Tight Bounds for Set Disjointness in the Message Passing Model

Mark Braverman [*] Faith Ellen [†] Rotem Oshman [‡] Toniann Pitassi [§]
Vinod Vaikuntanathan [¶]

January 5, 2014

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 multiparty computation, lower bounds on communication complexity in this model and related models have been somewhat scarce. In recent work [**?**, **?**, **?**], 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 tight lower bounds of the form $\Omega(n \cdot k)$ for the Set Disjointness problem in the message passing model. Our bounds are obtained by developing information complexity tools in the message-passing model, and then proving an information complexity lower bound for Set Disjointness. As a corollary, we show a tight lower bound for the task allocation problem [**?**] via a reduction from Set Disjointness.

[*] Department of Computer Science, Princeton University, mbraverm@cs.princeton.edu, Research supported in part by an Alfred P. Sloan Fellowship, an NSF CAREER award (CCF-1149888), and a Turing Centenary Fellowship.

[†] Department of Computer Science, University of Toronto, faith@cs.toronto.edu, Research supported in part by NSERC.

[‡] Department of Computer Science, University of Toronto, rotem@cs.toronto.edu, Research supported in part by NSERC.

[§] Department of Computer Science, University of Toronto, toni@cs.toronto.edu, Research supported in part by NSERC.

[¶] Department of Computer Science, University of Toronto, vinodv@cs.toronto.edu, Research supported in part by NSERC, DARPA award FA8750-11-2-0225 and an Alfred P. Sloan Fellowship.

1 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 finding a minimum-weight spanning tree [**?**] to computing functions of distributed data [**?**, **?**] and distributed computation and verification of network graph structures and properties [**?**, **?**].

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 [**?**]. 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. (One such example is described in [**?**].)

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 have been used extensively in distributed computing, for example, to study gossiping protocols [**?**], to compute various functions of distributed data [**?**], and to understand fundamental problems, such as achieving consensus in the presence of failures [**?**]. Message passing models are also used to study privacy and security in multi-party computation.

In this paper, we choose to focus on the Set Disjointness problem [**?**]. In Set Disjointness, denoted $\text{DISJ}_{n,k}$, 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 [**?**, **?**], 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 [**?**], showing that any protocol 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, namely the coordinator model of Dolev and Feder [**?**]. 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 [**?**], and which has recently played a pivotal role in several communication complexity lower bounds.

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

Theorem 1.1. *For every $\delta > 0$, $n \geq 1$ and $k = \Omega(\log n)$, there is a distribution ζ such that the information complexity of Set Disjointness is $IC_{\zeta,\delta}(DISJ_{n,k}) = \Omega(nk)$ and its communication complexity is $CC_{\delta}(DISJ_{n,k}) = \Omega(nk)$.*

We then apply this lower bound to obtain a lower bound of $\Omega(nk)$ on the Task Allocation problem, $\text{TASKALLOC}_{n,k}$. In this recently proposed problem [**?**], k players must partition n tasks among themselves.

Task Allocation is a useful primitive for distributed systems, where a number of tasks must be performed by the participants in the computation, but not every participant is able to carry out every task. We describe this problem more formally below.

Information complexity and communication complexity. Our main technical tool in this paper is *information complexity*. The main technical result of the present paper concerns the problem 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 [**?**]. In the randomized model, when error is allowed, an $\Omega(n)$ lower bound is also known, although it is considerably more difficult to prove [**?**, **?**]. This result was later improved using information-theoretic techniques by Bar-Yossef et al. [**?**]. Further advances in information complexity allow one to calculate the two-party communication complexity of disjointness precisely, up to additive $o(n)$ terms [**?**].

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 [**?**]. 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. Also notice that since the entire calculation in this case can be carried out by two participants, yielding a trivial $O(n)$ upper bound. The exact dependence of the communication complexity on n and k has been the subject of considerable investigation [**?**, **?**, **?**], with the currently strongest lower bound being $\Omega(\sqrt{n}/2^k k)$ [**?**]. 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. The information complexity approach usually proceeds in two steps: first, a direct-sum result shows that the amount of information the players convey about the problem is additive about the coordinates of the problem (i.e. scales with n in the case of disjointness); second, it is shown that to solve the one-coordinate version of the problem one needs to convey a non-trivial amount of information ($\Omega(1/k)$ in the case of the blackboard model).

In this paper, we consider *message-passing* models of communication complexity. In all of the multiparty models discussed so far, messages are *broadcast* to a centralized 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 pairwise channels. Unlike the other models, it is possible to achieve $\Omega(n \cdot k)$ lower bounds on problems in message passing models [**?**]. We will focus on the *coordinator* message passing model because it is most similar to standard communication complexity models, and lower bounds in this model imply similar lower bounds for other message passing models. In the *coordinator model* [**?**], the players communicate with a coordinator by sending and receiving messages on private channels.

A recent paper [**?**] developed a new technique, called *symmetrization*, for obtaining lower bounds of the form $\Omega(n \cdot k)$ via a 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 multiparty Set Disjointness problem, 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 [**?**, **?**]. 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 [**?**] or in a decentralized manner [**?**]. In [**?**], a lower bound of $\Omega(n \cdot k)$ 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 $\{\textsc{Disj}_{n,2}(S, T_i)\}_{i=1}^{k}$ are empty from the case where only a "small minority" are empty. (The precise values of "large majority" and "small minority" are parameters to the problem.) To obtain this lower bound, [**?**] first proves a direct-sum like result showing that, in order to compute the Gap-Majority of k bits $Z_1, \ldots, Z_k$ in the message-passing model, a protocol must leak $\Omega(k)$ bits of information about $Z_1, \ldots, Z_k$. The known $\Omega(n)$ lower bound on the information complexity of $\textsc{Disj}_{n,2}$ [**?**] is then applied to "lift" the $\Omega(k)$ one-bit lower bound on Gap-Majority to an $\Omega(nk)$ lower bound on Gap-Majority(2-DISJ). In [**?**], similar techniques are used to obtain optimal $\Omega(nk)$ 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.

The message-passing model and its history in distributed computing. The message-passing model is one of the fundamental models in the theory of distributed computing, and many variations of it have been studied. The famous consensus impossibility result [**?**] was originally proven for message-passing systems with faulty processors, and it is also a common setting for other forms of consensus (e.g., Byzantine consensus [**?**] and randomized consensus [**?**], among many others). Lamport's seminal paper introducing the causal order of events in a distributed system [**?**] and his later Paxos protocol for consensus [**?**] are also situated in the message-passing model. There is also much work on achieving data consistency through replication in message-passing systems (e.g., [**?**]). More recently, *gossip protocols* [**?**, **?**, **?**, **?**, **?**] have received considerable attention. In *gossip* (also called *rumor-spreading*), the goal is to quickly disseminate information throughout the network or compute some aggregate function of the information; this is achieved by having every node contact a small number of other nodes (typically, but not always, selected at random) to exchange information with them. Gossip protocols often use very large messages; for instance, a node might forward all the rumors it has collected so far in every round.

These problems and others are sometimes studied in fully-connected networks, and sometimes in networks with an arbitrary graph topology, where communication is more restricted. Lower bounds for the model that we focus on in this paper, the *coordinator model*, implies lower bounds for the basic message passing model where every node can directly communicate with every other node. It remains interesting future work to extend and apply our techniques in settings where communication is governed by an underlying network topology which is not fully connected.

Finally, we point out that the coordinator model is interesting in itself: although it does not model a fully-decentralized distributed system, it is appropriate for data centers or for sensor networks with centralized control. There is a growing body of work on streaming and sketching algorithms set in the coordinator model [**?**, **?**, **?**].

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 [**?**] showed multiparty protocols in the message-passing model for computing any function in an information-theoretically private way, assuming that the corruption threshold $t < k/2$.[1] In addition, we know 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 [**?**, **?**]. 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 3, 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 $\textsc{Disj}_{n,k}$ is at least n times the information cost of solving $\textsc{Disj}_{1,k} = \textsc{And}_k$, and then proving that it is at least $\Omega(k)$. Finally, in Section 7, we reduce set disjointness to the task allocation problem, to obtain an $\Omega(kn)$ lower bound on its communication complexity.

2 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 $\textsc{Disj}_{n,k}$, which might at first glance seem natural:

- *External information cost*, $\mathrm{I}(\mathbf{X}; \Pi(\mathbf{X}))$: 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 [**?**].

  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.

- *The players' 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 do not learn anything. 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 $X^i_j = 0$, by contacting the players in order $i = 1, \ldots, k$ and asking them to send X^i_j. If $X^i_j = 0$ for some i, then

[1] While our description focuses on the notion of semi-honest corruptions where the adversary corrupts t players who run the protocol as prescribed, but try to learn information about the other players' inputs from the transcript of the protocol execution. These results have also been extended to provide strong notions of security against malicious corruptions, sometimes at the expense of a smaller corruption threshold t.

$j \notin \bigcap_{i=1}^{k} X^i$, 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 $\ell < i$. 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, it is learned by *many players*. Because each message is sent to only one player, and we are interested in the *total* amount of communication between the coordinator and the players, we can separately charge each player that learns this bit of information, as this requires the coordinator to communicate separately with each of them.

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 trade-off 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 there should only a few players that have $X_j^i = 0$. On the other hand, our distribution should have *high entropy*, because, otherwise, the players can use Slepian-Wolf coding [**?**] to convey their joint input X to the coordinator using roughly $O(H(X))$ bits. In order to balance these two concerns, we follow [**?**], 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 larger than 2^k. This is necessary for our direct sum theorem (see below). 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 [**?**] and as mentioned above, any lower bound for Set Disjointness (or in the case of [**?**], bitwise-OR and other bitwise functions) 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.

3 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, for example, $\bar{e}_{i,j} = \bar{e}_{\{i,j\}}$.

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, 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

$$Y_{-i} = Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n$$

to denote the tuple obtained from Y by dropping the i-th coordinate. 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 $|m| = |X|$.

Models of computation. As mentioned in the Introduction, we will work in the asynchronous *coordinator* message passing model introduced in [**?**]. In this model, there is one additional participant, called the coordinator, which receives no input, and there is a private channel between every player and the coordinator. However, there are no channels between the players, so they cannot communicate directly with one another. The coordinator also has a private source of randomness. On each channel, the messages alternate between the coordinator and the player. Messages are required to be self-delimiting, so both the coordinator and the player know when one message (from coordinator to the player or vice-versa) has been completely sent. Each player i knows whether or not it is his/her turn to speak by looking at the transcript Π_i between player i and the coordinator. If the last message sent in this transcript was from the coordinator, then it is player i's turn to speak. The coordinator can communicate whenever he is no longer waiting for anyone to speak. This happens when in each transcript Π_j, $j \leq k$, the last message sent in this transcript was from the player. When it is the coordinator's turn to speak, he can send messages to as many players as he wishes. At the end of a protocol, the coordinator outputs the answer. Our complexity measure is the *total* number of bits sent on all channels. Since our complexity measure is the total number of bits, we can assume without loss of generality that the model is sequential and round based: in the first round, the coordinator speaks to exactly one player, and in the next round, this player responds, and so on.

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

$$\max_X \Pr\left[\text{player 1 outputs an incorrect answer}\right],$$

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

The *communication complexity* of a protocol Π is the maximum over all inputs X of the maximum number of bits exchanged between the players and the coordinators when Π is executed with input X. The *δ-error randomized communication complexity* of a problem $\mathcal{P}$ in the coordinator model, which we denote by $\mathrm{CC}_\delta(\mathcal{P})$, is the minimum communication complexity of any randomized protocol Π that solves $\mathcal{P}$ with error at most δ.

Useful classes of distributions. Let us define the class of distributions we use for our direct sum theorem and the lower bound for 1-bit AND. Fix an *input domain* $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_k$, and let $\mathcal{X} = (\mathcal{X}^1, \ldots, \mathcal{X}^k)$ be a random variable denoting the input. Our hard distribution 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 [**?**].

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 [**?**], 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 [**?**].)

Definition 2 (ϵ-collapsing distributions). *We say that 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

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

The square of the Hellinger distance is:

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

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

Lemma 3.1 ([**?**]). *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 facts will be useful to us in the sequel:

Fact 3.2 (Chain rule for mutual information [**?**]). *For any* $A_1, \ldots, A_n$, B *and* C *we have*

$$\mathrm{I}(A_1 \ldots A_n; B \mid C) = \sum_{i=1}^{n} \mathrm{I}(A_i; B \mid A_1 \ldots A_{i-1} C). \tag{1}$$

Lemma 3.3 ("Simplified chain rule"). *If A and B are independent given D, then* $\mathrm{I}(A; BC|D) = \mathrm{I}(A; C|B, D)$.

Proof. By the chain rule, $\mathrm{I}(A; BC \mid D) = \mathrm{I}(A; B \mid D) + \mathrm{I}(A; C \mid B, D)$. Since A and B are independent conditioned on D, we have $\mathrm{I}(A; B \mid D) = 0$, and the claim follows. □

Lemma 3.4 ([?]). *If A, B are independent given D, then* $\mathrm{I}(A; C \mid B, D) \geq \mathrm{I}(A; C \mid D)$.

Lemma 3.5 ([?]). *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 2, 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.*

In Sections 4 and 5 we show that the switched information cost of $\mathrm{D{\small ISJ}}_{n,k}$ under our hard distribution is $\Omega(nk)$. In Section 6 we use this fact to show that the internal information cost of $\mathrm{D{\small ISJ}}_{n,k}$ is also $\Omega(nk)$.

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 3.6. *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,

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

The claim follows. □

Problem statements. In the Set Disjointness problem, $\textsc{Disj}_{n,k}$, each player receives an input $X^i \in \{0,1\}^n$, and the goal is to compute

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

We also consider the *Task Allocation Problem*, $\textsc{TaskAlloc}_{n,k}$. Here we think of the elements $\{1, \ldots, n\}$ as *tasks* that need to be performed. Each player receives an input $X^i \subseteq [n]$ representing the set of tasks it is able to perform, and the coordinator must output an *assignment* $Y : [n] \to [k]$, such that for each $j \in [n]$, $j \in X^{Y(j)}$; that is, every task is assigned to a player that had that task in its input.

4 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 $[n] \setminus \{\mathbf{j}\}$ 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.

- 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.

- 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 $\mathbf{X}_\ell^i$ 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 $\mathbf{X}_\ell^i$.

4. The participants simulate the execution of Π using the joint input

$$\mathrm{embed}(\mathbf{X}, \mathbf{j}, \mathbf{U}) = \left\{(\mathbf{X}_1^i, \ldots, \mathbf{X}_{\mathbf{j}-1}^i, \mathbf{U}^i, \mathbf{X}_{\mathbf{j}+1}^i, \ldots, \mathbf{X}_n^i\right\}_{i=1}^k .$$

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 $\bigwedge_{i=1}^k \mathbf{X}_\ell^i = 0$. By union bound, the probability that $\bigvee_{\ell \neq \mathbf{j}} \bigwedge_{i=1}^k \mathbf{X}_\ell^i = 0$ 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 4.1. *For each player $i \in [k]$ we have*

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

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

Proof. We begin with the first inequality. For each player i, the player's view of the transcript of $\hat{\Pi}$ is given by

$$\hat{\Pi}^i(\mathbf{U}) = \mathbf{j}, \mathbf{Z}_{-\mathbf{j}}, \mathbf{M}_{[1,\mathbf{j}-1]}, \mathbf{X}^i_{[\mathbf{j}+1,n]}, \Pi(\mathrm{embed}(\mathbf{X}, \mathbf{j}, \mathbf{U})).$$

By Lemma 3.3, since the tuple $\langle \mathbf{j}, \mathbf{Z}_{-\mathbf{j}}, \mathbf{M}_{[1,\mathbf{j}-1]}, \mathbf{X}^i_{[\mathbf{j}+1,n]}\rangle$ is independent from $\mathbf{N}$ conditioned on $\mathbf{U}^i$ and $\mathbf{S}$ (or even without the conditioning), we can write

$$\begin{aligned}
&\underset{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi}{\mathrm{I}}\left(\mathbf{N}; \hat{\Pi}^i(\mathbf{U}) \mid \mathbf{U}^i, \mathbf{S}\right) \\
&= \underset{\substack{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi \\ (\mathbf{X}_{-\mathbf{j}},\mathbf{M}_{-\mathbf{j}},\mathbf{Z}_{-\mathbf{j}})\sim\xi^{n-1}}}{\mathrm{I}}\left(\mathbf{N}; \mathbf{j}, \mathbf{Z}_{-\mathbf{j}}, \mathbf{M}_{[1,\mathbf{j}-1]}, \mathbf{X}^i_{[\mathbf{j}+1,n]}, \Pi^i(\mathrm{embed}(\mathbf{X}_{-\mathbf{j}}, \mathbf{j}, \mathbf{U})) \mid \mathbf{U}^i, \mathbf{S}\right) \\
&= \underset{\substack{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi \\ (\mathbf{X}_{-\mathbf{j}},\mathbf{M}_{-\mathbf{j}},\mathbf{Z}_{-\mathbf{j}})\sim\xi^{n-1}}}{\mathrm{I}}\left(\mathbf{N}; \Pi^i(\mathrm{embed}(\mathbf{X}_{-\mathbf{j}}, \mathbf{j}, \mathbf{U})) \mid \mathbf{j}, \mathbf{M}_{[1,\mathbf{j}-1]}, \mathbf{X}^i_{[\mathbf{j}+1,n]}, \mathbf{U}^i, \mathbf{Z}_{-\mathbf{j}}, \mathbf{S}\right) \\
&= \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}}\left(\mathbf{M}_\mathbf{j}; \Pi^i(\mathbf{X}) \mid \mathbf{j}, \mathbf{M}_{[1,\mathbf{j}-1]}, \mathbf{X}^i_{[\mathbf{j},n]}, \mathbf{Z}\right). \qquad (2)
\end{aligned}$$

Next, since $\mathbf{X}^i_{[1,\mathbf{j}-1]}$ and $\mathbf{M}_{\mathbf{j}}$ (which we previously called $\mathbf{N}$) are independent, even given the conditioning in (2), we can apply Lemma 3.4 to add conditioning on $\mathbf{X}^i_{[1,\mathbf{j}-1]}$, yielding

$$\begin{aligned}
&\mathop{\mathrm{I}}_{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi}\left(\mathbf{N};\hat{\Pi}^i(\mathbf{U}) \mid \mathbf{U}^i,\mathbf{S}\right)\\
&\le \mathop{\mathrm{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}\left(\mathbf{M}_{\mathbf{j}};\Pi^i(\mathbf{X}) \mid \mathbf{j},\mathbf{M}_{[1,\mathbf{j}-1]},\mathbf{X}^i,\mathbf{Z}\right)\\
&= \frac{1}{n}\sum_{j=1}^{n} \mathop{\mathrm{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}\left(\mathbf{M}_j;\Pi^i(\mathbf{X}) \mid \mathbf{M}_{[1,j-1]},\mathbf{X}^i,\mathbf{Z}\right) = \frac{1}{n}\mathop{\mathrm{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}(\mathbf{M};\Pi^i(\mathbf{X}) \mid \mathbf{X}^i,\mathbf{Z}).
\end{aligned}$$

The last step uses the chain rule.

Now let us prove the second inequality, which is quite similar. We begin as before: by Lemma 3.3, since the tuple $\langle \mathbf{j},\mathbf{Z}_{-\mathbf{j}},\mathbf{M}_{[1,\mathbf{j}-1]},\mathbf{X}^i_{[\mathbf{j}+1,n]}\rangle$ is independent from $\mathbf{U}^i$ conditioned on $\mathbf{N}$ and $\mathbf{S}$,

$$\begin{aligned}
&\mathop{\mathrm{I}}_{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi}\left(\mathbf{U}^i;\hat{\Pi}^i(\mathbf{U}) \mid \mathbf{N},\mathbf{S}\right)\\
&= \mathop{\mathrm{I}}_{\substack{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi\\(\mathbf{X}_{-\mathbf{j}},\mathbf{M}_{-\mathbf{j}},\mathbf{Z}_{-\mathbf{j}})\sim\xi^{n-1}}}\left(\mathbf{U}^i;\mathbf{j},\mathbf{Z}_{-\mathbf{j}},\mathbf{M}_{[1,\mathbf{j}-1]},\mathbf{X}^i_{[\mathbf{j}+1,n]},\Pi^i(\mathrm{embed}(\mathbf{X}_{-\mathbf{j}},\mathbf{j},\mathbf{U})) \mid \mathbf{N},\mathbf{S}\right)\\
&= \mathop{\mathrm{I}}_{\substack{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi\\(\mathbf{X}_{-\mathbf{j}},\mathbf{M}_{-\mathbf{j}},\mathbf{Z}_{-\mathbf{j}})\sim\xi^{n-1}}}\left(\mathbf{U}^i;\Pi^i(\mathrm{embed}(\mathbf{X}_{-\mathbf{j}},\mathbf{j},\mathbf{U})) \mid \mathbf{j},\mathbf{M}_{[1,\mathbf{j}-1]},\mathbf{N},\mathbf{X}^i_{[\mathbf{j}+1,n]},\mathbf{Z}_{-\mathbf{j}},\mathbf{S}\right)\\
&= \mathop{\mathrm{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}\left(\mathbf{X}^i_{\mathbf{j}};\Pi^i(\mathbf{X}) \mid \mathbf{j},\mathbf{M}_{[1,\mathbf{j}]},\mathbf{X}^i_{[\mathbf{j}+1,n]},\mathbf{Z}\right). \qquad (3)
\end{aligned}$$

Next, since $\mathbf{M}_{[\mathbf{j}+1,n]}$ and $\mathbf{X}^i_{\mathbf{j}}$ (previously called $\mathbf{U}^i$) are independent given the conditioning in (3), we can apply Lemma 3.4 to add conditioning on $\mathbf{M}_{[\mathbf{j}+1,n]}$, yielding

$$\begin{aligned}
&\mathop{\mathrm{I}}_{(\mathbf{U},\mathbf{N},\mathbf{S})\sim\xi}\left(\mathbf{N};\hat{\Pi}^i(\mathbf{U}) \mid \mathbf{U}^i\mathbf{S}\right)\\
&\le \mathop{\mathrm{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}\left(\mathbf{X}^i_{\mathbf{j}};\Pi^i(\mathbf{X}) \mid \mathbf{j},\mathbf{M},\mathbf{X}^i_{[\mathbf{j}+1,n]},\mathbf{Z}\right)\\
&= \frac{1}{n}\sum_{j=1}^{n} \mathop{\mathrm{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}\left(\mathbf{X}^i_j;\Pi^i(\mathbf{X}) \mid \mathbf{M},\mathbf{X}^i_{[j+1,n]},\mathbf{Z}\right) = \frac{1}{n}\mathop{\mathrm{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}(\mathbf{X}^i;\Pi^i(\mathbf{X}) \mid \mathbf{M},\mathbf{Z}).
\end{aligned}$$

□

The direct sum theorem follows immediately from Lemma 4.1:

Theorem 4.2. *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*

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

5 The Information Complexity of One-Bit AND

By Theorem 4.2, 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 2):

- 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 similarly, $\Pi^i(X)$ denotes 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 $\mathbf{Z} = z$. Finally, given a sequence $i_1, \ldots, i_\ell \in [k]$, we use $\bar{e}_{i_1,\ldots,i_\ell}$ to denote the input in which players $i_1, \ldots, i_\ell$ receive zero, and all other players receive one.

5.1 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.

The basic structural property on which we rely is *rectangularity*, introduced in [**?**] for the two-player setting and the multi-player model with communication by shared blackboard. Rectangularity asserts, informally speaking, that if we partition the players into sets $A_1, \ldots, A_m \subseteq [k]$, the protocol's probability distribution over transcripts can be decomposed into a product of functions $f_1, \ldots, f_m$, such that each f_i depends only in the inputs to players in A_i. Here we require only a simple version where we use two sets, $A_1 = \{i\}$ and $A_2 = [k] \setminus \{i\}$ for some player $i \in [k]$. The lemma follows by reduction from two-player rectangularity [**?**], but for the sake of completeness we include a proof.

Lemma 5.1 (One-player rectangularity for the coordinator model)**.** *Let Π be a k-player private-coin protocol in the coordinator model, with inputs from $\mathcal{X} = \mathcal{X}^1 \times \ldots \times \mathcal{X}^k$. For $i \in [k]$, let $\mathcal{T}^i$ denote the set of possible transcripts observed by player i, so any transcript of Π is in $\mathcal{T}^1 \times \cdots \times \mathcal{T}^k$. Then, for all $i \in [k]$, there exist mappings $q^i : \mathcal{X}^i \times \mathcal{T}^i \to [0,1]$, $q^{-i} : \mathcal{X}^{-i} \times \mathcal{T}^i \to [0,1]$ and $p^{-i} : \mathcal{X}^{-i} \times \mathcal{T} \to [0,1]$ such that for any input $X \in \mathcal{X}$ and any transcript $\tau = (\tau^1, \ldots, \tau^k) \in \mathcal{T}^1 \times \cdots \times \mathcal{T}^k$,*

$$\Pr\left[\Pi^i(X) = \tau^i\right] = q^i(X^i, \tau^i) \cdot q^{-i}(X^{-i}, \tau^i) \text{ and}$$
$$\Pr\left[\Pi(X) = \tau\right] = q^i(X^i, \tau^i) \cdot p^{-i}(X^{-i}, \tau).$$

Proof. For any player i and any transcript $\tau^i \in \mathcal{T}^i$, let $\mathcal{A}(\tau^i) = \{(X, R) \mid \Pi_i^R(X) = \tau^i\}$ denote the set of inputs and random coin tosses such that τ^i is the transcript $\Pi_i^R(X)$ of the communication between player i and the coordinator in the deterministic protocol Π^R obtained from Π by fixing the outcome of the random coin tosses to R. Also, let $\mathcal{A}^i(\tau^i) = \{(X^i, R^i) \mid (X, R) \in \mathcal{A}(\tau^i)\}$ and $\mathcal{A}^{-i}(\tau^i) = \{(X^{-i}, R^{-i}) \mid (X, R) \in \mathcal{A}(\tau^i)\}$. Then, by the rectangular property for deterministic 2-player protocols, for all (X, R), $\Pi_i^R(X) = \tau^i$ if and only if $(X^i, R^i) \in \mathcal{A}^i(\tau^i)$ and $(X^{-i}, R^{-i}) \in \mathcal{A}^{-i}(\tau^i)$.

For any $X^i \in \mathcal{X}^i$, any $X^{-i} \in \mathcal{X}^{-i}$, and any $\tau^i \in \mathcal{T}^i$, define

$$q^i(X^i, \tau^i) = \Pr_{R^i}\left[(X^i, R^i) \in \mathcal{A}^i(\tau^i)\right] \text{ and}$$
$$q^{-i}(X^{-i}, \tau^i) = \Pr_{R^{-i}}\left[(X^{-i}, R^{-i}) \in \mathcal{A}^{-i}(\tau^i)\right].$$

On any input X, player i chooses R^i uniformly and the other players choose R^{-i} independently and uniformly. Therefore,

$$\begin{aligned}
&\Pr\left[\Pi^i(X) = \tau^i\right] \\
&\quad = \Pr_R\left[\Pi_i^R(X) = \tau^i\right] \\
&\quad = \Pr_{R^i}\left[(X^i, R^i) \in \mathcal{A}^i(\tau^i)\right] \cdot \Pr_{R^{-i}}\left[(X^{-i}, R^{-i}) \in \mathcal{A}^{-i}(\tau^i)\right] \\
&\quad = q^i(X^i, \tau^i) \cdot q^{-i}(X^{-i}, \tau^i).
\end{aligned}$$

For any transcript $\tau \in \mathcal{T}$, let $\mathcal{B}(\tau) = \{(X, R) \mid \Pi^R(X) = \tau\}$ denote the set of inputs and random coin tosses such that τ is the transcript $\Pi^R(X)$ of all communication (to and from the coordinator) in the deterministic protocol Π^R obtained from Π by fixing the outcome of the random coin tosses to R. Let $\mathcal{B}^{-i}(\tau) = \{(X^{-i}, R^{-i}) \mid (X, R) \in \mathcal{B}(\tau)\}$. By the rectangular property for deterministic protocols, for all (X, R), $\Pi^R(X) = \tau$ if and only if $(X^i, R^i) \in \mathcal{A}^i(\tau^i)$ and $(X^{-i}, R^{-i}) \in \mathcal{B}^{-i}(\tau)$.

For any $X^{-i} \in \mathcal{X}^{-i}$ and any $\tau \in \mathcal{T}$, define

$$p^{-i}(X^{-i}, \tau) = \Pr_{R^{-i}}\left[(X^{-i}, R^{-i}) \in \mathcal{B}^{-i}(\tau)\right].$$

On any input X, player i chooses R^i uniformly and the other players choose R^{-i} independently and uniformly. Therefore,

$$\begin{aligned}
&\Pr\left[\Pi(X) = \tau\right] \\
&\quad = \Pr_R\left[\Pi^R(X) = \tau\right] \\
&\quad = \Pr_{R^i}\left[(X^i, R^i) \in \mathcal{A}^i(\tau^i)\right] \cdot \Pr_{R^{-i}}\left[(X^{-i}, R^{-i}) \in \mathcal{B}^{-i}(\tau)\right] \\
&\quad = q^i(X^i, \tau^i) \cdot p^{-i}(X^{-i}, \tau).
\end{aligned}$$

□

For convenience, when we apply Lemma 5.1, we sometimes write

$$\Pr\left[\Pi(X) = \tau\right] = p^i(X^i, \tau) \cdot p^{-i}(X^{-i}, \tau),$$

where $p^i(X^i, \tau) = q^i(X^i, \tau^i)$.

Rectangularity, in turn, implies the Z-Lemma (or Pythagorean Lemma) of [**?**]. Here we use a simplified version (which omits one of the terms on the right-hand side):

Lemma 5.2 (Diagonal Lemma)**.** *For any $X, Y \in \mathcal{X}$ and $\ell \in [k]$ we have*

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

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 rectangular property, which, informally speaking, "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 5.3 (Conditional rectangularity for $\mathbf{M}$ and $\mathbf{X}^i$ under ξ)**.** *Let Π be a k-player private-coin protocol for AND_k. For $i \in [k]$, let $\mathcal{T}^i$ denote the set of possible transcripts observed by player i. Then there exists a function $c : \{0,1\} \times [k] \times \mathcal{T} \to [0,1]$ and, for all $i \in [k]$, there exists a function $c^i : \{0,1\} \times [k] \times \mathcal{T}^i \to [0,1]$ such that for any $x \in \mathcal{X}^i$, $m \in \{0,1\}$, $z \in [k] \setminus \{i\}$, $\tau \in \mathcal{T}$, and $\tau^i \in \mathcal{T}^i$,*

$$\Pr[\Pi(\mathbf{X}) = \tau \mid \mathbf{X}^i = x, \mathbf{M} = m, \mathbf{Z} = z] = p^i(x, \tau) \cdot c(d, z, \tau) \textit{ and}$$
$$\Pr[\Pi^i(\mathbf{X}) = \tau^i \mid \mathbf{X}^i = x, \mathbf{M} = m, \mathbf{Z} = z] = q^i(x, \tau^i) \cdot c^i(d, z, \tau^i),$$

where $p^i(x, \tau) = q^i(x, \tau^i)$ is the function from Lemma 5.1. Here the probability is over the protocol's own randomness as well as the input $\mathbf{X}$ drawn from ξ with the stated conditioning.

Proof. By Lemma 5.1 there exist functions q^i and q^{-i} such that, for any input $X \in \mathcal{X}$,

$$\Pr\left[\Pi^i(X) = \tau^i\right] = q^i(X^i, \tau^i) \cdot q^{-i}(X^{-i}, \tau^i).$$

Therefore we can write

$$\begin{aligned}
&\Pr\left[\Pi^i(\mathbf{X}) = \tau^i \mid \mathbf{M} = m, \mathbf{Z} = z, \mathbf{X}^i = x\right] \\
&= \sum_{X \in \mathcal{X}} q^i(X^i, \tau^i) \cdot q^{-i}(X^{-i}, \tau^i) \cdot \Pr\left[\mathbf{X} = X \mid \mathbf{M} = m, \mathbf{Z} = z, \mathbf{X}^i = x\right] \\
&= q^i(x, \tau^i) \cdot \sum_{X^{-i} \in \mathcal{X}^{-i}} q^{-i}(X^{-i}, \tau^i) \cdot \Pr\left[\mathbf{X}^{-i} = X^{-i} \mid \mathbf{M} = m, \mathbf{Z} = z\right].
\end{aligned}$$

Here we use the fact that the inputs $\mathbf{X}^1, \ldots, \mathbf{X}^k$ are independent conditioned on $\mathbf{M}$ and $\mathbf{Z}$. The second claim follows by setting

$$c^i(d, z, \tau^i) = \sum_{X^{-i} \in \mathcal{X}^{-i}} q^{-i}(X^{-i}, \tau^i) \cdot \Pr\left[\mathbf{X}^{-i} = X^{-i} \mid < = m, \mathbf{Z} = z\right].$$

Similarly, Lemma 5.1 implies there exist functions p^i and p^{-i} such that, for any input $X \in \mathcal{X}$,

$$\Pr\left[\Pi(X) = \tau\right] = p^i(X^i, \tau) \cdot p^{-i}(X^{-i}, \tau),$$

so

$$\begin{aligned}
&\Pr\left[\Pi(\mathbf{X}) = \tau \mid \mathbf{M} = m, \mathbf{Z} = z, \mathbf{X}^i = x\right] \\
&= p^i(x, \tau) \cdot \sum_{X^{-i} \in \mathcal{X}^{-i}} p^{-i}(X^{-i}, \tau) \cdot \Pr\left[\mathbf{X}^{-i} = X^{-i} \mid \mathbf{M} = m, \mathbf{Z} = z\right]
\end{aligned}$$

and the first claim follows by setting

$$c(d, z, \tau) = \sum_{X^{-i} \in \mathcal{X}^{-i}} p^{-i}(X^{-i}, \tau) \cdot \Pr\left[\mathbf{X}^{-i} = X^{-i} \mid \mathbf{M} = m, \mathbf{Z} = z\right].$$

□

Lemma 5.3 yields the following variant of the Diagonal Lemma (Lemma 5.2).

Lemma 5.4 (Diagonal Lemma for $\mathbf{M}$ and $\mathbf{X}^i$)**.** *For any $i \neq z$ we have*

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

Proof. The proof closely follows the proof of the original Z-Lemma from [**?**], but we include it here for completeness.

Recall that $\Pi^i[1,1,z] = \Pi^i(\bar{e}_z)$. By Lemmas 5.1 and 5.3, we can decompose the distributions from the lemma statement as follows:

$$\begin{aligned}
&\Pr\left[\Pi^i[0,0,z] = \tau^i\right] = q^i(0,\tau^i) \cdot c^i(0,z,\tau^i),\\
&\Pr\left[\Pi^i[1,1,z] = \tau^i\right] = \Pr\left[\Pi^i(\bar{e}_z) = \tau^i\right] = q^i(1,\tau^i) \cdot q^{-i}((\bar{e}_z)^{-i},\tau^i) = q^i(1,\tau^i) \cdot c^i(1,z,\tau^i), \text{ and}\\
&\Pr\left[\Pi^i(\bar{e}_{i,z}) = \tau^i\right] = q^i(0,\tau^i) \cdot q^{-i}((\bar{e}_{i,z})^{-i},\tau^i) = q^i(0,\tau^i) \cdot q^{-i}((\bar{e}_z)^{-i},\tau^i).
\end{aligned}$$

From the definition of Hellinger distance, it follows that

$$\begin{aligned}
1 - h^2(\Pi^i[0,0,z], \Pi^i[1,1,z]) &= \sum_{\tau^i} \sqrt{q^i(0,\tau^i) \cdot c^i(0,z,\tau^i) \cdot q^i(1,\tau^i) \cdot c^i(1,z,\tau^i)}\\
&\leq \sum_{\tau^i} \sqrt{q^i(0,\tau^i) q^i(1,\tau^i)} \left(\frac{c^i(0,z,\tau^i) + c^i(1,z,\tau^i)}{2}\right)\\
&= \frac{1}{2} \sum_{\tau^i} \sqrt{q^i(0,\tau^i) c^i(0,z,\tau^i) q^i(1,\tau^i) c^i(0,z,\tau^i)}\\
&\quad + \frac{1}{2} \sum_{\tau^i} \sqrt{q^i(0,\tau^i) q^{-i}((\bar{e}_z) - i,\tau^i) q^i(1,\tau^i) q^{-i}((\bar{e}_z)^{-i},\tau^i)}\\
&= \left(1 - h^2(\Pi^i[0,0,z], \Pi^i[1,0,z])\right)/2 + \left(1 - h^2(\Pi^i(\bar{e}_{i,z}), \Pi^i(\bar{e}_z)\right)/2\\
&= 1 - \left(h^2(\Pi^i[0,0,z], \Pi^i[1,0,z]) + h^2(\Pi^i(\bar{e}_{i,z}), \Pi^i(\bar{e}_z)\right)/2\\
&\leq 1 - h^2(\Pi^i(\bar{e}_{i,z}), \Pi^i(\bar{e}_z))/2.
\end{aligned}$$

□

Note that Lemma 5.2 concerns the complete transcript Π, while Lemma 5.4 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. What does it mean to "keep everything the same except $\mathbf{X}^i$"? One option is to fix $\mathbf{M} = m$ and a specific value $\mathbf{Z} = z \neq i$, and let $\mathbf{X}^i$ change from 0 to 1. This is well-defined only in the case where $m = 0$, because when $\mathbf{M} = 1$ and $\mathbf{Z} \neq i$, we must have $\mathbf{X}^i = 1$. The other option is to fix a specific input $\mathbf{X}^{-i} = X^{-i}$ for the rest of the players and let $\mathbf{X}^i$ change from 0 to 1. We are particularly interested in the case where all players receive 1, except for one player, $z \neq i$, and possibly player i itself.

Lemma 5.5 (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,

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

Proof. Given a complete transcript τ, let τ^i denote player i's part of the transcript. By Lemma 5.3,

$$\Pr[\Pi(\mathbf{X}) = \tau \mid \mathbf{X}^i = x, \mathbf{M} = 0, \mathbf{Z} = z] = p^i(x,\tau) \cdot c(0,z,\tau) \text{ and}$$
$$\Pr[\Pi^i(\mathbf{X}) = \tau^i \mid \mathbf{X}^i = x, \mathbf{M} = 0, \mathbf{Z} = z] = q^i(x,\tau^i) \cdot c^i(0,z,\tau^i).$$

Moreover,

$$\Pr[\Pi^i(\mathbf{X}) = \tau^i \mid \mathbf{X}^i = x, \mathbf{M} = 0, \mathbf{Z} = z] = \sum_{\substack{\tau' \in \mathcal{T} \\ \tau'^i = \tau^i}} \Pr[\Pi(\mathbf{X}) = \tau' \mid \mathbf{X}^i = x, \mathbf{M} = 0, \mathbf{Z} = z],$$

so,

$$c^i(d,z,\tau^i) = \sum_{\substack{\tau' \in \mathcal{T} \\ \tau'^i = \tau^i}} c(d,z,\tau').$$

Therefore,

$$\begin{aligned}
&1 - h^2(\Pi^i[0,0,z], \Pi^i[1,0,z]) \\
&= \sum_{\tau^i} \sqrt{q^i(0,\tau^i)c^i(d,z,\tau^i)q^i(1,\tau^i)c^i(d,z,\tau^i)} \\
&= \sum_{\tau^i} \left(\sqrt{q^i(0,\tau^i)q^i(1,\tau^i)} \sum_{\substack{\tau' \in \mathcal{T} \\ \tau'^i = \tau^i}} c(d,z,\tau') \right) \\
&= \sum_{\tau} \sqrt{q^i(0,\tau^i)q^i(1,\tau^i)} c(d,z,\tau) \\
&= \sum_{\tau} \sqrt{q^i(0,\tau^i)c(d,z,\tau)q^i(1,\tau^i)c(d,z,\tau)} \\
&= 1 - h^2(\Pi[i,0,0,z], \Pi[i,1,0,z]).
\end{aligned}$$

The other part of the lemma is similar: it is obtained by using Lemma 5.1 instead of Lemma 5.3 and replacing c, c^i with q^{-i}, p^{-i} (respectively). □

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

5.2 Step I: setting up a rectangle.

Fix a player i and a value $z \neq i$, and consider the following four distributions:

$$\begin{array}{cc} \Pi^i[0,0,z] & \Pi^i[1,0,z] \\ \\ \Pi^i(\bar{e}_{i,z}) & \Pi^i(\bar{e}_z) = \Pi^i[1,1,z] \end{array}$$

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 has the following nice property:

$$\begin{aligned}&\Pr\left[\mathbf{X}^i = 0 \mid \mathbf{M} = 0, \mathbf{Z} = z\right] = \Pr\left[\mathbf{X}^i = 1 \mid \mathbf{M} = 0, \mathbf{Z} = z\right] = 1/2, \quad \text{and}\\ &\Pr\left[\mathbf{M} = 0 \mid \mathbf{X}^i = 1, \mathbf{Z} = z\right] = \Pr\left[\mathbf{M} = 1 \mid \mathbf{X}^i = 1, \mathbf{Z} = z\right] = 1/2.\end{aligned}$$

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 3.5, which yields

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

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 3.5 again yields

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

Recall that $\Pr\left[\mathbf{M} = 0 \mid \mathbf{Z} = z\right] = 2/3$ (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 $\mathrm{I}\left(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z} = z\right) \geq (2/3)\, \mathrm{I}\left(\mathbf{X}^i; \Pi^i \mid \mathbf{M} = 0, \mathbf{Z} = z\right)$ and $\mathrm{I}\left(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = z\right) \geq (2/3)\, \mathrm{I}\left(\mathbf{M}; \Pi^i \mid \mathbf{X}^i = 1, \mathbf{Z} = z\right)$. It follows that

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

The last step uses the triangle inequality. Now we apply Lemma 5.4, which together with the above yields

$$\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = z) + \mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z} = z) \geq \frac{h^2(\Pi^i(\bar{e}_{i,z}), \Pi^i(\bar{e}_z))}{3}. \tag{4}$$

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

$$\begin{aligned}&\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z}) + \mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z}) \geq \frac{1}{k} \sum_{z \neq i} \left(\mathrm{I}(\mathbf{M}; \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = z) + \mathrm{I}(\mathbf{X}^i; \Pi^i \mid \mathbf{M}, \mathbf{Z} = z)\right)\\ &\overset{(4)}{\geq} \frac{k-1}{3k} \mathop{\mathbb{E}}_{\mathbf{Z} \neq i}\left[h^2(\Pi^i(\bar{e}_{i,\mathbf{Z}}, \Pi^i(\bar{e}_{\mathbf{Z}}))\right] \geq \frac{1}{6} \mathop{\mathbb{E}}_{\mathbf{Z} \neq i}\left[h^2(\Pi^i(\bar{e}_{i,\mathbf{Z}}, \Pi^i(\bar{e}_{\mathbf{Z}}))\right].\end{aligned} \tag{5}$$

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 (5) 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 AND_k is 0 in both cases? This will again follow from the structural properties of the protocol.

5.3 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 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 $\text{AND}_k(\bar{e}_i) = \text{AND}_k(\bar{e}_j) = 0$, but $\text{AND}_k(1^k) = 1$. By the correctness of the protocol and Lemma 3.1, 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

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

Now summing across all 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 (5), this yields our main result for this section:

Theorem 5.6. *For any $k > 1$, $\text{SIC}_{\xi,\delta}(\text{AND}_k) \geq (1-\delta)^2/96$.*

Combining Theorem 5.6 with our direct-sum theorem from Section 4, we obtain

Theorem 5.7. *For any $n \geq 1$ and for $k = \Omega(\log n)$, $\text{SIC}_{\eta,\delta}(\text{DISJ}_{n,k}) = \Omega(nk)$.*

6 Internal Information Complexity and Communication Complexity of Set Disjointness

Recall our definition of the internal information cost of a protocol from Section 3:

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

We will now use Theorem 5.7 to show that the internal information complexity of set disjointness is $\Omega(nk)$.

In Theorem 5.7 we showed that for all protocols Π we have

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

Let ζ be the distribution η restricted just to the input $\mathbf{X}$ (that is, ζ is the marginal distribution of $\mathbf{X}$ under η). We will show:

Theorem 6.1. *Let Π be a protocol in the coordinator model. Let η be a switched distribution and ζ be the marginal distribution of $\mathbf{X}$ under η, as above. Namely, ζ is the marginal distribution of $\mathbf{X}$ where $(\mathbf{X}, \mathbf{M}, \mathbf{Z}) \sim \eta$. Then,*

$$IC_\zeta(\Pi) > \underset{\eta}{\text{SIC}}(\Pi) - O(n \log k).$$

Proof. We consider the two terms in each sum separately. We start with the term corresponding to the amount of information learned by the coordinator. By the definition of η we have that $I(\mathbf{X}^i; \mathbf{X}^{[1..i-1]} | \mathbf{M}, \mathbf{Z}) = 0$ and thus by Lemma 3.4,

$$I(\mathbf{X}^i; \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}) \leq I(\mathbf{X}^i; \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}, \mathbf{X}^{[1..i-1]}).$$

Using the Chain Rule, we get:

$$\begin{aligned}
\sum_{i\in[k]} \left[\underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^i; \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}) \right] &\leq \sum_{i\in[k]} \left[\underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^i; \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}, \mathbf{X}^{[1..i-1]}) \right] \\
&\leq \sum_{i\in[k]} \left[\underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^i; \Pi(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}, \mathbf{X}^{[1..i-1]}) \right] \\
&= \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}; \Pi(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}) \\
&\leq \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}, \mathbf{M}, \mathbf{Z}; \Pi(\mathbf{X})) \\
&= \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}; \Pi(\mathbf{X})) + \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{M}, \mathbf{Z}; \Pi(\mathbf{X}) \mid \mathbf{X}) \\
&\leq \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}; \Pi(\mathbf{X})) + H(\mathbf{M}, \mathbf{Z}) \\
&\leq \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}; \Pi(\mathbf{X})) + O(n \log k)
\end{aligned}$$

Next, we consider the terms corresponding to what individual players learn. For each $i \in [k]$ we have

$$\begin{aligned}
\underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{M}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{Z}) &\leq \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^{-i}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{Z}) \\
&\leq \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^{-i}, \mathbf{Z}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i) \\
&= \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^{-i}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i) + \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{Z}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{X}^{-i}) \\
&\leq \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^{-i}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i) + H(\mathbf{Z}) \leq \underset{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\eta}{\mathrm{I}} (\mathbf{X}^{-i}; \Pi^i(\mathbf{X}) \mid \mathbf{X}^i) + \log k.
\end{aligned}$$

Putting these two calculations together we obtain that

$$\underset{\eta}{\mathrm{SIC}}(\Pi) < IC_\zeta(\Pi) + O(n \log k),$$

and thus $IC_\zeta(\Pi) > \mathrm{SIC}_\eta(\Pi) - O(n \log k)$, completing the proof.

□

We are now ready to prove our main theorem:

Theorem 6.2. *For any $\delta > 0$, $n \geq 1$ and for $k = \Omega(\log n)$,*

$$IC_\zeta(\textsc{Disj}_{n,k}) = \Omega(n \cdot k) \quad \textit{and} \quad CC_\delta(\textsc{Disj}_{n,k}) = \Omega(n \cdot k).$$

Proof. The first part of the theorem follows from Theorem 6.1 and Theorem 5.7. The second part follows from the first part as well as the connection between communication complexity and information complexity (Lemma 3.6). □

7 Lower Bound for Task Allocation

In the task allocation problem, there are k players and n tasks. Each player i receives as input a set X^i which specifies a subset of tasks that it is capable of performing. The goal is for the players to partition the tasks between them: each player i must output a subset of tasks such that each task is completed by exactly one player. To make this problem feasible, we consider only inputs for which each tasks has at least one player who is capable of performing it. Thus, task allocation is a promise problem. We require that, at the end of the protocol, the coordinator knows which player is assigned to each task.

Task allocation is a distributed one-shot variant of the well-known k-server problem, where a centralized online algortihm assigns tasks to k servers, minimizing the total cost of servicing all tasks. In the k-server problem, each (server, task) pair is associated with a cost for having the server perform the task, and the tasks arrive continually and must be assigned in an online manner. In the task allocation problem, all tasks are given in the beginning, and all have a cost of either 1 or infinity. Partitioning the tasks between the players corresponds to finding a minimum-weight assignment of tasks to servers. Task allocation is also closely related to the problem of finding a rooted spanning tree in directed broadcast networks [**?**].

Drucker, Kuhn and Oshman [**?**] showed tight communication complexity lower bounds for the two player task allocation problem. In this section, we generalize this lower bound to the k player setting by showing an $\Omega(nk)$ lower bound for task allocation in the message passing model.

Our reduction is similar in spirit to the reduction, due to Noga Alon [**?**], from two-party set disjointness to the promise task allocation problem in the two-player case.

Theorem 7.1. *There is a reduction from k-party Set Disjointness to k-party (Promise) Task Allocation with an overhead of $O(n \log n + k)$ bits. That is, given a task allocation protocol that communicates $C_{\mathsf{TA}}(n, k)$ bits, there is a protocol for set disjointness that communicates*

$$C_{\mathsf{SD}}(n, k) = C_{\mathsf{TA}}(n, k) + O(n \log n + k)$$

bits. Thus, for large enough n and large enough a, if $k \geq a \log n$, then the communication complexity of Task Allocation in the coordinator model is $\Omega(nk)$.

Proof. We now give the reduction from multiplayer set disjointness to multiparty (promise) task allocation in the coordinator model. Let the input to the set disjointness problem be $X^1, \ldots, X^k$. Define $Y^i = [n] - X^i$. As before, note that

$$\cap_{i=1}^{k} X^i \text{ is empty} \quad \textit{if and only if} \quad \cup_{i=1}^{k} Y^i = [n]$$

The players simulate the protocol for task allocation on the inputs Y^i, and the coordinator gets the output $(Z^1, Z^2, \ldots, Z^k)$ where Z^i is the set of tasks that player i is expected to complete.[2] The protocol then proceeds as follows:

1. The coordinator checks that the sets $Z^1, \ldots, Z^k$ form a partition of $[n]$. If not, the coordinator outputs "not disjoint" and halts. If the check passes, proceed to the next step.

2. The coordinator sends Z^i to player i. Each player i checks whether $Z^i \subseteq Y^i$. If $Z^i \nsubseteq Y^i$, player i sends a "not disjoint" message to the coordinator (and if $Z^i \subseteq Y^i$, it sends an "OK" message.) If the coordinator receives a "not disjoint" message from any player, it outputs "not disjoint" and halts. Otherwise, it outputs "disjoint" and halts.

[2] Our definition of the coordinator model postulates that the coordinator learns the result at the end of the protocol. A similar reduction works in the case where the players learn their respective outputs, namely each player i learns Z^i, the set of tasks that he is expected to complete.

Assume that $(X^1, \dots, X^k)$ is a YES instance of set disjointness, namely that $\cap_{i=1}^k X^i = \varphi$. Then, $\cup_{i=1}^k Y^i = [n]$, and the input $(Y^1, \dots, Y^k)$ satisfies the promise to the task allocation problem. By the correctness of the task allocation protocol, the output $(Z^1, \dots, Z^k)$ is a valid allocation of the tasks, namely, $(Z^1, \dots, Z^k)$ forms a partition of the universe $[n]$, and $Z^i \subseteq Y^i$ for each $i \in [k]$. Thus, the coordinator will output "disjoint" in the above protocol.

On the other hand, if $(X^1, \dots, X^k)$ is a NO instance of set disjointness, then we know that $\cup_{i=1}^k Y^i \neq [n]$ and either of the following two events happen:

- for some i, $Z^i \nsubseteq Y^i$; or
- $\cup_{i=1}^k Z^i \neq [n]$

Player i will detect the first of these two cases, and the coordinator will detect the second. In either case, the coordinator will output "not disjoint".

As for the complexity of the protocol, the coordinator runs step 2 of the protocol only if step 1 passes, namely if $(Z^1, \dots, Z^k)$ forms a partition of $[n]$. In this case, the overhead of the protocol is $O(n \log n + k)$ bits. Since the communication complexity of set disjointness in the coordinator model is $\Omega(nk)$, so is the communication complexity of task allocation. □