Near-Optimal Lower Bounds on the Multi-Party Communication Complexity of Set Disjointness

Amit Chakrabarti *
School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540
amitc@ias.edu

Subhash Khot †
Dept. of Computer Science
Princeton University
Princeton, NJ 08544
khot@cs.princeton.edu

Xiaodong Sun *
School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540
sunxd@ias.edu

Abstract

We study the communication complexity of the set disjointness problem in the general multi-party model. For t players, each holding a subset of a universe of size n, we establish a near-optimal lower bound of $\Omega(n/(t \log t))$ on the communication complexity of the problem of determining whether their sets are disjoint. In the more restrictive one-way communication model, in which the players are required to speak in a predetermined order, we improve our bound to an optimal $\Omega(n/t)$. These results improve upon the earlier bounds of $\Omega(n/t^2)$ in the general model, and $\Omega(\varepsilon^2 n/t^{1+\varepsilon})$ in the one-way model, due to Bar-Yossef, Jayram, Kumar, and Sivakumar [5]. As in the case of earlier results, our bounds apply to the unique intersection promise problem.

This communication problem is known to have connections with the space complexity of approximating frequency moments in the data stream model. Our results lead to an improved space complexity lower bound of $\Omega(n^{1-2/k}/\log n)$ for approximating the k^{th} frequency moment with a constant number of passes over the input, and a technical improvement to $\Omega(n^{1-2/k})$ if only one pass over the input is permitted.

Our proofs rely on the information theoretic direct sum decomposition paradigm of Bar-Yossef et al [5]. Our improvements stem from novel analytical techniques, as opposed to earlier techniques based on Hellinger and related distances, for estimating the information cost of protocols for one-bit functions.

1. Introduction

Communication complexity, introduced by Yao [13], quantifies the number of bits that two or more players need to communicate amongst themselves in order to compute a function whose input is split between them. This fundamental complexity measure has repeatedly proven to be a very useful tool for proving lower bounds. Its applications touch upon a wide range of topics, such as circuit complexity, cell probe complexity, space-bounded computation, and (more recently) combinatorial auctions. A comprehensive treatment of the subject and its many applications can be found in the book of Kushilevitz and Nisan [10]; an overview of more recent research can be found in the survey by Babai [4].

In this paper we study the complexity of functions with a large number t of arguments. Consequently, the communication models of interest involve t players, each of whom holds one of the arguments.[1] In the general *multi-party communication model*, the players may exchange several messages, taking turns to speak in an arbitrary sequence. A predesignated player (w.l.o.g., player t) then announces the output of the function. A more restricted model is the *one-way com-*

* Work supported by NSF grant CCR-9987845.

† Work supported by Prof. Arora's David and Lucile Packard Fellowship, NSF grant CCR-0098180, and NSF ITR Grant CCR-0205594. Part of this work was done while visiting IBM, Almaden Research Center.

[1] It is very important to note that the models we work with differ from the well-studied "number on the forehead" models, in which player i holds all of the arguments *except* the i^{th} one.

munication model, which requires that players speak exactly once each, and in the order $1, 2, \dots, t$, with player i speaking only to player $i+1$ (for $i < t$) and player t announcing the output of the function. Here, we obtain improved lower bounds for the set disjointness problem in both these models.

The above models of communication turn out to be naturally related to the *data stream model*, which has been at the centre of much recent research [3, 9, 2] on algorithms for massive data sets. In this model, an algorithm is allowed to make a few passes (usually one pass) over its massively long input and has only a limited amount of workspace in which to store and manipulate parts of this input. The connection between space complexity in the data stream model and multi-party communication complexity has been observed earlier [3, 5]. In this paper we again exploit this connection to translate our improved communication lower bounds into data stream space lower bounds.

1.1. Our Results

In the multi-party set disjointness problem, each of the t players is given a set from the universe $[n] = \{1, 2, \dots, n\}$ together with a promise that the sets are either pairwise disjoint or are *uniquely intersecting*, i.e., they have one element in common but are otherwise disjoint. The players must distinguish between these two cases, using a possibly randomised protocol.

The above problem was first studied by Alon, Matias, and Szegedy [3] who proved a lower bound of $\Omega(n/t^4)$ on its communication complexity. They also showed, via a reduction, that this implies an $\Omega(n^{1-5/k})$ lower bound on the space complexity of approximating the k^{th} frequency moment of an input sequence in the data stream model.[2] The communication lower bound was subsequently improved to $\Omega(n/t^2)$ by Bar-Yossef, Jayram, Kumar, and Sivakumar [5]; for the special case of one-way communication it was improved to the nearly optimal $\Omega(\varepsilon^2 n/t^{1+\varepsilon})$.

In this paper, we prove a lower bound of $\Omega(n/(t \log t))$ for the communication problem in the general model, which beats all of the aforementioned bounds. In the one-way model we can further improve our bound to an optimal $\Omega(n/t)$. Both these results hold in the so-called *blackboard model* in which the players write their messages on a blackboard for all other players to see; in the one-way case, the players write messages strictly in the order $1, 2, \dots, t$.

Using the reduction of Alon, Matias, and Szegedy [3], we can then show that approximating the k^{th} frequency moment of a sequence of integers from $[m]$ has space complexity $\Omega(m^{1-2/k}/\log m)$ if the algorithm is allowed a constant number of passes over its input. The best result obtainable previously was $\Omega(m^{1-3/k})$. If the algorithm is allowed only one pass over its input, we again improve earlier results and obtain an $\Omega(m^{1-2/k})$ bound; this improvement is largely technical, since Bar-Yossef et al. were already able to establish a $\Omega(\varepsilon^2 m^{1-(2+\varepsilon)/k})$ bound. It is worth noting that the best upper bound for the problem in the one pass model is $\tilde{O}(m^{1-1/k})$ for $k \geq 3$ and $O(\log m)$ for $k \in \{0, 1, 2\}$.

1.2. Our Techniques

Our proofs make crucial use of the notion of *information cost* which was formally introduced by Chakrabarti, Shi, Wirth, and Yao [6], and is also implicit in recent work by Ablayev [1] and Saks and Sun [12]. More precisely, we use the extension and generalisation of this notion due to Bar-Yossef et al [5]. Roughly speaking, the information cost of a protocol is the amount of information that can be gleaned about the players' inputs by examining the *transcript* of the protocol, constituted by the players' messages. When defined appropriately, the corresponding complexity measure, known as *information complexity*, can be shown to lower bound the actual communication complexity and satisfy direct-sum like properties. The key insight in the work of Chakrabarti et al. [6] as well as Bar-Yossef et al. [5] can be summed up as follows:

> To lower bound the actual complexity of a complex "direct-sum like" problem, it suffices to lower bound the information complexity of a simpler "primitive" problem.

In this work, the complex problem we are concerned with is the set disjointness (promise) problem. Suppose we require the players to output 0 to indicate that their sets are disjoint and 1 to indicate that they

[2] If the item j occurs ν_j times in the input sequence, the k^{th} frequency moment is defined as $\sum_{j \in [m]} \nu_j^k$.

intersect. Viewing the input of each player as an n-bit string (the characteristic vector of his set), we see that the function to be computed is the OR of n bits, each of which is the AND of t bits. Accordingly, we are interested in the information complexity of the function AND_t in which t players are given one bit each and have to compute the AND of their bits.

At this point our techniques deviate considerably from earlier ones. Bar-Yossef et al. [5] relate the information cost of a protocol to certain well-studied distance measures between probability distributions; specifically, Hellinger and related distances. We use analytical properties of functions arising in the definition of information cost, together with the special combinatorial structure of communication protocols, to obtain a tighter tradeoff between the information cost and the error probability of a protocol. As observed in [5], the earlier lower bounds were not tight due to limitations of the properties of the statistical distance measures used therein. Thus, our direct analytical approach appears necessary.

1.3. Optimality

Our lower bound for the one-way model is optimal, and therefore, our lower bound for the general model is nearly optimal. This is because there is a blackboard one-way protocol for the problem that uses $O(n/t + t + \log n)$ bits of communication. Without the one-way restriction the same upper bound holds even in the private message (i.e., non-blackboard) model. We now outline a proof of these upper bounds.

A result due to Håstad and Wigderson [8] says that the two-party set disjointness problem with input sets of size at most k has a public coin randomised communication protocol with cost $O(k)$, independent of the size of the universe. Now, if the universe has size n, applying the private versus public coin theorem of Newman [11] gives us a private coin protocol with cost $O(k + \log n)$.

Let $k = 3n/t$. Because of the unique intersection promise we know that at least $t/2$ of the players have input sets of size at most k each; call these players "good." Using $O(t)$ bits of private message communication, two good players can identify one another. By the unique intersection promise, the desired output is 1 iff the sets of these two players intersect. Thus, the two players may run the Håstad-Wigderson protocol and communicate the outcome to player t who then announces the output. The cost of this protocol is clearly $O(t + k + \log n)$, as desired. The protocol can be made one-way by exploiting the structure of the Håstad-Wigderson protocol and involving all of the good players in it; we omit the details.

1.4. Organisation of the Paper

The remainder of this paper is organised as follows. In Section 2 we outline the decomposition paradigm which is central to our proof and formally state our results. We then present our proofs in the next two sections. Section 3 handles general protocols and proves a near-optimal lower bound, and Section 4 improves the bound to an optimal one for one-way protocols.

2. Information Cost and the Decomposition Paradigm

We now formally introduce the notion of information cost and briefly review the decomposition paradigm of Bar-Yossef et al [5]. Since the main focus of this paper is the particular problem of set disjointness, we sometimes sacrifice full generality for ease of presentation.

Conventions and Basic Definitions. We shall use boldface letters to denote vectors and capital letters to denote random variables. We use "log" to denote logarithm to the base 2 and "ln" to denote the natural logarithm. $\mathbb{R}_+$ shall denote the set of non-negative reals. For discrete distributions X, Y we let $\mathrm{H}(X)$ denote the entropy of X and $\mathrm{I}(X : Y)$ the mutual information between X and Y. We recall that $\mathrm{I}(X : Y) = \mathrm{H}(X) - \mathrm{H}(X \mid Y) = \mathrm{H}(Y) - \mathrm{H}(Y \mid X)$. See [7] for further background.

We are interested in the computation of functions of the form $\phi_n \circ \psi_t$ where $\phi_n : \{0,1\}^n \to \{0,1\}$ and $\psi_t : \{0,1\}^t \to \{0,1\}$ are arbitrary Boolean functions. To be precise, we have n vectors $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(n)} \in \{0,1\}^t$ and t players; player i receives the i^{th} entry of each vector. Their goal is to compute $\phi_n\big(\psi_t(\mathbf{a}^{(1)}), \ldots, \psi_t(\mathbf{a}^{(n)})\big)$. In the case of the set disjointness problem, $\phi_n = \text{OR}_n$ and $\psi_t = \text{AND}_t$. We shall make the simplifying assumption that ϕ_n is a symmetric Boolean function.

Since we wish to prove lower bounds for *promise* problems, $(\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(n)})$ is required to lie in some restricted subset $\mathcal{L}_n \subseteq \{0,1\}^{nt}$ and thus, for symmetric ϕ_n, each $\mathbf{a}^{(j)}$ is required to lie in a set $\mathcal{L} \subseteq \{0,1\}^t$ of legal inputs. For the set disjointness problem with the unique intersection promise, $\mathcal{L}$ consists of the all-zeroes input $\mathbf{0}$, the all-ones input $\mathbf{1}$, and the unit vectors $\mathbf{e}_i$ (with a "1" in the i^{th} position) for all $i \in [t]$.

Let Π be a δ-error randomised protocol for $\phi_n \circ \psi_t$. We denote by $\Pi(\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(n)})$ the transcript of Π on input $(\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(n)})$; notice that this transcript is a random variable.

Definition 2.1 (Information cost and complexity) Let μ be a distribution on $\mathcal{L}$. The μ-information cost of Π, denoted $\operatorname{icost}_\mu(\Pi)$, is defined as $\mathrm{I}(\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(n)} : \Pi(\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(n)}))$ where the random vectors $\mathbf{A}^{(j)}$ are drawn independently from μ. The δ-error μ-information complexity of $\phi_n \circ \psi_t$, denoted $\mathrm{IC}_{\mu,\delta}(\phi_n \circ \psi_t)$, is defined to be the minimum of $\operatorname{icost}_\mu(\Pi)$ over all δ-error protocols Π for $\phi_n \circ \psi_t$.

For the decomposition paradigm, it is necessary to consider the following more general notion.

Definition 2.2 (Conditional information cost) Let D be a random variable. The D-conditional μ-information cost of Π, denoted $\operatorname{icost}_\mu(\Pi \mid D)$, is defined as $\mathrm{I}(\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(n)} : \Pi(\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(n)}) \mid D^n)$ where the $\mathbf{A}^{(j)}$ are as above.[3] The corresponding complexity measure, denoted $\mathrm{IC}_{\mu,\delta}(\phi_n \circ \psi_t \mid D)$, is defined in the obvious way.

We would like to relate the δ-error randomised communication complexity, $\mathrm{C}_\delta(\phi_n \circ \psi_t)$, to the conditional information complexity of the simpler function ψ_t with respect to some appropriate distributions. To enable this, μ and D are required to satisfy the following properties:

- μ must be a "collapsing" distribution for $\phi_n \circ \psi_t$. This means that for every $\mathbf{a} \in \mathcal{L}$, and all $\mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(n)}$ in the support of μ, we have $\phi_n(\psi_t(\mathbf{a}), \psi_t(\mathbf{a}^{(2)}), \ldots, \psi_t(\mathbf{a}^{(n)})) = \psi_t(\mathbf{a})$.
- D must "partition" μ. This means that for every d in the support of D, the conditional distribution $(\mu \mid D = d)$ must be a product distribution on $\{0,1\}^t$.

We can now state the direct sum theorem that is the basis of the decomposition paradigm.

Theorem 2.3 (Bar-Yossef et al. [5]) *Let μ be a collapsing distribution for $\phi_n \circ \psi_t$ and let D be a random variable that partitions μ. Then $\mathrm{C}_\delta(\phi_n \circ \psi_t) \geq \mathrm{IC}_{\mu,\delta}(\phi_n \circ \psi_t \mid D) \geq n \cdot \mathrm{IC}_{\mu,\delta}(\psi_t \mid D)$. The same holds for one-way protocols.*

For the particular case of the disjointness problem, let D be a uniform random integer from $[t]$ and let μ be distributed uniformly on $\{\mathbf{0}, \mathbf{e}_D\}$. It is easy to verify that this choice of distributions satisfies the required conditions, and so we have:

Corollary 2.4 *The set disjointness problem with the unique intersection promise, denoted $\mathrm{DISJ}_{n,t}$, satisfies $\mathrm{C}_\delta(\mathrm{DISJ}_{n,t}) \geq n \cdot \mathrm{IC}_{\mu,\delta}(\mathrm{AND}_t \mid D)$.*

Accordingly, we direct our efforts towards studying the one-bit function AND_t.

Communication Complexity Results. In the language just developed, our results for communication complexity can be stated as follows.

Theorem 2.5 *With μ and D as defined above, and any $\delta > 0$,*

$$\mathrm{IC}_{\mu,\delta}(\mathrm{AND}_t \mid D) \quad = \quad \Omega\left(\frac{1}{t \log t}\right).$$

Consequently, $\mathrm{C}_\delta(\mathrm{DISJ}_{n,t}) = \Omega(n/(t \log t))$.

When protocols are restricted to be one-way, we can tighten our analysis by making use of the additional combinatorial properties that result from this restriction. This yields:

Theorem 2.6 *With μ and D as above, and any $\delta > 0$,*

$$\mathrm{IC}^{\text{1-WAY}}_{\mu,\delta}(\mathrm{AND}_t \mid D) \quad = \quad \Omega\left(\frac{1}{t}\right).$$

Consequently, $\mathrm{C}^{\text{1-WAY}}_\delta(\mathrm{DISJ}_{n,t}) = \Omega(n/t)$.

[3] Our notation differs slightly from that of Bar-Yossef et al. [5] who would use something like $\operatorname{icost}_{\mu^n}(\Pi \mid D^n)$.

Results for the Data Stream Model. Let F_k^m denote the problem of computing, to within $(1 \pm \varepsilon)$ approximation, the k^{th} frequency moment of a sequence of items, each from $[m]$. By suitably modifying a reduction due to Alon, Matias, and Szegedy [3], Bar-Yossef et al. [5] show that a space efficient data stream algorithm for F_k^m can be used to construct an efficient t-party communication protocol for $\mathrm{DISJ}_{m,t}$ for an appropriately chosen t. This construction is such that the resulting protocol is one-way if the data stream algorithm makes only a single pass over its input. Using this construction and Theorems 2.5 and 2.6 we obtain the following results (we omit the quantitative details of the proofs):

Theorem 2.7 *Any ε-approximate randomised data stream algorithm for F_k^m that works with confidence $\geq \frac{3}{4}$ and makes a constant number of passes over its input requires $\Omega\left(\frac{m^{1-2/k}}{(1+3\varepsilon)^{2/k}\log m}\right)$ space. If the algorithm makes only one pass over its input, it requires $\Omega\left(\frac{m^{1-2/k}}{(1+3\varepsilon)^{2/k}}\right)$ space.*

3. The Information Complexity of AND_t

We now turn to the study of communication protocols for the function AND_t. This specific function itself plays very little role in the analysis in this section, though we do care that $\mathrm{AND}_t(\mathbf{0}) \neq \mathrm{AND}_t(\mathbf{1})$. Thus, the lower bound on the D-conditional μ-information complexity that we obtain here holds for a very general class of functions. However, this generalisation does not extend easily to the communication complexity world, and that is because the particular distributions D and μ used here were custom-made for the $\mathrm{DISJ}_{n,t}$ problem. Recall that D is a random integer chosen uniformly from $[t]$ and that μ is a random vector in $\{0,1\}^t$ chosen uniformly from $\{\mathbf{0}, \mathbf{e}_D\}$.

Let Π be a t-party randomised communication protocol for AND_t. Recall that $\Pi(\mathbf{a})$ denotes the (random) transcript of the protocol on input $\mathbf{a} \in \{0,1\}^t$. Let U be the set of all transcripts that are produced with positive probability on input $\mathbf{0}$, for which the protocol outputs "0". Let V be the corresponding set for output "1". For each transcript $\pi \in U \cup V$ and each $i \in [t]$, we define

$$\begin{aligned} p_\pi &= \Pr[\Pi(\mathbf{0}) = \pi]\,, \\ x_{i\pi} &= \Pr[\Pi(\mathbf{e}_i) = \pi]/p_\pi\,. \end{aligned}$$

The main combinatorial property of Π that we need is captured by the next lemma and its corollary.

Lemma 3.1 *Let $I \subseteq [t]$ be a nonempty set and let $\boldsymbol{\chi}_I \in \{0,1\}^t$ be its characteristic vector. For any transcript $\pi \in U \cup V$, $\Pr[\Pi(\boldsymbol{\chi}_I) = \pi] = p_\pi \prod_{i \in I} x_{i\pi}$.*

Proof (sketch): It is well known that deterministic communication protocols satisfy the following "rectangle" property: the set of all inputs that produce a fixed transcript form a combinatorial rectangle (see, e.g., [10, Ch. 1]). A randomised protocol can be thought of as a deterministic protocol in which each player's input is his original (one-bit) input plus the (multi-bit) contents of his random tape. With these observations, the statement of the lemma reduces to a simple combinatorial fact. ∎

Corollary 3.2 *For any nonempty $I \subseteq [t]$,*

$$\sum_{\pi \in U \cup V} p_\pi \prod_{i \in I} x_{i\pi} \leq 1\,. \tag{1}$$

Let us now define

$$\begin{aligned} \kappa &= t \cdot \mathrm{icost}_\mu(\Pi \mid D)\,; \\ \delta_0 &= \Pr[\Pi(\mathbf{0}) \in V] = \sum_{\pi \in V} p_\pi \qquad (2) \\ \delta_1 &= \Pr[\Pi(\mathbf{1}) \in U] = \sum_{\pi \in U} p_\pi \prod_{i=1}^{t} x_{i\pi}\,. \qquad (3) \end{aligned}$$

It is clear that $\max\{\delta_0, \delta_1\}$ lower bounds the error probability of Π. Thus, our goal is to prove that $\kappa = \Omega(1/\log t)$ for sufficiently small constants δ_0 and δ_1.

3.1. An Analytic Formulation

By definition of information cost, $\kappa = \sum_{i=1}^{t} \mathrm{I}(\mathbf{A} : \Pi(\mathbf{A}) \mid D = i) = \sum_{i=1}^{t} \mathrm{I}(\mathbf{A}_i : \Pi(\mathbf{A}_i))$, where $\mathbf{A}$ is drawn from μ and $\mathbf{A}_i$ is uniformly distributed on

$\{\mathbf{0}, \mathbf{e}_i\}$. Clearly, $\mathrm{H}(\mathbf{A}_i) = 1$, and it is not hard to verify that for any $\pi \in U \cup V$

$$\mathrm{H}(\mathbf{A}_i \mid \Pi(\mathbf{A}_i) = \pi) \quad = \quad h\left(\frac{1}{1 + x_{i\pi}}\right),$$

where h is the entropy function for two-valued distributions: $h(x) = -x\log x - (1-x)\log(1-x)$. Moreover, $\Pr[\mathbf{A}_i = \pi] = \frac{1}{2}p_\pi(1+x_{i\pi})$. Putting these facts together gives us the following analytic expression for κ:

$$\kappa \quad = \quad \sum_{\pi \in U \cup V} \sum_{i=1}^{t} p_\pi g(x_{i\pi}), \tag{4}$$

where the function $g : \mathbb{R}_+ \to \mathbb{R}$ is given by

$$g(x) \quad = \quad 1 - \frac{1+x}{2} \cdot h\left(\frac{1}{1+x}\right). \tag{5}$$

It will also be useful to define two auxiliary functions: $g_+(x) = g(x) - \frac{1}{2}(1-x)$ and $f(x) = 2g(x) + 2g_+(x)$. The following lemma collects together several useful facts about these functions which we shall repeatedly use.

Lemma 3.3 *The functions f, g, and g_+ satisfy the following properties:*
(i) *g_+ is nonnegative, decreasing in $[0,1]$, increasing in $[1,\infty)$, and convex in $[0,\infty)$.*
(ii) *$1 - x - g(x) < 4g_+(x)$, for $x \in [0, \frac{1}{2}]$.*
(iii) *$-g(x) < 4g_+(x)$, for $x \in [2,\infty)$.*
(iv) *$f(x) + \ln x \geq 0$, for $x \in [\frac{1}{2}, 2]$.*
(v) *f is decreasing in $[0,2]$.*

Proof : All of these facts can be proved via simple calculus. We omit the details. ∎

We note that replacing g with g_+ in (4) yields an expression which lower bounds κ:

$$\kappa \quad \geq \quad \sum_{\pi \in U \cup V} \sum_{i=1}^{t} p_\pi g_+(x_{i\pi}). \tag{6}$$

This is because $g(x) - g_+(x) = \frac{1}{2}(1-x)$ and, by a special case of (1), $\sum_\pi p_\pi x_{i\pi} \leq 1$. Combining this with (4) gives

$$\sum_{\pi \in U \cup V} \sum_{i=1}^{t} p_\pi f(x_{i\pi}) \quad \leq \quad 4\kappa, \tag{7}$$

3.2. Balancing

We call a protocol *balanced* if for each π and each i, $x_{i\pi} \in \{0\} \cup [\frac{1}{2}, 2]$. Intuitively, a protocol is balanced if any transcript which is produced on inputs $\mathbf{0}$ as well as $\mathbf{e}_i$ is "almost" equally likely to be produced on each of these inputs.

Lemma 3.4 *For every protocol Π there is a balanced protocol Π' which computes the same function as Π and satisfies $\mathrm{icost}_\mu(\Pi' \mid D) \leq 5 \cdot \mathrm{icost}_\mu(\Pi \mid D)$. Furthermore, if Π is one-way, then Π' can be designed to be one-way.*

Proof : If, for some π and i, $0 < x_{i\pi} < \frac{1}{2}$, then we can split π into two transcripts π_0 and π_1 such that π_0 is produced only on input $\mathbf{0}$ and π_1 is equally likely to be produced on inputs $\mathbf{0}$ and $\mathbf{e}_i$. The probability p_π is split amongst p_{π_0} and p_{π_1} as follows: $p_{\pi_0} = (1 - x_{i\pi})p_\pi$; $p_{\pi_1} = x_{i\pi}p_\pi$. For the new transcripts, $x_{i\pi_0} = 0$ and $x_{i\pi} = 1$. Such a split can be effected by appending to the existing protocol an extra message from player i that distinguishes π_0 from π_i. We have

$$\begin{aligned} & p_{\pi_0} g(x_{i\pi_0}) + p_{\pi_1} g(x_{i\pi_1}) - p_\pi g(x_{i\pi}) \\ = \quad & p_\pi(1 - x_{i\pi} - g(x_{i\pi})) \\ < \quad & 4p_\pi g_+(x_{i\pi}). \end{aligned}$$

Likewise, if, for some π and i, $x_{i\pi} > 2$, then we can split π into two transcripts π_0 and π_1 such that π_1 is produced only on input $\mathbf{e}_i$, π_0 is equally likely to be produced on inputs $\mathbf{0}$ and $\mathbf{e}_i$, $p_{\pi_0} = p_\pi$, $p_{\pi_1} = 0$, and $x_{i\pi_0} = 1$. We have

$$\begin{aligned} & p_{\pi_0} g(x_{i\pi_0}) + p_{\pi_1} g(x_{i\pi_1}) - p_\pi g(x_{i\pi}) \\ = \quad & -p_\pi g(x_{i\pi}) \\ < \quad & 4p_\pi g_+(x_{i\pi}). \end{aligned}$$

Repeatedly perform these two operations until the original protocol Π becomes a balanced protocol Π'. Letting $\kappa' = t \cdot \mathrm{icost}_\mu(\Pi' \mid D)$, we see that

$$\begin{aligned} \kappa' - \kappa \quad & < \quad \textstyle\sum_\pi \sum_{i=1, x_{i\pi} \in (0,\frac{1}{2}) \cup (2,\infty)}^{t} 4p_\pi g_+(x_{i\pi}) \\ & \leq \quad \sum_\pi \sum_{i=1}^{t} 4p_\pi g_+(x_{i\pi}) \\ & \leq \quad 4\kappa, \end{aligned}$$

where the final inequality follows from (6).

To prove a similar statement for one-way protocols we have to be more careful about how we split transcripts. Let π be a transcript and let i be the smallest index such that $x_{i\pi} \notin \{0\} \cup [\frac{1}{2}, 2]$. Since the protocol is one-way, π is simply a sequence $(u_1, \ldots, u_t)$, where u_i is player i's message. We split this *message* into two new ones, thereby simultaneously splitting all transcripts of the form $(u_1, \ldots, u_i, *, \ldots, *)$. We perform this operation repeatedly until the protocol becomes balanced. A calculation almost identical to the above completes the proof. ∎

In view of the above lemma, we shall henceforth assume that our protocol Π is balanced.

3.3. The Proof for Balanced Protocols

Consider, for a moment, the case when Π satisfies an additional nice property: $x_{i\pi}$ is never zero. Then, since Π is balanced, Part (iv) of Lemma 3.3 allows us to lower bound $f(x_{i\pi})$ by $-\ln x_{i\pi}$ for each i and π. By (7), an appropriate weighted sum of $f(x_{i\pi})$ lower bounds κ. On the other hand, by (3), the same weighted sum of $-\ln x_{i\pi}$ upper bounds $-\ln \delta_1$. Therefore, κ would be lower bounded by a constant!

The analysis below is based on this idea, but unfortunately the presence of zeroes among the $x_{i\pi}$ causes considerable complication. In particular, we are only able to show the slightly weaker result that $\kappa = \Omega(1/\log t)$.

For a transcript $\pi \in U \cup V$, define $F(\pi) = \sum_{i=1}^{t} f(x_{i\pi})$. We further partition the set U of transcripts as follows: $U = U' \cup U_0 \cup \cdots \cup U_t$, where

$$U' = \left\{ \begin{array}{c} \pi \in U : F(\pi) \geq 0 \text{ and } x_{i\pi} = 0 \\ \text{for some index } i \end{array} \right\},$$

$$U_0 = \{\pi \in U : \forall i\, (x_{i\pi} \neq 0)\},$$

$$U_r = \left\{ \begin{array}{c} \pi \in U : F(\pi) < 0 \text{ and } x_{i\pi} = 0 \\ \text{for exactly } r \text{ indices } i \end{array} \right\}, \quad \text{for } 1 \leq r \leq t\,.$$

Define $q' = \sum_{\pi \in U'} p_\pi$ and $q_r = \sum_{\pi \in U_r} p_\pi$ for all $r \geq 0$. The next lemma limits the profusion of zeroes amongst the $x_{i\pi}$.

Lemma 3.5 *The sets U_r and U' satisfy:*
(i) $\sum_{r=1}^{t} r q_r \leq 2\kappa$.
(ii) $q' + \sum_{r=1}^{t} q_r \leq 2\kappa$.
(iii) *U_r is nonempty only if $r \leq \frac{t}{5}$.*

Proof : Recall that g_+ is a nonnegative function (Lemma 3.3, Part (i)). Therefore, for any $\pi \in U_r$, $\sum_{i=1}^{t} g_+(x_{i\pi}) \geq r g_+(0) = \frac{r}{2}$ and for any $\pi \in U'$, $\sum_{i=1}^{t} g_+(x_{i\pi}) \geq g_+(0) = \frac{1}{2}$. Using (6), we now easily derive Parts (i) and (ii). For Part (iii), note that f is decreasing in $[\frac{1}{2}, 2]$; thus, for $\pi \in U_r$, $0 > F(\pi) \geq r f(0) + (t-r) f(2)$, which easily yields $r \leq \frac{t}{5}$. ∎

Notice that we may assume that $\delta_0 \leq \frac{1}{4}$ and that $\kappa \leq \frac{1}{8}$ since otherwise there is nothing to prove. Then, we have

$$\begin{aligned} q_0 &= 1 - q' - \textstyle\sum_{r=1}^{t} q_r - \sum_{\pi \in V} p_\pi \\ &\geq 1 - 2\kappa - \delta_0 \\ &\geq \tfrac{1}{2}\,. \end{aligned} \tag{8}$$

Using, in succession, (3), (1), Jensen's inequality for the function "ln", (8), Lemma 3.3 (iv), the definition of U', the definition of U_r, and (7), we derive

$$\begin{aligned} -\ln \delta_1 &= -\ln \sum_{\pi \in U_0} p_\pi \prod_{i=1}^{t} x_{i\pi} \\ &\leq -\ln \sum_{\pi \in U_0} p_\pi \prod_{i=1}^{t} x_{i\pi} - \frac{\delta_0}{q_0} \cdot \ln \sum_{\pi \in V} p_\pi \prod_{i=1}^{t} x_{i\pi} \\ &\leq -\frac{1}{q_0} \sum_{\pi \in U_0 \cup V} p_\pi \sum_{i=1}^{t} \ln x_{i\pi} - \ln q_0 - \frac{\delta_0}{q_0} \ln \delta_0 \\ &\leq 2 - 2 \sum_{\pi \in U_0 \cup V} p_\pi \sum_{i=1}^{t} \ln x_{i\pi} \\ &\leq 2 + 2 \sum_{\pi \in U_0 \cup V} p_\pi F(\pi) \\ &\leq 2 + 2 \sum_{\pi \in U_0 \cup V \cup U'} p_\pi F(\pi) \\ &= 2 + 2 \sum_{\pi \in U \cup V} p_\pi F(\pi) - 2 \sum_{r=1}^{t} \sum_{\pi \in U_r} p_\pi F(\pi) \\ &\leq 8\kappa + 2 - 2 \sum_{r=1}^{t} \sum_{\pi \in U_r} p_\pi F(\pi)\,. \end{aligned} \tag{9}$$

We shall now bound the latter sum from below. Fix an r such that U_r is nonempty. Since f is decreasing in $[0,2]$, lower bounding $F(\pi)$ for $\pi \in U_r$ roughly corresponds to upper bounding the quantities $x_{i\pi}$. Inequality (1), based on the combinatorial structure of the protocol, provides a tool for doing this, but due to the presence of zeroes amongst the $x_{i\pi}$, we need to apply it with a carefully chosen set $I \subseteq [t]$ which avoids most of these zeroes. In the sequel, we shall *average* over all I of a particular size s, where s depends on r.

Let $s = \lfloor \frac{t}{5r} \rfloor$. Since U_r is nonempty, Lemma 3.5 assures us that $s \geq 1$. Let us define the sets $U_r\langle I\rangle = \{\pi \in U_r : \prod_{i\in I} x_{i\pi} \neq 0\}$ and the quantities $q_r\langle I\rangle = \sum_{\pi\in U_r\langle I\rangle} p_\pi$. We apply Lemma 3.3, Part (iv), and Jensen's inequality to obtain

$$\begin{aligned}
S_r &:= \sum_{I\subseteq[t],|I|=s} \sum_{\pi\in U_r\langle I\rangle} \sum_{i\in I} p_\pi f(x_{i\pi}) \\
&\geq \sum_{I\subseteq[t],|I|=s} \sum_{\pi\in U_r\langle I\rangle} \sum_{i\in I} p_\pi \cdot (-\ln x_{i\pi}) \\
&= -\sum_{I\subseteq[t],|I|=s} q_r\langle I\rangle \sum_{\pi\in U_r\langle I\rangle} \frac{p_\pi}{q_r\langle I\rangle} \ln \prod_{i\in I} x_{i\pi} \\
&\geq -\sum_{I\subseteq[t],|I|=s} q_r\langle I\rangle \cdot \ln \sum_{\pi\in U_r\langle I\rangle} p_\pi \prod_{i\in I} x_{i\pi} \\
&\quad + \sum_{I\subseteq[t],|I|=s} q_r\langle I\rangle \ln q_r\langle I\rangle \\
&\geq \sum_{I\subseteq[t],|I|=s} q_r\langle I\rangle \ln q_r\langle I\rangle \,, \qquad (10)
\end{aligned}$$

where the final inequality follows by applying (1) for all the I's.

Consider a $\pi \in U_r$ and an $i \in [t]$ such that $x_{i\pi} \neq 0$. In the expression that defines S_r, the term $p_\pi f(x_{i\pi})$ appears exactly as many times as there are sets I which contain i but do not contain any j such that $x_{j\pi} = 0$. There are precisely r such j's, and so this latter number is $\binom{t-r-1}{s-1}$. Therefore,

$$\begin{aligned}
S_r &= \binom{t-r-1}{s-1} \sum_{\pi\in U_r} \sum_{i=1,x_{i\pi}\neq 0}^{t} p_\pi f(x_{i\pi}) \\
&\leq \binom{t-r-1}{s-1} \sum_{\pi\in U_r} F(\pi) \,, \qquad (11)
\end{aligned}$$

where the inequality holds because $f(0) > 0$. A similar counting argument yields

$$\begin{aligned}
\sum_{I\subseteq[t],|I|=s} q_r\langle I\rangle &= \binom{t-r-1}{s-1} \cdot \frac{t-r}{s} \cdot q_r \\
&\geq \binom{t-r-1}{s-1} \cdot \frac{tq_r}{2s} \,, \qquad (12)
\end{aligned}$$

since, by Lemma 3.5, $r \leq \frac{t}{5} \leq \frac{t}{2}$. Continuing from (10), we have

$$\begin{aligned}
S_r &\geq \left(\textstyle\sum_{I\subseteq[t],|I|=s} q_r\langle I\rangle\right) \cdot \ln \frac{\sum_{I\subseteq[t],|I|=s} q_r\langle I\rangle}{\binom{t}{s}} \\
&\geq \tbinom{t-r-1}{s-1} \tfrac{tq_r}{2s} \cdot \ln\left(\tfrac{q_r}{2} \cdot \tfrac{t-r-1}{t-1} \cdot \tfrac{t-r-2}{t-2} \cdots \tfrac{t-r-s+1}{t-s+1}\right) \\
&\geq \tbinom{t-r-1}{s-1} \tfrac{tq_r}{2s} \left(\ln q_r - \ln 2 + \ln\left(1 - \tfrac{r}{t-s+1}\right)^s\right) \\
&\geq \tbinom{t-r-1}{s-1} \tfrac{tq_r}{2s} \left(\ln q_r - \ln 2 + s\ln\left(1 - \tfrac{2r}{t}\right)\right) \\
&\geq \tbinom{t-r-1}{s-1} \tfrac{tq_r}{2s} \left(\ln q_r - \ln 2 - \tfrac{4rs}{t}\right) \\
&\geq \tbinom{t-r-1}{s-1} \cdot \left(\tfrac{5}{2} r q_r \ln q_r - 4 r q_r\right) \,,
\end{aligned}$$

where the first step follows from Jensen's inequality and the fourth follows from $s = \lfloor \frac{t}{5r} \rfloor$. Combining this with (11) yields

$$\sum_{\pi\in U_r} F(\pi) \geq \frac{5}{2} r q_r \ln q_r - 4 r q_r \,.$$

We now sum these inequalities over all r. Lemma 3.5 tells us that $\sum_{r=1}^{t} r q_r \leq 2\kappa$. On the other hand, by convexity of $(x \mapsto x \ln x)$,

$$\begin{aligned}
\sum_{r=1}^{t} r q_r \ln q_r &\geq \left(\sum_{r=1}^{t} r q_r\right) \cdot \ln \frac{\sum_{r=1}^{t} r q_r}{\sum_{r=1}^{t} r} \\
&\geq 2\kappa \ln(2\kappa) - 2\kappa \cdot \ln \sum_{r=1}^{t} r \\
&\geq -\frac{1}{e} - 4\kappa \ln t \,.
\end{aligned}$$

Therefore,

$$\sum_{r=1}^{t} \sum_{\pi\in U_r} p_\pi F(\pi) \geq -2 - 10\kappa \ln t - 8\kappa$$

Continuing from (9), we get

$$-\ln \delta_1 \leq 20\kappa \ln t + 24\kappa + 6$$

and, for δ_1 small enough, this gives $\kappa = \Omega(1/\log t)$. We have shown that for any protocol Π with error probability δ less than a sufficiently small constant,

$$\mathrm{icost}_{\mu}(\Pi \,|\, D) \;=\; \Omega\left(\frac{1}{\log t}\right)$$

and this completes the proof of Theorem 2.5.

4. A Tight Bound for One-Way Protocols

One-way communication protocols have special structural properties that are stronger than the "rectangle" property which was crucial in our arguments in Section 3. In particular, a one-way t-player protocol has a recursive structure given by the strategy of the first player and a collection of $(t-1)$-player protocols, one for each possible message of the first player. Moreover, this recursive structure gives us recursive formulae for the information cost as well as the error probability of the original protocol in terms of those of the sub-protocols. Using analytic tools (in particular, a novel inequality for probability distributions) we can then process these recursive formulae into an inductive proof of an $\Omega(1/t)$ lower bound on the information cost of the original protocol.

Fix a one-way protocol Π for AND_t. Let $M(a)$ denote the message of player 1 on input $a \in \{0,1\}$; note that $M(a)$ is a random variable. For each possible message u, let p_u (resp. q_u) be the probability that player 1 produces u on input 0 (resp. 1), and let Π_u be the $(t-1)$-player protocol that the rest of the players execute upon seeing u. Let P (resp. Q) denote the probability distributions given by $\{p_u\}$ (resp. $\{q_u\}$). The distributions P and Q and the collection $\{\Pi_u\}$ of $(t-1)$-player protocols completely specify Π. To finish this recursive description, we note that a 1-player protocol Π simply consists of the sole player announcing the protocol's output; it is therefore given by two pairs of probabilities $(p_0, q_0), (p_1, q_1)$ — with $p_0 + p_1 = q_0 + q_1 = 1$ — corresponding to the two messages "0" (the protocol outputs "0") and "1" (it outputs "1").

Let μ and D be the distributions specified in Section 2; we write them as μ_t and D_t to emphasise their dependence on t. Defining $\kappa(\Pi) = t \cdot \mathrm{icost}_{\mu_t}(\Pi \,|\, D_t)$ for a t-player protocol, we have:

Lemma 4.1 *Let A denote a uniform random bit. Then*

$$\begin{aligned}\kappa(\Pi) &= \mathrm{I}(A : M(A)) + \sum_u p_u \kappa(\Pi_u) \\ &= \overline{\mathrm{D}}(P, Q) + \sum_u p_u \kappa(\Pi_u)\end{aligned}$$

where $\overline{\mathrm{D}}(P,Q) = \sum_u \frac{p_u + q_u}{2}\left(1 - h\left(\frac{p_u}{p_u+q_u}\right)\right)$ is the Jensen-Shannon divergence between the distributions P and Q.

Proof : This follows in a straightforward fashion from the definitions. We omit the details. ∎

Let $\delta_0(\Pi)$ be the probability that Π outputs "1" on input $\mathbf{0}$, and let $\delta_1(\Pi)$ be the probability that it outputs "0" on input $\mathbf{1}$. In case Π is a protocol for AND_t, $\max\{\delta_0(\Pi), \delta_1(\Pi)\}$ lower bounds the error probability of Π. The following lemma, whose easy proof we omit, provides recursive formulae for these quantities:

Lemma 4.2 *The error probabilities of* Π *are given by*
(i) $\delta_0(\Pi) = \sum_u p_u \delta_0(\Pi_u)$, *and*
(ii) $\delta_1(\Pi) = \sum_u q_u \delta_1(\Pi_u)$. ∎

Let $g_+ : \mathbb{R}_+ \to \mathbb{R}$ be the function defined in Section 3.1. The following technical inequality is the key to our improved lower bound.

Lemma 4.3 *Suppose $P = (p_1, \ldots, p_r)$ and $Q = (q_1, \ldots, q_r)$ are discrete probability distributions such that for all u, we have $p_u q_u \neq 0$, and when $q_u \neq 0$, we have $p_u/q_u \leq 2$. Then, for any $\delta_1, \ldots, \delta_r \in [0,1]$, we have*

$$\overline{\mathrm{D}}(P,Q) + \textstyle\sum_{u=1}^{r} p_u g_+(\delta_u) \;\;\geq\;\; g_+\left(\sum_{u=1}^{r} q_u \delta_u\right) .$$

Proof : It will be necessary to treat messages u with $q_u = 0$ separately. Assume w.l.o.g. that $q_1 = \cdots = q_k = 0$ and that $q_u \neq 0$ for $u > k$; if no q_u is zero we put $k = 0$. For $u > k$ put $\lambda_u = p_u/q_u$; then $\lambda_u \leq 2$ whenever it is defined.

Consider the function $G : \mathbb{R}^2 \to \mathbb{R}$ given by $G(\lambda, \delta) = g_+(\lambda) + \lambda g_+(\delta)$. We claim that its Hessian

$$\nabla^2 G(\lambda, \delta) \;=\; \begin{bmatrix} g_+''(\lambda) & g_+'(\delta) \\ g_+'(\delta) & \lambda g_+''(\delta) \end{bmatrix}$$

is positive semidefinite in the rectangle $[0,2] \times [0,1]$. Indeed, a direct computation of derivatives gives

$g'_+(x) = \frac{1}{2}\left(1+\log\left(\frac{x}{x+1}\right)\right)$ and $g''_+(x) = ((2\ln 2)\cdot x(x+1))^{-1}$. So the trace of $\nabla^2 G$ is positive in $\mathbb{R}^2_+$ and the determinant is positive whenever

$$\frac{1}{\lambda+1} \quad \geq \quad (2\ln 2)^2 \cdot \delta(\delta+1) g'_+(\delta)^2 .$$

In the rectangle $[0,2]\times[0,1]$, the left hand side is at least $\frac{1}{3}$ whereas straightforward calculus shows that the right hand side is strictly less than $\frac{1}{3}$. This proves the claim. It follows that G is convex in the above rectangle.

Let $\hat{\delta} = \sum_{u=1}^{r} q_u\delta_u = \sum_{u=k+1}^{r} q_u\delta_u$ and $\hat{p} = \sum_{u=1}^{k} p_u$. Using the properties of g_+ from Lemma 3.3, and noting that $\hat{\delta}\leq 1$, we have

$$\begin{aligned}
\overline{\mathrm{D}}(P,Q) &= \sum_{u=1}^{k} p_u g_+(q_u/p_u) + \sum_{u=k+1}^{r} q_u g_+(p_u/q_u) \\
&\geq \hat{p}\cdot g_+(0) + \sum_{u=k+1}^{r} q_u g_+(\lambda_u) \\
&\geq \hat{p}\cdot g_+(\hat{\delta}) + \sum_{u=k+1}^{r} q_u g_+(\lambda_u) .
\end{aligned}$$

Now, using Jensen's inequality for the function G, we have

$$\begin{aligned}
&\overline{\mathrm{D}}(P,Q) + \sum_{u=1}^{r} p_u g_+(\delta_u) \\
&\geq \hat{p}\cdot g_+(\hat{\delta}) + \sum_{u=k+1}^{r} q_u(g_+(\lambda_u) + \lambda_u g_+(\delta_u)) \\
&= \hat{p}\cdot g_+(\hat{\delta}) + \sum_{u=k+1}^{r} q_u G(\lambda_u,\delta_u) \\
&\geq \hat{p}\cdot g_+(\hat{\delta}) + G\left(\sum_{u=k+1}^{r} q_u\lambda_u, \sum_{u=k+1}^{r} q_u\delta_u\right) \\
&= \hat{p}\cdot g_+(\hat{\delta}) + G(1-\hat{p},\hat{\delta}) \\
&\geq \hat{p}\cdot g_+(\hat{\delta}) + (1-\hat{p})\cdot g_+(\hat{\delta}) ,
\end{aligned}$$

which establishes the desired inequality. ∎

Lemma 4.4 *For a small enough error probability δ, we have $\kappa(\Pi) = \Omega(1)$.*

Proof : By Lemma 3.4, we may assume that Π is a balanced one-way protocol. Therefore, the probability distributions P and Q that define the strategy of player 1 in Π satisfy the conditions of Lemma 4.3. We shall use that lemma to prove the following claim:

$$\kappa(\Pi) + \frac{\delta_0(\Pi)}{2} \quad \geq \quad g_+(\delta_1(\Pi)) . \tag{13}$$

We prove our claim by induction on t. In the base case we have a 1-player protocol given by pairs (p_0,q_0) and (p_1,q_1) as described earlier. Then, $\kappa(\Pi) = \overline{\mathrm{D}}((p_0,p_1),(q_0,q_1)) \geq p_0 g_+(q_0/p_0)$; $\delta_0(\Pi) = p_1 = 1-p_0$ and $\delta_1(\Pi) = q_0$. Since $g_+(0) = \frac{1}{2}$, the claim now reduces to

$$p_0 g_+(q_0/p_0) + (1-p_0)g_+(0) \quad \geq \quad g_+(q_0) ,$$

which is immediate from the convexity of g_+.

For the inductive step, we have

$$\begin{aligned}
&\kappa(\Pi) + \frac{1}{2}\delta_0(\Pi) \\
&= \overline{\mathrm{D}}(P,Q) + \sum_u p_u\left(\kappa(\Pi_u) + \frac{1}{2}\delta_0(\Pi_u)\right) \\
&\geq \overline{\mathrm{D}}(P,Q) + \sum_u p_u g_+(\delta_1(\Pi_u)) \\
&\geq g_+\left(\sum_u q_u\delta_1(\Pi_u)\right) \\
&= g_+(\delta_1(\Pi)) ,
\end{aligned}$$

where the first step follows from Lemmas 4.1 and 4.2, the second from the inductive hypothesis, the third from Lemma 4.3, and the fourth from Lemma 4.2. This proves claim (13). Since g_+ is decreasing in $(0,1)$, if $\delta_1(\Pi)$ and $\delta_0(\Pi)$ are chosen sufficiently small, we can ensure that $\kappa(\Pi)\geq\frac{1}{2}$. ∎

The above lemma shows that $\mathrm{icost}_\mu(\Pi\,|\,D) = \Omega(1/t)$, for any Π which solves AND_t, and this in turn completes the proof of Theorem 2.6.

5. Acknowledgments

We thank Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, Oded Regev, D. Sivakumar, Benny Sudakov, and Avi Wigderson for several helpful discussions about the problems considered here.

References

[1] F. Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 175(2):139–159, 1996.

[2] M. Ajtai, T. Jayram, R. Kumar, and D. Sivakumar. Approximate counting of inversions in a data stream. In *STOC*, pages 370–379, 2002.

[3] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In *STOC*, pages 20–29, 1996.

[4] L. Babai. Communication complexity. Technical Report TR-2001-09, University of Chicago, 2001.

[5] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proc. 43rd Annu. IEEE Symp. Found. Comp. Sci.*, pages 209–218, 2002.

[6] A. Chakrabarti, Y. Shi, A. Wirth, and A. C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd Annu. IEEE Symp. Found. Comp. Sci.*, pages 270–278, 2001.

[7] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[8] J. Håstad and A. Wigderson. The probabilistic communication complexity of disjointness of k-sets is $O(k)$. Unpublished manuscript, 1990.

[9] P. Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. In *FOCS*, pages 189–197, 2000.

[10] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, 1997.

[11] I. Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 1991.

[12] M. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proc. 34th Annu. ACM Symp. Theory Comput.*, pages 360–369, 2002.

[13] A. C. Yao. Some complexity questions related to distributive computing. In *Proc. 11th Annu. ACM Symp. Theory Comput.*, pages 209–213, 1979.