Query-to-Communication Lifting for P^{NP} *†

Mika Göös[1], Pritish Kamath[2], Toniann Pitassi[3], and Thomas Watson[4]

1 **Harvard University, Cambridge, MA, USA**
`mika@seas.harvard.edu`
2 **Massachusetts Institute of Technology, Cambridge, MA, USA**
`pritish@mit.edu`
3 **University of Toronto, Toronto, ON, Canada**
`toni@cs.toronto.edu`
4 **University of Memphis, Memphis, TN, USA**
`Thomas.Watson@memphis.edu`

Abstract

We prove that the P^{NP}-type query complexity (alternatively, decision list width) of any boolean function f is quadratically related to the P^{NP}-type communication complexity of a lifted version of f. As an application, we show that a certain "product" lower bound method of Impagliazzo and Williams (CCC 2010) fails to capture P^{NP} communication complexity up to polynomial factors, which answers a question of Papakonstantinou, Scheder, and Song (CCC 2014).

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases Communication Complexity, Query Complexity, Lifting Theorem, P^{NP}

Digital Object Identifier 10.4230/LIPIcs.CCC.2017.12

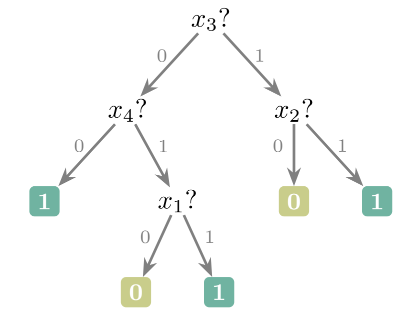
1 Introduction

Broadly speaking, a **query-to-communication lifting theorem** (a.k.a. communication-to-query simulation theorem) translates, in a black-box fashion, lower bounds on some type of *query complexity* (a.k.a. decision tree complexity) [38, 6, 19] of a boolean function $f\colon \{0,1\}^n \to \{0,1\}$ into lower bounds on a corresponding type of *communication complexity* [23, 19, 27] of a two-party version of f. Table 1 lists several known results in this vein.

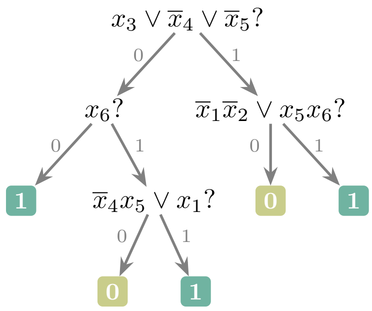
In this work, we provide a lifting theorem for P^{NP}-type query/communication complexity.

P^{NP} decision trees. Recall that a deterministic (i.e., P-type) decision tree computes an n-bit boolean function f by repeatedly querying, at unit cost, individual bits $x_i \in \{0,1\}$ of the input x until the value $f(x)$ is output at a leaf of the tree. A P^{NP} decision tree is more powerful: in each step, it can query/evaluate a width-k DNF of its choice, at the cost of k. Here k is simply the nondeterministic (i.e., NP-type) decision tree complexity of the predicate being evaluated at a node. The overall cost of a P^{NP} decision tree is the maximum over all inputs x of the sum of the costs of the individual queries that are made on input x. The P^{NP} query complexity of f, denoted $\mathsf{P}^{\mathsf{NPdt}}(f)$, is the least cost of a P^{NP} decision tree that computes f.

* A full version of the paper is available at `https://eccc.weizmann.ac.il/report/2017/024/`.
† P. Kamath was funded in parts by NSF grants CCF-1420956, CCF-1420692, CCF-1218547 and CCF-1650733. T. Watson was funded by NSF grant CCF-1657377.



Deterministic decision tree of cost 3



P^{NP} *decision tree of cost 4*

▶ **Example 1.** Consider the fabled *odd-max-bit* function [3, 7, 33, 36, 8] defined by $\textsc{Omb}(x) :=$ 1 iff $x \neq 0^n$ and the largest index $i \in [n]$ such that $x_i = 1$ is odd. This function admits an efficient $O(\log n)$-cost P^{NP} decision tree: we can *find* the largest i with $x_i = 1$ by using a binary search that queries 1-DNFs of the form $\bigvee_{a \leq j \leq n} x_j$ for different $a \in [n]$.

P^{NP} communication protocols. Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a two-party function, i.e., Alice holds $x \in \mathcal{X}$, Bob holds $y \in \mathcal{Y}$. A deterministic communication protocol can be viewed as a decision tree where in each step, at unit cost, it evaluates either an arbitrary predicate of Alice's input x or an arbitrary predicate of Bob's input y. A P^{NP} communication protocol [2, 15] is more powerful: in each step, it can evaluate an arbitrary predicate of the form $(x, y) \in \bigcup_{i \in [2^k]} R_i$ ("oracle query") at the cost of k (we always assume $k \geq 1$). Here each R_i is a rectangle (i.e., $R_i = X_i \times Y_i$ for some $X_i \subseteq \mathcal{X}$, $Y_i \subseteq \mathcal{Y}$) and k is just the usual nondeterministic communication complexity of the predicate being evaluated. The overall cost of a P^{NP} protocol is the maximum over all inputs (x, y) of the sum of the costs of the individual oracle queries that are made on input (x, y). The P^{NP} communication complexity of F, denoted $\mathsf{P}^{\mathsf{NPcc}}(F)$, is the least cost of a P^{NP} protocol computing F.

Note that if $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ can be written as a k-DNF on $2n$ variables, then the nondeterministic communication complexity of F, denoted $\mathsf{NP}^{\mathsf{cc}}(F)$, is at most $O(k \log n)$ bits: we can guess one of the $\leq 2^k\binom{n}{k}$ many terms in the k-DNF and verify that the term evaluates to true. Consequently, any P^{NP} decision tree for a function f can be simulated efficiently by a P^{NP} protocol, regardless of how the input bits of f are split between Alice and Bob. That is, letting F be f equipped with any bipartition of the input bits, we have

$$\mathsf{P}^{\mathsf{NPcc}}(F) \;\leq\; \mathsf{P}^{\mathsf{NPdt}}(f) \cdot O(\log n). \tag{1}$$

1.1 Main result

Our main result establishes a rough converse to inequality (1) for a special class of *composed*, or *lifted*, functions. For an n-bit function f and a two-party function $g\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ (called a *gadget*), their composition $F := f \circ g^n\colon \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ is given by $F(x, y) :=$ $f(g(x_1, y_1), \ldots, g(x_n, y_n))$. We use as a gadget the popular *index* function $\textsc{Ind}_m\colon [m] \times \{0,1\}^m$ defined by $\textsc{Ind}_m(x, y) := y_x$.

▶ **Theorem 2** (Lifting for P^{NP})**.** *Let $m = m(n) := n^4$. For every $f\colon \{0,1\}^n \to \{0,1\}$,*

$$\mathsf{P}^{\mathsf{NPcc}}(f \circ \textsc{Ind}_m^n) \;\geq\; \sqrt{\mathsf{P}^{\mathsf{NPdt}}(f) \cdot \Omega(\log n)}.$$

Table 1 Some query-to-communication lifting theorems. The first four are formulated in the language of boolean functions (as in this paper); the last two are formulated in the language of combinatorial optimization.

Query model	Communication model	References
deterministic	deterministic	[28, 14, 10, 17]
nondeterministic	nondeterministic	[13, 11]
polynomial degree	rank	[35, 34, 29, 31]
conical junta degree	nonnegative rank	[13, 22]
Sherali–Adams	LP extension complexity	[9, 22]
sum-of-squares	SDP extension complexity	[24]

The lower bound is tight up to the square root, since (1) can be adapted for composed functions to yield $\mathsf{P}^{\mathsf{NPcc}}(f \circ \textsc{Ind}_m^n) \le \mathsf{P}^{\mathsf{NPdt}}(f) \cdot O(\log m + \log n)$. The reason we incur a quadratic loss is because we actually prove a *lossless* lifting theorem for a related complexity measure that is known to capture P^{NP} query/communication complexity up to a quadratic factor, namely *decision lists*, discussed shortly in subsection 1.3.

1.2 Application

Impagliazzo and Williams [18] gave the following criteria – we call it the *product method* – for a function F to have large P^{NP} communication complexity. Here, a *product* distribution μ over $\mathcal{X} \times \mathcal{Y}$ is such that $\mu(x, y) = \mu_{\mathcal{X}}(x) \cdot \mu_{\mathcal{Y}}(y)$ for some distributions $\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}$. A rectangle $R \subseteq \mathcal{X} \times \mathcal{Y}$ is *monochromatic* (relative to F) if F is constant on R.

Product method [18]: *Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and suppose μ is a product distribution over $\mathcal{X} \times \mathcal{Y}$ such that $\mu(R) \le \delta$ for every monochromatic rectangle R. Then*

$$\mathsf{P}^{\mathsf{NPcc}}(F) \;\ge\; \Omega(\log(1/\delta)).$$

This should be compared with the well-known *rectangle size method* [20], [23, §2.4] (μ over $F^{-1}(1)$ such that $\mu(R) \le \delta$ for all monochromatic R implies $\mathsf{NP}^{\mathsf{cc}}(F) \ge \Omega(\log(1/\delta))$), which is known to characterize nondeterministic communication complexity up to an additive $\Theta(\log n)$ term.

Papakonstantinou, Scheder, and Song [25, Open Problem 1] asked whether the product method can yield a tight P^{NP} communication lower bound for every function. This is especially relevant in light of the fact that all existing lower bounds against $\mathsf{P}^{\mathsf{NPcc}}$ (proved in [18, 25]) have used the product method (except those lower bounds that hold against an even stronger model: unbounded error randomized communication complexity, $\mathsf{UPP}^{\mathsf{cc}}$ [26]). We show that the product method can fail exponentially badly, even for total functions.

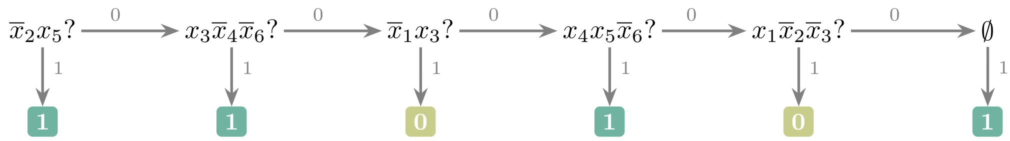
▶ **Theorem 3.** *There exists a total $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ satisfying the following.*

- *F has large P^{NP} communication complexity: $\mathsf{P}^{\mathsf{NPcc}}(F) \ge n^{\Omega(1)}$.*
- *For any product distribution μ over $\{0,1\}^n \times \{0,1\}^n$, there exists a monochromatic rectangle R that is large: $\log(1/\mu(R)) \le \log^{O(1)} n$.*

1.3 Decision lists (DLs)

Conjunction DLs. The following definition is due to Rivest [30]: a *conjunction decision list* of width k is a sequence $(C_1, \ell_1), \ldots, (C_L, \ell_L)$ where each C_i is a conjunction of $\le k$ literals and $\ell_i \in \{0,1\}$ is a label. We assume for convenience that C_L is the empty conjunction (accepting every input). Given an input x, the conjunction decision list finds the least $i \in [L]$ such that $C_i(x) = 1$ and outputs ℓ_i. We define the conjunction decision list width of f, denoted $\mathsf{DL}^{\mathsf{dt}}(f)$, as the minimum k such that f can be computed by a width-k conjunction decision list. For example, $\mathsf{DL}^{\mathsf{dt}}(\textsc{Omb}) = 1$. This complexity measure is quadratically related to P^{NP} query complexity (for details, see full version of this paper [12]).

▶ **Fact 4.** *For all $f\colon \{0,1\}^n \to \{0,1\}$, $\Omega(\mathsf{DL}^{\mathsf{dt}}(f)) \le \mathsf{P}^{\mathsf{NPdt}}(f) \le O(\mathsf{DL}^{\mathsf{dt}}(f)^2 \cdot \log n)$.*



A conjunction decision list of width 3

Rectangle DLs. A communication complexity variant of decision lists was introduced by Papakonstantinou, Scheder, and Song [25] (they called them *rectangle overlays*). A *rectangle decision list* of cost k is a sequence $(R_1, \ell_1), \ldots, (R_{2^k}, \ell_{2^k})$ where each R_i is a rectangle and $\ell_i \in \{0,1\}$ is a label. We assume for convenience that R_{2^k} contains every input. Given an input (x,y), the rectangle decision list finds the least $i \in [2^k]$ such that $(x,y) \in R_i$ and outputs ℓ_i. We define the rectangle decision list complexity of F, denoted $\mathsf{DL}^{\mathsf{cc}}(F)$, as the minimum k such that F can be computed by a cost-k rectangle decision list. We again have a quadratic relationship [25, Theorem 3] (for details, see full version of this paper [12]).

▶ **Fact 5.** *For all $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, $\Omega(\mathsf{DL}^{\mathsf{cc}}(F)) \le \mathsf{P}^{\mathsf{NPcc}}(F) \le O(\mathsf{DL}^{\mathsf{cc}}(F)^2)$.*

DLs are combinatorially slightly more comfortable to work with than P^{NP} decision trees/protocols. This is why our main lifting theorem (Theorem 2) is in fact derived as a corollary of a *lossless* lifting theorem for DLs.

▶ **Theorem 6** (Lifting for DL). *Let $m = m(n) := n^4$. For every $f\colon \{0,1\}^n \to \{0,1\}$,*

$$\mathsf{DL}^{\mathsf{cc}}(f \circ \textsc{Ind}_m^n) \;=\; \mathsf{DL}^{\mathsf{dt}}(f) \cdot \Theta(\log n).$$

Indeed, Theorem 2 follows because $\mathsf{P}^{\mathsf{NPcc}}(f \circ \textsc{Ind}_m^n) \ge \Omega(\mathsf{DL}^{\mathsf{cc}}(f \circ \textsc{Ind}_m^n)) \ge \Omega(\mathsf{DL}^{\mathsf{dt}}(f) \cdot \log n) \ge \Omega((\mathsf{P}^{\mathsf{NPdt}}(f)/\log n)^{1/2} \cdot \log n) = (\mathsf{P}^{\mathsf{NPdt}}(f) \cdot \Omega(\log n))^{1/2}$, where the first inequality is by Fact 5, the second is by Theorem 6, and the third is by Fact 4. We mention that Theorems 2 and 6, as well as Facts 4 and 5, in fact hold for all partial functions.

As a curious aside, we mention that a time-bounded analogue of decision lists (capturing P^{NP}) has also been studied in a work of Williams [39].

1.4 Separation between P^{NP} and DL

Facts 4 and 5 show that decision lists can be converted to P^{NP} decision trees/protocols with a quadratic overhead. Is this conversion optimal? In other words, are there functions that witness a quadratic gap between P^{NP} and DL? We at least show that *if a lossless lifting theorem holds for* P^{NP}, then such a quadratic gap indeed exists for communication complexity.

▶ **Conjecture 7.** *There is an $m = m(n) := n^{\Theta(1)}$ such that for every $f\colon \{0,1\}^n \to \{0,1\}$,*

$$\mathsf{P}^{\mathsf{NPcc}}(f \circ \textsc{Ind}_m^n) \;=\; \mathsf{P}^{\mathsf{NPdt}}(f) \cdot \Theta(\log n).$$

Our bonus contribution here (proof deferred to the full version [12]) shows that the simple $O(\log n)$-cost P^{NP} decision tree for the odd-max-bit function is optimal:

▶ **Theorem 8.** $\mathsf{P}^{\mathsf{NPdt}}(\textsc{Omb}) \geq \Omega(\log n)$.

▶ **Corollary 9.** *The second inequality of Fact 4 is tight (i.e., $\mathsf{P}^{\mathsf{NPdt}}(f) \geq \Omega(\mathsf{DL}^{\mathsf{dt}}(f)^2 \cdot \log n)$ for some f), and assuming Conjecture 7, the second inequality of Fact 5 is tight (i.e., $\mathsf{P}^{\mathsf{NPcc}}(F) \geq \Omega(\mathsf{DL}^{\mathsf{cc}}(F)^2)$ for some F).*

This corollary is witnessed by $f := \textsc{Omb}$ (which has $\mathsf{DL}^{\mathsf{dt}}(f) \leq O(1)$ and $\mathsf{P}^{\mathsf{NPdt}}(f) \geq \Omega(\log n)$) and its lifted version $F := \textsc{Omb} \circ \textsc{Ind}_m^n$ (which has $\mathsf{DL}^{\mathsf{cc}}(F) \leq O(\log n)$ and $\mathsf{P}^{\mathsf{NPcc}}(F) \geq \Omega(\log^2 n)$ under Conjecture 7). One caveat is that we have only shown the corollary for an extreme setting of parameters (constant $\mathsf{DL}^{\mathsf{dt}}(f)$ and logarithmic $\mathsf{DL}^{\mathsf{cc}}(F)$). It would be interesting to show a separation for functions of $n^{\Omega(1)}$ decision list complexity.

2 Preliminaries: Decision List Lower Bound Techniques

We present two basic lemmas in this section that allow one to prove lower bounds on conjunction/rectangle decision lists. First we recall the proof of the product method, which will be important for us, as we will extend the proof technique in both section 3 and section 4.

▶ **Lemma 10** (Product method for $\mathsf{DL}^{\mathsf{cc}}$)**.** *Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and suppose μ is a product distribution over $\mathcal{X} \times \mathcal{Y}$. Then F admits a monochromatic rectangle R with $\log(1/\mu(R)) \leq O(\mathsf{DL}^{\mathsf{cc}}(F))$.*

Proof (from [18, 25]). Let $(R_1, \ell_1), \ldots, (R_{2^k}, \ell_{2^k})$ be an optimal rectangle decision list of cost $k := \mathsf{DL}^{\mathsf{cc}}(F)$ computing F. Recall we assume that $R_{2^k} = \mathcal{X} \times \mathcal{Y}$ contains every input. We find a monochromatic R with $\mu(R) \geq 2^{-2k}$ via the following process.

We initialize $X := \mathcal{X}$ and $Y := \mathcal{Y}$ and iterate the following for $i = 1, \ldots, 2^k$ rounds, shrinking the rectangle $X \times Y$ in each round.

(†) *Round i:* (loop invariant: $R_i \cap X \times Y$ is a monochromatic rectangle)

Write $R_i \cap X \times Y = X_i \times Y_i$ and test whether $\mu(X_i \times Y_i) = \mu_{\mathcal{X}}(X_i) \cdot \mu_{\mathcal{Y}}(Y_i)$ is at least 2^{-2k}. Suppose not, as otherwise we are successful. Then either $\mu_{\mathcal{X}}(X_i) < 2^{-k}$ or $\mu_{\mathcal{Y}}(Y_i) < 2^{-k}$; say the former. We now "delete" the rows X_i from consideration by updating $X \leftarrow X \smallsetminus X_i$.

Note that since $R_i \cap X \times Y$ is removed from $X \times Y$ in each unsuccessful round, it must hold (inductively) that $\bigcup_{j<i} R_j$ is disjoint from $X \times Y$ at the start of the i-th round, and so $R_i \cap X \times Y$ is indeed monochromatic (since it only contains points for which R_i is the first rectangle in the decision list to contain them, which means F evaluates to ℓ_i). The process starts out with $\mu(X \times Y) = 1$ and in each unsuccessful round the quantity $\mu(X \times Y)$ decreases by $< 2^{-k}$. Some round must succeed, as otherwise the process would finish with $X \times Y = \emptyset$ and hence $\mu(X \times Y) = 0$ in 2^k rounds, which is impossible. ◀

Recall that our Theorem 3 states that the product method is not complete for the measure $\mathsf{DL}^{\mathsf{cc}}$. By contrast, we are able to give an alternative characterization for the analogous query complexity measure $\mathsf{DL}^{\mathsf{dt}}$. We do not know if this characterization has been observed in the literature before.

▶ **Lemma 11** (Characterization for $\mathsf{DL}^{\mathsf{dt}}$)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$. Then $\mathsf{DL}^{\mathsf{dt}}(f) \leq k$ iff for every nonempty $Z \subseteq \{0,1\}^n$ there exists an $\ell \in \{0,1\}$ and a width-k conjunction that accepts an input in $Z_\ell := Z \cap f^{-1}(\ell)$ but none in $Z_{1-\ell}$.*

Proof. Suppose f has a width-k conjunction decision list $(C_1, \ell_1), (C_2, \ell_2), \ldots, (C_L, \ell_L)$. The first C_i that accepts an input in Z (such an i must exist since the last C_L accepts every input) must accept an input in Z_{ℓ_i} but none in $Z_{1-\ell_i}$ (since all inputs in $C_i^{-1}(1) \cap Z$ are such that C_i is the first conjunction in the decision list to accept them).

Conversely, assume the right side of the "iff" holds. Then we can build a conjunction decision list for f iteratively as follows. Start with $Z = \{0,1\}^n$. Let C_1 be a width-k conjunction that accepts an input in some Z_{ℓ_1} but none in $Z_{1-\ell_1}$, and remove from Z all inputs accepted by C_1. Then continue with the new Z: let C_2 be a width-k conjunction that accepts an input in some Z_{ℓ_2} but none in $Z_{1-\ell_2}$, and further remove from Z all inputs accepted by C_2. Once Z becomes empty (this must happen since the right side of the iff holds for all nonempty Z), we have constructed a conjunction decision list $(C_1, \ell_1), (C_2, \ell_2), \ldots$ for f. ◀

3 Proof of the Lifting Theorem

In this section we prove Theorem 6, restated here for convenience.

▶ **Theorem 6** (Lifting for DL). *Let $m = m(n) := n^4$. For every $f \colon \{0,1\}^n \to \{0,1\}$,*

$$\mathsf{DL}^{\mathsf{cc}}(f \circ \textsc{Ind}_m^n) \;=\; \mathsf{DL}^{\mathsf{dt}}(f) \cdot \Theta(\log n).$$

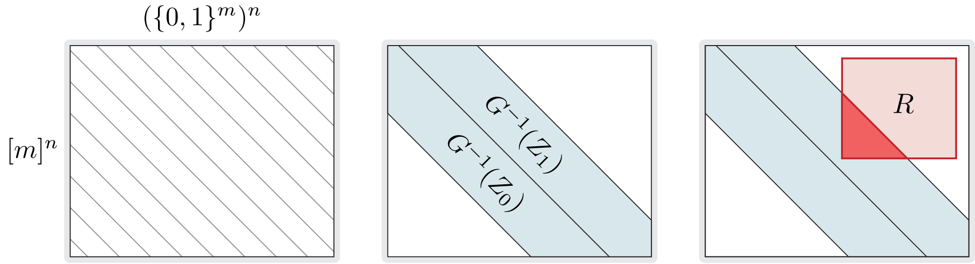
We use the abbreviations $g := \textsc{Ind}_m \colon [m] \times \{0,1\}^m \to \{0,1\}$ and $F := f \circ g^n$.

The upper bound of Theorem 6 is straightforward: given a width-k conjunction decision list for f (which necessarily has length $\leq 2^k \binom{n}{k} \leq n^{O(k)}$), we can form a rectangle decision list for F by transforming each labeled conjunction into a set of same-labeled rectangles (which can be ordered arbitrarily among themselves), one for each of the m^k ways of choosing a row from each of the copies of g corresponding to bits read by the conjunction – for a total of $n^{O(k)} \cdot m^k \leq n^{O(k)}$ rectangles and hence a cost of $k \cdot O(\log n)$. For example, if $k = 2$ and the conjunction is $z_1 \overline{z}_2$, then for each $x_1, x_2 \in [m]$ there would be a rectangle consisting of all inputs with that value of x_1, x_2 and with y_1, y_2 such that $g(x_1, y_1) = 1$ and $g(x_2, y_2) = 0$. For the rest of this section, we prove the matching lower bound.

3.1 Overview

Fix an optimal rectangle decision list $(R_1, \ell_1), \ldots, (R_{2^k}, \ell_{2^k})$ for F. By our characterization of $\mathsf{DL}^{\mathsf{dt}}$ (Theorem 11) it suffices to show that for every nonempty $Z \subseteq \{0,1\}^n$ there is a width-$O(k/\log n)$ conjunction that accepts an input in $Z_\ell := Z \cap f^{-1}(\ell)$ for some $\ell \in \{0,1\}$, but none in $Z_{1-\ell}$. Thus fix some nonempty Z henceforth.

Write $G := g^n$ for short. We view the communication matrix of F as being partitioned into *slices* $G^{-1}(z) = \{(x,y) : G(x,y) = z\}$, one for each $z \in \{0,1\}^n$; see (a) below. We focus naturally on the slices corresponding to Z, namely $G^{-1}(Z) = \bigcup_{z \in Z} G^{-1}(z)$, which is further partitioned into $G^{-1}(Z_0)$ and $G^{-1}(Z_1)$; see (b) below. Our goal is to find a rectangle R that touches $G^{-1}(Z_\ell)$ (for some ℓ) but not $G^{-1}(Z_{1-\ell})$, and such that $G(R) = C^{-1}(1)$ for a width-$O(k/\log n)$ conjunction C; see (c) below. Thus $C^{-1}(1)$ touches Z_ℓ but not $Z_{1-\ell}$, as desired.



(a) (b) (c)

We find such an R as follows. We maintain a rectangle $X \times Y$, which is initially the whole domain of F and which we iteratively shrink. In each round, we consider the next rectangle R_i in the decision list, and one of two things happens. Either:

- The round is declared unsuccessful, in which case we remove from $X \times Y$ a small number of rows and columns that together cover all of $R_i \cap X \times Y \cap G^{-1}(Z)$. This guarantees that throughout the whole execution, by the i-th round, $\bigcup_{j<i}(R_j \cap G^{-1}(Z))$ has been removed from $X \times Y$ – thus every input in $R_i \cap X \times Y \cap G^{-1}(Z)$ is such that R_i is the first rectangle in the decision list that contains it, so it is in $G^{-1}(Z_{\ell_i}) \subseteq F^{-1}(\ell_i)$ by the definition of decision lists.

Or,

- Success is declared, in which case it will hold that $R_i \cap X \times Y$ touches $G^{-1}(Z)$ – in fact, it touches $G^{-1}(Z_{\ell_i})$ but not $G^{-1}(Z_{1-\ell_i})$, by the above – and we can restrict $R_i \cap X \times Y$ to a subrectangle R that still touches $G^{-1}(Z_{\ell_i})$ but is such that $G(R)$ is fixed on $O(k/\log n)$ coordinates and has full support on the remaining coordinates. In other words, $G(R) = C^{-1}(1)$ for a width-$O(k/\log n)$ conjunction C.

This process is a variation of the process (†) from the product method (Theorem 10). The difference is that the Z-slices, $G^{-1}(Z)$, now play the role of the product distribution, and we maintain the monochromatic property for $R_i \cap X \times Y$ only inside the Z-slices. Another difference is that in each unsuccessful round we remove *both* rows *and* columns from $X \times Y$ (not *either–or* as in (†)).

To flesh out this outline, we need to specify how to determine whether a round is successful, which rows and columns to remove if not, and how to restrict to the desired R if so, and we need to argue that the process will terminate with success.

3.2 Tools

We will need to find a rectangle R such that $G(R)$ is fixed on few coordinates and has full support on the remaining coordinates. We now describe some tools that help us achieve this. First of all, under what conditions on $R = A \times B$ can we guarantee that $G(R)$ has full support over all n coordinates?

▶ **Definition 12** (Blockwise-density [13]). $A \subseteq [m]^n$ is called *δ-dense* if the uniform random variable $\boldsymbol{x}$ over A satisfies the following: for every nonempty $I \subseteq [n]$, the blocks $\boldsymbol{x}_I$ have min-entropy rate at least δ, that is, $\mathbf{H}_\infty(\boldsymbol{x}_I) \geq \delta \cdot |I| \log m$. Here, $\boldsymbol{x}_I$ is marginally distributed over $[m]^I$, and $\mathbf{H}_\infty(\boldsymbol{x}) := \min_x \log(1/\mathbf{Pr}[\boldsymbol{x} = x])$ is the usual min-entropy of a random variable (see, e.g., Vadhan's monograph [37] for an introduction).

▶ **Definition 13** (Deficiency). For $B \subseteq (\{0,1\}^m)^n$, we define $\mathbf{D}_\infty(B) := mn - \log|B|$ (equivalently, $|B| = 2^{mn - \mathbf{D}_\infty(B)}$), representing the log-size deficiency of B compared to the universe $(\{0,1\}^m)^n$. (The notation $\mathbf{D}_\infty$ was chosen partly because this corresponds to the Rényi max-divergence between the uniform distributions over B and over $(\{0,1\}^m)^n$.)

▶ **Lemma 14** (Full support). *If $A \subseteq [m]^n$ is 0.9-dense and $B \subseteq (\{0,1\}^m)^n$ satisfies $\mathbf{D}_\infty(B) \leq m^{0.3}$, then $G(A \times B) = \{0,1\}^n$ (i.e., for every $z \in \{0,1\}^n$ there are $x \in A$ and $y \in B$ with $G(x,y) = z$).*

We prove Theorem 14 in subsection 3.4 using the probabilistic method: we show for a suitably randomly chosen rectangle $U \times V \subseteq G^{-1}(z)$, (i) U intersects A with high probability, and (ii) V intersects B with high probability. The proof of (i) uses the second moment method (which is different from how blockwise-density was employed in previous work [13]). The proof of (ii) is a tightened analysis of a combination of arguments from [28, 14] (which were not stated in those papers with the high-probability guarantee we need). The latter papers proved the full support property under a different assumption on A, which they called "thickness".

Theorem 14 gives us the full support property assuming A is blockwise-dense and B has low deficiency. How can we get blockwise-density? Our tool for this is the following claim, which follows from [13]; we provide the simple argument.

▶ **Claim 15.** *If $A \subseteq [m]^n$ satisfies $|A| \geq m^n/2^{O(k)}$ then there exists an $I \subseteq [n]$ of size $|I| \leq O(k/\log n)$ and an $A' \subseteq A$ such that A' is fixed on I and 0.9-dense on $\overline{I} \coloneqq [n] \smallsetminus I$.*

Proof. If A is 0.9-dense, then we can take $I = \emptyset$ and $A' = A$, so assume not. Letting $\boldsymbol{x}$ be the uniform random variable over A, take $I \subseteq [n]$ to be a maximal subset for which there is a violation of blockwise-density: $\mathbf{H}_\infty(\boldsymbol{x}_I) < 0.9 \cdot |I| \log m$. From $\mathbf{H}_\infty(\boldsymbol{x}) \geq n \log m - O(k)$ we deduce $\mathbf{H}_\infty(\boldsymbol{x}_I) \geq |I| \log m - O(k)$ since marginalizing out $|\overline{I}| \log m$ bits may only cause the min-entropy to go down by $|\overline{I}| \log m$. Combining these, we get $|I| \log m - O(k) < 0.9 \cdot |I| \log m$, so $|I| \leq O(k/\log n)$.

Let $\alpha \in [m]^I$ be an outcome for which $\mathbf{Pr}[\boldsymbol{x}_I = \alpha] > 2^{-0.9 \cdot |I| \log m}$, and take $A' \coloneqq \{x \in A : x_I = \alpha\}$, which is fixed on I. To see that A' is 0.9-dense on $\overline{I}$, let $\boldsymbol{x}'$ be the uniform random variable over A' and note that if $\mathbf{H}_\infty(\boldsymbol{x}'_J) < 0.9 \cdot |J| \log m$ for some nonempty $J \subseteq \overline{I}$, a straightforward calculation shows that then $\boldsymbol{x}_{I \cup J}$ would also have min-entropy rate < 0.9, contradicting the maximality of I. ◀

3.3 Finding R

We initialize $X \coloneqq [m]^n$ and $Y \coloneqq (\{0,1\}^m)^n$ and iterate the following for $i = 1, \ldots, 2^k$ rounds.

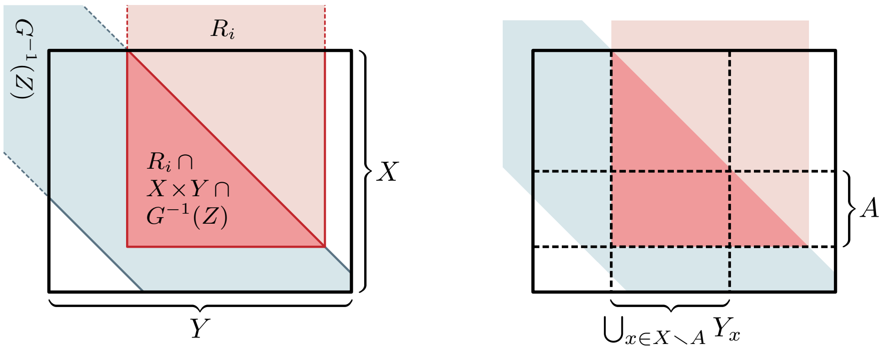
(‡) *Round i:* (loop invariant: $R_i \cap X \times Y \cap G^{-1}(Z)$ is monochromatic)

Define a set $A \subseteq X$ of *weighty rows* as

$$A \coloneqq \{x \in X : |Y_x| \geq 2^{mn - 3n \log m}\} \quad \text{where} \quad Y_x \coloneqq \{y \in Y : (x,y) \in R_i \cap G^{-1}(Z)\}.$$

Test whether there are many weighty rows: $|A| \geq m^n/2^{k+1}$?

- If *no*, we update $X \leftarrow X \smallsetminus A$ and $Y \leftarrow Y \smallsetminus \bigcup_{x \in X \smallsetminus A} Y_x$ and proceed to the next round. Since $R_i \cap G^{-1}(Z)$ has been removed from $X \times Y$, this ensures our loop invariant, as explained in subsection 3.1.
- If *yes*, we declare this round a *success* and halt.



We shortly argue that the process halts with success. First, we show how to find a desired R assuming the process is successful in round i (with associated sets R_i, $X \times Y$, A, and Y_x for $x \in X$). Using Claim 15, obtain $A' \subseteq A$ which is fixed to α on some $I \subseteq [n]$ of size $O(k/\log n)$ and is 0.9-dense on $\overline{I}$. Pick any $x' \in A'$, and define $\gamma \in \{0,1\}^I$ to be a value that maximizes the size of $B \coloneqq \{y \in Y_{x'} : g^I(\alpha, y_I) = \gamma\}$. Note that $|B| \geq |Y_{x'}|/2^{|I|} \geq 2^{mn-3n\log m - O(k/\log n)} \geq 2^{mn-m^{0.3}}$ since $x' \in A$ and $k \leq n\log(2m)$.

We claim that $R \coloneqq A' \times B$ can serve as our desired rectangle. Certainly, R touches $G^{-1}(Z_{\ell_i})$ (at (x', y) for any $y \in B$) but not $G^{-1}(Z_{1-\ell_i})$ by the loop invariant (since $R \subseteq R_i \cap X \times Y$). Also, $G(R)$ is fixed to γ on I. Defining

$$A'_{\overline{I}} \coloneqq \{x_{\overline{I}} \in [m]^{\overline{I}} : \alpha x_{\overline{I}} \in A'\} \qquad \text{and} \qquad B_{\overline{I}} \coloneqq \{y_{\overline{I}} \in (\{0,1\}^m)^{\overline{I}} : \exists y_I \text{ s.t. } y_I y_{\overline{I}} \in B\}$$

to be the projections of A' and B to the coordinates $\overline{I}$, we have that

$$A'_{\overline{I}} \text{ is 0.9-dense} \qquad \text{and} \qquad \mathbf{D}_\infty(B_{\overline{I}}) \leq \mathbf{D}_\infty(B) \leq m^{0.3}$$

(noting that $\mathbf{D}_\infty(B_{\overline{I}})$ is relative to $(\{0,1\}^m)^{\overline{I}}$). Applying Theorem 14 to $A'_{\overline{I}} \times B_{\overline{I}}$ shows that $G(R)$ has full support on $\overline{I}$. In summary, "$z_I = \gamma$" is the conjunction we were looking for.

We now argue that the process halts with success. In each unsuccessful round, we remove $|A| < m^n/2^{k+1}$ rows from X and at most $\sum_{x\in X\smallsetminus A} |Y_x| < m^n \cdot 2^{mn-3n\log m} \leq 2^{mn}/2^{k+1}$ columns from Y (since $k+1 \leq 2n\log m$). Suppose for contradiction that all 2^k rounds are unsuccessful; then at most half of the rows and half of the columns are removed altogether. Supposedly the set $X \times Y$ we finish with is disjoint from $\bigcup_{i\in[2^k]}(R_i \cap G^{-1}(Z)) = G^{-1}(Z)$. But since Z is nonempty, this contradicts the fact that $G(X \times Y)$ has full support by Theorem 14 (as it is straightforward to check that since $X \times Y$ contains at least half the rows and half the columns, it also satisfies the assumptions of the lemma).

This concludes the proof of Theorem 6, except for the proof of Theorem 14.

3.4 Full Support Lemma

▶ **Lemma 16** (Full support). *If $A \subseteq [m]^n$ is 0.9-dense and $B \subseteq (\{0,1\}^m)^n$ satisfies $\mathbf{D}_\infty(B) \leq m^{0.3}$, then $G(A \times B) = \{0,1\}^n$ (i.e., for every $z \in \{0,1\}^n$ there are $x \in A$ and $y \in B$ with $G(x,y) = z$).*

For coordinates $I \subseteq [n]$ we define $B_I \coloneqq \{y_I \in (\{0,1\}^m)^I : \exists y_{\overline{I}} \text{ s.t. } y_I y_{\overline{I}} \in B\}$ as the *projection* of B onto I. Moreover, for $V \subseteq \{0,1\}^m$ and $i \in [n]$ we let $B^{i,V} \coloneqq \{y \in B : y_i \in V\}$ be the *restriction* of the i-th coordinate to be in V. We will often use combinations of these notations; e.g., $B^{n,V}_{[n-1]}$ denotes the restriction of the n-th coordinate to be in V, subsequently projected on the coordinates in $[n-1]$.

We write random variables as bold letters. For a random variable $\boldsymbol{y}$ supported on B, $\boldsymbol{y}_I$ denotes the marginal distribution of $\boldsymbol{y}$ on the coordinates in I. In contrast, B_I only denotes the set obtained by projecting B on the coordinates in I, without any distribution associated to it. Note that while $\mathbf{D}_\infty(B)$ is the deficiency relative to $(\{0,1\}^m)^n$, the quantity $\mathbf{D}_\infty(B_I)$ is the deficiency relative to $(\{0,1\}^m)^I$; i.e., $\mathbf{D}_\infty(B_I) = m|I| - \log|B_I|$.

Theorem 14 follows from the following two claims.

▶ **Claim 17** (Alice side). *Suppose $A \subseteq [m]^n$ is 0.9-dense. Choose $\boldsymbol{U} \coloneqq \boldsymbol{U}_1 \times \cdots \times \boldsymbol{U}_n \subseteq [m]^n$ uniformly at random where each $\boldsymbol{U}_i \subseteq [m]$ is of size $|\boldsymbol{U}_i| = m^{0.36}$. Then*

$$\mathbf{Pr}[A \cap \boldsymbol{U} \neq \emptyset] \;\geq\; 1 - 2m^{-0.01}.$$

▶ **Claim 18** (Bob side). *Let $z \in \{0,1\}$ and suppose $B \subseteq (\{0,1\}^m)^n$ satisfies $\mathbf{D}_\infty(B) \leq m^{0.31}$. Choose $\boldsymbol{U} \subseteq [m]$, $|\boldsymbol{U}| = m^{0.36}$, uniformly at random and let $\boldsymbol{V} \coloneqq \{y \in \{0,1\}^m : \forall j \in \boldsymbol{U}, y_j = z\}$. Then*

$$\text{for } n \geq 2: \qquad \mathbf{Pr}\big[\mathbf{D}_\infty\big(B^{n,\boldsymbol{V}}_{[n-1]}\big) \leq \mathbf{D}_\infty(B) + 1\big] \;\geq\; 1 - 60m^{-0.28},$$

$$\text{for } n = 1: \qquad \mathbf{Pr}\left[B \cap \boldsymbol{V} \neq \emptyset\right] \;\geq\; 1 - 60m^{-0.28}.$$

We prove the Alice side claim shortly using the second moment method. The Bob side claim follows by a tightened analysis of arguments from [28, 14], which we provide in the full version of the paper [12]. Let us see why these two claims imply Theorem 14.

Proof of Theorem 14. Our goal is to show that for each $z \in \{0,1\}^n$ we have $A \times B \cap G^{-1}(z) \neq \emptyset$. Choose $\boldsymbol{U} \coloneqq \boldsymbol{U}_1 \times \cdots \times \boldsymbol{U}_n \subseteq [m]^n$, $|\boldsymbol{U}_i| = m^{0.36}$, uniformly at random. Correspondingly, define $\boldsymbol{V} \coloneqq \boldsymbol{V}_1 \times \cdots \times \boldsymbol{V}_n$ where $\boldsymbol{V}_i \coloneqq \{y \in \{0,1\}^m : \forall j \in \boldsymbol{U}_i, y_j = z_i\}$. We have $\boldsymbol{U} \times \boldsymbol{V} \subseteq G^{-1}(z)$ by construction so it suffices to show that $A \times B \cap \boldsymbol{U} \times \boldsymbol{V}$ is nonempty with positive probability. To this end, we show that the events $A \cap \boldsymbol{U} \neq \emptyset$ and $B \cap \boldsymbol{V} \neq \emptyset$ both happen with high probability, and hence, by a union bound, $A \times B \cap \boldsymbol{U} \times \boldsymbol{V}$ is nonempty with high probability. The Alice side claim (Claim 17) already shows $A \cap \boldsymbol{U} \neq \emptyset$ w.h.p., so it remains to consider $B \cap \boldsymbol{V}$.

Define $\boldsymbol{B}^{\rhd i} \coloneqq B \cap ((\{0,1\}^m)^i \times \boldsymbol{V}_{i+1} \times \cdots \times \boldsymbol{V}_n)$. That is, $\boldsymbol{B}^{\rhd i}$ is obtained by restricting the j-th coordinate to be in $\boldsymbol{V}_j$ for $i+1 \leq j \leq n$. Note that $\boldsymbol{B}^{\rhd n} = B$, $\boldsymbol{B}^{\rhd i-1} = (\boldsymbol{B}^{\rhd i})^{i,\boldsymbol{V}_i}$ and $\boldsymbol{B}^{\rhd 0} = B \cap \boldsymbol{V}$. Let $\widehat{\boldsymbol{B}}^{\rhd i} \coloneqq \boldsymbol{B}^{\rhd i}_{[i]}$ be the projection of $\boldsymbol{B}^{\rhd i}$ onto $[i]$. We define the following events E_i:

$$\text{for } i \geq 2: \qquad E_i \iff \mathbf{D}_\infty(\widehat{\boldsymbol{B}}^{\rhd i-1}) \;\leq\; \mathbf{D}_\infty(\widehat{\boldsymbol{B}}^{\rhd i}) + 1,$$

$$\text{for } i = 1: \qquad E_1 \iff \widehat{\boldsymbol{B}}^{\rhd 1} \cap \boldsymbol{V}_1 \neq \emptyset.$$

Note that $\widehat{\boldsymbol{B}}^{\rhd 1} \cap \boldsymbol{V}_1 \neq \emptyset$ implies that $\boldsymbol{B}^{\rhd 0} = B \cap \boldsymbol{V} \neq \emptyset$. Conditioned on $E_n \cap \cdots \cap E_{i+1}$, we have

$$\mathbf{D}_\infty\big(\widehat{\boldsymbol{B}}^{\rhd i}\big) \;\leq\; \mathbf{D}_\infty\big(\widehat{\boldsymbol{B}}^{\rhd n}\big) + n - i - 1 \;\leq\; m^{0.3} + n \;\leq\; m^{0.31}$$

and thus for $i \geq 2$, we have from Claim 18 that $\mathbf{D}_\infty(\widehat{\boldsymbol{B}}^{\rhd i-1}) \leq \mathbf{D}_\infty(\widehat{\boldsymbol{B}}^{\rhd i}) + 1$ holds with probability at least $1 - 60m^{-0.28}$. Thus

$$\mathbf{Pr}[E_i \mid E_n \cap \cdots \cap E_{i+1}] \;\geq\; 1 - 60m^{-0.28}.$$

Also, conditioned on $E_n \cap \cdots \cap E_2$, we have $\mathbf{D}_\infty(\widehat{\boldsymbol{B}}^{\rhd 1}) \leq m^{0.31}$, and hence using the case of $n = 1$ in Claim 18, $\mathbf{Pr}[\widehat{\boldsymbol{B}}^{\rhd 1} \cap \boldsymbol{V}_1 \neq \emptyset] \geq 1 - 60m^{-0.28}$. That is,

$$\mathbf{Pr}[E_1 \mid E_n \cap \cdots \cap E_2] \;\geq\; 1 - 60m^{-0.28}.$$

Now we are able to show $B \cap \boldsymbol{V} \neq \emptyset$ w.h.p., which concludes the proof:

$$\begin{aligned}
\mathbf{Pr}[B \cap \boldsymbol{V} \neq \emptyset] &\geq \mathbf{Pr}[E_1] \\
&\geq \mathbf{Pr}[E_n \cap \cdots \cap E_1] \\
&= \textstyle\prod_{i=1}^{n} \mathbf{Pr}[E_i \mid E_n \cap \cdots \cap E_{i+1}] \\
&\geq (1 - 60m^{-0.28})^n \\
&\geq 1 - 60nm^{-0.28} \\
&= 1 - 60m^{-0.03}.
\end{aligned}$$ ◀

Proof of Claim 17. For each $x \in A$ consider the indicator random variable $\mathbf{1}_x \in \{0,1\}$ indicating whether $x \in \boldsymbol{U}$. Let $\boldsymbol{s} := \sum_{x \in A} \mathbf{1}_x$ so that $\boldsymbol{s} = |A \cap \boldsymbol{U}|$ and $\mathbf{E}[\boldsymbol{s}] = \delta|A|$, where $\delta = |\boldsymbol{U}|/m^n = m^{-0.64n}$. We use the second moment method to estimate

$$\mathbf{Pr}[A \cap \boldsymbol{U} \neq \emptyset] \;=\; 1 - \mathbf{Pr}[\boldsymbol{s} = 0] \;\geq\; 1 - \frac{\mathbf{Var}[\boldsymbol{s}]}{\mathbf{E}[\boldsymbol{s}]^2}.$$

Thus, to prove the claim it suffices to show that $\mathbf{Var}[\boldsymbol{s}] \leq 2m^{-0.01} \cdot \mathbf{E}[\boldsymbol{s}]^2 = 2m^{-0.01} \cdot \delta^2|A|^2$. Since

$$\mathbf{Var}[\boldsymbol{s}] \;=\; \textstyle\sum_{x,x'} \mathbf{Cov}[\mathbf{1}_x, \mathbf{1}_{x'}] \;=\; \sum_{x,x'} \left(\mathbf{E}[\mathbf{1}_x \mathbf{1}_{x'}] - \mathbf{E}[\mathbf{1}_x]\mathbf{E}[\mathbf{1}_{x'}]\right),$$

it suffices to show that, for each fixed $x^* \in A$,

$$\textstyle\sum_{x \in A} \mathbf{Cov}[\mathbf{1}_x, \mathbf{1}_{x^*}] \;\leq\; 2m^{-0.01} \cdot \delta^2|A|.$$

Fix $x^* \in A$. Let $I_x \subseteq [n]$ denote the set of all blocks i such that $x_i = x_i^*$. First note that under $I_x = \emptyset$ it holds that $\mathbf{Cov}[\mathbf{1}_x, \mathbf{1}_{x^*}] < 0$, i.e., the events "$x \in \boldsymbol{U}$" and "$x^* \in \boldsymbol{U}$" are negatively correlated. The interesting case is thus $I_x \neq \emptyset$ when the two events are positively correlated. We note that

$$\mathbf{Pr}[x \in \boldsymbol{U} \mid x^* \in \boldsymbol{U}] \;=\; \left(\tfrac{m^{0.36}-1}{m-1}\right)^{n-|I_x|} \;\leq\; m^{0.64|I_x|} \cdot \delta. \tag{2}$$

Let $\boldsymbol{I}$ be the distribution of $I_{\boldsymbol{x}}$ when $\boldsymbol{x} \in A$ is chosen uniformly at random. We have

$$\begin{aligned}
\textstyle\sum_{x \in A} \mathbf{Cov}[\mathbf{1}_x, \mathbf{1}_{x^*}] &\leq \textstyle\sum_{x: I_x \neq \emptyset} \mathbf{Cov}[\mathbf{1}_x, \mathbf{1}_{x^*}] \\
&\leq \textstyle\sum_{x: I_x \neq \emptyset} \mathbf{E}[\mathbf{1}_x \mathbf{1}_{x^*}] \\
&= \textstyle\sum_{x: I_x \neq \emptyset} \mathbf{Pr}[x \in \boldsymbol{U} \text{ and } x^* \in \boldsymbol{U}] \\
&= \mathbf{Pr}[x^* \in \boldsymbol{U}] \cdot \textstyle\sum_{x: I_x \neq \emptyset} \mathbf{Pr}[x \in \boldsymbol{U} \mid x^* \in \boldsymbol{U}] \\
&= \delta \cdot \textstyle\sum_{x: I_x \neq \emptyset} \mathbf{Pr}[x \in \boldsymbol{U} \mid x^* \in \boldsymbol{U}] \\
&= \delta|A| \cdot \textstyle\sum_{\emptyset \neq I \subseteq [n]} \mathbf{Pr}[\boldsymbol{I} = I] \cdot \mathbf{E}_{\boldsymbol{x} \sim A | I_{\boldsymbol{x}} = I} \mathbf{Pr}[\boldsymbol{x} \in \boldsymbol{U} \mid x^* \in \boldsymbol{U}] \\
&\leq \delta|A| \cdot \textstyle\sum_{\emptyset \neq I \subseteq [n]} \mathbf{Pr}_{\boldsymbol{x} \sim A}[\boldsymbol{x}_I = x_I^*] \cdot \mathbf{E}_{\boldsymbol{x} \sim A | I_{\boldsymbol{x}} = I} \mathbf{Pr}[\boldsymbol{x} \in \boldsymbol{U} \mid x^* \in \boldsymbol{U}] \\
&\leq \delta|A| \cdot \textstyle\sum_{\emptyset \neq I \subseteq [n]} 2^{-0.9|I| \log m} \cdot m^{0.64|I|} \cdot \delta \qquad \text{(0.9-density and (2))} \\
&= \delta^2|A| \cdot \textstyle\sum_{\emptyset \neq I \subseteq [n]} 2^{-0.26|I| \log m} \\
&= \delta^2|A| \cdot \textstyle\sum_{k \in [n]} \binom{n}{k} 2^{-0.26k \log m} \\
&\leq \delta^2|A| \cdot \textstyle\sum_{k \in [n]} (m^{0.25})^k \cdot 2^{-0.26k \log m} \\
&\leq \delta^2|A| \cdot 2 \cdot 2^{-0.01 \log m} \\
&\leq 2m^{-0.01} \cdot \delta^2|A|.
\end{aligned}$$ ◀

4 Application

In this section we prove Theorem 3, restated here for convenience.

▶ **Theorem 3.** *There exists a total $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ satisfying the following.*
- *F has large P^{NP} communication complexity: $\mathsf{P}^{\mathsf{NPcc}}(F) \geq n^{\Omega(1)}$.*
- *For any product distribution μ over $\{0,1\}^n \times \{0,1\}^n$, there exists a monochromatic rectangle R that is large: $\log(1/\mu(R)) \leq \log^{O(1)} n$.*

The function witnessing the separation is $F := f \circ g^n$ where $g := \textsc{Ind}_m$ is the index function with $m := n^4$ and $f\colon \{0,1\}^n \to \{0,1\}$ is defined as follows. We interpret the input M to f as a $\sqrt{n} \times \sqrt{n}$ boolean matrix, and set

$$f(M) \;:=\; 1 \qquad \text{iff} \qquad \text{every row of } M \text{ contains a unique 1-entry.}$$

Complexity class aficionados [1] can recognize f as the canonical complete problem for the decision tree analogue of $\forall\cdot\mathsf{US}$ ($\subseteq \Pi_2\mathsf{P}$) where US is the class of functions whose 1-inputs admit a *unique* witness [5]. We have $F\colon \{0,1\}^{n\log m} \times \{0,1\}^{nm} \to \{0,1\}$, but we can polynomially pad Alice's input length to match Bob's (as in the statement of Theorem 3).

4.1 Lower bound

It is proved in several sources [32, 21, 16] that f cannot be computed by an efficient $\Sigma_2\mathsf{P}$-type decision tree (i.e., quasi-polynomial-size depth-3 circuit with an $\textsc{Or}$-gate at the top and small bottom fan-in), let alone an efficient P^{NP} decision tree. However, for completeness, we might as well give a simple proof using our characterization (Theorem 11). Applying the lifting theorem to the following lemma yields the lower bound.

▶ **Lemma 19.** $\mathsf{DL}^{\mathsf{dt}}(f) \geq \sqrt{n}$.

Proof. By Theorem 11 it is enough to exhibit a nonempty subset $Z \subseteq \{0,1\}^n$ of inputs such that each conjunction C of width $\sqrt{n}-1$ accepts an input in $Z_1 := Z \cap f^{-1}(1)$ iff it accepts an input in $Z_0 := Z \cap f^{-1}(0)$. We define Z as the set of $\sqrt{n} \times \sqrt{n}$ matrices with at most one 1-entry in each row. If C accepts an input $M \in Z_1$, then there is some row of M none of whose entries are read by C; we may modify that row to all-0 and conclude that C accepts an input in Z_0. If C accepts an input $M \in Z_0$, then for each all-0 row of M there is some entry that is not read by C; we may modify each of those entries to a 1 and conclude that C accepts an input in Z_1. ◀

4.2 Upper bound

Let μ be a product distribution over the domain of $F = f \circ g^n$. Call a matrix M *heavy* if it contains a row with at least two 1-entries. Hence $f(M) = 0$ for every heavy matrix M. There is an efficient nondeterministic protocol of cost $k \leq O(\log n)$, call it Π, that checks whether a particular (x, y) describes a heavy matrix $M = g^n(x, y)$. Namely, Π guesses a row index $i \in [\sqrt{n}]$ and two column indices $1 \leq j < j' \leq \sqrt{n}$, and then communicates $2\log m + 1 \leq O(\log n)$ bits to check that $M_{ij} = M_{ij'} = 1$. We view Π as defining a rectangle covering $\bigcup_{i \in [2^k]} R_i$ of all those (x, y) that describe heavy matrices. Note that each R_i is monochromatic for F.

If there is an R_i with $\mu(R_i) \geq 2^{-4k}$, the theorem is proved. So suppose not: $\mu(R_i) < 2^{-4k}$ for all i. Starting with $S :=$ domain of F and iterating over the R_i exactly as in the proof of Theorem 10, we can delete from S either the rows or the columns of each R_i, ending up with a rectangle S still of measure $\mu(S) \geq 1 - 2^k \cdot 2^{-2k} \geq 0.99$. We will complete the argument by showing that F_S (i.e., F restricted to the rectangle S) admits a large monochromatic rectangle relative to μ_S, the conditional distribution of μ given S (which is also product).

$$\begin{tikzcd}
\forall\cdot\mathsf{US} \arrow[d, "\text{restrict to } S"] \\
\forall\cdot\mathsf{UP} \arrow[d, "\text{Yannakakis}"] \\
\forall\cdot\mathsf{P} \arrow[d, "="] \\
\mathsf{coNP} \arrow[d, "\text{product method}"] \\
\text{Large monochr. rectangle}
\end{tikzcd}$$

All $(x, y) \in S$ are such that $M = g^n(x, y)$ is *not* heavy. This means that the function F_S is easier than the ($\forall\cdot\mathsf{US}$-complete) function F in the following sense: for each row $i \in [\sqrt{n}]$ there is an efficient $O(\log n)$-cost nondeterministic protocol, call it Π_i, to check whether the i-th row of $M = g^n(x, y)$ contains a 1-entry, and moreover, this protocol is *unambiguous* in that it has at most one accepting computation on any input. (In complexity lingo, F_S admits an efficient $\forall\cdot\mathsf{UP}$ protocol.) It is a well-known theorem of Yannakakis [40, Lemma 1] that any such unambiguous Π_i can be made deterministic with at most a quadratic blow-up in cost; let $\Pi_i^{\det}$ be that $O(\log^2 n)$-bit deterministic protocol. But now $\neg F_S$ (negation of F_S) is computed by the following $O(\log^2 n)$-bit nondeterministic protocol: on input (x, y) guess a row index $i \in [\sqrt{n}]$ and run $\Pi_i^{\det}$ accepting iff $\Pi_i^{\det}(x, y) = 0$. (That is, F_S admits an efficient $\forall\cdot\mathsf{P} = \mathsf{coNP}$ protocol.) We proved $\mathsf{NP}^{\mathsf{cc}}(\neg F_S) \leq O(\log^2 n)$; in particular,

$$\mathsf{DL}^{\mathsf{cc}}(F_S) \;\leq\; O(\mathsf{P}^{\mathsf{NPcc}}(F_S)) \;\leq\; O(\mathsf{NP}^{\mathsf{cc}}(\neg F_S)) \;\leq\; O(\log^2 n).$$

Hence we can apply (as a black box) the product method (Theorem 10) to find a monochromatic rectangle $R \subseteq S$ with $\log(1/\mu_S(R)) \leq O(\log^2 n)$ and hence $\log(1/\mu(R)) \leq O(\log^2 n)$. This completes the proof of Theorem 3.

5 Conclusion

Let $\mathsf{PM}(F)$ denote the best lower bound on $\mathsf{DL}^{\mathsf{cc}}(F)$ that can be derived by the product method (Theorem 10). For any communication complexity measure $\mathcal{C}(F)$, we use the convention that $\mathcal{C}$ by itself refers to the class of (families of) functions $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with $\mathcal{C}(F) \leq \mathrm{polylog}(n)$. Then our application (Theorem 3) shows that the inclusion $\mathsf{P}^{\mathsf{NPcc}} \subseteq \mathsf{PM}$ is strict: there is an $F \in \mathsf{PM} \smallsetminus \mathsf{P}^{\mathsf{NPcc}}$. Here are some open questions.

1. Is there an $F \in \mathsf{PM} \smallsetminus \mathsf{UPP}^{\mathsf{cc}}$? This would be a stronger result since $\mathsf{P}^{\mathsf{NPcc}} \subseteq \mathsf{UPP}^{\mathsf{cc}}$. Note that our $\forall\cdot\mathsf{US}$-complete function does not witness this, since it is in $\mathsf{PP}^{\mathsf{cc}}$. One way to see this is to note that it is the intersection of a $\mathsf{coNP}^{\mathsf{cc}}$ function (does each row have at most one 1?) and a $\mathsf{PP}^{\mathsf{cc}}$ function (is the number of 1's at least the number of rows?), and use the closure of PP under intersection [4].

2. Is there any reasonable upper bound for PM? For example, does $\mathsf{PM} \subseteq \mathsf{PSPACE}^{\mathsf{cc}}$ hold?
3. Does $\mathsf{BPP}^{\mathsf{cc}} \subseteq \mathsf{PM}$ or even $\mathsf{BPP}^{\mathsf{cc}} \subseteq \mathsf{P}^{\mathsf{NPcc}}$ hold for total functions? The separation $\mathsf{BPP}^{\mathsf{cc}} \not\subseteq \mathsf{PM}$ was shown for partial functions implicitly in [25].
4. Is there a lossless $\mathsf{P}^{\mathsf{NPdt}}$-to-$\mathsf{P}^{\mathsf{NPcc}}$ lifting theorem (Conjecture 7)?
5. Can the quadratic upper bounds in Facts 4 and 5 be shown tight for more general parameters (beyond constant $\mathsf{DL}^{\mathsf{dt}}(f)$ and logarithmic $\mathsf{DL}^{\mathsf{cc}}(F)$ as in subsection 1.4)?

Acknowledgments. We thank Paul Balister, Shalev Ben-David, Béla Bollobás, Robin Kothari, Nirman Kumar, Santosh Kumar, Govind Ramnarayan, Madhu Sudan, Li-Yang Tan, and Justin Thaler for discussions and correspondence.

References

1 Scott Aaronson, Greg Kuperberg, and Christopher Granade. Complexity zoo. Online, 2017. URL: https://complexityzoo.uwaterloo.ca.

2 László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th Symposium on Foundations of Computer Science (FOCS)*, pages 337–347. IEEE, 1986. doi:10.1109/SFCS.1986.15.

3 Richard Beigel. Perceptrons, PP, and the polynomial hierarchy. *Computational complexity*, 4(4):339–349, 1994. doi:10.1007/BF01263422.

4 Richard Beigel, Nick Reingold, and Daniel Spielman. PP is closed under intersection. *Journal of Computer and System Sciences*, 50(2):191–202, 1995. doi:10.1006/jcss.1995.1017.

5 Andreas Blass and Yuri Gurevich. On the unique satisfiability problem. *Information and Control*, 55(1–3):80–88, 1982. doi:10.1016/S0019-9958(82)90439-9.

6 Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288(1):21–43, 2002. doi:10.1016/S0304-3975(01)00144-X.

7 Harry Buhrman, Nikolai Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *Proceedings of the 22nd Conference on Computational Complexity (CCC)*, pages 24–32. IEEE, 2007. doi:10.1109/CCC.2007.18.

8 Mark Bun and Justin Thaler. Approximate degree and the complexity of depth three circuits. Technical Report TR16-121, Electronic Colloquium on Computational Complexity (ECCC), 2016. URL: https://eccc.weizmann.ac.il/report/2016/121/.

9 Siu On Chan, James Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. *Journal of the ACM*, 63(4):34:1–34:22, 2016. doi:10.1145/2811255.

10 Susanna de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication (and what it means for proof and circuit complexity). In *Proceedings of the 57th Symposium on Foundations of Computer Science (FOCS)*, pages 295–304. IEEE, 2016. doi:10.1109/FOCS.2016.40.

11 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.

12 Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-Communication Lifting for P^{NP}. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:24, 2017. URL: https://eccc.weizmann.ac.il/report/2017/024.

13 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.

14 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`.

15 Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. In *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 55 of *LIPIcs*, pages 86:1–86:15. Schloss Dagstuhl, 2016. `doi:10.4230/LIPIcs.ICALP.2016.86`.

16 Johan Håstad, Stasys Jukna, and Pavel Pudlák. Top-down lower bounds for depth-three circuits. *Computational Complexity*, 5(2):99–112, 1995. `doi:10.1007/BF01268140`.

17 Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. In *Proceedings of the 57th Symposium on Foundations of Computer Science (FOCS)*, pages 282–288. IEEE, 2016. `doi:10.1109/FOCS.2016.38`.

18 Russell Impagliazzo and Ryan Williams. Communication complexity with synchronized clocks. In *Proceedings of the 25th Conference on Computational Complexity (CCC)*, pages 259–269. IEEE, 2010. `doi:10.1109/CCC.2010.32`.

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

20 Mauricio Karchmer, Eyal Kushilevitz, and Noam Nisan. Fractional covers and communication complexity. *SIAM Journal on Discrete Mathematics*, 8(1):76–92, 1995. `doi:10.1137/S0895480192238482`.

21 Ker-I Ko. Separating and collapsing results on the relativized probabilistic polynomial-time hierarchy. *Journal of the ACM*, 37(2):415–438, 1990. `doi:10.1145/77600.77623`.

22 Pravesh Kothari, Raghu Meka, and Prasad Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs. In *Proceedings of the 49th Symposium on Theory of Computing (STOC)*. ACM, 2017. To appear.

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

24 James Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the 47th Symposium on Theory of Computing (STOC)*, pages 567–576. ACM, 2015. `doi:10.1145/2746539.2746599`.

25 Periklis Papakonstantinou, Dominik Scheder, and Hao Song. Overlays and limited memory communication. In *Proceedings of the 29th Conference on Computational Complexity (CCC)*, pages 298–308. IEEE, 2014. `doi:10.1109/CCC.2014.37`.

26 Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33(1):106–123, 1986. `doi:10.1016/0022-0000(86)90046-2`.

27 Anup Rao and Amir Yehudayoff. *Communication Complexity*. In preparation, 2017.

28 Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999. `doi:10.1007/s004930050062`.

29 Alexander Razborov and Alexander Sherstov. The sign-rank of AC^0. *SIAM Journal on Computing*, 39(5):1833–1855, 2010. `doi:10.1137/080744037`.

30 Ronald Rivest. Learning decision lists. *Machine Learning*, 2(3):229–246, 1987. `doi:10.1007/BF00058680`.

31 Robert Robere, Toniann Pitassi, Benjamin Rossman, and Stephen Cook. Exponential lower bounds for monotone span programs. In *Proceedings of the 57th Symposium on Foundations of Computer Science (FOCS)*, pages 406–415. IEEE, 2016. `doi:10.1109/FOCS.2016.51`.

32 Miklos Santha. Relativized Arthur–Merlin versus Merlin–Arthur games. *Information and Computation*, 80(1):44–49, 1989. `doi:10.1016/0890-5401(89)90022-9`.

33 Rocco Servedio, Li-Yang Tan, and Justin Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. In *Proceedings of the 25th Conference* *on Learning Theory (COLT)*, pages 14.1–14.19. JMLR, 2012. URL: `http://www.jmlr.org/proceedings/papers/v23/servedio12/servedio12.pdf`.

34 Alexander Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011. `doi:10.1137/080733644`.

35 Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information and Computation*, 9(5–6):444–460, 2009.

36 Justin Thaler. Lower bounds for the approximate degree of block-composed functions. In *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 55 of *LIPIcs*, pages 17:1–17:15. Schloss Dagstuhl, 2016. `doi:10.4230/LIPIcs.ICALP.2016.17`.

37 Salil Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1–3):1–336, 2012. `doi:10.1561/0400000010`.

38 Nikolai Vereshchagin. Relativizability in complexity theory. In *Provability, Complexity, Grammars*, volume 192 of *AMS Translations, Series 2*, pages 87–172. American Mathematical Society, 1999.

39 Ryan Williams. Brute force search and oracle-based computation. Technical report, Cornell University, 2001. URL: `https://web.stanford.edu/~rrwill/bfsearch-rev.ps`.

40 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`.