2013 IEEE 54th Annual Symposium on Foundations of Computer Science

A Tight Bound for Set Disjointness in the Message-Passing Model

Mark Braverman*, Faith Ellen†, Rotem Oshman‡, Toniann Pitassi‡ and Vinod Vaikuntanathan§
*Computer Science Department, Princeton University. Email: mbraverm@cs.princeton.edu
†Computer Science Department, University of Toronto. Email: faith@cs.toronto.edu
‡Computer Science Department, University of Toronto. Email: rotem@cs.toronto.edu
§Computer Science Department, University of Toronto. Email: toni@cs.toronto.edu
¶Computer Science Department, University of Toronto. Email: vinodv@cs.toronto.edu

***Abstract*—In a multiparty message-passing model of communication, there are k players. Each player has a private input, and they communicate by sending messages to one another over private channels. While this model has been used extensively in distributed computing and in secure multiparty computation, lower bounds on communication complexity in this model and related models have been somewhat scarce. In recent work [25], [29], [30], strong lower bounds of the form $\Omega(n \cdot k)$ were obtained for several functions in the message-passing model; however, a lower bound on the classical set disjointness problem remained elusive.**

In this paper, we prove a tight lower bound of $\Omega(n \cdot k)$ for the set disjointness problem in the message passing model. Our bound is obtained by developing information complexity tools for the message-passing model and proving an information complexity lower bound for set disjointness.

I. Introduction

One of the most natural application domains for communication complexity is distributed computing: When we wish to study the cost of computing in a network spanning multiple cores or physical machines, it is very useful to understand how much communication is necessary, since communication between machines often dominates the cost of the computation. Accordingly, lower bounds in communication complexity have been used to obtain many negative results in distributed computing, from the round complexity of computing functions of distributed data [23], [18] to distributed computation and verification of network graph structures and properties [27], [12].

To the best of our knowledge, however, all applications of communication complexity lower bounds in distributed computing to date have used *only two-player lower bounds*. The reason for this appears to be twofold: First, the models of multi-party communication favored by the communication complexity community, the *number-on-forehead model* and the *number-in-hand broadcast model*, do not correspond to most natural models of distributed computing. Second, two-party lower bounds are surprisingly powerful, even for networks with many players. A typical reduction from a two-player communication complexity problem to a distributed problem T finds a *sparse cut* in the network, and shows that, to solve T, the two sides of the cut must implicitly solve, say, set disjointness [19]. However, there are problems that cannot be addressed by reduction from a two-player problem, because such reductions must reveal almost the entire structure of the network to one of the two players (see, e.g., [18].)

In this paper we study communication complexity in *message-passing models*, where each party has a private input, and the parties communicate by sending messages to each other over private channels. These models are used extensively in distributed computing, for example, to study gossiping protocols [16], to compute functions of distributed data [17], and to understand fundamental problems, such as achieving consensus in the presence of failures [11]. Message-passing models are also used to study privacy and security in multi-party computation.

We have chosen to focus on the set disjointness problem [7] because of its many applications in the two-player setting. In set disjointness, denoted $\textsc{Disj}_{n,k}$, k players each receive a set $X_i \subseteq [n]$, and their goal is to determine whether the intersection $\bigcap_{i=1}^{k} X_i$ is empty or not. An $\Omega(n)$ lower bound on the two-player version of set disjointness, due to Kalyanasundaram, Schnitger and Razborov [15], [26], is one of the most widely applied lower bounds in communication complexity. The lower bound was recently re-proven as an information complexity lower bound [2], showing that any protocol

Mark Braverman is supported in part by an Alfred P. Sloan Fellowship, an NSF CAREER award (CCF-1149888), and a Turing Centenary Fellowship. Faith Ellen, Rotem Oshman and Toniann Pitassi are supported in part by NSERC. Vinod Vaikuntanathan is supported in part by DARPA award FA8750-11-2-0225, a Connaught New Researcher Award and an Alfred P. Sloan Fellowship.

0272-5428/13 $26.00 © 2013 IEEE
DOI 10.1109/FOCS.2013.77

for two-party set disjointness must "leak" a total of $\Omega(n)$ bits about the input.

Our main result is a tight lower bound on the communication complexity of the set disjointness problem in a multiparty message-passing model, specifically the coordinator model of Dolev and Feder [9]. This lower bound implies a corresponding bound in the "truly distributed" message-passing model, where there is no coordinator. Our main technical tool in this paper is *information complexity*, which has its origins in the work of Chakrabarthi, Shi, Wirth and Yao [6], and which has recently played a pivotal role in several communication complexity lower bounds.

Our main theorem is an $\Omega(nk)$ lower bound on the information complexity (and hence also the communication complexity) of the set disjointness function in the multi-party coordinator model.

Theorem 1. *For every $\delta > 0$, $n \geq 1$ and $k = \Omega(\log n)$, there is a distribution ζ such that the information complexity of $\textsc{Disj}_{n,k}$ with error probability δ is $\Omega(nk)$ and its communication complexity is $\Omega(nk)$.*

The communication and information complexity of set disjointness: Variants of set disjointness are perhaps the most studied problems in communication complexity. In the two-party case, it is not hard to see that evaluating the disjointness of two subsets of $[n]$ *deterministically* requires at least $n+1$ bits of communication, for example, using a fooling set argument [20]. In the randomized model, when error is allowed, an $\Omega(n)$ lower bound is also known, although it is considerably more difficult to prove [15], [26]. This result was later improved using information-theoretic techniques by Bar-Yossef et al. [2]. Further advances in information complexity allow one to calculate the two-party communication complexity of disjointness precisely, up to additive $o(n)$ terms [5].

In the multi-party case, there are three main models to consider, all with interesting applications. The first model is the *number on forehead (NOF)* model, where each player is given all inputs except for one. The NOF model has important connections to circuit lower bounds for the $\mathbf{ACC^0}$ class [3]. Since the disjointness problem has small $\mathbf{AC^0}$ circuits, this means that for $k > \log n$, the communication complexity of NOF disjointness is polylogarithmic. The second model is the *number in hand blackboard model*. In this model each party is only given her input, and the communication is carried out via a blackboard, so each message transmitted by a player is received by all other players. In this case, the communication complexity of disjointness might be as high as $\Theta(n \log k)$ (note that an $\Omega(n)$ lower bound is trivial). Due to applications in streaming computation lower bounds, the version where the sets are either fully disjoint or have a single element in common has been studied. A lower bound of $\Omega(n/k)$ has been shown in this case using information-theoretic techniques [14].

Message-passing models: In this paper we consider *message-passing* models of communication complexity. In all of the multi-party models discussed so far, messages are written on a shared blackboard, so that the entire communication transcript is seen by all players. In message-passing models (also known as private channel models), the players communicate to one another by sending and receiving messages through private point-to-point channels. This type of model is one of the most widely-used in distributed computing (see, e.g., [21]) and in cryptography. Unlike shared-blackboard models, it is possible to achieve $\Omega(nk)$ lower bounds on problems in message-passing models [24].

There are many variants of message-passing models, differing in the topology of the communication network, the synchrony or asynchrony, and other parameters. Here we will focus on the *coordinator* message-passing model [9], where the players communicate with a coordinator by sending and receiving messages on private channels. We chose the coordinator model for two main reasons: first, it is technically easier to formalize than other models; and second, introducing a coordinator allows us to overcome obstacles related to the existence of information-theoretically secure multi-party computation protocols (more on this below).

Although the coordinator model does not capture a fully-decentralized system, it is closely related to the more decentralized message-passing model in which all players can communicate directly with each other [25]; the lower bound we prove in this paper implies a lower bound of $\Omega(nk/\log k)$ in the decentralized message-passing model. The coordinator model is also interesting in itself; it is appropriate for data centers or for sensor networks with centralized control, and there is a growing body of work on streaming and sketching algorithms set in the coordinator model [8], [22].

Communication complexity in message-passing models has received some attention recently. [25] introduces a technique called *symmetrization* for obtaining lower bounds of the form $\Omega(nk)$ via reduction to the two-party case. The symmetrization technique works for coordinate-wise problems such as Set Intersection, where the parties need to compute the intersection of their sets; this amounts to coordinate-wise AND on the players' inputs. However, symmetrization seems to fall short of yielding results for the multi-party set disjointness prob-

lem, and the development of new information-theoretic machinery seems necessary.

Another recent line of work dealing with communication complexity in the message-passing setting appears in [29], [30]. In these papers, the main interest is in *distributed streaming* or *distributed data aggregation*: each of k machines holds some data set or receives an input stream, and we wish to compute or approximate some function of the joint input, either through a central coordinator [29] or in a decentralized manner [30]. In [29], a lower bound of $\Omega(nk)$ is proven for the Gap-Majority(2-DISJ) problem: here the coordinator holds a set S, each player $i \in [k]$ holds a set T_i, and the goal is to distinguish the case where a "large majority" of the intersections $\{\text{DISJ}_{n,2}(S, T_i)\}_{i=1}^{k}$ are empty from the case where only a "small minority" are empty. In [30], similar techniques are used to obtain optimal lower bounds on a variety of problems in the decentralized message-passing model, including computing the number of distinct elements in the joint input, and checking various graph properties when the input is interpreted as a graph. Our set disjointness lower bound resolves a conjecture from [30]: it proves that computing the exact diameter of a graph with n edges requires $\Omega(nk)$ bits.

Connection to secure multiparty computation: Our results also have applications to showing lower bounds on the "amount of privacy" that one can achieve in the context of secure multiparty computation.

In the field of secure multiparty computation, the goal is for k players to communicate over a network to compute a joint function f on their inputs $x_1, \ldots, x_k$ while ensuring that no coalition of t players learn any information about the remaining players' inputs (other than what is already implied by their own inputs and outputs). In the 1980s, the work of Ben-Or, Goldwasser and Wigderson [4] showed multiparty protocols in the message-passing model for computing any function in an information-theoretically private way, assuming that the corruption threshold $t < k/2$. The BGW protocol is an important obstacle for information-complexity lower bounds in models with private channels, since it shows that sometimes computation is possible without leaking *any* information. We circumvent this obstacle by introducing a *coordinator* into the model, thus rendering any information-theoretically secure computation impossible. Essentially, instead of measuring only the amount of information each player *learns* (which may be zero), introducing a coordinator allows us to also measure the amount of information each player *leaks* to all the other players together, through its communication with the coordinator. We show that the average player either *learns* or *leaks* a lot of information.

Even with private channels between every pair of players, it is known that information-theoretic *perfect* privacy is impossible to achieve if $t \geq k/2$; that is, the adversary must learn some information about the honest players' inputs in this setting. An important question that remains is: *how much information* must the parties reveal about their inputs in order to compute a function f?

Recently, a number of works investigated this quantitative question *in the two-party setting* from the framework of information complexity [10], [1]. We believe that the information complexity tools developed here will lead to a better quantitative understanding of privacy in multiparty computation. For example, our information complexity lower bound already shows that in any k-party protocol for set disjointness there is a constant fraction of players i for which either (a) player i learns $\Omega(n)$ bits of information about the collective inputs of the players in $[k] \setminus \{i\}$, or (b) player i ends up revealing $\Omega(n)$ bits of information about its own input to the other players. We leave a more thorough investigation of this connection as future work.

Organization of the paper: The remainder of the paper is organized as follows. We begin by giving some intuition about our approach for obtaining an $\Omega(kn)$ lower bound on the communication complexity of set disjointness. In Section III, we present necessary definitions and facts about information theory, Hellinger distance, and information complexity. The next two sections present our lower bound, first proving that the information cost of solving $\text{DISJ}_{n,k}$ is at least n times the information cost of solving $\text{DISJ}_{1,k} = \text{AND}_k$, and then proving that it is at least $\Omega(k)$.

II. Overview: Why is Set Disjointness Hard?

Before diving into the technical details, let us explain the motivation behind our definition of information cost and the hard distribution we use in the lower bound.

Choosing the "right" notion of information complexity: There are several possible ways to quantify the amount of information leaked by a protocol that solves $\text{DISJ}_{n,k}$, which might at first glance seem natural:

(1) *External information cost*, $\text{I}(\mathbf{X}; \Pi(\mathbf{X}))$: how much information an external observer gains about the input X by observing the transcript of all the players and the coordinator. External information cost was used to prove the optimal $\Omega(n/k)$ lower bound on Promise set disjointness in the broadcast model [13].

The external information cost can also be viewed as the *coordinator's information cost*, because the coordinator observes the entire transcript and does not initially know any of the inputs.

(2) *The players' internal information cost*, $\sum_i \mathrm{I}(\mathbf{X}^{-i}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i)$: how much the players together learn about the input X from their interactions with the coordinator, given their private input.

Unfortunately, neither of these is high enough to yield an $\Omega(kn)$ lower bound on set disjointness. It is easy to see that the players' information cost is not always high: In the trivial protocol where all players send their inputs to the coordinator, the players learn nothing. Of course, in this protocol, the coordinator learns the entire input.

Likewise, the coordinator's information cost is not always high. To see why, consider the following protocol: For each coordinate j, the coordinator searches for the smallest index i such that $X_j^i = 0$, by contacting the players in order $i = 1, \ldots, k$ and asking them to send X_j^i. If $X_j^i = 0$ for some i, then $j \notin \bigcap_{i=1}^k X^i$, and the coordinator moves on to coordinate $j+1$ without asking the remaining players $\ell > i$ for X_j^ℓ. Otherwise, all players $i \in [k]$ have $X_j^i = 1$) and the coordinator halts with output "no", as $j \in \bigcap_{i=1}^k X^i$.

The transcript of the protocol can be losslessly compressed into $O(n \log k)$ bits by simply writing, for each coordinate j, the index of the first player i that has $X_j^i = 0$, or writing 0 if there is no such player. Therefore the coordinator cannot learn more than $O(n \log k)$ bits about the input by observing the transcript. On the other hand, in this protocol the players gain a significant amount of information: each player i from which the coordinator requests X_j^i learns that $X_j^\ell = 0$ for all $\ell < i$. For a *single* player i, this is not necessarily a lot of information; in fact, in the distribution we design below, it will correspond to roughly one bit of information. However, this one bit is learned by *many players*, and each player must learn it separately, because it is not privy to the coordinator's communication with the other players. We can charge each player separately for the information that it learns, even if this information overlaps with the information learned by the other players.

As we have seen, there is a protocol where the players learn nothing, but the coordinator learns a lot, and there is a protocol where the coordinator learns very little, but the players learn a lot. We will show that this tradeoff is inherent, by bounding from below the sum of the information learned by the coordinator about the players' inputs and the information learned by each player from the coordinator (about the inputs of the other players).

Designing a hard distribution: From the example above, we see that a hard distribution should make it hard for the coordinator to find the players that have zeroes, forcing it to communicate with $\Omega(k)$ players about each coordinate $j \in [n]$. This means that with reasonably large probability, in each coordinate j, only a few players i should have $X_j^i = 0$. On the other hand, our distribution should have *high entropy*, because the players can use Slepian-Wolf coding [28] to convey their joint input X to the coordinator using roughly $O(H(X))$ bits. To order to balance these two concerns, we follow [2], and use a *mixture of product distributions*.

Our hard distribution is a product $\eta = \xi^n$, where ξ is a hard distribution for a single coordinate $j \in [n]$. Informally, ξ has two "modes", selected by a "switch" $\mathbf{M}_j \in \{0, 1\}$:

- An "easy" mode, $\mathbf{M}_j = 0$, where each $\mathbf{X}_j^i = 0$ with probability $1/2$ independently.
- A "hard" mode, $\mathbf{M}_j = 1$, where there is exactly one player i with $\mathbf{X}_j^i = 0$, and the remaining players $\ell \neq i$ have $\mathbf{X}_j^\ell = 1$. The identity of the player that receives a zero is a random variable $\mathbf{Z} \in_{\mathsf{U}} [k]$.

More formally, for each $j \in [n]$, there is an independent distribution ξ over triples $(\mathbf{X}_j, \mathbf{M}_j, \mathbf{Z}_j)$, where $\mathbf{X}_j \in \{0,1\}^k$, $\mathbf{M}_j \in \{0,1\}$, and $\mathbf{Z}_j \in [k]$, such that the components $\mathbf{X}_j^1, \ldots, \mathbf{X}_j^k$ of $\mathbf{X}_j$ are independent given $\mathbf{M}_j$ and $\mathbf{Z}_j$. Each player i is given the input $\mathbf{X}_1^i, \ldots, \mathbf{X}_n^i$.

It may seem surprising that, under our distribution η, the answer to set disjointness is *almost always* "yes": The probability that we get some coordinate $j \in \bigcap_{i=1}^n \mathbf{X}^i$ is roughly $n/2^k$, which is negligible when n is significantly smaller than 2^k. This is necessary for our direct sum theorem. However, it might seem to make η an easy distribution, rather than a hard one. The key to η's hardness lies in the fact that the protocol must succeed with high probability on *any* input, even inputs that are very unlikely under η. This means that for hard coordinates, the protocol must "convince itself" that there really is some player that had a zero. This is hard because it is difficult to find such a player.

Ruling out Slepian-Wolf coding: As observed in [25] and as mentioned above, any lower bound for set disjointness (or other functions in the case of [25]) must implicitly rule out an approach where the players use Slepian-Wolf or other clever coding techniques to convey their inputs to the coordinator efficiently. Our lower bound does this quite explicitly.

Under the distribution $\eta = \xi^n$, we think of the players as jointly "owning" the input $\mathbf{X}$, because they are the only ones that initially know it. On the other hand, we think of the coordinator as "owning" the switches, $\mathbf{M} = \mathbf{M}_1, \ldots, \mathbf{M}_n$: the coordinator can easily determine if a given coordinate is "easy" or "hard" by sampling $O(\log n)$ players' inputs—if it finds no zeroes, it can conclude that the coordinate is "hard" with very high probability (in n). Since we are aiming for an $\Omega(nk)$

lower bound and the coordinator can determine $\mathbf{M}$ using $O(n \log n)$ bits, we may as well give this information to the coordinator for free.

Given that a coordinate j is hard, its entropy is only $1/k$. If the coordinator could convey the set of hard coordinates (or enough information about this set) to the players, they could then use Slepian-Wolf coding to send this part of the input to the coordinator in roughly $O(n)$ total bits (one bit per hard coordinate). However, the entropy of the set of hard coordinates is $n/2$, so conveying it (or sufficient information about it) to the players requires the coordinator to send $\Omega(n)$ bits to each player, for a total of $\Omega(nk)$ bits. In the absence of this information, the overall entropy of the input is $\Omega(nk)$, ruling out this type of approach.

We will formalize this intuition by showing that any protocol for set disjointness is "bad" in one (or both) of the following ways.

(1) The players convey to the coordinator "useless" information about their inputs: in the easy case when $\mathbf{M}_j = 0$, the coordinator learns $\Omega(k)$ bits about coordinate j, $\mathbf{X}_j^1, \ldots, \mathbf{X}_j^k$. This information is "useless" for the coordinator because when $\mathbf{M}_j = 0$ it can safely ignore coordinate j: with overwhelming high probability the sets do not intersect there.
One example of this approach is the naive protocol where players send their entire input to the coordinator.

(2) If the players do not convey to the coordinator a lot of information when $\mathbf{M} = 0$, then we will show that the coordinator conveys to the players "useless" information about the set of hard coordinates: $\Omega(k)$ players must learn whether coordinate j is easy (more formally, they learn $\Omega(1)$ bits of information about coordinate j) even when their input is $\mathbf{X}_j^i = 1$, i.e., they are not the special player that the coordinator is searching for.
An example of this approach is the protocol where the coordinator first samples a few inputs to determine which coordinates are hard, then sends the set of hard coordinates to all the players; each player responds by sending the coordinator a list of the hard coordinates where its input is zero.

In our lower bound proof, we explicitly bound from below the sum of the information costs described above.

III. Preliminaries

Notation: We use boldface letters to denote random variables, and capital letters to denote vectors or sets. For a set $A \subseteq [k]$, we let $\bar{e}_A$ denote the complement of A's characteristic vector; that is, $\bar{e}_A$ has 1 in exactly those coordinates that are not elements of A. For convenience we drop the curly brackets, so that, e.g., $\bar{e}_{i,j} = \bar{e}_{\{i,j\}}$.

If $X \in \{0,1\}^{k \cdot n}$ is a k-tuple of n-bit inputs, then $X^i \in \{0,1\}^n$ denotes the input to the i-th player, $X_j \in \{0,1\}^k$ denotes the vector comprising j-th coordinate of each player's input, and $X_j^i \in \{0,1\}$ denotes the j-th coordinate of X^i. For an n-tuple $Y \in \{0,1\}^n$, we use Y_{-i} to denote the tuple obtained from Y by dropping the i-th coordinate (that is, $Y_{-i} = Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n$). We also let $Y_{[i,j]} := Y_i, \ldots, Y_j$. Finally, $\mathrm{embed}(X, i, x)$ denotes the vector obtained from X by inserting x in coordinate i: $\mathrm{embed}(X, i, x) = (X_1, \ldots, X_{i-1}, x, X_i, \ldots, X_m)$, where $m = |X|$.

The coordinator model: As mentioned in the Introduction, we will work in the asynchronous *coordinator* message-passing model introduced in [9]. In this model, there is one additional participant, called the coordinator, who receives no input. There is a private channel between every player and the coordinator, but the players cannot communicate directly with one another. For lack of space, we omit the formal description of the model.

For any protocol Π and any input $X \in \{0,1\}^{k \cdot n}$, we let $\Pi(X)$ denote the distribution of Π's transcript (as seen by the coordinator) when run with input X, and, for each player $i \in [k]$, we let $\Pi^i(X)$ denote the transcript of messages sent between player i and the coordinator (in both directions).

Communication complexity: Let Π be a protocol for solving a problem $\mathcal{P}$. The *error* of Π is given by

$$\max_X \Pr\left[\text{the coordinator outputs an incorrect answer}\right],$$

where the probability is taken over the private randomness of the coordinator and the players.

The *communication complexity* of a protocol Π is the worst-case number of bits exchanged between the players and the coordinator on any input. The *δ-error randomized communication complexity* of a problem $\mathcal{P}$ in the coordinator model, denoted $\mathrm{CC}_\delta(\mathcal{P})$, is the minimum communication complexity of any randomized protocol Π that solves $\mathcal{P}$ with error at most δ.

Useful classes of distributions: Our hard distribution for set disjointness uses an auxiliary "switch" $\mathbf{M}$, which determines if a coordinate is hard or easy, and another auxiliary variable $\mathbf{Z}$, which selects the player that receives zero in the hard case. Conditioned on $\mathbf{M}$ and $\mathbf{Z}$, the players' inputs are independent from each other. The value of $\mathbf{M}$ is assumed known to the coordinator, but the value of $\mathbf{Z}$ is hidden from all participants.

The following definition captures distributions that behave like our hard distribution. It is a special case of a mixture of product distributions [2].

Definition 1 (Switched distributions)**.** *We say that the joint distribution η of $(\mathbf{X}, \mathbf{M}, \mathbf{Z})$ is* switched by $\mathbf{M}$ and $\mathbf{Z}$ *if $\mathbf{X}^1, \ldots, \mathbf{X}^k$ are conditionally independent given $\mathbf{M}$ and $\mathbf{Z}$, and $\mathbf{M}$ is independent from $\mathbf{Z}$.*

Our hard distribution for a single coordinate also has the property that with very high probability, it produces a set disjointness instance on which the answer is "yes". This is important for our direct sum reduction. Adapting the definition of a *collapsing distribution* from [2], we capture this notion as follows. (The following definition is specifically for 1-bit AND; it is easy to generalize to arbitrary functions along the same lines as [2].)

Definition 2 (ϵ-collapsing distributions)**.** *A distribution $\mu : \{0,1\}^k \to [0,1]$ is* ϵ-collapsing for AND *if*

$$\Pr_{\mathbf{X}\sim\mu}\left[\bigwedge_{i=1}^{k} \mathbf{X}_k = 1\right] \leq \epsilon.$$

Information theory and Hellinger distance: Let μ be a distribution on a finite set D and let X, Y, Z be random variables. The *entropy* of X is defined by

$$H(X) = \sum_{\omega\in D} \mu(\omega) \log \frac{1}{\mu(\omega)}$$

The *conditional entropy* of X given Y is

$$H(X|Y) = \sum_{y} H(X|Y = y) Pr[Y = y],$$

where $H(X|Y = y)$ is the entropy of the conditional distribution of X given the event $\{Y = y\}$.
The *joint entropy* of X and Y is the entropy of their joint distribution and is denoted by $H(X, Y)$.
The *mutual information* between X and Y is

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

The *conditional mutual information* between X and Y conditioned on X is

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z).$$

The *Hellinger distance* between probability distributions P and Q on a domain $\mathcal{D}$ is defined by

$$h(P,Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{\omega\in\mathcal{D}} |\sqrt{P(\omega)} - \sqrt{Q(\omega)}|^2}.$$

Hellinger distance is a metric and, in particular, it satisfies the triangle inequality. Another useful property of the Hellinger distance is the following:

Lemma 2 ([2])**.** *Let $\mathcal{P}$ be a problem, and let Π be a δ-error protocol for $\mathcal{P}$. If X and Y are inputs such that $\mathcal{P}(X) \neq \mathcal{P}(Y)$, then $h(\Pi(X), \Pi(Y)) \geq (1-\delta)/\sqrt{2}$.*

Essentially, the lemma asserts that since the protocol must distinguish between the two inputs X and Y, the Hellinger distance of the respective distributions on the transcript must be large.

The following lemma from [2] addresses the converse direction—it shows that our ability to distinguish between samples from two distributions grows with their Hellinger distance.

Lemma 3 ([2])**.** *Let μ_0, μ_1 be two distributions. Suppose that $\mathbf{Y}$ is generated as follows: we first select $\mathbf{S} \in_{\mathsf{U}} \{0,1\}$, and then sample $\mathbf{Y}$ from $\mu_{\mathbf{S}}$. Then $\mathrm{I}(\mathbf{S};\mathbf{Y}) \geq h^2(\mu_0, \mu_1)$.*

Information cost: In general, we define the *internal information cost* of a protocol as follows.

Definition 3. *Let $\mathbf{X} \sim \zeta$ be a distribution. The* internal information cost *of a protocol Π with k parties communicating through a coordinator with respect to ζ is given by*

$$\underset{\zeta}{\mathrm{IC}}(\Pi) := \underset{\mathbf{X}\sim\zeta}{\mathrm{I}}(\mathbf{X}; \Pi(\mathbf{X})) + \sum_{i\in[k]} \left[\underset{\mathbf{X}\sim\zeta}{\mathrm{I}}(\mathbf{X}^{-i}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i)\right].$$

If $\mathcal{P}$ is a problem (formally, a Boolean predicate on $k \times n$-bit inputs and outputs from some domain), then we define the information complexity of $\mathcal{P}$ as

$$\underset{\zeta,\delta}{\mathrm{IC}}(\mathcal{P}) = \inf_{\Pi} \underset{\zeta}{\mathrm{IC}}(\Pi)$$

where the infimum is taken over all δ-error randomized protocols for $\mathcal{P}$.

This is a general definition which does not depend on the structure of the distribution ζ. However, our lower bound uses a switched distribution, and as we explained in Section II, we give a bound on the following, more fine-grained expression:

Definition 4. *Let $(\mathbf{X}, \mathbf{M}, \mathbf{Z}) \sim \eta$ be a distribution switched by $\mathbf{M}$ and $\mathbf{Z}$. The* switched information cost *of a protocol Π with respect to μ is given by*

$$\underset{\eta}{\mathrm{SIC}}(\Pi) := \sum_{i\in[k]} \left[\underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}}(\mathbf{X}^i; \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}) + \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}}(\mathbf{M}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{Z})\right].$$

The switched information cost *of a problem $\mathcal{P}$ is defined analogously.*

The two notions of information cost are closely related; indeed, our lower bound of $\Omega(nk)$ on

$\mathrm{SIC}_\eta(\textsc{Disj}_{n,k})$ also implies a lower bound of $\Omega(nk)$ on $\mathrm{IC}_\eta(\textsc{Disj}_{n,k})$ (the details are omitted here).

To obtain a lower bound on the communication cost of a problem $\mathcal{P}$, it is sufficient to give a lower bound on its internal information cost (or similarly, on its switched information cost):

Lemma 4. *For any problem $\mathcal{P}$, $\mathrm{CC}_\delta(\mathcal{P}) \geq 1/2 \cdot \mathrm{IC}_{\zeta,\delta}(\mathcal{P})$.*

Proof: For any δ-error protool Π,

$$\begin{aligned}\underset{\zeta}{\mathrm{IC}}(\Pi) &= \underset{\mathbf{X}\sim\zeta}{\mathrm{I}}(\mathbf{X};\Pi(\mathbf{X})) \\ &\qquad + \sum_{i\in[k]} \left[\underset{\mathbf{X}\sim\zeta}{\mathrm{I}}(\mathbf{X}^{-i};\Pi^i(\mathbf{X}) \mid \mathbf{X}^i)\right] \\ &\leq H(\Pi) + \sum_{i\in[k]} H(\Pi^i \mid \mathbf{X}^i) \\ &\leq H(\Pi) + \sum_{i\in[k]} H(\Pi^i) \leq |\Pi| + \sum_{i\in[k]} |\Pi^i| = 2|\Pi|.\end{aligned}$$

The claim follows. ∎

In Sections IV and V we show that the switched information cost of $\textsc{Disj}_{n,k}$ under our hard distribution is $\Omega(nk)$. This implies a lower bound of $\Omega(nk)$ on the communication complexity of $\textsc{Disj}_{n,k}$.

IV. Direct Sum Theorem

We begin by proving that the information cost of computing the set disjointness function

$$\textsc{Disj}_{n,k}(\mathbf{X}^1,\ldots,\mathbf{X}^k) = \bigvee_{j=1}^{n} \bigwedge_{i=1}^{k} \mathbf{X}^i_j$$

is as least n times the cost of solving the one-bit problem $\textsc{And}_k = \bigwedge_{i=1}^k \mathbf{X}^i_j$. The proof is by reduction: given a protocol Π for $\textsc{Disj}_{n,k}$ and a switched distribution $\eta = \xi^n$, where ξ itself is a switched and ϵ-collapsing distribution, we will construct a protocol $\hat{\Pi}$ for $\textsc{And}_k$, such that $\mathrm{SIC}_\xi(\hat{\Pi}) \leq (1/n)\,\mathrm{SIC}_\eta(\Pi)$.

The one-bit protocol $\hat{\Pi}$ uses Π by constructing an n-bit input, running Π on it, and returning Π's answer. However, the input to $\hat{\Pi}$ is only a single bit per player. To construct an n-bit input, the coordinator first selects a random coordinate $\mathbf{j} \in_{\mathsf{U}} [n]$, into which the one-bit input to $\hat{\Pi}$ will be embedded. Next we wish to randomly sample the other coordinates $[n]\backslash\{\mathbf{j}\}$ from ξ^{n-1}, in order to obtain an n-bit input on which we can run Π. We must do this carefully: we need $\hat{\Pi}$ to have an information cost proportionate to the information cost of Π, but we do not know where Π incurs the majority of its information cost—does the coordinator learn a lot about the inputs given the switch $\mathbf{M}$, or do the players learn a lot about the switch $\mathbf{M}$ given their inputs? One of these terms may be *small*, and we must ensure that $\hat{\Pi}$'s corresponding cost in the same term is also small.

I. If in Π the coordinator does not learn much about the input given $\mathbf{M}$ and $\mathbf{Z}$, then our new protocol $\hat{\Pi}$ should also not reveal too much about the input to the coordinator. A good solution is to have the coordinator sample $\mathbf{M}_{-\mathbf{j}}$ and $\mathbf{Z}_{-\mathbf{j}}$ and send them to the players, who can then sample their inputs independently using their private randomness.

II. If in Π the players do not learn much about $\mathbf{M}$ given their inputs and $\mathbf{Z}$, then we should not reveal $\mathbf{M}$ to the players in $\hat{\Pi}$. A good solution is to have the coordinator sample $\mathbf{M}_{-\mathbf{j}}, \mathbf{Z}_{-\mathbf{j}}$ and $\mathbf{X}_{-\mathbf{j}}$, and send to each player i its input $\mathbf{X}^i_{-\mathbf{j}}$. Thus the players do not know $\mathbf{M}$ before they execute Π (except what they can deduce from their inputs).

Since we do not know in advance how Π behaves on the average coordinate, our solution is to "hedge our bets" by using the first approach to sample the coordinates below $\mathbf{j}$, and the second approach to sample the coordinates above $\mathbf{j}$. More formally, on one-bit input $(\mathbf{U},\mathbf{N},\mathbf{S}) \sim \xi$, protocol $\hat{\Pi}$ works as follows:

1) The coordinator samples a random coordinate $\mathbf{j} \in_{\mathsf{U}} [n]$ and samples $\mathbf{Z}_{-\mathbf{j}} \in_{\mathsf{U}} [k]^{n-1}$, and sends them to all players.
2) For each $\ell < \mathbf{j}$, the coordinator samples $\mathbf{M}_\ell$ and sends it to all players. Each player i then samples $\mathbf{X}^i_\ell$ from its marginal distribution given $\mathbf{M}_\ell$ and $\mathbf{Z}_\ell$.
3) For each $\ell > \mathbf{j}$, the coordinator samples $\mathbf{X}_\ell, \mathbf{M}_\ell$ from their marginal distribution given $\mathbf{Z}_\ell$, and sends to each player i its input $\mathbf{X}^i_\ell$.
4) The participants simulate the execution of Π using the joint input

$$\begin{aligned}\mathrm{embed}(\mathbf{X},\mathbf{j},\mathbf{U}) = \big\{\big(&\mathbf{X}^i_1,\ldots,\mathbf{X}^i_{\mathbf{j}-1}, \\ &\mathbf{U}^i,\mathbf{X}^i_{\mathbf{j}+1},\ldots,\mathbf{X}^i_n\big)\big\}_{i=1}^k.\end{aligned}$$

5) The coordinator outputs the value output by Π.

The last step is the reason we require ξ to be ϵ-collapsing: for each coordinate $\ell \neq \mathbf{j}$, with probability at least $1-\epsilon$ we have $\bigwedge_{i=1}^k \mathbf{X}^i_\ell = 0$. By union bound, the probability that $\bigvee_{\ell\neq\mathbf{j}} \bigwedge_{i=1}^k \mathbf{X}^i_\ell = 0$ is at least $1-(n-1)\epsilon$. Whenever this occurs we have $\textsc{Disj}_{n,k}(\mathrm{embed}(\mathbf{X},\mathbf{j},\mathbf{U})) = \textsc{And}_k(\mathbf{U})$, that is, if Π succeeds then $\hat{\Pi}$ succeeds as well. Therefore the error probability of $\hat{\Pi}$ is at most $n\epsilon+\delta$, where δ is the error probability of Π.

The following lemma relates the information cost of $\hat{\Pi}$ to that of Π:

Lemma 5. *For each player $i \in [k]$ we have*

$$\underset{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi}{\mathrm{I}}\left(\mathbf{N}; \hat{\Pi}^i(\mathbf{U}) \mid \mathbf{U}^i, \mathbf{S}\right) \leq \frac{1}{n}\left[\underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}}(\mathbf{M}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{Z})\right]$$

and

$$\underset{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi}{\mathrm{I}}\left(\mathbf{U}^i; \hat{\Pi}^i(\mathbf{U}) \mid \mathbf{N}, \mathbf{S}\right) \leq \frac{1}{n}\left[\underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}}(\mathbf{X}^i; \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z})\right].$$

The direct sum theorem follows immediately:

Theorem 6. *Let ξ be an ϵ-collapsing distribution switched by $\mathbf{M}$ and $\mathbf{Z}$, where $\epsilon < (1-\delta)/n$, and let $\eta = \xi^n$. Then*

$$\underset{\eta,\delta}{\mathrm{SIC}}(\textsc{Disj}_{n,k}) \geq n \cdot \underset{\xi,\delta+n\epsilon}{\mathrm{SIC}}(\textsc{And}_k).$$

V. The Information Complexity of AND

By Theorem 6, in order to obtain an $\Omega(nk)$ lower bound on $\textsc{Disj}_{n,k}$ it is sufficient to show a lower bound of $\Omega(k)$ on the information complexity of $\textsc{And}_k$ under a hard one-bit distribution ξ, which is both switched and ϵ-collapsing. We will use the following distribution on $(\mathbf{X}, \mathbf{M}, \mathbf{Z})$ (informally described in Section II):

- First we select $\mathbf{Z} \in_{\mathsf{U}} [k]$ and, independently, the mode $\mathbf{M}$ is selected with $\Pr[\mathbf{M} = 0] = 2/3$ and $\Pr[\mathbf{M} = 1] = 1/3$.
- If $\mathbf{M} = 0$, then each player's input $\mathbf{X}^i$ is 0 or 1 with equal probability, independent of the other inputs. If $\mathbf{M} = 1$, then the joint input is $\bar{e}_{\mathbf{Z}} := 1^{\mathbf{Z}-1}01^{k-\mathbf{Z}}$.

The distribution is switched by $\mathbf{M}$ and $\mathbf{Z}$, and is ϵ-collapsing with $\epsilon = 1/(3 \cdot 2^{k-1})$.

Notation: In this section we let $\Pi(X)$ denote the distribution of the protocol's transcript when executed on input $X \in \{0,1\}^k$, and $\Pi^i(X)$ denote the distribution of player i's view of the transcript. We also let $\Pi^i[x, m, z]$ denote the distribution of player i's view when the input is drawn from ξ, conditioned on $\mathbf{X}^i = x, \mathbf{M} = m$ and $\mathbf{Z} = z$. For example, if $j \neq i$, then $\Pi^i[1, 1, j] = \Pi(\bar{e}_j)$. Notice that $\Pi^i[0, 1, j]$ for $i \neq j$ is not well-defined, because $\Pr\left[\mathbf{X}^i = 0, \mathbf{M} = 1, \mathbf{Z} \neq i\right] = 0$. Similarly, we let $\Pi[i, x, m, z]$ denote the distribution of Π's transcript, conditioned on $\mathbf{X}^i = x, \mathbf{M} = m$ and $\mathbf{Z} = z$.

Structural properties of protocols in the coordinator model: We prove that $\mathrm{SIC}_{\xi,\delta}(\textsc{And}_k) = \Omega(k)$ in several steps. The distribution ξ comes in only when we relate Hellinger distance to mutual information; for the most part we rely on the fact that Π has error at most δ on any input, and on the structural properties of Π. We begin by outlining these properties. For lack of space, the proofs of these properties are omitted here.

The first property we will use is a simplified version of the Z-Lemma (or Pythagorean Lemma) of [2], extended the coordinator model.

Lemma 7 (Diagonal Lemma)**.** *For any X, Y and $\ell \in [k]$ we have*

$$h^2(\Pi(X), \Pi(Y))) \geq \frac{1}{2}h^2(\Pi(X), \Pi(\mathrm{embed}(Y^{-\ell}, \ell, X^\ell))).$$

Under our distribution ξ, the inputs $\mathbf{X}^i$ are independent given $\mathbf{M}$ and $\mathbf{Z}$. This allows us to prove the following variant of the Diagonal Lemma, which considers player i's view and "abstracts away" all the inputs $\mathbf{X}^{-i}$ by grouping them together under the conditioning $\mathbf{M} = m, \mathbf{Z} = z$ (for some m and z).

Lemma 8 (Diagonal Lemma for $\mathbf{M}$ and $\mathbf{X}^i$)**.** *For any $i \neq z$ we have*

$$h^2(\Pi^i[0, 0, z], \Pi^i[1, 1, z]) \geq \frac{1}{2}h^2(\Pi^i(\bar{e}_{i,z}), \Pi^i(\bar{e}_z)).$$

Note that Lemma 7 concerns the complete transcript Π, while Lemma 8 concerns one player's local view, Π^i. To move between the two we use the following "localization" lemma, which shows that when we "keep everything the same" and change only $\mathbf{X}^i$, the distance between the transcript's distributions is caused entirely by player i's local view. We are interested in two cases: one where we fix $\mathbf{M} = 0$ and $\mathbf{Z} = z \neq i$, and let $\mathbf{X}^i$ change from 0 to 1; and the other where $i \neq z$, players $[k] \setminus \{i, z\}$ receive 1, player z receives 0, and we vary player i's input, yielding the two inputs $\bar{e}_{i,z}$ and $\bar{e}_z$.

Lemma 9 (Localizing the distance to a single player's transcript)**.** *For any $i \neq z$ we have*

$$h(\Pi[i, 0, 0, z], \Pi[i, 1, 0, z]) = h(\Pi^i[0, 0, z], \Pi^i[1, 0, z]),$$

and similarly,

$$h(\Pi(\bar{e}_{i,z}), \Pi(\bar{e}_z)) = h(\Pi^i(\bar{e}_{i,z}), \Pi^i(\bar{e}_z)).$$

Now we are ready to describe the main proof that the information complexity of $\textsc{And}_k$ is $\Omega(k)$.

Step I. Setting up a rectangle: Fix a player i and a value $z \neq i$, and consider a rectangle the following four distributions:

$$\Pi^i[0,0,z] \qquad \Pi^i[1,0,z]$$

$$\Pi^i(\bar{e}_{i,z}) \qquad \Pi^i(\bar{e}_z) = \Pi^i[1,1,z]$$

The two distributions in the top row differ only in the value of $\mathbf{X}^i$, which is 0 for the first column and 1 for the second; the same holds for the bottom row. The top-row distributions have $\mathbf{M} = 0$, and it is helpful to think of the bottom row as representing the hard case, $\mathbf{M} = 1$ (although $\Pi^i[0,1,z]$ is not well-defined, and moreover, the input $\bar{e}_{i,z}$ has probability 0 under ξ).

Notice that our distribution ξ satisfies:

$$\Pr\left[\mathbf{X}^i = 0 \mid \mathbf{M} = 0, \mathbf{Z} = z\right] = \Pr\left[\mathbf{X}^i = 1 \mid \mathbf{M} = 0, \mathbf{Z} = z\right] = 1/2,$$

and

$$\Pr\left[\mathbf{M} = 0 \mid \mathbf{X}^i = 1, \mathbf{Z} = z\right] = \Pr\left[\mathbf{M} = 1 \mid \mathbf{X}^i = 1, \mathbf{Z} = z\right] = 1/2.$$

In other words, given that we are in the top row of the rectangle ($\mathbf{M} = 0, \mathbf{Z} = z$), the distribution of the transcript Π^i is equally likely to be $\Pi^i[0,0,z]$ or $\Pi^i[1,0,z]$, the two top-row distributions. This means that *if the two top-row distributions have a large Hellinger distance*, then the conditional mutual information $\mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M} = 0, \mathbf{Z} = z)$ is large: although $\mathbf{X}^i$ is equally likely to be 0 or 1 *a priori* given $\mathbf{M} = 0, \mathbf{Z} = z$, because of the large Hellinger distance, the transcript Π^i allows us to distinguish the case $\mathbf{X}^i = 0$ from the case $\mathbf{X}^i = 1$. This is captured by Lemma 3, which yields

$$\mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M} = 0, \mathbf{Z} = z) \geq h(\Pi^i[0,0,z], \Pi^i[1,0,z]).$$

Similarly, given that we are in the rightmost column ($\mathbf{X}^i = 1, \mathbf{Z} = z$), the distribution of Π^i is equally likely to be $\Pi^i[1,0,z]$ or $\Pi^i[1,1,z]$. Therefore a large Hellinger distance between these distributions implies that $\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i = 1, \mathbf{Z} = z)$ is large: Lemma 3 again yields

$$\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i = 1, \mathbf{Z} = z) \geq h(\Pi^i[1,0,z], \Pi^i[1,1,z]).$$

Recall that $\Pr[\mathbf{M} = 0 \mid \mathbf{Z} = z] = 2/3$ (as $\mathbf{M}$ and $\mathbf{Z}$ are independent), and observe that when $z \neq i$ we have $\Pr\left[\mathbf{X}^i = 1 \mid \mathbf{Z} = z\right] = 2/3$. Therefore,

$$\mathrm{I}\left(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z} = z\right) \geq \frac{2}{3}\,\mathrm{I}\left(\mathbf{X}^i; \Pi^i \mid \mathbf{M} = 0, \mathbf{Z} = z\right)$$

and

$$\mathrm{I}\left(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = z\right) \geq \frac{2}{3}\,\mathrm{I}\left(\mathbf{M}; \Pi^i \mid \mathbf{X}^i = 1, \mathbf{Z} = z\right).$$

It follows that

$$\begin{aligned}
&\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = z) + \mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z} = z) \\
&\quad \geq \frac{2}{3}\left(h^2(\Pi^i[1,0,z], \Pi^i[1,1,z]) \right. \\
&\qquad \left. + h^2(\Pi^i[0,0,z], \Pi^i[1,0,z])\right) \\
&\quad \geq \frac{1}{3}\left(h(\Pi^i[1,0,z], \Pi^i[1,1,z]) \right. \\
&\qquad \left. + h(\Pi^i[0,0,z], \Pi^i[1,0,z])\right)^2 \\
&\quad \geq \frac{h^2(\Pi^i[0,0,z], \Pi^i[1,1,z])}{3}.
\end{aligned}$$

The last step uses the triangle inequality. Now we apply Lemma 8, which together with the above yields

$$\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = z) + \mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z} = z) \geq \frac{1}{3}h^2\left(\Pi^i(\bar{e}_{i,z}), \Pi^i(\bar{e}_z)\right). \quad (1)$$

This holds only for $z \neq i$. Taking the expectation over *all* $z \in [k]$, we obtain

$$\begin{aligned}
&\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z}) + \mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z}) \\
&\geq \frac{1}{k}\sum_{z \neq i}\left(\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = z) + \mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z} = z)\right) \\
&\overset{(1)}{\geq} \frac{k-1}{3k}\mathop{\mathbb{E}}_{\mathbf{Z} \neq i}\left[h^2(\Pi^i(\bar{e}_{i,\mathbf{Z}}, \Pi^i(\bar{e}_{\mathbf{Z}}))\right] \\
&\geq \frac{1}{6}\mathop{\mathbb{E}}_{\mathbf{Z} \neq i}\left[h^2(\Pi^i(\bar{e}_{i,\mathbf{Z}}, \Pi^i(\bar{e}_{\mathbf{Z}}))\right]. \qquad (2)
\end{aligned}$$

The last step uses the fact that $k - 1 \geq k/2$, as $k > 1$.

Let us define the *usefulness of player* i to be $\gamma_i := \mathbb{E}_{\mathbf{Z} \neq i}\left[h^2(\Pi^i(\bar{e}_{i,\mathbf{Z}}, \Pi^i(\bar{e}_{\mathbf{Z}}))\right]$. Roughly speaking, player i's usefulness corresponds to how sensitive the protocol is to the fact that $\mathbf{X}^i = 0$, *when some other player* $z \neq i$ *also has 0*. By (2) we see that in order to obtain our desired $\Omega(k)$ lower bound, it is sufficient to bound the sum $\sum_i \gamma_i$ (or the average, $\sum_i \gamma_i/k$). But why should γ_i be large on average? In other words, why should the protocol distinguish the case where only one player has zero from the case where two players have zero, when the answer to $\textsc{And}_k$ is 0 in both cases? This will again follow from the structural properties of the protocol.

A. Step II: bounding the average usefulness.

In order to show that the average player has a large usefulness γ_i, consider any two players $i \neq j$, and the rectangle consisting of the following four distributions:

$$\begin{array}{ll} \Pi(\bar{e}_i) & \Pi(1^k) \\ \Pi(\bar{e}_{i,j}) & \Pi(\bar{e}_j) \end{array}$$

We have $\textsc{And}_k(\bar{e}_i) = \textsc{And}_k(\bar{e}_j) = 0$, but $\textsc{And}_k(1^k) = 1$. By the correctness of the protocol and Lemma 2, the

statistical distance between $\Pi(\bar{e}_i)$ and $\Pi(1^k)$ must be at least $1-\delta$, which implies that $h(\Pi(\bar{e}_i),\Pi(1^k)) \geq (1-\delta)/\sqrt{2}$. By the diagonal lemma (with $\ell = j$), $h(\Pi(\bar{e}_i),\Pi(\bar{e}_j)) \geq h(\Pi(\bar{e}_i),\Pi(1^k))/\sqrt{2} \geq (1-\delta)/2$, that is, the protocol must distinguish $\bar{e}_i$ from $\bar{e}_j$. (Roughly speaking, this means that the protocol must *find* a player that has zero in the case where $\mathbf{M}=1$, an interesting fact in itself.) By the triangle inequality,

$$h(\Pi(\bar{e}_i),\Pi(\bar{e}_{i,j})) + h(\Pi(\bar{e}_j),\Pi(\bar{e}_{i,j})) \geq h(\Pi(\bar{e}_i),\Pi(\bar{e}_j)) \geq (1-\delta)/2,$$

and therefore

$$\begin{aligned} h^2(\Pi(\bar{e}_i),\Pi(\bar{e}_{i,j})) + h^2(\Pi(\bar{e}_j),\Pi(\bar{e}_{i,j})) &\geq \frac{\left(h(\Pi(\bar{e}_i),\Pi(\bar{e}_{i,j})) + h(\Pi(\bar{e}_j),\Pi(\bar{e}_{i,j}))\right)^2}{2} \\ &\geq \frac{(1-\delta)^2}{8}. \end{aligned}$$

Now summing across all pairs of players $i \neq j$, we see that $2\sum_i \sum_{j\neq i} h^2(\Pi(\bar{e}_i),\Pi(\bar{e}_{i,j})) \geq k(k-1)\cdot(1-\delta)^2/8$, which implies that $\sum_i \gamma_i \geq k\cdot(1-\delta)^2/16 = \Omega(k)$. Together with (2), this yields our main result for this section:

Theorem 10. *For any $k > 1$, $\mathrm{SIC}_{\xi,\delta}(\mathrm{AND}_k) \geq (1-\delta)^2/96$.*

Combining Theorem 10 with our direct-sum theorem from Section IV, we obtain

Theorem 11. *For any $n \geq 1$ and for $k = \Omega(\log n)$, $\mathrm{SIC}_{\eta,\delta}(\mathrm{DISJ}_{n,k}) = \Omega(nk)$.*

Theorem 11 implies a lower bound of $\Omega(nk)$ on the communication complexity of $\mathrm{DISJ}_{n,k}$.

REFERENCES

[1] Anil Ada, Arkadev Chattopadhyay, Stephen A. Cook, Lila Fontes, Michal Koucký, and Toniann Pitassi. The hardness of being private. In *CCC'12*, pages 192–202.

[2] 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.

[3] Richard Beigel and Jun Tarui. On acc. *Computational Complexity*, 4:350–366, 1994.

[4] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). In *STOC'88*, pages 1–10.

[5] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 19(171), 2012.

[6] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS'01*, pages 270–278.

[7] A. Chattopadhyay and T. Pitassi. The story of set disjointness. In *SIGACT News Complexity Theory Column 67*, 2011.

[8] Graham Cormode and Minos Garofalakis. Sketching streams through the net: distributed approximate query tracking. In *VLDB '05*, pages 13–24, 2005.

[9] Danny Dolev and Tomas Feder. Determinism vs. nondeterminism in multiparty communication complexity. *SIAM Journal on Computing*, 21(5):889–895, 1992.

[10] Joan Feigenbaum, Aaron D. Jaggard, and Michael Schapira. Approximate privacy: foundations and quantification (extended abstract). In *EC'10*.

[11] Michael J. Fischer, Nancy A. Lynch, and Michael S. Paterson. Impossibility of distributed consensus with one faulty process. *J. ACM*, 32(2):374–382, April 1985.

[12] Silvio Frischknecht, Stephan Holzer, and Roger Wattenhofer. Networks cannot compute their diameter in sublinear time. In *SODA '12*, pages 1150–1162.

[13] Andre Gronemeier. Asymptotically optimal lower bounds on the nih-multi-party information. *arXiv preprint arXiv:0902.1609*, 2009.

[14] T. Jayram. Hellinger strikes back: A note on the multi-party information complexity of and. pages 562–573.

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

[16] R. Karp, C. Schindelhauer, S. Shenker, and B. Vocking. Randomized rumor spreading. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, pages 565–, 2000.

[17] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 482–, 2003.

[18] F. Kuhn and R. Oshman. The complexity of data aggregation in directed networks. In *DISC'11*, pages 416–431.

[19] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 2006.

[20] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997.

[21] Nancy A. Lynch. *Distributed Algorithms*. Morgan Kaufmann, 1996.

[22] Amit Manjhi, Vladislav Shkapenyuk, Kedar Dhamdhere, and Christopher Olston. Finding (recently) frequent items in distributed data streams. In *ICDE '05*, pages 767–778.

[23] Boaz Patt-Shamir. A note on efficient aggregate queries in sensor networks. *Theor. Comput. Sci.*, 370(1-3):254–264, 2007.

[24] Jos D. P. Rolim Pavol Duris. Lower bounds on the multiparty communication complexity. *J. Comput. Syst. Sci.*, 56(1):90–95, 1998.

[25] Jeff M. Phillips, Elad Verbin, and Qin Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. In *SODA '12*, pages 486–501.

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

[27] 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 Journal on Computing*, 41(5):1235–1265, 2012.

[28] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *Information Theory, IEEE Transactions on*, 19(4):471–480, 1973.

[29] David P. Woodruff and Qin Zhang. Tight bounds for distributed functional monitoring. In *STOC 2012*, pages 941–960, 2012.

[30] David P. Woodruff and Qin Zhang. Distributed computation does not help. *CoRR*, abs/1304.4636, 2013.