INFORMATION THEORY METHODS IN COMMUNICATION COMPLEXITY

BY NIKOLAOS LEONARDOS

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Michael Saks

and approved by

New Brunswick, New Jersey

JANUARY, 2012

ABSTRACT OF THE DISSERTATION

Information theory methods in communication complexity

by Nikolaos Leonardos

Dissertation Director: Michael Saks

This dissertation is concerned with the application of notions and methods from the field of information theory to the field of communication complexity. It consists of two main parts.

In the first part of the dissertation, 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 dissertation 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 $d(f)$ is the depth of the formula, and they are optimal up to the constant in the base of the denominator.

In the second part of the dissertation, we explore the applicability of the information-theoretic method in the number-on-the-forehead model. The work of Bar-Yossef, Jayram, Kumar & Sivakumar [BYJKS04] revealed a beautiful connection between Hellinger distance and two-party randomized communication protocols. Inspired by their work and motivated by the open questions in the number-on-the-forehead model, we introduce the notion of Hellinger volume. We show that it lower bounds the information cost of multi-party protocols. We provide a small toolbox that allows one to manipulate several Hellinger volume terms and also to lower bound a Hellinger volume when the distributions involved satisfy certain conditions. In doing so, we prove a new upper bound on the difference between the arithmetic mean and the geometric mean in terms of relative entropy. Finally, we show how to apply the new tools to obtain a lower bound on the informational complexity of the AND_k function.

Acknowledgements

First, I would like to thank my advisor, Michael Saks. Mike has been teaching me all these years how to do research and how to think about a particular problem. By observing him teach, I learned to value intuition and try hard to reveal it when I'm presenting a mathematical proof. I'm glad he was my advisor.

I wish also to thank the rest of the members of my dissertation committee; Mark Braverman, William Steiger, Mario Szegedy. I thank our Graduate Program Director, William Steiger, for teaching me a theorem in geometry, and Mario Szegedy, who I'll remember for his passion for research and teaching.

I also thank Eric Allender, for being a member of my examination committee and being a Department Chair that students could rely on, and Martin Farach-Colton, for teaching me Algorithms in a lively and engaging way.

Going back to the time before I enrolled Rutgers, I would like to thank professor Stathis Zachos in the National Technical University of Athens (NTUA), who was the person that introduced me to Theoretical Computer Science. I think he was right to recommend reading Enderton's "Introduction to mathematical Logic." I also thank professor Aris Pagourtzis, Panos Cheilaris, and all my friends in the Computation and Reasoning Lab in NTUA.

Very important in forming a solid background in Theoretical Computer Science was also the teachings of the professors in MPLA, a graduate program in Greece. I wish to thank specially Yiannis Moschovakis, Elias Koutsoupias, and Constantinos Dimitracopoulos.

Finally, I wish to thank the people that are close to me; my parents, Panos and Eleni, for encouraging me to pursue education and knowledge, and Christina, for her love and support.

Table of Contents

Chapter 1

Introduction

The communication complexity model was introduced in [Yao79]. The standard variation involves two parties, Alice and Bob, who wish to compute a function $f : X \times Y \rightarrow Z$, where X, Y, and Z are finite sets. Alice knows $x \in X$, but has no knowledge of Bob's input. Similarly, Bob knows $y \in Y$, but has no knowledge of x. To correctly determine $f(x, y)$ they need to communicate. We are interested in the minimum amount of communication needed. In the first part of the dissertation, Chapter 3, we work with the randomized model, where Alice and Bob are equipped with random strings and have the power of making random choices. Furthermore, we only require to compute $f(x, y)$ correctly with probability 2/3. In the second part, Chapter 4, we consider the number-on-the-forehead model, introduced in [CFL83]. In this variation, there are k players, $P_1, \ldots, P_k$, that wish to compute a function $f : X_1 \times \cdots \times X_k \rightarrow Z$. The inputs to the players are $x_i \in X_i$ for $i \in \{1, \ldots, k\}$. In contrast to the previous model, player P_i, knows all inputs except his own.

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 & 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]. The first application of information-theoretic methods in communication complexity lower bounds can be traced to Ablayev [Abl96].

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 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, Kumar & Sivakumar [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*. These are 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. Heiman, Newman & Wigderson [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 $f^\vee$.

Denote by $R_\delta(f)$ the δ-error randomized communication complexity of f (see Section 2.2 and the paragraph on "communication complexity" in Section 3.1 for more details). We prove 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 with $d(f) \geq 1$,*

$$\max\{R_\delta(f^\wedge), R_\delta(f^\vee)\} \geq (1 - 2\sqrt{\delta}) \cdot \frac{n(f)}{8^{d(f)}}.$$

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 (in which each non-leaf node has t children and all leaves are at the same depth, and 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 Saks & Wigderson [SW86]) that f has a randomized decision tree algorithm using an expected number $O(n(f)/c^{d(f)})$ of queries, 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, 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 $d \geq 1$,*

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

Independently, Jayram, Kopparty & Raghavendra [JKR09], also using the informational complexity approach, obtained the weaker bound $\frac{(1-2\sqrt{\delta})\cdot n(f)}{d(f)!16^{d(f)}}$.

As a simple corollary of Theorem 1 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 with $d(f) \geq 1$,*

$$\max\{R_\delta(f^\wedge), R_\delta(f^\vee)\} \geq (1 - 2\sqrt{\delta}) \cdot \frac{n(f)}{16^{d(f)}}.$$

This result should be compared with the result of Heiman, Newman & 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, $(\mathrm{AND}_n)^n$ is equal to AND_{2n} that has randomized decision tree complexity $\Theta(n)$ but communication complexity 2.) 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.

Our results are interesting only for formulae of small depth. For example, for f that is represented by a binary uniform tree $n(f)/8^{d(f)} < 1$, while there is a simple $\sqrt{n(f)}$ lower bound that follows by embedding either a $\sqrt{n(f)}$-wise OR or a $\sqrt{n(f)}$-wise AND. Binary uniform trees require $\Omega(\sqrt{n(f)})$ communication even for quantum protocols. This is because $\sqrt{n(f)}$-wise PARITY can be embedded in such a tree (see Farhi, Goldstone & Gutmann [FGG08]), and then the bound follows from the lower bound for the generalized inner product function (see Cleve, Dam, Nielsen & Tapp [CDNT98] and Kremer [Kre95]). This can also be shown by methods of Lee, Shraibman & Zhang [LSZ09], which seem more promising towards a lower bound on the quantum communication complexity of arbitrary AND/OR trees.

Finally, we consider the more general setting, where $f(x, y)$ is a two-party read-once formula with its variables partitioned arbitrarily between Alice and Bob. This situation includes the case where the function is of the form $f^\vee$ or $f^\wedge$ and the variable partition is the natural one indicated earlier. As the case $f = \mathrm{AND}_n$ shows, we don't have a lower bound on $R_\delta(f)$ of the form $n(f)/c^{d(f)}$. However we can get an interesting general lower bound.

Consider the deterministic simultaneous message model, which is perhaps the weakest non-trivial communication complexity model. In this model Alice and Bob are trying to communicate $f(x, y)$ to a third party, the referee. Alice announces some function value $m_A(x)$ and simultaneously Bob announces a function value $m_B(y)$, and together $m_A(x)$ and $m_B(y)$ are enough for the referee to determine $f(x, y)$. The deterministic simultaneous message complexity, denoted $D^{||}(f)$, is the minimum number of bits (in worst case) that must be sent by Alice and Bob so that the referee can evaluate f. As a consequence of Theorem 15 we prove the following.

Theorem 4. *For any two-party read-once function f with $d(f) \geq 1$,*

$$R_\delta(f) \geq (1 - 2\sqrt{\delta}) \cdot \frac{D^{||}(f)}{d(f) \cdot 8^{d(f)-1}}.$$

In the second part of the dissertation, Chapter 4, we consider the number-on-the-forehead (NOF) model. Proving lower bounds on the number-on-the-forehead (NOF) communication complexity of functions, is one of the most important research areas in the theory of communication complexity, The NOF model was introduced in [CFL83], where it was used to prove lower bounds for branching programs. Subsequent papers revealed connections of this model to circuit complexity [BT94, HG90, Nis94, NW91] and proof complexity [BPS05]. In particular, an explicit function which requires super-polylogarithmic complexity in the NOF model with polylogarithmically many players would give an explicit function outside of the circuit complexity class ACC^0. Regarding proof complexity, it was shown in [BPS05], that $n^{\Omega(1)}$ lower bounds for k-party NOF disjointness imply

$2^{n^{\Omega(1)}}$ proof-size lower bounds for tree-like, degree $k-1$, threshold systems. Also, $\omega(\log^4 n)$ lower bounds for 3-party NOF disjointness imply $n^{\omega(1)}$ proof-size lower bounds for tree-like Lovász-Schrijver proof systems.

There are explicit functions known with NOF complexity in $\Omega(n/2^k)$ [BNS92, CT93, Raz00, FG05], which become trivial for logarithmic number of players. For disjointness, the general known lower bounds for k-players are of the form $n^{1/k}/2^{2^k}$ [LS09a, CA08] and $2^{\Omega(\sqrt{\log n}/\sqrt{k})-k}$ [BHN09]. It is interesting to note that the NOF complexity of disjointness is bounded above by $O(k^2 n/2^k)$, as follows from the work of Grolmusz [Gro94].

Most of the lower bounds are obtained by an upper bound on discrepancy, in a manner that was first shown in [BNS92]. In this dissertation we are interested in how information-theoretic methods might be applied to the NOF model. The first use of information theory in communication complexity lower bounds can be traced to [Abl96]. In [CSWY01] the notions of information cost and informational complexity were defined explicitly. Building on their work, a very elegant information-theoretic framework for proving lower bounds in randomized number-in-hand (NIH) communication complexity was established in [BYJKS04].

In [BYJKS04] a proof of the linear lower bound for two-party disjointness is given. The proof has two main stages. In the first stage, a direct-sum theorem for informational complexity is shown, which says that the informational complexity of disjointness, $\mathsf{DISJ}_{n,2}(x,y) = \bigvee_{j=1}^{n} \mathsf{AND}_2(x_j, y_j)$, is lower bounded by n times the informational complexity of the binary AND_2 function. Although it is not known how to prove such a direct-sum theorem directly for the classical randomized complexity, Bar-Yossef et al. prove it for the informational complexity with respect to a suitable distribution. A crucial property of the distribution is that it is over the zeroes of disjointness. At this point we should point out a remarkable characteristic of the method: even though the information cost of a protocol is analyzed with respect to a distribution over zeroes only, the protocol is required to be correct over all inputs. This requirement is essential in the second stage, where a constant lower bound is proved on the informational complexity of AND_2. This is achieved using properties of the Hellinger distance for distributions. Bar-Yossef et al. reveal a beautiful connection between Hellinger distance and NIH communication protocols. (More properties of Hellinger distance relative to the NIH model have been established in [Jay09].)

In this work we provide tools for accomplishing the second stage in the NOF model. We introduce the notion of Hellinger volume of $m \geq 2$ distributions and show that it can be useful for proving lower bounds on informational complexity in the NOF model, just as Hellinger distance is useful in the NIH model. However, as we point out in the last section, there are fundamental difficulties in proving a direct-sum theorem for informational complexity in the NOF model. Nevertheless, we believe that Hellinger volume and the related tools we prove, could be useful in an information-theoretic attack at NOF complexity.

The work in both parts of the dissertation is closely related to the work of Bar-Yossef, Jayram, Kumar & Sivakumar [BYJKS04]. In particular, we use their definition of *information cost* and *conditional information cost*. In more recent work by Barak, Braverman, Chen & Rao [BBCR10], the terms *external information cost* and *internal information cost* were introduced. External information cost quantifies the amount of information learned by an outside observer of the communication about the inputs, and it coincides with the definition of information cost in [BYJKS04]. Internal information cost was employed in [BBCR10] to establish direct-sum theorems for randomized communication complexity. It quantifies the amount of information the players learn about the other player's input upon execution of the protocol. In subsequent work by Braverman & Rao [BR11], the internal information cost of computing a function f according to a fixed distribution, was shown to be exactly equal to the amortized communication complexity of computing many copies of f.

The results in the first part of this dissertation (Chapter 3) have been presented in 2009, in the 24th IEEE Conference on Computational Complexity [LS09b], and invited in the issue entitled "Selected papers from the 24th Annual IEEE Conference on Computational Complexity (CCC 2009)" of Computational Complexity Journal [LS10]. The results in the second part (Chapter 4) have been submitted for publication.

Chapter 2
Preliminaries and previous work

In this chapter we state the basic definitions and facts of information theory that we will make use of and define the communication complexity models that we will be working with. Also, we discuss why information theory is relevant in the study of communication complexity and we provide the results from the work of Bar-Yossef, Jayram, Kumar & Sivakumar [BYJKS04] on which we build.

2.1 Information theory

The following definitions and facts can be found in the textbook by Cover & Thomas [CT06, Chapter 2],

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. Let $(\Omega_1, \zeta_1), \dots, (\Omega_n, \zeta_n)$ be 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 $\omega = (\omega_1, \dots, \omega_n) \in \Lambda$, $\nu(\omega) = \prod_{j=1}^{n} \zeta_j(\omega_j)$. 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.

Entropy and mutual information. 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, the *conditional entropy of X given A*, and the *conditional entropy of X given Y* are respectively (we use log for $\log_2$)

$$\begin{aligned} \mathrm{H}(X) &= -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log \Pr[X = x], \\ \mathrm{H}(X \mid A) &= -\sum_{x \in \mathcal{X}} \Pr[X = x \mid A] \cdot \log \Pr[X = x \mid A], \\ \mathrm{H}(X \mid Y) &= \sum_{y \in \mathcal{X}} \Pr[Y = y] \cdot \mathrm{H}(X \mid Y = y). \end{aligned}$$

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

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

The *relative entropy* or *divergence* of distributions P and Q over Ω is

$$D(P \| Q) = \sum_{x \in \Omega} P(x) \log \frac{P(x)}{Q(x)}.$$

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

Proposition 5. *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 6 ([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 the definition of mutual conditional information

$$\mathrm{I}(\mathbf{Z} \,;\, \Pi \mid \mathbf{D}) = \mathrm{H}(\mathbf{Z} \mid \mathbf{D}) - \mathrm{H}(\mathbf{Z} \mid \Pi, \mathbf{D}).$$

By Proposition 5(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. □

2.2 Communication complexity

For a proper introduction to the subject of communication complexity the reader should consult the textbook by Kushilevitz & Nisan [KN06].

Two-party private-coin model. The two-party private-coin randomized communication model was introduced by Yao [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 $\Pi_{\text{out}}(x, y)$ the outcome of the protocol. We call Π a *δ-error protocol for f* if, for all $\langle x, y \rangle$, $\Pr[\Pi_{\text{out}}(x, y) = f(x, y)] \geq 1 - \delta$. 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.

The number-on-the-forehead model. The multi-party private-coin randomized number-on-the-forehead communication model was introduced by Chandra, Furst & Lipton [CFL83]. There are k players, numbered $1, \ldots, k$, trying to compute a function $f : \mathcal{Z} \to \{0, 1\}$, where $\mathcal{Z} = \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_k$. On input $z \in \mathcal{Z}$, player j receives input z_j (conceptually, placed on his forehead), but he has access only to $z^{-j} = (z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_k)$. They wish to determine $f(z)$, by broadcasting messages according to a protocol Π. Let the random variable $\Pi(z)$ denote the transcript of the communication on input z (where the probability is over the random coins of the players) and $\Pi_{\text{out}}(z)$ the outcome of the protocol. We call Π a *δ-error protocol for f* if, for all z, $\Pr[\Pi_{\text{out}}(z) = f(z)] \geq 1 - \delta$. The *communication cost of* Π is $\max |\Pi(z)|$, where the maximum is over all inputs z and over all coin tosses of the players. The *δ-error randomized communication complexity of f*, denoted $R_\delta(f)$, is the cost of the best δ-error protocol for f.

2.3 Communication complexity lower bounds via information theory

The informational complexity paradigm, introduced by Chakrabarti, Shi, Wirth & Yao [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 inputs chosen randomly according to the distribution. 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 respectively. 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. The relevance of informational complexity comes from the following proposition.

Proposition 7. $\mathrm{IC}_{\zeta,\delta}(f) \geq R_\delta(f)$.

Proof. For any protocol Π,

$$\mathrm{IC}_{\zeta,\delta}(f) \leq \mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y})) = H(\Pi(\mathbf{X}, \mathbf{Y})) - H(\Pi(\mathbf{X}, \mathbf{Y}) | \mathbf{X}, \mathbf{Y}).$$

Applying in turn parts (1) and (2) of Proposition 5 gives

$$\mathrm{IC}_{\zeta,\delta}(f) \leq H(\Pi(\mathbf{X}, \mathbf{Y})) \leq R_\delta(f). \qquad \square$$

Mutual information may be easier to handle if one conditions on the appropriate random variables. To that end, Bar-Yossef et al. [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 $\mathbf{D}$ with respect to ζ, denoted $\mathrm{IC}_{\zeta,\delta}(f \mid \mathbf{D})$, is $\min_\Pi \mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \mid \mathbf{D})$, where the minimum is over all δ-error randomized protocols for f. Conditional informational complexity also provides a lower bound on randomized communication complexity.

We now give an alternate definition for informational complexity that we will use when considering the number-on-the-forehead model. It is not hard to see that conditional informational complexity as defined in the previous paragraph is essentially equivalent to the following definition of informational complexity with respect to a collection of distributions.

For a collection of distributions $\eta = \{\zeta_1, \ldots, \zeta_k\}$, we define the *$\delta$-error informational complexity* of f with respect to η, denoted $\mathrm{IC}_{\eta,\delta}(f)$, to be $\mathbf{E}_j[\mathrm{IC}_{\zeta_j,\delta}(f)]$, where j is a random variable uniformly distributed over $[k]$.

Remark. As discussed in the introduction, the authors of [BBCR10] define the *external information cost* to be $\mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}))$ and they introduce the *internal information cost* which they define as $\mathrm{I}(\mathbf{X} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y}) + \mathrm{I}(\mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X})$.

2.4 The methodology of Bar-Yossef, Jayram, Kumar & Sivakumar

Bar-Yossef, Jayram, Kumar & Sivakumar [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 Bar-Yossef et al. [BYJKS04] that we will use.

Their methodology has two main parts. In the first part they make use of Proposition 6 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 8 (Sensitive input)**.** *Consider $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \to \mathbb{R}$, a family of functions $\mathcal{H} = \langle h_j : \mathcal{Z}_j \to \mathcal{S}_j \rangle_{j\in[n]}$, and $\mathbf{z} = \langle z_1, \ldots, z_n \rangle \in \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_n$. For $j \in [n]$, $u \in \mathcal{Z}_j$, let $\mathbf{z}[j,u] = \langle z_1, \ldots, z_{j-1}, u, z_{j+1}, \ldots, z_n \rangle$. 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 9 (Collapsing distribution, Bar-Yossef et al. [BYJKS04])**.** *Let $f, \mathcal{H}$ be as in Definition 8. 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 10 (Bar-Yossef et al. [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 $\mathrm{IC}_{\zeta^n,\delta}(f^h \,|\, \mathbf{D}) \geq n \cdot \mathrm{IC}_{\zeta,\delta}(h \,|\, D)$.

Defining a distribution ζ satisfying the two requirements asked in Theorem 10, moves the attention from $\mathrm{IC}_{\zeta^n,\delta}(f^h \,|\, \mathbf{D})$ to $\mathrm{IC}_{\zeta,\delta}(h \,|\, D)$. For example, in Bar-Yossef et al. [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 one only has 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 11. *(Hellinger distance.) The* Hellinger distance *between probability distributions P and Q on a domain Ω is defined by*

$$\mathrm{h}(P,Q) = \sqrt{\tfrac{1}{2}\textstyle\sum_{\omega\in\Omega}\left(\sqrt{P_\omega} - \sqrt{Q_\omega}\right)^2}.$$

We write $\mathrm{h}^2(P,Q)$ *for* $(\mathrm{h}(P,Q))^2$.

Lemma 12 (Bar-Yossef et al. [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 13 (Properties of Hellinger distance, Bar-Yossef et al. [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:*

$$\mathrm{h}^2(P,Q) + \mathrm{h}^2(Q,R) \geq \tfrac{1}{2}\mathrm{h}^2(P,R).$$

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 12 we are left with a sum of Hellinger distance terms, which we need to lower bound. Applying properties 1 and 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 include terms to which Property 4 can be applied.

Chapter 3

Read-once functions

In this chapter we prove our main theorem, Theorem 15, from which the theorems 2, 3, and 4 that were stated in the introduction follow.

3.1 Notation, terminology, and preliminaries

In the first section we establish notation and terms that we will use to describe the basic objects that we will be dealing with.

Definitions pertaining to rooted trees. All trees in this work 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, $\mathrm{path}(u)$ 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 $j \in \{1, \ldots, k\}$, 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 node 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. We 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 $i \in [n]$, $h_i : \mathcal{Z}_i \to \mathcal{S}_i$. For $\mathcal{H} = \langle h_1, \ldots, h_n \rangle$, let $f^{\mathcal{H}} : \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_n \to \mathbb{R}$ denote 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 $\{\mathrm{AND}, \mathrm{OR}, \mathrm{NAND}, \mathrm{NOR}\}$. We write f_C for the function represented by a tree circuit C. 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 work 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 easy to show 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.

Communication problems associated with 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

$$f^g(x,y) = f(g(x_1,y_1), \ldots, g(x_n,y_n)).$$

Our goal is to prove Theorems 1 and 2, which say that for any read-once boolean function f, either $f^{\vee}$ or $f^{\wedge}$ has high randomized communication cost. To do this it will be more convenient to consider $f^{\bar{\wedge}}$ for functions f that come from trees using only NAND gates. We first prove the following lemma.

For $f_1, f_2 : \{0,1\}^n \to \{0,1\}$, we write $f_1 \equiv f_2$ when

$$(\exists \sigma \in \{0,1\}^n)(\forall x \in \{0,1\}^n)(f_1(x) = f_2(\sigma \oplus x)),$$

where $\sigma \oplus x$ is the bitwise XOR of σ and x.

Lemma 14. *Let C be an* AND/OR *tree in canonical form and let T be the underlying tree. Then, $f_C \equiv f_T$ when the root of C is labeled by an* OR *gate, and $f_C \equiv \neg f_T$ when the root of C is labeled by an* AND *gate.*

Proof. We proceed by induction on d_T. When $d_T = 1$, the case with an AND at the root is trivial. For OR we observe that $f_C(x) = \bigvee_j x_j = \neg \bigwedge_j^n \neg x_j = f_T(\neg x)$. Now suppose $d_T > 1$. Let $C = C_1 \wedge \cdots \wedge C_k$ and recall that C is in canonical form; thus, each C_j has an OR at the root. It follows by induction that $f_C(x) \equiv \bigwedge_j f_{T_j} = \neg f_T(x)$. If $C = C_1 \vee \cdots \vee C_k$, then we have $f_C = \bigvee_j f_{C_j} = \neg \bigwedge_j \neg f_{C_j} \equiv \neg \bigwedge_j f_{T_j} = f_T$. $\square$

Our lower bounds follow from the following main theorem.

Theorem 15. *1. Let T be a tree in standard form with $d_T \geq 1$.*

$$R_\delta(f_T^{\bar{\wedge}}) \geq (2 - 4\sqrt{\delta}) \cdot \frac{N_T}{8^{d_T}}.$$

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

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

To deduce Theorems 1 and 2 we use the following proposition.

Proposition 16. *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) $N_T \geq n(f)/2$, (3) $d_T \leq d(f)$. Moreover, if the canonical representation of f is a uniform tree, $N_T = n(f)$.*

Proof. Let C be the representation of f in canonical form. Define tree circuits C_1 and C_2 as follows. To obtain C_1 delete all leaves that feed into $\wedge$ gates, and introduce a new variable for any node that becomes a leaf. Let C_1 be the canonical form of the resulting tree. 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. If the underlying tree of C is uniform, then one of C_1, C_2 will have $n(f)$ leaves; so in the uniform case we have $N_T = n(f)$. Condition (1) follows immediately from the following claim.

Claim 17. *(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 $f_1^\wedge$. In particular, given an input (x, y) to $f_1^\wedge$ Alice and Bob can—without any communication—construct input (x', y') to $f^\wedge$ such that $f^\wedge(x', y') = f_1^\wedge(x, y)$. This is done 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 $f^\wedge(x', y') = f_1^\wedge(x, y)$. The second part of the claim follows from Lemma 14. Parts (3) and (4) follow similarly. □

3.2 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 $f_T^{\bar{\wedge}}$ would be to apply the following straightforward generalization of Theorem 10.

Theorem 18. *Consider a function $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \to \{0,1\}$, a family of functions $\mathcal{H} = \langle h_j : \mathcal{X}_j \times \mathcal{Y}_j \to \mathcal{S}_j \rangle_{j\in[n]}$, and random variables $\mathbf{X} = \langle X_1, \ldots, X_n\rangle \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n, \mathbf{Y} = \langle Y_1, \ldots, Y_n\rangle \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n, \mathbf{D} = \langle D_1, \ldots, D_n\rangle$, 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 18 to the function $f_T^{\bar{\wedge}}$, with f the n-bit NAND and $h_j = f_{T_j}$, for $j \in [n]$. However, this won't take us very far. The problem is that if μ—the marginal distribution of $\langle \mathbf{X}, \mathbf{Y}\rangle$—is collapsing for f_T, 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 18 repeatedly. This problem arose in Jayram, Kumar & Sivakumar [JKS03] when studying the function $\mathrm{TRIBES}_{m,n}(\mathbf{x}, \mathbf{y}) = \bigwedge_{k=1}^{m} \mathrm{DISJ}_n(\mathbf{x}_k, \mathbf{y}_k) = \bigwedge_{k=1}^{m} \bigvee_{j=1}^{n} (x_{kj} \wedge y_{kj})$. Jayram et al. [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.

3.2.1 Further definitions on trees

We proceed with definitions of 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

$$\mathrm{IC}_{\zeta,\delta}(f_T^{\bar{\wedge}} \mid \Gamma, \mathbf{D}) = \min_{\Pi} \mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi(\mathbf{X}, \mathbf{Y}) \mid \Gamma, \mathbf{D}).$$

Definition 19. *(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.*

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.

Consider now a t-uniform tree T, colored properly by γ. Each white node has exactly one red child that is the root of a red binary subtree. For $t > 2$ there will be two kinds of white leaves: those that have no red nodes on the path that connects them to the root, and those that have at least one red node on that path. Notice that the union of a white leaf of the first kind, the corresponding root-to-leaf path, and the red binary subtrees that are "hanging" from the white nodes on the path, form a full binary subtree S of T. Furthermore, the restriction of γ on S, denoted γ_S, is a valid coloring for S.

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 $S \propto \gamma$, if S has exactly one white leaf. (Notice that, since γ is valid, this leaf would have to be a principal leaf. Thus, $S \propto \gamma$ is equivalent to saying that the restriction of γ on V_S is a valid coloring for S.)

(5) Define $\mathrm{FBS}_T(\gamma) = \{S \in \mathrm{FBS}_T \mid S \propto \gamma\}$. 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 $\langle \mathrm{W}_\gamma \cap V_S, \mathrm{R}_\gamma \cap V_S \rangle$.

3.2.2 The input distribution

Our proof will have two main components, analogous to the ones in Jayram et al. [JKS03]. The distribution over the inputs that we shall define is carefully chosen so that each component of the proof can be carried out.

In the first part (Section 3.2.3) we prove a direct-sum theorem for arbitrary trees. Given an arbitrary tree T in standard form, we show how the information cost of a protocol for $f_T^{\bar{\wedge}}$ can be decomposed into a sum of information costs that correspond to full binary subtrees of T. In the second part of the proof (Section 3.2.4) we provide a lower bound on the informational complexity of $f_S^{\bar{\wedge}}$, where S is an arbitrary binary tree in standard form.

For a uniform binary tree with N_S leaves, there is a natural distribution for which one can prove an $\Omega(\sqrt{N_S})$ lower bound on information cost. However, this distribution is not useful for us because it does not seem to be compatible with the first part of the proof. It turns out that for our purposes it is sufficient to prove a much weaker lower bound on the information cost for binary trees, of the form $\Omega(1/c^d)$ for some fixed $c > 0$, which will be enough to give a lower bound of $\Omega(n/c^d)$ on the communication complexity for general trees. The distribution for binary trees that we choose gives such a bound and is also compatible with the first part of the proof. This allows us to show that the information cost of a tree of depth d is at least $\frac{n}{2^d}B(d)$, where $B(d)$ is a lower bound on the information cost of (a communication protocol on) a depth-d binary tree.

Given an arbitrary tree T in standard form, we now define a distribution over inputs to Alice and Bob for $f_T^{\bar{\wedge}}$.

First, we associate to each standard binary tree S a special input $\langle \alpha_S, \beta_S \rangle$. We will be interested in the value $f_S^{\bar{\wedge}}(\alpha_S, \beta_S)$. These inputs, which now seem arbitrary, introduce structure in the final distribution. This structure is crucial for the effectiveness of the second part of our proof.

Definition 20. *We define input $\langle \alpha_S, \beta_S \rangle$ to $f_S^{\bar{\wedge}}$ for a standard binary tree S. The definition is recursive on the depth d_S of the tree.*

$$\langle \alpha_S, \beta_S \rangle = \begin{cases} \langle 1, 1 \rangle & \text{if } d_S = 0, \\ \langle \alpha_{S_1}\overline{\alpha}_{S_2}, \overline{\beta}_{S_1}\beta_{S_2} \rangle & \text{if } S = S_1 \circ S_2. \end{cases}$$

We will need the following property of $\langle \alpha_S, \beta_S \rangle$.

Proposition 21. *For a standard binary tree S with $d_S > 0$, $f_S^{\bar{\wedge}}(\alpha_S, \beta_S) = f_S^{\bar{\wedge}}(\overline{\alpha}_S, \overline{\beta}_S) = 0$ and $f_S^{\bar{\wedge}}(\alpha_S, \overline{\beta}_S) = f_S^{\bar{\wedge}}(\overline{\alpha}_S, \beta_S) = 1$.*

Proof. The proof is by induction on d_S.

For $d_S = 1$ the (unique) tree results in the function $f_S^{\bar{\wedge}}(x_1x_2, y_1y_2) = (x_1 \bar{\wedge} y_1) \bar{\wedge} (x_2 \bar{\wedge} y_2)$. Clearly,

$$f_S^{\bar{\wedge}}(\alpha_S, \beta_S) = f_S^{\bar{\wedge}}(10, 01) = 0, \quad f_S^{\bar{\wedge}}(\overline{\alpha}_S, \overline{\beta}_S) = f_S^{\bar{\wedge}}(01, 10) = 0\,;$$
$$f_S^{\bar{\wedge}}(\alpha_S, \overline{\beta}_S) = f_S^{\bar{\wedge}}(10, 10) = 1, \quad f_S^{\bar{\wedge}}(\overline{\alpha}_S, \beta_S) = f_S^{\bar{\wedge}}(01, 01) = 1.$$

Suppose $d_S > 1$ and let $S = S_1 \circ S_2$. We have $f_S(\alpha_S, \beta_S) = f_{S_1}^{\bar{\wedge}}(\alpha_{S_1}, \overline{\beta}_{S_1}) \bar{\wedge} f_{S_2}^{\bar{\wedge}}(\overline{\alpha}_{S_2}, \beta_{S_2}) = 1 \bar{\wedge} 1 = 0$ (where we applied the inductive hypothesis on S_1 and S_2). The other 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) $\mathbf{D} = \langle D_1, \ldots, D_N \rangle \in_R \{\text{ALICE}, \text{BOB}\}^N$. Thus, for any $\mathbf{d} \in \{\text{ALICE}, \text{BOB}\}^N$, we have $\Pr[\mathbf{D} = \mathbf{d}] = 2^{-N}$.

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

The inputs $\mathbf{X} = \langle X_1, \ldots, X_N \rangle$ and $\mathbf{Y} = \langle Y_1, \ldots, Y_N \rangle$ are determined by values γ, $\mathbf{d} = \langle d_1, \ldots, d_N \rangle$, and $\mathbf{r} = \langle r_1, \ldots, r_N \rangle$ for Γ, $\mathbf{D}$, and $\mathbf{R}$ as follows.

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

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

The reader may think of the random variables $\mathbf{D}$ and $\mathbf{R}$ as labeling the leaves of the tree T. For a leaf $j \in [N]$, the corresponding variable D_j chooses the player whose j-th bit will be fixed to 0. The j-th bit of the other player is then set to be equal to the random bit R_j.

Example. At this point it might be useful for the reader to see how the input for a binary tree S is distributed. As remarked earlier, a coloring γ for S paints a root-to-leaf path white and all the other nodes red. For any such γ we have $\Pr[\Gamma = \gamma] = 1/N_S$. All the other input bits, besides the ones that correspond to the single white leaf, are fixed according to Definition 20 and the red forest determined by γ. Thus, the only entropy in the input (given a coloring γ) comes from the single white leaf. The mutual information of the transcript and this leaf is what we lower bound in Section 3.2.4.

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

Proposition 22. *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 $f^{\bar{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = 0$. If u is a white node, then $f^{\bar{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = 1$.*

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

When $d_{T_u} = 0$, u is a leaf. If u is red and its parent is white, then T_u is a (one-vertex) tree in the red forest determined by γ. Definition 20 then implies that $\langle \mathbf{x}_u, \mathbf{y}_u\rangle = \langle 1, 1\rangle$ and so $f^{\bar{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = 0$. If u is white, notice that either $\mathbf{x}_u = 0$ or $\mathbf{y}_u = 0$ (see item (ii) above).

When $d_{T_u} > 0$ and u is white, then u has a red child v. By induction $f^{\bar{\wedge}}_{T_v}(\mathbf{x}_v, \mathbf{y}_v) = 0$, and it follows that $f^{\bar{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = 1$. 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 $f^{\bar{\wedge}}_{T_u}(\mathbf{x}_u, \mathbf{y}_u) = f^{\bar{\wedge}}_F(\mathbf{x}_F, \mathbf{y}_F)$. The statement then follows by Proposition 21, because, according to the definition of ζ_T, $\langle \mathbf{x}_F, \mathbf{y}_F \rangle = \langle \alpha_F, \beta_F \rangle$. 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 $\bar{\wedge}$-gate that are equal to '1' are, in some sense, irrelevant to the output). $\square$

3.2.3 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 $f^{\bar{\wedge}}_T$ and we want a protocol for $f^{\bar{\wedge}}_S$. 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 $f^{\bar{\wedge}}_T(\mathbf{x}, \mathbf{y}) = f^{\bar{\wedge}}_S(\mathbf{x}_S, \mathbf{y}_S)$. Then by running Π on $\langle \mathbf{x}, \mathbf{y} \rangle$ they obtain the desired output.

Let Π be any protocol for $f^{\bar{\wedge}}_T$. For any $S \in \mathrm{FBS}_T$ we will construct a family of protocols for $f^{\bar{\wedge}}_S$. 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 values of X_j and Y_j are determined by the coloring γ, so Alice and Bob can each determine x_j and 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 preprocessing, 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 22 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. Recall that N_T is the number of leaves, and that m_T and ρ_T are integers related to the tree T defined in part (6) of the paragraph on "definitions related to colorings" in Paragraph 3.2.1. The distribution ξ_T for triples $\langle S, \gamma, \mathbf{d}\rangle$ is as follows,

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

This is indeed a distribution since

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

Lemma 23. *Consider any protocol* Π *for a tree* T*. Suppose* $(\langle \mathbf{X}, \mathbf{Y}\rangle, \langle \Gamma, \mathbf{D}\rangle) \sim \zeta_T$ *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 $S \in \mathrm{FBS}_T$.)

$$\begin{aligned}
&\mathbf{E}_{\langle S,\gamma,\mathbf{d}\rangle \sim \xi_T}[\mathrm{I}(\mathbf{X}', \mathbf{Y}'\,;\,\Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma', \mathbf{D}')] \\
&= \sum_{S,\gamma,\mathbf{d}} \xi_T(S, \gamma, \mathbf{d}) \sum_{\gamma',\mathbf{d}'} \nu_S(\gamma', \mathbf{d}') \cdot \mathrm{I}(\mathbf{X}', \mathbf{Y}'\,;\,\Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma', \mathbf{D}' = \mathbf{d}')] \\
&= \sum_{S,\gamma',\mathbf{d}'} \sum_{\gamma: S \propto \gamma} \sum_{\mathbf{d}} \frac{1}{m_T 2^{N_T}} \cdot \frac{1}{N_S 2^{N_S}} \cdot \mathrm{I}(\mathbf{X}', \mathbf{Y}'\,;\,\Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma', \mathbf{D}' = \mathbf{d}')] \qquad (3.1) \\
&= \sum_{\substack{S,\gamma: \\ S \propto \gamma}} \sum_{\mathbf{d}} \frac{1}{m_{\gamma,T}} \cdot \frac{m_{\gamma,T}}{m_T 2^{N_T}} \cdot \mathrm{I}(\mathbf{X}', \mathbf{Y}'\,;\,\Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S). \qquad (3.2)
\end{aligned}$$

The transition from (3.1) to (3.2) needs to be justified. Look first at equation (3.2). 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 (3.1). 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 $\langle \gamma, \mathbf{d} \rangle$, for which we have $\mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_{\widehat{S}}[\gamma, \mathbf{d}] \,|\, \Gamma' = \widehat{\gamma}_S, \mathbf{D}' = \widehat{\mathbf{d}}_S) = A$. 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 (3.1). 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 (Definition 3.2.1), the last equation gives

$$
\begin{aligned}
&\mathbf{E}_{\langle S,\gamma,\mathbf{d}\rangle \sim \xi_T}[\mathrm{I}(\mathbf{X}', \mathbf{Y}' \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma', \mathbf{D}')] \\
&\quad \le \tfrac{1}{\rho_T} \sum_{\substack{S,\gamma: \\ S \propto \gamma}} \sum_{\mathbf{d}} \tfrac{m_{\gamma,T}}{m_T 2^{N_T}} \cdot \mathrm{I}(X'_{j(\gamma,S)}, Y'_{j(\gamma,S)} \,;\, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S). \qquad (3.3)
\end{aligned}
$$

For the left-hand side we have

$$
\begin{aligned}
&\mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi \,|\, \Gamma, \mathbf{D}) \\
&\quad = \sum_{\gamma,\mathbf{d}} \nu_T(\gamma, \mathbf{d}) \cdot \mathrm{I}(\mathbf{X}, \mathbf{Y} \,;\, \Pi \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d}) \\
&\quad \ge \sum_{\gamma,\mathbf{d}} \tfrac{m_{\gamma,T}}{m_T 2^{N_T}} \sum_{j \in \mathrm{PL}_T(\gamma)} \mathrm{I}(X_j, Y_j \,;\, \Pi \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d}) \\
&\quad = \sum_{\substack{S,\gamma: \\ S \propto \gamma}} \sum_{\mathbf{d}} \tfrac{m_{\gamma,T}}{m_T 2^{N_T}} \cdot \mathrm{I}(X_{j(\gamma,S)}, Y_{j(\gamma,S)} \,;\, \Pi \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d}). \qquad (3.4)
\end{aligned}
$$

The inequality follows from Proposition 6, ignoring terms that correspond to nonprincipal leaves. The last equality follows from the bijection between $\mathrm{FBS}_T(\gamma)$ and $\mathrm{PL}_T(\gamma)$ as discussed in Definition 3.2.1.

In view of equations (3.3) and (3.4), to finish the proof one only needs to verify that the two distributions

$$
(X'_{j(\gamma,S)}, Y'_{j(\gamma,S)}, \Pi_S[\gamma, \mathbf{d}] \,|\, \Gamma' = \gamma_S, \mathbf{D}' = \mathbf{d}_S), \quad (X_{j(\gamma,S)}, Y_{j(\gamma,S)}, \Pi \,|\, \Gamma = \gamma, \mathbf{D} = \mathbf{d})
$$

are identical. To see this, notice first that $\Pr[X'_{j(\gamma,S)} = b_x \mid \Gamma' = \gamma_S, \mathbf{D'} = \mathbf{d}_S] = \Pr[X_{j(\gamma,S)} = b_x \mid \Gamma = \gamma, \mathbf{D} = \mathbf{d}]$, 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 $\Pr[\Pi_S[\gamma, \mathbf{d}](\mathbf{X'}, \mathbf{Y'}) = \tau \mid X'_{j(\gamma,S)} = b_x, Y'_{j(\gamma,S)} = b_y, \Gamma' = \gamma_S, \mathbf{D'} = \mathbf{d}_S] = \Pr[\Pi(\mathbf{X}, \mathbf{Y}) = \tau \mid X_{j(\gamma,S)} = b_x, Y_{j(\gamma,S)} = b_y, \Gamma = \gamma, \mathbf{D} = \mathbf{d}]$. ☐

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 $\rho_T \geq N_T/2^{d_T}$.

Lemma 24. *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) $R_\delta(f^{\bar\wedge}_T) \geq R_\delta(f^{\bar\wedge}_{\widehat{T}})$, (3) $\rho_{\widehat{T}} \geq N/2^d$.*

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 $N_j \geq N/2$, then we apply the procedure to T_j to obtain $\widehat{T}_j$, set $\widehat{T} = \widehat{T}_j$, and remove the remaining 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 $\widehat{T}$ is 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 applies and $\widehat{T}$ is 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 $\rho_{\widehat{T}_1} \leq \cdots \leq \rho_{\widehat{T}_n}$. By the definition of ρ_T (Definition 3.2.1, part (6) in "definitions related to colorings"),

$$\rho_{\widehat{T}} = \sum_{j=1}^{n-1} \rho_{\widehat{T}_j} \geq \sum_{j=1}^{n-1} N_j/2^{d-1} = (N - N_n)/2^{d-1} > (N - N/2)/2^{d-1} = N/2^d.$$

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'. □

3.2.4 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 25. *Let T be a standard binary tree, and let T_u be 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*

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

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

When $d_{T_u} = 0$ we have $f_{T_u}(x, y) = x \barwedge y$. This case was shown in Bar-Yossef et al. [BYJKS04, Section 6], but we redo it here for completeness. First, notice that Γ_u is constant and thus the left-hand side simplifies to $\mathrm{I}(X_u, Y_u\,;\,\Pi(X, Y) \mid D_u)$. Expanding on values of D_u this is equal to

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

because given $D_u = \textsc{Alice}$ we have $X_u = 0$ and given $D_u = \textsc{Bob}$ we have $Y_u = 0$. Also, given $D_u = \textsc{Alice}$ we have $Y_u \in_R \{0, 1\}$ 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 12 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

$$\begin{aligned}
&\mathrm{I}(X_u, Y_u\,;\,\Pi(X,Y) \mid D_u) \\
&\quad \geq \tfrac{1}{2}\big(\mathrm{h}^2(\Pi(a0b, c0d), \Pi(a0b, c1d)) + \mathrm{h}^2(\Pi(a0b, c0d), \Pi(a1b, c0d))\big) \\
&\quad \geq \tfrac{1}{4} \cdot \mathrm{h}^2(\Pi(a0b, c1d), \Pi(a1b, c0d)).
\end{aligned}$$

For the last inequality we used the triangle inequality of Hellinger distance (Proposition 13(1)). 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.

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

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

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

By induction, this is bounded from below by

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

Applying the cut-and-paste property (Proposition 13(2)) of Hellinger distance this becomes

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

Now, since the square of Hellinger distance satisfies the (weak) triangle inequality (see Proposition 13), we have

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

Recalling the definition of $\langle \alpha_T, \beta_T\rangle$, Definition 20, we get

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

This completes the inductive proof. □

Corollary 26. *For any binary tree T in standard form*

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

Proof. First apply Proposition 25 with the root of T as u and empty a, b, c, d.

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

The second inequality is an application of the Pythagorean property of Hellinger distance, Proposition 13(3). The last inequality follows from Propositions 21 and 13(4). □

3.2.5 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 27. *1. For any tree T in standard form,*

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

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

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

Proof. Let Π be a δ-error protocol for $f_T^{\bar{\wedge}}$. Lemma 23 holds for any Π, therefore

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

We now use the bound from Corollary 26 to obtain (1). For (2), we can compute ρ_T exactly to be $(t-1)^{d_T}$. $\square$

Corollary 28. *1. For any tree T in standard form,*

$$R_\delta(f_T^{\bar{\wedge}}) \geq (2 - 4\sqrt{\delta}) \cdot \frac{N_T}{8^{d_T+1}}.$$

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

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

Proof. Recalling that informational complexity is a lower bound for randomized complexity, (2) is immediate from Corollary 27(2). For (1), we apply Corollary 27(1) to $f_{\widehat{T}}$, where $\widehat{T}$ is as in Lemma 24. $\square$

The constants do not match the ones in Theorem 15. Let $T = T_1 \circ \cdots \circ T_t$. The slight improvements can be obtained by applying Theorem 18 with f being the t-variate NAND, and, for each $j \in [t]$, h_j and ζ_j being f_{T_j} and ζ_{T_j}, respectively. Applying Corollary 28(1) to each of the trees T_j gives part (1); similarly for part (2).

3.3 Lower bound for read-once threshold functions

In this section we prove Theorem 3, stated in the introduction.

A *threshold* gate, denoted T_k^n for $n > 1$ and $1 \le k \le n$, receives n boolean inputs and outputs '1' if and only if at least k of them are '1'. A *threshold tree* is a rooted tree in which every leaf corresponds to a distinct input variable and every gate is a threshold gate. A *read-once threshold function* f_E is a function that can be represented by a threshold tree E. As before, we define $f_E^\wedge$ and $f_E^\vee$ and we want to lower bound $\max\{R_\delta(f_E^\wedge), R_\delta(f_E^\vee)\}$. The following proposition shows that Alice and Bob can reduce a problem defined by an AND/OR tree to one defined by a threshold tree. Theorem 3 will then follow as a corollary of Theorem 1.

Proposition 29. *For any threshold tree E, there is an* AND/OR *tree T such that, for $g \in \{\wedge, \vee\}$, (1) $R_\delta(f_T^g) \le R_\delta(f_E^g)$, (2) $N_T \ge N_E/2^{d_E}$, 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 $N_{E_1} \ge \cdots \ge N_{E_n}$. Suppose the gate on the root is T_k^n. We consider cases on k. (1) If $1 < k \le n/2$, 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'. Afterwards, they simulate the protocol for f_E^g. $\square$

The tree T in the above proposition may not be a canonical representation of some function. However, transforming 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.

3.4 General form of main theorem

The lower bounds we obtained apply to functions of the (restricted) form $f^{\wedge}$ and $f^{\vee}$. In this section we consider arbitrary two-party read-once functions, and prove Theorem 4, stated in the introduction. Theorems 1 and 2 are deduced from our main result, Theorem 15. We also use Theorem 15 to deduce, communication complexity lower bounds for two-party read-once functions.

Consider an AND/OR tree-circuit C in canonical form, and suppose that its leaf-set is partitioned into two sets $\mathcal{X}_C = \{x_1, \ldots, x_s\}$ and $\mathcal{Y}_C = \{y_1, \ldots, y_t\}$ (thus, f_C is a two-party read-once function). We show that C can be transformed to a tree T in standard form, such that Alice and Bob can decide the value of f_T using any protocol for f_C. (The reader may have expected $f_T^{\bar{\wedge}}$ in the place of f_T. To avoid confusion we note that f_T will already be a two-party read-once function. In particular, for some tree T' with $d_{T'} = d_T - 1$ and $N_{T'} = N_T/2$, $f_T = f_{T'}^{\bar{\wedge}}$.)

Lemma 30. *For any two-party read-once function f, there is a tree T in standard form, such that (1) $R_\delta(f_T) \leq R_\delta(f)$, (2) $N_T \geq D^{||}(f)/d(f)$, and (3) $d_T \leq d(f)$.*

Proof. We use notation from the paragraph before the statement of the lemma. The transformation of C proceeds in three stages.

In the first stage we collapse subtrees to single variables. For a node w let $A_w = \{u \in V_C \mid u \text{ is a child of } w \text{ and } L_{C_u} \subseteq \mathcal{X}_C\}$. Define B_w with $\mathcal{Y}$ in the place of $\mathcal{X}$. Let $W_\mathcal{X} = \{w \in V_C \mid L_{C_w} \nsubseteq \mathcal{X}_C \text{ and } A_w \neq \emptyset\}$. Define $W_\mathcal{Y}$ similarly. For each $w \in W_\mathcal{X}$, collapse $\{C_u \mid u \in A_w\}$ to a single variable x_w. That is, we remove all C_u with $u \in A_w$ from the tree, and add a new leaf x_w as a child of w. Similarly with $\mathcal{Y}$ in the place of $\mathcal{X}$ and B_w in the place of A_w. Name the resulting tree C_1. We claim that $R_\delta(f_C) = R_\delta(f_{C_1})$ and $D^{||}(f_C) = D^{||}(f_{C_1})$. It is easy to see that $R_\delta(f_C) \geq R_\delta(f_{C_1})$ and $D^{||}(f_C) \geq D^{||}(f_{C_1})$. Alice, for each $w \in W_{\mathcal{X}}$, can set each $x \in \mathcal{X}_{A_w}$ equal to x_w. Bob, for each $w \in W_{\mathcal{Y}}$, can set each $y \in \mathcal{Y}_{B_w}$ equal to y_w. After this preprocessing that requires no communication, they run a protocol for f_C. For the other direction, suppose $w \in W_{\mathcal{X}}$ is labeled by an AND gate. Alice sets x_w equal to $\bigwedge_{u \in A_w} f_{C_u}(\mathbf{x}_u)$ (for an OR gate, replace $\bigwedge$ with $\bigvee$). Bob acts similarly and afterwords they run a protocol for f_{C_1}. Clearly, $N_C \geq N_{C_1}$ and $d_C \geq d_{C_1}$. Notice also that in C_1 each node has at most one leaf in $\mathcal{X}_{C_1}$ and at most one in $\mathcal{Y}_{C_1}$ (where the partition for L_{C_1} is the obvious one).

In the second stage, we remove every leaf of C_1 that has a non-leaf sibling. If after these two stages some nodes are left with only one child, we collapse them with their unique child and label the new node with the gate of the child. Name the resulting tree C_2. We have $R_\delta(f_{C_1}) \geq R_\delta(f_{C_2})$ and $D^{||}(f_{C_1}) \geq D^{||}(f_{C_2})$, since Alice and Bob can generate values ('1'/'0') for the truncated leaves according to the gate of the parent (AND/OR). Clearly, $d_{C_1} \geq d_{C_2}$. Observe also that $N_{C_2} \geq N_{C_1}/d_{C_1}$. This is because for every pair of leaves in C_1 that remain in C_2, there can be at most $2(d_{C_1} - 1)$ leaves that will be removed—one pair for each of the $d_{C_1} - 1$ nodes along the path to the root (see last sentence of previous paragraph).

For the final stage, let T be the tree-circuit that is otherwise identical to C_2, but every gate of C_2 has been replaced by a NAND gate. It follows from Lemma 14 that $f_T \equiv f_{C_2}$ or $f_T \equiv \neg f_{C_2}$. Thus, for the models of interest, the complexity of f_{C_2} is equal to that of f_T. Also, $N_T = N_{C_2}$ and $d_T = d_{C_2}$.

For part (2), observe that $D^{||}(f_{C_1}) \leq N_{C_1}$. Tracing the inequalities from each stage,

$$N_T = N_{C_2} \geq N_{C_1}/d_{C_1} \geq D^{||}(f_{C_1})/d_{C_1} = D^{||}(f)/d_{C_1} \geq D^{||}(f)/d(f).$$

Parts (1) and (3) are immediate. $\square$

The tree-circuit T is in standard form, and Theorem 15 can be applied, yielding $R_\delta(f_T) \geq 4(2-4\sqrt{\delta}) \cdot N_T/8^{d_T}$. (For the constants involved, recall the parenthetic remark before the statement of the lemma.) Then, Theorem 4,

$$R_\delta(f) \geq (8-16\sqrt{\delta}) \cdot \frac{D^{||}(f)}{d(f) \cdot 8^{d(f)}},$$

follows from the lemma.

Chapter 4

The number-on-the-forehead model

In this chapter we present a set of tools that could be useful in an information-theoretic attack at the number-on-the-forehead complexity of disjointness.

4.1 Notation, terminology, and preliminaries

We introduce the notion of Hellinger volume of m distributions. In the next section we show that it has properties similar in flavor to the ones of Hellinger distance.

Definition 31. *The m-dimensional Hellinger volume of distributions $p_1, \dots, p_m$ over Ω is*

$$h_m(p_1, \dots, p_m) = 1 - \sum_{\omega \in \Omega} \sqrt[m]{p_1(\omega) \cdots p_m(\omega)}.$$

Notice that in the case $m = 2$, $h_2(p_1, p_2)$ is the square of the Hellinger distance between distributions p_1 and p_2.

The following fact follows from the arithmetic-geometric mean inequality.

Fact 32. *For any distributions $p_1, \dots, p_m$ over Ω, $h_m(p_1, \dots, p_m) \geq 0$.*

We write $[n] = \{1, 2, \dots, n\}$. For a sequence $\langle a_1, \dots, a_n \rangle$ we let, for $j \in [n]$, $a_{<j} = \langle a_1, \dots, a_{j-1} \rangle$, and $a^{-j} = (a_1, \dots, a_{j-1}, a_{j+1}, \dots, a_k)$. We will denote subsets of $\{0,1\}^k$ as follows: $I = \{0,1\}^k$; for $j \in [k]$, I_j is the set of points in I such that the j-th coordinate is set to zero, i.e. $I_j = \{z \in I \mid z_j = 0\}$; I_{OZ} (resp. I_{EZ}) is the set of points in I with an odd (resp. even) number of zeros.

4.2 An upper bound on the difference between the arithmetic and geometric mean.

For a nonnegative real sequence $\alpha = (\alpha_1, \ldots, \alpha_m)$, let $A(\alpha)$ and $G(\alpha)$ denote its arithmetic and geometric mean respectively. That is

$$A(\alpha) = \sum \alpha_j \Big/ m \quad \text{and} \quad G(\alpha) = \sqrt[m]{\prod \alpha_j}.$$

Theorem 33. *For any distribution p over $[m]$,*

$$A(p) - G(p) \leq \ln 2 \cdot D(p\|u),$$

where u is the uniform distribution over $[m]$.

Proof. Let $x_j = mp(j)$, $x = \langle x_1, \ldots, x_n \rangle$, and define

$$f(x) = \sum x_j \ln x_j \; + \; \sqrt[m]{\prod x_j}.$$

Theorem 33 is equivalent to showing that, for $x_1, \ldots, x_n \geq 0$, if $\sum x_j = m$, then $f(x) \geq 1$.

We proceed using Lagrange multipliers. We first need to check that $f(x) \geq 1$ when x is on the boundary, i.e. $x_j = 0$ for some $j \in [n]$. Without loss of generality, assume $x_1 = 0$. By the convexity of $t \ln t$, the minimum is attained when $x_2 = \cdots = x_m = m/(m-1)$. Thus,

$$f(x) \geq (m-1)\frac{m}{m-1} \ln \frac{m}{m-1} > m\left(1 - \frac{m-1}{m}\right) = 1.$$

According to [Lue03, Theorem on page 300], it suffices to show that $f(x) \geq 1$ for any x that satisfies the following system of equations.

$$\partial f / \partial x_j = 1 + \ln x_j + \sigma/(m x_j) = \lambda, \quad \text{for } j \in [m], \tag{L}$$

where $\sigma = \sqrt[m]{x_1 \cdots x_m} \neq 0$. Without loss of generality, since $\sum x_j = m$, we may assume $x_m \le 1$. The system (L) implies

$$\sum_{j=1}^{m-1} x_j(\partial f/\partial x_j) = m - x_m + \sum_{j=1}^{m-1} x_j \ln x_j + \sigma(m-1)/m = \lambda(m - x_m),$$

$$(m-1)x_m(\partial f/\partial x_m) = (m-1)(x_m + x_m \ln x_m + \sigma/m) = (m-1)\lambda x_m.$$

Subtracting the second from the first we get

$$\sum_{j=1}^{m-1} x_j \ln x_j - (m-1)x_m \ln x_m = m(\lambda - 1)(1 - x_m).$$

We also have

$$\sum x_j(\partial f/\partial x_j) = m + f(x) = m\lambda.$$

Suppose $x = (x_1, \ldots, x_m)$ satisfies the system (L). Since $x_m \le 1$, we have $x_m \ln x_m \le 0$, and using the last two equations we have

$$f(x) = m(\lambda - 1) \ge \frac{\sum_{j=1}^{m-1} x_j \ln x_j}{1 - x_m} \ge \frac{\sum_{j=1}^{m-1} x_j(1 - 1/x_j)}{1 - x_m} = 1.$$

This completes the proof. □

Corollary 34. *For any nonnegative real sequence $\alpha = (\alpha_1, \ldots, \alpha_m)$,*

$$A(\alpha) - G(\alpha) \le \sum \alpha_j \ln \frac{\alpha_j}{A(\alpha)}.$$

Proof. Apply Theorem 33 with $p(j) = \alpha_j / \sum_j \alpha_j$. □

Remark. Let $\widehat{\alpha}$ to be a normalized version of α, with $\widehat{\alpha}_j = \alpha_j / \sum \alpha_j$. Let also u denote the uniform distribution on $[m]$. Then, the right-hand side takes the form $\sum \alpha_j \ln(m\widehat{\alpha}_j) = mA(\alpha) \sum \widehat{\alpha}_j \ln(\widehat{\alpha}_j / u_j)$, and the above inequality becomes

$$\frac{A(\alpha) - G(\alpha)}{A(\alpha)} \le m \ln 2 \cdot D(\widehat{\alpha} \| u).$$

4.3 Properties of Hellinger volume

Hellinger volume lower bounds mutual information. The next lemma shows that Hellinger volume can be used to lower bound mutual information.

Lemma 35. *Consider random variables $Z \in_R [m]$, $\Phi(Z) \in \Omega$, and distributions Φ_z, for $z \in [m]$, over Ω. Suppose that given $Z = z$, the distribution of $\Phi(Z)$ is Φ_z. Then*

$$\mathrm{I}(Z\,;\,\Phi(Z)) \geq \frac{h_m(\Phi_1, \ldots, \Phi_m)}{m \ln 2}.$$

Proof. The left-hand side can be expressed as follows (see [CT06, page 20]),

$$\begin{aligned}\mathrm{I}(Z\,;\,\Phi(Z)) &= \sum_{j,\omega} \Pr[Z = j] \cdot \Pr[\Phi(Z) = \omega | Z = j] \cdot \log \frac{\Pr[\Phi(Z) = \omega | Z = j]}{\Pr[\Phi(Z) = \omega]} \\ &= \sum_{j,\omega} \frac{1}{m} \Phi_j(\omega) \log \frac{\Phi_j(\omega)}{\frac{1}{m} \sum_j \Phi_j(\omega)},\end{aligned}$$

and the right-hand side

$$h_m(\Phi_1, \ldots, \Phi_m) = \sum_\omega \left(\frac{1}{m} \sum_j \Phi_j(\omega) - \Big(\prod_j \Phi_j(\omega) \Big)^{\frac{1}{m}} \right).$$

It suffices to show that for each $\omega \in \Omega$,

$$\sum_j \frac{1}{m} \Phi_j(\omega) \log \frac{\Phi_j(\omega)}{\frac{1}{m} \sum_j \Phi_j(\omega)} \geq \frac{1}{m \ln 2} \left(\frac{1}{m} \sum_j \Phi_j(\omega) - \Big(\prod_j \Phi_j(\omega) \Big)^{\frac{1}{m}} \right).$$

Let $s = \sum_j \Phi_j(\omega)$, and $\rho(j) = \Phi_j(\omega)/s$, for $j \in [m]$; thus, for all j, $\rho(j) \in [0, 1]$, and $\sum_j \rho(j) = 1$. Under this renaming of variables, the left-hand side becomes $\ln 2 \cdot \frac{s}{m} \sum_j \rho(j) \log(m\rho(j))$ and the right one $\frac{s}{m} \cdot (\frac{1}{m} - \sqrt[m]{\prod \rho(j)})$. Thus, we need to show

$$\ln 2 \cdot \sum_j \rho(j) \log(m\rho(j)) \geq \frac{1}{m} - \Big(\prod_j \rho(j) \Big)^{\frac{1}{m}}.$$

Observe that the left-hand side is $\ln 2 \cdot D(\rho \| u)$, and the inequality holds by Theorem 33. $\square$

Symmetric-difference lemma. Let $P = \{P_z\}_{z \in Z}$ be a collection of distributions over a common space Ω. For $A \subseteq Z$, the *Hellinger volume of A with respect to P*, denoted by $\psi(P; A)$, is

$$\psi(A; P) = 1 - \sum_{\omega \in \Omega} \Big(\prod_{z \in A} P_z(\omega) \Big)^{1/|A|}.$$

The collection P will be understood from the context and we'll say that the Hellinger volume of A is $\psi(A)$. Note that, from Fact 32, $\psi(A;P) \geq 0$.

The following lemma can be seen as an analog to the weak triangle inequality that is satisfied by the square of the Hellinger distance.

Lemma 36 (Symmetric-difference lemma)**.** *If* A, B *satisfy* $|A| = |B| = |A\Delta B|$, *where* $A\Delta B = (A \setminus B) \cup (B \setminus A)$. *Then*

$$\psi(A) + \psi(B) \geq \frac{1}{2} \cdot \psi(A\Delta B).$$

Proof. By our hypothesis, it follows that $A \setminus B$, $B \setminus A$ and $A \cap B$ all have size $|A|/2$. Define u, v, w to be the vectors in $\mathbb{R}^{\Omega}$ defined by

$$\begin{aligned} u(\omega) &= \Big(\prod_{z \in A \setminus B} P_z(\omega)\Big)^{1/|A|}, \\ v(\omega) &= \Big(\prod_{z \in B \setminus A} P_z(\omega)\Big)^{1/|A|}, \\ w(\omega) &= \Big(\prod_{z \in A \cap B} P_z(\omega)\Big)^{1/|A|}. \end{aligned}$$

By the definition of Hellinger volume,

$$\begin{aligned} \psi(A) &= 1 - u \cdot w, \\ \psi(B) &= 1 - v \cdot w, \\ \psi(A\Delta B) &= 1 - u \cdot v. \end{aligned}$$

Thus the desired inequality is

$$2 - (u + v) \cdot w \geq (1 - u \cdot v)/2,$$

which is equivalent to

$$3 + u \cdot v \geq 2(u + v) \cdot w. \tag{4.1}$$

Since

$$\psi(A \setminus B) = 1 - u \cdot u,$$

$$\psi(B \setminus A) = 1 - v \cdot v,$$

$$\psi(A \cap B) = 1 - w \cdot w,$$

it follows that $\|u\|, \|v\|$ and $\|w\|$ are all at most 1. Thus $2(u+v) \cdot w \leq 2\|u+v\|$, and so (4.1) follows from

$$3 + u \cdot v \geq 2\|u+v\|.$$

Squaring both sides, it suffices to show

$$9 + 6u \cdot v + (u \cdot v)^2 \geq 4(\|u\|^2 + \|v\|^2 + 2u \cdot v)$$

Using the fact that $\|u\| \leq 1$ and $\|v\| \leq 1$ this reduces to

$$(1 - u \cdot v)^2 \geq 0,$$

which holds for all u, v. $\square$

Let s_l, s_r be two disjoint subsets of $[k]$. Let $I_l \subseteq I$ (resp., I_r) be the set of strings with odd number of zeros in the coordinates indexed by s_l (resp., s_r). Let $s_p = s_l \cup s_r$ and $I_p = I_l \Delta I_r$. It is not hard to see that I_p is the set of strings with odd number of zeros in the coordinates indexed by s_p. By the symmetric-difference lemma,

$$\psi(I_l) + \psi(I_r) \geq \frac{\psi(I_p)}{2}. \tag{4.2}$$

For each $j \in [k]$, let $I_j \subseteq I$ be the set of strings where the j-th coordinate is set to zero. Applying the above observation inductively, we can obtain the following lemma.

Lemma 37. *Let $s \subseteq [k]$ be an arbitrary non-empty set and let $I_s \subseteq I$ be the set of strings with odd number of zeros in the coordinates indexed by s. Then,*

$$\sum_{j \in s} \psi(I_j) \geq \frac{\psi(I_s)}{2^{\lceil \log |s| \rceil}}.$$

Proof. We prove the claim via induction on the size of s. If s is a singleton set, it trivially holds. Otherwise, assume that for any subset of $[k]$ of size less than $|s|$, the claim is true.

Partition s into two non-empty subsets s_l, s_r with the property that $|s_l| = \lceil |s|/2 \rceil$ and $|s_l| = \lfloor |s|/2 \rfloor$. Then $\lceil \log |s| \rceil = 1 + \max\{\lceil \log |s_l| \rceil, \lceil \log |s_r| \rceil\}$. By the inductive hypothesis,

$$\sum_{j \in s_l} \psi(I_{s_l}) \geq \frac{\psi(I_{s_l})}{2^{\lceil \log |s_l| \rceil}} \quad \text{and} \quad \sum_{j \in s_r} \psi(I_{s_r}) \geq \frac{\psi(I_{s_r})}{2^{\lceil \log |s_r| \rceil}}.$$

Thus,

$$\begin{aligned}
\sum_{j \in s} \psi(I_{s_l}) &= \sum_{j \in s_l} \psi(I_{s_l}) + \sum_{j \in s_r} \psi(I_{s_r}) \\
&\geq \frac{\psi(I_{s_l})}{2^{\lceil \log |s_l|}} + \frac{\psi(I_{s_r})}{2^{\lceil \log |s_r| \rceil}} && \text{by the Inductive Hypothesis,} \\
&\geq \frac{1}{2^{\lceil \log |s| \rceil - 1}} [\psi(I_{s_l}) + \psi(I_{s_r})] && \text{by the choice of } s_l \text{ and } s_r, \\
&\geq \frac{1}{2^{\lceil \log |s| \rceil}} \psi(I_s) && \text{by Equation (4.2).}
\end{aligned}$$

□

Let $I_{OZ} \subseteq I$ be the set of strings which have odd number of zeros. The next corollary is an immediate consequence of Lemma 37 when $s = [k]$.

Lemma 38.

$$\sum_{j=1}^{k} \psi(I_j) \geq \frac{\psi(I_{OZ})}{2^{\lceil \log k \rceil}}.$$

NOF communication complexity and Hellinger volume. It was shown in Bar-Yossef, Jayram, Kumar & Sivakumar [BYJKS04], that the distribution of transcripts of a two-party protocol on a fixed input, is a product distribution. The same is true for a multi-party NOF protocol.

Lemma 39. *Let* Π *be a* k*-player NOF communication protocol with input set* $\mathcal{Z} = \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_k$ *and let* Ω *be the set of possible transcripts. For each* $j \in [k]$, *there is a mapping* $q_j : \Omega \times \mathcal{Z}^{-j} \to \mathbb{R}$*, such that for every* $z = (z_1, \ldots, z_k) \in \mathcal{Z}$ *and* $\omega \in \Omega$*,*

$$\Pr[\Pi(z) = \omega] = \prod_{j=1}^{k} q_j(\omega; z^{-j}).$$

Proof. Suppose $|\Pi(z)| \le l$. For $i = 1, \ldots, l$, let $\Pi_i(z)$ denote the i-th bit sent in an execution of the protocol. Let $\sigma_i \in [k]$ denote the player that sent the i-th bit. Then

$$\begin{aligned}
\Pr[\Pi(z) = \omega] &= \Pr[\Pi_1(z) = \omega_1, \ldots, \Pi_l(z) = \omega_l] \\
&= \prod_{i=1}^{l} \Pr[\Pi_i(z) = \omega_i | \Pi_{<i}(z) = \omega_{<i}], \\
&= \prod_{i=1}^{l} \Pr[\Pi_i(z^{-\sigma_i}; \omega_{<i}) = \omega_i],
\end{aligned}$$

because every bit send by player j depends only on z^{-j} and the transcript up to that point. We set

$$q_j(\omega; z^{-j}) = \prod_{i:\sigma_i = j} \Pr[\Pi_i(z^{-j}; \omega_{<i}) = \omega_i]$$

to obtain the expression of the lemma. □

As a corollary, we have the following cut-and-paste property for Hellinger volume.

Lemma 40. *Let* $I_{OZ} \subseteq I$ *be the set of inputs which have odd number of zeros, and let* $I_{EZ} = I \setminus I_{OZ}$*. Then*

$$\psi(I_{OZ}) = \psi(I_{EZ}).$$

Proof. Using the expression of the previous lemma, we have that for any $\omega \in \Omega$,

$$\prod_{v \in I_{OZ}} P_v(\omega) \;=\; \prod_{v \in I_{OZ}} \prod_{j=1}^{k} q_j(\omega; v^{-j}) \;=\; \prod_{u \in I_{EZ}} \prod_{j=1}^{k} q_j(\omega; u^{-j}) \;=\; \prod_{u \in I_{EZ}} P_u(\omega).$$

The middle equality holds, because for each $j \in [k]$ and $v \in I_{OZ}$ there is a unique $u \in I_{EZ}$ such that $v^{-j} = u^{-j}$. □

Lower bounding Hellinger volume. Eventually, we will need to provide a lower bound for the Hellinger volume of several distributions over protocol transcripts. In the two-party case, one lower bounds the Hellinger distance between the distribution of the transcripts on an accepting input and the distribution of the transcripts on a rejecting input. The following lemma will allow for similar conclusions in the multi-party case.

Lemma 41. *Let $A \subseteq I$ be of size $t \geq 2$. Suppose there is an event $T \subseteq \Omega$, a constant $0 \leq \delta \leq 1$ and an element v in A such that $P_v(T) \geq 1 - \delta$ and that for all $u \in A$ with $u \neq v$, $P_u(T) \leq \delta$. Then*

$$\psi(A) \geq \left(2 - 4\sqrt{\delta(1-\delta)}\right) \cdot \frac{1}{t}.$$

Proof. We need to show

$$1 - \sum_{\omega \in \Omega} \prod_{u \in A} P_u(\omega)^{\frac{1}{t}} \geq \left(2 - 4\sqrt{\delta(1-\delta)}\right) \cdot \frac{1}{t}.$$

Let $a = P_v(T) = \sum_{\omega \in T} P_v(\omega)$ and $b = \sum_{\omega \in T} \frac{1}{t-1} \sum_{u \neq v} P_u(\omega)$. Notice that by assumption $a \geq 1 - \delta$ and $b \leq \delta$.

Recall Hölder's inequality: for any nonnegative x_k, y_k, $k \in m$,

$$\sum_{k=1}^{m} x_k y_k \leq \left(\sum_{k=1}^{m} x_k^t\right)^{\frac{1}{t}} \left(\sum_{k=1}^{m} y_k^{\frac{t}{t-1}}\right)^{\frac{t-1}{t}}.$$

We first treat the sum over $\omega \in T$.

$$\begin{aligned}
\sum_{\omega \in T} \prod_{u \in A} P_u(\omega)^{\frac{1}{t}} &= \sum_{\omega \in T} P_v(\omega)^{\frac{1}{t}} \prod_{u \neq v} P_u(\omega)^{\frac{1}{t}} \\
&\leq \left(\sum_{\omega \in T} P_v(\omega)\right)^{\frac{1}{t}} \left(\sum_{\omega \in T} \prod_{u \neq v} P_u(\omega)^{\frac{1}{t-1}}\right)^{\frac{t-1}{t}} \\
&\leq \left(\sum_{\omega \in T} P_v(\omega)\right)^{\frac{1}{t}} \left(\sum_{\omega \in T} \frac{1}{t-1} \sum_{u \neq v} P_u(\omega)\right)^{\frac{t-1}{t}} \\
&= a^{\frac{1}{t}} b^{\frac{t-1}{t}},
\end{aligned}$$

where we first used Hölder's inequality and then the arithmetic-geometric mean inequality. We do the same steps for the sum over $\omega \notin T$ to find

$$\sum_{\omega \notin T} \prod_{u \in A} P_u(\omega)^{\frac{1}{t}} \leq (1-a)^{\frac{1}{t}} (1-b)^{\frac{t-1}{t}}.$$

Hence,

$$\sum_{\omega \in \Omega} \prod_{u \in A} P_u(\omega)^{\frac{1}{t}} \leq a^{\frac{1}{t}} b^{\frac{t-1}{t}} + (1-a)^{\frac{1}{t}} (1-b)^{\frac{t-1}{t}}.$$

Let $g(a,b,x) = a^x b^{1-x} + (1-a)^x (1-b)^{1-x}$. We will show that under the constraints $a \geq 1-\delta$ and $b \leq \delta$ where $\delta < 1/2$, for any fixed $0 \leq x \leq 1/2$, $g(a,b,x)$ is maximized for $a = 1-\delta$ and $b = \delta$. The partial derivatives for $g(a,b,x)$ with respect to a and b are

$$g_a(a,b,x) = x[a^{x-1} b^{1-x} - (1-a)^{x-1}(1-b)^{1-x}] = x\Big[\Big(\frac{b}{a}\Big)^{1-x} - \Big(\frac{1-b}{1-a}\Big)^{1-x}\Big]$$

$$g_b(a,b,x) = (1-x)[a^x b^{-x} - (1-a)^x (1-b)^{-x}] = (1-x)\Big[\Big(\frac{b}{a}\Big)^{-x} - \Big(\frac{1-b}{1-a}\Big)^{-x}\Big]$$

Under our constraints, $\frac{b}{a} < 1 < \frac{1-b}{1-a}$, $1-x > 0$ and $-x \leq 0$, thus, $g_a(a,b,x) < 0$ and $g_b(a,b,x) \geq 0$ for any such a, b, and x. This implies that for any fixed b, $g(a,b,x)$ is maximized when $a = 1-\delta$ and similarly for any fixed a, $g(a,b,x)$ is maximized when $b = \delta$. Therefore, for all a, b, and $0 \leq x \leq 1$, $g(a,b,x) \leq g(1-\delta,\delta,x)$.

For $0 \leq x \leq 1/2$, let

$$f(\delta,x) = 1 - g(1-\delta,\delta,x) = 1 - (1-\delta)^x \delta^{1-x} - \delta^x (1-\delta)^{1-x}.$$

Since $f(\delta,x)$ is convex for any constant $0 \leq \delta \leq 1$,

$$f(\delta,x) \geq \frac{f(\delta,1/2) - f(\delta,0)}{1/2 - 0} \cdot x = 2\big(1 - 2\sqrt{\delta(1-\delta)}\big) \cdot x. \qquad \square$$

4.4 An application

In this section we show how to derive a lower bound for the informational complexity of the AND_k function. Define a collection of distributions $\eta = \{\zeta_1, \ldots, \zeta_k\}$, where, for each $j \in [k]$, ζ_j is the uniform distribution over $I_j = \{0,1\}^k$ (recall that $I_j \subseteq I$, $j \in [k]$, is the set of k-bit strings with the j-th bit set to 0). We prove the following lower bound on the δ-error informational complexity of AND_k with respect to η.

Remark. The choice of the collection η is not arbitrary, but is suggested by the way the direct-sum theorem for informational complexity is proved in [BYJKS04] for the two-party setting. In particular, two properties of η seem crucial for such a purpose. First, for each $j \in [k]$, ζ_j is a distribution with support only on the zeroes of AND_k. Second, under any ζ_j, the input of each player is independent of any other input.

Theorem 42.

$$\mathrm{IC}_{\eta,\delta}(\mathsf{AND}_k) \geq \log e \cdot \left(1 - 2\sqrt{\delta(1-\delta)}\right) \cdot \frac{1}{k^2\, 4^{k-1}}.$$

Proof. Let Π be a δ-error protocol for AND_k. By Lemma 35 we have that,

$$I(Z; \Pi(Z)) \geq \frac{1}{2^{k-1} \ln 2} \cdot \psi(I_j),$$

where $Z \sim \zeta_j$, for any $j \in [k]$, Thus, by the definition of $\mathrm{IC}_{\eta,\delta}(\mathsf{AND}_k)$,

$$\mathrm{IC}_{\eta,\delta}(\mathsf{AND}_k) \geq \sum_{j=1}^{k} \frac{1}{k\, 2^{k-1} \ln 2} \cdot \psi(I_j).$$

Applying in turn Lemmas 38, 40, and 41 we have

$$\mathrm{IC}_{\eta,\delta}(\mathsf{AND}_k) > \frac{\psi(I_{OZ})}{k^2\, 2^k \ln 2} = \frac{\psi(I_{EZ})}{k^2\, 2^k \ln 2} \geq \log e \cdot \left(1 - 2\sqrt{\delta(1-\delta)}\right) \cdot \frac{1}{k^2\, 4^{k-1}},$$

where the application of Lemma 41 is with $A = I_{EZ}$, $t = 2^{k-1}$, T the set of transcripts that output “1”, and v the all-one vector in I. $\square$

It is of interest to note, that

$$\mathrm{IC}_{\eta,\delta}(\mathsf{AND}_k) \leq \frac{1}{k} \cdot H(1/2^{k-1}) = O(1/2^k).$$

This is achieved by the following protocol. The players, one by one, reveal with one bit whether they see a 0 or not. The communication ends with the first player that sees a 0. The amount of information revealed is $H(1/2^{k-1})$ under ζ_1 and 0 otherwise.

Chapter 5
Conclusions and future work

In this chapter we discuss related open problems regarding the two-party randomized communication complexity, and some difficulties in applying the information-theoretic framework to the number-on-the-forehead model.

5.1 Two-party randomized communication complexity

A problem that stands out after the lower bound presented in Chapter 3 for read-once formulae, is to determine the communication complexity of f_{U_d}, where U_d is the uniform binary tree of depth d. It is not hard to see, by embedding $\mathsf{DISJ}_{\sqrt{2}^d,2}$, that $R_\delta(f_{U_d}) = \Omega(\sqrt{2}^d)$. The corresponding question for the decision tree model was answered in the work of Saks & Wigderson [SW86], where it was shown that the randomized decision tree complexity of f_{U_d} is $\Theta\Big(\big(\frac{1+\sqrt{33}}{4}\big)^d\Big)$. The randomized decision tree for f_{U_d} can be transformed into a communication protocol with only doubling the length, showing that $R_\delta(f_{U_d}) = O\Big(\big(\frac{1+\sqrt{33}}{4}\big)^d\Big)$. We believe that the lower bound can be improved and it would be interesting if an information-theoretic approach could yield the improvements.

Progress in the complexity of the uniform binary tree would probably yield improvements on the bounds for the general trees. Note the $\Omega(\sqrt{n})$ bound, where n is the number of variables, which is trivial for uniform trees, was shown to hold also for arbitrary trees by Jain, Klauck & Zhang [JKZ10].

Another direction for future research would be to prove lower bounds on $R_\delta(f^\wedge)$ and $R_\delta(f^\vee)$, in the case where f is an arbitrary boolean function. This question has been considered before as intermediate step in understanding the relationship between randomized and quantum communication complexity. For example, Sherstov [She10]), shows that $\max\{R_{1/3}(f^{\wedge}), R_{1/3}(f^{\vee})\} \geq \Omega(\mathrm{bs}(f)^{1/4})$, where $\mathrm{bs}(f)$ is the block-sensitivity of f (see [BdW02]).

5.2 Number-on-the-forehead communication complexity

Proving lower bounds in the number-on-the-forehead model is a major research direction. Until now, the only method that has successfully been extended from the two-party to the multi-party NOF model is discrepancy. An interesting question is if the information-theoretic framework can be useful in proving lower bounds for the NOF model. However, there seem to be fundamental difficulties in proving a direct-sum theorem on informational complexity in the NOF model. The reader familiar with the techniques of Bar-Yossef, Jayram, Kumar & Sivakumar [BYJKS04], should recall that in the first part of the method a direct-sum for informational complexity of disjointness is proved. In particular, it is shown that with respect to suitable collections of distributions η and ζ for $\mathsf{DISJ}_{n,2}$ and AND_2 respectively, the information cost of $\mathsf{DISJ}_{n,2}$ is at least n times the informational complexity of AND_2 : $\mathrm{IC}_{\eta,\delta}(\mathsf{DISJ}_{n,2}) \geq n \cdot \mathrm{IC}_{\zeta,\delta}(\mathsf{AND}_2)$. This is achieved via a simulation argument in which the players, to decide the AND_2 function, use a protocol for disjointness by substituting their inputs in a special copy of AND_2 and using their random bits to generate the inputs for the rest $n-1$ copies of AND_2. In the NOF model the players can no longer perform such a simulation. This is because, with private random bits, they cannot agree on what the input on the rest of the copies should be without additional communication. This problem can be overcome if we think of their random bits as being not private, but on each player's forehead, just like the input. However, In such a case, although the direct-sum theorem holds, it is useless. This is because $\mathrm{IC}_{\zeta,\delta}(\mathsf{AND}_k) = 0$, as is shown by the protocol we describe in the next paragraph.

We describe a protocol that computes AND_k on every input, with one-sided error. It has the property that for any distribution over the zeroes of AND_k, no player learns anything about his own input. We give the details for three players. Let x_1, x_2, x_3 denote the input. Each player has two random bits on his forehead, denoted a_1, a_2, a_3 and b_1, b_2, b_3. The first player does the following: if $x_2 = x_3 = 1$, he sends $a_2 \oplus a_3$, otherwise he sends $a_2 \oplus b_3$. The other two players behave analogously. If the XOR of the three messages is '0', they answer '1', otherwise they know that the answer is '0'. Notice that any player learns nothing from another player's message. This is because the one-bit message is XOR-ed with one of his own random bits, which he cannot see.

References

[Abl96] Farid M. Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theor. Comput. Sci.*, 157(2):139–159, 1996.

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 67–76, New York, NY, USA, 2010. ACM.

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002.

[BHN09] Paul Beame and Dang-Trinh Huynh-Ngoc. Multiparty communication complexity and threshold circuit size of AC^0. In *FOCS*, pages 53–62. IEEE Computer Society, 2009.

[BNS92] László Babai, Noam Nisan, and Mario Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. Syst. Sci.*, 45(2):204–232, 1992.

[BPS05] Paul Beame, Toniann Pitassi, and Nathan Segerlind. Lower bounds for Lovász-Schrijver systems and beyond follow from multiparty communication complexity. In *In Proc. 32nd Int. Conf. on Automata, Languages and Programming (ICALP'05)*, pages 1176–1188, 2005.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, 2011.

[BT94] Richard Beigel and Jun Tarui. On acc. *Computational Complexity*, 4:350–366, 1994.

[BYJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *IEEE Conference on Computational Complexity*, pages 93–102, 2002.

[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.

[CA08] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. Technical Report TR-08-002, ECCC, 2008.

[CDNT98] Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp. Quantum entanglement and the communication complexity of the inner product function. In *QCQC '98: Selected papers from the First NASA International Conference on Quantum Computing and Quantum Communications*, pages 61–74, London, UK, 1998. Springer-Verlag.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, STOC '83, pages 94–99, New York, NY, USA, 1983. ACM.

[CKS03] Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *IEEE Conference on Computational Complexity*, pages 107–117. IEEE Computer Society, 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.

[CT93] Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM J. Discrete Math.*, 6(1):110–123, 1993.

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

[FG05] Jeff Ford and Anna Gál. Hadamard tensors and lower bounds on multiparty communication complexity. In *In ICALP*, pages 1163–1175, 2005.

[FGG08] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum algorithm for the hamiltonian nand tree. *Theory of Computing*, 4(1):169–190, 2008.

[Gro94] Vince Grolmusz. The BNS lower bound for multi-party protocols in nearly optimal. *Inf. Comput.*, 112(1):51–54, 1994.

[HG90] Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. In *FOCS*, volume II, pages 610–618. IEEE, 1990.

[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.

[Jay09] T. S. Jayram. Hellinger strikes back: A note on the multi-party information complexity of and. In *Proceedings of the 12th International Workshop and 13th International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, APPROX '09 / RANDOM '09, pages 562–573, Berlin, Heidelberg, 2009. Springer-Verlag.

[JKR09] T. S. Jayram, Swastik Kopparty, and Prasad Raghavendra. On the communication complexity of read-once AC^0 formulae. In *IEEE Conference on Computational Complexity*, pages 329–340, 2009.

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

[JKZ10] Rahul Jain, Hartmut Klauck, and Shengyu Zhang. Depth-independent lower bounds on the communication complexity of read-once boolean formulas. In My T. Thai and Sartaj Sahni, editors, *COCOON*, volume 6196 of *Lecture Notes in Computer Science*, pages 54–59. Springer, 2010.

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

[Kre95] Ilan Kremer. Quantum communication. Master's thesis, Computer Science Department, Hebrew University, 1995.

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

[LS09a] Troy Lee and Adi Shraibman. Disjointness is hard in the multi-party number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[LS09b] Nikos Leonardos and Michael Saks. Lower bounds on the randomized communication complexity of read-once functions. *Computational Complexity, Annual IEEE Conference on*, 0:341–350, 2009.

[LS10] Nikos Leonardos and Michael Saks. Lower bounds on the randomized communication complexity of read-once functions. *Computational Complexity*, 19(2):153–181, 2010.

[LSZ09] Troy Lee, Adi Shraibman, and Shengyu Zhang. Personal communication, 2009.

[Lue03] D.G. Luenberger. *Linear and nonlinear programming*. Kluwer Academic, 2003.

[Nis94] Noam Nisan. The communication complexity of threshold gates. In *Proceedings of "Combinatorics, Paul Erdos is Eighty"*, pages 301–315, 1994.

[NW91] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. In *STOC*, pages 419–429. ACM, 1991.

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

[Raz00] Ran Raz. The bns-chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[She10] Alexander A. Sherstov. On quantum-classical equivalence for composed communication problems. *Quantum Information & Computation*, 10(5&6):435–455, 2010.

[SS02] Michael Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *STOC '02: Proceedings of the thirty-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.