Electronic Colloquium on Computational Complexity, Report No. 71 (2024)

Lifting with Inner Functions of Polynomial Discrepancy

Yahel Manor* Or Meir†

April 10, 2024

Abstract

Lifting theorems are theorems that bound the communication complexity of a composed function $f \circ g^n$ in terms of the query complexity of f and the communication complexity of g. Such theorems constitute a powerful generalization of direct-sum theorems for g, and have seen numerous applications in recent years.

We prove a new lifting theorem that works for every two functions f, g such that the discrepancy of g is at most inverse polynomial in the input length of f. Our result is a significant generalization of the known direct-sum theorem for discrepancy, and extends the range of inner functions g for which lifting theorems hold.

1 Introduction

The direct-sum question is a fundamental question in complexity theory, which asks whether computing a function g on n independent inputs is n times harder than computing it on a single input. A related type of result, which is sometimes referred to as an "XOR lemma", says that computing the XOR of the outputs of g on n independent inputs is about n times harder than computing g on a single coordinate. Both questions received much attention in the communication complexity literature, see, e.g., [KRW91, FKNN95, KKN92, CSWY01, Sha01, JRS03, BPSW05, JRS05, LS09, BBCR10, Jai11, She11, BR11, Bra12].

A lifting theorem is a powerful generalization of both direct-sum theorems and XOR lemmas. Let $f\colon \{0,1\}^n \to \mathcal{O}$ and $g\colon \Lambda \times \Lambda \to \{0,1\}$ be functions (where Λ and $\mathcal{O}$ are some arbitrary sets). The block-composed function $f \circ g^n$ is the function that corresponds to the following task: Alice gets $x_1, \ldots, x_n \in \Lambda$, Bob gets $y_1, \ldots, y_n \in \Lambda$, and they wish to compute the output of f on the n-bit string whose i-th bit is $g(x_i, y_i)$. Lifting theorems say that the "natural way" for computing $f \circ g^n$ is more-or-less the best way. In particular, direct-sum theorems and XOR lemmas can be viewed as lifting theorems for the special cases where f is the identity function and the parity function respectively.

A bit more formally, observe that there is an obvious protocol for computing $f \circ g^n$: Alice and Bob jointly simulate a decision tree of optimal height for solving f. Any time the tree queries the i-th bit, they compute g on (x_i, y_i) by invoking the best possible communication protocol for g. A (query-to-communication) *lifting theorem* is a theorem that says that this protocol is roughly optimal. Specifically, let $D^{\mathrm{dt}}(f)$ and $D^{\mathrm{cc}}(g)$ denote the deterministic query complexity of f and

*Department of Computer Science, University of Haifa, Haifa 3498838, Israel. Supported by the Israel Science Foundation (grant No. 716/20).

†Department of Computer Science, University of Haifa, Haifa 3498838, Israel. ormeir@cs.haifa.ac.il. Partially supported by the Israel Science Foundation (grant No. 716/20).

ISSN 1433-8092

communication complexity of g respectively, and let $R^{\mathrm{dt}}(f)$ and $R^{\mathrm{cc}}(g)$ denote the corresponding randomized complexities. Then, a lifting theorem says that

$$D^{\mathrm{cc}}(f \circ g^n) = \Omega\left(D^{\mathrm{dt}}(f) \cdot D^{\mathrm{cc}}(g)\right) \qquad \text{(in the deterministic setting)} \tag{1}$$
$$R^{\mathrm{cc}}(f \circ g^n) = \Omega\left(R^{\mathrm{dt}}(f) \cdot R^{\mathrm{cc}}(g)\right) \qquad \text{(in the randomized setting).}$$

In other words, a lifting theorem says that the communication complexity of $f \circ g^n$ is close to the upper bound that is obtained by the natural protocol.

In recent years, lifting theorems found numerous applications, such as proving lower bounds on monotone circuit complexity and proof complexity (e.g. [RM97, GP14, RPRC16, PR17, GGKS18, PR18, dRMN$^+$20a, dRMN$^+$20b]), the separation of partition number and deterministic communication complexity [GPW15], proving lower bounds on data structures [CKLM18], and an application to the famous log-rank conjecture [HHL16], to name a few.

For most applications, it is sufficient to prove a lifting theorem that holds for every outer function f, but only for one particular choice of the inner function g. Moreover, it is desirable that the inner function g would be a simple as possible, and that its input length $b = \log|\Lambda|$ would be a small as possible in terms of in the input length n of the outer function f. For these reasons, the function g is often referred to as the "gadget".

On the other hand, if we view lifting theorems as a generalization of direct-sum theorems, then it is an important research goal to prove lifting theorems for as many inner functions g as possible, including "complicated" ones. This goal is not only interesting in its own right, but might also lead to additional applications. Indeed, this goal is a natural extension of the long line of research that attempts to prove direct-sum theorems for as many functions as possible. This is the perspective we take in this work, following Chattopadhyay et. al. [CKLM17, CFK$^+$19]. In particular, we intentionally avoid the term "gadget", since we now view the function g as the main object of study.

Previous work. The first lifting theorem, due to Raz and McKenzie [RM99], holds only when the inner function g is the index function. For a long time, this was the only inner function for which lifting theorems were known to hold for every outer function f. Then, the works of Chattopadhyay et. al. [CKLM17] and Wu et. al. [WYY17] proved a lifting theorem for the case where g is the inner product function. The work of [CKLM17] went further than that, and showed that their lifting theorem holds for any inner function g that satisfies a certain hitting property. This includes, for example, the gap-Hamming-distance problem.

All the above results are stated only for the deterministic setting. In the randomized setting, Göös, Pitassi, and Watson [GPW17] proved a lifting theorem with the inner function g being the index function. In addition, Göös et. al. [GLM$^+$15] proved a lifting theorem in the non-deterministic setting (as well as several related settings) with g being the inner product function.

More recently, Chattopadhyay et. al. [CFK$^+$19] proved a lifting theorem that holds for every inner function g that has logarithmic input length and exponentially small discrepancy. This theorem holds in both the deterministic and randomized setting, and includes the cases where g is the inner product function or a random function. Since our work builds on the lifting theorem of [CFK$^+$19], we discuss this result in more detail. The *discrepancy* of g, denoted $\mathrm{disc}(g)$, is a natural and widely-studied property of functions, and is equal to the maximum bias of g in any combinatorial rectangle. Formally, it is defined as follows:

Definition 1.1. Let $g : \Lambda \times \Lambda \to \{0, 1\}$ be a function, and let U, V be independent random variables that are uniformly distributed over Λ. Given a combinatorial rectangle $R \subseteq \Lambda \times \Lambda$, the *discrepancy* *of g with respect to R*, denoted $\mathrm{disc}_R(g)$, is defined as follows:

$$\mathrm{disc}_R(g) = \left|\Pr\left[g(U,V)=0 \text{ and } (U,V)\in R\right] - \Pr\left[g(U,V)=1 \text{ and } (U,V)\in R\right]\right|.$$

The *discrepancy of* g, denoted $\mathrm{disc}(g)$, is defined as the maximum of $\mathrm{disc}_R(g)$ over all combinatorial rectangles $R \subseteq \Lambda \times \Lambda$.

Informally, the main theorem of [CFK+19] says that if $b = \log|\Lambda|$, $\mathrm{disc}(g) = 2^{-\Omega(b)}$ and $b \geq c \cdot \log n$ for some constant c, then

$$D^{\mathrm{cc}}(f \circ g^n) = \Omega\left(D^{\mathrm{dt}}(f) \cdot b\right) \qquad \text{and} \qquad R^{\mathrm{cc}}_{1/3}(f \circ g^n) = \Omega\left(R^{\mathrm{dt}}_{1/3}(f) \cdot b\right).$$

We note that when $\mathrm{disc}(g) = 2^{-\Omega(b)}$, it holds that $D^{\mathrm{cc}}(g) \geq R^{\mathrm{cc}}(g) \geq \Omega(\log|\Lambda|)$, and therefore the latter result is equivalent to Equation (1).

The research agenda of [CFK+19]. As discussed above, we would like to prove a lifting theorem that holds for as many inner functions g as possible. Inspired by the literature on direct-sum theorems, [CFK+19] conjectured that lifting theorems should hold for every inner function g that has a sufficiently large information cost $\mathrm{IC}(g)$.

Conjecture 1.2 (special case of [CFK+19, Conj. 1.4])**.** *There exists a constant $c > 0$ such that the following holds. Let $f : \{0,1\}^n \to \mathcal{O}$ and $g : \Lambda \times \Lambda \to \{0,1\}$ be an arbitrary function such that* $\mathrm{IC}(g) \geq c \cdot \log n$*. Then*

$$R^{\mathrm{cc}}(f \circ g^n) = \Omega\left(R^{\mathrm{dt}}(f) \cdot \mathrm{IC}(g)\right).$$

Proving this conjecture is a fairly ambitious goal. As an intermediate goal, [CFK+19] suggested to prove this conjecture for complexity measures that are simpler than $\mathrm{IC}(g)$. In light of their result, it is natural to start with discrepancy. It has long been known that the quantity $\Delta(g) \stackrel{\mathrm{def}}{=} \log\frac{1}{\mathrm{disc}(g)}$ is a lower bound on $R^{\mathrm{cc}}(g)$ up to a constant factor. More recently, it has even been shown that $\Delta(g)$ is a lower bound on $\mathrm{IC}(g)$ up to a constant factor [BW12]. Motivated by this consideration, [CFK+19] suggested the following natural conjecture: for every function g such that $\Delta(g) \geq c \cdot \log n$, it holds that $R^{\mathrm{cc}}(f \circ g^n) = \Omega\left(R^{\mathrm{dt}}(f) \cdot \Delta(g)\right)$ (see Conjecture 1.5 there). The lifting theorem of [CFK+19] proves this conjecture for the special case where $\Delta(g) = \Omega(b)$.

Our result. In this work, we prove the latter conjecture of [CFK+19] in full, by waiving the limitation of $\Delta(g) = \Omega(b)$ from their result, where $b = \log|\Lambda|$. As in previous works, our result holds even if f is replaced with a general search problem $\mathcal{S}$. In what follows, we denote by $R^{dt}_\beta(\mathcal{S})$ and $R^{cc}_\beta(\mathcal{S} \circ g^n)$ the randomized query complexity of $\mathcal{S}$ with error β and the randomized communication complexity of $\mathcal{S} \circ g^n$ with error β respectively. We now state our result formally.

Theorem 1.3 (Main theorem)**.** *There exists a universal constant c such that the following holds: Let $\mathcal{S}$ be a search problem that takes inputs from $\{0,1\}^n$, and let $g : \Lambda \times \Lambda \to \{0,1\}$ be an arbitrary function such that $\Delta(g) \geq c \cdot \log n$. Then*

$$D^{\mathrm{cc}}(\mathcal{S} \circ g^n) = \Omega\left(D^{\mathrm{dt}}(\mathcal{S}) \cdot \Delta(g)\right),$$

and for every $\beta > 0$ it holds that

$$R^{\mathrm{cc}}_\beta(\mathcal{S} \circ g^n) = \Omega\left(\left(R^{\mathrm{dt}}_{\beta'}(\mathcal{S}) - O(1)\right) \cdot \Delta(g)\right),$$

where $\beta' = \beta + 2^{-\Delta(g)/50}$.

Discrepancy with respect to product distributions We note that Definition 1.1 is in fact a special case of the common definition of discrepancy. The general definition refers to an arbitrary distribution μ over $\Lambda \times \Lambda$. The *discrepancy of g with respect to μ* is defined similarly to Definition 1.1 except that the random variables U, V are distributed according to μ rather than the uniform distribution. We show that Theorem 1.3 holds even where the discrepancy is with respect to a product distribution. We do so by reducing the case of product distribution into the case of uniform distribution, more details can be found in Section 7.

Remark 1.4. It is interesting to note that one of the first direct-sum results in the randomized setting went along these lines. In particular, the work of Shaltiel [Sha01] implies that for every function g such that $\Delta(g) \geq c$ for some universal constant c, it holds that $R^{cc}(g^n) = \Omega\left(n \cdot \Delta(g)\right)$. Our main theorem can be viewed as a generalization of that result.

Remark 1.5. A natural question is whether the requirement that $\Delta(g) \geq c \cdot \log n$ is necessary. In principle, it is possible that this requirement could be relaxed. Any such relaxation, however, would imply a lifting theorem that allows inner functions of smaller input length than is currently known, which would be considered a significant breakthrough.

Remark 1.6. In order to facilitate the presentation, we restricted our discussion on the previous works to lifting theorems that hold for every outer function f (and indeed, every search problem S). If one is willing to make certain assumptions on the outer function f, it is possible to prove stronger lifting theorems that in particular allow for a wider variety of inner functions (see, e.g., [She09, SZ09, GP14, HHL16, dRMN+20b, ABK21]).

1.1 Our techniques

Following the previous works, we use a "simulation argument": We show that given a protocol that computes $f \circ g^n$ with communication complexity C, we can construct a decision tree that computes f with query complexity $O(\frac{C}{\Delta(g)})$. In particular, we follow the simulation argument of [CFK+19] and extend their main technical lemma. We now describe this argument in more detail, focusing on the main lemma of [CFK+19] and our extension of that lemma. For simplicity, we focus on the deterministic setting, but the proof in the randomized setting follows similar ideas.

The simulation argument. We assume that we have a protocol Π that computes $f \circ g^n$, and would like to construct a decision tree T that computes f. The basic idea is that given an input $z \in \{0,1\}^n$, the tree T uses the protocol Π to find a pair of inputs $(x, y) \in \Lambda^n \times \Lambda^n$ such that $(f \circ g^n)(x, y) = f(z)$, and then returns the output of Π on (x, y).

In order to find the pair (x, y), the tree T maintains a pair of random variables (X, Y). Initially, the variables (X, Y) are uniformly distributed over $\Lambda^n \times \Lambda^n$. Then, the tree gradually changes the distribution of (X, Y) until they satisfy $(f \circ g^n)(X, Y) = f(z)$ with probability 1, at which point the tree chooses (x, y) to be an arbitrary pair in the support of (X, Y). This manipulation of the distribution of (X, Y) is guided by a simulation of the protocol Π on (X, Y) (hence the name "simulation argument"). Throughout this process, the decision tree maintains the following structure of (X, Y):

- There is a set of coordinates, denoted $F \subseteq [n]$, such that for every $i \in F$ it holds that $g(X_i, Y_i) = z_i$ with probability 1.
- $X_{[n]\setminus F}$ and $Y_{[n]\setminus F}$ are *dense* in the following sense: for every $J \subseteq [n] \setminus F$, the variables X_J and Y_J have high min-entropy.

Intuitively, the set F is the set of coordinates i for which the simulation of Π has already computed $g(X_i, Y_i)$, while for the coordinates $i \in [n] \setminus F$ the value $g(X_i, Y_i)$ is unknown. Initially, the set F is empty, and then it is gradually expanded until it holds that $(f \circ g^n)(X, Y) = f(z)$.

The main lemma of [CFK$^+$19]. Suppose now that as part of the process described above, we would like expand the set F by adding a new set of coordinates $I \subseteq [n] \setminus F$. This means that we should condition the distribution of (X, Y) on the event that $g^I(X_I, Y_I) = z_I$. This conditioning, however, decreases the min-entropy of (X, Y), which might cause $X_{[n]\setminus(F\cup I)}$ and $Y_{[n]\setminus(F\cup I)}$ to lose their density.

In order to resolve this issue, [CFK$^+$19] defined a notion of "sparsifying values" of X and Y. Informally, a value x in the support of X is called *sparsifying* if after conditioning Y on the event $g^I(x_I, Y_I) = z_I$, the variable $Y_{[n]\setminus(F\cup I)}$ ceases to be dense. A sparsifying value of Y is defined similarly. It is not hard to see that if X and Y do not have any sparsifying values in their supports, then the density of $X_{[n]\setminus(F\cup I)}$ and $Y_{[n]\setminus(F\cup I)}$ is maintained after the conditioning on $g^I(X_I, Y_I) = z_I$. Therefore, [CFK$^+$19] design their decision tree such that before the conditioning on the event $g^I(X_I, Y_I) = z_I$, the tree first removes the sparsifying values from the supports of X and Y.

The removal of sparsifying values, however, raises another issue: when we remove values from the supports of X and Y, we decrease the min-entropy of X and Y. In particular, the removal of the sparsifying values might cause $X_{[n]\setminus F}$ and $Y_{[n]\setminus F}$ to lose their density. This issue is resolved by the main technical lemma of [CFK$^+$19]. Informally, this lemma says that if $X_{[n]\setminus F}$ and $Y_{[n]\setminus F}$ are dense, then the sparsifying values are very rare. This means that the removal of these values barely changes the min-entropy of X and Y, and in particular, does not violate the density property.

Our contribution. Recall that the lifting theorem of [CFK$^+$19] requires that $\Delta(g) = \Omega(b)$ (where $b = \log |\Lambda|$), and that our goal is to waive that requirement. Unfortunately, it turns out that main lemma of [CFK$^+$19] fails when $\Delta(g)$ is very small relatively to b. In fact, in Section 6 we provide an example in which *all* the values in the support of X are sparsifying.

Hence, unlike [CFK$^+$19], we cannot afford to remove the sparsifying values before conditioning on the event $g^I(X_I, Y_I) = z_I$. Therefore, in our simulation, the variables $X_{[n]\setminus F}$ and $Y_{[n]\setminus F}$ sometimes lose their density after the conditioning. Nevertheless, we observe that even if the density property breaks in this way, it can often be restored by removing some more values from the supports of X and Y. We formalize this intuition by defining a notion of "recoverable values". Informally, a value x in the support of X is called *recoverable* if after conditioning Y on the event $g^I(x_I, Y_I) = z_I$, the density of $Y_{[n]\setminus(F\cup I)}$ can be restored by discarding some values from its support.

Our main lemma says, informally, that if $X_{[n]\setminus F}$ and $Y_{[n]\setminus F}$ are dense, then almost all the values of X and Y are recoverable. In particular, we can afford to remove the unrecoverable values of X and Y without violating their density. Given our lemma, it is easy to fix the simulation argument of [CFK$^+$19]: whenever our decision tree is about to condition on an event $g^I(x_I, Y_I) = z_I$, it first discards the unrecoverable values of X and Y; then, after the conditioning, the decision tree restores the density property by discarding some additional values. The rest of our argument proceeds exactly as in [CFK$^+$19].

The proof of our main lemma. The definition of a sparsifying value of X can be stated as follows: the value x is sparsifying if there exists a value y_J such that the probability

$$\Pr\left[Y_J = y_J \mid g(x_I, Y_I) = z_I\right] \tag{2}$$

is too high. On the other hand, it can be showed that a value x is *unrecoverable* if there are *many* such corresponding values y_J. Indeed, if there are only few such values y_J, then we can recover the density of $Y_{[n]\setminus(F\cup I)}$ by discarding them.

Very roughly, the main lemma of [CFK$^+$19] is proved by showing that for every y_J, there is only a very small number of corresponding x's for which the latter probability is too high. Then, by taking union bound over all possible choices of y_J, it follows that there are only few values x for which there exists some corresponding y_J. In other words, there are only few sparsifying values.

This argument works in the setting of [CFK$^+$19] because they can prove a very strong upper bound on the number of values x for a single y_J — indeed, the bound is sufficiently strong to survive the union bound. In our setting, on the other hand, the fact that we assume a smaller value of $\Delta(g)$ translates to a weaker bound on the number of values x for a single y_J. In particular, we cannot afford to use the union bound. Instead, we take a different approach: we observe that, since for every y_J there is only a small number of corresponding x's, it follows by an averaging argument that there can only be a small number of x's that have *many* corresponding y_J's. In other words, there can only be a small number of *unrecoverable* x's.

Implementing this idea is more difficult than it might seem at a first glance. The key difficulty is that when we say "values x that have many corresponding y_J's" we do not refer to the absolute number of y_J's but rather to their probability mass. Specifically, the probability distribution according to which the y_J's should be counted is the probability distribution of Equation (2). Unfortunately, this means that for every value x, we count the y_J's according to a different distribution, which renders a simple averaging argument impossible. We overcome this difficulty by proving a finer upper bound on the number of x's for each y_J and using a careful bucketing scheme for the averaging argument.

2 Preliminaries

We assume familiarity with the basic definitions of communication complexity (see, e.g., [KN97]). For any $n \in \mathbb{N}$, we denote $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$. We denote by $c \in \mathbb{N}$ some large universal constant that will be chosen later ($c = 1000$ will do). For the rest of this paper, we fix some natural number $n \in \mathbb{N}$, a finite set Λ, and denote $b = \log |\Lambda|$. We fix $g : \Lambda \times \Lambda \to \{0, 1\}$ to be an arbitrary function such that $\Delta(g) \geq c \cdot \log n$ (where $\Delta(g) \stackrel{\text{def}}{=} \log \frac{1}{\operatorname{disc}(g)}$), and abbreviate $\Delta \stackrel{\text{def}}{=} \Delta(g)$. Since our main theorem holds trivially when $n = 1$, we assume that $n \geq 2$. Furthermore, throughout this paper, X and Y denote independent random variables that take values from Λ^n.

Let $I \subseteq [n]$ be a set of coordinates. We denote by Λ^I the set of strings over alphabet Λ of length $|I|$ and index the coordinates of the string by I. Given a string $x \in \Lambda^n$, we denote by $x_I \in \Lambda^I$ the projection of x to the coordinates in I (in particular, $x_\emptyset$ is defined to be the empty string). We denote by $g^I : \Lambda^I \times \Lambda^I \to \{0, 1\}^I$ the function that takes as inputs $|I|$ pairs from $\Lambda \times \Lambda$ that are indexed by I, and outputs the string in $\{0, 1\}^I$ whose i-th bit is the output of g on the i-th pair. In particular, we denote $g^n \stackrel{\text{def}}{=} g^{[n]}$, so g^n is the direct-sum function that takes as inputs $x, y \in \Lambda^n$ and outputs the binary string

$$g^n(x, y) \stackrel{\text{def}}{=} (g(x_1, y_1), \ldots, g(x_n, y_n)) .$$

We denote by $g^{\oplus I} : \Lambda^I \times \Lambda^I \to \{0, 1\}$ the function that given $x, y \in \Lambda^I$ outputs the parity of the string $g^I(x, y)$. The following bound is used throughout the paper.

Proposition 2.1. *Assume that $\beta, l \in \mathbb{R}$ such that $\beta \leq 1$ and $l \geq 1$. Then it holds that*

$$\sum_{S \subseteq [n], |S| \geq l} \beta^{|S|} \cdot \frac{1}{n^{|S|}} \leq 2\beta^l$$

.

Proof. It holds that

$$\begin{aligned}
\sum_{S \subseteq [n], |S| \geq l} \beta^{|S|} \frac{1}{n^{|S|}} &\leq \sum_{s=\lceil l \rceil}^{n} \binom{n}{s} \beta^s \frac{1}{n^s} \\
&\leq \sum_{s=\lceil l \rceil}^{n} \beta^s \frac{1}{s!} && (\binom{n}{s} \leq \frac{n^s}{s!}) \\
&\leq 2 \sum_{s=\lceil l \rceil}^{\infty} \left(\frac{\beta}{2}\right)^s && (s! \geq 2^{s-1}) \\
&\leq 2 \frac{\left(\frac{\beta}{2}\right)^{\lceil l \rceil}}{1 - \frac{\beta}{2}} && \text{(geometric sum)} \\
&\leq 2 \frac{\left(\frac{\beta}{2}\right)^{l}}{1 - \frac{\beta}{2}} && (\frac{\beta}{2} \leq 1 \text{ and } l \leq \lceil l \rceil) \\
&= 2\beta^l \cdot \frac{\left(\frac{1}{2}\right)^{l}}{1 - \frac{\beta}{2}} \\
&\leq 2\beta^l && (\frac{1}{2} \leq 1 - \frac{\beta}{2}). \quad \blacksquare
\end{aligned}$$

Search problems. Given a finite set of inputs $\mathcal{I}$ and a finite set of outputs $\mathcal{O}$, a *search problem* $\mathcal{S}$ is a relation between $\mathcal{I}$ and $\mathcal{O}$. Given $z \in \mathcal{I}$, we denote by $\mathcal{S}(z)$ the set of outputs $o \in \mathcal{O}$ such that $(z, o) \in \mathcal{S}$. Without loss of generality, we may assume that $\mathcal{S}(z)$ is always non-empty, since otherwise we can set $\mathcal{S}(z) = \{\perp\}$ where $\perp$ is some special failure symbol that does not belong to $\mathcal{O}$.

Intuitively, a search problem $\mathcal{S}$ represents the following task: given an input $z \in \mathcal{I}$, find a solution $o \in \mathcal{S}(z)$. In particular, if $\mathcal{I} = \mathcal{X} \times \mathcal{Y}$ for some finite sets $\mathcal{X}, \mathcal{Y}$, we say that a deterministic protocol Π *solves* $\mathcal{S}$ if for every input $(x, y) \in \mathcal{I}$, the output of Π is in $\mathcal{S}(x, y)$. We say that a randomized protocol Π *solves* $\mathcal{S}$ *with error* β if for every input $(x, y) \in \mathcal{I}$, the output of Π is in $\mathcal{S}(x, y)$ with probability at least $1 - \beta$. We denote by $D^{\mathrm{cc}}(\mathcal{S})$ the deterministic communication complexity of a search problem $\mathcal{S}$. Given $\beta > 0$, we denote by $R^{\mathrm{cc}}_{\beta}(\mathcal{S})$-th randomized (public-coin) communication complexity of $\mathcal{S}$ with error β. In case that β is omitted one should assume that $\beta = \frac{1}{3}$.

Given a search problem $\mathcal{S} \subseteq \{0,1\}^n \times \mathcal{O}$, we denote by $\mathcal{S} \circ g^n \subseteq (\Lambda^n \times \Lambda^n) \times \mathcal{O}$ the search problem that satisfies $\mathcal{S} \circ g^n(x, y) = \mathcal{S}(g^n(x, y))$ for every $x, y \in \Lambda^n$.

2.1 Decision trees

Informally, a decision tree is an algorithm that solves a search problem $\mathcal{S} \subseteq \{0,1\}^n \times \mathcal{O}$ by querying the individual bits of its input. The decision tree is computationally unbounded, and its complexity is measured by the number of bits it queries.

Formally, a *deterministic decision tree* T from $\{0,1\}^n$ to $\mathcal{O}$ is a binary tree in which every internal node is labeled with a coordinate in $[n]$ (which represents a query), every edge is labeled by a bit (which represents the answer to the query), and every leaf is labeled by an output in $\mathcal{O}$. Such a tree computes a function from $\{0,1\}^n$ to $\mathcal{O}$ in the natural way, and with a slight abuse of notation, we denote this function by T as well. The *query complexity* of T is the depth of the tree. We say that a tree T solves a search problem $\mathcal{S} \subseteq \{0,1\}^n \times \mathcal{O}$ if for every $z \in \{0,1\}^n$ it holds that $T(z) \in \mathcal{S}(z)$. The *deterministic query complexity of* $\mathcal{S}$, denoted $D^{\text{dt}}(\mathcal{S})$, is the minimal query complexity of a decision tree that solves $\mathcal{S}$.

A *randomized decision tree* T is a random variable that takes deterministic decision trees as values. The *query complexity* of T is the maximal depth of a tree in the support of T. We say that T *solves a search problem* $\mathcal{S} \subseteq \{0,1\}^n \times \mathcal{O}$ *with error* β if for every $z \in \{0,1\}^n$ it holds that

$$\Pr\left[T(z) \in \mathcal{S}(z)\right] \geq 1 - \beta.$$

The *randomized query complexity of* $\mathcal{S}$ *with error* β, denoted $R^{\text{dt}}_{\beta}(\mathcal{S})$, is the minimal query complexity of a randomized decision tree that solves $\mathcal{S}$ with error β. In case that β is omitted, one should assume that $\beta = \frac{1}{3}$.

2.2 Probability

Below we recall some standard definitions and facts from probability theory. Recall that the exponential distribution, denoted $\text{Ex}(\lambda)$, is defined by the following cumulative probability distribution

$$1 - e^{-\lambda x} \text{ for } x \geq 0.$$

The Erlang distribution, denoted $\text{Erl}(k, \lambda)$, is defined as the sum of k exponential variables with parameters λ and its cumulative probability distribution of $\text{Erl}(k, \lambda)$ is

$$1 - e^{-\lambda x} \sum_{i=0}^{k-1} \frac{(\lambda x)^i}{i!} \text{ for } x \geq 0.$$

Given two distributions μ_1, μ_2 over a finite sample space Ω, the *statistical distance* (or *total variation distance*) between μ_1 and μ_2 is

$$|\mu_1 - \mu_2| \stackrel{\text{def}}{=} \max_{\mathcal{E} \subseteq \Omega} \left\{ |\mu_1(\mathcal{E}) - \mu_2(\mathcal{E})| \right\}.$$

We say that μ_1 and μ_2 are ε*-close* if $|\mu_1 - \mu_2| \leq \varepsilon$.

Fact 2.2. *Let $\mathcal{E}$ be some event and μ some distribution. Then*

$$|\mu - (\mu \mid \mathcal{E})| \leq 1 - \Pr\left[\mathcal{E}\right].$$

Let V be a random variable that takes values from a finite set $\mathcal{V}$. The *min-entropy* of V, denoted $H_\infty(V)$, is the largest number $k \in \mathbb{R}$ such that for every value x it holds that $\Pr\left[V = v\right] \leq 2^{-k}$. The *deficiency* of V is defined as

$$D_\infty(V) \stackrel{\text{def}}{=} \log|\mathcal{V}| - H_\infty(V).$$

Intuitively, the deficiency of V measures the amount of information that is known about V relative to the uniform distribution. Deficiency has the following easy-to-prove properties.

Fact 2.3. *For every random variable V it holds that*

$$D_{\infty}(V) \geq 0.$$

Fact 2.4. *For every random variable V and an event $\mathcal{E}$ with positive probability it holds that*

$$D_{\infty}(V \mid \mathcal{E}) \leq D_{\infty}(V) + \log \frac{1}{\Pr[\mathcal{E}]}.$$

Fact 2.5. *Let V_1, V_2 be random variables. Then,*

$$D_{\infty}(V_1) \leq D_{\infty}(V_1, V_2).$$

2.2.1 Vazirani's Lemma

Given a boolean random variable V, we denote the bias of V by

$$\mathrm{bias}(V) \stackrel{\mathrm{def}}{=} |\Pr[V=0] - \Pr[V=1]|.$$

Vazirani's lemma is a useful result that says that a random string is close to being uniformly distributed if the parity of every set of bits in the string has a small bias. We use the following variants of the lemma.

Lemma 2.6 ([GLM+15])**.** *Let $\varepsilon > 0$, and let Z be a random variable taking values in $\{0,1\}^m$. If for every non-empty set $S \subseteq [m]$ it holds that*

$$\mathrm{bias}(\bigoplus_{i \in S} Z_i) \leq \varepsilon \cdot (2 \cdot m)^{-|S|} \tag{3}$$

then for every $z \in \{0,1\}^m$ it holds that

$$(1-\varepsilon) \cdot \frac{1}{2^m} \leq \Pr[Z = z] \leq (1+\varepsilon) \cdot \frac{1}{2^m}.$$

The following version of Vazirani's lemma bounds the deficiency of the random variable via a weaker assumption on the biases

Lemma 2.7 ([CFK+19])**.** *Let $t \in \mathbb{N}$ be such that $t \geq 1$, and let Z be a random variable taking values in $\{0,1\}^m$. If for every set $S \subseteq [m]$ such that $|S| \geq t$ it holds that*

$$\mathrm{bias}\left(\bigoplus_{i \in S} Z_I\right) \leq (2 \cdot m)^{-|S|}$$

then $D_{\infty}(Z) \leq t \log m + 1$.

2.2.2 Coupling

Let μ_1, μ_2 be two distributions over a sample space Ω. A *coupling of μ_1 and μ_2* is a distribution ν over the sample space Ω^2 whose marginals over the first and second coordinates are μ_1 and μ_2 respectively. The following standard fact characterizes the statistical distance between μ_1 and μ_2 using couplings.

Fact 2.8. *Let μ_1, μ_2 be two distributions over a sample space Ω. For every coupling ν of μ_1 and μ_2 it holds that*

$$|\mu_1 - \mu_2| = \min_{\nu} \Pr_{(V_1, V_2) \leftarrow \nu}[V_1 \neq V_2],$$

where the minimum is taken over all couplings ν of μ_1 and μ_2. In particular, any coupling gives an upper bound on the statistical distance.

2.3 Prefix-free codes

A set of strings $C \subseteq \{0,1\}^*$ is called a *prefix-free code* if no string in C is a prefix of another string in C. Given a string $w \in \{0,1\}^*$, we denote its length by $|w|$. We use the following simple corollary of Kraft's inequality.

Fact 2.9 (Corollary of Kraft's inequality)**.** *Let $C \subseteq \{0,1\}^*$ be a finite prefix-free code, and let W be a random string that takes values from C. Then, there exists a string $w \in C$ such that $\Pr[W = w] \geq 2^{-|w|}$.*

A simple proof of Fact 2.9 can be found in [CFK+19].

2.4 Properties of discrepancy

Recall that $\Delta \stackrel{\text{def}}{=} \Delta(g) = \log \frac{1}{\operatorname{disc}(g)}$, that X, Y are independent random variables that take values from Λ^n, and that $g^{\oplus S} : \Lambda^S \times \Lambda^S \to \{0,1\}$ is the function that given $x_S, y_S \in \Lambda^S$ outputs the parity of the string $g^S(x_S, y_S)$. We use the following properties of discrepancy. In what follows, the parameter λ controls $\operatorname{bias}\left(g^{\oplus S}(x_S, Y_S)\right)$ and the parameter γ controls the error probability.

Lemma 2.10 (see, e.g., [CFK+19, Cor. 2.12])**.** *Let $\lambda > 0$ and let $S \subseteq [n]$. If*

$$D_\infty(X_S) + D_\infty(Y_S) \leq (\Delta(g) - 6 - \lambda) \cdot |S|$$

then

$$\operatorname{bias}\left(g^{\oplus S}(X_S, Y_S)\right) \leq 2^{-\lambda|S|}.$$

Lemma 2.11 (see, e.g., [CFK+19, Cor. 2.13])**.** *Let $\gamma, \lambda > 0$ and let $S \subseteq [n]$. If*

$$D_\infty(X_S) + D_\infty(Y_S) \leq (\Delta(g) - 7 - \gamma - \lambda) \cdot |S|$$

then the probability that X takes a value $x \in \Lambda^n$ such that

$$\operatorname{bias}\left(g^{\oplus S}(x_S, Y_S)\right) > 2^{-\lambda|S|}$$

is less than $2^{-\gamma|S|}$.

2.5 Background from [CFK+19]

In this section, we review some definitions and results from [CFK+19] that we use in our proofs. We present those definitions and results somewhat differently than [CFK+19] in order to streamline the proofs in the setting where $\Delta \ll b$. The most significant deviation from the presentation of [CFK+19] is the following definition of a σ-sparse random variable, which replaces the notion of a δ-dense random variable from [GLM+15, GPW17, CFK+19]. Both notions are aimed to capture a random variable over Λ^n on which very little information is known.

Definition 2.12. Let X be a random variable taking values from Λ^n, and let $\sigma > 0$. We say that X is *σ-sparse* if for every set $S \subseteq [n]$ it holds that $D_\infty(X_S) \leq \sigma \cdot \Delta \cdot |S|$.

Remark 2.13. The relation between the above definition to the notion of δ-dense random variable of [GLM+15, GPW17, CFK+19] is the following: X is σ-sparse if and only if it is $(1 - \frac{\Delta}{b} \cdot \sigma)$-dense.

As explained in Section 1.1, our proof relies on a simulation argument that takes a protocol Π for $\mathcal{S} \circ g^n$ and constructs a decision tree T for $\mathcal{S}$. We use the following notion of restriction to keep track of the queries that the tree makes and their answers.

Definition 2.14. A *restriction* ρ is a string in $\{0,1,*\}^n$. We say that a coordinate $i \in [n]$ is *free* in ρ if $\rho_i = *$, and otherwise we say that i is *fixed*. Given a restriction $\rho \in \{0,1,*\}^n$, we denote by $\mathrm{free}(\rho)$ and $\mathrm{fix}(\rho)$ the sets of free and fixed coordinates of ρ respectively. We say that a string $z \in \{0,1\}^n$ is *consistent* with ρ if $z_{\mathrm{fix}(\rho)} = \rho_{\mathrm{fix}(\rho)}$.

Intuitively, $\mathrm{fix}(\rho)$ represents the queries that have been made so far, and $\mathrm{free}(\rho)$ represents the coordinates that have not been queried yet. As explained in Section 1.1, the decision tree maintains a pair of random variables X, Y with a certain structure, which is captured by the following definition.

Definition 2.15. Let $\rho \in \{0,1,*\}^n$ be a restriction, let $\sigma_X, \sigma_Y > 0$, and let X, Y be independent random variables that take values from Λ^n. We say that X and Y are $(\rho, \sigma_X, \sigma_Y)$*-structured* if there exist $\sigma_X, \sigma_Y > 0$ such that $X_{\mathrm{free}(\rho)}$ and $Y_{\mathrm{free}(\rho)}$ are σ_X-sparse and σ_Y-sparse respectively and

$$g^{\mathrm{fix}(\rho)}\left(X_{\mathrm{fix}(\rho)}, Y_{\mathrm{fix}(\rho)}\right) = \rho_{\mathrm{fix}(\rho)}.$$

Intuitively, this structure says that $(X_{\mathrm{fix}(\rho)}, Y_{\mathrm{fix}(\rho)})$ must be consistent with the queries that the decision tree has made so far, and that the simulated protocol Π does not know much about the free coordinates. The following proposition formalizes the intuition that the simulated protocol does not know the value of g on the free coordinates. In what follows, the parameter γ controls how close are the values of g on the free coordinates to being uniformly distributed.

Proposition 2.16 ([CFK$^+$19, Prop. 3.10])**.** *Let $\gamma \geq 0$. Let X, Y be random variables that are $(\rho, \sigma_X, \sigma_Y)$-structured for $\sigma_X + \sigma_Y \leq 1 - \frac{8}{c} - \gamma$, and let $I = \mathrm{free}(\rho)$. Then,*

$$\forall z_I \in \{0,1\}^I : \Pr\left[g^I\left(X_I, Y_I\right) = z_I\right] \in \left(1 \pm 2^{-\gamma\Delta}\right) \cdot 2^{-|I|}.$$

Next, we state the uniform marginals lemma of [CFK$^+$19] (which generalized an earlier lemma of [GPW17]). Intuitively, this lemma says that the simulated protocol Π cannot distinguish between the distribution (X, Y) and the same distribution conditioned on $g^n(X,Y) = z$. In what follows, the parameter γ controls the indistinguishability.

Lemma 2.17 (Uniform marginals lemma, [CFK$^+$19, Lemma 3.4])**.** *Let $\gamma \geq 0$, let ρ be a restriction, and let $z \in \{0,1\}^n$ be a string that is consistent with ρ. Let X, Y be $(\rho, \sigma_X, \sigma_Y)$-structured random variables that are uniformly distributed over sets $\mathcal{X}, \mathcal{Y} \subseteq \Lambda^n$ respectively such that $\sigma_X + \sigma_Y \leq 1 - \frac{10}{c} - \gamma$. Let (X', Y') be uniformly distributed over $(g^n)^{-1}(z) \cap (\mathcal{X} \times \mathcal{Y})$. Then, X and Y are $2^{-\gamma\Delta}$-close to X' and Y' respectively.*

The following folklore fact allows us to transform an arbitrary variable over Λ^n into a σ-sparse one by fixing some of its coordinates.

Proposition 2.18 (see, e.g., [CFK$^+$19, Prop. 3.6])**.** *Let X be a random variable, let $\sigma_X > 0$, and let $I \subseteq [n]$ be a maximal subset of coordinates such that $D_\infty(X_I) > \sigma_X \cdot \Delta \cdot |I|$. Let $x_I \in \Lambda^I$ be a value such that*

$$\Pr\left[X_I = x_I\right] > 2^{\sigma_X \cdot \Delta \cdot |I| - b \cdot |I|}.$$

Then, the random variable $X_{[n]-I} \mid X_I = x_I$ is σ_X-sparse.

Proposition 2.18 is useful in the deterministic setting, since in this setting the decision tree is free to condition the distributions of X, Y in any way that does not increase their sparsity. In the randomized setting, however, the decision tree is more restricted, and cannot condition the inputs on events such as $X_I = x_I$ which may have very low probability. In [GPW17], this issue was resolved by observing that the probability space can be partitioned to disjoint events of the form $X_I = x_I$, and that the randomized simulation can use such a partition to achieve the same effect of Proposition 2.18. This leads to the following lemma, which we use as well.

Lemma 2.19 (Density-restoring partition [GPW17]). *Let $\mathcal{X} \subseteq \Lambda^n$ denote the support of X, and let $\sigma_X > 0$. Then, there exists a partition*

$$\mathcal{X} \stackrel{\text{def}}{=} \mathcal{X}^1 \cup \cdots \cup \mathcal{X}^\ell$$

where each $\mathcal{X}^j$ is associated with a set $I_j \subseteq [n]$ and a value $x_j \in \Lambda^{I_j}$ such that:

- $X_{I_j} \mid X \in \mathcal{X}^j$ *is fixed to x_j.*
- $X_{[n]-I_j} \mid X \in \mathcal{X}^j$ *is σ_X-sparse.*

Moreover, if we denote $p_{\geq j} \stackrel{\text{def}}{=} \Pr\left[X \in \mathcal{X}^j \cup \ldots \cup \mathcal{X}^\ell\right]$, then it holds that

$$D_\infty(X_{[n]-I_j} \mid X \in \mathcal{X}^j) \leq D_\infty(X) + \sigma_X \cdot \Delta \cdot |I_j| + \log \frac{1}{p_{\geq j}}.$$

In what follows, we recall some definitions and the main lemma from [CFK$^+$19]. Those definitions and lemma are not used to prove our main result, but are given here so they can be compared to our definitions and main lemma.

Definition 2.20 ([CFK$^+$19]). Let Y be a random variable taking values from Λ^n. We say that a value $x \in \Lambda^n$ is *leaking* if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that

$$\Pr\left[g^I\left(x_I, Y_I\right) = z_I\right] < 2^{-|I|-1}.$$

Let $\sigma_Y, \varepsilon > 0$, and suppose that Y is σ_Y-sparse. We say that a value $x \in \Lambda^n$ is *ε-sparsifying* if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that the random variable $Y_{[n]-I} \mid g^I\left(x_I, Y_I\right) = z_I$ is not $(\sigma_Y + \varepsilon)$-sparse. We say that a value $x \in \Lambda^n$ is *ε-dangerous* if it is either leaking or ε-sparsifying.

Lemma 2.21 (main lemma of [CFK$^+$19]). *There exists universal constants h, c such that the following holds: Let b be some number such that $b \geq c \cdot \log n$ and let $\gamma, \varepsilon, \sigma_X, \sigma_Y > 0$ be such that $\varepsilon \geq \frac{4}{\Delta}$, and $\sigma_X + \sigma_Y \leq 1 - \frac{h \cdot b \cdot \log n}{\Delta^2 \cdot \varepsilon} - \gamma$. Let X, Y be $(\rho, \sigma_X, \sigma_Y)$-structured random variables. Then, the probability that $X_{\text{free}(\rho)}$ takes a value that is ε-dangerous for $Y_{\text{free}(\rho)}$ is at most $2^{-\gamma\Delta}$.*

3 The main lemma

In this section, we state and prove our main lemma. As discussed in the introduction, our simulation argument maintains a pair of random variables $X, Y \in \Lambda^n$. A crucial part of the simulation consists of removing certain "unsafe" values from the supports of these variables. Our main lemma says that almost all values are safe.

There are two criteria for a value $x \in \Lambda^n$ to be "safe": First, it should hold that after we condition Y on an event of the form $g^I\left(x_I, Y_I\right) = z_I$, the density of Y can be recovered by conditioning on a high probability event (such values are called *recoverable*). Second, it holds that $g^n(x, Y)$ is distributed almost uniformly (such values are called *almost uniform*). This guarantees that from the point of view of Alice, who knows X, the value of $g^n(X, Y)$ is distributed almost uniformly. For the simplicity of notation we denote:

Notation 3.1. For $x \in \Lambda^n$ we define the random variable $Z^x \stackrel{\text{def}}{=} g^n\left(x, Y\right)$.

We now formally define the notions of safe, recoverable and almost uniform values.

Definition 3.2 (safe values)**.** Let $\alpha \geq 0$. Let Y be a random variable in Λ^n, let $\sigma_Y > 0$ be the minimal value for which Y is σ_Y-sparse and let $x \in \Lambda^n$. We say that x is *almost uniform* (for Y) if for any assignment $z \in \{0,1\}^n$ it holds that

$$\Pr\left[Z^x = z\right] \in 2^{-n}\left(1 \pm 2^{-\frac{\Delta}{10}}\right).$$

We say that x is *α-recoverable* (for Y) if for all $I \subseteq [n]$ and z_I the following holds: there exists an event $\mathcal{E}$ that only depend on Y such that $\Pr\left[\mathcal{E} \mid Z_I^x = z_I\right] \geq 1 - 2^{-\alpha\Delta}$ and such that the random variable

$$Y_{[n]-I} \mid \mathcal{E} \text{ and } Z_I^x = z_I$$

is $(\sigma_Y + \frac{4}{c})$-sparse. We say that x is *α-safe* (for Y) if it is both almost uniform and α-recoverable. Almost uniform, recoverable, and safe values for X are defined analogously.

We turn to state our main lemma.

Lemma 3.3 (Main lemma)**.** *For every $c > 0$ the following holds. Let $\alpha \geq \frac{1}{\Delta}$ and $\gamma > 0$, and let X and Y be independent random variables such that X is σ_X-sparse. Let $\sigma_Y > 0$ be the minimal value for which Y is σ_Y-sparse. If $\sigma_X + 2\sigma_Y \leq \frac{9}{10} - \frac{25}{c} - \gamma - \alpha$, then*

$$\Pr_{x \sim X}\left[x \textit{ is not } \alpha\textit{-safe for } Y\right] \leq 2^{-\gamma \cdot \Delta}.$$

In the rest of this section, we prove the main lemma. Let $\alpha, \sigma_X, \sigma_Y$ be as in the lemma. Let X, Y be independent random variables such that X is σ_X-sparse and Y is σ_Y-sparse. The following two propositions, which are proved in Sections 3.1 and 3.2, upper bound the probabilities that X takes a value that is not almost uniform or not recoverable respectively. Lemma 3.3 follows immediately from the following propositions using the union bound.

Proposition 3.4. *Let $\gamma > 0$ be a real number such that $\sigma_X + \sigma_Y \leq \frac{9}{10} - \frac{11}{c} - \gamma$. The probability that X takes a value that is not almost uniform is at most $2^{-\gamma \cdot \Delta}$.*

Proposition 3.5. *Let $\gamma > 0$ and assume that $\sigma_X + 2 \cdot \sigma_Y \leq 1 - \frac{24}{c} - \gamma - \alpha$. Then, the probability that X takes an almost uniform value x that is not α-recoverable is at most $2^{-\gamma \cdot \Delta}$.*

Proof of Lemma 3.3 from Propositions 3.4 and 3.5 Any value that is not α-safe is either not almost uniform or almost uniform but not α-recoverable. By applying Proposition 3.4 with $\gamma = \gamma + \frac{1}{c}$, it follows that the probability that X takes a value that is not almost uniform is at most $2^{-(\gamma + \frac{1}{c}) \cdot \Delta} \leq 2^{-\gamma\Delta - 1}$. By applying Proposition 3.5 with $\gamma = \gamma + \frac{1}{c}$, and $\alpha = \alpha$, it follows that the probability that X takes an almost uniform value that is not α-recoverable is at most $2^{-(\gamma + \frac{1}{c}) \cdot \Delta} \leq 2^{-\gamma\Delta - 1}$. Therefore, the probability that X takes a value that is not α-safe value is at most $2^{-\gamma\Delta - 1} + 2^{-\gamma\Delta - 1} = 2^{-\gamma\Delta}$. ■

Remark 3.6. We now compare our notion of safe values to the notion of dangerous values in [CFK+19].

- Our requirement that safe values would be almost uniform corresponds to the requirement of [CFK+19] that safe values would be non-leaking. However, our definition is stronger: Both definitions bound the probability of $\Pr\left[Z_I^x = z_I\right]$. The definition of almost-uniform values bounds the probability tightly from both above and below while the definition of non-leaking values bounds it only from below. We use our stronger requirement to bound the probability of a certain event in Section 5.2.1.

- Our requirement that safe values would be recoverable corresponds to the requirement of [CFK$^+$19] that safe values would be non-sparsifying. Our requirement is weaker: Both definitions regard the sparsity of the random variable $Y_{[n]-I} \mid Z_I^x = z_I$. While the definition of [CFK$^+$19] requires the random variable to have low sparsity, our definition only requires that the sparsity of the random variable can be made low by conditioning on an additional event. As discussed in the introduction, this weakening is necessary as when $\Delta \ll b$, there might not be enough values x that satisfy the stronger requirement (see Section 6).

3.1 Proof of Proposition 3.4

In this section we prove Proposition 3.4, restated next, following the ideas of [CFK$^+$19]. Essentially, the proof uses Vazirani's lemma to reduce the problem to bounding $\text{bias}(g^{\oplus S}(x_S, Y_S))$ for most values of x. This is done using the fact that X and Y have low sparsity together with the low discrepancy of g.

Proposition 3.4. *Let $\gamma > 0$ be a real number such that $\sigma_X + \sigma_Y \le \frac{9}{10} - \frac{11}{c} - \gamma$. The probability that X takes a value that is not almost uniform is at most $2^{-\gamma\cdot\Delta}$.*

Proof. We start by observing that for every $x \in \Lambda^n$, if it holds that

$$\text{bias}(g^{\oplus S}(x_S, Y_S)) \le 2^{-\frac{\Delta}{10}} \cdot (2n)^{-|S|}$$

for every non-empty set $S \subseteq [n]$, then by first variant of Vazirani's lemma (Lemma 2.6) we get that x is almost uniform.

It remains to show that with probability at least $1 - 2^{-\gamma\cdot\Delta}$ the random variable X takes a value x that satisfies the latter condition on the biases. We start by lower bounding the probability that $\text{bias}(g^{\oplus S}(x_S, Y_S)) \le 2^{-\frac{\Delta}{10}} \cdot (2n)^{-|S|}$ for a specific set $S \subseteq [n]$. Fix a non-empty set $S \subseteq [n]$. By assumption, it holds that

$$\begin{aligned} D_\infty(X_S) + D_\infty(Y_S) &\le (1 - \frac{11}{c} - \gamma - \frac{1}{10}) \cdot \Delta \cdot |S| \\ &= \left(\Delta - \frac{7\Delta}{c} - \gamma\Delta - \frac{\Delta}{10} - \frac{4\Delta}{c}\right) \cdot |S| \\ &\le \left(\Delta - 7 - \gamma\Delta - \frac{\Delta}{10} - 2\log n - 2\right) \cdot |S| . \qquad (\Delta \ge c \log n) \end{aligned}$$

By applying Lemma 2.11 with $\gamma = \gamma\Delta + \log n + 1$ and $\lambda = \log n + 1 + \frac{\Delta}{10}$ it follows that with probability at least $1 - 2^{-\gamma\Delta-1} \cdot \frac{1}{n^{|S|}}$, the random variable X takes a value x such that

$$\text{bias}\left(g^{\oplus S}(x_S, Y_S)\right) \le 2^{-\frac{\Delta}{10}} \cdot (2n)^{-|S|} .$$

Next, by taking the union bound over all non-empty sets $S \subseteq [n]$, it follows that

$$\Pr\left[\exists S \ne \emptyset : \text{bias}\left(g^{\oplus S}(x_S, Y_S)\right) > 2^{-\frac{\Delta}{10}} \cdot (2n)^{-|S|}\right] \le \sum_{S\subseteq[n]:S\ne\emptyset} 2^{-\gamma\Delta-1} \cdot \frac{1}{n^{|S|}} < 2^{-\gamma\Delta-1} \cdot 2 = 2^{-\gamma\Delta}$$

by Proposition 2.1. Therefore with probability at least $1 - 2^{-\gamma\Delta}$, the random variable X takes a value x such that $\text{bias}(g^{\oplus S}(x_S, Y_S)) \le 2^{-\frac{\Delta}{10}} \cdot (2n)^{-|S|}$ for all non-empty sets $S \subseteq [n]$, as required. ∎

3.2 Proof of Proposition 3.5

In the following subsection we prove Proposition 3.5. To this end, we introduce the notion of light values. A value x is *light* if after conditioning on x the distribution $Y_J \mid Z_{I,x_I} = z_I$ is "mostly" $\left(\sigma_Y + \frac{4}{c}\right)$-sparse, that is, for almost all values y_J it holds that

$$\Pr\left[Y_J = y_J \mid Z_I^x = z_I\right] \le 2^{\left(\sigma_Y + \frac{4}{c}\right)\cdot\Delta\cdot|J| - b\cdot|J|}.$$

The term "light" is motivated by the intuitive idea that the values y_J that violate the sparsity requirement are "heavy" in terms of their probability mass, so a "light" value x is one that does not have many "heavy" values y_J. We show that when a values x is light then it also recoverable. Intuitively, the density of Y_J can be restored by conditioning on the high probability event that the heavy values are not selected. We now formally define light values. In the following definition of heavy value there is an additional parameter t, which measures by how much the heavy value y_J violates the $(\sigma_Y + \frac{4}{c})$-sparsity. This parameter is used later in the proof.

Definition 3.7. Let $x \in \Lambda^n$, $J \subseteq [n]$ and $t \in \mathbb{R}$. We say that y_J is *t-heavy* (for x and z_I) if

$$\Pr\left[Y_J = y_J \mid Z_I^x = z_I\right] > 2^{(\sigma_Y + \frac{4}{c})\cdot\Delta\cdot|J| - b\cdot|J| + t - 1}$$

We say that a value y_J is heavy if it is 0-heavy. We say that a value x is *α-light* with respect to J if for every $I \subseteq [n] - J$ and $z_I \in \{0,1\}^I$ it holds that

$$\Pr\left[Y_J \text{ is heavy for } x \text{ and } z_I \mid Z_I^x = z_I\right] \le 2^{-\alpha\Delta} \cdot \left(\frac{1}{2n}\right)^{|J|}.$$

A value x is *α-light* if it is *α-light* with respect to all sets J.

We now state the relationship between light values and recoverable values that was mentioned earlier.

Proposition 3.8. *If $x \in \Lambda^n$ is α-light then it is α-recoverable.*

We postpone the proof of Proposition 3.8 to the end of this subsection. To prove Proposition 3.5, we consider values x that are not light with respect to some specific J. The following proposition bounds the probability of such values. We then complete the proof by taking union bound over all sets J.

Proposition 3.9. *Assume that $\sigma_X + 2\cdot\sigma_Y \le 1 - \frac{22}{c} - \gamma - \alpha$. For every $J \subseteq [n]$, the probability that X takes an almost uniform value x that is not α-light with respect to J is at most $2^{-\gamma\cdot\Delta\cdot|J|}$.*

The proof Proposition 3.9 is provided in Section 3.3. We now prove Proposition 3.5 restated below.

Proposition 3.5. *Assume that $\sigma_X + 2\cdot\sigma_Y \le 1 - \frac{24}{c} - \gamma - \alpha$. Then, the probability that X takes an almost uniform value x that is not α-recoverable is at most $2^{-\gamma\cdot\Delta}$.*

Proof. By applying Proposition 3.9 with $\gamma = \gamma + \frac{2}{c}$, we obtain that for every set $J \subseteq [n]$, the probability that X takes an almost uniform value x that is not α-light with respect to J is at most $2^{-\gamma\cdot\Delta} \cdot \frac{1}{(2n)^{|J|}}$. By the union bound and Proposition 2.1, we obtain that with probability at least $1 - 2^{-\gamma\cdot\Delta}$, the random variable X takes a value x that is α-light with respect to all $J \subseteq [n]$. Such a value x is α-recoverable by Proposition 3.8, so the required result follows. ■

Proof of Proposition 3.8 Let $\alpha \geq \frac{1}{\Delta}$ and let $x \in \Lambda^n$ be α-light. We show that x is α-recoverable by showing that for every $I \subseteq [n]$ and $z_I \in \Lambda^I$ there exists an event $\mathcal{E}$ such that the following random variable is $(\sigma_Y + \frac{4}{c})$-sparse:

$$Y_{[n]-I} \mid \mathcal{E} \text{ and } Z_I^x = z_I.$$

We choose $\mathcal{E}$ to be the event such that y_J are not heavy for some non-empty set $J \subseteq [n] - I$. We first prove that $\Pr\left[\neg\mathcal{E} \mid Z_{I,x_I} = z_I\right] < 2^{-\alpha\Delta}$. By the union bound, it holds that

$$\begin{aligned}
\Pr\left[\neg\mathcal{E} \mid Z_{I,x_I} = z_I\right] &= \Pr\left[\exists J \neq \emptyset \subseteq [n] - I : Y_J \text{ is heavy for } x \text{ and } z_I \mid Z_I^x = z_I\right] \\
&\leq \sum_{\emptyset \neq J \subseteq [n]-I} \Pr\left[Y_J \text{ is heavy for } x \text{ and } z_I \mid Z_I^x = z_I\right] \\
&\leq \sum_{\emptyset \neq J \subseteq [n]-I} 2^{-\alpha\Delta} \cdot \left(\frac{1}{2n}\right)^{|J|} && (x \text{ is } \alpha\text{-light}) \\
&\leq 2^{-\alpha\Delta} && (\text{Proposition 2.1})
\end{aligned}$$

It remain to prove that the random variable

$$Y_{[n]-I} \mid \mathcal{E} \text{ and } Z_I^x = z_I$$

is $\left(\sigma_Y + \frac{4}{c}\right)$-sparse. For every $J \subseteq [n] - I$, let y_J be some value of Y_J then it holds that

$$\begin{aligned}
\Pr\left[Y_J = y_J \mid \mathcal{E} \text{ and } Z_{I,x_I} = z_I\right] &= \frac{\Pr\left[Y_J = y_J \text{ and } Y \in \mathcal{E} \mid Z_I^x = z_I\right]}{\Pr\left[\mathcal{E} \mid Z_I^x = z_I\right]} \\
&\leq \frac{2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| - 1}}{\Pr\left[\mathcal{E} \mid Z_I^x = z_I\right]} && (y_J \text{ is not heavy}) \\
&\leq \frac{2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| - 1}}{1 - 2^{-\alpha\Delta}} \\
&\leq \frac{2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| - 1}}{1 - \frac{1}{2}} && (\text{since } \alpha \geq \frac{1}{\Delta}) \\
&\leq 2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J|},
\end{aligned}$$

and therefore the above random variable is $(\sigma_Y + \frac{4}{c})$-sparse, as required. ∎

3.3 Proof of Proposition 3.9

The proof of Proposition 3.9 consist of two parts. In the first part, we prove that for any y_J there are only few x such that y_J is heavy for that x. In the second part, we complete the proof using an averaging and bucketing argument. The first part is encapsulated in the following result.

Proposition 3.10. *Let $\gamma \geq \frac{2}{c}$ and assume that $\sigma_X + \sigma_Y \leq 1 - \frac{15}{c} - \gamma$. Then, for every $J \subseteq [n]$ and for every $y_J \in \Lambda^J$, the probability that X takes an almost uniform value x such that y_J is t-heavy for x is at most $2^{-\gamma \cdot \Delta \cdot |J| - 2t}$.*

The proof of Proposition 3.10 is given in Section 3.4, and the rest of this section is devoted to proving Proposition 3.9 using Proposition 3.10. An important point about the proposition is that we get stronger bounds on the probability for higher values of t, that is, for values y that violate the sparsity more strongly.

It is tempting to try and deduce Proposition 3.9 directly from Proposition 3.10 by an averaging argument. Such an argument would first say that, for all y, there is only a small probability that X takes a value for which y is heavy by Proposition 3.10. Thus, for most values of x, the probability of choosing y such that y is heavy for x should be small as well. There is, however, a significant obstacle here: in order to prove Proposition 3.9, we need to bound probability that Y_J is heavy for x with respect to a distribution that *depends on* x, namely, the distribution Y_J conditioned on the value of $g^I(x_I, Y_I)$. In contrast, the naive averaging argument assumes that X and Y are independent. This complication renders a naive averaging argument impossible.

In order to overcome this obstacle, we note that conditioning on $Z_I^x = z_I$ can increase the probabilities of getting some y_J. This increase in probability is characterize by the maximal t such that y_J is t-heavy for x. Thus, for each x we consider all y_J such that y_J is t-heavy for x, and place them into buckets according to the value of t. Then, we bound the weight of each bucket separately, while making use of the fact that Proposition 3.10 provides a stronger upper bound for larger values of t. Using this bucketing scheme turns out to be sufficient for the averaging argument to go through. We now prove Proposition 3.9 restated below.

Proposition 3.9. *Assume that* $\sigma_X + 2 \cdot \sigma_Y \leq 1 - \frac{22}{c} - \gamma - \alpha$. *For every* $J \subseteq [n]$*, the probability that* X *takes an almost uniform value* x *that is not* α*-light with respect to* J *is at most* $2^{-\gamma \cdot \Delta \cdot |J|}$.

Proof. Let $J \subseteq [n]$, and let $\mathcal{X}$ and $\mathcal{Y}_J$ denote the supports of X and Y_J respectively. For every $x \in \mathcal{X}$ and $y_J \in \mathcal{Y}_J$, let $t(x, y_J, I, z_I) \in \mathbb{R}$ be the number for which

$$\Pr\left[Y_J = y_J \mid Z_I^x = z_I\right] = 2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| + t_{x,y_J} - 1}.$$

We define t_{x,y_J} as the maximum of $t(x, y_J, I, z_I)$ over all possible I, z_I. Next, consider a table whose rows and columns are indexed by $\mathcal{X}$ and $\mathcal{Y}_J$ respectively. For every row $x \in \mathcal{X}$ and column $y_J \in \mathcal{Y}_J$, we set the corresponding entry to be

$$\mathrm{ent}(x, y_J) \stackrel{\text{def}}{=} \begin{cases} 2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| + t_{x,y_J} - 1} & t_{x,y_J} > 0 \text{ and } x \text{ is almost uniform} \\ 0 & \text{otherwise.} \end{cases}$$

Now we use a bucketing argument. We take each pair (x, y_J) and put it in the bucket that is labeled by $\lceil t_{x,y_J} \rceil$. Then, for each y_J, we upper bound the probability of (X, y_J) to be in each bucket. Let $\gamma' = \gamma + \sigma_Y + \alpha + \frac{7}{c}$. Note that when $\lceil t_{X,y_J} \rceil = t$, we can bound t_{X,y_J} from below by $t - 1$. Then, by applying Proposition 3.10 with $\gamma = \gamma'$, we get that for every y_J and every $t \in \mathbb{Z}_{>0}$ it holds that

$$\Pr\left[\lceil t_{X,y_J} \rceil = t \text{ and } X \text{ is almost uniform}\right] \leq 2^{-\gamma' \cdot \Delta \cdot |J| - 2(t-1)}. \tag{4}$$

Therefore, for every $y_J \in \mathcal{Y}_J$, the expected entry in the y_J-th column (over the random choice of X)

is

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{ent}(X, y_J)\right] = & \sum_{t=1}^{\infty} \left(\Pr\left[\lceil t_{X,y_J} \rceil = t \text{ and } X \text{ is almost uniform}\right] \cdot \right. \\
& \left. \mathbb{E}\left[\mathrm{ent}(X, y_J) \mid \lceil t_{X,y_J} \rceil = t \text{ and } X \text{ is almost uniform}\right]\right) \\
\leq & 2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| - 1} \cdot \sum_{t=1}^{\infty} \Pr\left[\lceil t_{X,y_J} \rceil = t \text{ and } X \text{ is almost uniform}\right] \cdot 2^t \\
\leq & 2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| - 1} \cdot \sum_{t=1}^{\infty} 2^{-\gamma' \cdot \Delta \cdot |J| - 2(t-1)} \cdot 2^t & \text{(Equation (4))} \\
= & 2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| - b \cdot |J| - 1} \cdot 2^{-\gamma' \cdot \Delta \cdot |J| + 2} \cdot \sum_{t=1}^{\infty} 2^{-t} \\
\leq & 2^{(\sigma_Y + \frac{4}{c} - \gamma' + \frac{1}{c}) \cdot \Delta \cdot |J| - b \cdot |J|} \\
= & 2^{(\sigma_Y + \frac{4}{c} - \gamma - \sigma_Y - \alpha - \frac{7}{c} + \frac{1}{c}) \cdot \Delta \cdot |J| - b \cdot |J|} & \text{(definition of } \gamma'\text{)} \\
\leq & 2^{-(\gamma + \alpha + \frac{2}{c}) \cdot \Delta \cdot |J| - b \cdot |J|}.
\end{aligned}
$$

It follows that the expected sum of a random row of the table (over the random choice of X) is

$$
\begin{aligned}
\mathbb{E}\left[\sum_{y_J \in \mathcal{Y}_J} \mathrm{ent}(X, y_J)\right] &= \sum_{y_J \in \mathcal{Y}_J} \mathbb{E}\left[\mathrm{ent}(X, y_J)\right] \\
&\leq \sum_{y_J \in \mathcal{Y}_J} 2^{-(\gamma + \alpha + \frac{2}{c}) \cdot \Delta \cdot |J| - b \cdot |J|} \\
&= 2^{-(\gamma + \alpha + \frac{2}{c}) \cdot \Delta \cdot |J|}.
\end{aligned}
$$

By Markov's inequality, the probability that the sum of the X-th row is more than $2^{-\left(\alpha + \frac{2}{c}\right) \cdot \Delta |J|}$ is upper bounded by $2^{-\gamma \cdot \Delta \cdot |J|}$. We now prove that if a value $x \in \mathcal{X}$ is almost uniform and the sum in the x-th row is at most $2^{-\left(\alpha + \frac{2}{c}\right) \cdot \Delta |J|}$, then x is α-light with respect to J, and this will finish the proof of the proposition.

Let $x \in \mathcal{X}$ be such a value. We prove that x is α-light with respect to J. Let $I \subseteq [n] - J$ and let $z_I \in \{0,1\}^I$. We would like to prove that

$$
\Pr\left[Y_J \text{ is heavy for } x \text{ and } z_I \mid Z_I^x = z_I\right] \leq 2^{-\alpha \Delta} \cdot \left(\frac{1}{2n}\right)^{|J|}.
$$

Observe that for every heavy y_J, it holds that $\Pr\left[Y_J = y_J \mid Z_I^x = z_I\right] \leq \mathrm{ent}(x, y_J)$. It follows that

$$
\begin{aligned}
& \Pr\left[Y_J \text{ is heavy for } x \text{ and } z_I \mid Z_I^x = z_I\right] \\
\leq & \sum_{y_J \text{ is heavy for } x \text{ and } z_I} \Pr\left[Y_J = y_J \mid Z_I^x = z_I\right] \\
\leq & \sum_{y_J} \mathrm{ent}(x, y_J) \\
\leq & 2^{-\left(\alpha + \frac{2}{c}\right) \cdot \Delta |J|} \\
\leq & 2^{-\alpha \Delta} \cdot \left(\frac{1}{2n}\right)^{|J|} & (|J| \geq 1, \Delta \geq c \log n)
\end{aligned}
$$

as required. ■

3.4 Proof of Proposition 3.10

In this section, we prove Proposition 3.10, restated next.

Proposition 3.10. *Let $\gamma \geq \frac{2}{c}$ and assume that $\sigma_X + \sigma_Y \leq 1 - \frac{15}{c} - \gamma$. Then, for every $J \subseteq [n]$ and for every $y_J \in \Lambda^J$, the probability that X takes an almost uniform value x such that y_J is t-heavy for x is at most $2^{-\gamma \cdot \Delta \cdot |J| - 2t}$.*

Proposition 3.10 is essentially a more refined version of the analysis in [CFK+19]. An important point about this proposition is that it gives stronger bounds for larger values of t. The proof of Proposition 3.10 follows closely the equivalent proof in [CFK+19] while taking the parameter t into consideration. The proof consists of three main steps: first, we use Bayes' formula to reduce the upper bound of Proposition 3.10 to upper bounding the probability of a related type of values, called *skewing* values; then, we use Vazirani's lemma to reduce the latter task to the task of the upper bounding the biases of Z_I^x. Finally, we upper bound the biases of Z_I^x using the low deficiency of X and Y and the discrepancy of g. We start by formally defining skewing values, and then establish their connection to heavy values.

Definition 3.11. Let $J \subseteq [n]\,, y_J \in \operatorname{supp}(Y_J)$, and $I \subseteq [n] - J$. Let $e(y_J)$ be the real number that satisfies.

$$\Pr\left[Y_J = y_J\right] = 2^{\sigma_Y \cdot \Delta \cdot |J| - b \cdot |J| - e(y_J)}$$

We note that $e\left(y_J\right)$ is non-negative as we assume that Y is σ_Y-sparse. We say that x is *t-skewing* for y_J (with respect to I) if

$$D_\infty\left(Z_I^x \mid Y_J = y_J\right) > 4 \cdot \log n \cdot |J| + e\left(y_J\right) + t - 2$$

Proposition 3.12. *Let $x \in \Lambda^n$, $J \subseteq [n]$, and $y_J \in \Lambda^J$. If y_J is t-heavy for x and is almost uniform then x is t-skewing for y_J (with respect to some set I).*

Proof. By the assumption that y_J is t-heavy for x, there exist $I \subseteq [n]$ and $z_I \in \Lambda^I$ such that

$$\Pr\left[Y_J = y_J \mid Z_I^x = z_I\right] > 2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| + t - b \cdot |J| - 1}. \tag{5}$$

We show that x is t-skewing for y_J with respect to I. It follows that

$$\begin{aligned}
\Pr\left[Z_I^x = z_I \mid Y_J = y_J\right] &= \frac{\Pr\left[Y_J = y_J \mid Z_I^x = z_I\right] \cdot \Pr\left[Z_I^x = z_I\right]}{\Pr\left[Y_J = y_J\right]} && \text{(Bayes' formula)} \\
&> \frac{2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| + t - b \cdot |J| - 1} \cdot \Pr\left[Z_I^x = z_I\right]}{\Pr\left[Y_J = y_J\right]} && \text{(Equation (5))} \\
&\geq \frac{2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| + t - b \cdot |J| - 1} \cdot 2^{-|I| - 1}}{\Pr\left[Y_J = y_J\right]} && \text{(x is almost uniform)} \\
&= \frac{2^{(\sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |J| + t - b \cdot |J| - 1} \cdot 2^{-|I| - 1}}{2^{\sigma_Y \cdot \Delta \cdot |J| - b \cdot |J| - e(y_J)}} && \text{(definition of $e(y_J)$)} \\
&= 2^{4 \cdot \log n \cdot |J| + e(y_J) + t - 2 - |I|} && (\Delta \geq c \cdot \log n).
\end{aligned}$$

The latter inequality implies that

$$D_\infty\left(Z_I^x \mid Y_J = y_J\right) > 4 \cdot \log n \cdot |J| + e\left(y_J\right) + t - 2.$$

as required. ∎

Next, we formally define biasing values and relate them to skewing values via the usage of Vazirani lemma, thus allowing us to focus on the biases. Informally, biasing values are values x such that when conditioning on $Y_J = y_J$, the bias of $g^{\oplus S}(x_S, Y_S) = \bigoplus_{i \in S} Z_i^x$ is too high for some $S \subseteq I$.

Definition 3.13. Let $J \subseteq [n]$ and let $y_J \in \Lambda^J$. We say that x is *t-biasing* for y_J if there exists a set $S \subseteq [n] - J$ such that $|S| \geq 4 \cdot |J| + \frac{t+e(y_J)-3}{\log n}$ and

$$\operatorname{bias}\left(g^{\oplus S}(x_S, Y_S) \mid Y_J = y_J\right) > (2n)^{-|S|}.$$

Proposition 3.14. *Let $x \in \Lambda^n$, let $J \subseteq [n]$, and let $y_J \in \Lambda^J$. If x is not t-biasing for y_J then x is not t-skewing for y_J.*

Proof. Let x, J, y_J be as in the proposition, and assume that x is not t-biasing for y_J. We prove that x is not t-skewing for y_J. To this end, we prove that for every $I \subseteq [n] - J$ it holds that

$$D_\infty\left(Z_I^x \mid Y_J = y_J\right) \leq 4 \cdot \log n \cdot |J| + e\left(y_J\right) + t - 2.$$

Let $I \subseteq [n] - J$. By assumption, x is not t-biasing for y_J. Therefore we can apply the second variant of Vazirani's lemma (Lemma 2.7) to the random variable $Z_{I,x_I} \mid Y_J = y_J$ and obtain that

$$\begin{aligned} D_\infty\left(Z_I^x \mid Y_J = y_J\right) &\leq \left(4 \cdot |J| + \frac{t+e(y_J)-3}{\log n}\right) \cdot \log |I| + 1 \\ &\leq \left(4 \cdot |J| + \frac{t+e(y_J)-3}{\log n}\right) \cdot \log n + 1 \\ &= 4 \cdot \log n \cdot |J| + t + e(y_J) - 2, \end{aligned}$$

as required. ∎

We finally prove Proposition 3.10, restated next.

Proposition 3.10. *Let $\gamma \geq \frac{2}{c}$ and assume that $\sigma_X + \sigma_Y \leq 1 - \frac{15}{c} - \gamma$. Then, for every $J \subseteq [n]$ and for every $y_J \in \Lambda^J$, the probability that X takes an almost uniform value x such that y_J is t-heavy for x is at most $2^{-\gamma \cdot \Delta \cdot |J| - 2t}$.*

Proof. Let $J \subseteq [n]$ and let $y_J \in \Lambda^J$. We first observe that it suffices to prove that with probability at least $1 - 2^{-\gamma \cdot \Delta \cdot |J| - 2t}$, the random variable X takes a value x that is not t-biasing for y_J. Indeed, if x is a value that is not t-biasing for y_J, then by Proposition 3.14 it is not t-skewing for y_J, and then by Proposition 3.12 it cannot be the case that y_J is t-heavy for x and that x is almost uniform. It remains to upper bound the probability that x is t-biasing for y_J.

We start by upper bounding the probability that X takes a value x such that

$$\operatorname{bias}\left(g^{\oplus S}(x_S, Y_S) \mid Y_J = y_J\right) > (2n)^{-|S|}$$

for some fixed non-empty set $S \subseteq [n] - J$ such that $|S| \geq 4 \cdot |J| + \frac{t+e(y_J)-3}{\log n}$ (the case where S is empty is trivial). Let S be such a set. In order to upper bound the latter probability, we use Lemma 2.11, which in turn requires us to upper bound the deficiencies $D_\infty(X_S)$ and $D_\infty(Y_S \mid Y_J = y_J)$. By assumption, we know that $D_\infty(X_S) \le \sigma_X \cdot \Delta \cdot |S|$. We turn to upper bound $D_\infty(Y_S \mid Y_J = y_J)$. For every $y_S \in \Lambda^S$, it holds that

$$\begin{aligned}
\Pr\left[Y_S = y_S \mid Y_J = y_J\right] &= \frac{\Pr\left[Y_{S\cup J} = y_{S\cup J}\right]}{\Pr\left[Y_J = y_J\right]} \\
&= \frac{\Pr\left[Y_{S\cup J} = y_{S\cup J}\right]}{2^{\sigma_Y\cdot\Delta\cdot|J| - b\cdot|J| - e(y_J)}} && \text{(Definition of } e(y_J)\text{)} \\
&\le \frac{2^{\sigma_Y\cdot\Delta\cdot(|S|+|J|) - b\cdot(|S|+|J|)}}{2^{\sigma_Y\cdot\Delta\cdot|J| - b\cdot|J| - e(y_J)}} && (Y \text{ is } \sigma_Y\text{-sparse}) \\
&= 2^{\sigma_Y\cdot\Delta\cdot|S| + e(y_J) - b\cdot|S|}
\end{aligned}$$

and thus

$$D_\infty(Y_S \mid Y_J = y_J) \le \sigma_Y \cdot \Delta \cdot |S| + e(y_J).$$

By our assumption on the size of S, it follows that

$$e(y_J) \le \log n \cdot |S| + 3 \le 4 \cdot \log n \cdot |S| \le \frac{4}{c} \cdot \Delta \cdot |S|$$

and therefore

$$\begin{aligned}
D(X_S) + D_\infty(Y_S \mid Y_J = y_J) &\le (\sigma_X + \sigma_Y + \frac{4}{c}) \cdot \Delta \cdot |S| \\
&\le (1 - \frac{11}{c} - \gamma) \cdot \Delta \cdot |S| && (\sigma_X + \sigma_Y \le 1 - \frac{15}{c} - \gamma) \\
&= \left(\Delta - \frac{7\Delta}{c} - \gamma\Delta - \frac{4\Delta}{c}\right) \cdot |S| \\
&\le (\Delta - 7 - \gamma\Delta - 2\log n - 2) \cdot |S| \,.
\end{aligned}$$

Now, by applying Lemma 2.11 with $\gamma = \gamma\Delta + \log n + 1$ and $\lambda = \log n + 1$, it follows that the probability that X takes a value x such that

$$\operatorname{bias}\left(g^{\oplus S}\left(x_S, Y_S\right) \mid Y_J = y_J\right) > (2n)^{-|S|}$$

is at most

$$2^{-\gamma\cdot\Delta\cdot|S|} \cdot \frac{1}{(2n)^{|S|}},$$

where the inequality holds since S is assumed to be non-empty. By taking union bound over all relevant sets S, it follows that the probability that X takes a value x that is t-biasing for y_J is at most

$$\begin{aligned}
&\sum_{S\subseteq[n]:|S|\ge 4\cdot|J| + \frac{t+e(y_J)-3}{\log n}} 2^{-\gamma\cdot\Delta\cdot|S|} \cdot \frac{1}{(2n)^{|S|}} \\
&\le \sum_{S\subseteq[n]:|S|\ge 4\cdot|J| + \frac{t-3}{\log n}} 2^{-\gamma\cdot\Delta\cdot|S|} \cdot \frac{1}{(2n)^{|S|}} && (e(y_J) \ge 0) \\
&\le 2 \cdot 2^{-\gamma\cdot\Delta\cdot\left(4\cdot|J| + \frac{t-3}{\log n}\right)} \cdot \frac{1}{2} && \text{(Proposition 2.1)} \\
&\le 2^{-\gamma\cdot\Delta\cdot\left(4|J| + \frac{t-3}{\log n}\right)} \\
&\le 2^{-\gamma\cdot\Delta|J| - \gamma\cdot\Delta\frac{t}{\log n}} && (|J| \ge 1, \log n \ge 1) \\
&\le 2^{-\gamma\cdot\Delta\cdot|J| - 2t} && (\frac{\gamma\cdot\Delta}{\log n} \ge \frac{2\Delta}{c\log n} \ge 2)
\end{aligned}$$

as required. ■

4 The deterministic lifting theorem

In this section we prove the deterministic part of our main theorem, restated below.

Theorem 4.1 (Deterministic lifting theorem)**.** *There exists a universal constant $c \in \mathbb{N}$ such that the following holds. Let $\mathcal{S}$ be a search problem that takes inputs from $\{0,1\}^n$ and let $g : \Lambda \times \Lambda \to \{0,1\}$ be an arbitrary function such that $\Delta(g) \geq c \cdot \log n$. Then there is a deterministic decision tree T that solves $\mathcal{S}$ with complexity $O\left(\frac{D^{cc}(\mathcal{S}\circ g^n)}{\Delta(g)}\right)$.*

In what follows, we fix Π to be an optimal deterministic protocol for $\mathcal{S} \circ g^n$ and let $D^{\text{cc}}(\mathcal{S} \circ g^n)$ be its complexity. We construct a decision tree T that solves $\mathcal{S}$ using $O(\frac{D^{\text{cc}}(\mathcal{S}\circ g^n)}{\Delta})$ queries. We construct the decision tree T in Section 4.1, prove its correctness in Section 4.2, and analyze its query complexity in Section 4.3.

High-level idea of the proof. Intuitively, given an input $z \in \{0,1\}^n$, the decision tree T attempts to construct a full transcript π of Π that is consistent with some input (x, y) such that $g^n(x, y) = z$. Such a transcript must output a solution in

$$(\mathcal{S} \circ g^n)(x, y) = \mathcal{S}\left(g^n(x, y)\right) = \mathcal{S}(z)$$

thus solving $\mathcal{S}$ on z. The tree works iteratively, constructing the transcript π message-by-message. Throughout this process, the tree maintains random variables X, Y that are distributed over the inputs of Alice and Bob such that the input (X, Y) is consistent with π. In particular, the tree preserves the invariant that the variables (X, Y) are $(\rho, \sigma_X, \sigma_Y)$-structured where ρ is a restriction that is consistent with z and for some values σ_X and σ_Y. The restriction ρ keep track which of the input bits of z have been queried, and is maintained accordingly by the tree. By the uniform marginals lemma (Lemma 2.17), we get that the protocol Π cannot distinguish between the distribution of (X, Y) and the same distribution conditioned on $g^n(X, Y) = z$. Hence, when the protocol ends, π must be consistent with an input in $\left(g^n\right)^{-1}(z)$.

We now describe how the tree preserves the foregoing invariant. Suppose without loss of generality that the invariant is violated because X_I is not σ_X-sparse (the case where Y is not σ_Y-sparse is treated similarly). Specifically, assume that the deficiency of X_I is too large for some set of coordinates $I \subseteq \text{free}(\rho)$. To restore the structure, the tree queries z_I and conditions X and Y on $g^I(X_I, Y_I) = z_I$. The conditioning on $g^I(X_I, Y_I) = z_I$, however, could increase the deficiency of X and Y, which might violate their structure. In order to avoid this, the tree conditions X in advance on taking a safe value. Moreover, after the conditioning on $g^I(X_I, Y_I) = z_I$ the tree recovers the density of Y by conditioning it on a high-probability event, as in the the definition of a recoverable value (Definition 3.2).

In order to upper bound the query complexity of the tree, we keep track of the deficiency $D_\infty\left(X_{\text{free}(\rho)}, Y_{\text{free}(\rho)}\right)$ throughout the simulation of the protocol. We prove that the messages that the tree sends in the simulation increase the deficiencies by at most $O(D^{\text{cc}}(\mathcal{S} \circ g^n))$ overall. On the other hand, whenever the tree queries a set of coordinates I, the deficiency decreases by at least $\Omega(|I| \cdot \Delta)$. As a result we get that the tree cannot make more then $O\left(\frac{D^{\text{cc}}(\mathcal{S}\circ g^n)}{\Delta}\right)$ queries in total.

4.1 The construction of the deterministic decision tree

In this section we describe the construction of the deterministic decision tree T. For the construction we set $\sigma \stackrel{\mathrm{def}}{=} \frac{1}{4}, \alpha \stackrel{\mathrm{def}}{=} \frac{1}{c}$ and $c = 200$.

Invariants. The tree maintains a partial transcript π, a restriction ρ, and two independent random variables X, Y over Λ^n such that the input (X, Y) is consistent with π. The tree works iteratively. In each iteration the tree simulates a single round of the protocol Π. The tree maintains the invariant that at the beginning of each iteration, if it is Alice's turn to speak in the simulated protocol, then X and Y are $(\rho, \sigma + \frac{4}{c}, \sigma)$-structured. If it is Bob's turn to speak, it is the other way around.

The algorithm of the tree. When T starts the simulation, the tree sets the transcript π to be empty , the restriction ρ to $\{*\}^n$, and X, Y to uniform random variables over Λ^n. It is easy to verify that the invariant holds at the beginning of the simulation. We now describe a single iteration of the tree. Without lose of generality, we assume that this is Alice's turn. The tree performs the following steps:

1. The tree conditions $X_{\mathrm{free}(\rho)}$ on taking an α-safe value for $Y_{\mathrm{free}(\rho)}$.
2. Let $M(x, \pi)$ be the message that Alice sends on partial transcript π and input x. The tree chooses a message m such that $P[M(X, \pi) = m] \geq 2^{-|m|}$, adds it to π, and conditions X on the event that $M(X, \pi) = m$. We note that such a message m must exist (see explanation below).
3. Let $I \subseteq \mathrm{free}(\rho)$ be a maximal set that violates the σ-sparsity of $X_{\mathrm{free}(\rho)}$, and let x_I be a value such that $\Pr[X_I = x_I] > 2^{\sigma \cdot \Delta \cdot |I| - b \cdot |I|}$. The tree conditions X on $X_I = x_I$.
4. The tree query z_I and sets ρ_I to z_I.
5. The tree conditions Y on $g^I(x_I, Y_I) = z_I$.
6. The tree conditions Y on an event $\mathcal{E}$ such that $Y \mid \mathcal{E}$ is $\left(\sigma + \frac{4}{c}\right)$-sparse and $\Pr[\mathcal{E}] \geq \frac{1}{2}$. Such an event must exist since the value x is safe, and in particular, recoverable

When the protocol Π halts, the tree T halts as well and returns the output of the transcript π. A few additional comments are in order:

- We need to explain why we never condition X nor Y on an event with zero probability. Regarding Step 1, we need to prove that there exist α-safe values. To do so we apply Lemma 3.3 with $\gamma = \frac{1}{c}$ (and α as defined above). As required by the lemma, it holds that

$$\sigma_X + 2\sigma_Y \leq 3\sigma + \frac{4}{c} = 0.77 \leq 0.865 = 1 - \frac{27}{c} = 1 - \frac{25}{c} - \gamma - \alpha.$$

  Therefore we know that the conditioning is on an event with probability of at least $1 - 2^{-\frac{\Delta}{c}} > 0$. Regarding the conditioning at Step 5, we note that the value x_I is safe. Hence, x_I is almost uniform and the event $g^I(x_I, Y_I) = z_I$ is not empty.

- In Step 2, we claimed that there must be a message m such that $\Pr[M(X, \pi) = m] \geq 2^{-|m|}$. To this end, we note that the set of possible messages of Alice must be a prefix-free code. Otherwise, there would be two messages m_1, m_2 such that m_1 is prefix of m_2. In case that Bob got the message m_1 from Alice, Bob will not know whether Alice sent m_2 but the remaining bits have not arrived yet, or Alice sent the message m_1. We complete the proof of the claim by using Fact 2.9 that ensures that such a message exists.

- We need to prove that at the end of Step 6, the variable X is σ-sparse and Y is $\left(\sigma+\frac{4}{c}\right)$-sparse as required by the invariants. It holds that X is σ-sparse by Proposition 2.18, and Y is $\left(\sigma+\frac{4}{c}\right)$-sparse by the definition of the event $\mathcal{E}$ at end of Step 6.

4.2 Correctness of the deterministic decision tree

We now prove that the tree always returns a correct answer, that is, given an input z the tree outputs an answer o such that $(z,o)\in\mathcal{S}$. Let X,Y and π be the distributions and transcript at the end of the simulation. Let o be the output associated with the transcript π. As asserted in Section 4.1, every $(x,y)\in\operatorname{supp}(X,Y)$ is consistent with the transcript π. By the fact that Π solves $\mathcal{S}\circ g^n$, for every pair $(x,y)\in\operatorname{supp}(X,Y)$ it holds that

$$\left(g^n(x,y),o\right)\in\mathcal{S}.$$

To complete the proof of correctness, we show that there exists a pair $(x,y)\in\operatorname{supp}(X,Y)$ such that $g^n(x,y)=z$ and therefore $(z,o)\in\mathcal{S}$. As X,Y are (ρ,σ_X,σ_Y)-structured, it is guaranteed that

$$g^{\mathrm{fix}(\rho)}\left(X_{\mathrm{fix}(\rho)},Y_{\mathrm{fix}(\rho)}\right)=z_{\mathrm{fix}(\rho)}$$

with probability 1. Regarding the free part, we apply Proposition 2.16 with $\gamma=\frac{1}{c}$. (so the requirement $2\sigma+\frac{4}{c}\leq 1-\frac{8}{c}-\gamma$ holds) and we get that

$$\Pr\left[g^{\mathrm{free}(\rho)}\left(X_{\mathrm{free}(\rho)},Y_{\mathrm{free}(\rho)}\right)=z_{\mathrm{free}(\rho)}\right]>0.$$

Therefore there must be some pair $(x,y)\in\operatorname{supp}(X,Y)$ such that $g^{\mathrm{free}(\rho)}\left(x_{\mathrm{free}(\rho)},y_{\mathrm{free}(\rho)}\right)=z_{\mathrm{free}(\rho)}$. Both facts together ensure us that there exist a pair $(x,y)\in\operatorname{supp}(X,Y)$ such that $g^n(x,y)=z$, as required.

4.3 Query complexity of the deterministic decision tree

In this section, we upper bound the query complexity of the tree by $O\left(\frac{D^{cc}(\mathcal{S}\circ g^n)}{\Delta}\right)$. The upper bound is proven using a potential argument. We define our potential function to be the deficiency of X,Y, i.e.,

$$D_\infty\left(X_{\mathrm{free}(\rho)},Y_{\mathrm{free}(\rho)}\right)=D_\infty\left(X_{\mathrm{free}(\rho)}\right)+D_\infty\left(Y_{\mathrm{free}(\rho)}\right).$$

We prove that whenever the simulated protocol sends a message m, the deficiency increases by at most $O(|m|)$, and that whenever the tree makes a query, the deficiency decreased by at least $\Omega(\Delta)$. The deficiency is equal to zero at the beginning of the simulation and it is always non-negative by Fact 2.3. The length of the transcript is bounded by $D^{cc}(\mathcal{S}\circ g^n)$, and therefore we get a bound of $O\left(\frac{D^{cc}(\mathcal{S}\circ g^n)}{\Delta}\right)$ queries.

We now analyze in detail the changes in the deficiency during a single iteration of the tree step-by-step:

1. In Step 1 the tree conditions on the event that $X_{\mathrm{free}(\rho)}$ is safe. By applying Lemma 3.3 with $\gamma=\frac{1}{c}$, we obtain that $\Pr\left[X_{\mathrm{free}(\rho)}\text{ is not safe}\right]\leq 2^{-\frac{\Delta}{c}}\leq\frac{1}{2}$. Therefore, it follows that $D_\infty\left(X_{\mathrm{free}(\rho)}\right)$ increases at most by 1 by Fact 2.4.

2. In Step 2, the deficiency $D_\infty\left(X_{\mathrm{free}(\rho)}\right)$ increases by at most $\log\frac{1}{\Pr[m]}$ by Fact 2.4. As m is chosen such that $\Pr[m]\geq 2^{-|m|}$, it holds that $D_\infty\left(X_{\mathrm{free}(\rho)}\right)$ increases by at most $|m|$

3. In Step 3, the deficiency $D_\infty\left(X_{\mathrm{free}(\rho)}\right)$ increases by at most $b\cdot|I|-\sigma\cdot\Delta\cdot|I|$ by Fact 2.4.

4. In Step 4, we reduce $\mathrm{free}(\rho)$ by $|I|$. The deficiency $D_\infty\left(X_{\mathrm{free}(\rho)}\right)$ is decreased by $b\cdot|I|$ since X_I is a constant, while $D_\infty\left(Y_{\mathrm{free}(\rho)}\right)$ does not increase by Fact 2.5.

5. In Step 5, we get that $D_\infty\left(Y_{\mathrm{free}(\rho)}\right)$ increases by at most $\log\frac{1}{\Pr[g(x_I,Y_I)=z_I]}$ by Fact 2.4. Since x_I is almost uniform, we know that $\Pr\left[g\left(x_I,Y_I\right)=z_I\right]\geq 2^{-|I|-1}$, and therefore $D_\infty\left(Y_{\mathrm{free}(\rho)}\right)$ increases by at most $|I|+1$.

6. In Step 6 the tree conditions Y on the event $\mathcal{E}$ such that $\Pr[Y\in\mathcal{E}]\geq\frac{1}{2}$. By Fact 2.4 we get that $D_\infty(Y_{\mathrm{free}})$ increases by at most 1.

At the end, we have that any message m increase the deficiency by at most $|m|+1\in O(|m|)$, and for any set of queries I the deficiency decreases by at least

$$\begin{aligned}
b\cdot|I|-(b\cdot|I|-\sigma\cdot\Delta\cdot|I|)-|I|-2 &\geq \left(\sigma-\frac{3}{\Delta}\right)\cdot\Delta\cdot|I| && (\text{rearranging,}|I|\geq 1)\\
&\geq \left(\sigma-\frac{3}{c}\right)\cdot\Delta\cdot|I| && (\Delta>c\log n)\\
&\in\Omega\left(\Delta\cdot|I|\right) && (\sigma-\frac{3}{c}=0.235\text{ is a positive constant}).
\end{aligned}$$

5 The randomized lifting theorem

In this section we prove the randomized lifting theorem. We start by stating the following simulation result, which implies the lifting theorem as a simple consequence.

Notation 5.1. Let Π be some protocol that takes inputs in $\Lambda^n\times\Lambda^n$. For every $z\in\{0,1\}^n$ we let Π'_z denote the distribution of transcripts of the protocol Π on uniformly random inputs X,Y conditioned on the event $g^n(X,Y)=z$.

Theorem 5.2. *Let $g:\Lambda\times\Lambda\to\{0,1\}$ be a function such that $\Delta(g)\geq 1000\log n$. Let Π be a public-coin randomized protocol that takes inputs from $\Lambda^n\times\Lambda^n$ and uses at most C bits of communication. Then, there is a randomized decision tree given an input $z\in\{0,1\}^n$, samples from a distribution that is $\left(2^{-\frac{\Delta(g)}{20}}\cdot(1+C)\right)$-close to the distribution Π'_z and makes at most $80\left(\frac{C}{\Delta(g)}+1\right)$ queries.*

Before we prove Theorem 5.2, we show how Theorem 5.2 implies Theorem 1.3, restated next.

Theorem 1.3 (Randomized part)**.** *There exists a universal constant c such that the following holds: Let $\mathcal{S}$ be a search problem that takes inputs from $\{0,1\}^n$, let $g:\Lambda\times\Lambda\to\{0,1\}$ be an arbitrary function such that $\Delta(g)\geq c\cdot\log n$, and let $\beta>0$. Then, it holds that*

$$R^{\mathrm{cc}}_{\beta}(\mathcal{S}\circ g^n)\in\Omega\left(\left(R^{\mathrm{dt}}_{\beta'}(\mathcal{S})-80\right)\cdot\Delta(g)\right),$$

where $\beta'=\beta+2^{-\Delta(g)/50}$.

Proof. We choose $c = 1000$. Let Π be an optimal protocol that solves $S \circ g^n$ with error probability β, and denote by C the complexity of Π. In the case that $C \geq n \cdot \Delta(g)$, the lower bound holds trivially, and we therefore assume that $C < n \cdot \Delta(g)$. Let T be the tree obtained by applying Theorem 5.2 to Π. We construct a decision tree T' for $\mathcal{S}$ as follows: on input z, the tree T' simulates the tree T on z, thus obtaining a transcript π of Π, and returns the output associated with this transcript. For a transcript π of Π, we denote by $\mathcal{O}(\pi)$ the output associated with this transcript. By assumption, for every inputs $(x, y) \in \Lambda^n \times \Lambda^n$ such that $g^n(x, y) = z$, it holds that the output of Π on x, y is in $\mathcal{S}(z)$ with probability $1 - \beta$ as $\mathcal{S}(z) = (\mathcal{S} \circ g^n)(x, y)$. The error probability of T' on z is

$$\begin{aligned}
\Pr_{o \leftarrow T'(z)}[(z, o) \notin \mathcal{S}] &= \Pr_{\pi \leftarrow T}[(z, \mathcal{O}(\pi)) \notin \mathcal{S}] \\
&\leq \Pr_{\pi \leftarrow \Pi'_z}[(z, \mathcal{O}(\pi)) \notin \mathcal{S}] + 2^{-\frac{\Delta(g)}{20}} \cdot (1 + C) && \text{(by Theorem 5.2)} \\
&\leq \beta + 2^{-\frac{\Delta(g)}{20}} \cdot (1 + C) \\
&\leq \beta + 2^{-\frac{\Delta(g)}{20}} \cdot n \cdot \Delta(g) \\
&\leq \beta + 2^{-\frac{\Delta(g)}{20}} \cdot 2^{\frac{\Delta(g)}{c}} \cdot \Delta(g) && (\Delta(g) \geq c \cdot \log n) \\
&\leq \beta + 2^{-\frac{\Delta(g)}{50}}. && (\text{as } \Delta(g) \geq c = 1000)
\end{aligned}$$

where the second to last transition hold for sufficiently large (but constant) c. Therefore, the tree T' solves $\mathcal{S}$ with error probability at most β'. Note that the query complexity of T' is the same as the query complexity of T. Therefore, by Theorem 5.2, the query complexity of T' is at most $80\left(\frac{R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n)}{\Delta(g)} + 1\right)$. Together with the fact that T' solves $\mathcal{S}$ we get that

$$R^{dt}_{\beta'}(\mathcal{S}) \leq 80\left(\frac{R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n)}{\Delta(g)} + 1\right)$$

and therefore

$$\begin{aligned}
R^{dt}_{\beta'}(\mathcal{S}) - 80 &\leq 80\left(\frac{R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n)}{\Delta(g)}\right) \\
\left(R^{dt}_{\beta'}(\mathcal{S}) - 80\right) \cdot \Delta(g) &\leq 80\left(R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n)\right) \\
R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n) &\in \Omega\left(\left(R^{dt}_{\beta'}(\mathcal{S}) - 80\right) \cdot \Delta(g)\right).
\end{aligned}$$

■

In the rest of this section we prove Theorem 5.2. The proof is organized as follow: We first introduce the algorithm of tree T in Section 5.1 and prove some basic facts about the tree. Then, in Section 5.2 we prove the correctness of the tree, that is, we prove that the distribution of the outputs of the tree given z is indeed close to Π'_z. We end the section with proving the bound on the query complexity in Section 5.3.

5.1 Construction of the decision tree

The construction of the randomized decision tree is similar to the deterministic tree but has several important differences. The key difference is in the goal of the tree: in the randomized case, we should sample a transcript from Π'_z, whereas in the deterministic case we only need to find some transcript in the support of Π'_z.

As in the deterministic case, the tree maintains two random inputs X and Y. As a result of the key difference described above, the tree cannot condition on low-probability events, as this can drastically change the distribution of the returned transcript. This constraint results in number of changes to the way the tree samples messages and restores density. We now describe these changes:

- When the deterministic tree chooses a message for Alice in the simulation, it is sufficient for the tree to choose some high probability message m and to condition X on sending it. In contrast, the randomized tree needs to sample a message that is close to the distribution of the next message in Π'_z. In order to do so in the case that Alice speaks, the randomized decision tree samples the message m by first sampling x from X and then simulating the protocol on x to obtain m. The case where it is Bob's turn to speak is handled similarly. In Section 5.2 we prove that this distribution of the message is sufficiently close to its distribution in Π'_z using the uniform marginals lemma.

- The change to the selection of messages in the previous item creates a new problem. Denote by M be the random message that sampled by the simulation. In the deterministic case, when the tree conditions X on $M = m$, the deficiency of X grows by at most $|m|$, as the choice of m guarantees that $\Pr[M = m] \geq 2^{-|m|}$. In the randomized case, on the other hand, this does not hold anymore, since the chosen message m may have an arbitrarily low probability. In order to resolve this issue, we maintain a counter K_{msg} that keep track of the sum $\sum_{m\in\pi} \log \frac{1}{\Pr[M=m]}$, that is, the total increase in deficiency caused by sending messages. The tree halts if the counter K_{msg} surpasses $C + \Delta$ at any point. This way we ensure that sending messages contributes at most $C + \Delta$ to the deficiency. In Section 5.2, we prove that the tree halts in this way only with small probability, and therefore this modification does not contribute much to the error probability.

- In the deterministic case, the tree restores the density of the variable X by conditioning it on an event of the form $X_I = x_I$ for some value x_I. Unfortunately, the event $X_I = x_I$ might have too low probability. Instead, we use the density-restoring partition (Lemma 2.19). Specifically the tree sample some class $\mathcal{X}_j$ of the partition and conditions on $X \in \mathcal{X}_j$.

- The change to the density-restoration procedure create a new problem. Let J be the a random variable of the partition class, and let j be the index of the class partition that was chosen. Using the density-restoring partition increases the deficiency by an additional term of $\log \frac{1}{\Pr[J\geq j]}$. As with the messages, we create a counter K_{prt} that keeps track of the sum $\sum \log \frac{1}{\Pr[J\geq j]}$ and halt if $K_{\text{prt}} > 5C + 2\Delta$. We show that $K_{\text{prt}} > 5C + 2\Delta$ occurs with only with negligible probability.

The changes described above follow the previous works [GPW17, CFK+19]. In particular, our construction closely follows the construction in [CFK+19]. As in the deterministic case, the main difference is the addition of density-restoring step for Y at the end of the iteration. This, in turn, requires a non-trivial addition to the proof of correctness (see Section 5.2.1). We turn to formally describe the decision tree.

Parameters. Let set $c = 1000, \sigma \stackrel{\text{def}}{=} \frac{1}{10} + \frac{2}{c}$ and $\alpha \stackrel{\text{def}}{=} \frac{1}{10}$. Let $\Delta \stackrel{\text{def}}{=} \log \frac{1}{\text{disc}(g)}$ be such that $\Delta > c \log n$. Denote by C the worst-case complexity of the protocol Π.

Assertions and Invariants. Throughout the simulation the tree maintains two independent random variables $X, Y \in \Lambda^n$ that are uniformly distributed over some subsets $\mathcal{X}, \mathcal{Y} \subseteq \Lambda^n$ respectively. The tree also maintains a restriction ρ. It holds that the set $\text{fix}(\rho)$ is the set of all queried bits, and $\rho_{\text{fix}(\rho)} = z_{\text{fix}(\rho)}$. The tree simulates Π iteratively, where at each iteration the tree simulates single round. At the beginning of each iteration, if it is Alice's (respectively Bob's) turn to speak then X, Y are $\left(\rho, \sigma + \frac{4}{c}, \sigma\right)$-structured (respectively $\left(\rho, \sigma, \sigma + \frac{4}{c}\right)$-structured). Throughout the simulation, the tree maintains some partial transcript π. At any point in the simulation, it holds that all pairs of inputs (x, y) in $\mathcal{X} \times \mathcal{Y}$ are consistent with the partial transcript π.

The Algorithm Of The Tree. The tree starts by sampling public coins for the randomized protocol Π and fixing them. From this point on, the protocol can be thought of as a deterministic protocol. The tree then sets initial values to its variables. The distributions X, Y are initialized to be uniform over Λ^n. The transcript π is initialized to the empty transcript, the restriction ρ to $\{*\}^n$, and the counters $K_{\text{msg}}, K_{\text{prt}}$ to 0. We now describe a single iteration of the tree. Without lose of generality we assume here that it is Alice's turn to speak in π. An iteration where it is Bob's turn to speak in π is the same with the exception of swapping the roles X and Y.

1. The tree conditions $X_{\text{free}(\rho)}$ on taking a value x that is α-safe for $Y_{\text{free}(\rho)}$.

2. Let the random variable M be the message that Alice sends on input X and the current partial transcript π. The tree samples a message m according to M and appends it to π. Then, the tree adds $\log \frac{1}{\Pr[M=m]}$ to K_{msg} and conditions X on $M = m$,.

3. Let $\mathcal{X}_{\text{free}(\rho)} = \mathcal{X}^1 \cup \cdots \cup \mathcal{X}^l$ be the density-restoring partition we get from Lemma 2.19. The tree chooses a random class $\mathcal{X}^j$ such that it choose the i-th class with probability $\Pr\left[X_{\text{free}(\rho)} \in \mathcal{X}^i\right]$. The tree then conditions X on $X_{\text{free}(\rho)} \in \mathcal{X}^j$. Let I_j and x_j be the set of coordinates and value associated with $\mathcal{X}^j$ as defined by Lemma 2.19.

4. Recall that
$$p_{\geq i} \stackrel{\text{def}}{=} \Pr\left[X_{\text{free}(\rho)} \in \mathcal{X}^i \cup \cdots \cup \mathcal{X}^l\right],$$
where the random variable X here refers to the random variable as in the start of Step 3. The tree adds $\log \frac{1}{p_{\geq j}}$ to K_{prt}.

5. If $K_{\text{prt}} > 5C + 2\Delta$ or $K_{\text{msg}} > C + \Delta$ then the tree halts.

6. The tree queries the coordinates in I_j and sets $\rho_{I_j} = z_{I_j}$.

7. The tree conditions Y on $g^{I_j}\left(x_{I_j}, Y_{I_J}\right) = \rho_{I_j}$.

8. The tree conditions Y on an event $\mathcal{E}$ such that $\Pr[\mathcal{E}] > 1 - 2^{-\alpha\Delta}$ and $Y_{\text{free}(\rho)} \mid \mathcal{E}$ is $\left(\sigma + \frac{4}{c}\right)$-sparse. Such an event must exist as x is α-safe, and in particular, α-recoverable. For the sake of the analysis, we assume that there is a canonical choice of such the event $\mathcal{E}$.

When the simulation reaches the end of the protocol, the decision tree halts and outputs the transcript π. In order for the algorithm to be well-defined, it remains to explain two points.

- We explain why all the conditionings done in the tree are on non-empty events. Regarding Step 1, the probability that $X_{\text{free}(\rho)}$ is safe is lower bounded by the main lemma. We know that $X_{\text{free}(\rho)}, Y_{\text{free}(\rho)}$ are $\left(\sigma + \frac{4}{c}\right)$-sparse and σ-sparse respectively. We now apply the main lemma with $\gamma = \frac{1}{10}$, which is possible since

$$\sigma_X + 2\sigma_Y = \frac{3}{10} + \frac{10}{1000} = 0.310 \leq 0.675 = \frac{7}{10} - \frac{25}{1000} = \frac{9}{10} - \frac{25}{c} - \gamma - \alpha.$$

Thus the probability is lower bounded by $1 - 2^{-\frac{1}{10}\Delta} > 0$ and we can conclude that the event is not empty. In Step 7, the tree conditions on $g^{I_j}\left(x_{I_j}, Y_{I_J}\right) = \rho_{I_j}$, and it holds that $\Pr\left[g^{I_j}\left(x_{I_j}, Y_{I_J}\right) = \rho_{I_j}\right] > 2^{-|I|-1}$ as x is almost uniform.

- Earlier we asserted that if at the beginning of the iteration it is Alice's turn to speak then $X_{\text{free}(\rho)}, Y_{\text{free}(\rho)}$ are $\left(\rho, \sigma + \frac{4}{c}, \sigma\right)$-structured, and they are $\left(\rho, \sigma, \sigma + \frac{4}{c}\right)$-structured when it is Bob's turn to speak. Assume that in the start of the iteration it is the case that $X_{\text{free}(\rho)}, Y_{\text{free}(\rho)}$ are $\left(\rho, \sigma + \frac{4}{c}, \sigma\right)$-structured. After Step 3 it is guaranteed that X is σ-sparse, and Y is $\left(\sigma + \frac{4}{c}\right)$-sparse by Step 8. Therefore X, Y swap rules and the assertion holds.

5.2 Correctness of the decision tree

In this section we prove the correctness of the decision tree, that is, that on every input z the output distribution of the tree given z is $\left(2^{-\frac{\Delta(g)}{20}} \cdot (1 + C)\right)$-close to Π'_z (recall that Π'_z denotes the distribution of transcripts of the protocol Π on uniformly random inputs conditioned on the event $g^n(X, Y) = z$). For the rest of the proof we fix z to be some input. For simplicity, we will assume that the protocol always runs for exactly C rounds. In case that the protocol finishes earlier, we consider all the remaining rounds as containing empty messages.

We now set up some additional notation. Let $\Pi^{\dagger}$ be a random transcript constructed by T when invoked on the input z. In the case that T halts early we let $\Pi^{\dagger}$ to be some special symbol $\perp$. Instead of bounding the distance between $\Pi^{\dagger}$ and Π'_z directly, it is easier to bound the distance of each of them to an intermediate transcript. In order to define the intermediate transcript, we introduce an intermediate decision tree T^*, which is the same as the tree T except that Step 5 is removed. We now define the intermediate transcript Π^* to be the transcript generated by T^* when invoked on the input z. The correctness of T now follows immediately from the next two claims, that are proved in Sections 5.2.1 and 5.2.2 respectively.

Claim 5.3. Π'_z *is* $C \cdot 2^{-\frac{\Delta}{20}}$*-close to* Π^*.

Claim 5.4. Π^* *is* $2 \cdot 2^{-\Delta}$*-close to* $\Pi^{\dagger}$.

5.2.1 Proof Of Claim 5.3

In order to prove Claim 5.3, we define a notion called the *extended protocol*, which is an augmented version of Π that imitates the action of the tree T^*. The extended protocol works in iterations, such that each iteration corresponds to a single round of the original protocol, and is analogous to a single iteration of the tree T^*.

In the following description of the protocol, we denote by $\mathcal{X}$ and $\mathcal{Y}$ the set of all inputs of Alice and Bob that are compatible with the current (partial) transcript of the extended protocol. Additionally, let X and Y be uniform variables over $\mathcal{X}$ and $\mathcal{Y}$ respectively. Throughout the protocol, Alice and Bob maintain a restriction ρ, which is initialized with $\{*\}^n$ and is kept consistent with $g^n(x, y)$. Alice and Bob also maintain shared partial transcript π of the original protocol.

We now describe a single iteration of the extended protocol. Without loss of generality we assume that it is Alice's turn to speak in this round. In the case where it is Bob's turn to speak, the roles of Alice (respectively X) and Bob (respectively Y) are swapped. In a single iteration, the extended protocol performs the following steps:

1. Alice sends 0 if $x_{\text{free}(\rho)}$ is α-safe for $Y_{\text{free}(\rho)}$, and 1 otherwise.

2. Alice sends the message m that Alice would send in protocol Π on the input x in this round. The message m is appended to π by both Alice and Bob.

3. Let $\mathcal{X}^j$ be the density-restoring partition of $\mathcal{X}_{\text{free}(\rho)}$. Alice sends the index j such that $x_{\text{free}(\rho)} \in \mathcal{X}^j$.

4. Bob sends $g^{I_j}\left(x_{I_j}, y_{I_j}\right)$. Both Alice and Bob update $\rho_{I_j} = g^{I_j}\left(x_{I_j}, y_{I_j}\right)$.

5. Bob sends 0 if $y \in \mathcal{E}$, where $\mathcal{E}$ is the event from Step 8 of the tree, and 1 otherwise.

The tree T^* can be naturally modified to simulate not only the protocol but also the extended protocol. We call this modified tree the *extended tree*. More formally, we change T^* such that it maintains extended transcript π^e as follows:

- in Step 1 the tree appends 0 to π^e.
- in Step 2 the tree appends the same message m to π^e as it appends to π.
- in Step 3 the tree appends the index j of the partition class to π^e.
- in Step 7 the tree appends the queried bits to π^e.
- in Step 8 the tree appends 0 to π^e.

After those changes, the sets $\mathcal{X} = \text{supp}(X)$ and $\mathcal{Y} = \text{supp}(Y)$ that are maintained by the tree are equal to the set of all inputs that are consistent with the partial transcript π^e.

Let E^* be a random extended transcript constructed by T^* when invoked on the input z. Let E' be a transcript of the extended protocol on random inputs X', Y' uniformly chosen from $(g^n)^{-1}(z)$. We prove Claim 5.3 by bounding the statistical distance between the transcripts E^* and E'. Then, as the transcripts of random transcript of Π^* and Π'_z can be extracted from E^* and E', the statistical distance $|\Pi^* - \Pi'_z|$ is bounded by $|E^* - E'|$.

We bound the distance $|E^* - E'|$ by constructing a coupling for E^* and E'. Let $E^*_{\leq i}$ be the prefix of E^* that corresponds for the first i iterations (the same goes for $E'_{\leq i}$). The coupling is constructed round-by-round, that is, we iteratively construct couplings of $E'_{\leq i}$ and $E^*_{\leq i}$ from a coupling of $E'_{\leq i-1}$ and $E^*_{\leq i-1}$. We formally state this in the following claim.

Claim 5.5. *For every i exists a coupling of $E'_{\leq i}$ and $E^*_{\leq i}$ such that*

$$\Pr\left[E'_{\leq i} \neq E^*_{\leq i}\right] \leq 2^{-\frac{\Delta}{20}} \cdot i.$$

Proof. We prove the claim by induction on i. In the base case of $i = 0$ we set both E' and E^* to the empty transcript , and then $\Pr\left[E'_0 \neq E^*_0\right] = 0$ as desired. Assume by induction that there exists a coupling of $E'_{\leq i-1}$ and $E^*_{\leq i-1}$. We describe an algorithm that samples a coupling of $E'_{\leq i}$ and $E^*_{\leq i}$ using the coupling of $E'_{\leq i-1}$ and $E^*_{\leq i-1}$. The algorithm maintains two partial transcripts e^* and e', which constitute the parts of $E'_{\leq i}$ and $E^*_{\leq i}$ that were constructed until the current step. We will sometimes say that the algorithm fails, in which case the output for $E^*_{\leq i}$ is sampled by sampling from the distribution $\left(E^*_{\leq i} \mid e^* \text{ is a prefix of } E^*_{\leq i}\right)$ and the output for $E'_{\leq i}$ is sampled similarly and independently.

We assume without loss of generality that it is Alice's turn to speak in the original protocol. The algorithm first samples transcripts e' and e^* from the coupling of of $E'_{\leq i-1}$ and $E^*_{\leq i-1}$ respectively. If $e' \neq e^*$ the algorithm fails immediately. Let X be the uniform distribution over the set $\mathcal{X}$ of Alice's inputs that are consistent with the partial transcript e^*. The random variable Y is defined analogously for $\mathcal{Y}$ and Bob. Let X' and Y' be random variables that are jointly distributed like X and Y conditioned on $g^n(X,Y) = z$ (respectively). The algorithm now follows the steps of the extended tree and extended protocol and constructs e' and e^* by appending messages corresponding to those steps.

- The algorithm appends 0 to e^* since T^* always appends 0 to its transcript. The algorithm also appends 0 to the transcript e' with probability $\Pr\left[X_{\mathrm{free}(\rho)} \text{ is } \alpha\text{-safe} \mid g^n(X,Y) = z\right]$, and otherwise it appends 1 to e' and fails. To bound the failure probability, we use the uniform marginals lemma (Lemma 2.17) with $\gamma = \frac{1}{10}$ and get that

  $$\Pr\left[X_{\mathrm{free}(\rho)} \text{ is not } \alpha\text{-safe} \mid g^n(X,Y) = z\right] \leq \Pr\left[X_{\mathrm{free}(\rho)} \text{ is not } \alpha\text{-safe}\right] + 2^{-\frac{\Delta}{10}}.$$

  Then, by applying the main lemma (Lemma 3.3) with $\gamma = \frac{1}{10}$, we get that $\Pr\left[X_{\mathrm{free}(\rho)} \text{ is not } \alpha\text{-safe}\right]$ is at most $2^{-\frac{\Delta}{10}}$. Note that we can apply the main lemma and the uniform marginals lemma as X and Y are $(\sigma + \frac{4}{c})$-sparse and σ-sparse, respectively. Thus, the coupling fails at this step with probability at most $2 \cdot 2^{-\frac{\Delta}{10}}$.

- Let $M(x)$ be the next message in Π on input x and let $J(x)$ be the partition class of the density restoring partition that contains x. We define the random variables $M^* = M(X)$,$J^* = J(X)$,$M' = M(X')$, and $J' = J(X')$. As we will explain momentarily, there exists a coupling of the pairs (M^*, J^*) and (M', J') s.t $\Pr[(M^*, J^*) \neq (M', J')] \leq 2^{-\frac{\Delta}{10}}$. The algorithm samples from this coupling and appends the resulting samples to e^* and e'. If the samples are different, the algorithm fails.
  To show that the required coupling exists, we use the uniform marginals lemma (Lemma 2.17) with $\gamma = \frac{1}{10}$ to show that X and X' are $2^{-\frac{\Delta}{10}}$-close. Therefore $M(X), J(X)$ and $M(X'), J(X)$ are $2^{-\frac{\Delta}{10}}$-close , which implies the existence of the desired coupling by Fact 2.8. In order to apply the uniform marginals lemma, we need to show that X and Y are sufficiently sparse. The random variable Y is σ-sparse by the invariant of the tree. To see why X is sufficiently sparse, recall that X was $(\sigma + \frac{4}{c})$-sparse at the beginning at the beginning of the iteration, and the last step only conditioned it on an event of probably $\geq 1 - 2^{-\frac{\Delta}{10}}$.

- The algorithm appends z_I to both e^* and e', where I is the set associated with the class $J' = J^*$.

- For the last message the algorithm appends 0 to e^*(since in the last step of the extended tree it always appends 0). The algorithm appends 0 to e' with probability $\Pr[Y' \in \mathcal{E}]$, otherwise it appends 1 and fails.
  We turn to bound the probability to fail in this step. As $\Pr[Y \notin \mathcal{E}]$ is at most $2^{-\alpha\Delta}$, it is tempting to apply the uniform marginals lemma to relate $\Pr[Y' \in \mathcal{E}]$ and $\Pr[Y \in \mathcal{E}]$. Unfortunately the uniform marginals lemma cannot be applied as Y is not necessarily sparse at

this point. Yet, we are still able relate $\Pr[Y' \in \mathcal{E}]$ and $\Pr[Y \in \mathcal{E}]$ as follows:

$$\begin{aligned}
\Pr\left[Y' \in \mathcal{E}\right] =& \Pr\left[Y \in \mathcal{E} \mid g^n(X,Y) = z\right] \\
=& \frac{\Pr\left[g^n(X,Y) = z \mid Y \in \mathcal{E}\right]}{\Pr\left[g^n(X,Y) = z\right]} \cdot \Pr\left[Y \in \mathcal{E}\right] && \text{(Bayes' rule)} \\
=& \frac{\Pr\left[g^{\mathrm{free}(\rho)}\left(X_{\mathrm{free}(\rho)}, Y_{\mathrm{free}(\rho)}\right) = z_{\mathrm{free}(\rho)} \mid Y \in \mathcal{E}\right]}{\Pr\left[g^{\mathrm{free}(\rho)}\left(X_{\mathrm{free}(\rho)}, Y_{\mathrm{free}(\rho)}\right) = z_{\mathrm{free}(\rho)}\right]} \cdot \Pr\left[Y \in \mathcal{E}\right] \\
\geq& \frac{\Pr\left[g^{\mathrm{free}(\rho)}\left(X_{\mathrm{free}(\rho)}, Y_{\mathrm{free}(\rho)}\right) = z_{\mathrm{free}(\rho)} \mid Y \in \mathcal{E}\right]}{2^{-|\mathrm{free}(\rho)|} \cdot \left(1 + 2^{-\frac{\Delta}{10}}\right)} \cdot \Pr\left[Y \in \mathcal{E}\right] && (X \text{ is almost uniform})
\end{aligned}$$

Regarding the last inequality, recall that the tree removes all unsafe values from the support of X, thus all the values in the support of X are safe and thus almost uniform. Observe that by the definition of $\mathcal{E}$, it hold that $Y \mid Y \in \mathcal{E}$ is $\left(\sigma + \frac{4}{c}\right)$-sparse. Thus, we can bound the numerator from below using Proposition 2.16. It follows that

$$\begin{aligned}
&\Pr\left[Y \in \mathcal{E} \mid g^n(X,Y) = z\right] \\
\geq& \frac{2^{-|\mathrm{free}(\rho)|} \cdot \left(1 - 2^{-\frac{\Delta}{10}}\right)}{2^{-|\mathrm{free}(\rho)|} \cdot \left(1 + 2^{-\frac{\Delta}{10}}\right)} \cdot \Pr\left[Y \in \mathcal{E}\right] && \text{(By Proposition 2.16)} \\
\geq& \left(1 - 2 \cdot 2^{-\frac{\Delta}{10}}\right) \cdot \Pr\left[Y \in \mathcal{E}\right] \\
\geq& \Pr\left[Y \in \mathcal{E}\right] - 2 \cdot 2^{-\frac{\Delta}{10}} \geq 1 - \left(2 \cdot 2^{-\frac{\Delta}{10}} + 2^{-\alpha\Delta}\right).
\end{aligned}$$

Thus, the probability that the algorithm fails at this step is at most $\left(2 \cdot 2^{-\frac{\Delta}{10}} + 2^{-\alpha\Delta}\right)$.

At last, the sum of all the failure probabilities is

$$\begin{aligned}
&\overbrace{2 \cdot 2^{-\frac{\Delta}{10}}}^{\text{First Message}} + \overbrace{2^{-\frac{\Delta}{10}}}^{\text{Second Message}} + \overbrace{2^{-\frac{\Delta}{10}}}^{\text{Third Message}} + \overbrace{2 \cdot 2^{-\frac{\Delta}{10}} + 2^{-\alpha\Delta}}^{\text{Fifth message}} \\
\leq& 7 \cdot 2^{-\frac{\Delta}{10}} && (\alpha = \frac{1}{10}) \\
\leq& 2^{-\frac{\Delta}{20}}. && (\Delta \geq c = 1000)
\end{aligned}$$

The probability that $E'_{\leq i} \neq E^*_{\leq i}$ is at most the failure probability of the algorithm, which is at most

$$\begin{aligned}
\Pr\left[E'_{\leq i} \neq E^*_{\leq i}\right] &\leq \Pr\left[E'_{\leq i-1} \neq E^*_{\leq i-1}\right] + 2^{-\frac{\Delta}{20}} \\
&\leq 2^{-\frac{\Delta}{20}}(i-1) + 2^{-\frac{\Delta}{20}} \\
&= 2^{-\frac{\Delta}{20}} \cdot i
\end{aligned}$$

thus completing the proof. ■

As the number of rounds is bounded by the communication complexity of the protocol we get that there exists a coupling such that

$$\Pr\left[E' \neq E^*\right] \leq 2^{-\frac{\Delta}{20}} \cdot C.$$

Using Fact 2.8 and this coupling, we get that the statistical distance between E' and E^* is upper bounded by $2^{-\frac{\Delta}{20}} \cdot C$.

Remark 5.6. We now explain why we use the notion of almost-uniform values which was not present in [CFK+19]. Recall the work of [CFK+19] did not use the notion of "almost uniform" and instead used the notion of "non-leaking". The notion of "almost uniform" bounds the probability

$$\Pr\left[g^{\mathrm{free}(\rho)}\left(x_{\mathrm{free}(\rho)}, Y_{\mathrm{free}(\rho)}\right) = z_{\mathrm{free}(\rho)}\right]$$

from both below and above, while the notion of "non-leaking" only bounds the probability from below. In the above proof, in order to bound $\Pr\left[Y \in \mathcal{E} \mid g^n\left(X, Y\right) = z\right]$ we need to bound the latter probability from above, and this is the reason that we use the stronger notion of "almost uniform" instead of the weaker notion of "non-leaking". The reason that this issue did not come up in [CFK+19] is that the step of restoring the density of Y(Step 8) does not exist in [CFK+19], and therefore their analysis does not require to bound $\Pr\left[Y \in \mathcal{E} \mid g^n\left(X, Y\right) = z\right]$.

5.2.2 Proof of Claim 5.4

We now prove that the transcripts $\Pi^{\dagger}$ and Π^{*} are sufficiently close. By definition $\Pi^{\dagger}$ differs from Π^{*} if and only if the tree T halts on Step 5. For the ease of notation, we denote the event that the tree halts due to K_{msg} by $\mathcal{H}_{\mathrm{msg}}$, and the event that the tree halts due to K_{prt} by $\mathcal{H}_{\mathrm{prt}}$. Then, by Fact 2.2 and the union bound it hold that

$$\left|\Pi^{\dagger} - \Pi^{*}\right| \leq \Pr\left[\mathcal{H}_{\mathrm{msg}} \text{ or } \mathcal{H}_{\mathrm{prt}}\right] \leq \Pr\left[\mathcal{H}_{\mathrm{msg}}\right] + \Pr\left[\mathcal{H}_{\mathrm{prt}}\right].$$

In what follows we upper bound the probabilities of $\mathcal{H}_{\mathrm{msg}}$ and $\mathcal{H}_{\mathrm{prt}}$ separately.

It suffices to upper bound the probabilities of $\mathcal{H}_{\mathrm{msg}}$ and $\mathcal{H}_{\mathrm{prt}}$ conditioned on every fixed choice of the random coins of the protocol, since $\Pr\left[\mathcal{H}_{\mathrm{msg}}\right]$ and $\Pr\left[\mathcal{H}_{\mathrm{prt}}\right]$ are a convex combinations of those probabilities. For the rest of this section, we fix a choice of the random coins , and assume that all the probabilities below are conditioned on this choice.

We first set up some notation. Let $\bar{M} = (M_1, \ldots, M_r)$ be the random messages that are chosen by the tree and let $\bar{J} = (J_1, \ldots, J_r)$ be the random variable of the indices of the partition class chosen by the tree. We denote by $\bar{m} = (m_1, \ldots, m_r)$ and $\bar{j} = (j_1, \ldots, j_r)$ some specific values of $\bar{M}$ and $\bar{J}$ respectively.

Bound on the probability of $\mathcal{H}_{\mathrm{msg}}$. Before formally proving the bound on $\Pr\left[\mathcal{H}_{\mathrm{msg}}\right]$ we first give a high-level description of the argument. For every transcript let K_{msg} be the value of K_{msg} at the end of the simulation of this transcript. Then, it holds that the probability of getting this transcript is at most $2^{-K_{\mathrm{msg}}}$. Recall that the tree halts if $K_{\mathrm{msg}} > C + \Delta$, so every specific halting transcript has a probability of at most $2^{-C-\Delta}$. By taking a union bound over all of them, we get a probability of $2^{-\Delta}$ as required. In the formal proof, we also need to take into account in the union bound the different choices of the partition classes, see below. Note that the following proof is identical to the corresponding proof in [CFK+19] and we provide it here for completeness.

We define $\mathcal{B}$ to be the set of all pairs $(\bar{m}, \bar{j})$ for which T halts on Step 5 because of K_{msg}, then it holds that

$$\begin{aligned}
\Pr\left[\mathcal{H}_{\mathrm{msg}}\right] &= \sum_{(\bar{m},\bar{j})\in\mathcal{B}} \Pr\left[\bar{M} = \bar{m}, \bar{J} = \bar{j}\right] \\
&= \sum_{(\bar{m},\bar{j})\in\mathcal{B}} \prod_{i=1}^{r} \Pr\left[M_i = m_i \mid \bar{M}_{<i} = \overline{m}_{<i}, \bar{J}_{<i} = \bar{j}_{<i}\right] \cdot \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \overline{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right].
\end{aligned}$$

Now, recall that conditioned on the event $\mathcal{H}_{\text{msg}}$, we know that at the end of the simulation we have

$$K_{\text{msg}} = \sum \log \frac{1}{\Pr\left[M_i = m_i \mid \bar{M}_{<i} = \overline{m}_{<i}, \bar{J} = \bar{j}_{<i}\right]} > C + \Delta.$$

In other words,

$$\prod_{i=1}^{r} \Pr\left[M_i = m_i \mid \bar{M}_{<i} = \overline{m}_{<i}, \bar{J}_{<i} = \bar{j}_{<i}\right] < 2^{-C-\Delta},$$

and therefore we get

$$\begin{aligned}
\Pr\left[\mathcal{H}_{\text{msg}}\right] &= \sum_{(\bar{m},\bar{j})\in\mathcal{B}} \prod_{i=1}^{r} \Pr\left[M_i = m_i \mid \bar{M}_{\leq j} = \overline{m}_{<i}, \bar{J}_{<i} = \bar{j}_{<i}\right] \cdot \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \bar{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right] \\
&< \sum_{(\bar{m},\bar{j})\in\mathcal{B}} 2^{-C-\Delta} \cdot \prod_{i=1}^{r} \Pr\left[J_i = j_i \mid \bar{M}_{\leq j} = \overline{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right] \\
&\leq 2^{-C-\Delta} \sum_{(\bar{m},\bar{j})} \prod_{i=1}^{r} \Pr\left[J_i = j_i \mid \bar{M}_{\leq j} = \overline{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right].
\end{aligned}$$

We claim that the sum $\sum_{(\bar{m},\bar{j})} \prod_{i=1}^{r} \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \bar{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right]$ is equal to $\sum_{\bar{m}} 1$, which in our case is upper bounded by 2^C. To see it, observe that for every round t it holds that

$$\begin{aligned}
&\sum_{\bar{m},\bar{j}_{\leq t}} \prod_{i=1}^{t} \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \bar{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right] \\
&= \sum_{\bar{m},\bar{j}_{<t}} \left(\left(\prod_{i=1}^{t-1} \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \bar{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right] \right) \cdot \sum_{j_t} \Pr\left[J_t = j_t \mid M_{\leq t} = \bar{m}_{\leq t}, \bar{J}_{<i} = \bar{j}_{<t}\right] \right) \\
&= \sum_{\bar{m},\bar{j}_{<t}} \prod_{i=1}^{t-1} \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \bar{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right],
\end{aligned}$$

where the last transition hold as

$$\sum_{j_t} \Pr\left[J_t = j_t \mid \bar{M}_{\leq t} = \bar{m}_{\leq t}, \bar{J}_{<i} = \bar{j}_{<t}\right] = 1.$$

By induction, we get that

$$\sum_{(\bar{m},\bar{j})} \prod_{i=1}^{r} \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \bar{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right] = \sum_{\bar{m}} 1 \leq 2^C.$$

We now get that

$$\Pr\left[\mathcal{H}_{\text{msg}}\right] < 2^{-C-\Delta} \sum_{\bar{m},J} \prod_{i=1}^{r} \Pr\left[J_i = j_i \mid \bar{M}_{\leq i} = \bar{m}_{\leq i}, \bar{J}_{<i} = \bar{j}_{<i}\right] \leq 2^{-\Delta},$$

as required.

Bound on the probability of $\mathcal{H}_{\mathbf{prt}}$. This part of the proof is a variant of the analysis of [GPW17]. Let $p^{(i)}$ be the probability $p_{\geq j}$ in the i-th iteration. Using this notation, we can write $K_{\mathrm{prt}} = \sum_{i=1}^{C} \log \frac{1}{p_{\geq}^{(i)}}$. For the next claim, we will need the notion of *stochastic domination.* Given two real-valued random variables X, Y, we say that X is *stochastically dominant* over Y if for all t it hold that

$$\Pr[X \geq t] \geq \Pr[Y \geq t].$$

We will also use the following fact, whose proof is deferred to the end of this section.

Fact 5.7. *Let $A_1 \dots A_n$ and $B_1 \dots B_n$ be random variables over $\mathbb{R}$ such that A_i are i.i.d and independent from $B_1 \dots B_n$. Assume that for all $i \leq n$ and $b_1, \dots, b_{i-1} \in \mathbb{R}$ the random the random variable B_i is stochastically dominant A_i conditioned on $B_1 = x_1, \dots, B_{i-1} = x_{i-1}$. Then, $\sum_{i=1}^{n} B_i$ is stochastically dominant over $\sum_{i=1}^{n} A_i$.*

Claim 5.8. *The Erlang distribution $Erl(C, \ln 2)$ is stochastically dominant over K_{prt}.*

Proof. Let $\left(U^{(1)}, \dots, U^{(C)}\right)$ be independent uniform random variables over $[0,1]$. We start by proving that $\Pr\left[\forall i : p^{(i)} \geq t_i\right] \geq \Pr\left[\forall i : U^{(i)} \geq t_i\right]$ for every $t_i \in [0,1]$. We do so by analyzing the distribution of

$$p^{(i)} \mid M_{<i} = \bar{m}_{<i}, J_{<i} = \bar{j}_{<i}.$$

Let $p_1 \dots p_l$ be the probabilities assigned to the different partition classes in this round. Then, the probability $\Pr\left[p^{(i)} = p_{\geq j} \mid M_{<i} = \bar{m}_{<i}, J_{<i} = \bar{j}_{<i}\right]$ is equal to $p_{\geq j} - p_{\geq (j+1)}$. Thus, it easy to see that

$$\Pr\left[p^{(i)} \geq t \mid M_{<i} = \bar{m}_{<i}, J_{<i} = \bar{j}_{<i}\right] \geq 1 - t = \Pr\left[U^{(i)} \geq t\right]$$

for every choice of $\bar{m}_{<i}$ and $\bar{j}_{<i}$ and every $t \in [0,1]$. Let $\delta^{(i)} = \log \frac{1}{U^{(i)}}$. As the function $\log \frac{1}{x}$ is monotonically decreasing, we can get that

$$\Pr\left[\log \frac{1}{p^{(i)}} \geq t \mid M_{<i} = \bar{m}_{<i}, J_{<i} = \bar{j}_{<i}\right] \leq \Pr\left[\delta^{(i)} \geq t\right].$$

for every choice of $\bar{m}_{<i}$ and $\bar{j}_{<i}$. Note that for all real $a_1 \dots a_{i-1}$ it holds that

$$\Pr\left[\log \frac{1}{p^{(i)}} \geq t \mid p^{(1)} = a_1, \dots, p^{(i-1)} = a_{i-1}\right]$$

is a convex combination of the probabilities $\Pr\left[\log \frac{1}{p^{(i)}} \geq t \mid M_{<i} = \bar{m}_{<i}, J_{<i} = \bar{j}_{<i}\right]$ for different $\bar{m}_{<i}$ and $\bar{j}_{<i}$, and thus

$$\Pr\left[\log \frac{1}{p^{(i)}} \geq t \mid p^{(1)} = a_1, \dots, p^{(i-1)} = a_{i-1}\right] \leq \Pr\left[\delta^{(i)} \geq t\right].$$

By applying Fact Fact 5.7 we get that

$$\Pr\left[\sum_{i=1}^{C} \log \frac{1}{p^{(i)}} \geq t\right] \leq \Pr\left[\sum_{i=1}^{C} \delta^{(i)} \geq t\right].$$

The left-hand side of the equation is just $\Pr[K_{\mathrm{prt}} \geq t]$. In the right-hand side, the random variable $\delta^{(i)}$ is distributed like the exponential distribution $\mathrm{Ex}(\ln 2)$ as

$$\Pr\left[\delta^{(i)} \leq t\right] = \Pr\left[\log \frac{1}{U^{(i)}} \leq t\right] = \Pr\left[U^{(i)} \geq 2^{-t}\right] = 1 - 2^{-t}.$$

As the Erlang random variable is a sum of exponential random variables, we get that $\sum_{i=1}^{C} \delta^{(i)}$ is distributed like $\mathrm{Erl}(C, \ln 2)$, and thus we complete the proof. ■

As a result of the above claim, it holds that in order to bound the probability that $K_{\text{prt}} > 5C + 2\Delta$, it suffices to bound the probability that $\text{Erl}(C, \ln 2) > 5C + 2\Delta$. For convenience, we denote $t \stackrel{\text{def}}{=} 5C + 2\Delta$ and $\lambda \stackrel{\text{def}}{=} \ln 2$.

$$\Pr\left[\text{Erl}(C, \lambda) > t\right] = e^{-\lambda \cdot t} \sum_{i=0}^{C-1} \frac{1}{i!} \cdot (\lambda t)^i \,.$$

By our choice of t, it easy to see that $\frac{1}{(i+1)!} \cdot (\lambda t)^{i+1}$ is larger than $\frac{1}{i!} \cdot (\lambda t)^i$ by a factor of at least 2. Therefore

$$\begin{aligned}
\sum_{i=0}^{C-1} \frac{1}{i!} \cdot (\lambda t)^i &\leq \sum_{i=0}^{C-1} \left(\frac{1}{2}\right)^{C-i} \frac{1}{C!} \cdot (\lambda t)^C \\
&< \frac{1}{C!} \cdot (\lambda t)^C \cdot \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{-i} \\
&= \frac{1}{C!} \cdot (\lambda t)^C \,.
\end{aligned}$$

Then

$$\begin{aligned}
\Pr\left[\text{Erl}(C, \lambda) > t\right] &\leq 2^{-t} \cdot (\lambda t)^C \cdot \frac{1}{C!} \\
&\leq 2^{-t} \cdot (\lambda t)^C \cdot \left(\frac{e}{C}\right)^C && \left(C! \geq \left(\frac{C}{e}\right)^C\right).
\end{aligned}$$

Substituting $t = 5C + 2\Delta$ we get

$$\begin{aligned}
\Pr\left[\text{Erl}(C, \lambda) > 5C + 2\Delta\right] &\leq 2^{-5C-2\Delta} \cdot (5\lambda C + 2\lambda\Delta)^C \cdot \left(\frac{e}{C}\right)^C \\
&= 2^{-5C-2\Delta} \cdot 5^C \cdot (\lambda e)^C \cdot \left(1 + \frac{2\Delta}{5 \cdot C}\right)^C \\
&\leq 2^{-5C-2\Delta} \cdot 5^C \cdot (\lambda e)^C \cdot e^{C + \frac{2\Delta}{5}} && ((1+x)^y \leq e^{x \cdot y}) \\
&= 2^{-5C-2\Delta} \cdot 2^{(\log 5) \cdot C} \cdot 2^{(\log(\lambda e)) \cdot C} \cdot 2^{(\log e) \cdot C + (\log e) \cdot \frac{2}{5} \cdot \Delta} \\
&= 2^{-\left(5 - \log(5\lambda e^2)\right)C - \left(2 - \frac{2 \cdot \log e}{5}\right) \cdot \Delta}.
\end{aligned}$$

As $5 - \log(5\lambda e^2) \geq 0$ and $2 - \frac{2 \cdot \log e}{5} \geq 1$ we have that

$$\Pr\left[K_{\text{prt}} > 5C + 2\Delta\right] \leq \Pr\left[\text{Erl}_{C,\lambda} > 5C + 2\Delta\right] \leq 2^{-\Delta}.$$

Proof of Fact 5.7 We will prove by induction. The base case is trivial as $\sum_{i=1}^{1} B_i = B_1$. Assume that $\Pr\left[\sum^{n-1} B_i \geq x\right] \geq \Pr\left[\sum^{n-1} A_i \geq x\right]$. Then,

$$\begin{aligned}
\Pr\left[\sum^{n} B_i \geq x\right] &= \mathbb{E}_{b_1,\ldots,b_{n-1}\leftarrow B_1,\ldots,B_{n-1}}\left[\Pr\left[B_n \geq x - \sum^{n-1} b_i \mid B_1 = b_1,\ldots,B_{n-1} = b_{n-1}\right]\right] \\
&\geq \mathbb{E}_{b_1\ldots b_{n-1}\leftarrow B_1\ldots B_{n-1}}\left[\Pr\left[A_n \geq x - \sum^{n-1} b_i\right]\right] \\
&= \Pr\left[A_n + \sum^{n-1} B_i \geq x\right] \\
&= \mathbb{E}_{a_n\leftarrow A_n}\left[\Pr\left[\sum^{n-1} B_i \geq x - a_n\right]\right] \\
&\geq \mathbb{E}_{a_n\leftarrow A_n}\left[\Pr\left[\sum^{n-1} A_i \geq x - a_n\right]\right] \\
&= \Pr\left[\sum^{n} A_i \geq x\right].
\end{aligned}$$

■

5.3 Query complexity of the decision tree

The following analysis of the query complexity of the randomized decision tree is very similar to the analysis of the deterministic decision tree. As in the analysis of deterministic case we define our potential function to be

$$D_\infty\left(X_{\mathrm{free}(\rho)}, Y_{\mathrm{free}(\rho)}\right) = D_\infty\left(X_{\mathrm{free}(\rho)}\right) + D_\infty\left(Y_{\mathrm{free}(\rho)}\right).$$

The main difference between the analysis of the deterministic and randomized cases is as follows. While in the deterministic case we could bound the increase in the deficiency caused by sending a message m by $|m|$, in the randomized setting this does not hold. Nevertheless, K_{msg} bounds the increase in deficiency. Another difference is that Step 3 can increase the deficiency by extra term of $\log\frac{1}{p_{\geq j}}$. Due to Step 5 of the tree, it holds that $K_{\mathrm{msg}} + K_{\mathrm{prt}} \leq 6C + 3\Delta$ (otherwise the tree halts).

With the latter issue resolved, the rest of the analysis is similar to the deterministic case. We prove that any query decreases the deficiency by at least $\Omega(\Delta)$ per queried bit. Thus, we get an upper bound of $O\left(\frac{C+\Delta}{\Delta}\right) = O\left(\frac{C}{\Delta} + 1\right)$ on the number of queries. We now analyze a single iteration of the tree. Without loss of generality, we assume that it is Alice's turn to speak. We go through the iteration step by step. We start by noting that Steps 4 and 5 do not change the deficiency, and therefore are ignored in the following list.

- In Step 1, the tree conditions on an event with probability $1 - 2^{-\frac{1}{10}\Delta} \geq \frac{1}{2}$ by Lemma 3.3 with $\gamma = \frac{1}{10}$. Therefore, the deficiency increases at by most 1 by Fact 2.4.

- In Step 2, the tree samples a message and then conditions on sending it. Therefore, the conditioning increases the deficiency by at most $\log\frac{1}{p_m}$ by Fact 2.4.

- In Step 3, the tree chooses a partition class $\mathcal{X}^j$ of X. By Lemma 2.19 we know that the deficiency increases by at most $(b - \sigma\cdot\Delta)\cdot|I_j| + \log\frac{1}{p_{\geq j}}$. For the rest of the analysis we will denote the probability $p_{\geq j}$ and set I_j at the i-th step by $p^{(i)}$ and $I^{(i)}$ respectively.

- In Step 6, the tree decreases the size of $\text{free}(\rho)$ by $\left|I^{(i)}\right|$. It holds that $D_\infty\left(Y_{\text{free}(\rho)}\right)$ does not increase by Fact 2.5, while $D_\infty\left(X_{\text{free}(\rho)}\right)$ decreases by $b\cdot\left|I^{(i)}\right|$ as $X_{I^{(i)}}$ is fixed to $x_{I^{(i)}}$. Altogether we get that the deficiency decreases by at least $b\cdot\left|I^{(i)}\right|$.

- In Step 7, the tree conditions Y on $g^{I^{(i)}}\left(x_{I^{(i)}}, Y_{I^{(i)}}\right)=\rho_I$. As x is safe, we know that the probability of this event is at least $2^{-\left|I^{(i)}\right|-1}\geq 2^{-2\left|I^{(i)}\right|}$, and therefore the deficiency increases by at most $2\left|I^{(i)}\right|$ by Fact 2.4.

- In Step 8, the tree conditions on the event $\mathcal{E}$ such that $\Pr\left[\mathcal{E}\right]\geq\frac{1}{2}$. Thus, the deficiency increases at most by 1 by Fact 2.4.

Steps 1 and 8 can be executed at most C times and therefore contribute at most $2C$ to the deficiency. Step 2 contributes $\sum_{m\in\pi}\log\frac{1}{p_m}=K_{\text{msg}}$,which is at most $C+\Delta$ due to Step 4. Similarly, the contribution of the term $\log\frac{1}{p_{\geq j}}$ from Step 3 is $\sum\log\frac{1}{p^{(i)}}=K_{\text{prt}}$, which is at most $5C+2\Delta$ due to Step 5. We note that in fact when the tree halts at Steps 4 or 5 the counters K_{msg} or K_{prt} may be bigger then $C+\Delta$ and $5C+2\Delta$ respectively, but such an iteration does not make queries and therefore it does not affect the analysis.

We turn to bound the change in deficiency that is caused by Step 3 (without the $\log\frac{1}{p^{(i)}}$), and Steps 6 and 7. Assume that the tree queries I in some iteration. In Steps 3, and 7 the deficiency increases by at most

$$(b-\sigma\cdot\Delta)\cdot\left|I^{(i)}\right|+2\left|I^{(i)}\right|,$$

and Step 6 decreases the deficiency by at least $b\cdot\left|I^{(i)}\right|$. Altogether, we get that the deficiency decreases by at least

$$\begin{aligned}
b\cdot\left|I^{(i)}\right|-(b-\sigma\cdot\Delta)\cdot\left|I^{(i)}\right|-2\left|I^{(i)}\right| &= \Delta\cdot\left|I^{(i)}\right|\left(\sigma-\frac{2}{c}\right)\\
&=\frac{1}{10}\cdot\Delta\cdot\left|I^{(i)}\right|\in\Omega\left(\Delta\cdot\left|I^{(i)}\right|\right). \qquad (\sigma=\frac{1}{10}+\frac{2}{c})
\end{aligned}$$

Summing over all iterations of the simulation we get that

$$\begin{aligned}
D_\infty\left(X_{\text{free}(\rho)}\right)+D_\infty\left(X_{\text{free}(\rho)}\right) &\leq 8C+3\Delta-\sum_i\frac{\Delta}{10}\cdot\left|I^{(i)}\right|\\
&\leq 8C+3\Delta-\frac{\Delta}{10}\cdot\sum_i\left|I^{(i)}\right|.
\end{aligned}$$

By Fact 2.3, $D_\infty\left(X_{\text{free}(\rho)}\right)+D_\infty\left(X_{\text{free}(\rho)}\right)\geq 0$ therefore we get that

$$\begin{aligned}
\frac{\Delta}{10}\cdot\sum_i\left|I^{(i)}\right| &\leq 8C+3\Delta\\
\sum_i\left|I^{(i)}\right| &\leq 80\left(\frac{C}{\Delta}+1\right)\in O\left(\frac{C}{\Delta}+1\right).
\end{aligned}$$

The term $\sum_i\left|I^{(i)}\right|$ is the equal to query complexity of the tree, and thus we complete the proof.

6 The tightness of the main lemma of [CFK$^+$19]

Our work expands upon the results of [CFK$^+$19] and generalizes them into a wider regime of parameters. Despite the similarities between our work and their work, a crucial difference between the two works is that we use a weaker notion of safe values. To motivate the need for a weaker notion of safety, we show that the conclusion of the main lemma of [CFK$^+$19] cannot hold when $\Delta \ll b$. We first recall the notion of dangerous values of [CFK$^+$19] and restate their main lemma in our terms.

Definition 2.20. *Let Y be a random variable taking values from Λ^n. We say that a value $x \in \Lambda^n$ is* leaking *if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that*

$$\Pr\left[g^I\left(x_I, Y_I\right) = z_I\right] < 2^{-|I|-1}.$$

Let $\sigma_Y, \varepsilon > 0$, and suppose that Y is σ_Y-sparse. We say that a value $x \in \Lambda^n$ is ε-sparsifying *if there exists a set $I \subseteq [n]$ and an assignment $z_I \in \{0,1\}^I$ such that the random variable $Y_{[n]-I} \mid g^I\left(x_I, Y_I\right) = z_I$ is not $(\sigma_Y + \varepsilon)$-sparse. We say that a value $x \in \Lambda^n$ is* ε-dangerous *if it is either leaking or ε-sparsifying.*

Remark 6.1. We note that even when $\varepsilon \geq 1$, Definition 2.20 remains non-trivial, since the sparsity is measured with respect to Δ rather than b. On the other hand, from the perspective of the simulation, a value x that is 1-dangerous is "useless". The reason is that the discrepancy of g (and in particular, Lemmas 2.10 and 2.11) cannot give any effective bound on the biases of ε-sparse distributions for $\varepsilon \geq 1$.

Lemma 2.21 (Main lemma of [CFK$^+$19])**.** *There exists universal constants h, c such that the following holds: Let b be some number such that $b \geq c \cdot \log n$ and let $\gamma, \varepsilon, \sigma_X, \sigma_Y > 0$ be such that $\varepsilon \geq \frac{4}{\Delta}$, and $\sigma_X + \sigma_Y \leq 1 - \frac{h \cdot b \cdot \log n}{\Delta^2 \cdot \varepsilon} - \gamma$. Let X, Y be $(\rho, \sigma_X, \sigma_Y)$-structured random variables. Then, the probability that $X_{\mathrm{free}(\rho)}$ takes a value that is ε-dangerous for $Y_{\mathrm{free}(\rho)}$ is at most $2^{-\gamma\Delta}$.*

It can be seen that the lemma is only applicable when $\Delta = \Omega(\sqrt{b \cdot \log n})$. In this section we show this is almost optimal. The main result of this section is that in the case where $\Delta \in O\left(\sqrt{b}\right)$, the conclusion of the main lemma of [CFK$^+$19] completely fails: that is, it may happen that $X_{\mathrm{free}(\rho)}$ takes only values that are 2-dangerous for $Y_{\mathrm{free}(\rho)}$. As noted in Remark 6.1, such values are useless for the simulation.

Proposition 6.2. *For every $b \geq 1000$ and $n \geq \left\lfloor 3\sqrt{b} \right\rfloor + 1$ the following holds: There exist an inner function $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ with $\Delta(g) \in \Theta\left(\sqrt{b}\right)$ and two random variables X, Y over $\left(\{0,1\}^b\right)^n$ that are $\frac{2}{\Delta}$-sparse such that every $x \in supp(X)$ is 2-dangerous for Y.*

The rest of this section is dedicated to proving Proposition 6.2. We start by introducing some notation. Let b, n be as in the proposition and let $\Lambda = \{0,1\}^b$. We denote by I the set $\left[\left\lfloor 3\sqrt{b}\right\rfloor\right]$ and define $d \overset{\text{def}}{=} \left\lfloor \frac{1}{3}\sqrt{b} \right\rfloor$. For each value $y \in \Lambda^n$, we view the last block $y_n \in \Lambda$ as consisting of $|I|$ substrings of length d and of the remaining $b - d|I|$ bits. We call each such string of d bits a *cell*, and denote the i-th cell by $y_{n,i}$ for every $i \in I$. Note that the definition of cells applies only to the last block y_n. For each $i \in I$, we refer to the first d bits of the i-th block y_i as the *prefix* of y_i, and denote it by y_i^{pre}. Let U be the uniform distribution over Λ^n, and for each $i \in I$, let $\mathcal{A}_i$ be the event that $\langle U_i^{\mathrm{pre}}, U_{n,i} \rangle = 1$.

We now choose the function g to be $g(v,w) = \langle v^{\text{pre}}, w^{\text{pre}} \rangle$ for $v, w \in \Lambda$. It hold that $\frac{d}{2} \leq \Delta \leq d+1$, where the lower bound is by Lindsey's lemma and the upper bound is trivial. We choose the random variable X to be the uniform distribution over Λ^n. We define Y to be equal to the random variable U conditioned on the events $\mathcal{A}_i$ for every $i \in I$. In order to prove Proposition 6.2 we need to show that X, Y are $\frac{2}{\Delta}$-sparse and that every $x \in \operatorname{supp}(X)$ is 2-dangerous for Y. The variable X is trivially 0-sparse. We turn to prove that Y is $\frac{2}{\Delta}$-sparse.

Claim 6.3. *Y is $\frac{2}{\Delta}$-sparse*

Proof. For any set S, we prove that the deficiency $D_\infty(Y_S) \leq 2|S| = \frac{2}{\Delta} \cdot \Delta \cdot |S|$, and this would imply that Y is $\frac{2}{\Delta}$-sparse. Let U be the uniform distribution over Λ^n. In order to bound $D_\infty(Y_S)$, we start by expressing $\Pr[Y_S = y_S] \cdot 2^{-b|S|}$ as follows

$$\begin{aligned}
\frac{\Pr[Y_S = y_S]}{2^{b|S|}} &= \frac{\Pr[U_S = y_S \mid \forall i \in I : \mathcal{A}_i]}{\Pr[U_S = y_S]} && (Y = U \mid \forall i \in I : U \in \mathcal{A}_i) \\
&= \frac{\Pr[\forall i \in I : \mathcal{A}_i \mid U_S = y_S]}{\Pr[\forall i \in I : \mathcal{A}_i]}. && \text{(Bayes' formula)}
\end{aligned}$$

We note that $\Pr[\forall i \in I : \mathcal{A}_i] = \left(\frac{1-2^{-d}}{2}\right)^{|I|}$: this holds as each event $\mathcal{A}_i$ has probability $\frac{1-2^{-d}}{2}$ and they are all independent. We turn to bound the probability in the above numerator. The events $\mathcal{A}_i$ are independent even conditioned on $U_S = y_S$, and therefore we can bound their probabilities separately. In general, the probability of $\mathcal{A}_i$ conditioned on $U_S = y_S$ can be as large as 1 (for example, if both i and n are in S, and $\langle y_i^{\text{pre}}, y_{n,i} \rangle = 1$). Nevertheless, it is easy to see that for each $i \in I \backslash S$, the probability of $\mathcal{A}_i$ conditioned on $U_S = y_S$ is at most $\frac{1}{2}$ even if $n \in S$. Therefore we can get a bound as follows

$$\frac{\Pr[\forall i \in I : \mathcal{A}_i \mid U_S = y_S]}{\Pr[\forall i \in I : \mathcal{A}_i]} \leq \frac{2^{-|I \backslash S|}}{\left(\frac{1-2^{-d}}{2}\right)^{|I|}} = 2^{|S|} \cdot \left(1 - 2^{-d}\right)^{-|I|}.$$

The second term can be bounded as follows

$$\left(1 - 2^{-d}\right)^{-|I|} \leq e^{|I| 2^{-d}} \leq e^{3\sqrt{b} \cdot 2^{-\left\lfloor \frac{1}{3}\sqrt{b} \right\rfloor}},$$

which is at most 2 for $b \geq 1000$. Therefore

$$D_\infty(Y_S) = \log \max_{y_S} \frac{\Pr[Y_S = y_S]}{2^{b|S|}} \leq 1 + |S| \leq 2|S|. \qquad \blacksquare$$

It remain to prove that every $x \in \operatorname{supp}(X)$ is 2-dangerous for Y. In order to do so, we use the following claim on the distribution of Y.

Claim 6.4. *Let y_n be such that $y_{n,i} \neq \overline{0}$ for every i. Then* $\Pr[Y_n = y_n] \geq 2^{-b}$.

Proof. First, observe that the event $Y_{n,i} = y_{n,i}$ is independent of the event $\mathcal{A}_j$ for every $j \neq i$ and thus it holds that

$$\begin{aligned}
\Pr[Y_{n,i} = y_{n,i}] &= \Pr[U_{n,i} = y_{n,i} \mid \mathcal{A}_i] \\
&= \frac{\Pr[\mathcal{A}_i \mid U_{n,i} = y_{n,i}] \Pr[U_{n,i} = y_{n,i}]}{\Pr[\mathcal{A}_i]} && \text{(Bayes' rule).}
\end{aligned}$$

As $y_{n,i} \neq \bar{0}$ it easy to see that $\Pr[\mathcal{A}_i \mid U_{n,i} = y_{n,i}] \geq \Pr[\mathcal{A}_i]$, and therefore

$$\frac{\Pr[\mathcal{A}_i \mid U_{n,i} = y_{n,i}] \Pr[U_{n,i} = y_{n,i}]}{\Pr[\mathcal{A}_i]} \geq \Pr[U_{n,i} = y_{n,i}] = 2^{-d}.$$

Finally by combining this inequality for all cells and and the fact that the remaining $b - d \cdot |I|$ bits are uniformly distributed and independent, we get the desired result. ■

We finally show that all values in $\operatorname{supp}(X)$ are 2-dangerous. It holds that every x such that $x_i^{\text{pre}} = \bar{0}$ for some $i \in I$ is leaking as $\Pr[g^n(x, Y) = 1^n] = 0$, and thus such an x is dangerous. It remains to handle the case where $x_i^{\text{pre}} \neq 0^d$ for all $i \in I$. Let $x \in \Lambda^n$ be such a value. We show that every such x is 2-sparsifying. Specifically, we show that there exists a value y_n such that $\Pr\left[Y_n = y_n \mid g^I(x_I, Y_I) = 1^{|I|}\right]$ is too high. We choose y_n be equal to the concatenation of all prefixes x_i^{pre} appended by $b - d|I|$ zeros. For this choice of y_n it hold that

$$\begin{aligned}
Y_n = y_n &\Rightarrow \forall i \in I : \langle y_{n,i}, Y_i^{\text{pre}} \rangle = 1 && \text{(definition of } Y\text{)}\\
&\Rightarrow \forall i \in I : \langle x_i^{\text{pre}}, Y_i^{\text{pre}} \rangle = 1 && \text{(definition of } y_{n,1}\text{)}\\
&\Rightarrow g^I(x_I, Y_I) = 1^I.
\end{aligned}$$

Thus we have that

$$\begin{aligned}
&\Pr\left[Y_n = y_n \mid g^I(x_I, Y_I) = 1^{|I|}\right]\\
=&\frac{\Pr[Y_n = y_n] \cdot \Pr\left[g^I(x_I, Y_I) = 1^{|I|} \mid Y_n = y_n\right]}{\Pr\left[g^I(x_I, Y_I) = 1^{|I|}\right]} && \text{(Bayes' formula)}\\
=&\frac{\Pr[Y_n = y_n]}{\Pr\left[g^I(x_I, Y_I) = 1^{|I|}\right]} && (Y_n = y_n \Rightarrow g^I(x, Y) = 1^I).
\end{aligned}$$

We note that the events $g(x_i, Y_i) = 1$ are independent for each $i \in I$, and that each of them occurs with probability at most $\leq \frac{1}{2 \cdot (1 - 2^{-d})}$, thus

$$\Pr\left[g^I(x_I, Y_I) = 1^{|I|}\right] \leq \left(\frac{1}{2 \cdot (1 - 2^{-d})}\right)^{|I|}.$$

Therefore we get

$$\Pr\left[Y_n = y_n \mid g^I(x_I, Y_I) = 1^{|I|}\right] \geq \Pr[Y_n = y_n] \cdot \left(2 - 2^{1-d}\right)^{|I|}.$$

It holds that $\Pr[Y_n = y_n] \geq 2^{-b}$ by Claim 6.4 as $y_{n,i} = x_i^{\text{pre}} \neq \bar{0}$ for all i. Thus, we get that

$$\begin{aligned}
\Pr\left[Y_n = y_n \mid g^I(x_I, Y_I) = 1^{|I|}\right] &\geq 2^{-b} \cdot \left(2 - 2^{1-d}\right)^{|I|}\\
&= 2^{|I|-b} \cdot \left(1 - 2^{-d}\right)^{|I|}\\
&\geq 2^{|I|-b} \cdot \left(1 - 2^{-\left\lfloor \frac{1}{3}\sqrt{b} \right\rfloor}\right)^{3\sqrt{b}}\\
&\geq 2^{|I|-b-1} && (b \geq 1000).
\end{aligned}$$

Hence, the distribution $Y_n \mid g^I(x_I, Y_I) = 1^{|I|}$ is not $\frac{|I|-1}{\Delta}$-sparse and thus x is $\frac{|I|-3}{\Delta}$-sparsifying. As a result we can see that all values of x are $\left(\frac{|I|-2}{\Delta}\right)$-dangerous. We have that

$$\begin{aligned}
\left(\frac{|I|-3}{\Delta}\right) &\geq \frac{\left\lfloor 3\sqrt{b}\right\rfloor - 3}{\left\lfloor \frac{1}{3}\sqrt{b}\right\rfloor + 1} && \left(\Delta \leq d+1, d = \left\lfloor \frac{1}{3}\sqrt{b}\right\rfloor, |I| = \left\lfloor 3\sqrt{b}\right\rfloor\right) \\
&\geq \frac{3\sqrt{b}-4}{\frac{1}{3}\sqrt{b}+1} && (x \geq \lfloor x \rfloor \geq x-1) \\
&= 9\left(1 - \frac{13}{3\sqrt{b}+9}\right) \\
&\geq 2 && (b \geq 9),
\end{aligned}$$

as required.

Remark 6.5. We note that the choice of the constant 2 in Proposition 6.2 is arbitrary. For any constant ε one can construct an example such that all values are ε-dangerous, where $\Delta \in \Theta\left(\sqrt{\frac{b}{\varepsilon}}\right)$ and $n \geq b = \Omega(\varepsilon)$.

7 Discrepancy with respect to product distributions.

Discrepancy is commonly defined with respect to an underlying distribution as follows.

Definition 7.1. Let μ be a distribution over $\Lambda \times \Lambda$ and let $g : \Lambda \times \Lambda \to \{0,1\}$ be a function. Let $(V, W) \in \Lambda \times \Lambda$ be a pair of random variables that are distributed according to μ. Given a combinatorial rectangle $R \subseteq \Lambda \times \Lambda$, the *discrepancy of g with respect to R* (and μ), denoted $\mathrm{disc}_{\mu,R}(g)$, is defined as follows:

$$\mathrm{disc}_{\mu,R}(g) = \left| \Pr_{\mu}\left[g(V,W) = 0 \text{ and } (V,W) \in R\right] - \Pr_{\mu}\left[g(V,W) = 1 \text{ and } (V,W) \in R\right] \right|.$$

The *discrepancy of g* with respect to μ, denoted $\mathrm{disc}_{\mu}(g)$, is defined as the maximum of $\mathrm{disc}_{\mu,R}(g)$ over all combinatorial rectangles $R \subseteq \Lambda \times \Lambda$. We define

$$\Delta_{\mu}(g) \stackrel{\text{def}}{=} \log \frac{1}{\mathrm{disc}_{\mu}(g)}$$

Note that the definition of discrepancy that we used throughout this paper is the special case of the above definition for the uniform distribution. A natural question is to ask whether our main result (Theorem 1.3) can be generalized to other distributions. We show that the theorem holds with respect to every *product* distribution μ.

Theorem 7.2. *There exists a universal constant c such that the following holds: Let $\mathcal{S}$ be a search problem that takes inputs from $\{0,1\}^n$, and let $g : \Lambda \times \Lambda \to \{0,1\}$ be an arbitrary function such that $\Delta_{\mu}(g) \geq c \cdot \log n$ for some product distribution μ. Then*

$$D^{\mathrm{cc}}(\mathcal{S} \circ g^n) \in \Omega\left(D^{\mathrm{dt}}(\mathcal{S}) \cdot \Delta_{\mu}(g)\right),$$

and for every $\beta > 0$ it holds that

$$R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n) \in \Omega\left(\left(R^{\mathrm{dt}}_{\beta'}(\mathcal{S}) - O(1)\right) \cdot \Delta_{\mu}(g)\right),$$

where $\beta' = \beta + 2^{-\Delta_{\mu}(g)/50}$.

We start by defining a simple kind of reduction between two communication problems,

Definition 7.3. We say that a function $g' : \Lambda' \times \Lambda' \to \mathcal{O}$ is reducible to $g : \Lambda \times \Lambda \to \mathcal{O}$ if there exist functions $r_A, r_B : \Lambda' \to \Lambda$ such that for every $x', y' \in \Lambda'$ it holds that $g'(x', y') = g(r_A(x), r_B(y))$.

It easy to see that if g' is reducible to g then $D^{\mathrm{cc}}(g') \le D^{\mathrm{cc}}(g)$ and $R^{\mathrm{cc}}_{\beta}(g') \le R^{\mathrm{cc}}_{\beta}(g)$.

Claim 7.4. *Given a search problem $S : \{0,1\}^n \to \mathcal{O}$ and functions $g' : \Lambda' \times \Lambda' \to \{0,1\}$ that is reducible to $g : \Lambda \times \Lambda \to \{0,1\}$ it hold that $S \circ (g')^n$ is reducible to $S \circ g^n$.*

Proof. Let r'_A, r'_B the functions that reduce g' to g, and let $r_A = (r'_A)^n$and let $r_B = (r'_B)^n$. Then, it easy to see that

$$(S \circ g^n) \circ \left(r'_A \times r'_B\right)^n = S \circ \left(g \circ \left(r'_A \times r'_B\right)\right)^n = S \circ \left(g'\right)^n \qquad \blacksquare$$

To prove Theorem 7.2, we use the following lemma that relates discrepancy of a function g with respect to the some product distribution μ to the discrepancy with respect to the uniform distribution for some function g' that is reducible to g.

Lemma 7.5. *Let $g : \Lambda \times \Lambda \to \{0,1\}$ be a function and let $\mu = \mu_X \times \mu_Y$ be a product distribution. Then, for every constant $\varepsilon > 0$ there exists a function $g' : \Lambda' \times \Lambda' \to \{0,1\}$ reducible to g such that $\mathrm{disc}_U(g') \le \mathrm{disc}_{\mu}(g) + \varepsilon$, where U is the uniform distribution over $\Lambda' \times \Lambda'$.*

We prove Lemma 7.5 later in this section. We turn to prove Theorem 7.2 by applying Theorem 1.3 with a inner function g' that is constructed by Lemma 7.5.

Proof of Theorem 7.2 from Lemma 7.5 Let c' be the universal constant c from Theorem 1.3. We choose the universal constant c to be equal to $\max(c'+1, 2)$. Let n, μ, $\mathcal{S}$, and g be as in the theorem. Let $\varepsilon = \mathrm{disc}_{\mu}(g)$ and let g' be the function obtained from Lemma 7.5 for g and μ. As g' is reducible to g, it hold that $S \circ g^n$ is reducible to $S \circ (g')^n$. Therefore, it hold that $D^{\mathrm{cc}}(\mathcal{S} \circ (g')^n) \le D^{\mathrm{cc}}(\mathcal{S} \circ g^n)$ and $R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ (g')^n) \le R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n)$.

We note that the discrepancy of g' is bounded from above as follows

$$\mathrm{disc}_U\left(g'\right) \le \mathrm{disc}_{\mu}(g) + \mathrm{disc}_{\mu}(g) \le 2n^{-c} \le n^{-c'}$$

In other words, it holds that $\Delta_U(g') \ge \Delta_{\mu}(g) - 1 \ge c' \cdot \log n$. Hence, we can apply Theorem 1.3 on $S \circ g'$ and get

$$D^{\mathrm{cc}}(\mathcal{S} \circ g^n) \ge D^{\mathrm{cc}}\left(\mathcal{S} \circ \left(g'\right)^n\right) \in \Omega\left(D^{\mathrm{dt}}(\mathcal{S}) \cdot \Delta_U\left(g'\right)\right) = \Omega\left(D^{\mathrm{dt}}(\mathcal{S}) \cdot \Delta_{\mu}(g)\right),$$

and for every $\beta > 0$ it holds that

$$R^{\mathrm{cc}}_{\beta}(\mathcal{S} \circ g^n) \ge R^{\mathrm{cc}}_{\beta}\left(\mathcal{S} \circ \left(g'\right)^n\right) \in \Omega\left(\left(R^{\mathrm{dt}}_{\beta'}(\mathcal{S}) - O(1)\right) \cdot \Delta_U\left(g'\right)\right) = \Omega\left(\left(R^{\mathrm{dt}}_{\beta'}(\mathcal{S}) - O(1)\right) \cdot \Delta_{\mu}(g)\right),$$

where $\beta' = \beta + 2^{-\Delta(g)/50}$. $\blacksquare$

It only remain to prove Lemma 7.5. In order to prove the lemma, we first introduce a notion of a *canonical rectangle* with respect to function g and then we then show that the discrepancy is maximized with respect to such rectangles.

Definition 7.6. Let $g : \Lambda \times \Lambda \to \{0,1\}, g' : \Lambda' \times \Lambda'$ be functions such that g' is reducible to g and let $r_A, r_B : \Lambda' \to \Lambda$ be the reductions. A rectangle $R' \subseteq \Lambda' \times \Lambda'$ is *canonical* if and only if there exists a rectangle $R = A \times B \subseteq \Lambda \times \Lambda$ such that $R' = r_A^{-1}(A) \times r_B^{-1}(B)$.

Claim 7.7. *Let $g : \Lambda \times \Lambda \to \{0, 1\}$ and $g' : \Lambda' \times \Lambda' \to \{0, 1\}$ be functions such that g' is reducible to g. The discrepancy of g' with respect to any product distribution μ is maximized by a canonical rectangle.*

Proof. We show that for every rectangle $R' = M \times N$ of g', there exists a canonical rectangle of g' with at least the same discrepancy and this will imply the claim. The discrepancy of a rectangle R' can be written as follow

$$\mathrm{disc}_{\mu,R'}(g') = \left| \sum_{(x',y') \in R} (-1)^{g'(x',y')} \Pr_{(V,W) \leftarrow \mu} \left[(V, W) = (x', y') \right] \right|.$$

Assume without loss of generality that the above sum inside the absolute value is positive. It hold that each row $x' \in M$ either has a positive contribution to the sum or not. If the contribution is positive then we add all the rows v' such that $r_A(v') = r_A(x')$. For each such row v' it hold that $g'(x', y') = g'(v', y')$ for every $y' \in N$ and thus the contributions to the sum from x' and v' are equal. As a result we get that adding v' increases the discrepancy. Similarly, if the contribution of a row x' is not positive we remove this row from M, and the new rectangle has at least the same discrepancy as the one with the row x'. The same can be done for the columns. Therefore we get a canonical rectangle that has at least the same discrepancy as R'. ■

We also use the following folklore fact

Fact 7.8 (Folklore)**.** *For every distribution ν over finite set Λ and $\varepsilon > 0$ there exists a distribution ν' over Λ such that all the probabilities that μ' assigns are rational and $|\nu' - \nu| \leq \varepsilon$.*

Finally, we prove Lemma 7.5. Let $\mu = \mu_X \times \mu_Y$ be a product distribution. The high-level idea of the proof is that we choose Λ' and $r_A, r_B : \Lambda' \to \Lambda$ such that for every $x \in \Lambda$ it holds that $\mu_X(x) \approx \frac{|r_A^{-1}(x)|}{|\Lambda'|}$ (the same is done for r_B and μ_Y). As the discrepancy is maximized by canonical rectangles, we get that for each input x the set $r_A^{-1}(x)$ is either contained in the rows of the rectangle or disjoint from them. Therefore when calculating the discrepancy we get that the contribution of x is multiplied by about $\Pr[\mu_X = x]$ as in the definition of discrepancy with respect to μ.

Proof of Lemma 7.5 Let g, $\mu = \mu_X \times \mu_Y$, and ε be as in the lemma. Let $\varepsilon' = \frac{\varepsilon}{4}$. Let μ'_X (respectively, μ'_Y) be some distribution that is ε'-close to μ_X (respectively, μ_Y) and is rational, whose existence guaranteed by Fact 7.8. As μ'_X and μ'_Y are rational then exists an integer $l > 0$ such that $l \cdot \mu'_X(x)$ and $l \cdot \mu'_Y(y)$ are integers for all $x, y \in \Lambda$. We choose $\Lambda' = [l]$, and choose the function r_A such that for every $x \in \Lambda$, the function r_A maps $l \cdot \mu'_X(x)$ values to x. The function r_B is chosen in the same way with respect to μ'_Y.

We prove that $\mathrm{disc}_U(g') \leq \mathrm{disc}(g) + \varepsilon$. Let $R' = M \times N$ be rectangle that maximizes the discrepancy of g', and recall that we can assume R' is a canonical rectangle by Claim 7.7. Let

$R = r_A(M) \times r_B(N)$ be the rectangle of g that corresponds to R'. Then it holds that

$$\begin{aligned}
\mathrm{disc}_U\left(g'\right) &= \mathrm{disc}_{U,R'}\left(g'\right) \\
&= \left| \sum_{(x',y')\in R'} (-1)^{g'(x',y')} \Pr\left[U = \left(x',y'\right)\right] \right| \\
&= \left| \sum_{(x,y)\in R} \sum_{(x',y')\in r_A^{-1}(x)\times r_B^{-1}(y)} (-1)^{g'(x',y')} \frac{1}{l^2} \right| \\
&= \left| \sum_{(x,y)\in R} (-1)^{g(x,y)} \sum_{(x',y')\in r_A^{-1}(x)\times r_B^{-1}(y)} \frac{1}{l^2} \right| \\
&= \left| \sum_{(x,y)\in R} (-1)^{g(x,y)} \frac{\left|r_A^{-1}(x)\times r_B^{-1}(y)\right|}{l^2} \right| \\
&= \left| \sum_{(x,y)\in R} (-1)^{g(x,y)} \cdot \mu'_X(x) \cdot \mu'_Y(y) \right|. \\
&= \left| \sum_{(x,y)\in R} (-1)^{g(x,y)} \Pr_{(V,W)\leftarrow\mu'} \left[(V,W) = (x,y)\right] \right| \\
&= \mathrm{disc}_{\mu',R}(g).
\end{aligned}$$

Next, we show that as μ and μ' are ε'-close then $\mathrm{disc}_{\mu',R}(g)$ is only bigger then $\mathrm{disc}_{\mu,R'}(g)$ by ε as follows

$$\begin{aligned}
\mathrm{disc}_U\left(g'\right) =& \mathrm{disc}_{\mu',R}(g) \\
=& \left| \Pr_{(V,W)\leftarrow\mu'} \left[g(V,W) = 0 \text{ and } (V,W)\in R\right] \right. \\
& \left. - \Pr_{(V,W)\leftarrow\mu'} \left[g(V,W) = 1 \text{ and } (V,W)\in R\right] \right| \\
\leq& 4\varepsilon' + \left| \Pr_{(V,W)\leftarrow\mu} \left[g(V,W) = 0 \text{ and } (V,W)\in R\right] \right. && (\mu' \text{ is } 2\varepsilon'\text{-close to } \mu) \\
& \left. - \Pr_{(V,W)\leftarrow\mu} \left[g(V,W) = 1 \text{ and } (V,W)\in R\right] \right| \\
=& \varepsilon + \mathrm{disc}_{\mu,R}(g) \\
\leq& \varepsilon + \mathrm{disc}_{\mu}(g)
\end{aligned}$$

■

References

[ABK21] Anurag Anshu, Shalev Ben-David, and Srijita Kundu. On query-to-communication lifting for adversary bounds. In Valentine Kabanets, editor, *36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference)*, volume 200 of *LIPIcs*, pages 30:1–30:39. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.

[BPSW05] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A direct sum theorem for corruption and the multiparty NOF communication complexity of set disjointness. In *20th Annual IEEE Conference on Computational Complexity (CCC 2005), 11-15 June 2005, San Jose, CA, USA*, pages 52–66. IEEE Computer Society, 2005.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011.

[Bra12] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012.

[BW12] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *APPROX-RANDOM*, pages 459–470, 2012.

[CFK+19] Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. Query-to-communication lifting using low-discrepancy gadgets. *Electronic Colloquium on Computational Complexity (ECCC)*, 26:103, 2019.

[CKLM17] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudorandom properties. *CoRR*, abs/1704.06807, 2017.

[CKLM18] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation beats richness: new data-structure lower bounds. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1013–1020, 2018.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS*, pages 270–278, 2001.

[dRMN+20a] Susanna F. de Rezende, Or Meir, Jakob Nordström, Toniann Pitassi, and Robert Robere. KRW composition theorems via lifting. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 43–49. IEEE, 2020.

[dRMN+20b] Susanna F. de Rezende, Or Meir, Jakob Nordström, Toniann Pitassi, Robert Robere, and Marc Vinyals. Lifting with simple gadgets and applications to circuit and proof complexity. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS '20)*, November 2020. Also available as ECCC TR19-186.

[FKNN95] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995.

[GGKS18] Ankit Garg, Mika Göös, Pritish Kamath, and Dmitry Sokolov. Monotone circuit lower bounds from resolution. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 902–911, 2018.

[GLM$^+$15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 257–266. ACM, 2015.

[GP14] Mika Göös and Toniann Pitassi. Communication lower bounds via critical block sensitivity. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 847–856. ACM, 2014.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS '17)*, pages 1077–1088, 2015.

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

[HHL16] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 282–288. IEEE Computer Society, 2016.

[Jai11] Rahul Jain. New strong direct product results in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:24, 2011.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *ICALP*, pages 300–315, 2003.

[JRS05] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *20th Annual IEEE Conference on Computational Complexity (CCC 2005), 11-15 June 2005, San Jose, CA, USA*, pages 285–296. IEEE Computer Society, 2005.

[KKN92] Mauricio Karchmer, Eyal Kushilevitz, and Noam Nisan. Fractional covers and communication complexity. In *Proceedings of the Seventh Annual Structure in Complexity Theory Conference, Boston, Massachusetts, USA, June 22-25, 1992*, pages 262–274. IEEE Computer Society, 1992.

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

[KRW91] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via direct sum in communication coplexity. In *Proceedings of the Sixth Annual Structure in Complexity Theory Conference, Chicago, Illinois, USA, June 30 - July 3, 1991*, pages 299–304. IEEE Computer Society, 1991.

[LS09] Nikos Leonardos and Michael E. Saks. Lower bounds on the randomized communication complexity of read-once functions. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 341–350. IEEE Computer Society, 2009.

[PR17] Toniann Pitassi and Robert Robere. Strongly exponential lower bounds for monotone computation. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC '17)*, pages 1246–1255, 2017.

[PR18] Toniann Pitassi and Robert Robere. Lifting Nullstellensatz to monotone span programs over any field. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC '18)*, pages 1207–1219. ACM, 2018.

[RM97] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. In *38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997*, pages 234–243, 1997.

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999.

[RPRC16] Robert Robere, Toniann Pitassi, Benjamin Rossman, and Stephen A. Cook. Exponential lower bounds for monotone span programs. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS '16)*, pages 406–415, 2016.

[Sha01] Ronen Shaltiel. Towards proving strong direct product theorems. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity, Chicago, Illinois, USA, June 18-21, 2001*, pages 107–117. IEEE Computer Society, 2001.

[She09] Alexander A. Sherstov. The pattern matrix method (journal version). *CoRR*, abs/0906.4291, 2009.

[She11] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 41–50. ACM, 2011.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[WYY17] Xiaodi Wu, Penghui Yao, and Henry S. Yuen. Raz-McKenzie simulation with the inner product gadget. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:10, 2017.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il