Randomized Communication vs. Partition Number*

Mika Göös[1], T. S. Jayram[2], Toniann Pitassi[3], and Thomas Watson[4]

1 **Harvard University, Cambridge, MA, USA**
`mika@seas.harvard.edu`
2 **IBM Almaden, San Jose, CA, USA**
`jayram@us.ibm.com`
3 **University of Toronto, Toronto, Canada**
`toni@cs.toronto.edu`
4 **University of Memphis, Memphis, TN, USA**
`thomas.watson@memphis.edu`

Abstract

We show that *randomized* communication complexity can be superlogarithmic in the partition number of the associated communication matrix, and we obtain near-optimal *randomized* lower bounds for the Clique vs. Independent Set problem. These results strengthen the deterministic lower bounds obtained in prior work (Göös, Pitassi, and Watson, FOCS 2015). One of our main technical contributions states that information complexity when the cost is measured with respect to only 1-inputs (or only 0-inputs) is essentially equivalent to information complexity with respect to all inputs.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases communication complexity, partition number, information complexity

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.52

1 Introduction

A prior work [16] exhibited a boolean function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ whose deterministic communication complexity is superlogarithmic in the *partition number*

$$\chi(F) \;\coloneqq\; \chi_0(F) + \chi_1(F)$$

where $\chi_i(F)$ is the least number of rectangles (sets of the form $A \times B$ where $A \subseteq \mathcal{X}$, $B \subseteq \mathcal{Y}$) needed to partition the set $F^{-1}(i)$. In this follow-up work, we upgrade the lower-bound results from [16] to hold against randomized protocols – here the notation $\tilde{\Omega}(m)$ hides factors polylogarithmic in m.

▶ **Theorem 1.** *There is an F with randomized communication complexity $\tilde{\Omega}(\log^{1.5} \chi(F))$.*

▶ **Theorem 2.** *There is an F with randomized communication complexity $\tilde{\Omega}(\log^{2} \chi_1(F))$.*

A main technical contribution of our paper – which is key to both the proofs of Theorem 1 as well as the subsequent strengthening by [5] – informally states that the information complexity of a function (as defined by [9]) remains essentially unchanged if the cost is

* The full version of this work is available at [14], `https://eccc.weizmann.ac.il/report/2015/169/`.

44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).
Editors: Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl;
Article No. 52; pp. 52:1–52:15

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

measured with respect to only 1-inputs (or only 0-inputs) rather than all inputs. We say a protocol Π is ϵ-correct if it succeeds with probability at least $1-\epsilon$ on each input, and define $\mathrm{IC}(\Pi)$ as the maximum over all input distributions of the information cost (defined later), and define $\mathrm{IC}^b(\Pi)$ similarly but with the maximum over all distributions over b-inputs (for some $b \in \{0,1\}$).

▶ **Theorem 3.** *Fix any F and $b \in \{0,1\}$. For every 1/3-correct protocol Π there is a 1/3-correct protocol Π' such that $\mathrm{IC}(\Pi') \leq O\big(\mathrm{IC}^b(\Pi) + \log(\mathrm{CC}(\Pi)+2)\big)$. Moreover, $\mathrm{CC}(\Pi') \leq O\big(\mathrm{CC}(\Pi)\cdot\log(\mathrm{CC}(\Pi)+2)\big)$.*

In the theorem statement above, the additional lower order term involving the communication cost appears due to technical reasons. This makes the statement slightly weaker but this is mitigated in the aforementioned applications due to the additional fact that we can also bound the communication cost of the new protocol.

1.1 Applications and discussion

Theorem 1

Prior to this work, no examples of F were known with randomized communication complexity larger than $\log\chi(F)$. In fact, such a separation cannot be obtained using the usual rectangle-based lower-bound methods, as catalogued by Jain and Klauck [17]. In particular, Theorem 1 shows that randomized complexity can be polynomially larger than the *partition bound* [17, 19], which is one of the most powerful general lower bound methods for randomized communication. (Consequently, our proof of Theorem 1 has to exploit another powerful lower-bound method, namely *information complexity*.) Note also that every F has deterministic communication complexity at least $\log\chi(F)$ and at most $O(\log^2\chi(F))$, where the latter upper bound is a classical result of [2]. Theorem 1 shows that the upper bound cannot be improved much even if we allow randomization.

Theorem 2

The relationship between $\chi_1(F)$ and the communication complexity of F can be equivalently formulated in the language of the *Clique vs. Independent Set* game, played on a graph derived from F (Alice holds a clique, Bob holds an independent set: do they intersect?). See [34, §4] or [21, §4.4] for the equivalence. Yannakakis [34] (extending [2]) proved that every F has deterministic communication complexity at most $O(\log^2\chi_1(F))$. Our Theorem 2 shows that this upper bound is essentially tight even if we allow randomized protocols, and it implies that there is a graph on n nodes for which Clique vs. Independent Set requires $\tilde{\Omega}(\log^2 n)$ randomized communication. (The deterministic upper bound $O(\log^2 n)$ holds for all graphs.)

Extension complexity. In fact, we prove Theorem 2 by showing that (the negation of) the function F has high *approximate nonnegative rank* (a.k.a. smooth rectangle bound; see Section 2 for definitions). One consequence in the field of *extended formulations* (see [34, 11] for definitions) is that we obtain a graph G such that the polytope generated by the so-called "clique inequalities" of G has extension complexity $n^{\tilde{\Omega}(\log n)}$. (The slack matrix associated with the clique inequalities is simply (the negation of) the Clique vs. Independent Set game. These inequalities capture the independent set polytope of G when G is perfect – our graph G however is not.) The previous bound in this direction was $n^{\Omega(\log^{0.128} n)}$ from a related work [13]. Technically speaking, the lower bound from [13] was proved for *nondeterministic* communication complexity, so the full result remains incomparable with Theorem 2.

Log-rank conjecture. The famous log-rank conjecture of Lovász and Saks [30] postulates that the deterministic communication complexity of F is polynomially related to $\log \operatorname{rank}(F)$. Gavinsky and Lovett [12] have shown that the conjecture is equivalent to asking whether the randomized communication complexity of F is polynomially bounded in $\log \operatorname{rank}(F)$. Here our Theorem 2 gives at least a near-quadratic separation between the randomized communication complexity of F and $\log \operatorname{rank}(F) \leq \log \chi_1(F)$; the previous best lower bound was $\Omega(\log^{1.63} \operatorname{rank}(F))$ due to Kushilevitz [26]. Furthermore, Troy Lee has pointed out to us that our construction underlying Theorem 2 exhibits nearly a 4-th power separation between the logarithms of *approximate nonnegative rank* and *approximate rank*. This gives lower bounds for the so-called *log-approximate-rank conjecture* [28, Conjecture 42], which is the randomized analogue of the log-rank conjecture. The previous best separation was quadratic (as witnessed by the set-disjointness problem).

Theorem 3

One-sided information complexity satisfies a famous direct sum property ([6, 9]): for any protocol Π computing $\mathrm{AND}_k \circ F^k$ (i.e., the AND of k copies of F) there exists a protocol Π' computing F with $\mathrm{IC}^1(\Pi') \leq O(\mathrm{IC}^1(\Pi)/k)$ (see, e.g., [5, Claim 37]). One can also formulate a dual lemma for $\mathrm{OR}_k \circ F^k$ in terms of IC^0. This is the context where our Theorem 3 relating IC and IC^1 (and IC^0) is useful: it implies that analogous direct sum lemmas hold for *two-sided* information complexity, up to low order terms. Iterating such a two-sided lemma some constantly many times, one obtains an alternative proof for the result that every n-bit constant-depth balanced read-once AND–OR tree with binary bottom fan-in (defining an Alice–Bob bipartition of input bits) has randomized communication complexity $\Omega(n)$; this result was first proved in [20, 29] even for *unbalanced* trees.

Another application of Theorem 3 appears in the recent work [5]. They improved our 1.5-th power separation in Theorem 1 to near-quadratic (which is optimal) by iteratively applying Theorem 3 to analyze a communication analogue of a query-complexity construction due to Ambainis, Kokainis, and Kothari [4] (which is a variation of usual AND–OR trees).

1.2 Our techniques

The basic strategy in [16] for obtaining the deterministic versions of Theorems 1–2 was to first obtain analogous gaps in the easier-to-understand world of query complexity, then "lift" the results to communication complexity using a so-called *simulation lemma*. For getting randomized lower bounds, two obstacles immediately present themselves: (i) The functions studied in [16] are too easy for randomized protocols (as shown by [31]). (ii) There is no known simulation lemma for the bounded-error randomized setting.

To handle obstacle (i), we modify the functions from [16] in a way that preserves their low partition numbers while eliminating the structure that was exploitable by randomized protocols. (Similar constructions have been given by [3, 1].) To handle obstacle (ii) for Theorem 2, we actually prove a lower bound for a model that is stronger than the standard randomized model, but for which there *is* a known simulation lemma [15]. This idea alone does not handle obstacle (ii) for Theorem 1, though. For that, we start by giving a proof of the query complexity analogue of Theorem 1, then develop a way to *mimic* that argument using communication complexity, by going through information complexity (exploiting machinery from [23] and [10]). In the process, this yields our Theorem 3 (one-sided is equivalent to two-sided information complexity), which is of independent interest.

2 Complexity Measures

We study the following communication complexity models/measures; see Figure 1. For any complexity measure $\mathcal{C}$ we write $\mathsf{co}\mathcal{C}(F) \coloneqq \mathcal{C}(\neg F)$ and $2\mathcal{C}(F) \coloneqq \max\{\mathcal{C}(F), \mathsf{co}\mathcal{C}(F)\}$ for short.

- **P^{cc}:** The deterministic communication complexity of F is denoted $\mathsf{P}^{\mathsf{cc}}(F)$.
- **$\mathsf{BPP}^{\mathsf{cc}}$:** The randomized communication complexity of F is denoted $\mathsf{BPP}^{\mathsf{cc}}(F)$.
- **$\mathsf{UP}^{\mathsf{cc}}$:** Recall (e.g., [27, 21]) that a cost-c nondeterministic protocol for F corresponds to a covering (allowing overlaps) of $F^{-1}(1)$ with 2^c rectangles. A nondeterministic protocol is *unambiguous* if on every 1-input there is a unique accepting computation; combinatorially, this means we have a disjoint covering (partition) of $F^{-1}(1)$. We define $\mathsf{UP}^{\mathsf{cc}}(F) \coloneqq \lceil \log \chi_1(F) \rceil$. Thus $\mathsf{coUP}^{\mathsf{cc}}(F) = \lceil \log \chi_0(F) \rceil$, and $2\mathsf{UP}^{\mathsf{cc}}(F) \in \lceil \log \chi(F) \rceil \pm 1$.
- **$\mathsf{WAPP}^{\mathsf{cc}}$:** Abstractly speaking, a WAPP computation (*Weak Almost-Wide* PP; introduced in [8]) is a randomized computation that accepts 1-inputs with probability in $[(1-\epsilon)\alpha, \alpha]$, and 0-inputs with probability in $[0, \epsilon\alpha]$, where $\epsilon < 1/2$ is an error parameter and $\alpha = \alpha(n) > 0$ is arbitrary.
  Instantiating this for protocols, we define $\mathsf{WAPP}^{\mathsf{cc}}_{\epsilon}(F)$ as the least "cost" of a randomized (public-coin) protocol Π that computes F in the above sense; the "cost" of a protocol Π with parameter α is defined as the usual communication cost (number of bits communicated) plus $\log(1/\alpha)$. In this definition, we may assume w.l.o.g. that Π is *zero-communication* [23]: Π is simply a probability distribution over rectangles R, and Π accepts an input (x, y) iff $(x, y) \in R$ for the randomly chosen R. Such a protocol Π exchanges only 2 bits to check the condition $(x, y) \in R$, and the rest of the cost is coming from having a tiny α.
  We note that $\mathsf{WAPP}^{\mathsf{cc}}$ corresponds to the (one-sided) *smooth rectangle bound* of [17], which is known to be equivalent to *approximate nonnegative rank* [24]. A consequence of this equivalence is that $\mathsf{WAPP}^{\mathsf{cc}}$ could alternatively be defined without charging anything for $\alpha > 0$, as long as we restrict our protocols to be *private-coin*; see also [15, Theorem 9]. Also, $2\mathsf{WAPP}^{\mathsf{cc}}$ is equivalent to the *relaxed partition bound* of [23] (we elaborate on this in Section 4.2). We remark that $\mathsf{WAPP}^{\mathsf{cc}}$ is not amenable to efficient amplification of the error parameter; there can be an exponential gap between $\mathsf{WAPP}^{\mathsf{cc}}_{\epsilon}$ and $\mathsf{WAPP}^{\mathsf{cc}}_{\delta}$ for different constants ϵ and δ, at least for partial functions [15, Theorem 6].

Define the following decision tree models/measures for a boolean function $f\colon \{0,1\}^n \to \{0,1\}$:

- **P^{dt}:** The deterministic decision tree complexity of f is denoted $\mathsf{P}^{\mathsf{dt}}(f)$.
- **$\mathsf{BPP}^{\mathsf{dt}}$:** The randomized decision tree complexity of f is denoted $\mathsf{BPP}^{\mathsf{dt}}(f)$.
- **$\mathsf{UP}^{\mathsf{dt}}$:** A nondeterministic decision tree is a DNF formula. We think of the conjunctions in the DNF formula as *certificates* – partial assignments to inputs that force the function to be 1. The cost is the maximum number of input bits read by a certificate. A nondeterministic decision tree is *unambiguous* if on every 1-input there is a unique accepting certificate. We define $\mathsf{UP}^{\mathsf{dt}}(f)$ as the least cost of an unambiguous decision tree for f. Other works that have studied unambiguous decision trees include [33, 7, 13, 16, 25].
- **$\mathsf{WAPP}^{\mathsf{dt}}$:** We define $\mathsf{WAPP}^{\mathsf{dt}}_{\epsilon}(f)$ as the least height of a randomized decision tree that accepts 1-inputs with probability in $[(1-\epsilon)\alpha, \alpha]$, and 0-inputs with probability in $[0, \epsilon\alpha]$, where $\alpha = \alpha(n) > 0$ is arbitrary. (Note that only the number of queries matters; we do not charge for α being small.) Like the communication version, this measure is not amenable to efficient amplification of the error parameter [15].
  The analogue of a $\mathsf{WAPP}^{\mathsf{cc}}$ protocol being w.l.o.g. a distribution over rectangles is that a $\mathsf{WAPP}^{\mathsf{dt}}$ decision tree is w.l.o.g. a distribution over conjunctions. This implies that we may characterize $\mathsf{WAPP}^{\mathsf{dt}}_{\epsilon}(f)$ using *conical juntas*: A *conical junta* h is a nonnegative

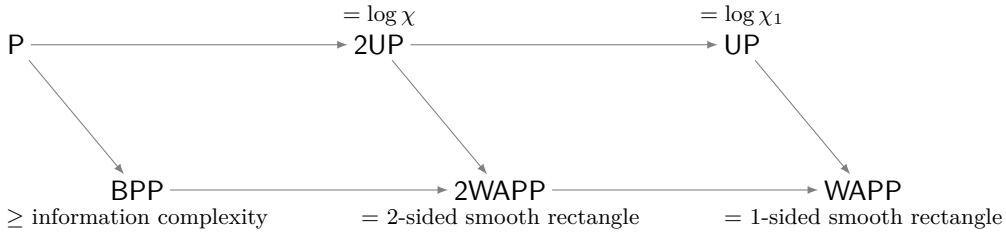


Figure 1 Models of computation that can be instantiated for both communication and query complexity. Here A ⟶ B means that model B can simulate model A without any overhead.

linear combination of conjunctions. That is, $h = \sum w_C C$ where the sum ranges over conjunctions $C\colon \{0,1\}^n \to \{0,1\}$ and $w_C \geq 0$ for all C. Then $\mathsf{WAPP}^{\mathsf{dt}}_{\epsilon}(f)$ is the least *degree* (maximum width of a conjunction with positive weight in h) of a conical junta h that ϵ-approximates f in the sense that $h(z) \in [1-\epsilon, 1]$ for all $z \in f^{-1}(1)$, and $h(z) \in [0, \epsilon]$ for all $z \in f^{-1}(0)$. Other works have studied conical juntas under such names as the (one-sided) *partition bound for query complexity* [17] and *query complexity in expectation* [22].

3 Overview

In this section we give an outline for obtaining our main results, Theorems 1–2. For complexity models/measures $\mathcal{C}$ and $\mathcal{C}'$, we informally say "$\mathcal{C}$-vs-$\mathcal{C}'$ gap" to mean the existence of a function whose $\mathcal{C}$ complexity is significantly higher than its $\mathcal{C}'$ complexity. Using the notation defined in Section 2, we can rephrase our main results as follows.

▶ **Theorem 1 ($\mathsf{BPP}^{\mathsf{cc}}$-vs-$\mathsf{2UP}^{\mathsf{cc}}$).** *There is an F such that* $\mathsf{BPP}^{\mathsf{cc}}(F) \geq \tilde{\Omega}(\mathsf{2UP}^{\mathsf{cc}}(F)^{1.5})$.

▶ **Theorem 2 ($\mathsf{BPP}^{\mathsf{cc}}$-vs-$\mathsf{UP}^{\mathsf{cc}}$).** *There is an F such that* $\mathsf{BPP}^{\mathsf{cc}}(F) \geq \tilde{\Omega}(\mathsf{UP}^{\mathsf{cc}}(F)^{2})$.

1. **Tribes-List (Section 3.1):** Our starting point is to define *Tribes-List*, a variant of a function introduced in [16]. Its purpose is to witness a BPP-vs-UP gap for query complexity.
2. **Composition (Section 3.2):** Next, we modify Tribes-List using two types of function composition, which we call *lifting* and AND-*composition*, to obtain candidate functions for BPP-vs-2UP gaps in both query and communication complexity.
3. **Overview of proofs (Section 3.3):** With the candidate functions defined, we outline our strategy to prove the desired communication lower bounds.

3.1 Tribes-List

The *Tribes-List* function $\mathrm{TL}\colon \{0,1\}^n \to \{0,1\}$ is defined on $n \coloneqq \Theta(k^3 \log k)$ bits where k is a parameter. We think of the input as a $k \times k$ matrix M with entries M_{ij} taking values from the alphabet $\Sigma \coloneqq \{0,1\} \times ([k]^{k-1} \cup \{\perp\})$. Here each entry is encoded with $\Theta(k \log k)$ bits, and we assume that the encoding of $M_{ij} = (m_{ij}, p_{ij}) \in \Sigma$ is such that a single bit is used to encode the value $m_{ij} \in \{0,1\}$ and another bit is used to encode whether or not $p_{ij} = \perp$. If $p_{ij} \neq \perp$, then we can learn its exact value in $[k]^{k-1}$ by querying all the $\Theta(k \log k)$ bits.

Informally, we have $\mathrm{TL}(M) = 1$ iff M has a unique all-$(1,*)$ column (here $*$ is a wildcard) that also contains an entry with $k-1$ pointers to entries of the form $(0,*)$ in all other

Unambiguous decision tree for TL:

Nondeterministically guess a column index $j \in [k]$. Consider the entries $M_{ij} = (m_{ij}, p_{ij})$ for $i \in [k]$: check that $m_{ij} = 1$ for all i and that $p_{ij} \neq \bot$ for at least one i (this is $\leq 2k$ queries). Let i be the first row index for which $p_{ij} \neq \bot$ and read the full value of p_{ij} (this is $\Theta(k \log k)$ queries). Interpret $p_{ij} \in [k]^{[k] \setminus \{j\}}$ as a *list of pointers*, describing a row index for all columns other than j. For each of these $k-1$ pointed-to entries $M_{i'j'}$, check that $m_{i'j'} = 0$ (this is $k-1$ queries).

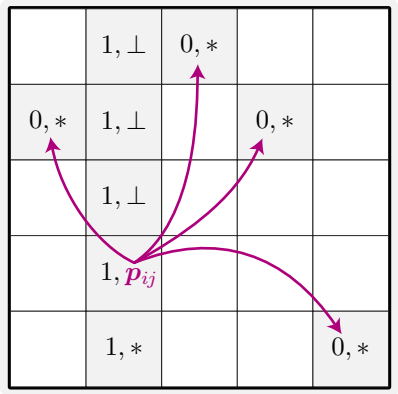


Figure 2 The unambiguous decision tree that defines the Tribes-List function.

columns. More formally, we define TL in Figure 2 by describing an unambiguous decision tree of cost $\Theta(k \log k)$ computing it.

3.2 Composition

Given a base function witnessing some complexity gap, we will establish a different but related complexity gap by transforming the function into a more complex one via one (or both) of the following operations involving function composition: *lifting* and AND-*composition*. Lifting is used to go from a query complexity gap to an analogous communication complexity gap. AND-composition is used to go from a gap with a UP upper bound to a gap with a 2UP upper bound. To show that an operation indeed converts one gap to another gap, we need two types of results: an observation showing how the relevant upper bounds behave under the operation, and a more difficult lemma showing how the relevant lower bounds behave under the operation.

Lifting

Let $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ be a fixed two-party function (called the *gadget*). We can *lift* $f\colon \{0,1\}^n \to \{0,1\}$ via the gadget g to obtain a two-party composed function $f \circ g^n\colon (\{0,1\}^b)^n \times (\{0,1\}^b)^n \to \{0,1\}$ where Alice is given $x = (x_1, \ldots, x_n)$ and Bob is given $y = (y_1, \ldots, y_n)$ (with each $x_i, y_i \in \{0,1\}^b$) and the goal is to compute $(f \circ g^n)(x,y) \coloneqq f(g(x_1,y_1), \ldots, g(x_n,y_n))$.

A decision tree for f generally yields a corresponding type of communication protocol for $f \circ g^n$: whenever the decision tree queries the i-th bit, Alice and Bob communicate $b+1$ bits to evaluate the corresponding bit $g(x_i, y_i)$. By counting conjunctions, it can be verified that such a connection holds for the 2UP and UP models as well:

▶ **Observation 4.** *For all* $f\colon \{0,1\}^n \to \{0,1\}$, $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$, *and* $\mathcal{C} \in \{\mathsf{2UP}, \mathsf{UP}\}$, *we have* $\mathcal{C}^{\mathsf{cc}}(f \circ g^n) \leq \mathcal{C}^{\mathsf{dt}}(f) \cdot O(b + \log n)$.

For any model $\mathcal{C}$, a result in the converse direction (giving a black-box method of converting a communication protocol for $f \circ g^n$ into a comparably efficient decision tree for

f) is highly nontrivial and is called a *simulation lemma*. In this work, we use a simulation lemma for $\mathcal{C} = \mathsf{WAPP}$:

▶ **Lemma 5** (Simulation for **WAPP** [15]). *For all* $f\colon \{0,1\}^n \to \{0,1\}$ *and constants* $0 < \epsilon < \delta < 1/2$, *we have* $\mathsf{WAPP}^{\mathsf{dt}}_{\delta}(f) \leq O\big(\mathsf{WAPP}^{\mathsf{cc}}_{\epsilon}(f \circ g^n)/\log n\big)$ *where* $g\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ *is the inner-product gadget defined as follows:* $b = b(n) := 100 \log n$, *and* $g(x_i, y_i) := \langle x_i, y_i \rangle \bmod 2$.

AND-composition

Given $f\colon \{0,1\}^n \to \{0,1\}$ we can compose it with the k-bit AND function to obtain $\mathrm{AND} \circ f^k\colon (\{0,1\}^n)^k \to \{0,1\}$ defined by $(\mathrm{AND} \circ f^k)(z_1, \ldots, z_k) = 1$ iff $f(z_i) = 1$ for all i. Similarly, given $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ we can obtain $\mathrm{AND} \circ F^k\colon \mathcal{X}^k \times \mathcal{Y}^k \to \{0,1\}$ defined by $(\mathrm{AND} \circ F^k)(x, y) = 1$ iff $F(x_i, y_i) = 1$ for all i.

AND-composition converts a UP upper bound into a $\mathsf{2UP}$ upper bound [16]:

▶ **Observation 6.** *For all* f *and* k, *we have* $\mathsf{2UP}^{\mathsf{dt}}(\mathrm{AND} \circ f^k) \leq k \cdot \mathsf{UP}^{\mathsf{dt}}(f) + O(\mathsf{UP}^{\mathsf{dt}}(f)^2)$. *Similarly, for all* F *and* k, *we have* $\mathsf{2UP}^{\mathsf{cc}}(\mathrm{AND} \circ F^k) \leq k \cdot \mathsf{UP}^{\mathsf{cc}}(F) + O(\mathsf{UP}^{\mathsf{cc}}(F)^2 + \log k)$.

The two parts of Observation 6 are analogous, so we describe the idea only in terms of the query complexity part. Since $\mathsf{coUP}^{\mathsf{dt}}(f) \leq \mathsf{P}^{\mathsf{dt}}(f) \leq O(\mathsf{UP}^{\mathsf{dt}}(f)^2)$, it suffices to have $\mathsf{coUP}^{\mathsf{dt}}(f)$ as the second term on the right side. The idea is to let a 1-certificate for $\mathrm{AND} \circ f^k$ be comprised of 1-certificates for each of the k copies of f, and a 0-certificate for $\mathrm{AND} \circ f^k$ be comprised of a 0-certificate for the first copy of f that evaluates to 0, together with 1-certificates for each of the preceding copies of f.

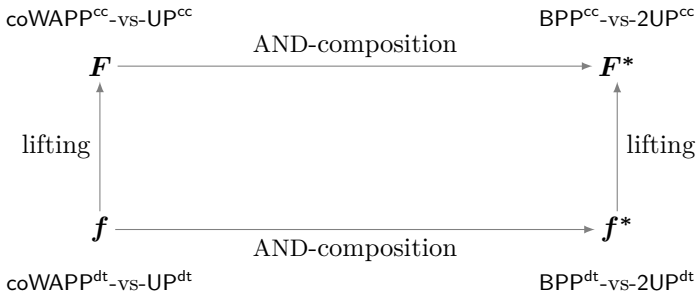
On the other hand, the following lemma (proven in Section 4.1) shows that randomized query complexity goes up by a factor of k under AND-composition.

▶ **Lemma 7.** *For all* f *and* k, *we have* $\mathsf{BPP}^{\mathsf{dt}}(f) \leq O\big(\mathsf{BPP}^{\mathsf{dt}}(\mathrm{AND} \circ f^k)/k\big)$.

We note that Lemma 7 qualitatively strengthens the tight direct sum result for randomized query complexity in [18] since computing the outputs of all k copies of f is at least as hard as computing the AND of the outputs. Similarly, if we could prove an analogue of Lemma 7 for communication complexity, it would qualitatively strengthen the notoriously-open tight direct sum conjecture for randomized communication complexity.

3.3 Overview of proofs

The following diagram shows how we construct the functions used to witness our gaps. Starting with some f, we can lift it to obtain F, or we can apply AND-composition to obtain f^*. We can obtain F^* by either lifting f^* or equivalently applying AND-composition to F.



Proof of Theorem 2

We start by discussing the proof of Theorem 2 as it will be used in the proof of Theorem 1. We actually prove the following stronger version of Theorem 2 that gives a lower bound even against $\mathsf{coWAPP}^{\mathsf{cc}}_{\epsilon}(F) \leq O(\mathsf{BPP}^{\mathsf{cc}}(F))$:

▶ **Theorem 2⋆ (coWAPP$^{\mathsf{cc}}$-vs-UP$^{\mathsf{cc}}$).** *There is an F s.t.* $\mathsf{coWAPP}^{\mathsf{cc}}_{0.04}(F) \geq \tilde{\Omega}(\mathsf{UP}^{\mathsf{cc}}(F)^2)$.

Our proof follows the same outline as in [16] and only requires us to lift the following analogous result for query complexity (proved in the full version [14]):

▶ **Lemma 8 (coWAPP$^{\mathsf{dt}}$-vs-UP$^{\mathsf{dt}}$).** $\mathsf{coWAPP}^{\mathsf{dt}}_{0.05}(\mathrm{TL}) \geq \tilde{\Omega}(\mathsf{UP}^{\mathsf{dt}}(\mathrm{TL})^2)$.

To derive Theorem 2⋆, set $f := \mathrm{TL}$ and $F := f \circ g^n$, where g is the gadget from Lemma 5 and n is the input length of f. Recall that $\mathsf{UP}^{\mathsf{dt}}(f) \geq n^{\Omega(1)}$. Thus by Observation 4, $\mathsf{UP}^{\mathsf{cc}}(F) \leq \mathsf{UP}^{\mathsf{dt}}(f) \cdot O(\log n) \leq \tilde{O}(\mathsf{UP}^{\mathsf{dt}}(f))$, and by Lemma 5, $\mathsf{coWAPP}^{\mathsf{cc}}_{0.04}(F) \geq \Omega(\mathsf{coWAPP}^{\mathsf{dt}}_{0.05}(f) \cdot \log n) \geq \Omega(\mathsf{coWAPP}^{\mathsf{dt}}_{0.05}(f))$. Thus $\mathsf{coWAPP}^{\mathsf{cc}}_{0.04}(F) \geq \tilde{\Omega}(\mathsf{UP}^{\mathsf{cc}}(F)^2)$.

Proof of Theorem 1

An "obvious" strategy for Theorem 1 would be again to first prove the analogous query complexity result and then lift it to communication complexity. (This is the outline used for the analogous result in [16].) In other words, we would follow the lower-right path in the above diagram:

Obvious strategy

(a) Start with f witnessing a $\mathsf{BPP}^{\mathsf{dt}}$-vs-$\mathsf{UP}^{\mathsf{dt}}$ gap.
(b) Obtain f^* witnessing a $\mathsf{BPP}^{\mathsf{dt}}$-vs-$2\mathsf{UP}^{\mathsf{dt}}$ gap by applying AND-composition to f.
(c) Obtain F^* witnessing a $\mathsf{BPP}^{\mathsf{cc}}$-vs-$2\mathsf{UP}^{\mathsf{cc}}$ gap by lifting f^*.

We have the tools to complete steps (a) and (b):

▶ **Lemma 9 (BPP$^{\mathsf{dt}}$-vs-2UP$^{\mathsf{dt}}$).** *There is an f such that* $\mathsf{BPP}^{\mathsf{dt}}(f) \geq \tilde{\Omega}(2\mathsf{UP}^{\mathsf{dt}}(f)^{1.5})$.

Proof. This is witnessed by $f^* := \mathrm{AND} \circ \mathrm{TL}^k$ where $k := \mathsf{UP}^{\mathsf{dt}}(\mathrm{TL})$. By Observation 6, $2\mathsf{UP}^{\mathsf{dt}}(f^*) \leq O(k^2)$, and by Lemmas 7–8,

$$\mathsf{BPP}^{\mathsf{dt}}(f^*) \geq \Omega(k \cdot \mathsf{BPP}^{\mathsf{dt}}(\mathrm{TL})) \geq \Omega(k \cdot \mathsf{coWAPP}^{\mathsf{dt}}_{0.05}(\mathrm{TL})) \geq \tilde{\Omega}(k^3)\,. \qquad \blacktriangleleft$$

Unfortunately, we do not know how to carry out step (c), because we currently lack a simulation lemma for BPP. (We believe that such a lemma is true, and it is an interesting open problem to prove this!) We get around this obstacle by reversing the order of steps (b) and (c), that is, we instead follow the upper-left path in the diagram:

Modified strategy

(a′) Start with f witnessing a $\mathsf{coWAPP}^{\mathsf{dt}}$-vs-$\mathsf{UP}^{\mathsf{dt}}$ gap.
(b′) Obtain F witnessing a $\mathsf{coWAPP}^{\mathsf{cc}}$-vs-$\mathsf{UP}^{\mathsf{cc}}$ gap by lifting f.
(c′) Obtain F^* witnessing a $\mathsf{BPP}^{\mathsf{cc}}$-vs-$2\mathsf{UP}^{\mathsf{cc}}$ gap by applying AND-composition to F.

Steps (a′) and (b′) are just Theorem 2⋆. For step (c′) it would suffice to have an analogue of Lemma 7 for communication complexity. This is open, but fortunately we have some wiggle room since it suffices to have coWAPP_ϵ instead of BPP on the left side of Lemma 7. For this, we *can* prove a communication analogue (indeed, with $\mathsf{2WAPP}_\epsilon$ instead of coWAPP_ϵ):

▶ **Lemma 10.** *For all F, k, and constants $0 < \epsilon < 1/2$, we have*

$$\mathsf{2WAPP}^{\mathsf{cc}}_\epsilon(F) \;\leq\; O\big(\mathsf{BPP}^{\mathsf{cc}}(\mathrm{AND} \circ F^k)/k + \log \mathsf{BPP}^{\mathsf{cc}}(\mathrm{AND} \circ F^k)\big).$$

To derive Theorem 1, let F be the function in Theorem 2⋆, and let $F^* \coloneqq \mathrm{AND} \circ F^k$ where $k \coloneqq \mathsf{UP}^{\mathsf{cc}}(F)$. Then F^* witnesses Theorem 1: By Observation 6, $\mathsf{2UP}^{\mathsf{cc}}(F^*) \leq O(k^2)$, and by Lemma 10, $\mathsf{BPP}^{\mathsf{cc}}(F^*) \geq \Omega\big(k \cdot (\mathsf{2WAPP}^{\mathsf{cc}}_{0.04}(F) - O(\log k))\big) \geq \Omega\big(k \cdot (\mathsf{coWAPP}^{\mathsf{cc}}_{0.04}(F) - O(\log k))\big) \geq \tilde{\Omega}(k^3)$.

Proof of Lemma 10

We start with the intuition for the proof of Lemma 7, which is a warmup for Lemma 10. For brevity let $f^* \coloneqq \mathrm{AND} \circ f^k$. Given an input z for f, the basic idea is to plant z into a random coordinate of $f^*(z_1, \ldots, z_k)$, and plant random 1-inputs into the other coordinates, and then run the randomized decision tree for f^*. If q is the query complexity of f^*, the expected number of bits of z that are queried (over a random 1-input) will be at most q/k. Our new randomized decision tree will simulate this but abort after $8q/k$ queries to z have been made. If an answer is returned, we output the same value for $f(z)$, and if no answer is returned within this many queries, then we output 0. A simple analysis shows that we succeed with high probability in the average-case (which is equivalent to worst-case by the minimax theorem).

To prove Lemma 10, we would like to mimic this argument in the communication world, using the fact that internal information complexity is sandwiched between $\mathsf{BPP}^{\mathsf{cc}}$ and $\mathsf{2WAPP}^{\mathsf{cc}}$ [23] and satisfies a sort of AND-composition analogous to Lemma 7 using well-known properties (by planting the input into a random coordinate, and planting random 1-inputs into the other coordinates). However there is a significant barrier to this idea "just working": the AND-composition property (direct sum lemma) requires a distribution over 1-inputs of F (one-sided), while the relation to $\mathsf{2WAPP}^{\mathsf{cc}}$ requires an arbitrary distribution over inputs to F (two-sided). To bridge this divide, we prove a new property of information complexity: the one-sided version is essentially equivalent to the two-sided version. A key ingredient in showing the latter is the "information odometer" of [10], which allows us to keep track of the amount of information that has been revealed, and abort the protocol once we have reached our limit, and argue that we can carry this out without revealing too much extra information. We note that this one-vs-two sided information complexity lemma is the only component of the proof of Theorem 1 that distinguishes between arbitrary rectangle partitions ($\mathsf{2UP}^{\mathsf{cc}}$) and rectangle partitions induced by protocols (P^{cc}).

Organization

The only ingredients that remain to be proved are Lemma 8 (which we prove in the full version [14] and Lemma 7 and Lemma 10 (both of which we prove in Section 4).

4 AND-Composition Lemmas

In this section we prove Lemma 7 and Lemma 10, restated here for convenience.

▶ **Lemma 11.** *For all f and k, we have* $\mathsf{BPP}^{\mathsf{dt}}(f) \le O\big(\mathsf{BPP}^{\mathsf{dt}}(\mathrm{AND} \circ f^k)/k\big)$.

▶ **Lemma 12.** *For all F, k, and constants $0 < \epsilon < 1/2$, we have*

$$\mathsf{2WAPP}^{\mathsf{cc}}_{\epsilon}(F) \;\le\; O\big(\mathsf{BPP}^{\mathsf{cc}}(\mathrm{AND} \circ F^k)/k + \log \mathsf{BPP}^{\mathsf{cc}}(\mathrm{AND} \circ F^k)\big).$$

4.1 AND-composition for query complexity

We now prove Lemma 7. For brevity let $f^* \coloneqq \mathrm{AND} \circ f^k$. Let T^* be a height-q randomized decision tree for f^* with error $1/8$. We design a height-$8q/k$ randomized decision tree for f with error $1/4$.

Let D be an arbitrary distribution over $f^{-1}(1)$. Consider the following randomized decision tree T that takes $z \in \{0,1\}^n$ as input:

1. Pick $i \in [k]$ uniformly at random and let $z_i \coloneqq z$.
2. For $j \in [k] \setminus \{i\}$ sample $z_j \sim D$ independently.
3. Run $T^*(z_1, \ldots, z_k)$ until it has made $8q/k$ queries in the i-th component.
4. If T^* already produced an output in Step 3, output the same bit; else output 0.

Note that with probability 1 we have $f^*(z_1, \ldots, z_k) = f(z)$. Let R_T denote T's randomness and R_{T^*} denote T^*'s randomness. If $f(z) = 0$ then

$$\mathbb{P}_{R_T}[T(z) = 1] \;\le\; \max_{(z_1,\ldots,z_k) \in (f^*)^{-1}(0)} \mathbb{P}_{R_{T^*}}[T^*(z_1, \ldots, z_k) = 1] \;\le\; 1/8 \;\le\; 1/4.$$

Furthermore,

$$\begin{aligned}
\mathbb{P}_{z\sim D,\, R_T}[T(z) = 0] \;&=\; \mathbb{P}_{z_1,\ldots,z_k \sim D,\, i\in[k],\, R_{T^*}}\begin{bmatrix} T^*(z_1,\ldots,z_k) \text{ outputs 0 or makes more} \\ \text{than } 8q/k \text{ queries in the } i\text{-th component} \end{bmatrix} \\
&\le\; \max_{(z_1,\ldots,z_k)\in(f^*)^{-1}(1)} \left(\begin{array}{l} \mathbb{P}_{R_{T^*}}[T^*(z_1,\ldots,z_k) = 0] + \\ \max_{R_{T^*}} \mathbb{P}_{i\in[k]} \begin{bmatrix} T^*(z_1,\ldots,z_k) \text{ makes more than} \\ 8q/k \text{ queries in the } i\text{-th component} \end{bmatrix} \end{array} \right) \\
&\le\; 1/8 + 1/8 \;=\; 1/4.
\end{aligned}$$

Now let D be an arbitrary distribution over $\{0,1\}^n$ and define T w.r.t. $(D \,|\, f^{-1}(1))$. We have

$$\begin{aligned}
\mathbb{P}_{z\sim D,\, R_T}[T(z) \neq f(z)] \;&=\; \textstyle\sum_{b\in\{0,1\}} \mathbb{P}_{z\sim(D\,|\,f^{-1}(b)),\, R_T}[T(z) \neq b] \cdot \mathbb{P}_{z\sim D}[f(z) = b] \\
&\le\; \textstyle\sum_{b\in\{0,1\}} (1/4) \cdot \mathbb{P}_{z\sim D}[f(z) = b] \;=\; 1/4.
\end{aligned}$$

By the minimax theorem, there is a height-$8q/k$ randomized decision tree (a mixture of the T's) that on any input produces the wrong output with probability $\le 1/4$.

4.2 Definitions

We adopt the following conventions throughout the proof of Lemma 10. We denote random variables with upper-case letters, and we denote particular outcomes of the random variables with the corresponding lower-case letters. All communication protocols are randomized and mixed-coin, and we use (R, R_A, R_B) to denote the public randomness, Alice's private randomness, and Bob's private randomness, respectively. We say a protocol Π is ϵ-correct for F if for all (x, y), $\mathbb{P}_{R,R_A,R_B}[\Pi(x, y) = F(x, y)] \ge 1 - \epsilon$. For a distribution D over

inputs, we say Π is (ϵ, D)-correct for F if $\mathbb{P}_{(X,Y)\sim D,\, R,R_A,R_B}[\Pi(X,Y) = F(X,Y)] \geq 1-\epsilon$. The internal information cost of a protocol Π with respect to $(X,Y) \sim D$ is defined as $\mathrm{IC}_D(\Pi) := \mathbb{I}(R, M\,; X \,|\, Y) + \mathbb{I}(R, M\,; Y \,|\, X) = \mathbb{I}(M\,; X \,|\, Y, R) + \mathbb{I}(M\,; Y \,|\, X, R)$ where the random variable M is the concatenation of all messages. We also let $\mathrm{CC}(\Pi)$ denote the worst-case communication cost of Π.

It is convenient for us to work with a measure $\mathsf{2WAPP}^{\mathsf{cc}*}$ that is defined slightly differently from $\mathsf{2WAPP}^{\mathsf{cc}}$ but is equivalent in the sense that for all F and $0 < \epsilon < 1/2$, $\mathsf{2WAPP}^{\mathsf{cc}}_{\epsilon}(F) \leq \mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon}(F) \leq O(\mathsf{2WAPP}^{\mathsf{cc}}_{\epsilon/2}(F))$. We note that $\mathsf{2WAPP}^{\mathsf{cc}}$ directly expresses the two-sided smooth rectangle bound of [17], while $\mathsf{2WAPP}^{\mathsf{cc}*}$ directly expresses the relaxed partition bound of [23] and was the definition used in [15].

▶ **Definition 13.** We define $\mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon}(F)$ as the minimum of $\mathrm{CC}(\Pi)+\log(1/\alpha)$ over all $\alpha > 0$ and all protocols Π with output values $\{0,1,\perp\}$ such that for all (x,y), $\mathbb{P}[\Pi(x,y) \neq \perp] \leq \alpha$ and $\mathbb{P}[\Pi(x,y) = F(x,y)] \geq (1-\epsilon)\alpha$ (i.e., Π is $(1-(1-\epsilon)\alpha)$-correct).

We also need the distributional version of $\mathsf{2WAPP}^{\mathsf{cc}*}$.

▶ **Definition 14.** For an input distribution D, we define $\mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon,D}(F)$ as the minimum of $\mathrm{CC}(\Pi) + \log(1/\alpha)$ over all $\alpha > 0$ and all protocols Π with output values $\{0,1,\perp\}$ such that $\mathbb{P}[\Pi(x,y) \neq \perp] \leq \alpha$ for all (x,y), and $\mathbb{P}[\Pi(X,Y) = F(X,Y)] \geq (1-\epsilon)\alpha$ for $(X,Y) \sim D$ (i.e., Π is $(1-(1-\epsilon)\alpha, D)$-correct).

4.3 AND-composition for communication complexity

We now outline the proof of Lemma 10. Recall that the proof of Lemma 7 involved these steps:

(i) embedding the input into a random coordinate of a k-tuple and filling the other coordinates with random 1-inputs (to cut the cost on 1-inputs by a factor k),

(ii) aborting the execution if the cost became too high (to ensure low cost also on 0-inputs while maintaining average-case correctness on 1-inputs),

(iii) using the minimax theorem to go from average-case to worst-case correctness.

We start by noting that an analogue of (i) holds for information complexity (which lower bounds $\mathsf{BPP}^{\mathsf{cc}}$). Then as one of our main technical contributions we prove an analogue of (ii) for information complexity. Then inbetween (ii) and (iii) we insert a step applying the known result that information complexity upper bounds $\mathsf{2WAPP}^{\mathsf{cc}*}$ in the distributional setting. Finally we use the analogue of (iii) for $\mathsf{2WAPP}^{\mathsf{cc}*}$. Formally, Lemma 10 follows by stringing together the following lemmas.

▶ **Lemma 15.** *Fix any F, k, $0 < \epsilon < 1/2$, and distribution D over $F^{-1}(1)$. For every ϵ-correct protocol Π for $\mathrm{AND} \circ F^k$ there is an ϵ-correct protocol Π' for F with $\mathrm{IC}_D(\Pi') \leq \mathrm{CC}(\Pi)/k$ and $\mathrm{CC}(\Pi') \leq \mathrm{CC}(\Pi)$.*

▶ **Lemma 16.** *Fix any F, constants $0 < \epsilon < \delta < 1/2$, and input distribution D, and let $D^1 := (D \,|\, F^{-1}(1))$. For every (ϵ, D)-correct protocol Π there is a (δ, D)-correct protocol Π' with $\mathrm{IC}_D(\Pi') \leq O\big(\mathrm{IC}_{D^1}(\Pi) + \log(\mathrm{CC}(\Pi) + 2)\big)$.*

▶ **Lemma 17.** *Fix any F, constants $0 < \epsilon < \delta < 1/2$, and input distribution D. For every (ϵ, D)-correct protocol Π we have $\mathsf{2WAPP}^{\mathsf{cc}*}_{\delta,D}(F) \leq O(\mathrm{IC}_D(\Pi) + 1)$.*

▶ **Lemma 18.** *Fix any F and $0 < \epsilon < 1/2$. Then $\mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon}(F) \leq 2 + \max_D \mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon,D}(F)$.*

Lemma 15 is a standard application of the "direct sum" property of information cost. Lemma 16 is proved in Section 4.4 and relies on [10]. Lemma 17 is due to [23, Theorem 1.1 of the ECCC version]. Lemma 18 follows from an argument in [23, Appendix A of the ECCC version] that uses LP duality.

The moral conclusion of Lemma 16 is that "one-sided information complexity" is essentially equivalent to "two-sided information complexity" for average-case protocols. Combining Lemma 16 with [9, Theorem 3.5 of the ECCC version] shows that a similar equivalence holds for worst-case protocols. More specifically, a distribution-independent definition of information complexity for bounded-error protocols can be obtained by maximizing over all input distributions; our corollary shows that this measure is essentially unchanged if we maximize only over distributions over 1-inputs (or symmetrically, 0-inputs).

▶ **Corollary 19.** *Fix any F, constants $0 < \epsilon < \delta < 1/2$, and $b \in \{0,1\}$. Then*

$$\inf_{\substack{\delta\text{-correct} \\ \text{protocols } \Pi}} \max_{\substack{D \text{ over} \\ \text{all inputs}}} \mathrm{IC}_D(\Pi) \;\leq\; \max_{\substack{D \text{ over} \\ b\text{-inputs}}} \inf_{\substack{\epsilon\text{-correct} \\ \text{protocols } \Pi}} O\big(\mathrm{IC}_D(\Pi) + \log(\mathrm{CC}(\Pi) + 2)\big).$$

Theorem 3 follows by swapping the quantifiers on the right side of the inequality in Corollary 19 (which only weakens the statement), and by straightforwardly accounting for the communication cost in the proof. We can also assume the protocol Π' has error $\leq 1/3$ by a standard error reduction technique (take a majority vote of several runs of the protocol), which does not affect information complexity except by constant factors. We do not directly employ this worst-case version of Lemma 16, but it is used in the follow-up work [5].

4.4 One-sided information vs. two-sided information

Intuition for Lemma 16

Recall the following idea, which was implicit in the proof of Lemma 7. Suppose we have a randomized decision tree computing some function, and we have a bound b on the expected number of queries made over a random 1-input. Then to obtain a randomized decision tree with a worst-case query bound, we can keep track of the number of queries made during the execution and halt and output 0 if it exceeds, say, $8b$. Correctness on 0-inputs is maintained since we either run the original decision tree to completion and thus output 0 with high probability, or we abort and output 0 anyway. We get average-case correctness on 1-inputs since by Markov's inequality, with probability at least $7/8$ the original decision tree uses at most $8b$ queries, in which case we run it to completion and output 1 with high probability.

The high-level intuition is to mimic this idea for information complexity. We have a protocol with a bound on the information cost w.r.t. the distribution D^1 over 1-inputs. The "information odometer" of [10] allows us to "keep track of" information cost, so we can halt and output 0 if it becomes too large. This will guarantee that the information cost is low w.r.t. the input distribution D, and correctness on 0-inputs is maintained. However, there is a complication with showing the average-case correctness on 1-inputs.

For each computation path specified by an input (x, y), an outcome of public randomness r, and a full sequence of messages m, there is a contribution $c_{x,y,r,m}$ such that the information cost w.r.t. D is the expectation of $c_{x,y,r,m}$ over a random computation path with $(x, y) \sim D$. Similarly, there is a contribution $c^1_{x,y,r,m}$ such that the information cost w.r.t. D^1 is the expectation of $c^1_{x,y,r,m}$ over a random computation path with $(x, y) \sim D^1$. These contributions play the role of "number of queries" along a computation path in the decision tree setting, but a crucial difference is that $c_{x,y,r,m} \neq c^1_{x,y,r,m}$ in general; i.e., the contribution to information cost depends on the input distribution (whereas number of queries did not). To show the

average-case correctness on 1-inputs, we need a bound on the typical value of $c_{x,y,r,m}$, whereas the assumption that information cost w.r.t. D^1 is low gives us a bound on the typical value of $c^1_{x,y,r,m}$.

Thus the heart of the argument is to show that typically, $c_{x,y,r,m}$ is not much larger than $c^1_{x,y,r,m}$. Intuitively, one might expect the difference to be at most 1, since the only additional information that can be revealed (beyond what is revealed under D^1) should be the fact that (x, y) is a 1-input (which is 1 bit of information). More precisely, we show that for given (x, y), the expected difference depends on how balanced F is on the x row and the y column. Then we just need to note that F is typically reasonably balanced for both the x row and the y column.

The formal proof of Lemma 16 is deferred to the full version [14] due to space constraints.

Acknowledgments. We thank Mark Braverman, Troy Lee, and Omri Weinstein for discussions. Work done by M.G. while at IBM Research Almaden.

References

1 Scott Aaronson, Shalev Ben-David, and Robin Kothari. Separations in query complexity using cheat sheets. In *Proceedings of the 48th Symposium on Theory of Computing (STOC)*, pages 863–876. ACM, 2016. doi:10.1145/2897518.2897644.

2 Alfred Aho, Jeffrey Ullman, and Mihalis Yannakakis. On notions of information transfer in VLSI circuits. In *Proceedings of the 15th Symposium on Theory of Computing (STOC)*, pages 133–139. ACM, 1983. doi:10.1145/800061.808742.

3 Andris Ambainis, Kaspars Balodis, Aleksandrs Belovs, Troy Lee, Miklos Santha, and Juris Smotrovs. Separations in query complexity based on pointer functions. In *Proceedings of the 48th Symposium on Theory of Computing (STOC)*, pages 800–813. ACM, 2016. doi:10.1145/2897518.2897524.

4 Andris Ambainis, Martins Kokainis, and Robin Kothari. Nearly optimal separations between communication (or query) complexity and partitions. In *Proceedings of the 31st Computational Complexity Conference (CCC)*, pages 4:1–4:14. Schloss Dagstuhl, 2016. doi:10.4230/LIPIcs.CCC.2016.4.

5 Anurag Anshu, Aleksandrs Belovs, Shalev Ben-David, Mika Göös, Rahul Jain, Robin Kothari, Troy Lee, and Miklos Santha. Separations in communication complexity using cheat sheets and information complexity. In *Proceedings of the 57th Symposium on Foundations of Computer Science (FOCS)*, pages 555–564. IEEE, 2016. doi:10.1109/FOCS.2016.66.

6 Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004. doi:10.1016/j.jcss.2003.11.006.

7 Aleksandrs Belovs. Non-intersecting complexity. In *Proceedings of the 32nd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, pages 158–165. Springer, 2006. doi:10.1007/11611257_13.

8 Elmar Böhler, Christian Glaßer, and Daniel Meister. Error-bounded probabilistic computations between MA and AM. *Journal of Computer and System Sciences*, 72(6):1043–1076, 2006. doi:10.1016/j.jcss.2006.05.001.

9 Mark Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015. doi:10.1137/130938517.

10 Mark Braverman and Omri Weinstein. An interactive information odometer and applications. In *Proceedings of the 47th Symposium on Theory of Computing (STOC)*, pages 341–350. ACM, 2015. doi:10.1145/2746539.2746548.

11 Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *Journal of the ACM*, 62(2):17:1–17:23, 2015. doi:10.1145/2716307.

12 Dmitry Gavinsky and Shachar Lovett. En Route to the Log-Rank Conjecture: New Reductions and Equivalent Formulations. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 514–524. Springer, 2014. doi:10.1007/978-3-662-43948-7_43.

13 Mika Göös. Lower bounds for clique vs. independent set. In *Proceedings of the 56th Symposium on Foundations of Computer Science (FOCS)*, pages 1066–1076. IEEE, 2015. doi:10.1109/FOCS.2015.69.

14 Mika Göös, T. S. Jayram, Toniann Pitassi, and Thomas Watson. Randomized communication vs. partition number. Technical Report TR15-169, Electronic Colloquium on Computational Complexity (ECCC), 2015. URL: https://eccc.weizmann.ac.il/report/2015/169/.

15 Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM Journal on Computing*, 45(5):1835–1869, 2016. doi:10.1137/15M103145X.

16 Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 56th Symposium on Foundations of Computer Science (FOCS)*, pages 1077–1088. IEEE, 2015. doi:10.1109/FOCS.2015.70.

17 Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 25th Conference on Computational Complexity (CCC)*, pages 247–258. IEEE, 2010. doi:10.1109/CCC.2010.31.

18 Rahul Jain, Hartmut Klauck, and Miklos Santha. Optimal direct sum results for deterministic and randomized decision tree complexity. *Information Processing Letters*, 110(20):893–897, 2010. doi:10.1016/j.ipl.2010.07.020.

19 Rahul Jain, Troy Lee, and Nisheeth Vishnoi. A quadratically tight partition bound for classical communication complexity and query complexity. Technical report, arXiv, 2014. arXiv:1401.4512.

20 T.S. Jayram, Swastik Kopparty, and Prasad Raghavendra. On the communication complexity of read-once AC^0 formulae. In *Proceedings of the 24th Conference on Computational Complexity (CCC)*, pages 329–340. IEEE, 2009. doi:10.1109/CCC.2009.39.

21 Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*, volume 27 of *Algorithms and Combinatorics*. Springer, 2012.

22 Jedrzej Kaniewski, Troy Lee, and Ronald de Wolf. Query complexity in expectation. In *Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 761–772. Springer, 2015. doi:10.1007/978-3-662-47672-7_62.

23 Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *SIAM Journal on Computing*, 44(5):1550–1572, 2015. doi:10.1137/130928273.

24 Gillat Kol, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Approximate nonnegative rank is equivalent to the smooth rectangle bound. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 701–712. Springer, 2014. doi:10.1007/978-3-662-43948-7_58.

25 Robin Kothari, David Racicot-Desloges, and Miklos Santha. Separating decision tree complexity from subcube partition complexity. In *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*, pages 915–930. Schloss Dagstuhl, 2015. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.915.

26 Eyal Kushilevitz. Unpublished. Cited in [32], 1994.

27 Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

28 Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–399, 2007. doi:10.1561/0400000040.

29 Nikos Leonardos and Michael Saks. Lower bounds on the randomized communication complexity of read-once functions. *Computational Complexity*, 19(2):153–181, 2010. doi:10.1007/s00037-010-0292-2.

30 László Lovász and Michael Saks. Lattices, Möbius functions and communication complexity. In *Proceedings of the 29th Symposium on Foundations of Computer Science (FOCS)*, pages 81–90. IEEE, 1988. doi:10.1109/SFCS.1988.21924.

31 Sagnik Mukhopadhyay and Swagato Sanyal. Towards better separation between deterministic and randomized query complexity. In *Proceedings of 35th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 206–220. Schloss Dagstuhl, 2015. doi:10.4230/LIPIcs.FSTTCS.2015.206.

32 Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995. doi:10.1007/BF01192527.

33 Petr Savický. On determinism versus unambiguous nondeterminism for decision trees. Technical Report TR02-009, Electronic Colloquium on Computational Complexity (ECCC), 2002. URL: http://eccc.hpi-web.de/report/2002/009/.

34 Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991. doi:10.1016/0022-0000(91)90024-Y.