Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 193 (2025)

Pseudodeterministic Communication Complexity

Mika Göös
EPFL

Nathaniel Harms
University of British Columbia

Artur Riazanov
EPFL

Anastasia Sofronova
EPFL

Dmitry Sokolov
EPFL & Université de Montréal

Weiqiang Yuan
EPFL

June 1, 2026

Abstract

We exhibit an n-bit partial function with randomized communication complexity $O(\log n)$ but such that any completion of this function into a total one requires randomized communication complexity $n^{\Omega(1)}$. In particular, this shows an exponential separation between randomized and *pseudodeterministic* communication protocols. Previously, Gavinsky (2025) showed an analogous separation in the weaker model of parity decision trees. We use lifting techniques to extend his proof idea to communication complexity.

Contents

ISSN 1433-8092

1 Introduction

Here is a simple question about the behavior of randomized algorithms. A basic statistical task is to distinguish "few" vs. "many", formalized by the GAP MAJORITY problem:

$$\textsc{GapMaj}(x) \coloneqq \begin{cases} 1 & \text{if } |x| \geq \frac{2}{3}n, \\ 0 & \text{if } |x| \leq \frac{1}{3}n, \\ * & \text{otherwise,} \end{cases}$$

where $|x|$ denotes the number of 1s in $x \in \{0,1\}^n$ and $*$ indicates that we put no requirement on the output of an algorithm (that is, GAPMAJ is a search problem with 0 and 1 both being acceptable outputs). This problem is difficult to solve deterministically: It requires $\Omega(n)$ queries for decision trees and parity decision trees, and $\Omega(n)$ bits of deterministic communication when we turn it into a two-player problem by composing it with an appropriate gadget. For example, composing with XOR yields the GAP HAMMING DISTANCE problem given by $\textsc{GapHD}(x, y) \coloneqq \textsc{GapMaj}(x \oplus y)$. On the other hand, for *randomized* decision trees (and the other models as well), the cost of GAPMAJ is only 1 query because the algorithm can sample a random coordinate and output it; when $|x| \geq \frac{2}{3}n$ or $|x| \leq \frac{1}{3}n$ this is correct with probability at least 2/3. Our simple question is:

What is the randomized algorithm doing on the $*$ inputs?

On inputs with $|x| = n/2$, the sampling algorithm will output 0 or 1 with equal probability. But *must it* do this, or can we ask that the randomized algorithm produce consistent outputs for every input x? That is, can we ask that the algorithm computes some *completion* of GAPMAJ into a total function: on any input x it should output either value 0 with probability at least 2/3, or value 1 with probability at least 2/3. Such an algorithm is called *pseudodeterministic*. Generally, a pseudodeterministic algorithm is a randomized algorithm that is required to output (with probability at least 2/3) a single consistent output for every input; this is desirable not only for partial boolean functions, but for any problem where there is more than 1 acceptable output for each input, that is, for any search problem or relation. The goal is to combine the efficiency of randomized algorithms with the consistency of deterministic algorithms. So, are pseudodeterministic algorithms nearly as efficient as randomized ones?

We show that, for communication protocols, the answer is *no*: there is a partial boolean function with an efficient randomized protocol but such that every total completion of that function has large randomized complexity.

Theorem 1. *There is a partial communication problem $\{0,1\}^n \times \{0,1\}^n \to \{0,1,*\}$ with randomized communication complexity $O(\log n)$ but pseudodeterministic communication complexity $\Omega(\sqrt{n})$.*

Partial n-bit boolean functions with poly log n randomized communication cost are a type of "BPP search problems": relations $R \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ computed by an efficient randomized protocol. By analogy to Turing machine complexity classes, the class of "BPP search problems" would be called FBPP. If we use FPs to denote the class of relations admitting pseudodeterministic protocols with cost poly log n, Theorem 1 is the first explicit example witnessing the separation

$$\mathsf{FPs} \subsetneq \mathsf{FBPP}, \tag{1}$$

and it does so in the most restricted setting, where the output is only a single bit. We remark, however, that (1) has a caveat: just like the Turing machine analogue, the communication class

FBPP has several natural definitions, which are not equivalent [Aar10, Gol11, ILW23, ABK24]. For one of these definitions (which requires the output to be efficiently verifiable), the inclusion (1) holds in the *opposite direction* for partial boolean functions; our Theorem 1 gives an explicit witness for (1) under the remaining definitions. Weaker versions of (1), which allow the outputs to be large and do not require an explicit example, can be proved by counting; see Appendix A for a discussion of these nuances.

Several recent works [GGMW20, GIPS21, Gav25, FGHH25, BHH+25] asked for lower bounds on pseudodeterministic communication complexity for BPP search problems. Many of them proved versions of (1) for weaker models of computation: Goldwasser, Grossman, Mohanty, and Woodruff [GGMW20] proved it for one-way communication protocols; Goldwasser, Impagliazzo, Pitassi, and Santhanam [GIPS21] proved it for decision trees; and Gavinsky [Gav25] proved it for parity decision trees. Our result can thus be viewed as a qualitative strengthening of these prior works (the upper bound in Theorem 1 holds even for randomized non-adaptive decision trees).

For partial functions, the only known prior separation was due to Blondal, Hatami, Hatami, Lalov, and Tritiak [BHH+25]. They showed an $O(1)$-vs-$\Omega(\log\log n)$ separation for Gap Hamming Distance, i.e., the XOR-lift $\textsc{GapMaj}(x \oplus y)$. This would be the ideal function to witness the separation (1) for partial functions, and we suspect it exhibits the maximum possible separation:

Conjecture 2. *The pseudodeterministic communication complexity of* GapHD *is* $\Omega(n)$.

This was one of the principal inspirations for our present work, but our Theorem 1 ultimately does not prove the separation using GapHD; we instead use a slightly more complicated lift of GapMaj, as explained next.

1.1 Our techniques

Our proof extends the work of Gavinsky [Gav25]. He proved an $\Omega(\sqrt{n})$ query lower bound for pseudodeterministic parity decision trees computing the GapMaj partial function. To turn this function into a communication problem, we compose it with the Inner Product gadget on $m = O(\log n)$ bits defined by $\mathrm{IP}_m(a,b) \coloneqq \sum_{i=1}^m a_i b_i \bmod 2$. This produces a communication problem on $O(n \log n)$ bits:

$$\textsc{GapMaj} \circ \mathrm{IP}_m^n(x_1,\ldots,x_n,y_1,\ldots,y_n) \coloneqq \textsc{GapMaj}\left(\mathrm{IP}_m(x_1,y_1),\ldots,\mathrm{IP}_m(x_n,y_n)\right).$$

Here $x_i, y_i \in \{0,1\}^m$, Alice has all the x_i inputs, and Bob has all the y_i inputs. This problem has randomized communication complexity $O(\log n)$ because the players can select $O(1)$ random indices $i \in [n]$ and solve $\mathrm{IP}_m(x_i,y_i)$ deterministically.

Standard randomized lifting theorems [GPW20, CFK+21] show that for every partial function $f\colon \{0,1\}^n \to \{0,1,*\}$ the randomized communication complexity of $f \circ \mathrm{IP}_m^n$ equals roughly the randomized query complexity of f. It is open to prove such a general lifting theorem for pseudodeterministic complexity. The immediate obstacle to applying a randomized lifting theorem to a pseudodeterministic protocol for $f \circ \mathrm{IP}_m^n$ is that the protocol may produce different consistent outputs for inputs (x,y) and (x',y') such that $\mathrm{IP}_m^n(x,y) = \mathrm{IP}_m^n(x',y') \in f^{-1}(*)$. Since we cannot invoke known lifting theorems as a black box, we employ a *white-box* approach: We use tools from lifting theory to adapt Gavinsky's argument for communication protocols. This adaptation introduces new technical challenges not present in the setting of parity decision trees. We give an overview of the proof in Section 2, including an exposition of Gavinsky's original argument.

1.2 History and Motivations

Pseudodeterministic algorithms were introduced by Gat and Goldwasser [GG11] (under the name *Bellagio* algorithms), and independently by Huynh and Nordström [HN12] in the context of proof complexity (they called them *consistent* algorithms). Pseudodeterministic algorithms have applications to cryptography and distributed computing, and have been studied in several computational models including Turing machines [GGR13, OS17, LOS21, CLO+23], decision trees [GGR13, GIPS21, CDM23], and parity decision trees [Gav25]. They are natural and interesting in their own right, they provide an intermediary between randomized and deterministic complexity, and they are also related to notions of *replicability* in machine learning and the natural sciences [GIPS21, ILPS22, CMY23, BHH+25].

Aside from intrinsic interest in pseudodeterminism, [Gav25] points out its connection to structural questions about randomized communication (which we discuss in the open problems below), and [GIPS21] argues that understanding pseudodeterministic communication is an important step towards understanding the communication complexity of search problems (i.e., relations). Whereas the communication complexity of *functions* benefits from connections to well-understood query complexity measures (e.g., sensitivity and block-sensitivity) via lifting theorems, analogues of these measures for search problems are not well understood. Pseudodeterministic algorithms must compute *some* function, and therefore serve as an intermediate between functions and relations.

Pseudodeterminism was introduced in [HN12] for applications in proof complexity. That paper shows that *cutting plane proofs* of CNF unsatisfiability can be turned into pseudodeterministic communication protocols for the *falsified-clause* search problem where two players search for a clause that is not satisfied by a (distributed) assignment of variables. Their lower bound for this communication problem allowed them to prove time–space tradeoffs for cutting planes.

1.3 Open Problems

Improved quantitative bounds. Let us start by reiterating that we conjecture Gap Hamming Distance to require pseudodeterministic complexity $\Omega(n)$ (Conjecture 2). Proving this seems to require two steps. First, improve Gavinsky's lower bound of $\Omega(\sqrt{n})$ for parity decision trees computing GapMaj to $\Omega(n)$. Second, lift this to GapHD. Another strategy may be to directly improve the current lower bound for the pseudodeterministic communication complexity of GapHD, which is $\Omega(\log\log n)$, and which uses very different techniques [BHH+25].

It is important to note that Gap Hamming Distance is, in a precise sense, the *only* partial boolean communication problem where we can hope to prove an $O(1)$-vs-poly(n) separation between randomized and pseudodeterministic communication. This is because *all* n-bit partial boolean matrices with randomized cost $O(1)$ are a submatrix of GapHD on $O(n)$ bits (with some constant $\alpha < 1/2$ in place of $1/3$ in the gap) [LS09, FGHH25]. So a pseudodeterministic lower bound for any of these matrices implies the same lower bound for GapHD.

Separation for a TFNP problem. The most outstanding *qualitative* problem that is left open is to prove pseudodeterministic lower bounds for a *total* NP *search problem* (TFNP). A relation $R \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ is in (communication analogue of) TFNP if

- R is *total*, meaning that for all inputs (x,y) there is a solution z such that $(x,y,z) \in R$; and
- R is *efficiently checkable*, meaning that there is a deterministic verifier protocol V of cost poly $\log n$ such that $V(xz, yz) = 1$ iff $(x,y,z) \in R$ for all (x,y,z).

Conjecture 3. *There exists some* TFNP *communication problem with randomized communication complexity* $\text{poly}\log n$ *but pseudodeterministic complexity* $n^{\Omega(1)}$.

The analogous separation for query complexity was the main result of the paper [GIPS21]. The communication analogue of TFNP has been studied explicitly in, e.g., [GKRS19, BFI23, GGJL25], but, implicitly, TFNP problems are ubiquitous in communication complexity: this includes all (monotone) Karchmer–Wigderson games as well as the aforementioned falsified-clause search problems, which arise when applying communication lower bounds to proof complexity. The survey [dRGR22] explains these connections and more: TFNP problems give a unified lens to study the interconnections between communication protocols, propositional proofs, and boolean circuits.

Separation for a BPP-verifiable problem. In fact, it is still open to prove a weaker result than Conjecture 3. If we define BPP search problems for communication complexity by adapting the definition of Goldreich [Gol11] for Turing machines, we get the class of relations $R \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ such that

- There is a randomized $\text{poly}\log(n)$ cost protocol solving R; and
- There is a randomized $\text{poly}\log(n)$ cost *verifier* protocol V: $\Pr[V(xz, yz) = 1] \geq 2/3$ if $(x, y, z) \in R$ and $\Pr[V(xz, yz) = 1] \leq 1/3$ otherwise, for all (x, y, z).

We can then replace the TFNP requirement in Conjecture 3 with these BPP-verifiable problems and the conjecture is still open.

Our result implies this separation for a weaker, more specialized type of verification. The partial function $\text{GapMaj} \circ \text{IP}$ does *not* satisfy the above definition, because the verifier cannot exist (we could use it to solve *exact* majority efficiently, which cannot be done). However, we can define the relation $R \subseteq \{0,1\}^{nm} \times \{0,1\}^{nm} \times [n]$ where for each x, y the valid outputs are the numbers $t \in [n]$ that are within $\pm n/10$ of the number of 1-valued IP_m gadgets. This relation can be solved in cost $O(m) = O(\log n)$ the same way as $\text{GapMaj} \circ \text{IP}_m$. It can also be *sort-of* verified: given $(x, y, \widehat{t})$, if the correct number of 1-valued gadgets is t, there is a randomized verifier V which satisfies

- If $|t - \widehat{t}| \leq n/10$ (i.e., $(x, y, \widehat{t}) \in R$) then $\Pr\big[V(x, y, \widehat{t}) = 1\big] \geq 2/3$; and
- If $|t - \widehat{t}| \geq n/9$ then $\Pr\big[V(x, y, \widehat{t}) = 1\big] \leq 1/3$.

An efficient pseudodeterministic protocol for this estimation problem could be used as a pseudodeterministic protocol for $\text{GapMaj} \circ \text{IP}_m$, so Theorem 1 shows that it cannot exist. However, the necessary gap between $n/10$ and $n/9$ means that this does not satisfy the definition of [Gol11].

Structure of communication protocols. Communication complexity is closely related to the size of monochromatic rectangles within the communication matrices: efficient deterministic protocols imply large monochromatic rectangles inside the matrix. The question of whether the same is true for randomized protocols that compute a total boolean function was raised in [GKPW18]. A striking version of this question is obtained for constant-cost protocols: Chattopadhyay, Lovett, and Vinyals [CLV19] and Hambardzumyan, Hatami, and Hatami [HHH23] conjecture that

Conjecture 4 ([CLV19, HHH23])**.** *There exists a function* η *such that every* $N \times N$ *boolean matrix with randomized communication cost* c *has an* $\eta(c) \cdot N \times \eta(c) \cdot N$ *monochromatic rectangle.*

Completions of the GapHD matrix *cannot* have such large monochromatic rectangles [FF81], and therefore Conjecture 4 already implies lower bounds on the pseudodeterministic cost of GapHD, as noted in [FGHH25]. Gavinsky [Gav25] points out similar implications for parity decision trees, with monochromatic affine subspaces in place of rectangles. The best progress so far towards Conjecture 4 is to find such monochromatic rectangles in matrices of bounded γ_2-norm [BHT25].

2 Warm-Up and Proof Overview

Our proof starts with the lower bound for parity decision trees by Gavinsky [Gav25] and upgrades it to a communication lower bound using techniques from query-to-communication lifting [GLM+16, GPW20]. As a warm-up to our main proof, we will present an exposition of Gavinsky's argument, but simplified to the special case of decision trees.

Of course, for decision trees (rather than parity decision trees), there is a simpler proof of a superior $\Theta(n)$ bound: in any completion f of GapMaj, find the input $x \in f^{-1}(0)$ with largest weight; then, observe that by fixing the 1-valued coordinates of x and letting the 0-valued coordinates vary, we obtain the Or problem on $\Omega(n)$ bits, which implies a lower bound of $\Omega(n)$ queries. However, the more complicated warm-up proof gives a technique that can be lifted to communication complexity.

2.1 Warm-Up: Lower Bound for Decision Trees

Theorem 5 (Simplified version of [Gav25])**.** *Any completion $f\colon \{0,1\}^n \to \{0,1\}$ of* GapMaj *requires randomized query complexity $\Omega(\sqrt{n})$.*

Proof sketch. The proof has two parts, encapsulated in the *Closeness Lemma* and the *Stage Lemma*. The Closeness Lemma formalizes the key observation that a shallow decision tree cannot distinguish between the uniform distribution over $\{0,1\}^n$ and the distribution where the bits are slightly biased towards 1. Specifically, suppose $\boldsymbol{u} \sim \{0,1\}^n$ is uniform random and $\boldsymbol{x}$ is obtained from $\boldsymbol{u}$ by setting to 1 a uniformly random set of $\sqrt{n}$ coordinates; that is, we "sprinkle" some $\sqrt{n}$ many 1s into $\boldsymbol{u}$. Then the probability that a depth-$o(\sqrt{n})$ decision tree T can distinguish $\boldsymbol{u}$ from $\boldsymbol{x}$ is $o(1)$; in other words, the output distributions $T(\boldsymbol{u})$ and $T(\boldsymbol{x})$ are close to each other.

Using this Closeness Lemma, we break down the argument into stages, each stage handled by the Stage Lemma. In each stage, we start with a subcube $X \subseteq \{0,1\}^n$ where at least half the strings $x \in X$ satisfy $f(x) = 0$. A subcube X is equivalent to a partial assignment of variables, where a set $F \subseteq [n]$ are "free", while variables $[n] \setminus F$ are "fixed", i.e., all $x \in X$ agree on the coordinates $[n] \setminus F$. Our goal in each stage is to find $\approx \sqrt{n}$ free variables to fix (i.e., to remove from F to obtain a new subcube); essentially, these newly fixed variables are the ones queried by the decision tree, along with the $\sqrt{n}$ "sprinkled" 1-bits. Our choice should satisfy the property that, after fixing these new variables, we maintain the invariant that at least half the remaining strings x have $f(x) = 0$. If we can accomplish this goal in each stage, repeating this process over $0.9\sqrt{n}$ stages yields a contradiction as illustrated in Figure 1: on one hand, we have fixed $0.9\sqrt{n} \cdot \sqrt{n} = 0.9n$ bits to 1, so f is identically 1 on all remaining strings. On the other hand, we maintained the invariant that half the remaining strings x have $f(x) = 0$, forcing a contradiction.

We find the $\approx \sqrt{n}$ variables to fix as follows. The key trick is that, by Yao's principle, there is a deterministic $o(\sqrt{n})$-depth decision tree T that errs with probability at most ε on the distribution

$$\tfrac{1}{2}(\mathsf{unif}(X) + \sigma),$$

where σ is the "sprinkled-1s" distribution: start with a uniform $\boldsymbol{u} \sim X$ and construct $\boldsymbol{x}$ by fixing a random set of $\sqrt{n}$ free variables to 1. Crucially, T has error probability 2ε over each component distribution $\mathsf{unif}(X)$ and σ. Since $\Pr_{\boldsymbol{x}\sim X}[f(\boldsymbol{x}) = 0] \geq 1/2$ and T has error 2ε over $\mathsf{unif}(X)$, we can easily find a 0-leaf ℓ where $f(x) = 0$ for most x reaching that leaf; by fixing the variables queried by T on the path to ℓ, we obtain a subcube L with $\Pr_{\boldsymbol{x}\sim L}[f(\boldsymbol{x}) = 0] \geq 1/2$. We can "upgrade" the last step to let us fix an additional $\sqrt{n}$ variables to 1, by using the fact that T also has error 2ε over σ, and the Closeness Lemma, which says that the distribution over leaves of T is similar for σ as for $\mathsf{unif}(X)$.

Full proof (skippable on first reading). We now give a more formal proof that serves as a blueprint for our communication lower bound. The following lemma formalizes the process for each stage, that is, the act of assigning $(1 + o(1))\sqrt{n}$ bits. To better match with the communication proof we view the conditioning on a partial assignment as zooming in to a subcube of $\{0, 1\}^n$.

Notation: Throughout the paper, for any string $x \in \Sigma^n$ and a set $S \subseteq [n]$, we write $x_F \in \Sigma^F$ for the substring of x on coordinates F. For $X \subseteq \Sigma^n$, we write $X_F \coloneqq \{x_F \mid x \in X\}$.

Lemma 6 (Stage Lemma for Decision Trees)**.** *Suppose $f\colon \{0, 1\}^n \to \{0, 1\}$ is a completion of* GapMaj *with randomized query complexity $o(\sqrt{n})$. Let $X \subseteq \{0, 1\}^n$ be a subcube of the boolean cube, that is, there exists $F \subseteq [n]$ be such that $X_{[n]\smallsetminus F}$ is fixed and X_F is free. Suppose that $|F| \geq 0.1n$ and $\Pr_{\boldsymbol{u}\sim X}[f(\boldsymbol{u}) = 0] \geq 1/2$. Then there exists a subcube $X' \subseteq X$ and $F' \subseteq F$ such that*

$$\Pr_{\boldsymbol{u}'\sim X'}[f(\boldsymbol{u}') = 0] \geq \frac{1}{2} \qquad \textit{and} \qquad X_{F'} \textit{ is free.}$$

Moreover, there is a set $R \subseteq F \smallsetminus F'$ such that: $X'_R = 1_R$ and $|R| \geq 0.9|F \smallsetminus F'|$.

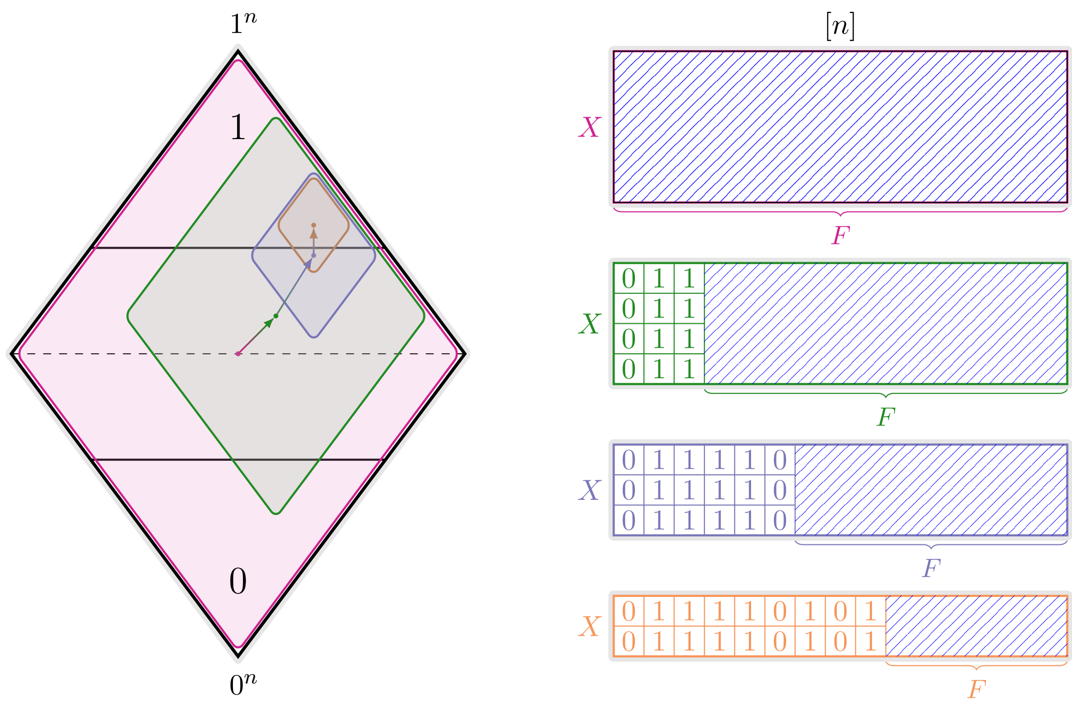
This lemma suffices to prove the lower bound, as follows (see Figure 1 for an illustration):

Proof of Theorem 5 given Lemma 6. We may assume that $|f^{-1}(0)| \geq 2^{n-1}$, as otherwise we can swap the roles of 0/1 output values in the upcoming argument. Begin with $X = \{0, 1\}^n$, $F = [n]$, and $S = \varnothing$. We will apply the stage lemma iteratively as long as $|F| \geq 0.1n$ and update X, F, S in each stage, as follows: assuming the invariant $\Pr_{\boldsymbol{u}\sim X}[f(\boldsymbol{u}) = 0] \geq 1/2$, we apply the stage lemma and update

- $X \leftarrow X'$ with X' from the stage lemma, maintaining the invariant $\Pr_{\boldsymbol{u}'\sim X'}[f(\boldsymbol{u}') = 0] \geq 1/2$;
- $F \leftarrow F'$ from the stage lemma;
- $S \leftarrow S \cup R$ from the stage lemma, maintaining the invariant that all $x \in X'$ have value 1 on coordinates S.

On termination, we have preserved the invariant $\Pr_{\boldsymbol{u}\sim X}[f(\boldsymbol{u}) = 0] \geq 1/2$, but on the other hand, since $|S| \geq 0.9|[n] \smallsetminus F| \geq 0.8n$ and all $x \in X$ have coordinates in S fixed to 1, we have $f(x) = 1$ for all $x \in X$, a contradiction. □

To prove the stage lemma, we define the following distribution over inputs. Given any subcube $X \subseteq \{0, 1\}^n$ with free variables $F \subseteq [n]$, we define σ as the distribution over X obtained by choosing a uniformly random $\boldsymbol{u} \sim X$ and a uniformly random set $\boldsymbol{I} \sim \binom{F}{\sqrt{n}}$ of $\sqrt{n}$ free variables. Then we take $\boldsymbol{x} \in X$ to be the string obtained from $\boldsymbol{u}$ by setting all coordinates $i \in \boldsymbol{I}$ to 1.



(a) Evolution of subcube X; it ends up contained in $f^{-1}(1)$.

(b) Evolution of free variables F; each stage fixes at least $\sqrt{n}$ new 1s.

Figure 1: Decision tree stages

This distribution has two important properties: first, it has $\sqrt{n}$ "planted" 1-valued coordinates, so that it is more biased towards 1 than a uniformly random string. Second, it is indistinguishable from the uniform distribution by $o(\sqrt{n})$-query decision trees, as established in the *Closeness Lemma*:

Lemma 7 (Closeness Lemma for Decision Trees)**.** *Let $X' \subseteq X$ be a subcube of X and $F' \subseteq F$ the set of its free variables. Suppose $|F \setminus F'| \leq o(\sqrt{n})$ and $|F| \geq 0.1n$. Then*

$$\Pr_{\boldsymbol{x}\sim\sigma}[\boldsymbol{x} \in X'] \geq (1 - o(1)) \Pr_{\boldsymbol{u}\sim X}[\boldsymbol{u} \in X'].$$

Proof. Let $\boldsymbol{I} \sim \binom{F}{\sqrt{n}}$ be as in the definition of σ. Then $\boldsymbol{x} \sim \sigma$ is obtained by choosing $\boldsymbol{u} \sim X$ and then fixing variables in $\boldsymbol{I}$ to 1. If $\boldsymbol{u} \in X'$ and the set $\boldsymbol{I} \subseteq F$ does not intersect the set $F \setminus F'$ of variables which are fixed by X', then it must be the case that $\boldsymbol{x} \in X'$ as well. Therefore

$$\Pr_{\boldsymbol{x}\sim\sigma}[\boldsymbol{x} \in X'] \geq \Pr_{\boldsymbol{u}\sim X}[\boldsymbol{u} \in X' \wedge (\boldsymbol{I} \cap (F' \setminus F) = \varnothing)] = \Pr_{\boldsymbol{u}\sim X}[\boldsymbol{u} \in X'] \cdot \Pr\big[\boldsymbol{I} \cap (F' \setminus F) = \varnothing\big].$$

By the union bound,

$$\Pr\big[\boldsymbol{I} \cap (F' \setminus F) \neq \varnothing\big] \leq \sqrt{n} \cdot \frac{|F' \setminus F|}{|F|} \leq \sqrt{n} \cdot \frac{o(\sqrt{n})}{0.1n} = o(1). \qquad \square$$

We now conclude the lower bound by proving the stage lemma:

Proof of Lemma 6. Suppose that f has randomized decision tree complexity $o(\sqrt{n})$; we may assume that the randomized decision tree has error probability ε. Let X be any subcube with free variables F satisfying $|F| \geq 0.1n$. Recall the distribution σ over X where $\boldsymbol{x} \sim \sigma$ is sampled by choosing $\boldsymbol{I} \sim \binom{F}{\sqrt{n}}$ and $\boldsymbol{x} \sim \mathsf{unif}(\{x \in X \mid x_{\boldsymbol{I}} = 1_{\boldsymbol{I}}\})$.

By Yao's principle, there exists a deterministic decision tree T of depth $o(\sqrt{n})$, computing f with error ε over the mixture distribution

$$\tfrac{1}{2}(\mathsf{unif}(X) + \sigma).$$

Observe that tree T has error at most 2ε over each individual distribution σ and $\mathsf{unif}(X)$.

Note that each leaf ℓ of T corresponds to a subcube $L \subseteq X$ obtained by fixing the $o(\sqrt{n})$ bits queried by T on the path to ℓ; therefore the free variables F_L of L satisfy $|F \smallsetminus F_L| = o(\sqrt{n})$. By applying the closeness lemma, Lemma 7, to each subcube L corresponding to the 0-leaves of T, we obtain

$$\Pr_{\boldsymbol{x}\sim\sigma}[T(\boldsymbol{x}) = 0] = \sum_{0-\text{leaf } L} \Pr_{\boldsymbol{x}\sim\sigma}[\boldsymbol{x} \in L] \geq (1 - o(1)) \Pr_{\boldsymbol{u}\sim X}[T(\boldsymbol{u}) = 0] \geq (1 - o(1)) \left(\tfrac{1}{2} - 2\varepsilon\right) \geq 1/4,$$

where the penultimate inequality uses the assumption $\Pr_{\boldsymbol{u}\sim X}[f(\boldsymbol{u}) = 0] \geq 1/2$. Now, since T errs with probability at most 2ε on σ, we have

$$\Pr_{\boldsymbol{x}\sim\sigma}[f(\boldsymbol{x}) = 1 \mid T(\boldsymbol{x}) = 0] = \frac{\Pr_{\boldsymbol{x}\sim\sigma}[f(\boldsymbol{x}) = 1 \land T(\boldsymbol{x}) = 0]}{\Pr_{\boldsymbol{x}\sim\sigma}[T(\boldsymbol{x}) = 0]} \leq 8\varepsilon.$$

Hence, according to the total probability law, there exists a 0-leaf ℓ in T such that $\Pr_{\boldsymbol{x}\sim\sigma}[f(\boldsymbol{x}) = 1 \mid \boldsymbol{x} \in L] \leq 8\varepsilon \leq 1/2$. Again by total probability law, there exists $I \in \binom{F}{\sqrt{n}}$ such that $\Pr_{\boldsymbol{x}\sim\sigma}[f(\boldsymbol{x}) = 1 \mid \boldsymbol{x} \in L, \boldsymbol{I} = I] \leq 1/2$. We then observe that for $\boldsymbol{u} \sim X$ the distributions of $(\boldsymbol{u} \mid \boldsymbol{u}_I = 1)$ and $(\boldsymbol{x} \mid \boldsymbol{I} = I)$ coincide, therefore $\Pr_{\boldsymbol{u}\sim X}[f(\boldsymbol{u}) = 1 \mid \boldsymbol{u} \in \ell, \boldsymbol{u}_I = 1] \leq 1/2$. We then conclude by defining $X' \coloneqq \{x \in L \mid x_I = 1_I\}$ with $R \coloneqq I$ being the variables forced to 1, and the new set of free variables being $F' \coloneqq F_L \smallsetminus R$ (the free variables remaining in the leaf L, minus those forced to 1), which satisfies $|R| = \sqrt{n} \geq 0.9|F \smallsetminus F'| = 0.9(\sqrt{n} + o(\sqrt{n}))$. □

2.2 Proof Overview

Let us describe how to lift the argument for decision trees in Section 2.1 to prove a lower bound for the communication complexity of $\textsc{GapMaj}\circ\mathrm{IP}_m^n$. For the sake of brevity, we will write $\Sigma \coloneqq \{0,1\}^m$ for the alphabet of inputs to the IP_m gadget, so that both inputs to $\textsc{GapMaj}\circ\mathrm{IP}_m^n$ are in domain Σ^n. For the remainder of the paper, $m = \log|\Sigma|$ indicates the bit-size of the IP_m gadget. Below, we use $\mathsf{R}(\cdot)$ to denote randomized communication complexity (to within error 1/3).

Theorem 8. *Any completion $f\colon \Sigma^n \times \Sigma^n \to \{0,1\}$ of $\textsc{GapMaj}\circ\mathrm{IP}_m^n$, with $m \coloneqq 100\log n$, has*

$$\mathsf{R}(f) = \Omega(\sqrt{n}\log n).$$

Our proof proceeds analogously to the proof for query complexity in Section 2.1. At a high level, we have the following analogies:

- For decision trees we tracked a subcube $X \subseteq \{0,1\}^n$, because leaves of a decision tree correspond to subcubes. Now we track a rectangle $X \times Y \subseteq \Sigma^n \times \Sigma^n$ because leaves of a communication protocol correspond to rectangles. (Gavinsky's argument for parity decision trees tracks an affine subspace $X \subseteq \mathbb{F}_2^n$.)

- We still have a set F of "free" variables, but these take on a different meaning, because the rectangle $X \times Y$ no longer corresponds exactly to a set of inputs obtained by fixing variables $[n] \setminus F$ and leaving those in F completely unrestricted; this is because communication protocols may exchange information about variables without entirely "fixing" them. Our "free variables" F instead satisfy the condition that "only a small amount of information about them is known".
- We still have a *Stage Lemma*, which suffices to prove the theorem in almost the same way: starting with a rectangle $X \times Y$ where $f(x, y) = 0$ for most $(x, y) \in X \times Y$, we iteratively restrict to a smaller rectangle $X' \times Y'$ by "fixing" some free variables, including $\sqrt{n}$ variables fixed to 1. Unlike the decision tree proof, we can no longer fix $\sqrt{n}$ variables to 1 in *every* stage, but only *most* stages (which we call "safe stages").
- We still have a *Closeness Lemma*, which we use to find $\sqrt{n}$ variables to fix to 1; however, the simple birthday paradox argument in Lemma 7 no longer works, because a protocol does not simply query $o(\sqrt{n})$ bits, so $\sqrt{n}$ random coordinates will not simply "miss" the queried coordinates with high probability. Indeed, our new Closeness Lemma will only apply in the "safe stages".

Let's give a little more detail on how the Stage Lemma and Closeness Lemma differ from their decision tree analogues. We also provide a diagram of the Stage Lemma in Figure 3.

Stage Lemma: Suppose for the sake of contradiction that there exists a randomized communication protocol with error ε and communication cost $o(\sqrt{n}m)$. We proceed as follows:

1. Start each stage with a rectangle $X \times Y$ (originally $\Sigma^n \times \Sigma^n$) which contains mostly 0-valued inputs to f, and a set $F \subseteq [n]$ of "free variables". We keep track of the "deficit"

$$\mathcal{D}(X \times Y, F) := 2|F|m - \mathrm{H}_\infty(\boldsymbol{x}_F) - \mathrm{H}_\infty(\boldsymbol{y}_F),$$

   where $(\boldsymbol{x}, \boldsymbol{y}) \sim X \times Y$ and $\mathrm{H}_\infty(\boldsymbol{x}) := \min_x \log(1/\Pr[\boldsymbol{x} = x])$ denotes min-entropy. Intuitively, the deficit quantifies how many bits of information Alice and Bob know about the free variables F. For intuition, note that if Alice and Bob were merely simulating a decision tree (solving individual gadgets one at a time) then the deficit would always be 0.

2. By Yao's principle we obtain a deterministic protocol Π with cost $|\Pi| = o(\sqrt{n}m)$ and error ε over the distribution

$$\frac{1}{2}\left(\mathsf{unif}(X \times Y) + \sigma(X \times Y, F)\right),$$

   where $\sigma(X \times Y, F)$ is a *sprinkled-1s* distribution (Definition 14). In the sprinkled-1s distribution, an input $(\boldsymbol{x}, \boldsymbol{y})$ is chosen uniformly from $X \times Y$, and then a random set of $\sqrt{n}$ gadgets in F are fixed to have value 1, i.e., we force the input distribution to increase the number of 1-valued gadgets by at least $\sqrt{n}$.

3. We use the fact that Π has error 2ε on each of $\mathsf{unif}(X \times Y)$ and $\sigma(X \times Y, F)$, together with the Closeness Lemma (which says that distribution over leaves induced by $\sigma(X \times Y, F)$ and $\mathsf{unif}(X \times Y)$ is similar), to find a smaller subrectangle $X' \times Y' \subseteq X \times Y$ and smaller set of free variables $F' \subseteq F$, where $\sqrt{n}$ new IP gadgets are fixed to 1. However, the Closeness Lemma will now apply only in "safe stages" where the deficit is low; in the unsafe stages, we skip the step of fixing $\sqrt{n}$ gadgets to 1. This is OK, because most stages will be safe.

Closeness Lemma: The Closeness Lemma should show that the distribution over leaves of the protocol is similar under the sprinkled-1s distribution $\sigma(X \times Y, F)$ and $\mathsf{unif}(X \times Y)$, so that we can safely sprinkle $\sqrt{n}$ 1-values "without the protocol noticing". However, to prove this claim, there are two caveats, which constitute the main technical novelty of the proof, and do not have an analogue in Gavinsky's parity decision tree proof:

1. In each stage, we must first transform Π into a more structured protocol Π' using techniques from lifting and properties of the IP_m gadget.
2. We can prove the claim only when the deficit $\mathcal{D}(X \times Y, F)$ is *small* (i.e., there is little information known about the free variables F, so that intuitively they behave similarly to the completely free variables in the decision tree argument).

In the next section, we formalize the main claims required for this argument.

3 Proof of the Main Theorem

For the sake of contradiction, we make the following assumption throughout the proof:

Assumption 9. *Assume that there is a completion of* $\textsc{GapMaj} \circ \mathrm{IP}_m^n$*, denoted* $f : \Sigma^n \times \Sigma^n \to \{0,1\}$*, such that the randomized communication cost of* f *is* $o(\sqrt{n}m)$ *and* $\Pr_{(\boldsymbol{x},\boldsymbol{y})}[f(\boldsymbol{x},\boldsymbol{y}) = 0] \geq 1/2$*.*

As in the decision tree proof, the main theorem will follow from a Stage Lemma. To state the communication version of the Stage Lemma, we need a few definitions. First, we will use the *deficit* as a potential function throughout the stage procedure:

Definition 10 (Deficit). For a set $F \subseteq [n]$ and a random variable $\boldsymbol{x}$ over Σ^n, we write

$$\mathrm{D}_\infty(\boldsymbol{x}_F) \coloneqq |F|m - \mathrm{H}_\infty(\boldsymbol{x}_F).$$

For a rectangle $X \times Y \in \Sigma^n \times \Sigma^n$ and $F \subseteq [n]$ we define the *deficit*:

$$\mathcal{D}(X \times Y, F) \coloneqq \mathrm{D}_\infty(\boldsymbol{x}_F) + \mathrm{D}_\infty(\boldsymbol{y}_F) = 2|F|m - \mathrm{H}_\infty(\boldsymbol{x}_F) - \mathrm{H}_\infty(\boldsymbol{y}_F),$$

where $(\boldsymbol{x},\boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$. Observe that the deficit is non-negative.

Next, we will maintain the property that, for the current rectangle $X \times Y$ in the process, the sets X and Y are pseudorandom in sense of being γ-*spread*:

Definition 11 (Spread Variables). A random variable $\boldsymbol{x} \in \Sigma^J$ is γ-spread if $\forall I \subseteq J$ it holds that $\mathrm{H}_\infty(\boldsymbol{x}_I) \geq \gamma|I|m$. A set $X \subseteq \Sigma^J$ is γ-spread if $\boldsymbol{x} \sim \mathsf{unif}(X)$ is γ-spread.

The reason we make this definition and maintain the spreadness property throughout the stage process is that it lets us take advantage of near-uniformity of the Inner Product gadget outputs on the free variables. Let us state this lemma now to make the motivation of this definition clear:

Lemma 12 ([GLM+16, Lemma 13])**.** *Suppose* $m \geq 100\log n$*. Let* $X \times Y \subseteq \Sigma^n \times \Sigma^n$ *and* $F \subseteq [n]$ *be such that the random variables* $\boldsymbol{x}_F$ *and* $\boldsymbol{y}_F$ *are* 0.9*-spread, for* $(\boldsymbol{x},\boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$*. Then*

$$\forall z \in \{0,1\}^F : \qquad \Pr\left[(\mathrm{IP}_m(\boldsymbol{x}_i,\boldsymbol{y}_i))_{i\in F} = z\right] \in 2^{-|F|} \cdot (1 \pm 2^{-m/20}).$$

We may now state the communication version of the Stage Lemma.

Lemma 13 (Stage Lemma)**.** *Assume Assumption 9. Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$ be a rectangle and $F \subseteq [n]$, such that: $|F| \geq 0.1n$,*

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim X\times Y}[f(\boldsymbol{x},\boldsymbol{y})=0] \geq 1/2, \qquad \text{and} \qquad X_F \text{ and } Y_F \text{ are } 0.9\text{-spread.}$$

Then there exist $X' \times Y' \subseteq X \times Y$ and $F' \subsetneq F$, such that: $|F \setminus F'| \geq \Omega(\sqrt{n})$,

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim X'\times Y'}[f(\boldsymbol{x},\boldsymbol{y})=0] \geq 1/2, \qquad \text{and} \qquad X'_{F'} \text{ and } Y'_{F'} \text{ are } 0.9\text{-spread.}$$

Moreover, there exists a set $R \subseteq F \setminus F'$ such that either $|R| = \sqrt{n}$ ("safe stage") or $R = \varnothing$ ("unsafe stage"), and

1. *For every $x, y \in X' \times Y'$, and every $i \in R$, $\mathrm{IP}_m(x_i, y_i) = 1$ (i.e., the output of every gadget inside R is fixed to 1);*
2. *$\mathcal{D}(X' \times Y', F') \leq \mathcal{D}(X \times Y, F) + o(\sqrt{n}m) - \Omega(|F \setminus (F' \cup R)| \cdot m)$.*

Here, the constants under $\Omega(\cdot)$ are universal.

To give some intuition behind the deficit inequality, it comes from

$$\mathcal{D}(X' \times Y', F') \leq \mathcal{D}(X \times Y, F) + O(|\Pi| + |R|) - \Omega(|F \setminus (F' \cup R)| \cdot m),$$

where Π refers to the protocol obtained from Yao's principle, and the deficit increases in each stage by $O(|\Pi| + |R|)$ because the entropy is reduced by $O(1)$ for each bit of communication, as well as for each gadget in R that we fix to 1 (we lose $O(1)$ bits of entropy instead of $O(m)$ because we leave the $2m$ input bits random); meanwhile, the process will also fix the $2m$ input bits for each gadget in $F \setminus (F' \cup R)$ in an uncontrolled way, to "restore" spreadness and decreases the deficit.

The remainder of this paper is dedicated to proving the Stage Lemma. Assuming this lemma, here is the proof of the main theorem:

Proof of Theorem 8 given Lemma 13. As in the proof for decision trees, we may assume without loss of generality that $\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\Sigma^n\times\Sigma^n}[f(\boldsymbol{x},\boldsymbol{y})=0] \geq \frac{1}{2}$, and we will apply the Stage Lemma iteratively over many stages,

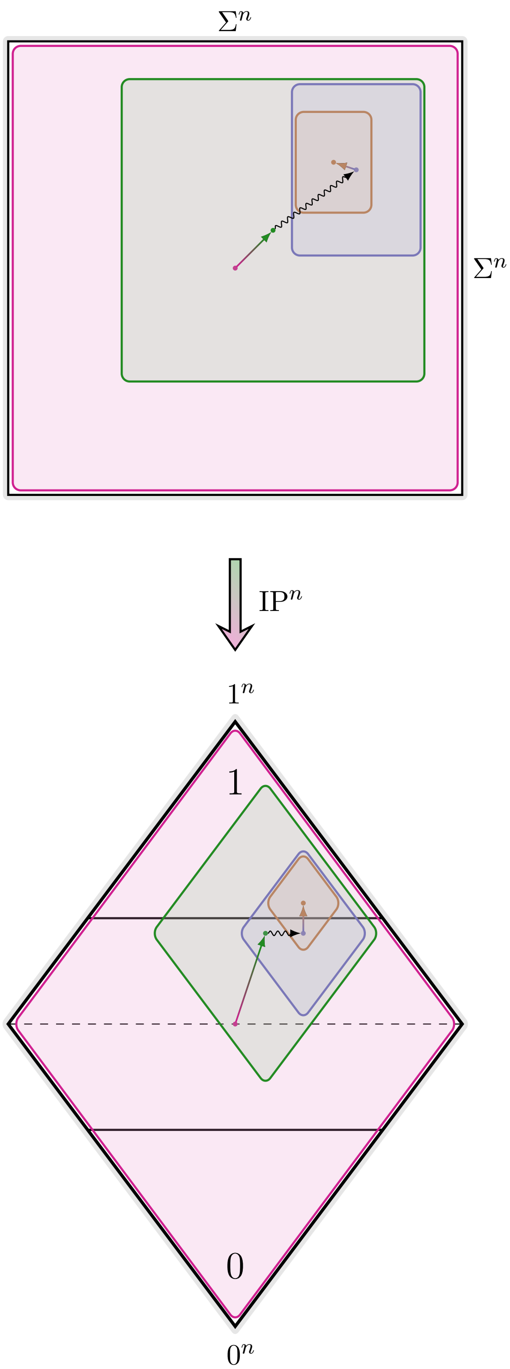
Observe that the initial value $X \times Y = \Sigma^n \times \Sigma^n$ and $F = [n]$ satisfies the condition of Lemma 13, and that $X' \times Y'$ and F' obtained from Lemma 13 maintain the conditions to reapply the lemma, as long as $|F'| \geq 0.1n$. We may therefore perform the following procedure.

Initialize

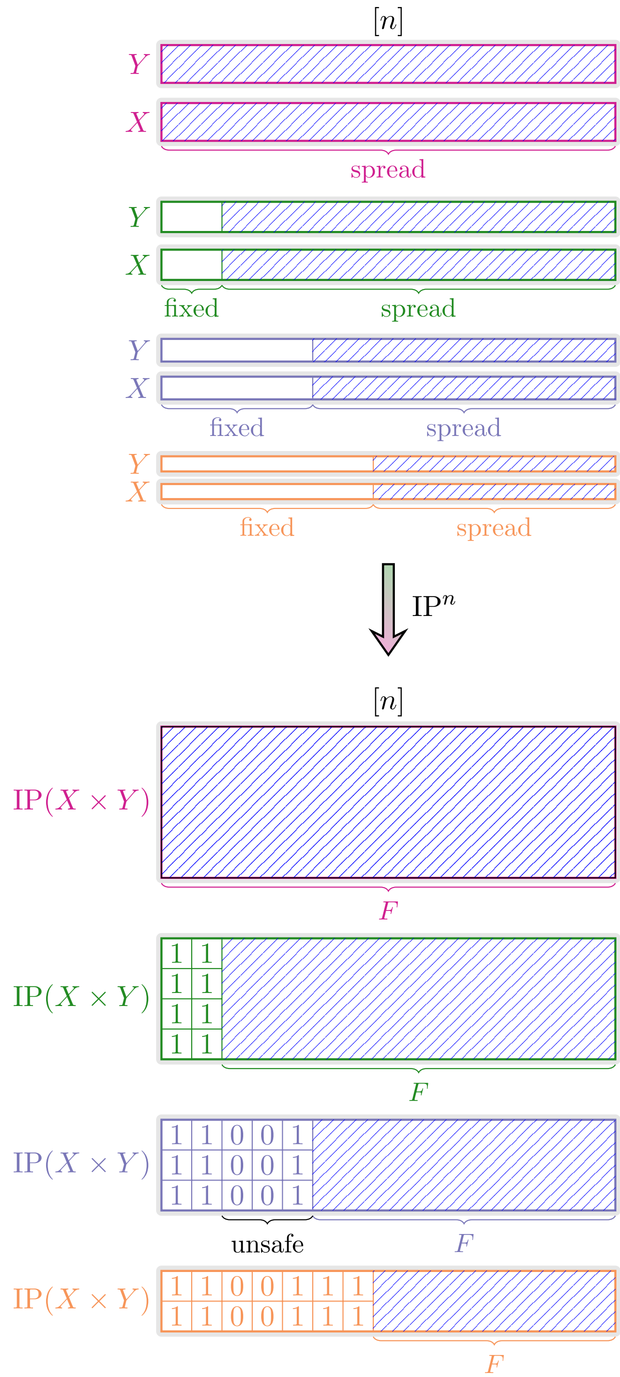
- $X \times Y = \Sigma^n \times \Sigma^n$ (the current rectangle);
- $F = [n]$ (the current "free variables");
- $T = \emptyset$ (the current set of gadgets fixed to 1);
- $U = \emptyset$ (the current set of "unfree variables" other than T).

While $|F| \geq 0.1n$, apply Lemma 13 to obtain $X' \times Y' \subseteq X \times Y$, $F' \subsetneq F$, and $R \subseteq F \setminus F'$, and update

$$X \times Y \leftarrow X' \times Y', \qquad U \leftarrow U \cup (F \setminus (F' \cup R)), \qquad F \leftarrow F', \qquad T \leftarrow T \cup R.$$



(a) Evolution of $X \times Y$



(b) Evolution of F

Figure 2: Communication stages. The second stage is *unsafe*, we assign too many coordinates

Since F shrinks in each iteration, the process halts. Upon halting, the final rectangle $X \times Y$ satisfies

$$\Pr[f(\boldsymbol{x}, \boldsymbol{y}) = 0] \geq 1/2 \tag{2}$$

for uniformly random $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$.

Note that the total number of stages is $O(\sqrt{n})$, since $|F \setminus F'| = \Omega(\sqrt{n})$, and so the number of "free variables" decreases by $\Omega(\sqrt{n})$ on every stage.

Now consider the deficit $\mathcal{D}(X \times Y, F)$ of the final rectangle with respect to the final set of free variables. By property (2) of Lemma 13, this satisfies

$$0 \leq \mathcal{D}(X \times Y, F) \leq o(\sqrt{n}m) \cdot O(\sqrt{n}) - \Omega(|U|m) \leq o(nm) - \Omega(|U|m),$$

implying $|U| \leq o(n)$ for sufficiently large n.

When the process halts, the number of free variables is at most $|F| < 0.1n$, meaning that $|T| + |U| = n - |F| \geq 0.9n$, so

$$|T| \geq 0.9n - |U| \geq 0.8n.$$

For each $i \in T$, we have ensured in property (1) of Lemma 13 that $\mathrm{IP}_m(x_i, y_i) = 1$ for all $(x, y) \in X \times Y$. Since $|T| > \frac{2}{3}n$, this means that $f(x, y) = 1$ for all $(x, y) \in X \times Y$, contradicting Equation (2). This concludes the proof. □

3.1 Proving the Stage Lemma

The structure of the proof of the Stage Lemma is shown in Figure 3, which explains how the 4 main ingredients fit together: the Protocol Transformation lemma (Lemma 15), the communication version of the Closeness Lemma (Lemma 18), the Safe Stage Lemma (Lemma 16), and the Unsafe Stage Lemma (Lemma 17). We state these lemmas formally below. Let us begin with the definition of the sprinkled-1s distribution, similar to the one we used for decision trees:

Definition 14 (Sprinkled 1s Distribution). For $X \times Y \subseteq \Sigma^n \times \Sigma^n$ and $F \subseteq [n]$, let $\sigma(X \times Y, F)$ be the *sprinkled-1s distribution* obtained by

1. Choosing a uniformly random subset $\boldsymbol{I} \sim \binom{F}{\sqrt{n}}$ of free variables, and
2. Returning $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$ conditional on the event $\mathrm{IP}_m(\boldsymbol{x}_i, \boldsymbol{y}_i) = 1$ for all $i \in \boldsymbol{I}$.

We prove the following Protocol Transformation lemma in Section 4.

Lemma 15 (Protocol Transformation Lemma)**.** *Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$, let $F \subseteq [n]$, and let Π be any deterministic protocol with cost $|\Pi| = o(\sqrt{n}m)$ and error ε over an arbitrary distribution μ. Suppose that X_F and Y_F are both 0.9-spread. Then there exists a protocol Π' with error at most ε over the same distribution, such that each leaf $L = X^{(L)} \times Y^{(L)}$ of Π' is associated with a subset $F_L \subseteq F$ of free variables, where:*

1. *For every leaf L, the random variable $(\boldsymbol{x}^{(L)}, \boldsymbol{y}^{(L)}) \sim \mathsf{unif}(X^{(L)} \times Y^{(L)})$ has $\boldsymbol{x}^{(L)}_{F_L}$ and $\boldsymbol{y}^{(L)}_{F_L}$ both 0.9-spread; and*
2. *For a random leaf $\boldsymbol{L} := L(\boldsymbol{x}, \boldsymbol{y})$ defined as the leaf of a random $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$, with probability at least $1 - 4\varepsilon$,*

$$\mathcal{D}(\boldsymbol{L}, F_{\boldsymbol{L}}) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F \setminus F_{\boldsymbol{L}}| \cdot m),$$

where the constants under $O(\cdot)$ and $\Omega(\cdot)$ are universal.

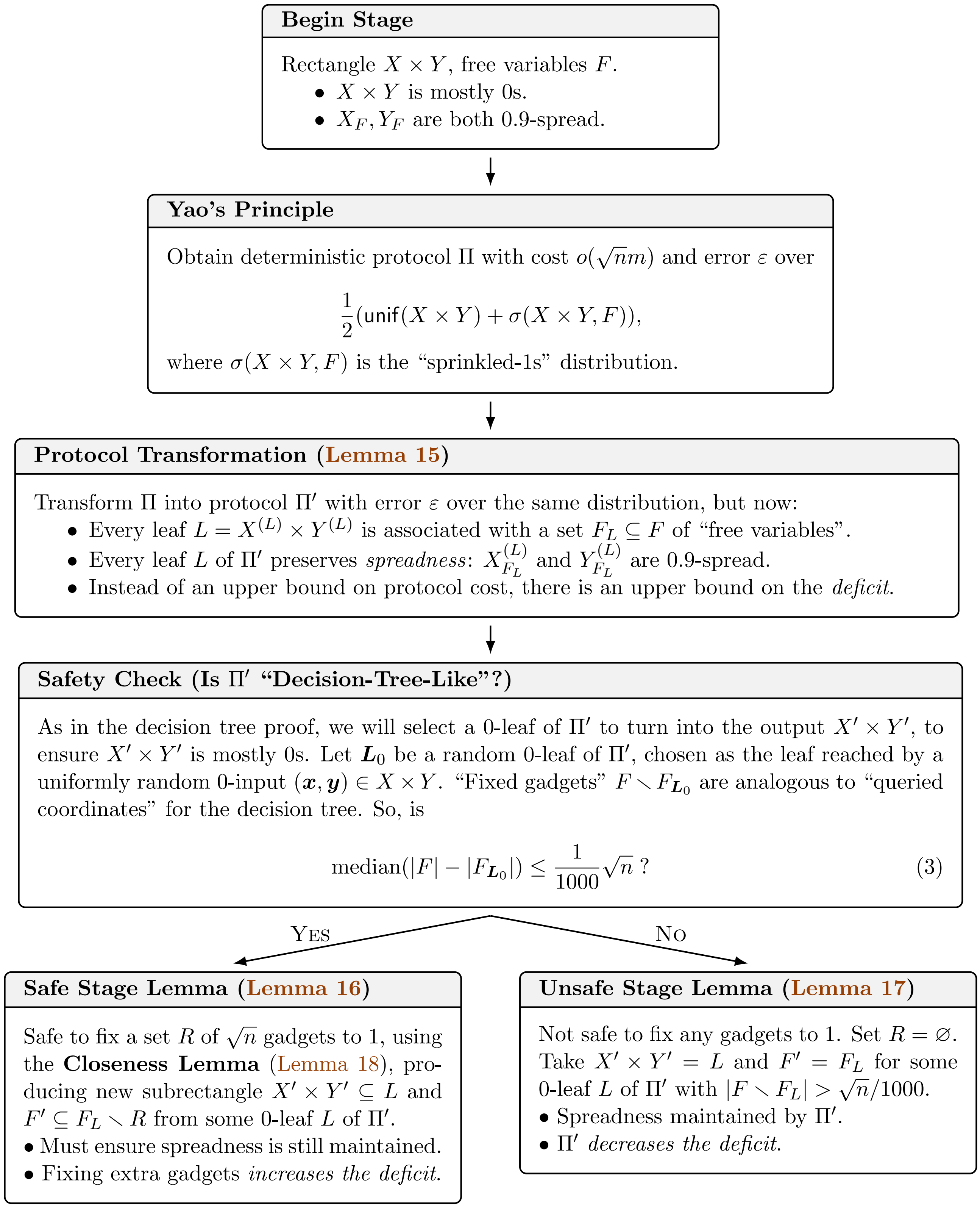


Figure 3: Outline of the proof of the Stage Lemma.

Lemma 16 (Safe Stage Lemma)**.** *Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$ and $F \subseteq [n]$ satisfy the conditions of Lemma 13, and assume the current stage is* safe*, i.e., Equation (3) holds. Then the conclusion of Lemma 13 holds with $|R| = \sqrt{n}$.*

Lemma 17 (Unsafe Stage Lemma)**.** *Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$ and $F \subseteq [n]$ satisfy the conditions of Lemma 13, and assume the current stage is* unsafe*, i.e., Equation (3) does not hold. Then the conclusion of Lemma 13 holds with $R = \emptyset$.*

The Stage Lemma (Lemma 13) follows immediately from Lemmas 16 and 17.

3.2 Proof of the Safe Stage Lemma

In this section we state the main ingredients required for the Safe Stage Lemma, including the communication version of the Closeness Lemma, and we prove the lemma using those ingredients. We defer the proofs of each ingredient to Section 5.

We start with the rectangle $X \times Y$ and free variables $F \subseteq [n]$. Recall that Π' is the transformed protocol from the Protocol Transformation Lemma, which has error ε over the distribution

$$\frac{1}{2}(\mathsf{unif}(X \times Y) + \sigma(X \times Y, F)).$$

Suppose that the current stage is safe, i.e., Equation (3) holds. We want to find a subset $R \subseteq F$ of the free variables where we can fix all of the IP gadget values to 1, and produce a new rectangle $X' \times Y' \subseteq X \times Y$ and new free variables $F' \subseteq F \setminus R$ satisfying the conditions to repeat the process in Lemma 13.

The crucial step is the "birthday paradox" step, where we show that we can sprinkle $\sqrt{n}$ random 1-valued gadgets without the protocol Π' noticing. This is formalized in the next lemma, which uses spreadness of $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$ on the free variables F, together with the pseudorandomness of the Inner Product gadget under the spreadness condition (Lemma 12). We prove the Closeness Lemma in Section 5.1.

Lemma 18 (Closeness Lemma)**.** *Assume $m \geq 100 \log n$. Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$ and let $F \subseteq [n]$ be such that the random variables $\boldsymbol{x}_F$ and $\boldsymbol{y}_F$ are both 0.9-spread, where $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$. Now let $X' \times Y' \subseteq X \times Y$ and $F' \subseteq F$ also be such that $\boldsymbol{x}'_{F'}, \boldsymbol{y}'_{F'}$ are both 0.9-spread, for $(\boldsymbol{x}', \boldsymbol{y}') \sim \mathsf{unif}(X' \times Y')$, and suppose $|F| \geq 0.1n$ and $|F| - |F'| \leq \sqrt{n}/1000$. Then*

$$\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \sigma(X \times Y, F)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y'] \in (1 \pm 0.1) \Pr_{(\boldsymbol{x},\boldsymbol{y}) \in \mathsf{unif}(X \times Y)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y'].$$

Moreover,

$$\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \sigma(X \times Y, F)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y' \mid \boldsymbol{I} \subseteq F'] \in (1 \pm 0.1) \Pr_{(\boldsymbol{x},\boldsymbol{y}) \in \mathsf{unif}(X \times Y)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y'],$$

where $\boldsymbol{I} \sim \binom{F}{\sqrt{n}}$ denotes the random set of coordinates used to generate $(\boldsymbol{x}, \boldsymbol{y})$ in the distribution $\sigma(X \times Y, F)$.

Informally, this lemma claims that if we start from a big enough spread rectangle in $\Sigma^n \times \Sigma^n$, a typical leaf of a shallow protocol will not be able to distinguish between uniform distribution and "sprinkled 1s" distribution.

With the Closeness Lemma, finding the target rectangle $X' \times Y'$ together with its free variables $F' \subseteq F$ and set $R \subseteq F \setminus F'$ of the 1-valued gadgets is done in three steps, given by the next three propositions. The first proposition claims that it is possible to find a 0-leaf in the protocol, such that it is mostly correct under the sprinkled distribution, and the deficiency would not grow too much.

Proposition 19 (Sprinkle 1s)**.** *Let $\varepsilon > 0$ be a sufficiently small constant. Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$ and $F \subseteq [n]$ satisfy the conditions of Lemma 13. If the current stage is safe, then there exists a leaf L of Π' such that:*

1. *L contains mostly 0 inputs under the sprinkled-1s distribution:*

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\sigma(X\times Y,F)}[f(\boldsymbol{x},\boldsymbol{y}) = 0 \mid (\boldsymbol{x},\boldsymbol{y}) \in L] \geq 1 - 10\varepsilon,$$

2. $|F| - |F_L| \leq \sqrt{n}/1000.$
3. $\mathcal{D}(L, F_L) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|).$

The next proposition states that we can fix a particular outcome for sprinkling it 1s; again, without increasing the deficiency too much.

Proposition 20 (Fix 1s)**.** *Under the conditions of Proposition 19, let L be a leaf of Π' obtained from that proposition, and let F_L be its associated set of free variables (as defined in Lemma 15). Then there exists $X' \times Y' \subseteq L$, $R \subseteq F_L$ with $|R| = \sqrt{n}$, and $F' := F_L \setminus R$ such that*

1. $\Pr[f(\boldsymbol{x},\boldsymbol{y}) = 0] \geq 0.6$ *for* $(\boldsymbol{x},\boldsymbol{y}) \sim \mathsf{unif}(X' \times Y')$;
2. *For all $j \in R$ and all $x, y \in X' \times Y'$, $\mathrm{IP}_m(x_j, y_j) = 1$, and*
3. $\mathcal{D}(X' \times Y', F') \leq \mathcal{D}(L, F_L) + 2|R|.$

While the leaf L of Π' is spread (due to the protocol transformation), in the the previous step we moved to a subrectangle $X' \times Y' \subseteq L$ by fixing the R gadgets to 1, which may not have preserved spreadness. We regain spreadness by the same technique as in the Protocol Transformation Lemma.

Proposition 21 (Spreadify)**.** *Let $A \times B \subseteq \Sigma^n \times \Sigma^n$ and $H \subseteq [n]$ such that $\Pr_{\boldsymbol{a},\boldsymbol{b}\sim A\times B}[f(\boldsymbol{a},\boldsymbol{b}) = 0] \geq 0.6$. Then there exists $H' \subseteq H$ and a rectangle $A' \times B' \subseteq A \times B$ such that*

1. $\Pr_{\boldsymbol{a},\boldsymbol{b}\sim A'\times B'}[f(\boldsymbol{a},\boldsymbol{b}) = 0] \geq 1/2.$
2. *$A'_{H'}$ and $B'_{H'}$ are 0.9-spread.*
3. $\mathcal{D}(A' \times B', H') \leq \mathcal{D}(A \times B, H) - \Omega(|H \setminus H'|m).$

We prove these propositions in Section 5.2. With these tools, we can complete the proof of the safe stage lemma.

Proof of the Safe Stage Lemma. Start with $X \times Y$ and F satisfying the conditions of Lemma 13. Let L be the leaf of Π' with free variables F_L obtained from Proposition 19. Let $X' \times Y'$, F', and $R = F \setminus F'$ be from Proposition 20 and let $X'' \times Y'' \subseteq X' \times Y'$ and $F'' \subseteq F'$ be obtained from applying Proposition 21 to $X' \times Y'$ and F'.

From Proposition 20 we guarantee that $\mathrm{IP}_m(x_i, y_i) = 1$ for every $(x, y) \in X'' \times Y''$ and every $i \in R$. From Propositions 20 and 21 we get

$$\begin{aligned}
\mathcal{D}(X'' \times Y'', F'') &\leq \mathcal{D}(X' \times Y', F') - \Omega(|F' \setminus F''|m) && \text{(Proposition 21)} \\
&\leq \mathcal{D}(L, F_L) + 2|R| - \Omega(|F' \setminus F''|m) && \text{(Proposition 20)} \\
&\leq \mathcal{D}(X \times Y, F) + O(|\Pi|) + 2|R| - \Omega(|F' \setminus F''|m) && \text{(Proposition 19)} \\
&\leq \mathcal{D}(X \times Y, F) + o(\sqrt{n}m) - \Omega(|F' \setminus F''|m) && (|\Pi|, |R| = o(\sqrt{n}m)) \\
&\leq \mathcal{D}(X \times Y, F) + o(\sqrt{n}m) - \Omega(|F \setminus (F'' \cup R)|m)
\end{aligned}$$

This concludes the proof of Lemma 16. □

3.3 Proof of the Unsafe Stage Lemma

A stage is *unsafe* if Equation (3) is false. In this case, we choose any 0-valued leaf L_0 with $|F| - |F_{L_0}|$ at least the median in Equation (3), and which also contains mostly 0-valued inputs of f; we must show this exists.

Recall that $\boldsymbol{L}_0 = (\boldsymbol{L} \mid \boldsymbol{L} \text{ outputs } 0)$ is a random leaf drawn by taking the unique 0-leaf containing $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$ conditional on $(\boldsymbol{x}, \boldsymbol{y})$ ending up in a 0-valued leaf. Suppose Equation (3) is false. Let $R = \emptyset$. Then, to conclude the lemma, we want to show that there exists a 0-valued leaf L_0 of Π' with $|F| - |F_{L_0}| \geq \text{median}\,(|F| - |F_{\boldsymbol{L_0}}|) > \sqrt{n}/1000$, such that

- $\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \mathsf{unif}(L_0)}[f(\boldsymbol{x}, \boldsymbol{y}) = 0] \geq 1/2$, and
- $\mathcal{D}(L_0, F_{L_0}) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F \setminus F_{L_0}| \cdot m)$.

Let $\boldsymbol{L}_0'$ be an independent copy of $\boldsymbol{L}_0$. We upper bound the probabilities for the following events:

(E1) $|F| - |F_{\boldsymbol{L}_0'}| < \text{median}\,(|F| - |F_{\boldsymbol{L_0}}|)$;
(E2) $\text{ERROR}_{\boldsymbol{L}_0'} > 1/2$, where $\text{ERROR}_{\boldsymbol{L}_0'} \coloneqq \Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \mathsf{unif}(\boldsymbol{L}_0')}[f(\boldsymbol{x}, \boldsymbol{y}) = 1]$.
(E3) $\mathcal{D}(\boldsymbol{L}_0', F_{\boldsymbol{L}_0'}) > \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F \setminus F_{\boldsymbol{L}_0'}| \cdot m)$.

For (E1), by definition, $\Pr_{\boldsymbol{L}_0'}[|F| - |F_{\boldsymbol{L}_0'}| < \text{median}\,(|F| - |F_{\boldsymbol{L_0}}|)] < 1/2$.

For (E2), since Π' has error ε over the mixture $\frac{1}{2}\,(\mathsf{unif}(X \times Y) + \sigma(X \times Y, F))$, it has error at most 2ε over $\mathsf{unif}(X \times Y)$. Therefore $\mathbb{E}_{\boldsymbol{L}_0'}[\text{ERROR}_{\boldsymbol{L}_0'}] \leq 4\varepsilon$, so it follows by Markov inequality that $\Pr_{\boldsymbol{L}_0'}[\text{ERROR}_{\boldsymbol{L}_0'} \geq 20\varepsilon] \leq \frac{1}{5}$. For a small enough ε, we can conclude that:

$$\Pr_{\boldsymbol{L}_0'}\left[\text{ERROR}_{\boldsymbol{L}_0'} > \frac{1}{2}\right] = \Pr_{\boldsymbol{L}_0'}\left[\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \mathsf{unif}(\boldsymbol{L}_0')}[f(\boldsymbol{x}, \boldsymbol{y}) = 0] < 1/2\right] \leq \frac{1}{5}.$$

For (E3), we are guaranteed from Lemma 15 that

$$\mathcal{D}(\boldsymbol{L}, F_{\boldsymbol{L}}) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F \setminus F_{\boldsymbol{L}}|m) \tag{4}$$

with probability at least $1 - 4\varepsilon$ over a random leaf $\boldsymbol{L}$ chosen as the leaf containing $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$. We want this to hold for the random 0-leaf $\boldsymbol{L}_0'$. Since $\Pr[f(\boldsymbol{x}, \boldsymbol{y}) = 0] \geq 1/2$ and the protocol has error at most 2ε over $(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{L}$ is a 0-leaf with probability at least $1/2 - 2\varepsilon$. Then Equation (4) must hold with probability at least $1 - 10\varepsilon$ for $\boldsymbol{L}_0'$ instead of $\boldsymbol{L}$, when ε is sufficiently small.

Applying the union bound, we can conclude that with positive probability, none of the three events occur. Then there exists a 0-leaf L_0 that satisfies the properties stated in the claim.

4 Density Restoring Partitions and Protocol Transformation

4.1 Density Restoring Partitions

The following is [GPW20, Lemma 3.5] (with straightforward adaptation to allow for $\boldsymbol{x}$ non-uniform, see also [CFK+21]):

Lemma 22 (Density Restoring Partition)**.** *Let $\boldsymbol{x} \in \Sigma^n$ be a random variable with support X, let $F \subseteq [n]$, and let $\gamma \in (0,1)$. Then there exists a partition $X = \bigsqcup_{j=1}^{r} X_j$ with associated sets $I_j \subseteq F$ and values $\alpha_j \in \Sigma^{I_j}$ such that:*

- $X_j \coloneqq \{x \in X \mid x_{I_j} = \alpha_j\} \smallsetminus \bigcup_{i<j} X_i$*;*
- $(\boldsymbol{x}_{F \smallsetminus I_j} \mid \boldsymbol{x} \in X_j)$ *is γ-spread;*
- $\mathrm{D}_\infty(\boldsymbol{x}_{F \smallsetminus I_j} \mid \boldsymbol{x} \in X_j) \leq \mathrm{D}_\infty(\boldsymbol{x}_F) - (1-\gamma)|I_j|m + \delta_j$ *where* $\delta_j = \log\left(\frac{1}{\Pr[\boldsymbol{x} \in \cup_{k \geq j} X_k]}\right)$.

We define a procedure to apply the density restoring partition to rectangles $X \times Y$:

Lemma 23 (Density Restoring Partition for Rectangles)**.** *Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$, $F \subseteq [n]$, $\varepsilon, \gamma \in (0,1)$. Then there exist partitions $X = \bigsqcup_{i=1}^{a} X_i$ and $Y = \bigsqcup_{j=1}^{b} Y_j$ where each rectangle $X_i \times Y_j \subseteq X \times Y$ is associated with a set $F_{i,j} \subseteq F$ of "free variables" which satisfy the following properties:*

1. *For all i, j, the random variables $(\boldsymbol{x}_{F_{i,j}} \mid \boldsymbol{x} \in X_i)$ and $(\boldsymbol{y}_{F_{i,j}} \mid \boldsymbol{y} \in Y_j)$ are both γ-spread, where $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$;*
2. *With probability at least $1 - 2\varepsilon$ over $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$, the unique rectangle $X_{\boldsymbol{i}} \times Y_{\boldsymbol{j}}$ containing $(\boldsymbol{x}, \boldsymbol{y})$ satisfies*

$$\mathcal{D}(X_{\boldsymbol{i}} \times Y_{\boldsymbol{j}}, F_{\boldsymbol{i},\boldsymbol{j}}) \leq \mathcal{D}(X \times Y, F) - (1-\gamma)|F \smallsetminus F_{\boldsymbol{i},\boldsymbol{j}}|m + 2\log(1/\varepsilon).$$

Proof. Begin by applying the density restoring partition of Lemma 22 to each of X and Y to obtain

$$X = \bigsqcup_i X_i, \qquad \text{and} \qquad Y = \bigsqcup_j Y_j,$$

where for each i, j there exist sets $I_i, J_j \subseteq F$ and assignments $\alpha_i \in \Sigma^{I_i}$, $\beta_j \in \Sigma^{J_j}$ such that

$$X_i \coloneqq \{x \in X \mid x_{I_i} = \alpha_i\} \smallsetminus \bigcup_{k<i} X_k, \qquad \text{and} \qquad Y_j \coloneqq \{y \in Y \mid y_{J_j} = \beta_j\} \smallsetminus \bigcup_{k<j} Y_k,$$

where $(\boldsymbol{x}_{F \smallsetminus I_i} \mid \boldsymbol{x} \in X_i)$ and $(\boldsymbol{y}_{F \smallsetminus J_j} \mid \boldsymbol{y} \in Y_j)$ are both γ-spread for $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$. We define the set $F_{i,j}$ associated with $X_i \times Y_j$ as

$$F_{i,j} \coloneqq F \smallsetminus (I_i \cup J_j),$$

so that the "free variables" are those which are not assigned by either one of the density restoring partitions. Since $F_{i,j} \subseteq F \smallsetminus I_i$ and $F_{i,j} \subseteq F \smallsetminus J_j$, the random variables

$$(\boldsymbol{x}_{F_{i,j}} \mid \boldsymbol{x} \in X_i) \qquad \text{and} \qquad (\boldsymbol{y}_{F_{i,j}} \mid \boldsymbol{y} \in Y_j)$$

remain γ-spread, establishing the first property of the lemma. Let us verify the second property. For every fixed rectangle $X_i \times Y_j$, we have from Lemma 22 that

$$\mathrm{D}_\infty(\boldsymbol{x}_{F\setminus I_i} \mid \boldsymbol{x} \in X_i) \le \mathrm{D}_\infty(\boldsymbol{x}_F) - (1-\gamma)|I_j|m + \log\left(\frac{1}{\Pr\left[\boldsymbol{x'} \in \bigcup_{k\ge i} X_k\right]}\right),$$

where $\boldsymbol{x'} \sim \mathsf{unif}(X)$. Since $F_{i,j} = F \setminus (I_i \cup J_j)$, we can use Proposition 25 (stated below) to bound $\mathrm{D}_\infty(\boldsymbol{x}_{F_{i,j}}) \le \mathrm{D}_\infty(\boldsymbol{x}_{F\setminus I_i})$. Then the remaining task to bound the final term.

Claim 24. *With probability at least $1-\varepsilon$ over $\boldsymbol{x} \sim \mathsf{unif}(X)$, the unique value $\boldsymbol{i} \in [a]$ such that $\boldsymbol{x} \in X_{\boldsymbol{i}}$ satisfies $\log\left(\frac{1}{\Pr\left[\boldsymbol{x'}\in\bigcup_{k\ge \boldsymbol{i}} X_k\right]}\right) < \log(1/\varepsilon)$.*

Proof of claim. We may think of the distribution over the index $\boldsymbol{i} \in [a]$ defined by $p(i) := \Pr[\boldsymbol{i} = i] = \Pr[\boldsymbol{x} \in X_i]$. Then for all $t \in [0,1]$,

$$\Pr_{\boldsymbol{x}}\left[\log\left(\frac{1}{\Pr\left[\boldsymbol{x'} \in \bigcup_{k\ge i} X_k\right]}\right) > t\right] = \Pr_{\boldsymbol{i}}\left[\log\left(\frac{1}{\sum_{k\ge \boldsymbol{i}} p(k)}\right) > t\right] = \Pr_{\boldsymbol{i}}\left[\sum_{k\ge \boldsymbol{i}} p(k) < 2^{-t}\right].$$

Let i^* be the smallest value such that $\sum_{k\ge i^*} p(k) < 2^{-t}$. If $\sum_{k\ge \boldsymbol{i}} p(k) < 2^{-t}$ then $\boldsymbol{i} \ge i^*$ so the probability of this event is at most $\sum_{k\ge i^*} p(k) < 2^{-t}$. Setting $t = \log(1/\varepsilon)$ produces the desired bound. □

Applying the same reasoning to $\boldsymbol{y}$ and using the union bound over $\boldsymbol{x}$ and $\boldsymbol{y}$, we have with probability at least $1-2\varepsilon$ the inequality

$$\mathrm{D}_\infty(\boldsymbol{x}_{F_{\boldsymbol{i},\boldsymbol{j}}} \mid \boldsymbol{x} \in X_{\boldsymbol{i}}) + \mathrm{D}_\infty(\boldsymbol{y}_{F_{\boldsymbol{i},\boldsymbol{j}}} \mid \boldsymbol{y} \in Y_{\boldsymbol{j}}) \le \mathrm{D}_\infty(\boldsymbol{x}_F) + \mathrm{D}_\infty(\boldsymbol{y}_F) - (1-\gamma)(|I_{\boldsymbol{i}}| + |J_{\boldsymbol{j}}|)m + 2\log(1/\varepsilon),$$

which concludes the proof since $|I_{\boldsymbol{i}}| + |J_{\boldsymbol{j}}| \ge |I_{\boldsymbol{i}} \cup J_{\boldsymbol{j}}|$. □

We used the following proposition in the proof above:

Proposition 25. *Let $\boldsymbol{x}$ be a random variable over Σ^n and let $J \subseteq F \subseteq [n]$. Then*

$$\mathrm{D}_\infty(\boldsymbol{x}_{F\setminus J}) \le \mathrm{D}_\infty(\boldsymbol{x}_F).$$

Proof. First, observe

$$\max_{z\in\Sigma^{F\setminus J}} \Pr[\boldsymbol{x}_{F\setminus J} = z] = \max_{z\in\Sigma^{F\setminus J}} \sum_{w\in\Sigma^F, w_J = z} \Pr[\boldsymbol{x}_F = w] \le |\Sigma|^{|J|} \max_{w\in\Sigma^F} \Pr[\boldsymbol{x}_F = w],$$

so that

$$\mathrm{H}_\infty(\boldsymbol{x}_{F\setminus J}) \ge \mathrm{H}_\infty(\boldsymbol{x}_F) - |J|m.$$

Then

$$\mathrm{D}_\infty(\boldsymbol{x}_{F\setminus J}) = (|F| - |J|)m - \mathrm{H}_\infty(\boldsymbol{x}_{F\setminus J}) \le |F|m - \mathrm{H}_\infty(\boldsymbol{x}_F) = \mathrm{D}_\infty(\boldsymbol{x}_F).$$ □

4.2 Protocol Transformation

To prove the Protocol Transformation Lemma, we must relate the growth of the deficit function to the communication complexity of the protocol Π.

Proposition 26. *Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$, let $F \subseteq [n]$, and let Π be any communication protocol with cost d. Then for any $\varepsilon > 0$, with probability at least $1 - 2\varepsilon$ over a random leaf $\boldsymbol{L}$ of Π chosen as the leaf reached by $(\boldsymbol{x}', \boldsymbol{y}') \sim X \times Y$,*

$$\mathcal{D}(\boldsymbol{L}, F) \leq \mathcal{D}(X \times Y, F) + 2d + 2\log(1/\varepsilon).$$

To prove the statement, we will need a chain rule for min-entropy:

Lemma 27 (see e.g. [Vad12, Lemma 6.30])**.** *Let $(\boldsymbol{a}, \boldsymbol{b})$ be distributed over $A \times B$ with $|A| = 2^\ell$ and $\mathrm{H}_\infty(\boldsymbol{a}, \boldsymbol{b}) \geq k$, and let $\mathcal{A}$ denote the marginal distribution of $\boldsymbol{a}$. Then for every $\varepsilon > 0$*

$$\Pr_{\boldsymbol{a}' \sim \mathcal{A}}[\mathrm{H}_\infty(\boldsymbol{b} \mid \boldsymbol{a} = \boldsymbol{a}') \geq k - \ell - \log 1/\varepsilon] \geq 1 - \varepsilon.$$

Proof. Fix some $a \in A, b \in B$ then $\Pr[\boldsymbol{a} = a \wedge \boldsymbol{b} = b] / \Pr[\boldsymbol{a} = a] = \Pr[\boldsymbol{b} = b \mid \boldsymbol{a} = a]$, so $\mathrm{H}_\infty(\boldsymbol{b} \mid \boldsymbol{a} = a) \geq \mathrm{H}_\infty(\boldsymbol{a}, \boldsymbol{b}) - \log 1/\Pr[\boldsymbol{a} = a]$. Then defining $p(a) \coloneqq \Pr[\boldsymbol{a} = a]$ we have

$$\Pr[\log 1/p(\boldsymbol{a}) \geq \ell + \log 1/\varepsilon] = \Pr[p(\boldsymbol{a}) \leq \varepsilon/|A|] = \sum_{a\colon p(a) \leq \varepsilon/|A|} p(a) \leq \varepsilon. \qquad \square$$

Proof of Proposition 26. The claim is equivalent to the statement that, with probability at least $1 - 2\varepsilon$ over a random leaf $\boldsymbol{L}$,

$$\mathrm{H}_\infty(\boldsymbol{x}_F \mid (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{L}) + \mathrm{H}_\infty(\boldsymbol{y}_F \mid (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{L}) \geq \mathrm{H}_\infty(\boldsymbol{x}_F) + \mathrm{H}_\infty(\boldsymbol{y}_F) - 2d - 2\log(1/\varepsilon).$$

Let $t(x, y)$ denote the transcript of protocol Π on input x, y, so that $t(x, y) \in \{0, 1\}^d$, and observe that transcript $t(x, y)$ is in one-to-one correspondence with the leaf L of (x, y). Consider the random variable $(t(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{x}_F)$. By the chain rule for min-entropy (Lemma 27), with probability at least $1-\varepsilon$ over a random transcript $\boldsymbol{t}' = t(\boldsymbol{x}', \boldsymbol{y}')$

$$\mathrm{H}_\infty(\boldsymbol{x}_F \mid t(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{t}') \geq \mathrm{H}_\infty(\boldsymbol{x}_F) - d - \log(1/\varepsilon).$$

Using the same argument for $\boldsymbol{y}$ and applying the union bound, we get the desired conclusion. $\square$

We may now prove the protocol transformation lemma.

Proof of Protocol Transformation Lemma. Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$, let $F \subseteq [n]$, and let Π be a deterministic protocol with cost $o(\sqrt{n}m)$ and error ε over μ. We define the protocol Π' in Algorithm 1, which we think of as outputting a transcript (i.e., a leaf L) along with a value.

By definition, on any input the protocol outputs the same value as the original protocol Π, so its output is correct with the probability $1 - \varepsilon$ on μ.

For each leaf $L' = U_i^{(L)} \times V_j^{(L)}$ of Π', we associate the free variables $F_{L'} = F_{i,j}^{(L)} \subseteq F$ given by Lemma 23. The first desired property of Lemma 15 is that for every leaf $U_i^{(L)} \times V_j^{(L)}$ of Π', the variables

$$(\boldsymbol{x}_{F_{i,j}^{(L)}} \mid (\boldsymbol{x}, \boldsymbol{y}) \sim U_i^{(L)} \times V_j^{(L)}) \qquad \text{and} \qquad (\boldsymbol{y}_{F_{i,j}^{(L)}} \mid (\boldsymbol{x}, \boldsymbol{y}) \sim U_i^{(L)} \times V_j^{(L)})$$

Algorithm 1 The transformed protocol Π'

Input: $(x, y) \in X \times Y$

Output: Leaf $L'(x, y)$ and value $b = \Pi(x, y)$

Run protocol Π on (x, y) to obtain a leaf $L = L(x, y)$ and a value b.

Apply the density restoring partition for rectangles (Lemma 23) to $L = U^{(L)} \times V^{(L)}$ and F, to obtain

$$U^{(L)} = \bigsqcup_i U_i^{(L)}, \qquad \text{and} \qquad V^{(L)} = \bigsqcup_j V_j^{(L)}.$$

Alice sends value i for the unique $U_i^{(L)}$ containing x.

Bob sends value j for the unique $V_j^{(L)}$ containing y.

The players output leaf $U_i^{(L)} \times V_j^{(L)}$ and value b.

---

are 0.9-spread. This is immediately guaranteed by Lemma 23. The second desired property of Lemma 15 is that a random leaf $\boldsymbol{L'} = U_{\boldsymbol{i}}^{(\boldsymbol{L})} \times V_{\boldsymbol{j}}^{(\boldsymbol{L})}$ satisfies

$$\mathcal{D}(\boldsymbol{L'}, F_{\boldsymbol{L'}}) \le \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F \setminus F_{\boldsymbol{L'}}|)m$$

with probability at least $1 - 4\varepsilon$, where constants under $O(\cdot)$ and $\Omega(\cdot)$ (C_1 and C_2, respectively) are universal. The random leaf $\boldsymbol{L'}$ is obtained by first selecting a random leaf $\boldsymbol{L}$ of Π by taking the leaf containing $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$, and then conditional on $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{L}$ taking the part $U_{\boldsymbol{i}}^{(\boldsymbol{L})} \times V_{\boldsymbol{j}}^{(\boldsymbol{L})}$ containing $(\boldsymbol{x}, \boldsymbol{y})$.

By Proposition 26, with probability at least $1 - 2\varepsilon$ over $\boldsymbol{L}$, we have

$$\mathcal{D}(\boldsymbol{L}, F) \le \mathcal{D}(X \times Y, F) + 2|\Pi| + 2\log(1/\varepsilon).$$

Now for any fixed leaf L of Π, Lemma 23 guarantees that with probability at least $1 - 2\varepsilon$ over $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$ conditional on $(\boldsymbol{x}, \boldsymbol{y}) \in L$, we have

$$\mathcal{D}(U_{\boldsymbol{i}}^{(L)} \times V_{\boldsymbol{j}}^{(L)}, F_{\boldsymbol{i},\boldsymbol{j}}^{(L)}) \le \mathcal{D}(L, F) - 0.1|F \setminus F_{\boldsymbol{i},\boldsymbol{j}}^{(L)}|m + 2\log(1/\varepsilon).$$

Therefore, by the union bound, we may combine these inequalities to obtain, with probability at least $1 - 4\varepsilon$ over the random leaf $\boldsymbol{L'}$ of Π',

$$\mathcal{D}(\boldsymbol{L'}, F_{\boldsymbol{L'}}) \le \mathcal{D}(X \times Y, F) + 2|\Pi| - 0.1|F \setminus F_{\boldsymbol{L'}}|m + 4\log(1/\varepsilon).$$

For any fixed $\varepsilon > 0$ and sufficiently large n, $4\log(1/\varepsilon) < |\Pi|$, so we get the conclusion with $C_1 = 3$ and $C_2 = 0.1$. □

5 Completing the Safe Stage Lemma

5.1 Closeness Lemma

In this section, we show that the sprinkled-1s distribution in Definition 14 is close to uniform for the "good leaves", i.e., the leaves L where we fix at most $\sqrt{n}/1000$ new variables ($|F| - |F_L| \le \sqrt{n}/1000$). The intuition of this proof is that if the communication protocol "fixes" at most $\sqrt{n}/1000$ variables within one stage, then this is similar to "querying" at most $\sqrt{n}/1000$ gadgets, and therefore by a birthday paradox argument we may sprinkle a random set of $\sqrt{n}$ gadgets fixed to 1 without affecting the distribution over leaves of the protocol. We require the following birthday paradox bounds:

Claim 28. *Let $\boldsymbol{I} \sim \binom{[\ell]}{s}$ be a uniformly random set of s elements drawn from $[\ell]$ and let $T \subseteq [\ell]$ be such that $s \cdot |T| \leq \ell/100$. Then*

$$\Pr[\boldsymbol{I} \cap T = \emptyset] \geq 1 - 1/100 \qquad \textit{and} \qquad \mathbb{E}[2^{|\boldsymbol{I} \cap T|}] \leq 1 + 1/20.$$

Proof. Write $\boldsymbol{I} = \{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_k\}$. For any fixed k we estimate

$$\Pr[|\boldsymbol{I} \cap T| \geq k] = \Pr[\exists K \subseteq [s] : |K| = k \wedge \ \forall j \in K, \ \ \boldsymbol{i}_j \in T] \leq \binom{s}{k}\left(\frac{|T|}{\ell}\right)^k \leq \frac{(s \cdot |T|)^k}{\ell^k} \leq 100^{-k}.$$

Then $\Pr[\boldsymbol{I} \cap T = \emptyset] \geq 1 - 1/100$ and

$$\mathbb{E}[2^{|\boldsymbol{I} \cap T|}] \leq 1 + \sum_{k=1}^{\ell} 2^k \Pr[|\boldsymbol{I} \cap T| \geq k] \leq 1 + \sum_{k=1}^{\ell} \frac{1}{50^k} \leq 1 + \frac{1}{20}.$$ □

Lemma 18 (Closeness Lemma)**.** *Assume $m \geq 100 \log n$. Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$ and let $F \subseteq [n]$ be such that the random variables $\boldsymbol{x}_F$ and $\boldsymbol{y}_F$ are both 0.9-spread, where $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$. Now let $X' \times Y' \subseteq X \times Y$ and $F' \subseteq F$ also be such that $\boldsymbol{x}'_{F'}, \boldsymbol{y}'_{F'}$ are both 0.9-spread, for $(\boldsymbol{x}', \boldsymbol{y}') \sim \mathsf{unif}(X' \times Y')$, and suppose $|F| \geq 0.1n$ and $|F| - |F'| \leq \sqrt{n}/1000$. Then*

$$\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \sigma(X \times Y, F)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y'] \in (1 \pm 0.1) \Pr_{(\boldsymbol{x},\boldsymbol{y}) \in \mathsf{unif}(X \times Y)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y'].$$

Moreover,

$$\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \sigma(X \times Y, F)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y' \mid \boldsymbol{I} \subseteq F'] \in (1 \pm 0.1) \Pr_{(\boldsymbol{x},\boldsymbol{y}) \in \mathsf{unif}(X \times Y)}[(\boldsymbol{x}, \boldsymbol{y}) \in X' \times Y'],$$

where $\boldsymbol{I} \sim \binom{F}{\sqrt{n}}$ denotes the random set of coordinates used to generate $(\boldsymbol{x}, \boldsymbol{y})$ in the distribution $\sigma(X \times Y, F)$.

Proof. Within this proof we will use measure notation. For any subset $I \subseteq F$ of size $|I| = \sqrt{n}$, let ψ_I denote the measure on rectangles defined by our distributions $\sigma(X, Y, F)$ conditioned on choosing the set I of gadgets to fix to 1:

$$\forall A \times B \subseteq X \times Y \colon \qquad \psi_I(A \times B) \coloneqq \Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim \sigma(X \times Y, F)}[(\boldsymbol{x}, \boldsymbol{y}) \in A \times B \mid \boldsymbol{I} = I].$$

According the definition of $\sigma(X, Y, F)$, the lemma's conclusion is equivalent to:

$$\mathbb{E}_{\boldsymbol{I} \sim \binom{F}{\sqrt{n}}}[\psi_{\boldsymbol{I}}(X' \times Y')] \in \frac{|X' \times Y'|}{|X \times Y|}(1 \pm 0.1), \text{ and} \tag{5}$$

$$\mathbb{E}_{\boldsymbol{I} \sim \binom{F}{\sqrt{n}}}[\psi_{\boldsymbol{I}}(X' \times Y') \mid \boldsymbol{I} \subseteq F'] \in \frac{|X' \times Y'|}{|X \times Y|}(1 \pm 0.1). \tag{6}$$

To prove the equations, we first observe that ψ_I is uniform over its support

$$\text{supp}(\psi_I) = \{(x, y) \in X \times Y \mid \forall i \in I, \text{IP}_m(x_i, y_i) = 1\},$$

so

$$\psi_I(X' \times Y') = \frac{|\text{supp}(\psi_I) \cap X' \times Y'|}{|\text{supp}(\psi_I)|}.$$

Next, we calculate a formula for $|\text{supp}(\psi_I) \cap A \times B|$ for any rectangle $A \times B \subseteq X \times Y$. For convenience, we define $\delta = 2^{-m/20} = n^{-5}$ for $m = 100 \log n$. Observe that $(1 - 2^{-m})^{\sqrt{n}} \geq 1 - \sqrt{n} \cdot 2^{-m} \geq 1 - \delta$. Consider $(\boldsymbol{a}, \boldsymbol{b}) \sim \mathsf{unif}(A \times B)$. Let's assume that $\boldsymbol{a}_S, \boldsymbol{b}_S$ are both 0.9-spread, for some set $S \subseteq F$. Then, due to the pseudorandomness lemma, Lemma 12, the sequence of inner products $(\text{IP}_m(\boldsymbol{a}_i, \boldsymbol{b}_i))_{i \in S}$ is nearly uniform; specifically,

$$\forall z \in \{0,1\}^S: \qquad \Pr[(\text{IP}_m(\boldsymbol{a}_i, \boldsymbol{b}_i))_{i \in S} = z] \in \frac{1}{2^{|S|}}(1 \pm 2^{-m/20}) = \frac{1}{2^{|S|}}(1 \pm \delta).$$

Therefore,

$$\begin{aligned} |\text{supp}(\psi_I) \cap A \times B| &= |A \times B| \cdot \Pr[(\boldsymbol{a}, \boldsymbol{b}) \in \text{supp}(\psi_I)] \\ &= |A \times B| \cdot \Pr\big[(\text{IP}_m(\boldsymbol{a}_i, \boldsymbol{b}_i))_{i \in I} = 1^I\big] \\ &\leq |A \times B| \cdot \Pr\big[(\text{IP}_m(\boldsymbol{a}_i, \boldsymbol{b}_i))_{i \in I \cap S} = 1^{I \cap S}\big] \\ &\leq |A \times B| \cdot 2^{-|I \cap S|} \cdot (1 + \delta). \end{aligned} \tag{7}$$

In the case $I \subseteq S$, repeating the above calculation also yields

$$|\text{supp}(\psi_I) \cap A \times B| \in |A \times B| \cdot 2^{-|I|} \cdot (1 \pm \delta). \tag{8}$$

With these calculations, we may now complete the proof. For the rectangle $X \times Y$, $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$ have $\boldsymbol{x}_F, \boldsymbol{y}_F$ both 0.9-spread and that $I \subseteq F$ always, so by Equation (8), we have

$$|\text{supp}(\psi_I)| = |\text{supp}(\psi_I) \cap X \times Y| \in |X \times Y| \cdot 2^{-|I|} \cdot (1 \pm \delta),$$

using the fact that $\delta < 0.1$. For the rectangle $X' \times Y'$, the random variables $\boldsymbol{x}'_{F'}$, $\boldsymbol{y}'_{F'}$ are both 0.9-spread for $(\boldsymbol{x}', \boldsymbol{y}') \sim \mathsf{unif}(X' \times Y')$. When $I \subseteq F'$, using the same argument, we have

$$|\text{supp}(\psi_I) \cap X' \times Y'| \in |X' \times Y'| \cdot 2^{-|I|} \cdot (1 \pm \delta).$$

Therefore, the left-hand side of Equation (6) can be bounded as

$$\mathbb{E}_{\boldsymbol{I} \sim \binom{F}{\sqrt{n}}}[\psi_{\boldsymbol{I}}(X' \times Y') \mid \boldsymbol{I} \subseteq F'] = \mathbb{E}_{\boldsymbol{I} \sim \binom{F}{\sqrt{n}}}\left[\frac{|\text{supp}(\psi_{\boldsymbol{I}}) \cap X' \times Y'|}{|\text{supp}(\psi_{\boldsymbol{I}})|} \mid \boldsymbol{I} \subseteq F'\right] \in \frac{|X' \times Y'|}{|X \times Y|} \cdot (1 \pm 2\delta),$$

as desired. To prove Equation (5), we can similarly express the left-hand side as

$$\begin{aligned} \mathbb{E}_{\boldsymbol{I} \sim \binom{F}{\sqrt{n}}}[\psi_I(X' \times Y')] &= \mathbb{E}_{\boldsymbol{I} \sim \binom{F}{\sqrt{n}}}\left[\frac{|\text{supp}(\psi_{\boldsymbol{I}}) \cap X' \times Y'|}{|\text{supp}(\psi_{\boldsymbol{I}})|}\right] \\ &\in \mathbb{E}_{\boldsymbol{I} \sim \binom{F}{\sqrt{n}}}[|\text{supp}(\psi_{\boldsymbol{I}}) \cap X' \times Y'|] \cdot \frac{2^{\sqrt{n}}}{|X \times Y|} \cdot (1 \pm \delta). \end{aligned} \tag{9}$$

In this case, we do not have $\boldsymbol{I} \subseteq F'$ always, but we do have it with large enough probability, so we can write the following lower bound

$$\mathbb{E}_{\boldsymbol{I}\sim\binom{F}{\sqrt{n}}}[|\mathrm{supp}(\psi_{\boldsymbol{I}}) \cap X' \times Y'|] \geq \Pr\big[\boldsymbol{I} \subseteq F'\big] \cdot \frac{|X' \times Y'|}{2^{\sqrt{n}}} \cdot (1-\delta) \geq (1-1/100) \cdot \frac{|X' \times Y'|}{2^{\sqrt{n}}} \cdot (1-\delta),$$

where we use the fact that $|\boldsymbol{I}| \cdot |F \setminus F'| \leq n/1000 \leq |F|/100$, together with Claim 28, to obtain

$$\Pr\big[\boldsymbol{I} \subseteq F'\big] = \Pr\big[\boldsymbol{I} \cap (F \setminus F') = \emptyset\big] \geq 1 - \frac{1}{100}.$$

Using Equation (7) and Claim 28, we also get an upper bound

$$\mathbb{E}_{\boldsymbol{I}\sim\binom{F}{\sqrt{n}}}[|\mathrm{supp}(\psi_{\boldsymbol{I}}) \cap X' \times Y'|] \leq \frac{|X' \times Y'|}{2^{\sqrt{n}}} \cdot (1+\delta) \cdot \mathbb{E}[2^{|\boldsymbol{I} \setminus F'|}] \leq \frac{|X' \times Y'|}{2^{\sqrt{n}}} \cdot (1+\delta) \cdot (1+\tfrac{1}{20}).$$

By combining these bounds with Equation (9), we establish Equation (5). □

5.2 Sprinkling and Fixing the 1s

Throughout this section, we are in the context of the Safe Stage Lemma, meaning that we have rectangle $X \times Y$ and free variables $F \subseteq [n]$ satisfying the conditions of the Stage Lemma; Π is the protocol assumed in Assumption 9, Π' is the transformed protocol from the Protocol Transformation Lemma, and Equation (3) holds.

5.2.1 Finding a Leaf

Proposition 19 (Sprinkle 1s). *Let $\varepsilon > 0$ be a sufficiently small constant. Let $X \times Y \subseteq \Sigma^n \times \Sigma^n$ and $F \subseteq [n]$ satisfy the conditions of Lemma 13. If the current stage is safe, then there exists a leaf L of Π' such that:*

1. *L contains mostly 0 inputs under the sprinkled-1s distribution:*
$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\sigma(X\times Y,F)}[f(\boldsymbol{x},\boldsymbol{y}) = 0 \mid (\boldsymbol{x},\boldsymbol{y}) \in L] \geq 1 - 10\varepsilon,$$
2. *$|F| - |F_L| \leq \sqrt{n}/1000$.*
3. *$\mathcal{D}(L, F_L) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|)$.*

Proof. We say a leaf L of the protocol Π' is *good* if it satisfies the following three conditions:

1. F is a 0-leaf of Π' (i.e., the protocol outputs 0 if it reaches L);
2. $|F| - |F_L| \leq \sqrt{n}/1000$;
3. $\mathcal{D}(L, F_L) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F \setminus F_L| m)$, from Lemma 15.

Let $L_1, \ldots, L_N$ be all the good leaves of Π', with associated variables F_{L_i}. First, we lower bound the probability to be in a good leaf, and then prove that a typical good leaf does not contain too many errors, which yields the proposition.

Let $\boldsymbol{L}$ be a random leaf of Π', chosen as the unique leaf containing $(\boldsymbol{x},\boldsymbol{y}) \sim \mathsf{unif}(X \times Y)$. The current stage is safe, meaning that Equation (3) holds, so $\Pr[|F| - |F_{\boldsymbol{L}}| \leq \sqrt{n}/1000 \mid \boldsymbol{L} \text{ prints } 0] \geq 1/2$. By Lemma 15,

$$\Pr[\mathcal{D}(\boldsymbol{L}, F_{\boldsymbol{L}}) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F \setminus F_{\boldsymbol{L}}| m)] \geq 1 - 4\varepsilon.$$

Since Π' errs with probability at most ε over the mixture distribution $\frac{1}{2}(\mathsf{unif}(X \times Y) + \sigma(X \times Y, F))$, it errs with probability at most 2ε over $\mathsf{unif}(X \times Y)$. By assumption, $\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathsf{unif}(X\times Y)}[f(\boldsymbol{x},\boldsymbol{y}) = 0] \geq 1/2$, so by the union bound,

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathsf{unif}(X\times Y)}[L(\boldsymbol{x},\boldsymbol{y}) \text{ is good}] \geq \Pr\big[|F| - |F_{L(\boldsymbol{x},\boldsymbol{y})}| \leq \sqrt{n}/1000 \text{ and } \Pi'(\boldsymbol{x},\boldsymbol{y}) = 0\big] - 4\varepsilon \geq 1/4 - 5\varepsilon.$$

We rewrite the inequality as

$$\sum_{i\in[N]} \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim X\times Y}[(\boldsymbol{x},\boldsymbol{y}) \in L_i] \geq 1/4 - 5\varepsilon. \tag{10}$$

For brevity, we write $\sigma \coloneqq \sigma(X \times Y, F)$. By the Closeness Lemma, for every $i \in [N]$,

$$\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u},\boldsymbol{v}) \in L_i] \in (1 \pm 1/10) \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim X\times Y}[(\boldsymbol{x},\boldsymbol{y}) \in L_i], \tag{11}$$

$$\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u},\boldsymbol{v}) \in L_i \mid \boldsymbol{I} \subseteq F_{L_i}] \in (1 \pm 1/10) \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim X\times Y}[(\boldsymbol{x},\boldsymbol{y}) \in L_i]. \tag{12}$$

Let $p_L \coloneqq \Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[f(\boldsymbol{u},\boldsymbol{v}) = 1 \mid (\boldsymbol{u},\boldsymbol{v}) \in L]$. Since Π' has error ε over the mixture, it has error 2ε over σ, so

$$\begin{aligned}
2\varepsilon &\geq \Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[f(\boldsymbol{u},\boldsymbol{v}) = 1 \wedge \Pi'(\boldsymbol{u},\boldsymbol{v}) = 0] \\
&\geq \sum_{L \text{ 0-leaf of } \Pi'} p_L \Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u},\boldsymbol{v}) \in L] \\
&\geq \sum_{i\in[N]} p_{L_i} \Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u},\boldsymbol{v}) \in L_i] \\
&\geq 0.9 \sum_{i\in[N]} p_{L_i} \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathsf{unif}(X\times Y)}[(\boldsymbol{x},\boldsymbol{y}) \in L_i] && \text{(by Equation (11).)}
\end{aligned}$$

Thus, for a random $L_{\boldsymbol{i}}$ chosen as the unique leaf containing $(\boldsymbol{x}',\boldsymbol{y}')$ drawn from the uniform distribution on $\bigcup_i L_i$,

$$\begin{aligned}
\mathbb{E}[p_{L_{\boldsymbol{i}}}] &= \sum_{i\in[N]} p_{L_i} \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mathsf{unif}(X\times Y)}\Big[(\boldsymbol{x},\boldsymbol{y}) \in L_i \mid (\boldsymbol{x},\boldsymbol{y}) \in \bigcup_i L_i\Big] \\
&\leq 2\varepsilon \cdot \frac{10}{9} \cdot \Pr\left[(\boldsymbol{x},\boldsymbol{y}) \in \bigcup_i L_i\right]^{-1} \\
&\leq 2\varepsilon \cdot \frac{10}{9} \cdot \frac{1}{1/4 - 5\varepsilon} && \text{(by Equation (10))} \\
&\leq 10\varepsilon,
\end{aligned}$$

for sufficiently small $\varepsilon > 0$. Therefore, there exists a good leaf L_i with $p_{L_i} \leq 10\varepsilon$. For this leaf,

$$\mathcal{D}(L_i, F_{L_i}) \leq \mathcal{D}(X \times Y, F) + O(|\Pi|) - \Omega(|F| - |F_{L_i}|)m \leq \mathcal{D}(X \times Y, F) + O(|\Pi|),$$

as desired.

□

5.2.2 Fixing the Gadgets

Proposition 20 (Fix 1s)**.** *Under the conditions of Proposition 19, let L be a leaf of Π' obtained from that proposition, and let F_L be its associated set of free variables (as defined in Lemma 15). Then there exists $X' \times Y' \subseteq L$, $R \subseteq F_L$ with $|R| = \sqrt{n}$, and $F' \coloneqq F_L \setminus R$ such that*

1. $\Pr[f(\boldsymbol{x}, \boldsymbol{y}) = 0] \geq 0.6$ *for* $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(X' \times Y')$*;*
2. *For all $j \in R$ and all $x, y \in X' \times Y'$, $\mathrm{IP}_m(x_j, y_j) = 1$, and*
3. $\mathcal{D}(X' \times Y', F') \leq \mathcal{D}(L, F_L) + 2|R|$*.*

Proof. Let $\sigma \coloneqq \sigma(X \times Y, F)$. By Proposition 19 there exists a leaf $L = X^{(L)} \times Y^{(L)}$ of Π' satisfying $|F| - |F_L| \leq \sqrt{n}/1000$ and

$$\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[f(\boldsymbol{u}, \boldsymbol{v}) = 1 \mid (\boldsymbol{u}, \boldsymbol{v}) \in L] \leq 1/5.$$

Recall from Definition 14 that $(\boldsymbol{u}, \boldsymbol{v}) \sim \sigma$ is obtained by choosing $\boldsymbol{I} \sim \binom{F}{\sqrt{n}}$ and taking $(\boldsymbol{u}, \boldsymbol{v})$ uniform over the set $\{(u, v) \in X \times Y \mid \mathrm{IP}_m(u_i, v_i) = 1,\ \forall i \in \boldsymbol{I}\}$. Observe that

$$\begin{aligned}
&\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[f(\boldsymbol{u}, \boldsymbol{v}) = 1 \mid (\boldsymbol{u}, \boldsymbol{v}) \in L \wedge \boldsymbol{I} \subseteq F_L] \cdot \Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[\boldsymbol{I} \subseteq F_L \mid (\boldsymbol{u}, \boldsymbol{v}) \in L] \\
=&\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[f(\boldsymbol{u}, \boldsymbol{v}) = 1 \wedge \boldsymbol{I} \subseteq F_L \mid (\boldsymbol{u}, \boldsymbol{v}) \in L] \\
\leq&\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[f(\boldsymbol{u}, \boldsymbol{v}) = 1 \mid (\boldsymbol{u}, \boldsymbol{v}) \in L] \\
\leq&1/5.
\end{aligned}$$

Moreover, by Bayes' rule, we have

$$\begin{aligned}
\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[\boldsymbol{I} \subseteq F_L \mid (\boldsymbol{u}, \boldsymbol{v}) \in L] &= \Pr[\boldsymbol{I} \subseteq F_L] \cdot \frac{\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u}, \boldsymbol{v}) \in L \mid \boldsymbol{I} \subseteq F_L]}{\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u}, \boldsymbol{v}) \in L]} \\
&\geq 0.99 \cdot \frac{\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u}, \boldsymbol{v}) \in L \mid \boldsymbol{I} \subseteq F_L]}{\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[(\boldsymbol{u}, \boldsymbol{v}) \in L]} && \text{(by Claim 28)} \\
&\geq 0.99 \cdot \frac{0.9 \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim X\times Y}[(\boldsymbol{x}, \boldsymbol{y}) \in L]}{1.1 \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim X\times Y}[(\boldsymbol{x}, \boldsymbol{y}) \in L]} && \text{(by Equations (11) and (12))} \\
&> 0.8.
\end{aligned}$$

If follows that

$$\Pr_{(\boldsymbol{u},\boldsymbol{v})\sim\sigma}[f(\boldsymbol{u}, \boldsymbol{v}) = 1 \mid (\boldsymbol{u}, \boldsymbol{v}) \in L \wedge \boldsymbol{I} \subseteq F_L] \leq (1/5)/0.8 = 1/4.$$

So there exists a choice $\boldsymbol{I} = R \in \binom{F_L}{\sqrt{n}}$ such that

$$\Pr[f(\boldsymbol{u}, \boldsymbol{v}) = 1 \mid (\boldsymbol{u}, \boldsymbol{v}) \in L \wedge \boldsymbol{I} = R] \leq 1/4.$$

Let $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(L)$ and let $(\boldsymbol{u}', \boldsymbol{v}')$ be distributed identically to $(\boldsymbol{u}, \boldsymbol{v} \mid (\boldsymbol{u}, \boldsymbol{v}) \in L \wedge \boldsymbol{I} = R)$, which is equivalent to $(\boldsymbol{u}', \boldsymbol{v}')$ distributed as $(\boldsymbol{x}, \boldsymbol{y} \mid \mathrm{IP}_m(x_i, y_i) = 1,\ \forall i \in R)$. In calculations below, we write $F \coloneqq F_L$ for brevity. We first want to show

$$\mathrm{H}_\infty(\boldsymbol{u}'_F, \boldsymbol{v}'_F) \geq \mathrm{H}_\infty(\boldsymbol{x}_F, \boldsymbol{y}_F) - |R| - 1. \tag{13}$$

From Lemma 12,

$$\Pr[\forall i \in R:\ \mathrm{IP}_m(\boldsymbol{x}_i, \boldsymbol{y}_i) = 1] \in 2^{-|R|}(1 \pm 2^{-m/20}),$$

so

$$\max_{(u,v)\in\Sigma^F\times\Sigma^F} \Pr\big[(\boldsymbol{u}'_F, \boldsymbol{v}'_F) = (u,v)\big] \le \max_{(x,y)\in\Sigma^F\times\Sigma^F} \Pr[(\boldsymbol{x}_F, \boldsymbol{y}_F) = (x,y)] \cdot 2^{|R|}(1-2^{-m/20})^{-1},$$

which establishes Equation (13). The set $\{x, y \mid \mathrm{IP}_m(x_i, y_i) = 1, \forall i \in R\}$ is not a rectangle; to obtain a rectangle we will find variable settings $x^0_R, y^0_R \in \Sigma^R$ so that the rectangle $X' \times Y' :=$ $\{(x,y) \in L \mid (x_R, y_R) = (x^0_R, y^0_R)\}$ satisfies the desired condition. Let $(\boldsymbol{x}^0_R, \boldsymbol{y}^0_R)$ be distributed identically to $(\boldsymbol{u}'_R, \boldsymbol{v}'_R)$. Then by the chain rule for min-entropy (Lemma 27), for all $\delta \in (0,1)$,

$$\Pr_{(\boldsymbol{x}^0_R, \boldsymbol{y}^0_R)}\Big[\mathrm{H}_\infty\left(\boldsymbol{u}'_{F\setminus R}, \boldsymbol{v}'_{F\setminus R} \mid (\boldsymbol{u}'_R, \boldsymbol{v}'_R) = (\boldsymbol{x}^0_R, \boldsymbol{y}^0_R)\right) \ge \mathrm{H}_\infty(\boldsymbol{u}'_F, \boldsymbol{v}'_F) - 2|R|m - \log(1/\delta)\Big] \ge 1-\delta.$$

The last step is to ensure that the rectangle will be mostly 0-valued. Observe that

$$\mathbb{E}_{\boldsymbol{x}^0_R, \boldsymbol{y}^0_R}\left[\Pr_{\boldsymbol{u}', \boldsymbol{v}'}[f(\boldsymbol{u}', \boldsymbol{v}') = 1 \mid (\boldsymbol{u}'_R, \boldsymbol{v}'_R) = (\boldsymbol{x}^0_R, \boldsymbol{y}^0_R)]\right] = \Pr\big[f(\boldsymbol{u}', \boldsymbol{v}') = 1\big] \le 1/4,$$

so by Markov's inequality we have

$$\Pr_{\boldsymbol{x}^0_R, \boldsymbol{y}^0_R}\big[\Pr\big[f(\boldsymbol{u}', \boldsymbol{v}') = 1 \mid (\boldsymbol{u}'_R, \boldsymbol{v}'_R) = (\boldsymbol{x}^0_R, \boldsymbol{y}^0_R)\big] > 0.4\big] \le \frac{1}{4}\cdot\frac{10}{4} = 5/8.$$

By the union bound (using $\delta = 1/4$), there exist x^0_R, y^0_R such that $\Pr\big[f(\boldsymbol{u}', \boldsymbol{v}') = 0 \mid (\boldsymbol{u}'_R, \boldsymbol{v}'_R) = (x^0_R, y^0_R)\big] \ge$ $6/10$ and

$$\mathrm{H}_\infty\left(\boldsymbol{x}'_{F\setminus R}, \boldsymbol{v}'_{F\setminus R} \mid (\boldsymbol{u}'_R, \boldsymbol{v}'_R) = (x^0_R, y^0_R)\right) \ge \mathrm{H}_\infty(\boldsymbol{u}'_F, \boldsymbol{v}'_F) - 2|R|m - 2 \ge \mathrm{H}_\infty(\boldsymbol{x}_F, \boldsymbol{y}_F) - 2|R|m - |R| - 3.$$

Therefore the rectangle $X' \times Y'$ for $X' := \{x \in X^{(L)} \mid x_R = x^0_R\}$ and $Y' := \{y \in Y^{(L)} \mid y_R = y^0_R\}$ satisfies

$$\begin{aligned}
\mathcal{D}(X' \times Y', F \setminus R) &= 2(|F| - |R|)m - \mathrm{H}_\infty(\boldsymbol{u}'_{F\setminus R} \mid \boldsymbol{u}'_R = x^0_R) - \mathrm{H}_\infty(\boldsymbol{v}'_{F\setminus R} \mid \boldsymbol{v}'_R = y^0_R)\\
&\le 2(|F| - |R|)m - \mathrm{H}_\infty(\boldsymbol{x}_F, \boldsymbol{y}_F) + 2|R|m + |R| + 3\\
&\le 2|F|m - \mathrm{H}_\infty(\boldsymbol{x}_F, \boldsymbol{y}_F) + 2|R|\\
&= \mathcal{D}(L, F) + 2|R|. \square
\end{aligned}$$

5.2.3 Restoring Spreadness

Proposition 21 (Spreadify). *Let $A \times B \subseteq \Sigma^n \times \Sigma^n$ and $H \subseteq [n]$ such that $\Pr_{\boldsymbol{a},\boldsymbol{b}\sim A\times B}[f(\boldsymbol{a}, \boldsymbol{b}) =$ $0] \ge 0.6$. Then there exists $H' \subseteq H$ and a rectangle $A' \times B' \subseteq A \times B$ such that*

1. $\Pr_{\boldsymbol{a},\boldsymbol{b}\sim A'\times B'}[f(\boldsymbol{a}, \boldsymbol{b}) = 0] \ge 1/2$.
2. $A'_{H'}$ *and* $B'_{H'}$ *are* 0.9*-spread.*
3. $\mathcal{D}(A' \times B', H') \le \mathcal{D}(A \times B, H) - \Omega(|H \setminus H'|m)$.

Proof. For parameter $\varepsilon > 0$ to be fixed later, we apply the density restoring partition for the rectangle $A \times B$ (Lemma 23) to obtain partitions $A = \bigsqcup_i A_i$ and $B = \bigsqcup_j B_j$ with each rectangle $A_i \times B_j$ associated with a set $H_{i,j} \subseteq H$ of free variables. For each i, j, the random variables $\boldsymbol{x}_{H_{i,j}}$ and $\boldsymbol{y}_{H_{i,j}}$ are 0.9-spread, where $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(A_i \times B_j)$. With probability at least $1 - 2\varepsilon$ over $(\boldsymbol{x}, \boldsymbol{y}) \sim \mathsf{unif}(A \times B)$, the unique rectangle $A_{\boldsymbol{i}} \times B_{\boldsymbol{j}}$ containing $(\boldsymbol{x}, \boldsymbol{y})$ satisfies

$$\mathcal{D}(A_{\boldsymbol{i}} \times B_{\boldsymbol{j}}, H_{\boldsymbol{i},\boldsymbol{j}}) \leq \mathcal{D}(A \times B, H) - 0.1|H \setminus H_{\boldsymbol{i},\boldsymbol{j}}|m + 2\log(1/\varepsilon). \tag{14}$$

For each i, j, define

$$p_{i,j} \coloneqq \Pr_{(\boldsymbol{a}', \boldsymbol{b}') \sim \mathsf{unif}(A_i \times B_j)}[f(\boldsymbol{a}', \boldsymbol{b}') = 1].$$

Let $(\boldsymbol{i}, \boldsymbol{j})$ the random variable chosen as the unique values such that $A_{\boldsymbol{i}} \times B_{\boldsymbol{j}}$ contains $(\boldsymbol{a}, \boldsymbol{b}) \sim \mathsf{unif}(A \times B)$. By assumption,

$$\mathbb{E}[p_{\boldsymbol{i},\boldsymbol{j}}] = \sum_{i,j} \Pr[(\boldsymbol{x}, \boldsymbol{y}) \in A_i \times B_j] \cdot p_{i,j} = \Pr_{(\boldsymbol{a}, \boldsymbol{b}) \sim \mathsf{unif}(A \times B)}[f(\boldsymbol{a}, \boldsymbol{b}) = 1] \leq 4/10,$$

so by Markov's inequality, $\Pr[p_{\boldsymbol{i},\boldsymbol{j}} > 1/2] \leq 8/10$. Setting $\varepsilon = 1/20$, the probability that $p_{\boldsymbol{i},\boldsymbol{j}} > 1/2$ or that Equation (14) fails for i, j is at most $8/10 + 1/10 = 9/10$, so there exists rectangle $A' \times B' = A_i \times B_j$ with free variables $H' \coloneqq H_{i,j}$ satisfying the desired conditions. □

A Appendix: Counting Arguments and Variations of FBPP

A.1 Variations of FBPP

As mentioned in the introduction, there is a variety of natural ways to define a class FBPP of "BPP-search problems", i.e., search problems with efficient randomized protocols. Here are four definitions of increasing restrictiveness:

(i) Direct translation of the definition of BPP for decision problems. A sequence of relations $R_n \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ is in FBPP if and only if $\forall n$ there exists a randomized protocol Π with cost $\operatorname{poly}\log n$ such that $\forall x, y \in \{0,1\}^n$, $\Pr[(x, y, \Pi(x, y)) \in R_n] \geq 2/3$.

  The issue with this definition is that, unlike for BPP, the constant 2/3 is no longer arbitrary: choosing a different constant will change the class, because for search problems, the error cannot generally be boosted (see examples below).

(ii) Fix the issue by demanding that there is an efficient protocol achieving any error $\varepsilon > 0$. The sequence $R_n \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ is in FBPP if and only if $\forall n, \forall \varepsilon > 0$ there exists a randomized protocol with cost $\operatorname{poly}(\log n, 1/\varepsilon)$ such that $\forall x, y, \Pr[(x, y, \Pi(x, y)) \in R_n] \geq 1 - \varepsilon$. This is the approach taken in [Aar10, ABK24] for Turing machines.

(iii) Demand that the cost of the protocol is $\operatorname{poly}(\log n, \log(1/\varepsilon))$ instead of $\operatorname{poly}(\log n, 1/\varepsilon)$, so that dependence on ε is the same as it is for decision problems.

(iv) Require that solutions are not only efficient to find, but also efficient to verify. This is the approach taken by [Gol11], and this is also analogous to the definition of efficient deterministic search problems[1] in the context of TFNP. According to this definition, R_n is in FBPP if and only if both the following conditions hold:

[1] We remark that, confusingly, the name FP is used for two different classes of problems: relations where a solution can be found efficiently, and relations where a solution can be found *and* verified efficiently.

- The condition from definition (i) above, where the relation can be computed with cost $\operatorname{poly}\log n$ and success probability is $2/3$; and,
- For every n there exists a randomized *verifier* protocol V which outputs a value in $\{0,1\}$ such that, for all $(x, y, z) \in \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$, when Alice is given (x, z) and Bob is given (y, z), they can run protocol V which satisfies $\Pr[V(x, y, z) = 1_{(x,y,z)\in R_n}] \geq 2/3$.

Note that each of these definitions is more strict than the previous one: If a relation satisfies definition (ii) then it also satisfies definition (i), etc.

Pseudodeterminism. It is most natural to compare pseudodeterminism with definition (iii). This is because, if a problem admits an efficient $\operatorname{poly}\log n$ pseudodeterministic protocol, it also satisfies definition (iii). With a pseudodeterministic protocol, we can take a majority vote on the output of $O(\log(1/\varepsilon))$ independent runs; with probability $1 - \varepsilon$, the canonical output for given input (x, y) will be the majority of outputs.

Partial boolean functions. Each of the above definitions is *more strict* than the previous one, but this is only true for problems with large outputs. When we consider partial boolean functions, i.e., problems where the valid outputs are single bits (the main topic of this paper), these 4 definitions collapse into only 2, because the first 3 definitions are equivalent. Given a protocol with the guarantees in definition (i), we can perform majority-vote error boosting. If there is only 1 valid output for given (x, y) then the majority vote will take this value; if there are 2 valid outputs for given (x, y) then the output is valid with probability 1.

If we then take definition (iii) as the definition of FBPP for partial functions, our Theorem 1 proves $\mathsf{FBPP} \subsetneq \mathsf{FPs}$ for partial functions.

Furthermore, when the number of valid outputs is small, definition (iv) is a special case of efficient pseudodeterminism. This is because, for relations $R_n \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ with only a small $\leq \operatorname{poly}\log(n)$ number of possible outputs, the verifier can be used to create a pseudodeterministic protocol, by iterating over every possible output in a fixed order:

Claim 29. *Let $R_n \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ be a relation that satisfies definition (iv) above, and also $|\{z \in \{0,1\}^* \mid \exists x, y : (x, y, z) \in R_n\}| \leq \operatorname{poly}\log n$. Then there is a pseudodeterministic protocol for R_n with cost $\operatorname{poly}\log n$.*

A.2 Separations via counting

Let us explain how to prove Equation (1) for definitions (i)–(iii) of FBPP given above, using a counting argument that works equally well for boolean circuits and other models of computation. An initial sketch of the argument for definition (i) is as follows:

> Define a relation $S \subseteq \{0,1\}^n \times \{0,1\}^n \times [3]$ by choosing $S(x, y) := \{i : (x, y, i) \in S\}$ as a random subset of [3] of size 2 (i.e., we forbid one random output element). On any input, outputting a random number in [3] will solve S with probability $2/3$. A simple counting argument shows that the pseudodeterministic complexity is $\Omega(n)$ with high probability over the choice of S.

For definition (i), we can make this argument explicit (see Appendix A.3), i.e., we exhibit a specific relation satisfying definition (i) but which does not have an efficient pseudodeterministic protocol. But let us now give the full counting argument for definition (iii).

Proposition 30. *There exists a sequence $R_n \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^*$ of relations such that for every $\varepsilon > 0$, $\mathsf{R}_\varepsilon(R_n) = O(\log(1/\varepsilon))$, but which has pseudodeterministic cost $\Omega(n)$.*

Proof. Fix any n. Let $\mathcal{R}$ be the set of all relations R such that $|R(x,y) \cap \{0,1\}^k| = 2^k - 1$ for all $(x,y) \in \{0,1\}^n \times \{0,1\}^n$ and $k \in [n]$. That is, in any relation $R \in \mathcal{R}$ we have removed exactly one of the 2^k valid outputs for every output length $k \in [n]$. Each relation $R \in \mathcal{R}$ has randomized cost $O(\log(1/\varepsilon))$ since for every ε we can take $k = \lceil \log(1/\varepsilon) \rceil$ and output a random value $\boldsymbol{z} \sim \{0,1\}^k$; then $(x, y, \boldsymbol{z}) \in R$ with probability at least $1 - 2^{-k} \geq 1 - \varepsilon$.

Fix any function $f \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^*$ such that the length of any output is at most $|f(x,y)| \leq c$ for some c. Let $\mathcal{R}_f \subseteq \mathcal{R}$ be the set of relations consistent with f (i.e., $f(x,y) \in R$ for all x, y). Then

$$\frac{|\mathcal{R}_f|}{|\mathcal{R}|} = \prod_{x,y} \left(1 - 2^{-|f(x,y)|}\right) \leq \left(1 - 2^{-c}\right)^{2^{2n}} \leq e^{-2^{2n-c}}.$$

Every pseudodeterministic protocol with cost c computes one of these functions f. Using Newman's theorem to bound the number of random bits in any randomized protocol by $\log n + O(1)$, the number of pseudodeterministic protocols with cost c is at most

$$\left(\left(2^{2^n}\right)^{2^c} \cdot \left(2^{c+1}\right)^{2^c} \right)^{O(n)}.$$

Then the fraction of relations consistent with any pseudodeterministic protocol is at most

$$2^{O(n(2^{n+c} + c2^c)) - \Omega(2^{2n-c})} < 1$$

when $c = o(n)$. ∎

A.3 Explicit Separation for Definition (i)

In this section, we provide an explicit relation satisfying definition (i) but does not admit an efficient pseudodeterministic protocol. In fact, we separate randomized communication from the stronger quantum pseudoterministic model.

Theorem 31. *There exists an explicit two-party search problem $f \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^2$ with randomized complexity $O(1)$ but quantum pseudodeterministic complexity $\Omega(n)$.*

Before presenting the proof, we remark that in concurrent work, Aaronson, Gur, and Li [AGL26] showed an analogous separation for query complexity. Indeed, the two-party search problem we exhibit here is a composition of their search problem with the two-bit AND gadget.

Let R_ε denote the ε-error classical communication complexity, Q_ε the ε-error quantum communication complexity, and QPs_ε denote the ε-error quantum pseudodeterministic communication complexity. We will omit ε for the above-mentioned measures when $\varepsilon = 1/3$.

We first claim that it suffices to find an f and two different constant error parameters $0 < \varepsilon_2 < \varepsilon_1 < \frac{1}{3}$, such that $\mathsf{R}_{\varepsilon_1}(f) = O(1)$ while $\mathsf{Q}_{\varepsilon_2}(f) = \Omega(n)$. Indeed, if we can find such an f, we have $\mathsf{R}(f) \leq \mathsf{R}_{\varepsilon_1}(f) = O(1)$, and on the other hand,

$$\mathsf{QPs}(f) \geq \Omega(\mathsf{QPs}_{\varepsilon_2}(f)/\log(1/\varepsilon_2)) \geq \Omega(\mathsf{Q}_{\varepsilon_2}(f)/\log(1/\varepsilon_2)) = \Omega(n),$$

where we used the fact that quantum pseudodeterministic complexity admits error reduction.

We now construct f as follows: Let $n = 2m$ without loss of generality. Given $x, y \in \{0,1\}^n$ as the input for Alice and Bob respectively, we interpret $x = (x^1, x^2)$ as the concatenation of two m-bit strings $x^1, x^2 \in \{0,1\}^m$, and similarly for $y = (y^1, y^2)$. Then we define $f(x,y) \coloneqq \{0,1\}^2 \setminus \{(\mathrm{IP}(x^1,y^1), \mathrm{IP}(x^2,y^2))\}$ as the set of all pairs of bits excluding $(\mathrm{IP}(x^1,y^1), \mathrm{IP}(x^2,y^2))$. We will chose $\varepsilon_1 = 1/4, \varepsilon_2 = 0.1$ and prove that $\mathsf{R}_{1/4}(f) = O(1)$, while $\mathsf{Q}_{0.1}(f) = \Omega(n)$.

We first show $\mathsf{R}_{1/4}(f) = O(1)$: Observe that $|f(x,y)| = 3$ for all $x, y \in \{0,1\}^n$, the simple protocol that outputs a uniform random pair $(\boldsymbol{u}, \boldsymbol{v}) \sim \{0,1\}^2$ succeeds with probability 3/4.

It remains to show $\mathsf{Q}_{0.1}(f) = \Omega(n)$. For the sake of contradiction, suppose that $\mathsf{Q}_{0.1}(f) = o(n)$. Then there exists a quantum protocol Π of cost $d = o(n)$ that computes f with error at most 0.1 with respect to $\mathsf{unif}(\{0,1\}^n \times \{0,1\}^n)$. We will then construct a quantum protocol Π' of cost $d' = O(d) = o(n)$ that computes n-bit IP with error at most 0.4 with respect to $\mathsf{unif}(\{0,1\}^m \times \{0,1\}^m)$, a contradiction to the following folklore lower bound for the quantum distributional communication complexity of IP.

Lemma 32 ([Kre95]). *The quantum distributional communication complexity of the n-bit* IP *with respect to $\mu \coloneqq \mathsf{unif}(\{0,1\}^{n/2} \times \{0,1\}^{n/2})$ is $\mathsf{Q}_{0.4}(\mathrm{IP}_{n/2}, \mu) = \Omega(n)$.*

Before presenting the protocol Π', we define the following notation:

$$\begin{aligned} p_{00} &\coloneqq \Pr\big[\boldsymbol{u} \neq \mathrm{IP}(\boldsymbol{x}^1, \boldsymbol{y}^1) \wedge \boldsymbol{v} \neq \mathrm{IP}(\boldsymbol{x}^2, \boldsymbol{y}^2)\big], \\ p_{01} &\coloneqq \Pr\big[\boldsymbol{u} \neq \mathrm{IP}(\boldsymbol{x}^1, \boldsymbol{y}^1) \wedge \boldsymbol{v} = \mathrm{IP}(\boldsymbol{x}^2, \boldsymbol{y}^2)\big], \\ p_{10} &\coloneqq \Pr\big[\boldsymbol{u} = \mathrm{IP}(\boldsymbol{x}^1, \boldsymbol{y}^1) \wedge \boldsymbol{v} \neq \mathrm{IP}(\boldsymbol{x}^2, \boldsymbol{y}^2)\big], \end{aligned}$$

where $\boldsymbol{x} = (\boldsymbol{x}^1, \boldsymbol{x}^2), \boldsymbol{y} = (\boldsymbol{y}^1, \boldsymbol{y}^2)$ are uniform n-bit strings and $(\boldsymbol{u}, \boldsymbol{v}) \coloneqq \Pi(\boldsymbol{x}, \boldsymbol{y})$. It follows that $p_{00} + p_{01} + p_{10} = \Pr[(\boldsymbol{u}, \boldsymbol{v}) \in f(\boldsymbol{x}, \boldsymbol{y})] \geq 0.9$.

Next, we specify the protocol Π' as follows: Alice holds $\boldsymbol{x'} \sim \{0,1\}^m$ and Bob holds $\boldsymbol{y'} \sim \{0,1\}^m$. To compute $\mathrm{IP}(\boldsymbol{x'}, \boldsymbol{y'})$, they first sample two uniform random m-bit strings $\boldsymbol{x''}, \boldsymbol{y''} \sim \{0,1\}^m$. Consider the following cases:

- $p_{00} \geq 0.3$: In this case, we have either $p_{00} + p_{01} \geq 0.6$ or $p_{00} + p_{10} \geq 0.6$. Suppose the former holds without loss of generality, as the other case can be dealt with a similar argument. Both parties simulate Π on $\boldsymbol{x} \coloneqq (\boldsymbol{x'}, \boldsymbol{x''})$ and $\boldsymbol{y} \coloneqq (\boldsymbol{y'}, \boldsymbol{y''})$, obtain $(\boldsymbol{u}, \boldsymbol{v}) \coloneqq \Pi(\boldsymbol{x}, \boldsymbol{y})$. Finally, they output $\neg \boldsymbol{u}$. We conclude that Π' succeeds with probability

$$\Pr\big[\Pi'(\boldsymbol{x'}, \boldsymbol{y'}) = \mathrm{IP}(\boldsymbol{x'}, \boldsymbol{y'})\big] = \Pr\big[\boldsymbol{u} \neq \mathrm{IP}(x', y')\big] = p_{00} + p_{01} \geq 0.6.$$

- $p_{00} < 0.3$: In this case, we have $p_{01} + p_{10} \geq 0.6$. Similar to the previous case, both parties simulate Π on $\boldsymbol{x} \coloneqq (\boldsymbol{x'}, \boldsymbol{x''})$ and $\boldsymbol{y} \coloneqq (\boldsymbol{y'}, \boldsymbol{y''})$, obtain $(\boldsymbol{u}, \boldsymbol{v}) \coloneqq \Pi(\boldsymbol{x}, \boldsymbol{y})$. Since $(\boldsymbol{x''}, \boldsymbol{y''})$ is the common information between both parties, they can compute $\boldsymbol{v'} \coloneqq \mathrm{IP}(\boldsymbol{x''}, \boldsymbol{y''})$ without extra communication. Finally, they output $\neg(\boldsymbol{u} \oplus \boldsymbol{v} \oplus \boldsymbol{v'})$. We conclude that Π' succeeds with probability

$$\Pr\big[\Pi(\boldsymbol{x'}, \boldsymbol{y'}) = \mathrm{IP}(\boldsymbol{x'}, \boldsymbol{y'})\big] = \Pr\big[\boldsymbol{u} \oplus \boldsymbol{v} \oplus \mathrm{IP}(\boldsymbol{x'}, \boldsymbol{y'}) \oplus \mathrm{IP}(\boldsymbol{x''}, \boldsymbol{y''}) = 1\big] = p_{01} + p_{10} \geq 0.6.$$

Acknowledgements

We thank Tom Watson and an anonymous STOC reviewer for pointing out a mistake in an earlier version of this paper. This work was supported by the Swiss State Secretariat for Education, Research, and Innovation (SERI) under contract number MB22.00026.

References

[Aar10] Scott Aaronson. BQP and the polynomial hierarchy. In *Proceedings of the 42nd Symposium on Theory of Computing (STOC)*, pages 141–150, Cambridge, MA, USA, 2010. ACM. doi:10.1145/1806689.1806711.

[ABK24] Scott Aaronson, Harry Buhrman, and William Kretschmer. A qubit, a coin, and an advice string walk into a relational problem. In *Proceedings of the 15th Innovations in Theoretical Computer Science (ITCS)*, volume 287 of *LIPIcs*, pages 1:1–1:24, Berkeley, CA, USA, 2024. Schloss Dagstuhl. doi:10.4230/LIPIcs.ITCS.2024.1.

[AGL26] Hugo Aaronson, Tom Gur, and Jiawei Li. Pseudo-deterministic quantum algorithms. Technical report, arXiv, 2026. doi:10.48550/arXiv.2602.17647.

[BFI23] Sam Buss, Noah Fleming, and Russell Impagliazzo. TFNP characterizations of proof systems and monotone circuits. In *Proceedings of the 14th Innovations in Theoretical Computer Science (ITCS)*, volume 251 of *LIPIcs*, pages 30:1–30:40, Cambridge, MA, USA, 2023. Schloss Dagstuhl. doi:10.4230/LIPIcs.ITCS.2023.30.

[BHH+25] Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Borsuk-Ulam and replicable learning of large-margin halfspaces. Technical report, arXiv, 2025. doi:10.48550/arXiv.2503.15294.

[BHT25] Igor Balla, Lianna Hambardzumyan, and Istvá Tomon. Factorization norms and an inverse theorem for maxcut. Technical report, arXiv, 2025. doi:10.48550/arXiv.2506.23989.

[CDM23] Arkadev Chattopadhyay, Yogesh Dahiya, and Meena Mahajan. Query complexity of search problems. In *Proceedings of the 48th Mathematical Foundations of Computer Science (MFCS)*, pages 34–1, Bordeaux, France, 2023. Schloss Dagstuhl. doi:10.4230/LIPIcs.MFCS.2023.34.

[CFK+21] Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. Query-to-communication lifting using low-discrepancy gadgets. *SIAM Journal on Computing*, 50(1):171–210, 2021. doi:10.1137/19M1310153.

[CLO+23] Lijie Chen, Zhenjian Lu, Igor Oliveira, Hanlin Ren, and Rahul Santhanam. Polynomial-time pseudodeterministic construction of primes. In *Proceedings of the 64th Symposium on Foundations of Computer Science (FOCS)*, pages 1261–1270, Santa Cruz, CA, USA, 2023. IEEE. doi:10.1109/FOCS57990.2023.00074.

[CLV19] 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, New Brunswick, NJ, USA, 2019. Schloss Dagstuhl. doi:10.4230/LIPIcs.CCC.2019.14.

[CMY23] Zachary Chase, Shay Moran, and Amir Yehudayoff. Stability and replicability in learning. In *Proceedings of the 64th Symposium on Foundations of Computer Science (FOCS)*, pages 2430–2439, Santa Cruz, CA, USA, 2023. IEEE. doi:10.1109/FOCS57990.2023.00148.

[dRGR22] Susanna de Rezende, Mika Göös, and Robert Robere. Proofs, circuits, and communication. *SIGACT News*, 53(1):59–82, 2022. doi:10.1145/3532737.3532745.

[FF81] Peter Frankl and Zoltán Füredi. A short proof for a theorem of Harper about Hamming-spheres. *Discrete Mathematics*, 34(3):311–313, 1981. doi:10.1016/0012-365X(81)90009-1.

[FGHH25] Yuting Fang, Mika Göös, Nathaniel Harms, and Pooya Hatami. Constant-cost communication is not reducible to k-hamming distance. In *Proceedings of the 57th Symposium on Theory of Computing (STOC)*, pages 565–571, Prague, Czechia, 2025. ACM. doi:10.1145/3717823.3718129.

[Gav25] Dmytro Gavinsky. Unambiguous parity-query complexity. *Random Structures & Algorithms*, 66(3):e70010, 2025. doi:10.1002/rsa.70010.

[GG11] Eran Gat and Shafi Goldwasser. Probabilistic search algorithms with unique answers and their cryptographic applications. Technical report, Electronic Colloquium on Computational Complexity (ECCC), 2011. URL: https://eccc.weizmann.ac.il/report/2011/136/.

[GGJL25] Mika Göös, Tom Gur, Siddhartha Jain, and Jiawei Li. Quantum communication advantage in TFNP. In *Proceedings of the 57th Symposium on Theory of Computing (STOC)*, pages 1465–1475, Prague, Czechia, 2025. ACM. doi:10.1145/3717823.3718155.

[GGMW20] Shafi Goldwasser, Ofer Grossman, Sidhanth Mohanty, and David Woodruff. Pseudo-deterministic streaming. In *Proceedings of the 11th Innovations in Theoretical Computer Science (ITCS)*, volume 151 of *LIPIcs*, pages 79:1–79:25, Seattle, WA, USA, 2020. Schloss Dagstuhl. doi:10.4230/LIPIcs.ITCS.2020.79.

[GGR13] Oded Goldreich, Shafi Goldwasser, and Dana Ron. On the possibilities and limitations of pseudodeterministic algorithms. In *Proceedings of the 4th Innovations in Theoretical Computer Science (ITCS)*, pages 127–138, Berkeley, CA, USA, 2013. ACM. doi:10.1145/2422436.2422453.

[GIPS21] Shafi Goldwasser, Russell Impagliazzo, Toniann Pitassi, and Rahul Santhanam. On the pseudo-deterministic query complexity of NP search problems. In *Proceedings of the 36th Computational Complexity Conference (CCC)*, pages 36:1–36:22, Toronto, ON, Canada (Virtual Conference), 2021. Schloss Dagstuhl. doi:10.4230/LIPIcs.CCC.2021.36.

[GKPW18] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for P^{NP}. *Computational Complexity*, 28(1):113–144, 2018. doi:10.1007/s00037-018-0175-5.

[GKRS19] Mika Göös, Pritish Kamath, Robert Robere, and Dmitry Sokolov. Adventures in monotone complexity and TFNP. In *Proceedings of the 10th Innovations in Theoretical Computer Science (ITCS)*, volume 124 of *LIPIcs*, pages 38:1–38:19, San Diego, California, USA, 2019. Schloss Dagstuhl. doi:10.4230/LIPICS.ITCS.2019.38.

[GLM+16] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM Journal on Computing*, 45(5):1835–1869, 2016. doi:10.1137/15M103145X.

[Gol11] Oded Goldreich. In a world of P=BPP. In *Studies in Complexity and Cryptography*, pages 191–232. Springer, Berlin, Germany, 2011. doi:10.1007/978-3-642-22670-0_20.

[GPW20] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. *SIAM Journal on Computing*, 49(4):FOCS17–441–FOCS17–461, 2020. doi:10.1137/17M115339X.

[HHH23] Lianna Hambardzumyan, Hamed Hatami, and Pooya Hatami. Dimension-free bounds and structural results in communication complexity. *Israel Journal of Mathematics*, 253(2):555–616, 2023. doi:10.1007/s11856-022-2365-8.

[HN12] Trinh Huynh and Jakob Nordström. On the virtue of succinct proofs: Amplifying communication complexity hardness to time-space trade-offs in proof complexity. In *Proceedings of the 44th Symposium on Theory of Computing (STOC)*, pages 233–248, New York, NY, USA, 2012. ACM. doi:10.1145/2213977.2214000.

[ILPS22] Russell Impagliazzo, Rex Lei, Toniann Pitassi, and Jessica Sorrell. Reproducibility in learning. In *Proceedings of the 54th Symposium on Theory of Computing (STOC)*, pages 818–831, Rome, Italy, 2022. ACM. doi:10.1145/3519935.3519973.

[ILW23] Rahul Ilango, Jiatu Li, and Ryan Williams. Indistinguishability obfuscation, range avoidance, and bounded arithmetic. In *Proceedings of the 55th Symposium on Theory of Computing (STOC)*, pages 1076–1089, Orlando, FL, USA, 2023. ACM. doi:10.1145/3564246.3585187.

[Kre95] Ilan Kremer. Quantum communication. Master's thesis, Hebrew University of Jerusalem, Jerusalem, Israel, 1995.

[LOS21] Zhenjian Lu, Igor Oliveira, and Rahul Santhanam. Pseudodeterministic algorithms and the structure of probabilistic time. In *Proceedings of the 53rd Symposium on Theory of Computing (STOC)*, pages 303–316, Virtual Event, Italy, 2021. ACM. doi:10.1145/3406325.3451085.

[LS09] Nati Linial and Adi Shraibman. Learning complexity vs communication complexity. *Combinatorics, Probability and Computing*, 18(1-2):227–245, 2009. doi:10.1017/S0963548308009656.

[OS17] Igor Oliveira and Rahul Santhanam. Pseudodeterministic constructions in subexponential time. In *Proceedings of the 49th Symposium on Theory of Computing (STOC)*, pages 665–677, Montreal, QC, Canada, 2017. ACM. doi:10.1145/3055399.3055500.

[Vad12] Salil Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1–3):1–336, 2012. doi:10.1561/0400000010.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il