Nondeterministic and Randomized Boolean Hierarchies in Communication Complexity

Toniann Pitassi
University of Toronto, Canada
Institute for Advanced Study, Princeton, NJ, USA
toni@cs.toronto.edu

Morgan Shirley
University of Toronto, Canada
shirley@cs.toronto.edu

Thomas Watson
University of Memphis, TN, USA
Thomas.Watson@memphis.edu

Abstract

We investigate the power of randomness in two-party communication complexity. In particular, we study the model where the parties can make a constant number of queries to a function with an efficient one-sided-error randomized protocol. The complexity classes defined by this model comprise the Randomized Boolean Hierarchy, which is analogous to the Boolean Hierarchy but defined with one-sided-error randomness instead of nondeterminism. Our techniques connect the Nondeterministic and Randomized Boolean Hierarchies, and we provide a complete picture of the relationships among complexity classes within and across these two hierarchies. In particular, we prove that the Randomized Boolean Hierarchy does not collapse, and we prove a query-to-communication lifting theorem for all levels of the Nondeterministic Boolean Hierarchy and use it to resolve an open problem stated in the paper by Halstenberg and Reischuk (CCC 1988) which initiated the study of this hierarchy.

2012 ACM Subject Classification Theory of computation → Communication complexity

Keywords and phrases Boolean hierarchies, lifting theorems, query complexity

Digital Object Identifier 10.4230/LIPIcs.ICALP.2020.92

Category Track A: Algorithms, Complexity and Games

Related Version A full version of this paper [27] is available at `https://eccc.weizmann.ac.il/report/2019/043`.

Funding *Toniann Pitassi*: Supported by NSERC.
Morgan Shirley: Supported by NSERC.
Thomas Watson: Supported by NSF grant CCF-1657377.

Acknowledgements We thank Benjamin Rossman for helpful comments and discussions.

1 Introduction

A classic example of the power of randomness in communication is the Equality function: Alice gets an n-bit string x, Bob gets an n-bit string y, and they want to know whether x equals y. Though Equality is maximally hard for deterministic communication [35], it can be solved by a randomized protocol using $O(1)$ bits of communication (in the public-coin model) using the fingerprinting technique. Although this example (known for over 40 years) demonstrates the power of randomized communication in the standard two-party setting, many questions remain about the exact power of randomness in communication.

Much is still not understood about the power of randomness in other important communication settings beyond the standard two-party model. For example, in the Number-on-Forehead (NOF) model, even for three parties, no explicit function is known to exhibit a superpolylogarithmic separation between randomized and deterministic communication [25] (despite the fact that linear lower bounds for randomized NOF protocols were proven over 30 years ago in the seminal paper [3]). Another example concerns randomness in the context of *nondeterministic* two-party protocols (so-called Arthur–Merlin and Merlin–Arthur models). While strong lower bounds are known for Merlin–Arthur protocols [22] (though even here, explicit linear lower bounds remain elusive), no strong lower bounds are known for Arthur–Merlin protocols computing any explicit function – such bounds are necessary for making progress on rigidity bounds and circuit lower bounds, and also important for delegation [29, 12, 1].

We wish to highlight that even in the standard setting of plain randomized two-party communication protocols, many fundamental questions remain poorly understood. Our goal in this paper is to make progress on some of these questions. Much is known about the *limitations* of randomness – e.g., strong (indeed, linear) lower bounds are known for the classic Set-Disjointness function [2, 21, 30, 4], which can be viewed as showing that $\mathsf{coNP}^{\mathsf{cc}} \not\subseteq \mathsf{BPP}^{\mathsf{cc}}$ (where we use $^{\mathsf{cc}}$ superscripts to indicate the communication analogues of classical complexity classes). However, surprisingly little is known about the *power* of randomness. Most known efficient randomized protocols for other functions, such as Greater-Than, can be viewed as oracle reductions to the aforementioned Equality upper bound: Greater-Than $\in \mathsf{P}^{\text{Equality}}$. Until recently, it was not even known whether $\mathsf{BPP}^{\mathsf{cc}} = \mathsf{P}^{\text{Equality}}$ (assuming the classes are defined to contain only total two-party functions), i.e., whether Equality is the "only" thing randomness is good for. Chattopadhyay, Lovett, and Vinyals [8] answered this question in the negative by exhibiting a total function that is in $\mathsf{BPP}^{\mathsf{cc}}$ (indeed, in $\mathsf{coRP}^{\mathsf{cc}}$) but not in $\mathsf{P}^{\text{Equality}}$ (though the upper bound is still a form of fingerprinting). Since Equality $\in \mathsf{coRP}^{\mathsf{cc}}$, we have $\mathsf{P}^{\text{Equality}} \subseteq \mathsf{P}^{\mathsf{RP}^{\mathsf{cc}}}$ where the latter class contains functions with efficient deterministic protocols that can make (adaptive) oracle queries to any function in $\mathsf{RP}^{\mathsf{cc}}$. In fact, [8] exhibited a strict infinite hierarchy of classes within $\mathsf{P}^{\mathsf{RP}^{\mathsf{cc}}}$, with $\mathsf{P}^{\text{Equality}}$ at the bottom, and with subsequent levels having increasingly powerful specific oracle functions.

However, it remains open whether $\mathsf{BPP}^{\mathsf{cc}} = \mathsf{P}^{\mathsf{RP}^{\mathsf{cc}}}$ for total functions (intuitively, whether two-sided error can be efficiently converted to oracle queries to one-sided error). It is even open whether $\mathsf{BPP}^{\mathsf{cc}} \subseteq \mathsf{P}^{\mathsf{NP}^{\mathsf{cc}}}$ for total functions [15], although this is known to be false if the classes are defined to allow partial functions [26]. We obtain a more detailed understanding of the structure of $\mathsf{P}^{\mathsf{RP}^{\mathsf{cc}}}$ by focusing on *restricting the number of oracle queries* (rather than restricting the $\mathsf{RP}^{\mathsf{cc}}$ function as in [8]). For constants $q = 0, 1, 2, 3, \ldots$, the class $\mathsf{P}_{\parallel}^{\mathsf{RP}[q]^{\mathsf{cc}}}$ consists of all two-party functions with an efficient (polylogarithmic communication cost) deterministic protocol that can make q many nonadaptive queries to an oracle for a function in $\mathsf{RP}^{\mathsf{cc}}$. Even if partial functions are allowed, it was not known whether these classes form a strict infinite hierarchy, i.e., whether $\mathsf{P}_{\parallel}^{\mathsf{RP}[q]^{\mathsf{cc}}} \subsetneq \mathsf{P}_{\parallel}^{\mathsf{RP}[q+1]^{\mathsf{cc}}}$ for all q. One of our main contributions (Theorem 1) implies that this is indeed the case, even for total functions. Our proof introduces new lower bound techniques.

With $\mathsf{NP}^{\mathsf{cc}}$ in place of $\mathsf{RP}^{\mathsf{cc}}$, $\bigcup_q \mathsf{P}_{\parallel}^{\mathsf{NP}[q]^{\mathsf{cc}}}$ forms the communication version of the Boolean Hierarchy from classical complexity, which was previously studied by Halstenberg and Reischuk [16, 17, 18]. We also prove results (Theorem 2 and Theorem 3) that resolve a 31-year-old open question posed in their work. $\bigcup_q \mathsf{P}_{\parallel}^{\mathsf{RP}[q]^{\mathsf{cc}}}$ can be viewed as the communication version of the Randomized Boolean Hierarchy, which has not been studied explicitly in previous works. Overall, we obtain a complete understanding of the relationships among the classes within and across the Nondeterministic and Randomized Boolean Hierarchies in communication complexity. In the following subsection we discuss the relevant communication complexity classes and describe our theorems in detail.

1.1 Background and our contributions

Forgetting about communication complexity for a moment, the Boolean Hierarchy in classical complexity theory consists of problems that have a polynomial-time algorithm making a constant number of queries to an NP oracle. This hierarchy has an intricate relationship with other complexity classes, and its second level (DP) captures the complexity of certain "exact" versions of optimization problems. It consists of an infinite sequence of complexity classes $\mathsf{NP}(q)$ for $q = 1, 2, 3, \ldots$ (where $\mathsf{NP}(1) = \mathsf{NP}$ and $\mathsf{NP}(2) = \mathsf{DP}$). Among the several equivalent ways of defining these classes [34, 7, 23, 32], perhaps the simplest is that $\mathsf{NP}(q)$ consists of all decision problems that can be computed by taking the parity of the answers to a sequence of q NP problems on the given input. As illustrated in Figure 1, it is known that these levels are intertwined with the classes $\mathsf{P}_{\|}^{\mathsf{NP}[q]}$ of all decision problems solvable in polynomial time using q nonadaptive NP queries (for constant q) [23, 32, 5]:

$$\mathsf{NP}(q) \subseteq \mathsf{P}_{\|}^{\mathsf{NP}[q]} \subseteq \mathsf{NP}(q+1) \qquad \text{and} \qquad \mathsf{coNP}(q) \subseteq \mathsf{P}_{\|}^{\mathsf{NP}[q]} \subseteq \mathsf{coNP}(q+1)$$

(by closure of $\mathsf{P}_{\|}^{\mathsf{NP}[q]}$ under complementation). Here, $\mathsf{coNP}(q)$ means $\mathsf{co}(\mathsf{NP}(q))$ rather than $(\mathsf{coNP})(q)$.

Analogous to the above *Nondeterministic* Boolean Hierarchy, one can define the *Randomized* Boolean Hierarchy by using RP (one-sided error randomized polynomial time) instead of NP in the definitions [6]. The analogous inclusions like in Figure 1 hold among all the classes $\mathsf{RP}(q)$, $\mathsf{coRP}(q)$, and $\mathsf{P}_{\|}^{\mathsf{RP}[q]}$, by similar arguments. Although the (suitably defined) *Polynomial* Hierarchy over RP is known to collapse to its second level, which equals BPP [36], the *Boolean* Hierarchy over RP has not been widely studied.

Recall the basic (deterministic) model of communication [35, 24], where Alice is given an input x and Bob is given an input y, and they wish to collaboratively evaluate some function $F(x, y)$ of their joint input by engaging in a protocol that specifies how they exchange bits of information about their inputs. Many classical complexity classes (P, RP, NP, and so on) have natural two-party communication analogues [2] (including the classes in the Nondeterministic and Randomized Boolean Hierarchies). The area of structural communication complexity, which concerns the properties of and relationships among these classes, is undergoing a renaissance and has turned out to yield new techniques and perspectives for understanding questions in a variety of other areas (circuit complexity, proof complexity, data structures, learning theory, delegation, fine-grained complexity, property testing, cryptography, extended formulations, etc.) [15]. For any classical time-bounded complexity class $\mathcal{C}$, we use $\mathcal{C}^{\mathsf{cc}}$ to denote its communication complexity analogue – the class of all two-party functions on n bits that admit a protocol communicating at most polylog(n) bits, in a model defined by analogy with the classical $\mathcal{C}$.

Halstenberg and Reischuk [16, 17] initiated the study of the Nondeterministic Boolean Hierarchy in two-party communication complexity. They observed that the inclusions shown in Figure 1 hold for the communication versions of the classes, by essentially the same proofs as in the time-bounded setting. They also proved that $\mathsf{NP}(q)^{\mathsf{cc}} \neq \mathsf{coNP}(q)^{\mathsf{cc}}$, which simultaneously implies that each of the inclusions is strict: $\mathsf{NP}(q)^{\mathsf{cc}} \subsetneq \mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}} \subsetneq \mathsf{NP}(q+1)^{\mathsf{cc}}$.

The communication version of the Randomized Boolean Hierarchy has not been explicitly studied as far as we know, but as mentioned earlier it is interesting since EQUALITY $\in \mathsf{coRP}^{\mathsf{cc}}$ and many randomized protocols have been designed by reduction to this fact (such as

$$\begin{tikzcd}
 & \begin{array}{c}\mathsf{NP} = \\ \mathsf{NP}(1)\end{array} \arrow[dr] & & \begin{array}{c}\mathsf{DP} = \\ \mathsf{NP}(2)\end{array} \arrow[dr] & & \mathsf{NP}(3) \arrow[dr] & & \mathsf{NP}(4) \arrow[dr] & & & & \\
\mathsf{P} \arrow[ur] \arrow[dr] & & \mathsf{P}^{\mathsf{NP}[1]} \arrow[ur] \arrow[dr] & & \mathsf{P}_{\|}^{\mathsf{NP}[2]} \arrow[ur] \arrow[dr] & & \mathsf{P}_{\|}^{\mathsf{NP}[3]} \arrow[ur] \arrow[dr] & & \mathsf{P}_{\|}^{\mathsf{NP}[4]} & \cdots & \mathsf{P}_{\|}^{\mathsf{NP}} \arrow[r] & \mathsf{P}^{\mathsf{NP}} \\
 & \begin{array}{c}\mathsf{coNP}(1) \\ = \mathsf{coNP}\end{array} \arrow[ur] & & \begin{array}{c}\mathsf{coNP}(2) \\ = \mathsf{coDP}\end{array} \arrow[ur] & & \mathsf{coNP}(3) \arrow[ur] & & \mathsf{coNP}(4) \arrow[ur] & & & &
\end{tikzcd}$$

Figure 1 Relations between classes in the Boolean Hierarchy. Here, $\mathcal{C}_1 \to \mathcal{C}_2$ represents $\mathcal{C}_1 \subseteq \mathcal{C}_2$.

GREATER-THAN $\in \mathsf{P}^{\mathsf{RP}^{\mathsf{cc}}}$). What can we say about the power of a fixed number of queries to an $\mathsf{RP}^{\mathsf{cc}}$ oracle? Our first contribution strengthens the aforementioned separation due to Halstenberg and Reischuk.

▶ **Theorem 1.** *For total functions,* $\mathsf{coRP}(q)^{\mathsf{cc}} \not\subseteq \mathsf{NP}(q)^{\mathsf{cc}}$ *for every constant q.*

Since $\mathsf{RP}^{\mathsf{cc}} \subseteq \mathsf{NP}^{\mathsf{cc}}$, Theorem 1 simultaneously implies that each of the inclusions in the Randomized Boolean Hierarchy is strict: $\mathsf{RP}(q)^{\mathsf{cc}} \subsetneq \mathsf{P}_{\|}^{\mathsf{RP}[q]\mathsf{cc}} \subsetneq \mathsf{RP}(q+1)^{\mathsf{cc}}$, and thus the hierarchy does not collapse. Previously, no separation beyond the first level seemed to be known in the literature. Our proof of Theorem 1 is completely different from (and more involved than) Halstenberg and Reischuk's proof of $\mathsf{coNP}(q)^{\mathsf{cc}} \not\subseteq \mathsf{NP}(q)^{\mathsf{cc}}$, which used the "easy-hard argument" of [20].

In [16, 17], Halstenberg and Reischuk asked whether the inclusion $\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}} \subseteq \mathsf{NP}(q+1)^{\mathsf{cc}} \cap \mathsf{coNP}(q+1)^{\mathsf{cc}}$ is strict. When $q = 0$, this is answered by the familiar results that $\mathsf{P}^{\mathsf{cc}} = \mathsf{NP}^{\mathsf{cc}} \cap \mathsf{coNP}^{\mathsf{cc}}$ when the classes are defined to contain only total functions [18], whereas $\mathsf{P}^{\mathsf{cc}} \subsetneq \mathsf{NP}^{\mathsf{cc}} \cap \mathsf{coNP}^{\mathsf{cc}}$ (indeed, $\mathsf{P}^{\mathsf{cc}} \subsetneq \mathsf{ZPP}^{\mathsf{cc}}$) holds when partial functions (promise problems) are allowed. For $q > 0$, we resolve this 31-year-old open question by proving that the situation is analogous to the $q = 0$ case.

▶ **Theorem 2.** *For total functions,*

$$\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}} = \mathsf{NP}(q+1)^{\mathsf{cc}} \cap \mathsf{coNP}(q+1)^{\mathsf{cc}} \qquad \textit{and} \qquad \mathsf{P}_{\|}^{\mathsf{RP}[q]\mathsf{cc}} = \mathsf{RP}(q+1)^{\mathsf{cc}} \cap \mathsf{coRP}(q+1)^{\mathsf{cc}}$$

for every constant q.

▶ **Theorem 3.** *For partial functions,* $\mathsf{RP}(q+1)^{\mathsf{cc}} \cap \mathsf{coRP}(q+1)^{\mathsf{cc}} \not\subseteq \mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}}$ *for every constant q.*

Since $\mathsf{RP}^{\mathsf{cc}} \subseteq \mathsf{NP}^{\mathsf{cc}}$, Theorem 3 implies that

$$\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}} \subsetneq \mathsf{NP}(q+1)^{\mathsf{cc}} \cap \mathsf{coNP}(q+1)^{\mathsf{cc}} \qquad \text{and} \qquad \mathsf{P}_{\|}^{\mathsf{RP}[q]\mathsf{cc}} \subsetneq \mathsf{RP}(q+1)^{\mathsf{cc}} \cap \mathsf{coRP}(q+1)^{\mathsf{cc}}$$

for partial functions. Taken together, Theorem 1, Theorem 2, and Theorem 3 complete the picture of the relationships among the classes within and across both hierarchies, for both total and partial functions.

1.2 Query-to-communication lifting

Our proof of Theorem 3 uses the paradigm of *query-to-communication lifting* [28, 11, 9, 14, 10, 13, 33]. This approach to proving communication lower bounds has led to breakthroughs on fundamental questions in communication complexity and many of its application areas. The idea consists of two steps:

(1) First prove an analogous lower bound in the simpler setting of decision tree depth complexity (a.k.a. query complexity). This step captures the combinatorial core of the lower bound argument without the burden of dealing with the full power of communication protocols.

(2) Then apply a *lifting theorem*, which translates the query lower bound into a communication lower bound for a related two-party function. This step encapsulates the general-purpose machinery for dealing with protocols, and can be reused from one argument to the next.

The availability of a lifting theorem greatly simplifies the task of proving certain communication lower bounds, because it divorces the problem-specific aspects from the generic aspects. The format of a lifting theorem is that if $f\colon \{0,1\}^n \to \{0,1\}$ is any partial function and $g\colon \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ is a certain "small" two-party function called a gadget, then the communication complexity of the two-party composed function $f\circ g^n\colon \mathcal{X}^n\times\mathcal{Y}^n\to\{0,1\}$ – in which Alice gets $x=(x_1,\dots,x_n)$, Bob gets $y=(y_1,\dots,y_n)$, and their goal is to evaluate $(f\circ g^n)(x,y) \coloneqq f(g(x_1,y_1),\dots,g(x_n,y_n))$ – should be approximately the query complexity of the outer function f. One direction is generally straightforward: given a query upper bound for f, a communication upper bound for $f\circ g^n$ is witnessed by a protocol that simulates the decision tree for f and evaluates $g(x_i,y_i)$ whenever it queries the i^{th} bit of the input to f; the number of bits of communication is at most the number of queries made by the decision tree times the (small) cost of evaluating a copy of g. The other direction is the challenging part: despite Alice and Bob's ability to send messages that depend in arbitrary ways on all n coordinates, they nevertheless cannot do much better than just simulating a decision tree, which involves "talking about" one coordinate at a time.

A lifting theorem must be stated with respect to a particular model of computation, such as deterministic, one-sided error randomized, nondeterministic, etc., which we associate with the corresponding complexity classes. Indeed, lifting theorems are known for P [28, 14], RP [13], NP [11, 9], and many other classes. It is convenient to recycle complexity class names to denote the complexity of a given function in the corresponding model, e.g., $\mathsf{P}^{\mathsf{dt}}(f)$ is the minimum worst-case number of queries made by any decision tree that computes f, and $\mathsf{P}^{\mathsf{cc}}(F)$ is the minimum worst-case communication cost of any protocol that computes F. With this notation, the deterministic lifting theorem from [28, 14] can be stated as: for all f, $\mathsf{P}^{\mathsf{cc}}(f\circ g^n)=\mathsf{P}^{\mathsf{dt}}(f)\cdot\Theta(\log n)$ where $g\colon [m]\times\{0,1\}^m\to\{0,1\}$ is the "index" gadget defined by $g(x,y)=y_x$ with $m\coloneqq n^{20}$. (Note that $\mathsf{P}^{\mathsf{cc}}(g)=O(\log n)$ since Alice can send her $\log m$-bit "pointer" to Bob, who responds with the pointed-to bit from his string.) The index gadget has also been used in lifting theorems for several other complexity classes.

We prove lifting theorems for all classes in the Nondeterministic Boolean Hierarchy, with the index gadget.

▶ **Theorem 4.** *For every partial function $f\colon\{0,1\}^n\to\{0,1\}$ and every constant q,*

(i) $\mathsf{NP}(q)^{\mathsf{cc}}(f\circ g^n) \;=\; \mathsf{NP}(q)^{\mathsf{dt}}(f)\cdot\Theta(\log n)$

(ii) $\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}}(f\circ g^n) \;=\; \mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{dt}}(f)\cdot\Theta(\log n)$

where $g\colon[m]\times\{0,1\}^m\to\{0,1\}$ is the index gadget defined by $g(x,y)=y_x$ with $m\coloneqq n^{20}$.

Only part *(ii)* is needed for proving Theorem 3, but part *(i)* forms an ingredient in the proof of *(ii)* and is of independent interest.

The most closely related lifting theorem to Theorem 4 is the one for P^{NP} [10], corresponding to computations that make an unbounded number of adaptive queries to an NP oracle. In that paper, the overall idea was to approximately characterize P^{NP} complexity in terms of *decision lists* (DL), and then prove a lifting theorem directly for DLs. Briefly, a conjunction

DL (introduced by [31]) is a sequence of small-width conjunctions each with an associated output bit, and the output is determined by finding the first conjunction in the list that accepts the given input. A rectangle DL is similar but with combinatorial rectangles instead of conjunctions. The proof from [10] shows how to convert a rectangle DL for $f \circ g^n$ into a conjunction DL for f.

The gist of our arguments for both parts of Theorem 4 is to approximately characterize (via different techniques than for P^{NP}) these classes in terms of DLs with a bounded number of *alternations* (how many times the associated output bit flips as we walk down the entire DL). The DL lifting argument from [10] does not preserve the number of alternations, but we show how it can be adapted to do this. Our techniques also yield an approximate lifting theorem for $\mathsf{P}_{\|}^{\mathsf{NP}}$ (corresponding to computations that make an unbounded number of nonadaptive NP oracle queries), but we omit the details.

2 Preliminaries

We assume familiarity with deterministic computation in query and communication complexity [19, 24]. Recall the following standard definitions of nondeterministic and one-sided error randomized models:

- An $\mathsf{NP}^{\mathsf{dt}}$ decision tree is a DNF formula Φ. Given an input z, the output of such a decision tree is Φ evaluated on z. A function f is computed by Φ if $f(z) = \Phi(z)$ on all inputs z for which $f(z)$ is defined. The cost of Φ is the maximum width (number of literals) in any conjunction in Φ.
- An $\mathsf{NP}^{\mathsf{cc}}$ protocol is a set $\mathcal{R}$ of combinatorial rectangles. Given an input (x, y), the output of such a protocol is $\mathcal{R}(x, y) \coloneqq 1$ iff there exists an $R \in \mathcal{R}$ containing (x, y). A two-party function F is computed by $\mathcal{R}$ if $F(x, y) = \mathcal{R}(x, y)$ on all inputs (x, y) for which $F(x, y)$ is defined. The cost of $\mathcal{R}$ is $\lceil \log(|\mathcal{R}| + 1) \rceil$, which intuitively represents the number of bits required to specify a rectangle in $\mathcal{R}$ or indicate that the input is in no such rectangle.
- An $\mathsf{RP}^{\mathsf{dt}}$ decision tree is a distribution $\boldsymbol{T}$ over deterministic decision trees. Given an input z, the output of such a decision tree is computed by sampling a deterministic decision tree T from $\boldsymbol{T}$ and evaluating $T(z)$. A function f is computed by $\boldsymbol{T}$ if for all $z \in f^{-1}(0)$, $\Pr_{T \sim \boldsymbol{T}}[T(z) = 1] = 0$ and for all $z \in f^{-1}(1)$, $\Pr_{T \sim \boldsymbol{T}}[T(z) = 1] \geq 1/2$. The cost of $\boldsymbol{T}$ is the maximum number of bits queried in any T in its support.
- An $\mathsf{RP}^{\mathsf{cc}}$ protocol is a distribution $\boldsymbol{\Pi}$ over deterministic communication protocols. Given an input (x, y), the output of such a protocol is computed by sampling a deterministic protocol Π from $\boldsymbol{\Pi}$ and evaluating $\Pi(x, y)$. A function F is computed by $\boldsymbol{\Pi}$ if for all $(x, y) \in F^{-1}(0)$, $\Pr_{\Pi \sim \boldsymbol{\Pi}}[\Pi(x, y) = 1] = 0$ and for all $(x, y) \in F^{-1}(1)$, $\Pr_{\Pi \sim \boldsymbol{\Pi}}[\Pi(x, y) = 1] \geq 1/2$. The cost of $\boldsymbol{\Pi}$ is the maximum number of bits exchanged in any Π in its support.

Let $\mathcal{C}$ be an arbitrary complexity class name representing a model of computation (such as NP or RP). We let $\mathcal{C}^{\mathsf{cc}}(F)$ denote the communication complexity of a two-party function F in the corresponding model: the minimum cost of any $\mathcal{C}^{\mathsf{cc}}$ protocol computing F. We let $\mathcal{C}^{\mathsf{dt}}(f)$ denote the query complexity of a Boolean function f in the corresponding model: the minimum cost of any $\mathcal{C}^{\mathsf{dt}}$ decision tree computing f. Often we will abuse notation by having F or f refer to an infinite *family* of functions, where there is at most one function in the family for each possible input length. In this case, $\mathcal{C}^{\mathsf{cc}}(F)$ or $\mathcal{C}^{\mathsf{dt}}(f)$ will be the complexity parameterized by the input length n; we typically express this with asymptotic notation. When written by itself, $\mathcal{C}^{\mathsf{cc}}$ or $\mathcal{C}^{\mathsf{dt}}$ denotes the class of all families of functions with complexity at most polylogarithmic in n, in the corresponding model. We will always clarify whether a class $\mathcal{C}^{\mathsf{cc}}$ or $\mathcal{C}^{\mathsf{dt}}$ is meant to contain partial functions or just total functions, since this is not explicit in the notation.

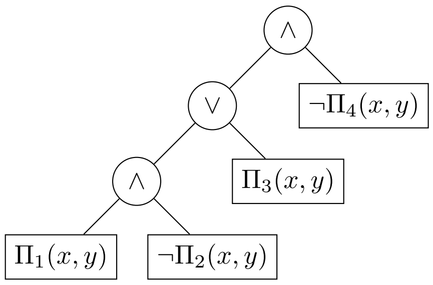

Figure 2 A visualization of a $\mathcal{C}(4)^{\mathsf{cc}}$ protocol, where each Π_i is a $\mathcal{C}^{\mathsf{cc}}$ protocol.

For $\mathsf{RP}^{\mathsf{cc}}$ and $\mathsf{RP}^{\mathsf{dt}}$, the constant $1/2$ in the success probability is arbitrary: by amplification, choosing a different positive constant in the definition would only affect the complexity of any function by a constant factor. Also note that $\mathsf{NP}^{\mathsf{dt}}(f) \leq \mathsf{RP}^{\mathsf{dt}}(f)$ for all f, and since we defined $\mathsf{RP}^{\mathsf{cc}}$ using the public-coin model, we have $\mathsf{NP}^{\mathsf{cc}}(F) \leq \mathsf{RP}^{\mathsf{cc}}(F) + O(\log n)$ for all F (by decreasing the number of random bits for sampling a protocol to $O(\log n)$ and using nondeterminism to guess an outcome that results in output 1).

2.1 Nondeterministic and Randomized Boolean Hierarchies

In the following definitions, restrict $\mathcal{C}$ to be either NP or RP. We will use two different but equivalent definitions of the constituent levels of the Nondeterministic and Randomized Boolean Hierarchies. Our "official" definition is in terms of the following "decision list functions" (also known as "odd-max-bit"):

▶ **Definition 5.** $\Delta_q \colon \{0,1\}^q \to \{0,1\}$ *is defined inductively as follows:*

- $\Delta_1(z_1) \coloneqq z_1$.
- *If q is odd,* $\Delta_q(z_1,\ldots,z_{q-1},z_q) \coloneqq \Delta_{q-1}(z_1,\ldots,z_{q-1}) \vee z_q$.
- *If q is even,* $\Delta_q(z_1,\ldots,z_{q-1},z_q) \coloneqq \Delta_{q-1}(z_1,\ldots,z_{q-1}) \wedge (\neg z_q)$.

In other words, letting $\oplus \colon \mathbb{N} \to \{0,1\}$ denote the parity function, we have $\Delta_q(z) \coloneqq \oplus(i)$ where i is the greatest index such that $z_i = 1$ (or $i = 0$ if z is all zeros).

▶ **Definition 6.** *A $\mathcal{C}(q)^{\mathsf{cc}}$ protocol is an ordered list of q many $\mathcal{C}^{\mathsf{cc}}$ protocols $\Pi = (\Pi_1,\ldots,\Pi_q)$. Given an input (x,y), the output of the protocol is $\Pi(x,y) \coloneqq \Delta_q(\Pi_1(x,y),\ldots,\Pi_q(x,y))$. The cost of a $\mathcal{C}(q)^{\mathsf{cc}}$ protocol is the sum of the costs of the component $\mathcal{C}^{\mathsf{cc}}$ protocols.*

See Figure 2 for an example of such a protocol. The Nondeterministic Boolean Hierarchy is $\bigcup_{\text{constant } q} \mathsf{NP}(q)^{\mathsf{cc}}$ and the Randomized Boolean Hierarchy is $\bigcup_{\text{constant } q} \mathsf{RP}(q)^{\mathsf{cc}}$. We are also interested in the complement classes $\mathsf{coNP}(q)^{\mathsf{cc}}$ and $\mathsf{coRP}(q)^{\mathsf{cc}}$. As is standard, when we write $\mathsf{co}\mathcal{C}(q)^{\mathsf{cc}}$ we refer to the class $\mathsf{co}(\mathcal{C}(q)^{\mathsf{cc}})$ (that is, functions that are the negations of functions in $\mathcal{C}(q)^{\mathsf{cc}}$) as opposed to $(\mathsf{co}\mathcal{C})(q)^{\mathsf{cc}}$ (that is, where the component protocols are $\mathsf{co}\mathcal{C}^{\mathsf{cc}}$ protocols).

There are analogous definitions in query complexity:

▶ **Definition 7.** *A $\mathcal{C}(q)^{\mathsf{dt}}$ decision tree is an ordered list of q many $\mathcal{C}^{\mathsf{dt}}$ decision trees $T = (T_1,\ldots,T_q)$. Given an input z, define the output as $T(z) \coloneqq \Delta_q(T_1(z),\ldots,T_q(z))$. The cost of a $\mathcal{C}(q)^{\mathsf{dt}}$ decision tree is the sum of the costs of the component $\mathcal{C}^{\mathsf{dt}}$ decision trees.*

Our alternative definition of the Nondeterministic and Randomized Boolean Hierarchies simply replaces Δ_q with the parity function $\oplus_q \colon \{0,1\}^q \to \{0,1\}$.

▶ **Lemma 8.** *For $\mathcal{C} \in \{\mathsf{NP}, \mathsf{RP}\}$, if the definitions of $\mathcal{C}(q)^{\mathsf{cc}}$ and $\mathcal{C}(q)^{\mathsf{dt}}$ are changed to use $\oplus_q$ in place of Δ_q, it only affects the complexity measures $\mathcal{C}(q)^{\mathsf{cc}}(F)$ and $\mathcal{C}(q)^{\mathsf{dt}}(f)$ by a constant factor (depending on q).*

Wagner [32] showed that these alternative characterizations are equivalent for the classical Nondeterministic Boolean Hierarchy, and Halstenberg and Reischuk [17] observed the same (up to a constant factor) in communication complexity. This latter proof uses only the property that $\mathsf{NP}^{\mathsf{cc}}$ is closed under intersection and union; that is, if $\mathsf{NP}^{\mathsf{cc}}(F_1), \mathsf{NP}^{\mathsf{cc}}(F_2) \leq k$, then $\mathsf{NP}^{\mathsf{cc}}(F_1 \wedge F_2)$ and $\mathsf{NP}^{\mathsf{cc}}(F_1 \vee F_2)$ are both $O(k)$. Observe that since this property also holds for $\mathsf{RP}^{\mathsf{cc}}$, $\mathsf{NP}^{\mathsf{dt}}$, and $\mathsf{RP}^{\mathsf{dt}}$, their proof works for these models of computation as well. In fact, in all of these models, the cost of the intersection or union of i cost-k computations is at most ik.

We use both definitions in this paper. We found that the Δ_q definition makes it easier to prove the lifting theorems, and the $\oplus_q$ definition makes it easier to prove concrete upper and lower bounds.

2.2 Parallel queries

In the following definitions, restrict $\mathcal{C}$ to be either NP or RP.

▶ **Definition 9.** *A $\mathsf{P}_{\|}^{\mathcal{C}[q]\mathsf{cc}}$ protocol consists of a deterministic protocol $\Pi_{\det}$ that maps an input (x, y) to two things: a function $\mathsf{out}\colon \{0,1\}^q \to \{0,1\}$ and an ordered list of q many $\mathcal{C}^{\mathsf{cc}}$ protocols $(\Pi_1, \ldots, \Pi_q)$. The output is then $\mathsf{out}(\Pi_1(x,y), \ldots, \Pi_q(x,y))$. The cost of a $\mathsf{P}_{\|}^{\mathcal{C}[q]\mathsf{cc}}$ protocol is the communication cost (depth) of $\Pi_{\det}$ plus the maximum over (x,y) of the sum of the costs of the $\mathcal{C}^{\mathsf{cc}}$ protocols produced by $\Pi_{\det}(x,y)$.*

▶ **Definition 10.** *A $\mathsf{P}_{\|}^{\mathcal{C}[q]\mathsf{dt}}$ decision tree consists of a deterministic decision tree $T_{\det}$ that maps an input z to two things: a function $\mathsf{out}\colon \{0,1\}^q \to \{0,1\}$ and an ordered list of q many $\mathcal{C}^{\mathsf{dt}}$ decision trees $(T_1, \ldots, T_q)$. The output is then $\mathsf{out}(T_1(z), \ldots, T_q(z))$. The cost of a $\mathsf{P}_{\|}^{\mathcal{C}[q]\mathsf{dt}}$ decision tree is the query cost (depth) of $T_{\det}$ plus the maximum over z of the sum of the costs of the $\mathcal{C}^{\mathsf{dt}}$ decision trees produced by $T_{\det}(z)$.*

The following lemma states that at each leaf of $\Pi_{\det}$ or $T_{\det}$, we can replace the q "$\mathcal{C}$ oracle queries" with one "$\mathcal{C}(q)$ oracle query" (where some leaves may output the oracle's answer, while other leaves output the negation of it). This was shown in classical time-bounded complexity using the so-called "mind-change argument" [5], and this argument can be translated directly to communication and query complexity. For example, [17] used this method to show that $\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}} \subseteq \mathsf{NP}(q+1)^{\mathsf{cc}} \cap \mathsf{coNP}(q+1)^{\mathsf{cc}}$. We will only need to use the result for $\mathcal{C} = \mathsf{NP}$.

▶ **Lemma 11.** *For $\mathcal{C} \in \{\mathsf{NP}, \mathsf{RP}\}$, we have $\mathsf{P}_{\|}^{\mathcal{C}[q]\mathsf{cc}}(F) = \Theta(\mathsf{P}^{\mathcal{C}(q)[1]\mathsf{cc}}(F))$ and $\mathsf{P}_{\|}^{\mathcal{C}[q]\mathsf{dt}}(f) = \Theta(\mathsf{P}^{\mathcal{C}(q)[1]\mathsf{dt}}(f))$ for every constant q.*

3 Separations

In this section we prove our separation results (Theorem 1 and Theorem 3).

3.1 Proof of Theorem 1

▶ **Theorem** (Restatement of Theorem 1). *For total functions, $\mathsf{coRP}(q)^{\mathsf{cc}} \not\subseteq \mathsf{NP}(q)^{\mathsf{cc}}$ for every constant q.*

Fix any constant q. Let $\oplus_q\textsc{NonEq}\colon (\{0,1\}^n)^q \times (\{0,1\}^n)^q \to \{0,1\}$ be the two-party total function where Alice's and Bob's inputs are divided into q blocks $x = (x_1, \dots, x_q)$ and $y = (y_1, \dots, y_q)$ with each $x_i, y_i \in \{0,1\}^n$, and which is defined by $\oplus_q\textsc{NonEq}(x,y) \coloneqq 1$ iff there are an odd number of blocks i such that $x_i \neq y_i$. Note that $\mathsf{RP}(q)^{\mathsf{cc}}(\oplus_q\textsc{NonEq}) = O(1)$ by Lemma 8 and the standard fact that $\mathsf{RP}^{\mathsf{cc}}(\textsc{NonEq}) = O(1)$. Thus $\overline{\oplus_q\textsc{NonEq}} \in \mathsf{coRP}(q)^{\mathsf{cc}}$. We will now prove that $\mathsf{NP}(q)^{\mathsf{cc}}(\overline{\oplus_q\textsc{NonEq}}) = \Omega(n)$.

Suppose for contradiction $\overline{\oplus_q\textsc{NonEq}}$ has an $\mathsf{NP}(q)^{\mathsf{cc}}$ protocol of cost $k \leq n/2^q$, say $\Pi = (\mathcal{R}_1, \dots, \mathcal{R}_q)$ where each $\mathcal{R}_i$ is a nonempty set of rectangles. By Lemma 8 we may assume the protocol outputs 1 iff the input is contained in an odd number of the rectangle unions $\bigcup_{R \in \mathcal{R}_i} R$ for $i \in [q]$, in other words, $\Pi(x,y) \coloneqq \oplus_q(\mathcal{R}_1(x,y), \dots, \mathcal{R}_q(x,y))$. Note that, assuming $\mathcal{R}_i$ has cost k_i, the total number of rectangles in these unions is at most $\sum_i |\mathcal{R}_i| \leq \prod_i (|\mathcal{R}_i| + 1) \leq \prod_i 2^{k_i} = 2^k$.

We will iteratively construct a sequence of rectangles Q^j for $j = 0, \dots, q$ such that *(i)* there are at least j many values of i for which $Q^j \subseteq \bigcup_{R \in \mathcal{R}_i} R$, and *(ii)* Q^j contains an input where exactly j blocks are unequal. We obtain the contradiction when $j = q$: by *(ii)* some input $(x,y) \in Q^q$ has $x_i \neq y_i$ for all i and thus $\overline{\oplus_q\textsc{NonEq}}(x,y) = 1 - \oplus(q)$, yet by *(i)* we have $\mathcal{R}_i(x,y) = 1$ for all i and thus $\Pi(x,y) = \oplus(q)$, contradicting the supposed correctness of Π.

We will actually maintain stronger invariants than the above *(i)* and *(ii)*: For *(i)*, we will actually have for some j values of i – we assume they are $1, \dots, j$ for notational convenience – some individual rectangle $R_i \in \mathcal{R}_i$ contains Q^j. For *(ii)*, Q^j will actually have the following form: for some fixed strings $a^j = (a_1, \dots, a_j) \in (\{0,1\}^n)^j$ and $b^j = (b_1, \dots, b_j) \in (\{0,1\}^n)^j$ such that $a_i \neq b_i$ for all $i \in [j]$, and for some nonempty set $S^j \subseteq (\{0,1\}^n)^{q-j}$, we have $Q^j \coloneqq \{a^j s : s \in S^j\} \times \{b^j s : s \in S^j\}$, which we abbreviate as $a^j S^j \times b^j S^j$. Defining a *diagonal* input in Q^j to be one of the form $(a^j s, b^j s)$ for any particular $s \in S^j$, we see that each diagonal input has exactly j unequal blocks, as needed for *(ii)*.

In fact, we will maintain not just that S^j is nonempty, but that it is sufficiently large. Specifically, the *deficiency* of S^j, defined as $\mathbf{D}_\infty(S^j) \coloneqq n(q-j) - \log|S^j|$, will be at most $(2^j - 1)(2k+1)$. At the end, since $(2^q - 1)(2k+1) < \infty$, this guarantees that S^q will contain at least one element from $(\{0,1\}^n)^{q-q}$. The latter set only has one element, namely the empty tuple, so this means Q^q will contain the single input (a^q, b^q), which has all blocks unequal.

We start with $S^0 = (\{0,1\}^n)^q$, which indeed has $\mathbf{D}_\infty(S^0) = 0 = (2^0 - 1)(2k+1)$, and thus Q^0 is the rectangle of all possible inputs. Now supposing we already have a^j, b^j, and S^j satisfying all the properties from the previous two paragraphs, we explain how to obtain $a_{j+1}, b_{j+1} \in \{0,1\}^n$ and $S^{j+1} \subseteq (\{0,1\}^n)^{q-(j+1)}$ so these properties again hold (with $a^{j+1} \coloneqq a^j a_{j+1}$ and $b^{j+1} \coloneqq b^j b_{j+1}$).

We first observe that each diagonal input in Q^j must be contained in at least one rectangle from $\mathcal{R}_{j+1} \cup \dots \cup \mathcal{R}_q$. This is because such an input (x,y) is already contained in the rectangles $R_1 \in \mathcal{R}_1, \dots, R_j \in \mathcal{R}_j$, and these cannot be the only values of i such that $\mathcal{R}_i(x,y) = 1$ since otherwise we would have $\Pi(x,y) = \oplus(j)$ while $\overline{\oplus_q\textsc{NonEq}}(x,y) = 1 - \oplus(j)$, contradicting the supposed correctness of Π. Now pick one of the (at most 2^k) rectangles from $\mathcal{R}_{j+1} \cup \dots \cup \mathcal{R}_q$ that contains the largest fraction (at least $1/2^k$) of diagonal inputs from Q^j, and assume this rectangle is $R_{j+1} \in \mathcal{R}_{j+1}$ for notational convenience. Defining $\tilde{S}^j \coloneqq \{s \in S^j : (a^j s, b^j s) \in R_{j+1}\}$, we see that $\mathbf{D}_\infty(\tilde{S}^j) \leq \mathbf{D}_\infty(S^j) + k \leq (2^j - 1)(2k+1) + k$.

Since R_{j+1} is a rectangle, it must in fact contain the entire rectangle $a^j\tilde{S}^j \times b^j\tilde{S}^j$. Since $a^j\tilde{S}^j \times b^j\tilde{S}^j \subseteq a^j S^j \times b^j S^j = Q^j$, by assumption it is also contained in each of $R_1, \dots, R_j$. In the end, we will ensure Q^{j+1} is a subrectangle of $a^j\tilde{S}^j \times b^j\tilde{S}^j$, which will maintain property *(i)*: Q^{j+1} is contained in each of $R_1, \dots, R_{j+1}$.

To maintain *(ii)*, we will find some $a_{j+1} \neq b_{j+1}$ and then define $S^{j+1} := \{s : a_{j+1}s \in \tilde{S}^j \text{ and } b_{j+1}s \in \tilde{S}^j\}$. Then $a_{j+1}S^{j+1} \subseteq \tilde{S}^j$ and $b_{j+1}S^{j+1} \subseteq \tilde{S}^j$ ensure that $Q^{j+1} := a^{j+1}S^{j+1} \times b^{j+1}S^{j+1}$ is indeed a subrectangle of $a^j\tilde{S}^j \times b^j\tilde{S}^j$, as we needed for *(i)*. The fact that this can be done with a not-too-small S^{j+1} is encapsulated in the following technical lemma, which we prove shortly:

▶ **Lemma 12.** *Consider any bipartite graph with left nodes U and right nodes V, and suppose $1 \geq \varepsilon \geq 2/|U|$. If an ε fraction of all possible edges are present in the graph, then there exist distinct nodes $u, u' \in U$ that have at least $(\varepsilon^2/2) \cdot |V|$ common neighbors.*

Specifically, take $U := \{0,1\}^n$ and $V := (\{0,1\}^n)^{q-(j+1)}$ (so $|V| = 1$ if $j = q-1$, but that is fine), put an edge between $u \in U$ and $v \in V$ iff $uv \in \tilde{S}^j$, and let $\varepsilon := |\tilde{S}^j|/2^{n(q-j)} = 1/2^{\mathbf{D}_\infty(\tilde{S}^j)}$. Notice that $\varepsilon \geq 2/|U|$ holds since $\mathbf{D}_\infty(\tilde{S}^j) \leq (2^j - 1)(2k+1) + k \leq 2^{j+1}k - 1 \leq n - 1$ follows from our assumption that $k \leq n/2^q$. Thus Lemma 12 guarantees we can pick strings $a_{j+1} \neq b_{j+1}$ (corresponding to the nodes u, u') such that S^{j+1} (the set of common neighbors) has size at least $(\varepsilon^2/2) \cdot 2^{n(q-(j+1))}$. Thus

$$\begin{aligned} \mathbf{D}_\infty(S^{j+1}) &:= n(q-(j+1)) - \log|S^{j+1}| \;\leq\; \log(2/\varepsilon^2) \;=\; 2\mathbf{D}_\infty(\tilde{S}^j) + 1 \\ &\leq 2\big((2^j-1)(2k+1) + k\big) + 1 \;=\; \big(2(2^j-1)+1\big)(2k+1) \\ &= (2^{j+1}-1)(2k+1) \end{aligned}$$

as we needed for *(ii)*. This finishes the proof of Theorem 1.

Proof of Lemma 12. Let d_u and d_v denote the degrees of nodes $u \in U$ and $v \in V$, and let $d_{u,u'}$ denote the number of common neighbors of $u, u' \in U$. Summing over ordered pairs u, u' of not-necessarily-distinct left nodes, we have

$$\sum_{u,u' \in U} d_{u,u'} \;=\; \sum_{v \in V} d_v^2 \;\geq\; \Big(\sum_{v \in V} d_v\Big)^2/|V| \;=\; \varepsilon^2 \cdot |U|^2 \cdot |V|$$

by Cauchy–Schwarz and the assumption $\sum_{v \in V} d_v = \varepsilon \cdot |U| \cdot |V|$. Now sampling u, u' independently uniformly at random from U, we have

$$\varepsilon^2 \cdot |V| \;\leq\; \mathop{\mathrm{E}}_{u,u'}[d_{u,u'}] \;\leq\; \mathop{\mathrm{E}}_{u,u'}[d_{u,u'} \mid u \neq u'] + \mathop{\mathrm{E}}_{u}[d_u] \cdot \mathop{\Pr}_{u,u'}[u = u']$$

(the conditioning is valid by the assumption $|U| \geq 2$). Since $\mathrm{E}_u[d_u] = \varepsilon \cdot |V|$ and $\Pr_{u,u'}[u = u'] = 1/|U|$, rearranging gives

$$\mathop{\mathrm{E}}_{u,u'}[d_{u,u'} \mid u \neq u'] \;\geq\; \varepsilon^2 \cdot |V| - \varepsilon \cdot |V|/|U| \;\geq\; (\varepsilon^2/2) \cdot |V|$$

where the last inequality holds by the assumption $1/|U| \leq \varepsilon/2$. Thus there must be some $u \neq u'$ such that $d_{u,u'}$ is at least this large. ◀

3.2 Proof of Theorem 3

▶ **Theorem** (Restatement of Theorem 3). *For partial functions,* $\mathsf{RP}(q+1)^{\mathsf{cc}} \cap \mathsf{coRP}(q+1)^{\mathsf{cc}} \not\subseteq \mathsf{P}_{\parallel}^{\mathsf{NP}[q]\mathsf{cc}}$ *for every constant q.*

To prove this result, we require the query complexity separation $\mathsf{coRP}(q)^{\mathsf{dt}} \not\subseteq \mathsf{NP}(q)^{\mathsf{dt}}$.

▶ **Definition 13.** *Fix any constant q. Let $\oplus_q\textsc{GapOr}\colon (\{0,1\}^n)^q \to \{0,1\}$ be the partial function where the input is divided into q blocks $z = (z_1, \ldots, z_q)$ having the promise that each $z_i \in \{0,1\}^n$ is either all zeros or at least half ones (call such an input* valid*), and which is defined by $\oplus_q\textsc{GapOr}(z) := 1$ iff an odd number of blocks i are such that z_i is nonzero.*

Note that $\mathsf{RP}(q)^{\mathsf{dt}}(\oplus_q\textsc{GapOr}) = O(1)$ by Lemma 8 and the fact that $\mathsf{RP}^{\mathsf{dt}}(\textsc{GapOr}) = 1$. The full version of this paper [27] contains a proof of the following lemma:

▶ **Lemma 14.** $\mathsf{NP}(q)^{\mathsf{dt}}(\overline{\oplus_q\textsc{GapOr}}) = \Omega(n)$.

We now proceed to the proof of Theorem 3.

Fix any constant q. Let $\textsc{Which}\oplus_{q+1}\textsc{GapOr}\colon (\{0,1\}^{2n})^{2(q+1)} \to \{0,1\}$ be the following partial function: The input is divided into two halves $z = (z^0, z^1)$, and each half is divided into $q+1$ blocks $z^h = (z^h_1, \dots, z^h_{q+1})$ having the promise that each $z^h_i \in \{0,1\}^{2n}$ is either all zeros or at least a *quarter* ones, and moreover it is promised that the number of nonzero blocks in z^0 has the opposite parity as the number of nonzero blocks in z^1 (call such an input *valid*). The partial function is defined by

$$\textsc{Which}\oplus_{q+1}\textsc{GapOr}(z) = \begin{cases} 1 & \text{if the number of nonzero blocks is odd in } z^0 \text{ and even in } z^1 \\ 0 & \text{if the number of nonzero blocks is even in } z^0 \text{ and odd in } z^1 \end{cases}$$

We henceforth abbreviate $\textsc{Which}\oplus_{q+1}\textsc{GapOr}$ as f. Note that $\mathsf{RP}(q+1)^{\mathsf{dt}}(f) = O(1)$ by applying the $\mathsf{RP}(q+1)^{\mathsf{dt}}$ decision tree for $\oplus_{q+1}\textsc{GapOr}$ on z^0 (adapted for the different block length and different threshold for fraction of ones in a block). By symmetry (focusing on z^1), we also have $\mathsf{RP}(q+1)^{\mathsf{dt}}(\overline{f}) = O(1)$. Letting $g\colon [m] \times \{0,1\}^m \to \{0,1\}$ be the index gadget with $m := N^{20}$ where $N := 4(q+1)n$, this implies that

$$\mathsf{RP}(q+1)^{\mathsf{cc}}(f \circ g^N) = O(\log n) \qquad \text{and} \qquad \mathsf{coRP}(q+1)^{\mathsf{cc}}(f \circ g^N) = O(\log n)$$

(by the "easy direction" of $\mathsf{RP}(q+1)$ lifting) and thus $f \circ g^N \in \mathsf{RP}(q+1)^{\mathsf{cc}} \cap \mathsf{coRP}(q+1)^{\mathsf{cc}}$. We will now prove that $\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{dt}}(f) = \Omega(n)$, which by Theorem 4.*(ii)* implies that $\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}}(f \circ g^N) = \Omega(n \log n)$.

We show this by reduction from Lemma 14. We henceforth abbreviate $\overline{\oplus_q\textsc{GapOr}}$ as f'. Supposing f has a $\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{dt}}$ decision tree T of cost $k \le n/2$, say with deterministic phase T_{det}, we will use it to construct an $\mathsf{NP}(q)^{\mathsf{dt}}$ decision tree T' of cost at most k for f'.

By Lemma 11 we may assume that each leaf of T_{det} produces a single $\mathsf{NP}(q)^{\mathsf{dt}}$ decision tree and chooses whether to output the same or opposite answer as that decision tree. Follow the root-to-leaf path in T_{det} where all queries are answered with zero. Let $\rho \in (\{0,*\}^{2n})^{2(q+1)}$ be the partial assignment with at most $k \le n/2$ zeros that records these queries (so an input leads to this leaf iff it is consistent with ρ). Let $T_{\mathrm{leaf}} = (\Phi_1, \dots, \Phi_q)$ be the $\mathsf{NP}(q)^{\mathsf{dt}}$ decision tree of cost at most k produced at this leaf, where each Φ_i is a DNF. By symmetry, we assume (without loss of generality) that this leaf chooses to output the same answer as T_{leaf}.

Given any valid input z' to f', we show how to map it to a valid input z to f such that *(i)* $f'(z') = f(z)$, *(ii)* z is consistent with ρ, and *(iii)* each bit of z either is fixed or is some preselected bit of z'. Since *(iii)* implies that $T_{\mathrm{leaf}}(z)$ can be viewed as an $\mathsf{NP}(q)^{\mathsf{dt}}$ decision tree $T'(z')$ by substituting a constant or variable of z' in for each variable of z (which does not increase the width of any conjunction), and T' would correctly compute f' since

$$f'(z') \;=\; f(z) \;=\; T(z) \;=\; T_{\mathrm{leaf}}(z) \;=\; T'(z')$$

by *(i)*, correctness of T, *(ii)*, and *(iii)* respectively, this would show that $\mathsf{NP}(q)^{\mathsf{dt}}(f') \le k \le n/2$, which we know is false from Lemma 14.

To define z, we start with ρ (that is, we place zeros everywhere ρ requires, thus ensuring *(ii)*). Since ρ has at most n zeros in each block (indeed, at most $n/2$ zeros total), we can then place more zeros in such a way that each block now has exactly n zeros and n stars.

Next we replace the stars in z^0_{q+1} with ones and replace the stars in z^1_{q+1} with zeros. Finally, for each $i \in [q]$, we fill in the stars of z^0_i with a copy of z'_i and fill in the stars of z^1_i with another copy of z'_i. This construction satisfies *(iii)*.

To verify *(i)*, first observe that since each block of z' is either all zeros or at least half $(n/2)$ ones, this ensures each block of z is either all zeros or at least a quarter $(2n/4)$ ones. Furthermore, if z' has exactly ℓ nonzero blocks then the number of nonzero blocks is $\ell + 1$ in z^0 (since z^0_{q+1} is nonzero) and ℓ in z^1 (since z^1_{q+1} is all zeros). Hence if $f'(z') = 1$ (ℓ is even) then $f(z) = 1$ (since $\ell + 1$ is odd and ℓ is even), and if $f'(z') = 0$ (ℓ is odd) then $f(z) = 0$ (since $\ell + 1$ is even and ℓ is odd). Thus $f'(z') = f(z)$, and this finishes the proof of Theorem 3.

4 Total function collapse

▶ **Theorem** (Restatement of Theorem 2). *For total functions,*

$$\mathsf{P}_{\|}^{\mathsf{NP}[q]\mathsf{cc}} = \mathsf{NP}(q+1)^{\mathsf{cc}} \cap \mathsf{coNP}(q+1)^{\mathsf{cc}} \qquad \textit{and} \qquad \mathsf{P}_{\|}^{\mathsf{RP}[q]\mathsf{cc}} = \mathsf{RP}(q+1)^{\mathsf{cc}} \cap \mathsf{coRP}(q+1)^{\mathsf{cc}}$$

for every constant q.

In this section we will present intuition for the proof of the nondeterministic case of Theorem 2. For the complete proof, see the full version of this paper [27].

This proof is similar to the argument for the $q = 0$ case (that is, for total functions, $\mathsf{P}^{\mathsf{cc}} = \mathsf{NP}^{\mathsf{cc}} \cap \mathsf{coNP}^{\mathsf{cc}}$). In that proof, Alice and Bob can use the fact that the rectangles in the $\mathsf{NP}^{\mathsf{cc}}$ protocol's 1-monochromatic covering of F are disjoint from the rectangles in the $\mathsf{coNP}^{\mathsf{cc}}$ protocol's 0-monochromatic covering. Specifically, if $F(x, y) = 1$, then (x, y) is in some 1-rectangle, which is row-disjoint or column-disjoint from each 0-rectangle. (If a 1-rectangle and 0-rectangle shared a row and a column, they would intersect, which is not possible for a total function.) Therefore, Alice and Bob can repeatedly eliminate from consideration at least half of the remaining 0-rectangles, by identifying a 1-rectangle that either has x in its row set but is row-disjoint from at least half the remaining 0-rectangles, or has y in its column set but is column-disjoint from at least half the remaining 0-rectangles. If (x, y) is indeed in a 1-rectangle, then this process can always continue until there are no 0-rectangles left. If (x, y) is in a 0-rectangle, then this process will never eliminate that rectangle, so the process will halt with a nonempty set of 0-rectangles.

We repeat a similar argument, but using the "top level" of the $\mathsf{NP}(q+1)^{\mathsf{cc}}$ and the $\mathsf{coNP}(q+1)^{\mathsf{cc}}$ protocols for F as our monochromatic rectangle sets. Here we think of a $\mathsf{coNP}(q+1)^{\mathsf{cc}}$ protocol as computing F by applying $\overline{\Delta}_{q+1}$ (with negations pushed to the leaves) to the indicators for $q + 1$ rectangle unions. Depending on the parity of q, the rectangle union $\mathcal{R}_{q+1}$ queried at depth 1 of the $\mathsf{NP}(q+1)^{\mathsf{cc}}$ protocol will correspond to either 1-monochromatic rectangles or 0-monochromatic rectangles for F. The rectangle union $\mathcal{R}'_{q+1}$ queried at depth 1 of the $\mathsf{coNP}(q+1)^{\mathsf{cc}}$ protocol will be the opposite color of monochromatic rectangles. Crucially, this means that no input is in a rectangle from both of these sets (as we are assuming F is total). See Figure 3 for an illustration.

A key observation is that a deterministic protocol similar to the one used in the $q = 0$ case, ran using these top-level rectangle sets, will return the correct answer *under the promise that (x, y) is in one of these rectangles.* Say, for example, that (x, y) is in some rectangle in the 1-monochromatic top-level set. Then the deterministic protocol will successfully eliminate all 0-rectangles from the other top-level set, and will announce that the answer is 1. If (x, y) was in one of the 0-rectangles, then that rectangle will never be eliminated, and so the protocol would announce that the answer is 0.

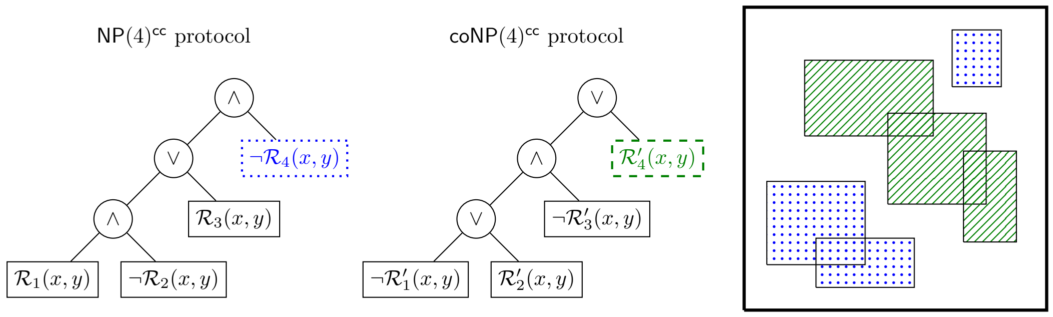

Figure 3 If a total function has an $\mathsf{NP}(4)^{\mathsf{cc}}$ protocol and a $\mathsf{coNP}(4)^{\mathsf{cc}}$ protocol, then the rectangle unions from the $\mathsf{NP}^{\mathsf{cc}}$ functions at depth one of each protocol are disjoint.

If (x, y) is in the top-level rectangle union for one of the protocols, then (x, y) is not in the top-level rectangle union of the *other* protocol, so $F(x, y)$ can be computed by the other protocol but where the top level is skipped (resulting in only q many $\mathsf{NP}^{\mathsf{cc}}$ oracle queries). This boils down to the observation that $\Delta_{q+1}(z_1, \ldots, z_q, 0) = \Delta_q(z_1, \ldots, z_q)$.

What if (x, y) is in neither top-level rectangle union? Then we can make no guarantees about the behavior of the deterministic protocol – it might answer 0 or 1 (which we interpret as merely a "guess" for $F(x, y)$). However, in this case *both* protocols correctly compute $F(x, y)$ even if the top level is skipped. Therefore, we will still get the correct answer no matter which guess is produced by the deterministic protocol.

5 Query-to-communication lifting theorems

▶ **Theorem** (Restatement of Theorem 4.). *For every partial function $f : \{0,1\}^n \to \{0,1\}$ and every constant q,*

(i) $\mathsf{NP}(q)^{\mathsf{cc}}(f \circ g^n) = \mathsf{NP}(q)^{\mathsf{dt}}(f) \cdot \Theta(\log n)$

(ii) $\mathsf{P}_{\parallel}^{\mathsf{NP}[q]\mathsf{cc}}(f \circ g^n) = \mathsf{P}_{\parallel}^{\mathsf{NP}[q]\mathsf{dt}}(f) \cdot \Theta(\log n)$

where $g : [m] \times \{0,1\}^m \to \{0,1\}$ is the index gadget defined by $g(x, y) = y_x$ with $m := n^{20}$.

In this section we present the proof for Theorem 4.*(i)*, as well as the necessary background. The full version of this paper [27] contains the proof of Theorem 4.*(ii)*. The high-level idea of the latter proof is to convert a $\mathsf{P}^{\mathsf{NP}(q)[1]\mathsf{cc}}$ protocol for $f \circ g^n$ into a $\mathsf{P}^{\mathsf{NP}(q)[1]\mathsf{dt}}$ decision tree for f by using the P lifting theorem of Raz and McKenzie [28, 14] to handle the deterministic phase, followed by our $\mathsf{NP}(q)$ lifting theorem to handle the single $\mathsf{NP}(q)$ oracle query.

5.1 Decision lists

The reason we call Δ_q a "decision list function" is that it highlights the connection between the Boolean Hierarchy classes and the *decision list* models of computation:

▶ **Definition 15.** *A rectangle decision list $\mathcal{L}_{\mathrm{R}}$ is an ordered list of pairs $(R_1, \ell_1), (R_2, \ell_2), \ldots$ where each R_i is a combinatorial rectangle, $\ell_i \in \{0,1\}$ is a label, and the final rectangle in the list contains all inputs in the domain. For an input (x, y), the output $\mathcal{L}_{\mathrm{R}}(x, y)$ is ℓ_i where i is the first index for which $(x, y) \in R_i$. The cost of $\mathcal{L}_{\mathrm{R}}$ is the* log *of the length of the list.*

▶ **Definition 16.** *A conjunction decision list $\mathcal{L}_{\mathrm{C}}$ is an ordered list of pairs $(C_1, \ell_1), (C_2, \ell_2), \ldots$ where each C_i is a conjunction, $\ell_i \in \{0,1\}$ is a label, and the final conjunction in the list accepts all inputs in the domain. For an input z, the output $\mathcal{L}_{\mathrm{C}}(z)$ is ℓ_i where i is the first index for which $C_i(z) = 1$. The cost of $\mathcal{L}_{\mathrm{C}}$ is the maximum width of any conjunction in the list.*

Note that the restriction on the final rectangle/conjunction is without loss of generality. The complexity measures $\mathsf{DL}^{\mathsf{cc}}(F)$ and $\mathsf{DL}^{\mathsf{dt}}(f)$ are the minimum cost of any rectangle/conjunction decision list computing F or f, and the classes $\mathsf{DL}^{\mathsf{cc}}$ and $\mathsf{DL}^{\mathsf{dt}}$ contain those functions with complexity at most polylog(n).

We now define *q-alternating decision lists* to have the additional restriction that the sequence of output labels $\ell_1, \ell_2, \ldots$ only flips between 0 and 1 at most q times, and furthermore the last label is 0. This restriction partitions the list into contiguous *levels* where all labels in the same level are equal; without loss of generality the last level consists only of the final "catch-all" entry. For convenience, in the list entries we replace the labels with the level numbers themselves.

▶ **Definition 17.** *A q-alternating rectangle decision list $\mathcal{L}_{\mathrm{R}}$ is an ordered list of pairs $(R_1, \ell_1), (R_2, \ell_2), \ldots$ where each R_i is a combinatorial rectangle, $\ell_i \in \{0, 1, \ldots, q\}$ is a level such that $\ell_i \geq \ell_{i+1}$ for all i, and the final rectangle in the list contains all inputs in the domain and is the only rectangle at level 0. For an input (x, y), the output $\mathcal{L}_{\mathrm{R}}(x, y)$ is $\oplus(\ell_i)$ where i is the first index for which $(x, y) \in R_i$. The cost of $\mathcal{L}_{\mathrm{R}}$ is the* log *of the length of the list.*

▶ **Definition 18.** *A q-alternating conjunction decision list $\mathcal{L}_{\mathrm{C}}$ is an ordered list of pairs $(C_1, \ell_1), (C_2, \ell_2), \ldots$ where each C_i is a conjunction, $\ell_i \in \{0, 1, \ldots, q\}$ is a level such that $\ell_i \geq \ell_{i+1}$ for all i, and the final conjunction in the list accepts all inputs in the domain and is the only conjunction at level 0. For an input z, the output $\mathcal{L}_{\mathrm{C}}(z)$ is $\oplus(\ell_i)$ where i is the first index for which $C_i(z) = 1$. The cost of $\mathcal{L}_{\mathrm{C}}$ is the maximum width of any conjunction in the list.*

The complexity measures $\mathsf{DL}(q)^{\mathsf{cc}}(F)$ and $\mathsf{DL}(q)^{\mathsf{dt}}(f)$ are the minimum cost of any q-alternating rectangle/conjunction decision list computing F or f, and the classes $\mathsf{DL}(q)^{\mathsf{cc}}$ and $\mathsf{DL}(q)^{\mathsf{dt}}$ contain those functions with complexity at most polylog(n).

It turns out that q-alternating decision lists are equivalent to $\mathsf{NP}(q)$ in both communication and query complexity. As this follows almost immediately from the definition of Δ_q, we omit the proof here.

▶ **Lemma 19.** $\mathsf{DL}(q)^{\mathsf{cc}}(F) = \Theta(\mathsf{NP}(q)^{\mathsf{cc}}(F))$ *and* $\mathsf{DL}(q)^{\mathsf{dt}}(f) = \Theta(\mathsf{NP}(q)^{\mathsf{dt}}(f))$ *for every constant q. Thus,* $\mathsf{DL}(q)^{\mathsf{cc}} = \mathsf{NP}(q)^{\mathsf{cc}}$ *and* $\mathsf{DL}(q)^{\mathsf{dt}} = \mathsf{NP}(q)^{\mathsf{dt}}$ *for partial functions.*

This can be contrasted with the lemma from [10] stating that $\mathsf{DL}^{\mathsf{cc}} = \mathsf{P}^{\mathsf{NPcc}}$ and $\mathsf{DL}^{\mathsf{dt}} = \mathsf{P}^{\mathsf{NPdt}}$ for partial functions.

5.2 Query-to-communication lifting for $\mathsf{NP}(q)$

The big-O direction of Theorem 4.*(i)* follows immediately from the same fact for NP: for every f, $\mathsf{NP}^{\mathsf{cc}}(f \circ g^n) = \mathsf{NP}^{\mathsf{dt}}(f) \cdot O(\log n)$ holds by replacing each of the $n^{O(k)}$ conjunctions in a width-k DNF with m^k rectangles (each of which contains inputs where the gadget outputs satisfy the conjunction), for a total of $n^{O(k)} m^k = 2^{k \cdot O(\log n)}$ rectangles. In the rest of this section we prove the big-Ω direction. By Lemma 19 it suffices to show

$$\mathsf{DL}(q)^{\mathsf{cc}}(f \circ g^n) = \mathsf{DL}(q)^{\mathsf{dt}}(f) \cdot \Omega(\log n).$$

5.2.1 Technical preliminaries

Our proof is closely related to the P^{NP} lifting theorem of Göös, Kamath, Pitassi, and Watson [10], so we start by recalling some definitions and lemmas that were used in that work. We will need to tweak some of the statements and parameters, though.

Define $G\colon [m]^n \times (\{0,1\}^m)^n \to \{0,1\}^n$ as $G \coloneqq g^n$. This partitions the input domain into 2^n slices $G^{-1}(z) = \{(x,y) : g(x_i, y_i) = z_i \text{ for all } i \in [n]\}$, one for each $z \in \{0,1\}^n$. For a set $Z \subseteq \{0,1\}^n$, let $G^{-1}(Z) \coloneqq \bigcup_{z\in Z} G^{-1}(z)$. Consider sets $A \subseteq [m]^n$ and $B \subseteq (\{0,1\}^m)^n$. For $I \subseteq [n]$, we let $A_I \coloneqq \{x_I : x \in A\}$ and $B_I \coloneqq \{y_I : y \in B\}$ be the *projections* onto the coordinates of I. The *min-entropy* of a random variable $\boldsymbol{x}$ is $\mathbf{H}_\infty(\boldsymbol{x}) \coloneqq \min_x \log(1/\Pr[\boldsymbol{x} = x])$. We say A is *δ-dense* if the uniform random variable $\boldsymbol{x}$ over A satisfies the following: for every nonempty $I \subseteq [n]$, $\mathbf{H}_\infty(\boldsymbol{x}_I) \geq \delta |I| \log m$ (that is, the min-entropy of the marginal distribution of $\boldsymbol{x}$ on coordinates I is at least a δ fraction of the maximum possible for a distribution over $[m]^I$). The *deficiency* of B is $\mathbf{D}_\infty(B) \coloneqq mn - \log|B|$.

▶ **Lemma 20** ([10, Lemma 11]). *If $A \subseteq [m]^n$ is 0.8-dense and $B \subseteq (\{0,1\}^m)^n$ has deficiency at most n^4, then $G(A \times B) = \{0,1\}^n$, that is, for every $z \in \{0,1\}^n$ there are $x \in A$ and $y \in B$ with $G(x,y) = z$.*

Here the density parameter is $\delta = 0.8$ and the deficiency is $\mathbf{D}_\infty(B) \leq n^4$, instead of $\delta = 0.9$ and $\mathbf{D}_\infty(B) \leq n^2$ as in the original. Lemma 20 still holds because our gadget size has increased: we use $m \coloneqq n^{20}$, whereas [10] used $m \coloneqq n^4$. This can be verified by a simple substitution in the proof.

The next lemma is altered somewhat from the original. For a proof, see the full version [27].

▶ **Lemma 21** (A more general version of [10, Claim 12]). *Let $\mathcal{X} \subseteq [m]^n$ be 0.85-dense. If $A' \subseteq \mathcal{X}$ satisfies $|A'| \geq |\mathcal{X}|/2^{k+1}$ then there exist an $I \subseteq [n]$ of size $|I| < 20(k+1)/\log m$ and an $A \subseteq A'$ such that A is fixed on coordinates I and 0.8-dense on all other coordinates.*

5.2.2 The simulation

We exhibit an algorithm that takes a q-alternating rectangle decision list $\mathcal{L}_\mathrm{R}$ for $f \circ g^n$ of cost k, and converts it to a q-alternating conjunction decision list $\mathcal{L}_\mathrm{C}$ for f of cost $O(k/\log n)$. The argument from [10] does exactly this except without preserving the bound on the number of alternations. In [10] the argument is formulated using a "dual" characterization of $\mathsf{DL}^{\mathsf{dt}}$, but it has the effect of building $\mathcal{L}_\mathrm{C}$ in order, obtaining each conjunction by "extracting" it from one of the rectangles in $\mathcal{L}_\mathrm{R}$. The trouble is that the rectangles are not necessarily "extracted from" in order: after extracting a conjunction from some rectangle, the *next* conjunction that gets put in $\mathcal{L}_\mathrm{C}$ may be extracted from a rectangle that is *earlier* in $\mathcal{L}_\mathrm{R}$. Thus $\mathcal{L}_\mathrm{C}$ may end up with more alternations than $\mathcal{L}_\mathrm{R}$.

To fix this, we convert the argument to a "primal" form and argue that it still works when we force the rectangles to be extracted from in order. The high-level view is that we iterate through the rectangles of $\mathcal{L}_\mathrm{R}$ in order, and for each we extract as many conjunctions as we can until the rectangle becomes "exhausted", at which time we remove the remaining "error" portion of the rectangle (by deleting few rows and columns) and move on to the next rectangle. With this modification, the rest of the technical details from [10] continue to work, and it now preserves the number of alternations.

At any step of this process, we let $X \times Y$ be the remaining rows and columns (after having removed the error portion of all previous rectangles in $\mathcal{L}_\mathrm{R}$), and we let $Z \subseteq \{0,1\}^n$ be the remaining inputs to f (which have not been accepted by any previous conjunctions we put in $\mathcal{L}_\mathrm{C}$). Suppose (R_i, ℓ_i) is our current entry in $\mathcal{L}_\mathrm{R}$. The goal is to find a subrectangle $A \times B \subseteq R_i \cap (X \times Y)$ that is "conjunction-like" in the sense that $G(A \times B)$ is exactly the inputs accepted by some small-width conjunction C, and such that among all remaining inputs $z \in Z$, C only accepts those with $f(z) = \oplus(\ell_i)$. These properties would ensure it is safe to put (C, ℓ_i) next in $\mathcal{L}_\mathrm{C}$.

Algorithm 1 Simulation algorithm.

```
In: $\mathcal{L}_\mathrm{R} = (R_1, \ell_1), \ldots, (R_{2^k}, \ell_{2^k})$ and $\mathcal{X} \subseteq [m]^n$, $\mathcal{Y} \subseteq (\{0,1\}^m)^n$
Out: $\mathcal{L}_\mathrm{C}$
1: initialize $X \leftarrow \mathcal{X}$, $Y \leftarrow \mathcal{Y}$, $Z \leftarrow$ domain of $f$, $\mathcal{L}_\mathrm{C} \leftarrow$ empty list
2: for $i = 1$ to $2^k$ do
3:     while $Z \neq \emptyset$ do
4:         for each $x \in X$, let $Y_x := \{y \in Y : (x,y) \in R_i \cap G^{-1}(Z)\}$
5:         let $A' := \{x \in X : |Y_x| \geq 2^{mn-n^4}\}$
6:         if $|A'| \leq |\mathcal{X}|/2^{k+1}$ then
7:             update $X \leftarrow X \smallsetminus A'$
8:             update $Y \leftarrow Y \smallsetminus \bigcup_{x \in X \smallsetminus A'} Y_x$
9:             break out of inner loop
10:        else $|A'| > |\mathcal{X}|/2^{k+1}$
11:            let $A \subseteq A'$, $I \subseteq [n]$, $\alpha \in [m]^I$ be such that:
12:                $|I| = O(k/\log n)$, $A_I$ is fixed to $\alpha$, and $A_{[n] \smallsetminus I}$ is 0.8-dense
13:            pick any $x' \in A$ and choose $\beta \in (\{0,1\}^m)^I$ to maximize $|\{y \in Y_{x'} : y_I = \beta\}|$
14:            let $C$ be the conjunction "$z_I = g^I(\alpha, \beta)$"
15:            update $\mathcal{L}_\mathrm{C}$ by appending $(C, \ell_i)$ to it
16:            update $Z \leftarrow Z \smallsetminus C^{-1}(1)$
```

Combining Lemma 20 and Lemma 21 (using $\mathcal{X} = [m]^n$) suggests an approach for finding a conjunction-like subrectangle: If A' is not too small, we can restrict it to A that is fixed on few coordinates I and dense on the rest (by Lemma 21). If B is also not too small (low deficiency) and fixed on coordinates I, then $G(A \times B)$ is fixed on I and takes on all possible values on the remaining coordinates (by Lemma 20, which still works with $[n] \smallsetminus I$ in place of $[n]$). In other words, $G(A \times B) = C^{-1}(1)$ for a small-width conjunction C, as desired.

The other property we needed to ensure is that if this C is the first conjunction in $\mathcal{L}_\mathrm{C}$ to accept a particular z, then $f(z) = \oplus(\ell_i)$ (so $\mathcal{L}_\mathrm{C}$ is correct). This will follow if we know there is some $(x, y) \in G^{-1}(z)$ such that R_i is the first rectangle in $\mathcal{L}_\mathrm{R}$ to contain (x, y), as that guarantees $f(z) = f(G(x,y)) = (f \circ g^n)(x,y) = \oplus(\ell_i)$. It turns out this will hold automatically if $A \times B \subseteq R_i \cap (X \times Y)$, because $A \times B$ touches the slice of every z that is accepted by C, and all inputs $(x, y) \in G^{-1}(Z)$ that were in some R_j with $j < i$ have already been removed from $X \times Y$.

Our algorithm for building $\mathcal{L}_\mathrm{C}$ from $\mathcal{L}_\mathrm{R}$ is shown in Algorithm 1. It is described as starting from some arbitrary initial rectangle $\mathcal{X} \times \mathcal{Y}$. For the purpose of proving Theorem 4.*(i)* we only need to take $\mathcal{X} = [m]^n$ and $\mathcal{Y} = (\{0,1\}^m)^n$, but when we invoke this as a component in the proof of Theorem 4.*(ii)* we will need to start from some $\mathcal{X} \times \mathcal{Y}$ that is merely "dense $\times$ large" rather than the full input domain, so we state this more general version now.

▶ **Lemma 22.** *If $\mathcal{L}_\mathrm{R}$ computes $f \circ g^n$ on $\mathcal{X} \times \mathcal{Y}$ and has cost k, and if $\mathcal{X}$ is 0.85-dense and $\mathbf{D}_\infty(\mathcal{Y}) \leq n^3$, then $\mathcal{L}_\mathrm{C}$ produced by Algorithm 1 computes f and has cost $O(k/\log n)$. Moreover, if $\mathcal{L}_\mathrm{R}$ is q-alternating then so is $\mathcal{L}_\mathrm{C}$.*

Proof. To verify the cost, just note that lines 11 and 12 always succeed by Lemma 21 (since $\mathcal{X}$ is 0.85-dense and $|A'| \geq |\mathcal{X}|/2^{k+1}$), so when a conjunction is added to $\mathcal{L}_\mathrm{C}$ on lines 14 and 15, it has width $|I| < 20(k+1)/\log m = O(k/\log n)$. On line 13, defining $B := \{y \in Y_{x'} : y_I = \beta\}$ we have $|B| \geq |Y_{x'}|/2^{m|I|} \geq 2^{mn-n^4-m|I|} = 2^{m(n-|I|)-n^4}$ (since $x' \in A \subseteq A'$) and therefore

$\mathbf{D}_\infty(B_{[n]\smallsetminus I}) \leq n^4$ (relative to $(\{0,1\}^m)^{[n]\smallsetminus I}$). Thus by applying Lemma 20 to $A_{[n]\smallsetminus I}$ (which is 0.8-dense) and $B_{[n]\smallsetminus I}$ we have $g^{[n]\smallsetminus I}(A_{[n]\smallsetminus I} \times B_{[n]\smallsetminus I}) = \{0,1\}^{n-|I|}$ and therefore $G(A \times B) = C^{-1}(1)$. (Lemma 20 works with the same parameters even though the sets are now on fewer than n coordinates.)

The algorithm terminates because Z always shrinks on line 16: for any $y \in B$ we have $G(x', y) \in Z$ (from the definition of $Y_{x'}$) and $C(G(x', y)) = 1$ (since $x'_I = \alpha$ and $y_I = \beta$ and thus $G(x', y)_I = g^I(\alpha, \beta)$).

The algorithm maintains the invariant that for all $j < i$, $R_j \cap (X \times Y) \cap G^{-1}(Z) = \emptyset$. This vacuously holds at the beginning, and is clearly maintained in the **else** case because i stays the same and nothing gets added to X, Y, or Z. Lines 7 and 8 maintain the invariant in the **if** case because the removed rows and columns cover all of $R_i \cap (X \times Y) \cap G^{-1}(Z)$ and i goes up by 1.

Next we argue that when the algorithm terminates, Z must be empty. In each iteration of the outer loop, we throw out at most $|\mathcal{X}|/2^{k+1}$ rows and at most $|\mathcal{X}| \cdot 2^{mn-n^4} \leq m^n \cdot 2^{mn-n^4} \leq 2^{mn-n^3}/2^{k+1} \leq |\mathcal{Y}|/2^{k+1}$ columns. (We throw out columns in Y_x for $x \notin A'$, all of these Y_x had the property $|Y_x| < 2^{mn-n^4}$, we do this for at most $|\mathcal{X}|$ values of x, and $n^4 - n \log m \geq n^3 + k + 1$.) Since the outer loop executes 2^k times, by the end at most half the rows of $\mathcal{X}$ and half the columns of $\mathcal{Y}$ have been discarded, so $|X| \geq |\mathcal{X}|/2$ and $|Y| \geq |\mathcal{Y}|/2$. This means X is essentially as dense as $\mathcal{X}$ (only a -1 loss in any $\mathbf{H}_\infty(\boldsymbol{x}_I)$) and Y is essentially as low-deficiency as $\mathcal{Y}$ (only a $+1$ loss in $\mathbf{D}_\infty$). Thus Lemma 20 (with a tiny perturbation of the parameters, which does not affect the result) shows that $G(X \times Y) = \{0,1\}^n$. However, the last rectangle that is processed, R_{2^k}, contains all of $\mathcal{X} \times \mathcal{Y}$ by definition (since we assume $\mathcal{L}_\mathrm{R}$ is correct on $\mathcal{X} \times \mathcal{Y}$). So, the invariant guarantees $(X \times Y) \cap G^{-1}(Z) = \emptyset$ at termination. This can only happen if $G^{-1}(Z) = \emptyset$ and thus $Z = \emptyset$ (since $G(X \times Y) = \{0,1\}^n$).

We now argue that $\mathcal{L}_\mathrm{C}$ is correct. Consider any z in the domain of f. Since Z is empty at termination, z must be accepted by some conjunction in $\mathcal{L}_\mathrm{C}$. Let (C, ℓ_i) be the first entry such that $C(z) = 1$, so $z \in Z$ during the iteration of the inner loop when this entry was added. Since in this iteration we have $G(A \times B) = C^{-1}(1)$ and $z \in C^{-1}(1)$, there is some $(x, y) \in A \times B$ with $G(x, y) = z$. Since $A \times B \subseteq R_i$ we have $(x, y) \in R_i$. Since $A \times B \subseteq X \times Y$, we have $(x, y) \in (X \times Y) \cap G^{-1}(Z)$ and thus (x, y) cannot be in R_j for any $j < i$ since $R_j \cap (X \times Y) \cap G^{-1}(Z) = \emptyset$ by the invariant. In summary, R_i is the first rectangle in $\mathcal{L}_\mathrm{R}$ that contains (x, y). By correctness of $\mathcal{L}_\mathrm{R}$ on $\mathcal{X} \times \mathcal{Y}$, we have $\oplus(\ell_i) = (f \circ g^n)(x, y) = f(G(x, y)) = f(z)$. Thus $\mathcal{L}_\mathrm{C}$ also correctly outputs $\oplus(\ell_i)$ on input z.

The "moreover" part is straightforward to verify: the levels assigned to conjunctions in $\mathcal{L}_\mathrm{C}$ come from the levels assigned to rectangles in $\mathcal{L}_\mathrm{R}$ (namely $\{0, \ldots, q\}$), in the same order (which is non-increasing). ◀

References

1 Josh Alman and Ryan Williams. Probabilistic rank and matrix rigidity. In *Proceedings of the 49th Symposium on Theory of Computing (STOC)*, pages 641–652. ACM, 2017. `doi:10.1145/3055399.3055484`.

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 László Babai, Noam Nisan, and Mario Szegedy. Multiparty protocols and logspace-hard pseudorandom sequences. In *Proceedings of the 21st Symposium on Theory of Computing (STOC)*, pages 1–11. ACM, 1989. `doi:10.1145/73007.73008`.

4 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.

5 Richard Beigel. Bounded queries to SAT and the Boolean hierarchy. *Theoretical Computer Science*, 84(2):199–223, 1991. doi:10.1016/0304-3975(91)90160-4.

6 Alberto Bertoni, Danilo Bruschi, Deborah Joseph, Meera Sitharam, and Paul Young. Generalized Boolean hierarchies and Boolean hierarchies over RP. In *Proceedings of the 7th International Conference on Fundamentals of Computation Theory (FCT)*, pages 35–46. Springer, 1989. doi:10.1007/3-540-51498-8_4.

7 Jin-Yi Cai and Lane Hemachandra. The Boolean hierarchy: Hardware over NP. In *Proceedings of the 1st Structure in Complexity Theory Conference (STRUCTURES)*, pages 105–124. Springer, 1986. doi:10.1007/3-540-16486-3_93.

8 Arkadev Chattopadhyay, Shachar Lovett, and Marc Vinyals. Equality alone does not simulate randomness. In *Proceedings of the 34th Computational Complexity Conference (CCC)*, pages 14:1–14:11. Schloss Dagstuhl, 2019. doi:10.4230/LIPIcs.CCC.2019.14.

9 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.

10 Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for P^{NP}. *Computational Complexity*, 28(1):113–144, 2019. doi:10.1007/s00037-018-0175-5.

11 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.

12 Mika Göös, Toniann Pitassi, and Thomas Watson. Zero-information protocols and unambiguity in Arthur–Merlin communication. *Algorithmica*, 76(3):684–719, 2016. doi:10.1007/s00453-015-0104-9.

13 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.2017.21.

14 Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. *SIAM Journal on Computing*, 47(6):2435–2450, 2018. doi:10.1137/16M1059369.

15 Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Computational Complexity*, 27(2):245–304, 2018. doi:10.1007/s00037-018-0166-6.

16 Bernd Halstenberg and Rüdiger Reischuk. Relations between communication complexity classes. In *Proceedings of the 3rd Structure in Complexity Theory Conference (STRUCTURES)*, pages 19–28. IEEE, 1988. doi:10.1109/SCT.1988.5259.

17 Bernd Halstenberg and Rüdiger Reischuk. Relations between communication complexity classes. *Journal of Computer and System Sciences*, 41(3):402–429, 1990. doi:10.1016/0022-0000(90)90027-I.

18 Bernd Halstenberg and Rüdiger Reischuk. Different modes of communication. *SIAM Journal on Computing*, 22(5):913–934, 1993. doi:10.1137/0222057.

19 Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*, volume 27 of *Algorithms and Combinatorics*. Springer, 2012. doi:10.1007/978-3-642-24508-4.

20 Jim Kadin. The polynomial time hierarchy collapses if the Boolean hierarchy collapses. *SIAM Journal on Computing*, 17(6):1263–1282, 1988. doi:10.1137/0217080.

21 Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992. doi:10.1137/0405044.

22 Hartmut Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proceedings of the 18th Conference on Computational Complexity (CCC)*, pages 118–134. IEEE, 2003. doi:10.1109/CCC.2003.1214415.

23 Johannes Köbler, Uwe Schöning, and Klaus Wagner. The difference and truth-table hierarchies for NP. *Theoretical Informatics and Applications*, 21(4):419–435, 1987. doi:10.1051/ita/1987210404191.

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

25 Nati Linial, Toniann Pitassi, and Adi Shraibman. On the communication complexity of high-dimensional permutations. In *Proceedings of the 10th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 54:1–54:20. Schloss Dagstuhl, 2019. doi:10.4230/LIPIcs.ITCS.2019.54.

26 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.

27 Toniann Pitassi, Morgan Shirley, and Thomas Watson. Nondeterministic and randomized boolean hierarchies in communication complexity. Technical Report TR19-043, Electronic Colloquium on Computational Complexity (ECCC), 2019. URL: https://eccc.weizmann.ac.il/report/2019/043.

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. On rigid matrices. Technical report, Steklov Mathematical Institute, 1989. In Russian.

30 Alexander Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992. doi:10.1016/0304-3975(92)90260-M.

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

32 Klaus Wagner. Bounded query computations. In *Proceedings of the 3rd Structure in Complexity Theory Conference (STRUCTURES)*, pages 260–277. IEEE, 1988. doi:10.1109/SCT.1988.5286.

33 Thomas Watson. A $\mathrm{ZPP}^{\mathrm{NP}[1]}$ lifting theorem. In *Proceedings of the 36th International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 59:1–59:16. Schloss Dagstuhl, 2019. doi:10.4230/LIPIcs.STACS.2019.59.

34 Gerd Wechsung. On the Boolean closure of NP. In *Proceedings of the 5th International Conference on Fundamentals of Computation Theory (FCT)*, pages 485–493. Springer, 1985. doi:10.1007/BFb0028832.

35 Andrew Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th Symposium on Theory of Computing (STOC)*, pages 209–213. ACM, 1979. doi:10.1145/800135.804414.

36 Stathis Zachos and Hans Heller. A decisive characterization of BPP. *Information and Control*, 69(1-3):125–135, 1986. doi:10.1016/S0019-9958(86)80044-4.