following input distribution μ: with probability $\epsilon' > \epsilon$, all players receive 1, and with probability $1-\epsilon'$, one random player receives 0 and the other players receive 1.

Fix a deterministic protocol Π for $\textsc{And}_k$. Let $p_1, \ldots, p_\ell$ be the order in which players speak when the input is 1^k, and assume for the sake of contradiction that $\ell < (1 - \epsilon/(1-\epsilon')) \cdot k$.

If $\Pi(1^k) = 0$, then Π's error under μ is $\epsilon' > \epsilon$, so we can assume that $\Pi(1^k) = 1$. Let $\mathcal{E}$ be the event that the input is not 1^k, but all of the players $p_1, \ldots, p_\ell$ receive 1. We have $\Pr_\mu[\mathcal{E}] = (1-\epsilon')\cdot(1-\ell/k) > \epsilon$. But when $\mathcal{E}$ occurs, the transcript of Π is identical to its transcript on 1^k, as Π is deterministic and all players that speak when the input is 1^k still receive 1 (in particular, the order of players that speak also remains the same under $\mathcal{E}$). Therefore with probability $> \epsilon$ the protocol outputs the wrong answer. □

4 Matching Upper Bound

We now describe a deterministic protocol for $\textsc{Disj}_{n,k}$ with communication complexity $O(n \log k + k)$, which, in light of the proof above, is optimal even for randomized algorithms.

High-level outline. In the protocol, players attempt to prove that $\bigcap_{i=1}^k X_i = \emptyset$ by writing on the board the indices of coordinates $i \in [n]$ where their input is zero; these coordinates cannot be in the intersection. If for each coordinate there is some player that can rule it out, then the intersection is empty. To reduce the amount of communication, players never write on the board a coordinate that already appears on the board. A naive implementation of this idea is to have players go in order $1, \ldots, k$, and have each player in its turn write on the board the coordinates where it got 0, and which do not already appear on the board. This leads to a communication complexity of $O(n \log n + k)$, as each coordinate requires $O(\log n)$ bits to be written on the board (the additive k term comes from players that need to indicate that they have no new zeroes to contribute). We can lower the communication complexity by "packing" the coordinates together, encoding sets of coordinates more compactly than the cost of writing them on the board one-by-one; for this to work, only players that have "many" new coordinates to write at once should contribute, and the other players should simply pass when it is their turn to speak. When we have reduced the number of coordinates that do not appear on the board to $\mathrm{poly}(k)$, we can afford to go back to the naive approach and simply have each player add all its zero coordinates to the board if they do not already appear there.

Detailed description. The protocol runs in *cycles*, where in each cycle some prefix of the players $1, \ldots, k$ each speak exactly once, in order, and the remaining players do not speak. Let Z_i be the set of coordinates that do not appear on the board at the beginning of cycle i, and let $z_i := |Z_i|$. Notice that if the input sets are indeed disjoint, then by the pigeonhole principle, at least one player has at least z_i/k zero coordinates that do not appear on the board ("new zeroes").

Suppose that at the beginning of cycle i we still have $z_i \geq k^2$. When it is the turn of player j to speak, if player j has at least z_i/k new zeroes, then it chooses z_i/k of them and writes them on the board, encoded as a subset of Z_i. The number of possible subsets is $\binom{z_i}{z_i/k} \leq ((z_i e)/(z_i/k))^{z_i/k} = (ek)^{z_i/k}$, so encoding one subset requires $(z_i/k)\log(ek)$ bits. If player j does not have z_i/k new zeroes to contribute, it writes a single bit on the board indicating this ("pass"). To simplify the analysis slightly, after the first player that does *not* pass, we immediately begin a new cycle (skipping over the remaining players in the cycle).

When at the beginning of some cycle i we have $z_i < k^2$, each player simply writes all its new zeroes on the board in the naive encoding (as elements of Z_i).

The protocol ends when one of the following occurs: if at any point all coordinates appear on the board, then the players halt and output "disjoint". Otherwise, if a complete cycle goes by in which all players pass, then the players halt and output "non-disjoint". Also, if we reach a cycle i with $z_i < k^2$, and at the end of the cycle not all coordinates appear on the board, then the players output "non-disjoint".

Correctness. If the players output "disjoint" then clearly the inputs *are* disjoint. For the other case, assume for the sake of contradiction that the players announce "not disjoint" at the end of cycle i, but the inputs are in fact disjoint. If cycle i has $z_i < k^2$ then the player only announce "not disjoint" if not all coordinates are written on the board at the end of the cycle. But in this cycle all players write their remaining zeroes on the board, so any coordinate that does not appear on the board at the end of the cycle is in the intersection, and the sets are not disjoint. Assume therefore that $z_i \geq k^2$, and the players output "not disjoint" because all players passed when it was their turn to speak in cycle i. Since the inputs are disjoint, for each coordinate $j \in Z_i$, some player received a zero in coordinate j. It follows that some player has at least z_i/k zeroes in coordinates from Z_i, and the smallest such player would not have passed when it was its turn to speak. This is a contradiction.

Communication complexity. In each cycle i with $z_i \geq k^2$, at most k players pass, and at most $(z_i/k)\log(ek)$ bits are added to the board to represent new zero coordinates. The total is $(z_i/k)\log(ek)+k$ bits, and since $z_i \geq k^2$, we have $(z_i/k)\log(ek) + k \leq (z_i/k)(\log(ek)+1)$. Also, $z_{i+1} = z_i - z_i/k = ((k-1)/k)z_i$, so in general we have $z_i = ((k-1)/k)^i n$ (with the first cycle being cycle 0). Therefore the total number of bits written on the board, while $z_i \geq k^2$, is bounded from above by

$$\sum_{i=0}^{\infty} \frac{\left(\frac{k-1}{k}\right)^i n}{k} \left(\log(ek)+1\right) = \frac{n}{k}\left(\log(ek)+1\right) \cdot \frac{1}{1-\frac{k-1}{k}} = n\left(\log(ek)+1\right).$$

If we reach a cycle i with $z_i < k^2$, there can be at most one such cycle (if we reach the end and not all coordinates appear on the board, the players output "non-disjoint"). The total number of bits written during this cycle is bounded by $z_i \cdot \log(z_i) + k \leq n\log(k^2) + k$. The total communication complexity of the protocol is $O(n\log k + k)$.

References

[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.

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 151–160. ACM, 2013.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, FOCS '11, pages 748–757, 2011.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC '83, pages 94–99, 1983.

[CP10] Arkadev Chattopadhyay and Toniann Pitassi. The story of set disjointness. *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.

[Gro09] André Gronemeier. Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the AND-function and disjointness. In *Proc. 26th Symp. on Theor. Aspects of Comp. Sc. (STACS)*, pages 505–516, 2009.

[HW07] Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(11):211–219, 2007.

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

[Pat11] Mihai Patrascu. Unifying the landscape of cell-probe lower bounds. *SIAM J. Comput.*, 40(3):827–847, 2011.

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

[She13] Alexander A. Sherstov. Communication lower bounds using directional derivatives. In *Proc. 45th Symp. on Theory of Comp. (STOC)*, pages 921–930, 2013.

[SHK$^+$12] Atish Das Sarma, Stephan Holzer, Liah Kor, Amos Korman, Danupon Nanongkai, Gopal Pandurangan, David Peleg, and Roger Wattenhofer. Distributed verification and hardness of distributed approximation. *SIAM J. Comput.*, 41(5):1235–1265, 2012.