Randomized Communication vs. Partition Number

Mika Göös[†] T.S. Jayram Toniann Pitassi Thomas Watson
Harvard University *IBM Almaden* *University of Toronto* *University of Memphis*

January 4, 2018

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.

1 Introduction

A prior work [GPW15] 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 [GPW15] 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 [ABB+16b]—informally states that the information complexity of a function (as defined by [Bra15]) remains essentially unchanged if the cost is 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 we define $\mathrm{CC}(\Pi)$ as the worst-case communication cost, 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 $\mathrm{IC}(\Pi') \le O\big(\mathrm{IC}^b(\Pi) + \log(\mathrm{CC}(\Pi)+2)\big)$. Moreover, $\mathrm{CC}(\Pi') \le O\big(\mathrm{CC}(\Pi) \cdot \log(\mathrm{CC}(\Pi)+2)\big)$.*

[†]Work done while at IBM Research Almaden.

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 [JK10]. In particular, Theorem 1 shows that randomized complexity can be polynomially larger than the *partition bound* [JK10, JLV14], 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 [AUY83]. 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 [Yan91, §4] or [Juk12, §4.4] for the equivalence. Yannakakis [Yan91] (extending [AUY83]) 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 [Yan91, FMP+15] 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 [Göö15]. Technically speaking, the lower bound from [Göö15] 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 [LS88] postulates that the deterministic communication complexity of F is polynomially related to $\log \operatorname{rank}(F)$. Gavinsky and Lovett [GL14] 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 $\log \operatorname{rank}(F) \le \log \chi_1(F)$; the previous best lower bound was $\Omega(\log^{1.63} \operatorname{rank}(F))$ due to Kushilevitz [Kus94]. 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* [LS07, 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 ([BJKS04, Bra15]): for any protocol Π computing $\text{AND}_k \circ F^k$ (i.e., the AND of k copies of F) there exists a protocol Π' computing F with $\text{IC}^1(\Pi') \leq O(\text{IC}^1(\Pi)/k)$ (see, e.g., [ABB+16b, Claim 37]). One can also formulate a dual lemma for $\text{OR}_k \circ F^k$ 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 [JKR09, LS10] even for *unbalanced* trees.

Another application of Theorem 3 appears in the recent work [ABB+16b]. 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 [AKK16] (which is a variation of usual AND–OR trees).

Our one-vs.-two-sided information complexity equivalence has also been used in [Ass17] to prove certain streaming lower bounds for the set cover problem.

1.2 Our techniques

The basic strategy in [GPW15] 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 [GPW15] are too easy for randomized protocols (as shown by [MS15]). (ii) There is no known simulation lemma for the bounded-error randomized setting (though since this paper was written, such a lemma has, in fact, been proven [GPW17]).

To handle obstacle (i), we modify the functions from [GPW15] 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 [ABB+16a, ABK16].) 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 [GLM+16]. 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 [KLL+15] and [BW15a]). 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 $\mathsf{co}\mathcal{C}(F) := \mathcal{C}(\neg F)$ and $2\mathcal{C}(F) := \max\{\mathcal{C}(F), \mathsf{co}\mathcal{C}(F)\}$ 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., [KN97, Juk12]) 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

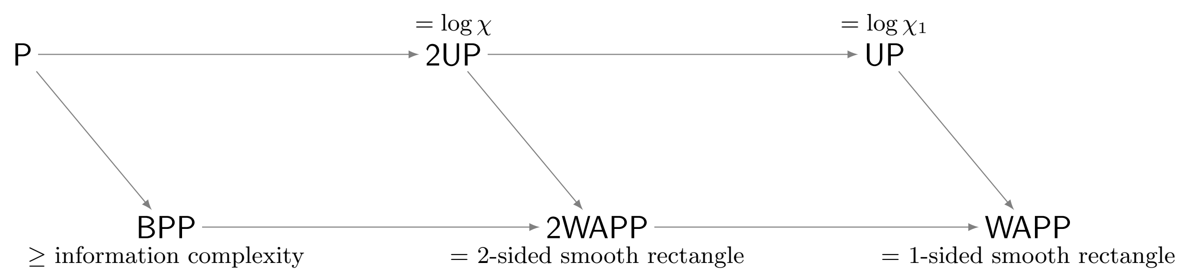

Figure 1: Models of computation that can be instantiated for both communication and query complexity. Here $\mathsf{A} \longrightarrow \mathsf{B}$ means that model B can simulate model A without any overhead.

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 $\mathsf{2UP}^{\mathsf{cc}}(F) \in \lceil \log \chi(F) \rceil \pm 1$.

– **$\mathsf{WAPP}^{\mathsf{cc}}$:** Abstractly speaking, a WAPP computation (*Weak Almost-Wide* PP; introduced in [BGM06]) 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* [KLL+15]: 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 [JK10], which is known to be equivalent to *approximate nonnegative rank* [KMSY14]. 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 [GLM+16, Theorem 9]. Also, $\mathsf{2WAPP}^{\mathsf{cc}}$ is equivalent to the *relaxed partition bound* of [KLL+15] (we elaborate on this in Section 5.2). 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 [GLM+16, Theorem 6].

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

– **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 [Sav02, Bel06, Göö15, GPW15, KRS15].

– **$\mathsf{WAPP}^{\mathsf{dt}}$:** 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 [GLM+16].

  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 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* [JK10] and *query complexity in expectation* [KLdW15].

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 $\mathsf{BPP}^{\mathsf{cc}}(F) \geq \tilde{\Omega}(\mathsf{UP}^{\mathsf{cc}}(F)^2)$.*

(§ 3.1) **Tribes-List:** Our starting point is to define *Tribes-List*, a variant of a function introduced in [GPW15]. Its purpose is to witness a BPP-vs-UP gap for query complexity.

(§ 3.2) **Composition:** 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-$\mathsf{2UP}$ gaps in both query and communication complexity.

(§ 3.3) **Overview of proofs:** 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 $n := \Theta(k^3 \log k)$ 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 $\Sigma := \{0,1\} \times ([k]^{k-1} \cup \{\perp\})$. 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 columns. More formally, we define TL in Figure 2 by describing an unambiguous decision tree of cost $\Theta(k \log k)$ computing it.

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 $p_{ij} \neq \perp$ for at least one i (this is $\leq 2k$ queries). Let i be the first row index for which $p_{ij} \neq \perp$ 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).

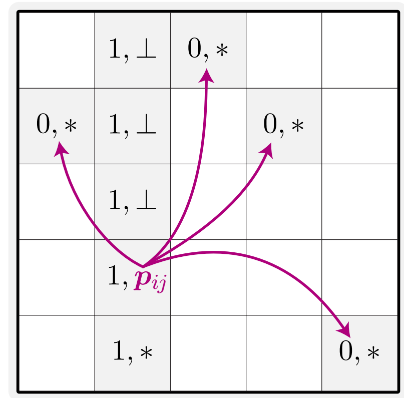

Figure 2: The unambiguous decision tree that defines the Tribes-List function.

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 $\mathsf{2UP}$ 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 $\mathsf{2UP}$ 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 [GLM+16]). *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:* $b = b(n) \coloneqq 100 \log n$, *and* $g(x_i, y_i) \coloneqq \langle x_i, y_i \rangle \bmod 2$.

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

AND-composition converts a UP upper bound into a $2\mathsf{UP}$ upper bound [GPW15]:

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

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 $\text{AND} \circ f^k$ be comprised of 1-certificates for each of the k copies of f, and a 0-certificate for $\text{AND} \circ f^k$ 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 5.1) shows that randomized query complexity goes up by a factor of k under AND-composition.

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

We note that Lemma 7 qualitatively strengthens the tight direct sum result for randomized query complexity in [JKS10] 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 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.

$$\begin{tikzcd}
\mathsf{coWAPP}^{\mathsf{cc}}\text{-vs-}\mathsf{UP}^{\mathsf{cc}} & & \mathsf{BPP}^{\mathsf{cc}}\text{-vs-}2\mathsf{UP}^{\mathsf{cc}} \\
\boldsymbol{F} \arrow[rr, "\text{AND-composition}"] & & \boldsymbol{F^*} \\
\boldsymbol{f} \arrow[u, "\text{lifting}"] \arrow[rr, "\text{AND-composition}"'] & & \boldsymbol{f^*} \arrow[u, "\text{lifting}"'] \\
\mathsf{coWAPP}^{\mathsf{dt}}\text{-vs-}\mathsf{UP}^{\mathsf{dt}} & & \mathsf{BPP}^{\mathsf{dt}}\text{-vs-}2\mathsf{UP}^{\mathsf{dt}}
\end{tikzcd}$$

Proof sketch 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* ($\mathsf{coWAPP}^{\mathsf{cc}}$-vs-$\mathsf{UP}^{\mathsf{cc}}$). *There is an F such that $\mathsf{coWAPP}^{\mathsf{cc}}_{0.04}(F) \geq \tilde{\Omega}(\mathsf{UP}^{\mathsf{cc}}(F)^2)$.*

Our proof follows the same outline as in [GPW15] and only requires us to lift the following analogous result for query complexity (proved in Section 4):

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

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 sketch 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 [GPW15].) 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-$\mathsf{2UP}^{\mathsf{dt}}$ gap by applying AND-composition to f.
(c) Obtain F^* witnessing a $\mathsf{BPP}^{\mathsf{cc}}$-vs-$\mathsf{2UP}^{\mathsf{cc}}$ gap by lifting f^*.

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

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

Proof. This is witnessed by $f^* := \mathrm{AND} \circ \mathrm{TL}^k$ where $k := \mathsf{UP}^{\mathsf{dt}}(\mathrm{TL})$. By Observation 6, $\mathsf{2UP}^{\mathsf{dt}}(f^*) \leq O(k^2)$, and by Lemmas 7–8, $\mathsf{BPP}^{\mathsf{dt}}(f^*) \geq \Omega(k \cdot \mathsf{BPP}^{\mathsf{dt}}(\mathrm{TL})) \geq \Omega(k \cdot \mathsf{coWAPP}^{\mathsf{dt}}_{0.05}(\mathrm{TL})) \geq \tilde{\Omega}(k^3)$. □

When this paper was written we did not know how to carry out step (c), because we lacked a simulation lemma for BPP. (Such a lemma is now known [GPW17].) 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-$\mathsf{2UP}^{\mathsf{cc}}$ 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 $F^* := \mathrm{AND} \circ F^k$ where $k := \mathsf{UP}^{\mathsf{cc}}(F)$. 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 sketch 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 $\mathsf{BPP^{cc}}$ and $\mathsf{2WAPP^{cc}}$ [KLL+15] 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 $\mathsf{2WAPP^{cc}}$ 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 [BW15a], 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 ($\mathsf{2UP^{cc}}$) and rectangle partitions induced by protocols ($\mathsf{P^{cc}}$). The log term in Lemma 10 comes from the information odometer ingredient, and it is known than by composing with the k-bit majority function instead of AND, a version of Lemma 10 holds without the log term [Wat17].

Organization. The only ingredients that remain to be proved are Lemma 8 (which we prove in Section 4) and Lemma 7 and Lemma 10 (both of which we prove in Section 5).

4 Decision Tree Lower Bound

In this section we prove Lemma 8, restated here for convenience.

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

Recall that $\mathsf{UP^{dt}}(\mathrm{TL}) \leq O(k \log k)$ by definition. To prove Lemma 8 we show that there is no $o(k^2)$-degree conical junta $h = \sum w_C C$ that outputs values in $[0.95, 1]$ on inputs from $\mathrm{TL}^{-1}(0)$ and outputs values in $[0, 0.05]$ on inputs from $\mathrm{TL}^{-1}(1)$. A similar lower bound for the plain $k \times k$ *Tribes* function was proved by [JK10, Theorem 4] using LP duality; our argument is more direct.

To illustrate the basic style of argument, we start gently by proving an $\Omega(n)$ conical junta degree bound for approximating the NAND function—this lower bound will be used in the proof of Lemma 8, too.

4.1 Warm-up: Lower bound for NAND

Suppose for contradiction that $h = \sum w_C C$ is a conical junta of degree $o(n)$ computing the n-bit NAND function to within error $1/5$. We will argue that if h is correct on inputs of Hamming weights n and $n-1$, then it must mess up on inputs of Hamming weight $n-2$: h will output a value larger than 1, which is a contradiction. We now give the details.

To begin, we have $h(\vec{1}) \leq 1/5$ by the correctness of h (here $\vec{1}$ is the all-1 input). This means that the total weight (sum of w_C's) associated with conjunctions that read only 1's (i.e., have only positive literals) is at most 1/5. Let $X \in \mathrm{NAND}^{-1}(1)$ be a uniformly random string of Hamming weight $n-1$. By correctness,

$$\mathbb{E}[h(X)] \;=\; \textstyle\sum w_C \mathbb{E}[C(X)] \;=\; \sum w_C \mathbb{P}[C(X)=1] \;\geq\; 4/5.$$

In the above sum, there are two types of conjunctions that contribute with a positive acceptance probability: those that read only 1's, and those that read a single 0 and some $o(n)$ many 1's. Since the first type has total weight $\leq 1/5$ we must have $\sum_{C\in\mathscr{C}} w_C \mathbb{P}[C(X)=1] \geq 3/5$ where $\mathscr{C}$ is the set of conjunctions of the second type. Consider the acceptance probability of any $C \in \mathscr{C}$ on a uniformly random string $Y \in \mathrm{NAND}^{-1}(1)$ of Hamming weight $n-2$: if the width of C is d, then $\mathbb{P}[C(Y)=1] = (n-d)/\binom{n}{2}$, which is $(2-o(1))/n$ for $d = o(n)$. Since $\mathbb{P}[C(X)=1] = 1/n$ we conclude that

$$\mathbb{P}[C(Y)=1] \;=\; (2-o(1)) \cdot \mathbb{P}[C(X)=1]. \tag{1}$$

We now arrive at the desired contradiction:

$$\mathbb{E}[h(Y)] \;\geq\; \textstyle\sum_{C\in\mathscr{C}} w_C \mathbb{P}[C(Y)=1] \;=\; (2-o(1)) \sum_{C\in\mathscr{C}} w_C \mathbb{P}[C(X)=1] \;\geq\; (2-o(1))\cdot 3/5 \;>\; 1.$$

4.2 Proof of Lemma 8

We prove a lower bound for $\mathrm{TL}\colon \Sigma^{k\times k} \to \{0,1\}$ by arguing that $\Omega(k^2)$ entries must be touched: We only charge one query for reading a whole matrix entry in $\Sigma = \{0,1\} \times ([k]^{k-1} \cup \{\perp\})$. That is, we assume each conjunction either reads nothing from an entry or reads one fully. The width of a conjunction is then understood as the number of entries it reads.

We study three types of *random* inputs to TL:

- $X \in \mathrm{TL}^{-1}(0)$ is defined so that the columns in X are independent, and in each column all entries are $(1,\perp)$ except we plant a single $(0,\perp)$ entry in a random row index. Hence there are altogether k many $(0,\perp)$ entries in X.
- $Y \in \mathrm{TL}^{-1}(0)$ is defined like X except we replace a random $(1,\perp)$ entry in X with a $(0,\perp)$ entry. Hence there are altogether $k+1$ many $(0,\perp)$ entries in Y, two of them sharing a column.
- $Z \in \mathrm{TL}^{-1}(1)$ is defined like X except we replace a random $(0,\perp)$ entry (k different choices) in X with a $(1,p)$ entry, where p is a list of pointers to all other positions of $(0,\perp)$ entries (making Z indeed a 1-input).

The crux of the argument is contained in the following claim.

Claim 11. *For every conjunction C of width $o(k^2)$, either $\mathbb{P}[C(Y)=1] \geq 1.4 \cdot \mathbb{P}[C(X)=1]$ or $\mathbb{P}[C(Z)=1] \geq 0.5 \cdot \mathbb{P}[C(X)=1]$.*

Before proving Claim 11, let us see how to finish the proof of Lemma 8 assuming it. We have a similar claim for conical juntas:

Claim 12. *For every conical junta h of degree $o(k^2)$, either $\mathbb{E}[h(Y)] \geq 1.1 \cdot \mathbb{E}[h(X)]$ or $\mathbb{E}[h(Z)] \geq 0.1 \cdot \mathbb{E}[h(X)]$.*

Proof. Let $h = \sum w_C C$. By linearity, $\mathbb{E}[h(X)] = \sum w_C \mathbb{P}[C(X) = 1]$ and similarly for Y and Z. By Claim 11, let $\mathscr{C}$ be a set of conjunctions such that for each $C \in \mathscr{C}$, $\mathbb{P}[C(Y) = 1] \geq 1.4 \cdot \mathbb{P}[C(X) = 1]$, and for each $C \notin \mathscr{C}$, $\mathbb{P}[C(Z) = 1] \geq 0.5 \cdot \mathbb{P}[C(X) = 1]$. Either $\sum_{C \in \mathscr{C}} w_C \mathbb{P}[C(X) = 1] \geq 0.8 \cdot \mathbb{E}[h(X)]$, in which case

$$\mathbb{E}[h(Y)] \;\geq\; \textstyle\sum_{C \in \mathscr{C}} w_C \mathbb{P}[C(Y) = 1] \;\geq\; \textstyle\sum_{C \in \mathscr{C}} w_C \cdot 1.4 \cdot \mathbb{P}[C(X) = 1] \;\geq\; 1.4 \cdot 0.8 \cdot \mathbb{E}[h(X)],$$

or $\sum_{C \notin \mathscr{C}} w_C \mathbb{P}[C(X) = 1] \geq 0.2 \cdot \mathbb{E}[h(X)]$, in which case

$$\mathbb{E}[h(Z)] \;\geq\; \textstyle\sum_{C \notin \mathscr{C}} w_C \mathbb{P}[C(Z) = 1] \;\geq\; \textstyle\sum_{C \notin \mathscr{C}} w_C \cdot 0.5 \cdot \mathbb{P}[C(X) = 1] \;\geq\; 0.5 \cdot 0.2 \cdot \mathbb{E}[h(X)]. \qquad \square$$

Now to prove Lemma 8, suppose for contradiction that h is a conical junta of degree $o(k^2)$ computing $\neg\mathrm{TL}$ to within error 0.05. That is, the value of h is in $[0.95, 1]$ on 0-inputs of TL and in $[0, 0.05]$ on 1-inputs of TL. In particular, $\mathbb{E}[h(X)] \in [0.95, 1]$, $\mathbb{E}[h(Y)] \in [0.95, 1]$, and $\mathbb{E}[h(Z)] \in [0, 0.05]$. This directly contradicts Claim 12.

Proof of Claim 11. We may assume that C accepts X with positive probability for otherwise the claim is trivial. Hence C reads at most a single $(0, \perp)$ entry from each column. We analyze two cases depending on how many $(0, \perp)$ entries C reads in total.

The first (easy) case is when C reads less than $k/2$ many $(0, \perp)$ entries. Here C cannot detect us replacing a random $(0, \perp)$ entry with a $(1, p)$ entry with probability better than $1/2$. That is, $\mathbb{P}[C(Z) = 1] \geq \mathbb{P}[C(Z) = 1 \,|\, C(X) = 1] \cdot \mathbb{P}[C(X) = 1] \geq 0.5 \cdot \mathbb{P}[C(X) = 1]$ where X, Z are jointly distributed as described above.

The second case is when C reads at least $k/2$ many $(0, \perp)$ entries. Because C has width $o(k^2)$ there is some $S_1 \subseteq [k]$ of size $|S_1| \geq (1 - o(1))k$ such that C reads $o(k)$ entries from each of the columns indexed by S_1. (More precisely, if C has width δk^2, then there is a set of $(1 - \sqrt{\delta})k$ columns from each of which C reads at most $\sqrt{\delta}k$ entries.) Let $S_2 \subseteq [k]$, $|S_2| \geq k/2$, be the set of columns where C reads a $(0, \perp)$. Let $i \in [k]$ denote the unique column where the jointly distributed X and Y differ. Note that i is a uniform random variable; for example, $\mathbb{P}[i \in S_1] = 1 - o(1)$. In what follows, we take $\approx$ to mean *up to a* $(1 \pm o(1))$ *factor*. We calculate:

$$\begin{aligned}
\mathbb{P}[C(Y) = 1] \;&\geq\; \mathbb{P}[C(Y) = 1 \text{ and } i \in S_1] \\
&\approx\; \mathbb{P}[C(Y) = 1 \,|\, i \in S_1] \\
&=\; \mathbb{P}[C(Y) = 1 \text{ and } i \in S_2 \,|\, i \in S_1] \;+\; \mathbb{P}[C(Y) = 1 \text{ and } i \notin S_2 \,|\, i \in S_1] \\
&=\; \lambda \cdot \underbrace{\mathbb{P}[C(Y) = 1 \,|\, i \in S_1 \cap S_2]}_{\textbf{(I)}} \;+\; (1 - \lambda) \cdot \underbrace{\mathbb{P}[C(Y) = 1 \,|\, i \in S_1 \smallsetminus S_2]}_{\textbf{(II)}},
\end{aligned}$$

where $\lambda := \mathbb{P}[i \in S_2 \,|\, i \in S_1] \geq 1/2 - o(1)$. In the first term, the condition $(i \in S_1 \cap S_2)$ means that C reads a single $(0, \perp)$ and $o(k)$ many $(1, \perp)$'s from the i-th column. Hence we are in a situation analogous to that in (1), and the same argument yields

$$\textbf{(I)} \;\geq\; (2 - o(1)) \cdot \mathbb{P}[C(X) = 1 \,|\, i \in S_1 \cap S_2] \;\approx\; 2 \cdot \mathbb{P}[C(X) = 1]$$

since X and i are independent. In the second term, the condition $(i \in S_1 \smallsetminus S_2)$ means that C reads $o(k)$ many $(1, \perp)$'s from the i-th column. Hence C cannot detect our planting of an additional $(0, \perp)$ entry in that column with probability better than $o(1)$:

$$\textbf{(II)} \;\geq\; (1 - o(1)) \cdot \mathbb{P}[C(X) = 1 \,|\, i \in S_1 \smallsetminus S_2] \;\approx\; \mathbb{P}[C(X) = 1]$$

since X and i are independent. In summary, we get that for some $\lambda \geq 1/2 - o(1)$,

$$\begin{aligned}\mathbb{P}[C(Y)=1] &\geq (2\lambda + (1-\lambda) - o(1)) \cdot \mathbb{P}[C(X)=1] \\ &\geq (3/2 - o(1)) \cdot \mathbb{P}[C(X)=1] \\ &\geq 1.4 \cdot \mathbb{P}[C(X)=1]. \end{aligned}$$

□

5 AND-Composition Lemmas

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

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

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

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

5.1 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 $f^{-1}(1)$. (We need to go through average-case complexity in order to "hide" from T^* which coordinate contains the true input to f.) 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 $j \in [k] \smallsetminus \{i\}$ 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; otherwise 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

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

Furthermore,

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

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

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

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 $\leq 1/4$.

5.2 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), $\mathbb{P}_{R,R_A,R_B}[\Pi(x, y) = F(x, y)] \geq 1 - \epsilon$. 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 $\mathrm{IC}_D(\Pi) \coloneqq \mathbb{I}(R, M\,; X \mid Y) + \mathbb{I}(R, M\,; Y \mid X) = \mathbb{I}(M\,; X \mid Y, R) + \mathbb{I}(M\,; Y \mid X, R)$ 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 [JK10], while $\mathsf{2WAPP}^{\mathsf{cc}*}$ directly expresses the relaxed partition bound of [KLL+15] and was the definition used in [GLM+16].

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).

5.3 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 $D^1 \coloneqq (D \mid F^{-1}(1))$. 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* $\mathsf{2WAPP}^{\mathsf{cc}*}_{\delta,D}(F) \le O(\mathrm{IC}_D(\Pi) + 1)$.

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

Lemma 15 is a standard application of the "direct sum" property of information cost; for completeness we sketch the argument in Appendix A. Lemma 16 is proved in Section 5.4 and relies on [BW15a]. Lemma 17 is due to [KLL+15, Theorem 1.1 of the ECCC version]. Lemma 18 follows from an argument in [KLL+15, Appendix A of the ECCC version] that uses LP duality; for completeness, in Appendix A we give a more intuitive version of the argument phrased in terms of the minimax theorem.

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 [Bra15, 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*

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

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 $\le 1/3$ 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 [ABB+16b].

5.4 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 [BW15a] 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.

Formal proof of Lemma 16. Assume w.l.o.g. that every execution of Π communicates exactly the same number of bits, and that Alice always sends a bit in odd rounds and Bob always sends a bit in even rounds (by inserting dummy coin flip rounds if necessary). As shown in [BW15a], we can also assume that Π is "smooth" (i.e., in every step, the bit to be communicated is 1 with probability between 1/3 and 2/3)—this is needed in order to apply Lemma 20 below.

Consider a probability space with random variables X, Y, R, R_A, R_B, M, F where $(X, Y) \sim D$ is the input, (R, R_A, R_B) is Π's randomness, $M \coloneqq M_1, \ldots, M_{\mathrm{CC}(\Pi)}$ is the sequence of bits communicated by Π, and $F \coloneqq F(X, Y)$ is the function value. For convenience of notation, if we condition on "x", this is shorthand for conditioning on "$X = x$". Letting $t \in \{1, \ldots, \mathrm{CC}(\Pi)\}$ and letting $\mathbb{D}$ denote KL-divergence (relative entropy), if we define

$$\begin{aligned} d_{x,y,r,m_{<t}} &\coloneqq \mathbb{D}\left(\frac{M_t \mid x, y, r, m_{<t}}{M_t \mid y, r, m_{<t}}\right) + \mathbb{D}\left(\frac{M_t \mid x, y, r, m_{<t}}{M_t \mid x, r, m_{<t}}\right), \\ c_{x,y,r,m} &\coloneqq \textstyle\sum_t d_{x,y,r,m_{<t}}, \\ c_{x,y} &\coloneqq \mathbb{E}[c_{X,Y,R,M} \mid x, y], \end{aligned}$$

then it can be seen [BW15a, Appendix C of the ECCC version] that

$$\mathrm{IC}_D(\Pi) \;=\; \mathbb{E}[c_{X,Y,R,M}] \;=\; \mathbb{E}[c_{X,Y}]. \tag{2}$$

Note that if t is odd the second term of $d_{x,y,r,m_{<t}}$ is 0, and if t is even the first term is 0; hence we think of $d_{x,y,r,m_{<t}}$ as defined by a single term (depending on who communicates in round t).

Although the following lemma was not explicitly stated in this way in [BW15a], it follows immediately from the corresponding part of the argument for the "conditional abort theorem" in that paper [BW15b].

Lemma 20 (Odometer). *For every smooth protocol Π, constant $\gamma > 0$, input distribution D, and $I > 0$, there is a protocol Π^* with $\mathrm{IC}_D(\Pi^*) \leq O\big(I + \log(\mathrm{CC}(\Pi) + 2)\big)$ that simulates Π in the following sense: Π^* uses the same randomness (R, R_A, R_B) as Π and some additional, independent randomness Q. Consider any fixed outcome x, y, r, r_A, r_B, and let m be Π's messages. Then* *(i) for every q, Π^* outputs either $\perp$ or the same bit that Π does, and*

(ii) if $c_{x,y,r,m} \le I$ then $\mathbb{P}_Q[\Pi^* \text{ outputs } \perp] \le \gamma$.

Define $\gamma := (\delta-\epsilon)/5$. To obtain Π' witnessing Lemma 16, we obtain Π^* from Lemma 20 with $I := (\mathrm{IC}_{D^1}(\Pi)/\gamma + 2\log(1/\gamma))/\gamma$ and replace the output $\perp$ with 0. Then we have $\mathrm{IC}_D(\Pi') = \mathrm{IC}_D(\Pi^*) \le O\big(\mathrm{IC}_{D^1}(\Pi) + \log(\mathrm{CC}(\Pi)+2)\big)$, so we just need to verify that Π' is (δ, D)-correct. In the following, we use Π, Π^*, Π' to denote random variables (jointly distributed with $X, Y, R, R_A, R_B, M, F, Q$) representing the outputs of the protocols.

Claim 21. $\mathbb{P}[c_{X,Y,R,M} > I \text{ and } F = 1] \le 4\gamma$.

Assuming Claim 21, we have

$$\begin{aligned}
\mathbb{P}[\Pi' \ne \Pi = F] &= \mathbb{P}[\Pi^* = \perp \text{ and } \Pi = F = 1] \\
&\le \mathbb{P}[\Pi^* = \perp \text{ and } F = 1] \\
&\le \mathbb{P}\big[c_{X,Y,R,M} > I \text{ and } F = 1\big] + \mathbb{P}\big[\Pi^* = \perp \,\big|\, c_{X,Y,R,M} \le I \text{ and } F = 1\big] \\
&\le 4\gamma + \gamma \\
&= 5\gamma
\end{aligned}$$

where the first line follows by construction of Π' and part (i) of Lemma 20, and the fourth line follows by Claim 21 and part (ii) of Lemma 20. Finally,

$$\mathbb{P}[\Pi' \ne F] \;\le\; \mathbb{P}[\Pi \ne F] + \mathbb{P}[\Pi' \ne \Pi = F] \;\le\; \epsilon + 5\gamma \;=\; \delta$$

since Π is (ϵ, D)-correct. This finishes the proof of Lemma 16.

To prove Claim 21, we first need to state another claim. Analogously to the notation leading up to (2), if for $(x,y) \in F^{-1}(1)$ we define

$$\begin{aligned}
d^1_{x,y,r,m_{<t}} &:= \mathbb{D}\left(\frac{M_t \,|\, x,y,r,m_{<t}}{M_t \,|\, y,r,m_{<t}, F=1}\right) + \mathbb{D}\left(\frac{M_t \,|\, x,y,r,m_{<t}}{M_t \,|\, x,r,m_{<t}, F=1}\right), \\
c^1_{x,y,r,m} &:= \textstyle\sum_t d^1_{x,y,r,m_{<t}}, \\
c^1_{x,y} &:= \mathbb{E}[c^1_{X,Y,R,M} \,|\, x,y],
\end{aligned}$$

then we have

$$\mathrm{IC}_{D^1}(\Pi) \;=\; \mathbb{E}[c^1_{X,Y,R,M} \,|\, F=1] \;=\; \mathbb{E}[c^1_{X,Y} \,|\, F=1]. \tag{3}$$

Claim 22. *For $(x,y) \in F^{-1}(1)$, we have $c_{x,y} - c^1_{x,y} \le \log\big(1/\mathbb{P}[F=1\,|\,y]\big) + \log\big(1/\mathbb{P}[F=1\,|\,x]\big)$.*

Proof of Claim 21. For any (x,y), by Markov's inequality we have

$$\mathbb{P}\big[c_{X,Y,R,M} > c_{X,Y}/\gamma \,\big|\, x,y\big] \;\le\; \gamma. \tag{4}$$

Say y is bad if $\mathbb{P}[F=1\,|\,y] \le \gamma$, and x is bad if $\mathbb{P}[F=1\,|\,x] \le \gamma$. By Claim 22 and a union bound,

$$\begin{aligned}
\mathbb{P}\big[c_{X,Y} > c^1_{X,Y} + 2\log(1/\gamma) \text{ and } F=1\big] &\le \mathbb{P}\big[(Y \text{ is bad or } X \text{ is bad}) \text{ and } F=1\big] \\
&\le \mathbb{P}[F=1\,|\,Y \text{ is bad}] + \mathbb{P}[F=1\,|\,X \text{ is bad}] \\
&\le 2\gamma.
\end{aligned} \tag{5}$$

By Markov's inequality and (3) we have

$$\mathbb{P}\big[c^1_{X,Y} > \mathrm{IC}_{D^1}(\Pi)/\gamma \text{ and } F=1\big] \;\le\; \mathbb{P}\big[c^1_{X,Y} > \mathrm{IC}_{D^1}(\Pi)/\gamma \,\big|\, F=1\big] \;\le\; \gamma. \tag{6}$$

Claim 21 follows by combining (4), (5), and (6) using a union bound. $\square$

Proof of Claim 22. Fix $(x, y) \in F^{-1}(1)$. Let $M_A \coloneqq M_1, M_3, \ldots$ be the bits sent by Alice, and let $M_B \coloneqq M_2, M_4, \ldots$ be the bits sent by Bob. Let $M_{A,<t} \coloneqq M_1, M_3, \ldots, M_k$ where k is the largest odd value $< t$, and let $M_{B,<t} \coloneqq M_2, M_4, \ldots, M_k$ where k is the largest even value $< t$.

For the moment, also consider any fixed r, r_B. Consider a separate probability space with random variables X^*, M^* distributed as $(X, M \mid y, r, r_B)$, and note that for even t, M^*_t is a deterministic function of $M^*_{A,<t}$. For the conditioning notation in the following, let $x^* \coloneqq x$. We have

$$\begin{aligned}
\textstyle\sum_{\text{odd } t} \mathbb{E}\big[d_{X,Y,R,M_{<t}} \,\big|\, x, y, r, r_B\big] \;&=\; \textstyle\sum_{\text{odd } t} \mathbb{E}_{M^*_{A,<t}}\left[\mathbb{D}\left(\frac{M^*_t \mid x^*, m^*_{A,<t}}{M^*_t \mid m^*_{A,<t}}\right) \,\middle|\, x^*\right] \\
&=\; \mathbb{D}\left(\frac{M^*_A \mid x^*}{M^*_A}\right) \\
&=\; \mathbb{D}\left(\frac{M_A \mid x, y, r, r_B}{M_A \mid y, r, r_B}\right)
\end{aligned}$$

where the middle equality is a direct application of the chain rule for $\mathbb{D}$. Similarly, for any fixed r, r_A, we have

$$\textstyle\sum_{\text{even } t} \mathbb{E}\big[d_{X,Y,R,M_{<t}} \,\big|\, x, y, r, r_A\big] \;=\; \mathbb{D}\left(\frac{M_B \mid x, y, r, r_A}{M_B \mid x, r, r_A}\right).$$

Then (no longer fixing any of r, r_A, r_B) we have

$$\begin{aligned}
c_{x,y} \;&=\; \mathbb{E}\big[\textstyle\sum_t d_{X,Y,R,M_{<t}} \,\big|\, x, y\big] \\
&=\; \mathbb{E}_{R,R_B}\Big[\textstyle\sum_{\text{odd } t} \mathbb{E}\big[d_{X,Y,R,M_{<t}} \,\big|\, x, y, r, r_B\big]\Big] + \mathbb{E}_{R,R_A}\Big[\textstyle\sum_{\text{even } t} \mathbb{E}\big[d_{X,Y,R,M_{<t}} \,\big|\, x, y, r, r_A\big]\Big] \\
&=\; \mathbb{E}_{R,R_B}\left[\mathbb{D}\left(\frac{M_A \mid x, y, r, r_B}{M_A \mid y, r, r_B}\right)\right] + \mathbb{E}_{R,R_A}\left[\mathbb{D}\left(\frac{M_B \mid x, y, r, r_A}{M_B \mid x, r, r_A}\right)\right] \qquad (7)
\end{aligned}$$

and similarly,

$$c^1_{x,y} \;=\; \mathbb{E}_{R,R_B}\left[\mathbb{D}\left(\frac{M_A \mid x, y, r, r_B}{M_A \mid y, r, r_B, F=1}\right)\right] + \mathbb{E}_{R,R_A}\left[\mathbb{D}\left(\frac{M_B \mid x, y, r, r_A}{M_B \mid x, r, r_A, F=1}\right)\right]. \qquad (8)$$

Note that

$$\begin{aligned}
&\mathbb{D}\left(\frac{M_A \mid x, y, r, r_B}{M_A \mid y, r, r_B}\right) - \mathbb{D}\left(\frac{M_A \mid x, y, r, r_B}{M_A \mid y, r, r_B, F=1}\right) \\
&=\; \textstyle\sum_{m_A} \mathbb{P}[m_A \mid x, y, r, r_B] \cdot \log\left(\frac{\mathbb{P}[m_A \mid y, r, r_B, F=1]}{\mathbb{P}[m_A \mid y, r, r_B]}\right) \\
&\leq\; \textstyle\sum_{m_A} \mathbb{P}[m_A \mid x, y, r, r_B] \cdot \log\big(1/\mathbb{P}[F=1 \mid y]\big) \\
&=\; \log\big(1/\mathbb{P}[F=1 \mid y]\big) \qquad (9)
\end{aligned}$$

and similarly,

$$\mathbb{D}\left(\frac{M_B \mid x, y, r, r_A}{M_B \mid x, r, r_A}\right) - \mathbb{D}\left(\frac{M_B \mid x, y, r, r_A}{M_B \mid x, r, r_A, F=1}\right) \;\leq\; \log\big(1/\mathbb{P}[F=1 \mid x]\big). \qquad (10)$$

Claim 22 follows by combining (7), (8), (9), and (10) using linearity of expectation. $\square$

A Appendix: Basic Lemmas

A.1 Proof of Lemma 15

Write the input to $\text{AND} \circ F^k$ as $\big((X_1, Y_1), \dots, (X_k, Y_k)\big) \sim D^k$. Let (R, R_A, R_B) be Π's randomness and M be Π's messages. It is known (see [BR14, Lemma 3.14 of the ECCC Revision #1 version] and [BM13, Fact 2.3 of the ECCC Revision #1 version]) that

$$\mathrm{CC}(\Pi) \;\geq\; \mathrm{IC}_{D^k}(\Pi) \;\geq\; \sum_{i=1}^{k} \mathbb{I}\big(R, M \,;\, X_i \,\big|\, X_{1,\dots,i-1}, Y_i, Y_{i+1,\dots,k}\big) + \mathbb{I}\big(R, M \,;\, Y_i \,\big|\, X_{1,\dots,i-1}, X_i, Y_{i+1,\dots,k}\big).$$

Therefore there exists i and $x_{1,\dots,i-1}, y_{i+1,\dots,k}$ such that

$$\mathrm{CC}(\Pi)/k \;\geq\; \mathbb{I}\big(R, M \,;\, X_i \,\big|\, x_{1,\dots,i-1}, Y_i, y_{i+1,\dots,k}\big) + \mathbb{I}\big(R, M \,;\, Y_i \,\big|\, x_{1,\dots,i-1}, X_i, y_{i+1,\dots,k}\big)$$

which is exactly $\mathrm{IC}_D(\Pi')$ where Π' is the following protocol with input denoted (X_i, Y_i):

1. Sample the same public randomness R as Π.
2. Alice privately samples R_A and $X_{i+1,\dots,k}$ according to D^{k-i} conditioned on $y_{i+1,\dots,k}$.
3. Bob privately samples R_B and $Y_{1,\dots,i-1}$ according to D^{i-1} conditioned on $x_{1,\dots,i-1}$.
4. Run Π on input $(x_{1,\dots,i-1}, X_i, X_{i+1,\dots,k})$, $(Y_{1,\dots,i-1}, Y_i, y_{i+1,\dots,k})$ with randomness (R, R_A, R_B).

Trivially, $\mathrm{CC}(\Pi') \leq \mathrm{CC}(\Pi)$. The ϵ-correctness of Π' follows from the ϵ-correctness of Π since with probability 1, $F(x_j, Y_j) = 1$ for $j < i$ and $F(X_j, y_j) = 1$ for $j > i$ and thus

$$(\text{AND} \circ F^k)\big((x_{1,\dots,i-1}, X_i, X_{i+1,\dots,k}),\, (Y_{1,\dots,i-1}, Y_i, y_{i+1,\dots,k})\big) \;=\; F(X_i, Y_i).$$

A.2 Proof of Lemma 18

Define α^* such that $\log(1/\alpha^*) = \max_D \mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon,D}(F)$. Consider the following two-player zero-sum game.

- Each pure row strategy is an input (x, y) to F.
- Each pure column strategy is a distribution μ over pairs (S, b), where S is a rectangle and $b \in \{0, 1, \perp\}$, such that $\mathbb{P}_{(S,b)\sim\mu}\big[(x, y) \in S \text{ and } b \neq \perp\big] \leq \alpha^*$ holds for each (x, y).
- The payoff to the column player is $P((x, y), \mu) \coloneqq \mathbb{P}_{(S,b)\sim\mu}\big[(x, y) \in S \text{ and } b = F(x, y)\big]$.

We claim that for every mixed row strategy D there exists a pure column strategy μ such that $\mathbb{E}_{(x,y)\sim D}[P((x, y), \mu)] \;\geq\; (1 - \epsilon)\alpha^*$. By assumption, there exists a $\mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon,D}$ protocol Π with communication cost c and associated α satisfying $c + \log(1/\alpha) \leq \log(1/\alpha^*)$. Assume Π only uses public randomness (by making any private randomness public). Consider the distribution μ over pairs (S, b) sampled as follows:

- with probability $1 - \alpha^* \cdot 2^c/\alpha$, let S be arbitrary and $b = \perp$;
- otherwise, sample the randomness of Π and a uniformly random transcript (of which we may assume there are exactly 2^c many) from the induced deterministic protocol, and let (S, b) be the rectangle and output of that transcript.

Then for each (x, y),

$$\begin{aligned}\mathbb{P}_{(S,b)\sim\mu}\big[(x,y)\in S \text{ and } b\neq\bot\big] &= (\alpha^*\cdot 2^c/\alpha)\cdot\mathbb{P}_{\Pi\text{'s randomness}}[\Pi(x,y)\neq\bot]\cdot\\ &\qquad \mathbb{P}_{\text{uniform transcript}}[\Pi(x,y) \text{ has that transcript}]\\ &\leq (\alpha^*\cdot 2^c/\alpha)\cdot\alpha\cdot(1/2^c)\\ &= \alpha^*\end{aligned}$$

so μ is a valid pure column strategy. Similarly, for each (x, y) we have $P((x, y), \mu) = (\alpha^*/\alpha)\cdot\mathbb{P}_{\Pi\text{'s randomness}}[\Pi(x, y) = F(x, y)]$, and thus

$$\mathbb{E}_{(x,y)\sim D}[P((x,y),\mu)] \;=\; (\alpha^*/\alpha)\cdot\mathbb{P}_{(x,y)\sim D,\,\Pi\text{'s randomness}}[\Pi(x,y)=F(x,y)] \;\geq\; (1-\epsilon)\alpha^*.$$

Since the set of all pure column strategies μ forms a polytope, and since $P((x, y), \mu)$ is an affine function of μ for each (x, y), we may consider w.l.o.g. only the finitely-many pure column strategies that are vertices of the polytope. Thus we may employ the minimax theorem to find a mixed column strategy ν such that for every pure row strategy (x, y) we have $\mathbb{E}_{\mu\sim\nu}[P((x, y), \mu)] \geq (1-\epsilon)\alpha^*$. Consider a protocol Π that publicly samples $\mu\sim\nu$ and $(S, b)\sim\mu$, then checks whether $(x, y)\in S$ (with 2 bits of communication) and outputs b if so and $\bot$ if not. Then for each (x, y),

- $\mathbb{P}[\Pi(x,y)\neq\bot] = \mathbb{E}_{\mu\sim\nu}\big[\mathbb{P}_{(S,b)\sim\mu}\big[(x,y)\in S \text{ and } b\neq\bot\big]\big] \leq \mathbb{E}_{\mu\sim\nu}[\alpha^*] = \alpha^*$ by the definition of pure column strategies, and
- $\mathbb{P}[\Pi(x,y) = F(x,y)] = \mathbb{E}_{\mu\sim\nu}\big[\mathbb{P}_{(S,b)\sim\mu}\big[(x,y)\in S \text{ and } b = F(x,y)\big]\big] = \mathbb{E}_{\mu\sim\nu}[P((x,y),\mu)] \geq (1-\epsilon)\alpha^*$.

Thus Π witnesses that $\mathsf{2WAPP}^{\mathsf{cc}*}_{\epsilon}(F) \leq 2 + \log(1/\alpha^*)$.

Acknowledgments

We thank Mark Braverman, Troy Lee, and Omri Weinstein for discussions. M.G. is partially supported by the Simons Award for Graduate Students in TCS. This work is also supported by NSF grant CCF-1657377.

References

[ABB+16a] 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.

[ABB+16b] 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.

[ABK16] 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.

[AKK16] 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.

[Ass17] Sepehr Assadi. Tight space-approximation tradeoff for the multi-pass streaming set cover problem. In *Proceedings of the 36th Symposium on Principles of Database Systems (PODS)*, pages 321–335. ACM, 2017. doi:10.1145/3034786.3056116.

[AUY83] 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.

[Bel06] 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.

[BGM06] 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.

[BJKS04] 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.

[BM13] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proceedings of the 45th Symposium on Theory of Computing (STOC)*, pages 161–170. ACM, 2013. doi:10.1145/2488608.2488629.

[BR14] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014. doi:10.1109/TIT.2014.2347282.

[Bra15] Mark Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015. doi:10.1137/130938517.

[BW15a] 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.

[BW15b] Mark Braverman and Omri Weinstein. Personal communication, 2015.

[FMP$^+$15] 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.

[GL14] 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.

[GLM+16] 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.

[Göö15] 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.

[GPW15] 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.

[GPW17] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. In *Proceedings of the 58th Symposium on Foundations of Computer Science (FOCS)*, pages 132–143. IEEE, 2017. doi:10.1109/FOCS.2017.2.

[JK10] 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.

[JKR09] 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.

[JKS10] 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.

[JLV14] 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.

[Juk12] Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*, volume 27 of *Algorithms and Combinatorics*. Springer, 2012.

[KLdW15] 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.

[KLL+15] 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.

[KMSY14] 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.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[KRS15] 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.

[Kus94] Eyal Kushilevitz. Unpublished. Cited in [NW95], 1994.

[LS88] 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.

[LS07] 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.

[LS10] 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.

[MS15] Sagnik Mukhopadhyay and Swagato Sanyal. Towards better separation between deterministic and randomized query complexity. In *Proceedings of the 35th Conference on Foundation of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 206–220. Schloss Dagstuhl, 2015. doi:10.4230/LIPIcs.FSTTCS.2015.206.

[NW95] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995. doi:10.1007/BF01192527.

[Sav02] 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/.

[Wat17] Thomas Watson. Communication complexity of statistical distance. In *Proceedings of the 21st International Workshop on Randomization and Computation (RANDOM)*, pages 49:1–49:10. Schloss Dagstuhl, 2017. doi:10.4230/LIPIcs.APPROX-RANDOM.2017.49.

[Yan91] 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.