Electronic Colloquium on Computational Complexity, Report No. 10 (2009)

Lower bounds on the randomized communication complexity of read-once functions[*]

Nikos Leonardos Michael Saks

January 28, 2009

Abstract

We prove lower bounds on the randomized two-party communication complexity of functions that arise from read-once boolean formulae.

A read-once boolean formula is a formula in propositional logic with the property that every variable appears exactly once. Such a formula can be represented by a tree, where the leaves correspond to variables, and the internal nodes are labeled by binary connectives. Under certain assumptions, this representation is unique. Thus, one can define the depth of a formula as the depth of the tree that represents it.

The complexity of the evaluation of general read-once formulae, has attracted interest mainly in the decision tree model. In the communication complexity model, many interesting results deal with specific read-once formulae, such as DISJOINTNESS and TRIBES. In this paper we use information theory methods to prove lower bounds that hold for any read-once formula. Our lower bounds are of the form $n(f)/c^{d}(f)$, where $n(f)$ is the number of variables and $d(f)$ the depth of the formula, and they are optimal up to the constant c in the denominator.

1 Introduction

A landmark result in the theory of two-party communication complexity is the linear lower bound on the randomized communication complexity of set-disjointness proved by Kalyanasundaram and Schnitger [KS92]. Razborov [Raz92] gave a simplified proof, and Bar-Yossef et al. [BYJKS04] gave an elegant information theory proof, building on the informational complexity framework of Chakrabarti et al. [CSWY01].

Let us define a *two-party boolean function* to be a boolean function f together with a partition of its variables into two parts. We usually refer to the variables in the two classes as x and y and write $f(x, y)$ for the function. A two-party function is associated with the following communication problem: Given that Alice gets x and Bob gets y, compute $f(x, y)$.

If f is any n-variate boolean function and g is a 2-variate boolean function, we define f^g to be the two-party function taking two n bit strings x and y and defined to be $f^g(x, y) = f(g(x_1, y_1), \ldots, g(x_n, y_n))$. The disjointness communication problem can be reformulated

[*]This work was supported in part by NSF under grants CCF-0515201 and CCF-0832787.

ISSN 1433-8092

as a boolean function computation problem: Alice gets $x \in \{0,1\}^n$, Bob gets $y \in \{0,1\}^n$ and they want to compute $(\mathrm{OR}_n)^\wedge(x,y)$, where OR_n is the n-wise OR function.

Jayram et al. [JKS03], extended the techniques for disjointness in order to prove a linear lower bound for the randomized complexity on the function $(\mathrm{TRIBES}_{s,t})^\wedge$ where $\mathrm{TRIBES}_{s,t}$ is the function taking input $(z_{i,j} : 1 \leq i \leq s, 1 \leq j \leq t)$ and equal to $\mathrm{TRIBES}_{s,t}(z) = \bigwedge_{i=1}^{s} \bigvee_{j=1}^{t} z_{i,j}$.

The functions OR_n and $\mathrm{TRIBES}_{s,t}$ are both examples of *read-once boolean functions*, functions that can be represented by boolean formulae involving $\vee$ and $\wedge$, in which each variable appears (possibly negated) at most once. Such a formula can be represented by a rooted ordered tree, with nodes labeled by $\vee$ and $\wedge$, and the leaves labeled by variables. It is well known (see e.g. [HNW93]) that for any read-once function f, f has a unique representation (which we call the *canonical representation* of f) as a tree in which the labels of nodes on each root-to-leaf path alternate between $\wedge$ and $\vee$. The depth of f, $d(f)$, is defined to be the maximum depth of a leaf in the canonical representation, and $n(f)$ is the number of variables.

We want to consider communication problems derived from arbitrary read-once formulae. Based on the examples of OR_n and $\mathrm{TRIBES}_{s,t}$ mentioned above it seems natural to consider the function $f^\wedge$, but in the case that f is the n-wise AND, $f^\wedge$ trivializes (and can be computed with a two-bit protocol), and the more interesting function to consider is $f^\vee$.

Our main result says that for any read-once function f, at least one of the functions $f^\vee$ and $f^\wedge$ has high δ-error communication complexity.

Theorem 1. *For any read-once function f of depth at least* 1*,*

$$\max\{R_\delta(f^\wedge), R_\delta(f^\vee)\} \geq (2 - 4\sqrt{\delta}) \cdot \frac{n(f)}{8^{d(f)}}.$$

This result is, in some sense, best possible (up to the constant 8 in the base of $d(f)$). That is, there is a constant $c > 1$, such that if f is given by a t-uniform tree of depth d (so $n = t^d$), then $f^\wedge$ and $f^\vee$ both have randomized communication protocols using $O(n(f)/c^{d(f)})$ bits. This follows from the fact (see [SW86]) that f has a randomized decision tree algorithm using an expected number of queries $O(n(f)/c^{d(f)})$, and any decision tree algorithm for f is easily converted to a communication protocol for $f^\vee$ or $f^\wedge$ having comparable complexity. In fact, for t-uniform trees (in which each non-leaf node has t children and all leaves are at the same depth), we can improve the lower bound.

Theorem 2. *For any read-once function f that can be represented by a t-uniform* AND/OR *tree of depth $d \geq 1$,*

$$\max\{R_\delta(f^\wedge), R_\delta(f^\vee)\} \geq (2 - 4\sqrt{\delta}) \cdot \frac{t(t-1)^{d-1}}{4^d}.$$

As a simple corollary of our main result we obtain a similar lower bound for the more general class of *read-once threshold functions*. Recall that a *t-out-of-k threshold gate* is the boolean function with k inputs that is one if the sum of the inputs is at least t. A threshold tree is a rooted tree whose internal nodes are labeled by threshold gates, and whose leaves are labeled by distinct variables (or their negations). A read-once threshold function is a function representable by a threshold tree. We prove the following bound.

Theorem 3. *For any read-once threshold function f of depth at least 1,*

$$\max\{R_\delta(f^\wedge), R_\delta(f^\vee)\} \geq (2 - 4\sqrt{\delta}) \cdot \frac{n(f)}{16^{d(f)}}.$$

This result should be compared with the result of Heiman, Newman, and Wigderson [HNW93] that every read-once threshold function f has randomized decision tree complexity at least $n(f)/2^{d(f)}$. A lower bound on communication complexity of $f^\vee$ or $f^\wedge$ gives the same lower bound on decision tree complexity for f, however, the implication goes only one way, since communication protocols for $f^\vee$ and $f^\wedge$ do not have to come from a decision tree algorithm for f, and can be much faster. (For example, $f^\wedge$ when $f = \mathrm{AND}_n$ has randomized decision tree complexity $\Theta(n)$ but communication complexity 1.) Thus, up to the constant in the base of the denominator, our result can be viewed as a strengthening of the decision tree lower bound.

As we were completing this paper we learned of independent work of Jayram, Kopparty, and Raghavendra [JKR09], also based on the informational complexity approach, that gives a weaker lower bound of $n(f)/2^{2^{d(f)}}$ for formulae coming from balanced trees.

2 Notation, terminology, and preliminaries

In this section we establish notation and terms that we will use to describe the basic objects that we will be dealing with. We list standard definitions and state some basic inequalities in information theory. We discuss communication complexity and set up its connection with information theory.

Definitions pertaining to rooted trees. All trees in this paper are rooted. For a tree T we write V_T for the set of vertices, L_T for the set of leaves, $N_T = |L_T|$ for the number of leaves, and d_T for the depth of T. For a vertex u, path(u) is the set of vertices on a path from u to the root (including both the root and u).

We write $T = T_1 \circ \cdots \circ T_k$ when for each $j \in [k]$, T_j is the subtree rooted at the j-th child of the root of T.

A tree is called *t-uniform* if all its leaves are at the same depth d, and every non-leaf has exactly t children.

A tree is in *standard form* if there are no nodes with exactly one child. For example, a standard binary tree is one where every internal node has exactly two children.

A *full binary subtree* of a tree T is a binary tree in standard form that is contained in T, contains the root of T, and whose leaf-set is a subset of the leaf-set of T. Denote by FBS_T the set of full binary subtrees of T.

Definitions pertaining to boolean functions. Denote by $[n]$ the set $\{1, \ldots, n\}$ of integers. Let $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \to \mathbb{R}$ be a function and suppose that, for $i \in [n]$, $h_i : \mathcal{Z}_i \to \mathcal{S}_i$. Define $f^{\mathcal{H}} : \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_n \to \mathbb{R}$, where $\mathcal{H} = \langle h_1, \ldots, h_n \rangle$, to be the function defined by $f^{\mathcal{H}}(z_1, \ldots, z_n) = f(h_1(z_1), \ldots, h_n(z_n))$. When $h_j = h$ for all $j \in [n]$, we write $f^h = f^{\mathcal{H}}$.

A *tree circuit* is a rooted tree in which every leaf corresponds to an input variable (or its negation), and each gate comes from the set {AND,OR,NAND,NOR}. An AND/OR tree is a tree circuit with gates AND and OR. The tree circuit is *read-once* if the variables occurring at leaves are distinct; all tree circuits in this paper are assumed to be read-once. A Boolean function f is read-once if it can be represented by a read-once tree circuit. The depth of a read-once function f, denoted $d(f)$, is the minimum depth of a read-once tree circuit that computes it. As mentioned in the introduction, it is well-known that every read-once function f has a unique representation, called the *canonical representation of f* whose tree is in standard form and such that the gates along any root to leaf path alternate between $\wedge$ and $\vee$. It is also known that the depth of the canonical representation is $d(f)$, that is, the canonical representation has minimum depth over all read-once tree circuits that represent f.

If T is any rooted tree, we write f_T for the boolean function obtained by associating a distinct variable x_j to each leaf j and labeling each gate by a NAND gate. We use symbol '$\overline{\wedge}$' for NAND.

Random variables and distributions. We consider discrete probability spaces (Ω, ζ), where Ω is a finite set and ζ is a nonnegative valued function on Ω summing to 1. If $(\Omega_1, \zeta_1), \ldots, (\Omega_n, \zeta_n)$ are such spaces, their product is the space (Λ, ν), where $\Lambda = \Omega_1 \times \cdots \times \Omega_n$ is the Cartesian product of sets, and for $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$, $\nu(\omega) = \zeta_1(\omega_1) \cdots \zeta_n(\omega)$. In the case that all of the (Ω_i, ζ_i) are equal to a common space (Ω, ζ) we write $\Lambda = \Omega^n$ and $\nu = \zeta^n$.

We use uppercase for random variables, as in $X, Y, \mathbf{D}$, and write in bold those that represent vectors of random variables. For a variable X with range $\mathcal{X}$ that is distributed according to a probability distribution μ, i.e. $\Pr[X = x] = \mu(x)$, we write $X \sim \mu$. If X is uniformly distributed in $\mathcal{X}$, we write $X \in_R \mathcal{X}$.

Unless otherwise stated, all random variables take on values from finite sets.

Information theory. Let X, Y, Z be random variables on a common probability space, taking on values, respectively, from finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$. Let A be any event. The *entropy* of X and the *conditional entropy of X given A* and the *conditional entropy of X given Y* are respectively

$$\mathrm{H}(X) = -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log \Pr[X = x],$$

$$\mathrm{H}(X \mid A) = -\sum_{x \in \mathcal{X}} \Pr[X = x \mid A] \cdot \log \Pr[X = x \mid A], \text{ and}$$

$$\mathrm{H}(X \mid Y) = \sum_{y} \Pr[Y = y] \cdot \mathrm{H}(X \mid Y = y).$$

The *mutual information* between X and Y is

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

and the *conditional mutual information* of X and Y given Z is

$$\begin{aligned}\mathrm{I}(X\,;\,Y \mid Z) &= \mathrm{H}(X \mid Z) - \mathrm{H}(X \mid Y, Z) = \mathrm{H}(Y \mid Z) - \mathrm{H}(Y \mid X, Z) \\ &= \sum_{z} \Pr[Z = z] \cdot \mathrm{I}(X\,;\,Y \mid Z = z).\end{aligned}$$

We will need the following facts about the entropy. (See [CT06, Chapter 2] for proofs and more details.)

Proposition 4. *Let X, Y, Z be random variables.*

1. $\mathrm{H}(X) \geq \mathrm{H}(X \mid Y) \geq 0$.

2. If $\mathcal{X}$ is the range of X then $\mathrm{H}(X) \leq \log |\mathcal{X}|$.

3. $\mathrm{H}(X, Y) \leq \mathrm{H}(X) + \mathrm{H}(Y)$ *with equality if and only if X and Y are independent. This holds for conditional entropy as well.* $\mathrm{H}(X, Y \mid Z) \leq \mathrm{H}(X \mid Z) + \mathrm{H}(Y \mid Z)$ *with equality if and only if X and Y are independent given Z.*

The following proposition makes mutual information useful in proving direct-sum theorems.

Proposition 5 ([BYJKS04])**.** *Let $\mathbf{Z} = \langle \mathbf{Z}_1, \ldots, \mathbf{Z}_n \rangle, \Pi, \mathbf{D}$ be random variables. If the $\mathbf{Z}_j$'s are independent given $\mathbf{D}$ then* $\mathrm{I}(\mathbf{Z}\,;\,\Pi \mid \mathbf{D}) \geq \sum_{j=1}^{n} \mathrm{I}(\mathbf{Z}_j\,;\,\Pi \mid \mathbf{D})$.

Proof. By definition $\mathrm{I}(\mathbf{Z}\,;\,\Pi \mid \mathbf{D}) = \mathrm{H}(\mathbf{Z} \mid \mathbf{D}) - \mathrm{H}(\mathbf{Z} \mid \Pi, \mathbf{D})$. By Proposition 4(3), $\mathrm{H}(\mathbf{Z} \mid \mathbf{D}) = \sum_j \mathrm{H}(\mathbf{Z}_j \mid \mathbf{D})$ and $\mathrm{H}(\mathbf{Z} \mid \Pi, \mathbf{D}) \leq \sum_j \mathrm{H}(\mathbf{Z}_j \mid \Pi, \mathbf{D})$. The result follows. □

Communication complexity. In this work we will be dealing with the two-party private-coin randomized communication model [Yao79]. Alice is given $x \in \mathcal{X}$ and Bob $y \in \mathcal{Y}$. They wish to compute a function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ by exchanging messages according to a protocol Π. Let the random variable $\Pi(x, y)$ denote the transcript of the communication on input $\langle x, y \rangle$ (where the probability is over the random coins of Alice and Bob) and $\Pi_{\mathrm{out}}(x, y)$ the outcome of the protocol. We call Π a *δ-error protocol for f* if, for all $\langle x, y \rangle$, $\Pr[\Pi_{\mathrm{out}}(x, y) = f(x, y)] \geq 1 - \delta$. The *communication cost of* Π is $\max |\Pi(x, y)|$, where the maximum is over all input pairs $\langle x, y \rangle$ and over all coin tosses of Alice and Bob. The *δ-error randomized communication complexity of f*, denoted $R_\delta(f)$, is the cost of the best δ-error protocol for f. (See [KN06] for more details.)

Communication complexity lower bounds via information theory. The *informational complexity* paradigm, introduced by [CSWY01], and used in [SS02, BYJKS02, CKS03, BYJKS04, JKS03] provides a way to prove lower bounds on communication complexity via information theory. We are given a two-party function f and we want to show that any δ-error randomized communication protocol Π for f requires high communication. We introduce a probability distribution over the inputs to Alice and Bob. We then analyze the behavior of Π when run on the random distribution of inputs. The informational complexity is the mutual information of the string of communicated bits (the *transcript* of Π) with Alice and Bob's inputs, and provides a lower bound on the amount of communication.

More precisely, let $\Omega = (\Omega, \zeta)$ be a probability space over which are defined random variables $\mathbf{X} = \langle X_1, \ldots, X_n \rangle$ and $\mathbf{Y} = \langle Y_1, \ldots, Y_n \rangle$ representing Alice and Bob's inputs. The *information cost* of a protocol Π with respect to ζ is defined to be $\mathrm{I}(\mathbf{X}, \mathbf{Y}\,;\,\Pi(\mathbf{X}, \mathbf{Y}))$, where $\Pi(\mathbf{X}, \mathbf{Y})$ is a random variable following the distribution of the communication transcripts when the protocol Π runs on input $\langle \mathbf{X}, \mathbf{Y} \rangle \sim \zeta$. The *$\delta$-error informational complexity* of f with respect to ζ, denoted $\mathrm{IC}_{\zeta,\delta}(f)$, is $\min_\Pi \mathrm{I}(\mathbf{X}, \mathbf{Y}\,;\,\Pi(\mathbf{X}, \mathbf{Y}))$, where the minimum is over all δ-error randomized protocols for f.

Mutual information may be easier to handle if one conditions on the appropriate random variables. To that end, [BYJKS04] introduced the notion of *conditional information cost* of a protocol Π with respect to an auxiliary random variable. Let (Ω, ζ) be as above, and let $\mathbf{D}$ be an additional random variable defined on Ω. The *conditional information cost* of Π conditioned on $\mathbf{D}$ with respect to ζ, is defined to be $\mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \,|\, \mathbf{D})$, where $\Pi(\mathbf{X}, \mathbf{Y})$ is as above and $(\langle \mathbf{X}, \mathbf{Y} \rangle, \mathbf{D}) \sim \zeta$. The *$\delta$-error conditional informational complexity* of f conditioned on D with respect to ζ, denoted $\mathrm{IC}_{\zeta,\delta}(f \,|\, \mathbf{D})$, is $\min_\Pi \mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \,|\, \mathbf{D})$, where the minimum is over all δ-error randomized protocols for f.

Informational complexity provides a lower bound on randomized communication complexity, as shown by the following calculation. By definition of mutual information $\mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \,|\, \mathbf{D}) = \mathrm{H}(\Pi(\mathbf{X}, \mathbf{Y}) \,|\, \mathbf{D}) - \mathrm{H}(\Pi(\mathbf{X}, \mathbf{Y}) \,|\, \mathbf{X}, \mathbf{Y}, \mathbf{D})$. Applying in turn parts (1) and (2) of Proposition 4 gives that, for any δ-error protocol Π, $\mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \,|\, \mathbf{D}) \leq \mathrm{H}(\Pi(\mathbf{X}, \mathbf{Y})) \leq R_\delta(f)$.

Communication problems associated to boolean functions. If f is an arbitrary n-variate boolean function, and g is a 2-variate boolean function, we denote by f^g the two-party boolean function given by $f^g(x, y) = f(g_1(x_1, y_1), \ldots, g_n(x_n, y_n))$. Our goal is to prove Theorems 1 and 2, which say that for any read-once boolean function f, $f^\vee$ or $f^\wedge$ is large. To do this it will be more convenient to consider $f^{\bar\wedge}$ for functions f that come from trees using only NAND gates.

Theorem 6. *1. Let T be a tree in standard form with $d_T \geq 1$.*

$$R_\delta(f_T^{\bar\wedge}) \geq (4 - 8\sqrt{\delta}) \cdot \frac{N_T}{8^{d_T}}.$$

2. If T is, in addition, a t-uniform tree of depth $d \geq 1$ then

$$R_\delta(f_T^{\bar\wedge}) \geq (2 - 4\sqrt{\delta}) \cdot \frac{t(t-1)^{d_T - 1}}{4^{d_T}}.$$

To deduce Theorems 1 and 2 from this, we use the following lemma.

Proposition 7. *Let f be a read-once formula. Then there is a tree T in standard form such that (1) $R_\delta(f_T^{\bar\wedge}) \leq \max\{R_\delta(f^\wedge), R_\delta(f^\vee)\}$, (2) $N_T \geq n(f)/2$, and (3) $d_T \leq d(f)$.*

Proof. Let C be the representation of f in canonical form. Define tree circuits C_1 and C_2 as follows. C_1 is obtained by deleting all leaves that feed into $\wedge$ gates, and introducing a new variable for any node that becomes a leaf after this pruning. Let C_2 be obtained similarly by deleting all leaves that feed into $\vee$ gates. Let f_1 and f_2, respectively, be the functions computed by C_1 and C_2. Let T_1 and T_2 be the trees underlying C_1 and C_2 respectively. We take T to be whichever of T_1 and T_2 has more leaves. Clearly conditions (2) and (3) above will hold. Condition (1) follows immediately from the following claim.

Claim 8. *(1) $R_\delta(f^\wedge) \geq R_\delta(f_1^\wedge)$. (2) $R_\delta(f_1^\wedge) = R_\delta(f_{T_1}^{\bar\wedge})$. (3) $R_\delta(f^\vee) \geq R_\delta(f_2^\vee)$. (4) $R_\delta(f_2^\vee) = R_\delta(f_{T_2}^{\bar\wedge})$.*

To prove the first part of the claim, it suffices to observe that any communication protocol for $f^{\wedge}$ can be used as a protocol for $f_1^{\wedge}$. Given inputs (x, y) to $f_1^{\wedge}$ Alice and Bob can construct inputs (x', y') to $f^{\wedge}$ such that $f^{\wedge}(x', y') = f_1^{\wedge}(x, y)$, as follows. If j is a leaf of C that is also a leaf of C_1 then Alice sets $x'_j = x_j$ and Bob sets $y'_j = y_j$. Suppose j is a leaf of C that is not a leaf of C_1. If the parent $p(j)$ of j is a leaf of C_1 then Alice sets $x'_j = x_{p(j)}$ and Bob sets $y'_j = y_{p(j)}$. If $p(j)$ is not a leaf of C_1, then Alice sets $x'_j = 1$ and Bob sets $y'_j = 1$. It is easy to verify that $f^{\wedge}(x', y') = f_1^{\wedge}(x, y)$.

To prove the second part of the claim, we observe that $f_1^{\wedge} = f_{T_1}^{\overline{\wedge}}$ if the top gate of C_1 is $\vee$ and $f_1^{\wedge} = \neg f_{T_1}^{\overline{\wedge}}$ if the top gate of C_1 is $\wedge$. In either case, they have identical communication complexity.

The proofs of parts 3 and 4 follow similarly. □

Notice that if T is a uniform tree, then one of T_1 and T_2 above will have N_T leaves. Thus, in the case of uniform trees we have $N_T = n(f)$ and save a factor of 2.

3 The methods of [BYJKS04]

The authors of [BYJKS04] introduced new techniques for proving lower bounds on information cost. In this section we summarize their method and list the results and definitions from [BYJKS04] that we will use.

Their methodology has two main parts. In the first part they make use of Proposition 5 to obtain a direct-sum theorem for the informational complexity of the function. This works particularly well with functions of the form $f^h(\mathbf{x}, \mathbf{y}) = f(h(x_1, y_1), \ldots, h(x_n, y_n))$. Before stating the direct-sum theorem, we need some definitions.

Definition 9 (Sensitive input)**.** *Let $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \to \mathbb{R}$, $\mathcal{H} = \langle h_j : \mathcal{Z}_j \to \mathcal{S}_j \rangle_{j \in [n]}$, and $\mathbf{z} = \langle z_1, \ldots, z_n \rangle \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$. For $j \in [n]$, $u \in S_j$, define $\mathbf{z}[j, u] = \langle z_1, \ldots, z_{j-1}, u, z_{j+1}, \ldots, z_n \rangle$. We say that $\mathbf{z}$ is* sensitive *for $f^{\mathcal{H}}$ if $(\forall j \in [n])(\forall u \in \mathcal{Z}_j)(f^{\mathcal{H}}(\mathbf{z}[j, u]) = h_j(u))$.*

For an example, consider the function $\mathrm{DISJ}_n(\mathbf{x}, \mathbf{y}) = \bigvee_{j=1}^{n} (x_j \wedge y_j)$. Any input $\langle \mathbf{x}, \mathbf{y} \rangle$ such that, for all $j \in [n]$, $x_j \wedge y_j = 0$, is sensitive.

Definition 10 (Collapsing distribution)([BYJKS04])**.** *Let $f, \mathcal{H}$ be as in Definition 9. Call a distribution μ over $\mathcal{Z}_1 \times \cdots \times \mathcal{Z}_n$* collapsing *for $f^{\mathcal{H}}$, if every $\mathbf{z}$ in the support of μ is sensitive.*

Theorem 11 ([BYJKS04])**.** *Let $f : \mathcal{S}^n \to \{0, 1\}$, and $h : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}$. Consider random variables $\mathbf{X} = \langle X_1, \ldots, X_n \rangle \in \mathcal{X}^n, \mathbf{Y} = \langle Y_1, \ldots, Y_n \rangle \in \mathcal{Y}^n, \mathbf{D} = \langle D_1, \ldots, D_n \rangle$, and $\mathbf{Z} = \langle Z_1, \ldots, Z_n \rangle$, where $Z_j = \langle X_j, Y_j, D_j \rangle$ for $j \in [n]$.*

Assume that $\{Z_j\}_{j \in [n]}$ is a set of mutually independent variables, and $Z_j \sim \zeta$ for all $j \in [n]$ (thus, $\mathbf{Z} \sim \zeta^n$). If, for all $j \in [n]$, X_j and Y_j are independent given D_j, and the marginal distribution of $(\mathbf{X}, \mathbf{Y})$ is a collapsing distribution for f^h, then $\mathrm{IC}_{\zeta^n, \delta}(f^h \mid \mathbf{D}) \geq n \cdot \mathrm{IC}_{\zeta, \delta}(h \mid D)$.

Defining a distribution ζ satisfying the two requirements asked in Theorem 11, moves the attention from $\mathrm{IC}_{\zeta^n,\delta}(f^h \,|\, \mathbf{D})$ to $\mathrm{IC}_{\zeta,\delta}(h \,|\, D)$. For example, in [BYJKS04] it is shown how to define ζ when f^h is $\mathrm{DISJ}_n(\mathbf{x},\mathbf{y}) = \bigvee_{j=1}^{n}(x_j \wedge y_j)$. Then they only have to deal with $\mathrm{IC}_{\zeta,\delta}(h \,|\, D)$, where $h(x,y) = x \wedge y$.

The second part of the method is a framework for proving lower bounds on information cost. The first step consists of a passage from mutual information to Hellinger distance.

Definition 12. *(Hellinger distance.) The* Hellinger distance *between probability distributions P and Q on a domain Ω is defined by* $\mathrm{h}(P,Q) = \sqrt{\frac{1}{2}\sum_{\omega\in\Omega}\left(\sqrt{P_\omega} - \sqrt{Q_\omega}\right)^2}$. *We write* $\mathrm{h}^2(P,Q)$ *for* $(\mathrm{h}(P,Q))^2$.

Lemma 13 ([BYJKS04])**.** *Let $\Phi(z_1)$, $\Phi(z_2)$, and $Z \in_R \{z_1, z_2\}$ be random variables. If $\Phi(z)$ is independent of Z for each $z \in \{z_1, z_2\}$ then* $\mathrm{I}(Z\,;\,\Phi(Z)) \geq \mathrm{h}^2(\Phi(z_1), \Phi(z_2))$.

The following proposition states useful properties of Hellinger distance. They reveal why Hellinger distance is better to work with than mutual information.

Proposition 14 (Properties of Hellinger distance)([BYJKS04])**.**

1. (Triangle inequality.) *Let P, Q, and R be probability distributions over domain Ω; then* $\mathrm{h}(P,Q) + \mathrm{h}(Q,R) \geq \mathrm{h}(P,R)$. *It follows that the square of the Hellinger distance satisfies a weak triangle inequality;* $\mathrm{h}^2(P,Q) + \mathrm{h}^2(Q,R) \geq \frac{1}{2}\mathrm{h}^2(P,R)$.

2. (Cut-and-paste property.) *For any randomized protocol* Π *and for any $x, x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$,* $\mathrm{h}(\Pi(x,y), \Pi(x',y')) = \mathrm{h}(\Pi(x,y'), \Pi(x',y))$.

3. (Pythagorean property.) *For any randomized protocol* Π *and for any $x, x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$,* $\mathrm{h}^2(\Pi(x,y), \Pi(x',y)) + \mathrm{h}^2(\Pi(x,y'), \Pi(x',y')) \leq 2\mathrm{h}^2(\Pi(x,y), \Pi(x',y'))$.

4. For any δ-error randomized protocol Π *for a function f, and for any two input pairs (x,y) and (x',y') for which $f(x,y) \neq f(x',y')$,* $\mathrm{h}^2(\Pi(x,y), \Pi(x',y')) \geq 1 - 2\sqrt{\delta}$.

After an application of Lemma 13 we are left with a sum of Hellinger distance terms, which we need to lower bound. Applying some of the properties 1–3 several times we can arrive at a sum of terms different than the ones we started with. To obtain a lower bound we would like the final terms to be such as the one in Property 4 and take advantage of the constant lower bound.

4 Read-once boolean formulae

Let $T = T_1 \circ \cdots \circ T_n$ be a tree in standard form computing a function f_T. A first step towards simplifying the informational complexity of $f_T^{\wedge}$, would be to apply the following straightforward generalization of Theorem 11.

Theorem 15. *Let $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \to \{0,1\}$, and $\mathcal{H} = \langle h_j : \mathcal{X}_j \times \mathcal{Y}_j \to \mathcal{S}_j \rangle_{j\in[n]}$. Consider random variables $\mathbf{X} = \langle X_1, \ldots, X_n \rangle \in \mathcal{X}_1 \times \mathcal{X}_n, \mathbf{Y} = \langle Y_1, \ldots, Y_n \rangle \in \mathcal{Y}_1 \times \mathcal{Y}_n, \mathbf{D} = \langle D_1, \ldots, D_n \rangle$, and $\mathbf{Z} = \langle Z_1, \ldots, Z_n \rangle$, where $Z_j = \langle X_j, Y_j, D_j \rangle$ for $j \in [n]$.*

Assume that $\{Z_j\}_{j\in[n]}$ is a set of mutually independent variables, and $Z_j \sim \zeta_j$ for all $j \in [n]$ (thus, $\mathbf{Z} \sim \zeta_1 \cdots \zeta_n$). If, for all $j \in [n]$, X_j and Y_j are independent given D_j, and the *marginal distribution of* $(\mathbf{X}, \mathbf{Y})$ *is a collapsing distribution for* $f^{\mathcal{H}}$*, then* $\mathrm{IC}_{\zeta_1 \cdots \zeta_n, \delta}(f^{\mathcal{H}} \mid \mathbf{D}) \geq \sum_{j=1}^{n} \mathrm{IC}_{\zeta_j, \delta}(h_j \mid D_j)$.

One can apply Theorem 15 to the function $f_T^{\overline{\wedge}}$, with f the n-bit NAND and $h_j = f_{T_j}^{\overline{\wedge}}$, for $j \in [n]$. However, this won't take as very far. The problem is that if μ—the marginal distribution of $\langle \mathbf{X}, \mathbf{Y} \rangle$—is collapsing for $f_T^{\overline{\wedge}}$ then the support of μ is a subset of $(f^{\mathcal{H}})^{-1}(0)$. Therefore, we will inherit for each subtree a distribution μ_j with a support inside $h_j^{-1}(1)$. But the support of a collapsing distribution should lie inside $h_j^{-1}(0)$. This means that we cannot apply Theorem 15 repeatedly. This problem arose in [JKS03] when studying the function $\mathrm{TRIBES}_{m,n}(\mathbf{x}, \mathbf{y}) = \bigwedge_{k=1}^{m} \mathrm{DISJ}_n(\mathbf{x}_k, \mathbf{y}_k) = \bigwedge_{k=1}^{m} \bigvee_{j=1}^{n} (x_{kj} \wedge y_{kj})$. The authors of [JKS03] managed to overcome this problem by proving a more complicated direct-sum theorem for a non-collapsing distribution for DISJ. Inspired by their idea, we show how to do the same for arbitrary read-once boolean functions.

The information cost of a protocol Π that we will employ for our proof will have the form $\mathrm{I}(\mathbf{X}, \mathbf{Y}\,;\, \Pi(\mathbf{X}, \mathbf{Y}) \mid \Gamma, \mathbf{D})$, where random variables Γ and $\mathbf{D}$ are auxiliary variables that will be used to define the distribution over the inputs.

4.1 Further definitions on trees

We now proceed with a series of definitions for objects that will be needed to finally define a distribution ζ for $(\langle \mathbf{X}, \mathbf{Y} \rangle, \langle \Gamma, \mathbf{D} \rangle)$, which will give meaning to $\mathrm{IC}_{\zeta,\delta}(f_T^{\overline{\wedge}} \mid \Gamma, \mathbf{D}) = \min_{\Pi} \mathrm{I}(\mathbf{X}, \mathbf{Y}\,;\, \Pi(\mathbf{X}, \mathbf{Y}) \mid \Gamma, \mathbf{D})$.

Definition 16. *(Valid coloring.) For our purposes, a* coloring *of a tree* T *is a partition of* V_T *into two sets* $\gamma = \langle \mathrm{W}_\gamma, \mathrm{R}_\gamma \rangle$*. The vertices of* W_γ *are said to be* white *and the vertices of* R_γ *are said to be* red*. A coloring is* valid *if it satisfies the following conditions.*

1. *The root is white.*
2. *A white node is either a leaf or exactly one of its children is red.*
3. *A red node is either a leaf or exactly two of its children are red.*

For example, for a standard binary tree, a valid coloring paints all nodes on some root-to-leaf path white and all the rest red. Thus, the number of valid colorings equals the number of leaves.

Definitions related to colorings. We note some properties of valid colorings and give further definitions of related objects. Consider a tree T and a valid coloring $\gamma = \langle \mathrm{W}_\gamma, \mathrm{R}_\gamma \rangle$.

(1) The red nodes induce a forest of binary trees in standard form called the *red forest.*

(2) We can define a one-to-one correspondence between the trees in the red forest and internal white nodes of T as follows. For each white node w, its unique red child is the root of one of the full binary trees. We let $\mathrm{RT}(w) = \mathrm{RT}_{\gamma,T}(w)$ denote the set of vertices in the red binary tree rooted at the red child of w. (For convenience, if w is a leaf, $\mathrm{RT}(w)$ is empty.)

(3) The *principal component* of γ is the set of white nodes whose path to the root consists only of white nodes. A *principal leaf* of γ is a leaf belonging to the principal component. Let $\mathrm{PL}_T(\gamma)$ denote the set of principal leaves of γ.

(4) A full binary subtree S of T (i.e. $S \in \mathrm{FBS}_T$) is said to be *compatible* with γ, written $S \sim \gamma$, if S has exactly one white leaf. (Notice that, since γ is valid, this leaf would have to be a principal leaf. Thus, $S \sim \gamma$ is equivalent to saying that the restriction of γ on V_S is a valid coloring for S.)

(5) Define $\mathrm{FBS}_T(\gamma) = \{S \in \mathrm{FBS}_T \mid S \sim \gamma\}$. This set is in one-to-one correspondence with the set $\mathrm{PL}_T(\gamma)$ of principal leaves. If u is a principal leaf then the set $\mathrm{path}(u) \cup \bigcup_{w \in \mathrm{path}(u)} \mathrm{RT}(w)$ induces a tree $\mathrm{F}_\gamma(u)$ that belongs to $\mathrm{FBS}_T(\gamma)$, and conversely if S is in $\mathrm{FBS}_T(\gamma)$ then its unique white leaf u is principal and $S = \mathrm{F}_\gamma(u)$.

(6) Define the positive integers $m_{\gamma,T} = |\mathrm{FBS}_T(\gamma)| = |\mathrm{PL}_T(\gamma)|$, $m_T = \sum_\gamma m_{\gamma,T}$, and $\rho_T = \min_\gamma m_{\gamma,T}$, where the min is over all valid colorings γ. (Notice that if $T = T_1 \circ \cdots \circ T_n$ then $\rho_T = \sum_j \rho_{T_j} - \max_j \rho_{T_j}$.)

On notation. Consider a tree T, $u \in V_T$, and a coloring γ of T. We write T_u for the subtree of T rooted at u. Consider a vector $\mathbf{z} \in \Sigma^{N_T}$, where each coordinate corresponds to a leaf. We write $\mathbf{z}_u$ for the part of $\mathbf{z}$ that corresponds to the leaves of T_u. For $S \in \mathrm{FBS}_T$ we write $\mathbf{z}_S$ for the part of $\mathbf{z}$ that corresponds to the leaves of S. We treat colorings similarly. For example, γ_S stands for $\langle \mathrm{W}_\gamma \cap V_S, \mathrm{R}_\gamma \cap V_S \rangle$.

4.2 The input distribution

Given an arbitrary tree T in standard form, we now define a distribution over inputs to Alice and Bob.

First, we associate to each standard binary tree T a special input $\langle \alpha_T, \beta_T \rangle$ defined recursively as follows.

Definition 17. *We define input $\langle \alpha_T, \beta_T \rangle$ for a standard binary tree T. The definition is recursive on the depth d_T of the tree.*

$$\langle \alpha_T, \beta_T \rangle = \begin{cases} \langle 1, 1 \rangle & \text{if } d_T = 0, \\ \langle \alpha_{T_1} \overline{\alpha}_{T_2}, \overline{\beta}_{T_1} \beta_{T_2} \rangle & \text{if } T = T_1 \circ T_2. \end{cases}$$

We will need the following property of $\langle \alpha_T, \beta_T \rangle$.

Proposition 18. *For a standard binary tree T with $d_T > 0$ we have $f_T^{\barwedge}(\alpha_T, \beta_T) = f_T^{\barwedge}(\overline{\alpha}_T, \overline{\beta}_T) = 0$ and $f_T^{\barwedge}(\alpha_T, \overline{\beta}_T) = f_T^{\barwedge}(\overline{\alpha}_T, \beta_T) = 1$.*

Proof. The proof is by induction on d_T.

For $d_T = 1$ the (unique) tree results in the function $f_T^{\barwedge}(x_1 x_2, y_1 y_2) = (x_1 \barwedge y_1) \barwedge (x_2 \barwedge y_2)$. Clearly, $f_T^{\barwedge}(\alpha_T, \beta_T) = f_T^{\barwedge}(10, 01) = 0$, $f_T^{\barwedge}(\overline{\alpha}_T, \overline{\beta}_T) = f_T^{\barwedge}(01, 10) = 0$, $f_T^{\barwedge}(\alpha_T, \overline{\beta}_T) = f_T^{\barwedge}(10, 10) = 1$, $f_T^{\barwedge}(\overline{\alpha}_T, \beta_T) = f_T^{\barwedge}(01, 01) = 1$.

For $d_T > 1$ we have $f_T^{\barwedge}(\alpha_T, \beta_T) = f_{T_1}^{\barwedge}(\alpha_{T_1}, \overline{\beta}_{T_1}) \barwedge f_{T_2}^{\barwedge}(\overline{\alpha}_{T_2} \beta_{T_2}) = 1 \barwedge 1 = 0$ (where we applied the inductive hypothesis on T_1 and T_2). The rest of the cases can be verified in a similar manner. □

An input will be determined by three independent random variables $\Gamma, \mathbf{D}, \mathbf{R}$, which are defined as follows.

(i) Γ ranges over valid colorings γ for T, according to a distribution that weights each γ by the number of principal leaves it has. More precisely

$$\Pr[\Gamma = \gamma] = m_{\gamma,T}/m_T.$$

(ii) $\mathbf{D} = \langle D_1, \ldots, D_N\rangle \in_R \{\textsc{Alice}, \textsc{Bob}\}^N$. Thus, for any $\mathbf{d} \in \{\textsc{Alice}, \textsc{Bob}\}^N$, $\Pr[\mathbf{D} = \mathbf{d}] = 1/2^N$.

(iii) $\mathbf{R} = \langle R_1, \ldots, R_N\rangle \in_R \{0,1\}^N$. Thus, for any $\mathbf{r} \in \{0,1\}^N$, $\Pr[\mathbf{R} = \mathbf{r}] = 1/2^N$.

The inputs $\mathbf{X} = \langle X_1, \ldots, X_N\rangle$ and $\mathbf{Y} = \langle X_1, \ldots, X_N\rangle$ to Alice and Bob are determined by values γ, $\langle d_1, \ldots, d_N\rangle$, and $\langle r_1, \ldots, r_N\rangle$ for $\Gamma, \mathbf{D}$, and $\mathbf{R}$ as follows.

(i) Let $F_1, \ldots, F_k$ be the trees of the red forest determined by γ. The inputs to subtree F_j, for $j \in [k]$, are set to $\langle \alpha_{F_j}, \beta_{F_j}\rangle$.

(ii) For a white leaf j, the corresponding input $\langle X_j, Y_j\rangle$ is determined as follows. If $d_j = \textsc{Alice}$ then we set $X_j = 0$ and $Y_j = r_j$; if $d_j = \textsc{Bob}$ then we set $Y_j = 0$ and $X_j = r_j$.

Let ζ_T be the resulting distribution on $\langle \mathbf{X}, \mathbf{Y}, \Gamma, \mathbf{D}\rangle$. Let μ_T (resp. ν_T) be the marginal distribution of $\langle \mathbf{X}, \mathbf{Y}\rangle$ (resp. $\langle \Gamma, \mathbf{D}\rangle$). We will often drop subscript T and write ζ, μ, and ν.

Proposition 19. *Consider a tree T and let $\langle \mathbf{x}, \mathbf{y}, \gamma, \mathbf{d}\rangle$ be in the support of ζ. If u is a red node with a white parent then $f^{\overline{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = 0$. If u is a white node then $f^{\overline{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = 1$.*

Proof. The proof is by induction on d_{T_u}.

When $d_{T_u} = 0$, u is a leaf. If u is red, the statement follows from Definition 17. If u is white, the statement follows from the definition of the distribution.

When $d_{T_u} > 0$ and u is white, then u has a red child v. By induction $f^{\overline{\wedge}}_{T_v}(\mathbf{x}_v, \mathbf{y}_v) = 0$, and it follows that $f^{\overline{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = 1$. If u is red and its parent is white then there is a tree F rooted at u in the red forest. We claim that $f^{\overline{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = f^{\overline{\wedge}}_{F}(\mathbf{x}_F, \mathbf{y}_F)$. The statement then follows by Proposition 18, because according to the definition of ζ_T, $\langle \mathbf{x}_F, \mathbf{y}_F\rangle = \langle \alpha_R, \beta_F\rangle$. The claim holds, because every $v \in V_F$ has only white children outside F, and—by the induction hypothesis—their values do not affect the value of v (since the inputs to a $\overline{\wedge}$-gate that are equal to '1' are, in some sense, irrelevant to the output). □

4.3 A direct-sum theorem for read-once boolean formulae

Let T be an arbitrary tree in standard form and $S \in \mathrm{FBS}_T$. Suppose we have a communication protocol Π for $f^{\overline{\wedge}}_T$ and we want a protocol for $f^{\overline{\wedge}}_S$. One natural way to do this is to have Alice extend her input $\mathbf{x}_S$ for S to an input $\mathbf{x}$ for T, and Bob extend his input $\mathbf{y}_S$ for S to an input $\mathbf{y}$ for T, in such a way that $f^{\overline{\wedge}}_T(\mathbf{x}, \mathbf{y}) = f^{\overline{\wedge}}_S(\mathbf{x}_S, \mathbf{y}_S)$. Then by running Π on $\langle \mathbf{x}, \mathbf{y}\rangle$ they obtain the desired output.

Let Π be any protocol for f_T. For any $S \in \mathrm{FBS}_T$ we will construct a family of protocols for S. Each protocol in the family will be specified by a pair $\langle \gamma, \mathbf{d}\rangle$ where γ is a valid coloring of T that is compatible with S, and $\mathbf{d} \in \{\textsc{Alice}, \textsc{Bob}\}^{N_T}$

Alice and Bob plug their inputs in T, exactly where S is embedded. To generate the rest of the input bits for T, they first use γ to paint the nodes of T not in S. For a red leaf j, the value of $\mathbf{X}_j$ and $\mathbf{Y}_j$ are determined by the coloring γ, so Alice and Bob can each determine $\mathbf{X}_j$ and $\mathbf{Y}_j$ without communication. For a white leaf j outside S they have to look at the value of d_j. If $d_j = \textsc{Alice}$, Alice sets $x_j = 0$, and Bob uses a random bit of his own to (independently) set his input bit y_j. If $d_j = \textsc{Bob}$, Bob sets $y_j = 0$, and Alice uses a random bit to set x_j. After this prepossessing, they simulate Π. Denote this protocol by $\Pi_S[\gamma, \mathbf{d}]$.

To argue the correctness of $\Pi_S[\gamma, \mathbf{d}]$ for any S, γ, and $\mathbf{d}$, notice that any node in S has only white children outside S (this follows from the conditions that a coloring satisfies). From Proposition 19, we know that a white node does not affect the value of its parent.

We now define a distribution over the triples $\langle S, \gamma, \mathbf{d}\rangle$ so that the average of the information cost of $\Pi_S[\gamma, \mathbf{d}]$ will be related to the information cost of Π. We do this by defining a distribution ξ_T for triples $\langle S, \gamma, \mathbf{d}\rangle$,

$$\xi_T(S, \gamma, \mathbf{d}) = \begin{cases} \frac{1}{m_T 2^{N_T}} & \text{if } S \sim \gamma, \\ 0 & \text{otherwise.} \end{cases}$$

This is indeed a distribution since

$$\sum_{S,\gamma,\mathbf{d}} \xi_T(S, \gamma, \mathbf{d}) = \sum_{S\sim\gamma} \sum_{\mathbf{d}} \frac{1}{m_T 2^{N_T}} = \sum_{S\sim\gamma} \frac{1}{m_T} = 1.$$

Lemma 20. *Consider any protocol* Π *for a tree* T. *Let* $(\langle \mathbf{X}, \mathbf{Y}\rangle, \langle \Gamma, \mathbf{D}\rangle) \sim \zeta_T$ *and* $(\langle \mathbf{X}', \mathbf{Y}'\rangle, \langle \Gamma', \mathbf{D}'\rangle) \sim \zeta_S$*; then*

$$\mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi \,|\, \Gamma, \mathbf{D}) \geq \rho_T \cdot \mathbf{E}_{\langle S,\gamma,\mathbf{d}\rangle\sim\xi_T}[\mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma', \mathbf{D}')].$$

Proof. We start by evaluating the right-hand side. (Recall that for γ and $\mathbf{d}$ we write γ_S and $\mathbf{d}_S$ for their restrictions in $S \in \mathrm{FBS}_T$.)

$$\begin{aligned}
&(1)\quad \mathbf{E}_{\langle S,\gamma,\mathbf{d}\rangle\sim\xi_T}[\mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma', \mathbf{D}')] \\
&(2)\quad\quad = \sum_{S,\gamma,\mathbf{d}} \xi_T(S, \gamma, \mathbf{d}) \sum_{\gamma',\mathbf{d}'} \nu_S(\gamma', \mathbf{d}') \\
&\quad\quad\quad\quad \cdot \mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma', \mathbf{D}' = \mathbf{d}')] \\
&(3)\quad\quad = \sum_{S} \sum_{\gamma',\mathbf{d}'} \sum_{\gamma:\gamma\sim S} \sum_{\mathbf{d}} \frac{1}{m_T 2^{N_T}} \cdot \frac{1}{N_S 2^{N_S}} \\
&\quad\quad\quad\quad \cdot \mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma', \mathbf{D}' = \mathbf{d}')] \\
&(4)\quad\quad = \sum_{S} \sum_{\gamma:\gamma\sim S} \sum_{\mathbf{d}} \frac{1}{m_{\gamma,T}} \cdot \frac{m_{\gamma,T}}{m_T 2^{N_T}} \\
&\quad\quad\quad\quad \cdot \mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S);
\end{aligned}$$

The transition from (3) to (4) needs to be justified. Look first at equation (4). Fix values $\widehat{S}, \widehat{\gamma}$, and $\widehat{\mathbf{d}}$ for the summation indices S, γ, and $\mathbf{d}$ respectively. Consider the corresponding term $A = \mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_{\widehat{S}}[\widehat{\gamma}, \widehat{\mathbf{d}}] \,|\, \Gamma' = \widehat{\gamma}_S, \mathbf{D}' = \widehat{\mathbf{d}}_S)$ in the sum. Now look at (3). Fix indices S, γ', and $\mathbf{d}'$ to $\widehat{S}$, $\widehat{\gamma}_S$, and $\widehat{\mathbf{d}}_S$ respectively. We claim that there are $N_S 2^{N_S}$ values $\langle \gamma, \mathbf{d} \rangle$, such that $\mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_{\widehat{S}}[\gamma, \mathbf{d}] \,|\, \Gamma' = \widehat{\gamma}_S, \mathbf{D}' = \widehat{\mathbf{d}}_S) = A$. Indeed, any $\langle \gamma, \mathbf{d} \rangle$ such that γ agrees with $\widehat{\gamma}$ outside S, and $\mathbf{d}$ agrees with $\widehat{\mathbf{d}}$ outside S, contributes A to the sum in equation (3). There are N_S such γ and 2^{N_S} such $\mathbf{d}$.

Let us define $j(\gamma, S)$ to be the white leaf of S which is colored white by γ. Recalling the definition of ρ_T (Section 4.1), the last equation gives

$$\begin{aligned}
(5)\quad & \mathbf{E}_{\langle S,\gamma,\mathbf{d}\rangle \sim \xi_T}[\mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma', \mathbf{D}')] \\
& \qquad \le \frac{1}{\rho_T} \sum_{S \sim \gamma, \mathbf{d}} \frac{m_{\gamma,T}}{m_T 2^{N_T}} \cdot \mathrm{I}(X'_{j(\gamma,S)}, Y'_{j(\gamma,S)} \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S).
\end{aligned}$$

For the left-hand side we have

$$\begin{aligned}
(6)\quad & \mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi \,|\, \Gamma, \mathbf{D}) \\
(7)\quad & \quad = \sum_{\gamma, \mathbf{d}} \nu_T(\gamma, \mathbf{d}) \cdot \mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi \,|\, G = \gamma, \mathbf{D} = \mathbf{d}) \\
(8)\quad & \quad \ge \sum_{\gamma, \mathbf{d}} \frac{m_{\gamma,T}}{m_T 2^{N_T}} \sum_{j \in \mathrm{PL}_T(\gamma)} \mathrm{I}(X_j, Y_j \,;\, \Pi \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d}) \\
(9)\quad & \quad \ge \sum_{S \sim \gamma, \mathbf{d}} \frac{m_{\gamma,T}}{m_T 2^{N_T}} \cdot \mathrm{I}(X_{j(\gamma,S)}, Y_{j(\gamma,S)} \,;\, \Pi \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d}),
\end{aligned}$$

The first inequality follows from Proposition 5 (notice that we are ignoring terms that correspond to white, but nonprincipal leaves). The second one follows from the bijection between $\mathrm{FBS}_T(\gamma)$ and $\mathrm{PL}_T(\gamma)$ as discussed in Section 4.1.

In view of (5) and (9), to finish the proof one only needs to verify that the distributions $(X'_{j(\gamma,S)}, Y'_{j(\gamma,S)}, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S)$ and $(X_{j(\gamma,S)}, Y_{j(\gamma,S)}, \Pi \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d})$ are identical. To see this, notice first that $\Pr[X'_{j(\gamma,S)} = b_x \,|\, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S] = \Pr[X_{j(\gamma,S)} = b_x \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d}]$, because S is colored the same in both cases and $j(\gamma, S)$ is the white leaf of S. Similarly for $Y'_{j(\gamma,S)}$ and $Y_{j(\gamma,S)}$. Finally, it follows immediately from the definition of $\Pi_S[\gamma, \mathbf{d}]$, that $\Pr[\Pi_S[\gamma, \mathbf{d}](\mathbf{X}', \mathbf{Y}') = \tau \,|\, X'_{j(\gamma,S)} = b_x, Y'_{j(\gamma,S)} = b_y, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S] = \Pr[\Pi(\mathbf{X}, \mathbf{Y}) = \tau \,|\, X_{j(\gamma,S)} = b_x, Y_{j(\gamma,S)} = b_y, \Gamma = \gamma, \mathbf{D} = \mathbf{d}]$. □

To obtain a lower bound from this lemma, we want to lower bound ρ_T and the informational complexity of standard binary trees. The later is done in the next section. The following lemma shows that we can assume $\rho_T \ge N_T / 2^{d_T}$.

Lemma 21. *For any tree T with N leaves and depth d, there is a tree $\widehat{T}$ with the following properties. (1) $\widehat{T}$ is in standard form, (2) $R_\delta(T) \ge R_\delta(\widehat{T})$, (3) $\rho_{\widehat{T}} \ge N/2^d$.*

Proof. First, we describe the procedure which applied on T produces $\widehat{T}$. If T is a single node we set $\widehat{T} = T$. Otherwise, assume $T = T_1 \circ \cdots \circ T_n$ and denote N_j the number of leaves in each T_j. We consider two cases.

A. If there is a j such that the leaves in T_j are at least $N/2$ then we apply the procedure to T_j to obtain $\widehat{T}_j$, set $\widehat{T} = \widehat{T}_j$, and remove the rest subtrees.

B. Otherwise, for each $j \in [n]$ apply the procedure on T_j to get $\widehat{T}_j$, and set $\widehat{T} = \widehat{T}_1 \circ \cdots \circ \widehat{T}_n$.

Now we prove by induction on d that $\widehat{T}$ has properties (1) and (3). When $d = 0$ and T is a single node, $\rho_T = 1$ and all properties are easily seen to be true. Otherwise, if $\widehat{T}$ was created as in case A then clearly property (1) holds. For property (3) assume $\widehat{T} = \widehat{T}_j$. By induction, $\rho_{\widehat{T}_j} \geq N_j/2^{d-1}$. It follows that $\rho_{\widehat{T}} = \rho_{\widehat{T}_j} \geq N/2^d$ (since $N_j \geq N/2$). Now suppose case B took place and $\widehat{T}$ was created from $\widehat{T}_1, \ldots, \widehat{T}_n$. The restructuring described in case B preserves property (1). For property (3) assume—without loss of generality—that $\rho_{\widehat{T}_1} \leq \cdots \leq \rho_{\widehat{T}_n}$. We have $\rho_{\widehat{T}} = \sum_{j=1}^{n-1} \rho_{\widehat{T}_j} \geq \sum_{j=1}^{n-1} N_j/2^{d-1} > (N - N/2)/2^{d-1} = N/2^d$.

Finally, property (2) is true because Alice and Bob can simulate the protocol for f_T after they set their bits below a truncated tree to '1'. □

4.4 Bounding the informational complexity of binary trees

In this section we concentrate on standard binary trees. Our goal is to prove a lower bound of the form $\mathrm{I}(\mathbf{X}, \mathbf{Y} ; \Pi \mid \Gamma, \mathbf{D}) \geq 2^{-\Theta(d_T)}$. We prove such an inequality using induction on d_T. The following statement provides the needed strengthening for the inductive hypothesis.

Proposition 22. *Let T be a standard binary tree, and T_u a subtree rooted at an internal node u of T. Assume that $(\langle \mathbf{X}_u, \mathbf{Y}_u \rangle, \langle \Gamma_u, \mathbf{D}_u \rangle) \sim \zeta_{T_u}$ and $\langle \mathbf{X}, \mathbf{Y} \rangle = \langle a\mathbf{X}_u b, c\mathbf{Y}_u d \rangle$, where a, b, c, d are fixed bit-strings. Then, for any protocol Π, we have*

$$\mathrm{I}(\mathbf{X}_u, \mathbf{Y}_u ; \Pi(\mathbf{X}, \mathbf{Y}) \mid \Gamma_u, \mathbf{D}_u) \quad \geq \frac{1}{N_{T_u} 2^{d_{T_u}+1}} \cdot \mathrm{h}^2(\Pi(a\alpha_{T_u} b, c\overline{\beta}_{T_u} d), \Pi(a\overline{\alpha}_{T_u} b, c\beta_{T_u} d)).$$

Proof. The proof is by induction on the depth d_{T_u} of T_u.

When $d_{T_u} = 0$ we have $f^{\bar{\wedge}}_{T_u}(x, y) = x \bar{\wedge} y$. This case was shown in [BYJKS04, Section 6], but we redo it here for completeness. First, notice that Γ_u is constant and thus the left-hand side can be written as $\mathrm{I}(X_u, Y_u ; \Pi(X, Y) \mid D_u)$. Expanding on values of D_u this is equal to

$$\tfrac{1}{2} \left(\mathrm{I}(Y_u ; \Pi(a0b, cYd) \mid D_u = \textsc{Alice}) + \mathrm{I}(X_u ; \Pi(aX_u b, c0d) \mid D_u = \textsc{Bob}) \right),$$

because given $D_u = \textsc{Alice}$ we have $X = 0$ and given $D_u = \textsc{Bob}$ we have $Y = 0$. Also, given $D_u = \textsc{Alice}$ we have $Y \in_R \{0, 1\}$ and thus the first term in the expression above can be written as $\mathrm{I}(Z ; \Pi(a0b, cZd))$, where $Z \in_R \{0, 1\}$. Now we apply Lemma 13 to bound this from below by $\mathrm{h}^2(\Pi(a0b, c0d), \Pi(a0b, c1d))$. Bounding the other term similarly and putting it all together we get

$$\mathrm{I}(X_u, Y_u ; \Pi(X, Y) \mid D_u) \tag{10}$$

$$\geq \tfrac{1}{2} \left(\mathrm{h}^2(\Pi(a0b, c0d), \Pi(a0b, c1d)) + \mathrm{h}^2(\Pi(a0b, c0d), \Pi(a1b, c0d)) \right) \tag{11}$$

$$\geq \tfrac{1}{4} \cdot \mathrm{h}^2(\Pi(a0b, c1d), \Pi(a1b, c0d)). \tag{12}$$

For the last inequality we used the triangle inequality of Hellinger distance (Proposition 14). Since $\langle \alpha_{T_u}, \beta_{T_u} \rangle = \langle 1, 1 \rangle$ this is the desired result.

Now suppose $d_{T_u} > 0$ and let $T_u = T_{u_1} \circ T_{u_2}$. Either $u_1 \in \mathrm{W}_{\Gamma_u}$ (i.e. u_1 is white), or $u_2 \in \mathrm{W}_{\Gamma_u}$. Thus, expanding on Γ_u, the left-hand side can be written as follows.

$$\begin{aligned}(13)\quad & \tfrac{N_{T_{u_1}}}{N_{T_u}} \cdot \mathrm{I}(\mathbf{X}_u, \mathbf{Y}_u\,;\, \Pi(a\mathbf{X}_u b, c\mathbf{Y}_u d) \,|\, \Gamma_u, u_1 \in \mathrm{W}_{\Gamma_u}, \mathbf{D}_u) \\ & \qquad + \tfrac{N_{T_{u_2}}}{N_{T_u}} \cdot \mathrm{I}(\mathbf{X}_u, \mathbf{Y}_u\,;\, \Pi(a\mathbf{X}_u b, c\mathbf{Y}_u d) \,|\, \Gamma_u, u_2 \in \mathrm{W}_{\Gamma_u}, \mathbf{D}_u).\end{aligned}$$

When u_1 is white, $\langle \mathbf{X}_{u_2}, \mathbf{Y}_{u_2} \rangle = \langle \alpha_{T_{u_2}}, \beta_{T_{u_2}} \rangle$, and $(\langle \mathbf{X}_{u_1}, \mathbf{Y}_{u_1} \rangle, \langle \Gamma_{u_1}, \mathbf{D}_{u_1} \rangle) \sim \zeta_{T_{u_1}}$. Similarly, given that u_2 is white, $\langle \mathbf{X}_{u_1}, \mathbf{Y}_{u_1} \rangle = \langle \alpha_{T_{u_1}}, \beta_{T_{u_1}} \rangle$, and $(\langle \mathbf{X}_{u_2}, \mathbf{Y}_{u_2} \rangle, \langle \Gamma_{u_2}, \mathbf{D}_{u_2} \rangle) \sim \zeta_{T_{u_2}}$. Thus, the above sum simplifies to

$$\begin{aligned}(14)\quad & \tfrac{N_{T_{u_1}}}{N_{T_u}} \cdot \mathrm{I}(\mathbf{X}_{u_1}, \mathbf{Y}_{u_1}\,;\, \Pi(a\mathbf{X}_{u_1}\alpha_{T_{u_2}} b, c\mathbf{Y}_{u_1}\beta_{T_{u_2}} d) \,|\, \Gamma_{u_1}, \mathbf{D}_{u_1}) \\ & \qquad + \tfrac{N_{T_{u_2}}}{N_{T_u}} \cdot \mathrm{I}(\mathbf{X}_{u_2}, \mathbf{Y}_{u_2}\,;\, \Pi(a\alpha_{T_{u_1}}\mathbf{X}_{u_2} b, c\beta_{T_{u_1}}\mathbf{Y}_{u_2} d) \,|\, \Gamma_{u_2}, \mathbf{D}_{u_2}).\end{aligned}$$

By induction, this is bounded from below by

$$\begin{aligned}(15)\quad & \tfrac{N_{T_{u_1}}}{N_{T_u}} \cdot \tfrac{1}{N_{T_{u_1}} 2^{d_{T_u}}} \cdot \mathrm{h}^2(\Pi(a\alpha_{T_{u_1}}\alpha_{T_{u_2}} b, c\overline{\beta}_{T_{u_1}}\beta_{T_{u_2}} d), \Pi(a\overline{\alpha}_{T_{u_1}}\alpha_{T_{u_2}} b, c\beta_{T_{u_1}}\beta_{T_{u_2}} d)) \\ & + \tfrac{N_{T_{u_2}}}{N_{T_u}} \cdot \tfrac{1}{N_{T_{u_2}} 2^{d_{T_u}}} \cdot \mathrm{h}^2(\Pi(a\alpha_{T_{u_1}}\alpha_{T_{u_2}} b, c\beta_{T_{u_1}}\overline{\beta}_{T_{u_2}} d), \Pi(a\alpha_{T_{u_1}}\overline{\alpha}_{T_{u_2}} b, c\beta_{T_{u_1}}\beta_{T_{u_2}} d)).\end{aligned}$$

Applying the cut-and-paste property (Proposition 14) of Hellinger distance this becomes

$$\begin{aligned}(16)\quad & \tfrac{N_{T_{u_1}}}{N_{T_u}} \cdot \tfrac{1}{N_{T_{u_1}} 2^{d_{T_u}}} \cdot \mathrm{h}^2(\Pi(a\alpha_{T_{u_1}}\alpha_{T_{u_2}} b, c\beta_{T_{u_1}}\beta_{T_{u_2}} d), \Pi(a\overline{\alpha}_{T_{u_1}}\alpha_{T_{u_2}} b, c\overline{\beta}_{T_{u_1}}\beta_{T_{u_2}} d)) \\ & + \tfrac{N_{T_{u_2}}}{N_{T_u}} \cdot \tfrac{1}{N_{T_{u_2}} 2^{d_{T_u}}} \cdot \mathrm{h}^2(\Pi(a\alpha_{T_{u_1}}\alpha_{T_{u_2}} b, c\beta_{T_{u_1}}\beta_{T_{u_2}} d), \Pi(a\alpha_{T_{u_1}}\overline{\alpha}_{T_{u_2}} b, c\beta_{T_{u_1}}\overline{\beta}_{T_{u_2}} d)).\end{aligned}$$

Now, since the square of Hellinger distance satisfies the (weak) triangle inequality (see Proposition 14), we have

$$(17)\quad \geq \tfrac{1}{N_{T_u} 2^{d_{T_u}+1}} \cdot \mathrm{h}^2(\Pi(a\alpha_{T_{u_1}}\overline{\alpha}_{T_{u_2}} b, c\beta_{T_{u_1}}\overline{\beta}_{T_{u_2}} d), \Pi(a\overline{\alpha}_{T_{u_1}}\alpha_{T_{u_2}} b, c\overline{\beta}_{T_{u_1}}\beta_{T_{u_2}} d)).$$

Recalling the Definition 17 of $\langle \alpha_T, \beta_T \rangle$ we get

$$(18)\quad = \tfrac{1}{N_{T_u} 2^{d_{T_u}+1}} \cdot \mathrm{h}^2(\Pi(a\alpha_T b, c\overline{\beta}_T d), \Pi(a\overline{\alpha}_T b, c\beta_T d)).$$

This completes the inductive proof. □

Corollary 23. *For any binary tree T in standard form*

$$\mathrm{IC}_{\zeta_T,\delta}(f_T^{\overline{\wedge}} \,|\, \Gamma, \mathbf{D}) \geq \frac{1}{4^{d_T}} \left(\tfrac{1}{2} - \sqrt{\delta} \right).$$

Proof. First apply Proposition 22 with the root of T as u and empty a, b, c, d.

$$\begin{aligned}
&\mathrm{IC}_{\zeta_T,\delta}(f_T^{\overline{\wedge}} \,|\, \Gamma, \mathbf{D}) \geq \tfrac{1}{2\cdot 4^{d_T}} \mathrm{h}^2(\Pi(\alpha_T, \overline{\beta}_T), \Pi(\overline{\alpha}_T, \beta_T)) \\
&\geq \tfrac{1}{2\cdot 4^{d_T}} \left(\tfrac{1}{2}\mathrm{h}^2(\Pi(\alpha_T, \overline{\beta}_T), \Pi(\overline{\alpha}_T, \overline{\beta}_T)) + \tfrac{1}{2}\mathrm{h}^2(\Pi(\alpha_T, \beta_T), \Pi(\overline{\alpha}_T, \beta_T)) \right) \\
&\geq \tfrac{1}{2\cdot 4^{d_T}} (1 - 2\sqrt{\delta}).
\end{aligned}$$

The second inequality is an application of the Pythagorean property of Hellinger distance. The last inequality follows from Proposition 18 and Proposition 14(4). □

4.5 Lower bounds for read-once boolean functions

In this section we use the main lemmas we have proved to obtain bounds for read-once boolean functions.

Corollary 24. *1. For any tree T in standard form,*

$$\mathrm{IC}_{\zeta_T,\delta}(f_T^{\overline{\wedge}} \,|\, \Gamma, \mathbf{D}) \geq \left(\tfrac{1}{2} - \sqrt{\delta} \right) \frac{N_T}{8^d_T}.$$

2. If, in addition, T is t-uniform,

$$\mathrm{IC}_{\zeta_T,\delta}(f_T^{\overline{\wedge}} \,|\, \Gamma, \mathbf{D}) \geq \left(\tfrac{1}{2} - \sqrt{\delta} \right) \frac{(t-1)^{d_T}}{4^{d_T}}.$$

Proof. Let Π be a δ-error protocol for $f_T^{\overline{\wedge}}$. Lemma 20 holds for any Π, therefore

$$\mathrm{IC}_{\zeta_T,\delta}(f_T^{\overline{\wedge}} \,|\, \Gamma, \mathbf{D}) \geq \rho_T \cdot \min_{S \in \mathrm{FBS}_T} \mathrm{IC}_{\zeta_S,\delta}(f_S^{\overline{\wedge}} \,|\, \Gamma, \mathbf{D}).$$

We now use the bound from Corollary 23 to obtain

$$\mathrm{IC}_{\zeta_T,\delta}(f_T^{\overline{\wedge}} \,|\, \Gamma, \mathbf{D}) \geq \rho_T \cdot \frac{1}{4^{d_T}} \left(\tfrac{1}{2} - \sqrt{\delta} \right).$$

For (1), we now use Lemma 21 to bound ρ_T from below. For (2), we can compute ρ_T exactly to be $(t-1)^{d_T}$. □

We are now ready to prove Theorem 6. Let $T = T_1 \circ \cdots \circ T_k$. Apply Theorem 15 with f being the k-variate NAND, and, for each $j \in [k]$, h_j and ζ_j being $f_{T_j}^{\overline{\wedge}}$ and ζ_{T_j}, respectively. Applying the first part of Corollary 24 to each of the trees T_j gives the first part of Theorem 6, and applying the second part of Corollary 24 gives the second part of Theorem 6.

5 Lower bound for read-once threshold functions

A *threshold* gate, denoted T_k^n for $n > 1$ and $1 \leq k \leq n$, receives n boolean inputs and outputs '1' if and only if at least k of them are '1'. A *threshold circuit* is a rooted tree in which every leaf corresponds to a distinct input variable, and every gate being a threshold gate. A *read-once threshold function* f_E is a function that can be represented by a threshold circuit E. As before, we define $f_E^{\overline{\wedge}}$ and we want to lower bound $R_\delta(f_E^{\overline{\wedge}})$, when E is a threshold circuit.

Proposition 25. *For any threshold circuit E, there is an* AND/OR *tree T such that, for $g \in \{\wedge, \vee\}$, (1) $R_\delta(f_T^g) \leq R_\delta(f_E^g)$, (2) $N_T \geq N_E/2^{d_E}$, and (3) $d_T = d_E$.*

Proof. We define T by recursion on d_E. When $d_E = 0$ we set $T = E$. Otherwise, let $E = E_1 \circ \cdots \circ E_n$, and assume $N_{E_1} \geq \cdots \geq N_{E_n}$. Suppose the gate on the root is T_k^n. We consider cases on k. (1) If $1 < k \leq n/2$, build $T_1, \ldots, T_{n-k+1}$ recursively, set $T = T_1 \circ \cdots \circ T_{n-k+1}$, and put an $\vee$-gate on the root. (2) If $n/2 < k < n$, build $T_1, \ldots, T_k$ recursively, set $T = T_1 \circ \cdots \circ T_k$, and put an $\wedge$-gate on the root. (3) Otherwise, if $k = 1$ or $k = n$, the threshold gate is equivalent to an $\vee$ or $\wedge$-gate respectively. We build $T_1, \ldots, T_n$ recursively and we set $T = T_1 \circ \cdots \circ T_n$. The gate on the root remains as is.

Properties (2) and (3) are easily seen to hold. For (1), it is not hard to show that a protocol for f_E^g can be used to compute f_T^g. Alice and Bob need only to fix appropriately their inputs in the subtrees that where cut of from E. If an input bit belongs to a subtree T_j that was cut of in case (1), then Alice and Bob set their inputs in T_j to '0'. If T_j was cut of in case (2), then Alice and Bob set their inputs in T_j to '1'. Afterwords, they simulate the protocol for f_E^g. □

The tree T in the above proposition may not be a canonical representation of some function. However, transorming to the canonical representation will only decrease its depth, and thus strengthen our lower bound. Thus, by this Proposition and Theorem 1 we obtain Theorem 3 as a corollary.

Acknowledgments

The authors thank Fengming Wang for helpful discussions.

References

[BYJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *CCC '02: Proceedings of the 17th IEEE Annual Conference on Computational Complexity*, page 93, Washington, DC, USA, 2002. IEEE Computer Society.

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

[CKS03] Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party commnunication complexity of set disjointness. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity, CCC2003*, 2003.

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

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience, 2006.

[HNW93] Rafi Heiman, Ilan Newman, and Avi Wigderson. On read-once threshold formulae and their randomized decision tree complexity. *Theor. Comput. Sci.*, 107(1):63–76, 1993.

[JKR09] T. S. Jayram, Swastik Kopparty, and Prasad Raghavendra. Personal communication, 2009.

[JKS03] T. S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *STOC*, pages 673–682. ACM, 2003.

[KN06] Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, New York, NY, USA, 2006.

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

[Raz92] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[SS02] Michael Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 360–369, New York, NY, USA, 2002. ACM.

[SW86] Michael Saks and Avi Wigderson. Probabilistic boolean decision trees and the complexity of evaluating game trees. In *SFCS '86: Proceedings of the 27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, pages 29–38, Washington, DC, USA, 1986. IEEE Computer Society.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing(preliminary report). In *STOC '79: Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213, New York, NY, USA, 1979. ACM.